Université Fédérale



Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut Supérieur de l'Aéronautique et de l'Espace

Présentée et soutenue par :
Xiaoqi XU

le vendredi 27 janvier 2023

Titre :

In search of invariants in varying electroencephalography signals for brain-computer interfaces

---

École doctorale et discipline ou spécialité :

EDSYS : Informatique

Unité de recherche :

Équipe d'accueil ISAE-ONERA DECISIO

Directeur(s) de Thèse :

Mme Raphaëlle ROY (directrice de thèse)
M. Nicolas DROUGARD (co-directeur de thèse)

Jury :

M. Fabien LOTTE Directeur de recherche INRIA Bordeaux - Président
M. Nicolas DROUGARD Enseignant-chercheur ISAE-SUPAERO - Co-directeur de thèse
M. Alexandre GRAMFORT Directeur de recherche INRIA Saclay - Examinateur
M. Moritz GROSSE-WENTRUP Professeur University of Vienna - Rapporteur
Mme Kathryn HESS BELLWALD Professeur EPFL - Examinatrice
M. Théodore PAPADOPOULO Directeur de recherche INRIA Sophia Antipolis - Rapporteur
Mme Raphaëlle ROY Enseignante-chercheuse ISAE-SUPAERO - Directrice de thèse
Mme Martina SCOLAMIERO Professeur adjoint KTH, Stockholm - Examinatrice

Thesis advisor: Professor Raphaëlle N. Roy & Nicolas Drougard Xiaoqi XU

In search of invariants in varying electroencephalography signals for brain-computer interfaces

Abstract

A Brain-computer interface (BCI) creates a direct pathway between brain activity and actions of a machine. Brain signals are collected and processed to estimate a given mental state of the user which is used either as an explicit command for communication or control purposes, or implicitly for system adaptation. Despite many impressive achievements, several big challenges remain, among which the variability of the electroencephalography (EEG) signal between sessions, tasks and users. This doctoral thesis proposes three ways to tackle this problem. All proposed methods were tested on publicly available datasets in both intra- and inter-subject classification settings and compared with state-of-the-art methods such as Riemannian classifiers.

The first method, from a spatial point of view, consists in decomposing the EEG signals spatially over the eigenvectors of the Laplacian and to reduce the dimension of the data by dropping high frequency components. The Laplacian based on EEG montage (geometric information) and functional connectivity graph were both tested and compared. Experiments on a large open dataset demonstrated that classification accuracy was improved even using small dimensions.

The second one, from a temporal point of view, was based on a nonparametric tool for time series analysis: path signature. It provides features independent of time parametrization and the lead-lag relationship between channels captured by the second level signature turned out to be more robust to inter-subject variability than covariance matrices when combined with Riemannian classifiers.

The third one studied the spatio-temporal dynamics of the EEG signals via topological data analysis (TDA). A complete review of TDA applied to EEG processing was performed to identify the general pipelines and several promising features. Experiments were done to fill the gap on its application for BCI in the literature. The results showed a significant improvement on inter-subject classification (10% compared to Riemannian classifiers) and opens exciting research problems.

In conclusion, these less explored mathematical tools have interesting potential to tackle the difficult problem of EEG variability and suggest a network and dynamic viewpoint on EEG signals.

Contents

1 Prelude 3
- 1.1 Brain computer interface . . . 4
- 1.2 Electroencephalography . . . 7
- 1.3 EEG variability . . . 10
  - 1.3.1 What . . . 11
  - 1.3.2 Why . . . 13
  - 1.3.3 How . . . 14
- 1.4 Baseline methods . . . 15
  - 1.4.1 Common Spatial Pattern . . . 15
  - 1.4.2 Riemannian classifier . . . 16
- 1.5 Datasets . . . 17
  - 1.5.1 BCI competition IV 2a dataset . . . 18
  - 1.5.2 Physionet MI dataset . . . 19
  - 1.5.3 Passive BCI competition . . . 19
  - 1.5.4 Good practices . . . 21
- 1.6 Objective . . . 22

2 Toccata: Laplacian 24
- 2.1 Laplacian . . . 25
  - 2.1.1 Theory . . . 25
  - 2.1.2 Application . . . 28
- 2.2 Dimensionality reduction . . . 30
  - 2.2.1 Geometry-based method . . . 32
  - 2.2.2 Connectivity-based method . . . 35
  - 2.2.3 Dimensionality reduction . . . 36
- 2.3 Further analyses . . . 39
- 2.4 Conclusion . . . 43

3 Canon: Path signature 45

3.1 Path signature . . . 46
3.2 Application . . . 50
3.2.1 On raw data . . . 51
3.2.2 On spectral power . . . 56
3.2.3 On covariance matrices . . . 56
3.3 Lead matrix . . . 58
3.3.1 Cyclicity analysis . . . 59
3.3.2 Riemann-signature classifier . . . 63
3.4 Conclusion . . . 66

4 Rondo: Topological data analysis **69**
4.1 Motivation . . . 70
4.2 Topological data analysis . . . 72
4.2.1 Topology and homology in a nutshell . . . 72
4.2.2 Definitions . . . 73
4.2.3 Persistent homology . . . 75
4.3 Topological features . . . 76
4.4 Application . . . 77
4.5 Experiments on EEG-based BCI . . . 82
4.5.1 Toy example . . . 83
4.5.2 Methods . . . 85
4.5.3 Results and discussions . . . 90
4.6 Conclusion . . . 101

5 Postlude **104**

Appendix A **112**
A.1 Introduction . . . 113
A.2 Methods . . . 113
A.2.1 Topological data analysis . . . 113
A.2.2 Convolutional neural networks . . . 114
A.3 Results . . . 116
A.4 Discussion . . . 116

References **134**

Listing of figures

1.1.1 Components of a usual BCI pipeline. 6
1.1.2 Example of the screen seen by subjects for (a) P300 speller[92] and (b) MI BCI[101]. 7
1.2.1 Illustration of (a) an EEG recording system (from `https://www.biosemi.com/new.htm`) and (b) an example of EEG signal during resting state. 8
1.2.2 Electrodes locations in 10-20 and 10-10 systems (from Wikipedia). 9
1.3.1 2D t-SNE embedded features learned by common dictionary learning over 20 subjects, four rounds, eight sessions. Each letter represents an identifier of a single subject and the center coordinates for each character correspond to the data point (here, session-wise transformation matrix) from a single session. From Nishimoto et al.[121] 12

2.2.1 EEG data can be expressed as the weighted sum of eigenvectors of the Laplacian. Data dimension can be reduced by dropping high frequency components. 31
2.2.2 Applying Laplacian spectral decomposition to EEG-based BCI. The Laplacian matrix can be computed either from the triangle mesh constructed with the 3D coordinates of the electrodes (green arrow) or from the functional connectivity graph estimated from the EEG data (orange arrow). Note that the first way is data-independent while the second is not. Then the EEG data is projected on the space spanned by the eigenvectors of the Laplacian or by the principal components (PCA used for comparison). Finally, two gold standard classifiers are applied to give the classification accuracy. The baseline (purple arrow) is the normal classification pipeline without dimensionality reduction. 32
2.2.3 Triangle mesh data used to compute the discrete Laplacian. In the picture on the left, the angles α_{ij} and β_{ij} used in the cotangent matrix are displayed here for indices i and j corresponding to the vertices p_i and p_j in the mesh. The value of s_i is the area of the shaded region defined as the union of two triangles for each face containing p_i. These triangles are depicted in the right picture: their vertices are p_i, the middle of an adjacent edge, and the circumcenter represented by a cross. 33

2.2.4 (a) Topomaps of the first 9 eigenvectors and eigenvalues of the Laplace-Beltrami operator. (b) Underlying mesh constructed using the Physionet MI dataset montage[150,151]. The eigenvectors corresponding to increasing eigenvalues have increasing spatial variation. 34
2.2.5 Topomaps of the first 9 eigenvalues and eigenvectors of the connectivity graph Laplacian (computed with the data from one subject of the Physionet MI dataset). 37
2.2.6 Intra-subject classification accuracy as a function of the number of principal components or Laplacian spectral components used for dimensionality reduction, averaged over 105 subjects. Baseline: CSP+SVM and RC without dimensionality reduction. 38
2.2.7 Inter-subject classification accuracy as a function of the number of principal components or Laplacian spectral components used for dimensionality reduction, averaged over 105 subjects. Baseline: CSP+SVM and RC without dimensionality reduction. 39
2.3.1 EEG signal reconstructed with 10 (a) and 25 (b) Laplacian eigenvectors respectively. Original signal is extracted from one subject in the database. 40
2.3.2 The first 4 Laplacian eigenvectors of the structural connectivity graph. From Preti & Van De Ville[135]. 41
2.3.3 Average spectral power using the mesh Laplacian (a) and the graph Laplacian (b) respectively. The shaded area denotes standard deviation (Physionet MI dataset). 42
2.3.4 Computation time depending on the number of kept Laplacian spectral components. 42

3.1.1 Geometric interpretation of level 2 signature. The shaded area in the subfigures represents (a) $S(X)^{1,2}_{a,b}$, (b) $S(X)^{2,1}_{a,b}$, and (c) the signed area A_{signed} (the difference between areas "+" and "−") respectively. 48
3.2.1 Different pipelines used in this chapter. Path signature could be used directly on multi-variate EEG time series, or alternatively on the time series of spectral power or covariance matrices extracted using a sliding window. Then either the truncated path signature can be used directly as a feature vector, or the lead matrix can be computed from the level 2 signature and then turned into an SPD matrix to be classified by a Riemannian classifier. 50
3.2.2 Pipeline of applying path signature on raw data. First a windowed FIR filter is applied to filter out the task related signal. Then the signature of multivariate EEG time series is computed and this is the features that are fed to the classifiers to get the classification results. 51

3.2.3 Electrode montage of BCI competition IV 2a dataset. In blue are the electrodes that appear in the table at the right. 53
3.2.4 Topographic maps of coefficients over level 1 signatures. 53
3.2.5 Intra-subject classification accuracy using up to level 2 signature as features and with dimensionality reduction (PCA or ANOVA). The baseline without dimensionality reduction is marked by the green dashed line. 54
3.2.6 Intra-subject classification accuracy of various variants of Riemannian classifier. The first one is the MDM classifier with covariance matrices as features and Riemannian metric to estimate the distance. The second one is the MDM classifier with covariance matrices as features and Euclidean metric to estimate the distance. The third one is the LR classfier with flattened covariance matrices as features. 55
3.3.1 The signals are $\sin(t-\alpha_i)$ shown at the left. In the middle is the lead matrix in which yellow corresponds to positive values and blue to negative values. If L_{ij} is positive it means signal i is leading signal j. At the right is the first eigenvector of the lead matrix. We see a perfect recovery of the cyclic order from the phase of the elements of the first pair of eigenvectors. 60
3.3.2 First row: on the left is the plot of 2 sets of sin waves; in the middle is the corresponding lead matrix; on the right is the absolute value of the eigenvalues of the lead matrix. Second row: the 2 eigenvectors corresponding to the 2 pairs of non zero eigenvalues. 61
3.3.3 Absolute value of the eigenvalues of the lead matrix averaged across trials. . 62
3.3.4 Frequency that a channel appears in the top 10 largest elements of the first (a) and second (b) leading eigenvectors of lead matrix in left/right MI conditions separately (over all subjects). 62
3.3.5 Pipeline of the Riemann-signature classifier. 64
3.3.6 Inter-subject classification accuracy for each subject for the Physionet dataset, with leave one out cross validation. A Riemannian classifier is applied on signature-based SPD matrix as described in Sec. 3.3.2 and on covariance matrices respectively. 65

4.2.1 A coffee mug is equivalent to a donut in a topologist's eyes. Image made by Henry Segerman (`http://www.segerman.org/`). 73

4.3.1 Examples of filtration and the associated barcode. (A) shows the sublevel filtration of function f(x). The first and second connected components appear at point a and b respectively and merge together at point c, giving rise to two bars: one from a till infinity and the other from b to c. (B) shows several stages of Rips filtration. At first, all points are isolated and we have the same number of bars and points. As balls grow larger, some balls start to merge together, giving death to certain bars. Then one loop appears, marked as a 1-dimensional feature corresponding to the blue line in the barcode. Finally the loop is filled so the blue bar stops and every point merges into one component living forever. 78
4.4.1 Pipeline of applying TDA to EEG signal processing. 80
4.5.1 Time delay embedding (a), the attractor (b) and the persistence diagram (c) of a periodic signal. The color in the attractor figure indicates time. 83
4.5.2 A quasiperiodic signal (a), its attractor after a 3 dim. embedding (b) and the corresponding persistence diagram (c). The color in the attractor figure indicates time. 84
4.5.3 A white noise drawn from a normal distribution (a), its attractor (b) and the corresponding persistence diagram (c). The color in the attractor figure indicates time. 85
4.5.4 Effects of embedding dimension d on the shape of the attractor. 86
4.5.5 Persistence diagram with different embedding dimension. 87
4.5.6 Effects of time delay τ on the shape of the attractor. 87
4.5.7 Persistence diagram with different time delay τ. 88
4.5.8 Mutual information between $\sin(t)$ and $\sin(t-\tau)$ with different time delay τ. 89
4.5.9 Persistence landscapes of a certain persistence diagram. (from Bubenik et al. [21]). The top row shows the rotation of the persistence diagram. The bottom row shows 2D and 3D view of λ_i, i.e. the i-th layer of the persistence landscape. . 90
4.5.10 Different dynamics in different frequency bands. The color in the attractor figure indicates time. 91
4.5.11 Average optimal parameters of time delay embedding computed using 30 randomly chosen trials from each subject in BCI competition IV 2a dataset. . . 91
4.5.12 Average classification accuracy using different parameters. 95
4.5.13 Benchmark of inter-subject classification on the BCI Competition IV 2a dataset. Here the TDA method is represented by the combination of persistent entropy and random forest. CSP is combined with SVM. 98
4.5.14 Leave-one-out cross validation accuracy on each subject. The classifiers used with topological features and the combined features are the random forest classifier for (a) to (c) and SVM for (d). 100

4.5.15 Intra-subject classification accuracy of each subject of the BCI competition IV 2a dataset with the CSP+SVM pipeline. 100

5.0.1 Different phases in the dégagé à la seconde in ballet (From Larousse encyclopedia). 109

A.2.1 Network Structure. 115
A.4.1 Results. 117

To those who fly.

"*There is nothing permanent except change.*"

Heraclitus

Acknowledgments

I feel very lucky to be supervised by my two supervisors, Raphaëlle and Nicolas, who gave me a lot of freedom to explore, supported and guided me throughout my doctoral studies. They taught me how to do research and they set an excellent example for me. I am extremely grateful for their kindness, encouragement and help. This thesis can not be realized without them.

I would also like to thank all the members of the jury: Alexandre Gramfort, Moritz Grosse-Wentrup, Kathryn Hess, Fabien Lotte, Théodore Papadopoulo, Martina Scolamiero, especially Moritz and Théodore for reviewing this thesis, and Kathryn for hosting me at the laboratory for topology and neuroscience at EPFL. Part of the work in this thesis is the fruit of this enriching visit and discussions with Darrick Lee at the lab. I also thank my doctoral school, fondation ISAE SUPAERO and ANITI for their financial support for this visit.

I would like to thank all the mathematicians that I met during my excursion in the mathematical world, in particular Alano Ancona, Jean-Michel Bismut, Agnès Desolneux, Dominique Hulin, Hans Henrik Rugh, and Stéphane Sabourau. Thanks to Gabriel for having finished the thesis that I could have done in another life. I am also thankful to my colleagues and students at ISAE SUPAERO and ANITI.

My thanks also go to Covid and the confinement during which I discovered my passion for ballet. I appreciate the dance classes of my amazing dance professor Emilie and the videos of so many great dancers in the world. I have learned a lot from ballet about balance, confidence and perseverance. I am grateful for my friends, old and new, far and near, for sharing laughter but also tears. I also thank those who abandoned me for letting me know that I am stronger than I thought. Last but not the least, I would like to express my deepest gratitude to my parents and my family for their tremendous love and support.

Nomenclature

BCI	Brain computer interface
CNS	Central nervous system
CSP	Common spatial pattern
EEG	Electroencephalography
fMRI	Functional magnetic resonance imaging
ICA	Independent component analysis
LDA	Linear discriminant analysis
LR	Logistic regression
MDM	Minimum distance to mean
MEG	Magnetoencephalography
MI	Motor imagery

MLP Multilayer perceptron

MRI Magnetic resonance imaging

PCA Principle component analysis

RC Riemannian classifier

RF Random forest

SPD Symmetric positive definite

SVM Support vector machine

TDA Topological data analysis

Adagio

1

Prelude

Brain-computer interfaces (BCIs) are systems that allow interacting with machines using only brain activity and no muscular activity. BCI research is a highly multidisciplinary domain that involves knowledge and insights from neuroscience, psychology, ergonomics, engineering and mathematics. In this first chapter, the basic concepts about BCI and elec-

troencephalography (EEG) are introduced as well as the main problem that this thesis tries to tackle – the EEG variability problem. The datasets used in the entire thesis and some practical guidelines are detailed in order to facilitate the reproduction of the results. The next three chapters propose three tools, each dealing with one aspect of the central problem relying on specific invariances. And finally conclusions and perspectives are outlined in the last chapter. We hope that through this work the researchers working on BCI could become interested in using mathematical tools seemingly faraway, and that mathematicians could become aware of the interesting underlying theoretical problems and great challenges in the BCI domain, thus gathering the expertise from both sides to make impossible possible.

1.1 Brain computer interface

"With our thoughts we make the world". This verse in the Avataṃsaka Sūtra was talking about the abstract interactions of our thoughts and the environment around us. Perhaps the author did not think that one day brain-computer interfaces would make this more direct. Informally, we could think of a BCI as an intermediary that extends our nerves and connects our mind to a broader range of objects. Formally, BCI is defined as "a system that measures central nervous system (CNS) activity and converts it into artificial output that replaces, restores, enhances, supplements, or improves natural CNS output and thereby changes the ongoing interactions between the CNS and its external or internal environment. "[180]. The major motivation of early BCIs was restoring the ability of the severely paralyzed people to communicate and interact with the environment[15].

A BCI is composed of several parts as shown in Fig. 1.1.1. The signal acquisition sys-

tem records amplifies and digitizes brain signals, in most cases EEG signals which we will discuss more in depth in Section 1.2. There exist BCIs based on other brain signals such as oxygenation measures acquired through functional Near Infrared Spectroscopy (fNIRS), or electrocorticography (ECoG), an invasive method to get electrical brain signals directly from the cortex. Due to the high signal-to-noise ratio of ECoG compared to EEG, this type of BCI could have spectacular performance with great precision and speed in movement or speech decoding[178,117], and can even close the loop of communication by providing tactile feedback to the CNS[49]. However in this thesis, we will mainly focus on EEG-based BCI applications because they are non-invasive, more accessible and have high temporal resolution.

Since the EEG signal is very noisy, preprocessing steps are essential; they consist in removing noise and artifacts in order to improve the signal-to-noise ratio[89]. Then according to different applications , relevant features are extracted. A supervised classifier takes as input the features and associate the sample with a label according to the predefined rules learned on labeled data. The labels are then sent to the computer as information for explicit or implicit control . This thesis focuses on the feature extraction and classification parts colored in rose in Fig. 1.1.1.

There are many ways to categorize BCIs. According to the acquisition device used to record brain signals, there are EEG-based BCI, fNIRS-based BCI, etc. According to the way that the user interacts with the system – explicit or implicit – there are active BCI and reactive BCIs. Moreover, BCIs can be grouped depending on the paradigm used for control.

A BCI paradigm is the set of protocols that concretely implement a BCI, defining the task the user should perform to use the system (i.e. to control a device or communicate).

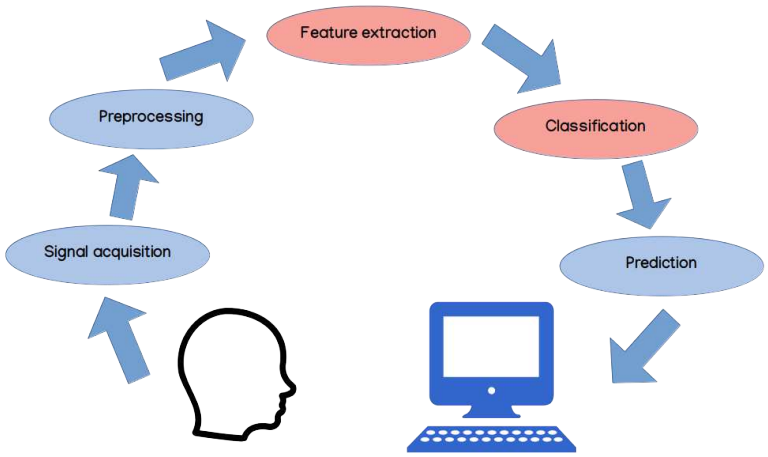
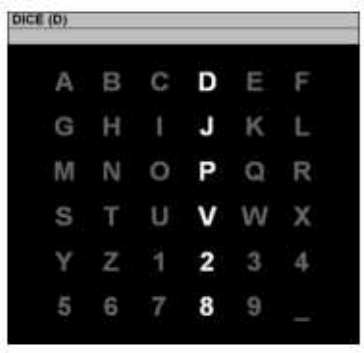


Figure 1.1.1: Components of a usual BCI pipeline.

The first and one of the most famous BCI paradigms is the P300 speller[47]. It is based on the P300 wave, an event-related potential (ERP) component – a positive deflection after 300 ms post-stimulation – which is strongly linked to attentional engagement and that is triggered by the expected stimuli[132]. The subject sits in front of a screen with a 6x6 matrix of letters as shown in Fig. 1.1.2 (a). The rows and columns flash randomly. The user focuses their attention on the letter that he/she wants to spell. When the row/column containing this letter is flashed, a P300 wave is produced by the flash in the EEG signals. Temporal features can be extracted from this ERP component, e.g. its peak amplitude, such that the letter could be chosen with high precision after several runs of flash.

Another major BCI paradigm is based on sensorimotor rhythms (SMR)[129]. The sensorimotor cortex is activated during motor task performance as well as motor imagery (MI, imagination of movement without actually moving), resulting in a decrease in power in the mu band (7-12 Hz) and the beta band (12-25 Hz), also known as event-related desynchronization (ERD), followed by an increase in the beta band, also known as beta rebound or

a 

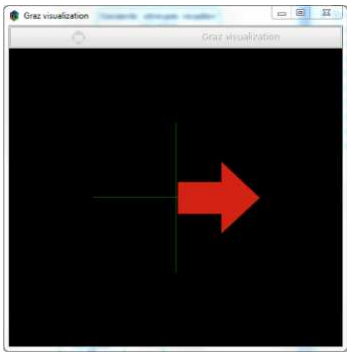
b 

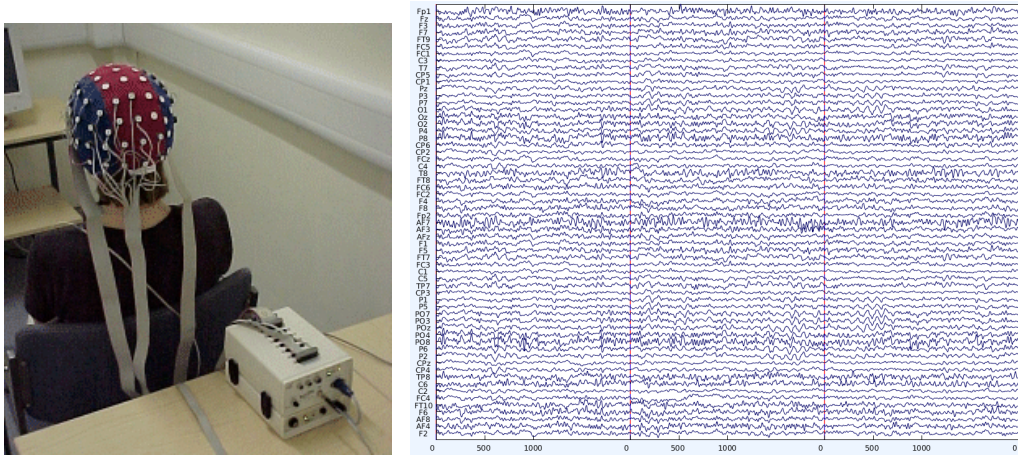
Figure 1.1.2: Example of the screen seen by subjects for (a) P300 speller[92] and (b) MI BCI[101].

event-related synchronization(ERS), after the imagination of movement[130]. SMR rhythms are generated by the sensorimotor cortex, located over the C3 and C4 electrodes (see Fig. 1.2.2), making these two electrodes very relevant for MI BCI. BCIs that are based on an automatic detection of this motor or MI activity can for instance enable users to move a cursor on the screen or control a wheelchair.

These two BCI paradigms need a certain degree of active participation of the subject even though the first one is less demanding and requires no user training. Only recently another type of BCI called passive BCI has been developed which needs no voluntary efforts from the user, mostly designed for non clinical applications. Passive BCI could be used to get an estimation of users' mental states (cognitive and affective) during their interaction with complex systems such as cars, planes and control rooms[188,143,144,4].

1.2 Electroencephalography

In many cultures, strong emotions like love are often described as electric shocks. This analogy is reasonable because they do cause fireworks in the brain. Electrical activity is the basis

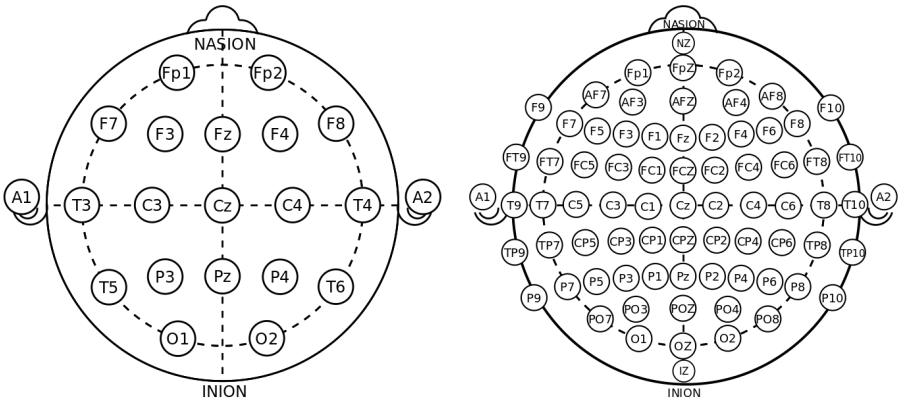


(a) recording system (b) EEG signal

Figure 1.2.1: Illustration of (a) an EEG recording system (from https://www.biosemi.com/new.htm) and (b) an example of EEG signal during resting state.

of communication between neurons, the information processing cells that constitute the CNS of which the brain and the spinal cord are the main constituents. Hence, the CNS constantly produces electrical signals to process information gathered from the sensory systems (e.g. visual system) and to coordinate the function of the whole body including motor activity to interact with one's environment[7]. The electrical signal generated by the brain can be recorded on the scalp using electrodes to capture differences in electrical potential in the order of microvolts. This technique is called *electroencephalography*, abbreviated as EEG.

EEG has always been a subject of fascination since its invention by Hans Berger in 1924. It is said that he destroyed all his personal EEG recordings, fearing that in the future people will recover his thoughts from it[160]. Yet, almost 100 years after the first EEG recordings, we are still far from that. EEG has mainly been used in clinical diagnosis in neural disorders like epilepsy, Alzheimer's disease, autism spectrum disorder (ASD), and sleep disorders.



(a) 10-20 system (b) 10-10 system

Figure 1.2.2: Electrodes locations in 10-20 and 10-10 systems (from Wikipedia).

For reproducibility purposes, the electrodes are placed on a cap with precise standards of positioning and labeling rules. The letter indicates the lobe over which the electrode is placed and the number indicates hemisphere (even numbers for right side, odd numbers for left side and 'z' for the midline). Fig. 1.2.2 shows the layout of the most used 10–20 system and 10-10 system[76]. It is based on the line from nasion to inion and the one from left ear to right ear. 10 or 20 refers to the proportion of the distance between adjacent electrodes over the total front-back or left-right distance.

EEG measures the superimposed currents generated by population of neurons. The signal to noise ratio is very low, which is mainly due to the current being distorted and attenuated by the soft and hard tissues between the current source and the electrode[23], a phenomenon also known as volume conduction, resulting in the poor spatial resolution of the EEG signal. However the temporal resolution of EEG signal is excellent, at the order of millisecond or less, making it suitable for real-time applications or analysis requiring high time precision. EEG recording system is easy to use and not expensive compared to other

brain imaging techniques like magnetoencephalography (MEG) or magnetic resonance imaging (MRI).

Another reason for the low signal to noise ratio of EEG, is that the EEG signal is very noisy. Indeed, it is often contaminated by various sources of artifacts. Artifacts can be categorized as internal/physiological, e.g. ocular or muscular artifacts, or external/non-physiological e.g. power line artifact. In order to deal with these various artifacts, a denoising step has to be implemented in BCI pipelines. This step is often referred to as a preprocessing step. The power line artifact can be easily removed by applying a spectral filter eliminating the corresponding frequency component. The physiological artifacts need more careful treatment. A common method is to identify and cancel artefactual components in the signal, for instance using blind source separation based on independent component analysis (ICA)[80,35].

Moreover, to enhance the signal to noise ratio, various techniques can be used, including spatial filtering methods, which can be used early in the pipeline to perform denoising, but also later on to allow for feature extraction [184]. Preprocessing steps and spatial filtering will be discussed in Section 1.5.4 and Chapter 2 respectively.

1.3 EEG variability

Compared to other brain recording methods, EEG has many advantages for BCI applications such as its high temporal resolution and cheap price, but one major obstacle to bring EEG-based BCI into everyday life for ordinary people is its lack of robustness due to the high variability of EEG signals which makes it is difficult to establish a robust classifier that works across time and users. In the following, first we clarify what we mean by EEG vari-

ability. Then various sources are identified to understand why EEG signal varies. Finally different natural and artificial techniques are mentioned on how to tackle the variability.

1.3.1 What

There are various ways to define and understand the EEG variability, for example as a covariate shift (i.e. a change of data distributions)[149] or as a morphological change of waves[73,74]. They are just different points of view to see the same underlying phenomenon, the change of characteristics of the signal. The discussions and methods in this thesis are general and are not constrained to a specific type of "variability".

In the BCI literature specifically two types of variability are mostly discussed: intra- and inter-subject variability[149,91]. The intra-subject variability includes in particular data changes for a given individual between different trials (short-term) and sessions (medium- and long-term), but also between tasks, user-state, and contexts. The inter-subject variability refers to data changes between different individuals.

Research works showed a possible higher impact of inter-subject variability compared to inter-session or intra-subject variability in a general manner. For example, Nishimoto et al. [121] showed a rate of successful personal identification much higher than chance level using EEG data recorded from multiple subjects and sessions (different days). It is further illustrated by Fig. 1.3.1. In this figure, each letter represents an identifier of a single subject and the center coordinates for each character correspond to the data point (here, session-wise transformation matrix) from a single session. It is clear that the distance between clusters of each subject is much larger than that between each session within one subject.

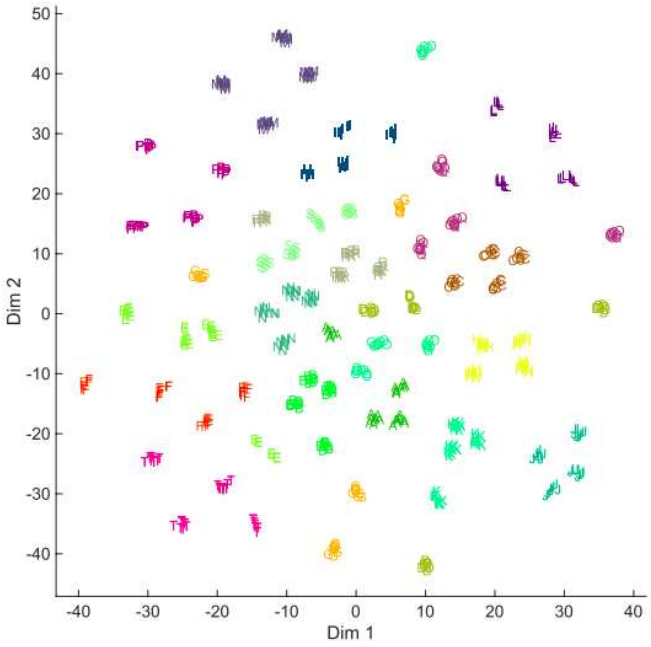


Figure 1.3.1: 2D t-SNE embedded features learned by common dictionary learning over 20 subjects, four rounds, eight sessions. Each letter represents an identifier of a single subject and the center coordinates for each character correspond to the data point (here, session-wise transformation matrix) from a single session. From Nishimoto et al. [121]

1.3.2 Why

The neural system is extremely complicated. Ruling out extrinsic noise, the variability comes from different initial states and intrinsic neuronal noise including sensory noise, cellular noise, electrical noise and synaptic noise[45]. Neurons respond differently to repeated stimuli[166,167,155], and neuronal variability is not equal in all neurons neither constant for a single neuron[45,107].

There is no consensus in the neuroscience community on the role of neural noise[107]. It is even not clear whether some randomness should be called "noise" or is actually information that we did not fully understand. What is certain is that we still arrive to act in a relatively reliable manner despite the seemingly noisy brain signal, demonstrating that there should be processes that allow the CNS to enhance information and diminish noise. Averaging is used when information is encoded by a population of neurons and is effective when the sources of noise are independent[45]. When signal is highly structured, prior knowledge is helpful to detect signal[11].

Furthermore, there are also many ways to group sources of variability depending on their origin, such as exterior (environmental) vs interior (subject-dependent) or microscopic (cellular level) vs macroscopic (network level). In accordance with the methods that will be presented in this thesis, we chose to group relevant sources of variability by spatial and temporal factors.

Spatially, the brain topographical variability is the main source of inter-subject variability[149]. Besides, variations can be found in electrical placement and underlying brain regions even with a fixed cap due to inter-subject differences in brain morphology[153].

Temporally, we can consider that there exists variability at three time-scales: short, medium

and long term. The short term variability corresponds to the variability across trials and is related to the task at hand, including with the stimuli and mental workload [54], and is modulated by attention [3]. As heart rate variability (HRV) could be seen as an indicator of mental workload [171], the neural variability has been shown to be informative about information processing capacity and functional integrity [3,161,51,43]. The medium term variability corresponds to the variability across sessions and can be a result of circadian rhythms [39,99], neural plasticity [48] and learning [3,90,109]. The long term variability can be related to age [93,109].

1.3.3 How

A lot of efforts have been made by the BCI community to overcome the challenge of EEG variability. Most of the work related to this issue borrows transfer learning techniques [103,82,181], the machine learning techniques that use the knowledge learned on one dataset to improve the performance on another usually related dataset. The transfer of learned knowledge is often done by domain adaptation, e.g. linear transformation [6], optimal transport [52], adaptive waveform learning [74] or Riemannian procrustes analysis [141] based on basic geometric operations, just to name a few. Some other kind of methods such as dictionary learning [115] have been proposed in the aim of finding some common features. However, compared to the domain adaptation methods, the search for robust features is much less explored. This is the main theme of this thesis, further developed in Section 1.6. In the next section, the baseline methods used as comparison are presented.

1.4 Baseline methods

While in theory all machine learning methods are suitable for EEG-based BCI, some methods stand out as gold standard methods since they fit the specificity of the EEG signal. As discussed in Section 1.2, the EEG signal has a very low signal-to-noise ratio, so successful methods must be able to reduce noise in order to extract useful information. Interestingly the gold standard human designed methods employ the same strategy as the brain. Recall that the neural noise can be diminished by averaging and using prior knowledge.

Common spatial pattern (CSP)[17] searches for a linear combination of channels that best distinguishes two classes by making the combined time series having a small variance for one class and a large variance for the other . Some independent noise is averaged out through linear combination.

A Riemannian classifier[9] combines the two strategies. It classifies the covariance matrices of all channels, taking into account the intrinsic geometry of the space where the covariance matrices live in. The averaging comes in the step of transforming raw signal into covariance matrix. It reduces especially noise in the temporal domain by condensing information in a time window. In the next step the prior knowledge of the structure of covariance matrices is useful to enhance information (this will be further discussed at the end of 3.2.1). The Riemannian distance is proved to be more accurate than Euclidean distance.

The detailed description of the two methods is as follows.

1.4.1 Common Spatial Pattern

Let $X \in \mathbb{R}^{n \times t}$ denote EEG signal of n channels and sampled with t time points. A linear combination of the original signal X can be written as $w^{\intercal}X$, where w is the weight vec-

tor also known as the spatial filter in EEG-based BCI domain. As said at the beginning of this section, CSP searches for the linear combination that maximizes the difference of variance (which is equal to band power) under two conditions. Let X_1, X_2 be the signal of the 2 classes respectively and assume that they are already centered, i.e. the mean equal to zero. Then CSP searches for w that either maximize or minimize the ratio

$$\frac{w^\intercal X_1 X_1^\intercal w}{w^\intercal X_2 X_2^\intercal w}.$$

In common practice, several filters that have the maximal or minimal ratios are kept to generate features. Note that $X_i X_i^\intercal$ is the covariance matrix Σ_i.

This problem can be reformulated as a generalized eigenvalue problem

$$\Sigma_1 w = \lambda \Sigma_2 w,$$

and solved by simultaneous diagonalization of Σ_1 and Σ_2.

1.4.2 Riemannian classifier

We have seen that the covariance matrices appear in the CSP algorithm. The motivation behind Riemannian classifier is to classify directly the covariance matrices. When estimated with enough samples, covariance matrices are symmetric positive definite (SPD) matrices, i.e. they belong to the set $\{A | A^\intercal = A, \lambda_i > 0, \forall i\}$, where λ_i are eigenvalues of A. This set is an open subset of $\mathbb{R}^{n \times n}$, so it is naturally a manifold. However, the geodesic of its embedding Euclidean space might go out of the SPD manifold. To remedy this, we have to equip the space with an intrinsic metric. There are many ways to do this. The most popular ones

in the medical imaging domain is the affine-invariant metric [1,2,3], and the Log-Euclidean metrics [5] provides a fast approximation when the data are concentrated with respect to the curvature.

Riemannian classifier got its name because it uses Riemannian geometry to compute the distance between covariance matrices, but the core of the classification algorithm could be any classical machine learning algorithm. For example with the minimum distance to mean (MDM) classifier, it computes the distance between a sample and the mean of each class and label it with the label of the nearest class. The particularity is just that the distance and mean are in the Riemannian sense.

It is convenient to project all covariance matrices to the tangent space of a reference point (e.g. identity matrix, geometric mean) via logarithmic map and then apply classical classifiers such as logistic regression or support vector machine in the Euclidean tangent space [10]. This is called the Tangent Space Riemannian Classifier and is the one used throughout this thesis.

1.5 Datasets

This section described the datasets that we are going to use for testing our proposed methods in this thesis. Data hold an important place in BCI research. Collecting EEG data is an expensive and time-consuming task despite the rapid advance in sensor technologies. This may be one of the causes of the scarcity of publicly available data in this domain. The systems used to collect data are quite diverse which adds complexity to the data processing. Moreover, there is no standard pipeline for data preprocessing and details regarding the preprocessing are not always shared or sufficiently described [72]. Some efforts made to rem-

edy this are presented in the following sections. Details of preprocessing for our tests are exhibited in order to facilitate the reproduction of our results.

In accordance with the principles of open science, we have exclusively worked with open datasets. We have further chosen the most used datasets among open EEG-based BCI datasets for active BCI applications: the BCI competition IV 2a dataset [163] and the Physionet motor imagery dataset [150]. We have also used the EEG dataset for mental workload estimation published in the first passive BCI competition [70] hosted by the Neuroergonomics conference 2021 [142]. The additional preprocessing steps applied to these datasets through the chapters are band-pass filters shown in the last row of Tab. 1.5.1.

1.5.1 BCI competition IV 2a dataset

The BCI competition IV 2a dataset is an indispensable classical MI BCI dataset which consists of data collected from 9 subjects on 2 sessions on different days [163]. Each session contains a total of 288 trials of 4 tasks: imagination of movement of left hand, right hand, feet and tongue. Each trial lasts for 6 s with 2 s of resting state at the beginning. After the cue appears on the screen at $t = 2$s, participants have to perform the motor imagery task.

EEG signals were recorded using 22 Ag/AgCl electrodes placed according to the international 10-20 system with the left mastoid as reference and right mastoid as ground electrode. Signals were down-sampled to 250 Hz and band-pass filtered into the 0.5-100 Hz band. A 50 Hz notch filter was applied. There were also 3 EOG channels, but they were not used in our experiments. Trials were checked by human experts and those containing artifacts were marked.

1.5.2 Physionet MI dataset

It is quite common that BCI datasets involve very few subjects. The great advantage of Physionet MI dataset is that it contains data recorded from 109 subjects. The original dataset contains EEG data of motor movement and imagery, but only motor imagery data were used in our experiments. There is only 1 session consisting of 14 runs, each including 2 trials of 1 minute resting state (eyes open/closed), and 3 runs of 2 minutes with the following 4 tasks: open and close left or right fist, imagine opening and closing left or right fist, open and close both fists or both feet, imagine opening and closing both fists or both feet. EEG signals were recorded using the BCI2000 system with 64 channels placed according to the international 10-20 system and were down-sampled to 160 Hz. The data of 4 subjects are rejected in inter-subject classification in this thesis because they do not have the same number of trials and time steps as other subjects.

1.5.3 Passive BCI competition

Open passive BCI datasets are very rare compared to active BCI ones[72]. Hinss et al. initiated an effort to fill the gap by creating a large passive BCI dataset with 30 subjects, 4 different tasks, and 3 sessions each separated by 7 days[71]. They organized the first passive BCI competition following which the data from 15 subjects for one mental workload task were made public[70,142].

Hence, this dataset presents the data acquired during a task designed to modulate mental workload. This task was the Multi-Atribute Task Battery-II (MATB-II) developed by NASA to evaluate operator performance and workload[118]. It is comprised of 4 sub-tasks: monitoring, dynamic resource management, tracking, and communication. Different

	BCI competition IV 2a	Physionet MI-BCI	Passive BCI competition
#subjects	9	105	15
#classes	2	2	3
#trials	72	22	149
#channels	22	64	61
band-pass filter	8-30 Hz	8-30 Hz	4-15 Hz

Table 1.5.1: Summary about the datasets (in the way they are used in this thesis). Only 2 classes were used for the BCI competition IV 2a and Physionet MI-BCI datasets, even through these datasets have 4 classes. For the Physionet dataset we have dropped 4 subjects who have different number of trials and time samples. The number of trials in the table refers to the number of trials per class. Band-pass filters are those that we applied as preprocessing steps and chosen according to tasks.

numbers of sub-tasks can be performed simultaneously under different conditions: easy, medium and difficult in order to elicit different levels of mental workload.

EEG signals were recorded using 64 active Ag-AgCl electrodes placed according to the international 10-20 system with the right mastoid as reference electrode. Leaving one channel not usable and one EOG channel results in 62 channels. Before each session, resting state data were recorded. For this work, the data from the first session only were used. The entire data during the resting state and the task were cut into 2 s non-overlapping epochs, filtered into the 1-40 Hz frequency band, and down-sampled to 250 Hz. Artifacts (muscle, heart, eye movement) were removed automatically using the SOBI algorithm [14]. We applied additionally a 4-15 Hz band-pass filter to make sure the classification results are not impacted by the muscle artifacts since the task involves a lot of movements especially in the difficult condition.

1.5.4 Good practices

Prevention is better than cure [*]. It is also true for EEG-based BCI. Paying attention to ensure the quality of data when recording EEG signals is much better than using sophisticated methods in the following processing steps to remedy. Farrens et al. [46] provided very complete guidelines to get the best data possible, including the details from equipment to procedures during the human experiments and ending with solutions to common problems.

Yet, even with good quality EEG signals, preprocessing steps are still necessary for numerous applications due to the inevitable external and physiological artifacts discussed in Section 1.2. Makoto Miyakoshi [112] gave his personal recipe of preprocessing steps according to needs of applications. His recommendations addressed the most common steps such as downsampling, filtering, removing/interpolating channels, epoching, running ICA, etc.

Last but not least, it is better to make the processing steps as general as possible. Throughout the experiments in this thesis, the standard classifiers from the scikit-learn toolkit [122] are used with default parameters (data/feature are rescaled if needed). The cross validation is applied both in intra- and inter-subject classification without shuffling the trials to approximate the realistic scenario of application. It is also important that researchers share their data and code to push the field forward together. Mother Of All BCI Benchmarks (MOABB) [82] is an open-source framework based on MNE [64] and the scikit-learn toolkit [122] aiming to standardize the pipelines from reading and preprocessing data to the evaluation of the results, providing also the tools for visualization and statistical analysis of results.

[*] Desiderius Erasmus

1.6 Objective

As mentioned earlier in Section 1.3, the main objective of this thesis is to find robust EEG features for BCI applications. We have chosen to focus on inter-subject variability since the inter-subject variability has a stronger impact on EEG data compared to inter-session variability [121]. To tackle this variability, three methods were proposed and tested.

Firstly, from a spatial point of view, we could use the Laplacian of the mesh constructed from EEG montage or the functional connectivity graph estimated from the EEG data to decompose the EEG signal as a weighted sum of the eigenvectors of the Laplacian. This could be interesting to reduce the sensibility of EEG signals to the precise placement of electrodes. Besides, dimensionality reduction could be done by dropping high frequency components. This also produces a smoothing and noise reduction effect.

Secondly, from a temporal point of view, path signature could be employed as a feature vector of the EEG signal. These features are translation invariant and independent from time parametrization. The first property deals with the covariate shift of the EEG signal and the second property could remove the variability related to temporal change (e.g. the different speed of executing a motor imagery task within/across subjects).

Lastly, since change is inevitable, instead of considering features designed for stationary data, we adopted a dynamical system point of view in the hope that the pattern of change is unchanged. The EEG data is embedded into phase space by time delay embedding and the topological features of its attractor are extracted via topological data analysis.

The three following chapters are dedicated to the three methods respectively with a detailed description of each method, as well as experiments on both intra- and inter-subject

classification compared with traditional methods. At the end is a general conclusion with comments on the results as well as the perspectives.

Andante, con dolcezza

2

Toccata: Laplacian

Harmonics have different yet related definitions in music and mathematics. The common point is that they represent a basic oscillation pattern. The frequencies and vibrations that an instrument can produce are prescribed by the Laplacian. Similarly, behind the one thousand and one brain waves, there is a set of harmonics that form a basis for the compo-

sitions, and these harmonics are related to neural processes, and therefore mental states and behaviours.

2.1 Laplacian

2.1.1 Theory

Definition The Laplacian is ubiquitous in the physical world for it describes one of the most common process: diffusion. It measures the difference between the value at the center and the average value of a small ball. So if the Laplacian is zero everywhere for a function, it means that the value of the function is equal to the average value over a small ball around it everywhere. These functions are called *harmonic* functions and correspond to equilibrium states.

In the Euclidean space $\mathbb{R}^n$, the *Laplacian* is defined as the divergence of the gradient:

$$\Delta f = \nabla \cdot \nabla f = \sum_i \frac{\partial^2 f}{\partial x_i^2}.$$

It can be generalized to a Riemannian manifold: $\Delta f = \operatorname{tr}(\operatorname{Hess} f)$, called *Laplace-Beltrami operator* defined by the trace of the Hessian with respect to the metric[83]. It can also be generalized to a graph[66]: $L = D - A$, where D is the degree matrix, A the adjacency matrix, and L the *graph Laplacian* of the graph. Note that the definition for a graph is the negative of its continuous counterpart.

The Laplacian is also known by the image processing and BCI community as a spatial

filter[156,25,108], which is in fact the first-order approximation of the Laplacian in $\mathbb{R}^2$:

$$\begin{aligned}
\Delta f(x_0, y_0) &= \frac{\partial^2 f}{\partial x^2}(x_0, y_0) + \frac{\partial^2 f}{\partial y^2}(x_0, y_0) \\
&\approx \frac{\partial f}{\partial x}(x_0 + 1, y_0) - \frac{\partial f}{\partial x}(x_0, y_0) + \frac{\partial f}{\partial y}(x_0, y_0 + 1) - \frac{\partial f}{\partial y}(x_0, y_0) \\
&\approx \big(f(x_0 + 1, y_0) - f(x_0, y_0)\big) - \big(f(x_0, y_0) - f(x_0 - 1, y_0)\big) \\
&\quad + \big(f(x_0, y_0 + 1) - f(x_0, y_0)\big) - \big(f(x_0, y_0) - f(x_0, y_0 - 1)\big) \\
&= f(x_0 + 1, y_0) + f(x_0 - 1, y_0) + f(x_0, y_0 + 1) + f(x_0, y_0 - 1) - 4f(x_0, y_0).
\end{aligned}$$

Hence we have the well-known convolutional kernel $\begin{pmatrix} & 1 & \\ 1 & -4 & 1 \\ & 1 & \end{pmatrix}$. Note that this is also equivalent up to sign with the graph Laplacian of the square grid graph with unit distance.

Laplacian eigenvalue problem The eigenvalues and eigenfunctions of the Laplacian are the solutions of the Helmholtz equation

$$-\Delta f = \lambda f. \tag{2.1.1}$$

The trivial solution is a constant function with $\lambda = 0$. The other eigenvalues are all positive. For a bounded domain $\Omega \in \mathbb{R}^n$, the eigenfunctions form an orthonormal basis for the Hilbert space $L^2(\Omega)$. This also holds for the Laplace-Beltrami operator on a compact Riemannian manifold with boundary. This provides a tool to decompose a generic function in the domain.

The eigenvalue λ indicates a notion of frequency or spatial variation. In fact, it is directly related to the Dirichlet energy. Suppose f is the normalized eigenfunction associated with the eigenvalue λ and vanishes on the boundary, then

$$\lambda = \int_\Omega f \cdot \lambda f \, dV = \int_\Omega f \cdot (-\Delta f) \, dV = \int_\Omega f \cdot (-\nabla \cdot \nabla f) \, dV = \int_\Omega < \nabla f, \nabla f > \, dV = 2E(f)$$

where $E(f)$ denotes the Dirichlet energy of f and the second last equality follows from the Green's first identity.

The graph counterpart for a graph $\mathcal{G} = (V, E)$ is

$$\lambda = f^\intercal \lambda f = f^\intercal L f = \sum_{i,j} f_i f_j L_{i,j} = \sum_{e \in E} (f_{e_0} - f_{e_1})^2.$$

Remember there is a sign reversal in the definition of the graph Laplacian, so $Lf = \lambda f$. This formula shows that λ measures the smoothness of f, expressed by the sum of the square of the difference between the values of f over vertices connected by an edge.

The Laplacian is intimately connected to the Fourier transform. The Fourier transform decomposes a function according to the frequency. Its basis functions are in fact the eigenfunctions of the Laplacian (and more generally of any linear operator that commutes with translations on the Euclidean space):

$$\Delta e^{-2\pi i \xi x} = -4\pi^2 \xi^2 e^{-2\pi i \xi x}.$$

The idea of the Fourier transform could be generalized to the sphere via spherical harmonics, which are eigenfunctions of the Laplace-Beltrami operator, or more generally of

all linear operators that commute with rotations. The generalization could go further to general spaces without symmetry or even discrete spaces like graphs, always via the eigenfunctions of the Laplacian.

The Laplacian grasps important geometric information of the underlying space. The geometry of the domain can almost be recovered from the eigenvalues of its Laplacian. In 1966, Mark Kac published a paper with the famous title "Can One Hear the Shape of a Drum ?"[84]. The vibration modes of a drum are given by the eigenfunctions of the Laplacian and its frequencies are proportional to the square root of eigenvalues of the Laplacian[138]. The question is whether conversely the eigenvalues of the Laplacian completely determine the shape of the domain. In 1992 Gordon et al.[58] gave the negative answer for this question. However, the answer could be yes with some reasonable restrictions on the convexity and the boundary of the domain[189].

The first non trivial eigenvalue (also called Fiedler value in spectral graph theory) is related to the existence of a "bottleneck" in the shape by Cheeger's inequality. The Cheeger constant is the infimum of the volume of the boundary of a cut over the volume of the submanifold (the smaller half of the two submanifolds generated by the cut). The Cheeger's inequality gives a tight bound on the first eigenvalue: $\lambda_1 \geq \frac{h^2}{4}$, where λ_1 is the smallest non-zero eigenvalue of Laplacian and h the Cheeger constant of the manifold. It is of great interest in low dimensional topology but also in graph theory to search for a sparse cut[113,33].

2.1.2 Application

Beyond mathematics and physics, the Laplacian is also used in applied fields for example in the geometry processing field in which the eigenvectors serve as a functional basis over

a triangle mesh [100]. In the EEG processing domain, the Laplacian (also known as Surface Laplacian (SL) in this domain) is a well-known filter in order to improve spatial resolution and diminish the volume conduction effect [25]. Specifically, in BCI research, the first or second order approximation of the Laplacian is often used to improve accuracy and signal orthogonality [75,108]. The "global" Laplacian involving all electrodes is seldom used, since the effect as a spatial filter is comparable between the approximate and global Laplacian [34].

Recently, the graph signal processing (GSP), a method based on analyzing graph signals using eigenfunctions of the Laplacian as a functional basis, has started to be applied in neuroscience [102]. There are evidences that showed correspondence between the harmonics of the brain networks and the resting-state brain waves [8], as well as the relevance between different spectral components, and the levels of adaptability throughout learning and the levels of task familiarity [78].

The idea is analogous to the Fourier transform of time series into the frequency domain. The difference here is that the frequency is in a spatial instead of a temporal sense. The eigenfunctions of the Laplacian carry geometric or topological information of the underlying network depending on the nature of the network (involving distance or not). Representing the brain signal as a continuum of spectral components is more beneficial than the traditional parcellation methods because it suits better the functional heterogeneity and multiplicity [67]. Besides, it allows multimodal analysis of brain activity using different sources of data with complementary advantages, for example projecting EEG data on harmonics of the white matter connectivity network [57].

Common applications of Fourier analysis like dimensionality reduction also apply here. Graichen et al. [61,62] constructed a mesh from the coordinates of EEG electrodes and decom-

posed the EEG signals into the eigenspaces of the Laplacian of the mesh. The capability of dimensionality reduction of this method was compared with principal component analysis (PCA) and independent component analysis (ICA). Gramfort and Clerc embedded EEG data into a 2-dimensional space using graph Laplacian based on Gaussian kernel and showed that the time-lag information could be captured in this low-dimensional representation[63].

2.2 Dimensionality reduction

The spatial information contained in EEG signals is much less explored compared to the time-frequency information. Even though the CSP filter belongs to the category of spatial filters, it does not really take into consideration the spatial information. Only a linear combination of channels is searched without caring for the location of the channels. However, if we consider the EEG signals as a function sampled over the manifold of the scalp, then the geometric information could be employed via the Laplacian of the scalp.

The (normalized) eigenfunctions of the Laplacian form an orthonormal basis for the square integrable functions over the manifold of the scalp, a family of functions including EEG signals. This means we can rewrite the EEG signals as a weighted sum of the eigenfunctions of the Laplacian (see Fig. 2.2.1). In the discrete setting, note the matrix of Laplacian eigenvectors by $\Phi = [\varphi_1, \ldots, \varphi_n] \in \mathbb{R}^{n \times n}$. Then, the transformation of the signal X from the spatial domain to the spectral domain is done by:

$$\widehat{X} = \Phi^{\mathsf{T}} X. \tag{2.2.1}$$

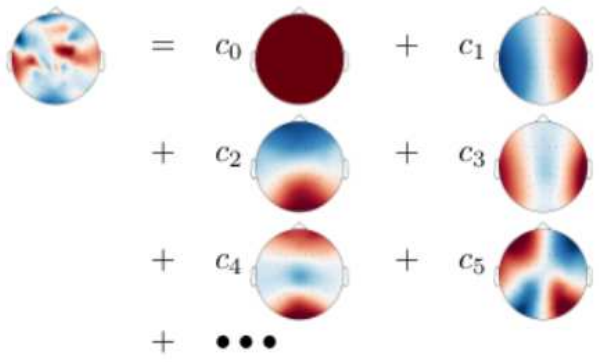


Figure 2.2.1: EEG data can be expressed as the weighted sum of eigenvectors of the Laplacian. Data dimension can be reduced by dropping high frequency components.

Here $\widehat{X}$ corresponds to the vector of coefficients $(c_0, \ldots, c_{n-1})^\intercal$ in Fig. 2.2.1. The original signal X can be reconstructed via the inverse transform:

$$X = \Phi\widehat{X}. \tag{2.2.2}$$

As in most natural phenomena, the energy is mainly concentrated on low frequency components. By dropping high frequency components, the essential information is still retained and the dimension of the data is reduced. An extra bonus is a low-pass filtering effect that removes high spatial frequency noise.

Putting this idea into practice comes in two ways. The difference is the underlying structure: one being the triangle mesh constructed from the 3D positions of the electrodes, the other being the functional connectivity graph estimated from the EEG data. Note that the first way is data-independent while the second is not. The illustration of the processing steps is shown in Fig. 2.2.2. In the following, Section 2.2.1 and 2.2.2 detail the computation of the Laplacian matrix in these two ways and the results are presented in Section

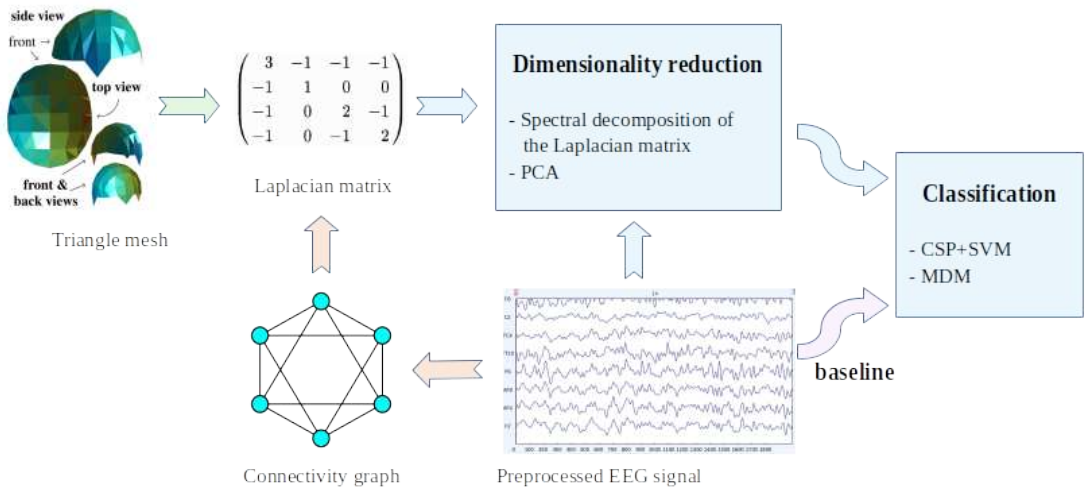


Figure 2.2.2: Applying Laplacian spectral decomposition to EEG-based BCI. The Laplacian matrix can be computed either from the triangle mesh constructed with the 3D coordinates of the electrodes (green arrow) or from the functional connectivity graph estimated from the EEG data (orange arrow). Note that the first way is data-independent while the second is not. Then the EEG data is projected on the space spanned by the eigenvectors of the Laplacian or by the principal components (PCA used for comparison). Finally, two gold standard classifiers are applied to give the classification accuracy. The baseline (purple arrow) is the normal classification pipeline without dimensionality reduction.

2.2.3.

For researchers in the BCI domain, attention should be payed to the difference between the traditional use of the Laplacian as a filter, and the method discussed here whose main ingredient is the spectral decomposition of the Laplacian.

2.2.1 Geometry-based method

The real scalp can be approximated by a triangle mesh constructed from the 3D coordinates of the electrodes in the standard 10-20 system. The approximation of the Laplacian on discrete meshes has been extensively studied by the geometry processing community[182,42,111]. The well-known cotangent formula[183,147] is used here for its good theoretical properties (convergence) and fast computation time.

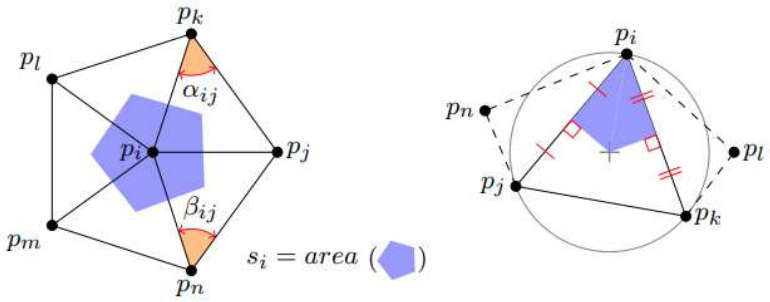


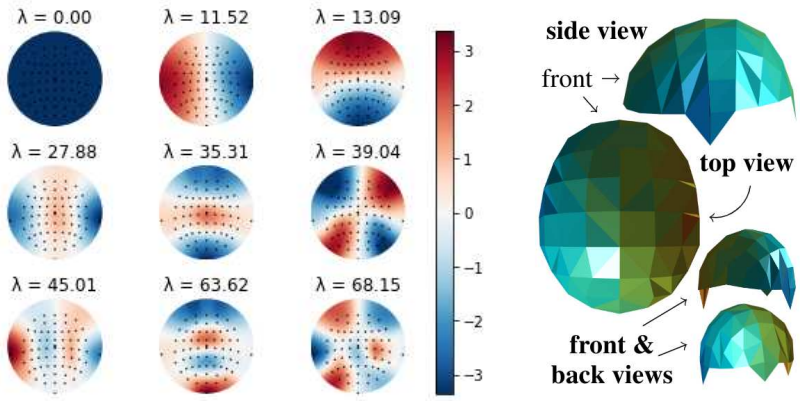
Figure 2.2.3: Triangle mesh data used to compute the discrete Laplacian. In the picture on the left, the angles α_{ij} and β_{ij} used in the cotangent matrix are displayed here for indices i and j corresponding to the vertices p_i and p_j in the mesh. The value of s_i is the area of the shaded region defined as the union of two triangles for each face containing p_i. These triangles are depicted in the right picture: their vertices are p_i, the middle of an adjacent edge, and the circumcenter represented by a cross.

In the discrete setting, the Laplacian is represented by a $n \times n$ matrix L where n is the number of nodes. The formula is given by $L = S^{-1}M$, where S is the *area matrix*, the diagonal matrix with the i^{th} diagonal value equals to the sum of area of small patches surrounding the i^{th} vertex (the area of the shaded region in Fig. 2.2.3) and M is the *cotangent weights matrix* defined by:

$$M_{ij} = \begin{cases} \sum_{k \neq i} m_{ik} & \text{if } i = j, \\ -m_{ij} & \text{if } i \text{ and } j \text{ adjacent}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.2.3}$$

m_{ij} are cotangent weights: $m_{ij} = \frac{1}{2}\big(\cot(\alpha_{ij}) + \cot(\beta_{ij})\big)$ when i and j are adjacent and $m_{ij} = 0$ otherwise. The angles α_{ij} and β_{ij} are the two angles on face of the edge connecting the i^{th} and j^{th} vertices as shown in Fig. 2.2.3.

The discrete version of the Helmholtz equation (Eq. 2.1.1) is $-L\varphi = \lambda\varphi$ and can be rewritten as a generalized eigenvalue problem $M\varphi = \lambda S\varphi$. This can be solved efficiently



(a) Topomaps of the first 9 Laplacian eigenvectors and eigenvalues

(b) Triangle mesh

Figure 2.2.4: (a) Topomaps of the first 9 eigenvectors and eigenvalues of the Laplace-Beltrami operator. (b) Underlying mesh constructed using the Physionet MI dataset montage [150,151]. The eigenvectors corresponding to increasing eigenvalues have increasing spatial variation.

by the Arnoldi method of ARPACK (implemented inside MATLAB). Fig. 2.2.4 shows the topomaps of the first 9 eigenvectors and corresponding eigenvalues of the Laplacian computed as described in Algorithm 1. The first eigenvector that corresponds to the eigenvalue $\lambda = 0$ is a constant vector. Ordered according to the corresponding eigenvalues in increasing order, the eigenvectors vary from the most smooth to the least smooth.

Algorithm 1 Computation of the Laplacian and its spectral decomposition

---

Require: EEG electrodes' 3D coordinates (x, y, z)
Ensure: Eigenvectors and eigenvalues of the Laplacian
1: Create a mesh with electrodes as vertices using the Delaunay triangulation [96]
2: Compute the area matrix S and the cotangent matrix $\mathcal{M}$
3: Solve the eigenvalue problem $\mathcal{M}\varphi = \lambda S\varphi$ using the `eigs` MATLAB function

---

2.2.2 Connectivity-based method

As discussed in Section 1.2, EEG signals are not optimal to estimate cerebral functional connectivity since spurious connections might appear when two electrodes record signals issued from one common source. Among many functional connectivity measures proposed in the literature, the Phase Lag Index (PLI) is known to be less influenced by this problem [159]. The PLI is a phase-based measure. To estimate the phase of a 1-dimensional time series $x(t)$, the analytical signal $z(t)$ based on the Hilbert transform is computed,

$$z(t) = x(t) + i\tilde{x}(t) = A(t)e^{i\varphi(t)},$$

where

$$\tilde{x}(t) = \frac{1}{\pi} PV\left(\int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau}\, d\tau \right)$$

is the Hilbert transform of $x(t)$, and PV refers to the Cauchy principal value. Then the phase is obtained by $\varphi(t) = \arctan \frac{\tilde{x}(t)}{x(t)}$.

The key idea of PLI is to omit the phase differences around $0 \mod \pi$ by measuring the asymmetry of the phase differences, which implies a consistent non-zero phase difference thus exclude the spurious connections coming from a common source. PLI is defined by

$$PLI = \left| \frac{1}{N} \sum_{k=1}^{N} \operatorname{sgn}\big(\Delta\varphi(t_k)\big) \right|,$$

where sgn is the sign function and $\Delta\varphi(t_k)$ is the phase difference at time t_k. Note that the value of PLI is between 0 and 1.

To further reduce the sensitivity of PLI due to its discontinuity, Vinck et al. proposed

the weighted phase lag index (wPLI) by scaling contributions of phase differences by their magnitudes[172]. The wPLI is chosen for its robustness in this study to estimate the functional connectivity from EEG data during MI task.

The functional connectivity graph is a non-directed weighted graph with $w = 1 - \text{wPLI}$ as weights between nodes. The reasoning behind this setting is that the larger the wPLI is, the more the two electrodes are "connected", and therefore the smaller the "distance" or weight. Thresholding is not applied to avoid introducing artificial parameters. The eigen-decomposition of the Laplacian is done with a standard function that one can find in many scientific computing packages (e.g. SciPy or NumPy in Python). Taking one subject from the Physionet MI dataset as an example, the topographic maps of the first 9 eigenvectors (ordered with increasing eigenvalues) are shown in Fig. 2.2.5. Since the underlying network is the functional connectivity during an MI task, the patterns of dipoles above the sensori-motor areas are quite consistent. The Fiedler vector (eigenvector associated with the smallest non-zero eigenvalue, in Fig. 2.2.5 with $\lambda = 19.45$) is the smoothest non-constant vector over the graph. So the value (expressed by color in Fig. 2.2.5) of the Fiedler vector goes from one extremity of the graph to the other, and we can see two distinct groups of channels: one over the motor cortex in the center (in red) and the other around it (in blue).

2.2.3 Dimensionality reduction

Two gold standard classification methods for BCI are used as baseline to test the classification and dimensionality reduction in the Laplacian spectral domain: the common spatial pattern (CSP) filter with a Support Vector Machine (SVM)[38] , and the Tangent Space Riemannian Classifier (RC)[10] (cf. Section 1.4 for detailed description). Using the above clas-

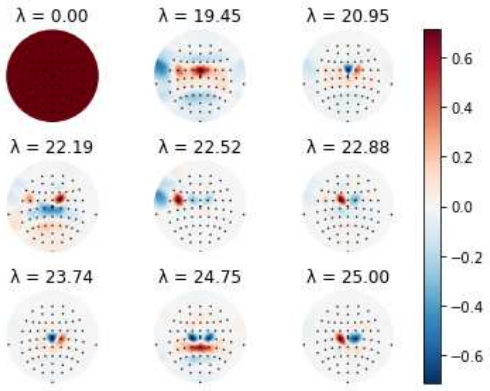
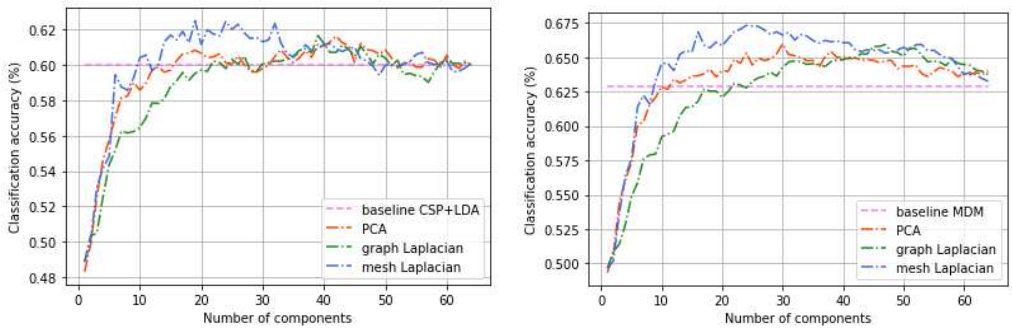


Figure 2.2.5: Topomaps of the first 9 eigenvalues and eigenvectors of the connectivity graph Laplacian (computed with the data from one subject of the Physionet MI dataset).

sification methods in the spectral domain has a different interpretation compared to using them in the spatial domain. CSP in the spectral domain could be seen as looking for a linear combination of spectral components that best differentiate between classes. The covariance matrices used in RC contain the correlation coefficients of pairs of spectral components instead of pairs of spatial components (i.e. electrodes).

Dimensionality reduction is done by keeping only the first spatial frequency components, as most information is retained by the smoothest components. This induces at the same time a low-pass effect on signals thus naturally suppressing noise and leading to better classification results as detailed hereafter. Two versions of the Laplacian, i.e. based on triangle mesh and functional connectivity graph, are tested and compared with principal component analysis (PCA) which is a standard unsupervised dimensionality reduction method.

The pipelines are shown in Fig. 2.2.2. The dataset used in this study is the Physionet MI dataset[150,151] (cf. Section 1.5 for descriptions of the dataset and the preprocessing steps). Only left and right classes were used. The intra-subject classification accuracy using an in-

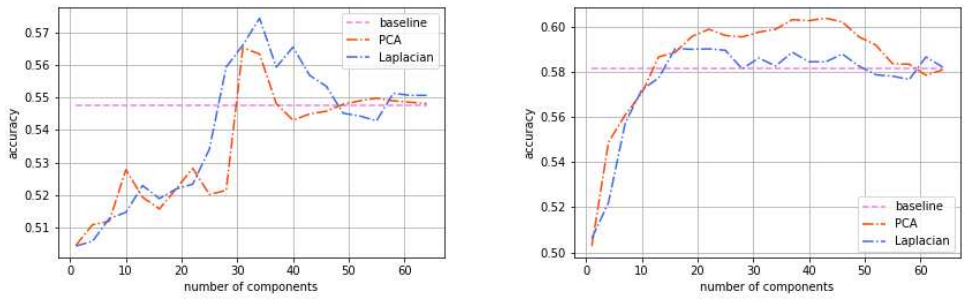


(a) CSP+SVM (b) RC

Figure 2.2.6: Intra-subject classification accuracy as a function of the number of principal components or Laplacian spectral components used for dimensionality reduction, averaged over 105 subjects. Baseline: CSP+SVM and RC without dimensionality reduction.

creasing number of principal components or Laplacian spectral components is shown in Fig. 2.2.6 as well as the performance of the baseline methods without dimensionality reduction. The classification accuracy is estimated using a 10-fold non-shuffled cross-validation procedure. The classification accuracy of both methods, i.e. CSP+SVM and RC, is improved with dimensionality reduction. The mesh Laplacian method works the best, improving the accuracy of CSP+SVM and RC by 2% and 5% respectively with a number of components equal to only 1/3 of the total number of channels.

The same pipelines were also tested on inter-subject classification. Only the mesh Laplacian method was compared to the PCA method since the connectivity graph Laplacian did not work well in intra-subject classification. The results are shown in Fig. 2.2.7. The Laplacian method works a bit better than the PCA method when combined with CSP+SVM and it is the inverse with RC. However the improvement in both cases are merely around 2%.



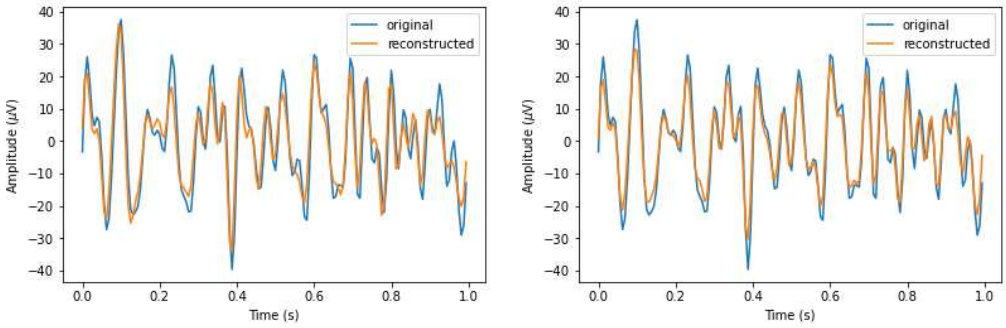
(a) CSP+SVM (b) RC

Figure 2.2.7: Inter-subject classification accuracy as a function of the number of principal components or Laplacian spectral components used for dimensionality reduction, averaged over 105 subjects. Baseline: CSP+SVM and RC without dimensionality reduction.

2.3 Further analyses

It is clear from the Fig. 2.2.6 that the mesh Laplacian performs best. It starts to get better results than the baseline (with 64 channels) from using only 10 eigenvectors. This shows the utility of the geometric information often neglected. Using the Eq. 2.2.2 the signal could be reconstructed after the truncation in spectral domain. Fig. 2.3.1 shows the reconstructed signals using 10 and 25 (which gives in average the best accuracy) Laplacian eigenvectors. The recovered signal with 10 eigenvectors (Fig. 2.3.1 (a)) fits well the original signal despite some discrepancy. The original signal is almost perfectly recovered with 25 eigenvectors (Fig. 2.3.1 (b)) and the sharp transitions as well as extreme values are attenuated.

The noise reduction effect is both similar and different from the usual low-pass filtering. Both are based on the Fourier transform but one in the time domain and the other in the spatial domain. What we did here is equivalent to requiring the signal of each channel



(a) (b)

Figure 2.3.1: EEG signal reconstructed with 10 (a) and 25 (b) Laplacian eigenvectors respectively. Original signal is extracted from one subject in the database.

to be more coherent with the neighboring channels while the low-pass filter demands the smoothness in time domain thus keeping only slower oscillations.

While the functional connectivity graph is useful for other purposes, its performance on dimensionality reduction is not good especially on low dimensions. There are several channels less connected to the entire network and thus the low frequency eigenvectors tend to have peak values on these channels (see Fig. 2.2.5), which reduces their capacity to capture global information. In consequence the energy is not concentrated in low frequency components as in the case using the mesh Laplacian. Fig. 2.3.3 shows the average power (over subjects and trials) of the MI EEG signal in the spectral domain. We see that the power per spectral component as well as its variation decrease very quickly for the mesh Laplacian (a) but not for the graph Laplacian (b). Since the dimensionality reduction is done via a low-pass filter, the information loss is much higher for the graph Laplacian than for the mesh Laplacian.

On the contrary, the Laplacian eigenvectors of structural connectivity graphs have been shown to be quite relevant to various mental activities. Basic cognitive functions such as

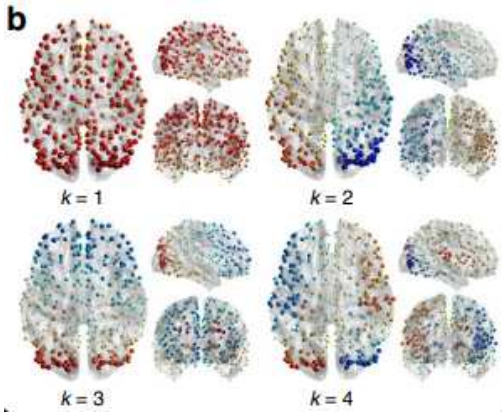
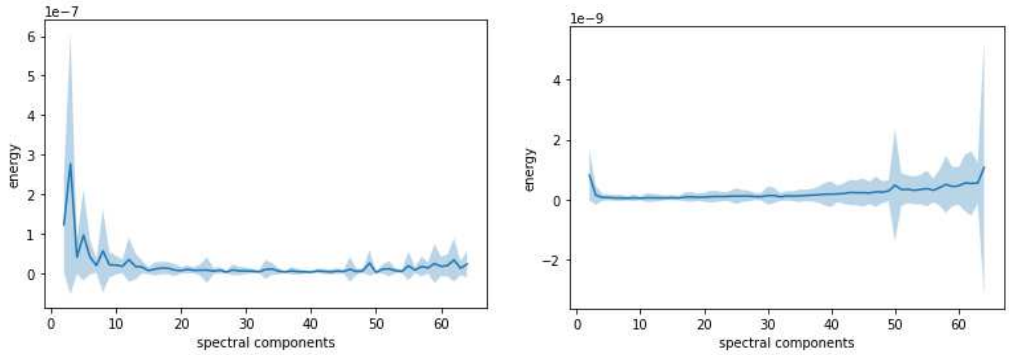


Figure 2.3.2: The first 4 Laplacian eigenvectors of the structural connectivity graph. From Preti & Van De Ville [135].

visual perception, motor movements and auditory processing are more aligned with low frequency eigenvectors, while high level cognitive functions like emotion, memory and cognitive control have more energy on high frequency eigenvectors [135]. Interestingly, the first few low frequency Laplacian eigenvectors of the structural connectivity graph are quite similar to that obtained from mesh (see Fig. 2.3.2), both being quite smooth and capturing lateral information (e.g. left/right, front/back).

Besides the improvement of classification accuracy, the dimensionality reduction also reduces considerably the computation time, especially for the Riemannian classifier, see Fig. 2.3.4 (computation done on a desktop computer with Inter Core i7-4790 CPU and 15.6 GB of RAM memory). The best number of components to use can be chosen by cross validation. Since the mesh is constructed using the standard 3D coordinates of the 10-20 system, it is subject-independent and does not require solving an eigenvalue problem as for the PCA, which saves the computation time by 33.7% (evaluation based on the computation time provided by the MOABB framework).



(a) (b)

Figure 2.3.3: Average spectral power using the mesh Laplacian (a) and the graph Laplacian (b) respectively. The shaded area denotes standard deviation (Physionet MI dataset).

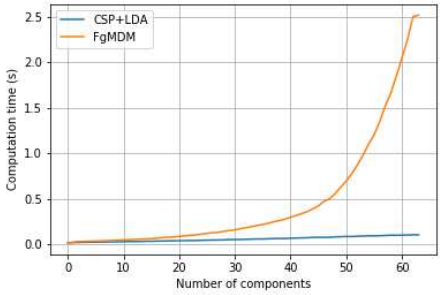


Figure 2.3.4: Computation time depending on the number of kept Laplacian spectral components.

2.4 Conclusion

This work considers the underlying geometric structure of the multichannel EEG signal by incorporating channel location or functional connectivity information. The geometric information is captured via the Laplacian of the triangle mesh or the connectivity graph respectively. The eigenfunctions of the Laplacian form an orthogonal basis for the square integrable functions over the considered domain. Similar to the Fourier transform, the original signal can be represented in the Laplacian spectral domain. The representation is adapted to the geometry of the domain and is sparse, especially in the mesh case. Thus, dimensionality reduction was made possible by applying a low-pass filter in the spectral domain.

Even though this operation is a linear transformation – hence would not turn a non-stationary signal to a stationary one – it is still beneficial for practical applications. The experiments show that the classification performance of gold standard methods could be improved by 2-5% in intra- and inter-subject classification while reducing significantly the computation time. This method might also be used as a noise reduction method and be compared with other existing denoising methods.

While it was impossible in this study to use the structural connectivity since we do not have structural connectome data and the source reconstruction from EEG data is also very complicated, there still remains many interesting avenues to explore even with these restrictions for example the relationship between temporal and spatial frequencies in order to leverage the high temporal resolution of EEG. Or even better, a joint spatial-temporal filtering could be designed as Perraudin et al. [127] showed that a joint approach performed better

in denoising and recovering temperature time series from multi-stations since it leverages the correlation both in the time and the spatial domain.

This work was promoted at an international conference: X. Xu, N. Drougard and R. N. Roy, "Dimensionality Reduction via the Laplace-Beltrami Operator: Application to EEG-based BCI," 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER), 2021, pp. 347-350, doi: 10.1109/NER49283.2021.9441404.

Laplacian spatial-spectral filtering for EEG-based BCI:

- Different from the traditional use of Laplacian as a spatial filter in EEG-based BCI, this method leverages the spectral decomposition of the Laplacian to provide a multiscale and sparse representation of the EEG signal.
- Applying a low-pass filter in the spectral domain reduces dimension and slightly improves both the intra- and inter-subject classification accuracy in BCI applications.
- The mesh Laplacian is purely geometric and independent of the brain signal, thus requiring less computation time.

Larghetto, con amore

3

Canon: Path signature

Canon is one of the musical forms that apply the contrapuntal compositional technique. It is characterised by having several voices imitating the initial melody in a time-delayed manner. The musicians use the latin terms *dux* and *comes* to denote the leader and the follower voices. In the music composed by the multichannel EEG signal there also exists this

kind of structure that can be perfectly captured by the path signature, even with tempo changes.

3.1 Path signature

A *path* is a map from a time interval to a certain domain. The discretized version of a path is also known as time series in the machine learning community and it can be found everywhere in our daily life for example the meteorological data or the stock market data. We can find various ways to characterize a path. For example the mean value of a path of temperature is a good indicator of the weather and the standard deviation of the price of a stock informs the fluctuation and the risk that an investor may take in the future. The *path signature* is also such a kind of feature that is associated with a path, just in a bit more complicated way. To put it simply, it is a series of iterated integrals that compute the volume of a singular simplex bounded by the axes and the projection of a multidimensional path on a subspace.

Path signature was originally introduced by Chen [28] in the framework of piece-wise smooth curves, then developed by Lyons et al. [104] to study stochastic differential equations. Recently it has been used as a feature generation method for time series in various applications such as handwritten character recognition [187], diagnosis of bipolar disorder and borderline personality disorder based on daily mood ratings [125], and diagnosis of Alzheimer's disease based on hippocampal and brain volume time series [114], just to name a few. To our knowledge, it has never been applied to EEG data nor BCI applications. Interested readers can find more details of the theory and also the practical use cases of path signature in Chevyrev & Kormilitzin [29]. Here we only give a simple introduction of basic concepts and

properties relevant to our application on BCI.

DEFINITION Consider a d-dimensional path $X_t : [a,b] \to \mathbb{R}^d$, which is a map from $[a,b]$ to $\mathbb{R}^d$ and suppose it to be continuous. We can also note it by $X_t = \{X_t^1, X_t^2, \ldots, X_t^d\}$ where X_t^i denotes its component in each dimension $i \in \{1,\ldots,d\}$. Its *path signature* is defined as an infinite series:

$$S(X)_{a,b} = (1, S(X)_{a,b}^1, \ldots, S(X)_{a,b}^n, S(X)_{a,b}^{1,1}, S(X)_{a,b}^{1,2}, \ldots)$$

where a path signature at level k is recursively defined as an integral over a k-dimensional simplex:

$$\begin{aligned} S(X)_{a,b}^{i_1,\ldots,i_k} &:= \int_{a<t<b} S(X)_{a,t}^{i_1,\ldots,i_{k-1}}\, dX_t^{i_k} \\ &= \int_{a<t_k<b} \int_{a<t_{k-1}<t_k} \cdots \int_{a<t_1<t_2} dX_{t_1}^{i_1}\, dX_{t_2}^{i_2} \ldots\, dX_{t_k}^{i_k} \end{aligned}$$

the multi-indexes are drawn from $\left\{(i_1,\ldots,i_k) | i_1,\ldots,i_k \in \{1,\ldots,d\}\right\}$ and the first level signature is defined as $S(X)_{a,b}^i = \int_a^b dX_t^i$. In the following we will use the terms path signature or signature interchangeably.

EXAMPLE Here we explicit the first two levels of signature to have a more intuitive understanding. By definition,

$$S(X)_{a,b}^i = \int_a^b dX_t^i = X_b^i - X_a^i = \Delta X^i. \tag{3.1.1}$$

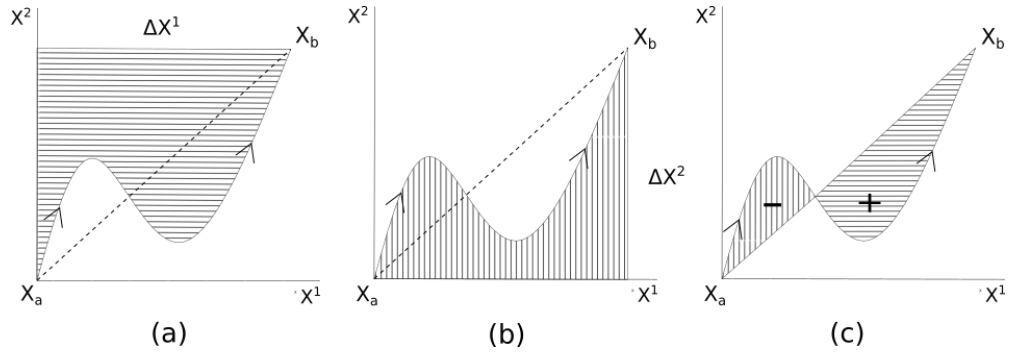


Figure 3.1.1: Geometric interpretation of level 2 signature. The shaded area in the subfigures represents (a) $S(X)^{1,2}_{a,b}$, (b) $S(X)^{2,1}_{a,b}$, and (c) the signed area A_{signed} (the difference between areas "+" and "−") respectively.

So the first level signature is totally determined by the increment of the path along each coordinate, (i.e. the difference between the ending and starting point on one dimension).

The second level signature captures richer information about the relationship between the pair of path components and it has a beautiful geometric interpretation. As shown in Fig. 3.1.1, $S(X)^{i_1,i_2}_{a,b}$ is the area bounded by the path $X_{[a,b]}$ and the X^{i_2}-axis and the difference between $S(X)^{1,2}_{a,b}$ and $S(X)^{2,1}_{a,b}$ is proportional to the signed area (also known as Lévy area).

$$S(X)^{1,2}_{a,b} = \int_a^b (X^1_t - X^1_a)\, dX^2_t = \frac{1}{2}\Delta X^1 \Delta X^2 + A_{signed}, \tag{3.1.2}$$

$$S(X)^{2,1}_{a,b} = \int_a^b (X^2_t - X^2_a)\, dX^1_t = \frac{1}{2}\Delta X^1 \Delta X^2 - A_{signed}, \tag{3.1.3}$$

and thus

$$\frac{1}{2}(S(X)^{1,2}_{a,b} - S(X)^{2,1}_{a,b}) = A_{signed}.$$

Properties It is immediate from the definition that path signature is invariant with respect to time reparametrization (the integration is on X_t^i). If we consider the path signature as a feature of multivariate time series, we might want the mapping between the two to be one-to-one. However, it is not the case because of the invariance under time reparametrization. Nevertheless, path signature fully characterizes paths up to tree-like equivalence[68]. In practice, we seldom meet tree-like paths (paths who travel backward on itself in some part), so path signature remains a characteristic feature.

The shuffle product identity states that the product of two terms of lower level signature can be expressed as a linear combination of higher level signature. Following the previous example we can see that $S(X)^1_{a,b}S(X)^2_{a,b} = S(X)^{1,2}_{a,b} + S(X)^{2,1}_{a,b}$ from equations 3.1.1, 3.1.2, and 3.1.3. In fact, the non-linearity of the path signature is sufficient to approximate any continuous bounded function in path space by a linear functional in signature space[30], which reduces a non-linear classification problem to a linear one.

There are some other properties that attract the machine learning community[55]. In practice, the signature is truncated at a certain level so it provides a fixed length feature regardless of the number of time steps. Besides, path signature is fast to compute and easy to be used online since the integrals are additive over a partition of time steps.

The application of path signature for EEG-based BCI is strongly motivated by its property of *translation invariance*, which is a direct consequence of its definition (a translation of the path doesn't change the resulting volume), and its invariance under time reparametrization. Indeed, as we mentioned in Section 1.2, EEG data varies greatly between sessions and subjects. Thanks to path signature, we hope that some of these variations (e.g. the covariate shift or the temporal difference) might be removed and that we could find some robust

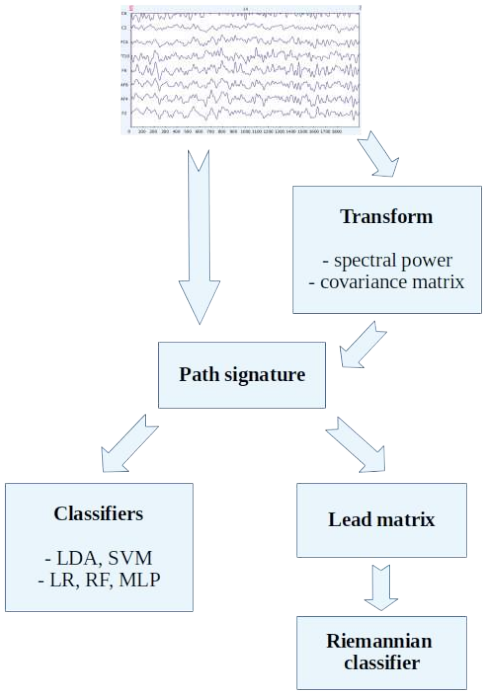


Figure 3.2.1: Different pipelines used in this chapter. Path signature could be used directly on multi-variate EEG time series, or alternatively on the time series of spectral power or covariance matrices extracted using a sliding window. Then either the truncated path signature can be used directly as a feature vector, or the lead matrix can be computed from the level 2 signature and then turned into an SPD matrix to be classified by a Riemannian classifier.

features.

3.2 Application

In this section, several methods for applying path signature to EEG data are presented, along with the results and analyses obtained using the data presented in Section 1.5. Path signature could be used directly on multi-variate EEG time series (cf. Sec. 3.2.1), or alternatively on the time series of spectral power (cf. Sec. 3.2.1) or covariance matrices (cf. Sec. 3.2.3) extracted using a sliding window. Then either the truncated path signature can be used directly as a feature vector, or the lead matrix can be computed from the level 2 sig-



Figure 3.2.2: Pipeline of applying path signature on raw data. First a windowed FIR filter is applied to filter out the task related signal. Then the signature of multivariate EEG time series is computed and this is the features that are fed to the classifiers to get the classification results.

nature and then turned into an SPD matrix to be classified by a Riemannian classifier (cf. Sec. 3.3.2). Fig. 3.2.1 illustrates the various pipelines in the whole chapter. Except when stated explicitly, the BCI competition IV 2a dataset was used throughout the study. Signature computation was done using the Signatory package[88].

3.2.1 On raw data

The most natural way is applying path signature directly on EEG time series and to use it as a feature map. More precisely, we considered the EEG signal X_t^n as a n-dimensional path $X : [0, T] \mapsto \mathbb{R}^n$ and computed the truncated signature up to level k. Then a classical classifier was applied on this feature vector. The pipeline is shown in Fig. 3.2.2.

Cross validation (10-fold for intra-subject and leave-one-out for inter-subject) was performed on the data with the left vs right motor-imagery paradigm. The results were summarized in Table 3.2.1 (cf. Nomenclature for the abbreviation of the name of the classifiers). Note that even though we call it "raw data", the data was still filtered into the 8-30 Hz frequency band relevant to motor imagery.

The results in Tab. 3.2.1 demonstrated a significant advantage of level 2 signatures over other levels for decoding mental states. The best accuracy is achieved at level 2 for all classifiers in intra-subject classification with a significant margin. For example, with the logistic

	level	1	2	3	4
intra	SVM	50.5(12.9)	66.1(15.2)	55.7(13.3)	56.9(13.9)
	LDA	51.0(12.6)	63.4(14.0)	54.7(14.2)	54.6(14.2)
	LR	50.9(11.3)	**67.1**(13.5)	56.6(13.3)	56.6(14.0)
	RF	49.7(11.6)	59.7(16.1)	58.0(16.8)	56.6(18.0)
	MLP	52.3(11.2)	61.6(16.4)	54.8(13.3)	56.8(15.1)
inter	SVM	52.6(3.6)	53.9(6.1)	54.2(6.2)	52.8(4.9)
	LDA	52.5(4.9)	53.5(6.2)	53.2(4.9)	53.9(6.0)
	LR	53.5(4.4)	54.7(5.8)	54.4(6.4)	53.6(4.9)
	RF	51.0(4.5)	54.6(6.0)	54.2(7.8)	52.0(4.6)
	MLP	53.1(5.6)	**58.7**(8.3)	56.5(5.2)	52.2(5.7)

Table 3.2.1: Mean cross validation score of various classifiers and truncation levels of path signature applied on raw data of the BCI competition IV 2a dataset (left vs right). Refer to the nomenclature at the beginning of thesis for the abbreviation of the algorithms (standard deviation between parentheses). The hyperparameters were the default ones in the scikit-learn library.

regression classifier the classification accuracy hits 67.1% at level 2 which is improved by more than 10% compared to other levels. Regarding inter-subject classification, the advantage is weakened especially for non-linear classifiers which are prone to over-fitting.

Feature importance analysis We examined feature importance by looking at the coefficients of the two best classifiers in intra-subject classification, SVM and LR, after scaling the features. The 5 signatures with the largest magnitude of coefficient are shown in Tab. 3.2.2. There is consistency of the sets of signatures most used by the two classifiers, and they are mostly level 2 signature of a pair of channels or a level 1 signature around C3 or C4, which makes sense since we used motor imagery data (C3 and C4 are respectively located directly above the right and left hand sensorimotor areas).

Great similarity between the two classifiers is also found in coefficients of level 1 signatures, see Fig. 3.2.4. Unfortunately, no clear pattern across subjects was present here, since

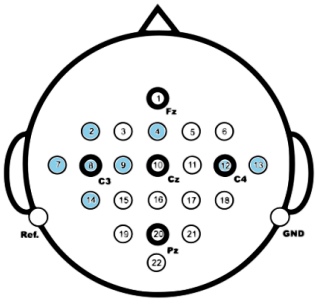
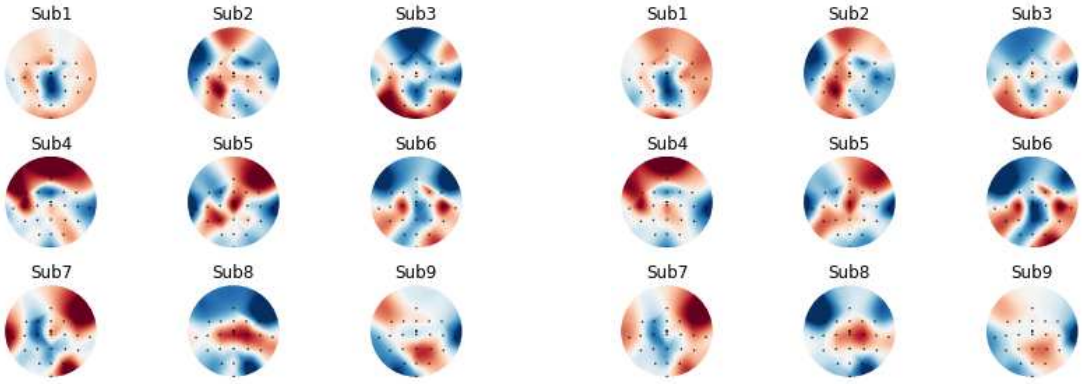


Figure 3.2.3: Electrode montage of BCI competition IV 2a dataset. In blue are the electrodes that appear in the table at the right.

SVM	LR
(8, 2)	(2, 8)
(2, 8)	(8, 2)
13	(9, 9)
(9, 9)	13
(4, 7)	(14, 12)

Table 3.2.2: Index of the top 5 most important signatures

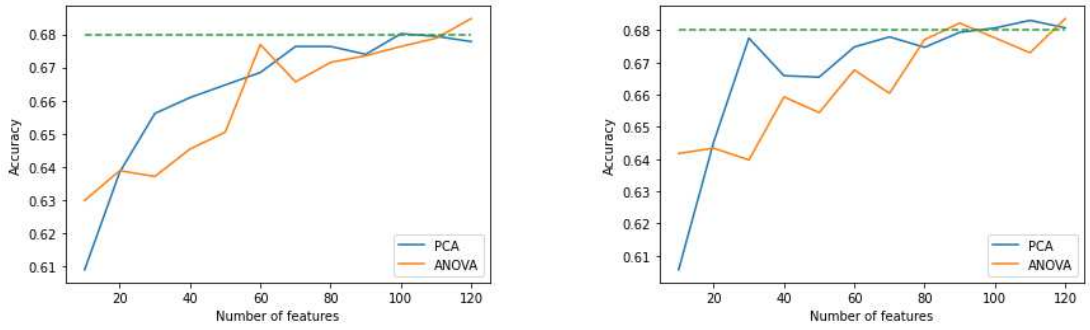
we have seen in Tab. 3.2.2 that the two classifiers rely mostly on level 2 signatures. Recall that level 1 signature is just the difference between the ending and starting point, while level 2 signature captures richer information about the relationship between the pair of channels. This will be further discussed in Section 3.3.



(a) SVM coefficients (b) LR coefficients

Figure 3.2.4: Topographic maps of coefficients over level 1 signatures.

Dimensionality reduction The dimension of signature up to level k is at the order of $O(n^k)$. In practice the log signature transform [29] can be used to remove some redun-



(a) SVM (b) LR

Figure 3.2.5: Intra-subject classification accuracy using up to level 2 signature as features and with dimensionality reduction (PCA or ANOVA). The baseline without dimensionality reduction is marked by the green dashed line.

dancy and reduce the dimension. Even though the dimension of the feature vector did not explode if the signature is truncated at level 2, it was still far above the number of available trials and could lead to overfitting. Hence, we still tried several dimensionality reduction methods in the hope of improving the results. Both unsupervised and supervised methods are tried, one using PCA and the other using a feature selection based on the ANOVA F-value, with the two best classifiers for intra-subject classification. The accuracy using different number of features can be found in Fig. 3.2.5, the green slashed line being the baseline without using dimensionality reduction. As can be observed dimensionality reduction is not quite helpful here, with a lower performance than the baseline reached as soon as a few features are dropped. Interestingly, we also observe that the PCA quickly surpasses the ANOVA-based selection method, even though the former being unsupervised.

Riemannian classifier revisited While working with path signature, we wondered why the Riemannian classifier worked well. Is it because the covariance matrices are good features, or that the Riemannian metric plays an important role?

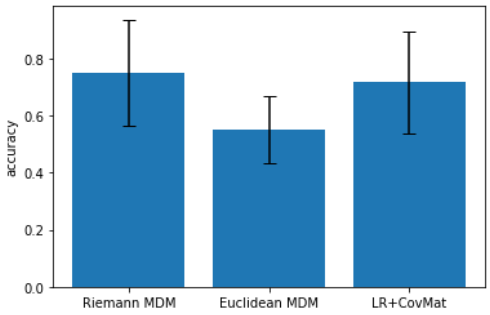


Figure 3.2.6: Intra-subject classification accuracy of various variants of Riemannian classifier. The first one is the MDM classifier with covariance matrices as features and Riemannian metric to estimate the distance. The second one is the MDM classifier with covariance matrices as features and Euclidean metric to estimate the distance. The third one is the LR classfier with flattened covariance matrices as features.

To answer this question, the covariance matrices were simply flattened (all entries were kept) and a logistic regression classifier was applied. Surprisingly, the result is relatively good. The classification accuracy is 71.6%, which is comparable with the same pipeline applied on path signature. Another test was done by using either a Euclidean or Riemannian metric in the MDM classifier. The results are 55.1% and 74.9% with Euclidean and Riemannian metric respectively (see Fig. 3.2.6).

The conclusion is that the Riemannian metric is important, especially when working with an MDM classifier. However, covariance matrices alone already contain much information. Also, there seems to be some similarity between covariance matrices and the level 2 signatures in terms of classification accuracy. Both encode pair-wise relationships between channels, the former being non directed while the latter being directed. If there is a way to add some structure and the Riemannian metric into level 2 signature, the results might be further improved. This is the motivation of Section 3.3.2 on the *lead matrix*. The next section is devoted to the use of path signature on frequency bands.

# channels	Accuracy
2	51.6(14.0)
22	59.1(13.2)

Table 3.2.3: Intra-subject classification accuracy with different numbers of channels. The features are path signatures of spectral power time series in the μ and β bands.

3.2.2 On spectral power

Inspired by traditional successful methods, we also considered using path signatures over the features proven to be discriminatory. The most classical feature used in EEG-based BCI is the spectral power. As seen in Section 1.1, motor imagery takes place in the sensorimotor cortex and is characterised by the mu-beta event-related (de)-synchronization [120], so we started by looking at the signature of the spectral power in the μ and β bands for channels C3 and C4 around the sensorimotor cortex at the left and right hemispheres. A sliding window of $0.2\,s$ goes over the entire trial of $2\,s$ to get the spectral power in each frequency band. Following the results of Section 3.2.1, the path signature up to level 2 of the 4-dimensional path (for two electrodes and two bands) is computed and classified by a linear SVM classifier. The 10-fold cross validation score of intra-subject classification is 51.6%, almost equal to chance level. It seems that the information provided by only 2 channels is too limited. The same method was then tested using all channels and this gives a classification accuracy of 58.6% (see Tab. 3.2.3).

3.2.3 On covariance matrices

The current gold standard classification method for EEG-based MI-BCIs is the Riemannian classifier. As discussed in the last paragraph in Section 3.2.1, the covariance matrices

alone already capture some information about the mental states. We wondered if the path of covariance matrices of the EEG signal inside a sliding window over one trial would also be discriminatory. To respect the intrinsic metric of SPD manifolds, two methods were tested: i) one is to project the covariance matrices onto the tangent space of a reference matrix, and compute the path signature in the tangent space; ii) the other one is based on the work of Darrick Lee and Robert Ghrist[95], embedding the SPD matrices into $GL_n(\mathbb{R})$ and compute the generalized path signature.

In the first method, the reference matrix can be simply the identity matrix or the mean matrix that minimizes the distance to all the covariance matrices where the metric used can be either Euclidean or Riemannian. Tab. 3.2.4 shows the results using identity, Euclidean and Riemannian mean as reference matrix respectively. There is almost no difference using different reference matrices and the results are all around 55%.

Identity	Euclidean mean	Riemannian mean
54.1(15.1)	55.4(15.6)	55.4(15.7)

Table 3.2.4: Intra-subject 10-fold cross validation score of the first method: projecting covariance matrices onto the tangent space of the reference matrix, then compute the Euclidean path signature (level 2) and classify them using a linear SVM classifier.

Deformation might be introduced during the projection onto the tangent space when the covariance matrices are scattered far away on the manifold, so it is of interest to consider computing the path signature directly on the SPD manifold. Lee & Ghrist[95] paved the way by generalizing path signature to Lie group valued time series and also provided a Julia package for implementation. With the same preprocessing (8-30 band-pass filter, 0.2 s sliding window with 100 time steps) and the same classifier, we got the classification accuracies shown in Tab. 3.2.5. The result is slightly better at level 2, but remains far inferior to

level 1	level 2	level 3	level 4
55.1(9.8)	57.0(13.7)	55.7(12.6)	55.1(12.7)

Table 3.2.5: Intra-subject 10-fold cross validation score of the second method: computing generalized path signature in SPD manifold at several truncation levels and classify them using a linear SVM classifier.

the results obtained using EEG time series directly (cf. Tab. 3.2.1).

3.3 Lead matrix

One possible explanation of the advantage of level 2 signatures may be the lead-lag relationship that they encode. If we go back to the definition of path signature at level 2, and we assume that all the paths are of class C^1 (i.e. their derivatives are continuous) and start from 0 (i.e. $X^i_a = 0$), then $S(X)^{1,2}_{a,b} = \int_a^b S(X)^1_{a,t}\, dX^2_t = \int_a^b X^1_t (X^2)'_t\, dt$, which can be interpreted as a measure of correlation between X^1 and $(X^2)'$, so how much X^1 leads X^2. An idealized example, i.e. a situation that would maximize the Pearson's correlation coefficient between X^1 and $(X^2)'$, would be $X^1_t = \cos(t)$ and $X^2_t = \sin(t)$ since we have a perfect alignment $X^1_t = (X^2)'_t$.

Based on level 2 signature, the *lead matrix* L is defined as:

$$L := \frac{1}{2}(S - S^\intercal)$$

where S refers to the square matrix formed by level 2 signatures $S_{i,j} = S(X)^{i,j}$. Note that the elements are exactly the signed area. The lead-lag relationship that lies in the lead matrix can be decoded through cyclicity analysis as detailed in Section 3.3.1 and a new way of combining the path signature and the Riemannian classifier is introduced in Section 3.3.2.

3.3.1 Cyclicity analysis

Cyclicity is a notion related to periodicity but without a fixed period. The underlying state space of a cyclic signal is a circle, so a cyclic signal can be turned periodic by time reparametrization. Examples of cyclic but not periodic signal could be the heartbeat signal or spikes of neurons. The cyclic order refers to the temporal ordering of cyclic signals, for example one firing neuron triggers another so that the signal of the former leads that of the latter. Baryshnikov and Schlafly showed that given a multidimensional path, one can recover the cyclic order under the hypothesis that all the components of the path are equivalent (up to time reparametrization) processes[13].

Consider the most simple case where the path components are created by shifting and rescaling a periodic function φ: $X_t^i = s_i\varphi(t-\alpha_i), i \in \{1,\dots,n\}$. We assume that φ is periodic since cyclicity and periodicity are equivalent up to time reparametrization and the path signature is invariant under time reparametrization. To simplify the computation, further assume that φ is just a sin wave, i.e. $\varphi(t) = \sin t$.

The goal is to recover the cyclic order of the components, or equivalently the system of offsets α_i from the lead matrix. The lead matrix over one period T (with length 2π) can be computed explicitly by noting that $\Delta X = 0$ on T. Its i,j-th element is given by:

$$\begin{aligned} L_{ij} &= \frac{1}{2}\int_T X_t^i (X^j)_t' - (X^i)_t' X_t^j \, dt \\ &= \frac{s_i s_j}{2}\int_T \sin(t-\alpha_i)\cos(t-\alpha_j) - \cos(t-\alpha_i)\sin(t-\alpha_j)\, dt \\ &\propto \sin(\alpha_i - \alpha_j). \end{aligned}$$

To let it simple, suppose that $s_i = i\pi^{-\frac{1}{2}}$. Therefore, $L = xy^\mathsf{T} - yx^\mathsf{T}$, with $x =$

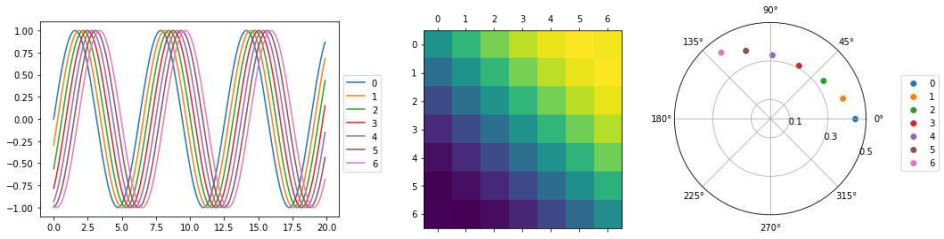


Figure 3.3.1: The signals are $\sin(t - \alpha_i)$ shown at the left. In the middle is the lead matrix in which yellow corresponds to positive values and blue to negative values. If L_{ij} is positive it means signal i is leading signal j. At the right is the first eigenvector of the lead matrix. We see a perfect recovery of the cyclic order from the phase of the elements of the first pair of eigenvectors.

$(\sin\alpha_1, \ldots, \sin\alpha_n)^\mathsf{T}$ and $y = (\cos\alpha_1, \ldots, \cos\alpha_n)^\mathsf{T}$. If x and y are linearly independent, the eigenvectors of L are $v_1 = -e^{-i\theta}\frac{x}{\|x\|} + \frac{y}{\|y\|}$ and $v_2 = \bar{v_1}$, where θ is the angle between x and y, i.e. $x^\mathsf{T}y = \|x\|\|y\|\cos\theta$. The i^{th} component of v_1 is a complex number, that is a linear transformation of $e^{i\alpha_i}$ in the complex plane described by the matrix $\begin{pmatrix} \frac{1}{\|y\|} & -\frac{\cos\theta}{\|x\|} \\ 0 & -\frac{\sin\theta}{\|x\|} \end{pmatrix}$. In this way, we know that the arguments of the complex components of v_1 (or v_2) have the same ordering as the system of offsets α_i.

Therefore, the cyclic order can be completely recovered from the phase of the eigenvector. Fig. 3.3.1 shows a numerical illustration of this example.

In general, if φ has more than one sin wave, there will be the same number of non zero conjugate pairs of eigenvalues as the number of sin waves. The cyclic order of each set of sin waves can still be recovered from eigenvectors. The magnitude of the eigenvalues indicates the magnitude of the sin wave and the magnitude of the elements in each eigenvector indicates the correlation with the corresponding sin wave.

Fig. 3.3.2 shows an example with 2 sets of sin waves. There are 2 conjugate pairs of eigenvalues and 2 blocks in the lead matrix. With the eigenvectors corresponding to the 2 non

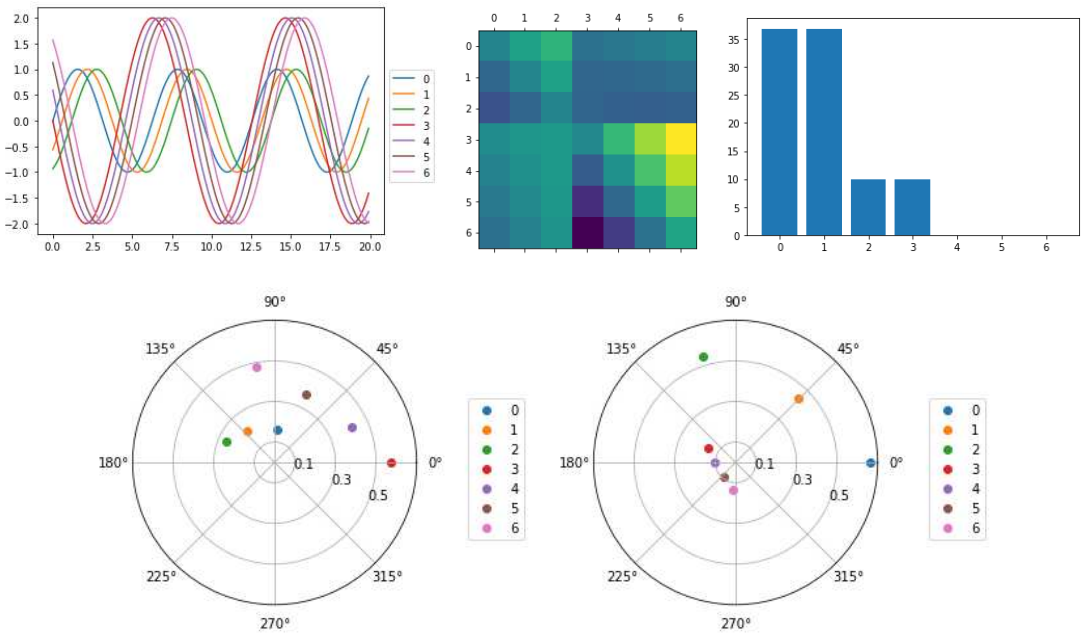


Figure 3.3.2: First row: on the left is the plot of 2 sets of sin waves; in the middle is the corresponding lead matrix; on the right is the absolute value of the eigenvalues of the lead matrix. Second row: the 2 eigenvectors corresponding to the 2 pairs of non zero eigenvalues.

zero eigenvalues, the cyclic order is again completely recovered.

Based on this analysis, we computed the eigenvalues and eigenvectors of the lead matrix. Fig. 3.3.3 showed that the absolute value of the eigenvalues averaged across trials. The first 2 conjugate pairs of eigenvalues dominate, which guarantees a good power of explicability of the corresponding eigenvectors.

Fig. 3.3.4 shows the number of times that one channel appeared in the top 10 largest elements of the first 2 eigenvectors corresponding to the 2 largest pairs of eigenvalues. The color of bars indicates left/right MI. For the first eigenvector, the channel POz (EEG-16) from the parieto-occipital region had the largest counts. This might be due to the fact that during a task, brain signals are dominated by visual related signal processed in the occipital

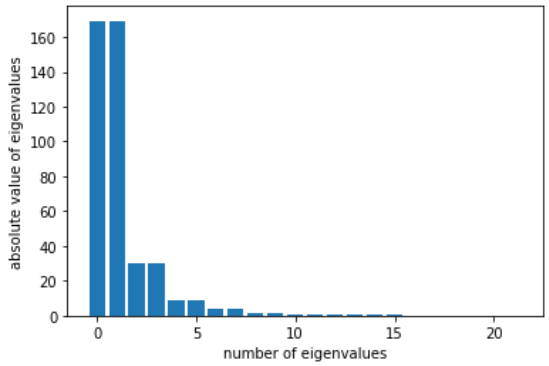


Figure 3.3.3: Absolute value of the eigenvalues of the lead matrix averaged across trials.

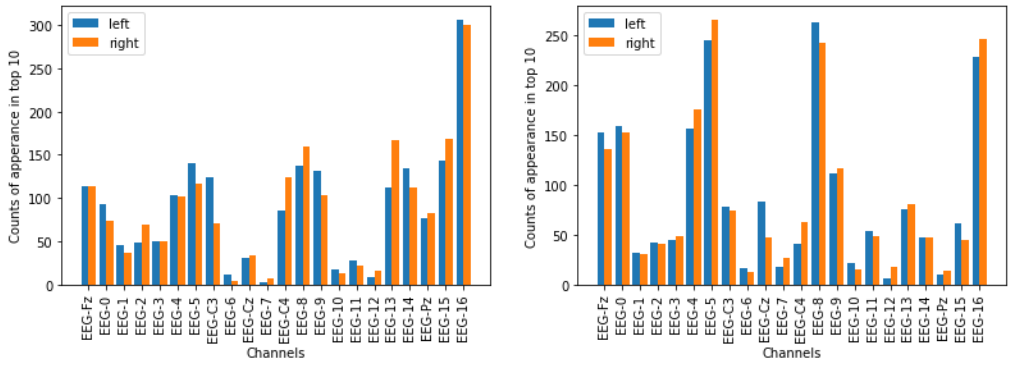


Figure 3.3.4: Frequency that a channel appears in the top 10 largest elements of the first (a) and second (b) leading eigenvectors of lead matrix in left/right MI conditions separately (over all subjects).

lobe and attentional networks related signals from the parietal lobe. Note that there was little difference between left/right conditions. For the second eigenvector, there was a large contribution from channels located above the sensorimotor cortex (e.g. EEG-5, EEG-8) and a clearer difference between left/right conditions.

3.3.2 Riemann-signature classifier

Since the path signature and the flattened covariance matrix have similar discriminative power when used as feature vectors as seen in Section 3.2.1, the lead matrix is also expected to have comparable results with covariance matrices applied on Riemannian classifier. The missing piece to achieve this is to make an SPD matrix from the lead matrix. It is clear that the lead matrix L is skew-symmetric, i.e. $L^\intercal = -L$. We observe that the matrix defined by:

$$A := -L^2 = L^\intercal L$$

is symmetric positive semi-definite. In fact, let v be an arbitrary vector, we have $v^\intercal A v = v^\intercal L^\intercal L v = \|Lv\|^2 \geq 0$.

The matrix A can be further turned into an SPD matrix by adding a small value on its diagonal. Now that we have the SPD matrix $A+\varepsilon I$, we can use them as features and leverage the advantages of the Tangent Space Riemannian classifiers. This idea is tested on a number of BCI datasets, both in intra- and inter-subject classification to evaluate the usefulness of signature features. The pipeline and results are summarized in Fig. 3.3.5 and Tab. 3.3.1 (ε was set to 0.001).

Although the signature-based method falls behind traditional Riemannian method for intra-subject classification, the difference is not big and it shows better robustness on inter-

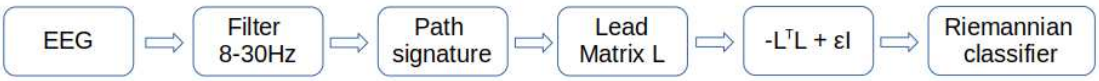


Figure 3.3.5: Pipeline of the Riemann-signature classifier.

			Signature		Covariance	
	#channels	#classes	intra	inter	intra	inter
BCI Competition IV 2a	22	2	71.4(18.1)	66.1(11.8)	81.1(16.6)	69.2(15.9)
Physionet MI-BCI	64	2	60.1(23.9)	47.0(11.0)	63.8(24.2)	46.2(14.8)
Passive BCI competition	61	3	88.9(10.5)	41.4(7.4)	90.9(9.3)	42.0(5.6)
BCI Competition IV 2a	22	2	71.4(18.1)	66.1(11.8)	81.1(16.6)	69.2(15.9)
Physionet MI-BCI	22	2	58.2(24.5)	51.6(12.4)	64.1(24.7)	51.5(16.0)
Passive BCI competition	21	3	58.0(17.7)	39.7(6.1)	71.2(15.9)	36.9(5.4)

Table 3.3.1: Intra- and inter-subject cross validation accuracy (standard deviation between parentheses) using Riemannian tangent space classifier on signature based SPD matrices and covariance matrices respectively on open BCI datasets.

subject classification, especially on Physionet dataset with a much larger number of subjects and a poor quality of data. Not only is the mean accuracy better, the standard deviation is also smaller: 11% for signature based method and 14.8% for traditional Riemannian method.

Fig. 3.3.6 shows more clearly the advantage of the signature based method for subjects with low classification accuracy. If we divide subjects into "good" and "bad" subjects according to their score under traditional Riemannian classifier with a 50% threshold (with 64 channels), we were able to see the clear difference between the two methods from Tab. 3.3.2. Note that 67 out 105 subjects are categorized as "bad".

By comparing the results using signature-based matrix and covariance matrix at the top half of Tab. 3.3.1, it seems that the number of channels also influences the result (e.g. the difference in intra-subject classification is 10% for BCI Competition IV 2a dataset with 22

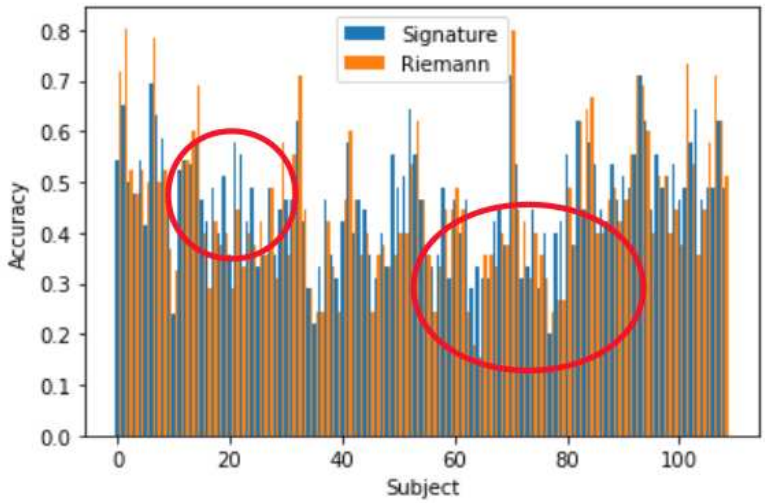


Figure 3.3.6: Inter-subject classification accuracy for each subject for the Physionet dataset, with leave one out cross validation. A Riemannian classifier is applied on signature-based SPD matrix as described in Sec. 3.3.2 and on covariance matrices respectively.

	Signature	Covariance
#channel=64	42.7(10.2)	37.2(8.5)
#channel=22	47.7(11.6)	44.7(12.6)

Table 3.3.2: Mean accuracy for "bad" subjects (defined by having an accuracy below 50% with Riemannian classifier applied on covariance matrices for the Physionet MI dataset).

channels compared to 2-3% for the other two datasets both with more than 60 channels). The signature-based method tends to be better to cope with a higher number of channels. To validate this conjecture, we selected the same 22 channels for the Physionet MI-BCI dataset and the passive BCI competition dataset (where the Cz channel is missing, so only 21 channels selected) as in the BCI competition IV 2a dataset, and the results are shown at the bottom part of Tab. 3.3.1. Comparing the upper and lower half of the table, a dramatic drop of performance is observed for the signature-based method, especially on the passive BCI competition dataset. The effects of changing montage are multiple and are very related to the paradigm of BCI. Since the characteristic brain activities are different for motor imagery task and mental workload task, using a montage with less channels might have different consequences for data with different paradigms. But in general, the Riemannian-signature classifier reacted more strongly than the traditional Riemannian classifier to the reduction of number of channels in intra-subject classification but remains robust in inter-subject classification.

3.4 Conclusion

In this chapter, we investigated the use of path signature for EEG-based BCI. The initial motivation was taking advantage of the translation invariance and time reparametrization invariance of path signature to compensate the distribution shift and the different time scales (e.g. fast/slow in performing motor imagery task) across sessions and subjects. This is in some degree demonstrated by the better results of the Riemann-signature classifier for inter-subject classification compared to the traditional Riemannian classifier shown in Tab. 3.3.1 and Fig. 3.3.6.

The lead-lag relationship between channels was investigated through the cyclicity analysis. Even though lead-lag relationships in EEG signals could be established by estimating phase differences via a traditional time-frequency analysis, the power of signature is that it does not rely on the hypothesis on periodicity which is more realistic. Moreover, there is no need to first filter the signal into a narrow frequency band to get an interpretable phase. The lead matrix contains the phase information of oscillations of different frequencies. Moreover, there are evidences that the lead-lag relationship, i.e. the ordering information, does encode information in the brain. For example, the neural sequence (the sequence of neuronal activity in which neurons are transiently active during task trials with different neurons active at different parts of the trial) has been used as the representation for a sensorimotor task of a mouse navigating a T-maze[44]. Zhou et al. argued that the neural sequences represent an ideal and flexible dynamical regime for the brain to read out time information[190]. However the link between mental states and the lead-lag relationship between channels needs to be further investigated and validated. The lead-lag relationship should also be interpreted with caveat since the implication of lead/lag based on the sign of signed area assumes that the signals are consubstantial. This is not the case with EEG signal which might have opposite trends in different channels.

Three ways of applying path signature were tested: directly on EEG multivariate time series, on spectral power, and on covariance matrices. The direct way worked best and indeed this is the most explicable. The notion of order that is captured by path signature is more obscure for extracted features like the elements in the covariance matrices. We also found that path signature performs better when we have more channels, longer time windows and more time steps. In line with the network point of view of modern neuroscience and

the success of Riemannian classifier, it is again confirmed that it is the relationship between EEG channels rather than the signal of a certain channel that is important to decode mental states. We found a way to integrate the signature into the Riemannian classifier and this Riemann-signature classifier showed better robustness than the Riemannian classifier for inter-subject classification.

In the future, some general techniques to improve the performance of the signature method [116] could be tried such as the lead-lag augmentation of the times series. It would also be interesting to investigate the link between the lead-lad relationship and other connectivity measures.

This work will will be promoted by a journal article currently being written.

Path signature for EEG-based BCI:

- Given a multidimensional time series, the path signature generates a feature vector that is invariant under time reparametrization and translation.
- Path signature is fast to compute and easy to be used online thanks to the additivity of the integrals.
- Level 2 signature captures lead-lag relationships between multi-channel EEG signals.
- The Riemannian classifier applied on the signature-based SPD matrices outperforms baseline methods for inter-subject classification.

Allegro con fuoco

4

Rondo: Topological data analysis

Circular structure is often employed in music such as the recurrence of the main theme in the main key in a rondo. Life is also composed of longer or smaller cycles. However this does not mean that everything is reduced to zero because these cycles are not contractible due to the voids that they delineate. The interesting things lie exactly in the voids, precisely

described in the language of homology.

4.1 Motivation

A dynamical system studies the evolution of a system over time in a state space (also called phase space, composed of all the possible values for the tuple of variables describing the system) according to the rules in the form of differential equations. The central problem is to determine how the system behaves in the long run from its behaviors in short periods of time described by the differential equations. Most natural phenomena are dissipative systems (i.e. with energy loss in contrast to the conservative systems). After a sufficiently long time, the state of the system will converge to a subspace of the state space, and this subspace is called the attractor. The evolution of a dynamical system is characterized by its attractor. The attractor of a linear dissipative system is simply a point. For nonlinear dissipative systems, the types of attractor are much more interesting; they can be a limit cycle, a manifold or a subset with fractal structure (called strange attractor).

Dynamical systems are suitable tools to study the macroscopic spatio-temporal dynamics of the electric field of the brain captured by EEG at a high temporal resolution. The recovery of the brain dynamics from one EEG time series is guaranteed by Takens' theorem [162] which states that through a time delay embedding with large enough embedding dimension (see Section 4.4 for more details) the attractor of a dynamical system can be reconstructed up to diffeomorphism from a single observation of the system. The EEG dynamics has been of great interest for neuroscientists to understand the links between brain functions and the large scale synchronization of neuronal networks [18]. The measures of the attractor (see paragraph below) are shown to be correlated with mental states, both

in normal conditions (e.g. sleep, emotion, cognitive tasks) and abnormal conditions (e.g. epilepsy, schizophrenia, coma, dementia)[158].

Classical measures to characterize the attractors are the correlation dimension, the Lyapunov exponents and the entropy. The correlation dimension[65] is the most frequently used measure. It measures the spatial extensiveness of the attractor, thus indicating the complexity of the dynamics. Lyapunov exponents and entropy are closely related, the entropy being equal to the sum of all positive Lyapunov exponents. They measure the stability of the system by measuring the growth rate of nearby trajectories.

There has been many new measures proposed to adapt to noisy non-stationary time series in the real world[158]. However the computation of both the classic and the new measures are based on the metric defined on phase space while the reconstruction of the attractor is just up to diffeomorphism (i.e. the metric might change) according to Takens' theorem. Topological data analysis, a new field of applied mathematics that studies the shape of the data using topology, allows the extraction of topological information of the attractor, which is independent of the choice of metric and is preserved by diffeomorphism.

The variability of the EEG signal (cf. Section 1.3) concerns the static measures of the signal. In contrast, dynamical system deals with the change of the system over time. Hence it might be an appropriate point of view to overcome the variability problem of EEG. In the following, first we introduce in Section 4.2 the main tool of this chapter: topological data analysis. Then, its application for EEG signal processing is reviewed in Section 4.4. Finally, some examples and tests on BCI competition IV 2a dataset are presented in Section 4.5.

4.2 Topological data analysis

As its name implies, topological data analysis (TDA) is the field that employs topology to analyse data. So the first step to understand TDA is to understand the key ideas of topology. Then the main technique in TDA is detailed in Section 4.2.3. The interested readers are referred to the books of Hatcher[69] and of Ghrist[53] for a mathematical introduction to the algebraic topology, and the tutorials of TDA for applications in data science[27] and neuroscience[157]. All the definitions are gathered in Section 4.2.2 to facilitate reading.

4.2.1 Topology and homology in a nutshell

Topology is the mathematical branch which studies the properties that are preserved by continuous deformation, such as scaling, twisting but not tearing. It could be viewed as "geometry without metric". Distance is of no importance in topology, instead the whole theory is based on the notion of "closeness". It is a popular joke among mathematicians that for a topologist a coffee mug is equal to a donut since one can be continually deformed into the other without tearing (Fig. 4.2.1).

More precisely, what we call topological properties are those invariant under homeomorphism (continuous map whose inverse is also continuous). However, homeomorphism is often too strict. A looser but useful notion is homotopy (see Definition 1). It is not only useful in the theory, but also in practice since some noise in the real world data can also be viewed in a homotopic (and non homeomorphic) way[56].

Based on the notion of homotopy, homotopy groups can be defined which distinguish different topological spaces. But they are difficult to compute in general. An alternative notion is homology (see Definition 5), which is based on simplicial complexes (see Definition



Figure 4.2.1: A coffee mug is equivalent to a donut in a topologist's eyes. Image made by Henry Segerman (http://www.segerman.org/).

2) and can be computed effectively using the "divide and conquer" strategy.

Intuitively, the k-dimensional homology catches k-dimensional "holes", i.e. independent cycles that are not filled. Homology is a homotopy invariant and the dimensions of the homology group, i.e. Betti numbers, are among the first topological statistics applied in real world applications[56].

4.2.2 Definitions

Definition 1. *Two functions $f, g : X \to Y$ are said to be **homotopic** if there exists a continuous function $H : X \times [0,1] \to Y$ such that $H(\cdot, 0) = f$ and $H(\cdot, 1) = g$. Two topological spaces X and Y are **homotopic** if there exist continuous functions $f : X \to Y$ and $g : Y \to X$ such that $f \circ g$ and $g \circ f$ are homotopic to id_Y and id_X respectively.*

Definition 2. *A k-**simplex**, noted as $\Delta_k = [v_0, \ldots, v_k]$, is the convex hull of a set of $k+1$ linearly independent points. A **simplicial complex** is a collection of simplices satisfying the*

following conditions: every subset and their intersections are also simplices in the collection.

Definition 3. *The space of k-**chains** ($k \in \mathbb{Z}_+$) on a simplicial complex K, noted as $C_k(K)$, is the set of the formal finite sums of the k-simplices in K. It is a vector space with the addition and the scalar multiplication defined in the natural way.*

Definition 4. *The **boundary** map takes a k-simplex $\Delta_k = [v_0, \ldots, v_k]$ to the $(k-1)$-chain: $\partial_k \Delta_k = \sum_i (-1)^i [v_0, \ldots, \hat{v}_i, \ldots, v_k]$ where $\hat{v}_i$ means omitting v_i. Since k-simplices form a basis of $C_k(K)$, the boundary map ∂_k can be extended as a linear operator on k-chains in $C_k(K)$. The kernel of ∂_k is called the k-**cycles** and the image of ∂_{k+1} is called the k-**boundaries**. They are noted as $Z_k(K)$ and $B_k(K)$ respectively.*

Example 1. *Take a 2-simplex $\Delta_2 = [v_0, v_1, v_2]$ for example: $\partial_2 \Delta_2 = [v_1, v_2] - [v_0, v_2] + [v_0, v_1]$. Its boundary is in fact the loop formed by its edges.*

Remark 1. *A simple yet important property is that the boundary of a boundary is always zero: $\partial_k \circ \partial_{k+1} = 0, \forall k \geq 0$. In other words, k-boundaries are k-cycles, i.e. $\operatorname{im} \partial_{k+1} \subseteq \ker \partial_k$. Homology is defined based on this property.*

Example 2. *Take the same example above, $\partial_1 \circ \partial_2 \, \Delta_2 = \partial_1([v_1, v_2] - [v_0, v_2] + [v_0, v_1]) = v_2 - v_1 - v_2 + v_0 + v_1 - v_0 = 0$.*

Definition 5. *The k-th **(simplicial) homology group** of K is defined as the quotient vector space $H_k(K) := Z_k(K)/B_k(K)$. Two k-cycles differ by a k-boundary are **homologous** and considered as the same element in $H_k(K)$. Hence the elements of $H_k(K)$ are the independent k-cycles not filled ("k-dim holes"). The dimension of $H_k(K)$ is called the k-th **Betti number**, and corresponds to the number of independent "k-dim holes".*

4.2.3 Persistent homology

What makes TDA powerful and particularly suitable to capture hierarchical features is the persistent homology, a method that extracts persistent topological features across scales. The key notion is the filtration, which is a nested family of subcomplexes indexed by a parameter. The homology of these subcomplexes evolves as the parameter grows, giving rise to the barcode or persistence diagram as a description of persistence of connected components (described with components' birth and death) by 0-dimensional homology, and of multidimensional holes by higher dimensional homology.

Fig. 4.3.1 illustrates two kinds of filtration that are mostly used. Fig. 4.3.1 (A) is an example of sublevel filtration. Let f be a map from a simplicial complex S to real numbers $\mathbb{R}$. Then the *sublevel filtration* consists of the family of subcomplexes $S_l = \{\Delta \in S : f(\Delta) < l\}$. In this example with a function of one variable, the only non trivial topological information is carried by a 0-dimensional homology since we have only segments and no loops. We see that, as the parameter l increases, new components emerge, marked as birth, and also a component merges with another one, leading to its death (by convention the elder rule is applied that keeps the older one). The parameters corresponding to birth and death form a family of intervals. It can be shown that the family of intervals is unique up to reordering[26] and can be used as a topological feature named *barcode*. We could also mark on the plan $\mathbb{R}^2$ the point at which the x and y coordinates correspond to birth and death of a feature respectively. These points, together with the diagonal, are called *persistence diagram*.

Fig. 4.3.1 (B) is an example of Rips filtration. Given a set of points X and a positive number α, the *Rips complex* $\mathrm{Rips}_\alpha(X)$ is the simplicial complex including all simplices in which the distance between any two of their vertices is smaller than α. *Rips filtration* consists of

the family of Rips complexes $\{\text{Rips}_\alpha(X)\}_\alpha$ indexed by α. In the example in $\mathbb{R}^2$, we could imagine balls with increasing radius around the initial point cloud. As the balls become bigger, they merge with each other, leading to the death of certain connected components but also the birth of some loops (only one for the example in Fig. 4.3.1 (B)). This also leads to a barcode or persistence diagram. The larger the scale of the loop is, the more persistent it is in the barcode. In most cases, long bars correspond to significant features while short bars correspond to noise. As α goes from 0 to ∞, the topology of each Rips complex goes from one trivial case, disjoint unions of points, to another trivial case, all points merged to one connected component. The persistence homology records the evolution between the two extreme cases, in the middle of which interesting features come to light. So, unlike hard thresholding methods usually used in the construction of graphs, TDA preserves more information since it is based on the filtration, i.e. a hierarchical collection of simplicial complexes.

4.3 Topological features

A metric structure is still needed in order to measure differences between barcodes or persistence diagrams. The Bottleneck distance, or Wasserstein distance, measures the distance between persistence diagrams by pairing points in the two diagrams.

There are many other ways to extract topological features from persistence diagrams other than measuring distance in the barcode space. The simplest way might be extracting Betti numbers. Another simple method is persistent entropy[31] defined as the Shannon entropy of the persistence barcode thus giving a scalar description of the barcode. Additionally, persistence landscapes introduced by Bubenik[20] smartly embed persistence diagrams

into a Hilbert space where most machine learning algorithms can be applied, thus making them a popular choice. More detailed definition of these features can be found in Section 4.5.2.

4.4 Application

Even though TDA is a very versatile tool that can be applied on time series [154,169], gene expression [137,124], neural code [40], functional connectivity [148] and structural connectivity [139], we restrict our discussions here to its application on EEG processing. In order to retrieve and evaluate in a comprehensive manner all research works related to TDA and EEG processing, a systematic search was conducted as described here after. Hence, using ((topological data analysis) OR (TDA)) AND ((electroencephalography) OR (EEG)) as search terms, we found 70 publications on PubMed and 85 publications on Web of Science (in 2021). Then we examined the title and abstract of each non-redundant publication and excluded those concerning recording methods other than EEG (e.g. MEG, fMRI) or using graph theory analysis instead of TDA as tools. The remaining publications are summarized in Table 4.4.1.

General pipeline From the papers that were collected, the general pipeline of the application of TDA on EEG signal processing consists of the preprocessing steps, the filtration (sublevel or Rips) to get a persistence diagram/barcode, the extraction of topological features, and further a comparison/classification step (see Fig. 4.4.1).

Besides standard preprocessing steps, e.g. band-pass filtering, downsampling and artifacts removal, few authors have chosen to transform data into another space. Independent

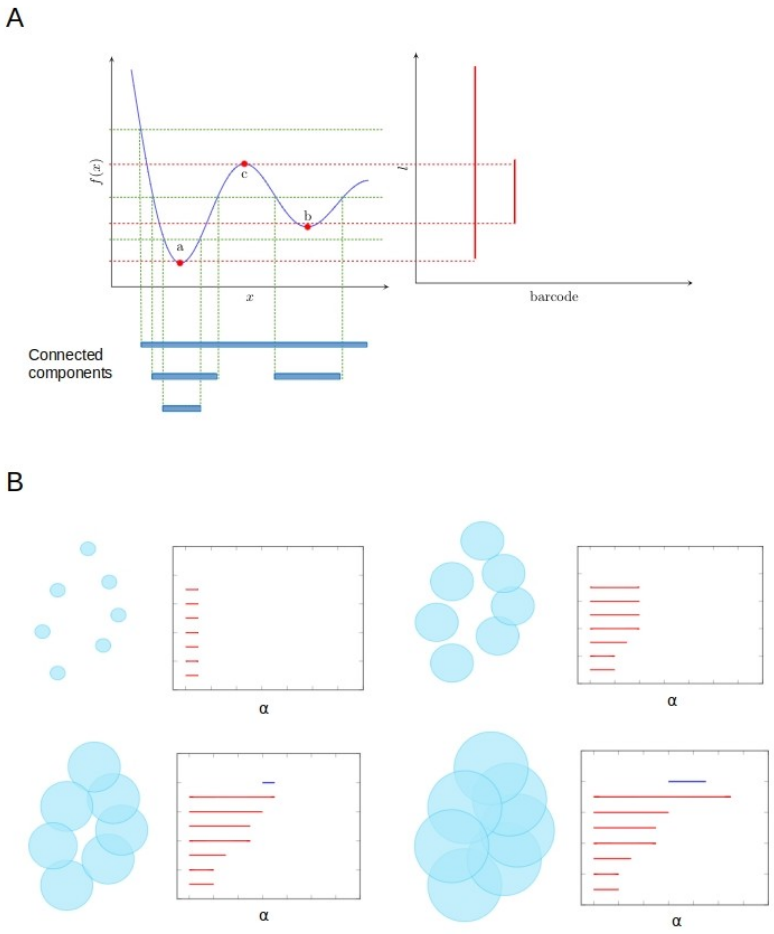


Figure 4.3.1: Examples of filtration and the associated barcode. (A) shows the sublevel filtration of function f(x). The first and second connected components appear at point a and b respectively and merge together at point c, giving rise to two bars: one from a till infinity and the other from b to c. (B) shows several stages of Rips filtration. At first, all points are isolated and we have the same number of bars and points. As balls grow larger, some balls start to merge together, giving death to certain bars. Then one loop appears, marked as a 1-dimensional feature corresponding to the blue line in the barcode. Finally the loop is filled so the blue bar stops and every point merges into one component living forever.

Article	Domain	Transform	Filtration	Features	Classifier	Dataset
Altındiş et al. [2]	MI-BCI	-	Rips filtration (time delay embedding)	Persistence diagram	-	Graz dataset
Bischof & Bunch [16]	Eyes-open/eyes-closed	-	Rips filtration (time delay embedding)	Betti-numbers	CNN	Bonn dataset
Yamanashi et al. [185]	Delirium	-	Rips filtration (time delay embedding)	Area of the 1-dimensional Betti curve	-	Private dataset
Majumder et al. [105]	Autism	-	Sublevel filtration	Persistent entropy	SVM	Private dataset
Wang et al. [174]	Aphasia	ICA	Sublevel filtration	Persistence landscape	-	Private dataset
Wang et al. [173]	Aphasia	ICA	Sublevel filtration	Persistence landscape	-	Private dataset
Ibáñez-Marcelo et al. [79]	Hypnotizability	ICA	Rips filtration (spatial embedding)	Persistence diagram	-	Private dataset
Nasrin et al. [119]	Brain states	-	Sublevel filtration	Persistence diagram	Bayesian learning	US Army APG simulation dataset
Wang et al. [176]	Seizure localization	Fourier transform	Sublevel filtration	Persistence landscape	-	Private dataset
Altindis et al. [1]	MI-BCI	-	Rips filtration (time delay embedding)	Betti numbers	kNN	Graz dataset
Piangerelli et al. [131]	Epileptic seizure	-	Sublevel filtration	Persistent entropy	Linear classifier	Physionet
Wang et al. [175]	Epilepsy	Fourier transform	Sublevel filtration	Persistence landscape	-	Private dataset

Table 4.4.1: Summary of applications of TDA to EEG analysis (ordered by publication date)

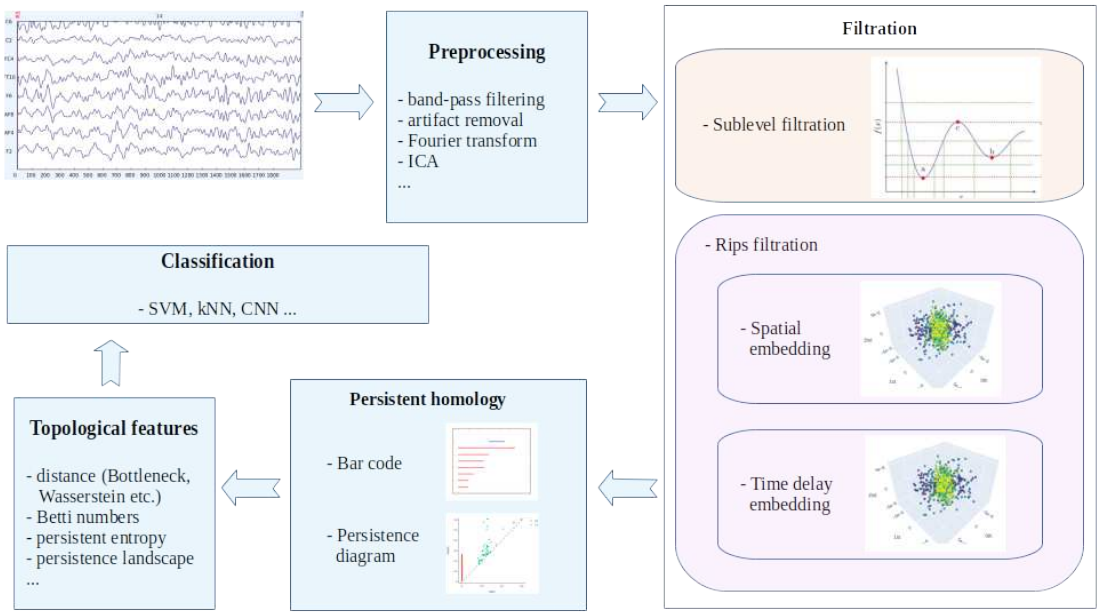


Figure 4.4.1: Pipeline of applying TDA to EEG signal processing.

component analysis (ICA) is the one mostly used with the purpose of getting source components, then followed by Fourier transform for the purpose of denoising data.

In the filtration step, three methods emerge: the first one applies sublevel filtration illustrated in Fig.4.3.1 (A) directly onto the EEG time series of each channel; the second one applies Rips filtration as illustrated in Fig.4.3.1 (B) onto the spatially embedded signal; and the third one onto the time delay embedded signal. Hence, both the second and third ones use Rips filtration. The difference is on the way of embedding signals. For the second group, data is embedded spatially so that each point in the state space represents a channel or a source component, and the distance is measured by connectivity measures such as Pearson correlation. For the third group, the data is embedded via time delay embedding which consists in keeping d equally spaced observations before current time $\{(x(t-(d-1)\tau), x(t-(d-2)\tau), \ldots, x(t-\tau), x(t))^{\intercal}\}_t$ given a time series $\{x(t)\}_t$. Most authors have

chosen the third one. One reason for this choice might be that the time delay embedding is well backed by Takens' theorem which gives the maximum embedding dimension needed in order to reconstruct the attractors (a set of points in the phase space that is invariant under dynamics and "attracts" neighbouring trajectories) up to diffeomorphism (so that the topological properties of the dynamical system are preserved).

A barcode or persistence diagram is constructed along with the filtration process as shown in Section 4.2.3. The most popular topological features for EEG processing (see Tab.4.4.1 column "Features") are persistence landscape, persistent entropy and Betti numbers. Finally, after the feature extraction step, a classifier might be applied or not depending on the application at hand (e.g. diagnostic, active brain-computer interface or mental state estimation via a passive brain-computer interface). The most used ones are linear classifiers due to the small size of clinical datasets. However, when the amount of data allows, deep learning methods show promising performance [16]. TDA can also be combined with other frameworks such as Bayesian networks [119].

Lacks and objective TDA is a young and dynamic field that attracts more and more attention. It has many advantages. Firstly, topological features are by nature robust and invariant to transformations such as translation, amplitude and frequency scaling [175]. Secondly, TDA can capture global and higher dimensional features where other methods such as graph theory fails. Thirdly, the information that TDA extracts is different and complementary to most machine learning methods.

Although there are theoretical results showing the robustness of persistence diagrams under perturbations [36,20], in practice the technical details of applying TDA for EEG signals need to be further investigated. Altındiş et al. [2] started in this direction by trying to find the

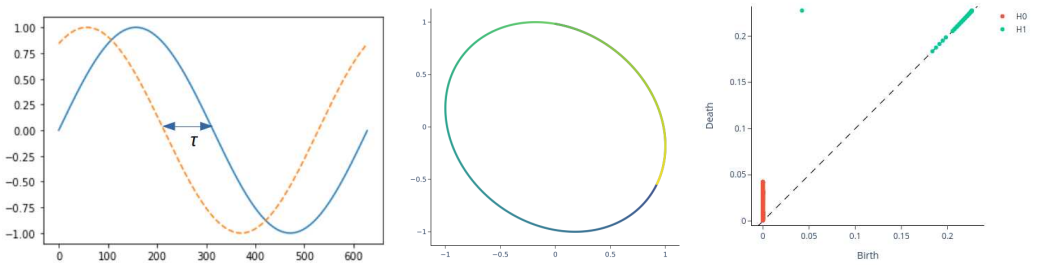
optimal embedding dimension, time delay and time window size using false nearest neighbor (FNN) test. The paired t-test showed that the significance level of extracted topological features was very sensitive to the choice of embedding parameters and hence it was important to use the optimal parameters.

The current domain of application is still quite restricted to clinical studies to improve the diagnostic of neurological diseases. However, TDA is also quite suitable for other non-clinical EEG related areas, e.g. EEG based brain-computer interface (BCI; Wolpaw et al. [179]) which allows the explicit control of machines or implicit mental state estimation using only EEG signals. Further, the datasets used in current publications are mostly private datasets, which if possible should be replaced by publicly available ones for increased reproducibility and comparison with other works.

With the limitations identified, the following studies focus on the application of TDA for EEG-based BCI with experiments on open dataset, with the main goal being to determine if TDA can provide a robust estimation across subjects. Attention is payed to the choice of parameters. Several variants of the basic pipeline will be tested with different numbers of channels, features, classifiers and frequency bands.

4.5 Experiments on EEG-based BCI

The domains of application of TDA on EEG signal processing remain largely clinic (cf. column "domain" in Tab. 4.4.1). More studies are needed to test its utility on EEG-based BCIs. We start by providing several toy examples to understand the pipeline in simple artificial cases. Then various methods are tested on the BCI competition IV 2a dataset described in Section 1.5. The time delay embedding, persistence diagram as well as its derived



(a) time delay embedding (b) attractor (c) persistence diagram

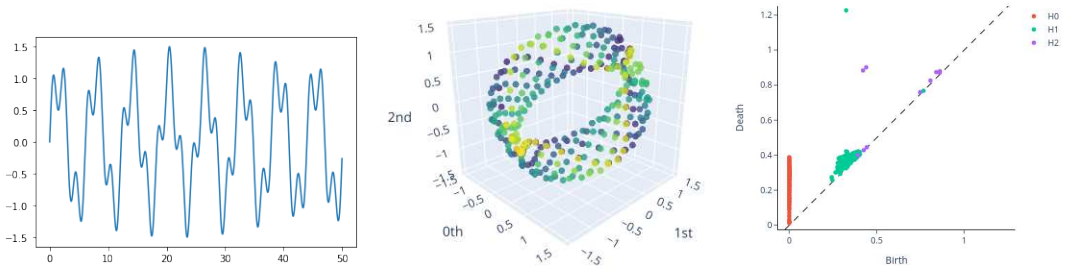
Figure 4.5.1: Time delay embedding (a), the attractor (b) and the persistence diagram (c) of a periodic signal. The color in the attractor figure indicates time.

features are computed using the giotto-tda toolbox in Python [165]. The classifers from the scikit-learn [122] toolbox are used with default parameters throughout all experiments below.

4.5.1 Toy example

Periodic signal Take the simplest case where $x = \sin(t)$ and the embedding dimension $n = 2$. In this setting, the time delay embedding consists in shifting x once by the time delay τ (Fig. 4.5.1 (a)) to get the 2D orbit $(x(t), x(t+\tau))$ in the phase space (Fig. 4.5.1 (b)). The orbit of a sine wave is an ellipse. In terms of persistence diagram, a significant loop is characterized by a point associated with a 1-dimensional persistent homology far from the diagonal (Fig. 4.5.1 (c)).

Quasiperiodic signal A quasiperiodic signal is a time series "almost periodic". An example could be the sum of two periodic functions with incommensurate frequencies, for example $x = \sin(t) + 0.5\sin(\pi t)$ (Fig. 4.5.2 (a)). Its orbit is a torus (Fig. 4.5.2 (b)). In terms of persistent diagram a torus shape point cloud is characterized by two points associated with 1-dimensional persistent homology and one point associated with 2-dimensional per-



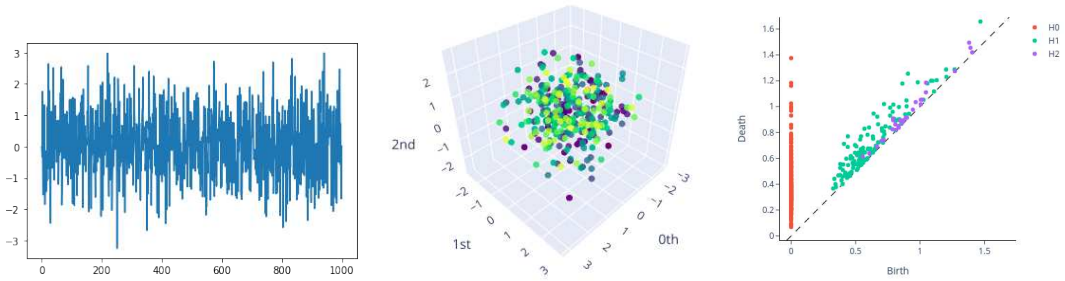
(a) signal (b) attractor (c) persistence diagram

Figure 4.5.2: A quasiperiodic signal (a), its attractor after a 3 dim. embedding (b) and the corresponding persistence diagram (c). The color in the attractor figure indicates time.

sistent homology since the homology group of torus is $H_k(\mathbb{T}^2) = \begin{cases} \mathbb{Z} & k = 0, \\ \mathbb{Z} \oplus \mathbb{Z} & k = 1, \\ \mathbb{Z} & k = 2, \\ 0 & \text{otherwise.} \end{cases}$.

In Fig. 4.5.2 (c) we only see one point associated with 1-dimensional persistent homology (in green) far from the diagonal because the torus is quite "thin" so that the small cycle is quickly filled in the filtration, which corresponds to the group of points near the diagonal.

White noise On the other hand, given a white noise drawn independently from a normal distribution with zero mean, the orbit has no clear structure (see Fig. 4.5.3 (b)) and the points in the persistent diagram are all close to the diagonal (see Fig. 4.5.3 (c)) i.e. without significant features.



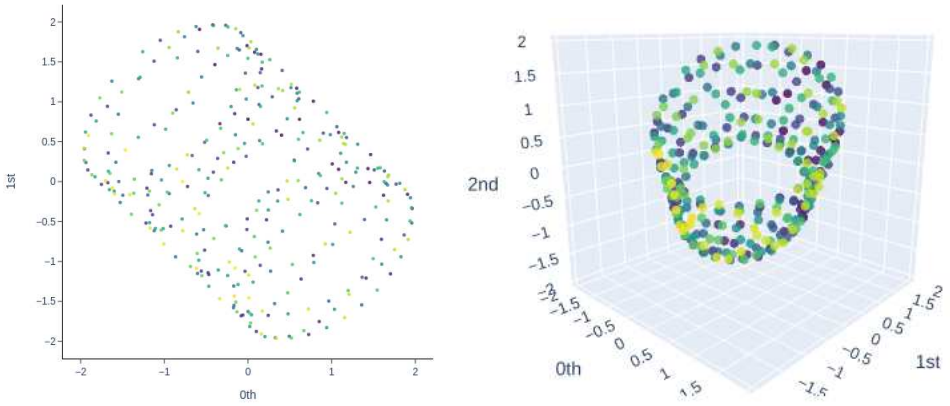
(a) signal (b) attractor (c) persistence diagram

Figure 4.5.3: A white noise drawn from a normal distribution (a), its attractor (b) and the corresponding persistence diagram (c). The color in the attractor figure indicates time.

4.5.2 Methods

In the same spirit as the toy examples, the idea is to study the "shape" of the attractor of the EEG signal using TDA. The first step is the time delay embedding of the signal, followed by the Rips filtration to get the persistence diagram. Then various types of features could be extracted from the persistent diagram and finally a classifier is applied to get the classification result. Although there are other possibilities in the filtration step as shown in Fig. 4.4.1, we have chosen the time delay embedding for its reliability to reconstruct the attractor of the EEG dynamics, while the spatial embedding is controversial and debated in the literature [136,128]. Sublevel filtration was not considered either, as it is quite sensible to noise whereas EEG is well known to be noisy.

Time delay embedding The time delay embedding is the first and a crucial step in the whole pipeline. Recall that the time delay embedding of a time series $\{x(t)\}_t$ is $\left\{\left(x(t-(d-1)\tau), x(t-(d-2)\tau), \ldots, x(t-\tau), x(t)\right)^\intercal\right\}_t$. Even though Takens' theorem guarantees an upper bound for the needed embedding dimension (twice of the dimension of the



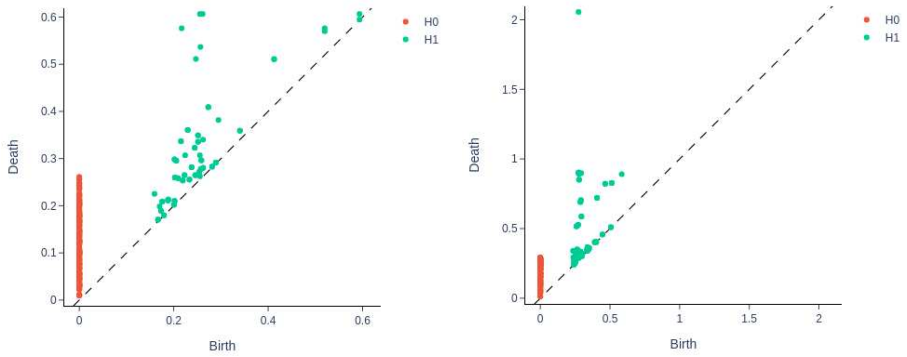
(a) dim = 2 (b) dim = 3

Figure 4.5.4: Effects of embedding dimension d on the shape of the attractor.

attractor) to recover the dynamics of the system from one observation, in practice it is not quite useful since the "true" dimension of the attractor is unknown. Note that both the choice of the embedding dimension d and of the time delay τ affect the persistent diagram and thus the features extracted from it.

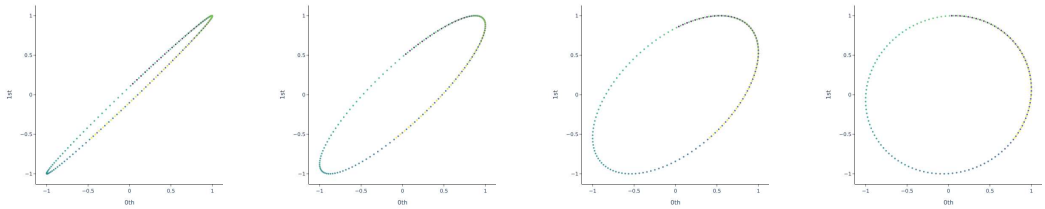
The effect of the embedding dimension is more straightforward. The lower dimensional embedding is simply a projection of the higher dimensional embedding. It is better to use the least possible embedding dimension for computational cost. However some critical features might be missed if the embedding dimension is not high enough. For example in Fig. 4.5.4 the cylinder structure is totally invisible with 2-dimensional embedding while being clear with 3-dimensional embedding. This is further confirmed by the associated persistence diagrams shown in Fig. 4.5.5 with a H_1 point far from the diagonal indicating a significant one dimensional loop.

To understand the effect of the time delay τ, we reuse the first example in Section 4.5.1, $x = \sin(t)$. Fig. 4.5.6 shows the influence of τ on the shape of the attractor. When τ is



(a) dim = 2 (b) dim = 3

Figure 4.5.5: Persistence diagram with different embedding dimension.

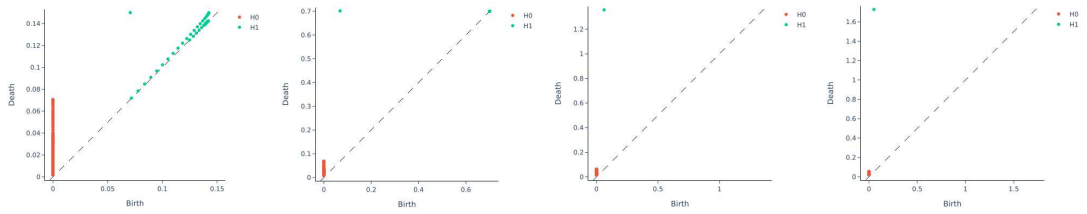


(a) $\tau = 10$ (b) $\tau = 50$ (c) $\tau = 100$ (d) $\tau = 150$

Figure 4.5.6: Effects of time delay τ on the shape of the attractor.

small (e.g. in (a)), the two dimensions are very correlated which results in a "flat" shape of the attractor. When τ is appropriate (e.g. in (d)) the loop is more remarkable and more easily detected by TDA as shown in Fig. 4.5.7 (d) where the H_1 point corresponding to the loop is much further from the diagonal (note that the scale for all the 4 subfigures is not the same, cf. labels on each axis) and there is no "false" H_1 points near the diagonal as in Fig. 4.5.7 (a).

In practice, there are some guidelines to choose the optimal parameters. False nearest neighbor[87] has been widely used to determine the optimal embedding dimension. The idea



(a) $\tau = 10$ (b) $\tau = 50$ (c) $\tau = 100$ (d) $\tau = 150$

Figure 4.5.7: Persistence diagram with different time delay τ.

is to find the lowest dimension that keeps the change of the distance to nearest neighbors under a threshold during the unfolding of the attractor successively into higher dimensions. To determine the optimal time delay the mutual information is used as an indicator. The τ that achieves minimum mutual information between $\{x(t)\}_t$ and $\{x(t-\tau)\}_t$ is chosen because most information will be added by adding the second delayed time series. Fig. 4.5.8 shows the mutual information with different time delay for the first example in Section 4.5.1. The attractor reconstructed with the τ achieving minimum mutual information has optimal shape (cf. Fig. 4.5.6 (d) with $\tau = 150$). In our study, the two parameters are computed with the above methods for randomly chosen trials of each subject and then the average is used as fixed parameters for practical purposes. An analysis of the estimation and the impact of these parameters is provided in Section 4.5.3.

Features Following Tab. 4.4.1 we picked the most popular features: Betti numbers, persistent entropy, persistence landscape and also the average persistence in each dimension. The formal definition and the intuition of each feature are detailed below.

Given a persistence diagram $D = \{(b_i, d_i)\}_{i \in I}$ where b_i and d_i correspond to the birth and death of homology classes, Betti numbers can be approximated by the number of

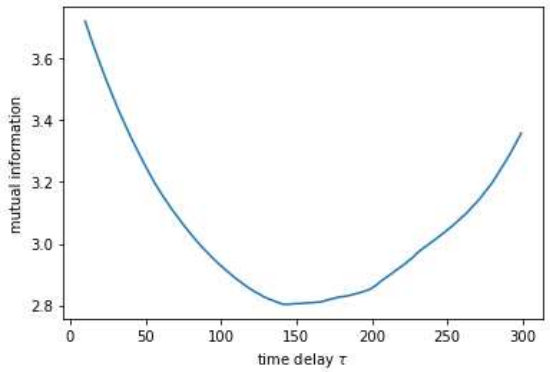


Figure 4.5.8: Mutual information between $\sin(t)$ and $\sin(t-\tau)$ with different time delay τ.

points off diagonal in each dimension. Persistence is the lifespan of each point, i.e. $d_i - b_i$ and the average persistence in each dimension is used as a feature vector. Persistent entropy is the Shannon entropy of the normalized persistence p_i, i.e. $-\sum_{i\in I} p_i \log p_i$ where $p_i = \frac{d_i - b_i}{\sum_{i\in I}(d_i - b_i)}$. Persistence landscape is the function $\lambda : \mathbb{N}\times\mathbb{R} \mapsto \bar{\mathbb{R}}$ defined by $\lambda(k,t)$ being the k-th largest value of the family $\{(\min\{t-b_i, d_i - t\})_+\}_{i\in I}$ where a_+ means $\max\{a, 0\}$. The L^p distance between the persistence landscapes of the given persistence diagram D and the trivial diagonal persistence diagram in each homology dimension is used as a feature.

Less formally, Betti numbers count the number of holes in each dimension. Persistence measures how long theses holes endure. Persistent entropy tells how much the persistence is ordered with the maximal entropy corresponding to the case where all the holes have the same lifespan. Persistence landscape maps the persistence diagram into the vector space of real-valued functions on $\mathbb{N}\times\mathbb{R}$ with its k-th layer $\lambda(k,t)$ being the k-th highest "tent" after rotating the persistence diagram by 45 degrees clockwise as shown in Fig. 4.5.9.

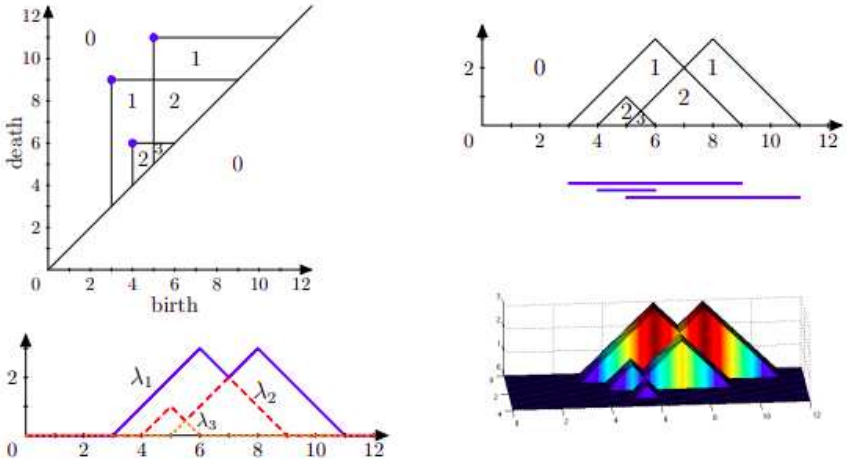


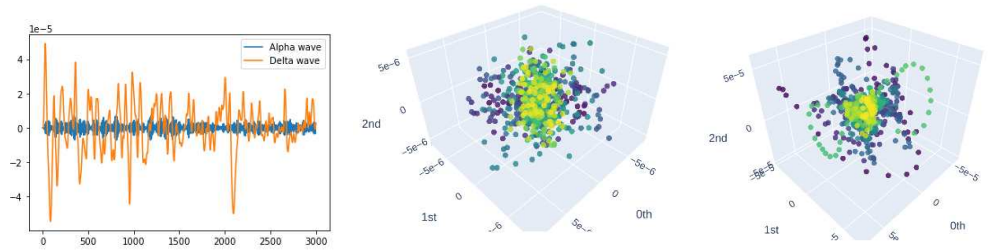
Figure 4.5.9: Persistence landscapes of a certain persistence diagram. (from Bubenik et al.[21]). The top row shows the rotation of the persistence diagram. The bottom row shows 2D and 3D view of λ_i, i.e. the i-th layer of the persistence landscape.

EEG PARTICULARITIES The previous pipeline addresses one dimensional time series while EEG records brain electrical signal at multiple sites. The natural way to tackle this is to concatenate features extracted from multiple channels.

Another particularity concerning the EEG signal is that it makes sense to study the dynamics of signals in different frequency bands. Fig. 4.5.10 shows the delta wave (1-4 Hz) and the alpha wave (8-13 Hz) filtered from one trial as well as their attractors. Separating different frequency bands allows to catch more clearly the different dynamics in different frequency bands.

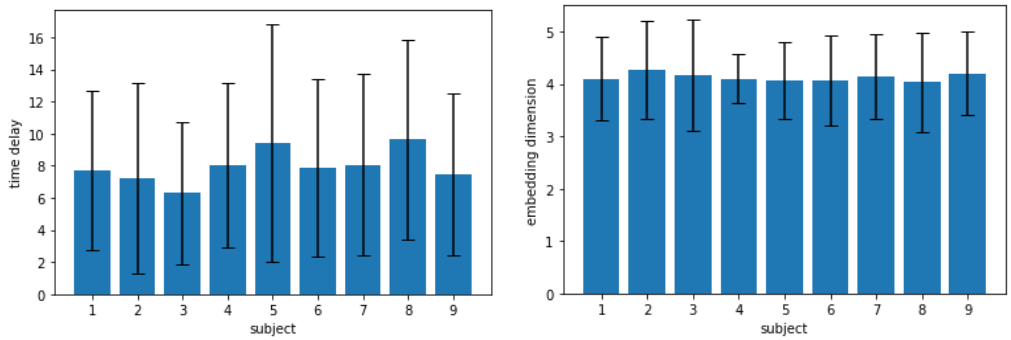
4.5.3 RESULTS AND DISCUSSIONS

TIME DELAY EMBEDDING PARAMETERS The optimal parameters of time delay embedding were computed for 30 randomly chosen trials from each subject using the methods discussed in Section 4.5.2. Fig. 4.5.11 shows the average optimal parameters for each sub-



(a) alpha and delta wave filtered from one trial (b) attractor of alpha wave (c) attractor of delta wave

Figure 4.5.10: Different dynamics in different frequency bands. The color in the attractor figure indicates time.



(a) time delay (b) embedding dimension

Figure 4.5.11: Average optimal parameters of time delay embedding computed using 30 randomly chosen trials from each subject in BCI competition IV 2a dataset.

ject. The embedding dimension is quite stable across subjects and has little variance within subjects while the time delay exhibits more variance both across and within subjects. To simplify the code and also exclude the influence of different embedding parameters during cross validation, the embedding parameters are fixed as the average across all subjects (8 for time delay and 4 for embedding dimension) in the following except when stated explicitly. Then the influence of parameters is further investigated in Fig. 4.5.12.

Features and classifiers With the parameters fixed in the previous paragraph and up to 3-dimensional homology computed, we studied in the first place the classification accuracy using different features, classifiers and number of channels. The length of each of the topological features corresponding to one channel equals to the homology dimension. For example, if the homology dimension is 3 and there are 2 channels, then the length of the feature vector is 6. According to the literature study in Section. 4.4, many studies used only one or two EEG channels. We selected the most relevant channels for the task of the dataset, i.e. C3 and C4 for motor imagery left vs right. The same classifiers as in Chapter 3 were applied on the topological features extracted from these channels. The results are summarized in Tab. 4.5.1. Most results are below the adjusted chance level (57% according to Combrisson & Jerbi [37]) except those with persistence as features which reach 58% in average with the Logistic Regression. We have seen in Section 3.2.2 with the signature method, that the classification accuracy could be improved by adding channels so we used the same pipeline and the same parameters as for the previous test except we used all available EEG channels instead of only the C3 and C4 channels. Tab. 4.5.2 shows that the results remain almost the same except again those with persistence as features that see an increase in accuracy up to 63% with Logistic Regression.

In terms of classifiers, the logistic regression classifier has a relatively good performance in both cases while the multilayer perceptron starts to show its power when more channels are used. In terms of features, the Betti numbers and the persistent entropy seem not helpful to distinguish left/right motor imagery while the persistence landscape and the persistence perform better. This is not a coincidence: the first two features are independent of the amplitude of the signal, while the last two features are not. This means that we might

	Betti	Entropy	Landscape	Persistence	mean
LDA	51.7(11.7)	53.2(14.4)	54.6(14.6)	57.8(17.4)	54.1(14.8)
SVM	51.2(12.4)	51.1(12.4)	55.6(14.9)	**58.2(17.1)**	54.0(14.7)
LR	51.5(11.6)	52.5(12.8)	54.9(15.8)	58.0(17.1)	**54.2(14.7)**
RF	50.5(14.3)	51.3(10.8)	55.7(13.5)	56.5(13.3)	53.5(13.3)
MLP	50.3(11.9)	51.4(10.4)	54.3(13.1)	54.9(15.2)	52.7(12.9)
mean	51.0(12.4)	51.7(12.3)	55.0(14.4)	**57.1(16.1)**	

Table 4.5.1: Average 10-fold cross-validation classification accuracy (%) with features extracted from only C3 and C4 channels and various classifiers of intra-subject classification on BCI competition IV 2a dataset (standard deviation between parentheses).

	Betti	Entropy	Landscape	Persistence	mean
LDA	49.2(13.3)	47.8(14.3)	52.7(13.1)	59.1(17.2)	52.2(15.2)
SVM	50.4(13.4)	48.9(13.8)	56.6(15.1)	62.5(17.1)	54.6(15.8)
LR	49.2(11.9)	49.5(13.0)	57.4(15.3)	**63.2(17.3)**	**54.8(15.6)**
RF	52.0(12.7)	51.7(13.3)	57.3(15.4)	55.9(14.9)	53.4(14.4)
MLP	49.9(10.0)	50.8(13.3)	56.8(15.2)	63.1(16.0)	55.2(14.9)
mean	49.6(12.3)	50.1(13.7)	55.8(14.9)	**60.7(16.7)**	

Table 4.5.2: Average 10-fold cross-validation classification accuracy (%) with features extracted from all channels and various classifiers of intra-subject classification on BCI competition IV 2a dataset (standard deviation between parentheses).

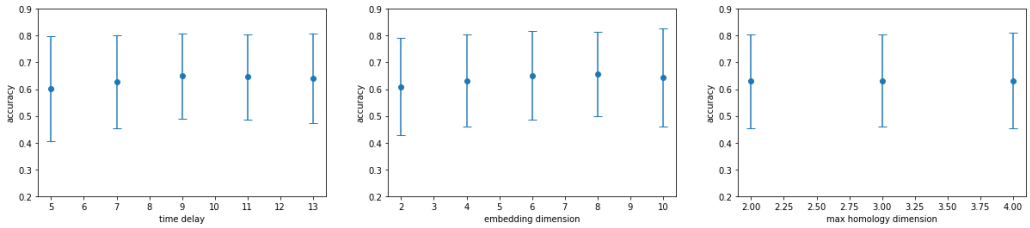
	Betti	Entropy	Landscape	Persistence	mean
LDA	49.8(11.7)	48.6(13.8)	51.2(13.3)	**54.2(13.4)**	50.9(13.2)
SVM	51.2(12.3)	49.3(14.2)	50.1(13.7)	46.7(8.5)	49.3(12.5)
LR	49.5(12.7)	50.0(13.9)	49.9(13.4)	52.5(13.7)	50.5(13.5)
RF	52.8(14.0)	52.0(11.5)	49.6(14.1)	51.7(15.0)	51.5(13.8)
MLP	51.8(10.4)	51.1(12.3)	50.6(13.0)	53.2(12.8)	**51.7(12.2)**
mean	51.0(12.4)	50.2(13.2)	50.3(13.5)	**51.6(13.2)**	

Table 4.5.3: Average 10-fold cross-validation classification accuracy (%) with features extracted from all channels and various classifiers of intra-subject classification on BCI competition IV 2a dataset. Signals are normalized before extraction of topological features (standard deviation between parentheses).

actually be using the spectral power of the signal in the guise of topological features, or at least in part.

To validate this conjecture, the signal is normalized to have zero mean and unit standard deviation before being embedded into the phase space. With the same parameters as for the Tab. 4.5.2, the new results are shown in Tab. 4.5.3. Compared to Tab. 4.5.2, the performance of persistence landscape as well as persistence drop below the adjusted chance level as the other two features with for instance LDA applied on persistence that reach only 54%, which confirms our conjecture.

Validation of parameters To further investigate the choice of parameters we varied the time delay, the embedding dimension and the homology dimension using the best classifier, logistic regression, and the best feature, persistence. Only one parameter was varied each time and the others were kept the same as before. Fig. 4.5.12 shows that the TDA method has a relatively stable performance with respect to change of parameters. The time delay chosen with mutual information is indeed optimal, while increasing the embedding dimension can continue to improve performance. However, the improvement is not signif-



(a) time delay (b) embedding dimension (c) homology dimension

Figure 4.5.12: Average classification accuracy using different parameters.

icant while the variance starts to get larger with a high embedding dimension. The performances using 2 or higher dimensional homology features have no considerable differences. That means the essential information is captured by 1-dimensional homology, i.e. loops, which correspond to the cyclicity of the EEG signal. This is in some way similar to the signature feature where level 2 signature is the key to the performance and it is also related to cyclicity.

FREQUENCY BANDS We have seen in Fig. 4.5.10 that separating frequency bands help to better capture different dynamics of each frequency bands, so the basic feature extraction steps shown in Fig. 4.4.1 are duplicated and applied on different frequency bands. Then the features extracted from each frequency band are concatenated to form the final feature vector. With the same parameters as in Tab. 4.5.2, we got the results in Tab. 4.5.4. The highest accuracy is slightly better (65.3% compared to 63.2% in Tab. 4.5.2) and above the adjusted chance level (57% according to Combrisson & Jerbi [37]), but still remains far from the gold standard methods (81% using a Riemannian classifier, cf. Tab. 3.3.1).

	Betti	Entropy	Landscape	Persistence	mean
LDA	49.0(14.2)	50.7(14.4)	54.2(13.9)	59.7(14.3)	53.4(14.8)
SVM	47.5(12.9)	49.9(13.5)	58.4(16.0)	**65.3(17.2)**	**55.3(16.6)**
LR	48.6(14.0)	48.7(13.3)	56.8(15.9)	63.6(16.7)	54.4(16.3)
RF	48.6(12.8)	52.7(13.1)	52.6(16.6)	58.1(15.8)	53.0(15.0)
MLP	48.3(10.4)	50.4(12.2)	56.3(14.6)	58.5(18.3)	53.4(14.8)
mean	48.4(13.0)	50.5(13.4)	55.6(15.6)	**61.0(16.8)**	

Table 4.5.4: Average 10-fold cross-validation classification accuracy (%) with features extracted from all channels and various classifiers of intra-subject classification on BCI competition IV 2a dataset. Signals are filtered into different frequency bands and the features extracted from each frequency band are concatenated. In parenthesis is the standard deviation.

Inter-subject classification To test the robustness of topological features across subjects, inter-subject classification is performed and the results are shown in Tab. 4.5.5. Compared to Tab. 4.5.2 which is the intra-subject classification with exactly the same setting, the performance is much better for all classifiers (especially random forest classifier which reaches 76% with persistent entropy) and features (except the persistence which reach at most 58% with logistic regression classifier). Traditional methods suffer from a significant drop of performance by passing from intra- to inter-subject classification, for example from 80% to 61% with CSP+SVM. However, it is the opposite for the TDA method which increases from 52% to 76% for example with persistence entropy and random forest classifier. Since both linear and non-linear classifiers have seen an improvement and the parameters of all classifiers were not tuned, this unusual trend might be a result of an insufficient number of training samples in the intra-subject case. The feature vector is of dimension 66 (3 for each 22 EEG channels) and the number of labeled trials for each subject is 144. In inter-subject classification, we have in total $144 \times 9 = 1296$ labeled trials which is significantly larger than in the intra-subject case.

	Betti	Entropy	Landscape	Persistence	mean
LDA	65.7(10.4)	70.3(13.1)	65.0(7.9)	57.9(7.3)	64.7(10.9)
SVM	66.0(10.1)	67.4(11.7)	67.5(8.5)	57.7(7.5)	64.7(10.4)
LR	67.1(11.3)	66.4(10.6)	63.0(7.8)	58.0(8.7)	63.6(10.4)
RF	**76.4(21.5)**	**76.4(21.5)**	75.9(21.1)	55.9(7.9)	**71.1(20.9)**
MLP	57.2(7.1)	62.8(8.4)	69.8(16.0)	57.4(8.1)	61.8(11.7)
mean	66.5(14.4)	**68.7(14.5)**	68.2(14.1)	57.4(8.0)	

Table 4.5.5: Inter-subject classification with topological features. The accuracy is obtained through the leave-one-out cross validation. Parameters are as follows: time delay is 8, embedding dimension is 4, homology dimension is 3 (standard deviation between parentheses).

	Betti	Entropy	Landscape	Persistence	mean
LDA	68.6(17.0)	65.6(17.5)	68.8(15.7)	66.3(16.6)	67.3(16.8)
SVM	68.4(19.5)	74.9(18.4)	73.2(16.2)	**78.9(15.8)**	73.8(17.9)
LR	65.7(17.6)	76.1(17.3)	74.7(16.0)	78.6(16.0)	73.8(17.4)
RF	76.9(17.2)	75.3(17.3)	75.5(17.4)	75.0(17.7)	**75.7(17.4)**
MLP	62.5(16.6)	76.4(16.6)	73.3(16.0)	78.5(15.2)	72.7(17.2)
mean	68.4(18.2)	**73.7(17.9)**	73.1(16.4)	75.4(17.0)	

Table 4.5.6: Intra-subject classification with the combination of TDA and CSP features. Same parameter setting with Tab. 4.5.2. The mean accuracy of CSP+SVM is 80.2(15.4) (standard deviation between parentheses).

Combination with CSP As mentioned in the motivation section of this chapter, the evolution information is complementary to traditional features used for EEG-based BCI such as the spectral power. So by combining the two, the performance might get improved and be more robust. We tested the combination of CSP features (6 components) and the above 4 topological features from all EEG channels. The intra-subject and inter-subject classification results are summarized in Tab. 4.5.6 and 4.5.7 respectively.

Interestingly, the intra-subject classification results are improved compared to that of the previous pure TDA pipeline, but the best combination (78.9% with persistence+SVM) is still below the baseline of CSP+SVM (80.2%).

	Betti	Entropy	Landscape	Persistence	mean
LDA	70.4(13.0)	73.6(15.0)	68.2(12.4)	60.3(10.5)	68.1(13.8)
SVM	72.7(14.2)	73.8(15.1)	71.4(13.3)	63.0(12.9)	70.2(14.6)
LR	70.3(12.6)	73.1(16.2)	68.8(13.5)	63.2(13.3)	68.8(14.4)
RF	76.8(16.8)	76.9(17.8)	**78.2(18.5)**	60.0(10.4)	**73.0(17.8)**
MLP	66.4(12.8)	70.1(13.6)	77.2(16.3)	61.5(13.8)	68.8(15.3)
mean	71.3(14.4)	**73.5(15.7)**	72.8(15.5)	61.6(12.3)	

Table 4.5.7: Inter-subject classification with the combination of TDA and CSP features. Same parameter setting with Tab. 4.5.5. The mean accuracy of CSP+SVM is 61.3(15.1) (standard deviation between parentheses).

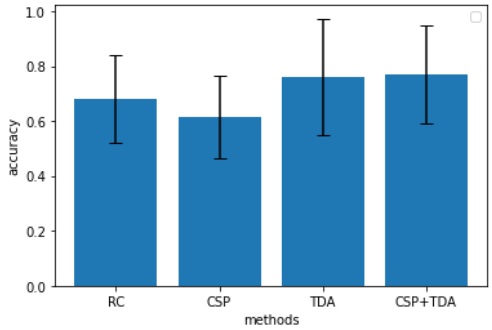


Figure 4.5.13: Benchmark of inter-subject classification on the BCI Competition IV 2a dataset. Here the TDA method is represented by the combination of persistent entropy and random forest. CSP is combined with SVM.

method	RC	CSP+SVM	TDA	CSP+TDA
accuracy	68.0(16.2)	61.3(15.1)	76.4(21.5)	76.9(17.8)

Table 4.5.8: Benchmark of inter-subject classification on the BCI Competition IV 2a dataset. Here the TDA method is represented by the combination of persistent entropy and random forest (standard deviation between parentheses).

However, the inter-subject classification is further improved compared to both TDA method and CSP+SVM. The CSP+TDA method reaches 78.2% with Landscape+RF, while the TDA method arrives at most 76.4%. Both are far above the CSP+SVM with a mean accuracy of 61.3%. To study this improvement further, another gold standard method, a Riemannian classifier (RC), is also added to the comparison. The averaged Leave-One-Out cross validation accuracy for four classification pipelines – RC, CSP, TDA and CSP+TDA – are summarized in Tab. 4.5.8 and Fig. 4.5.13. Here the results of persistent entropy+RF are used to represent the TDA method as both persistent entropy and RF have the best mean accuracy through the combinations in Tab. 4.5.5 and Tab. 4.5.7. From both the Fig. 4.5.13 and the Tab. 4.5.8, we could see that the combination of TDA and CSP features combines the high average accuracy of TDA and the low variance of CSP. Fig. 4.5.14 further shows the accuracy on each subject and for more topological features. It is clear that topological features except persistence and spectral power features are complementary since they behave differently on different subjects. For example on subject 2, 3 and 5 topological features perform much better than spectral power features while on subject 9 it is the inverse. Fortunately the combination of the features combines the strengths of the two features and preserves or even improves the best performance of the two. Interestingly, some subjects on which the TDA method performs well in inter-subject classification are also those who are not well classified by spectral power features in intra-subject classification (e.g. subject 2 and 5 as shown in Fig. 4.5.15).

Finally, concatenated features from different frequency bands were tested for inter-subject classification. The results are shown in Tab. 4.5.9. The performance decreases for all features except persistence which is much more correlated with spectral power. The

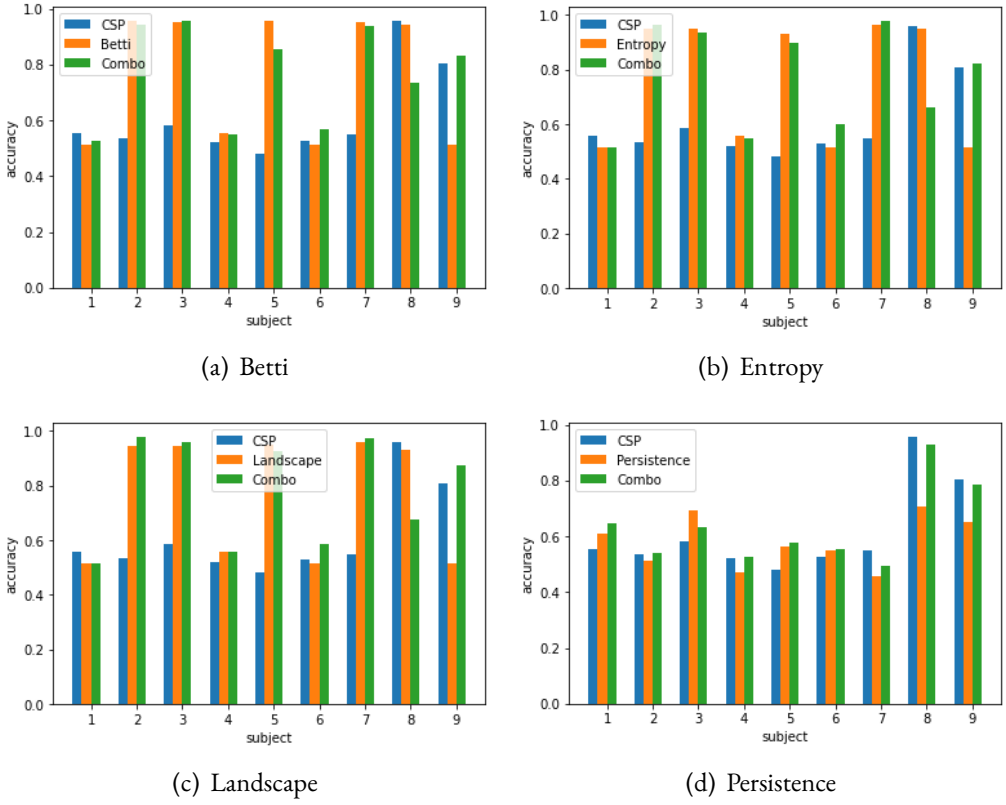


Figure 4.5.14: Leave-one-out cross validation accuracy on each subject. The classifiers used with topological features and the combined features are the random forest classifier for (a) to (c) and SVM for (d).

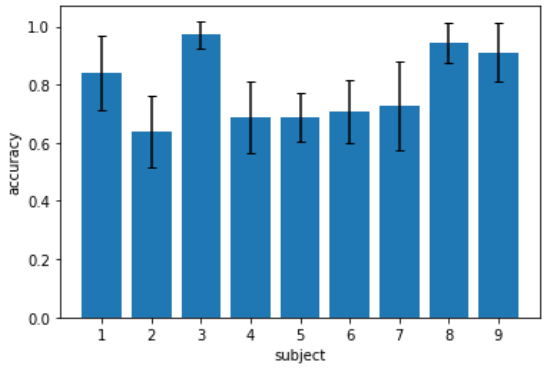


Figure 4.5.15: Intra-subject classification accuracy of each subject of the BCI competition IV 2a dataset with the CSP+SVM pipeline.

	Betti	Entropy	Landscape	Persistence	mean
LDA	61.3(9.1)	61.7(6.7)	61.3(6.5)	58.8(7.4)	60.8(7.6)
SVM	64.3(12.0)	65.7(9.7)	62.9(12.2)	63.9(11.2)	**64.2(11.4)**
LR	60.6(10.0)	64.7(11.3)	64.3(11.4)	**66.8(12.4)**	64.1(11.5)
RF	63.3(12.2)	62.5(10.2)	62.7(11.6)	62.7(9.9)	62.8(11.0)
MLP	55.6(4.0)	64.6(12.2)	60.7(8.7)	64.4(11.4)	61.3(10.3)
mean	61.0(10.4)	**63.8(10.3)**	62.4(10.4)	63.3(10.9)	

Table 4.5.9: Inter-subject classification with the combination of TDA and CSP features in alpha and beta bands separately. The performance of the same pipeline without TDA features is 66.1(13.5) (standard deviation between parentheses).

best performance reaches only 66.8% with Persistence+LR, 10% less than that in the whole frequency band case. The drop of performance might also be the lack of training samples since the length of the feature vector is doubled due to the separation of the alpha and beta frequency bands. Combining the results from all the experiments performed in this chapter, it seems that topological features are more robust but also need more training samples to be well learned by the classifier (linear or not).

4.6 Conclusion

In this chapter, we applied a new tool, topological data analysis, on EEG data for an MI EEG-based BCI application. First, a complete literature review is done to clarify the current research on the use of TDA for EEG processing. Following the pipeline the most theoretically sound, the EEG signal is embedded into the phase space using the time delay embedding, and then the Rips filtration is applied on its attractor to extract the persistent homology features.

However our results are not totally in line with those in some articles even though attention was payed to the choice of parameters. Yan et al. [186] reported very high classification

accuracy (more than 99%) on EEG emotion recognition with persistence landscape features combined with random forest classifier. With a dataset[97] of the same type (MI EEG-based BCI) as in our study, Altındiş et al.[2] concluded that the reconstructed attractor with optimal parameters showed significantly different 1st Betti number. However the difference is not quite remarkable and they did not show classification performance. Besides the comparison is between optimal and arbitrary parameters whereas in practice the choice of parameters is more approximate than arbitrary.

The experiments presented here demonstrate that the classification performance of TDA is robust to the choice of parameters. Also, using all channels available improve the performance. Interestingly, even though we get a performance below that of the gold standard method for intra-subject classification, it appears that this does not mean that topological features have no distinguishing power. Indeed, the inter-subject classification results are very promising and exceed that of the Riemannian classifier with an increment of 8% of accuracy, and even 10% when combined with CSP filtering. This shows the robustness and also the complexity (so that more training samples are needed) of topological features.

There are still many aspects to evaluate to fully understand and validate the conclusions. Perspectives include testing the TDA pipeline on datasets of larger size and of different BCI paradigms to see if the above observations hold. A theoretical explanation of why TDA features are robust across subjects is also needed.

Part of this work – the systematic review– has been promoted by a journal article: Xu X, Drougard N and Roy RN (2021). Topological Data Analysis as a New Tool for EEG Processing. Front. Neurosci. 15:761703. doi: 10.3389/fnins.2021.761703

The results will be promoted by a journal article currently being written. Preliminary

work was submitted to the NeurIPS 2021 BEETL Competition and is presented in Appendix A.

Topological Data Analysis for EEG-based BCI:

- TDA can be used to extract topological information on EEG dynamics, thus giving complementary information to traditional features of the EEG signal such as spectral power.
- The random forest classifier applied on persistent entropy features showed consistent good performance for inter-subject classification and outperformed gold standard methods such as CSP+SVM by 15% and Riemannian classifier by 8%.
- Topological features combined with CSP features can further improve the accuracy as well as the robustness of the estimation.

Maestoso

5

Postlude

The world is ever changing and so do we. The changes in our body are reflected by our physiological measures among which the EEG signal is an excellent example. These changes can be interpreted, at least in part, as a reflection of our learning and adapting to the changing environment. However, this variability becomes a major issue and a great challenge to

address when it comes to the problem of the classification of EEG signals, especially for BCI applications. The ideal classifier should be able to give a "correct" answer regardless of the person and the time. As we have seen, it is particularly difficult to handle the differences between individuals (the cross-subject variability). In addition, the signal is also non-stationary for instance for one person across time (the cross-session variability), tasks, and contexts.

Most of the human knowledge is obtained through finding something invariant under varying observations. For example the phase of the moon is always changing, but our ancestors found the period and the pattern of this change are invariant. The position of an object that we throw into the air traces a parabola. However if we derive it two times we get the acceleration which is constant.

Instead of using adaptive methods to chase after the changing features of the EEG signal, such as the transfer learning techniques mentioned at the end of Section 1.3, this PhD thesis explores the other overlooked avenue of extracting more robust features and tackles the problem of variability with the idea which is central in mathematics, the idea of distilling invariants that remain unchanged under transformation. First, a spatial method based on the Laplacian is proposed. By concentrating on the energy over low frequency Laplacian eigenvectors, some of the high spatial frequency variability could be reduced which leads to a minor improvement in classification. Second, a temporal method based on path signature gives features that are invariant under time reparametrization, which might remove the variability related to the speed of performing a task. Finally, the Topological Data Analysis technique is applied on the point cloud generated by time-delay embedded EEG signals with the hypothesis that the topological features of the EEG dynamics would be relatively

stable.

In an open science framework, we made several choices to promote the reproducibility and replicability of our results, and to have an ecological validity. All experiments were performed on open datasets and compared with the state-of-the-art BCI algorithms. Throughout our experiments we did not tune the parameters of the classifiers. The trials were not shuffled. The methods are also general enough to work on multiple BCI paradigms.

Main contributions First, the Laplacian-based method proved to be efficient for dimensionality reduction.On the Physionet MI dataset it improved the classification accuracy of the Riemannian classifier by 5% in intra-subject classification with approximately 1/3 of the original number of channels (which is 64). Second, the signature method combined with a Riemannian classifier was shown to be more robust in inter-subject classification. The advantage is even clearer when applied on users who have difficulty using MI BCI. Third, the TDA method outperformed gold standard methods for inter-subject classification (15% and 8% higher than CSP+SVM and RC respectively).

The findings of all the three methods point to the same direction: the relationship between channels is the key. In the first method, the spatial smoothness imposed by the low-pass filter enhanced the signal-to-noise ratio. In the second and third methods, the classification accuracies were improved when considering all the channels available instead of the only the "related" ones (such as C3, C4 for motor imagery). This finding is in line with the consensus of neuroscientists that cognitive functions are fulfilled not by single regions, but by networks at a system level[24,50].

Even though the signature and TDA methods did not outperform the gold-standard methods for intra-subject classification, they do provide complementary information to

traditional features. Indeed, it is encouraging to see that both showed better robustness in inter-subject classification, especially on subjects with poor classification accuracy by traditional methods. It might be a good choice to use the two methods as training methods over a large number of subjects in the initialization of a BCI system to reduce calibration time. In terms of computational cost, all three methods are fast to compute especially since it has been shown that the level 2 signature and the 3 dimensional homology are sufficient for a good performance. It might be especially helpful for people having difficulty to use BCIs and in clinical settings where calibration is particularly tiring for patients.

Limitations Although we benchmarked our methods with gold standards of the literature, a lack in our study is the missing comparison with transfer learning techniques [103,81] or deep learning methods [145]. However this could be remedied since we have chosen two intensively used publicly available datasets, the BCI competition IV 2a dataset and Physionet MI dataset. If time permits, we will gather the results in the literature on these datasets to perform this comparison.

Other limitations come from both methods and data. In the first Laplacian based method, the Laplacian depends either on the montage or the functional connectivity. In the former case, the position of the electrodes has to be picked from the standard montage or measured with specific equipment. In the latter case, our results showed that the functional graph estimated with EEG data is not optimal. In the second signature based method, the cyclicity analysis depends on the hypothesis that signals of all the channels are homogeneous while brain signals are known to be highly complex and heterogeneous [126,12]. The third TDA-based method relies on the hypothesis that the topology of the attractor is relatively stable. Even though it did show a certain robustness in inter-subject classification, the

characteristics of EEG dynamics such as fractal dimension have been shown to be modulated by circadian rhythm[39].

There are also limitations from data due to the specificity of EEG, such as the volume conduction which impairs its ability to estimate reliable functional connectivity graphs. Other limitations are more general in the BCI domain, for example the limited amount of data. It is impossible to get to know a person by meeting him only 2 or 3 times. It is the same when it comes to "knowing" the EEG signal of a person. However in most current BCI datasets there are only 2 or 3 sessions for each user in the best cases. The number of total users in most datasets is also far from what is needed to get reproducible results and valid conclusions[22].

Perspectives Technical breakthroughs will surely improve EEG-based BCIs and bring new insights to neuroscience. EEG recording system did not change a lot since the invention of EEG[168]. Perhaps it might be possible to improve its spatial resolution while always being non-invasive, as it has been reported last year that the temporal resolution of fMRI could be greatly improved with the fast fMRI[134]. However for EEG-based BCIs this would mean working with a very high number of electrodes – undesirable for real life use – or drastically improving source reconstruction algorithms.

Methodological advances are also needed, especially the joint effort of researchers from different fields. For instance, recently there has been some works on modeling working memory based on neural spike patterns and network properties[77,110]. Working memory tasks are often employed to elicit mental workload levels. However to this day the computational models on mental workload itself are still rare. One of the reasons might be that mental workload is a transactional concept involving the interaction between the individual

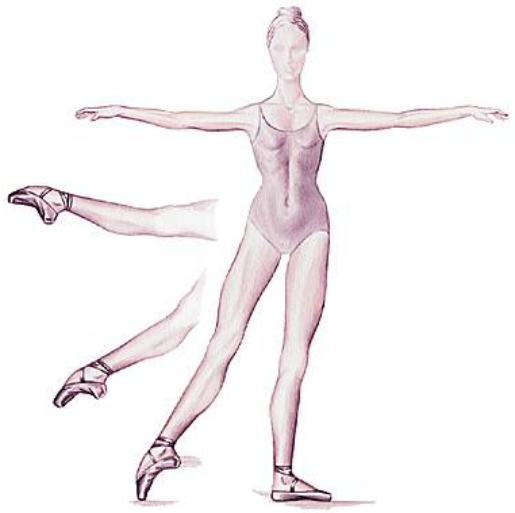


Figure 5.0.1: Different phases in the dégagé à la seconde in ballet (From Larousse encyclopedia).

capacities and the ongoing task[170,41].

In sport science or physical therapy, muscle contraction is inferred from a single measure, and the different stages of a "dégagé à la seconde" (ballet, Fig.5.0.1) might for instance be easily estimated from the tension of the involved group of muscles. A simple, direct, reliable and robust across subjects measure of cognitive effort based on EEG still remains to be found.

On another note, even though we have focused our work on inter-subject variability, EEG variability in the context of time should also be investigated thoroughly. Indeed, cycles of different scales ranging from ultradian (within one day), circadian (dayly), circaseptan (weekly), circatrigintan (monthly) to circannual (yearly) rhythms affect the human body hence the physiological measures. Particularly the strong impact of ultradian and circadian rhythms on EEG signals has been known for more than 50 years[99]. However it is seldom taken into consideration in BCI applications. Inter-session data, when available, are not

necessarily recorded at the same hour and the variability due to the circadian rhythm is not quantified.

Moreover, EEG variability should also be understood in the context of the brain, since EEG is just one measurement of brain activity. Similar (but not exactly the same) variability could also be observed on other brain signals such as fMRI (called BOLD variability in the literature)[60,93]. Recent advances in neuroscience reveal that continual reorganization of neuronal responses (called *neural* or *representational drift*) can be found in the sensorimotor cortex[146], hippocampus[191,86], primary olfactory cortex[152], and posterior parietal cortex[44]. The BCI inter-session variability might partly reflect this continual reorganization of the brain even if the task is fully learned.

Final note Variability is pervasive at all scales: a healthy ecosystem has a variety of species; a healthy society is composed of people of different opinions; a healthy brain is also a network of neurons of different morphological and functional classes. The fundamental reason for being variable is that it permits the system to have a rich repertoire of equilibrium states so that the system is versatile and robust. Dictatorship happens in a society having not enough variability, in the same way that the epilepsy seizure is characterized by an over synchronized brain[98]. On the other hand, a society with too much variability is not efficient, and an autistic brain present an excessive neural variability[43].

Garrett et al.[51] reported that a high (but not excessive) level of variability of the brain is associated with a high consistency of performance. It is ironic that what makes the natural system robust makes the artificial classifier not robust. Nature is always right and maybe we should reexamine our knowledge and methods. The root of the problem is that we still do not fully understand the underlying neural mechanism. We can observe the spiking

of the single neurons, the changes of connections between neurons and the activation of the neural networks. But the connection between these observations and the high level cognitive functions is still not clear. The "variability" might be due to the fact that we are working with the wrong representation. As Laplace said "ignorance of the different causes involved in the production of events, as well as their complexity, taken together with the imperfection of analysis, prevents our reaching the same certainty about the vast majority of phenomena"[94].

At the very end of this thesis I permit myself some personal point of view concerning the main subject: the endless change of EEG signals. The Buddhists have no knowledge on EEG but through meditation they have known and observed the ever changing nature of thoughts (and of the world) since more than two thousand years ago. According to Buddhists, everything is empty of intrinsic existence and nature, and the existence of things relies on the relationships that they have with other things[59]. All phenomenon that we observe are merely consequences of causes and conditions[140]. The world of phenomena changes under the constant change of interactions of things connected together and it is impossible to catch a phenomenon.

This interdependence point of view is perhaps the most important thing I have learned in doing this PhD. It is perfectly encapsulated by Poincaré in his book *Science and Hypothese*: "the aim of science is not things themselves, as the dogmatists in their simplicity imagine, but the relations between things; outside those relations there is no reality knowable"[133].

Life is a fabulous journey full of surprising changes. No matter how neurons drift in the brain, the memories and the love remain.

A

Technical report for the NeurIPS 2021 BEETL Competition: Benchmarks for EEG Transfer Learning

This appendix presents work submitted to the NeurIPS 2021 BEETL Competition that apply the Topological Data Analysis method presented in Chapter 4 on a sleep dataset provided by the competition.

A.1 Introduction

Electroencephalography (EEG), combined with other physiological data, have been proved to be a reliable source to indicate sleep stages. However, the sleep stage annotation from EEG data still relies heavily on human experts[106]. There are two main obstacles to automate this procedure: the low signal-to-noise ratio (SNR) and non-stationarity across sessions and subjects. More robust features and transfer learning strategies are needed to overcome these obstacles. Topological features are robust by nature, hence they constitute a promising candidate to partially solve the problem, even without any transfer learning techniques.

A.2 Methods

We propose a very simple method based on topological data analysis (TDA) and convolutional neural network (CNN). This method has been shown to be promising to detect gravitational waves[19], which also have a very low SNR as EEG. The code is based on keras[32] and the topological machine learning toolbox giotto-tda[164]. The raw data is first subsampled using a fast Fourier transform to 1000 points per trial in order to speed up the computation of the persistence diagrams and to perform feature dimensionality reduction.

A.2.1 Topological data analysis

Time delay embedding Time delay embedding is a technique widely used in physics for the estimation of invariants of nonlinear time series[85]. Let $X(t)$ denote the time series of one EEG channel. Then time delay embedding consists in mapping the time series from

its original space into its phase space by keeping several equally spaced observations, *i.e.* $X(t) \mapsto \big[X(t), X(t+\tau), X(t+2\tau), \ldots, X(t+(d-1)\tau)\big] \in \mathbb{R}^d$, where τ is the time delay and d is the embedding dimension. We chose the time delay τ that gives the minimum mutual information (so that adding $X(t+\tau)$ gives maximum information) and the d using the False Nearest Neighbours algorithm to have an optimal embedding.

Persistent homology Due to the article length restriction, we refer the interested readers to [27,157] for detailed explanation of the following notions. Topological data analysis is a young field aiming to tackle data analysis problems using topological tools which get the essence of a shape (invariant under diffeomorphisms). One effective way to distinguish different objects from a topological point of view is to compute their homology groups. Two objects with different homology groups are different (the inverse is not true). Persistent homology records a series of homology groups based on a process called filtration which provides a multi-scale summary of the data.

Our hypothesis is that we could classify EEG time series by computing the persistent homology of the point cloud formed by time delay embedding that preserves the dynamics of the EEG time series. More specifically, first we constructed the Rips complex corresponding to different levels of filtration based on the point cloud. Then persistent homology was computed and the length of the 1-dimensional persistence interval was extracted as topological features. Intuitively, they capture the periodic properties of the EEG time series.

A.2.2 Convolutional neural networks

Finally the subsampled raw signal of each channel was concatenated with the topological features and fed into a convolutional network proposed by [177]. The network stacked

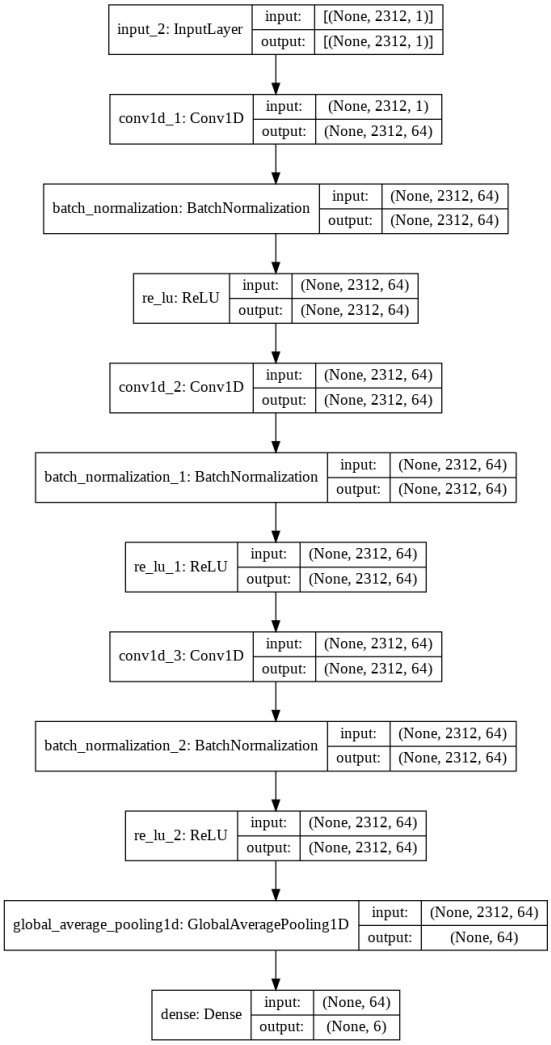


Figure A.2.1: Network Structure.

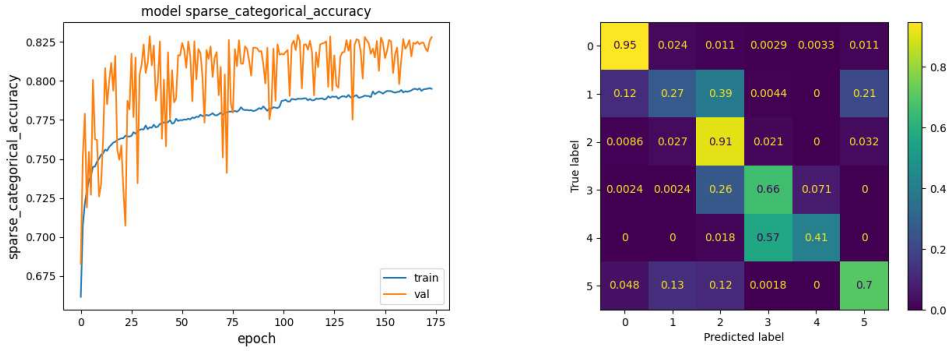
three convolutional blocks followed by a softmax layer. Each of the convolutional block was comprised of a 1-D convolutional layer (64 filters of size 3) followed by a batch normalization layer and a ReLU activation layer. The structure of the network is shown in Fig. A.2.1.

A.3 Results

Target group data were evenly divided into two parts: subject 0-2 were used as validation data to select the best model; subject 3-5 were used as test data. The network was trained with Adam optimizer and a learning rate of 10^{-4}. The training/validation accuracy and confusion matrix are shown in Fig. A.4.1. The -unbalanced- test accuracy is 80.8%. Interestingly, a Student's t-test showed that the topological features between source and target groups are not significantly different.

A.4 Discussion

Since our main goal was to evaluate the relevance of EEG topological features for cross-subject sleep stage classification, we focused on feature extraction and did not tune the network hyperparameters. This can explain the training/validation accuracy in Fig. A.4.1, as the training accuracy is still increasing while the validation accuracy is not yet stable. Besides the validation accuracy is above the training accuracy. There might be two reasons. Firstly, the sleep stage distributions are different. We used data from older people as validation data which contain less samples from stage 3, 4 and REM where the algorithm performs less well compared to stage 2 and wake (see confusion matrix in Fig. A.4.1). Secondly, the validation data size is relatively small compared to training data, which leads to



(a) Training/validation accuracy

(b) Confusion matrix

Figure A.4.1: Results.

more randomness.

Despite this need for fine tuning the network in future work, these preliminary results show that topological features are relevant for sleep stage estimation and might help tackle the non-stationary issues of EEG.

References

[1] Altindis, F., Yilmaz, B., Borisenok, S., & Icoz, K. (2018). Use of topological data analysis in motor intention based brain-computer interfaces. In *2018 26th European Signal Processing Conference (EUSIPCO)* (pp. 1695–1699).

[2] Altındiş, F., Yılmaz, B., Borisenok, S., & İçöz, K. (2021). Parameter investigation of topological data analysis for eeg signals. *Biomedical Signal Processing and Control*, 63, 102196.

[3] Arazi, A., Yeshurun, Y., & Dinstein, I. (2019). Neural variability is quenched by attention. *Journal of Neuroscience*, 39(30), 5975–5985.

[4] Aricò, P., Borghini, G., Flumeri, G. D., Sciaraffa, N., & Babiloni, F. (2018). Passive BCI beyond the lab: current trends and future directions. *Physiological Measurement*, 39(8), 08TR02.

[5] Arsigny, V., Fillard, P., Pennec, X., & Ayache, N. (2007). Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 29(1), 328–347.

[6] Arvaneh, M., Guan, C., Ang, K. K., & Quek, C. (2013). EEG Data Space Adaptation to Reduce Intersession Nonstationarity in Brain-Computer Interface. *Neural Computation*, 25(8), 2146–2171.

[7] Ashcroft, F. (2012). *The spark of life: electricity in the human body*. WW Norton & Company.

[8] Atasoy, S., Donnelly, I., & Pearson, J. (2016). Human brain networks function in connectome-specific harmonic waves. *Nature Communications*, 7(1), 10340.

[9] Barachant, A., Bonnet, S., Congedo, M., & Jutten, C. (2012). Multiclass brain–computer interface classification by riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4), 920–928.

[10] Barachant, A., Bonnet, S., Congedo, M., & Jutten, C. (2013). Classification of covariance matrices using a riemannian-based kernel for bci applications. *Neurocomputing*, 112, 172–178. Advances in artificial neural networks, machine learning, and computational intelligence.

[11] Barlow, H. B. (1999). The neurologic of matching filters. *J. Opt. Technol.*, 66(9), 776.

[12] Baroni, F. & Mazzoni, A. (2014). Heterogeneity of heterogeneities in neuronal networks. *Frontiers in Computational Neuroscience*, 8.

[13] Baryshnikov, Y. & Schlafly, E. (2016). Cyclicity in multivariate time series and applications to functional mri data. In *2016 IEEE 55th Conference on Decision and Control (CDC)* (pp. 1625–1630).

[14] Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F., & Moulines, E. (1997). A blind source separation technique using second-order statistics. *IEEE Transactions on signal processing*, 45(2), 434–444.

[15] Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., Perelmouter, J., Taub, E., & Flor, H. (1999). A spelling device for the paralysed. *Nature*, 398(6725), 297–298.

[16] Bischof, B. & Bunch, E. (2021). Geometric feature performance under downsampling for eeg classification tasks. *arXiv*, cs.LG, 2102.07669.

[17] Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., & Muller, K.-R. (2008). Optimizing spatial filters for robust eeg single-trial analysis. *IEEE Signal Processing Magazine*, 25, 41–56.

[18] Bressler, S. L. (2002). Understanding cognition through large-scale cortical networks. *Current Directions in Psychological Science*, 11(2), 58–61.

[19] Bresten, C. & Jung, J.-H. (2019). Detection of gravitational waves using topological data analysis and convolutional neural network: An improved approach.

[20] Bubenik, P. (2015). Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.*, 16(1), 77–102.

[21] Bubenik, P. et al. (2015). Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.*, 16(1), 77–102.

[22] Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365–376.

[23] Buzsáki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents — eeg, ecog, lfp and spikes. *Nature Reviews Neuroscience*, 13(6), 407–420.

[24] Caligiore, D., Pezzulo, G., Baldassarre, G., Bostan, A. C., Strick, P. L., Doya, K., Helmich, R. C., Dirkx, M., Houk, J., Jörntell, H., Lago-Rodriguez, A., Galea, J. M., Miall, R. C., Popa, T., Kishore, A., Verschure, P. F. M. J., Zucca, R., & Herreros, I. (2017). Consensus paper: Towards a systems-level view of cerebellar function: the interplay between cerebellum, basal ganglia, and cortex. *The Cerebellum*, 16(1), 203–229.

[25] Carvalhaes, C. & de Barros, J. A. (2015). The surface laplacian technique in eeg: Theory and methods. *International Journal of Psychophysiology*, 97(3), 174–188.

[26] Chazal, F., De Silva, V., Glisse, M., & Oudot, S. (2012). *The structure and stability of persistence modules*. Springer International Publishing. working paper or preprint.

[27] Chazal, F. & Michel, B. (2021). An introduction to topological data analysis: fundamental and practical aspects for data scientists. *arXiv math.ST 1710.04019*.

[28] Chen, K.-T. (1958). Integration of paths – a faithful representation of paths by noncommutative formal power series. *Transactions of the American Mathematical Society*, 89, 395–407.

[29] Chevyrev, I. & Kormilitzin, A. (2016). A primer on the signature method in machine learning.

[30] Chevyrev, I. & Oberhauser, H. (2018). Signature moments to characterize laws of stochastic processes.

[31] Chintakunta, H., Gentimis, T., Gonzalez-Diaz, R., Jimenez, M.-J., & Krim, H. (2015). An entropy-based persistence barcode. *Pattern Recognition*, 48(2), 391–401.

[32] Chollet, F. et al. (2015). Keras.

[33] Chung, F. R. (1997). *Spectral graph theory*, volume 92. American Mathematical Soc.

[34] Cincotti, F., Bianchi, L., del R Millan, J., Mourino, J., Salinari, S., Marciani, M., & Babiloni, F. (2001). Brain computer interface: the use of low resolution surface laplacian and linear classifiers for the recognition of imagined hand movements. In *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 1 (pp. 655–658 vol.1).

[35] Clerc, M., Bougrain, L., & Lotte, F. (2016). *Brain-computer interfaces 1: Methods and perspectives*. John Wiley & Sons.

[36] Cohen-Steiner, D., Edelsbrunner, H., & Harer, J. (2007). Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1), 103–120.

[37] Combrisson, E. & Jerbi, K. (2015). Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of Neuroscience Methods*, 250, 126–136. Cutting-edge EEG Methods.

[38] Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273–297.

[39] Croce, P., Quercia, A., Costa, S., & Zappasodi, F. (2018). Circadian rhythms in fractal features of eeg signals. *Frontiers in Physiology*, 9.

[40] Curto, C. (2017). What can topology tell us about the neural code? *Bulletin of the American Mathematical Society*, 54(1), 63–78.

[41] Dehais, F., Lafont, A., Roy, R., & Fairclough, S. (2020). A neuroergonomics approach to mental workload, engagement and human performance. *Frontiers in Neuroscience*, 14.

[42] Desbrun, M., Meyer, M., Schröder, P., & Barr, A. H. (1999). Implicit fairing of irregular meshes using diffusion and curvature flow. In *Proc Annual Conf Computer Graphics and Interactive Techniques* (pp. 317–324). USA: ACM Press/Addison-Wesley Publishing Co.

[43] Dinstein, I., Heeger, D. J., & Behrmann, M. (2015). Neural variability: friend or foe? *Trends in Cognitive Sciences*, 19(6), 322–328.

[44] Driscoll, L. N., Pettit, N. L., Minderer, M., Chettih, S. N., & Harvey, C. D. (2017). Dynamic reorganization of neuronal activity patterns in parietal cortex. *Cell*, 170(5), 986–999.

[45] FaiFaisal2008sal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience*, 9(4), 292–303.

[46] Farrens, J. L., Simmons, A. M., Luck, S. J., & Kappenman, E. S. (2021). Electroencephalogram (eeg) recording protocol for cognitive and affective human neuroscience research.

[47] Farwell, L. A. & Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr Clin Neurophysiol*, 70(6), 510–523.

[48] Feldman, D. E. (2009). Synaptic mechanisms for plasticity in neocortex. *Annual Review of Neuroscience*, 32(1), 33–55. PMID: 19400721.

[49] Flesher, S. N., Downey, J. E., Weiss, J. M., Hughes, C. L., Herrera, A. J., Tyler-Kabara, E. C., Boninger, M. L., Collinger, J. L., & Gaunt, R. A. (2021). A brain-computer interface that evokes tactile sensations improves robotic arm control. *Science*, 372(6544), 831–836.

[50] Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Essen, D. C. V., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, 102(27), 9673–9678.

[51] Garrett, D. D., Kovacevic, N., McIntosh, A. R., & Grady, C. L. (2011). The importance of being variable. *Journal of Neuroscience*, 31(12), 4496–4503.

[52] Gayraud, N. T. H., Rakotomamonjy, A., & Clerc, M. (2017). Optimal Transport Applied to Transfer Learning For P300 Detection. In *BCI 2017 - 7th Graz Brain-Computer Interface Conference* (pp.6). Graz, Austria.

[53] Ghrist, R. (2014). *Elementary Applied Topology*. CreateSpace Independent Publishing Platform.

[54] Gibson, E., Lobaugh, N. J., Joordens, S., & McIntosh, A. R. (2022). Eeg variability: Task-driven or subject-driven signal of interest? *NeuroImage*, 252, 119034.

[55] Giusti, C. & Lee, D. (2020). Iterated integrals and population time series analysis. In N. A. Baas, G. E. Carlsson, G. Quick, M. Szymik, & M. Thaule (Eds.), *Topological Data Analysis* (pp. 219–246). Cham: Springer International Publishing.

[56] Giusti, C., Pastalkova, E., Curto, C., & Itskov, V. (2015). Clique topology reveals intrinsic geometric structure in neural correlations. *Proceedings of the National Academy of Sciences*, 112(44), 13455–13460.

[57] Glomb, K., Rué Queralt, J., Pascucci, D., Defferrard, M., Tourbier, S., Carboni, M., Rubega, M., Vulliémoz, S., Plomp, G., & Hagmann, P. (2020). Connectome spectral analysis to track eeg task dynamics on a subsecond scale. *NeuroImage*, 221, 117137.

[58] Gordon, C., Webb, D., & Wolpert, S. (1992). Isospectral plane domains and surfaces via riemannian orbifolds. *Inventiones mathematicae*, 110(1), 1–22.

[59] Gowans, C. W. (2014). *Buddhist moral philosophy: An introduction*. Routledge.

[60] Grady, C. L. & Garrett, D. D. (2014). Understanding variability in the bold signal and why it matters for aging. *Brain imaging and behavior*, 8(2), 274–283. 24008589[pmid].

[61] Graichen, U., Eichardt, R., Fiedler, P., Strohmeier, D., & Haueisen, J. (2011). Adaptive spatial harmonic analysis of eeg data using laplacian eigenspace. In *Int Symp Noninvasive Functional Source Imaging Brain Heart* (pp. 18–21).

[62] Graichen, U., Eichardt, R., Fiedler, P., Strohmeier, D., Zanow, F., & Haueisen, J. (2015). Sphara–a generalized spatial fourier analysis for multi-sensor systems with non-uniformly arranged sensors: application to eeg. *PloS one*, 10(4), e0121741.

[63] Gramfort, A. & Clerc, M. (2007). Low dimensional representations of meg/eeg data using laplacian eigenmaps. In *2007 Joint Meeting of the 6th International Symposium on Noninvasive Functional Source Imaging of the Brain and Heart and the International Conference on Functional Biomedical Imaging* (pp. 169–172).

[64] Gramfort, A., Luessi, M., Larson, E., Engemann, D., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). Meg and eeg data analysis with mne-python. *Frontiers in Neuroscience*, 7.

[65] Grassberger, P. & Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica D: Nonlinear Phenomena*, 9(1), 189–208.

[66] Grone, R., Merris, R., & Sunder, V. S. (1990). The laplacian spectrum of a graph. *SIAM Journal on Matrix Analysis and Applications*, 11(2), 218–238.

[67] Haak, K. V. & Beckmann, C. F. (2020). Understanding brain organisation in the face of functional heterogeneity and functional multiplicity. *NeuroImage*, 220, 117061.

[68] Hambly, B. & Terry, L. (2010). Uniqueness for the signature of a path of bounded variation and the reduced path group. *Annals of mathematics*.

[69] Hatcher, A. (2000). *Algebraic topology*. Cambridge: Cambridge Univ. Press.

[70] Hinss, M. F., Darmet, L., Somon, B., Jahanpour, E., Lotte, F., Ladouce, S., & Roy, R. N. (2021a). An EEG dataset for cross-session mental workload estimation: Passive BCI competition of the Neuroergonomics Conference 2021. The project was validated by the local ethical committee of the University of Toulouse (CER number 2021-342).

[71] Hinss, M. F., Jahanpour, E. S., Somon, B., Pluchon, L., Dehais, F., & Roy, R. N. (2022). COG-BCI database: A multi-session and multi-task EEG cognitive dataset for passive brain-computer interfaces. This database has been created as part of a project validated by the local ethical committee of the University of Toulouse (CER number 2021-342).

[72] Hinss, M. F., Somon, B., Dehais, F., & Roy, R. N. (2021b). Open EEG datasets for passive brain-computer interface applications: Lacks and perspectives. In *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)* (pp. 686–689).

[73] Hitziger, S. (2015). *Modeling the variability of electrical activity in the brain*. Theses, Université Nice Sophia Antipolis.

[74] Hitziger, S., Clerc, M., Saillet, S., Bénar, C., & Papadopoulo, T. (2017). Adaptive waveform learning: A framework for modeling variability in neurophysiological signals. *IEEE Transactions on Signal Processing*, 65(16), 4324–4338.

[75] Hjorth, B. (1975). An on-line transformation of eeg scalp potentials into orthogonal source derivations. *Electroencephalography and clinical neurophysiology*, 39(5), 526–530.

[76] Homan, R. W., Herman, J., & Purdy, P. (1987). Cerebral location of international 10–20 system electrode placement. *Electroencephalography and clinical neurophysiology*, 66(4), 376–382.

[77] Huang, Q.-S. & Wei, H. (2021). A computational model of working memory based on spike-timing-dependent plasticity. *Frontiers in Computational Neuroscience*, 15.

[78] Huang, W., Goldsberry, L., Wymbs, N. F., Grafton, S. T., Bassett, D. S., & Ribeiro, A. (2016). Graph frequency analysis of brain signals. *IEEE Journal of Selected Topics in Signal Processing*, 10(7), 1189–1203.

[79] Ibáñez-Marcelo, E., Campioni, L., Phinyomark, A., Petri, G., & Santarcangelo, E. L. (2019). Topology highlights mesoscopic functional equivalence between imagery and perception: The case of hypnotizability. *NeuroImage*, 200, 437–449.

[80] Islam, M. K., Rastegarnia, A., & Yang, Z. (2016). Methods for artifact detection and removal from scalp eeg: A review. *Neurophysiologie Clinique/Clinical Neurophysiology*, 46(4), 287–305.

[81] Jayaram, V., Alamgir, M., Altun, Y., Scholkopf, B., & Grosse-Wentrup, M. (2016). Transfer learning in brain-computer interfaces. *IEEE Computational Intelligence Magazine*, 11(1), 20–31.

[82] Jayaram, V. & Barachant, A. (2018). MOABB: trustworthy algorithm benchmarking for BCIs. *Journal of Neural Engineering*, 15(6), 066011.

[83] Jost, J. (2011). *Riemannian Geometry and Geometric Analysis*. Springer Berlin, Heidelberg.

[84] Kac, M. (1966). Can one hear the shape of a drum? *The American Mathematical Monthly*, 73(4), 1–23.

[85] Kantz, H. & Schreiber, T. (2003). *Nonlinear Time Series Analysis*. Cambridge University Press, 2 edition.

[86] Keinath, A. T., Mosser, C.-A., & Brandon, M. P. (2022). The representation of context in mouse hippocampus is preserved despite neural drift. *Nature Communications*, 13(1), 2415.

[87] Kennel, M. B., Brown, R., & Abarbanel, H. D. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical review A*, 45(6), 3403.

[88] Kidger, P. & Lyons, T. (2021). Signatory: differentiable computations of the signature and logsignature transforms, on both CPU and GPU. In *International Conference on Learning Representations*. https://github.com/patrick-kidger/signatory.

[89] Kim, S.-P. (2018). *Preprocessing of EEG*, (pp. 15–33). Springer Singapore: Singapore.

[90] Krogh, A. & Hertz, J. A. (1992). Generalization in a linear perceptron in the presence of noise. *Journal of Physics A: Mathematical and General*, 25(5), 1135–1147.

[91] Krumpe, T., Baumgaertner, K., Rosenstiel, W., & Spüler, M. (2017). Non-stationarity and inter-subject variability of eeg characteristics in the context of bci development. In *GBCIC*.

[92] Krusienski, D. J., Sellers, E. W., Cabestaing, F., Bayoudh, S., McFarland, D. J., Vaughan, T. M., & Wolpaw, J. R. (2006). A comparison of classification techniques for the p300 speller. *Journal of Neural Engineering*, 3(4), 299–305.

[93] Kumral, D., Şansal, F., Cesnaite, E., Mahjoory, K., Al, E., Gaebler, M., Nikulin, V., & Villringer, A. (2020). Bold and eeg signal variability at rest differently relate to aging in the human brain. *NeuroImage*, 207, 116373.

[94] Laplace, S. (1825). *Essai Philosophique sur les Probabilités*. Paris: Gauthier-Villars.

[95] Lee, D. & Ghrist, R. (2020). Path signatures on lie groups.

[96] Lee, D. & Schachter, B. (1980). Two algorithms for constructing a delaunay triangulation. *Int J Comput Information Sci*.

[97] Leeb, R., Brunner, C., Müller-Putz, G., Schlögl, A., & Pfurtscheller, G. (2008). Bci competition 2008–graz data set b. *Graz University of Technology, Austria*, (pp. 1–6).

[98] Lehnertz, K., Bialonski, S., Horstmann, M.-T., Krug, D., Rothkegel, A., Staniek, M., & Wagner, T. (2009). Synchronization phenomena in human epileptic brain networks. *Journal of Neuroscience Methods*, 183(1), 42–48. BrainModes: A Principled Approach to Modeling and Measuring Large-Scale Neuronal Activity.

[99] Lehnertz, K., Rings, T., & Bröhl, T. (2021). Time in brain: How biological rhythms impact on eeg signals and on eeg-derived brain networks. *Frontiers in Network Physiology*, 1.

[100] Levy, B. (2006). Laplace-beltrami eigenfunctions towards an algorithm that "understands" geometry. In *IEEE Int Conf Shape Modeling and Applications* (pp. 13–13).

[101] Lindgren, J. (July 12, 2018). Tutorial – level 1 – choosing the bci paradigm. http://openvibe.inria.fr/tutorial-level-1-choosing-the-bci-paradigm/.

[102] Lioi, G., Gripon, V., Brahim, A., Rousseau, F., & Farrugia, N. (2021). Gradients of connectivity as graph Fourier bases of brain activity. *Network Neuroscience*, 5(2), 322–336.

[103] Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F. (2018). A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update. *Journal of Neural Engineering*, 15(3), 031005.

[104] Lyons, T. J., Caruana, M., & Lévy, T. (2007). *Differential equations driven by rough paths*. Springer.

[105] Majumder, S., Apicella, F., Muratori, F., & Das, K. (2020). Detecting autism spectrum disorder using topological data analysis. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1210–1214).

[106] Malafeev, A., Laptev, D., Bauer, S., Omlin, X., Wierzbicka, A., Wichniak, A., Jernajczyk, W., Riener, R., Buhmann, J., & Achermann, P. (2018). Automatic human sleep stage scoring using deep neural networks. *Frontiers in Neuroscience*, 12, 781.

[107] Masquelier, T. (2013). Neural variability, or lack thereof. *Frontiers in Computational Neuroscience*, 7.

[108] McFarland, D. J. (2015). The advantages of the surface laplacian in brain-computer interface research. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, 97(3), 271–276. 25091286[pmid].

[109] McIntosh, A. R., Kovacevic, N., & Itier, R. J. (2008). Increased brain signal variability accompanies lower behavioral variability in development. *PLOS Computational Biology*, 4(7), 1–9.

[110] Mejías, J. F. & Wang, X.-J. (2022). Mechanisms of distributed working memory in a large-scale network of macaque neocortex. *eLife*, 11, e72136.

[111] Meyer, M., Desbrun, M., Schröder, P., & Barr, A. H. (2003). Discrete differential-geometry operators for triangulated 2-manifolds. In H.-C. Hege & K. Polthier (Eds.), *Visualization and Mathematics III* (pp. 35–57). Berlin, Heidelberg: Springer Berlin Heidelberg.

[112] Miyakoshi, M. (2022). Makoto's preprocessing pipeline.

[113] Mohar, B. (1989). Isoperimetric numbers of graphs. *Journal of Combinatorial Theory, Series B*, 47(3), 274–291.

[114] Moore, P. J., Lyons, T. J., Gallacher, J., & Initiative, A. D. N. (2019). Using path signatures to predict a diagnosis of alzheimer's disease. *PloS one*, 14(9).

[115] Morioka, H., Kanemura, A., ichiro Hirayama, J., Shikauchi, M., Ogawa, T., Ikeda, S., Kawanabe, M., & Ishii, S. (2015). Learning a common dictionary for subject-transfer decoding with resting calibration. *NeuroImage*, 111, 167–178.

[116] Morrill, J., Fermanian, A., Kidger, P., & Lyons, T. (2020). A generalised signature method for multivariate time series feature extraction.

[117] Moses, D. A., Metzger, S. L., Liu, J. R., Anumanchipalli, G. K., Makin, J. G., Sun, P. F., Chartier, J., Dougherty, M. E., Liu, P. M., Abrams, G. M., Tu-Chan, A., Ganguly, K., & Chang, E. F. (2021). Neuroprosthesis for decoding speech in a paralyzed person with anarthria. *New England Journal of Medicine*, 385(3), 217–227.

[118] NASA (July 13, 2016). Matb-ii. https://matb.larc.nasa.gov/.

[119] Nasrin, F., Oballe, C., Boothe, D., & Maroulas, V. (2019). Bayesian topological learning for brain state classification. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)* (pp. 1247–1252).

[120] Neuper, C., Müller-Putz, G. R., Scherer, R., & Pfurtscheller, G. (2006). Motor imagery and eeg-based control of spelling devices and neuroprostheses. In C. Neuper & W. Klimesch (Eds.), *Event-Related Dynamics of Brain Oscillations*, volume 159 of *Progress in Brain Research* (pp. 393–409). Elsevier.

[121] Nishimoto, T., Higashi, H., Morioka, H., & Ishii, S. (2020). Eeg-based personal identification method using unsupervised feature extraction and its robustness against intra-subject variability. *Journal of Neural Engineering*, 17(2), 026007.

[122] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

[123] Pennec, X., Fillard, P., & Ayache, N. (2006). A riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1), 41–66.

[124] Perea, J. A., Deckard, A., Haase, S. B., & Harer, J. (2015). Sw1pers: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data. *BMC bioinformatics*, 16(1), 1–12.

[125] Perez Arribas, I., Goodwin, G. M., Geddes, J. R., Lyons, T., & Saunders, K. E. A. (2018). A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder. *Translational psychiatry*, 8(1), 274–274.

[126] Perez-Nieves, N., Leung, V. C. H., Dragotti, P. L., & Goodman, D. F. M. (2021). Neural heterogeneity promotes robust learning. *Nature Communications*, 12(1), 5791.

[127] Perraudin, N., Loukas, A., Grassi, F., & Vandergheynst, P. (2016). Towards stationary time-vertex signal processing.

[128] Pezard, L., Lachaux, J.-P., Thomasson, N., & Martinerie, J. (1999). Why bother to spatially embed eeg? comments on pritchard et al., psychophysiology, 33, 362–368, 1996. *Psychophysiology*, 36(4), 527–531.

[129] Pfurtscheller, G., Flotzinger, D., & Kalcher, J. (1993). Brain-computer interface—a new communication device for handicapped persons. *Journal of Microcomputer Applications*, 16(3), 293–299.

[130] Pfurtscheller, G. & Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin Neurophysiol*, 110(11), 1842–1857.

[131] Piangerelli, M., Rucco, M., Tesei, L., & Merelli, E. (2018). Topological classifier for detecting the emergence of epileptic seizures. *BMC Research Notes*, 11(1), 392.

[132] Picton, T. W. (1992). The p300 wave of the human event-related potential. *Journal of clinical neurophysiology*, 9(4), 456–479.

[133] Poincaré, H. (1952). *Science and hypothesis*. Courier Corporation.

[134] Polimeni, J. R. & Lewis, L. D. (2021). Imaging faster neural dynamics with fast fmri: A need for updated models of the hemodynamic response. *Progress in Neurobiology*, 207, 102174. How high spatiotemporal resolution fMRI can advance neuroscience.

[135] Preti, M. G. & Van De Ville, D. (2019). Decoupling of brain function from structure reveals regional behavioral specialization in humans. *Nature Communications*, 10(1), 4747.

[136] PRITCHARD, W. S., KRIEBLE, K. K., & DUKE, D. W. (1996). On the validity of estimating eeg correlation dimension from a spatial embedding. *Psychophysiology*, 33(4), 362–368.

[137] Rabadán, R., Mohamedi, Y., Rubin, U., Chu, T., Alghalith, A. N., Elliott, O., Arnés, L., Cal, S., Obaya, Á. J., Levine, A. J., & Cámara, P. G. (2020). Identification of relevant genetic alterations in cancer using topological data analysis. *Nature Communications*, 11(1), 3808.

[138] Rayleigh, B. J. W. S. (1878). *The theory of sound*. Macmillan.

[139] Reimann, M. W., Nolte, M., Scolamiero, M., Turner, K., Perin, R., Chindemi, G., Dłotko, P., Levi, R., Hess, K., & Markram, H. (2017). Cliques of neurons bound into cavities provide a missing link between structure and function. *Frontiers in Computational Neuroscience*, 11.

[140] Revel, J.-F. & Ricard, M. (2012). *Le moine et le philosophe: le bouddhisme aujourd'hui*. NiL.

[141] Rodrigues, P. L. C., Jutten, C., & Congedo, M. (2019). Riemannian procrustes analysis: Transfer learning for brain–computer interfaces. *IEEE Transactions on Biomedical Engineering*, 66(8), 2390–2401.

[142] Roy, R., Hinss, M., Darmet, L., Ladouce, S., Jahanpour, E., Somon, B., Xu, X., Drougard, N., Dehais, F., & Lotte, F. (2022). Retrospective on the first passive brain-computer interface competition on cross-session workload estimation. *Frontiers in Neuroergonomics*, 3.

[143] Roy, R. N., Bonnet, S., Charbonnier, S., & Campagne, A. (2013). Mental fatigue and working memory load estimation: interaction and implications for EEG-based passive BCI. *Annu Int Conf IEEE Eng Med Biol Soc*, 2013, 6607–6610.

[144] Roy, R. N. & Frey, J. (2016). Neurophysiological markers for passive brain–computer interfaces. *Brain–computer interfaces 1: Foundations and methods*, (pp. 85–100).

[145] Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., & Faubert, J. (2019). Deep learning-based electroencephalography analysis: a systematic review. *Journal of Neural Engineering*, 16(5), 051001.

[146] Rule, M. E., O'Leary, T., & Harvey, C. D. (2019). Causes and consequences of representational drift. *Current opinion in neurobiology*, 58, 141–147. 31569062[pmid].

[147] Rustamov, R. M. (2007). Laplace-Beltrami Eigenfunctions for Deformation Invariant Shape Representation. In A. Belyaev & M. Garland (Eds.), *Geometry Processing*: The Eurographics Association.

[148] Saggar, M., Sporns, O., Gonzalez-Castillo, J., Bandettini, P. A., Carlsson, G., Glover, G., & Reiss, A. L. (2018). Towards a new approach to reveal dynamical organization of the brain using topological data analysis. *Nature Communications*, 9(1), 1399.

[149] Saha, S. & Baumert, M. (2020). Intra- and inter-subject variability in eeg-based sensorimotor brain computer interface: A review. *Frontiers in Computational Neuroscience*, 13.

[150] Schalk, G., McFarland, D., Hinterberger, T., Birbaumer, N., & Wolpaw, J. (2004). Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Transactions on Biomedical Engineering*, 51(6), 1034–1043.

[151] Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N., & Wolpaw, J. R. (2004). Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Trans Biomed Eng*, 51(6), 1034–1043.

[152] Schoonover, C. E., Ohashi, S. N., Axel, R., & Fink, A. J. (2021). Representational drift in primary olfactory cortex. *Nature*, 594(7864), 541–546.

[153] Scrivener, C. L. & Reader, A. T. (2022). Variability of eeg electrode positions and their underlying brain regions: visualizing gel artifacts from a simultaneous eeg-fmri dataset. *Brain and Behavior*, 12(2), e2476.

[154] Seversky, L. M., Davis, S., & Berger, M. (2016). On time-series topological data analysis: New data and opportunities. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 59–67).

[155] Shadlen, M. N. & Newsome, W. T. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *Journal of Neuroscience*, 18(10), 3870–3896.

[156] Shrivakshan, G. & Chandrasekar, C. (2012). A comparison of various edge detection techniques used in image processing. *International Journal of Computer Science Issues (IJCSI)*, 9(5), 269.

[157] Sizemore, A. E., Phillips-Cremins, J. E., Ghrist, R., & Bassett, D. S. (2019). The importance of the whole: Topological data analysis for the network neuroscientist. *Network Neuroscience*, 3(3), 656–673.

[158] Stam, C. (2005). Nonlinear dynamical analysis of eeg and meg: Review of an emerging field. *Clinical Neurophysiology*, 116(10), 2266–2301.

[159] Stam, C. J., Nolte, G., & Daffertshofer, A. (2007). Phase lag index: Assessment of functional connectivity from multi channel eeg and meg with diminished bias from common sources. *Human Brain Mapping*, 28(11), 1178–1193.

[160] Stewart, I., Mandelbrot, B. B., & Robert, M. (1994). *Dieu joue-t-il aux dés?: les nouvelles mathématiques du chaos*. Flammarion.

[161] Szostakiwskyj, J. M. H., Willatt, S. E., Cortese, F., & Protzner, A. B. (2017). The modulation of eeg variability between internally- and externally-driven cognitive states varies with maturation and task performance. *PLOS ONE*, 12(7), 1–27.

[162] Takens, F. (1981). Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980* (pp. 366–381). Springer.

[163] Tangermann, M., Müller, K.-R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., Leeb, R., Mehring, C., Miller, K., Mueller-Putz, G., Nolte, G., Pfurtscheller, G., Preissl, H., Schalk, G., Schlögl, A., Vidaurre, C., Waldert, S., & Blankertz, B. (2012). Review of the bci competition iv. *Frontiers in Neuroscience*, 6.

[164] Tauzin, G., Lupo, U., Tunstall, L., Pérez, J. B., Caorsi, M., Medina-Mardones, A. M., Dassatti, A., & Hess, K. (2021). giotto-tda: A topological data analysis toolkit for machine learning and data exploration. *Journal of Machine Learning Research*, 22(39), 1–6.

[165] Tauzin, G., Lupo, U., Tunstall, L., Pérez, J. B., Caorsi, M., Medina-Mardones, A., Dassatti, A., & Hess, K. (2020). giotto-tda: A topological data analysis toolkit for machine learning and data exploration.

[166] Tolhurst, D., Movshon, J., & Dean, A. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research*, 23(8), 775–785.

[167] Tomko, G. J. & Crapper, D. R. (1974). Neuronal variability: non-stationary responses to identical visual stimuli. *Brain Research*, 79(3), 405–418.

[168] Tudor, M., Tudor, L., & Tudor, K. I. (2005). Hans berger (1873-1941)–povijest elektroencefalografije [hans berger (1873-1941)–the history of electroencephalography. *Acta medica Croatica : casopis Hravatske akademije medicinskih znanosti*, 59(4), 307–313.

[169] Umeda, Y. (2017). Time series classification via topological data analysis. *Information and Media Technologies*, 12, 228–239.

[170] Van Acker, B. B., Parmentier, D. D., Vlerick, P., & Saldien, J. (2018). Understanding mental workload: from a clarifying concept analysis toward an implementable framework. *Cognition, Technology & Work*, 20(3), 351–365.

[171] Veltman, J. A. & Gaillard, A. W. K. (1993). Indices of mental workload in a complex task environment. *Neuropsychobiology*, 28(1-2), 72–75.

[172] Vinck, M., Oostenveld, R., van Wingerden, M., Battaglia, F., & Pennartz, C. M. (2011). An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *NeuroImage*, 55(4), 1548–1565.

[173] Wang, Y., Behroozmand, R., Johnson, L. P., Bonilha, L., & Fridriksson, J. (2020a). Topological signal processing in neuroimaging studies. In *2020 IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops)* (pp. 1–4).

[174] Wang, Y., Behroozmand, R., Johnson, L. P., & Fridriksson, J. (2020b). Topology highlights neural deficits of post-stroke aphasia patients. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* (pp. 754–757).

[175] Wang, Y., Ombao, H., & Chung, M. K. (2018). Topological data analysis of single-trial electroencephalographic signals. *The annals of applied statistics*, 12(3), 1506–1534.

[176] Wang, Y., Ombao, H., & Chung, M. K. (2019). Statistical persistent homology of brain signals. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1125–1129).

[177] Wang, Z., Yan, W., & Oates, T. (2016). Time series classification from scratch with deep neural networks: A strong baseline.

[178] Willett, F. R., Avansino, D. T., Hochberg, L. R., Henderson, J. M., & Shenoy, K. V. (2021). High-performance brain-to-text communication via handwriting. *Nature*, 593(7858), 249–254.

[179] Wolpaw, J., Birbaumer, N., Heetderks, W., McFarland, D., Peckham, P., Schalk, G., Donchin, E., Quatrano, L., Robinson, C., & Vaughan, T. (2000). Brain-computer interface technology: a review of the first international meeting. *IEEE Transactions on Rehabilitation Engineering*, 8(2), 164–173.

[180] Wolpaw, J. & Wolpaw, E. W. (2012). *Brain–Computer Interfaces: Principles and Practice*. Oxford University Press.

[181] Wu, D., Xu, Y., & Lu, B.-L. (2020). Transfer learning for eeg-based brain-computer interfaces: A review of progress made since 2016.

[182] Xu, G. (2004). Convergence of discrete laplace-beltrami operators over surfaces. *Comput Math Appl*, 48(3-4), 347–360.

[183] Xu, G. (2006). Discrete laplace–beltrami operator on sphere and optimal spherical triangulations. *Int J Comput Geom & Appl*, 16(01), 75–93.

[184] Xu, X., Drougard, N., & Roy, R. N. (2021). Dimensionality reduction via the laplace-beltrami operator: Application to eeg-based bci. In *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)* (pp. 347–350).

[185] Yamanashi, T., Kajitani, M., Iwata, M., Crutchley, K. J., Marra, P., Malicoat, J. R., Williams, J. C., Leyden, L. R., Long, H., Lo, D., Schacher, C. J., Hiraoka, K., Tsunoda, T., Kobayashi, K., Ikai, Y., Kaneko, K., Umeda, Y., Kadooka, Y., & Shinozaki, G. (2021). Topological data analysis (tda) enhances bispectral eeg (bseeg) algorithm for detection of delirium. *Scientific Reports*, 11(1), 304.

[186] Yan, Y., Wu, X., Li, C., He, Y., Zhang, Z., Li, H., Li, A., & Wang, L. (2022). Topological eeg nonlinear dynamics analysis for emotion recognition. *IEEE Transactions on Cognitive and Developmental Systems*, (pp. 1–1).

[187] Yang, W., Jin, L., & Liu, M. (2016). Deepwriterid: An end-to-end online text-independent writer identification system. *IEEE Intelligent Systems*, 31(2), 45–53.

[188] Zander, T. O. & Kothe, C. (2011). Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *J Neural Eng*, 8(2), 025005.

[189] Zelditch, S. (2000). Spectral determination of analytic bi-axisymmetric plane domains. *Geometric & Functional Analysis GAFA*, 10(3), 628–677.

[190] Zhou, S., Masmanidis, S. C., & Buonomano, D. V. (2020). Neural sequences as an optimal dynamical regime for the readout of time. *Neuron*, 108(4), 651–658.e5.

[191] Ziv, Y., Burns, L. D., Cocker, E. D., Hamel, E. O., Ghosh, K. K., Kitch, L. J., Gamal, A. E., & Schnitzer, M. J. (2013). Long-term dynamics of ca1 hippocampal place codes. *Nature neuroscience*, 16(3), 264–266.