

POLITECNICO DI MILANO

Master of Science in Mathematical Engineering



Persistent homology and fractal  
dimension for the detection of Sleep  
Stages and K-complexes in EEGs

Supervisor: Prof. Paolo Lella

Co-examiner: Prof. Francesco Vaccarino

Stefano Spaziani

Matricola 921351

Academic year 2020-2021



# Ringraziamenti

Assurdo, ho sempre letto i ringraziamenti nelle tesi degli altri (in quanto contengono le cose piu' divertenti) e non mi sembra vero che sia arrivato il mio turno. Finire questo percorso al Poli genera tante emozioni. I ricordi piu' belli sono anche legati ai suoi luoghi: le serate in Piazza Leo, le lezioni di teatro in Nave, i tornei di basket con i Patatinaikos, i Comecichiamiamo o le Forme Deboli, in quel campo irregolarissimo nella Casa dello Studente. In questi 5 anni non siamo riusciti a portare il corso di Inge Mate in cima al podio. Peccato, perché eravamo carichi e pieni d'esperienza per quest'ultimo anno: senza il Covid ce l'avremmo SICURAMENTE fatta. Tutte queste esperienze si ripeteranno, in un modo o nell'altro. Finché avrò con me il tesserino, anche scaduto, so che il Poli sarà pronto ad accogliermi.

Questi luoghi sono stati riempiti da un sacco di persone che sono davvero grato di aver incontrato. Ringrazio i professori del dipartimento di Matematica. Siete davvero molto forti in mate e generalmente bravissimi insegnanti. Non posso davvero lamentarmi del livello della didattica in questi anni. Ringrazio particolarmente il prof. Grasselli, la prof. Sangalli e il mio relatore, il dott. Paolo Lella. Con questi professori sono riuscito ad avere un rapporto piu' personale, ed in particolare sono stati sempre disponibili ad ascoltare i miei dubbi e a darmi consigli e spiegazioni sugli studi e sul futuro. Ringrazio i miei compagni di corso: mi dispiace di non essere riuscito a conoscervi come col senno del poi avrei voluto. Avete (abbiamo) creato un ambiente incredibilmente motivante, anche per le classi future.

Vorrei inoltre salutare e ringraziare i miei insegnanti del liceo, in particolare il prof. Cavalli e la prof. Joannes. Avete destato in me la curiosità e la voglia di studiare; io e gli altri miei compagni vi ricorder-

emo sempre.

Ringrazio un casino Bruno, Benaz, VidueMarti: i Polibresciani. E' un peccato che non ci vediamo da un po' (rimedieremo!), ma sono contento di aver passato gli studi della triennale con voi. E' stato brutto quando abbiamo iniziato a scegliere esami diversi, mi piaceva troppo seguire le lezioni con voi che non portavate mai NESSUNA distrazione in classe. Grazie a tutti i gruppi di Brescia con cui sono riuscito a condividere studio, campetti, uscite, viaggi. Saluto tutti i gnari di Assistenza Senescenti, le SGAMMA, gli amici delle Partitelle Audaci, Abba. L'estate ci farà incontrare sicuramente.

E' il momento di ringraziare i compagni del Teatro delle Biglie. Essere una Biglia mi ha cambiato in meglio, vi voglio troppo bene. Sono felicissimo di aver trovato persone con questo tipo di interessi in un luogo come il Poli.

Grazie a tutto il mondo legato a Casa Sassi, per avermi ospitato (e sopportato) quelle due o tre volte.

Come è giusto che sia, ringrazio il gruppo dei Come è giusto che sia. Tuttavia, vi esorto a smettere di paccare, grazie.

Grazie a tutto il gruppo dei Castegnatesi. Purtroppo sarà troppo tardi quando il mondo si accorgerà che a Castegnato è nata una corrente musicale fortissima, ma non importa.

Ora è il momento della *gang*, o la *cumpa* insomma. Grazie alle LuGnare, quindi Ro, Meri, Marta, Ste, GiuTof per tutte le uscite da Iaio, i giochi, le Feste Radio, i viaggi, le letture del futuro, le tenzoni interminabili! A stare con voi ci si sente a casa, quindi grazie davvero per questo legame. Infine ciao rega delle Carriole, quindi Nino, Ballo, Picchio, Dario, Tomma e Misha. Sono cresciuto con voi e spero di continuare a farlo.

Azie Ico ♡

Infine ringrazio di cuore la family. Mamma, papà, questa tesi la dedico a voi che mi avete dato tutto e mi avete spronato a impegnarmi al massimo.

Grazie Sarì, Roby e Alex per le ore infinite di gioco assieme passate e future.

E poi ciao anche a Giugi, Clacla e a tutte le zie e zii, cugini e cugine.

Grazie per esserci sempre stati!

E' stato bello scrivere questi ringraziamenti, anche se sono noiosi e ripetitivi, perché mi sono accorto di quante bellissime persone ho conosciuto e ho accanto. Spero di non aver dimenticato nessuno. Buona lettura!



# Sommario

Il contributo principale di questo lavoro è lo sviluppo un algoritmo di rilevamento automatico delle fasi del sonno e dei complessi K a partire da segnali cerebrali collezionati tramite elettroencefalogramma (EEG). La classificazione delle polisonnografie è un compito oneroso che viene eseguito a occhio ancora oggi da molti medici e ricercatori. I complessi K sono forme d'onda presenti nella fase 2 di sonno non-REM , e che si caratterizzano per una primo stadio negativo seguita da un secondo stadio di rapida risalita, ad alta tensione. La densità di complessi K durante il sonno potrebbe essere collegata all'insorgenza di malattie neurodegenerative.

Il rilevamento viene effettuato tramite un metodo di supervised learning; nello specifico, sono state implementate due foreste casuali allenate con features provenienti dall'omologia persistente per la classificazione dei complessi K, e dalla teoria dei sistemi dinamici caotici per la classificazione delle fasi del sonno. Il teorema di Takens permette di recuperare la topologia dello spazio delle fasi generato dai segnali tramite l'immersione conosciuta come Sliding Window. I segnali, raccolti nella forma di serie temporali a tempo discreto, vengono trasformati in nuvole di punti. La tesi si propone di presentare alcuni risultati fondamentali di topologia algebrica computazionale che permettono il calcolo ottimizzato delle classi di omologia persistente da una nuvola di punti. Inoltre vengono esposte le basi della teoria dei sistemi dinamici caotici e discussa la nozione di dimensione frattale, utile per il riconoscimento delle fasi del sonno.

L'estrazione delle features viene eseguita studiando le classi di omologia persistente e la caoticità degli spazi topologici ottenuti come immagini di queste immersioni. I risultati numerici ottenuti su di un dataset

composto da varie polisonnografie fornite dall'ospedale San Raffaele di Milano sono paragonabili con quelli della letteratura.







# Abstract

The main contribution of this work is the development of an algorithm for the automatic detection of sleep stages and K-complexes from brain signals collected with electroencephalograms (EEGs). The classification of polysomnographies is an onerous operation that is still performed at sight by physicians and researchers. K-complexes are waveforms that appear in the NREM-2 sleep stage, and are characterized by an initial negative phase followed by a rapid raising phase, with high voltage. The density of K-complexes during sleep could be connected with the onset of neurodegenerative diseases.

The detection is carried out through a supervised learning method; specifically, a random forest scheme. The classifiers have been trained with features from persistent homology and chaotic dynamical systems for the detection of K-complexes and sleep stages, respectively. Takens' Theorem permits the recovery of the topology of the state space generated by the signals via the embedding known as Sliding Window. The signals, that are collected in the form of discrete time series, are embedded into point clouds. The dissertation presents some fundamental results of computational algebraic topology on the optimized computation of the persistent homology classes from a point cloud. In addition, the basic definitions of chaotic dynamical systems are introduced together with the notion of fractal dimension; the latter is useful in the detection of sleep stages.

Feature extraction is performed by studying the persistent homology classes and the chaoticness of the topological spaces which are obtained as images of these embeddings. The numerical results are obtained over a dataset of multiple polysomnographies provided by San Raffaele Hospital of Milan. The precision of the methods are comparable with the

ones present in the scientific literature.





# Contents

<b>Sommario</b>	<b>VII</b>
<b>Abstract</b>	<b>XI</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Why TDA . . . . .	3
1.2 Structure of the dissertation . . . . .	5
<b>2 Preliminary Notions</b>	<b>7</b>
2.1 Shapes for an algebraic topologist . . . . .	7
2.2 Point clouds and manifolds . . . . .	10
2.3 Some concepts of dynamical systems and chaos . . . . .	11
<b>3 Simplicial Complexes</b>	<b>15</b>
3.1 Simplices . . . . .	16
3.2 Simplicial Complexes . . . . .	17
3.3 Abstract Simplicial Complexes . . . . .	19
3.4 Useful Simplicial Complexes . . . . .	21
<b>4 Simplicial Homology</b>	<b>25</b>
4.1 Chain Complexes . . . . .	26
4.1.1 The boundary operator . . . . .	28
4.2 Simplicial Homology Groups . . . . .	30
4.3 Topological invariance of the homology groups . . . . .	33
4.4 Computation of simplicial homology groups via Matrix Reduction . . . . .	35
<b>5 Persistence</b>	<b>39</b>
5.1 Persistence Diagrams . . . . .	43
5.2 Stability of Persistence Diagrams . . . . .	47

<b>6</b>	<b>Sliding Window embedding</b>	<b>49</b>
6.1	Dynamical systems: delay coordinates and Takens' Theorem . . . . .	50
6.2	Delay embedding application in persistent homology . . . . .	52
6.2.1	Basic properties . . . . .	52
6.2.2	Examples . . . . .	54
<b>7</b>	<b>Automatic recognition of Sleep Phase and K-complexes</b>	<b>59</b>
7.1	Sleep and Sleep Stages . . . . .	59
7.1.1	Automatic classification of sleep stages: state of the art . . . . .	62
7.2	K-complexes (KCs) . . . . .	63
7.2.1	Automatic detection of K-complexes: state of the art . . . . .	64
7.3	From Polysomnography to Point Clouds . . . . .	64
7.3.1	Embedding parameters selection . . . . .	65
7.4	Embedding of Epoques . . . . .	68
7.5	Embedding of Waveforms . . . . .	70
<b>8</b>	<b>Methodology</b>	<b>75</b>
8.1	Main concepts of supervised Machine Learning . . . . .	75
8.2	Minimum Description Length . . . . .	77
8.3	Decision Trees . . . . .	80
8.4	Random Forests . . . . .	85
<b>9</b>	<b>Feature extraction and numerical results</b>	<b>89</b>
9.1	Feature extraction for sleep stage detection: chaotic dynamical systems . . . . .	89
9.1.1	Fractal dimension . . . . .	90
9.1.2	Minimum embedding dimension through through False Nearest Strand . . . . .	92
9.2	Feature extraxtion for K-complex classification: persistent homology . . . . .	93
9.2.1	Landscapes and silhouettes . . . . .	93
9.2.2	Euler characteristic . . . . .	94
9.3	Numerical Results . . . . .	96
9.3.1	Detection of sleep stages . . . . .	98
9.3.2	Detection of K-complexes . . . . .	101



<b>10 Conclusions</b>	<b>105</b>
<b>Bibliography</b>	<b>107</b>



# List of Figures

2.1	The torus (left) and the coffee mug (right) are topologically equivalent. . . . .	7
2.2	Example: the Möbius strip (left) and an untwisted band (right) are homotopically equivalent, but not homeomorphic. . . . .	9
2.3	Left: a torus-shaped point cloud. Right: a sphere-shaped point cloud. . . . .	11
2.4	the Lorentz system . . . . .	11
2.5	Left: a stationary point. Right: a transient converging into a limit cycle. . . . .	14
3.1	simplices . . . . .	16
3.2	Example: the 3-simplex is homeomorphic to a ball in $\mathbb{R}^3$ . The boundary of the 3-simplex is homeomorphic to the spherical surface. . . . .	17
3.3	Example: the object to the left is a simplicial complex, the one to the right is not. . . . .	17
3.4	From left to right: the 2,1,0-skeleton of a simplicial complex of dimension 2. . . . .	18
3.5	Example: an abstract simplicial complex of dimension 2. . . . .	20
3.6	Geometrical realization of an abstract simplicial complex. . . . .	20
3.7	Example: geometrical realization of the nerve of an open cover. . . . .	21
3.8	Example: Čech complex of a finite set of points. . . . .	22
3.9	Example: Vietoris-Rips complex of a finite set of points. . . . .	22
3.10	A triangulated dolphin shape. . . . .	24
4.1	Example: a non-oriented 1-simplex (left) and its corresponding oriented versions (centre, right). Notice that there is a difference in the notation. . . . .	26

4.2	Example: a non-oriented 2-simplex (left) and its corresponding oriented versions (centre, right). Notice that there is a difference in the notation. . . . .	26
4.3	Example: group of $p$ -chains for a simplicial complex, with coefficients in $\mathbb{Z}$ . . . . .	27
4.4	Example: boundary homomorphism applied to a 1 dimensional simplex. . . . .	29
4.5	Example: boundary homomorphism applied to a 2 dimensional simplex. . . . .	29
4.6	Example: boundary and cycle. . . . .	30
4.7	Images of the boundary operator: every chain gets mapped to a boundary; every cycle and boundary gets mapped to 0. . . . .	31
4.8	Example: The homology groups of the space to the left can be obtained by computing the simplicial homology groups of the complex to the right. As there are four connected components and two 1-dimensional holes, the 0-th dimensional homology group is isomorphic to $\mathbb{Z}^4$ and the 1-st dimensional homology group is isomorphic to $\mathbb{Z}^2$ . . . . .	32
4.9	Structure of a matrix in Smith normal form . . . . .	37
5.1	Example-1: element of a Čech filtration. The point clouds suggests a shape with one connected component, one small hole and one big hole. At this stage of the filtration, only 13 0-dimensional homology classes (connected components are alive). . . . .	40
5.2	Example-2: as the radius of the Čech complex grows, all connected components die apart from one. Two one dimensional homology classes (the two holes) are born. . . . .	41
5.3	Example-3: suppose this is the last element of the filtration. We remain with a 0-th dimensional homology class. One of the two classes representing the holes has died. This suggests that the shape behind the point cloud has 1 connected component, 1 small hole and 1 big hole. . . . .	41
5.4	Example: the homology class $\gamma$ is born in the complex $K_i$ of the filtration. Its image $f_p^{i,j}(\gamma)$ persists in the complex $K_{j-1}$ , and finally $f_p^{j-1,j}(\gamma)$ dies in the complex $K_{j-1}$ . The persistence index of $\gamma$ is $j - i$ . . . . .	43

5.5	Example: a juxtaposition of persistence diagrams of dimension 0, 1 and 2. The black dots represent the 0-th dimensional classes. The red triangles and the blue rhombuses constitute 1 and 2 dimensional classes, respectively. . . . .	44
6.1	Left: $\sin(t)$ . Center: 3-dimensional embedding with function $G(x) = \sin(10x)$ . Right: 2-dimensional delay coordinates. . . . .	51
6.2	Example: 3-dimensional embedding of a filtered signal. The band-pass filter keeps only frequencies between 0.5 and 2.3 Hz. On the right we have the corresponding persistence diagram. . . . .	54
6.3	Example: 3-dimensional embedding of a filtered signal. The band-pass filter keeps only frequencies between 0.5 and 13 Hz. On the right we have the corresponding persistence diagram. . . . .	55
6.4	Example: $f(t) = \cos(0.01t) \cdot e^{\frac{-t}{1000}}$ ; recovery of a stationary point. . . . .	55
6.5	Example: $f(t) = \cos(0.01t)$ ; recovery of a limit cycle. . . . .	56
6.6	Example: $f(t) = 2\sin(0.01t) + 1.8\sin(0.01\sqrt{3}t)$ ; recovery of a limit torus. . . . .	57
6.7	Example: persistence diagram of the reconstructed phase space of Figure 6.6. The two 1-dimensional and the single 2-dimensional high persistence classes suggest that the attractor is a torus. . . . .	57
6.8	Example: recovery of a strange attractor (the Lorentz system). . . . .	57
7.1	EEG signal during stage 1. . . . .	60
7.2	EEG signal during stage 2. . . . .	60
7.3	EEG signal during stage 3. . . . .	61
7.4	EEG signal during REM stage. . . . .	61
7.5	EEG signal during wakefulness. . . . .	62
7.6	A K complex. . . . .	63
7.7	A Wakefulness and the corresponding 3-dimensional embedding . . . . .	68

7.8	A Stage NREM1 epoque and the corresponding 3-dimensional embedding. . . . .	68
7.9	A Stage NREM2 epoque and the corresponding 3-dimensional embedding. . . . .	69
7.10	A Stage NREM3 epoque and the corresponding 3-dimensional embedding. . . . .	69
7.11	A REM stage epoque and the corresponding 3-dimensional embedding. . . . .	69
7.12	An epoque with a very sharp K-complex and the corresponding 3-dimensional embedding. . . . .	69
7.13	K1: example of a K-complex waveform (left), its three dimensional embeddting (center) and the corresponding persistence diagram (right.) . . . . .	71
7.14	K2: example of a K-complex waveform (left), its three dimensional embeddting (center) and the corresponding persistence diagram (right.) . . . . .	71
7.15	K3: example of a K-complex waveform (left), its three dimensional embeddting (center) and the corresponding persistence diagram (right.) . . . . .	71
7.16	K4: example of a K-complex waveform (left), its three dimensional embeddting (center) and the corresponding persistence diagram (right.) . . . . .	72
7.17	non-K1: example of a non-K-complex waveform (left), its three dimensional embeddting (center) and the corresponding persistence diagram (right.) . . . . .	72
7.18	non-K2: example of a K-complex waveform (left), its three dimensional embeddting (center) and the corresponding persistence diagram (right.) . . . . .	72
7.19	non-K3: example of a non-K-complex waveform (left), its three dimensional embeddting (center) and the corresponding persistence diagram (right.) . . . . .	73
7.20	non-K4: example of a non-K-complex waveform (left), its three dimensional embeddting (center) and the corresponding persistence diagram (right.) . . . . .	73
8.1	Example: a simple binary classification decision tree with binary instances. . . . .	80
8.2	Scheme of the random forest classifier. . . . .	85

9.1	Example: the Hausdorff measure of a portion of surface is approximated by the sum of median section of small spheres (with diameter less than $\delta$ ) that provide a covering of the surface. . . . .	91
9.2	Example: the Sierpinski gasket is a chaotic attractor $G$ with $\dim_{\mathfrak{H}}(G) = \ln 3 / \ln 2$ . . . . .	91
9.3	Example: the Koch snowflake is a chaotic attractor $G$ with $\dim_{\mathfrak{H}}(G) = \ln 4 / \ln 3$ . . . . .	91
9.4	Example: the Julia sets, which are the basin of attraction for systems of the type $z \mapsto z^2 + c \in \mathbb{C}$ , have a fractal dimension that satisfies $1 + C^{-1}\sqrt{c+2} \leq \dim_{\mathfrak{H}}(J) \leq 1 + C \log(c+2) ^{3/2}\sqrt{c+2}$ for some constant $C > 1$ and suitable $c$ . [18]. . . . .	92
9.5	Example: The first persistence landscape of dimension 1, $\lambda_D(1, t)$ , for a K-complex (left) and for a non-K-complex (right). K complexes are characterized by less classes with low persistence and more classes with high persistence. . . . .	94
9.6	Example: The weighted silhouette of dimension 1 and power 1, $\phi^1(t)$ , for a K-complex (left) and for a non-K-complex (right). K complexes are characterized by less classes with low persistence and more classes with high persistence. . . . .	94
9.7	$F_3$ : box-counting dimension for the different sleep stages. . . . .	99
9.8	$F_2$ : critical value of mutual information function for the different sleep stages. . . . .	99
9.9	$F_2$ : the number of 1-dimensional persistent homology classes is lower for K-complexes. . . . .	102
9.10	$F_6$ : the quotient between the highest 1-dimensional persistent class and the third one is higher for K-complexes. . . . .	103





# List of Tables

4.1	Homology groups of simple topological spaces. . . . .	33
5.1	Persistence diagrams of some point clouds. Notice the consistency with the homology groups of Table 4.2. . .	45
9.1	For each polysomnography we present the total number of 30 seconds epoques, partitioned into the 5 different Sleep Stages. In addition we display the number of K complexes (KCs) appearing during the night. . . . .	96
9.2	Detection of sleep stages. . . . .	99
9.3	Detection of state of sleep. . . . .	100
9.4	Detection of K complexes. . . . .	104



# Chapter 1

## Introduction

### 1.1 Why TDA

Computational topology and topological data analysis (TDA) are recently inaugurated fields of applied mathematics. The former is a subfield of topology which makes extensive use of notions from topics as computer science and geometry, among others. Two cornerstone texts that led to the foundation of the subject in the early 2000s have been written by Edelsbrunner and Harer [21] and by Zomorodian and Carlsson [65]. One of the objectives of computational topology is the efficient calculation and storing of topological features for opportunely discretized shapes. Such features include connected sum decomposition of manifolds, knots, links, homotopy and homology groups. Treating algebraic objects computationally can be arduous in general. This dissertation assesses the simpler case of simplicial homology groups, which can be obtained through linear algebra. These tools can be employed in multiple fields as numerical analysis or computational geometry. In addition, as it will be underlined in this work, there is a strong connection between topology, geometry and dynamical systems theory [57].

Topological data analysis aims at applying the results of computational topology in the investigation of the shape of data. Data often comes in the form of a signal, or a point cloud. TDA offers the possibility to associate to data topological, geometrical and algebraic quantities that are not accessible from other techniques. Two fundamental tools used in applications are the *mapper algorithm*, which summarizes shapes into graphs, and *persistent homology*, that permits multiscale

topological analysis of point clouds [10]. The literature shows that both of these techniques are suitable for the qualitative inspection of data. In addition to that, TDA points at providing precise quantitative results. For that to be possible, it is necessary to combine the computational topology framework with a robust statistical theory. Computational topology ensures that theoretical concepts as shapes, holes and knots among other objects can be conveniently modeled and stored in computers. Data, on the other hand, presents noise, missing values, outliers, and is represented as a random variable coming from a probability distribution. Summarizing, TDA is computational topology endowed with mathematical structures that open up the possibility to compute stable means, variance, errors, distance functions, on topologically derived objects [39, 9, 11].

TDA asserted itself as a valid scheme to study possibly complex data. Some practical applications of TDA are the study of cities or spider web networks [23, 8], 3D objects recognition [55], and in biology, the inspection of RNA sequences and proteins [50, 17]. Another field of application that we want to underline, as it is the one covered in this dissertation, is signal analysis [62, 48].

The computed topological quantities are often employed in machine learning algorithms, making TDA a new and reliable source of features for both regression and classification tasks [49].

This dissertation has three different aims. The first objective is somewhat didactic: we hope that reading this text can be useful to understand the theories behind simplicial homology and chaotic dynamical systems, without entering too much into the details but giving to the reader all the references in which these topics are covered with higher formalism and competence. The other two goals are of greater importance since they distinguish this work from the others. We desire to fascinate the reader by showing that topology is intrinsically present in signals we collect in real life.

As presented in this work, brain waves can be modeled as dynamical systems endowed with highly complex topological features, as high dimensional nontrivial homology groups and chaotic sets of fractal dimension. Thus, a great amount of information can be extrapolated from a scalar function. For the task of feature extraction, we assert

that for some applications inspecting the topological aspect can lead to high or even better precision in classification tasks. Specifically, for EEG signals, the frequency spectrum or the precise value of a time series in each instant can be of less interest since these quantities depend on the use of filters and the position of the electrodes. For the detection of Sleep Stages a smart idea could be to seek for the *complexity* of the trajectories described by the signal; for the detection of K-complexes a useful feature can be given by the *holes* in the phase space. These are topological concepts.

Finally, there is the interest in the classification tasks. Other than the personal passion in sleep functioning and brain waves, the scientific literature proposes highly-functioning methods and this work aims at finding comparable or even better results. The methodology in this thesis could be shared with the Research Center for Sleep of San Raffaele Hospital, in Milan, and employed both for research purposes and as a routine methodology, with the hope of reaching gains in classification precision and time savings. In particular the K-complexes waveforms have been reported to decrease in density in Alzheimer's disease (AD) patients [38]; it would be a great fulfilment to be a part of this research field.

## 1.2 Structure of the dissertation

This work is structured in the following way. Chapter 1 introduces the motivation and the description of the topics. Chapter 2 defines basic mathematical objects (groups, embeddings, dynamical systems) that reappear throughout the whole thesis. Chapter 3 is devoted to the presentation of simplicies and simplicial complexes, which are fundamental notions in the field of computational topology. Simplicial homology, its interpretations and properties are defined in Chapter 4. Chapter 5 describes persistent homology, the main feature-extracting tool for the detection of K-complexes in EEGs. In Chapter 6 the topics of signal analysis, simplicial homology, and dynamical systems are connected by the Takens' embedding theorem. Chapter 7 interrupts the theoretical aspects of the dissertation and delineates the classification problems which are assessed by exploiting machine learning algorithms. The two problems are the detection of Sleep Stages and K-complex waveforms in polysomnographies. Chapter 8 introduces to the statistical learn-

ing scheme, and in particular, to decision trees and random forests. Chapter 9 discusses feature selection for the two classification tasks: features are chosen from chaotic dynamical systems and persistent homology theories to assess the problems of detection of Sleep Stages and K-complexes, respectively. The same Chapter presents the numerical results. Finally, Chapter 10 contains a brief conclusive discussion with some hypotheses for future developments.

## Chapter 2

# Preliminary Notions

### 2.1 Shapes for an algebraic topologist

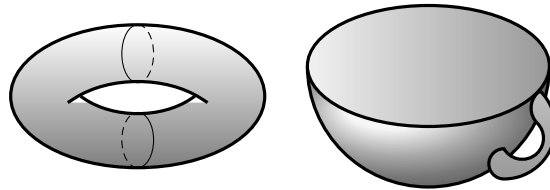


Figure 2.1: The torus (left) and the coffee mug (right) are topologically equivalent.

One of the main ideas of algebraic topology is to consider two spaces to be equivalent if they have "the same shape". The first section of this Chapter is devoted in presenting the basic intuitions and fundamental definitions behind the concepts of simplicial homology and homotopy. These notions are deeply correlated and are used to measure the shape of a topological space. The following statements are useful to get an informal description of the two mathematical objects.

***Simplicial homology** is the sequence of homology groups of a particular type of set - the simplicial complex. It formalizes the idea of the number of **holes** of a given dimension in the complex.*

***Homotopy equivalent** topological shapes can be transformed into one another by bending, shrinking and expanding operations.*

Simplicial homology is endowed with an algebraic structure: for this reason is necessary to introduce the basic concept of free abelian group.

**Definition 1** (Free abelian group). A group  $(G, +)$  is a set  $G$  together with a binary operator on  $G$ , here denoted on  $\cdot$ , with the following properties:

- *Associativity*: for all  $a, b, c \in G$  it holds  $(a + b) + c = a + (b + c)$ ;
- *Identity element*: there exists a unique element  $e$  in  $G$  such that, for every  $a \in G$ , we have  $e + a = a$  and  $a + e = a$ ;
- *Inverse element*: for each  $a \in G$ , there exists an element  $b \in G$  such that  $a + b = e$  and  $b + a = e$ , where  $e$  is the identity element.

A *free abelian group* is an abelian group with a basis. Being an abelian group means that it is a set with an addition operation that is associative, commutative, and invertible. A basis is a subset such that every element of the group can be uniquely expressed as a linear combination of basis elements with integer coefficients.

Free abelian groups have properties which make them similar to vector spaces. Now we introduce homomorphisms, that are structure-preserving maps between groups.

**Definition 2** (Group homomorphism). A *homomorphism* from a group  $(G, +_G)$  to a group  $(H, +_H)$  is a function  $\phi : G \rightarrow H$  such that for all elements  $a, b \in G$   $\phi(a +_G b) = \phi(a) +_H \phi(b)$ .

As we will see in the next Chapters, simplicial homology groups count the holes and the connected components in a topological space. Since both the torus and the coffee mug have 1 2-dimensional hole, 2 1-dimensional holes, and 1 connected component, their simplicial homology groups are the same.

We now proceed and present homotopies and homeomorphisms [28], [21].

**Definition 3** (Homotopy). A homotopy between two continuous function  $f$  and  $g$  from a topological space  $X$  to a topological space  $Y$  is defined to be a continuous function  $H : X \times [0, 1] \rightarrow Y$  such that  $H(x, 0) = f(x)$  and  $H(x, 1) = g(x)$  for all  $x \in X$ . Given two topological spaces  $X$  and  $Y$ , a *homotopy equivalence* between  $X$  and  $Y$  is a pair of continuous maps  $f : X \rightarrow Y$  and  $g : Y \rightarrow X$ , such that  $g \circ f$



is homotopic to the identity map  $id_X$  and  $f \circ g$  is homotopic to  $id_Y$ . If such a pair exists, then  $X$  and  $Y$  are said to have the same *homotopy type*. Finally,  $g$  is said to be a *homotopy inverse* to  $f$ .

**Definition 4** (Homeomorphism). A *homeomorphism* is a continuous function between topological spaces that has a continuous inverse function.

Two spaces with a homeomorphism between them are called homeomorphic, and from a topological viewpoint they are the same. A homeomorphism is a special case of a homotopy equivalence, in which  $g \circ f$  is **equal** to the identity map  $id_X$  and  $f \circ g$  is equal to  $id_Y$ . Therefore, if  $X$  and  $Y$  are homeomorphic then they are homotopy equivalent, but the opposite is not true.

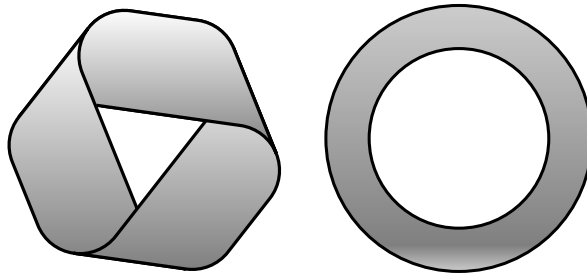


Figure 2.2: Example: the Möbius strip (left) and an untwisted band (right) are homotopically equivalent, but not homeomorphic.

**Definition 5** (Isomorphism). An *isomorphism* is a structure-preserving mapping between two structures of the same type that can be reversed by an inverse mapping.

- An *isometry* is an isomorphism of metric spaces;
- An *homeomorphism* is an isomorphism of topological spaces;
- A group *homomorphism* is an isomorphism iff it is bijective.

In the following sections we will see that if two topological spaces are homotopically equivalent, then they have isomorphic homology groups.

At this point we cover the foundational notion for the understanding and the practical applications of this thesis: embeddings. Embeddings

are structure-preserving maps. We will see two definitions of embedding: the topological one here, and one adapted for dynamical systems in the next Chapters. Embeddings will be used in this work to map time series into (higher dimensional) point clouds.

**Definition 6** (Embedding). In topology, an *embedding* is an homeomorphism onto its image. Specifically, an injective continuous map  $f : X \rightarrow Y$  between topological spaces is an embedding if  $X$  and  $f(X)$  (considered with the subspace topology inherited by  $Y$ ) are homeomorphic.

## 2.2 Point clouds and manifolds

In the field of topological data analysis, collection of data can be seen as the evaluation of the **relative** or **collective position** of the observed quantities, for example in an euclidean space. Often, once data collection is carried out, a theoretical object as a topological space is not available; instead we are given a discrete approximation of such a space in the form of a point cloud.

**Definition 7** (Point cloud). A *point cloud* is a set of data points in  $\mathbb{R}^n$ , with  $n > 0$ .

Now, it is crucial to understand the many interpretations of a point cloud in an informal way, since it is the conceptual core of this thesis. We state again the first of three ways to look at a point cloud:

**1:** *point cloud = **data**.*

A second possibility is to look at all the points of a point cloud altogether and construct some sort of shape around them; a manifold, for instance.

**Definition 8** (Manifold). A *topological manifold* is a second countable Hausdorff space that is locally homeomorphic to an Euclidean space.

**2:** *under suitable conditions, a point cloud can be seen as the vertex set of a **simplicial manifold**.*

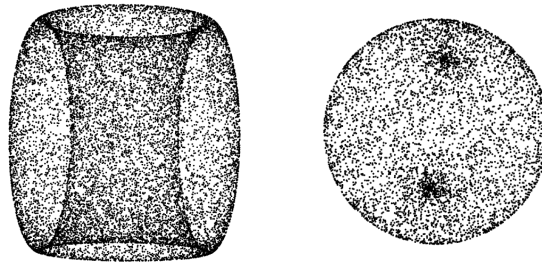


Figure 2.3: Left: a torus-shaped point cloud. Right: a sphere-shaped point cloud.

Manifolds are useful for two reasons, the first being that as topological spaces, they can be studied by themselves with the instruments of algebraic topology. The second, more subtle reason comes from the fact that manifolds can be seen as a domain of functions whose output is the observed data.

**3:** *a point cloud can be seen as the evolution of a discrete-time dynamical system in an appropriate state-space.*

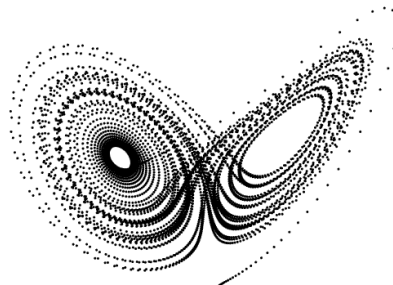


Figure 2.4: the Lorenz system

We will see how these three interpretations are interconnected and can help to solve real-life problems as the classification of signals.

## 2.3 Some concepts of dynamical systems and chaos

In Chapter 3 and Chapter 4 will show how computational algebraic topology is suited to describe data altogether, and to count or quan-

tify the entity of the topological features of a set. On the other set, dynamical systems theory is an intriguing tool to study the geometry of a point cloud, this time viewed as a trajectory in an euclidean space. We topic is introduced through some basic definitions.

**Definition 9** (Phase space). In the theory of dynamical systems, the *phase space* is a space in which all possible states of a system are represented, with each possible state corresponding to one unique point in the phase space.

**Definition 10** (Dynamical system, orbit). A *dynamical system* is a manifold  $M$  (the phase space) endowed with a family of smooth functions  $\phi_t : M \rightarrow M$  called *evolution functions*. Calling  $T$  the set of all possible  $t$ , we can write a dynamical system as a triplet  $(T, \mathcal{M}, \phi)$ .

An *orbit* is a collection of points related to a single evolution function.

An orbit can be understood as the subset of the phase space covered by the trajectory of the dynamical system under a particular set of initial conditions.

Time, denoted here with  $t$ , can be either a continuous or a discrete integer-valued variable. An archetypal example of continuous time dynamical system is a system of  $N$  first-order, autonomous, ordinary differential equations

$$\begin{aligned} \frac{d\phi^{(1)}}{dt} &= F_1(\phi^{(1)}, \phi^{(2)}, \dots, \phi^{(N)}), \\ \frac{d\phi^{(2)}}{dt} &= F_2(\phi^{(1)}, \phi^{(2)}, \dots, \phi^{(N)}), \\ &\vdots \\ \frac{d\phi^{(N)}}{dt} &= F_N(\phi^{(1)}, \phi^{(2)}, \dots, \phi^{(N)}), \end{aligned}$$

which can be written in vector form as

$$\frac{d\boldsymbol{\phi}}{dt} = \mathbf{F}[\boldsymbol{\phi}(t)] \quad (2.1)$$

where  $\boldsymbol{\phi}$  is an  $N$ -dimensional vector.

The dynamics of a systems can be of many different types. Some of the points or subsets of the phase space of particular interests are *attractors*, *stable and unstable manifolds* and *saddle points*. In this

study the focus is not on these characteristics. The feature that we are going to remark is the complexity, or the disorder, of the orbits of a dynamical system. This aspect is addressed by *chaos theory*. Chaotic dynamical systems are characterized by their unpredictability. Given an arbitrary list of consecutive values of an evolution function, we cannot infer its future values. Another factor to keep in mind is the "butterfly effect": a small change in the initial conditions leads to macroscopic differences of the evolution functions.

We now give the definition of chaotic dynamical system introduced by Robert Devaney in [15], since it intuitively highlights the fundamental aspects :

**Definition 11** (Chaotic dynamical system). A dynamical system is said to be *chaotic* if it has the following two properties:

- It must be *topologically transitive*: given two open sets  $A, B$  in  $\mathcal{M}$ , there exist an integer  $T$  such that, for all  $t > T$  it holds  $\phi_t(A) \cap B \neq \emptyset$ ;
- Its periodic orbits must form a dense set.

A consequence of these two properties is the *sensitivity to initial conditions*: for any  $x$  in  $\mathcal{M}$  and any  $\delta > 0$  there are  $y$  in  $\mathcal{M}$  such that  $0 < d(x, y) < \delta$  and  $d(\phi_t(x), \phi_t(y)) > e^{at}d(x, y)$  for some  $a > 0$ .

Sensitivity to initial condition is famous with the name of "butterfly effect": chaotic orbits can start really close to each other and end up arbitrarily far after a certain amount of time. Chaos can appear in the phase space of a system in many different ways. Two of them are *chaotic attractors* and *chaotic transients*. In this thesis these notions are going to be loosely interchanged. This is due to the fact that the chaotic orbits that are going to be studied are observed for a limited amount of time, neglecting the importance of the attractiveness of the chaotic set in which they lay.

To clarify what chaotic sets are we need the following definitions:

**Definition 12** (Limit set, invariant set, attractor). Given a dynamical system  $(T; \mathcal{M}, \phi)$ , we call  $y \in \mathcal{M}$  a  *$\omega$ -limit point* of  $x \in \mathcal{M}$  (or of an orbit containing  $x$ ) if there exists a sequence  $(t_n)_{n \in \mathbb{N}}$  such that  $\lim_{n \rightarrow \infty} t_n = \infty$  and  $\lim_{n \rightarrow \infty} \phi_{t_n}(x) = y$

An *attractor* is a subset  $A$  of the phase space characterized by the following three conditions:

- $A$  is *forward invariant* under  $\phi$ : if  $a \in A$  then so is  $\phi_t(a)$  for all  $t > 0$ ;
- There exists a *basin of attraction* of  $A$ , composed by all the points  $b$  in the phase space such that, for any open neighborhood  $N$  of  $A$ , there is a positive constant  $\bar{t}$  such that  $\phi_t(b) \in N$  for all  $t > \bar{t}$ ;
- There is no proper subset of  $A$  having the first two properties.

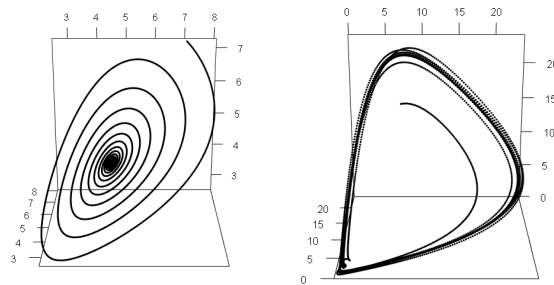


Figure 2.5: Left: a stationary point. Right: a transient converging into a limit cycle.

Attractors can have multiple natures. Figure 2.5 shows two important examples of attractors: the stationary point, in which an orbit collapses and never leaves, and the limit cycle, a set that is traveled periodically by an orbit. We can appreciate the topological difference of the two sets: the stationary point is a contractible set (homotopically equivalent to a point), while the limit cycle is homeomorphic to a circle.

**Definition 13** (Chaotic attractor). A *chaotic set* is the limit set (the collection of the limit points) of a chaotic orbit.

If this set is an attractor, we call it a *chaotic attractor*.

The term *chaotic transient* refers to the fact that an orbit can spend a long time in the vicinity of a nonattracting chaotic set before leaving it. So, for a limited amount of time, the motion can appear very irregular and indistinguishable from a motion on a chaotic attractor.

## Chapter 3

# Simplicial Complexes

This Chapter describes one of the key objects of computational topology: the simplicial complex. We can think of it as a set composed of points, line segments, triangles, and their  $n$ -dimensional counterparts. They are used as bricks to construct shapes. We now take a look at two conceptually different examples in which simplicial complexes are employed:

1. *Mesh*: we want to discretize a given manifold (to solve numerically a PDE on it, to calculate topological invariants,...). We construct a computational mesh (simplicial complex) that is homeomorphic to the abstract manifold. The resulting operative scheme is: manifold  $\rightarrow$  mesh (simplicial complex)  $\rightarrow$  vertices (point cloud);
2. *Filtration*: given a point cloud, we decide some criterion in order to connect two points with an edge, or three points with a triangle, and so on. The resulting operative scheme is: point cloud  $\rightarrow$  simplicial complex (mesh)  $\rightarrow$  manifold.

The operative scheme followed in this work is the second one. Data is going to be embedded in a high dimensional space, obtaining a point cloud. Its points will be connected with simplicial complexes, obtaining a topological space. The simplicial nature of this space is crucial as it permits to compute numerically its homology groups.

The following three Chapters are inspired by two books, that are used worldwide to treat this subject: *Elements of algebraic topology* by James Munkres [43], and *Computational Topology: an Introduction*, by Edelsbrunner and Harer [21]. Also the elements of the course *Computational algebraic topology* in Politecnico di Milano has been object of

study.

The starting point for this discussion is the notion of simplex.

### 3.1 Simplices

Simplices are convex sets created by connecting points in the euclidean space, as illustrated by the following definitions.

**Definition 14** (Affinely independent points). Let  $P_0, \dots, P_k$  be points in  $\mathbb{R}^d$ .

A point  $x = \sum_{i=0}^k \lambda_i P_i$ , with each  $\lambda_i \in \mathbb{R}$ , is an *affine combination* of the  $P_i$  if the  $\sum_{i=0}^k \lambda_i = 1$ .

The *affine hull* is the set of affine combinations.

It is a  $k$ -plane if the  $k + 1$  points are *affinely independent*, that is, the  $k$  vectors  $P_i - P_0$ , for  $1 \leq i \leq k$ , are linearly independent.

**Remark 1.** In  $\mathbb{R}^d$  we can have at most  $d$  linearly independent vectors and therefore at most  $d + 1$  affinely independent points.

**Definition 15** (Convex combination, convex hull). An affine combination,  $x = \sum \lambda_i P_i$ , is a *convex combination* if  $\sum_{i=0}^k \lambda_i = 1$  and all  $\lambda_i$  are non-negative.

The *convex hull* is the set of convex combinations.

**Definition 16** (Simplex). A  $k$ -simplex is the convex hull of  $k + 1$  affinely independent points,  $\sigma = \text{conv}\{P_0, P_1, \dots, P_k\}$ .

Its *dimension* is  $\dim \sigma = k$ .

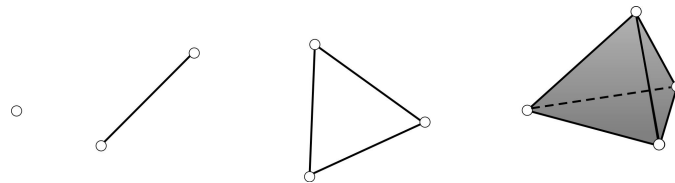


Figure 3.1: simplices

**Definition 17** (Face, boundary, interior, star). Let  $\sigma$  be a simplex. A *face*  $\tau$  of  $\sigma$  is the convex hull of a non-empty subset of the  $P_i$  ( $\tau \leq \sigma$ ), and it is *proper* if the subset is not the entire set ( $\tau < \sigma$ ).

The *boundary* of  $\sigma$  ( $\text{bd } \sigma$ ), is the union of all its proper faces, and the *interior* of  $\sigma$  ( $\text{int } \sigma$ ) is defined as  $\text{int } \sigma = \sigma - \text{bd } \sigma$ . The *star* of  $\sigma$  is defined as  $\text{St } \sigma = \bigcup \text{int } \rho$ , with  $\sigma \leq \rho$ .



The following Theorem is important as it lays the basis of the idea that we can discretize homeomorphically a topological space, and hence store a computational version of this space with the same topological invariants.

**Theorem 1.** *Let  $\sigma$  be a simplex such that  $\dim \sigma = k$ . Call  $\mathcal{B}^k$  and  $\mathcal{S}^{k-1}$  the unit ball and the unit sphere in  $\mathbb{R}^k$ . Then there exist a homeomorphism  $\varphi : \sigma \rightarrow \mathcal{B}^k$  such that  $\varphi|_{\text{bd}\sigma} : \text{bd}\sigma \rightarrow \mathcal{S}^{k-1}$  is an homeomorphism.*

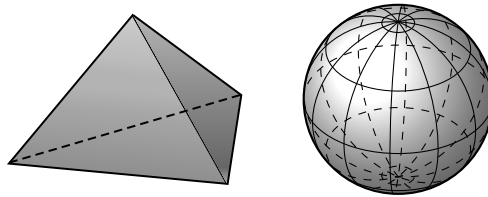


Figure 3.2: Example: the 3-simplex is homeomorphic to a ball in  $\mathbb{R}^3$ . The boundary of the 3-simplex is homeomorphic to the spherical surface.

## 3.2 Simplicial Complexes

We have seen that simplices are "bricks" of various dimensions. We can join these bricks to construct more involved shapes, called simplicial complexes. These are the objects over which we can compute homology groups numerically.

**Definition 18** (Simplicial complex). A *simplicial complex* is a finite collection of simplices  $K$  such that  $\sigma \in K$  and  $\tau \leq \sigma$  implies  $\tau \in K$ , and  $\sigma, \sigma_0 \in K$  implies  $\sigma \cap \sigma_0$  is either empty or a face of both.

The *dimension* of  $K$  is the maximum dimension of any of its simplices.

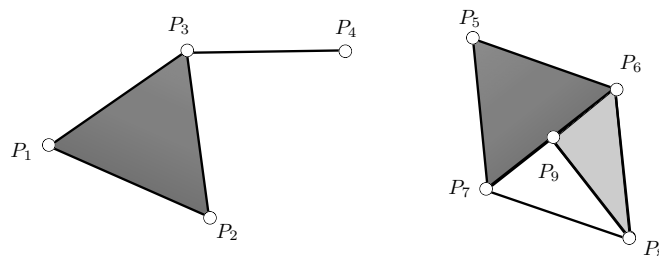


Figure 3.3: Example: the object to the left is a simplicial complex, the one to the right is not.

In figure 7.3 the left geometrical shape can be written as

$$\{\{P_1\}, \{P_2\}, \{P_3\}, \{P_4\}, \{P_1, P_2\}, \{P_1, P_3\}, \{P_2, P_3\}, \{P_3, P_4\}, \{P_1, P_2, P_3\}\}$$

and it is a simplicial complex of dimension 2 since all the requirements of the definition are satisfied. To the right, the set

$$\begin{aligned} &\{\{P_5\}, \{P_6\}, \{P_7\}, \{P_8\}, \{P_9\}, \{P_5, P_6\}, \{P_5, P_9\}, \{P_6, P_7\}, \{P_6, P_9\}, \\ &\{P_7, P_8\}, \{P_8, P_9\}, \{P_5, P_6, P_7\}, \{P_6, P_8, P_9\}\} \end{aligned}$$

is not a simplicial complex as the intersection  $\{P_5, P_6, P_7\} \cap \{P_6, P_8, P_9\} = \{P_6, P_9\}$  is not a face of  $\{P_5, P_6, P_7\}$ .

**Definition 19** (Subcomplex). Let  $K$  be a simplicial complex. A collection  $L$  of simplices of  $K$  that contains the faces of all its elements is a simplicial complex, called *subcomplex* of  $K$ .

**Definition 20** ( $p$ -skeleton). Let  $K$  be a simplicial complex. We call  $p$ -skeleton of  $K$  the subcomplex  $K^{(p)} = \{\sigma \in K \mid \dim \sigma \leq p\}$ .

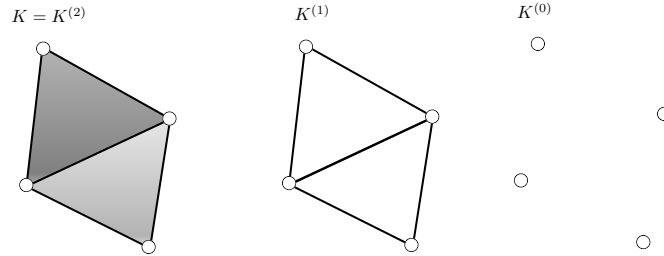


Figure 3.4: From left to right: the 2,1,0-skeleton of a simplicial complex of dimension 2.

**Definition 21** (Star). The *star* of a simplicial complex  $K$  is the union of all the stars of each simplex in  $S$ .

**Definition 22** (Topology of a simplicial complex, polytope). Let  $K$  be a simplicial complex in  $\mathbb{R}^N$ . The *underlying space*, or *polytope*, of  $K$ , is the subset of  $\mathbb{R}^N$  containing all the points in the simplices of  $K$  and it is denoted by  $|K|$ .

On  $|K|$  we can define a topology in this way:  $A \subset |K|$  is a closed set if and only if  $A \cap \sigma$  is a closed set for each  $\sigma$  belonging to  $K$ .

**Proposition 1.**  $C$  closed set in the topology inherited by  $\mathbb{R}^N \Rightarrow C$  closed set in the  $|K|$  topology.

Moreover, if  $K$  contains only a finite number of simplices, then the two topologies coincide.

**Proposition 2.** A function  $f : |K| \rightarrow X$  is continuous  $\iff f|_{\sigma} : \sigma \rightarrow X$  is continuous  $\forall \sigma \in K$ .

The underlying space will often appear in the following. In fact, it is used to think of a simplicial complex as a topological space "by itself", that is without looking at the euclidean space in which it is embedded (in the same way as topological manifolds can be seen by themselves and embedded in  $\mathbb{R}^n$  for some  $n$ ).

Now we cover the definition of simplicial map. This notion will be key for the functioning of simplicial homology and persistent homology.

**Definition 23** (Simplicial map). Let  $K, L$  be two simplicial complexes and let  $f : K^{(0)} \rightarrow L^{(0)}$  be a function such that for each set of vertices of  $K$   $\{P_0, P_1, \dots, P_k\}$  that generates a simplex of  $K$ , the set  $\{f(P_0), f(P_1), \dots, f(P_k)\}$  generates a simplex of  $L$ .

Then  $f$  can be extended to a continuous map  $\tilde{f} : |K| \rightarrow |L|$  such that given  $x \in \sigma \in K$ ,  $x = \sum_{i=0}^k t_i P_i$ , with  $P_0, \dots, P_k$  vertices of  $\sigma$ ,  $\tilde{f}(x) = \sum_{i=0}^k t_i f(P_i)$ . We call  $\tilde{f}$  the *simplicial map* induced by  $f$ .

### 3.3 Abstract Simplicial Complexes

In this section we introduce the abstract counterpart of the simplicial complex with its properties. Abstract simplicial complexes have a greater synthetic and explanatory potential, as classic simplicial complexes have always to be written with the tools of analytic geometry.

**Definition 24** (Abstract simplicial complex). An *abstract simplicial complex* is a finite collection of sets  $\mathcal{A}$  such that  $\alpha \in \mathcal{A}$  and  $\beta \subseteq \alpha$  implies  $\beta \in \mathcal{A}$ .

The sets in  $\mathcal{A}$  are its *simplices*. The *dimension* of a simplex is  $\dim \alpha = \text{card } \alpha - 1$ , and the dimension of a complex is the maximum dimension of any of its simplices.

A *face* of  $\alpha$  is a non-empty subset  $\beta \subseteq \alpha$ , which is proper if  $\beta \neq \alpha$ .

Two abstract simplicial complexes  $\mathcal{S}, \mathcal{T}$  are said to be *isomorphic*

if there is a bijection between the vertices of  $\mathcal{S}$  (the faces of cardinality 1) and the vertices of  $\mathcal{T}$  such that  $\{a_0, \dots, a_k\} \in \mathcal{S} \iff \{f(a_0), \dots, f(a_k)\} \in \mathcal{T}$ .

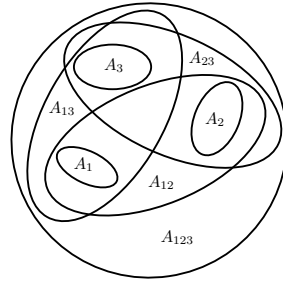


Figure 3.5: Example: an abstract simplicial complex of dimension 2.

**Definition 25** (Vertex Scheme). Let  $K$  be a simplicial complex and let  $V = K^{(0)}$  the set of its vertices. The collection of subsets of  $V$  defined by  $\mathcal{K} = \{\{v_0, \dots, v_k\} \subseteq V \mid [v_0, \dots, v_k] \in K\}$  is an abstract simplicial complex called the *vertex scheme* of  $K$ .

In the following definition we show how to switch from an abstract simplicial complex to a geometric one and viceversa, when possible:

**Definition 26** (Geometric Realization). If the abstract simplicial complex  $\mathcal{S}$  is isomorphic to the vertex scheme of a simplicial complex  $K$ , we say that  $K$  is a *geometric realization* of  $\mathcal{S}$ .

The geometric realization is unique up to homeomorphisms.

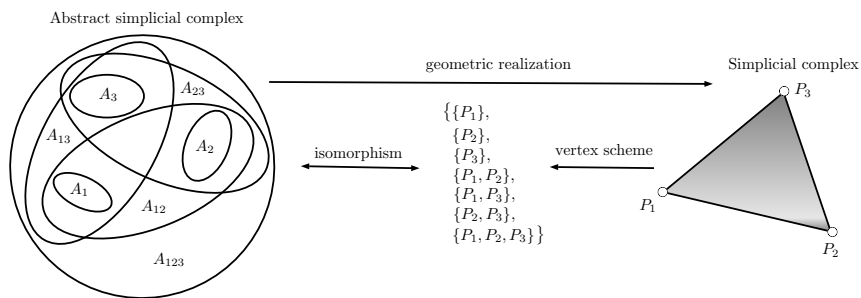


Figure 3.6: Geometric realization of an abstract simplicial complex.

Below, we define a particular kind of abstract simplicial complex: the nerve of a finite collection of sets. This object catches the underlying structure of the intersections between the sets of the collection.

**Definition 27** (Nerve). Let  $F$  be a finite collection of sets. We define the *nerve* to consist of all non-empty subcollections of  $F$  whose sets have a non-empty common intersection:

$$\text{Nrv}F = \{X \subseteq F \mid \bigcap X \neq \emptyset\}$$

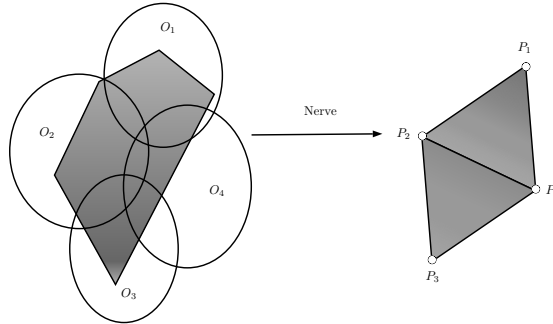


Figure 3.7: Example: geometrical realization of the nerve of an open cover.

**Theorem 2** (Nerve Theorem). Let  $F$  be a finite collection of closed, convex sets in Euclidean space. Then the nerve of  $F$  and the union of the sets of  $F$  have the same homotopy type.

### 3.4 Useful Simplicial Complexes

We present two particular categories of simplicial complex that are easy to compute on a machine and that are intrinsically related to the concept of filtration. Filtrations will be defined here and used extensively (sometimes implicitly) in Chapters 5, 6, 7, and 9.

**Definition 28** (Čech complex). Let  $S$  be a finite set of points in  $\mathbb{R}^d$  and write  $B_x(r) = x + r\mathbb{B}$  for the closed ball with center  $x$  and radius  $r$ . The *Čech complex* of  $S$  and  $r$  is the nerve of this collection of balls, substituting in the intersection the center for each ball:

$$\check{\text{Cech}}(r) = \{\sigma \subseteq S \mid \bigcap_{x \in \sigma} B_x(r) \neq \emptyset\}$$

Clearly, a set of balls has a non-empty intersection iff their centers lie inside a common ball of the same radius.

**Remark 2.** The nerve Theorem asserts that the homotopy type of a sufficiently nice topological space is encoded in the Čech nerve of an opportunely defined cover.

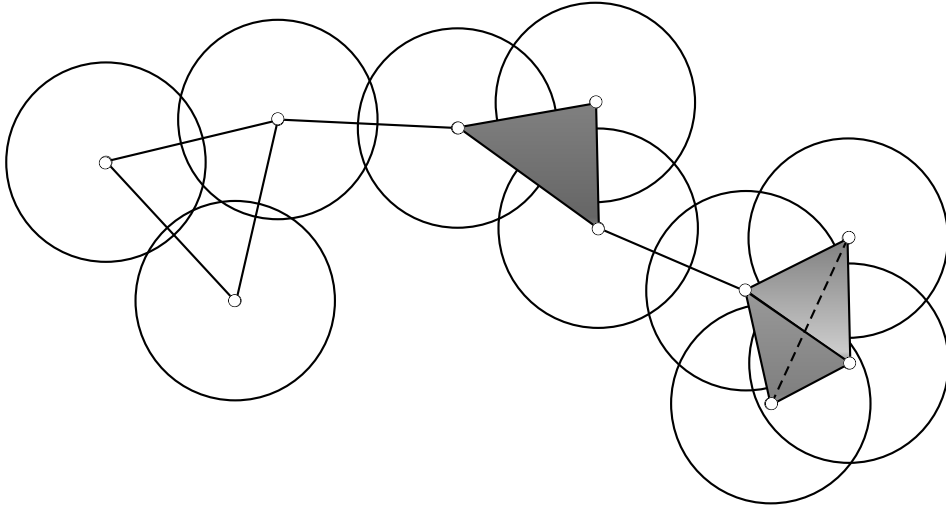


Figure 3.8: Example: Čech complex of a finite set of points.

**Definition 29** (Vietoris-Rips complex). The *Vietoris-Rips complex* of  $S$  and  $r$  consists of all subsets of diameter at most  $2r$ :

$$\text{Vietoris-Rips}(r) = \{\sigma \subseteq S \mid \text{diam } \sigma \leq 2r\}$$

The edges in the Vietoris-Rips complex are the same as in the Čech complex. Furthermore it is trivial to see that  $\check{C}ech(r) \subseteq \text{Vietoris-Rips}(r)$ .

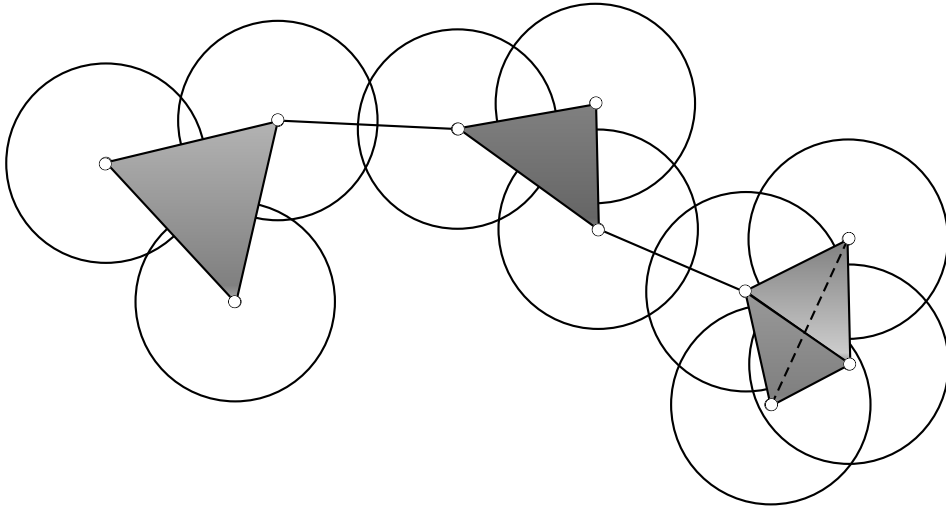


Figure 3.9: Example: Vietoris-Rips complex of a finite set of points.

The two simplicial complexes in figures 3.8 and 3.9 are the same apart from the left triangle, which is not filled for the Čech complex

while it is for the Vietoris-Rips complex. This is because the three closed balls to the left do not have a common intersection, but the diameter of the simplex generated by the three centers is smaller than two times the radius of the balls.

Both the Čech and the Vietoris-Rips complexes depend on the free parameter  $r$ . In both cases, as  $r$  varies we observe the following behaviour:

- If  $r < 0$ , then the resulting complex is the empty set.
- If  $r = 0$ , then the resulting complex is for both cases the starting set of vertices in  $\mathbb{R}^d$ .
- As  $r$  grows, new simplices get gradually added to the simplicial complex. In particular, if any simplex  $\sigma \in \check{\text{Cech}}(r_0)$  or  $\text{Vietoris-Rips}(r_0)$ , then for any  $r > r_0$ ,  $\sigma \in \check{\text{Cech}}(r)$  or  $\text{Vietoris-Rips}(r)$ .

This kind of parameter-dependent simplicial complexes are widely used in *persistent homology*, which is the main topic of Chapter 5. This theory does not study a single Čech or Vietoris-Rips simplicial complex, but the *evolution of the complexes as the free parameter varies*. The following definition of filtration describes exactly this notion.

**Definition 30** (Filtration of a monotonic function). Consider a simplicial complex,  $K$ , and a function  $f : K \rightarrow \mathbb{R}$ . We require  $f$  to be *monotonic*, that is,  $f(\sigma) \leq f(\tau)$  whenever  $\sigma$  is a face of  $\tau$ . Monotonicity implies that the sublevel set  $K_a = f^{-1}(-\infty, a]$  is a subcomplex of  $K$  for every  $a \in \mathbb{R}$ .

Given  $-\infty = a_0 < a_1 < a_2 < \dots < a_n$  the *filtration* of  $f$  is the sequence of  $n + 1$  complexes defined as:

$$\emptyset = K_0 \subseteq K_1 \subseteq K_2 \subseteq \dots \subseteq K_n = K$$

For the Čech or Vietoris-Rips complexes the filtration can be defined in this manner:  $f(\check{\text{Cech}}(r)) = r$ . The same reasoning holds for the Vietoris-Rips complexes. As aforementioned, starting from  $r < 0$ , the sublevel sets of  $f$  generate a sequence of simplicial complexes such that

$$\emptyset = K_0 \subseteq K_1 \subseteq K_2 \subseteq \dots \subseteq K_n.$$

To end the chapter, we present the definition of *simplicial manifold*, or *triangulation*, which was introduced in Chapter 2 without explanation. These objects are commonly used in other fields of applied mathematics as numerical analysis.

**Definition 31** (Triangulation). *A triangulation of a topological space  $X$  is a simplicial complex  $K$ , homeomorphic to  $X$ , together with a homeomorphism  $h : K \rightarrow X$ .*

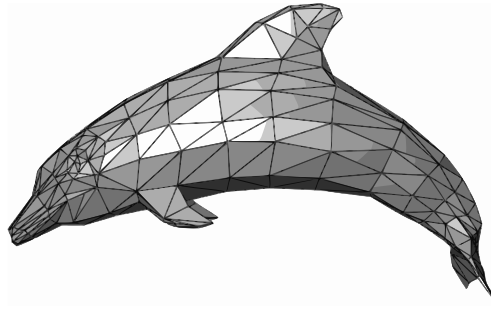


Figure 3.10: A triangulated dolphin shape.



## Chapter 4

# Simplicial Homology

Topological invariants are a handy tool. In fact, given two spaces, a fundamental topological problem is to determine whether they are homeomorphic or not. Using the definition the process can become particularly hard, especially when it has to be proven that a map between the two spaces that is continuous, bijective and with continuous inverse does not exist. A much simpler scheme is often to compute and confront topological invariants. The introduction of a sequence of topologically invariant groups for a space is due to Enrico Betti. Although simplicial homology groups can be computed over a limited set of topological spaces (the simplicial complexes), they are really useful in practice. For example, it can be proven that if a space is homeomorphic to a simplicial complex, then the homology groups of the former space are isomorphic to the simplicial homology groups of the latter. The simplicial homology groups have a clear topological interpretation, as they count the connected components and the holes of a space.

A newer application, highlighted in this thesis, comes from the field of computational topology: simplicial complexes can be used to approximate homeomorphically various shapes and to construct filtrations that permit the inspection of the nature of a point cloud of data. In this Chapter we will see the definitions of simplicial homology groups and the algorithms that are involved to compute them in practice.

### 4.1 Chain Complexes

**Definition 32** (Oriented simplex). Let  $K$  be a simplex (geometric or abstract). We say that two orderings of the vertices are *equivalent* if one can be obtained from the other with an even permutation. The two equivalence classes obtained in this way are called *orientations* of the simplex. An *oriented simplex* is a simplex with a choice of an orientation.

We denote by  $\sigma$  and  $\sigma'$  the same simplex chosen with different orientations.

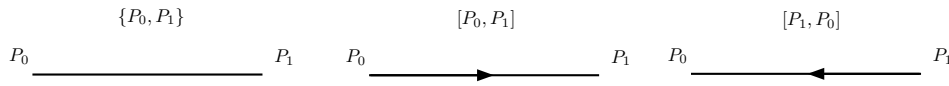


Figure 4.1: Example: a non-oriented 1-simplex (left) and its corresponding oriented versions (centre, right). Notice that there is a difference in the notation.

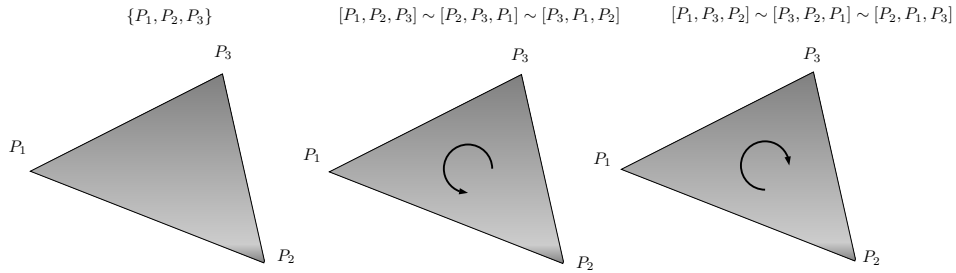


Figure 4.2: Example: a non-oriented 2-simplex (left) and its corresponding oriented versions (centre, right). Notice that there is a difference in the notation.

**Definition 33** (Group of  $p$ -chains). Let  $K$  be a simplicial complex and  $p$  a dimension. We first define the following set:

$$\mathcal{S}_p = p\text{-oriented simplices of } K$$

A  $p$ -chain on  $K$ , with *coefficients* in the group  $G$  is a function  $c : \mathcal{S}_p \rightarrow G$  such that:

1.  $c(\sigma) + c(\sigma') = 0_G$
2.  $c^{-1}(G/\{0_G\})$  contains a finite number of elements.

The set  $\mathcal{C}_p(K, G)$  of the  $p$ -chains of  $K$  with the sum operation  $+$  :  $\mathcal{C}_p(K, G) \times \mathcal{C}_p(K, G) \rightarrow \mathcal{C}_p(K, G)$ ,  $(c_1 + c_2)(\sigma) = c_1(\sigma) + c_2(\sigma) \forall \sigma \in \mathcal{S}_p$  is a group called the *group of oriented  $p$ -chains of  $K$* , with coefficients in  $G$ .

With a little abuse of notation, we can confuse the simplices in  $K$  with the elements of  $\mathcal{C}_p(K, G)$ , expressing them as a formal sum of simplices  $c = \sum a_i \sigma_i$ , where the  $\sigma_i$  are the  $p$ -simplices and the  $a_i$  are the *coefficients*.

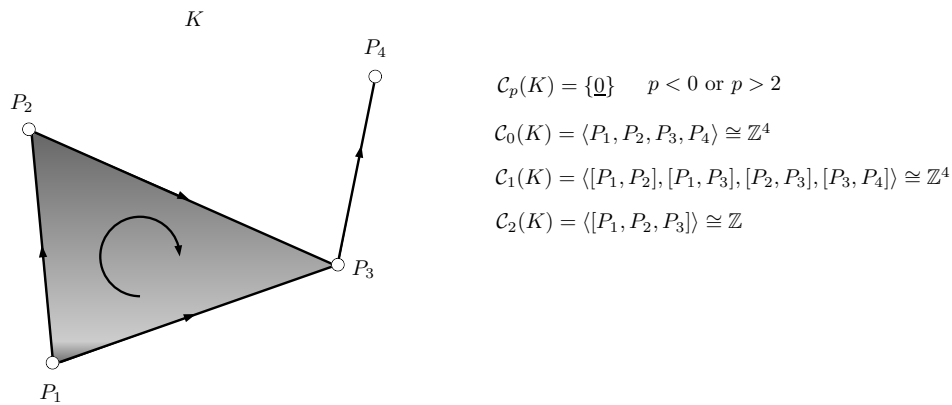


Figure 4.3: Example: group of  $p$ -chains for a simplicial complex, with coefficients in  $\mathbb{Z}$ .

The group of  $p$ -chains permits us to write combinations of faces of a simplicial complex, endowing an algebraic structure on it. Take as an example Figure 4.3. We can add vertices:  $P_1 + 3P_2 + P_4 \in \mathcal{C}_0(K)$ ; the same reasoning holds for the other dimensions.

To fix the ideas, we state some important considerations:

- If  $c_1 = \sum a_i \sigma_i$  and  $c_2 = \sum b_i \sigma_i$ , then  $c_1 + c_2 = \sum (a_i + b_i) \sigma_i$ ;
- For simplicity, we will omit the group  $G$  and write  $\mathcal{C}_p(K)$
- We have a group of  $p$ -chains for each integer  $p$ . For  $p$  less than 0 and greater than the dimension of  $K$  this group is trivial, consisting only of the neutral element.  
 $\mathcal{C}_p(K)$  is an abelian group whose neutral element is  $0 = \sum 0\sigma_i$  and the inverse of  $c$  is  $-c = c$  since  $c + c = 0$ .
- $\mathcal{C}_p(K)$  is isomorphic to the direct sum of  $G$  with itself  $\text{card}(\mathcal{S}_p)$  times.

- If  $G = \mathbb{Z}$ ,  $\mathcal{C}_p(K)$  is a free group: a base for such a group can be obtained considering every  $p$ -simplex of  $K$ , fixing an orientation for each one of them.

**Corollary 2.1.** *Let  $K$  be a simplicial complex,  $G$  be a group and  $f$  be a function such that  $f : \{p\text{-oriented simplices of } K\} \rightarrow G$  and  $f(-\sigma) = -f(\sigma) \forall \sigma \in K$ . Then  $f$  can be extended univocally to a homomorphism  $\tilde{f} : \mathcal{C}_p(K) \rightarrow G$ .*

#### 4.1.1 The boundary operator

We said that simplicial homology counts the holes of a topological space. But what is a "hole"? Think of a circle: it has a 1-dimensional hole, a void enclosed by the curve describing the circle. In the language of simplicial homology, a *hole* is present when a  $p$ -chain encloses a  $p - 1$  dimensional void. We can formalize this idea starting from the definitions of boundary operator and chain complex.

**Definition 34** (Boundary homomorphism). The *boundary* of a  $p$ -simplex is the alternating sum of its  $(p - 1)$ -dimensional faces. Writing  $\sigma = [u_0, u_1, \dots, u_p]$  for the simplex spanned by the listed vertices, its boundary is

$$\partial_p \sigma = \sum_{j=0}^p (-1)^j [P_0, \dots, \hat{P}_j, \dots, P_p]$$

where the hat indicates that  $P_j$  is omitted. We notice that  $\partial_p : \mathcal{C}_p \rightarrow \mathcal{C}_{p-1}$ : for a  $p$ -chain,  $\partial_p c = \sum a_i \partial_p \sigma_i$ ; moreover the boundary operator commutes with addition,  $\partial_p(c_1 + c_2) = \partial_p(c_1) + \partial_p(c_2)$ . Therefore, since  $\partial_p$  commutes with the group operation, we refer to it as the *boundary homomorphism*, or for short, the *boundary map* for chains.

**Definition 35** (Chain complex). A *chain complex* is denoted as  $\mathfrak{C} = \{\mathcal{C}_p, \partial_p\}_p$  and it is the sequence of chain groups connected by boundary homomorphisms:

$$\dots \xrightarrow{\partial_{p+2}} \mathcal{C}_{p+1} \xrightarrow{\partial_{p+1}} \mathcal{C}_p \xrightarrow{\partial_p} \mathcal{C}_{p-1} \xrightarrow{\partial_{p-1}} \dots$$

Last, we introduce exact sequences, as there is a link between the rank simplicial homology groups and the degree of non-exactness of a sequence.

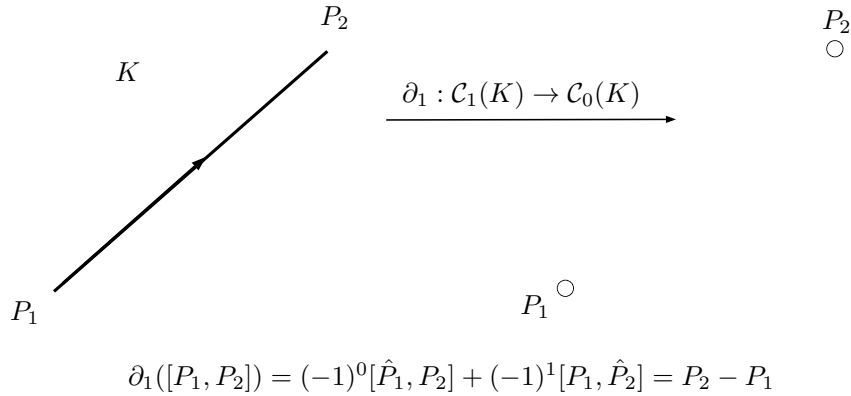


Figure 4.4: Example: boundary homomorphism applied to a 1 dimensional simplex.

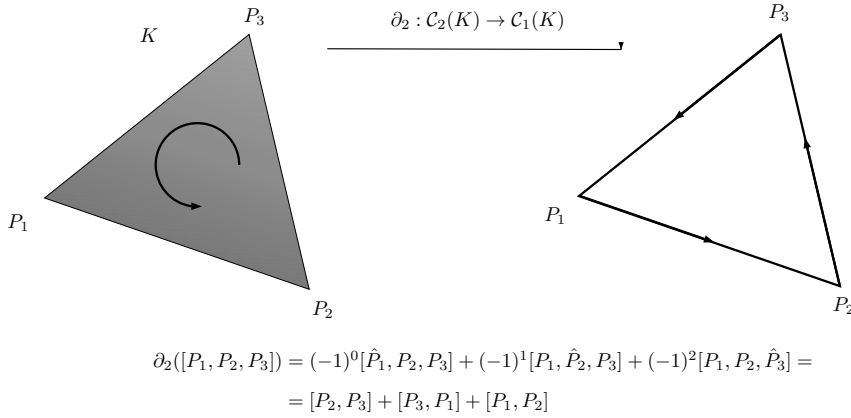


Figure 4.5: Example: boundary homomorphism applied to a 2 dimensional simplex.

**Definition 36** (Exact sequence). A possibly infinite sequence

$$\dots \xrightarrow{\phi_1} G_1 \xrightarrow{\phi_2} G_2 \xrightarrow{\phi_3} \dots \xrightarrow{\phi_n} G_n \xrightarrow{\phi_{n+1}} \dots$$

of groups and groups homomorphisms is called *exact* if the image of each homomorphism is equal to the kernel of the next:

$$\text{im}(\phi_k) = \ker(\phi_{k+1}).$$

A *short exact sequence* is an exact sequence of the form

$$0 \rightarrow G_1 \xrightarrow{\phi} G_2 \xrightarrow{\varphi} G_3 \rightarrow 0.$$

To shorten the notation, we'll refer to  $\partial_p$  as  $\partial$ , since the index of the boundary homomorphism is implied by the dimension of the chain it applies to.

## 4.2 Simplicial Homology Groups

Simplicial homology groups are defined as quotient of two other groups: the cycles, that are chains whose boundary is the trivial group, and the boundaries, that are chains resulted from boundary operations.

**Definition 37** (Cycles). The *group of  $p$ -cycles* is defined as  $\mathcal{Z}_p = \ker \partial_p$ . So, a  *$p$ -cycle* is a  $p$ -chain with empty boundary:  $\partial c = 0$ .

**Definition 38** (Boundaries). The *group of  $p$ -boundaries* is defined as  $\mathcal{B}_p = \text{im } \partial_{p+1}$ . So, a  *$p$ -boundary* is a  $p$ -chain that is the boundary of a  $p+1$ -chain:  $c = \partial d$  with  $d \in \mathcal{C}_{p+1}$ .

**Remark 3.** Cycles and boundaries are subgroups of the chain groups. Since the chain groups are abelian, so are these subgroups.

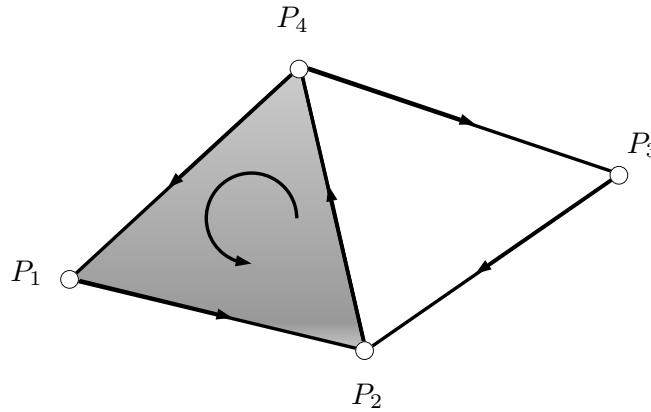


Figure 4.6: Example: boundary and cycle.

In figure 4.6 we have a simplicial complex containing both a cycle and a boundary of dimension 1:

- The 1-dimensional cycle is the chain  $[P_2, P_4] + [P_4, P_3] + [P_3, P_2]$ . In fact  $\partial_1([P_2, P_4] + [P_4, P_3] + [P_3, P_2]) = P_4 - P_2 + P_3 - P_4 + P_2 - P_3 = 0$ .
- The 1-dimensional boundary is the chain  $[P_1, P_2] + [P_2, P_4] + [P_4, P_1]$ . In fact applying the boundary operator  $\partial_2$  to the 2-dimensional complex  $[P_1, P_2, P_3]$  we get the 1-dimensional chain mentioned above.

**Lemma 3** (Fundamental Lemma of Homology).  $\partial_p \partial_{p+1} d = 0$  for every integer  $p$  and for every  $(p+1)$ -chain  $d$ .

*Proof.* Since  $\partial$  commutes with the chain sum operation, we just have to prove  $\partial_p \partial_{p+1} \tau = 0$  for a  $(p+1)$  simplex  $\tau$ .  $\partial_{p+1} \tau$  consists of all the  $p$ -faces of  $\tau$ . As every  $(p-1)$ -face of  $\tau$  belongs to exactly two  $p$  faces, so  $\partial_p(\partial_{p+1} \tau) = 0$ .  $\square$

We are finally ready to define the simplicial homology groups. An important consequence of Lemma 3 is that every  $p$ -boundary is also a  $p$ -cycle; that is,  $\mathcal{B}_p$  is a subgroup of  $\mathcal{Z}_p$ . This means that not every cycle encloses a hole: a chain whose boundary is the trivial group could be an element of  $\mathcal{B}_p$ . The quotient operation permits to choose only the cycles that enclose a hole.

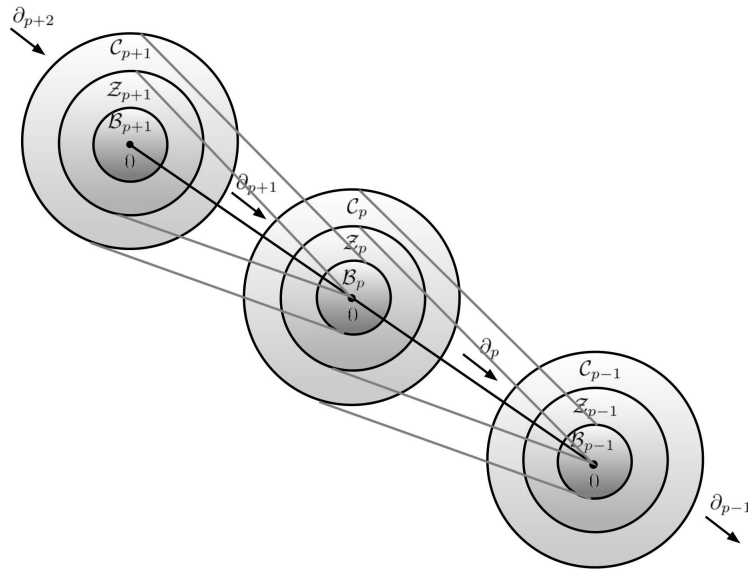


Figure 4.7: Images of the boundary operator: every chain gets mapped to a boundary; every cycle and boundary gets mapped to 0.

**Definition 39** (Homology groups and Betti numbers). The  $p$ -th homology group is the  $p$ -th cycle group modulo the  $p$ -th boundary group:  $\mathcal{H}_p = \mathcal{Z}_p / \mathcal{B}_p$ .

The  $p$ -th Betti number is the rank of this group:  $\beta_p = \text{rank} \mathcal{H}_p$ .

The two following Theorems have to be pondered with particular attention. They provide the basic topological interpretations of homology groups: 0-th dimensional homology counts the *connected components* of a set; higher dimensional homology groups count the *holes* of the

space. Two other results will state that homeomorphic (Theorem 7) and homotopically equivalent (Theorem 8) underlying spaces of simplicial complexes determine isomorphic homology groups.

Singular homology, that will not be treated in this dissertation, works on generic topological spaces and a fundamental result of this theory is the invariance under homotopy of homology groups. Thus, a way to compute the homology groups of a space is to find a homotopically equivalent simplicial complex and to compute its simplicial homology groups.

**Theorem 4** (0-th dimensional homology). *Let  $K$  be a simplicial complex.*

1. *The group  $\mathcal{H}_0(K)$  is a free group.*
2. *Let  $\{P_\alpha\}$  be a collection of vertices, one for each connected component of  $|K|$ . Then the homology classes of the points  $P_\alpha$  form a basis of  $\mathcal{H}_0(K)$ .*

**Theorem 5** (Homology of the boundary of a simplex). *Let  $\Delta_n$  the complex composed by a  $n$ -simplex and its faces, with  $n > 0$ . Let  $\Sigma^{n-1}$  be the subcomplex of  $\Delta_n$  given by  $\Sigma^{n-1} = \Delta_n - \Delta_n^{(n)}$ . ( $|\Sigma^{n-1}| = \bar{\Delta}_n - \text{int}(\Delta_n)$ ). Then*

$$\mathcal{H}_p(\Sigma^{n-1}) = \begin{cases} \langle \partial\Delta_n \rangle = \mathbb{Z}, & \text{if } p = n - 1 \text{ or } p = 0 \\ \{0\}, & \text{otherwise.} \end{cases}$$

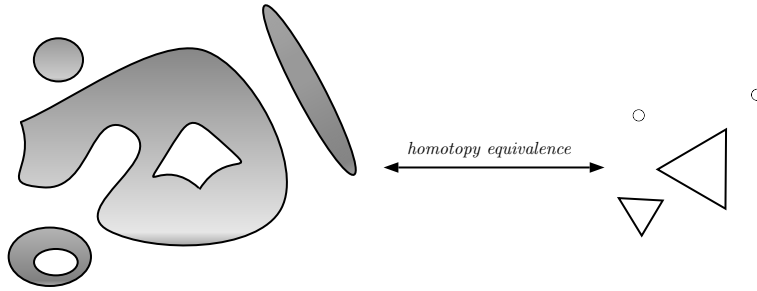


Figure 4.8: Example: The homology groups of the space to the left can be obtained by computing the simplicial homology groups of the complex to the right. As there are four connected components and two 1-dimensional holes, the 0-th dimensional homology group is isomorphic to  $\mathbb{Z}^4$  and the 1-st dimensional homology group is isomorphic to  $\mathbb{Z}^2$ .



**Remark 4.** Homology groups can also be interpreted as a measure of how much a sequence of chain complex is non exact. For instance, if we assume that this sequence is exact, we have:

$$\dots \xrightarrow{\partial_{p+2}} \mathcal{C}_{p+1} \xrightarrow{\partial_{p+1}} \mathcal{C}_p \xrightarrow{\partial_p} \mathcal{C}_{p-1} \xrightarrow{\partial_{p-1}} \dots$$

with the restriction

$$\mathcal{B}_p = \text{im} \partial_{p+1} = \text{ker} \partial_p = \mathcal{Z}_p \quad \forall p,$$

which suggests that "every cycle is a boundary": from the definition of homology groups we get that  $\mathcal{H}_p = \mathcal{Z}_p / \mathcal{B}_p \cong \underline{0} \quad \forall p$ . On the other end, if the chain complex is not exact, the rank of the homology groups tells us "how many cycles are not boundaries": these particular cycles enclose a "hole".






Homology groups of some simple shapes.				
Space ( $S$ )	Name	$\mathcal{H}_0(S)$	$\mathcal{H}_1(S)$	$\mathcal{H}_2(S)$
	Point	$\mathbb{Z}$	$\underline{0}$	$\underline{0}$
	Circle	$\mathbb{Z}$	$\mathbb{Z}$	$\underline{0}$
	Möbius Strip	$\mathbb{Z}$	$\mathbb{Z}$	$\underline{0}$
	Torus	$\mathbb{Z}$	$\mathbb{Z}^2$	$\mathbb{Z}$
	Sphere	$\mathbb{Z}$	$\underline{0}$	$\mathbb{Z}$

Table 4.1: Homology groups of simple topological spaces.

### 4.3 Topological invariance of the homology groups

Simplicial homology groups are topological invariants. This means that if two simplicial complexes are homeomorphic, then the corresponding groups are isomorphic. To prove this important property (as said before, the same holds with homotopies), the first step is to show that simplicial maps, defined in 23, can *induce homomorphism* between the chain groups and consequently, between simplicial homology groups.

**Definition 40** (Chain map). Let  $f : K \rightarrow L$  be a simplicial map. If  $P_0, \dots, P_p$  is a simplex of  $K$ , then the points  $f(P_0), \dots, f(P_p)$  describe a simplex of  $L$ . The *chain map* induced by the simplicial map  $f$  is the homomorphism  $(f_\#)_p : \mathcal{C}_p(K) \rightarrow \mathcal{C}_p(L)$  defined as:

$$(f_\#)_p([P_0, \dots, P_p]) = \begin{cases} [f(P_0), \dots, f(P_p)] & \text{if } f(P_0), \dots, f(P_p) \text{ are distinct,} \\ 0 & \text{otherwise} \end{cases}$$

**Lemma 6.** *The homomorphism  $(f_\#)_p$  commutes with  $\partial$ . Therefore,  $(f_\#)_p$  induces a homomorphism  $(f_*)_p : \mathcal{H}_p(K) \rightarrow \mathcal{H}_p(L)$ .*

*Proof.* We prove only the second statement. Consider a chain of order  $p$ ,  $c_p$ . Since  $(f_\#)_p$  commutes with  $\partial$ , it follows that  $(f_\#)_p$  carries cycles to cycles, as

$$\partial(f_\#)_p(c_p) = (f_\#)_p(\partial c_p) = 0.$$

Moreover, boundaries are carried to boundaries, as assuming it exists  $d_{p+1}$  such that  $c_p = \partial d_{p+1}$  we get

$$(f_\#)_p(c_p) = (f_\#)_p(\partial d_{p+1}) = \partial(f_\#)_p(d_{p+1}).$$

Thus  $(f_\#)_p$  induces a homomorphism  $(f_*)_p : \mathcal{H}_p(K) \rightarrow \mathcal{H}_p(L)$  of homology groups.  $\square$

The second step is not covered entirely in the following due to its technicality. We have to show that any continuous map between underlying spaces generates a simplicial maps between simplicial complexes, as homotopies and homeomorphisms are not simplicial maps in general. Then it can be proved that *the simplicial maps relative to structure-preserving transformations induce isomorphism between homology groups*, as stated in Theorems 7 and 8.

**Definition 41** (Star condition). Let  $h : |K| \rightarrow |L|$  be a continuous map. We say that  $h$  satisfies the *star condition* with respect to  $K$  and  $L$  if for each vertex  $P$  of  $K$ , there is a vertex  $Q$  of  $L$  such that  $h(\text{St}P) \subset \text{St}Q$ .

**Definition 42** (Simplicial approximation). Let  $h : |K| \rightarrow |L|$  be a continuous map. If  $f : K \rightarrow L$  is a simplicial map such that  $h(\text{St}P) \subset \text{St}P$  for each vertex  $P$  of  $K$ ,  $f$  is called a *simplicial approximation* to  $h$ .

If  $h : |K| \rightarrow |L|$  satisfies the star condition (relative to  $K$  and  $L$ ), we can define a homomorphism

$$(h_*)_p : \mathcal{H}_p(K) \rightarrow \mathcal{H}_p(L)$$

by setting  $h_* = f_*$ , where  $f$  is any simplicial approximation to  $h$ .

**Theorem 7** (Topological invariance of the homology groups). *If  $h : |K| \rightarrow |L|$  is a homeomorphism, then  $(h_*)_p : \mathcal{H}_p(K) \rightarrow \mathcal{H}_p(L)$  is an isomorphism.*

**Theorem 8** (Homotopy and Homology). *If  $|K| \rightarrow |L|$  are homotopically equivalent, then  $(f_*)_p$  is an isomorphism for all  $p$ .*

## 4.4 Computation of simplicial homology groups via Matrix Reduction

In this section we describe the natural context of computational topology, and we show how the machine calculates the homology groups of a simplicial complex.

Recall the definition of group of  $p$ -chains [33](#). If the coefficients of our group of  $p$ -chains lie in  $\mathbb{Z}_2$ , we say that we are working with *modulo 2 coefficients*. As the only elements of  $\mathbb{Z}_2$  are 0 and 1, the  $p$ -simplices in a modulo 2 chain can be only "switched on" ( $1 \cdot \sigma$ ) or "switched off" ( $0 \cdot \sigma$ ), obtaining a configuration that resembles the boolean nature of the bit. From now on in this work, we suppose to be using  $\mathbb{Z}_2$  coefficients. This choice brings some important consequences:

- Any sum between  $p$ -chains corresponds to their symmetric difference:  $c_1 + c_2 = \sum (a_i \sigma_i + b_i \sigma_i) = \sum (a_i + b_i) \sigma_i$  where the coefficients satisfy  $1 + 1 = 0$ ;
- As for any simplex it holds  $\sigma + \sigma = 0$ , then  $\sigma = \sigma'$ . This identity means that with modulo 2 coefficients we lose the information regarding the *orientation* of the simplices.
- The boundary map defined in [34](#) assumes the easier form  $\partial_p \sigma = \sum_{j=0}^p [P_0, \dots, \hat{P}_j, \dots, P_p]$ .

The computer can easily reckon homology groups of a simplicial complex using linear algebra, exploiting the linearity of the boundary operator and the resemblances between free abelian groups and vector spaces.

Let  $f : U \rightarrow V$  be a linear transformation between vector spaces: we have that the dimension of  $U$  equals to the sum of the dimension of the kernel of  $f$  and the dimension of the image of  $f$ . That is  $\dim U = \dim \ker f + \dim \operatorname{im} f$ . In the same manner, considering  $f = \partial$  to be the boundary map and writing  $n_p = \operatorname{rank} \mathcal{C}_p$  for the number of  $p$ -simplices in  $K$ ,  $z_p = \operatorname{rank} \mathcal{Z}_p$  and  $b_p = \operatorname{rank} \mathcal{B}_p$  for the ranks of the cycle and boundary groups, we have  $n_p = z_p + b_{p-1}$ .

**Definition 43** (Boundary matrices). Let  $K$  be a simplicial complex. Its  $p$ -th boundary matrix inserts the  $(p-1)$ -simplices as rows and the  $p$ -simplices as columns. Fix an ordering for the simplices. For each dimension,  $\partial_p = [a_i^j]$ , where  $i$  ranges from 1 to  $n_{p-1}$ ,  $j$  ranges from 1 to  $n_p$ , and  $a_i^j = 1$  if the  $i$ -th  $(p-1)$  simplex is a face of the  $j$ -th  $p$ -simplex and  $a_i^j = 0$  otherwise.

Given a  $p$ -chain  $c = \sum a_i \sigma_i$ , the boundary can be computed by matrix multiplication,

$$\partial_p c = \begin{bmatrix} a_1^1 & a_1^2 & \cdots & a_1^{n_p} \\ a_2^1 & a_2^2 & \cdots & a_2^{n_p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_{p-1}}^1 & a_{n_{p-1}}^2 & \cdots & a_{n_{p-1}}^{n_p} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n_p} \end{bmatrix}$$

The boundary matrix is such that its rows form a basis of  $\mathcal{C}_{p-1}$ , and its columns form a basis of  $\mathcal{C}_p$ . To compute homology groups,  $\partial_p$  has to be reduced to an easier configuration, the Smith normal form. This can be done with two types of matrix multiplications, called row and column operations, that translate into these possible effects:

- Both row and column operations do not affect the rank of  $\partial_p$ ;
- Row operations consist in adding or exchanging two rows;
- column operations consist in adding or exchanging two columns.

**Definition 44** (Smith normal form). The  $p$ -th boundary matrix is reduced to the Smith normal form if it is obtained using row and column operations, so it has the form  $N_p = U_{p-1} \partial_p V_p$ , and all its elements are 0 with the exception of an initial segment of the diagonal.

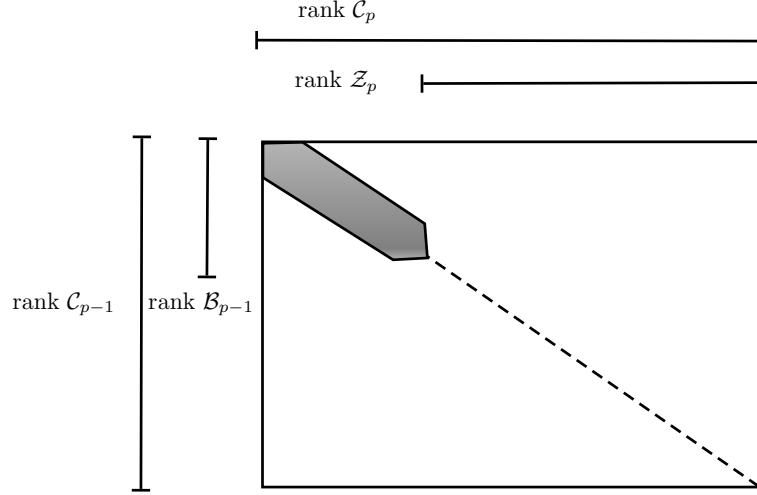


Figure 4.9: Structure of a matrix in Smith normal form

**Reduction algorithm.** Row and column operations can be performed optimally via the following Reduction algorithm, that requires a running time of  $C \cdot 2n_{p-1}n_p \min\{n_{p-1}, n_p\}$  and an allocation of memory of  $C \cdot (n_{p-1} + n_p)^2$ :

---

Initialize the boundary matrix to  $N_p[i, j] = a_i^j$ .  
 Call Reduce(1), structured as follows:

---

```

    Reduce(x)
    if  $\exists k \geq x, l \geq x$  with  $N_p[k, l] = 1$  {
    exchange rows  $x$  and  $k$ ; exchange columns  $x$  and  $l$ ;
    for  $i = x + 1 : n_{p-1}$  {
    if  $N_p[i, x] = 1$  { add row  $x$  to row  $i$ ; } }
    for  $j = x + 1 : n_p$  {
    if  $N_p[x, j] = 1$  { add column  $x$  to column  $j$ ; } }
    Call Reduce( $x + 1$ ); }
    
```

---

**Computation of Betti numbers.** Once we have all boundary matrices in normal form, we can extract the Betti numbers as differences between ranks,  $\beta = \text{rank } Z_p - \text{rank } B_p$ , for  $p \geq 0$ , in this way:

- recall that  $n_p = \text{rank } \mathcal{C}_p$  is the number of columns of the  $p$ -th boundary matrix;
- let  $n_p = b_{p-1} + z_p$  so that the leftmost  $b_{p-1}$  columns have ones in the diagonal and the rightmost  $z_p$  columns are zero. The former represents  $p$ -chains whose non-zero boundaries generate the group of  $(p-1)$ -boundaries. The latter represents  $p$ -cycles that generate  $\mathcal{Z}_p$ ;
- it is possible to obtain the bases of  $\mathcal{Z}_p$  and  $\mathcal{B}_p$  using the matrices  $V_p$  and  $U_{p-1}$ , respectively. The new basis for the cycle group is given in the last  $z_p$  columns of  $V_p$ . The new basis for the boundary group is made by the first  $b_{p-1}$  columns of the inverse of  $U_{p-1}$ .

## Chapter 5

# Persistence

This Chapter is devoted to persistent homology. This tool is of huge practical importance, as standard simplicial homology cannot be used univocally to recover the shape of data. In fact, as data comes in the form of a point cloud. From this point cloud we can build infinite triangulations, each one endowed with different topological features. This arbitrariness hides the true shape of data.

Persistent homology comes in aid in the following way. The point cloud is seen as a set of vertices, which become increasingly connected through higher dimensional simplices using a filtration. For example, Čech or Vietoris-Rips filtrations can be built with increasing values of the free parameter,  $r$ . Then, simplicial homology is computed for each simplicial complex in the filtration. The key concept is that **some homology classes will persist when passing from a simplicial complex in the filtration to the next; in the same way, some new classes will appear, and some will vanish**. The more a class is present while going through the filtration, the more the geometrical hole represented by that class is big. From the persistence of the homology classes we can reconstruct the topological space hidden in a point cloud.

We begin by recalling the definition of filtration (already seen in Chapter 3), as it is crucial for the understanding of the chapter.

**Definition 45** (Filtration of a monotonic function). Consider a simplicial complex,  $K$ , and a function  $f : K \rightarrow \mathbb{R}$ . We require  $f$  to be *monotonic*, that is,  $f(\sigma) \leq f(\tau)$  whenever  $\sigma$  is a face of  $\tau$ . Monotonic-

ity implies that the sublevel set  $K_a = f^{-1}(-\infty, a]$  is a subcomplex of  $K$  for every  $a \in \mathbb{R}$ .

Given  $-\infty = a_0 < a_1 < a_2 < \dots < a_n$  the *filtration* of  $f$  is the sequence of  $n + 1$  complexes defined as:

$$\emptyset = K_0 \subseteq K_1 \subseteq K_2 \subseteq \dots \subseteq K_n = K$$

More than in the sequence of complexes, we are interested in the corresponding sequence of **homology groups**. For every  $i \leq j$  we have an inclusion map from the underlying space of  $K_i$  to that of  $K_j$  and therefore an induced homomorphism,  $f_p^{i,j} : H_p(K_i) \rightarrow H_p(K_j)$ , for each dimension  $p$ . The filtration thus corresponds to a sequence of homology groups connected by homomorphisms,

$$0 = \mathcal{H}_p(K_0) \rightarrow \mathcal{H}_p(K_1) \rightarrow \dots \rightarrow \mathcal{H}_p(K_n) = \mathcal{H}_p(K)$$

for each dimension  $p$ . Going from  $K_{i-1}$  to  $K_i$ , we might:

- gain new homology classes;
- lose some homology classes when they become trivial or merge with each other.

We collect the classes that ”**are born**” at or before a given point of the sequence, and ”**die**” after another point in groups.

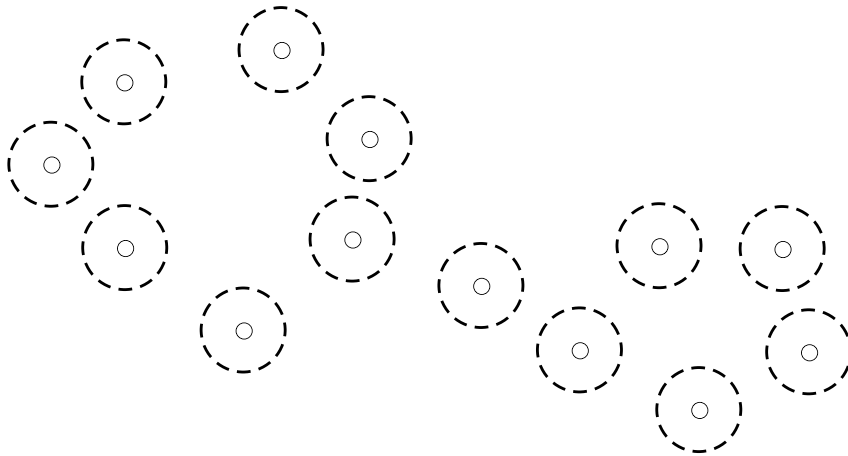


Figure 5.1: Example-1: element of a Čech filtration. The point clouds suggests a shape with one connected component, one small hole and one big hole. At this stage of the filtration, only 13 0-dimensional homology classes (connected components are alive).



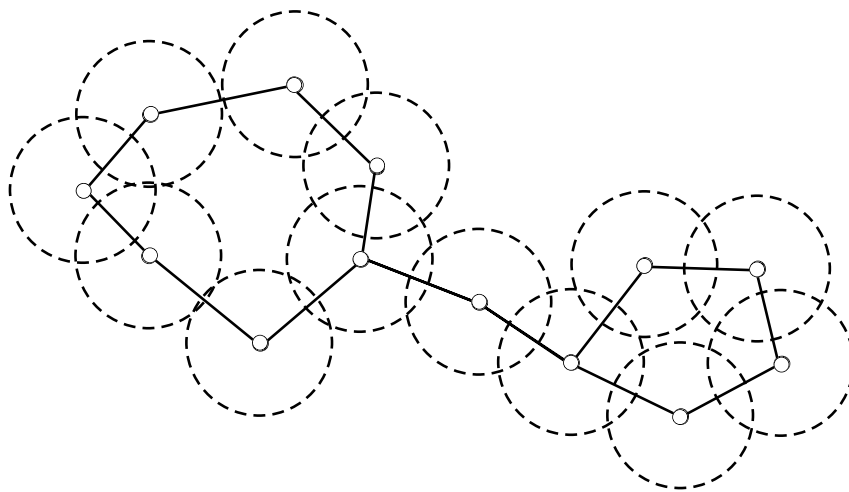


Figure 5.2: Example-2: as the radius of the Čech complex grows, all connected components die apart from one. Two one dimensional homology classes (the two holes) are born.

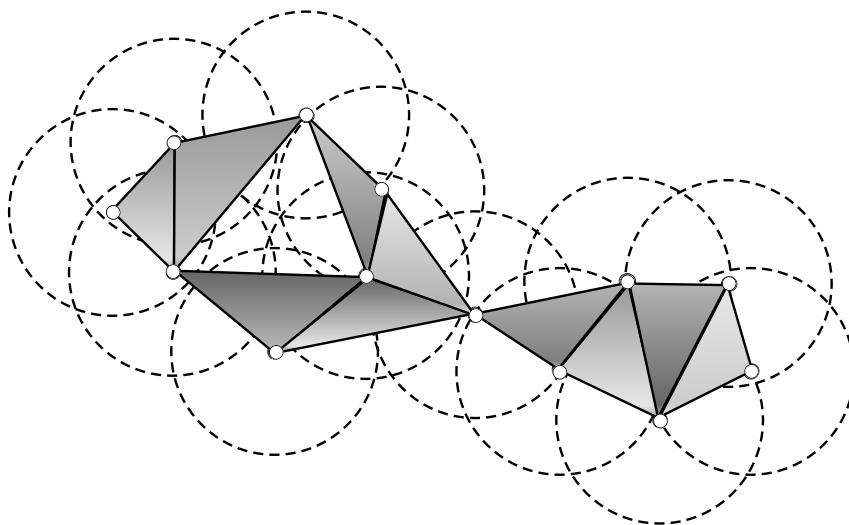


Figure 5.3: Example-3: suppose this is the last element of the filtration. We remain with a 0-th dimensional homology class. One of the two classes representing the holes has died. This suggests that the shape behind the point cloud has 1 connected component, 1 small hole and 1 big hole.

**Definition 46** (Persistent homology groups). The  $p$ -th persistent homology groups are defined as  $\mathcal{H}_p^{i,j} = \text{im } f_p^{i,j}$ , for  $0 \leq i \leq j \leq n$ , so they are the images of the homomorphism induced by inclusion.

The corresponding  $p$ -th persistent Betti numbers are the ranks of these

groups,  $\beta_p^{i,j} = \text{rank } \mathcal{H}_p^{i,j}$

**Remark 5** (Properties and nomenclature).

- $\mathcal{H}_p^{i,i} = \mathcal{H}_p(K_i)$ .
- From the definition, the persistent homology groups consist of the homology classes of  $K_i$  that are still alive at  $K_j$ ; thus they can be written in the form

$$\mathcal{H}_p^{i,j} = \mathcal{Z}_p(K_i) / (\mathcal{B}_p(K_j) \cap \mathcal{Z}_p(K_i)).$$

In fact, if the homology class is persistent it means that it encloses a hole both in  $K_i$  and in  $K_j$ . A cycle in  $K_i$  is automatically a cycle in  $K_j$  due to Lemma 6. But if this class happens to be also a boundary in  $K_j$ , then the application of the boundary operator on this class returns 0, but there is no hole enclosed in  $K_j$ .

- Letting  $\gamma$  be a class in  $\mathcal{H}_p^{i,j}$ , we say that it is *born* at  $K_i$  if  $\gamma \notin \mathcal{H}_p^{i-1,i}$ .
- If  $\gamma$  is born at  $K_i$ , then it *dies* entering  $K_j$  if it merges with an older class after the homomorphism induced by the inclusion of  $K_{j-1}$  in  $K_j$  is applied. That is:

$$f_p^{i,j}(\gamma) \notin \mathcal{H}_p^{i-1,j-1} \cap f_p^{i,j}(\gamma) \in \mathcal{H}_p^{i-1,j}$$

**Definition 47** (Persistence). If  $\gamma$  is born at  $K_i$  and dies entering  $K_j$ , then we call the difference of the values of  $f$  the *persistence*:

$$\text{pers } \gamma = a_j - a_i.$$

If  $\gamma$  is born at  $K_i$  but never dies, then we set its persistence at  $\infty$ .

Another quantity, similar to persistence, is the *persistence index*  $i-j$ , which tells for how many complexes in a filtration the homology class remains alive. This dissertation covers the case in which all filtrations are composed by Čech or Vietoris-Rips complexes; each one of these complexes corresponds to a *critical value* of  $r$ , a value for which at least one new homology class is born or dies, causing a change in the topology. A class with high persistence is generally considered to be relevant for the description of the topology of the point cloud (**with respect to the sublevel sets of the function  $f$ !**). To classes with low

persistences correspond to the effects generated by noise and topological features with small size.

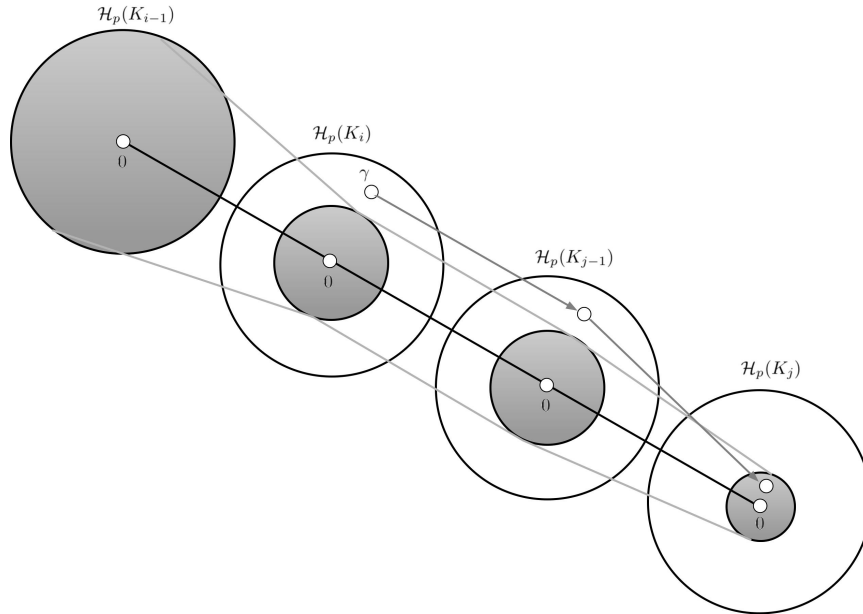


Figure 5.4: Example: the homology class  $\gamma$  is born in the complex  $K_i$  of the filtration. Its image  $f_p^{i,j}(\gamma)$  persists in the complex  $K_{j-1}$ , and finally  $f_p^{j-1,j}(\gamma)$  dies in the complex  $K_{j-1}$ . The persistence index of  $\gamma$  is  $j - i$ .

## 5.1 Persistence Diagrams

In this section we introduce the mathematical objects that permits the comprehensive analysis of a filtration: the persistence diagram. The persistence diagram collects all the aforementioned topological information: birth, death and persistence of all the persistent homology classes. Let us see how.

Let  $\mu_p^{i,j}$  be the number of  $p$ -dimensional classes born at  $K_i$  and dying entering  $K_j$ . This number can be expressed as:

$$\mu_p^{i,j} = (\beta_p^{i,j-1} - \beta_p^{i,j}) - (\beta_p^{i-1,j-1} - \beta_p^{i-1,j}),$$

for all  $i < j$  and all  $p$ . Indeed,  $(\beta_p^{i,j-1} - \beta_p^{i,j})$  counts the classes that are born **at or before**  $K_i$  and die **entering**  $K_j$ , while  $(\beta_p^{i-1,j-1} - \beta_p^{i-1,j})$  counts the classes that are born **at or before**  $K_{i-1}$  and die **entering**  $K_j$ .

**Definition 48** ( $p$ -th persistence diagram - 1). Drawing each point  $(a_i, a_j)$  in the extended real plane  $\mathbb{R}^2$  with multiplicity  $\mu_p^{i,j}$ , we get a multiset of points called the  $p$ -th persistence diagram of the filtration, denoted as  $Dgm_p(f)$ .

- In the persistence diagram, a class is represented by a point whose **vertical distance** from the diagonal is the **persistence**;
- Since the multiplicities are defined only for  $i < j$ , all points lie above the diagonal;
- All points on the diagonal are added to the diagonal, each with infinite multiplicity. This is needed to define distances between persistence diagrams, as we shall see later;

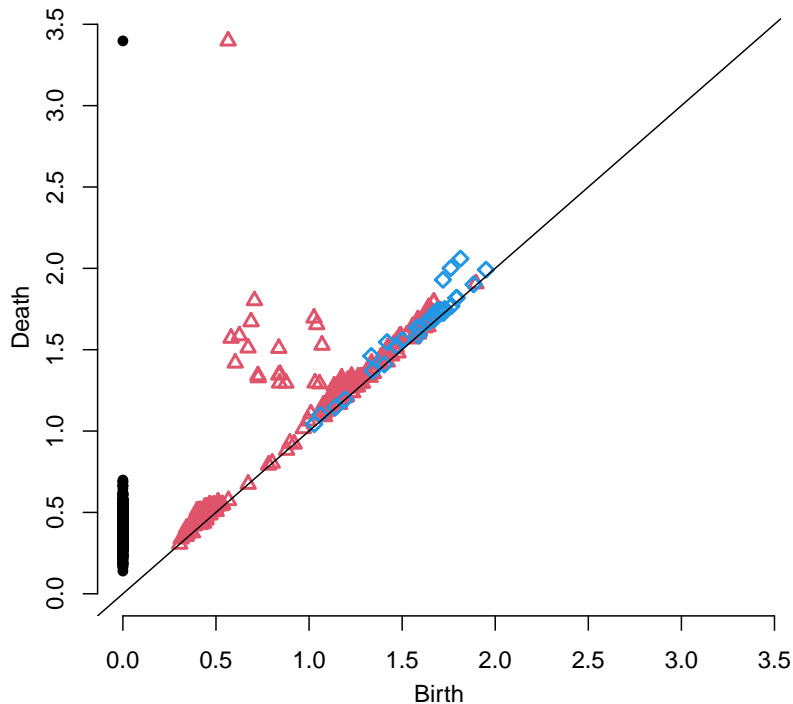


Figure 5.5: Example: a juxtaposition of persistence diagrams of dimension 0, 1 and 2. The black dots represent the 0-th dimensional classes. The red triangles and the blue rhombuses constitute 1 and 2 dimensional classes, respectively.

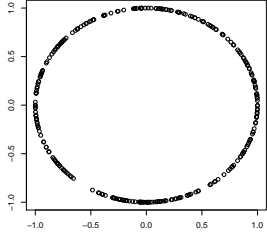
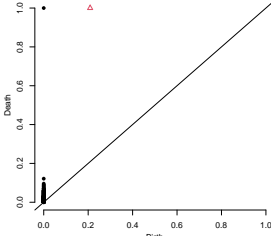
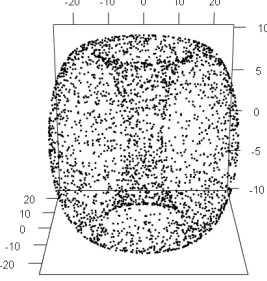
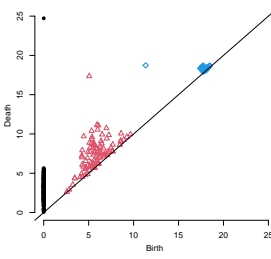
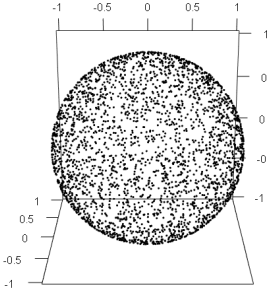
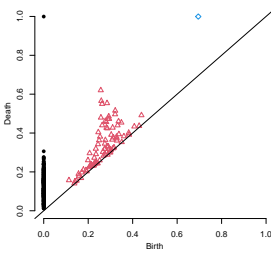
<i>Persistence diagrams of some point clouds.</i>		
Point cloud	Shape	Persistence diagram
 <p>A scatter plot showing a circle of points in a 2D coordinate system. The x and y axes both range from -1.0 to 1.0 with major ticks every 0.5 units.</p>	Circle	 <p>A persistence diagram with Birth on the x-axis and Death on the y-axis, both ranging from 0.0 to 1.0. A diagonal line represents the identity function. There is a single point at (0, 1) and a single point at (0.2, 0.2). A vertical bar is present at Birth = 0 from Death = 0 to 0.1.</p>
 <p>A scatter plot showing a torus (donut shape) of points in a 3D coordinate system. The x and y axes range from -20 to 20, and the z-axis ranges from -10 to 10.</p>	Torus	 <p>A persistence diagram with Birth on the x-axis and Death on the y-axis, both ranging from 0 to 25. A diagonal line is shown. There is a point at (0, 25) and a point at (18, 18). A vertical bar is present at Birth = 0 from Death = 0 to 6. A cluster of red points follows the diagonal from (0, 0) to (10, 10). A blue point is at (18, 18).</p>
 <p>A scatter plot showing a sphere of points in a 3D coordinate system. The x and y axes range from -1 to 1, and the z-axis ranges from -1 to 1.</p>	Sphere	 <p>A persistence diagram with Birth on the x-axis and Death on the y-axis, both ranging from 0.0 to 1.0. A diagonal line is shown. There is a point at (0, 1) and a point at (0.7, 0.7). A vertical bar is present at Birth = 0 from Death = 0 to 0.2. A cluster of red points follows the diagonal from (0, 0) to (0.5, 0.5). A blue point is at (0.7, 0.7).</p>

Table 5.1: Persistence diagrams of some point clouds. Notice the consistency with the homology groups of Table 4.2.

The following lemma asserts that the persistent diagram actually encodes all the information about persistence homology groups. This is

true as each section of the diagrams presents the multiplicities  $\mu_p^{i,j}$  corresponding to a specific complex in the filtration. Adding them together we obtain the persistent Betti numbers of that same complex.

**Lemma 9** (Fundamental Lemma of Persistent Homology). *Let  $\emptyset = K_0 \subseteq K_1 \subseteq \dots \subseteq K_n = K$  be a filtration. For every pair of indices  $0 \leq k \leq l \leq n$  and every dimension  $p$ , the  $p$ -th persistent Betti number is  $\beta_p^{k,l} = \sum_{i \leq k} \sum_{j < l} \mu_p^{i,j}$ .*

Being a multiset of points, the persistence diagram can also be defined without any link with algebraic topology. This approach is used to study spaces of persistent diagrams, metrics, or statistical properties.

**Definition 49** (Metrics on the space of persistence diagrams). The set of all persistence diagrams is denoted by  $\mathcal{D}$ . One metric on  $\mathcal{D}$  is the  $L^2$ -Wasserstein metric

$$d_{L^2}(X, Y)^2 = \inf_{\phi: X \rightarrow Y} \sum_{x \in X} \|x - \phi(x)\|^2$$

where all the possible bijections  $\phi$  between  $X$  and  $Y$  are considered. Let's remark that these bijections always exist as any point can be paired to the diagonal, that contains infinitely many points.

Another metric is the  $q$ -Wasserstein distance between two persistence diagrams, defined in this way:

$$d_{W_q}(X, Y) = \left( \inf_{\phi: X \rightarrow Y} \sum_{x \in X} \|x - \phi(x)\|_\infty^q \right)^{\frac{1}{q}}.$$

A third one is the so called *bottleneck distance*:

$$d_\infty(X, Y) = \inf_{\phi: X \rightarrow Y} \sup_{x \in X} \|x - \phi(x)\|_\infty.$$

As shown in [39], the space of persistence diagrams as defined in 48 together with the  $q$ -Wasserstein distance is not complete. A counterexample of a non-converging Cauchy sequence is given by defining  $D_n$  as the diagram containing the points  $x_1, \dots, x_n$ , where  $x_k = (0, 2^{-k}) \in \mathbb{R}^2$ . Then  $d_{W_q}(D_n, D_{n+k}) \leq \frac{1}{2^{n+k}}$ , so  $\{D_n\}$  is Cauchy. However, as  $n \rightarrow \infty$  the number of off-diagonal points tends to  $\infty$ , so the limit object cannot be a persistence diagram. This inconvenience leads to a more general definition of persistence diagram.

**Definition 50** (Generalized persistence diagram - 2). A *persistence diagram* is a countable multiset of points in  $\mathbb{R}^2$  along with the infinitely many copies of the diagonal  $\Delta = \{(x, y) \in \mathbb{R}^2 \mid x = y\}$ . For the countably many points  $x_j \in \mathbb{R}^2$  not lying on the diagonal, it is required that  $\sum_j \|x_j - \Delta\| < \infty$ .

This definition permits the introduction of a space of persistence diagrams with much better properties.

**Theorem 10** (Space of persistence diagrams). *Let  $D_0$  be the empty persistence diagram, that is, the diagram with only the diagonal. The space  $\mathfrak{D}_q = \{D \mid d_{W_q}(D, D_0) < \infty\}$  is complete and separable.*

The space  $\mathfrak{D}_q$  has other relevant properties as the possibility to define adequate concepts of expectation, variance, and conditional probability, giving to topological persistence a theoretical base for numerous statistical applications.

As we will briefly see in Chapter 9, another approach to extract statistical information from the persistence diagrams is the implementation of *summary functions*, which are defined on much more handy spaces and condense the relevant features of a diagram.

## 5.2 Stability of Persistence Diagrams

For practical purposes as classification tasks it is crucial to enjoy stability results for persistence diagrams. First, if we consider the same topological space with two filtrations generated by similar functions, we expect similar persistence diagrams. Second, when we fix the monotonic function and we compute the filtration on similar point clouds, we also expect the persistence diagrams to be similar.

For the results of this thesis the second consideration is of great importance: if a noiseless and noisy version of the same point cloud have dissimilar diagrams then persistent feature extraction for machine learning purposes becomes useless, as also similar point clouds would give rise to contrasting features.

Below we show the needed hypotheses for stability and corresponding important results.

**Definition 51** (Tame functions). Let  $\mathbb{X}$  be triangulable and  $f : \mathbb{X} \rightarrow \mathbb{R}$  continuous. Given the sublevel sets  $\mathbb{X}_a (= f^{-1}(-\infty, a])$  and  $\mathbb{X}_b$  we can

define a map connecting the  $p$ -th dimensional homology groups of the sublevel sets,  $f_p^{a,b} : \mathcal{H}_p(\mathbb{X}_a) \rightarrow \mathcal{H}_p(\mathbb{X}_b)$ . We call  $a$  an *homological critical value* if there is no  $\epsilon > 0$  for which  $f_p^{a-\epsilon, a+\epsilon}$  is an isomorphism for each dimension  $p$ . Finally, we call  $f$  *tame* if it has only finitely many homological critical values and all homology groups of all sublevel sets have finite rank.

**Theorem 11** (Bottleneck stability for tame functions). *Let  $\mathbb{X}$  be a triangulable topological space and  $f, g : \mathbb{X} \rightarrow \mathbb{R}$  two tame functions. For each dimension  $p$ , let  $D_{f,p}$  and  $D_{g,p}$  the persistence diagrams generated by the sublevel sets of  $f, g$  on  $\mathbb{X}$ . Then:*

$$\forall p \quad d_\infty(D_{f,p}, D_{g,p}) \leq \|f - g\|_\infty. \quad (5.1)$$

**Theorem 12** (Wasserstein stability for Lipschitz functions). *Let  $f, g : \mathbb{X} \rightarrow \mathbb{R}$  be tame Lipschitz functions on a metric space. With some hypotheses on the triangulations generated by  $f, g$ , there exist constants  $C$  and  $k > 1$  such that:*

$$d_{W_q}(D_f, D_g) \leq C \|f - g\|_\infty^{k/q} \quad \forall q \geq k. \quad (5.2)$$

**Proposition 3** (Bottleneck-Hausdorff Stability). *Let  $PC_X, PC_Y$  two point and  $X, Y$ , their corresponding persistent diagrams. Then  $d_\infty(X, Y) \leq 2d_H(PC_X, PC_Y)$ .*



## Chapter 6

# Sliding Window embedding

This chapter is devoted to a famous and widely used tool in signal analysis. It is known as *Takens' embedding* in dynamical systems theory and as *Sliding Window Embedding* in computational topology. Suppose our data is a continuous-time scalar function. From the topological point of view, as any loop is forbidden in the graph of a scalar function, the only relevant feature is the number of connected components, given by the number and the magnitude of the jump discontinuities. From the point of view of dynamical systems we can detect some basic properties by sight: a stationary point is present if the function becomes indefinitely constant. A limit cycle is present if the function becomes periodic. The key concept here is that **the topology of these attractors is contracted**: in 1D, there is no trace of a point in which an the orbit collapses. Same for the limit cycle, that should appear in the phase space as a closed curve. In addition, a circular limit cycle can be the graph of a function only in a 3 (or more)-dimensional space, whereas a scalar function's graph lives in  $\mathbb{R}^2$ .

The Sliding Window Embedding is a transformation from a scalar signal to a higher dimensional vector. It is employed to **recover the topology of the attractors**, or the transients, of a phase space. We will see that the image of a constant scalar function through this embedding is a stationary point; the image of a periodic scalar function is a limit cycle.

Notice that once we embed a scalar signal in a higher dimensional space we end up with a point cloud that can be studied with the tools

of persistent homology and dynamical systems.

## 6.1 Dynamical systems: delay coordinates and Takens' Theorem

Let  $\phi(t) : \mathcal{M} \rightarrow \mathbb{R}^n$  be a vectorial evolution function of our dynamical system. We assume that we can only observe a component, or more generally, one scalar function of the state vector,

$$g(t) = G(\phi(t)).$$

In this context we aim at obtaining information on the phase space. For example we want to determine if we are in presence of a chaotic transient or not, or reveal some fractal structure in the dynamics of the system. To do that we can rely on the so called *delay coordinate vector*, defined as

$$\boldsymbol{\psi}(t) = \begin{bmatrix} g(t) \\ g(t + \tau) \\ g(t + 2\tau) \\ \vdots \\ g(t + M\tau) \end{bmatrix}. \quad (6.1)$$

where  $\tau$  is some fixed time interval.  $\tau$  should be chosen to be of the order of the characteristic time over which  $g(t)$  varies. Some techniques for the choice of  $\tau$  are presented in Chapter 7. Thanks to equation (2.1), for each initial time  $t_0$  the evaluation  $\phi(t_0 - m\tau)$  is uniquely determined by  $\phi(t_0)$  and thus can be written as

$$\phi(t_0 - m\tau) = \phi(t_0) + \int_{t_0}^{(t_0 - m\tau)} \mathbf{F}[\phi(t)] dt = \mathbf{L}_m(\phi(t)).$$

Hence,  $g(t - m\tau) = G(\mathbf{L}_m(\phi(t)))$  and so we can view the vector  $\boldsymbol{\psi}(t)$  as a function of  $\phi(t)$ :

$$\boldsymbol{\psi} = \mathbf{H}(\phi). \quad (6.2)$$

The advantage of this representation comes from the fact that if the number of delays  $M$  is sufficiently large and if the function  $G$  is sufficiently regular, then we might inspect the phase space of  $\boldsymbol{\psi}$  to recover the dynamics of  $\phi$ .

## 6.1. Dynamical systems: delay coordinates and Takens' Theorem

Let us examine the situation from a more practical point of view. Say we use delay coordinates to construct a  $d_\psi$ -dimensional vector  $\psi$ . Then we assume that there exist an actual smooth low dimensional system which describes the dynamics:

$$\frac{d\phi}{dt} = \mathbf{F}(\phi),$$

where  $\phi$  has some dimensionality  $d_\phi$ . We now employ the aforementioned reasoning 6.2 and write  $\psi = \mathbf{H}(\phi)$ . Now we have to impose that  $\psi$  have to be a dynamical system: its orbits cannot overlap. This translates into

$$\phi \neq \phi' \text{ implies } \mathbf{H}(\phi) \neq \mathbf{H}(\phi'). \quad (6.3)$$

If (6.3) holds, then we say that  $\mathbf{H}$  is an *embedding* of the  $d_\phi$  dimensional space into the  $d_\psi$ .

Consider the example in the figures below. The original dynamical system is a sinusoid. Applying a 3-dimensional embedding we find a limit cycle in the 3-d space. There is no overlapping in this case, even if the cycle is winded. If we try to apply the same transformation with one less delay coordinate, we find an overlapping plot in the 2-dimensional space. This cannot be a dynamical sytem. Thus the



Figure 6.1: Left:  $\sin(t)$ . Center: 3-dimensional embedding with function  $G(x) = \sin(10x)$ . Right: 2-dimensional delay coordinates.

questions that becomes evident is: how large does  $d_\psi$  typically have to be to ensure that we avoid self-intersections of the  $d_\phi$ -dimensional  $\phi$  when we attempt to embed it in a  $d_\psi$ -dimensional delay coordinate  $\psi$ -space? The answer is due to Floris Takens [61]: his famous theorem says that generically  $d_\psi \geq d_\phi + 1$  is sufficient.

**Theorem 13** (Takens). *Let  $\mathcal{M}$  be a compact manifold of dimension  $m$ . For pairs  $(\chi, y)$  with  $\chi \in \text{Diff}^2(\mathcal{M})$ ,  $y \in C^2(\mathcal{M}, \mathbb{R})$ , it is a generic property that the map  $\chi_{\chi, y} : \mathcal{M} \rightarrow \mathbb{R}^{2m+1}$ , defined by*

$$\chi_{\chi, y}(x) = (y(x), y(\chi(x)), \dots, y(\chi^{2m}(x))) \quad (6.4)$$

*is an embedding.*

In this context the expression "generic property" stands for a property that holds for "almost any" map. We can try to give a simple justification to Theorem 13 in this way: say we have a smooth surface of dimension  $d_1$  and another of dimension  $d_2$ , both lying in an  $N$ -dimensional space. If these surface intersect (in a "good" intersection), then the dimension  $d_0$  of the intersection is

$$d_0 = d_1 + d_2 - N. \quad (6.5)$$

If (6.5) yields  $d_0 < 0$ , then the two sets do not intersect. Theorem 13 comes from equation (6.5) by taking  $d_1 = d_2 = d_\phi$  and requiring  $d_0 < 0$ . Thus, the smallest possible  $N$  is indeed  $2 * d_\phi + 1$ .

In other words Takens' theorem gives a condition on the minimum number of delay coordinates, so that the resulting embedding recovers the topology of a smooth attractor (or of a smooth chaotic nonattractor set) in a chaotic dynamical system. The condition is that the dimension of the embedding should be greater than two times that of the attractor. In Chapter 9 we will see that the standard notion of dimension is not appropriate for certain chaotic sets, thus we will introduce more general definitions of dimension.

## 6.2 Delay embedding application in persistent homology

### 6.2.1 Basic properties

The previous section showed how an attractor can be recovered from the application of Takens' theorem and delay coordinates. In the following we will study how the topology of these attractors can be studied through persistent homology, in order to obtain information on the original scalar signal.

**Definition 52** (Sliding window embedding). Suppose  $f$  is a function defined on an interval of the real numbers. Choose an integer  $M$  and a real number  $\tau$ , both greater than 0. The *sliding window embedding* of  $f$  based at  $t \in \mathbb{R}$  into  $\mathbb{R}^{M+1}$  is the point

$$SW_{M,\tau}f(t) = \begin{bmatrix} f(t) \\ f(t + \tau) \\ \vdots \\ f(t + M\tau) \end{bmatrix}. \quad (6.6)$$

Choosing different values of  $t$  gives a collection of points called *sliding window point cloud* for  $f$ . A critical parameter for this embedding is the *window-size*  $M\tau$ .

Perea and Harer studied in [48] the characteristic of  $SW_{M,\tau}$  as a linear operator. Then, we present an important approximation Theorem from the same paper. The Theorem states that the persistent homology of the sliding window point cloud of a function  $f \in C^k(\mathbb{T}, \mathbb{R})$  is stable when higher frequencies are filtered.

**Proposition 4.** *Let  $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ . Then for all  $M \in \mathbb{N}$  and  $\tau > 0$ , the mapping  $SW_{M,\tau} : C(\mathbb{T}, \mathbb{R}) \rightarrow C(\mathbb{T}, \mathbb{R}^{M+1})$  is a bounded linear operator with norm  $\|SW_{M,\tau}\| \leq \sqrt{M+1}$ .*

*Proof.* Linearity is obvious. For any  $f \in C(\mathbb{T}, \mathbb{R})$  and  $t \in \mathbb{T}$  it holds:

$$\|SW_{M,\tau}f(t)\|_{\mathbb{R}^{M+1}}^2 = |f(t)|^2 + |f(t+\tau)|^2 + \dots + |f(t+M\tau)|^2 \leq (M+1)\|f\|_{\infty}^2.$$

Thus also boundedness is proven  $\square$

To introduce the approximation Theorem, we now imagine to approximate a function  $f$  through its Fourier polynomials and study how the sliding windows change with respect to the threshold frequency. In particular, let

$$f(t) = S_N f(t) + R_N f(t)$$

where

$$S_N f(t) = \sum_{n=0}^N a_n \cos(nt) + b_n \sin(nt)$$

is the  $N$ -truncated Fourier series expansion of  $f$  and  $R_N f$  is the remainder.

**Definition 53.** Let  $D$  be a persistence diagram. Define  $\text{pers}(x, y) = y - x$  for  $(x, y) \in \mathbb{R}^2$ , and as  $\inf$  otherwise. We let

$$\text{mp}(D) = \max_{x \in D} \text{pers}(x)$$

**Theorem 14** (Approximation). *Let  $T \subset \mathbb{T}$ , and  $f \in C^k(\mathbb{T}, \mathbb{R})$ . Call  $PC = SW_{M, \tau} f(T)$  and  $PC_N = SW_{M, \tau} S_N f(T)$ . Finally, let  $D, D_N$  be the persistence diagrams relative to  $PC$  and  $PC_N$ , respectively. Then;*

$$d_H(PC, PC_N) \leq \sqrt{4k - 2} \|R_N f^{(k)}\|_2 \quad (6.7)$$

$$|\text{mp}(D) - \text{mp}(D_N)| \leq 2d_\infty(D, D_N) \quad (6.8)$$

$$d_\infty(D, D_N) \leq 2\sqrt{4k - 2} \|R_N f^{(k)}\|_2 \frac{\sqrt{M + 1}}{(N + 1)^{k-1/2}} \quad (6.9)$$

It follows that in signal analysis, and in particular in feature extraction for machine learning, filters can be used reasonably without altering the persistence diagrams too much.

### 6.2.2 Examples

1) : *Approximation theorem in action.* Two similar signals, filtered differently, have a similar persistence diagram.

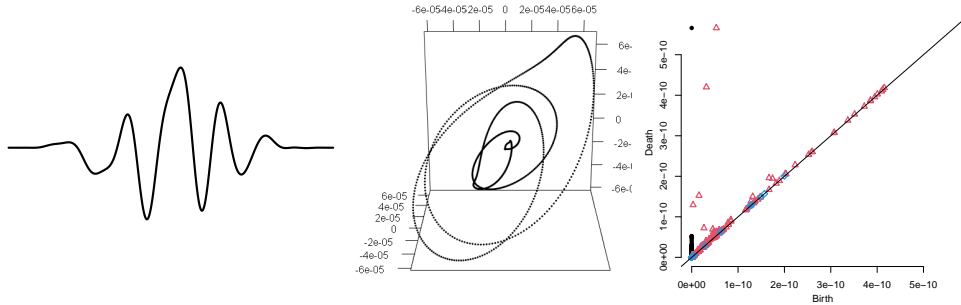


Figure 6.2: Example: 3-dimensional embedding of a filtered signal. The band-pass filter keeps only frequencies between 0.5 and 2.3 Hz. On the right we have the corresponding persistence diagram.

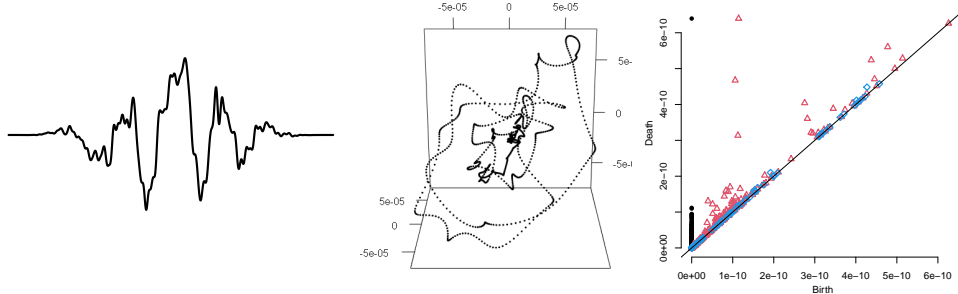


Figure 6.3: Example: 3-dimensional embedding of a filtered signal. The band-pass filter keeps only frequencies between 0.5 and 13 Hz. On the right we have the corresponding persistence diagram.

2) : Recovering simple attractors. The sliding window embedding unfolds with great clarity attractors as stationary points and limit cycles. Their topology can be studied with persistent homology.

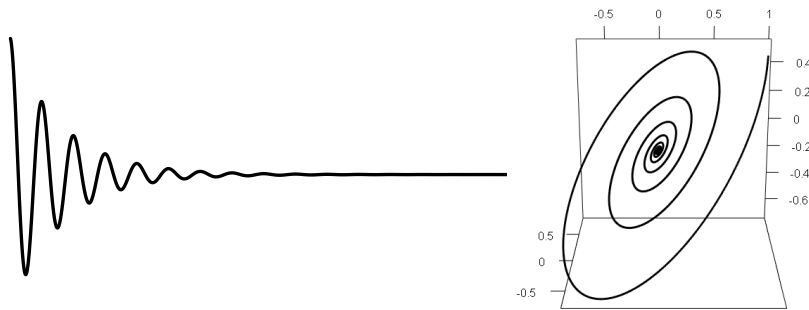


Figure 6.4: Example:  $f(t) = \cos(0.01t) \cdot e^{\frac{-t}{1000}}$ ; recovery of a stationary point.

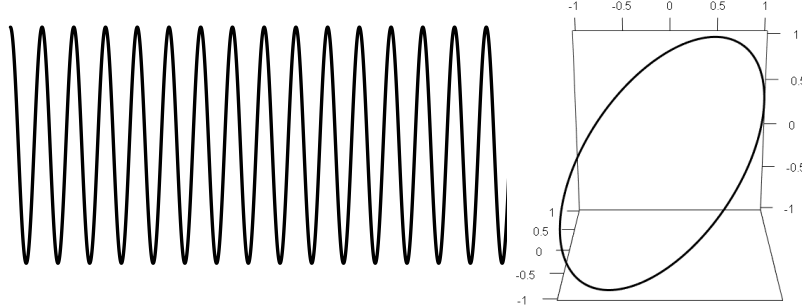


Figure 6.5: Example:  $f(t) = \cos(0.01t)$ ; recovery of a limit cycle.

3) : *Recovering complex attractors.* Attractors can have the shape of more complicated topological spaces. For example, in [25], [62] and [47] it is shown that quasiperiodic signals give rise to toroidal phase spaces through the sliding window embedding. Moreover, persistent homology can be exploited as a very precise tool (more precise than spectrum analysis, for example) to quantify periodicity, as in Figure 6.5, and quasiperiodicity, as shown in the Figure below. The phase space can also be a chaotic set or even a strange attractor.

**Definition 54** (Quasiperiodic function). A scalar *quasiperiodic* function  $f : \mathbb{R} \rightarrow \mathbb{C}$  has the form  $f(t) = \sum_{n=0}^N c_n e^{i\omega_n t}$ , where  $N \in \mathbb{N}$ , the  $c_n$  are non-zero complex numbers and the  $\omega_n$  are *incommensurate* positive real numbers. Incommensurate means that  $1, \omega_0, \dots, \omega_N$  are linearly independent over  $\mathbb{Q}$ .



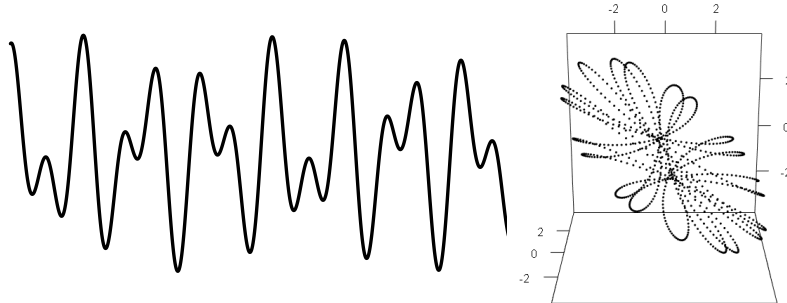


Figure 6.6: Example:  $f(t) = 2\sin(0.01t) + 1.8\sin(0.01\sqrt{3}t)$ ; recovery of a limit torus.

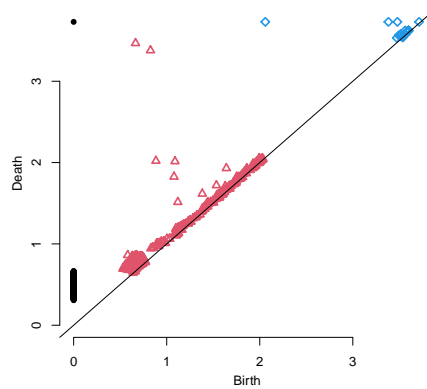


Figure 6.7: Example: persistence diagram of the reconstructed phase space of Figure 6.6. The two 1-dimensional and the single 2-dimensional high persistence classes suggest that the attractor is a torus.

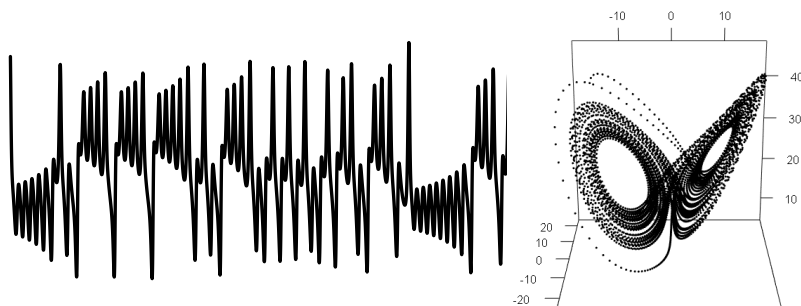


Figure 6.8: Example: recovery of a strange attractor (the Lorenz system).



## Chapter 7

# Automatic recognition of Sleep Phase and K-complexes

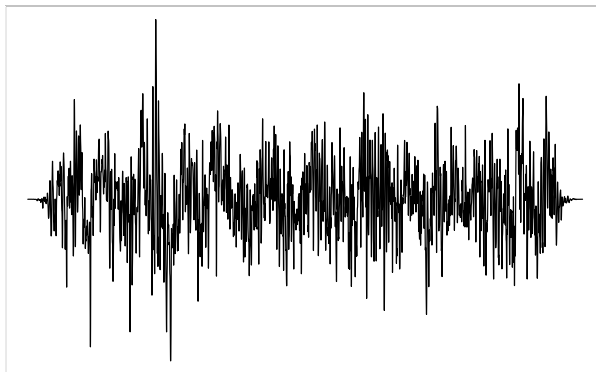
### 7.1 Sleep and Sleep Stages

Following a simple behavioral definition given in [14], sleep is a reversible behavioral state of perceptual disengagement from and unresponsiveness to the environment. Sleep is typically accompanied by a combination of physiologic and behavioral processes, as postural recumbence, behavioral quiescence, closed eyes. Nonetheless, other behaviors and anomalies can occur during sleep. Examples are given by sleepwalking, sleeptalking, teeth grinding, falling asleep and waking up, and dream imaginery.

Within sleep, two separate states have been defined over a multitude of physiologic parameters. These states, the rapid eye movement (*REM*) and non-rapid eye movement (*NREM*) are as distinct from one another as each is from wakefulness. *NREM* sleep, also known as quiescent sleep, is the collection of sleep stages 1-3. The EEG pattern in *NREM* sleep is commonly described as synchronous, with characteristic waveforms as sleep spindles, K-complexes, and high voltage slow waves. According to studies, the mental activity that takes place during *NREM* sleep is believed to be thought-like, whereas *REM* sleep includes hallucinatory and bizarre content. Dreaming is rare during *NREM* sleep. As mentioned, *NREM* sleep is divided into three stages:

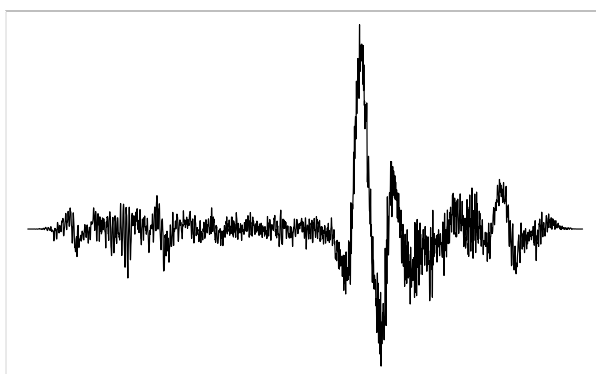
- *Stage 1*: occurs mostly at the onset of sleep. It is characterized by

the disappearance of alpha waves (8-12 Hz) and the appearance of theta waves (4-7 Hz). People aroused from this stage often believe that they have been fully awake; for this reason this state is sometime referred to as relaxed wakefulness [27].



*Figure 7.1: EEG signal during stage 1.*

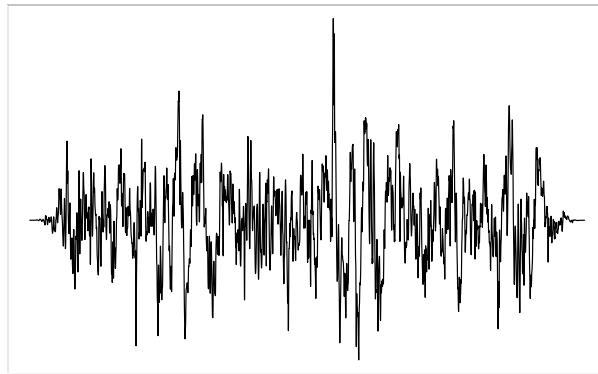
- *Stage 2*: EEG recordings tend to show characteristic sleep spindles, which are short bursts of high frequency brain activity, and K-complexes during this stage. Dreaming is rare during this stage.



*Figure 7.2: EEG signal during stage 2.*

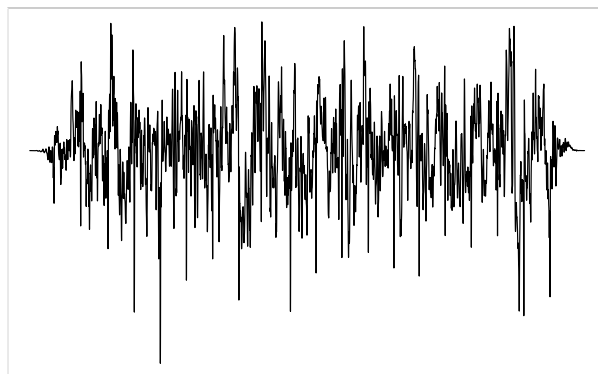
- *Stage 3*: it is also called deep sleep or slow-wave sleep (SWS). In

this stage delta waves (0.5-4 Hz) begin to occur and then become predominant. Dreaming is more common in this stage than in other stages of NREM sleep though not as common as in REM sleep.



*Figure 7.3: EEG signal during stage 3.*

*REM* sleep, by contrast, is defined by EEG activation, muscle atonia, and episodic bursts of rapid eye movements. The mental activity of human REM sleep is associated with dreaming.



*Figure 7.4: EEG signal during REM stage.*

The *polysomnography* (PSG) is a diagnostic test in which various bio-paramters of the patient are measured during sleep. We are going

to concentrate on brain activity, quantified by *electroencephalograms* (EEGs). PSG is used to detect sleep disorders including narcolepsy, idiopathic hypersomnia, periodic limb movement disorder (PLMD), REM behavior disorder, parasomnias, and sleep apnea. Some conditions are revealed by labeling the sleep stages of the patient and analyzing the equilibrium in the sleep cycles.

The assigning of sleep stages is currently done by sleep technicians that must manually label the sleep state for each epoch (a 30 second interval) based on the PSG multi-channel time series. The labeling of the stage is performed every 30 seconds of polysomnography. This process is onerous and subjected to human error.

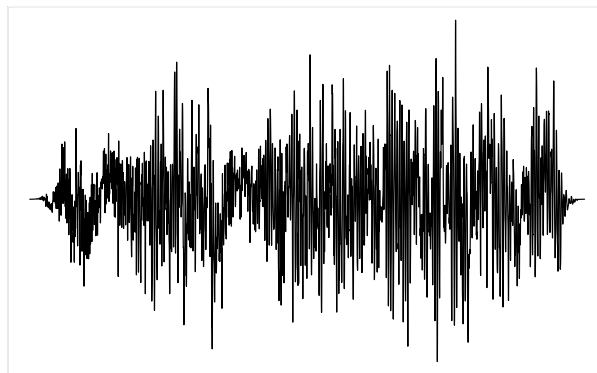


Figure 7.5: EEG signal during wakefulness.

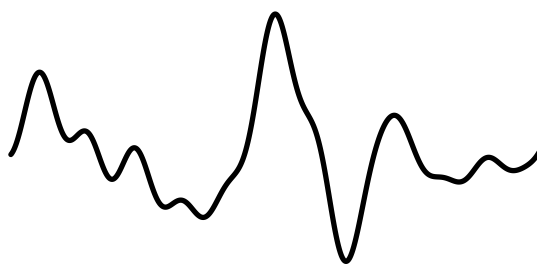
### 7.1.1 Automatic classification of sleep stages: state of the art

In recent years, many models for the automatic sleep scoring have been proposed. The division in sleep stages changes in the literature, both because each different division serves different applications and also to obtain the best numerical results. For example Persistent, Sinha [56] proposed a division in epoques of sleep spindles, REM, and awakeness, obtaining a classification accuracy of 95.35%. Subasi et al. [59] trained an artificial neural network based on wavelet features, with a precision of over 90 % for epoques with the labels: “drowsy”, “alert ”and “sleep ”. Wavelet and ANN have been used also by Ebrahimi et al. [20], using

the standard division of sleep stages and with an accuracy of 93%. Hsu et al. [30] employed an ANN trained with energy-based features, with a precision of 90.93%. Also Hidden Markov Models have been exploited for this task, by Doroshenkov et al [19]:their method is really good in classifying the REM stage. A combination of persistent homology and random forests is has been performed by Tymochko et al, in [63], with a maximum performance of 90%.

## 7.2 K-complexes (KCs)

KC is one of the distinctive features of NREM sleep and represents the EEG grapho-element with highest amplitude during normal sleep[13]. It was discovered more than 70 years ago by Loomis et al. It is characterized by a short positive transient followed by a slower, larger surface-negative complex with peaks at 350 and 550 ms, and then a final positivity peaking near 900 ms [58]. The frequency of a K-complex is estimated to be between 8 and 16 Hz. The K-complex represents the synchronized output of a network of cortical cells and therefore has the potential to be used as a marker for the functional integrity of the central nervous system. [12]



*Figure 7.6: A K complex.*

The presence of K-complexes is one of the characteristic markers of Stage 2 sleep. However, Stage 2 can also be recognized by other transient waveforms, such as sleep spindles. Although the detection of these particular kind of waveforms can be done by visual inspection

(following the guidelines of the American Association of Sleep Medicine, [31]), the procedure is time consuming and requires experience.

### 7.2.1 Automatic detection of K-complexes: state of the art

Automatic detection of K-complex combines machine learning techniques as SVM, decision trees and artificial neural network with various types of feature extraction and signal filtering [4]. A successful scheme has been to extract features from the frequency spectra through the wavelet transform, as shown by Tang and Ishii in [35], by Krohne et al. in [46], and by Patti et al. in [36], with precisions of 87%, 84% and 81.57% respectively. An example of neural network trained for the task is presented by Bankman et al. in [5]. They reached a peak accuracy of 90%. On the other hand, statistical methods comprehend Hidden Markov Models, implemented by Kam et al. in [32] (they found a sensitivity of 85%) and likelihood thresholding, performed by Devuyt et al. in [16], with an average sensitivity of 60.94%. Another approach comes from a plainer signal analysis; relevant features and connected outcomes are analyzed by Erdamar et al. in [29], reaching an accuracy of 91%. Finally, fractal-based features have been exploited by Al-Salman et al. with good results in [4], with an estimated precision of around 95% over their dataset.

## 7.3 From Polysomnography to Point Clouds

Data has been provided by the sleep research group of San Raffaele Hospital, in Milan. The dataset consists of 10 polysomnographies; each of them traces brain activity for the duration of an entire night. The file extension for the EEG data is `.edf`.

Each EEG has been imported in R [1] through the package `edfReader` [64], and divided in *epoques*, that is, periods of 30 seconds. Each `.edf` file contains multiple signals, coming from the different electrodes applied to the patients. The choice of the electrodes has been the following:

- For the automatic classification of sleep stages we used the electrode `O2-A1`, as it is positioned near the eye of the patient. The idea is to use this vicinity to have a better classification of NREM vs REM stages;



- For the automatic classification of K-complex we employed the electrode F4-A1, as it is the nearest to the frontal cortex. The idea is to catch K-complexes at the highest possible voltage in order to have a sharper definition of the characteristic transient.

The values of the EEG at a give instant suffer from measurement noise; we can eliminate part of it by applying a band-pass filter, getting rid of non-important highest and lowest frequencies. A theoretical result ensures that we do not lose additional information due to the low sampling frequency of the signal.

**Theorem 15** (Nyquist-Shannon sampling Theorem). *If a function  $f(t)$  contains no frequencies higher than  $W$  cps (characters per second), it is completely determined by giving its ordinates at a series of points spaced  $W/2$  seconds apart.*

The Nyquist-Shannon sampling Theorem [54] states that a band-limited signal can be recovered (exactly) from a sequence of observations whenever the sampling frequency is greater than twice the position, in the frequency domain, of the limiting band. `.edf` files provide a sampling frequency of 256 Hz. As we will see in greater detail, the maximum frequency allowed by the band-pass filter is 30 Hz. In this context the sampling Theorem holds.

### 7.3.1 Embedding parameters selection

The EEG dataset is composed by a collection of brain signals, that can be viewed as functions  $f$  from  $\mathbb{R}$  (time) to  $\mathbb{R}$  (voltage). We aim at embedding these signals in a higher dimensional space, in order to catch topological and chaotic features from the shape of the phase space. We remind that the sliding window vector takes the form  $SW_{M,\tau}f(t) = [f(t), f(t + \tau), f(t + 2\tau), \dots, f(t + M\tau)]$ . The parameters for this transformations are therefore the *embedding dimension*  $M + 1$  and the *delay*  $\tau$ . There are multiple ways to select satisfactory values for these parameters. Let us investigate some.

**Embedding dimension.** The embedding dimension must be large enough so that Takens' theorem holds. At the same time, for computational purposes it could be convenient to choose a reasonably small embedding dimension. This tradeoff can be performed optimally by knowing the dimensionality  $d$  of the chaotic set described by the phase

space of the dynamical system, and choose  $M > d$ . As we will see in Chapter 9, the appropriate dimension of the attractors can be non-integer. Two quantities that estimate this fractal dimension are the box-counting dimension and the correlation dimension.

**Definition 55** (Box-counting dimension). Let  $F \subseteq \mathbb{R}^d$  and  $\delta > 0$ . Let  $N_\delta(F)$  be the smallest number of sets with diameter less than or equal to  $\delta$  that are needed for covering  $F$ . Then the *box-counting dimension* of  $F$  is the limit

$$\dim_{\mathcal{B}}(F) = \lim_{\delta \rightarrow 0^+} \frac{\ln N_\delta(F)}{-\ln \delta}$$

if it exists.

The limit of the box-counting dimension is that it is not sensible to the density of the orbits, but only to the overall geometrical shape of the set. For this reason the correlation dimension can be preferable in specific situations.

**Definition 56** (Correlation dimension). Let  $\psi(i) = [\phi(i), \phi(i+\tau), \phi(i+2\tau), \dots, \phi(i+(m-1)\tau)] \in \mathbb{R}^m$  be a time delay embedding,  $N$  be the number of considered states  $\psi(i)$ , and  $\epsilon$  be a threshold distance. The *correlation sum* is defined as

$$C_N(\epsilon) = \frac{1}{N} \sum_{i,j=1, i \neq j}^N \Theta(\epsilon - \|\psi(i) - \psi(j)\|).$$

The *correlation integral* is the mean probability that the states at two different times are close:

$$C(\epsilon) = \lim_{N \rightarrow \infty} C_N(\epsilon).$$

As  $N \rightarrow \infty$  and  $\epsilon \rightarrow 0$  the correlation integral will take the form:

$$C(\epsilon) \sim \epsilon^\nu$$

We call  $\nu$  the *correlation dimension* of the phase space generated by  $\phi$ .

**Delay.** The delay  $\tau$  quantifies the "unfolding" of the phase space in  $\mathbb{R}^{M+1}$ . [33]. If  $\tau$  is too small, we observe the phenomenon of *redundance*: the attractor is compressed along the identity line. On the other end, if  $\tau$  is too big, we may lose the connectedness of the dynamics, thus ruining also its topological features. This phenomenon is

called *irrelevance* [33]. A good delay value has to be reasonably small, to avoid the folding of the attractor onto itself; at the same time it has to be large enough to avoid redundancy. This tradeoff can be addressed adequately using the notion of autocorrelation.

**Definition 57** (Autocorrelation). For a discrete-time deterministic real signal  $f$ , the *autocorrelation*  $R$  at *delay*  $\tau$  is defined as

$$R_{ff}(\tau) = \sum_{n \in \mathbb{Z}} f(n)f(n - \tau)$$

Autocorrelation is the correlation of a signal with a delayed copy of itself as a function of delay. A good choice for  $\tau$  in Takens' embedding is therefore the value for which the autocorrelation becomes smaller than a certain threshold ( $\frac{1}{e}$  in our case), as  $R_{ff}(\tau) = 0 \Rightarrow f(t)$  and  $f(t - \tau)$  are linearly independent on average, avoiding redundancy.

A better criterion to choose  $\tau$  is to select the value for which the auto mutual information (AMI) becomes smaller than a certain threshold. In fact, as the autocorrelation is a linear statistic, it ignores nonlinear dynamical correlations [24]. AMI can be thought as a generalized non-linear version of the autocorrelation.

**Definition 58** (Average Mutual Information). Given a univariate discrete-time signal  $f$ , divide the range of possible values of the signal in  $N$  bins. Let  $p_i$  the probability that  $f(t)$  is in bin  $i$ , and  $p_{i,j}(\tau)$  the probability that  $f(t)$  is in bin  $i$  and  $f(t + \tau)$  is in bin  $j$ . The *auto mutual information* is then defined as

$$I_{ff}(\tau) = \sum_{i,j} p_{i,j}(\tau) \log \left( \frac{p_{i,j}(\tau)}{p_i p_j} \right).$$

## 7.4 Embedding of Epoques

For the problem of detection of sleep stages, the entire epoques (30 seconds) are embedded in a 5 dimensional space. This choice is supported by the computation of the box-counting and correlation dimensions and permits the unfolding of all the chaotic sets described by the brain waves. The Sliding Window delay is chosen with the autocorrelation method 57, for computational purposes. For this problem we apply a band-pass filter (using package `seewave` [60]) with a band between 0.5 and 30 Hz, in order to distinguish the features of each sleep stage. In the images below the phase spaces have been projected in a 3D space and plotted with the package `rgl` [3]. The labels for the different sleep stages are: WAKE (wakefulness), NREM1, NREM2, NREM3 (non-REM stages 1,2,3), and REM.

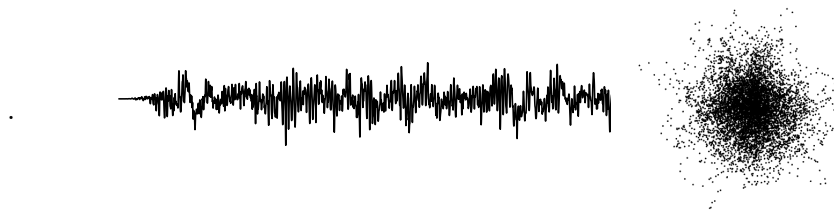


Figure 7.7: A Wakefulness and the corresponding 3-dimensional embedding



Figure 7.8: A Stage NREM1 epoqe and the corresponding 3-dimensional embedding.

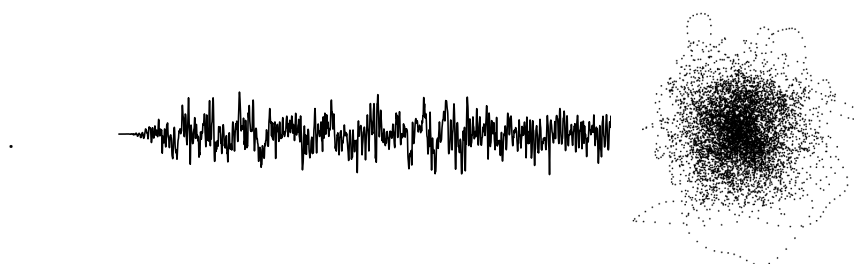


Figure 7.9: A Stage NREM2 epoch and the corresponding 3-dimensional embedding.

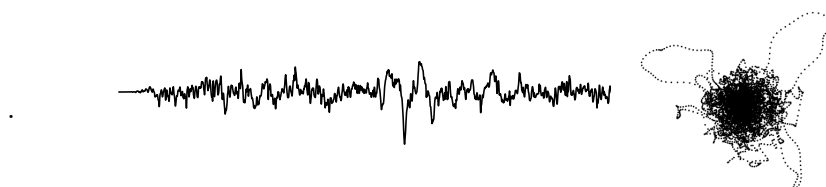


Figure 7.10: A Stage NREM3 epoch and the corresponding 3-dimensional embedding.



Figure 7.11: A REM stage epoch and the corresponding 3-dimensional embedding.

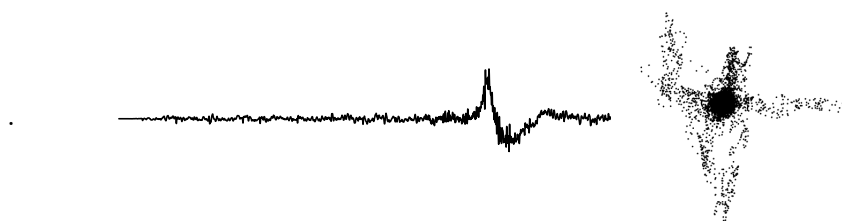


Figure 7.12: An epoch with a very sharp K-complex and the corresponding 3-dimensional embedding.

We can appreciate at sight some differences between the different phase spaces. In wakefulness epochs we have a sparse point cloud as

the signal is strong rich of frequencies. Point clouds from stage NREM1 is characterized by more compact point clouds. In stage NREM2 K-complexes and sleep spindles result in big cycles that leave the center and then come back. Point clouds from stage NREM3 are very definite as higher frequencies disappear. Finally, point clouds from REM stages present some sort of "spikes" due to fast and low-voltage oscillations.

## 7.5 Embedding of Waveforms

For the problem of detection of K complexes, from each epoch we extracted the 5 highest peaks. Each peak is the center of a signal fraction with length equal to 6 seconds. This permits also the recognition of longer K-complexes. The 6-second signals are embedded in a 5 dimensional space, in order to recover the corresponding phase spaces. The Sliding Window delay is chosen with the autocorrelation method [57](#). For this problem we apply a band-pass filter (using package `seewave` [\[60\]](#)) with a band between 0.5 and 2.3 Hz, in order to remove noise and catch only the characteristic shape of the waveform. The figures below are again projected in a 3D space.

The K-complex waveform is always embedded in a large cycle. This cycle results in a high persistence 1-dimensional homology class. Moreover, the time in which the K-complex happens is "stolen" from other smaller oscillations. This means that point clouds in which the K-complex is not present possess a multitude of low persistence 1-dimensional homology classes. This feature will be crucial in the classification task.

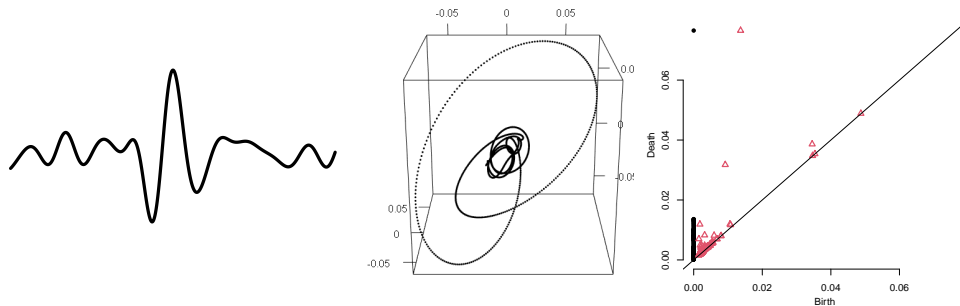


Figure 7.13: *K1*: example of a *K*-complex waveform (left), its three dimensional embedding (center) and the corresponding persistence diagram (right).

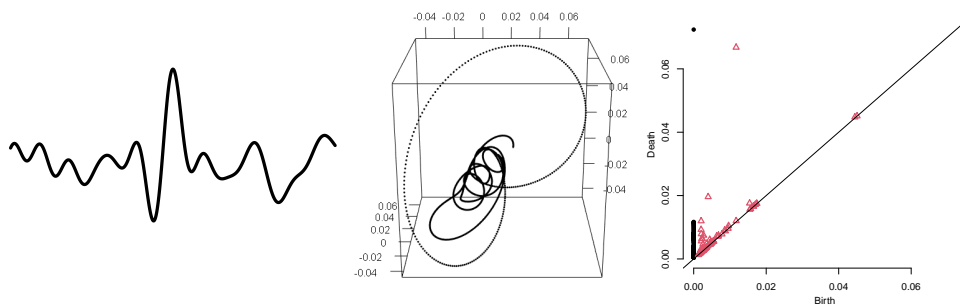


Figure 7.14: *K2*: example of a *K*-complex waveform (left), its three dimensional embedding (center) and the corresponding persistence diagram (right).

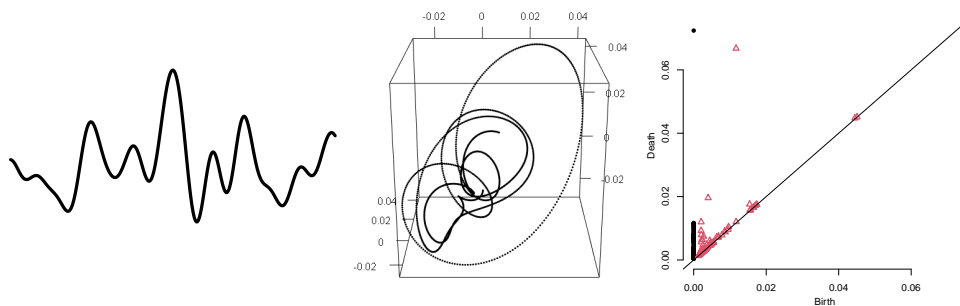


Figure 7.15: *K3*: example of a *K*-complex waveform (left), its three dimensional embedding (center) and the corresponding persistence diagram (right).

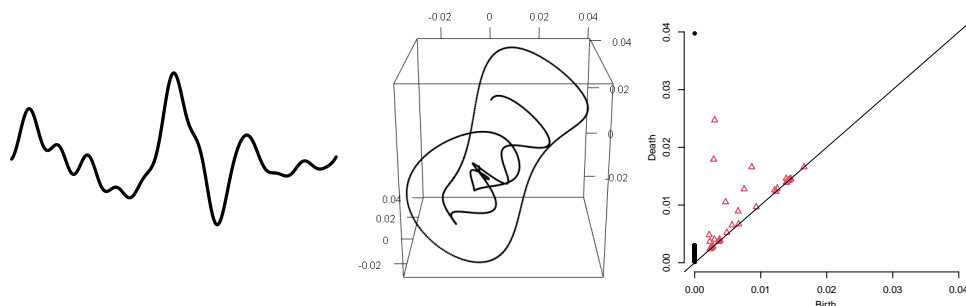


Figure 7.16:  $K_4$ : example of a K-complex waveform (left), its three dimensional embedding (center) and the corresponding persistence diagram (right).

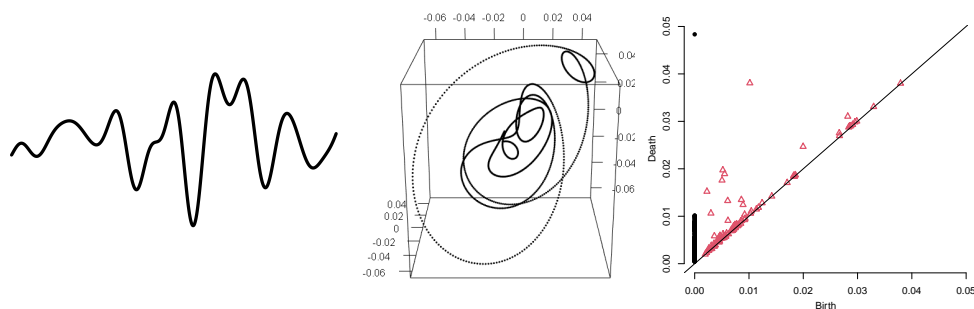


Figure 7.17: non-K1: example of a non-K-complex waveform (left), its three dimensional embedding (center) and the corresponding persistence diagram (right).

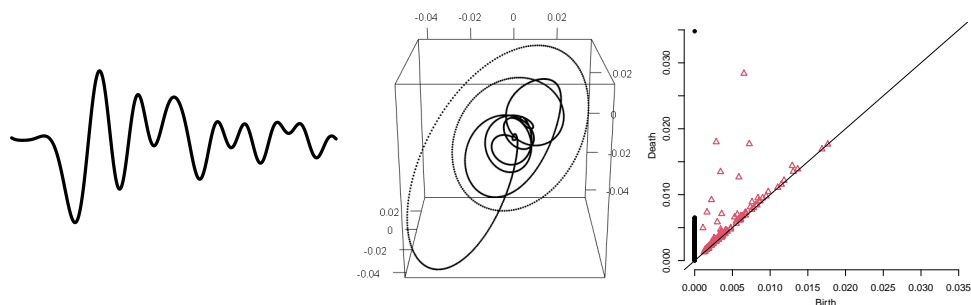


Figure 7.18: non-K2: example of a non-K-complex waveform (left), its three dimensional embedding (center) and the corresponding persistence diagram (right).



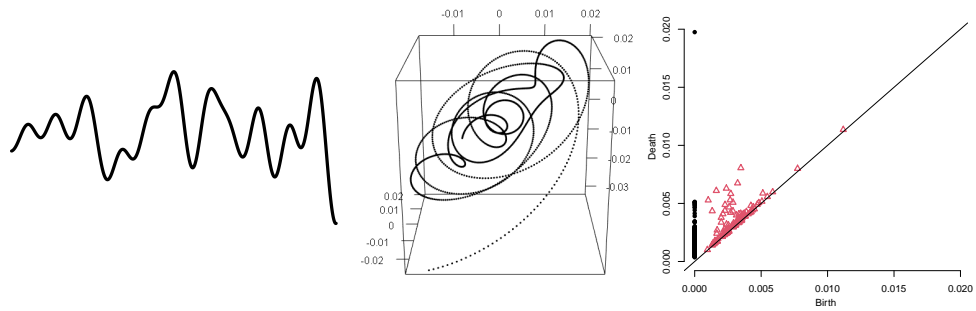


Figure 7.19: non-K3: example of a non-K-complex waveform (left), its three dimensional embedding (center) and the corresponding persistence diagram (right).

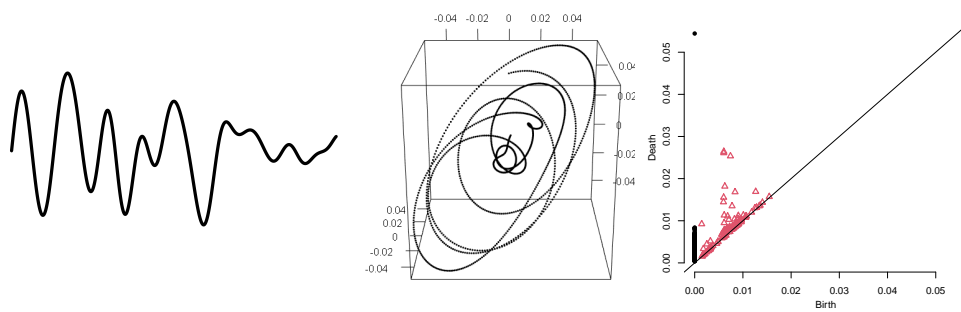


Figure 7.20: non-K4: example of a non-K-complex waveform (left), its three dimensional embedding (center) and the corresponding persistence diagram (right).



## Chapter 8

# Methodology

This Chapter presents the operative scheme used to solve the detection problem. We choose to implement *random forest* classifiers for two reasons. First, random forest and in general, decision trees, provide a fascinating theoretical foundation that connects basic concepts of machine learning to physical and philosophical ones, such as entropy and description length. The second and most important reason is their operativeness, consequent to their ability to not overfit and automatically perform cross-validation. The texts that inspired this Chapter are [53, 40, 6].

### 8.1 Main concepts of supervised Machine Learning

**Definition 59** (Statistical learning framework). The statistical learning model used in this thesis makes use of the following concepts and definitions:

1. **Learner's input:** the set of objects we wish to label is called *instance space*  $\mathcal{X}$ . Usually the elements of  $\mathcal{X}$  are vectors of *features*, or *instances*, and belong to  $\mathbb{R}^N$  for some  $N \in \mathbb{N}$ .

Our *label set* is  $\mathcal{Y} = \{0, 1\}$ .

The *training data* is a finite sequence of labeled domain points and belongs to the *training set*  $\mathcal{S} = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ .

2. **Learning algorithm:** a function  $\mathcal{A} : \mathcal{S} \rightarrow \{h : \mathcal{X} \rightarrow \mathcal{Y}\}$  that given a training set, outputs a rule to classify data in the instance space.

3. **Learner's output:** the learner outputs a *prediction rule*, or *classifier*  $\mathcal{A}(\mathcal{S}) = h : \mathcal{X} \rightarrow \mathcal{Y}$ . This function can be used to predict the label of new domain points.
4. **Data generating distribution:** we assume there is some joint probability distribution over domain points and labels  $\mathcal{D} : \mathcal{X} \times \mathcal{Y} \rightarrow [0, 1]$ .
5. **The i.i.d. assumption:** we assume that the examples in the training set are independently and identically distributed according to the marginal distribution  $\mathcal{D}_{\mathcal{X}}$ .
6. **Measures of success:** for a probability distribution  $\mathcal{D}$  over  $\mathcal{X} \times \mathcal{Y}$  one can measure how likely  $h$  is to predict wrongly when labeled points are randomly drawn from  $\mathcal{D}$ . We define the *true error*, or *true risk* of  $h$  as:

$$L_{\mathcal{D}}(h) \stackrel{\text{def}}{=} \mathbb{P}_{(x,y) \sim \mathcal{D}}(h(x) \neq y) \stackrel{\text{def}}{=} \mathcal{D}(\{(x, y) : h(x) \neq y\}).$$

We are looking for a predictor  $h$  such that this error is minimized. However, the learner does not know the data generating distribution  $\mathcal{D}$ . For this reason we are brought to define another error, called *empirical error* of  $h$ , based on the training data  $\mathcal{S}$ :

$$L_{\mathcal{S}}(h) \stackrel{\text{def}}{=} \frac{|\{i \in [m] : h(x_i) \neq y_i\}|}{m}.$$

Given  $\mathcal{S}$ , a learner can compute  $L_{\mathcal{S}}(h)$  for any function  $h : \mathcal{X} \rightarrow \mathcal{Y}$ .

7. **Informal goal:** to find a prediction rule,  $h$ , such that with high probability  $h$  approximately minimizes  $L_{\mathcal{D}}$ . We will search  $h$  over a restricted space,  $\mathcal{H} \subseteq \{\text{all possible functions } f : \mathcal{X} \rightarrow \mathcal{Y}\}$ .  $\mathcal{H}$  is called *hypothesis class*.

At this point, we have to decide a criterion to select our predictor,  $h$ . The obvious best choice would be the minimizer of the true error  $L_{\mathcal{D}}$  but as we said, our only knowledge is the training data. A possible criterion is therefore the *empirical error minimization*, or ERM rule:

**Definition 60** (ERM rule). For a given hypothesis class  $\mathcal{H}$ , and a training sample  $\mathcal{S}$ , the  $ERM_{\mathcal{H}}$  learner uses the *ERM rule* to choose a predictor  $h \in \mathcal{H}$  with the lowest possible error over  $\mathcal{S}$ :

$$ERM_{\mathcal{H}}(\mathcal{S}) \in \underset{h \in \mathcal{H}}{\operatorname{argmin}} L_{\mathcal{S}}(h)$$

As we do not know the real data distribution  $\mathcal{D}$ , minimizing the empirical error could lead to a classifier that fits well the data in  $\mathcal{S}$  but that does not catch the patterns of the real distribution. The phenomenon of low empirical risk and high true risk is commonly known as overfitting. Formally, given a hypothesis class  $\mathcal{H}$ , a hypothesis  $h$  is said to *overfit* the data if there exists some alternative hypothesis  $h' \in \mathcal{H}$  such that  $h$  has a smaller error than  $h'$  over the training examples, but  $h'$  has a smaller error than  $h$  over the entire distribution of instances.

Next section introduces the idea that if new information or assumptions are added, then the criterion to choose the best classifier  $h$  in the hypothesis class  $\mathcal{H}$  can be changed. SRM (Structural Risk Minimization) is an example of alternative of this kind to ERM.

## 8.2 Minimum Description Length

**The Structural Risk Minimization Paradigm.** We have seen that the choice of the hypothesis class  $\mathcal{H}$  results in different trade-offs between approximation and estimation errors. The Structural Risk Minimization (SRC) learner is used in decision trees theory and it is based on two concepts:

1. We assume that the set  $\mathcal{H}$  can be written as a countable union of classes  $\mathcal{H} = \bigcup_{n \in \mathbb{N}} \mathcal{H}_n$ . Moreover we assume that for each  $n$ , the class  $\mathcal{H}_n$  enjoys the *uniform convergence property* with *sample complexity*  $m_{\mathcal{H}_n}^{UC}(\epsilon, \delta)$ . This means that if  $\mathcal{S}$  is a sample of  $m \geq m_{\mathcal{H}_n}^{UC}(\epsilon, \delta)$  examples drawn i.i.d. according to  $\mathcal{D}$ , then, with probability of at least  $1 - \delta$ ,

$$\forall h \in \mathcal{H}, |L_{\mathcal{S}}(h) - L_{\mathcal{D}}(h)| \leq \epsilon.$$

2. We specify preferences over the different classes  $\mathcal{H}_n$  through a weight function  $w : \mathbb{N} \rightarrow [0, 1]$ : the higher is the weight of a class, the more we prefer it. Finally, we impose the condition  $\sum_{n=1}^{\infty} w(n) \leq 1$ .

Once we fix a sample size  $m$ , we are interested in the lowest possible upper bound on the difference between empirical and true risk. To do that we define the function:

$$\epsilon_n : \mathbb{N} \times (0, 1) \rightarrow (0, 1), \quad \epsilon_n(m, \delta) = \min\{\epsilon \in (0, 1) : m_{\mathcal{H}_n}^{UC}(\epsilon, \delta) \leq m\}.$$

From the definition of uniform convergence and  $\epsilon_n$ , it can be proved that for every  $m$  and  $\delta$ , with probability at least  $1 - \delta$  over the choice of  $S \sim \mathcal{D}$  we have that

$$\forall h \in \mathcal{H}_n, |L_{\mathcal{D}}(h) - L_S(h)| \leq \epsilon_n(m, w(n) \cdot \delta).$$

**SRM for singleton classes.** Let's stick to a case that is useful to us. Let's assume that  $\mathcal{H} = \bigcup_{n \in \mathbb{N}} \{h_n\}$  is a countable union of singleton classes : then each singleton class satisfies  $\epsilon_n(m, \delta) = \sqrt{\frac{\log(2/\delta)}{2m}}$ . Moreover, as now we can think of  $w$  as a function from  $\mathcal{H}$  to  $[0,1]$ , we can state that the output of the SRM paradigm for singleton classes is the following prediction rule:

$$SRM_{\mathcal{H}}(S) \in \operatorname{argmin}_{h \in \mathcal{H}} \left[ L_S(h) + \sqrt{\frac{-\log(w(h)) + \log(2/\delta)}{2m}} \right].$$

The difference between the ERM and the SRM paradigm is that with SRM we are willing to trade part of the bias towards a low empirical risk towards classes with a small  $\epsilon_n(m, w(n) \cdot \delta)$ , hoping to get a smaller estimation error.

**Minimum description length.** The last step needed to have an operative paradigm is to define the weight function in a convenient way. The Minimum Description Length (MDL) paradigm is a particular type of SRM in which the weight function is derived from the *length* of the hypothesis  $h$ , once they are translated in some description language.

**Definition 61** (Description language). Let  $\mathcal{H}$  be a hypothesis class. An *alphabet* is a finite set of symbols and it is denoted by  $\Sigma$ . A *string*  $\sigma$  is a finite sequence of symbols from  $\Sigma$ . The *length* of a string is denoted by  $|\sigma|$  and  $\Sigma^*$  is the set of all finite length strings.

A *description language* for  $\mathcal{H}$  is a function  $d : \mathcal{H} \rightarrow \Sigma^*$ .  $d(h)$  is called the *description of  $h$*  and it is a string of length  $|h|$ .

A description language is *prefix-free* if for every distinct  $h, h'$ ,  $d(h)$  is not a prefix of  $d(h')$ .

**Lemma 16** (Kraft Inequality). *If  $\mathcal{J} \subseteq \{0,1\}^*$  is a prefix-free set of strings, then  $\sum_{\sigma \in \mathcal{J}} \frac{1}{2^{|\sigma|}} \leq 1$ .*

So, once we choose a description language on  $\mathcal{H}$ , Kraft's inequality gives us the possibility to define a weight function  $w$  over that hypothesis class, simply taking  $w(h) = \frac{1}{2^{|h|}}$ . This yields the following:

**Theorem 17.** *Let  $\mathcal{H}$  be a hypothesis class and let  $d : \mathcal{H} \rightarrow \{0, 1\}^*$  be a prefix-free description language for  $\mathcal{H}$ . Then, for every sample size  $m$  every  $\delta > 0$ , and every probability distribution  $\mathcal{D}$ , with probability greater than  $1 - \delta$  over the choice of  $\mathcal{S} \sim \mathcal{D}^m$  we have that*

$$\forall h \in \mathcal{H}, \quad L_{\mathcal{D}}(h) \leq L_{\mathcal{S}}(h) + \sqrt{\frac{|h| + \log(2/\delta)}{2m}}$$

where  $|h|$  is the length of  $d(h)$ .

This theorem suggests that the output of the MDL paradigm is the result of the tradeoff between empirical risk and shorter description length:

$$MDL_{\mathcal{H}}(\mathcal{S}) \in \operatorname{argmin}_{h \in \mathcal{H}} \left[ L_{\mathcal{S}}(h) + \sqrt{\frac{|h| + \log(2/\delta)}{2m}} \right].$$

**Occam's Razor.** To end the section we try to answer to this naturally arising question: why it is better to classification rules with shorter description length? A philosophical answer can be given by the Occam's razor principle (stated around the year 1320), which is believed to belong to William of Ockam, and reads:

*A short explanation tends to be better than a long one.*

The MDL paradigm can be seen as application of this principle. We make two considerations on the issue. The first is that we decide what "short" means through a description language. The advantages of MDL depend totally on our capacity to model the complexities of the hypotheses  $h$ . The second aspect is that we still rely on the presumption that nature generally acts in this way, with its own description language which we try to mimic. This assumption can be true on average but false in specific cases.

### 8.3 Decision Trees

Previous sections were devoted in presenting the theoretical aspect of SRM and MLD for a generic hypothesis class  $\mathcal{H}$ . In this section we apply the theory to the hypotheses  $h$  involved in the classification tasks introduced in Chapter 7: decision trees. A **decision tree** is a predictor  $h : \mathcal{X} \rightarrow \mathcal{Y}$ , that uses a tree structure to predict the label associated with the instance  $\mathbf{x}$  by travelling from a root node of a tree to a leaf. In general, decision trees represent a disjunction of conjunctions of constraints on the attribute values of instances. Each path from the tree root to a leaf corresponds to a conjunction of attribute test. Each test, and so, every path split is given by comparing the attribute values to a threshold value.

In the following we will consider the case of binary classification, that is  $\mathcal{Y} = \{0, 1\}$ .

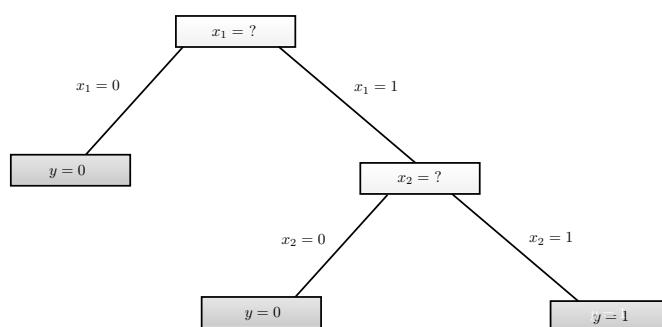


Figure 8.1: Example: a simple binary classification decision tree with binary instances.

**Decision trees with binary instances.** For simplicity we'll start by assuming that each instance is a vector of  $k$  bits, that is,  $\mathcal{X} = \{0, 1\}^k$ . We notice that any classifier  $h : \{0, 1\}^k \rightarrow \{0, 1\}$  can be represented by a tree with  $2^k$  leaves and depth  $k + 1$ . This means that the size of the hypothesis class  $\mathcal{H}$  grows exponentially with  $k$ , and a ERM paradigm is not suggested. Moreover, in this simplified setting thresholding the value of a single attribute value corresponds to the splitting rule  $\mathbb{1}_{\{x_i=1\}}$  for some  $i \in \{1, \dots, k\}$ .

To partially overcome this issue we rely on the MDL scheme. The set



of all possible decision trees is in fact countable, so we can construct a suitable description language on the basis of the intuition that *smaller trees should be preferred over larger ones*.

A possible description language is the following: if  $h$  is a tree with  $n$  nodes, we build  $d(h)$  as a sequence of  $n + 1$  vectors, each one of size  $\log_2(k + 3)$  bits. Each block indicates the state of each node:

- The node is an internal node of the form  $\mathbb{1}_{\{x_i=1\}}$  for some  $i \in \{1, \dots, k\}$ ;
- The node is a 1-valued leaf;
- The node is a 0-valued leaf;
- End of the code ( $k + 1$  - th block).

Since there are  $d + 3$  possibilities for each node, we use  $\log_2(k + 3)$ -sized blocks. Assuming without loss of generality that every internal node has two children, we have that the previous is a prefix-free encoding of the tree, and with this encoding a tree with  $k$  nodes has a length of  $\log_2(k + 3)$ . By Theorem 17, it follows that for every  $m$  and every decision tree  $h \in \mathcal{H}$  with  $k$  nodes, the bound

$$L_{\mathcal{D}}(h) \leq L_{\mathcal{S}}(h) + \sqrt{\frac{\log_2(k + 3) + \log(2/\delta)}{2m}}.$$

holds with a probability of at least  $1 - \delta$ .

This equation represents a tradeoff: from one hand we want to minimize the empirical error as in the ERM paradigm, but from the other hand we prioritize small trees; this leads to a raise of our empirical error but hopefully to a decrease of the true error.

**The CART algorithm.** The MDL scheme suggests a theoretical learning rule for decision trees. Unfortunately, if  $NP \neq P$ , it can be proven that no algorithm can minimize the right side of the bound in polynomial time.

For this reason, general algorithms that aim to treat trees of different sizes are based on heuristics: the tree is constructed gradually, and locally optimized decisions are made at the construction of each node. With these *greedy approach* algorithms we are not obtaining the best tree but hopefully one that works well in practice.

A possible implementation of this strategy is the CART algorithm by Breiman, Friedman Olshen, and Stone [7]. It is based on these fundamental steps:

- We start with a single leaf, the root, labeled with a majority vote among all labels over the training set;
- We perform a series of iterations, and at each iteration, we evaluate the effect of splitting a single node. To do that, we define a *gain measure* that quantifies the improvement due to the split.
- Among all the possible splits, we choose to perform the one that maximizes the gain, or we can decide to not perform the split at all.

CART works by recursive calls, with the initial inputs being the training set  $\mathcal{S}$  and a set of features  $\mathcal{F}$ . The Gain function, used to evaluate the gain of a split of the tree, receives as inputs  $\mathcal{S}$  and a feature index  $i$ .

**CART** ( $\mathcal{S}, \mathcal{F}$ )

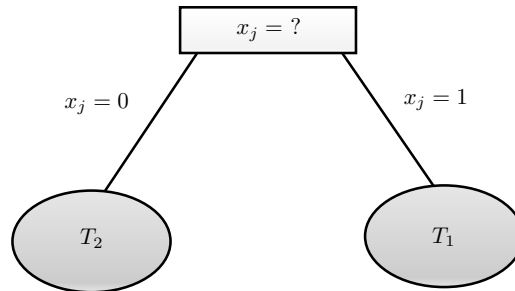
Breiman, Friedman Olshen, and Stone.

**Input:** Training set  $\mathcal{S}$ , set of features  $\mathcal{F} \in \{1, \dots, k\}$ 


---

**if** all samples in  $\mathcal{S}$  are labeled by 1, **return** 1.  
**if** all samples in  $\mathcal{S}$  are labeled by 0, **return** 0.  
**if**  $\mathcal{F} = \emptyset$ , **return** a leaf whose value = majority of labels in  $\mathcal{S}$   
**else**  
    Let  $j = \operatorname{argmax}_{i \in \mathcal{F}} \operatorname{Gain}(\mathcal{S}, i)$   
**if** all examples in  $\mathcal{S}$  have the same label, **return** a leaf whose value  
    = majority of labels in  $\mathcal{S}$  **else**  
    Let  $T_1$  be the tree returned by  $\operatorname{CART}(\{(\mathbf{x}, y) \in \mathcal{S} : x_j = 1\}, \mathcal{F}/\{j\})$ ;  
    Let  $T_2$  be the tree returned by  $\operatorname{CART}(\{(\mathbf{x}, y) \in \mathcal{S} : x_j = 0\}, \mathcal{F}/\{j\})$ ;  
**return** the tree:

---



**Gain Measure for CART.** Different algorithms use different implementations for  $\operatorname{Gain}(\mathcal{S}, i)$ . We present the one for CART.

Let  $\mathbb{P}_{\mathcal{S}}[B]$  be the probability that the event  $B$  holds with respect to the uniform distribution over  $\mathcal{S}$ . We now define a gain measure (the Gini impurity) based on statistics, and that can be interpreted as a measure of entropy of a physical system.

**Definition 62** (Entropies). The macroscopic state of a system is characterized by a distribution on the microstates. For a collection of classical particles with a discrete set of microstates, if  $E_i$  is the energy of the microstate  $i$  and  $p_i$  is the probability that it occurs during the

system's fluctuations (call  $p$  the set of all these probabilities), then the *Boltzmann-Gibbs entropy* of the system is

$$S(p) = -k \sum_i p_i \ln p_i.$$

A generalization of this entropy measure is the *Tsallis entropy*, which is dependent on a real parameter  $q$  (the *deformation coefficient*):

$$S_q = \frac{k}{q-1} \left( 1 - \sum_i p_i^q \right).$$

Notice that, in the limit as  $q \rightarrow 1$ , the Boltzmann-Gibbs entropy is recovered, namely  $S = S_1$ .

**Definition 63** (Gini impurity). The *Gini impurity* is a measure of how often a randomly chosen element from the set  $\mathcal{X}$  would be incorrectly labeled if it was randomly labeled according to the distribution of labels in the subset  $\mathcal{S}$ . For a set of items with 2 classes ( $i \in \{0, 1\}$ ), let  $p_i$  the fraction of items labeled with class  $i$  in  $\mathcal{S}$ . Then we have:

$$I_G = \sum_{i=1}^2 \left( p_i \sum_{k \neq i} p_k \right) = \sum_{i=1}^2 p_i (1 - p_i) = 2p_i(1 - p_i).$$

The Gini impurity corresponds to Tsallis Entropy with deformation coefficient  $p = 2$ .

Now we are ready to define the Gain function. Notice that the training error *before* a split on the feature  $i$  is equal to

$$e_b(\mathcal{S}, i) = I_G(\mathbb{P}_{\mathcal{S}}[y = 1]),$$

whereas the error *after* the split has the form

$$e_a(\mathcal{S}, i) = \mathbb{P}_{\mathcal{S}}[x_i = 1] I_G(\mathbb{P}_{\mathcal{S}}[y = 1 | x_i = 1]) + \mathbb{P}_{\mathcal{S}}[x_i = 0] I_G(\mathbb{P}_{\mathcal{S}}[y = 1 | x_i = 0]).$$

We define the Gain function to be the difference between the two errors:

$$\text{Gain}(\mathcal{S}, i) = e_b(\mathcal{S}, i) - e_a(\mathcal{S}, i).$$

**Remark 6.** Until now we covered *binary trees*. This model can be easily generalized to include real-valued features, thus using split rules of the form  $\mathbb{1}_{\{x_i < \gamma\}}$ .

## 8.4 Random Forests

The MDL paradigm has the goal of trading off empirical error in order to get a smaller true error, that is, to reduce the chance of overfitting. Another technique that can be used in parallel to reduce the possibility of overfitting is using an ensemble of trees over different parts of the datasets and considering the "average decision tree" with the *random forest* method, introduced by Breiman in 2008 [6].

**Definition 64** (Random forest). A *random forest* is a classifier consisting of a collection of decision trees  $\{h(\mathbf{x}, \boldsymbol{\theta}_t, t = 1, \dots, T)\}$ , where the  $\{\boldsymbol{\theta}_t\}$  are i.i.d. random vectors and each tree outputs a vote for the classification of each new data sample, given the input  $\mathbf{x}$ . Then each new sample is labeled by the class that obtains the most votes.

Let  $k$  be the number of features. We generate  $\{\boldsymbol{\theta}_t\}$  in this way: first, we sample a new training set  $\mathcal{S}'$  using the uniform distribution over  $\mathcal{S}$ , and elements from  $\mathcal{S}$  with repetition. Then we construct a sequence  $I_1, I_2, \dots, I_T$ , where each  $I_t$  is a subset of  $\{1, \dots, k\}$  of size  $k'$  generated uniformly at random. Finally we grow a decision tree over  $\mathcal{S}'$  and with the subset of features  $I_t$ . Intuitively, if  $k'$  is reasonably small, this method should prevent overfitting.

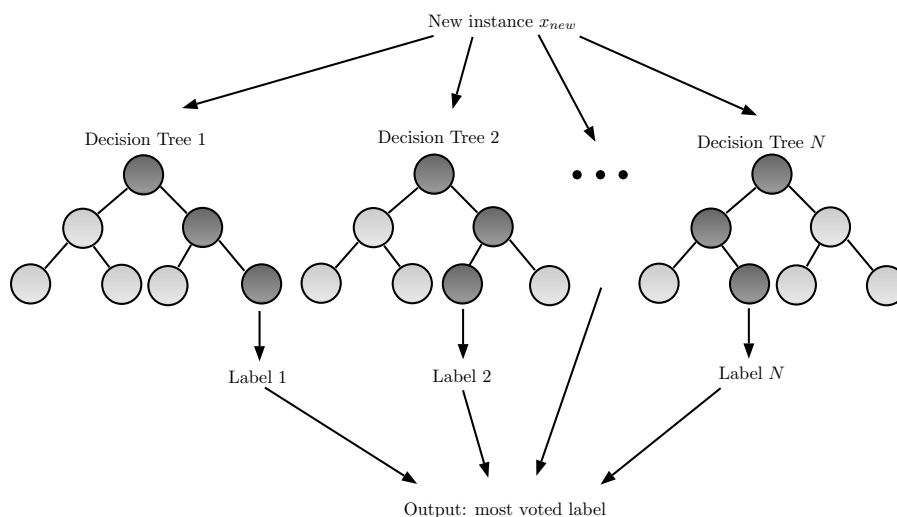


Figure 8.2: Scheme of the random forest classifier.

**Definition 65** (Margin). Given an ensemble of classifiers  $h_1(\mathbf{x}), h_2(\mathbf{x}), \dots, h_T(\mathbf{x})$  and with the training set  $\mathbf{X}, \mathbf{Y} = \{1, \dots, J\}$  drawn at random from  $\mathcal{D}$ ,

we define the margin function as

$$\text{mg}(\mathbf{X}, \mathbf{Y}) = \text{avg}_t \mathbb{1}_{\{h_t(\mathbf{X})=\mathbf{Y}\}} - \max_{j \neq \mathbf{Y}} \text{avg}_t \mathbb{1}_{\{h_t(\mathbf{X})=j\}}.$$

The margin measures how much the average number of votes at  $\mathbf{X}, \mathbf{Y}$  for the right class exceeds the average vote for any other class. The higher the margin, the higher is the confidence in our classification. For our specific case, that is, random forest with label set  $\mathbf{Y} = \{0, 1\}$ , the margin has the simpler form

$$\text{mg}(\mathbf{X}, \mathbf{Y}) = 2\mathbb{P}_{\Theta}(h(\mathbf{X}, \Theta) = \mathbf{Y}) - 1.$$

The following result shows how random forests do not overfit as the number of trees increases, as the error converges to a limiting value.

**Theorem 18.** *Let us define the generalization error as*

$PE^* = \mathbb{P}_{\mathbf{X}, \mathbf{Y}}(\text{mg}(\mathbf{X}, \mathbf{Y}) < 0)$ . *As the number of trees increases, for almost surely all sequences  $\{\Theta_1, \Theta_2, \dots\}$*

$$PE^* \rightarrow \mathbb{P}_{\mathbf{X}, \mathbf{Y}} [\mathbb{P}_{\Theta}(h(\mathbf{X}, \Theta) = \mathbf{Y}) - \max_{j \neq \mathbf{Y}} \mathbb{P}_{\Theta}(h(\mathbf{X}, \Theta) = j) < 0]$$

*Proof.* To prove the theorem we have to show that there is a set  $C$  of probability 0 on the sequence space  $\{\Theta_1, \Theta_2, \dots\}$  such that outside of  $C$  and for all  $\mathbf{x}$ ,

$$\frac{1}{N} \sum_{n=1}^N \mathbb{1}_{h(\Theta_n, \mathbf{x})=j} \rightarrow \mathbb{P}_{\Theta}(h(\Theta, \mathbf{x}) = j).$$

For a fixed training set  $\mathbf{X}$  and fixed  $\Theta$ , the set of all  $\mathbf{x}$  such that  $h(\Theta, \mathbf{x}) = j$  is an union of hyper-rectangles. For all  $h(\Theta, \mathbf{x})$  there is only a finite number  $K$  of such hyper-rectangles, denoted by  $S_1, \dots, S_k$ . Define  $\phi(\Theta) = k$  if  $\{\mathbf{x} : h(\Theta, \mathbf{x}) = j\} = S_k$ . Let  $N_k$  be the number of times that  $\phi(\Theta_n) = k$  in the first  $N$  trials. Then

$$\frac{1}{N} \sum_{n=1}^N \mathbb{1}_{h(\Theta_n, \mathbf{x})=j} = \frac{1}{N} \sum_k N_k \mathbb{1}_{\mathbf{x} \in S_k}.$$

By the law of Large Numbers,

$$\frac{1}{N} \sum_{n=1}^N \mathbb{1}_{\phi(\Theta_n)=k}$$

converges a.s. to  $\mathbb{P}(\phi(\Theta) = k)$ . Taking unions of all sets on which convergence does not occur for some value of  $k$  gives a set  $C$  of zero probability such that outside of  $C$ ,

$$\frac{1}{N} \sum_{n=1}^N \mathbb{1}_{h(\Theta_n, \mathbf{x})=j} \rightarrow \sum_k \mathbb{P}_{\Theta}(\phi(\Theta) = k) \mathbb{1}_{\mathbf{x} \in S_k}. \quad (8.1)$$

The right hand side of equation (8.1) is equal to  $\mathbb{P}_{\Theta}(h(\Theta, \mathbf{x}) = j)$ . This ends the proof.  $\square$





## Chapter 9

# Feature extraction and numerical results

The classification problems and the relative classifiers have been described in Chapters 7 and 8. In the following, we engage persistent homology and chaos theory to try to obtain precise classifiers. We develop six random forests. The first five are trained on epoques, and makes use of features coming from the computation of fractal dimension and signal correlations. Their task is to detect single sleep stages in an EEG and more in general to recognize the state of wakefulness or sleep. The last random forest is trained on shorter signals (6 seconds) and employs features coming from persistent homology.

### 9.1 Feature extraction for sleep stage detection: chaotic dynamical systems

By applying the Sliding Window Embedding to the EEG epoques, we uncover the true state spaces on which the trajectories run. The aesthetic complexity of these spaces varies with the sleep stage; the notion of fractal dimension can help in quantifying this concept. Fractals are *self-similar* sets, meaning that zooming in a specific points we can appreciate the repetition of the large scale structures, infinite times.

On the other side, a measure of the complexity is given by the minimum embedding dimension. We saw that a limit cycle can be drawn in  $\mathbb{R}^2$ . More complicated attractors present higher dimensionalities.

### 9.1.1 Fractal dimension

We begin by defining the Lebesgue dimension, that is the commonly known notion of dimension. Then, passing from the Hausdorff measure we will introduce non-integer dimensions.

**Definition 66** (Order, refinement). The *order* of an open cover of a topological space  $X$  is the smallest number  $n$  such that each point of the space belongs to, at most,  $n$  sets in the cover. A *refinement* of a cover  $C$  is another cover such that each of its sets is a subset of a set in  $C$ .

**Definition 67** (Lebesgue covering dimension). The *covering dimension* of a topological space  $X$  is defined to be the minimum value of  $n$  such that every open cover  $C$  of  $X$  has an open refinement with order  $n + 1$  or less. If no such minimal  $n$  exists, the space is said to be of *infinite covering dimension*.

**Definition 68** (Hausdorff measure). For any subset  $F$  of  $\mathbb{R}^n$  and any  $s \geq 0$  we define the *s-dimensional Hausdorff measure* of  $F$  as

$$\mathfrak{H}^s(F) = \lim_{\delta \rightarrow 0^+} \mathfrak{H}_\delta^s(F) = \sup_{\delta > 0} \mathfrak{H}_\delta^s(F)$$

where, for  $0 < \delta < +\infty$

$$\mathfrak{H}_\delta^s(F) = \frac{\omega_s}{2^s} \inf \left\{ \sum_{j=1}^{+\infty} (\text{diam} U_j)^s : F \subset \bigcup_{j=1}^{\infty} U_j, \text{diam} U_j \leq \delta \right\}.$$

For any  $s$  the constant

$$\omega_s = \pi^{s/2} \left( \int_0^\infty e^{-x} x^{s/2} dx \right)^{-1} = \frac{(\Gamma(1/2))^s}{\Gamma(1 + s/2)}$$

is positive and finite;  $\Gamma$  denotes the Euler Gamma function:  $\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$ .

**Definition 69** (Hausdorff dimension). The *Hausdorff dimension* of a set  $F \subset \mathbb{R}^n$  is defined as

$$\dim_{\mathfrak{H}}(F) = \inf\{s \geq 0 : \mathfrak{H}^s(F) = 0\} = \sup\{s \geq 0 : \mathfrak{H}^s(F) = +\infty\}$$

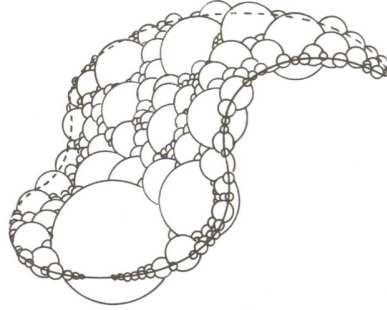


Figure 9.1: Example: the Hausdorff measure of a portion of surface is approximated by the sum of median section of small spheres (with diameter less than  $\delta$ ) that provide a covering of the surface.

**Definition 70** (Fractal Set). A set  $F \subset \mathbb{R}^d$  is called *fractal* if its Hausdorff dimension is non integer.

**Definition 71** (Strange attractor). A *strange attractor* is a chaotic attractor with non integer Hausdorff dimension.



Figure 9.2: Example: the Sierpinski gasket is a chaotic attractor  $G$  with  $\dim_{\mathfrak{H}}(G) = \ln 3 / \ln 2$ .

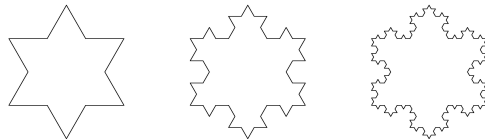


Figure 9.3: Example: the Koch snowflake is a chaotic attractor  $G$  with  $\dim_{\mathfrak{H}}(G) = \ln 4 / \ln 3$ .

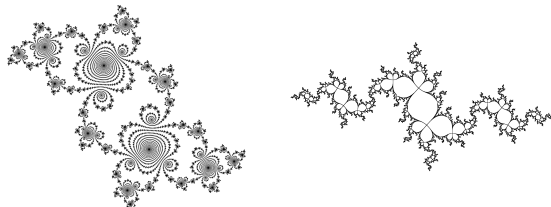


Figure 9.4: Example: the Julia sets, which are the basin of attraction for systems of the type  $z \mapsto z^2 + c \in \mathbb{C}$ , have a fractal dimension that satisfies  $1 + C^{-1}\sqrt{c+2} \leq \dim_{\mathfrak{H}}(J) \leq 1 + C|\log(c+2)|^{3/2}\sqrt{c+2}$  for some constant  $C > 1$  and suitable  $c$ . [18]

We have seen in Chapter 7 that there are multiple algorithms for the computation of fractal dimensions. It can be proved that for most practical applications, both the box-counting dimension and the correlation dimension are a good measure of the "fractalness" of a set. In the trivial cases as sets of integer dimension and for simple fractals, (as Hutchinson fractals like the Cantor set, the Sierpinski gasket, the Koch snowflake,...) these measures are equal to the Hausdorff dimension of the set.

### 9.1.2 Minimum embedding dimension through through False Nearest Strand

We have seen ways to compute the fractal dimension of a set. Then the minimum embedding dimension can be taken as the nearest integer greater than the fractal dimension. Another idea is to compute directly the minimum embedding dimension.

The False Nearest Strands (FNS) method is predisposed to execute this tasks in this manner: first, a tolerance  $\epsilon$  is decided. Then consider  $A$  to be the sample standard deviation of the original time series. This quantity is connected to the size of the attractor. For each  $d = 1, 2, \dots$  let  $S(d)$  be the mean Euclidean distance in the projected  $(d + 1)$  coordinate between strand pairs (segments between consecutive samples) found to be nearest neighbors in dimension  $d$ . If  $S(d)/A > \epsilon$ , where  $A$  is the estimated attractor size, then the strand is considered to be a false strand. The  $S(d)$  statistic is a measure of the average additional Euclidean distance we gain by embedding the strand in the next di-

mension, and is used to assess when this extra distance has grown too large, indicating a false strand.

## 9.2 Feature extraction for K-complex classification: persistent homology

In this section we present two examples of two *summary functions* of persistence diagrams. Summary functions are useful as, being from  $\mathbb{R}$  to  $\mathbb{R}$  and regular, they can be studied easily and are prone to feature extraction. Moreover, as shown in [9], they possess statistical properties (a notable one being the weak convergence of the sample average summary function) that make them suitable for data analysis.

After that, we will see the connection between Euler Characteristic and Homology groups. Euler Characteristic will prove to be an important feature for the classification of K-complexes.

### 9.2.1 Landscapes and silhouettes

**Definition 72** (Persistence landscape). Let  $D$  be a persistence diagram and  $p = (x, y) = \left(\frac{b+d}{2}, \frac{d-b}{2}\right)$  a point of  $D$ . Define then the following function:

$$\Lambda_p(t) = \begin{cases} t - x + y & t \in [x - y, x] \\ x + y - t & t \in (x, x + y] \\ 0 & \text{otherwise} \end{cases} = \begin{cases} t - b & t \in [b, (b + d)/2] \\ d - t & t \in ((b + d)/2, d] \\ 0 & \text{otherwise} \end{cases}$$

$p$  is itself on the graph of  $\Lambda_p(t)$ . Given the arrangement of curves obtained by overlaying the graphs of the functions  $\{\Lambda_p\}_{p \in D}$ , the *persistence landscape* of  $D$  is the collection of functions

$$\lambda_D(k, t) = k\max_{p \in D} \Lambda_p(t) \quad t \in [0, T], k \in \mathbb{N},$$

where  $k\max$  is the  $k$ -th largest value in the set.

From the definition of persistence landscape we immediately observe that  $\lambda_D(k, \cdot)$  is one-Lipschitz, since  $\Lambda_p$  is one-Lipschitz.

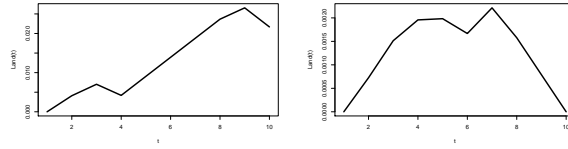


Figure 9.5: Example: The first persistence landscape of dimension 1,  $\lambda_D(1, t)$ , for a  $K$ -complex (left) and for a non- $K$ -complex (right).  $K$  complexes are characterized by less classes with low persistence and more classes with high persistence.

**Definition 73** (Power weighted silhouette). Consider a persistence diagram  $D$  with  $m$  off diagonal points. Let each point of  $D$  be represented by a pair  $(b_j, d_j)$ . For every  $0 < p \leq \infty$  we define the *power-weighted silhouette* as

$$\phi^p(t) = \frac{\sum_{j=1}^m |d_j - b_j|^p \Lambda_j(t)}{\sum_{j=1}^m |d_j - b_j|^p}$$

For each positive weight  $p$ , also the power weighted silhouette is a one-Lipschitz function.

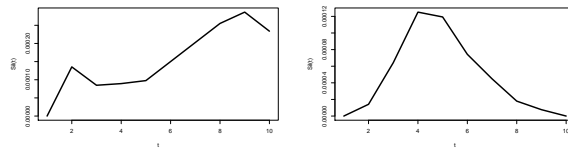


Figure 9.6: Example: The weighted silhouette of dimension 1 and power 1,  $\phi^1(t)$ , for a  $K$ -complex (left) and for a non- $K$ -complex (right).  $K$  complexes are characterized by less classes with low persistence and more classes with high persistence.

### 9.2.2 Euler characteristic

The Euler characteristic of a topological space is a number related its shape. Its importance comes from being a topological invariant: two homeomorphic topological spaces have the same Euler characteristic. Moreover, this quantity is also invariant under homotopies.

**Definition 74** (Euler characteristic). Let  $K$  be a simplicial complex, and let  $n_p$  be the number of  $p$ -simplices,  $n \geq 0$ . The *Euler characteristic*

of  $K$  is equal to the alternating sum

$$\chi = n_0 - n_1 + n_2 - n_3 + \dots$$

We can compute the Euler characteristic of a simplicial complex in two ways: with the definition, or exploiting homology groups. Although the latter seems an unintuitive procedure, simplicial complexes and filtrations can be practically inconvenient to store in memory.

The Euler-Poincaré theorem elegantly connects homology groups with the Euler characteristic.

Recall that from Chapter 4 we learnt that writing  $n_p = \text{rank } \mathcal{C}_p$  for the number of  $p$ -simplices in  $K$ ,  $z_p = \text{rank } \mathcal{Z}_p$  and  $b_p = \text{rank } \mathcal{B}_p$  for the ranks of the cycle and boundary groups, we have  $n_p = z_p + b_{p-1}$ .

**Theorem 19** (Euler-Poincaré). *The Euler characteristic of a topological space is the alternating sum of its Betti numbers:*

$$\chi = \sum_{p \geq 0} \beta_p.$$

*Proof.*

$$\chi = \sum_{p \geq 0} n_p = \sum_{p \geq 0} (-1)^p (z_p + b_{p-1}) = \sum_{p \geq 0} (-1)^p \beta_p.$$

□

**Remark 7** ("Truncated" Euler characteristic). Once a persistence diagram is obtained, we have access to the persistent Betti numbers and not to the normal Betti numbers. We are interested in the evolution of the Euler characteristic when passing from a complex in the filtration to the next. To do that, the diagram can be divided in vertical stripes, mimicking the difference in classes births. On these stripes we can compute the *persistent Euler characteristic* using the definition. Unfortunately, computing the persistent Betti numbers for dimension  $\geq 2$  is computationally demanding. For this reason, we make use of the *truncated persistent Euler characteristic*  $\tilde{\chi}_s = \beta_{0,s} - \beta_{1,s}$ , where  $s$  indicates the number of the stripe considered. This quantity equals the persistent Euler characteristic we would have found by always limiting the embedding dimension to 2.

### 9.3 Numerical Results

This section presents the numerical results for the two classification tasks. The adherence of the estimated labels and the true ones is summarized in a confusion matrix, containing the number of *true positives*  $TP$ , *true negatives*  $TN$ , *false positives*  $FP$ , and *false negatives*  $FN$ . From these quantities we estimate the overall performance of the classifier, by defining:

- *Accuracy*:  $(TP + TN)/(TP + TN + FP + FN)$ ;
- *Sensitivity*:  $(TP)/(TP + FN)$ ;
- *Precision*:  $(TP)/(TP + FP)$ ;
- *False positive rate*:  $(FP)/(TN + FP)$ .

The performance is evaluated by cross-validation over the whole dataset, consisting of 10 polysomnographies. Each polysomnography belongs to a different patient. Every EEG spans an entire night, starting from the onset of sleep until the patient is fully awake again the next morning.

Dataset Structure							
Dataset	#Epoques	#Awake	#NREM1	#NREM2	#NREM3	#REM	#KCs
<b>EEG1</b>	916	240	68	343	112	153	589
<b>EEG2</b>	918	136	121	296	158	207	447
<b>EEG3</b>	905	190	80	370	84	181	1059
<b>EEG4</b>	809	345	52	188	57	167	312
<b>EEG5</b>	810	101	62	318	188	141	536
<b>EEG6</b>	935	130	81	490	110	124	957
<b>EEG7</b>	876	181	111	328	90	166	509
<b>EEG9</b>	964	109	145	441	59	210	859
<b>EEG11</b>	950	97	80	557	95	121	1028
<b>EEG14</b>	879	183	96	387	108	105	782

Table 9.1: For each polysomnography we present the total number of 30 seconds epoques, partitioned into the 5 different Sleep Stages. In addition we display the number of K complexes (KCs) appearing during the night.

The features have been extracted employing the following R packages: TDA [22], tibble [41], Rfast [45], distances [52], moments [34], and nonlinearTseries [26]. The random forests have been created with the package randomForest [37]. Feature selection has been executed



choosing the most *important* variables. Importance is automatically computed by analyzing the precision gain for the splitting of each variable in the decision trees, using the Gini index.

**Remark 8** (Same size samples). In the binary classification case, random forest methods often provide a better performance when the set of instances labeled with 0 and 1 have the same or similar size.

1. For the detection of *sleep stages*, each of the five random forests is trained on a set of labels equal to 1 if we are dealing with an epoche of a particular stage (for example, the REM stage) and equal to 0 in any other case. Thus each stage-detecting random forest learns from the totality the stages in of a particular kind in the dataset and a randomly chosen collection of other stages of the same size.
2. For the detection of *K complexes*, the random forest is trained over all the KCs (corresponding to instances labeled with 1) and a randomly chosen collection of non-K complexes of the same size.

### 9.3.1 Detection of sleep stages

**Feature extraction.** For each epoche we extracted a series of features connected to the self-correlation, the complexity and the self-similarity of the signal. Self-correlation is measured through the average mutual information 58 and the autocorrelation function 57. The first values resulted to have greater importance. We also considered the optimized Sliding Window delays  $\tau$  relative to these two functions. Self similarity has been assessed by computing various types of dimension and fractal dimension (box-counting 55 and correlation 56 dimensions, minimum embedding dimension using FNS 9.1.2). Also the output of the *detrended fluctuation analysis* method [42] happened to be an important feature. DFA is another measure of self-affinity of a signal.

*Lyapunov exponents* [51] are one of the main quantifiers of the chaoticness of a signal, as they measure the scale of divergence of close trajectories. We also computed the *Kolmogorov-Sinai entropy* [44] of the signals to evaluate the carried information. As the importance of these two quantities resulted to be low, we discarded this set of features.

Finally we determined standard features as maximum, minimum, mean, variance, higher moments and frequency spectrum of the signals. Luckily also these quantities have been rejected, making the results more specifically connected to the signals' chaoticness.

Below we present the features that overcame the selection process.

---

**Detection of sleep stages:** Complexity

---

**Selected features:**

$F_1$ : Mutual information function  $I_{ff}$ ;

$F_2$ : Autocorrelation function  $R_{ff}$ ;

$F_3$ : Box-counting dimension of the 5D embedding  $\dim_B(SW_{5,\tau}(f))$ ;

$F_4$ : Correlation dimension  $\nu$ ;

$F_5$ : Detrended Fluctuation Analysis (DFA): long range correlation between values;

$F_6$ : Dimension with False Nearest Strands method (FNS);

$F_7$ : Delay with autocorrelation method  $\tau_{AC}$ ;

$F_8$ : Delay with mutual information method  $\tau_{AMI}$ ;

---

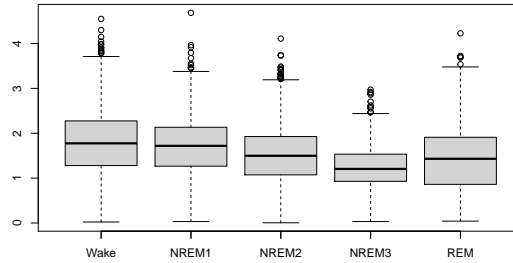


Figure 9.7:  $F_3$ : box-counting dimension for the different sleep stages.

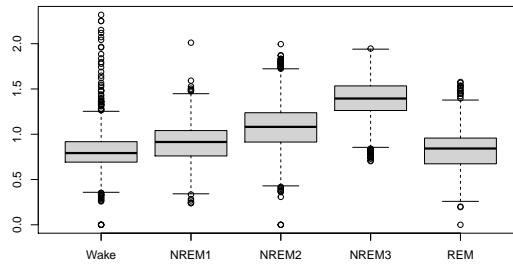


Figure 9.8:  $F_2$ : critical value of mutual information function for the different sleep stages.

<i>Detection of single Sleep Stages</i>				
Sleep Stage	Accuracy	Sensitivity	Precision	False positive rate
<b><i>AWAKE</i></b>	0.87	0.88	0.85	0.15
<b><i>NREM1</i></b>	0.65	0.73	0.63	0.43
<b><i>NREM2</i></b>	0.77	0.78	0.78	0.22
<b><i>NREM3</i></b>	0.89	0.89	0.89	0.11
<b><i>REM</i></b>	0.82	0.84	0.8	0.21

Table 9.2: Detection of sleep stages.

The results for this detection problem show that the NREM1 stage is the hardest to classify correctly. This can be explained as this stage is very similar by its nature to a awakesness stages and NREM2 stages. In the onset of sleep a person alternates rapidly between wakefulness and stage NREM1; for this reason the labeling of this period has an higher arbitrariness. Stage NREM3 and awakesness, on the other hand,

are classified with great precision.

<i>Classification of state of consciousness</i>				
State	Accuracy	Sensitivity	Precision	False positive rate
<i>Asleep</i>	0.81	0.84	0.80	0.21
<i>Deep Sleep</i>	0.87	0.89	0.86	0.14

*Table 9.3: Detection of state of sleep.*

The classification of the state of sleeps resents on the problem of stage NREM1 labeling. When awakeness and Stage NREM1 are considered together the task becomes the detection of deep sleep (Stages NREM2, NREM3, and REM). In this case the precision is significantly higher.

9.3.2 Detection of K-complexes

We resume the procedure to extract persstent homology features.

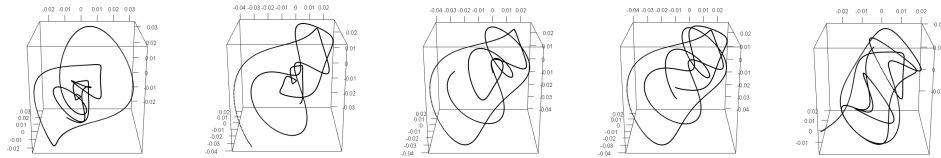
1. We select only NREM2 epoques (30 seconds signals), as KCs appear during this Stage.



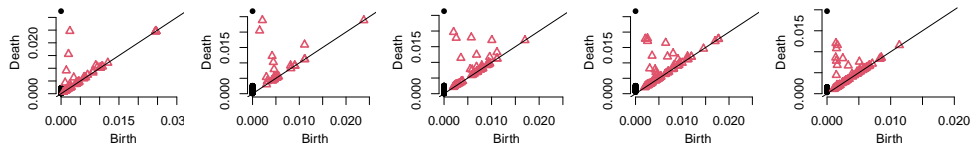
2. The 5 highest peaks are centered in 6 seconds signals, which are then filtered.



3. We apply the Sliding Window Embedding to the signals to obtain point clouds.



4. We compute the persistent diagram (and landscape, silhouette) from each point cloud.



5. Features are extracted and used as input to classify the signals. In this case the output is

1                      0                      0                      0                      0

**Feature extraction.** For the detection of  $K$  complexes, the objects employed in the production feature values are three: persistence diagrams 48, persistence landscapes and silhouettes 9.2.1. A first set of instances is obtained by summarizing the persistences, births and deaths in a sequence of vectors. The procedure is similar to the one showed in [2]. As shown in Chapter 7, diagrams corresponding to  $K$  complexes often display few (1-dimensional) classes with high persistence and few classes with medium persistence. On the other side, diagrams corresponding signals without a  $K$  complex present no (1-dimensional) classes with high persistence and a high number of classes with medium persistence. Small persistence classes are also very important:  $K$  complexes possess only few of them since there are less small scale oscillations in the signal.

The same reasoning can be repeated with 0-th dimensional classes as  $K$ -complexes result in more disconnected state spaces and other signals translate in more compact state spaces.

In addition we considered the norms of the diagrams 49 and the Euler Characteristic of the 2-dimensional embedding 7.

Finally, important features were extracted from specific values of the persistence landscape and silhouette.

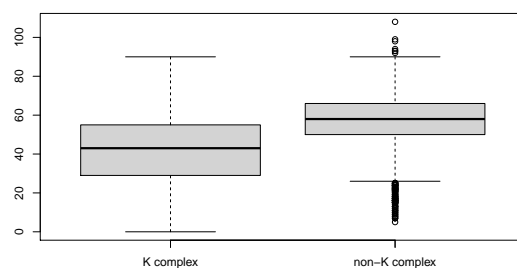


Figure 9.9:  $F_2$ : the number of 1-dimensional persistent homology classes is lower for  $K$ -complexes.

---

**Detection of K-complexes:** Persistent homology
 

---

**Selected features:**

$F_1$ : # of persistence classes of dimension 0;

$F_2$ : # of persistence classes of dimension 1;

$F_3$ :  $\text{mp}_0(D, 1)/\text{mp}_0(D, 2)$  (max 0-dimensional persistence / 2nd max 0-dimensional persistence);

$F_4$ :  $\text{mp}_0(D, 1)/\text{mp}_0(D, 3)$  (max 0-dimensional persistence / 3rd max 0-dimensional persistence);

$F_5$ :  $\text{mp}_1(D, 1)/\text{mp}_1(D, 2)$  (max 1-dimensional persistence / 2nd max 1-dimensional persistence);

$F_6$ :  $\text{mp}_1(D, 1)/\text{mp}_1(D, 3)$  (max 1-dimensional persistence / 3rd max 1-dimensional persistence);

$F_7$ : 0 and 1-dimensional silhouette  $\phi^1(t)$ ;

$F_8$ : 0 and 1-dimensional landscape  $\lambda_D(1, t)$ ;

$F_9$ : Early births and deaths for classes of dimension 1;

$F_{10}$ : Truncated Euler characteristic  $\tilde{\chi}_s$ .

$F_{11}$ : Bottleneck distance of dimension 1 from empty diagram  $d_\infty(D, D_0)$ .

$F_{12}$ : 1-Wasserstein distance of dimension 1 from empty diagram  $d_{W_1}(D, D_0)$ .

---

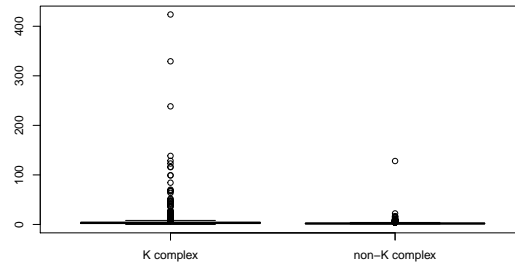


Figure 9.10:  $F_6$ : the quotient between the highest 1-dimensional persistent class and the third one is higher for K-complexes.

<i>Detection of K-complexes</i>				
Label	Accuracy	Sensitivity	Precision	False positive rate
<b><i>KC</i></b>	0.92	0.93	0.91	0.09

*Table 9.4: Detection of K complexes.*

The set of features from topological data analysis results in a very good overall performance of the classification method. The characteristic shape of the K-complexes is translated effectively in higher dimensions through the Sliding Window Embedding. In addition, the band-filtering and the parameter of the embedding ensure that 0 and 1 dimensional homology groups are sufficient to achieve a satisfactory level of precision.



## Chapter 10

# Conclusions

**Final considerations.** This dissertation presented an application of TDA and dynamical systems theory for the classification of bio-signals provided from EEGs. As in other works from the scientific literature, these fields of mathematics are revealed to be suitable for the recognition of shapes and patterns. The topology of the phase spaces obtained through the Sliding Window Embedding contains a great amount of information on the nature of biosignals. The study of the complexity of the orbits and of the persistent homology groups is less affected on the positioning of electrodes and possible noise in the measurements. The numerical results show good precision of the classification method for the detection of sleep stages, and very good precision for the detection of K complexes.

**Future developments.** There are multiple ideas that can be brought forward starting from the analysis of this dissertation, the first being elaborate a scheme to avoid the high computational costs of persistent homology. The application of TDA to sleep stages has been unfeasible: every attempt in summarizing the 30 seconds point clouds resulted in fatal loss of information. Concerning K-complexes and possibly other waveforms or patterns, an interesting idea could be that of employing higher dimensional persistent homology groups and study their interpretation.

The effectiveness of the K-complex classifier could be exploited to quantify the strength, or other detailed characteristic of the waveforms: this feature, together with density, are shown to depend on age and pos-

sibly be connected to neurodegenerative illnesses.

An interesting theoretical topic to elaborate on is the connection between fractals (Hutchinson fractals, for example) and their corresponding persistent homology patterns. It seems reasonable that fractals can be encoded via persistent diagram as their holes present self-similar and precise behaviour.

Finally, as numerous publications keep apart the fields of TDA and chaos when studying biosignals, it might be compelling to understand the causes of the fractal nature of brain waves with the continuous support of the mapper algorithm and persistent homology.

# Bibliography

- [1]
- [2] Henry Adams, Sofya Chepushtanova, Tegan Emerson, Eric Hanson, Michael Kirby, Francis Motta, Rachel Neville, Chris Peterson, Patrick Shipman, and Lori Ziegelmeier. Persistence images: A stable vector representation of persistent homology, 2016. [arXiv:1507.06217](https://arxiv.org/abs/1507.06217).
- [3] Daniel Adler, Duncan Murdoch, et al. *rgl: 3D Visualization Using OpenGL*, 2020. R package version 0.100.54. URL: <https://CRAN.R-project.org/package=rgl>.
- [4] Wessam AL-Salman, Yan Li, and Peng Wen. K-complexes detection in eeg signals using fractal and frequency features coupled with an ensemble classification model. *Neuroscience*, 422:119–133, 2019. URL: <https://www.sciencedirect.com/science/article/pii/S0306452219307262>, doi:<https://doi.org/10.1016/j.neuroscience.2019.10.034>.
- [5] I. Bankman, V. G. Sigillito, R. Wise, and P. L. Smith. Feature-based detection of the k-complex wave in the human electroencephalogram using neural networks. *IEEE Transactions on Biomedical Engineering*, 39:1305–1310, 1992.
- [6] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. URL: <http://dx.doi.org/10.1023/A:1010933404324>, doi:[10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- [7] Leo Breiman, Jerome Friedman, Richard Olshen, and Charles Stone. *Classification And Regression Trees*. 10 2017. doi:[10.1201/9781315139470](https://doi.org/10.1201/9781315139470).

- 
- [8] Gunnar Carlsson. Topology and data. *Bulletin of The American Mathematical Society - BULL AMER MATH SOC*, 46:255–308, 04 2009. doi:[10.1090/S0273-0979-09-01249-X](https://doi.org/10.1090/S0273-0979-09-01249-X).
- [9] Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, and Larry Wasserman. Stochastic convergence of persistence landscapes and silhouettes, 2013. arXiv:[1312.0308](https://arxiv.org/abs/1312.0308).
- [10] Frédéric Chazal and Bertrand Michel. An introduction to topological data analysis: fundamental and practical aspects for data scientists, 2021. arXiv:[1710.04019](https://arxiv.org/abs/1710.04019).
- [11] Frédéric Chazal, Brittany Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo, Alessandro Rinaldo, and Larry Wasserman. Robust topological inference: Distance to a measure and kernel distance. *J. Mach. Learn. Res.*, 18(1):5845–5884, January 2017.
- [12] Ian M. Colrain. The K-Complex: A 7-Decade History. *Sleep*, 28(2):255–273, 02 2005. arXiv:<https://academic.oup.com/sleep/article-pdf/28/2/255/13688327/sleep-28-2-255.pdf>, doi:[10.1093/sleep/28.2.255](https://doi.org/10.1093/sleep/28.2.255).
- [13] Luigi De Gennaro, Maurizio Gorgoni, Flaminia Reda, Giulia Lauri, Iliaria Truglia, Susanna Cordone, Serena Scarpelli, Anastasia Mangiaruga, Aurora D’Atri, Giordano Lacidogna, Michele Ferrara, Camillo Marra, and Paolo Rossini. The fall of sleep k-complex in alzheimer disease. *Scientific Reports*, 7:39688, 01 2017. doi:[10.1038/srep39688](https://doi.org/10.1038/srep39688).
- [14] William C. Dement, Meir Kryger, and Thomas Roth. Contributors. In Meir Kryger, Thomas Roth, and William C. Dement, editors, *Principles and Practice of Sleep Medicine (Sixth Edition)*, pages ix–xxxii. Elsevier, sixth edition edition, 2017. URL: <https://www.sciencedirect.com/science/article/pii/B9780323242882001768>, doi:<https://doi.org/10.1016/B978-0-323-24288-2.00176-8>.
- [15] Robert L. Devaney. *An introduction to chaotic dynamical systems / Robert L. Devaney*. Addison-Wesley Redwood City, Calif, 2nd ed. edition, 1989.

- [16] S Devuyst, Thierry Dutoit, Patricia Stenuit, and M Kerkhofs. Automatic k-complexes detection in sleep eeg recordings using likelihood thresholds. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2010:4658–61, 08 2010. doi:[10.1109/IEMBS.2010.5626447](https://doi.org/10.1109/IEMBS.2010.5626447).
- [17] T. Dey and S. Mandal. Protein classification with improved topological data analysis. In *WABI*, 2018.
- [18] Neil Dobbs, Jacek Graczyk, and Nicolae Mihalache. Hausdorff dimension of julia sets in the logistic family, 2020. arXiv:2007.07661.
- [19] L Doroshenkov, VA Konyshv, and Sergey Selishchev. Classification of human sleep stages based on eeg processing using hidden markov models. *Meditssinskaia tekhnika*, 41:24–8, 01 2007. doi:[10.1007/s10527-007-0006-5](https://doi.org/10.1007/s10527-007-0006-5).
- [20] Farideh Ebrahimi, Mohammad Mikaeili, and Homayoun Nazeran. Automatic sleep stage classification based on eeg signals by using neural networks and wavelet packet coefficients. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2008:1151–4, 02 2008. doi:[10.1109/IEMBS.2008.4649365](https://doi.org/10.1109/IEMBS.2008.4649365).
- [21] Herbert Edelsbrunner and John Harer. *Computational Topology: An Introduction*. 01 2010. doi:[10.1007/978-3-540-33259-6\\_7](https://doi.org/10.1007/978-3-540-33259-6_7).
- [22] Brittany T. Fasy, Jisu Kim, Fabrizio Lecci, Clement Maria, David L. Millman, Vincent Rouvreau. The included GUDHI is authored by Clement Maria, Dionysus by Dmitriy Morozov, PHAT by Ulrich Bauer, Michael Kerber, and Jan Reininghaus. *TDA: Statistical Tools for Topological Data Analysis*, 2019. R package version 1.6.9. URL: <https://CRAN.R-project.org/package=TDA>.
- [23] Michelle Feng and Mason A. Porter. Spatial applications of topological data analysis: Cities, snowflakes, random structures, and spiders spinning under the influence. *Physical Review Research*, 2(3), Sep 2020. URL: <http://dx.doi.org/10.1103/>

- [PhysRevResearch.2.033426](#), [doi:10.1103/physrevresearch.2.033426](#).
- [24] Andrew M. Fraser and Harry L. Swinney. Independent coordinates for strange attractors from mutual information. *Phys. Rev. A*, 33:1134–1140, Feb 1986. URL: <https://link.aps.org/doi/10.1103/PhysRevA.33.1134>, [doi:10.1103/PhysRevA.33.1134](#).
- [25] Hitesh Gakhar and Jose A. Perea. Sliding window persistence of quasiperiodic functions, 2021. [arXiv:2103.04540](#).
- [26] Constantino A. Garcia. *nonlinearTseries: Nonlinear Time Series Analysis*, 2020. R package version 0.2.10. URL: <https://CRAN.R-project.org/package=nonlinearTseries>.
- [27] S. Green. *Biological Rhythms, Sleep and Hypnosis*. Palgrave Insights in Psychology series. Palgrave Macmillan, 2011. URL: <https://books.google.it/books?id=JfpdAQAQBAJ>.
- [28] Allen Hatcher. *Algebraic topology*. Cambridge Univ. Press, Cambridge, 2000. URL: <https://cds.cern.ch/record/478079>.
- [29] Elena Hernández-Pereira, Veronica Bolón-Canedo, Noelia Sánchez-Marroño, Diego Álvarez Estévez, Vicente Moret-Bonillo, and Amparo Alonso-Betanzos. A comparison of performance of k-complex classification methods using feature selection. *Information Sciences*, 328:1–14, 2016. URL: <https://www.sciencedirect.com/science/article/pii/S0020025515006088>, [doi:https://doi.org/10.1016/j.ins.2015.08.022](#).
- [30] Yu-Liang Hsu, Ya-Ting Yang, Jeen-Shing Wang, and Chung-Yao Hsu. Automatic sleep stage recurrent neural classifier using energy features of eeg signals. *Neurocomputing*, 104:105–114, 2013. URL: <https://www.sciencedirect.com/science/article/pii/S0925231212008387>, [doi:https://doi.org/10.1016/j.neucom.2012.11.003](#).
- [31] Conrad Iber, Sonia Ancoli-Israel, A.L. Chesson, and Stuart Quan. The aasm manual for the scoring of sleep and associated events: Rules, terminology and technical specifications. *Westchester, IL: American Academy of Sleep Medicine*, 01 2007.

- [32] A Kam, A Cohen, A Geva, and Ariel Tarasiuk. Detection of k-complexes in sleep eeg using cd-hmm. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 1:33–6, 02 2004. doi: [10.1109/IEMBS.2004.1403083](https://doi.org/10.1109/IEMBS.2004.1403083).
- [33] H.S. Kim, R. Eykholt, and J.D. Salas. Nonlinear dynamics, delay times, and embedding windows. *Physica D: Nonlinear Phenomena*, 127(1):48–60, 1999. URL: <https://www.sciencedirect.com/science/article/pii/S0167278998002401>, doi:[https://doi.org/10.1016/S0167-2789\(98\)00240-1](https://doi.org/10.1016/S0167-2789(98)00240-1).
- [34] Lukasz Komsta and Frederick Novomestky. *moments: Moments, cumulants, skewness, kurtosis and related tests*, 2015. R package version 0.14. URL: <https://CRAN.R-project.org/package=moments>.
- [35] L. K. Krohne, R. B. Hansen, J. A. E. Christensen, H. B. D. Sorensen, and P. Jennum. Detection of k-complexes based on the wavelet transform. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5450–5453, 2014. doi:[10.1109/EMBC.2014.6944859](https://doi.org/10.1109/EMBC.2014.6944859).
- [36] Tarek Lajnef, Sahbi Chaibi, Jb Eichenlaub, Perrine Ruby, Pierre-Emmanuel Aguera, Mounir Samet, Abdennaceur Kachouri, and Karim Jerbi. Sleep spindle and k-complex detection using tunable q-factor wavelet transform and morphological component analysis. *Frontiers in Human Neuroscience*, 9, 08 2015. doi: [10.3389/fnhum.2015.00414](https://doi.org/10.3389/fnhum.2015.00414).
- [37] Andy Liaw and Matthew Wiener. Classification and regression by randomforest. *R News*, 2(3):18–22, 2002. URL: <https://CRAN.R-project.org/doc/Rnews/>.
- [38] Shunjie Liu, Junhao Pan, Qingfeng Lei, Lu He, Bingting Zhong, Yangyang Meng, and Zhong Li. Spontaneous k-complexes may be biomarkers of the progression of amnesic mild cognitive impairment. *Sleep medicine*, 67:99–109, March 2020. doi:[10.1016/j.sleep.2019.10.015](https://doi.org/10.1016/j.sleep.2019.10.015).

- [39] Yuriy Mileyko, Sayan Mukherjee, and John Harer. Probability measures on the space of persistence diagrams. *Inverse Problems*, 27(12):124007, nov 2011. doi:10.1088/0266-5611/27/12/124007.
- [40] Thomas M. Mitchell. *Machine Learning*. McGraw-Hill, Inc., USA, 1 edition, 1997.
- [41] Kirill Müller and Hadley Wickham. *tibble: Simple Data Frames*, 2020. R package version 3.0.4. URL: <https://CRAN.R-project.org/package=tibble>.
- [42] L.F. Márton, S.T. Brassai, L. Bakó, and L. Losonczi. Detrended fluctuation analysis of eeg signals. *Procedia Technology*, 12:125–132, 2014. The 7th International Conference Interdisciplinarity in Engineering, INTER-ENG 2013, 10–11 October 2013, Petru Maior University of Tirgu Mures, Romania. URL: <https://www.sciencedirect.com/science/article/pii/S2212017313006506>, doi:<https://doi.org/10.1016/j.protcy.2013.12.465>.
- [43] James R. Munkres. *Elements of Algebraic Topology*. Addison Wesley Publishing Company, 1984. URL: <http://www.worldcat.org/isbn/0201045869>.
- [44] Edward Ott. *Chaos in Dynamical Systems*. Cambridge University Press, 2 edition, 2002. doi:10.1017/CB09780511803260.
- [45] Manos Papadakis, Michail Tsagris, Marios Dimitriadis, Stefanos Fafalios, Ioannis Tsamardinos, Matteo Fasiolo, Giorgos Bouboudakis, John Burkardt, Changliang Zou, Kleanthi Lakiotaki, and Christina Chatzipantsiou. *Rfast: A Collection of Efficient and Extremely Fast R Functions*, 2020. R package version 2.0.1. URL: <https://CRAN.R-project.org/package=Rfast>.
- [46] C. R. Patti, H. Abdullah, Y. Shoji, A. Hayley, C. Schilling, M. Schredl, and D. Cvetkovic. K-complex detection based on pattern matched wavelets. In *2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pages 470–474, 2016. doi:10.1109/IECBES.2016.7843495.
- [47] J. A. Perea. Persistent homology of toroidal sliding window embeddings. In *2016 IEEE International Conference on Acoustics*,



- Speech and Signal Processing (ICASSP)*, pages 6435–6439, 2016. doi:[10.1109/ICASSP.2016.7472916](https://doi.org/10.1109/ICASSP.2016.7472916).
- [48] Jose Perea and John Harer. Sliding windows and persistence: An application of topological methods to signal analysis, 2013. arXiv:[1307.6188](https://arxiv.org/abs/1307.6188).
- [49] Chi Seng Pun, Kelin Xia, and Si Xian Lee. Persistent-homology-based machine learning and its applications – a survey, 2018. arXiv:[1811.00252](https://arxiv.org/abs/1811.00252).
- [50] Abbas Rizvi, Pablo Camara, Elena Kandror, Thomas Roberts, Ira Schieren, Tom Maniatis, and Raul Rabadan. Single-cell topological rna-seq analysis reveals insights into cellular differentiation and development. *Nature Biotechnology*, 35, 05 2017. doi:[10.1038/nbt.3854](https://doi.org/10.1038/nbt.3854).
- [51] Marco Sandri. Numerical calculation of lyapunov exponents. *Math. J.*, 6, 01 1996.
- [52] Fredrik Savje. *distances: Tools for Distance Metrics*, 2019. R package version 0.1.8. URL: <https://CRAN.R-project.org/package=distances>.
- [53] Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning - From Theory to Algorithms*. Cambridge University Press, 2014.
- [54] C. E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949. doi:[10.1109/JRPROC.1949.232969](https://doi.org/10.1109/JRPROC.1949.232969).
- [55] Gurjeet Singh, Facundo Mémoli, and Gunnar Carlsson. Topological methods for the analysis of high dimensional data sets and 3d object recognition. pages 91–100, 01 2007. doi:[10.2312/SPBG/SPBG07/091-100](https://doi.org/10.2312/SPBG/SPBG07/091-100).
- [56] Rakesh Sinha. Artificial neural network and wavelet based automated detection of sleep spindles, rem sleep and wake states. *Journal of medical systems*, 32:291–9, 09 2008. doi:[10.1007/s10916-008-9134-z](https://doi.org/10.1007/s10916-008-9134-z).
- [57] P. Skraba, V. D. Silva, and M. Vejdemo-Johansson. Topological analysis of recurrent systems. In *NIPS 2012*, 2012.

- 
- [58] Cash SS, Halgren E, and Dehgani N. The human K-complex represents an isolated cortical down-state. *Science*, 324(5930):1084–1087, 2009. doi:[10.1126/science.1169626](https://doi.org/10.1126/science.1169626).
- [59] A. Subasi, M. K. Kiyimik, M. Akin, and O. Eroglu. Automatic recognition of vigilance state by using a wavelet-based artificial neural network. *Neural Computing and Applications*, 14:45–55, 2004.
- [60] J. Sueur, T. Aubin, and C. Simonis. Seewave: a free modular tool for sound analysis and synthesis. *Bioacoustics*, 18:213–226, 2008. URL: <http://www.tandfonline.com/doi/abs/10.1080/09524622.2008.9753600>.
- [61] Floris Takens. *Detecting Strange Attractors in Turbulence*. *Lecture Notes in Mathematics*, volume 898, pages 366–381. 11 2006. doi:[10.1007/BFb0091924](https://doi.org/10.1007/BFb0091924).
- [62] Christopher J. Tralie and Jose A. Perea. (quasi)periodicity quantification in video data, using topology, 2018. [arXiv:1704.08382](https://arxiv.org/abs/1704.08382).
- [63] Sarah Tymochko, Kritika Singhal, and Giseon Heo. Classifying sleep states using persistent homology and markov chain: a pilot study, 2020. [arXiv:2002.07810](https://arxiv.org/abs/2002.07810).
- [64] Jan Vis. *edfReader: Reading EDF(+) and BDF(+) Files*, 2019. R package version 1.2.1. URL: <https://CRAN.R-project.org/package=edfReader>.
- [65] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. In *Proceedings of the Twentieth Annual Symposium on Computational Geometry*, SCG '04, page 347–356, New York, NY, USA, 2004. Association for Computing Machinery. doi:[10.1145/997817.997870](https://doi.org/10.1145/997817.997870).