ANÁLISIS TOPOLÓGICO DE PRIMITIVAS CINEMÁTICAS PARA LA CARACTERIZACIÓN DEL PARKINSON

YESSICA CAROLINA CASTRILLÓN GAMBOA

UNIVERSIDAD INDUSTRIAL DE SANTANDER FACULTAD DE CIENCIAS ESCUELA DE MATEMÁTICAS BUCARAMANGA 2021

ANÁLISIS TOPOLÓGICO DE PRIMITIVAS CINEMÁTICAS PARA LA CARACTERIZACIÓN DEL PARKINSON

YESSICA CAROLINA CASTRILLON GAMBOA

Una tesis presentada en cumplimiento de los requisitos para el grado de: Pregrado en Matemáticas

> Director: Fabio Martínez Carrillo Doctor en Ingeniería de Sistemas y Computación

> > Codirector:

Francisco A. Gómez Doctor en Ciencias de la Computación

UNIVERSIDAD INDUSTRIAL DE SANTANDER FACULTAD DE CIENCIAS ESCUELA DE MATEMÁTICAS BUCARAMANGA 2021

AGRADECIMIENTOS

El autor expresa su agradecimiento:

Principalmente a Dios, a mis padres que han sido un apoyo incondicional personal y profesionalmente, por todas las virtudes y valores que me han inculcado, por apoyarme e impulsarme a seguir y cumplir mis sueños. Al grupo de investigación *Bivl2ab* de la escuela de ingeniería de sistemas por permitirme ser parte y ampliar conocimientos en diversas áreas. Al profesor Fabio Martínez Carrillo en la dirección del presente proyecto de investigación y Francisco Gómez en la formación en el área del análisis topológico de datos. Muchísimas gracias su paciencia, dedicación y acompañamiento. Por último, gracias a mi familia y a aquellos que siempre han estado para mí, acompañándome y apoyándome en este proceso.

CONTENIDO

p	ág.
INTRODUCCIÓN	12
1. PROBLEMA DE INVESTIGACIÓN	15
2. OBJETIVOS	17
2.1. Objetivo general	17
2.2. Objetivos específicos	17
3. ANÁLISIS TOPOLÓGICO DE DATOS	18
3.1. NOCIONES DE TOPOLOGÍA	18
3.2. COMPLEJOS SIMPLICIALES	25
3.2.1. Nervios	28
3.2.2. Ejemplos de complejos	29
3.2.3. complejos de \check{C} ech	30
3.2.4. Complejos de Vietoris-Rips:	30
3.3. HOMOLOGÍA SIMPLICIAL	31
3.4. PERSISTENCIA	35
3.5. HOMOLOGÍA PERSISTENTE	36
4. EL ANÁLISIS TOPOLÓGICO, LA MARCHA Y LA ENFERMEDAD DE	
PARKINSON	42
4.1. DESCRIPTORES DE SERIES TEMPORALES DESDE LA TOPOLOGÍA .	44
4.2. DESCRIPTORES DE SERIES TEMPORALES FUNDAMENTADOS EN	
LA ENTROPÍA	48
4.3. CONFIGURACIÓN EXPERIMENTAL	51

5. RESULTADOS Y DISCUSIÓN	56
5.1. CENTRO DE MASA	56
5.2. EVALUACIÓN DE PUNTOS ARTICULARES DE INTERÉS	58
5.2.1. Trayectoria de representación para el grupo de la cabeza	59
5.2.2. Trayectoria de representación para el grupo de la brazo	59
5.2.3. Trayectoria de representación para el grupo de la cadera	60
5.2.4. Trayectoria de representación para el grupo de la pie	61
5.2.5. Resultados de la estimación por grupos del esqueleto	61
6. CONCLUSIONES Y PERSPECTIVAS	64
BIBLIOGRAFÍA	67
ANEXOS	71

LISTA DE FIGURAS

Figura 1. Ejemplo de espacios topológicos simples con sus números de Betti .	19
Figura 2. Ejemplo de una vecindad ϵ -contrapeso	21
Figura 3. Los componentes del cálculo de la distancia Hausdorff entre la línea	
roja K' y la línea azul K	22
Figura 4. Representación de $0, 1, 2$ y 3 simplejos $\ldots \ldots \ldots \ldots \ldots \ldots \ldots$	26
Figura 5. a) un no complejo simplicial, b) complejo simplicial.	26
Figura 6. Representación del Ejemplo 3.2.4	28
Figura 7. Representación del nervio (derecha) de la unión de tres bolas abier-	
tas (izquierda).	29
Figura 8. El complejo \check{C} ech $\check{C}_{\alpha}(X)$ (izquierda) y el Vietoris-Rips $V_{2\alpha}(X)$ (dere-	
cha) ¹	31
Figura 9. Relación entre los grupos de q-ciclos y q-bordes.	33
Figura 10.Una representación de nube de puntos a), su respectivo complejo	
simplicial b) y el invariante correspondiente c)	34
Figura 11 Emparejamiento de máximos y mínimos de una función $f:[0,1] \rightarrow \mathbb{R}$	
y su diagrama de persistencia ²	36
Figura 12Nacimiento de la clase γ en K_i	39
Figura 13.Un ejemplo de filtración simple de una nube de puntos y el código	
de barras de homología persistente resultantes	40

¹ CHAZAL, Frédéric y MICHEL, Bertrand. *An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists.* 2021. arXiv: 1710.04019.

² CHAZAL, Frédéric y MICHEL, Bertrand. *An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists.* 2021. arXiv: 1710.04019.

Figura 14.Trayectoria del centro de masa y puntos clave (a) y diferencia de dis-	
tribución entre los grupos control y párkinson, para el índice de constancia,	
se observa que el conjunto parkinsoniano es menos periódico (b).	43
Figura 15A izquierda ventana deslizante y disposición de los fragmentos de	
acuerdo con la similitud a derecha	46
Figura 16Dataset del grupo investigación Bilv2Lab ³	51
Figura 17Puntos claves del esqueleto humano ⁴	53
Figura 18Distribución entre las trayectorias del centro de masa de los grupos	
control y párkinson para los índices de Colwell	57
Figura 19Diferencia de distribución entre las trayectorias del centro de masa	
de los grupos control y párkinson para los índices de constancia y contin-	
gencias basados en análisis topológico de datos	58
Figura 20 Ciclo de la marcha humana ⁵ .	72
Figura 21 Ilustración de secuencia de video.	74
Figura 22Representación del centro de masa de marcha.	75
Figura 23A izquierda representación de la marcha y a derecha formato de	
salida de pose	76

³ SALAZAR I., Pertuz S. Contreras W. & Martínez F. A convolutional oculomotor representation to model parkinsonian fixational patterns from magnified videos. Pattern Analysis and Application. 2020.

⁴ G. HIDALGO Z. Cao, T. Simon S.-E. Wei H. Joo y SHEIKH, Y. *OpenPose library*.

⁵ TUNCA, Can, *et al.* "Inertial Sensor-Based Robust Gait Analysis in Non-Hospital Settings for Neurological Disorders". En: *Sensors* 17.4 (2017). DOI: 10.3390/s17040825.

LISTA DE TABLAS

pág.

Tabla 1. Propuesta de índices topológicos	48
Tabla 2. Índices propuestos por Colwell para medir la predictibilidad de fenó-	
menos periódicos ⁶	50
Tabla 3. Índices de las series de tiempo correspondiente a un punto del grupo	
la cabeza, de una persona control y con EP	59
Tabla 4. Índices de las series de tiempo correspondiente a un punto del grupo	
del brazo, de una persona control y con EP.	60
Tabla 5. Índices para la trayectoria de movimiento del punto clave represen-	
tante del grupo de la cabeza.	60
Tabla 6. índices para la trayectoria de movimiento del punto clave represen-	
tante del grupo del pie.	61
Tabla 7. Media y nivel de significación de los índices de cada grupo de puntos	
clave, control y párkinson. Se puede observar que las trayectorias del pie y	
la cabeza son siempre significativas.	62

⁶ COLWELL, Robert. "Predictability, Constancy, and Contingency of Periodic Phenomena". En: *Ecology* 55 (ago. de 1974), págs. 1148-1153. DOI: 10.2307/1940366.

LISTA DE ANEXOS

pág.

Anexo A. Análisis de marcha y el párkinson	71
Anexo B. Análisis de video	73
Anexo C. Trayectorias de movimiento	75

RESUMEN

TÍTULO: ANÁLISIS TOPOLÓGICO DE PRIMITIVAS CINEMÁTICAS PARA LA CARACTERIZACIÓN DEL PARKINSON *

AUTOR: YESSICA CAROLINA CASTRILLON GAMBOA **

PALABRAS CLAVE: ANÁLISIS TOPOLÓGICO DE DATOS, HOMOLOGÍA PERSISTENTE, PAR-KINSON, PRIMITIVAS CINEMÁTICAS, PREDICTIBILIDAD.

DESCRIPCIÓN: Esta investigación se centra en el Análisis Topológico de Datos (TDA), una herramienta matemática que permite estudiar la estructura topológica de una nube de puntos (datos) con el objetivo de encontrar "agujeros" k-dimensionales, a través de la homología persistente. Por medio de esta herramienta se propuso una metodología alternativa para la caracterización de la enfermedad de Parkinson (EP), modelando puntos k-dimensionales que representan la periodicidad durante la marcha de cada paciente. Inicialmente, se registran videos de personas control y pacientes con EP y se extraen trayectorias de movimientos. Cada trayectoria se caracteriza por su periodicidad usando *SW1PerS (por sus siglas en ingles: Sliding Windows and 1-Persistence Scoringes)*. Entonces, un análisis de homología persistente es realizado sobre los puntos de cada trayectoria, calculando índices topológicos que permiten medir características de predictibilidad. Esta metodología fue evaluada en un conjunto de 22 pacientes (11 pacientes con la EP y 11 pacientes control). El índice topológico con mayor capacidad de discriminación entre las dos poblaciones de pacientes es el basado en la periodicidad con el método SW1PerS. En este sentido, se evidencia que los pacientes de párkinson y las personas control tienen diferencias estadísticamente significativas con respecto a los patrones de periodicidad definidos.

^{*} Trabajo de investigación

^{**} Facultad de Ciencias. Escuela de Matemáticas. Director: Fabio Martínez Carrillo, Ph.D. Codirector Francisco A. Gomez, Ph.D.

ABSTRACT

TITLE: TOPOLOGICAL ANALYSIS OF KINEMATIC PRIMITIVES FOR PARKINSON'S MODELING.

AUTHOR: YESSICA CAROLINA CASTRILLÓN GAMBOA **

KEYWORDS: TOPOLOGICAL DATA ANALYSIS, PERSISTENT HOMOLOGY, PARKINSON, PRE-DICTABILITY.

DESCRIPTION: This research focuses on Topological Data Analysis (TDA), a mathematical tool that allows studying the topological structure of a point cloud (data) with the aim of finding k-dimensional "holes" through persistent homology. By means of this tool, an alternative methodology for the characterization of Parkinson's disease (PD) was proposed, modeling k-dimensional points representing the periodicity during the gait of each patient. Initially, videos of control individuals and PD patients are recorded and movement trajectories are extracted. Each trajectory is characterized for periodicity using *SW1PerS* (*Sliding Windows and 1-Persistence Scoringes*). The resulting points from each trajectory are mapped to a common space that allows persistent homology analysis. From this analysis, topological indices are calculated that allow measurement of topological features related to predictability, such as the periodicity of the time series. This methodology was evaluated in a set of 22 patients (11 PD patients and 11 control patients). The topological index with the highest discrimination capacity between the two patient populations is the one based on periodicity with the SW1PerS method. In this sense, it is evident that Parkinson's patients and controls have statistically significant differences with respect to the defined periodicity patterns.

^{*} Research work

^{**} Faculty of science. School of mathematics. Advisor: Fabio Martínez Carrillo, Ph.D. Co-advisor: Francisco A. Gomez, Ph.D.

INTRODUCCIÓN

La enfermedad de párkinson (EP) es la segunda enfermedad neurodegenerativa más frecuente en el mundo que afecta a 1 de cada 100 personas mayores de 60 años según la Organización Mundial de la Salud (OMS)¹. Esta condición corresponde a un trastorno neurodegenerativo progresivo crónico que afecta principalmente la movilidad y el control de la marcha ¹. Actualmente, no hay marcadores bioquímicos o radiológicos específicos para su diagnóstico, siendo los criterios clínicos como la evaluación visual de la marcha por parte de expertos las alternativas más ampliamente utilizadas para esta tarea.

En los últimos años se han explorado estrategias cuantitativas para la evaluación de la marcha de pacientes con enfermedad de Parkinson²³. En particular, el análisis de video se ha propuesto como una alternativa cuantificable a la evaluación clínica por parte de expertos⁴. Estrategias basadas en reconocimiento de patrones y aprendizaje de máquina sobre características extraídas del video han sido recientemente exploradas el diagnóstico de pacientes con párkinson. No obstante, los rasgos cuantitativos del video que correlacionan con marchas alteradas por la patología todavía

¹ ORGANIZATION, World Health. *Neurological disorders : public health challenges*. World Health Organization, 2006, págs. 140-142.

² MORRIS, Meg E., *et al.* "The biomechanics and motor control of gait in Parkinson disease". En: *Clinical Biomechanics* 16.6 (2001), págs. 459 -470.

³ BLIN, O.; FERRANDEZ, A.M. y SERRATRICE, G. "Quantitative analysis of gait in Parkinson patients: increased variability of stride length". En: *Journal of the Neurological Sciences* 98.1 (1990), págs. 91 -97.

⁴ BAJAJ, Nin P S, *et al.* "Accuracy of clinical diagnosis in tremulous parkinsonian patients: a blinded video study". En: *Journal of Neurology, Neurosurgery & Psychiatry* 81.11 (2010), págs. 1223-1228.

son poco entendidos⁵, principalmente por la dificultad en la caracterización de las dinámicas temporales asociadas a la marcha en pacientes con párkinson⁶.

Recientemente, un conjunto de métodos basados en topología computacional han sido propuestos para la caracterización de datos complejos. Estos métodos, permiten describir los invariantes topológicos posiblemente subyacen a los datos, en un área denominada análisis topológico de datos (Topological Data Analysis - TDA por sus siglas en inglés). El TDA describe los datos como complejos simpliciales, los cuales posteriormente son analizados a través de homologías persistentes. TDA ha sido recientemente explorado en neurociencia⁷, caracterización del cáncer⁸, y genética⁹, entre otros problemas biológicos. En cuanto a análisis topológicos de la enfermedad de párkinson, Nawar et al. utilizó persistencia homológica para caracterizar los cambios posturales a partir de la estimación del centro de presión (CoP) del paciente¹⁰. Perfiles de fuerza de marcha real de grupos sanos y de pacientes con enfermedad de párkinson han sido explorados utilizado medidas de complejidad

⁸ DEWOSKIN, D., *et al.* "Applications of computational homology to the analysis of treatment response in breast cancer patients". En: *Topology and its Applications* 157.1 (2010), págs. 157 -164.

⁵ CHEN, Shih-Wei, *et al.* "Quantification and recognition of parkinsonian gait from monocular video imaging using kernel-based principal component analysis". En: *BioMedical Engineering OnLine* 10.1 (2011), pág. 99. DOI: 10.1186/1475-925x-10-99.

⁶ RAMÁREZ, Juan; MORENO, Franklin y MEDINA, Rubén. "Análisis de la serie temporal del balanceo en la marcha de pacientes con parkinson". es. En: *Universidad, Ciencia y Tecnología* 10 (sep. de 2006), págs. 205 -210.

⁷ A roadmap for the computation of persistent homology.

⁹ BOWMAN, Gregory R, *et al.* "Structural insight into RNA hairpin folding intermediates". En: ().

¹⁰ NAWAR, Afra, *et al. Topological Descriptors for Parkinson's Disease Classification and Regression Analysis.* 2020. arXiv: 2004.07384 [cs.LG].

heurística y entropía¹¹. Sin embargo, estos trabajos se basan en medidas cinéticas que requieren instrumentos sofisticados de captura, que posiblemente alteran el gesto natural de la marcha.

En este trabajo se propone una nueva caracterización de la predictibilidad de la serie de tiempo (trayectoria de movimiento) asociada a la marcha de personas control y con párkinson. En particular, se propone utilizar TDA para la caracterización de invariantes topológicos relacionados con periodicidad en una nube de puntos que representa la serie de tiempo asociada al video de la marcha. Adicionalmente, se caracteriza la entropía como un posible marcador asociado a marchas alteradas en pacientes de párkinson. Para esto, se caracterizó un conjunto con 22 videos de marcha, 11 personas control y 11 personas con enfermedad de párkinson. Los resultados sugieren que el índice topológico basado en la periodicidad puede caracterizar las trayectorias de movimiento anormal, logrando establecer diferencias entre la población de párkinson con respecto a la de control.

¹¹ AFSAR O, Tirnakli U y N, Marwan. "Recurrence Quantification Analysis at work: Quasi-periodicity based interpretation of gait force profiles for patients with Parkinson disease". En: *Scientific Reports* 8 (jun. de 2018), págs. 300 -306.

1. PROBLEMA DE INVESTIGACIÓN

La enfermedad de Parkinson es una enfermedad neurodegenerativa caracterizada por movimientos involuntarios de temblor y disminución de la fuerza entre otros. Esta enfermedad es de evolución prolongada y se presenta en edades medias o avanzadas. Uno de los parámetros de estudio del párkinson es la dinámica de la marcha dado que en los pacientes se observa con regularidad una disminución de la marcha, la longitud de zancada y el balanceo de las manos^{12 13}. Es importante anotar que un aumento en la variabilidad de la marcha puede indicar una disminución en el control motor, aumentando significativamente el riesgo de caída¹⁴. Se han ideado diferentes métodos de obtención de datos relacionados con la marcha de los pacientes con EP, y estos datos se procesan para obtener diversas variables relacionadas con la marcha. El análisis de series de tiempo en video puede caracterizar enfermedades neurodegenerativas¹⁵, sin embargo estos estudios todavía no proveen una descripción cuantitativa completa de los patrones asociados a la enfermedad.

En la actualidad, debido a que no existe ningún examen específico para diagnosticar la enfermedad, la gran mayoría los pacientes con EP se diagnostican mediante

¹² MORRIS, Meg E., *et al.* "Stride length regulation in Parkinson's disease: Normalization strategies and underlying mechanisms". En: *Brain* 119.2 (abr. de 1996), págs. 551-568.

¹³ MORRIS, Meg E., *et al.* "The pathogenesis of gait hypokinesia in Parkinson's disease". En: *Brain* 117.5 (oct. de 1994), págs. 1169-1181.

¹⁴ SCHAAFSMA, Joanna, *et al.* "Gait dynamics in Parkinson's disease: Relationship to Parkinsonian features, falls and response to levodopa". En: *Journal of the neurological sciences* 212 (ago. de 2003), págs. 47-53.

¹⁵ HAUSDORFF, Jeffrey M., *et al.* "Gait variability and basal ganglia disorders: Stride-to-stride variations of gait cycle timing in parkinson's disease and Huntington's disease". En: *Movement Disorders* 13.3 (1998), págs. 428-437.

evaluaciones en persona y análisis cualitativo de los datos del paciente. Esta alternativa diagnostica puede resultar en una gran cantidad de errores diagnósticos, por ejemplo, se estima que solamente el 75 % de los diagnósticos clínicos de EP se confirman por estudios anatomopatológicos en la autopsia de pacientes que recibieron este diagnóstico¹⁶. El presente estudio propone una metodología que incorpora TDA en el análisis de series temporales asociadas a la marcha de pacientes control y con la enfermedad de Parkinson haciendo uso de métodos topológicos para la caracterización de la periodicidad y la entropía de dichas señales. Para lograr esta caracterización se proponen los siguientes objetivos:

¹⁶ TAPIA-NÚÑEZ, John y CHANÁ, Pedro. "Diagnosis of Parkinson's disease". En: *Revista de neurología* 38 (ene. de 2004), págs. 61-7.

2. OBJETIVOS

2.1. Objetivo general

Estudiar el análisis topológico de puntos multidimensionales que representan cinemáticas de la marcha para la caracterización del párkison.

2.2. Objetivos específicos

- Construir un modelo de cuantificación de periodicidad de las trayectorias de movimiento asociadas a la marcha de personas con enfermedad de Parkinson.
- Construir un modelo topológico n-dimensional que caracterice la periodicidad de las trayectorias.
- Implementar computacionalmente el modelo de caracterización propuesto calculando la periodicidad de trayectorias de movimiento como series de tiempo.
- Analizar las características topológicas obtenidas para así evaluar la caracterización de la marcha de personas control y personas con enfermedad de Parkinson.
- Evaluar la capacidad de caracterización del párkison.

3. ANÁLISIS TOPOLÓGICO DE DATOS

A continuación, se introducen los conceptos básicos en TDA requeridos para el desarrollo de esta tesis. Una descripción más detallada de estas herramientas puede encontrarse en las Notas en Persistencia, Probabilidad e Inferencia Estadística para Análisis Topológico de Datos del Centro de Investigación en Matemáticas, A.C (CIMAT)¹⁷.

3.1. Nociones de topología

La topología estudia las propiedades que se preservan de los objetos y los espacios bajo deformaciones continuas. La caracterización de los rasgos topológicos de un espacio permite describir propiedades invariantes del mismo. Por ejemplo, los números de Betti β_n miden número de huecos n-dimensionales que tiene un espacio, definiéndose como β_0 el número de componentes conectadas, β_1 el número de agujeros «circulares» bidimensionales y β_2 el número de agujeros o «vacíos» tridimensionales. Estos huecos se preservan bajo transformaciones continuas. Para ilustrar este concepto consideremos los espacios topológicos de la esfera \mathbb{S}^2 y el toro \mathbb{T}^2 (ver Figura 1). Dado que la esfera y el toro tienen solo una pieza conectada, entonces $\beta_0(\mathbb{S}^2) = \beta_0(\mathbb{T}^2) = 1$, el toro tiene dos agujeros circulares, como se ilustra en la Figura (1), la esfera no tiene ninguno, por tanto $\beta_1(\mathbb{T}^2) = 2$ y $\beta_1(\mathbb{S}^2) = 0$. En el caso de los huecos 2-dimensionales, los dos espacios tienen un vacío tridimensional indicando que $\beta_2(\mathbb{S})^2 = \beta_2(\mathbb{T}^2) = 1$. Dado que las características topológicas de estos dos espacios no son las mismas, entonces se puede intuir que los dos espa-

¹⁷ REVELES, Fermín, *et al. Notas en Persistencia, Probabilidad e Inferencia Estadística para Análisis Topológico de Datos.* 2016.

cios son diferentes debido a que no se pueden deformar continuamente la esfera en el toro, ni viceversa. De esta forma, los números de Betti describen una propiedad invariante bajo deformaciones continuas en un espacio, en particular, la existencia de estos agujeros.

Figura 1. Ejemplo de espacios topológicos simples con sus números de Betti. De izquierda a derecha: un disco, un círculo, una esfera bidimensional y un toro bidimensional¹⁸. Estos tres espacios topológicos se conectan de formas diferentes, resultando en agujeros de naturaleza también diferente. Estos agujeros se preservan bajo deformaciones continuas.



En el caso de una nube finita de puntos n-dimensionales, donde cada punto representa un dato se puede aproximar un espacio métrico finito debido a que los datos se pueden comparar a través de una distancia. Teniendo en cuenta que la nube de puntos del espacio es completamente disconexa, debido a que ningún punto está conectado entre sí, entonces los β_0 serán el número de puntos que tenga el espacio. También considerando que no se pueden crear ciclos en la superficie del espacio entonces no habrían β_1 . No obstante, si se consideran los puntos tienen con una bola alrededor (abierto), estos abiertos inducen una topología, sobre la cual es posible caracterizar los números de Betti. Esta caracterización puede abordarse computacionalmente, utilizando una representación basada en complejos simplicidades, y herramientas de topología algebraica, como la homología.

A continuación, se abordará una descripción más detallada sobre estas herramientas, las cuales hacen parte del TDA. Primero se formalizará el concepto de espacio topológico, y se introducirán conceptos geométricos y topológicos, así como técnicas comúnmente utilizadas en TDA. En particular técnicas relacionadas con el cálculo de homología persistente.

Definición 3.1.1 Una topología en un conjunto X, es una colección \mathcal{U} de subconjuntos de X, tal que:

- $\emptyset, X \in \mathcal{U}.$
- Si $\{S_J \mid j \in J\} \subseteq \mathcal{U}$, entonces, $\cup_{j \in J} S_j \in \mathcal{U}$.
- Si $S_1, S_2 \in \mathcal{U}$ entonces, $S_1 \cap S_2 \in \mathcal{U}$.

Los elementos de \mathcal{U} se llaman abiertos y (X, \mathcal{U}) se denomina espacio topológico.

Definición 3.1.2 Un conjunto $S \in U$ se dice cerrado si su complemento en X es abierto.

Debido a que las características topológicas y geométricas generalmente se asocian con el espacio continuo (espacio topológico conexo y compacto), los datos expresados como un conjunto finito de observaciones no revelan directamente información topológica. Una forma de resaltar ciertas características topológicas de los datos es "conectar" puntos de datos que están cerca entre sí para mostrar la forma global de los datos.

La cuantificación del concepto de proximidad entre puntos de datos generalmente se realiza a través de la distancia. A menudo es conveniente tratar el conjunto de datos como un espacio métrico discreto o una muestra de espacio métrico.

Definición 3.1.3 Un espacio métrico (M, d) es un conjunto M con una función d : $M \times M \rightarrow \mathbb{R}_+$, llamada distancia, tal que $\forall a, b \in M$ se tiene que:

1. $d(x,y) \ge 0$ y d(x,y) = 0 si, y solo si, x = y,

2.
$$d(x, y) = d(y, x)$$
,
3. $d(x, z) \le d(x, y) + d(y, z)$

Para medir cuán lejos están dos subconjuntos compactos de un espacio métrico la distancia de *Hausdorff* es típicamente usada. Esta métrica toma como referencia la mayor de todas las distancias desde un punto en un conjunto al punto más cercano en el otro conjunto. Desde la perspectiva de TDA se proporciona una forma conveniente de cuantificar la proximidad entre diferentes conjuntos de datos del mismo espacio métrico.

Para definir la distancia de *Hausdorff*, introduciremos el concepto de vecindad tubular, como sigue:

Dado un conjunto compacto K de \mathbb{R}^n , definimos la vecindad tubular o contrapeso de K de radio ϵ ; como el conjunto:

$$K^{\epsilon} = \left\{ y \in \mathbb{R}^{n} : \inf_{x \in K} \|x - y\| \le \epsilon \right\} = \bigcup_{x \in K} \overline{B(x, \epsilon)}$$

donde inf representa el infimo de las distancias de los puntos $x \in K$ y el punto $y \in \mathbb{R}^n$ tal que sean menores o iguales a ϵ ; la distancia es definida por la norma d(x, y) = ||y - x||. Este conjunto es equivalente a la unión de todas la bolas cerradas con centro en x y radio ϵ , un ejemplo ilustrativo se aprecia en la Figura 2.

Figura 2. Ejemplo de una vecindad ϵ -contrapeso



De acuerdo con la definición de vecindad contrapeso se define la distancia de Hausdorff como:

Definición 3.1.4 Sean dos conjuntos cerrados K, K' en \mathbb{R}^n , la distancia Hausdorff entre K, y K', $d_H(K, K')$ está definida como el ínf de los ϵ tales que K este en la vecindad tubular de K' y K' en la vecindad tubular de K, es decir:

$$d_H(K, K') = \inf \left\{ \epsilon \ge 0 : K \subset (K')^{\epsilon}, K' \subset K^{\epsilon} \right\}$$
$$= \max \left(\sup_{y \in K'} \left(\inf_{x \in K} \|x - y\| \right), \sup_{x \in K} \left(\inf_{y \in K'} \|x - y\| \right) \right).$$

Donde sup representa el supremo, inf el infimo, y ||x - y|| cuantifica la distancia desde un punto en K' a un punto en subconjunto $K \subseteq K'$; o viceversa. En la Figura 3 se ilustra con detalle la distancia de Hausdorff.

Figura 3. Los componentes del cálculo de la distancia Hausdorff entre la línea roja K' y la línea azul K.



En general, para un compacto $K \subset \mathbb{R}^n$ se define la *función distancia a* K como el ínfimo de las normas de un punto $x \in \mathbb{R}^n$ y un punto $y \in K$ como:

$$d_K(x) := \inf_{y \in K} d(x, y) = \inf_{y \in K} ||x - y||, \quad x \in \mathbb{R}^n$$

En TDA la mayoría de los espacios que se abordan satisfacen la propiedad de Haus-

dorff, permitiendo el cálculo natural de las distancias entre los puntos (datos).

Definición 3.1.5 Sea *X* un espacio topológico, *X* es Hausdorff si para cualesquiera $x, y \in X$, existen vecindades *U* de *x* y *V* de *y* tales que $U \cap V = \emptyset$.

El punto clave para definir invariantes topológicas es el estudio de las relaciones entre dos espacios topológicos. Aquellas propiedades de los espacios que se preserven bajo deformaciones continuas corresponden a invariantes topológicos. Para relacionar dos espacios topológicos utilizamos el concepto de continuidad, esta se define como una aplicación $f : X \to Y$ si para cualquier abierto V de Y, la imagen inversa de V es abierto, es decir, el conjunto $f^{-1}(V) = \{x \in X : f(x) \in V\}$ es un abierto de X.

Dada una colección de abiertos $\{U_i : i \in I\}$, definimos como compacto a $K \subset X$ tal que $K \subseteq \bigcup_J U_i$. Un espacio métrico es compacto si y solo si, para cualquier sucesión de X existe una subsucesión convergente y para el espacio euclídeo d-dimensional \mathbb{R}^d la caracterización de compactos se tiene por el teorema de Heine-Borel, el cual establece que, un espacio $K \subset \mathbb{R}^d$ es compacto si y sólo si, K es cerrado y acotado.

Decimos que un espacio topológico X es conexo si dados cualesquiera abiertos U, V tales que $X \subseteq U \cup V$ y $U \cap V = \emptyset$; entonces $X \subseteq U$ o $X \subseteq V$, equivalente se tiene que un espacio topológico es conexo si y sólo si, los únicos subconjuntos que son abiertos y cerrados al tiempo son el vacío y X mismo.

Una vez que se define un objeto con una determinada estructura, naturalmente se considerará aplicaciones entre ellos compatibles con su estructura, como, por ejemplo, el homeomorfismo entre espacios topológicos.

Definición 3.1.6 Un homeomorfismo $h : X \to Y$ es una biyección tal que h y h^{-1}

son continuas. Dado un homeomorfismo $f : X \to Y$, decimos que X es homeomorfo a Y.

La relación de equivalencia homotópica, es una noción más débil que la noción de homeomorfismo, para definirla veamos el concepto de homotopía.

Si existe una aplicación continua $H : X \times [0,1] \to Y$, tal que para cada $x \in X :$ $H(x,0) = f_0(x)$ y $H(x,1) = f_1(x)$, con $f_0, f_1 : X \to Y$, entonces decimos que f y gson **homotópicas** y en tal caso escribimos $f_0 \simeq f_1$.

Definición 3.1.7 Dos espacios X, Y son **homotópicamente equivalentes** si existen aplicaciones continuas $f : X \to Y$ y $g : Y \to X$, tal que $f \circ g \simeq 1_Y$ y $g \circ f \simeq 1_X$, donde 1_X , 1_Y son las aplicaciones identidad en los correspondientes espacios.

Decimos que un espacio es **contraíble** si es homotópicamente equivalente a un punto. Un ejemplo de ello es el espacio \mathbb{R}^n , donde todo convexo de \mathbb{R}^n es contráctil. Más en general, todo espacio estrellado de \mathbb{R}^n es contráctil (un espacio estrellado X es un espacio que contiene todos los segmentos que unen los puntos de X con un punto base $x_0 \in X$).

La equivalencia topológica es fundamental en la topología algebraica puesto que busca definir invariantes algebraicos en espacios topológicos. Los invariantes topológicos, objetos matemáticos como números, grupos y polinomios nos ayudan a caracterizar las propiedades topológicas de un espacio, esto con el fin de que permanezcan iguales entre espacios homotópicamente equivalentes. Una idea para caracterizar estos invariantes es estudiar la equivalencia topológica entre vecindades contrapeso, estas pueden ser construidas a partir de puntos. Para las construcciones de dichas vecindades se recurre a abiertos centrados en los datos. Estos abiertos permiten definir relaciones de vecindad entre los puntos que son descritas como simplejos. Y que, a su vez, son agrupadas en complejos simpliciales que

describen todas las relaciones de vecindad entre los puntos. A continuación, se detallarán estas estructuras.

3.2. Complejos simpliciales

La homología simplicial se define para los poliedros, estos se pueden entender como un espacio construido a base de puntos, segmentos, triángulos y sus análogos en dimensiones mayores, son esencialmente un espacio topológico que admite una triangulación por un complejo simplicial. Los complejos simpliciales son objetos matemáticos que permiten introducir modelos finitos o combinatorios de espacios topológicos, son definidos por sus vértices, donde a cada complejo K se le asociará un espacio topológico, que llamaremos la realización geométrica de K, la cual se construye pegando convexos determinados por los simplejos. En esta sección definiremos de forma precisa estos conceptos. Recordemos en primer lugar la definición de puntos afínamente independientes.

La combinación lineal $\sum_{i=0}^{m} \lambda_i p_i$ donde para cada índice $i, \lambda_i \in \mathbb{R}$ y $p_i \in \mathbb{R}^n$, es una **combinación convexa** si y sólo si, $\sum_{i=0}^{m} \lambda_i = 1$, $\lambda_i \ge 0$, $i = 0, \ldots, m$. Definimos **convexo** al subconjunto de \mathbb{R}^n si este contiene al segmento de línea que une cualesquiera dos de sus puntos, **envolvente convexa** de $P = \{p_0, \ldots, p_r\}$ en \mathbb{R}^n al conjunto de todas la combinaciones convexas de puntos de P y **afínamente independientes** a los puntos $\{x_0, x_1, \ldots, x_r\} \subset \mathbb{R}^n$ si los vectores $x_1 - x_0, x_2 - x_0, \ldots, x_r - x_0$ son linealmente independientes.

Definición 3.2.1 Si $\{x_0, \ldots, x_q\} \subset \mathbb{R}^n$ es un conjunto de q + 1 puntos afínamente independientes, a la envolvente convexa de estos puntos se le llama simplejo q-dimensional o q-simplejo y se denota por (x_0, \ldots, x_q) .

Así, un q- simplejo es un punto, segmento de recta, triángulo, tetraedro, para q = 0, 1, 2, 3, respectivamente como se ilustra en la Figura 4.





Los simplejos, σ , de K son llamados **caras** y si $\sigma \in K$ tiene m + 1 elementos entonces, la dimensión de σ es m y se llama m - simplejo y la dimensión de K es $\dim(K) = \max{\dim(\sigma); \sigma \in K}$.

Definición 3.2.2 Un complejo simplicial K es una colección finita de simplejos en \mathbb{R}^n , tales que:

- 1. Si $\sigma_1, \sigma_2 \in K$ entonces $\sigma_1 \cap \sigma_2 = \emptyset$ ó $\sigma_1 \cap \sigma_2$ es una cara común de σ_1 y σ_2 .
- 2. Si $\sigma \in K$ y $\tau \leq \sigma$ entonces $\tau \in K$.

A continuación, en la Figura 5 vemos un ejemplo de una estructura que no satisface la definición de complejo simplicial (a) y una representa un complejo simplicial $K = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, (x_1, x_3), (x_1, x_2), (x_2, x_3), (x_2, x_4)\}$ (b).

Figura 5. a) un no complejo simplicial, b) complejo simplicial.



Un subconjunto K' de K que es complejo simplicial se denomina subsimplejo de K. Se llama q - esqueleto de K y se denota K^q al subsimplejo $K^q = \{\sigma \in K; \dim(\sigma) \leq$ q}, entonces K^0 es el conjunto de vértices de K.

Se considera que los simplejos están inmersos en un espacio euclidiano, por tanto *K* es un complejo simplicial geométrico.

Definición 3.2.3 Sea *P* un conjunto finito. Un **complejo simplicial abstracto** es una colección finita *K* de subconjuntos de *P*, tal que los elementos de *P* pertenecen a *K* y si $\sigma \in K$ y $\tau \subset \sigma$, entonces $\tau \in P$.

Ejemplo 3.2.4 Sea $S = \{s_i\}, i \in \{1, ..., 7\}$, vemos que el complejo simplicial contiene:

- Un 3- simplejo (s_2, s_3, s_5, s_6) .
- Un 2-simplejo (s_2, s_4, s_7) .
- Once 1-simplejos, que son los once segmentos de recta.
- Siete 0-simple $\{\{s_1\}, \{s_2\}, \{s_3\}, \{s_4\}, \{s_5\}, \{s_6\}, \{s_7\}\}$

En particular, tomando el 2-simplejo (s_2, s_4, s_7) vemos que tiene:

- (s_2, s_5, s_3, s_6)
- tres 1-simple jos (caras) $\{(s_2, s_5), (s_2, s_7), (s_5, s_7)\}$.
- *tres* 0-*simplejos* {{ s_2 }, { s_5 }, { s_7 }}.
- la cara del subconjunto vacío: (Ø).

Ejemplo 3.2.5 Dado *X* un conjunto no vacío cualquiera, se define un complejo simplicial a partir de *X* tomando como vértices todos los elementos de *X* y como simplejos todos los subconjuntos finitos y no vacíos de *X*. Figura 6. Representación del Ejemplo 3.2.4



Los subsimplejos de K no están pensados como objetos geométricos en \mathbb{R}^n , sino sólo como conjuntos de vértices. Usualmente, una nube de puntos discreta tiene asociado un complejo simplicial, dicho complejo será equivalente a un espacio métrico compacto formado por la unión de bolas cerradas.

3.2.1. Nervios Supongamos que se tiene un conjunto finito de datos *S*, en un espacio topológico *X*, la idea principal de TDA es aproximar *X*. Definimos como recubrimiento abierto de *S* en *X* a una familia $\mathcal{U} = {\mathcal{U}_i}_{i \in I}$ de subconjuntos abiertos de *X* tales que $S \subseteq \bigcup_{i \in I} \mathcal{U}_i$.

Definición 3.2.6 Sean X un espacio métrico y $\mathcal{U} = {\mathcal{U}_i}_{i \in I}$, con I finito, un recubrimiento de X. El **nervio** de \mathcal{U} , es el complejo simplicial abstracto cuyo conjunto de vértices es I y para el que una familia de índices $\{i_0, \ldots, i_q\} \subset I$ forma un q-simplejo si, y solo si, $\bigcap_{j=0}^q U_{i_j} \neq \emptyset$. Notaremos el nervio por $\mathcal{N}(\mathcal{U})$.

Note que los conjuntos unitarios pertenecen a $\mathcal{N}(\mathcal{U})$ y que si $\{i_0, \ldots, i_q\} \in \mathcal{N}(\mathcal{U})$, entonces todo subconjunto de $\{i_0, \ldots, i_q\}$ forma también un simplejo ya que la intersección de los conjuntos asociados a esos índices es no vacía al haber menos elementos. Así pues, el nervio de un recubrimiento es efectivamente un complejo simplicial abstracto.

Decimos que un espacio X es paracompacto si para cualquier cubierta abierta \mathcal{U} de

Figura 7. Representación del nervio (derecha) de la unión de tres bolas abiertas (izquierda).



X, existe un refinamiento \mathcal{U}' que es localmente finito; esto es, cualquier $x \in X$ tiene una vecindad tal que intersecta sólo un número finito de $U_i \in \mathcal{U}'$. En particular, un espacio compacto es paracompacto.

Teorema 3.2.7 (Teorema del Nervio) ¹⁹ Sea $\mathcal{U} = \{U_i\}_{i \in I}$ una cubierta abierta del espacio paracompacto *X*, tal que cualquier intersección finita de los U_i es, o bien vacía, o contraible. Entonces, *X* y $\mathcal{N}(\mathcal{U})$ son homotópicamente equivalentes.

Este teorema es importante para la inferencia topológico-geométrica y la topología computacional, puesto que posibilita unir la topología del espacio X a través de un complejo simplicial. Particularmente el espacio X es la unión de convexos (bolas) en \mathbb{R}^n , X será homotópicamente equivalente al nervio de la cubierta obtenida por la unión de dichos convexos. A continuación, se definen los dos complejos más ampliamente usados en TDA.

3.2.2. Ejemplos de complejos Dado un conjunto de datos, o más generalmente un espacio topológico o métrico, existen muchas formas de construir complejos simpliciales. Presentamos aquí dos ejemplos clásicos que se utilizan ampliamente en la práctica.

¹⁹ CHAZAL, Frédéric y MICHEL, Bertrand. *An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists.* 2021. arXiv: 1710.04019.

3.2.3. complejos de \check{C} **ech.** Como primer ejemplo de complejo usado en el análisis de datos, tenemos los complejos de \check{C} ech se construyen como el nervio de un recubrimiento formado por bolas abiertas. Sea *S* un conjunto finito contenido en un subespacio euclídeo $Y, S \subseteq Y \subseteq \mathbb{R}^n$.

Definición 3.2.8 Sea $B_{\epsilon}(x)$ la bola abierta de radio ε centrada en x:

$$B_{\varepsilon}(x) = \{ y \in Y \mid d(x, y) < \varepsilon \}, \text{ con } \varepsilon \in \mathbb{R}, y x \in Y.$$

Se define entonces el **complejo** \check{C} ech, $\check{C}_{\varepsilon}(S)$, como el nervio del recubrimiento de $S: U_{\varepsilon}(S) = \{B_{\varepsilon}(x) \mid x \in S\}$, donde $S \subseteq Y$ y $\varepsilon \in \mathbb{R}$.

En este complejo un k- simplejo es una colección de k + 1 puntos distintos de S tal que las bolas de radio ε con centro x_i tienen intersección no vacía. Asimismo, por el teorema del nervio, dado que las bolas son convexas, y en consecuencia, contráctiles, el nervio captura la topología del recubrimiento, es decir tienen el mismo tipo de homotopía. El cómputo del \breve{C} ech para dimensiones mayores a r se vuelve rápidamente intratable.

3.2.4. Complejos de Vietoris-Rips: El complejo de Vietoris-Rips es el complejo más grande que tiene como 1-esqueleto el complejo \check{C} ech, es decir, el complejo \check{C} ech es un subsimplejo de los Vietoris-Rips. Consideremos un subconjunto discreto S de un espacio euclídeo \mathbb{R}^n correspondiente a una nube de puntos.

Definición 3.2.9 Sean *S* un conjunto finito y *X* un subespacio euclídeo tal que $S \subseteq X \subseteq \mathbb{R}^n$. Dado un $\varepsilon \ge 0$, llamamos **complejo de Vietoris-Rips** al complejo simplicial cuyos vértices son los puntos de *S* y en el que los puntos $\{x_0, \ldots, x_q\}$ forman un q - simplejo si y solo si $d(x_i, x_j) \le \varepsilon$ para todo $0 \le i, j \le q$. Lo denotamos por $V_{\varepsilon}(S)$.

El complejo Vietoris-Rips es un complejo simplicial abstracto y es menos costoso computacionalmente que el *Č*ech en lo que respecta a la entrada y almacenamiento

de datos.



Figura 8. El complejo \check{C} ech $\check{C}_{\alpha}(X)$ (izquierda) y el Vietoris-Rips $V_{2\alpha}(X)$ (derecha)²⁰

En la Figura 8 la parte inferior de $\check{C}_{\alpha}(X)$ es la unión de dos triángulos adyacentes, mientras que la parte inferior de $V_{2\alpha}(X)$ es el tetraedro atravesado por los cuatro vértices y todas sus caras. La dimensión del complejo \check{C} ech es 2 y la dimensión del complejo Vietoris-Rips es 3.

3.3. HOMOLOGÍA SIMPLICIAL

Dotamos a los complejos simpliciales de una estructura algebraica para definir la homología simplicial, orientándolos mediante una relación de equivalencia. Esta relación genera dos clases de equivalencia para cada simplejo, lo cual permite establecer una cadena simplicial y opuesta (el signo de la clase). Definimos la clase de equivalencia como sigue:

Definición 3.3.1 Sea *K* un complejo simplicial. Dado un simplejo $\sigma = (v_0, ..., v_q)$, con q > 0, se define la siguiente relación de equivalencia sobre el conjunto de todas las ordenaciones de dichos vértices de σ :

$$(v_0,\ldots,v_q)\sim (v_{\pi(0)},\ldots,v_{\pi(q)})$$

Donde π es una permutación par de los índices. Esta es una relación de equivalencia y origina dos clases de equivalencia, cada una llamada orientación de σ .

Por tanto, si se puede pasar de uno a otro mediante un número par de trasposiciones en el orden de los vértices los dos simplejos con los mismos vértices poseen la misma orientación. Luego, si se escogen las orientaciones, σ se dice un simplejo orientado y los 0-simplejos van a tener solo una orientación debido a que están formados por un único vértice.

En adelante, escribiremos $[v_0, \ldots, v_q]$ para denotar el simplejo (v_0, \ldots, v_q) por comodidad y siguiendo la orientación definida por la ordenación de los índices.

Definimos las estructuras algebraicas de un espacio vectorial de q- cadenas y el operador borde de q a una aplicación lineal, para ello, consideremos un cuerpo \mathbb{F} .

Definición 3.3.2 Se define el \mathbb{F} -espacio vectorial de q-cadenas simpliciales orientadas y se denota por $C_q(K; \mathbb{F})$, como el cociente del \mathbb{F} -espacio vectorial generado por todos los q-simplejos σ , de K por el subespacio generado por los $\sigma^1 + \sigma^2$, donde σ^1 y σ^2 representan las dos posibles orientaciones de un q-simplejo. Los elementos de $C_q(K; \mathbb{F})$ se llaman q- cadenas simpliciales orientadas.

En adelante, por abuso de notación, escribiremos $C_q(K)$ para denotar $C_q(K; \mathbb{F})$.

Definición 3.3.3 Se define el operador borde de orden q a la transformación lineal:

$$\partial_q : C_q(K) \to C_{q-1}(K)$$

dada por extensión lineal de

$$\partial_q \left([v_0, \dots, v_q] \right) = \sum_{i=0}^q (-1)^i \left[v_0, \dots, v_{i-1}, v_{i+1}, \dots, v_q \right]$$

donde $[v_0, \ldots, v_{i-1}, v_{i+1}, \ldots, v_q]$ representa el (q-1)- símplice orientado resultante de eliminar en $[v_0, \ldots, v_q]$ el vértice que ocupa la posición *i*.

Definición 3.3.4 Se denominan *q*-ciclos a los elementos $x \in Z_q(K) = \ker(\partial_q) y$ *q*-bordes a los elementos de $B_q(K) = \operatorname{Im}(\partial_{q+1})$.

Estos grupos guardan entre sí una relación, ya que para todo entero q toda (q + 1)-cadena se tiene que $\partial_q \circ \partial_{q+1} = 0$ lo cual implica que Im $(\partial_{q+1}) \subseteq \ker(\partial_q)$ para todo q. Entonces, se tiene que toda q-borde es también un q-ciclo, es decir, $B_q(K) \subseteq Z_q(K)$, como se ilustra en la Figura 9.

Figura 9. Relación entre los grupos de q-ciclos y q-bordes.



Definición 3.3.5 Se define el q-ésimo \mathbb{F} -espacio vectorial de homología simplicial de K como el siguiente cociente:

$$H_q(K) = \frac{Z_q(K)}{B_q(K)}.$$

Se define también el q-ésimo número de Betti de K, denotado por β_q , como la dimensión de $H_q(K)$.

Los número de Betti como vimos a inicio de este capítulo, describen las características importantes, es decir, describe la homología de un espacio. De manera intuitiva miden el número de componentes conexas, ciclos y hoyos de un complejo simplicial *K*. Entonces, para q = 0, 1, 2 y q = n se tiene que:

- β_0 representa el número de componentes conexas del espacio.
- β_1 es el número de agujeros unidimensionales o "circulares".
- β_2 es el número de "vacíos" o "cavidades" bidimensionales.
- β_n es el número de "huecos" de dimensión n.

Como se ilustra en la Figura 1 en el caso de la esfera, esta tiene tiene $\beta_0 = 1, \beta_1 = 0, \beta_2 = 1$, puesto que tiene solo una única componente conexa, ningún túnel, y una sola cavidad.

Los números de Betti se pueden calcular en estructuras como complejos simpliciales, entonces dada una nube de puntos, si consideramos una bola fija de radio un ϵ en cada uno de los puntos y generamos el complejo simplicial podemos calcular las invariantes asociados a dicho complejo. En la Figura 10 se ilustra la representación de un conjunto de datos en un espacio, el complejo simplicial y la respectiva invariante asociada al complejo. Dicha invariante es el número de Betti para el complejo simplicial de la nube de puntos.





Para un ϵ fijo se pueden hallar las características topológicas, pero cómo estar seguros que dicho ϵ es el indicado para hallar los invariantes que mejor describen la nube de puntos. Para resolver esa inquietud introduciremos el concepto de homología persistente, que consiste en calcular los números de Betti sobre filtraciones de complejos simpliciales.

3.4. PERSISTENCIA

Antes de introducir la herramienta principal del TDA, la homología persistente, ilustraremos el concepto de persistencia con el siguiente ejemplo.

Específicamente, sea $f : \mathbb{R} \longrightarrow \mathbb{R}$ una función C^{∞} . Supongamos que x es un punto crítico y f(x) es un valor crítico si f'(x) = 0. Decimos además que un punto crítico es no-degenerado si $f''(x) \neq 0$. Supongamos que f tiene solamente puntos críticos no-degenerados con valores críticos distintos. Entonces, cualquier punto es un mínimo o un máximo local.

Consideremos los subconjuntos de nivel $\mathbb{R}_{\alpha} = f^{-1}((-\infty, \alpha])$ para cada cada $\alpha \in \mathbb{R}$. Mientras α crece hacia ∞ , la conectividad de \mathbb{R}_{α} permanece igual excepto cuando pasamos un valor crítico. En puntos mínimos locales surgen nuevas componentes y en los máximos locales dos componentes se unen en una sola.

Realizamos un emparejamiento de los puntos críticos de f y cuando una nueva componente conexa, decimos que el mínimo local que la crea representa dicha componente. Luego, cuando se pasa un máximo local y se unen dos componentes, emparejamos el máximo, con el mínimo local más reciente que representa esta componente. En ese momento, el otro mínimo local se vuelve el representante de la componente que resulte de la unión.

Cabe notar que los puntos críticos que están emparejados pueden no ser adyacentes. Definimos la persistencia del par (x, y) como f(y) - f(x). Esta persistencia es medida por el diagrama de persistencia al llevar cualquier punto (x, y) al punto con coordenadas los valores críticos respectivos,

$$\operatorname{dgm}(f) = \{(f(x), f(y))\} \subset \mathbb{R}^2.$$

En el diagrama de persistencia todos los puntos están sobre la diagonal y = x y la persistencia es la distancia vertical a esta línea diagonal, ver Figura 11.

Figura 11. Emparejamiento de máximos y mínimos de una función $f : [0,1] \rightarrow \mathbb{R}$ y su diagrama de persistencia ²¹.



La construcción formal matemática del diagrama de barras lo abordaremos en la siguiente sección.

3.5. HOMOLOGÍA PERSISTENTE

La homología persistente es una herramienta poderosa para calcular, estudiar y codificar de manera eficiente características topológicas multiescala de familias anidadas de complejos simpliciales y espacios topológicos. Entre otras este análisis proporciona algoritmos eficientes para calcular los números de Betti de cada complejo en las familias consideradas. También codifica evolución de los grupos de homología de los complejos anidados a través de las escalas.
Si se considera la construcción de un complejo simplicial usando bolas de diámetro ϵ centradas en puntos de un conjunto, se tiene que para un ϵ pequeño, el complejo simplicial es sólo un conjunto disjunto de vértices y para un ϵ grande, sería un gran conjunto de simplejos. Por lo que uno se cuestiona por el valor adecuado del ϵ de tal forma que nos brinde una buena estructura, para ello la homología persistente nos brinda una respuesta.

Se obtiene una secuencia de complejos simpliciales anidados, llamada filtración, cuando incrementamos el ϵ , de tal forma características topológicas aparecen y desaparecen, estas existen durante un rango de tiempos. Este planteamiento nos permite verlas como una colección de códigos de barras (barcodes), uno en cada dimensión, donde cada una es representada por una barra. Entonces, entre mayor sea la duración de vida de una característica (barra), mayor será la longitud de la misma. Más adelante abordaremos la construcción de dichas barras y cómo están relacionadas con el diagrama de persistencia.

La idea concreta de la homología persistente es reunir toda la información topológica en una única estructura de un espacio filtrado por subespacios. Para ello definimos una filtración como sigue.

Definición 3.5.1 *Sea K un complejo simplicial. Una filtración es una sucesión de subcomplejos crecientes de K*

$$\emptyset = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n = K.$$

Cada subcomplejo K_i se denomina i-ésimo nivel de la filtración.

Se puede interpretar una filtración como una manera de obtener el complejo total K a partir de una familia de subcomplejos siguiendo el criterio de cada nivel K_i al

estado de dicha construcción en el instante de tiempo *i*.

Un ejemplo de una de las diversas formas que se puede definir una filtración en un complejo es la filtración por esqueletos, tomando $K_0 = \emptyset$ y K_i el i - 1 -esqueleto de K. Otro ejemplo es la filtración usada comúnmente del complejo de Vietoris-Rips definido en 3.2.9, cuyos simplejos $V_{\epsilon}(P)$ construidos sobre P están definidos por

$$\sigma = \{p_0, \dots, p_m\} \in R(P, \epsilon) \quad \Longleftrightarrow \quad \|p_i - p_j\| \le \epsilon, \forall i, j \in \{0, \dots, m\}.$$

A continuación e introducirán los conceptos básicos sobre homología persistente, por tanto, dada una filtración $\{K_i\}_{i=0,...,n}$, para cada $i \leq j$, la inclusión canónica $K_i \hookrightarrow K_j$ induce inclusiones de *p*-ciclos $Z_p(K_i) \hookrightarrow Z_p(K_j)$ e inclusiones de *p* bordes $B_p(K_i) \hookrightarrow B_p(K_j)$ y para cada dimensión *p*, denotaremos por $\iota_p^{i,j} : H_p(K_i) \to$ $H_p(K_j)$ a la aplicación inducida en homología por dicha inclusión. Por tanto, se tiene el siguiente diagrama de espacios vectoriales:

$$0 = H_p(K_0) \to H_p(K_1) \to \cdots \to H_p(K_n) = H_p(K).$$

Para el estudio de clases de ciclos, el concepto de homología se asocia a una filtración de un complejo, entonces, a medida que nos movemos en la filtración, desde K_i hasta K_j , se pueden anular algunas clases de homología y aparecer nuevas clases, esto se logra al unirse con otras o al volverse triviales, veamos ahora el concepto de espacios vectoriales en homología persistente.

Definición 3.5.2 Se definen a las imágenes de los homomorfismos inducidos por la inclusión $H_p^{i,j}(K) = \operatorname{Im} \iota^{i,j}$, para $0 \le i \le j \le n$, como los *p*-ésimos espacios vectoriales de homología persistente asociados a una flitración de K. Y el correspondiente *p*-ésimo número de Betti de persistencia como la dimensión de estos espacios vectoriales, $\beta_p^{i,j}(K) = \dim H_p^{i,j}(K)$. Los p-ciclos creados en el nivel i y que no son nulos en el nivel j, serán en este contexto significativos.

Para definir la "vida" de una clase γ en $H_p(K_i)$, es decir, qué tanto persiste, se dice que nace en K_i si $\gamma \notin H_p^{i-1,j}(K)$, y si γ nace en K_i , entonces se dice que muere entrando en K_j si se une con una clase que proviene de un nivel anterior al movernos de K_{j-1} hasta K_j , es decir

$$\iota^{i,j-1} \notin H_p^{i-1,j-1}$$
, pero $\iota^{i,j}(\gamma) \in H_p^{i-1,j}$.

Para ilustrar este concepto, en la Figura 12 la clase γ nace en K_i ya que no aparece en la imagen sombreada de $H_p^{i-1} = H_p(K_{i-1})$, y γ muere entrando en K_j , debido a que es la primera vez que su imagen se une con la imagen de $H_p^{i-1} = H_p(K_{i-1})$.

Figura 12. La clase γ nace en K_i ya que no aparece en la imagen sombreada de $H_p^{i-1} = H_p(K_{i-1})$. Asimismo, γ muere entrando en K_j , puesto que es la primera vez que su imagen se une con la imagen de $H_p^{i-1} = H_p(K_{i-1})$.



Llamamos persistencia a la diferencia entre los valores de la función correspondientes, si γ nace en K_i y muere entrando en K_j , como pers $(\gamma) = a_j - a_i$; **intervalo de persistencia** de γ al intervalo semiabierto $[a_i, a_j)$. Es infinita tanto la persistencia como su índice si γ nace en K_i pero no muere nunca.

Para visualizar las nociones de nacimiento, muerte y persistencia se pueden hacer

de dos formas, mediante los llamados códigos de barras o través de diagramas de persistencia. A continuación, explicaremos de forma breve en qué consisten.

Definición 3.5.3 Sea $\{K_i\}$ una filtración y K un complejo simplicial, si $0 \le p \le \dim(K)$, entonces el p-ésimo código de barras de la filtración se obtiene como sigue:

- En el primer cuadrante del plano se marcan en el eje horizontal los niveles de la filtración.
- En el eje vertical una ordenación arbitraria de intervalos de persistencia [a_i, a_j) o semirrectas [a_i,∞).

En la Figura 13 se observa un ejemplo de filtración simple de una nube de puntos y el códigos de barras de homología persistente resultante, se puede ver cómo nace una característica en la filtración K_3 y muere en la K_6 , y otra que nace en la k_4 y muere en K_5 . Note que para cada nivel *i* de la filtración, $\beta_p(K_i)$ es el cardinal de la intersección de la vertical sobre dicho nivel con los intervalos.





Cada característica representada por una barra, cuanto mayor sea la duración de vida, mayor será la longitud de su barra. Entonces, el código de barras nos permite ver cuáles son las características que persisten más y que por tanto son las que más aportan información topológica.

4. EL ANÁLISIS TOPOLÓGICO, LA MARCHA Y LA ENFERMEDAD DE PARKINSON

En este trabajo, los patrones de marcha registrados en secuencias de video, serán modelados desde una perspectiva topológica. Esto con el fin de encontrar patrones asociados a la enfermedad de Parkinson, según el comportamiento dinámico de un conjunto de trayectorias (serie de tiempo) extraídas del video.

Específicamente en este trabajo se realizaron dos análisis para caracterizar patrones de la marcha parkinsoniana y la de población control. En una primera aproximación, se tomó el cálculo del centro de masa durante el desplazamiento de cada paciente y en una segunda, un conjunto de puntos asociados a diferentes articulaciones del cuerpo²² ver Figura 14(a). Para poder construir el espacio topológico se realizó un análisis de serie de tiempo a cada una de estas trayectorias (representadas como un vector $\{x_0, ..., x_i, ..., x_n\}$, donde cada x_i es la posición del centro de masa o punto clave en un tiempo t = i, los tiempos corresponden al número de cuadros que tenga el video) con el objetivo de representar, calcular y analizar las características de cada señal.

En la Figura 14(b) vemos la diferencia de distribución entre los grupos control y párkinson, para el cálculo de la constancia de las señales, evidencia que el grupo parkinsoniano es menos constante, es decir, es más aleatorio respecto al grupo control (debido al temblor y la inestabilidad en la marcha del párkinson), y adicionalmente, una muestra de serie temporal del centro de masa asociado a la marcha de

²² G. HIDALGO Z. Cao, T. Simon S.-E. Wei H. Joo y SHEIKH, Y. *OpenPose library*.

un sujeto control y uno con EP, ilustrando cómo son las señales de cada grupo.

Figura 14. Trayectoria del centro de masa y puntos clave (a) y diferencia de distribución entre los grupos control y párkinson, para el índice de constancia, se observa que el conjunto parkinsoniano es menos periódico (b).



Para la caracterización de los rasgos que describen trayectorias anormales asociadas a la enfermedad de párkinson, se propone la caracterización cuantitativa de periodicidad y la entropía. La periodicidad de la marcha comúnmente se ve alterada en pacientes con párkinson, principalmente por las afectaciones del sistema de control motor ¹¹. Por otro lado, la entropía busca caracterizar las variaciones en nivel de incertidumbre previamente observadas para la marcha de pacientes con párkinson²³. Estas dos características pueden estimarse directamente a partir de la serie de tiempo. Sin embargo, su cuantificación depende de un conjunto de parámetros de difícil escogencia. En contraste, en este trabajo proponemos utilizar una descripción basada en homología persistente para caracterizar estos dos rasgos en la serie

²³ "Measuring signal fluctuations in gait rhythm time series of patients with Parkinson's disease using entropy parameters". En: *Biomedical Signal Processing and Control* 31 (2017), págs. 265-271.

de tiempo.

Dado que la marcha en su naturaleza sigue un ciclo, se puede pensar en cuantificar qué tan predecible son los ciclos y cómo se ven afectados con la enfermedad de Parkinson. La marcha puede ser predecible (**predictibilidad**) ya que tiene cierta **constancia** en el tiempo (es decir, hay poco o ningún cambio de magnitud en cualquier escala de tiempo), o porque el grado en que el cambio de cantidad depende del ciclo, pero es constante en la marcha larga (**contingencia**).

En este capítulo, abordaremos los descriptores temporales haciendo uso del análisis topológico de datos como una herramienta para la caracterización de la periodicidad y entropía de una serie de tiempo. Estas dos características se pueden caracterizar alternativamente utilizando teoría de la información, particularmente a través los índices de Colwell²⁴. A continuación, se describen en detalle cómo se modelan tales descriptores temporales.

4.1. DESCRIPTORES DE SERIES TEMPORALES DESDE LA TOPOLOGÍA

En condiciones normales el centro de masa y ciertos puntos articulares del esqueleto humano durante la marcha siguen patrones periódicos, oscilatorios, descritos en periodos naturales del ciclo de marcha (ver Anexos 6). Haciendo uso de herramientas topológicas se puede caracterizar la periodicidad de las señales, definiendo conceptos como persistencia máxima del código de barras de dimensión 1, como indica la ecuación (1). Ya que, el concepto de entropía ayuda a la caracterización de las series temporales asociadas a la marcha de los dos grupos de estudio, ha-

²⁴ COLWELL, Robert. "Predictability, Constancy, and Contingency of Periodic Phenomena". En: *Ecology* 55 (ago. de 1974), págs. 1148-1153. DOI: 10.2307/1940366.

llando las invariantes de cada señal, y así, comparar los dos grupos. Por tanto, este concepto hace uso de código de barras de dimensión 0, mide la longitud de las diferentes barras de los códigos de barras y se define con base a la entropía de Shannon donde la probabilidad p_i se toma como la longitud de la barra entre la suma de las longitudes de las barras, como explica la ecuación (2).

En este trabajo, se hizo uso de SW1PerS (por sus siglas en ingles Sliding Windows and 1-Persistence Scoringes), un método para cuantificar la periodicidad de series de tiempo. Este método se compone por las herramientas de ventanas deslizantes (Sliding Windows) y la persistencia 1-dimensional. Este algoritmo transforma la serie temporal de entrada en una nube de puntos e interpreta la periodicidad de la señal original como "circularidad" de este conjunto²⁵. Uno de los objetivos principales de este método es hallar ventanas deslizantes y persistencia, para así calcular la periodicidad. Específicamente, este método consiste en tomar las trayectorias como una serie de tiempo g_0, g_1, \ldots, g_S medida en los tiempos t_0, t_1, \ldots, t_s , mapeada del intervalo $[t_0, t_s]$ linealmente en $[0, 2\pi]$ y aplicando interpolación cubic spline se obteniendo una función continua $g:[0,2\pi]\longrightarrow \mathbb{R}$ de modo que $g(0)=g_0$ y $g(2\pi)=g_s$. Se define tamaño de ventana a un $0 < w < 2\pi$, y para cada $t \in [0, 2\pi - w]$ se considera la gráfica de q restringida al intervalo [t, t + w] como la ventana deslizante. En la Figura 15.a) se ilustra la ventana deslizante de longitud w que inicia en $t \in [0, 2\pi]$ y la función g. Se desliza la ventana variando t de 0 a $2\pi - w$ y se obtiene un fragmento de g, ahora, si dichos fragmentos son similares (próximos), entonces la imagen emergente es análoga a la Figura 15.b), dando a entender que dicha señal es periódica.

²⁵ PEREA J.A., Deckard A. Haase S.B. et al. "SW1PerS: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data". En: *BMC Bioinformatics* 13.1 (nov. de 2015), pág. 257.





Para su construcción, se definen dos valores $M \in \mathbb{R}^+$ y $\tau = \frac{w}{M}$ para algún tamaño de ventana w y la ventana deslizante discretizada que inicia en t es dada por el vector

$$SW_{M,\tau}g(t) = \begin{bmatrix} g(t) \\ g(t+\tau) \\ \vdots \\ g(t+M\tau) \end{bmatrix}$$

Dado que *t* toma valores de un conjunto finito denso $T \,\subset [0, 2\pi - w]$, entonces da como resultado una colección de puntos $X \subset \mathbb{R}^{M+1}$ el cual se llamará "Nube de puntos de ventana deslizante" o "matriz de ventanas deslizantes" donde cada fila (vector) es una ventana. En dicha nube, dos puntos están cerca si y solo si los fragmentos correspondientes que discretizan son similares. De ello se deduce que la medida en que *X* puede considerarse como una muestra de una curva cerrada sin auto-intersecciones (un círculo topológico) que corresponde directamente con la periodicidad de *g* medida con ventanas de tamaño w^{25} .

La nube de puntos resultante se centra en la media y se normaliza; es decir, reem-

plazamos X por

$$\bar{X} = \left\{ \frac{\mathbf{x} - \operatorname{mean}(\mathbf{x})}{\|\mathbf{x} - \operatorname{mean}(\mathbf{x})\|} : \mathbf{x} \in X \right\}$$

con el sentido de ajustar la tendencia en la serie de tiempo original. Dados los datos normalizados \bar{X} , se extrae el diagrama de persistencia correspondiente a la nube de puntos, el cual indica las características homologicas H_0 y H_1 . Luego, tomando la persistencia máxima en H_1 y (b, d) el par que la alcanza, es decir la barra más larga (siendo *b* el nacimiento y *d* la muerte) el cual servirá para cuantificar la periodicidad de la nube de puntos de ventanas deslizantes cuando se mida con cierto tamaño de ventana.

Se establece el índice de periodicidad planteado por Perea y Harer (que puede estar asociado a la contingencia)²⁶ como el valor p entre 0 y 1 (no periódico y periódico):

$$p = 1 - \frac{\max(d-b)}{\sqrt{3}}.$$
(1)

Este valor de periodicidad es útil para poder construir una característica haciendo uso de la homología y así caracterizar la periodicidad de marcha parkinsoniana de la normal.

Con base a la persistencia H_0 , en *An entropy-based persistence barcode*²⁷ se propone realizar una medida de la entropía y caracterizar la constancia topológicamente. Entonces, para un complejo simplicial *K* de dimensión $m \operatorname{con} n+1$ simplejos, las barras que tiendan a infinito tendrán un valor [a, n+1) de esta manera se tendrán valores finitos que los representen. Luego, para el conjunto de barras $B = [a_j; b_j] : j \in J$,

²⁶ PEREA, Jose y HARER, John. Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis. 2013. arXiv: 1307.6188 [math.AT].

²⁷ "An entropy-based persistence barcode". En: *Pattern Recognition* 48.2 (2015), págs. 391-401.

se define la entropía de Shannon como:

$$H(B) = -\sum_{j \in J} p_j * \log(p_j)$$
⁽²⁾

donde $p_j = l_j/L$, $l_j = b_j - a_j$, y $L = \sum_j l_j$, es decir, l_j es la longitud de la barra j de B y L la suma de todas las longitudes de las barras de B, y así p_j es la probabilidad de ocurrencia de la barra j.

Intuitivamente, la entropía mide la longitud de las diferentes barras de los códigos de barras. Un código de barras con longitudes uniformes tiene una pequeña entropía. En la Tabla 1 ilustramos las propuestas de índices de constancia y contingencia basada en las persistencias H_0 y H_1 .

Tabla 1. Propuesta de índices topológicos

Índice	Propuesta topológica	Cálculo de Índice
Constancia TDA	Entropía con Persistencia H_0	$H(B) = -\sum_{j \in J} p_j * \log(p_j)$
Contingencia TDA	Periodicidad con SW1PerS ²⁵	$P_y = 1 - \frac{\max\{d-b\}}{\sqrt{3}}$

4.2. DESCRIPTORES DE SERIES TEMPORALES FUNDAMENTADOS EN LA EN-TROPÍA

El ciclo de la marcha se puede ver afectado de diversas formas en la enfermedad de párkinson (ver Anexo 1), estudiar qué tan predecibles y constantes son las trayectorias asociadas a la marcha normal y parkinsoniana brinda una caracterización de los dos grupos de estudio. Los siguientes conceptos brindan una propuesta de dicho planteamiento.

La predictibilidad es una medida de variación que permite relacionar el estado de un fenómeno periódico dado un tiempo específico, describe lo predecible que es una

señal, incluyendo la frecuencia e intensidad de la misma así como su distribución periódica dentro del régimen temporal ²⁴. Teniendo en cuenta esta definición, esta es entonces una medida apropiada para caracterizar las series temporales de un sujeto control y uno con párkinson, puesto que en la enfermedad de Parkinson, un fenómeno que caracteriza la marcha son las diversas alteraciones de la misma, modificando los ciclos de marcha y diferenciándola de la marcha normal.

Un método para aproximar estas características lo propone Robert Colwell con unos índices para medir la predictibilidad de fenómenos periódicos, el cual depende sustancialmente del concepto de entropía *H* propuesta por Shannon²⁸, quien considera la predictibilidad como lo opuesto a la incertidumbre y la compone por dos parámetros la constancia y la contingencia, la constancia mide el grado en que un estado permanece sin variaciones y la contingencia mide las diferencias estacionales y su constancia a lo largo del tiempo.

Para el cálculo de los índices partimos de la nube de puntos de ventanas deslizantes vista como una tabla de contingencia N, la cual se proyecta en forma de una matriz de frecuencia, con los tiempos como columnas (tiempos del ciclo) y los ventanas como filas, entonces se define N_{ij} como el número de ciclos durante los cuales la señal estuvo en el la ventana i en el tiempo j y las marginales por columna $X_j = \sum_{i=1}^{s} N_{ij}$, representa la variable aleatoria temporal, las marginales por fila $Y_i = \sum_{j=1}^{t} N_{ij}$, representan la variable aleatoria de los niveles y el gran total $Z = \sum_i \sum_j N_{ij} = \sum_j X_j = \sum_i Y_i$, el cual permite el cálculo de las frecuencias relativas representando las probabilidades para las respectivas variables aleatorias, por tanto la entropía con respecto al tiempo H(X), al estado H(Y), al tiempo y estado H(XY)

²⁸ SHANNON, C. E. "A mathematical theory of communication". En: *The Bell System Technical Journal* 27.3 (1948), págs. 379-423.

se definen como:

$$H(X) = -\sum_{j=1}^{t} \frac{X_j}{Z} \log \frac{X_j}{Z},$$
(3)

$$H(Y) = -\sum_{i=1}^{s} \frac{Y_i}{Z} \log \frac{Y_i}{Z},$$
(4)

$$H(XY) = -\sum_{j=1}^{t} \frac{N_{ij}}{Z} \log \frac{N_{ij}}{Z}.$$
(5)

La predictibilidad de un fenómeno periódico se establece como máxima cuando hay certeza con respecto al estado (fila) ²⁴, es decir $H_X(Y) = H(XY) - H(X)$, por tanto se puede definir predictibilidad *P*, constancia *C* y contingencia *M* en la Tabla 2.

Tabla 2. Índices propuestos por Colwell para medir la predictibilidad de fenómenos periódicos ²⁴

Índice	Cálculo del índice		
Constancia (C)	$C = 1 - \frac{H(Y)}{\log s}$		
Contingencia (M)	$M = \frac{H(X) + H(Y) - H(XY)}{\log s}$		
Predictibilidad (P)	$P = 1 - \frac{H_X(Y)}{\log s} = 1 - \frac{H(XY) - H(X)}{\log s}$		

Estas medidas brindan valores entre 0 y 1 puesto que el término de normalización $\log s$ lo permite. El valor 1 en predictibilidad y constancia indican las señales ser más predecibles y constantes, en el caso de la contingencia el valor 0 indica ser más persistente. Ahora, ya que la constancia y contingencia contribuyen a la predictibilidad de un fenómeno periódico, entonces la predictibilidad P se puede obtener como la suma de la constancia C y la contingencia M, es decir P = C + M. En vista de que la marcha de un paciente con EP es menos estable que la normal, se esperaría ver que los índices de predictibilidad y constancia para la marcha parkinsoniana sea menores que la control.

4.3. CONFIGURACIÓN EXPERIMENTAL

Para evaluación de las estrategias propuestas se utilizó un conjunto de datos con 22 vídeos de marcha de 22 participantes, 11 pacientes de control y 11 pacientes diagnosticados con la enfermedad de Parkinson²⁹. Cada video contiene la marcha de los sujetos en vista frontal, como se ilustra en la Figura 16. Cada vídeo contiene un total 240 frames, con 59.4 fotogramas por segundo. Respecto a la población bajo estudio, la edad media fue 72.2 ± 4.0 (media \pm desviación estandar) para los pacientes de control y 72.6 ± 7.7 para los pacientes. Los estadios de la enfermedad para los pacientes variaron entre 2,5 y 3,0 según la escala de Hoehn y Yahr ³⁰.

Figura 16. Dataset del grupo investigación Bilv2Lab³¹



²⁹ SALAZAR I., Pertuz S. Contreras W. & Martínez F. A convolutional oculomotor representation to model parkinsonian fixational patterns from magnified videos. Pattern Analysis and Application. 2020.

³⁰ Estandar clínico que categoriza a los pacientes en 4 diferentes nivele de la EP y control utilizando observaciones de experto

Una vez adquiridos los videos, se extrajeron un conjunto de trayectorias correspondientes a la dinámica del movimiento de cada paciente. Como referencia se tomó la trayectoria del centro de masa, de la silueta del cuerpo, segmentada automáticamente por una substracción del fondo (Ver Anexos 3). Las demás trayectorias se tomaron como puntos articulares del cuerpo, que componen la postura del cuerpo y son calculadas automáticamente con el software de OpenPose²². Las trayectorias son entonces la secuencia de puntos temporales que representan los centros articulares o el centro de masa, durante la secuencia del vídeo. Las trayectorias en total tienen una longitud de 240 cuadros, obteniendo así, una matriz de 25 filas (correspondientes a las trayectorias de cada punto de interés) y 240 columnas (240 fotogramas) por vídeo.

Dada la trayectoria de movimiento (serie de tiempo) asociada a la marcha de un control o paciente con EP, se definió un tamaño de ventana de 40 para el centro de masa, dado que en promedio los sujetos de estudios daban 6 pasos. Para las trayectorias de los puntos claves del esqueleto se tomaron ventanas deslizantes de 20 y 40 dependiendo de los mismos, esto debido a que comportamiento de las trayectorias no son iguales, por ejemplo, las trayectorias de la cabeza no tienen el mismo comportamiento que la del brazo. Se tomaron 13 puntos de los 25 que componen el esqueleto, puesto que la toma fue de vista frontal, estos puntos se agruparon de tal forma que representaran la cabeza, brazo, centro y pies de cada sujeto en estudio, estos grupos se conformaron como:

- La cabeza con los puntos 0,17 y 15.
- El brazo con los puntos 1, 2, 3 y 4.
- La pierna con 8,9 y 10
- El pie con los puntos 11, 22, 23 y 24.

Para ilustración y ubicación de estas referencias en el esqueleto ver la Figura 17.



Figura 17. Puntos claves del esqueleto humano³²

Seguidamente, se halló la nube de puntos de ventana deslizante establecida como un arreglo de $n \times k$, donde n es el número de ventanas y k el tamaño de ventana establecido, esto mediante la extracción de parches rectangulares del vector de entrada (serie de tiempo).

En el procedimiento de tratamiento de los datos se hizo uso de Javaplex³³, paquete de software Java de análisis de datos topológicos y homología persistente. Se establecieron las dimensiones 0 y 1 para el cálculo de la homología persistente de la nube de puntos de ventanas deslizantes. Se estableció el espacio métrico dada la nube de puntos bajo la métrica euclidiana, luego, con el espacio métrico crea el

³³ TAUSZ, Andrew; VEJDEMO-JOHANSSON, Mikael y ADAMS, Henry. "JavaPlex: A research software package for persistent (co)homology". En: *Proceedings of ICMS 2014*. Ed. por HONG, Han y YAP, Chee. Lecture Notes in Computer Science 8592. 2014, págs. 129-136.

complejo simplicial de Vietoris Rips y se halla el algoritmo de homología persistente simplicial para obtener los valores de persistencia en dimensión 0 y 1. Con dichos valores se procedió al cálculo de los índices de constancia y contingencia basados en topología establecidos en la Tabla 1.

Para calcular los índices de predictibilidad, constancia y contingencia de Colwell dada la matriz de ventanas deslizantes $N_{n\times m}$ donde n es el número de niveles establecidos (nLevels = 50, puesto que brindó mejores resultados) y m el tamaño de ventana, fila i de la matriz corresponde al histograma de la fila i de la matriz de ventanas deslizantes siguiendo la estructura establecida en 4.2. Se calcularon los marginales por columna, fila y total, X_j , Y_j y Z_j respectivamente y luego se implementaron las ecuaciones (3) para hallar la entropía con respecto al tiempo H(X), al estado H(Y), al tiempo y estado H(XY); ya establecidos estos valores se procedió a calcular los índices como se establece en la Tabla 2.

Se realizaron los cálculos de los índices para cada paciente Párkinson y para cada sujeto control. Para cuantificar la diferencia entre los dos grupos se realizó un estadístico mediante la prueba t-student para cada uno de los índices. Esta prueba de hipótesis valora las posibles diferencias estadísticas entre los promedios de los dos grupos (control y párkinson) del índice en cuestión. Se asumen dos hipótesis (nula H_0 y alternativa H_1) donde H_0 asume la igualdad de los promedios y H_1 la desigualdad. El valor t de la prueba de hipótesis se obtiene mediante la ecuación (6) en donde \bar{X}_1, \bar{X}_2 son las medias muéstrales (medias de los índices para control y párkinson); S_c^2 es la varianza común, calculada como $S_c^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$ y n_1, n_2 son la cantidad de muestras en cada población, puesto que tenemos 11 pacientes con párkinson y personas control, se tiene $n_1 = n_2 = 11$.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_c^2}{n_1} + \frac{S_c^2}{n_2}}}.$$
 (6)

Esta prueba sigue una distribución con $(n_1 + n_2 - 2)$ grados de libertad y se rechaza la hipótesis nula (H_0) cuando t es mayor al valor crítico. Se toma un margen de error del 5 % ($\alpha = 0.05$) ya que es una distribución de dos colas. El valor de t nos indica la diferencia entre los promedios de las dos poblaciones en estudio (control y párkinson) y una vez calculado, se halla el nivel de significación (p) para poder apreciar el grado de diferencias entre los controles y los párkinson para el cierto índice, el nivel para concluir que dos poblaciones tienen diferencias estadísticamente significativas entre sus muestras es cuando se rechaza H_0 indicando que las muestras son diferentes, es decir que se cumpla p < 0.05.

5. RESULTADOS Y DISCUSIÓN

Para la evaluación de la metodología propuesta se midió la capacidad del análisis topológico en la cuantificación y discriminación de patrones parkinsonianos con respecto a la población control. En particular, se calcularon trayectorias de video, se realizaron cálculos de los índices para la cuantificación de los pacientes evaluados. Luego se cuantificaron estadísticas, agrupando las dos poblaciones consideradas. Finalmente se procedió a estudiar posibles diferencias significativas en los patrones calculados para los patrones locomotores aquí descritos.

En la primera parte de este capítulo se presentarán los resultados obtenidos para el biomarcador del centro de masa y en la segunda, los resultados de la evaluación de la metodología para el conjunto de trayectorias que describen el esqueleto humano. En particular, se ilustrará la capacidad de caracterización para un punto en particular de cada grupo establecido como: cabeza, brazo, cadera y pie. Finalmente, se realizará la descripción de los resultados obtenidos para cada grupo.

5.1. CENTRO DE MASA

En primera instancia, se validaron los índices sobre la trayectoria del centro de masa, la cual en términos biomecánicos resumen la dinámica del cuerpo y puede tener mayor certeza en la discriminación de patrones anormales con respecto a los patrones control. En este sentido se calcularon los índices de Colwell por cada uno de los videos del conjunto de datos y se compararon los pacientes diagnosticados con párkinson y los controles utilizando estos descriptores. La Figura 18 muestra un gráfico de violín para cada uno de los tres índices (18.a predictibilidad, 18.b constancia y 18.c contingencia). También, para los tres índices se evidenciaron diferencias significativas con valores de $p \le 1e - 5$ para la contingencia y la constancia.

Como puede observarse en las figuras 18.a y 18.b, el grupo párkinson tiene una menor media evidenciando que la marcha parkinsoniana es menos predecible y constante respecto a la marcha control. Lo anterior resulta interesante en términos biomecánicos, resultando ser los patrones parkinsonianos valores atípicos, con menores niveles cíclicos como se determina típicamente para una locomoción bípeda. En 18.c se puede apreciar como el grupo de párkinson tiene mayor dispersión y valor de media, entonces el índice de contingencia índica que en las series asociadas a la marcha parkinsoniana, las ventanas deslizantes tienen un grado de diferencia en varias partes de la serie de tiempo pero el patrón es el mismo para el ciclo de marcha. Esto puede ser debido a que en la marcha de párkinson hay disminución en la velocidad y variabilidad entre pasos. También se puede observar en términos poblacionales distribuciones unimodales centradas en los valores medios, con ciertos valores atípicos que pueden corresponder a errores asociados al proceso de captura de las trayectorias.





En cuanto al segundo análisis sobre la trayectoria global de locomoción, el índice de la propuesta topológica de entropía con persistencia H_0 (TDA constancia) evidencia en la figura 19.a que el grupo control tiene menor entropía en las series del centro de masa y en 19.b que el grupo con EP el centro de masa asociado a la marcha es menos periódico que el grupo control. En este análisis poblacional se observan menores diferencias estadísticas entre las distribuciones asociadas a cada clase de pacientes, con medias cercanas, pero una mayor varianza para los patrones asociados al párkinson.

Figura 19. Gráfico de violín ilustrando la diferencia de distribución entre las trayectorias del centro de masa de los grupos control y párkinson para los índices de constancia y contingencias basados en análisis topológico de datos. El conjunto parkinsoniano tiene mayor entropía.



5.2. EVALUACIÓN DE PUNTOS ARTICULARES DE INTERÉS

En una segunda evaluación de la metodología propuesta, se seleccionaron puntos articulares claves en la descripción de la marcha, y su correspondiente trayectoria fue tratada como insumo para el análisis de periodicidad y topológico. Los puntos seleccionados fueron: para la cabeza $\{0, 17, 15\}$, brazo $\{1, 2, 3, 4\}$, pierna $\{8, 9, 10\}$ y

pie $\{11, 22, 23, 24\}$.

5.2.1. Trayectoria de representación para el grupo de la cabeza Para el grupo de la cabeza se tomó en particular un punto clave (número 17, ver figura 17). Dado que en la marcha parkinsoniana el paciente camina inclinando hacia adelante, con pasos cortos y rápidos, entonces la zancada (fase de apoyo) tiende a ser corta, ocasionando un movimiento de cabeza poco oscilante y rígido, por el contrario en la marcha control esta fase es más larga, por tanto la señal tiende a ser sinusoidal, como se puede observar en la tabla 3, la cual ilustra los índices y las series de tiempo asociadas al punto establecido del esqueleto para la persona con EP y una control. Los índices de Colwell indican que la señal correspondiente al control es más predecible, varia menos y es menos contante que la serie correspondiente al párkinson. Los índices topológicos evidencian que la serie del sujeto control es más periódica (TDA contingencia) y con menor entropía (TDA constancia).

Grupo: Cabeza, Keypoint: 17					
Índice	Control	párkinson	Serie de tiempo control	Serie de tiempo párkinson	
C. Predictibilidad	0.74	0.58	C01kp17	P10kp17	
C. Constancia	0.37	0.47			
C. Contingencia	0.22	0.12			
TDA Constancia	0.02	0.10	7 20 10 10 10 10 10 10 10 10 10 10 10 10 10		
TDA Contingencia	0.66	0.49			

Tabla 3. Índices de las series de tiempo correspondiente a un punto del grupo la cabeza, de una persona control y con EP.

5.2.2. Trayectoria de representación para el grupo de la brazo Se tomó como referencia del grupo de puntos del brazo el punto clave 2 el cual corresponde al hombro del sujeto de estudio. En la tabla 4 los índices señalan que el control es un

tanto más predecible y menos constante que el párkinson, los índices topológicos indican tener poca periodicidad y entropía. Por tanto, estas series de tiempo tienden a tener un comportamiento similar.

Grupo: Brazo, Keypoint: 2					
Índice	Control	párkinson	Serie de tiempo control	Serie de tiempo párkinson	
C. Predictibilidad	0.63	0.58	C01kp2	P10kp2	
C. Constancia	0.55	0.60	195.3		
C. Contingencia	0.02	0.03	300 0 507 5 566 2		
TDA Constancia	0.05	0.01	112 1 . 132 0 . 0	124- 102- 0 50 110 150 200 200	
TDA Contingencia	0.23	0.16			

Tabla 4. Índices de las series de tiempo correspondiente a un punto del grupo del brazo, de una persona control y con EP.

5.2.3. Trayectoria de representación para el grupo de la cadera Dado que el grupo de la cadera está compuesto por dos puntos clave, se tomó el punto 9 como referencia. Los índices en la tabla 5 indicaron que el comportamiento de la serie de tiempo de dicho punto asociada a la persona control tiene tiende a ser similar a la trayectoria representante de la cabeza (tabla 4) evidenciado por el argumento de la fase de marcha en pacientes con párkinson y control.

Tabla 5. Índices para la trayectoria de movimiento del punto clave representante del grupo de la cabeza.

Grupo: Cadera, Keypoint: 9					
Índice	Control	párkinson	Serie de tiempo control	Serie de tiempo párkinson	
C. Predictibilidad	0.62	0.59	C01kp9	P10kp9	
C. Constancia	0.57	0.47	205	2017 2019 2017 2019 2017	
C. Contingencia	0.02	0.15			
TDA Constancia	0.38	0.29	270- <u> <u> </u> </u>	2000 2000 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	
TDA Contingencia	0.30	0.21			

5.2.4. Trayectoria de representación para el grupo de la pie Se tomó el punto clave 11 del grupo del pie derecho, el cual corresponde al tobillo, teniendo en cuenta que en la marcha parkinsoniana una vez iniciada se suelen arrastrar los pies, dando pasos cortos, la trayectoria indica tener menos amplitud y/o periodicidad que la asociada a la marcha control. Esto se evidencia en el cálculo de los índices en la tabla 6, los índices de Colwell indican que el control es más periódico y menos constante que el párkinson, y de TDA la periodicidad de la serie del control.

Tabla 6. índices para la trayectoria de movimiento del punto clave representante del grupo del pie.

Grupo: Pie, Keypoint: 11					
Índice	Control	párkinson	Serie de tiempo control	Serie de tiempo párkinson	
C. Predictibilidad	0.60	0.49	C07kp11	Pl0kpl1	
C. Constancia	0.35	0.36			
C. Contingencia	0.14	0.24	343 - 342 - 385 -		
TDA Constancia	0.30	0.36	333 - 323 - 0 30 132 130 200 200	00- 0 50 110 150 200 200	
TDA Contingencia	0.44	0.33			

5.2.5. Resultados de la estimación por grupos del esqueleto Para la estimación de marcha usando los puntos clave del cuerpo, se calcularon los índices para cada punto clave de cada grupo propuesto, esto se llevó a cabo de igual forma que se procedió con el centro de masa, los resultados para cada grupo se ilustran en la tabla 7, en donde se evidenció que el índice de predictibilidad Colwell para los grupos de la cabeza, cadera y pie mostró que difieren los controles y los párkinson en cuan predecibles son. El índice usando TDA Contingencia (cuantificando periodicidad) para todos los grupos mostró un grado de diferencia entre el conjunto de controles y el conjunto de párkinson.

Tabla 7. Media y nivel de significación de los índices de cada grupo de puntos clave, control y párkinson. Se puede observar que las trayectorias del pie y la cabeza son siempre significativas.

Grupo	Índiaa	m	nedia	Nivel de
	indice	Control	Párkinson	significación (p)
	C. Predictibilidad	0.64	0.50	p<0.05
	C. Constancia	0.48	0.50	p<0.05
Cabeza	C. Contingencia	0.094	0.095	p > 0.05
	TDA Constancia	0.040	0.043	p > 0.05
	TDA Contingencia	0.57	0.41	p<0.05
	C. Predictibilidad	0.62	0.61	p > 0.05
	C. Constancia	0.57	0.58	p > 0.05
Brazo	C. Contingencia	0.036	0.033	p<0.05
	TDA Constancia	0.039	0.036	p > 0.05
	TDA Contingencia	0.22	0.19	p<0.05
	C. Predictibilidad	0,65	0,60	p<0.05
	C. Constancia	0.59	0.55	p<0.05
Cadera	C. Contingencia	0.02	0.09	p<0.05
	TDA Constancia	0.342	0.348	p > 0.05
	TDA Contingencia	0.29	0,20	p<0.05
Pie	C. Predictibilidad	0.59	0.54	p<0.05
	C. Constancia	0.43	0.46	p<0.05
	C. Contingencia	0.15	0.13	p<0.05
	TDA Constancia	0.32	0.36	p<0.05
	TDA Contingencia	0.44	0.23	p<0.05

También se puede evaluar de la tabla 7 a mayor detalle que para los grupo del pie y cabeza los resultados fueron óptimos, en el caso del grupo de pie se corrobora debido a que en la marcha parkinsoniana hay poca elevación de los pies del suelo, consecuente arrastre, disminución de la velocidad o incremento, entre otras; a diferencia del movimiento de los pies en el ciclo de marcha control. En cuanto al grupo de la cabeza, la marcha del párkinson adoptan una postura en anteflexión de tronco y cabeza, manteniendo cierta rigidez, por tanto la marcha es más constante y más predecible, lo cual se evidencia en los resultados. El grupo de la cadera dado que su posición es próxima al centro de masa, los resultados fueron similares; y en cuanto al grupo del brazo, la marcha parkinsoniana tiene cierta disminución del braceo y por esta razón, se deberían ver diferencias en los resultados, esto puede deberse a posiblemente los puntos presentan mucha oclusión, lo que impide un apropiado análisis dinámico, por lo tanto se recomienda utilizar métodos más apropiados de captura.

6. CONCLUSIONES Y PERSPECTIVAS

En este trabajo se presenta un enfoque metodológico novedoso para cuantificar patrones locomotores anormales con asociación a la enfermedad del Parkinson. El trabajo se fundamenta en el análisis periódico y topológico de un conjunto de trayectorias (puntos espaciales relacionados durante una secuencia) que tienen un sentido biomecánico. Por una parte, la metodología implementada con SW1PerS ayudó a localizar periodos sobre series de tiempos gracias al análisis de la persistencia. Las partes similares cercanas se agruparon en una componente conexa y los picos que estaban alejados formaron componentes conexas más persistentes, mostrando periodicidad. También, la entropía propuesta por Colwell midió la variabilidad entre períodos y compara el comportamiento de las series de tiempo periodo a periodo, por lo que, si no hay cambio entre ellos, la entropía es mínima. En este sentido se pudieron evidenciar diferencias significativas entre la población control y un conjunto de pacientes diagnosticados con párkinson, por lo cual, esta metodología puede resultar útil para definir nuevos biomarcadores que permitan predecir esta patología, pero también estudiar su avance y comportamiento ante la administración de un tratamiento en particular.

Por otra parte, en la metodología implementada mediante TDA se logró una caracterización entre la marcha de pacientes con enfermedad de Parkinson y sujetos sanos. En particular los intervalos de persistencia de series de tiempo del centro de masa evidenciaron diferencias estadísticamente significativas, mostrando que la marcha parkinsoniana es menos periódica y con mayor entropía que la de sujetos sanos. Además, los índices de Colwell mostraron que para pacientes con EP la marcha es menos predecible y permanece con poca variación. Esto debido a que los pacientes con párkinson presentan mayores fluctuaciones en el tiempo de duración del balanceo en comparación con las personas sanas. En conjunto el análisis de estos índices y la caracterización topológica del párkinson dan una nueva mirada a la caracterización de la enfermedad, lo cual, hoy en día es de especial interés, al no existir un método diagnóstico definitivo, ni conocimiento sobre las alteraciones y evolución de la enfermedad en cada paciente ¹⁶.

Para la estimación de marcha usando los puntos clave del cuerpo, el índice de predictibilidad de Colwell indicó que las trayectorias de los grupos: cabeza, cadera y pie, son menos predecibles para la marcha parkinsoniana. El índice de periodicidad TDA contingency asociado a la contingencia, describió para todos los grupos que las series de tiempo de los pacientes con EP son menos periódicas que las de control, esto es debido a que en la marcha del párkinson los pacientes inclinan la cabeza y al dar una zancada corta debido a los pasos cortos y rápidos, la cabeza tiene un movimiento poco sinusoidal. A pesar de los importantes avances obtenidos, en el trabajo se tiene una alta dependencia de la calidad de las trayectorias calculadas y las metodologías usadas para este propósito, lo cual puede alterar las medidas tomadas. En futuros trabajos se pretende entonces ahondar en alternativas robustas y reproducibles para el cálculo definitivo de trayectorias, así como también la descripción de protocolos específicos donde la metodología propuesta puede ser usada. También, trabajos futuros involucran la realización de estudios con mayores conjuntos de datos que permitan determinar estadísticamente la efectividad de los biomarcadores seleccionados desde la estadística propuesta.

Tomando como línea base los resultados encontrados en este trabajo de investigación, se plantea como trabajo futuro la evaluación de métricas como distancia cuello de botella, debido a que estas permiten realizar un análisis más robusto de los diagramas de persistencia, entre el conjunto control y párkinson, además de usar nuevos complejos simpliciales, tales como el complejo de testigo que ha sido propuesto para reducir la complejidad de los complejos de Vietoris-Rips, y así facilitar un cálculo eficiente, desde el punto de vista computacional, de la homología simplicial. Finalmente se espera evaluar el método propuesto sobre distintos conjuntos de datos que permitan determinar con mayor precisión la capacidad de caracterizar pacientes con la enfermedad de Parkinson, además de evaluar la capacidad para identificar el estadio de la misma.

BIBLIOGRAFÍA

A roadmap for the computation of persistent homology (vid. pág. 13).

AFSAR O, Tirnakli U y N, Marwan. "Recurrence Quantification Analysis at work: Quasi-periodicity based interpretation of gait force profiles for patients with Parkinson disease". En: *Scientific Reports* 8 (jun. de 2018), págs. 300 -306 (vid. págs. 14, 43).

"An entropy-based persistence barcode". En: *Pattern Recognition* 48.2 (2015), págs. 391-401 (vid. pág. 47).

BAJAJ, Nin P S, *et al.* "Accuracy of clinical diagnosis in tremulous parkinsonian patients: a blinded video study". En: *Journal of Neurology, Neurosurgery & Psychiatry* 81.11 (2010), págs. 1223-1228 (vid. pág. 12).

BLIN, O.; FERRANDEZ, A.M. y SERRATRICE, G. "Quantitative analysis of gait in Parkinson patients: increased variability of stride length". En: *Journal of the Neuro-logical Sciences* 98.1 (1990), págs. 91 -97 (vid. pág. 12).

BOBROWSKI, Omer y SKRABA, Primoz. "Homological Percolation and the Euler Characteristic". En: (oct. de 2019) (vid. pág. 19).

BOWMAN, Gregory R, *et al.* "Structural insight into RNA hairpin folding intermediates". En: () (vid. pág. 13).

CAO, Zhe, *et al.* "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields". En: *CoRR* abs/1812.08008 (2018). arXiv: 1812.08008 (vid. pág. 75).

CHAZAL, Frédéric y MICHEL, Bertrand. *An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists*. 2021. arXiv: 1710.04019 (vid. págs. 29, 31, 36).

CHEN, Shih-Wei, *et al.* "Quantification and recognition of parkinsonian gait from monocular video imaging using kernel-based principal component analysis". En: *Bio-Medical Engineering OnLine* 10.1 (2011), pág. 99. DOI: 10.1186/1475-925x-10-99 (vid. pág. 13).

COLWELL, Robert. "Predictability, Constancy, and Contingency of Periodic Phenomena". En: *Ecology* 55 (ago. de 1974), págs. 1148-1153. DOI: 10.2307/1940366 (vid. págs. 44, 49, 50).

DEWOSKIN, D., *et al.* "Applications of computational homology to the analysis of treatment response in breast cancer patients". En: *Topology and its Applications* 157.1 (2010), págs. 157 -164 (vid. pág. 13).

FORSYTH, David A. y PONCE, Jean. *Computer Vision - A Modern Approach, Second Edition.* Pitman, 2012, págs. 1-791 (vid. pág. 73).

G. HIDALGO Z. Cao, T. Simon S.-E. Wei H. Joo y SHEIKH, Y. *OpenPose library* (vid. págs. 42, 52, 53).

HAUSDORFF, Jeffrey M., *et al.* "Gait variability and basal ganglia disorders: Strideto-stride variations of gait cycle timing in parkinson's disease and Huntington's disease". En: *Movement Disorders* 13.3 (1998), págs. 428-437 (vid. pág. 15).

"Measuring signal fluctuations in gait rhythm time series of patients with Parkinson's disease using entropy parameters". En: *Biomedical Signal Processing and Control* 31 (2017), págs. 265-271 (vid. pág. 43).

MORRIS, Meg E., *et al.* "Stride length regulation in Parkinson's disease: Normalization strategies and underlying mechanisms". En: *Brain* 119.2 (abr. de 1996), págs. 551-568 (vid. pág. 15).

MORRIS, Meg E., *et al.* "The biomechanics and motor control of gait in Parkinson disease". En: *Clinical Biomechanics* 16.6 (2001), págs. 459 -470 (vid. pág. 12).

MORRIS, Meg E., *et al.* "The pathogenesis of gait hypokinesia in Parkinson's disease". En: *Brain* 117.5 (oct. de 1994), págs. 1169-1181 (vid. pág. 15).

NAWAR, Afra, et al. Topological Descriptors for Parkinson's Disease Classification and Regression Analysis. 2020. arXiv: 2004.07384 [cs.LG] (vid. pág. 13).

ORGANIZATION, World Health. *Neurological disorders : public health challenges*. World Health Organization, 2006, págs. 140-142 (vid. pág. 12).

PEREA J.A., Deckard A. Haase S.B. et al. "SW1PerS: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data". En: *BMC Bio-informatics* 13.1 (nov. de 2015), pág. 257 (vid. págs. 45, 46, 48).

PEREA, Jose y HARER, John. *Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis.* 2013. arXiv: 1307.6188 [math.AT] (vid. pág. 47).

RAMÁREZ, Juan; MORENO, Franklin y MEDINA, Rubén. "Análisis de la serie temporal del balanceo en la marcha de pacientes con parkinson". es. En: *Universidad, Ciencia y Tecnología* 10 (sep. de 2006), págs. 205 -210 (vid. pág. 13).

REVELES, Fermín, *et al. Notas en Persistencia, Probabilidad e Inferencia Estadística para Análisis Topológico de Datos.* 2016 (vid. pág. 18). SALAZAR I., Pertuz S. Contreras W. & Martínez F. *A convolutional oculomotor representation to model parkinsonian fixational patterns from magnified videos*. Pattern Analysis and Application. 2020 (vid. pág. 51).

SCHAAFSMA, Joanna, *et al.* "Gait dynamics in Parkinson's disease: Relationship to Parkinsonian features, falls and response to levodopa". En: *Journal of the neurological sciences* 212 (ago. de 2003), págs. 47-53 (vid. pág. 15).

SHANNON, C. E. "A mathematical theory of communication". En: *The Bell System Technical Journal* 27.3 (1948), págs. 379-423 (vid. pág. 49).

TAPIA-NÚÑEZ, John y CHANÁ, Pedro. "Diagnosis of Parkinson's disease". En: *Revista de neurología* 38 (ene. de 2004), págs. 61-7 (vid. págs. 16, 65, 72).

TAUSZ, Andrew; VEJDEMO-JOHANSSON, Mikael y ADAMS, Henry. "JavaPlex: A research software package for persistent (co)homology". En: *Proceedings of ICMS 2014*. Ed. por Han HONG y Chee YAP. Lecture Notes in Computer Science 8592. 2014, págs. 129-136 (vid. pág. 53).

TUNCA, Can, *et al.* "Inertial Sensor-Based Robust Gait Analysis in Non-Hospital Settings for Neurological Disorders". En: *Sensors* 17.4 (2017). DOI: 10.3390/s17040825 (vid. pág. 72).

WHITTLE, M.W. *Gait Analysis: An Introduction; Butterworth-Heinemann:* Oxford, UK, 2014. (Vid. pág. 71).

ANEXOS

Anexo A. Análisis de marcha y el párkinson

El análisis de marcha pretende evaluar particularidades de la locomoción desde un punto de vista cinemático y cinético para determinar y cuantificar anomalías asociadas con ciertas patologías. Este análisis se apoya en diferentes dispositivos y metodologías que tipifican patrones anormales, siendo indicadores cruciales del estado y la progresión de los trastornos que tienen síntomas motores. Esta herramienta es entonces esencial para soportar diagnósticos, evaluar la progresión de patologías y determinar la efectividad de ciertos tratamientos. En este análisis un componente crucial es el ciclo de la marcha, el cual brinda una descripción de las características de la marcha humana.

Los términos de los eventos para la marcha normal del pie derecho se presentan como círculos de acercamiento en la Figura 20. En condiciones normales, las fases de contacto inicial y de apoyo se denominan alternativamente como golpe de talón y despegue, respectivamente; el contacto inicial marca el comienzo de la transferencia de carga al pie en contacto con el suelo. El despegue marca el levantamiento completo de la carga finalizando la fase de apoyo. La fase donde ocurre el movimiento real del pie es la de balanceo y ésta comienza con el despliegue del pie y termina en el contacto inicial. El 60% del ciclo de la marcha se constituye por la fase de apoyo y el 40% restante por la de balanceo. Un solo ciclo de marcha se define como el período entre dos eventos sucesivos de marcha repetitiva³⁴, por tanto cada

³⁴ WHITTLE, M.W. *Gait Analysis: An Introduction; Butterworth-Heinemann:* Oxford, UK, 2014.

ciclo define un periodo en la marcha.

Este ciclo de movimiento puede variar drásticamente en condiciones atípicas y frente a ciertas patologías, revelando patrones que pueden conducir a un diagnóstico definitivo. Por ejemplo, en la marcha parkinsoniana el primer punto de contacto del pie puede no ser el talón ya que en esta marcha patológica se arrastran los pies, con pasos lentos y pequeños, es decir, el pie golpea el suelo de una vez³⁵. En el caso de la marcha parkinsoniana, la parte superior tiende a realizar un menor balanceo y el desplazamiento vertical del tronco se reduce, y en la parte inferior aparece variabilidad temporal y espacial en la regularidad y la estabilidad de los parámetros de la marcha, con acortamiento de la longitud del paso, poca elevación de los pies del suelo con un consecuente arrastre, disminución de la velocidad o incremento a expensas de aumentar la cadencia más que la longitud del mismo ¹⁶. Por tanto, es relevante estudiar los ciclos de la marcha parkinsoniana y normal.

Figura 20. Ciclo de la marcha humana³⁶.



³⁵ TUNCA, Can, *et al.* "Inertial Sensor-Based Robust Gait Analysis in Non-Hospital Settings for Neurological Disorders". En: *Sensors* 17.4 (2017). DOI: 10.3390/s17040825.
Anexo B. Análisis de video

El análisis de video es un mecanismo que permite examinar y analizar imágenes captadas en video, su principal objetivo es la descripción y clasificación de alto nivel de secuencias de videos a partir de patrones de movimiento y de apariencia³⁷. El análisis de imágenes y la visión por computador juegan un rol fundamental en este mecanismo aplicado en áreas como la vigilancia automática, biomedicina, deportes entre otras.

Usando técnicas de procesamiento de video se puede estudiar la secuencia de cuadros que componen un video de marcha de un paciente, con lo cual se obtiene una representación inicial de la estructura corporal en cada uno de los videos. El patrón de marcha se puede ver afectado en la EP, debido a que se afectan directamente las funciones motrices, por lo tanto el estudio de la marcha parkinsoniana es importante ya que el análisis de la marcha, siendo una herramienta en la evaluación del párkinson, basándose en el seguimiento de algunas variables dinámicas permiten de manera simultánea cuantificar el grado de la enfermedad y comparar la marcha con patrones normales.

En otro orden de ideas, una secuencia de video es una colección de imágenes $\{x_i\}_{i\in\Delta}$ tal que $\Delta = \{1, \dots, L\}$ y $x_i \in \mathbb{R}^{W \times H}$ donde L es el número de imágenes y cada una es codificada en arreglos de dimensión, correspondiente a su ancho W y alto H, ver figura 21. El proceso de análisis de marcha estima las relaciones funcionales de la estructura corporal durante el movimiento. Una estimación correcta de la marcha se obtiene calculando la posición del centro de masa y puntos claves

³⁷ FORSYTH, David A. y PONCE, Jean. *Computer Vision - A Modern Approach, Second Edition.* Pitman, 2012, págs. 1-791.

del cuerpo humano.

Figura 21. Ilustración de secuencia de video.



Anexo C. Trayectorias de movimiento

Las siguientes son dos estimaciones de la marcha extraídos de secuencias de video.

Centro de masa En el análisis clínico de la marcha humana es útil evaluar el rendimiento biomecánico de todo el cuerpo, por tanto, un biomarcador primario de la marcha es el centro de trayectoria de masa (CoM) el cual expresa los puntos de características de posturas musculoesqueléticas. Cada punto en una cuadrícula espaciada por *W* píxeles representa el centro de las masas de los segmentos de miembros inferiores, pelvis, tronco, cabeza y brazos. El centro de masa se estima para cada cuadro de acuerdo con la ecuación (7), formando la trayectoria del centro de masa ($P_0, P_1, \ldots, P_{W-1}$) mediante una segmentación corporal automática utilizando una mezcla de fondo de sustracción gaussiana, ver figura 22.

$$(x,y)_{cm} = \frac{\sum_{i} P_i(x_i, y_i)}{\sum_{i} P_i}$$

$$\tag{7}$$



Figura 22. Representación del centro de masa de marcha.

OpenPose Otra estimación de la marcha se obtiene calculando puntos claves del cuerpo humano. CAO, Zhe,et al³⁸ proponen una estimación 2D del esqueleto

 ³⁸ CAO, Zhe, *et al.* "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields".
 En: *CoRR* abs/1812.08008 (2018). arXiv: 1812.08008.

humano fijando 25 puntos o "partes" anatómicas, estos puntos se estiman para cada cuadro formando así 25 trayectorias describiendo la marcha del paciente en la secuencia de video. Ese trabajo se documenta bajo la biblioteca de código abierto OpenPose**OpenPose_library1**, es una biblioteca de detección de puntos clave de múltiples personas en tiempo real para la estimación del cuerpo, la cara, las manos y los pies en imágenes individuales, vea la figura 23.



Figura 23. A izquierda representación de la marcha y a derecha formato de salida de pose.