Análisis topológico de datos para el refinamiento de datasets

Carlos Lafuente Carreras

Trabajo de fin de grado de Matemáticas

Universidad de Zaragoza

Director del trabajo: Miguel Ángel Marco Buzunáriz

11 de julio de 2025

Abstract

The purpose of this paper is to explore the application of advanced concepts from algebraic topology to the practical problem of dataset refinement. In particular, we focus our study on point clouds, as they are the most direct data format for applying the tools of simplicial homology. Our goal is to develop a method capable of distinguishing the essential structure from noise in a point cloud, removing outliers or redundant elements without altering its fundamental topological features. To this end, the work is structured into three chapters that guide the reader from the theoretical foundations to the practical application and experimental validation.

The first chapter rigorously builds the language of simplicial homology, a tool that translates topological problems into a more manageable algebraic framework. We begin by introducing *simplicial complexes*, structures composed of *simplices* capable of discretizing triangulable topological spaces. We define *simplicial maps*, which relate simplicial complexes to one another, and present the *Simplicial Approximation Theorem*, which enables the approximation of continuous maps through simplicial ones, essential for faithfully translating problems from continuous spaces to the simplicial domain.

To operate algebraically with these geometric objects, we equip simplices with an *orientation*. This allows us to define *q-chain groups*, which are free abelian groups whose generators are the oriented q-simplices themselves. A q-chain is, in essence, a formal sum of simplices, the fundamental object upon which we define the *boundary operator* (∂). This operator assigns to each q-chain its boundary, which is a (q-1)-chain. The main property of this construction is that the boundary of a boundary is zero ($\partial\partial = 0$), which allows us to define the *homology groups* as the quotient of cycles (boundaryless chains) and boundaries (chains that are the boundary of something). These groups provide algebraic invariants that count the topological features of a space, such as its connected components, holes, and voids. We provide an interpretation of the elements of these groups and an example of their calculation on the Möbius Strip. The chapter concludes by proving the most important result of the theory: homotopy invariance, which demonstrates that homology groups are identical for all spaces that can be continuously deformed into one another. This ensures that they are a robust topological invariant, laying the groundwork for their reliable application in data analysis.

The second chapter bridges the gap between classical theory and Topological Data Analysis (TDA). We address the problem of constructing a simplicial complex from data by creating a *filtration*, which is an increasing sequence of complexes indexed by a scale parameter ε. To do this, we introduce *Vietoris-Rips complexes*, a computationally efficient construction that generates a

complex for each value of ε. To analyze this dynamic structure, we introduce *persistent homology*, a fundamental tool in TDA that allows us to study how homology groups change throughout the filtration. This enables us to distinguish robust topological features (those that "persist" over a long range of the filtration) from ephemeral ones (considered noise). We introduce *persistence diagrams* and *barcodes*, as visual tools to represent this information, and *bottleneck distance*, a metric for comparing persistence diagrams. The chapter closes with the *Stability Theorem*, a crucial result that establishes a formal connection between the geometry of the data and its topological signature. This rigorous connection is what confidently allows us to use the bottleneck distance as a significant measure of a point's impact, thereby legitimizing its use as the central criterion in our filtering algorithm.

Finally, the third chapter presents this work's original contribution: a topological-impact filtering algorithm. It details an iterative method that uses the bottleneck distance to quantify the distortion caused by removing a point from the persistence diagram. Evaluations on a noisy synthetic torus with outliers compare the effectiveness of filtering based on H_0, H_1, H_2. Results conclusively demonstrate that filtering via the first homology group (H_1) strikes the best balance, effectively eliminating outliers and irrelevant points while faithfully preserving the torus's global structure.

Resumen

Este Trabajo de Fin de Grado explora la aplicación de conceptos avanzados de topología algebraica al problema práctico del refinamiento de conjuntos de datos, en particular, centramos nuestro estudio en las nubes de puntos por ser el formato de datos más directo para la aplicación de las herramientas de la homología simplicial. El objetivo es construir un método capaz de discernir entre la estructura esencial y el ruido en una nube de puntos, eliminando elementos atípicos o redundantes sin alterar sus características topológicas fundamentales. Para ello, el trabajo se estructura en tres capítulos que guían al lector desde los fundamentos teóricos hasta la aplicación práctica y la validación experimental.

El primer capítulo se dedica a construir rigurosamente el lenguaje de la homología simplicial, una herramienta que traduce problemas topológicos a un marco algebraico más manejable. Empezaremos introduciendo los *complejos simpliciales*, estructuras formadas por *símplices* capaces de discretizar espacios topológicos triangulables. Definiremos las *aplicaciones simpliciales*, que nos permitirán relacionar complejos simpliciales entre sí. Presentaremos un resultado conocido como *Teorema de Aproximación Simplicial*, el cual nos permitirá aproximar aplicaciones continuas a través de aplicaciones simpliciales, esencial para trasladar problemas de espacios continuos al ámbito simplicial de forma fiel.

Para poder operar algebraicamente con estos objetos geométricos, dotaremos a los símplices de una *orientación*. Esto nos permite definir los *grupos de q-cadenas*, que son grupos abelianos libres donde los generadores son los propios símplices orientados de dimensión q. Una q-cadena es, en esencia, una suma formal de símplices, el objeto fundamental sobre el que definiremos el *operador de frontera* (∂). Este operador asigna a cada q-cadena su borde, que es una (q-1)-cadena. El resultado central de esta construcción es la propiedad de que el borde de un borde es nulo ($\partial\partial = 0$), lo que nos permite definir los *grupos de homología* como el cociente entre ciclos (cadenas sin borde) y fronteras (cadenas que son el borde de algo). Estos grupos nos proporcionan invariantes algebraicos que cuentan las características topológicas de un espacio, como sus componentes conexas, agujeros y cavidades. Daremos una interpretación de los elementos de dichos grupos y un ejemplo del cálculo de estos sobre la Banda de Möbius. El capítulo concluye demostrando el resultado más importante de la teoría: la invarianza homotópica, que demuestra que los grupos de homología son idénticos para todos los espacios que pueden deformarse continuamente uno en otro. Asegurando así que estos son un invariante topológico robusto y sentando las bases para su aplicación fiable en el análisis de datos.

El segundo capítulo da el salto desde la teoría clásica al Análisis Topológico de Datos (TDA).

Se aborda el problema de construir un complejo simplicial a partir de los datos mediante la creación de una *filtración*, una sucesión creciente de complejos indexada por un parámetro de escala ε. Para ello, se presentan los *complejos de Vietoris-Rips*, una construcción computacionalmente eficiente que genera un complejo para cada valor de ε. Para analizar esta estructura dinámica, introduciremos la *homología persistente*, una herramienta fundamental del TDA que nos permite estudiar cómo cambian los grupos de homología a lo largo de toda la filtración. Esto permite distinguir las características topológicas robustas (las que "persisten" durante gran cantidad de pasos de la filtración) de las efímeras (consideradas ruido). Introduciremos los *diagramas de persistencia* y los *códigos de barras* como herramientas visuales para representar esta información. Presentaremos la *distancia bottleneck*, una métrica para comparar diagramas de persistencia. Cerraremos el capítulo con el *Teorema de Estabilidad*, un resultado crucial que establece una conexión formal entre la geometría de los datos y la firma topológica de estos. Esta conexión rigurosa es la que nos permite, con confianza, usar la distancia bottleneck como una medida significativa del impacto de un punto y, por tanto, legitima su uso como el criterio central de nuestro algoritmo de filtrado.

Finalmente, el tercer capítulo presenta la contribución original de este trabajo: un algoritmo de filtrado por impacto topológico. Se detalla el diseño de un método iterativo que utiliza la distancia bottleneck para cuantificar la distorsión que la eliminación de un punto causa en el diagrama de persistencia. Se realizan experimentos sobre un toroide sintético con ruido y puntos atípicos (u outliers, del inglés), comparando la eficacia del filtrado basado en H_0, H_1 y H_2. Los resultados demuestran de manera concluyente que el filtrado mediante el primer grupo de homología (H_1) ofrece el mejor equilibrio, eliminando eficazmente los outliers y puntos irrelevantes mientras preserva con gran fidelidad la estructura global del toroide.

Índice general

Capítulo 1

Homología Simplicial

1.1. Complejos Simpliciales

Empezaremos construyendo nuestros poliedros. Notar que estamos trabajando con la Topología Usual en el Espacio Euclídeo $\mathbb{R}^n$.

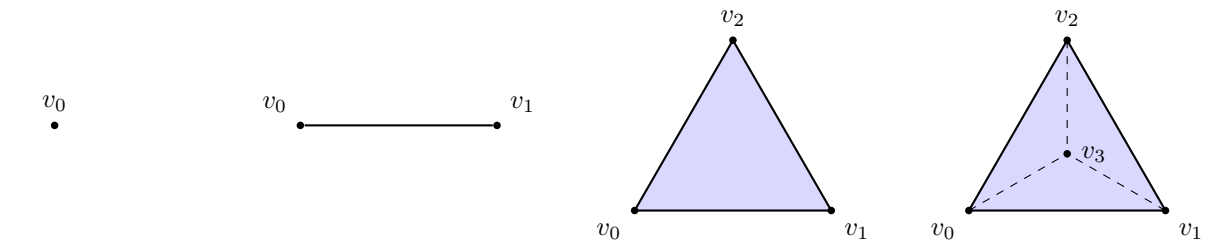
Definición 1.1.1. Se dice que un conjunto ordenado de puntos $\{x_0, x_1, \ldots, x_k\} \subset \mathbb{R}^n$ es *afínmente independiente* si $\{x_1 - x_0, x_2 - x_0, \ldots, x_k - x_0\}$ es un subconjunto de $\mathbb{R}^n$ linealmente independiente.
El conjunto $\{x_0, x_1, \ldots, x_k\}$ está en *posición general* si cada subconjunto de $n+1$ puntos forma un conjunto afínmente independiente.

Definición 1.1.2. Dado $\{x_0, \ldots, x_k\} \in \mathbb{R}^n$ conjunto de puntos afínmente independientes. Denominamos *k-símplex*, o k-símplice σ, al conjunto convexo más pequeño que los contiene a todos de forma que:

$$\sigma = \left\{ \sum_{i=0}^{k} \lambda_i x_i \;\middle|\; \sum_{i=0}^{k} \lambda_i = 1, \lambda_i \geq 0 \right\}$$

Los puntos $x_0, \ldots, x_k$ serán los *vértices* del k-símplex, denominamos al conjunto de vértices como $Vert(\sigma) = \{x_0, \ldots, x_k\}$. Para un x que cumpla $x = \sum_{i=0}^{k} \lambda_i x_i$ llamamos *coordenadas baricéntricas* de x a $\lambda_0, \ldots, \lambda_k$.

Figura 1.1: Símplices Básicos



0-símplex	1-símplex	2-símplex	3-símplex
$\{v_0\}$	$[v_0, v_1]$	$[v_0, v_1, v_2]$	$[v_0, v_1, v_2, v_3]$

Definición 1.1.3. Sean σ, τ dos símplices, decimos que τ es cara de σ si $\mathrm{Vert}(\tau) \subset \mathrm{Vert}(\sigma)$, se escribe $\tau \leq \sigma$. Si $\tau < \sigma$ (es decir, $\mathrm{Vert}(\tau) \subsetneq \mathrm{Vert}(\sigma)$), entonces decimos que τ es una *cara propia* de σ.

Definición 1.1.4. Dos k-símplices σ^m y σ^n están *correctamente unidos* si $\sigma^m \cap \sigma^n = \emptyset$ ó $\sigma^m \cap \sigma^n \leq \sigma^m \ _\wedge \ \sigma^m \cap \sigma^n \leq \sigma^n$.

Definición 1.1.5. Un *complejo simplicial*, o complejo, es una familia finita K tal que:

(i) Si $\sigma \in K$, entonces cada cara de σ también pertenece a K.

(ii) Si $s, t \in K$, entonces s y t están correctamente unidos.

La *dimensión* de K será el mayor número entero r tal que K contiene a un r-símplex.

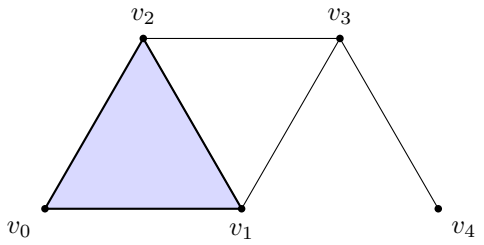


Figura 1.2: Complejo Simplicial
Este contiene cinco 0-símplices, seis 1-símplices y un 2-símplex.

Definición 1.1.6. La unión de los elementos de K, dotado de la Topología Euclídea del Subespacio, se denota por $|K|$ y se llama *poliedro asociado* de K.

Recordamos que dado un e.t. (X, τ) un subconjunto S de X, la Topología del Subespacio se define por $\tau_s = \{\ S \cap U \mid U \in \tau\}$.

Definición 1.1.7. Sea X un espacio topológico. Si existe un complejo simplicial K cuyo poliedro asociado $|K|$ es homeomorfo a X, entonces X se dice que es un *espacio triangulable*, y se denomina *triangulación* de X al complejo K.

Definición 1.1.8. La *clausura* de un k-símplex σ, $Cl(\sigma)$, es el complejo simplicial compuesto por σ y todas sus caras.

Definición 1.1.9. Sea σ un k-símplex con vértices $\{x_0, \ldots, x_k\}$.
Decimos que dos ordenaciones de sus vértices $(x_{i_0}, \ldots, x_{i_k})$, $(x_{j_0}, \ldots, x_{j_k})$ son equivalentes, es decir, que existe una relación de equivalencia entre ambas, si existe una permutación par $\rho \in S_{k+1}$ tal que $(x_{i_0}, \ldots, x_{i_k}) = (x_{j_{\rho(0)}}, \ldots, x_{j_{\rho(k)}})$.
Una *orientación* de σ es una clase de equivalencia de ordenaciones $(x_{i_0}, x_{i_1}, \ldots, x_{i_k})$ de sus vértices.
La clase de equivalencia de permutaciones pares de dicho ordenamiento determina la *orientación positiva*, mientras que la clase de equivalencia de permutaciones impares determina la *orientación negativa*.

Definición 1.1.10. Un *símplex orientado* es el par $\big(\sigma,\, [x_0, x_1, \ldots, x_k]\big)$ donde $[x_0, \ldots, x_k]$ denota la clase de equivalencia de la ordenación $(x_0, \ldots, x_k)$.
El cambio de orientación viene dado por permutaciones impares:

$$[\, x_{\tau(0)}, \ldots, x_{\tau(k)}] = \operatorname{sgn}(\tau)\, [\, x_0, \ldots, x_k\,], \quad \operatorname{sgn}(\tau) \in \{\pm 1\}.$$

1.2. Aproximaciones Simpliciales

Las aplicaciones simpliciales nos permitirán relacionar complejos de manera estructurada. Para profundizar en cómo estas aplicaciones preservan la estructura local, introduciremos las estrellas de un vértice. Esta herramienta será esencial para definir las aproximaciones simpliciales.

Definición 1.2.1. Sean K y L complejos simpliciales. Una *aplicación simplicial* $\varphi : K \to L$ es una función $\varphi : \operatorname{Vert}(K) \to \operatorname{Vert}(L)$ tal que, siempre que $\{x_0, \ldots, x_k\}$ genere un símplex de K, entonces $\{\varphi(x_0), \ldots, \varphi(x_k)\}$ genere un símplex de L.

Definición 1.2.2. Dado un vértice v en un complejo simplicial K, la *estrella* de x_i, denotada como $\operatorname{St}(x_i, K)$, es la unión de todos los interiores de los símplices de K que tienen a x_i como vértice.

Estos conjuntos abiertos actuarán como *entornos simpliciales*, nos serán útiles a continuación para controlar de forma local la imagen de una función continua al construir las aproximaciones simpliciales.

Definición 1.2.3. Sea K y L complejos simpliciales, y sea $h : |K| \to |L|$ una función continua. Una aplicación simplicial $f : K \to L$ es una *aproximación simplicial* a h si satisface la condición:

$$h(\operatorname{St}(x_i)) \subseteq \operatorname{St}(f(x_i))$$

para cada vértice x_i de K.

Las aproximaciones simpliciales son fundamentales en nuestro marco teórico, ya que trasladan el estudio de objetos topológicos continuos a un contexto discreto de complejos simpliciales. Permitiéndonos trabajar con herramientas algebraicas y combinatorias.

Definición 1.2.4. Una *subdivisión* de un complejo simplicial K es un complejo simplicial K' tal que:

(i) Cada símplex de K' está contenido en algún símplex de K.

(ii) Cada símplex de K es equivalente a la unión de finita de símplices de K'.

Las subdivisiones nos permiten refinar complejos simpliciales sin alterar su topología. Un caso particularmente importante es la subdivisión baricéntrica, la cual presentaremos más adelante. Esta técnica será clave en el Teorema de Aproximación Simplicial, donde necesitaremos ajustar los complejos para garantizar la existencia de aproximaciones.

Lema 1.2.5. Sea K un complejo simplicial:

(i) Si K' es una subdivisión de K, entonces $|K|$ y $|K'|$ son iguales como espacios topológicos.

(ii) si K' es una subdivisión de K y K'' es una subdivisión de K', entonces K'' es también una subdivisión de K.

Demostración:
(i) Obviamente, la unión de los símplices de K' coincide con la unión de los símplices de K. Se sigue que $|K|$ y $|K'|$ son iguales como conjuntos.
Sea A cerrado en K y $\tau \in K'$. Entonces $\tau \in \sigma$ para algún $\sigma \in K$, por lo tanto:

$$A \cap \tau = (A \cap \sigma) \cap \tau.$$

Pero $(A \cap \sigma)$ es cerrado en σ, por definición, por lo que $A \cap \tau$ es cerrado en τ, dado que la topología en τ es la topología de subespacio de σ. Por lo tanto, A es cerrado en K'. De manera inversa, sea A cerrado en K' y $\sigma \in K$. Entonces $\sigma = \tau_1 \cup \cdots \cup \tau_n$ para algunos $\tau_1, \ldots, \tau_n \in K'$, y por lo tanto:

$$A \cap \sigma = A \cap (\tau_1 \cup \cdots \cup \tau_n) = (A \cap \tau_1) \cup \cdots \cup (A \cap \tau_n)$$

Por definición, el conjunto $A \cap \tau_i$ es cerrado en τ_i. Sea m_i la dimensión de τ_i, entonces $A \cap \tau_i$ es cerrado en $\mathbb{R}^{m_i+1}$, porque τ_i es cerrado en $\mathbb{R}^{m_i+1}$. Si m es la dimensión de σ, entonces $m \geq m_i$, y por lo tanto $A \cap \tau_i$ está cerrado en $\mathbb{R}^{m+1}$. Pero σ lleva la topología de subespacio de $\mathbb{R}^{m+1}$, por lo que $A \cap \tau_i$ es cerrado en σ. Por lo tanto, como se ve en la igualdad anterior, $A \cap \sigma$ es la unión finita de conjuntos cerrados y, por lo tanto es cerrado.

(ii) Consecuencia inmediata de la Definición 1.2.4 □

De esta forma, podemos definir inductivamente la subdivisión n-ésima de K, garantizando que $|K| = |K^n| \quad \forall n \in \mathbb{N}$

Definición 1.2.6. Sea $\sigma = (x_0 \ldots x_k)$ k-símplex, definimos su *baricentro* como $\dot{\sigma} = \frac{1}{k+1} \sum_{i=0}^{k} x_i$
La *subdivisión baricéntrica* de un complejo simplicial K es una subdivisión $sd(K)$ de K donde cada símplex original se subdivide utilizando los baricentros de sus caras.
La *n-ésima subdivisión baricéntrica* de un complejo simplicial K, $sd^n(K)$. Se obtiene aplicando la subdivisión baricéntrica a $sd^{n-1}(K)$ para $n > 1$.

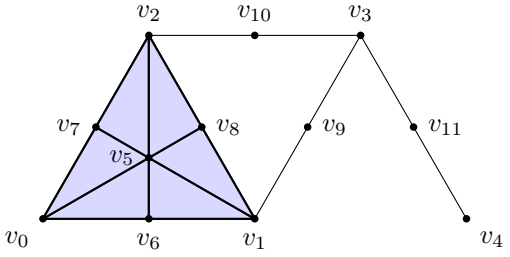


Figura 1.3: Subdivisión Baricéntrica de la Figura 1.2

Teorema 1.2.7 (Teorema de Aproximación Simplicial)**.**
Si K y L son complejos simpliciales y $f : |K| \to |L|$ es una función continua. Entonces, existe una aproximación simplicial $\phi : Sd^q(K) \to L$ de f para un entero $q \geq 1$.

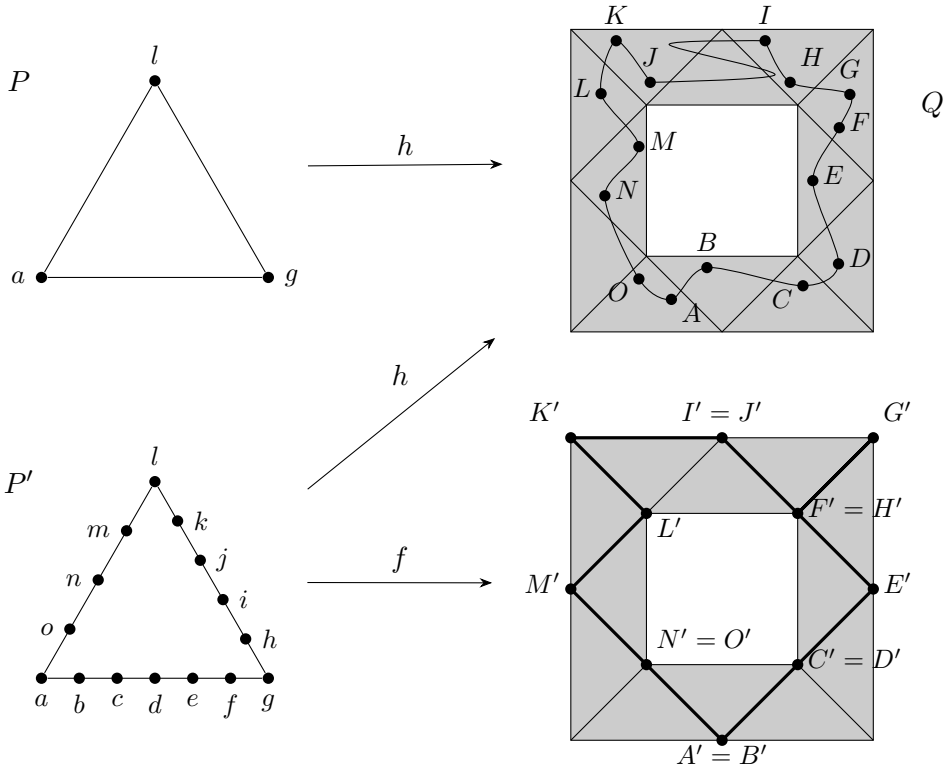


Figura 1.4: Ejemplo de Aproximación simplicial

El Teorema de Aproximación Simplicial garantiza que cualquier función continua entre los poliedros asociados de dos complejos simpliciales puede ser aproximada por una aplicación simplicial, siempre que se subdivida suficientemente el complejo original. Estableciendo así un puente riguroso entre la topología geométrica y el álgebra homológica.

1.3. Grupos de Homología

Los grupos de homología son una de las herramientas fundamentales de la topología algebraica, los cuales nos permiten estudiar y clasificar espacios topológicos mediante estructuras algebraicas. Estos grupos capturan información esencial sobre la forma y la conectividad de un espacio, identificando características como componentes conexas, agujeros y cavidades de diferentes dimensiones.

Definición 1.3.1. El *grupo de cadenas* $C_k(K)$ es el cociente del grupo abeliano libre sobre el conjunto de los k-símplices orientados de K, dado por $[\sigma] = -[\tau]$ siendo $\sigma = \tau$ orientados de distinta forma.
Un elemento $c \in C_k$ es una *cadena k-dimensional*, $c = \sum_i n_i[\sigma_i]$, donde $\sigma_i \in K$ con coeficientes $n_i \in \mathbb{Z}$.

Definición 1.3.2. El *operador de frontera* $\partial_k : C_k \to C_{k-1}$ es un homomorfismo definido linealmente sobre una cadena c por su acción sobre cualquier símplex $\sigma = [x_0, \ldots, x_k] \in c$:

$$\partial_k(\sigma) = \sum_i (-1)^i [x_0, x_1, \ldots, \hat{x_i}, \ldots, x_k]$$

donde $\hat{x_i}$ indica que x_i ha sido eliminado de la secuencia.

Teorema 1.3.3. La composición $C_{q+1}(K) \xrightarrow{\partial} C_q(K) \xrightarrow{\partial} C_{q-1}(K)$ es el homomorfismo nulo.

Demostración: Debemos probar que $\partial \circ \partial\,(c_{n+1}) = 0$ para cada $n+1$-cadena. Para ello, basta con demostrar que $\partial \circ \partial\,(\lambda\sigma^{n+1}) = 0$ para cada $(n+1)$-cadena elemental $\lambda\sigma^{n+1}$. Observamos que:

$$\partial \circ \partial\,(\sigma^{n+1}) = \partial\left(\partial\,(\sigma^{n+1})\right) = \partial\left(\sum_i (-1)^i [x_0, \ldots, \hat{x_i}, \ldots, x_{n+1}]\right) =$$

$$= \sum_{j \le i} (-1)^{i+j} [x_0, \ldots, \hat{x_j}, \ldots, \hat{x_i}, \ldots, x_{n+1}] + \sum_{j > i} (-1)^{i+j} [x_0, \ldots, \hat{x_i}, \ldots, \hat{x_j}, \ldots, x_{n+1}] =$$

$$= \sum_{i=0}^{q+1} (-1)^i \Big(\sum_{j=0}^{i} (-1)^j [x_0, \ldots, \hat{x_j}, \ldots, \hat{x_i}, \ldots, x_{n+1}] + \sum_{j=i+1}^{q+1} (-1)^{j-1} [x_0, \ldots, \hat{x_i}, \ldots, \hat{x_j}, \ldots, x_{n+1}] \Big) =$$

$$\overset{*}{=} \sum_{i=0}^{q+1} (-1)^i \cdot 0 = 0 \quad \Longrightarrow \quad \partial \circ \partial\,(c_{n+1}) = 0$$

(*): Vemos que los términos de la suma se cancelan por pares con signos opuestos. □

Definición 1.3.4. Si K es un complejo simplicial orientado, entonces:

- $Z_q(K) = \ker \partial_q$ es el *grupo de los q-ciclos simpliciales*
- $B_q(K) = \operatorname{Im} \partial_{q+1}$ es el *grupo de las q-fronteras simpliciales*

Teorema 1.3.5. Si K es un complejo orientado, entonces $B_q(K) \subset Z_q(K)$ para cada entero q tal que $0 \le q \le n$, donde n es la dimensión de K.

Demostración: Consecuencia directa del Teorema 1.3.3

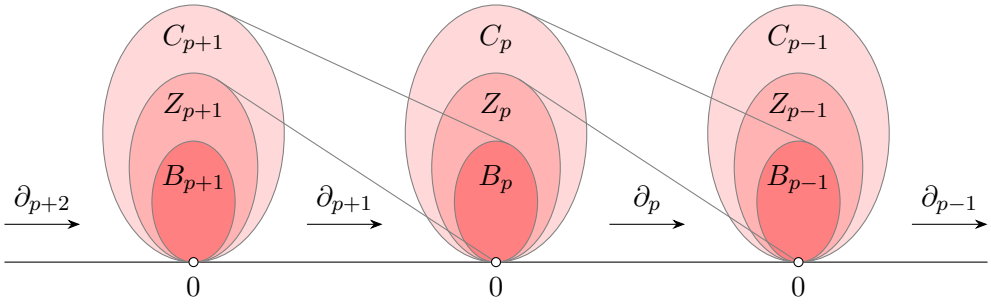


Figura 1.5

Definición 1.3.6. Dos q-ciclos w_q y z_q en un complejo K son *homólogos*, escrito $w_q \sim z_q$, siempre que exista una cadena $(q+1)$-dimensional c_{q+1} tal que

$$\partial(c_{q+1}) = w_q - z_q.$$

Esta relación de homología para los q-ciclos es una relación de equivalencia y particiona $Z_q(K)$ en *clases de homología*:

$$[z_q] = \{w_q \in Z_q(K) : w_q \sim z_q\}.$$

La clase de homología $[z_q]$ es en realidad la *clase lateral*:

$$z_q + B_q(K) = \{z_q + \partial(c_{q+1}) : \partial(c_{q+1}) \in B_q(K)\}.$$

Definición 1.3.7. Sea K es un complejo simplicial orientado,

$$H_q(K) = Z_q(K)/B_q(K) \quad \text{es el } \textit{grupo de homología} \text{ simplicial en dimensión } q$$

Por tanto, cada elemento de $H_q(K)$ está determinado por la clase de homología de un q-ciclo.

Los grupos de homología nos dan información sobre la estructura del complejo simplicial. Para dar un sentido más tangible a esta definición, vamos a ver una forma de interpretar los elementos de los grupos de homología de dimensión 0, 1 y 2:

- El grupo de homología de dimensión 0, $H_0(K) = Z_0(K)/B_0(K)$, tiene una interpretación geométrica intuitiva: Como todo 0-símplex tiene borde nulo, el cociente "colapsa" todos los 0-símplex (vértices) que estén conectados a través de cadenas de 1-símplex (aristas). Podemos ver fácilmente que H_0 será una suma directa de generadores libres, cada uno generado por cada componente conexa de nuestro conjunto. De forma que $H_0 \cong \mathbb{Z}^k$ siendo k el número de componentes conexas.

- En el caso de $H_1(K) = Z_1(K)/B_1(K)$ siguiendo el razonamiento anterior, los elementos de H_1 son clases de equivalencia de ciclos de cadenas de 1-símplices (aristas), tales que dichos ciclos no puedan expresarse como frontera de cadenas de 2-símplices (triángulos rellenos). Es decir, detecta agujeros 1-dimensionales, como por ejemplo el del interior de la 1-esfera (anillo) $\mathbb{S}^1$.

- Finalmente, los elementos de $H_2(K) = Z_2(K)/B_2(K)$ serán las clases de equivalencia de ciclos de cadenas de 2-símplices, que no puedan expresarse como frontera de cadenas de 3-símplices (tetraedros). Este último detecta cavidades 2-dimensionales, como podría ser la cavidad del interior de la 2-esfera $\mathbb{S}^2$

Estos Grupos de Homología son, por construcción, grupos abelianos finitamente generados. Y por tanto, se puede descomponer como $H_q(K) \cong \mathbb{Z}^n \oplus \mathbb{Z}_{p_1^{\alpha_1}} \oplus \ldots \oplus \mathbb{Z}_{p_k^{\alpha_k}}$ para $k, \alpha_1, \ldots, \alpha_k \in \mathbb{N}$ y $p_1, \ldots, p_k$ números primos no necesariamente distintos. Y siendo $n \in \mathbb{N}$ el rango de $H_q(K)$, lo cual motiva la siguiente definición.

Definición 1.3.8. El *número de Betti* β_q en la dimensión q es igual al rango del grupo de homología H_q, es decir:

$$\beta_q = \text{rank}(H_q),$$

Los números de Betti son invariantes topológicos. De forma paralela a los grupos de homología, β_0 medirá el número de componentes conexas, β_1 el número de agujeros 1-dimensionales y β_2 el número de cavidades 2-dimensionales de nuestro espacio topológico.

Ejemplo : Vamos a calcular de grupos de homología de la Banda de Möbius $\mathbb{M}$:

Sea la Figura 1.6 la triangulación de $\mathbb{M}$ con la orientación de sus vértices $(x_0, x_1, x_2, x_3, x_4, x_5)$,

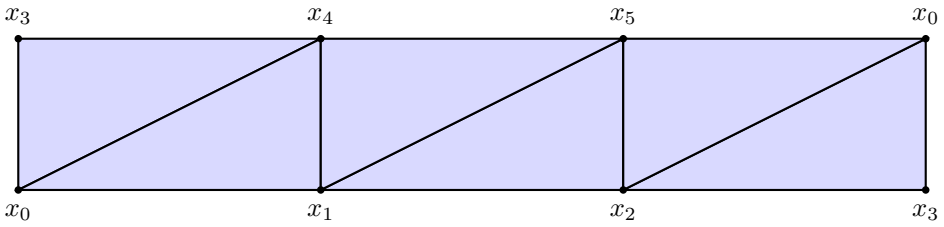


Figura 1.6

denotada por $\mathcal{M}$. Se ve a simple vista que no existe ningún 3-símplex, por lo tanto $B_2(\mathcal{M}) = 0$. Supongamos que

$$w = a_0[x_0, x_1, x_4] + a_1[x_0, x_3, x_4] + a_2[x_1, x_2, x_5] + a_3[x_1, x_4, x_5] + a_4[x_2, x_3, x_0] + a_5[x_2, x_5, x_0]$$

es un 2-ciclo. Al calcular el borde de w, como $\partial(w) = 0$, los coeficientes de las 2-cadenas $(x_0, x_1), (x_1, x_2), (x_2, x_3), (x_3, x_4), (x_4, x_5), (x_5, x_0)$ deberán ser 0 ya que solo aparecen una sola vez y no se pueden cancelar con nada. Por tanto $a_0, a_1, \ldots, a_5 = 0$ y $Z_2(\mathcal{M}) = 0 \therefore H_2(\mathcal{M}) = 0$. Ahora, se pueden identificar visualmente los siguientes 1-ciclos:

$$z = 1 \cdot [x_0, x_1] + 1 \cdot [x_1, x_2] + 1 \cdot [x_2, x_3] + 1 \cdot [x_3, x_0]$$

$$z' = 1 \cdot [x_0, x_3] + 1 \cdot [x_3, x_4] + 1 \cdot [x_4, x_5] + 1 \cdot [x_3, x_0]$$

Y como la resta de estos es el borde de una 2-cadena

$$z - z' = \partial(1 \cdot [x_0, x_1, x_4] + 1 \cdot [x_1, x_2, x_5] + 1 \cdot [x_2, x_3, x_0] - 1 \cdot [x_0, x_2, x_5] - 1 \cdot [x_5, x_1, x_4] - 1 \cdot [x_4, x_0, x_3])$$

$$\therefore \; z \sim z'$$

Realizando cálculos análogos podemos ver que todo 1-ciclo es homólogo a un múltiplo de z

$$\therefore \; H_1(\mathcal{M}) = \{ \; [az] \mid a \in \mathbb{Z} \; \} \implies H_1(\mathcal{M}) \sim \mathbb{Z}$$

En cuanto a $H_0(\mathcal{M})$, todo 0-símplex está unido a los demás a través de cadenas de 1-símplices.

Por tanto todo 0-símplex es homólogo a x_0 (por ejemplo).

$$H_0(\mathcal{M}) = \{\ [a[x_0]] \mid a \in \mathbb{Z}\ \} \implies H_0(\mathcal{M}) \sim \mathbb{Z}$$

1.4. Propiedades de la Homología Simplicial

Nuestro objetivo en esta sección será demostrar que los grupos de homología son invariantes homotópicos. Esto significa que espacios "deformables" el uno en el otro tienen los mismos grupos de homología. Este resultado es crucial, pues justifica que las características que mide la homología son robustas y no artefactos de una representación geométrica particular.
Para ello, empezaremos introduciendo los complejos de cadenas, que nos servirán para descomponer algebraicamente nuestro complejo orientado.

Definición 1.4.1. El operador de frontera conecta los grupos de cadenas en un *complejo de cadenas* $(\{C_i\}, \partial)$, una serie (posiblemente infinita) de grupos y homomorfismos de grupos:

$$\cdots \longrightarrow C_{q+1} \xrightarrow{\partial_{q+1}} C_q \xrightarrow{\partial_q} C_{q-1} \longrightarrow \cdots$$

Surge ahora una pregunta ¿Cómo podemos relacionar dos complejos de cadenas entre sí, manteniendo su estructura? La respuesta son las aplicaciones de cadenas.

Definición 1.4.2. Si $(\{C_i\}, \partial)$ y $(\{C'_i\}, \partial')$ son complejos de cadenas, una *aplicación de cadenas*

$$\phi : (\{C_i\}, \partial) \to (\{C'_i\}, \partial')$$

es una serie de homomorfismos $\{\phi_n : C_n \to C'_n\}$ tal que el siguiente diagrama conmuta:

$$\begin{array}{ccccccccc}
\cdots & \xrightarrow{\partial_{q+2}} & C_{q+1} & \xrightarrow{\partial_{q+1}} & C_q & \xrightarrow{\partial_q} & C_{q-1} & \xrightarrow{\partial_{q-1}} & \cdots \\
 & & \downarrow{\scriptstyle \phi_{q+1}} & & \downarrow{\scriptstyle \phi_q} & & \downarrow{\scriptstyle \phi_{q-1}} & & \\
\cdots & \xrightarrow{\partial'_{q+2}} & C'_{q+1} & \xrightarrow{\partial'_{q+1}} & C'_q & \xrightarrow{\partial'_q} & C'_{q-1} & \xrightarrow{\partial'_{q-1}} & \cdots
\end{array}$$

es decir, $\partial_n \phi_n = \phi_{n-1} \partial'_n \quad$ para todo $n \in \mathbb{Z}$.

Pero para probar la invarianza homotópica de los grupos de homología, necesitamos ver que las aplicaciones continuas entre espacios topológicos inducen homomorfimos entre sus grupos de homología. Para ello, lo veremos primero con aplicaciones de cadenas, para posteriormente extenderlo a aplicaciones entre poliedros asociados, y finalmente a cualquier aplicación.
Vemos si las las aplicaciones de cadenas también preservan las relaciones entre ciclos y fronteras y por tanto, los grupos de homología.

Teorema 1.4.3. Sea $\quad f : (\{C_i(K)\}, \partial) \to (\{C_i(L)\}, \partial') \quad$ una aplicación de cadenas, esta induce los homomorfismos sobre los grupos de homología $\quad f_{q*} : H_q(K) \to H_q(L) \quad$ para cada $q \in \mathbb{N}$.

Demostración: Si $b_q = \partial(c_{q+1}) \in B_q(K)$, entonces

$$f_q(b_q) = f_q(\partial(c_{q+1})) = \partial'(f_{q+1}(c_{q+1})),$$

por lo tanto, $f_q(b_q)$ es la frontera de la cadena $(q+1)$-dimensional $f_{q+1}(c_{q+1})$. Así, f_q lleva $B_q(K)$ en $B_q(L)$.
Vamos ahora a demostrar que f_q lleva $Z_q(K)$ en $Z_q(L)$. Esto es cierto para $q = 0$, ya que $Z_0(K) = C_0(K)$ y $Z_0(L) = C_0(L)$. Para $q \geq 1$, supongamos que $z_q \in Z_q(K)$. Notemos que

$$\partial'(f_q(z_q)) = f_{q-1}(\partial(z_q)) = f_{q-1}(0) = 0,$$

por lo tanto, $f_q(z_q)$ es un q-ciclo en L.
Dado que

$$H_q(K) = Z_q(K)/B_q(K), \qquad H_q(L) = Z_q(L)/B_q(L),$$

entonces el homomorfismo inducido $f_{q*} : H_q(K) \to H_q(L)$ puede definirse de la forma estándar: $f_{q*}(z_q + B_q(K)) = f_q(z_q) + B_q(L)$,
o, equivalentemente, $f_{q*}([z_q]) = [f_q(z_q)]$. □

Lema 1.4.4. Sea $s : |K| \to |L|$ una aplicación simplicial entre los complejos K y L, esta induce un homomorfismo de grupos $s_* : H_q(K) \to H_q(L)$.

Demostración: Vamos a construir un homomorfismo inducido $s_q : C_q(K) \to C_q(L)$ para cada dimensión $q \geq 0$. Dado un q-símplex orientado $\sigma = (v_0, \ldots, v_q)$ en $C_q(K)$, definimos la acción de s_q sobre σ mediante:

$$s_q(\sigma) := \begin{cases} (s(v_0), \ldots, s(v_q)) & \text{si } s(v_i) \neq s(v_j) \text{ para todo } i \neq j, \\ 0 & \text{en caso contrario.} \end{cases}$$

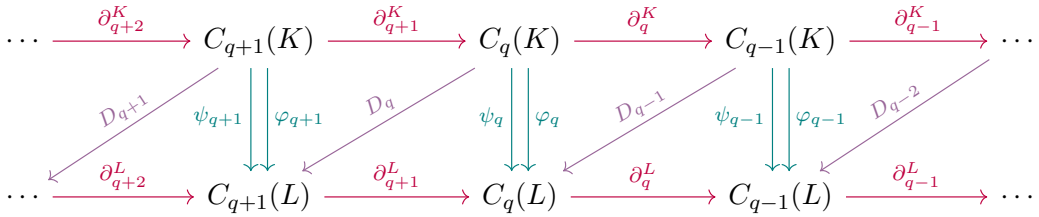
Para una q-cadena general $c = \sum_{i=1}^{n} a_i\sigma_i$ (donde $a_i \in \mathbb{Z}$ y σ_i son q-símplices orientados), extendemos s_q por linealidad: $s_q(c) = \sum_{i=1}^{n} a_i s_q(\sigma_i)$. Vemos que es un homomorfismo: Para dos q-cadenas $c_1, c_2 \in C_q(K)$,

$$s_q(c_1 + c_2) = s_q\left(\sum a_i\sigma_i + \sum b_j\tau_j\right) = \sum a_i s_q(\sigma_i) + \sum b_j s_q(\tau_j) = s_q(c_1) + s_q(c_2).$$

Ademas es trivial que σ, $s_q(-\sigma) = -s_q(\sigma)$.
Por tanto, La aplicación s_q así definida es aplicación de cadenas de $C_q(K)$ a $C_q(L)$, y ademas homomorfismo. Aplicando el Teorema anterior, obtenemos el homomorfismo de grupos de homología $s_* : H_q(K) \to H_q(L)$. □

Definición 1.4.5. Sean $\psi, \varphi : (\{C_i(K)\}, \partial^K) \to (\{C_i(L)\}, \partial^L)$ aplicaciones de cadenas. Estas dos aplicaciones son *homótopas* si existe $\mathscr{D} = \{D_i\}_{-1}^{+\infty}$ serie de homomorfismos denominada

homotopía de cadenas de la forma $D_q : C_q(K) \to C_{q+1}(L)$ tal que: $\partial^L \circ D_q + D_{q-1} \circ \partial^K = \psi_q - \varphi_q$



A continuación vamos a introducir dos resultados que nos servirán para demostrar los siguientes teoremas.

Teorema 1.4.6. Si $s, t : |K| \to |L|$ son aplicaciones simpliciales *cercanas*, en el sentido de que para cada símplice A de K podemos encontrar un símplex B en L tal que tanto $s(A)$ como $t(A)$ son caras de B, entonces $s_* = t_* : H_q(K) \to H_q(L)$ para todo q.

Teorema 1.4.7. Si $f, g : |K| \to |L|$ son aplicaciones homótopas, entonces podemos encontrar una subdivisión baricéntrica K^m y una sucesión de aplicaciones simpliciales

$$\chi_1, \ldots, \chi_n : |K^m| \to |L|$$

tal que χ_1 aproxima simplicialmente a f, χ_n aproxima simplicialmente a g, y cada par χ_i, χ_{i+1} son cercanas.

Teorema 1.4.8. Sea $f : |K| \to |L|$ cualquier aplicación continua. Entonces f induce un homomorfismo

$$f_* : H_q(K) \to H_q(L) \quad \text{en cada dimensión.}$$

Demostración: Sea $f : |K| \to |L|$ continua, aplicando el Teorema de Aproximación Simplicial 1.2.7, elegimos una aproximación simplicial $s : |K^m| \to |L|$. La cual, al ser una aplicación simplicial, induce un homomorfismo de grupos de homología, como hemos visto en el Lema 1.4.4.
Y sea $\chi : C(K) \to C(K^m)$ la aplicación de cadenas de la subdivisión baricéntrica, que también induce un homomorfismo de grupos de homología, como vimos en el Teorema 1.4.3. De esta forma, obtenemos el homomorfismo $f_* : H_q(K) \to H_q(L)$ de f por la composición

$$H_q(K) \xrightarrow{\chi_*} H_q(K^m) \xrightarrow{s_*} H_q(L)$$

Sin embargo, aún tenemos que ver que la elección de la aproximación simplicial no afecta a nuestro resultado. Para ello aplicaremos los Teoremas 1.4.6 y 1.4.7: Aproximamos simplicialmente f de dos formas, $s : |K^m| \to |L|$ y $t : |K^n| \to |L|$ de tal forma que m $\leq$ n.
Sean $\chi_1 : C(K) \to C(K^m)$ y $\chi_2 : C(K^m) \to C(K^n)$ las aplicaciones de las cadenas de las subdivisiones, y sea $\theta : |K^n| \to |K^m|$ la aplicación simplicial estándar. Tenemos que ver que:

$$s_* \chi_{1*} = t_* \chi_{1*} \chi_{2*} : H_q(K) \to H_q(L)$$

Sin embargo, al igual que t, también $s\theta : |K^n| \to |L|$ aproxima simplicialmente a $f : |K^n| \to |L|$.

Por lo tanto, $s\theta$ y t deben ser aplicaciones simpliciales cercanas y

$$s_*\theta_* = t_* : H_q(K^n) \to H_q(L).$$

Dado que también sabemos que θ_* y χ_2 son inversos el uno del otro, tenemos

$$t_*\chi_2\chi_2 x_1^* = s_*\chi_2\chi_2 x_1^* = s_* x_1^*,$$

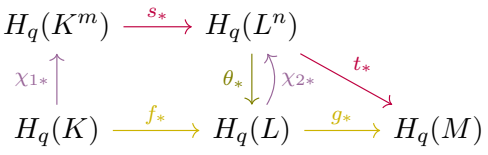
como necesitábamos.
Finalmente, tenemos un homomorfismo bien definido $f_* : H_q(K) \to H_q(L)$ □

Teorema 1.4.9. Si f es la aplicación identidad de $|K|$, entonces cada $f_* : H_q(K) \to H_q(K)$ es el homomorfismo identidad. Además, si tenemos dos aplicaciones

$$|K| \xrightarrow{f} |L| \xrightarrow{g} |M|,$$

entonces $(g \circ f)_* = g_* \circ f_* : H_q(K) \to H_q(M)$ para todo q.

Demostración: La primera parte del teorema sigue claramente por construcción. Supongamos que tenemos aplicaciones $K \xrightarrow{f} L \xrightarrow{g} M$. Sea $\chi_1 : C(K) \to C(K^m)$ la aplicación de cadenas de subdivisión, $\chi_2 : C(L) \to C(L^m)$, y sea $\theta : |L^n| \to |L|$ una aplicación simplicial estándar. Escoge una aproximación simplicial $t : |L^n| \to |M|$ para $g : |L| \to |M|$, entonces una aproximación simplicial $s : |K^m| \to |L^n|$ para $f\chi_2 : |K| \to |L^n|$. Ahora tenemos el siguiente diagrama de grupos de homología y homomorfismos:



Se verifica fácilmente que θs aproxima simplicialmente a $f : |K^m| \to |L|$ y que ts aproxima simplicialmente a $gf : |K^m| \to |M|$.

Por lo tanto, $g_* \circ f_* = t_*\chi_{2*}\theta_* s_*\chi_{1*} = t_* s_*\chi_{1*} = (ts)_*\chi_{1*} = (g \circ f)_*$ □

Teorema 1.4.10. Si $f, g : |K| \to |L|$ son aplicaciones homótopas, entonces

$$f_* = g_* : H_q(K) \to H_q(L) \quad \text{para todo } q.$$

Demostración: Se sigue directamente de los Teoremas 1.4.6 y 1.4.7, ya que, con la notación establecida en el primero,

$$f_* = s_{1*}\chi_* = s_{2*}\chi_* = \cdots = s_{n*}\chi_* = g_*$$

□

Finalmente, uniremos toda la teoría vista para demostrar la invarianza homotópica de los grupos de homología:

Definición 1.4.11. Dos e.t. $|K|, |L|$ son *homotópicamente equivalentes* si $\exists f : |K| \to |L|$, $\exists g : |L| \to |K|$ continuas tales que:

$$(f \circ g) \simeq \mathrm{id}_{|K|} \quad \wedge \quad (g \circ f) \simeq \mathrm{id}_{|L|}$$

Ahora, aplicando a los espacios de la definición anterior el Teorema 1.4.8, f y g inducen homomorfismos sobre los grupos de homología, de forma que $f_* : H_q(K) \to H_q(L)$ y $g_* : H_q(L) \to H_q(K)$. Ahora, aplicando el Teorema 1.4.10

$$(f \circ g)_* = (\mathrm{id}_{|K|})_* \quad \wedge \quad (g \circ f)_* = (\mathrm{id}_{|L|})_*$$

Haciendo uso del Teorema 1.4.9

$$(f \circ g)_* = f_* \circ g_* = (\mathrm{id}_{|K|})_* = \mathrm{id}_{H_q(|K|)} \quad \wedge \quad (g \circ f)_* = g_* \circ f_* = (\mathrm{id}_{|L|})_* = \mathrm{id}_{H_q(|L|)}$$

Por lo que f_* es inversa de g_* y viceversa, por lo que ambas son isomorfismos.
Por ende,

$$H_q(|K|) \cong H_q(|L|)$$

Capítulo 2

Análisis Topológico de Datos

El Análisis Topológico de Datos (TDA, por sus siglas en inglés) es un campo emergente que emplea herramientas topológicas y geométricas procedentes de la Topología Algebraica con el fin de obtener información de la estructura subyacente de nuestros conjuntos de datos (o datasets, del inglés). De esta forma, nos permite encontrar patrones y relaciones que pueden ser cruciales para entender la naturaleza de nuestros datos.

Una herramienta fundamental del Análisis Topológico de Datos es la homología persistente. A diferencia de la homología clásica que analiza un espacio en un momento fijo, la homología persistente describe cómo aparecen, persisten o desaparecen características topológicas, a medida que avanza el "tiempo". Sin embargo, como trabajar con un tiempo continuo sería muy complicado, crearemos las filtraciones para poder discretizarlo:

2.1. Filtraciones

Definición 2.1.1. Sea K un complejo simplicial finito, y sea $K_1 \subset K_2 \subset \cdots \subset K_N = K$ una sucesión creciente de complejos simpliciales de K. Dicha sucesión $\{K_i\}_{i=1}^N$ se denomina *filtración de complejos simpliciales*.
Definimos así la inclusión de cadenas $f_i : K_i \to K_{i+1}$, la cual induce el homomorfismo entre los grupos de homología $f_*^{i,i+1} : H_p(K_i) \to H_q(K_{i+1}) \quad \forall q \in \mathbb{N}$. Al estar la composición de aplicaciones de cadenas bien definida, obtenemos $f_{i,j} : K_i \to K_j \quad \forall i < j$ y su homomorfismo de grupos de homología $f_*^{i,j} : H_q(K_i) \to H_q(K_j) \quad \forall q \in \mathbb{N}$.

De forma que podemos ver el avance de k pasos en el tiempo como un avance de k niveles de la filtración, de K_i a K_{i+k}.

Definición 2.1.2. Sea $\mathcal{U} = \{U_i\}_{i \in I}$ una colección no vacía de conjuntos. El *nervio* de $\mathcal{U}$ es el complejo simplicial

$$\text{Nrv } \mathcal{U} := \left\{ J \subseteq I \;\middle|\; \bigcap_{j \in J} U_j \neq \emptyset \right\}$$

A continuación, presentaremos los complejos con los que vamos a trabajar.

Definición 2.1.3. Sea S un conjunto finito de puntos en $\mathbb{R}^d$ y escribimos $B_x(r) = x + r\mathbb{B}^d$ para denotar la bola cerrada con centro en x y radio r. El *complejo de Čech* de S y r es isomorfo al *nervio* de esta colección de bolas:

$$\check{\text{C}}ech(r) = \left\{ \sigma \subseteq S \;\middle|\; \bigcap_{x \in \sigma} B_x(r) \neq \emptyset \right\}.$$

De manera que si k+1 puntos tienen bolas con intersección no vacía , se formará un k-símplex entre ellos.

Teorema 2.1.4 (Teorema del Nervio)**.** Sea F una colección finita de conjuntos cerrados convexos en el espacio euclídeo. Entonces, el nervio de F y la unión de los conjuntos en F tienen el mismo tipo de homotopía.

Este teorema justifica el uso del complejo Čech en TDA: al aproximar un conjunto de datos con bolas, su nervio preserva la topología subyacente. Esta construcción, sin embargo presenta un costo computacional elevado.

Con el fin de reducir dicho costo computacional, en lugar de verificar todas las subcolecciones, podemos simplemente comprobar los puntos por pares y añadir símplices de dimensión 2 o superior siempre que sea posible. Esta simplificación conduce al complejo de Vietori-Rips:

Definición 2.1.5. El *complejo de Vietoris-Rips* de S y r, consiste en todos los subconjuntos tales que sus puntos están a distancia $2r$:

$$\text{VR}(r) = \{\sigma \subseteq S \mid \text{d}(x, y) \leq 2r \quad \forall x, y \in \sigma\}$$

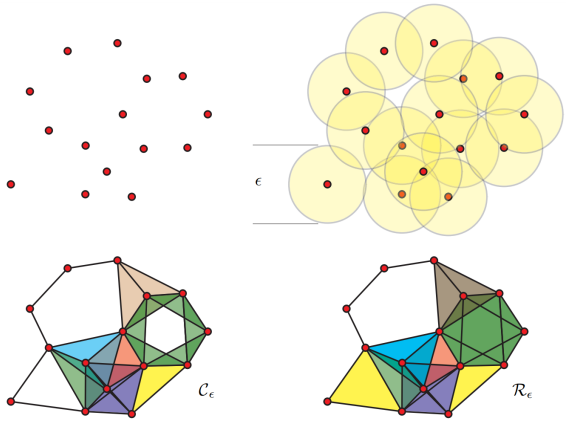


Figura 2.1: Comparativa entre complejos de Čech y Vietoris–Rips para el mismo ε. Créditos de imagen a [9].

Claramente, las aristas en el complejo de Vietoris-Rips son las mismas que en el complejo de Čech. Además, se cumple que Čech(r) $\subseteq$ Vietoris-Rips(r).

Lema 2.1.6. Sean S un conjunto finito de puntos en el Espacio Euclídeo y $r \geq 0$, tenemos que

$$\text{Čech(r)} \subseteq \text{Vietoris-Rips}(r) \subseteq \text{Čech}(\sqrt{2}\text{r})$$

Como hemos visto anteriormente, el complejo Čech es homotópicamente equivalente a la unión de bolas $\bigcup_{x \in S} B(x, r)$ por el Teorema 2.1.4. Al compartir el complejo Vietoris-Rips el mismo tipo de homotopía en escalas similares, preserva la información topológica esencial de los datos. Esto justifica su uso en TDA, ya que el complejo de Vietoris-Rips es más eficiente computacionalmente.

2.2. Homología Persistente

Sin embargo, ahora surge una pregunta ¿Cuál es el tamaño óptimo de ε? Ya que, si eligiéramos un ε demasiado pequeño, obtendríamos un conjunto de puntos aislados. Por el contrario, si eligiéramos un ε demasiado grande, obtendríamos un solo símplex de dimensión acorde al numero de puntos.

A pesar de ser tanto computable como reveladora, la homología de un complejo asociado a un conjunto de puntos en un valor particular de ε es insuficiente, es un error preguntar qué valor de ε es óptimo. Tampoco basta con hacer un "cuenta" del número y tipos de agujeros que aparecen en cada valor de parámetro ε. Se requiere un método para ver que agujeros son esenciales y cuáles pueden ser ignorados de manera segura. La persistencia va a ser nuestra respuesta a este problema:

Definición 2.2.1. Sea $\mathcal{C} = (C_i)_{i=1}^{N}$ una filtración, junto con las aplicaciones de cadenas $f^i : C_i \to C_{i+1}$. Denotaremos al par $\{C_i, f^i\}$ *complejo de persistencia*.
Para $i < j$, la (i,j)-*homología persistente* de $\mathcal{C}$ de dimensión q, denotada $H_q^{i\to j}(C)$, se define como la imagen del homomorfismo inducido $f_{q*}^{i,j} : H_q(C^i) \to H_q(C^j)$.
Podemos expresar el grupo de (i,j)-homología persistente de dimensión q como

$$H_q^{i\to j}(C) = \frac{Z_q^i}{B_k^j \cap Z_q^i}$$

Como acabamos de ver, la homología persistente captura agujeros que persisten entre pasos de la filtración, sin embargo, ahora nos surge la necesidad de una descripción más estructurada y manejable de esta información.

Esta necesidad nos lleva a introducir la noción de *módulo de persistencia*, una herramienta algebraica que unifica toda la homología a lo largo de la filtración en una sola estructura. Así, es posible descomponer la homología persistente y obtener una representación canónica de sus generadores y sus intervalos de persistencia.

Sin embargo, antes de eso vamos a recordar algunos conceptos algebraicos. Asumiremos que nuestro anillo R es conmutativo con elemento neutro.
Si además, R no tuviera divisores de cero y todo ideal de R fuera principal. Diriamos que R es un *dominio de ideales principales* (DIP).

Definición 2.2.2. Un *anillo graduado* es un anillo $\langle R, +, \cdot\rangle$ equipado con una descomposición de sumas directas de grupos abelianos $R \cong \bigoplus_{i\in\mathbb{Z}} R_i$ de tal forma que la multiplicación esta definida por pares bilineales $R_n \otimes R_m \to R_{n+m}$. Los elementos de un mismo R_i se denominan *homogéneos* de *grado* i.
Graduaremos no negativamente nuestro anillo polinómico $R[t]$ con la *graduación estándar* $(t^n) = t^n \cdot R[t]$.
Un *módulo graduado* M sobre un anillo graduado R es un modulo equipado con una descomposición de sumas directas, $M \cong \bigoplus_{i\in\mathbb{Z}} M_i$, de tal forma que la acción de R sobre M está definida por pares bilineales $R_n \otimes M_m \to M_{n+m}$.
Un anillo (o módulo) graduado está *graduado no negativamente* si $R_i = 0$ (o $M_i = 0$) $\quad \forall i < 0$.

Teorema 2.2.3. Si D es un DIP, entonces cada D-módulo finitamente generado es isomorfo a la suma directa de D-módulos cíclicos. Es decir, se descompone de manera única de la forma

$$D^{\beta} \oplus \left(\bigoplus_{i=1}^{m} D/d_i D\right)$$

para $\beta \in \mathbb{Z}$, $d_i \in D$, tal que $d_i | d_{i+1}$
De forma similar, cada módulo graduado M sobre un DIP graduado D se descompone de manera única de la forma

$$\left(\bigoplus_{i=1}^{n} \Sigma^{\alpha_i} D\right) \oplus \left(\bigoplus_{i=1}^{m} \Sigma^{\gamma_i} D/d_j D\right)$$

donde $d_i \in D$ son elementos homogéneos tal que $d_i | d_{i+1}, \alpha_i, \gamma_j \in \mathbb{Z}$ y Σ^{α} denota un α-*desplazamiento hacia arriba* en la graduación.

Definición 2.2.4. Un *módulo de persistencia* $\mathcal{M} = \{M^i, \varphi^i\}$ es una familia de R-módulos M^i, junto con los homomorfismos $\varphi^i : M^i \to M^{i+1}$.
Un módulo de persistencia es de *tipo finito* si cada componente es un R-módulo finitamente generado, y $\exists m \in \mathbb{Z}$ tal que las aplicaciones φ^i son isomorfismos $\forall i \geq m \in \mathbb{Z}$

Esta última definición implica que hay algún punto m a partir del cual se estabiliza la sucesión. Este concepto justifica toda la teoría algebraica introducida, como $H_q(C^i)$ son grupos abelianos, en particular $\mathbb{Z}$-módulos, $H_q(\mathcal{C}) = \{H_q(C^i), f_i\}$ son módulos de persistencia de tipo finito. Sin embargo, como $\mathbb{Z}[x]$ no es un DIP, vamos a tener que calcular nuestra homología con coeficientes en un cuerpo F, de forma que $F[x]$ sí sea un DIP. De esta forma vamos a realizar estra construcción:

Suponiendo un módulo de persistencia $\mathcal{M} = \{M^i, \varphi^i\}_{i \geq 0}$ sobre un cuerpo F, para asegurar que $F[x]$ sea un DIP. Graduamos a $F[x]$ con la graduación estándar, de forma que los ideales graduados sean de la forma $x^n \cdot F[x]$, y definimos un módulo graduado sobre $F[x]$ de la forma:

$$\alpha(\mathcal{M}) = \bigoplus_{i=0}^{\infty} M^i$$

y donde la acción de x está dada por $x \cdot (m^0, m^1, m^2, \ldots) = (0, \varphi^0(m^0), \varphi^1(m^1), \varphi^2(m^2), \ldots)$, de forma que x actúa como un desplazamiento hacia el módulo de arriba en la graduación.

De esta manera, hemos obtenido un módulo graduado $\alpha(\mathcal{M})$ sobre un DIP graduado $F[x]$, así, podemos descomponer los módulos de persistencia de la siguiente forma:

Teorema 2.2.5. Para un módulo de persistencia de tipo finito $\mathcal{C}$ con coeficientes en el cuerpo F

$$H_*(\mathcal{C}; F) \cong \bigoplus_i x^{t_i} \cdot F[x] \oplus \left(\bigoplus_j x^{r_j} \cdot \frac{F[x]}{x^{s_j} \cdot F[x]} \right)$$

Las componentes libres corresponden biyectivamente con aquellos generadores de homología que surgen en el paso t_i de la filtración y persisten infinitamente. Las componentes de torsión corresponden a aquellos generadores de homología que aparecen en el paso r_j y desaparecen en el paso $r_j + s_j$ de la filtración.

Definición 2.2.6. Un *código de barras de persistencia* es una representación gráfica de $H_k(C;F)$ como una colección de segmentos de líneas horizontales en un plano cuya eje horizontal corresponde al parámetro y cuyo eje vertical representa un orden arbitrario de los generadores de homología.

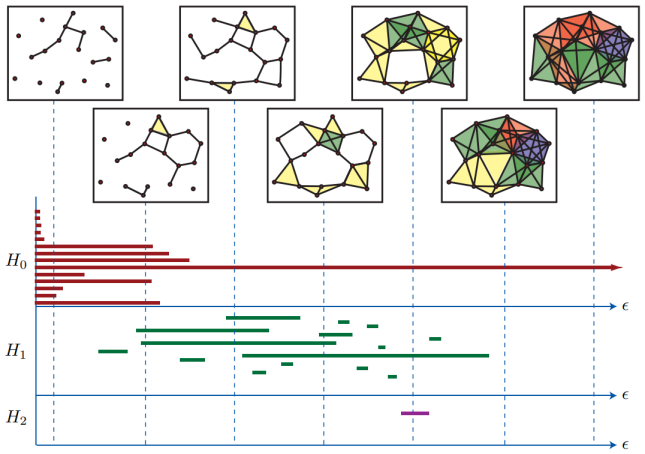


Figura 2.2: Código de barras de persistencia de un complejo $(\mathcal{R}_{\varepsilon_i})_i$.
Créditos de imagen a [9].

Como podemos ver al inicio de H_0 aparecen tantas barras como puntos, ya que al tener un valor de ε tan pequeño obtenemos un conjunto totalmente disconexo. A medida que va aumentando ε, se van uniendo los puntos. Si dos puntos se unen, la barra de uno de los dos acabará, y la componente conexa que formen seguirá en la otra. Por tanto, acabaremos obteniendo la barra de la componente conexa de todos los puntos, que persistirá infinitamente. De H_1 vemos que a medida que aparecen 1-ciclos aparecen barras que duran hasta que el valor de epsilon se haga mayor que el diámetro del 1-ciclo. Lo mismo para H_2. En este caso podemos ver que aparecen 2 agujeros significativos, pero ninguna cavidad significativa.

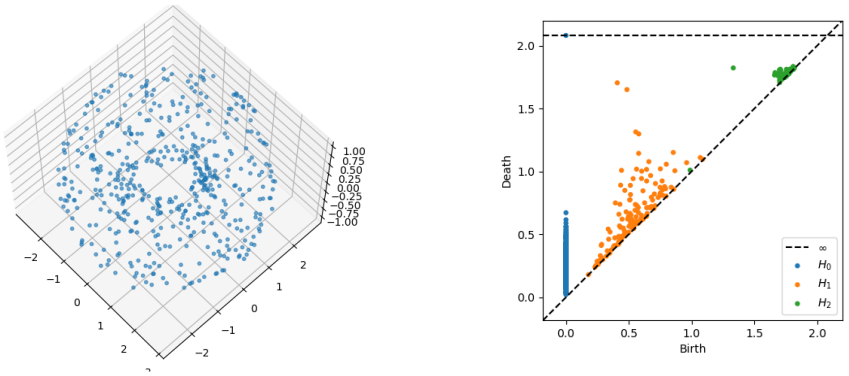


Figura 2.3: Nube de puntos en forma de Toroide
junto con su Diagrama de Persistencia

Además de con códigos de barras, podemos representar la homología persistente con *Diagramas de Persistencia*, estos representan en cada eje el nacimiento y muerte de cada elemento de

H_q, pudiéndose hacer también para cada grupo de homología por separado. De forma que cuanto más alejado esté un punto de la recta x=y, representa una clase de homología más persistente. Por otra parte, cuanto más cerca de la diagonal, más posible es que dicho elemento represente ruido. Notamos que en el diagrama de persistencia de la Figura 2.3 se observa un punto de H_0 en la línea horizontal superior, este simboliza un punto en el infinito, haciendo referencia a la componente conexa final. Gracias a los diagramas de persistencia, vamos a obtener una forma de comparar la homología persistente de dos conjuntos de datos. Lo haremos a través de la distancia bottleneck:

Definición 2.2.7. Sean D_1 y D_2 dos diagramas de persistencia. La *distancia bottleneck* entre D_1 y D_2 se define como:

$$d_B(D_1, D_2) = \inf_{\gamma} \sup_{p \in D_1} \|p - \gamma(p)\|_\infty$$

donde γ es el conjunto de biyecciones entre los multi-conjuntos D_1 y D_2 (Pueden tener varios puntos en las mismas coordenadas, si hay un punto con multiplicidad $m > 1$ se considera como m copias disjuntas). Definiendo $\|p - q\|_\infty = \text{máx}(\mid x_p - x_q \mid, \mid y_p - y_q \mid)$ y por convención para el caso $y_p = y_q = +\infty$ usaremos $\|p - q\|_\infty = \mid x_p - x_q \mid$

Esta definición motiva la inclusión de la diagonal de $\mathbb{R}^2$, de tal forma que añadimos tantos puntos de la diagonal como haga falta para igualar la cardinalidad de ambos diagramas de persistencia. De esta forma conseguimos la biyección global que minimice la máxima distancia. Pero, ¿Cómo sabemos que realmente esta distancia implica que dos nubes de datos sean distintos?

Definición 2.2.8. Sean $X, Y \subset \mathbb{R}^d$ dos conjuntos compactos y sea $\varepsilon \geq 0$. Una ε-correspondencia entre X y Y es un subconjunto $C \subseteq X \times Y$ tal que:

(i) Para todo $x \in X$, existe $y \in Y$ tal que $(x, y) \in C$;

(ii) Para todo $y \in Y$, existe $x \in X$ tal que $(x, y) \in C$;

(iii) Para todo $(x, y), (x', y') \in C$, se cumple que $|d(x, x') - d(y, y')| \leq \varepsilon$
donde $d(x, x') = \|x - x'\|$ es la distancia euclidiana.

La distancia de Gromov-Hausdorff entre X y Y se define por

$$d_{GH}(X, Y) = \text{ínf}\{\varepsilon \geq 0 : \text{existe una } \varepsilon\text{-correspondencia entre } X \text{ y } Y\}.$$

Teorema 2.2.9. Sean $P, Q \subset \mathbb{R}^d$ nubes de puntos finitas y $D_{P,k}, D_{Q,k}$ los diagramas de persistencia del grupo de homología de grado k, para una filtración cualquiera de complejos simpliciales de Čech o Vietoris-Rips. Se cumple $\forall k > 0$

$$d_B\big(D_P, D_Q\big) \leq d_{GH}(P, Q).$$

Gracias a esta desigualdad aseguramos cierta estabilidad de la distancia bottleneck, frente al ruido. Pequeños cambios geométricos capturados por la distancia Gromov-Hausdorff no afectan a las características topológicas capturadas por la distancia Bottleneck.
Además, una distancia bottleneck alta nos asegura que los e.t. son geométricamente distintos.

Capítulo 3

Aplicación

Nuestro objetivo en este último capítulo será aplicar la teoría introducida anteriormente para refinar conjuntos de datos o datasets de forma topológica. Es decir, queremos eliminar todo punto que sea redundante, atípico o sobrerrepresentado, pero manteniendo la estructura topológica de nuestro conjunto.

Para llevar a cabo esta tarea, se ha desarrollado una implementación propia en Python, cuyo código completo y comentado se puede consultar en el Apéndice A.
Dicha implementación se fundamenta en dos bibliotecas clave: `Ripser`, para el cálculo eficiente de la homología persistente mediante complejos de Vietoris-Rips y la generación de diagramas de persistencia; y `Persim`, para cuantificar la similitud entre estos diagramas a través de la distancia bottleneck.

3.1. El algoritmo

Algoritmo 1: Filtrado por Impacto Topológico

Input: $Puntos$, $Epsilon$, $MaxIter$

$S \leftarrow$ Estructura($Puntos$);
for $iter \leftarrow 0$ **to** $MaxIter - 1$ **do**
  $D \leftarrow$ DiagramasPersistencia(S);
  $Eliminar \leftarrow []$;
  for $i \leftarrow 0$ **to** $Longitud(S) - 1$ **do**
    $S_i \leftarrow$ EliminarElemento(S, i);
    $D_i \leftarrow$ DiagramasPersistencia(S_i);
    $d \leftarrow$ DistanciaBottleneck(D, D_i);
    if $d < Epsilon$ **then**
      Añadir i a $Eliminar$;
  if $Eliminar$ *está vacío* **then**
    break;
  $S \leftarrow$ EliminarElementos($S, Eliminar$);

Output: S

La idea básica del algoritmo es conseguir filtrar el conjunto de puntos según su relevancia topológica. Calcularemos el diagrama de persistencia de nuestro conjunto original, e iterativa-

mente compararemos este con el diagrama de persistencia del mismo conjunto pero sin cada uno de los puntos. De esta forma, si la distancia bottleneck entre los dos diagramas no supera cierto umbral, significa que este punto no afecta lo suficiente a la topología de nuestro conjunto, y por tanto podemos añadirlo a un a conjunto de puntos que serán eliminados en lote al final de la iteración. Vamos a imponer un umbral muy bajo, 0.02, de forma que nos aseguramos de que este punto es irrelevante a nivel topológico.

Para probar el funcionamiento del algoritmo, vamos a crear artificialmente una nube de puntos en forma de toroide como el de la Figura 2.4. Pero esta vez, vamos a añadir pequeñas perturbaciones que sigan una distribución normal a cada coordenada de cada punto del toro.

De esta forma, simularemos un caso más parecido al que nos podríamos encontrar en un dataset real. Además, esto nos dará la oportunidad de verificar empíricamente la robustez de nuestro algoritmo frente a pequeñas perturbaciones, ya que al estar basado en la distancia bottleneck, teóricamente, debido al Teorema 2.2.9 debería serlo.
También vamos a añadir datos atípicos (u outliers, del inglés) cuyas coordenadas seguirán una distribución uniforme sobre un cubo que contenga al toro. En este caso, el toro consistirá de 500 puntos, y añadiremos un 2 % de outliers.

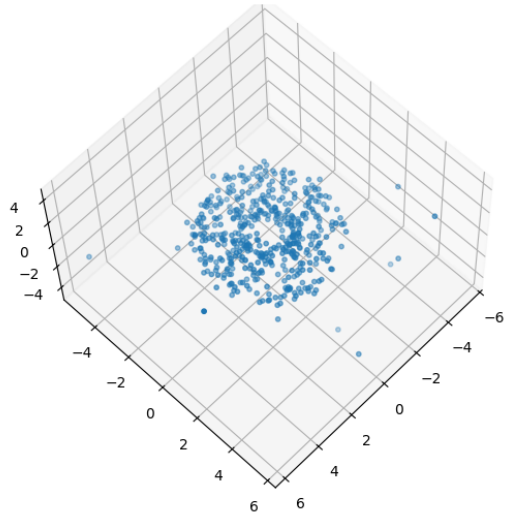


Figura 3.1: Toro con ruido normal y outliers

Pero, ¿Qué dimensión del grupo de homología de los diagramas de persistencia utilizamos para comparar utilizando la distancia bottleneck? Vamos a ver cuál de las 3 es la mejor opción. Empezaremos por H_0:

3.2. Filtrado mediante H_0

A priori, el filtrado basado en H_0 no debería ser óptimo, sin embargo, su análisis revela comportamientos interesantes. Por definición, la distancia bottleneck busca el emparejamiento que minimice la peor pareja. Por tanto, al quitar un punto p_i de nuestro diagrama de persistencia y emparejar los puntos del diagrama de persistencia original con los del diagrama de persistencia sin p_i, tenemos dos opciones, o emparejamos cada punto consigo mismo y a p_i con la diagonal, o intentamos emparejar cada punto con otro distinto a sí mismo. De forma que si la peor pareja tiene un coste menor de 0.02, ese punto será irrelevante topológicamente y podremos eliminarlo.

Vamos a centrarnos en la primera estrategia, está claro que el peor emparejamiento sería el de p_i con la diagonal, el cual tiene un coste de $\frac{persistencia}{2} = \frac{t_{muerte} - t_{nacimiento}}{2} = \frac{t_{muerte}}{2}$ dado que todos los elementos de H_0 nacen en el 0. Como en esta dimensión el tiempo de muerte de un punto es directamente proporcional a la distancia con su vecino mas cercano, el único escenario donde un punto tuviera una distancia menor a 0.02 a la diagonal, sería que tuviera a su vecino

mas cercano a 0.04 de distancia. Lo cual es improbable en nuestro conjunto, pero sería útil para eliminar puntos sobrerrepresentados.

Como siempre vamos a buscar la mejor opción entre las dos descritas, vamos a intentar buscar un caso donde la segunda estrategia pudiera ser mejor. Para ello necesitaremos que el peor emparejamiento, y por tanto el resto de emparejamientos, tuvieran un coste de menos de 0.02, con la condición de no emparejar los puntos consigo mismos (o al menos no todos). Esto solo se cumpliría en el caso de que hubiera pares de puntos de nuestro diagrama que estuvieran a una distancia menor a 0.02, además del punto emparejado con la diagonal a una distancia menor de 0.02 (que indicaría un punto sobrerrepresentado como hemos argumentado anteriormente), lo cual es extremadamente improbable. La otra opción de que esto sucediera sería que nuestro diagrama de persistencia consistiera de un clúster de radio menor de 0.04 donde se encontraran todos los puntos, que representaría una estructura regular casi perfecta, que no necesitaría filtrado de ningún tipo al no tener ruido ni outliers.

Vamos a comprobarlo compilando nuestro código utilizando diagramas de persistencia de H_0:

Como podemos observar en la Figura 3.2, no se elimina ningún punto de nuestro conjunto de datos, lo cual coincide con nuestras predicciones. Aunque este filtrado no sea útil para nuestros objetivos, y en este caso no dé ningún resultado, podría ser útil para eliminar puntos sobrerrepresentados que no aportan información topológica o simplificar nubes de datos densas.

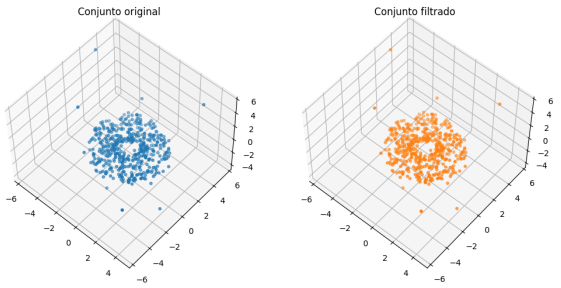


Figura 3.2: Filtrado mediante H_0.

3.3. Filtrado mediante H_1

Probaremos ahora con H_1, sin embargo la connotación de persistencia de los elementos de nuestro diagrama es sustancialmente distinta a la anterior, mientras que antes estaba ligada a la distancia al vecino más cercano, ahora está ligada a la robustez y escala de nuestros ciclos 1-dimensionales (agujeros). De esta forma, un ciclo de gran tamaño formado por un conjunto grande de puntos tendrá una persistencia muy alta, mientras que un pequeño ciclo formado por pocos puntos tendrá una persistencia baja.

Al comparar mediante H_1 deberíamos cumplir dos objetivos: El primero será eliminar outliers, ya que estos no participan en la formación de ciclos relevantes, posiblemente de algunos pequeños y efímeros, representados por puntos muy cercanos a la diagonal. Y por lo tanto, al buscar un emparejamiento entre los dos diagramas de persistencia, si el punto no interviene en ningún ciclo (es decir que al eliminarlo, no desaparece ningún punto del diagrama) está claro que el mejor emparejamiento será el de cada punto del diagrama consigo mismo, lo cual tiene coste 0 y podremos eliminar dicho punto. En el caso de que los outliers intervinieran en la formación de ciclos efímeros, al estar dichos ciclos representados en el diagrama como puntos muy cercanos a la diagonal, podríamos emparejar en los diagramas a cada punto consigo mismo y el punto de dicho ciclo con la diagonal, y al ser este tan cercano a ella, tendría un coste bajo y podríamos

eliminar el outlier.

El segundo objetivo será eliminar puntos irrelevantes. Si eliminamos un punto que participa en la creación de un ciclo, el ciclo en vez de desaparecer toma otro camino alternativo, si el punto que representa dicho ciclo en el diagrama de persistencia no se desplaza apenas, querrá decir que ese punto es irrelevante topológicamente y procederemos a eliminarlo. Por esta misma razón, el algoritmo también sería capaz de eliminar puntos sobrerrepresentados, aunque en nuestro conjunto de ejemplo no los haya. Veamos que resultados obtenemos:

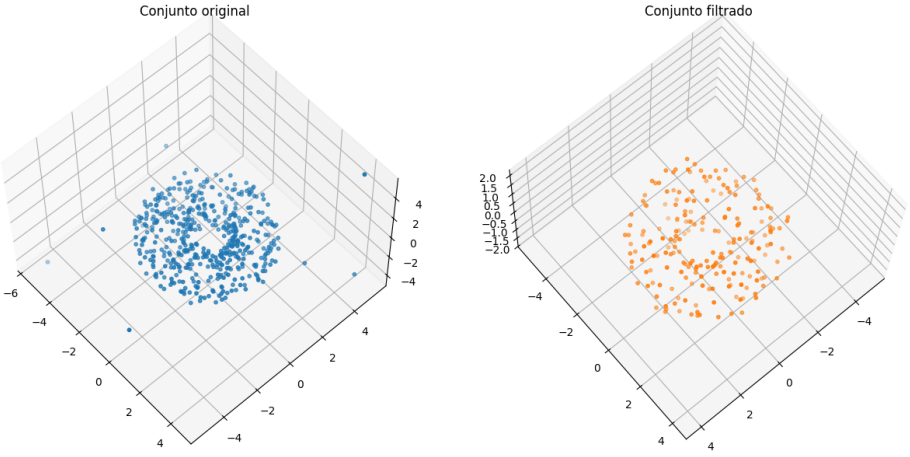


Figura 3.3: Filtrado mediante H_1.

El algoritmo ha eliminado 300 puntos de 510, un 58,8 % del total. Pero, ¿Ha eliminado los puntos correctamente sin romper su estructura? Vamos a comprobarlo comparando a través de la distancia bottleneck. Sean $\{D_{O,i}\}_i, \{D_{F,i}\}_i$ los diagramas del grupo de homología de dimensión i del conjunto original y el filtrado respectivamente, obtenemos estos resultados:

$$d_B(D_{O,0}, D_{F,0}) = 2{,}9721$$

$$d_B(D_{O,1}, D_{F,1}) = 0{,}1896$$

$$d_B(D_{O,2}, D_{F,2}) = 0{,}1036$$

Estos datos son reveladores, no solo estamos consiguiendo muy buenos valores para H_1 sino también para H_2, lo cual indica que se están preservando las características topológicas, tanto los ciclos como las cavidades.

Además, una distancia bottleneck tan alta entre los diagramas de persistencia de los grupos de homología de dimensión 0 confirman los resultados evidentes a simple vista en la Figura 3.3, estamos eliminando los outliers correctamente. Esto se hace más evidente al observar la Figura 3.4, las componentes conexas más alejadas que representan los outliers desaparecen tras el filtrado.

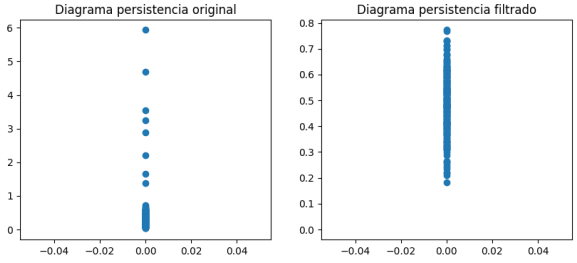


Figura 3.4: Diagramas de persistencia de H_0.

Los resultados experimentales demuestran de manera concluyente que el algoritmo basado en H_1 es capaz de cumplir nuestros objetivos de manera eficaz: identifica y elimina outliers y, a la vez, reduce el tamaño de nuestro dataset eliminando puntos irrelevantes sin destruir sus características topológicas esenciales.

3.4. Filtrado mediante H_2

A priori, dado que los elementos de H_2 tan frágiles (puesto que si cualquier elemento del recubrimiento de la cavidad despareciera, dicha cavidad también dejaría de existir) se esperaría que este algoritmo pudiera eliminar correctamente los outliers y dejar intactos los puntos que recubren la cavidad del toro. Sin embargo, la Figura 3.5 nos muestra un resultado muy distinto a nuestras predicciones, el algoritmo ha eliminado 494 puntos de los 510 originales.

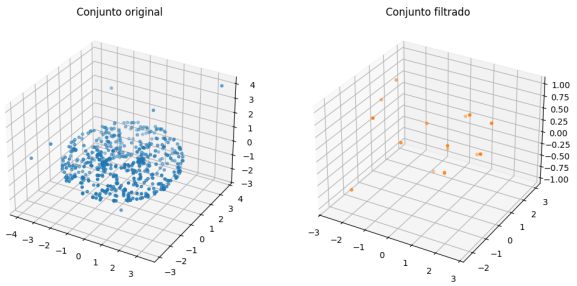


Figura 3.5: Filtrado mediante H_2 con $\varepsilon = 0{,}02$.

Con estos valores de ε aunque uno por uno los puntos a eliminar se pudieran considerar irrelevantes topológicamente, al eliminar en nuestro algoritmo todos a la vez tras evaluar todos los puntos, dada la fragilidad de las cavidades, provocaríamos un colapso de nuestra estructura. Con el fin de eliminar menos puntos vamos a reducir el tamaño de ε a 0,002. Obteniendo así estos resultados:

$$d_B(D_{O,0}, D_{F,0}) = 1{,}5305$$

$$d_B(D_{O,1}, D_{F,1}) = 0{,}3658$$

$$d_B(D_{O,2}, D_{F,2}) = 0{,}2098$$

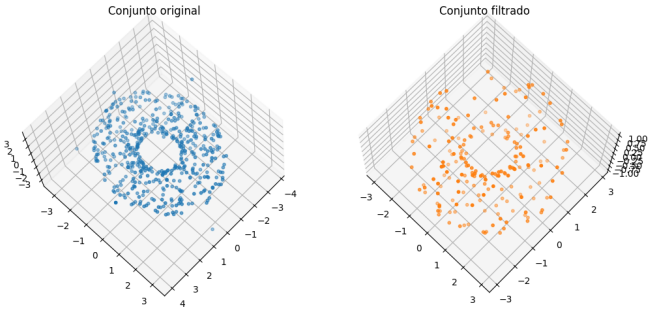


Figura 3.6: Filtrado mediante H_2 con $\varepsilon = 0{,}002$.

De esta forma sí estamos consiguiendo eliminar los outliers y puntos irrelevantes, de forma que nos quedamos con 203 puntos de los 510 originales, eliminado un 60,19 % de los puntos, pero a un costo computacional significativamente superior que comparando mediante H_1, dado que el algoritmo se ve forzado a construir y procesar no solo triángulos, sino también tetraedros, cuyo número potencial en el dataset es un orden polinómico superior. Además, hemos obtenido peores valores de distancia bottleneck entre los diagramas de los grupos de homología de dimensión 1 y 2 que con el filtrado mediante H_1.

3.5. Filtrado mediante H_1 y H_2

Como hemos visto anteriormente, comparar mediante H_1 no es solo la mejor manera de refinar nuestro dataset, ya que hemos conseguido los mejores resultados de las 3, sino que además es sustancialmente más rápido de calcular que mediante H_2 puesto que calcular los diagramas de persistencia de este son mucho más costosos. Sin embargo, si quisiéramos forzar obtener buenos valores tanto para los diagramas de persistencia de los grupos de homología H_1 como para los de H_2, podríamos, en cada iteración, calcular los diagramas de persistencia de ambas dimensiones para cada punto, y comparar con los diagramas originales utilizando los valores $\varepsilon_1 = 0{,}02$ y $\varepsilon_2 = 0{,}002$ para los de dimensión 1 y 2 respectivamente. De esta forma hemos obtenido estos resultados:

$$d_B(D_{O,0}, D_{F,0}) = 2{,}4207$$

$$d_B(D_{O,1}, D_{F,1}) = 0{,}139$$

$$d_B(D_{O,2}, D_{F,2}) = 0{,}0588$$

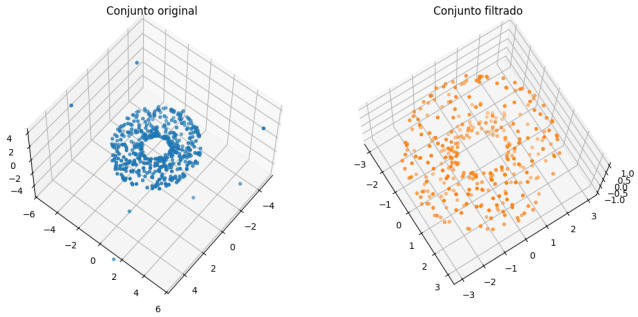


Figura 3.7: Filtrado mediante H_1 y H_2.

Hemos conseguido eliminar 180 puntos de los 510 originales, un 35,2 % del total. Además, hemos obtenido los mejores resultados en cuanto a las distancias bottleneck, lo cual indica que hemos preservado la mayor cantidad de puntos que son esenciales para nuestras características topológicas. Sin embargo, hay que destacar que este filtrado tiene un coste computacional ligeramente superior al anterior, ya que no solo estamos forzados a calcular el diagrama de persistencia hasta dimensión 2, lo cual dispara el coste computacional, sino que además le tenemos que añadir la comparacion bottleneck respecto de ambas dimensiones. Por tanto, aunque sea el más eficiente en cuanto a resultados, es con mucha diferencia el menos eficiente computacionalmente, teniendo este un coste desorbitado.

3.6. Conclusiones

En definitiva, el filtrado por impacto topológico se presenta como una herramienta potente y flexible, que va más allá de la simple eliminación de ruido. Su capacidad para simplificar estructuras complejas manteniendo su forma lo convierte en una técnica prometedora para el pre-procesamiento de datos.

Futuras líneas de trabajo podrían explorar métodos para la selección automática del umbral ε basados en las características intrínsecas del diagrama de persistencia o estrategias para reducir el elevado coste computacional del algoritmo, tales como el uso de submuestreos, la paralelización del proceso de evaluación de puntos o el uso de técnicas de aproximación como los complejos witness.

Bibliografía

[1] F.H. Croom, *Basic Concepts of Algebraic Topology*, Springer, Nueva York, 1978, https://doi.org/10.1007/978-1-4684-9475-4.

[2] A. Hatcher, *Algebraic Topology*, Cambridge University Press, Cambridge, 2002, https://pi.math.cornell.edu/~hatcher/AT/AT.pdf.

[3] J.J. Rotman, *An Introduction to Algebraic Topology*, Springer, Nueva York, 1988 https://doi.org/10.1007/978-1-4612-4576-6.

[4] M.A. Armstrong, *Basic Topology*, Springer, Nueva York, 1983, https://doi.org/10.1007/978-1-4757-1793-8.

[5] G.E. Bredon, *Topology and Geometry*, Springer, Nueva York, 1993, https://doi.org/10.1007/978-1-4757-6848-0.

[6] J.D. Boissonnat, F. Chazal, M. Yvinec, *Geometric and Topological Inference*, Cambridge University Press, Cambridge, 2018, https://doi.org/10.1017/9781108297806.

[7] H. Edelsbrunner, J. Harer, *Computational Topology: An Introduction*, American Mathematical Society, 2010, Disponible en: https://webhomes.maths.ed.ac.uk/~v1ranick/papers/edelcomp.pdf.

[8] A. Zomorodian, G. Carlsson, Computing Persistent Homology, *Discrete and Computational Geometry* **33**, 249-274, 2005, https://doi.org/10.1007/s00454-004-1146-y.

[9] R. Ghrist, Barcodes: The persistent topology of data, *Bulletin of the American Mathematical Society* **45**, 61-75, 2008, http://dx.doi.org/10.1090/S0273-0979-07-01191-3.

[10] N. Otter, M.A. Porter, U. Tillmann, P. Grindrod and H.A. Harrington, A roadmap for the computation of persistent homology, *EPJ Data Science* **6**, 17, 2017, https://doi.org/10.1140/epjds/s13688-017-0109-5.

[11] A. Zomorodian, *Topology for computing*, Cambridge University Press, 2005, https://doi.org/10.1017/CBO9780511546945.