

# Introducción al análisis topológico de datos

Luis Alberto

2025

# Índice general

<b>1. Complejos simpliciales</b>	<b>1</b>
1.1. Notación y nomenclatura estándar . . . . .	1
1.2. Realización geométrica . . . . .	3
1.3. Funciones simpliciales . . . . .	5
1.4. Subdivisiones . . . . .	5
1.5. Teorema de aproximación simplicial . . . . .	10
1.6. Miscelánea de ejercicios . . . . .	12
<b>2. Lema del nervio y el algoritmo MAPPER</b>	<b>14</b>
2.1. Homotopía . . . . .	14
2.2. Lema del nervio . . . . .	17
2.3. Algoritmo MAPPER . . . . .	18
<b>3. Complejos de Čech y Vietoris-Rips</b>	<b>21</b>
3.1. Complejo de Čech . . . . .	21
3.2. Complejo de Vietoris-Rips . . . . .	24
<b>4. Homología persistente</b>	<b>26</b>
4.1. Complejos de cadenas sobre $\mathbb{Z}_2$ . . . . .	26
4.2. Homología simplicial con coeficientes en $\mathbb{Z}_2$ . . . . .	27

# Capítulo 1

## Complejos simpliciales

En este capítulo seguimos [1] y [9] como introducción a los complejos simpliciales. Para cualquier conjunto  $V$ , su conjunto potencia es denotado por  $\mathcal{P}(V)$ . La cardinalidad de  $V$  es  $\#V$ .

**Definición 1.** Un *complejo simplicial* sobre un conjunto  $V$  es un conjunto finito  $K \subseteq \mathcal{P}(V)$  cerrado bajo subconjuntos.

Formalmente, la definición de arriba corresponde a la de los *complejos simpliciales abstractos y finitos* [9, Definition 2.1]; dado que no estudiaremos otro tipo de complejos simpliciales omitimos los otros adjetivos.

**Observación 1.** En la Definición 1 tenemos dos complejos simpliciales que son importantes. El primero es el complejo simplicial *vano* (*void* en inglés), es decir  $\emptyset \subseteq \mathcal{P}(V)$ . El segundo es el complejo simplicial *vacío* (*empty* en inglés):  $\{\emptyset\}$  [9, Remark 2.3].

### 1.1. Notación y nomenclatura estándar

Sea  $K$  un complejo simplicial. Cada elemento de  $K$  se llama *simplejo*<sup>1</sup> (*simplex* y plural *simplices* en inglés). La *dimensión* de  $\sigma$  es  $\dim(\sigma) = \#\sigma - 1$ . Si  $\dim(\sigma) = k$  decimos que  $\sigma$  es un *k-simplejo*. Un *vértice* de un complejo simplicial es un 0-simplejo; El conjunto de  $n$ -simplejos de  $K$  será denotado por  $K_n$ . Para simplificar notación, identificaremos  $K_0$  con  $\bigcup K_0$ , es decir, un vértice de  $K$  será simplemente un elemento de  $V$ .

---

<sup>1</sup>También es común en español, quizás lo es más, utilizar el término *cara*. No lo adopto porque viene del uso geométrico de esa palabra: *las caras del poliedro*. Aunque es verdad que los complejos simpliciales pueden considerarse objetos geométricos, su combinatoria es más natural en mi opinión. Véanse [9], [11]

La dimensión de  $K$  es  $\dim(K) = \max\{\dim(\sigma) \mid \sigma \in K\}$ . Una *faceta* (o *careta* si se usa cara) de un complejo simplicial es un simplejo maximal con respecto a la contención; es decir,  $\sigma$  es faceta si  $\sigma \subseteq \tau$ , entonces  $\tau = \sigma$ . Diremos que un complejo simplicial es *puro* siempre que todas sus facetas tengan la misma dimensión.

Los complejos simpliciales son objetos combinatorios. Usualmente se representan gráficamente como sigue. Un punto representa un vértice, las aristas, 1-simplejos; los triángulos representan 2-simplejos, etc. En la Figura 1.1 encontramos ejemplos de simplejos y en la Figura 1.2, de complejos simpliciales. Nótese que el triángulo hueco en la Figura 1.2b no es un simplejo. También, es importante notar que del dibujo solo es imposible determinar si el tetraedro de la Figura 1.1c es un 3-simplejo o es la unión de 4 2-simplejos. Es por esto último que sólo representamos con fines ilustrativos.

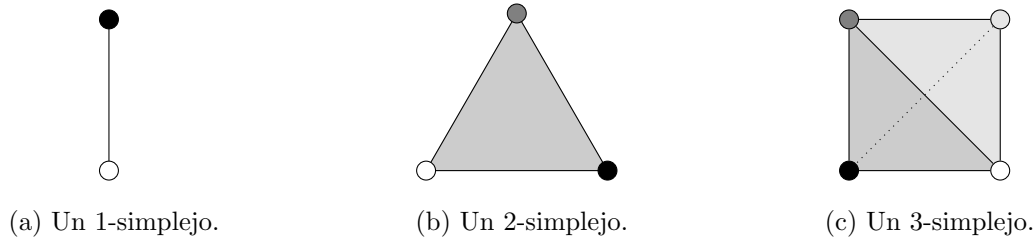


Figura 1.1: Representación gráfica de algunos simplejos. el tetraedro en (c) es un sólido 3-dimensional.

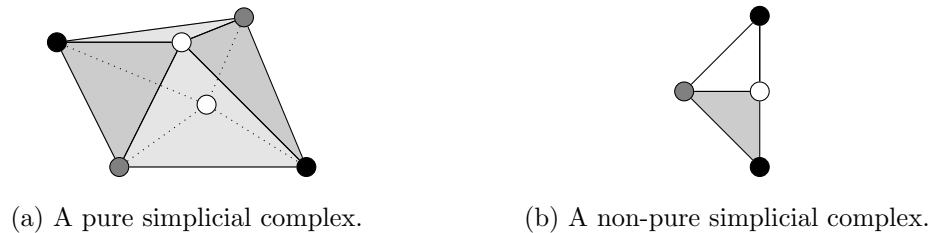


Figura 1.2: Examples of simplicial complexes. All triangles in (a) are filled but the simplicial complex itself is not a 3-dimensional solid.

Sea  $\sigma$  un simplejo. Vamos a sobrecargar la notación y usaremos  $\sigma$  para denotar al complejo simplicial  $\mathcal{P}(\sigma)$  como a la única faceta de tal complejo simplicial. En este sentido, todo simplejo es un complejo simplicial.

**Ejercicio 1.** ¿Cuántos  $k$ -simplejos tiene un  $n$ -simplejo?

El *complejo simplicial generado* por  $F \subseteq \mathcal{P}(V)$  es el complejo simplicial mínimo con respecto a contener todos los simplejos  $\sigma \in F$ . Tal complejo simplicial es denotado por

$$\bigcup F = \bigcup_{\sigma \in F} \sigma.$$

**Ejercicio 2.** Muestra que todo complejo simplicial es generado por sus facetas.

Un *subcomplejo* de un complejo simplicial  $K$  es un subconjunto de  $K$  que es un complejo simplicial. El  *$n$ -esqueleto* de  $K$  es el subcomplejo de  $K$  generado por todos los  $k$ -simplejos en  $K$  con  $k \leq n$  y es denotado por  $\text{skel}_n(K)$ .

**Ejercicio 3.** Muestra que en general  $\text{skel}_n(K)$  no es puro de dimensión  $n$ ; es decir,  $\text{skel}_n(K)$  puede contener facetas de dimensión  $k < n$ .

**Ejercicio 4.** Muestra que  $\text{skel}_n(K) = K_n$  si y sólo si  $n < -1$  o  $K$  es vano. ¿Qué pasaría con esta equivalencia si en la Definición 1 eliminamos los complejos simpliciales vacío y vano?

Un *simplejo frontera* de un  $k$ -simplejo  $\sigma$ , es un  $(k-1)$ -simplejo  $\tau \in \sigma$ . La *frontera* de  $\sigma$  es  $\partial\sigma = \bigcup \{\tau \mid \tau \text{ es un simplejo frontera de } \sigma\}$ .

**Ejercicio 5.**  $\partial\sigma$  es un complejo simplicial.

## 1.2. Realización geométrica

Intuitivamente, el complejo simplicial de la Figura 1.2a es una 2-esfera. Esta sección formaliza tal intuición. Usamos  $\mathbb{R}^V$  para denotar el  $\mathbb{R}$  espacio vectorial con base  $V$ , en particular,  $\mathbb{R}^n$  es el espacio euclidiano de  $n$  dimensiones. Siempre consideramos  $\mathbb{R}^V$  como espacio topológico con la topología inducida por la métrica euclidiana.

**Definición 2.** Un *conjunto afinmente independiente* en  $\mathbb{R}^n$  es un conjunto  $A = \{v_0, \dots, v_k\}$  tal que para cada  $t_i, s_i \geq 0$  con

$$\sum_{i=0}^k s_i = \sum_{i=0}^k t_i = 1,$$

si

$$\sum_{i=0}^k s_i v_i = \sum_{i=0}^k t_i v_i,$$

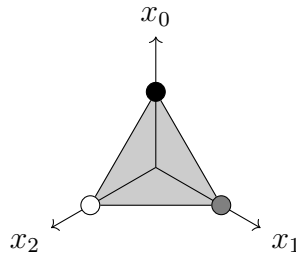


Figura 1.3: El  $V$ -simplejo estándar para  $V = \{x_0, x_1, x_2\}$ .

entonces  $s_i = t_i$  para cada  $i$ . Un vector  $x \in \mathbb{R}^n$  es una *combinación afín* de  $A$  si  $x = \sum_{i=0}^k s_i v_i$  con  $\sum_{i=0}^k s_i = 1$  y  $s_i \geq 0$  para cada  $i$ . La *cápsula convexa* (*convex hull* en inglés) de  $A$  es el conjunto de todas las combinaciones afines de  $A$ .

**Ejercicio 6.** Demuestra que todo conjunto linealmente independiente de  $\mathbb{R}^n$  es afín independiente.

Tomamos las definiciones 3 y 4 de [9, Section 2.2.1].

**Definición 3.** Sea  $V$  un conjunto finito. El  $V$ -simplejo estándar es la cápsula convexa de la base ortonormal estándar de  $\mathbb{R}^V$ .

En la Figura 1.3, tenemos el  $V$ -simplejo estándar para  $V = \{x_0, x_1, x_2\}$ .

**Definición 4.** Sea  $K$  un complejo simplicial. Consideremos la unión de todos los  $\sigma$ -simplejos estándar en  $\mathbb{R}^{K_0}$  con  $\sigma \in K$ . Este conjunto con la topología de subespacio de  $\mathbb{R}^{K_0}$  es la *realización geométrica estándar* de  $K$ . Cualquier espacio homeomorfo a la realización geométrica estándar de  $K$  como  $|K|$  y la llamaremos la *realización geométrica* de  $K$ .

**Observación 2.** La realización geométrica es definida mediante los  $V$ -simplejos estándar. Como ellos tienen la topología de subespacio de  $\mathbb{R}^V$  siempre supondremos que  $|K|$  tiene una métrica que le induce la topología adecuada. También, cualquier afirmación topológica hecha sobre un complejo simplicial estará hecha sobre su realización geométrica..

**Ejercicio 7.** Sea  $K$  un complejo simplicial. En este ejercicio veremos cómo podemos estudiar complejos simpliciales infinitos y sus realizaciones geométricas. La equivalencia de la última parte sirve como definición de la topología de la realización geométrica de complejos simpliciales infinitos [16, p. 8].

1. Prueba que si  $\sigma \in K$ , entonces  $|\sigma| \subseteq |K|$ .

2. Muestra que la realización geométrica de cualquier simplejo es un espacio compacto.
3. Demuestra que si  $\sigma \in K$ , entonces  $|\sigma|$  es un conjunto cerrado de  $|K|$ .
4. Demuestra que  $C \subseteq |K|$  es cerrado si y solo si  $C \cap |\sigma|$  es cerrado para cada  $\sigma \in K$ .

### 1.3. Funciones simpliciales

Si  $f: X \rightarrow Y$  es una función, usamos  $f(A)$  para denotar la imagen de  $A \subseteq X$  bajo  $f$ .

**Definición 5.** Dados dos complejos simpliciales  $K$  and  $L$ , una *función simplicial*<sup>2</sup> de  $K$  a  $L$  es una función  $f: K_0 \rightarrow L_0$  tal que si  $\sigma \in K$  entonces  $f(\sigma) \in L$ .

**Observación 3.** Toda función simplicial  $f: K_0 \rightarrow L_0$  induce una función  $f: K \rightarrow L$ . Así, cuando decimos que  $f: K \rightarrow L$  es una función simplicial, lo que queremos decir es que es inducida por un mapa simplicial.

**Ejercicio 8.** Demuestre que la clase de complejos simpliciales  $\mathcal{CS}$  junto con las funciones simpliciales es una categoría; es decir, demuestre que la composición de funciones simpliciales es simplicial y que para cada  $K \in \mathcal{CS}$  existe una función simplicial  $1_K: K \rightarrow K$  tal que  $1_K \circ f = f$  y  $g \circ 1_K = g$  para cualesquiera funciones simpliciales  $f: L \rightarrow K$  y  $g: K \rightarrow L$ .

**Ejercicio 9.** ¿Quiénes son  $1_\emptyset$  y  $1_{\{\emptyset\}}$ ? ¿Tu prueba del ejercicio anterior consideró esto?

**Ejercicio 10.** Muestra que no toda función entre los vértices de dos complejos simpliciales induce una función simplicial.

**Ejercicio 11.** Muestra que toda función simplicial  $f: K \rightarrow L$  induce una función continua  $|f|: |K| \rightarrow |L|$  dada por

$$|f|(\sum_{i=0}^k s_i v_i) = \sum_{i=0}^k s_i f(v_i).$$

### 1.4. Subdivisiones

**Definición 6.** Una *subdivisión* de un complejo simplicial  $K$  es un complejo simplicial  $K'$  tal que para una realización geométrica fija  $|K|$ , existe una realización geométrica  $|K'|$  tale que:

---

<sup>2</sup>Es usual en la literatura en inglés encontrar esto como *simplicial map*. Yo mismo les llamo así a veces.

1. La realización geométrica de cada simplejo en  $K'$  está contenida en la realización geométrica de algún simplejo de  $K$ .
2. Para cada  $\sigma \in K$ , hay un subcomplejo  $\sigma'$  de  $K'$  tal que  $|\sigma| = |\sigma'|$ .

Los siguientes ejercicios están diseñados para una mejor y pronta comprensión de la definición de arriba.

**Ejercicio 12.** Si  $K'$  es una subdivisión de  $K$ , entonces  $|K| \cong |K'|$ .

**Ejercicio 13.** Representa gráficamente los siguientes complejos simpliciales

1.  $K = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$ .
2.  $K' = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, c\}, \{c, b\}\}$

Muestra que  $K'$  es una subdivisión de  $K$ .

**Ejercicio 14.** Muestra que la imagen en la Figura 1.4 es una subdivisión de un 2-simplejo. Para ello usa la siguiente estrategia. Sean  $K$  el 2-simplejo generado por  $a, b, c$  y  $E$  su realización geométrica estándar.

1. Encuentra el baricentro de la realización geométrica de todos los simplejos no vacíos de  $K$ . Todos ellos deben tener coordenadas respecto a la base estándar  $\{a, b, c\}$  de  $\mathbb{R}^{\{a,b,c\}}$ . Aquí puedes optar por lo que te confunda menos, usar  $e_x$  o  $x$  para referirte al básico de  $\mathbb{R}^\beta$  que corresponde a  $x \in \beta$ , pero hazlo consistentemente durante el resto de los incisos.
2. Sea  $B$  el conjunto del baricentro de todos los simplejos no vacíos de  $K$ , sabemos que  $\#B = \#K - 1$ . Sea  $f: K \setminus \emptyset \rightarrow B$  la función que a cada simplejo le asocia su baricentro. Es claro que esta función es biyectiva. Sea  $\Delta(B, f)$  el complejo simplicial sobre  $B$  definido mediante  $\sigma \in \Delta(B, f)$  si y sólo si  $\sigma = f[\tau]$  para algún  $\tau \in \text{Bar}(K)$ . Demuestra que  $\Delta(B, f) \cong \text{Bar}(K)$ . Pista, un isomorfismo es  $f$ .
3. Sean  $R$  la realización geométrica estándar de  $\Delta(B, f)$  e  $i: B \rightarrow \mathbb{R}^{\{a,b,c\}}$  la inclusión. Si estás considerando  $e_x$  como el básico estándar asociado a  $x$ , sean  $\bar{B}$  la base estándar de  $\mathbb{R}^B$  e  $i_e: \bar{B} \rightarrow \mathbb{R}^{\{a,b,c\}}$  definida como  $i_{e_x} = x$ . Esta función está bien definida y es consistente con tu notación pues

$$i_e(e_{t_a e_a + t_b e_b + t_c e_c}) = t_a e_a + t_b e_b + t_c e_c$$



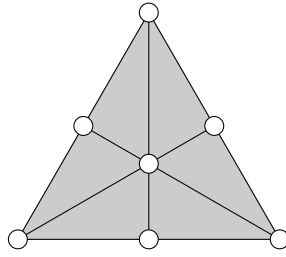


Figura 1.4: La primera subdivisión baricéntrica de un 2-simplejo.

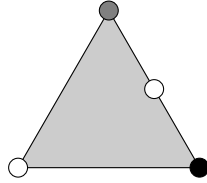


Figura 1.5: Esto no es una subdivisión de un 2-simplejo.

. Sea  $g$  la extensión lineal de  $i$  (o  $i_e$ ), así  $g$  es continua. Muestra que  $g[R] = E$  y de hecho es un homeomorfismo. Pista: si  $x \in E$ , entonces  $x$  es un elemento de la cápsula convexa de a lo más 3 elementos de  $B$ . Basta tomarse los tres más cercanos, si hay 4 más cercanos, entonces elimina de ellos los dos más lejanos (esta situación implica que  $x$  está en una de las mediatrices del  $E$ ). Las coordenadas de  $x$  en esta nueva cápsula convexa definen una función continua de  $E$  en  $R$  y es inversa de  $g|_R$ .

4. Para terminar, necesitas probar los dos puntos de la definición de subdivisión. Para ello, considera  $E$  como la realización de ambos complejos simpliciales. Para el punto 1, para  $\sigma' \in \Delta(B, f)$  toma cualquier punto interior de  $|\sigma'|$ . Ese punto se encuentra en la realización geométrica de exactamente un simplejo de  $K$  pues  $g[\sigma'] \subseteq \sigma$  para un único  $\sigma \in K$  con  $\dim(\sigma') = \dim(\sigma)$  (demuestra esto último). Para el punto 2, muestra que para  $\sigma \in K$  basta tomar  $\sigma' \leq \Delta(B, f)$  como el complejo simplicial generado por todos los baricentros de  $|\sigma|$ .

Intuitivamente, una subdivisión de  $K$  se obtiene agregando simplejos internos a (algunos de) los simplejos de  $K$ . Sin embargo, esto debe hacerse con cuidado. En la Figura 1.5 vemos que este procedimiento fue realizado sin **preservar la estructura combinatoria** del complejo simplicial original. Esto quiere decir que aunque el espacio topológico asociado al objeto en esa figura (es decir el triángulo) es la realización geométrica del 2-simplejo, como objeto combinatorio no es un complejo simplicial.

**Ejercicio 15.** Demuestra que si  $L$  es una subdivisión de  $K$  y  $K$  lo es de  $L$ , entonces existe un isomorfismo  $K \cong L$ .

*Prueba del Ejercicio 15.* Por el Ejercicio 12 supondremos que la realización geométrica de la Definición 6 es la realización geométrica estándar de  $K$ .

Por el segundo punto de la Definición 6, sabemos que para cada  $\sigma \in K$  existe un subcomplejo simplicial  $f(\sigma) \leq L$  tal que

$$|\sigma| = |f(\sigma)|. \quad (1.1)$$

Por el primer punto de la Definición 6, para cada  $\tau' \in f(\sigma)$ , existe  $\tau \in K$  tal que  $|\tau'| \subseteq |\tau|$ . De la Ecuación 1.1, podemos elegir  $\tau \subseteq \sigma$ .

A su vez, existe  $g(\tau) \leq K$  tal que

$$|\tau| = |g(\tau)|. \quad (1.2)$$

Es decir,  $g(\tau)$  es un subcomplejo de  $\sigma$ . Esto es válido para cualquier  $\tau \in f(\sigma)$ .

**Observación 4.** Recordemos que  $|\sigma|$  es la realización geométrica estándar de  $\sigma$  por nuestra suposición al inicio de la demostración. Por la Ecuación 1.1, existe  $\tau \in f(\sigma)$  tal que  $\dim(\tau) = \dim(\sigma)$ , ya que cualquier punto en el interior de  $|\sigma|$  se encuentra en la cápsula convexa de un conjunto afínmente independiente con  $\#\sigma$  puntos.

Tomemos  $\bar{\tau} \in f(\sigma)$  tal que  $\dim(\bar{\tau}) = \dim(\sigma)$ . Así,  $K(\bar{\tau})$  es un subcomplejo de  $\sigma$  que contiene, por la Observación 4, un simplejo cuya dimensión iguala la de  $\sigma$ . Esto sólo es posible si

$$g(\bar{\tau}) = \sigma. \quad (1.3)$$

Sustituyendo la Ecuación 1.3 en las ecuaciones 1.1 y 1.2, concluimos que

$$|\bar{\tau}| = |f(\sigma)|.$$

Dado que  $\bar{\tau} \in f(\sigma)$ , la ecuación anterior solo es posible si  $\bar{\tau} = f(\sigma)$ .

Dado que este argumento es simétrico, concluimos que  $g(\tau)$  debe ser un simplejo también y por la Ecuación 1.3, tanto  $g$  como  $f$  son funciones biyectivas y de hecho de aquí son inversas entre sí.

Para terminar, demostraremos que  $f$  y  $g$  son funciones simpliciales. Primero, notemos que  $\dim(f(\sigma)) = \dim(\sigma)$  (análogamente para  $g$ ). Por lo tanto, si  $v \in K_0$ , entonces  $f(v) \in L_0$ .

Luego, por el Ejercicio 7, podemos concluir que si  $\sigma \subseteq \sigma'$ , entonces  $f(\sigma) \subseteq f(\sigma')$  (análogamente para  $g$ ). Esto implica que  $f[\sigma] \in L$ , es más  $f[\sigma] = f(\sigma)$  (análogamente para  $g$ ). Con esto concluimos que  $K \cong L$ .  $\square$

Aunque existen diferentes subdivisiones generadas con un método, nosotros solamente estudiaremos la subdivisión baricéntrica. Para ello necesitamos algunas definiciones relacionadas a órdenes parciales.

**Definición 7.** Un *conjunto parcialmente ordenado* (abreviado *orden parcial*) es una pareja  $(P, <)$  donde  $P$  es un conjunto y  $<$  es una relación binaria sobre  $P$  que es transitiva, irreflexiva y asimétrica. Simplificaremos esto y diremos que  $P$  es un orden parcial y  $<$  es el *orden* de  $P$ . Una *cadena*  $\sigma$  en un orden parcial  $P$  es subconjunto de  $P$  tal que la restricción de  $<$  a  $\sigma$  satisface *tricotomía*, es decir  $(C, <)$  es un *conjunto totalmente ordenado* (abreviado *orden total*).

Aunque no profundizaremos en la teoría de órdenes parciales, para poder entender más fácilmente qué es el complejo de orden de un orden parcial, usaremos la representación gráfica de los órdenes parciales.

**Definición 8.** Sean  $P$  un orden parcial y  $x < y$  en  $P$ . Decimos que  $y$  *cubre*  $x$  si no existe  $z \in P$  tal que  $x < z < y$ . El *diagrama de Hasse*  $\mathcal{H}$  de  $P$  es la gráfica con conjunto de vértices  $P$  y aristas definidos por pares  $x, y$  tales que  $x$  cubre  $y$  o viceversa.

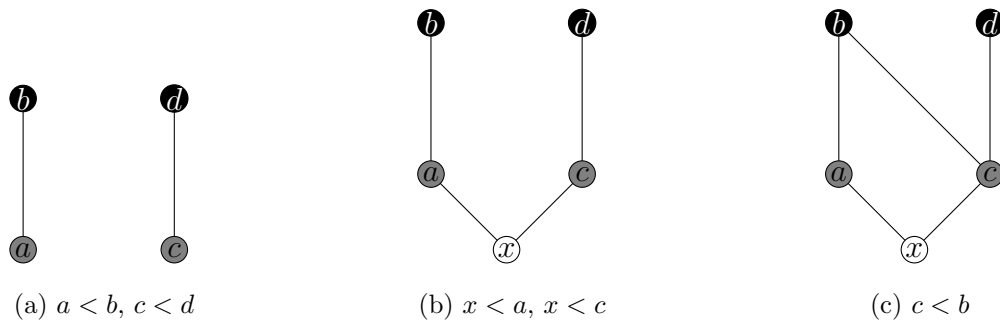


Figura 1.6: Diagramas de Hasse de tres órdenes parciales. En (a) tenemos dos relaciones de cobertura solamente. Agregamos dos relaciones más y obtenemos (b). Con una relación más se obtiene (c)

**Definición 9.** El *complejo de orden* de un orden parcial  $P$  es el complejo simplicial  $\Delta(P)$  generado por las cadenas maximales de  $P$ . En otras palabras  $\sigma \subseteq P$  es un simplejo de  $\Delta(P)$  si y solo si es un orden total con el orden heredado de  $P$ .

**Ejercicio 16.** Dibuja los complejos de orden de los órdenes de la Figura 1.6.

**Definición 10.** La *primera subdivisión baricéntrica* de un complejo simplicial  $K$  es el complejo de orden de  $K \setminus \{\emptyset\}$  visto como orden parcial y lo denotamos mediante  $\text{Bar}(K)$ . La  $N$ -ésima *subdivisión baricéntrica* de  $K$  es la primera subdivisión baricéntrica de  $\text{Bar}^{N-1}(K)$ .

En otras palabras, los simplejos de  $\text{Bar}(K)$  son cadenas de simplejos en  $K$ . En la Figura 1.4 encontramos la forma usual de representar gráficamente una subdivisión baricéntrica. El vértice central, el baricentro del triángulo, corresponde a la única faceta del 2-simplejo. En general, el baricentro de un simplejo corresponde a su única faceta.

## 1.5. Teorema de aproximación simplicial

El teorema de aproximación simplicial nos asegura que dada cualquier función continua  $f$  entre las realizaciones geométricas de dos complejos simpliciales, podemos encontrar una función simplicial de una subdivisión del dominio de  $f$  al codominio de  $f$  que se “parece” a  $f$ . Por supuesto, para formalizar esto necesitamos varias definiciones.

**Definición 11.** El *diámetro*  $\text{diam}(X)$  de un conjunto  $X \subseteq \mathbb{R}^k$  es  $\sup\{d(x, y) \mid x, y \in X\}$ .

Aunque no hemos definido qué es una operación para obtener subdivisiones, intuitivamente, son métodos que permiten subdividir cualquier simplejo iterativamente. Un ejemplo son las subdivisiones baricéntricas.

**Definición 12.** Una operación para obtener subdivisiones  $\text{Div}$  *refina* si para cada  $\epsilon > 0$  y cada complejo simplicial  $K$  existe  $N \in \mathbb{N}$  tal que

$$\sup\{\text{diam}(\sigma) \mid \sigma \in \text{Div}^N(K)\} < \epsilon.$$

La prueba del teorema de aproximación simplicial necesita el número de Lebesgue de una cubierta de espacio compacto.

**Definición 13.** Sean  $C$  un subespacio compacto de  $\mathbb{R}^k$  y  $\mathcal{U}$  una cubierta abierta de  $C$ . Decimos que  $\lambda > 0$  es un *número de Lebesgue* de  $\mathcal{U}$  si para cada  $x \in C$  y  $\epsilon < \lambda$  la bola abierta  $B_\epsilon(x)$  está contenida en algún elemento de  $\mathcal{U}$ .

**Definición 14.** La *estrella abierta* de un vértice  $v$  de un complejo simplicial  $K$  es  $\text{St}(v) = \bigcup_{v \in \sigma} |\sigma|^\circ$ .

**Ejercicio 17.** Demuestra que  $\bigcap_{v \in \sigma} \text{St}(v) \neq \emptyset$  si y solo si  $\sigma \in K$

Recordemos que nuestros complejos simpliciales son finitos. Una versión general del siguiente teorema puede encontrarse en [16, Theorem 16.5]

**Teorema 1** (Aproximación simplicial). Sea  $f: |K| \rightarrow |L|$  una función continua entre las realizaciones geométricas de dos complejos simpliciales. Si  $\text{Div}$  refina, entonces existen  $N \in \mathbb{N}$  y una función simplicial  $\bar{f}: \text{Div}^N(K) \rightarrow L$  tales que  $f(\text{St}(v)) \subseteq \text{St}(\bar{f}(v))$  para cada  $v \in S_0(\text{Div}^N(K))$ .

*Demostración.* Supongamos que tenemos métricas para  $|K|$  y  $|L|$  que nos dan la topología adecuada. Las estrellas abiertas  $\text{St}(w)$  con  $w \in L_0$  forman una cubierta abierta de  $|L|$ . Sea  $\lambda$  un número de Lebesgue de esa cubierta. Dado que  $f$  es continua y  $|K|$  es compacto existe  $\delta > 0$  such that  $f(B_\delta(x)) \subseteq B_\lambda(f(x)) \subseteq \text{St}(w)$  para algún  $w \in L_0$ . Elijamos  $N$  tal que  $\sup\{\text{diam}(\sigma) \mid \sigma \in \text{Div}^N(K)\} < \frac{\delta}{2}$ . Definamos  $\bar{f}: \text{Div}^N(K) \rightarrow L$  mediante  $\bar{f}(v) = w$  para algún  $w$  tal que  $f(\text{St}(v)) \subseteq \text{St}(w)$ . Este es una función simplicial bien definida por Ejercicio 17.  $\square$

Finalmente, tenemos que la subdivisión baricéntrica refina [16, Theorem 15.4]; por tanto, podemos usar el teorema anterior con subdivisiones baricéntricas.

**Proposición 1.** La subdivisión baricéntrica refina.

*Demostración.* Sea  $\sigma$  un  $k$ -simplejo. Por definición  $(\text{Bar}(\sigma))_0 = \{\tau \mid \tau \in \sigma\}$ . Identificamos  $\tau$  con  $\sum_{v \in \tau} \frac{v}{\#\tau} \in |\sigma|$ , en otras palabras  $\tau$  es el baricentro de  $|\tau|$ .

Para cada  $v \in \sigma$ ,

$$d(\sigma, v) = d\left(\sum_{w \in \sigma} \frac{1}{k+1} w, \sum_{w \in \sigma'} \frac{1}{k+1} v\right) \quad (1.4)$$

$$\leq \sum_{w \in \sigma} d\left(\frac{1}{k+1} w, \frac{1}{k+1} v\right) \quad (1.5)$$

$$= \frac{1}{k+1} \sum_{w \in \sigma} d(w, v) \quad (1.6)$$

$$\leq \frac{k}{k+1} \sup\{d(w, v) \mid w \in \sigma\} \quad (1.7)$$

$$= \frac{k}{k+1} \text{diam}(\sigma). \quad (1.8)$$

Dado que  $\text{diam}(\sigma)$  se alcanza para algún par  $v, w \in \sigma$ , de los cálculos de arriba concluimos que  $\sup\{\text{diam}(\tau) \mid \tau \in \text{Bar}(\sigma)\} \leq \frac{k}{k+1} \text{diam}(\sigma)$ . Debido a que  $\frac{k}{k+1} < 1$  la proposición queda probada.  $\square$

**Ejercicio 18.** Muestra con todo detalle la Proposición 1.

## 1.6. Miscelánea de ejercicios

En estos ejercicios vamos a explorar con mayor generalidad algunos de los conceptos de este capítulo

**Definición 15.** Un *complejo simplicial infinito* sobre un conjunto  $V$  es un subconjunto  $K \subseteq \mathcal{P}(V)$  cuyos elementos son todos finitos y es cerrado bajo subconjuntos.

La mayoría de los conceptos que se revisaron en este capítulo se pueden generalizar de manera directa para complejos simpliciales infinitos.

**Ejercicio 19.** Demuestra que todo complejo simplicial, no necesariamente finito, es generado por sus facetas.

**Definición 16.** Un  *$n$ -simplejo geométrico* es la cápsula convexa de un conjunto afinmente independiente de cardinalidad  $n + 1$ . Una *cara* de un simplejo geométrico  $\sigma$  generado por  $A$  es un simplejo geométrico generado por un subconjunto de  $A$ .

**Definición 17.** Un *complejo simplicial geométrico*,  $\tilde{K}$  en  $\mathbb{R}^N$  es una colección de simplejos geométricos en  $\mathbb{R}^N$  tales que

1.  $\tilde{K}$  es cerrado bajo caras.
2. La intersección de cualesquiera dos simplejos geométricos en  $\tilde{K}$  es un simplejo geométrico.

**Ejercicio 20.** Prueba que la realización geométrica estándar de un complejo simplicial es un complejo simplicial geométrico. (No necesitas la topología de la realización geométrica para probar esto).

**Ejercicio 21.** Da un ejemplo de una colección de simplejos geométricos que no sea un complejo simplicial geométrico porque falla el segundo punto de la definición. Concluye que en general es falso que la unión de realizaciones geométricas de los simplejos de un complejo simplicial  $K$  es la realización geométrica de  $K$ .

**Definición 18.** Una *cubierta* de un conjunto  $X$  es una familia de subconjuntos de  $X$ , digamos  $\mathcal{U} \subseteq \mathcal{P}(X)$ , tal que  $X = \bigcup \mathcal{U}$ .

**Definición 19.** El *nervio* de una cubierta  $\mathcal{U}$  de  $X$  es

$$\mathcal{N}(\mathcal{U}) = \{S \subseteq \mathcal{U} \mid \emptyset \neq \bigcap S \text{ y } \#S \in \mathbb{N}\}.$$

Para que esta definición sea consistente, debemos tener cuidado, pues si  $S = \emptyset$ , entonces  $\bigcap S$  no está definido. Así, hay que agregar  $\emptyset$  a la definición del nervio (considerando el complejo simplicial generado) o acordar que  $\bigcap \emptyset = X$ . En cualquier caso, tenemos el siguiente ejercicio.

**Ejercicio 22.** Prueba que el nervio de una cubierta es un complejo simplicial.

**Definición 20.** Una *cubierta con índices* de un conjunto  $X$  es una familia de subconjuntos de  $X$  indicada por un conjunto  $\mathcal{J}$ , digamos  $\mathcal{U} = (U_j)_{j \in \mathcal{J}}$ , tal que  $X = \bigcup_{j \in \mathcal{J}} U_j$ .

**Definición 21.** El *nervio* de una cubierta indicada  $\mathcal{U}$  de  $X$  es

$$\mathcal{N}(\mathcal{U}) = \{J \subseteq \mathcal{J} \mid \emptyset \neq U_J \text{ y } \#J \in \mathbb{N}\}$$

donde  $U_J = \bigcap_{j \in J} U_j$ . Aquí sucede lo mismo que con la Definición 19.

**Ejercicio 23.** Es claro que el nervio de una cubierta indicada es un complejo simplicial. También es claro que si olvidamos los índices de una cubierta indicada, obtenemos una cubierta. Da un ejemplo de una cubierta indicada  $(U_j)_{j \in \mathcal{J}}$  tal que la cubierta obtenida al olvidar los índices tenga un nervio no isomorfo al de la cubierta indicada original.

# Capítulo 2

## Lema del nervio y el algoritmo MAPPER

En este capítulo presentaremos una versión del Lema del Nervio. Este resultado es multifacético en el sentido de que hay diferentes enunciados no equivalentes que relacionan homotópicamente un espacio con un complejo simplicial. Cada uno de estos teoremas dice que algunas propiedades topológicas de un espacio son compartidas por un complejo simplicial que se obtiene de una cubierta del espacio original.

Podríamos tener, por tanto, un curso completo para demostrar algunas de las variantes de este lema; sin embargo, aquí no lo demostraremos. La razón para no proveer una prueba de este resultado es que cualquiera de sus enunciados requiere material de topología algebraica que no vamos a revisar y que no es trivial. A pesar de ello, para entender su enunciado necesitamos diferentes conceptos que ilustraremos antes. Referencias muy buenas para profundizar en este tema son las siguientes: [20, 16, 17].

### 2.1. Homotopía

El material presentado aquí no debe considerarse una introducción a este tema. Lo único que presentamos son las definiciones y algunos ejercicios para poder entender intuitivamente el Lema del Nervio. Para una comprensión más amplia recomendamos [20] o [17].

**Definición 22.** Sean  $X$  un espacio topológico,  $x, y \in X$  y  $I = [0, 1] \subseteq \mathbb{R}$ . Una *trayectoria* en  $X$  de  $x$  a  $y$  es una función continua  $u: I \rightarrow X$  tal que  $u(0) = x$  y  $u(1) = y$ . En este caso decimos que



$x$  es *conectable por trayectorias* con  $y$

**Ejemplo 1.** La realización geométrica de una gráfica  $K$  es conectable por trayectorias. Dados cualesquiera dos puntos en la realización geométrica o bien son vértices, o están en el interior de una arista o uno es un vértice y el otro está en el interior de una arista.

Si los puntos son vértices, sabemos que existe un camino entre ellos en la gráfica. Así basta con parametrizar la realización geométrica de este camino para obtener una trayectoria en  $|K|$ .

Si uno de ellos, digamos  $x$ , está en el interior de una arista  $e$ , podemos replicar el argumento anterior con el otro punto y cualquiera de los vértices de  $e$ . Finalmente, parametrizar la curva entre el último vértice y  $x$ .

El otro caso es similar.

**Ejercicio 24.** Usa lo siguiente para demostrar que ser conectable por trayectorias es una relación de equivalencia.

- La función constante es una trayectoria.
- La trayectoria inversa  $u^-$  de  $u$  es la composición  $u \circ \text{inv}$  donde  $\text{inv}(t) = 1 - t$  para cada  $t \in [0, 1]$ .
- La función que manda  $t$  a  $2t$  puede usarse para reparametrizar cualquier trayectoria de manera que la traza de  $u$  se recorra en la mitad del tiempo.

Las clases de equivalencia bajo esta relación se llaman las componentes conexas por trayectorias de  $X$ .

**Definición 23.** El conjunto de las componentes conexas por trayectorias de  $X$  es denotado por  $\pi_0(X)$ . Decimos que  $X$  es *0-conexo* si  $\#\pi_0(X) = 1$ .

**Definición 24.** Sean  $X$  y  $Y$  dos espacios topológicos. Dos funciones continuas  $f, g: X \rightarrow Y$  son *homotópicas* ( $f \simeq g$ ) si existe una *homotopía*  $H$  de  $f$  a  $g$ , es decir, si existe una función continua  $H: X \times I \rightarrow Y$  tal que las siguientes ecuaciones entre restricciones de  $H$  se cumplen:  $H|_{X \times \{0\}} = f$  y  $H|_{X \times \{1\}} = g$ . Usualmente escribimos  $H_t = H|_{X \times \{t\}}$ .

**Ejemplo 2.** Sean  $f, g: I \rightarrow \mathbb{R}^2$  definidas como  $f(s) = (s, s^2)$  y  $g(s) = (s, s)$ . Definamos  $H: I \times I \rightarrow \mathbb{R}^2$  como  $H(t, s) = (s, ts^2 + (1-t)s)$ .  $H$  es continua pues sus funciones coordenadas son continuas en ambas variables. Luego  $H_0 = g$  y  $H_1 = f$ . Por lo tanto,  $f \simeq g$ . Intuitivamente,  $H$  está deformando la gráfica de la identidad en la de la función  $x \mapsto x^2$  en el intervalo  $I$ .

**Definición 25.** Una homotopía como en el ejemplo anterior se llama *homotopía lineal*.

**Ejercicio 25.** Muestra, como en el Ejercicio 24, que ser homotópicas es una relación de equivalencia en las funciones continuas entre dos espacios topológicos. Formalmente debes tener cuidado puesto que las homotopías son funciones de dos variables, por lo que la continuidad se debe verificar utilizando la topología del producto: Un argumento de cálculo multivariado es suficiente.

**Ejercicio 26.** Si  $f \simeq f'$  y  $g \simeq g'$  con  $f: X \rightarrow Y$  y  $g: Y \rightarrow Z$ , entonces  $g \circ f \simeq g' \circ f'$ . Como antes, un argumento de cálculo multivariado ayudará a mostrar que la función que propones es continua.

**Definición 26.** Un *inverso homotópico* de una función continua  $f: X \rightarrow Y$  es una función continua  $g: Y \rightarrow X$  tal que  $f \circ g$  y  $g \circ f$  son homotópicas a la identidad. En tal caso, diremos que  $f$  es una *equivalencia homotópica* y  $X$  y  $Y$  serán llamados *homotópicamente equivalentes* o *del mismo tipo de homotopía*. Si  $X$  es homotópicamente equivalente a un punto, diremos que es *contráctil*.

**Ejercicio 27.** Muestra que  $X$  es contráctil si y solo sí,  $1_X$  es homotópica a una constante.

**Ejercicio 28.** Un *anillo* es el conjunto de puntos que se encuentran entre dos círculos concéntricos; en otras palabras un conjunto de la forma

$$A = \{(x, y) \mid r \leq \|(x, y)\| \leq R\}$$

Con este ejercicio demostrarás que un anillo es homotópicamente equivalente a  $S^1$ . Para ello supondremos que tenemos un anillo como arriba en el que  $r = 1$ .

1. Considera la función  $H: A \times I \rightarrow S^1$  definida mediante  $H(x, y, t) = \frac{t(x, y)}{\|(x, y)\|} + (1 - t)(x, y)$ . Demuestra que esta función es una homotopía entre  $1_A$  y  $H_1$ .
2. Si  $p = H_1$  y  $i: S^1 \rightarrow A$  es la inclusión, demuestra que  $p \circ i = 1_{S^1}$ .
3. Muestra que existe una homotopía  $\bar{H}$  entre  $i \circ p$  y  $1_{S^1}$ . Concluye que  $S^1$  es homotópicamente equivalente a  $A$  y por tanto,  $A$  no es contráctil.

**Ejercicio 29.** Muestra que todo subconjunto convexo de  $\mathbb{R}^n$  es contráctil. Pista: puedes suponer que el origen está contenido en el subconjunto.

## 2.2. Lema del nervio

Ya vimos (definiciones 18, 20, 19 y 21) que dado cualquier espacio topológico  $X$  junto con una cubierta, podemos recuperar un complejo simplicial asociado a la cubierta. En esta sección enunciaremos el Lema del nervio y algunas consecuencias de él.

**Teorema 2** (Lema del nervio). Sean  $X$  un espacio topológico y  $\mathcal{U} = (U_j)_{j \in \mathcal{J}}$  una cubierta indicada de  $X$  que es abierta y finita. Si para cada  $J \subseteq \mathcal{J}$  tal que  $U_J \neq \emptyset$  se cumple que  $U_J$  es contráctil, entonces  $X$  es homotópicamente equivalente a  $|\mathcal{N}(\mathcal{U})|$ .

Como mencionamos al inicio del capítulo, existen muchas otras variantes del teorema anterior. En particular, la prueba de esta versión puede encontrarse en [9, Capítulo 15]. En cuanto a los primeros enunciados de este teorema, tenemos un resultado que apareció exclusivamente para complejos simpliciales en [3]; sin embargo, ahí no se puede encontrar un enunciado tan claro como el que presentamos. En cambio, en [21, 12] encontramos los primeros enunciados que conocemos como lemas del nervio. Su importancia en topología y sus múltiples aplicaciones han inspirado el desarrollo de diferentes versiones del lema del nervio. Por ejemplo, el enunciado original trataba de homotopía pero recientemente se han demostrado versiones homológicas en [15, 14]. Por otro lado, también se ha investigado la funtorialidad del teorema [2]. Creo que el y la lectora podrá encontrar la última referencia particularmente atractiva, pues todo ese artículo está pensado en desarrollar un lema del nervio sólido para las aplicaciones.

También es importante notar que dimos el Lema del nervio para cubiertas indicadas y el Ejercicio 23 nos dice que si olvidamos los índices, entonces los nervios van a ser diferentes.

**Teorema 3.** Si  $\mathcal{U} = (U_j)_{j \in \mathcal{J}}$  es una cubierta indicada de  $X$ , y  $\mathcal{U}'$  es la cubierta (sin índices) asociada, entonces  $|\mathcal{N}(\mathcal{U})|$  es homotópicamente equivalente a  $|\mathcal{N}(\mathcal{U}')|$ .

No vamos a demostrar el teorema anterior pues requiere colapsos [18, Definición 1.36] pero referimos a [2, Remark 1.4] para los detalles. Sin embargo, sí usaremos este teorema para usar indistintamente cubiertas o cubiertas indicadas según nos convenga.

**Corolario 1.** Si  $\mathcal{U}$  es una cubierta abierta de  $X$  tal que todo elemento de  $\mathcal{U}$  es convexo, entonces  $X$  es homotópicamente equivalente a  $|\mathcal{N}(\mathcal{U})|$ .

*Demostración.* Como todo convexo es contráctil (Ejercicio 29) y la intersección de convexos es convexo, el Lema del nervio (Teorema 2) implica el resultado.  $\square$

### 2.3. Algoritmo MAPPER

En esta sección presentamos el algoritmo MAPPER tal cual aparece en [19]. Para ello motivaremos un poco su diseño aunque no lo analizaremos ni computacional ni estadísticamente (para un estudio más profundo referimos a [4, 6]).

Cabe mencionar que la siguiente motivación, aunque basada en [19, Sección 2.1], no trata de ser una traducción o explicación de la motivación de los creadores de MAPPER que no menciona el Lema del nervio. Sin embargo, usando ese resultado como se menciona en [5, Sección 3], es podemos darle una mejor motivación al algoritmo MAPPER.

Consideremos un espacio topológico  $X \subseteq \mathbb{R}^n$  y una función continua  $f: X \rightarrow \mathbb{R}^d$ . Si  $X$  es compacto, sabemos que  $\text{im}(f)$  es compacto. Así, dada cualquier cubierta abierta de  $\text{im}(f)$ , podemos encontrar una subcubierta finita y  $\text{im}(f)$  está contenida en un rectángulo. Es decir, si  $d = 1$ , entonces podemos cubrir a  $\text{im}(f)$  con una cantidad finita de intervalos abiertos. (Nota que para eso sólo necesitas que  $\text{im}(f)$  sea acotada, pero suponer compacidad nos asegura que podemos trabajar con cualquier cubierta que queramos).

Sea  $\mathcal{U}$  una cubierta abierta y finita de  $\text{im}(f)$ . Si  $d = 1$ , podemos, como en [19], suponer que esta cubierta consiste de intervalos abiertos de la misma longitud, digamos  $l$ , espaciados a una distancia fija, digamos  $s$ . Para cada  $U \in \mathcal{U}$ , sabemos que  $f^{-1}[U]$  es abierto. De manera que  $\mathcal{U}' = \{f^{-1}[U] \mid U \in \mathcal{U}\}$  es una cubierta abierta y finita de  $X$ .

Es falso que en general  $\mathcal{U}'$  satisfaga las hipótesis del Lema del nervio. Un ejemplo es  $p_2: S^1 \rightarrow \mathbb{R}$  definida como  $p_2(x, y) = y$  con la siguiente cubierta del intervalo unitario:  $\{[0, .6), (.4, 1]\}$ . El problema es que en este caso  $\mathcal{U}'$  resulta en dos semicírculos cuya intersección tiene dos componentes conexas.

**Ejercicio 30.** Muestra que el ejemplo anterior en efecto no satisface las hipótesis del Lema del nervio (Teorema 2) y que el nervio de  $\mathcal{U}'$  no es homotópicamente equivalente a  $S^1$  (para lo último tendrás que citar, sin prueba, algún teorema no mencionado en estas notas).

Algo que podemos hacer, es considerar, para cada  $U \in \mathcal{U}$ , el conjunto de componentes conexas de  $f^{-1}[U]$ . Llamemos a esta cubierta el *refinamiento de la imagen inversa de  $\mathcal{U}$* . Esto, al menos en el ejemplo anterior resuelve el problema.

**Ejercicio 31.** Muestra que el refinamiento de la imagen inversa de la cubierta del ejemplo del Ejercicio 30 sí satisface las hipótesis del Lema del nervio. Dibuja el nervio.

Si a la explicación que hemos expuesto le quitamos las menciones del Lema del nervio, recuperamos [19, Sección 2.1]. Veamos cómo. El refinamiento de la imagen inversa de  $\mathcal{U}$  nos induce un nervio. Pidiendo una partición de la unidad para esa cubierta, podemos parametrizar continuamente los elementos de  $X$  usando las coordenadas baricéntricas de la realización geométrica de tal nervio. Informalmente, podemos pensar que el nervio de esa cubierta es una deformación del espacio original. Esto parece ser suficiente para los autores de [19] para justificar el siguiente algoritmo.

1. Consideremos una nube de puntos  $Y$  tomada de un espacio  $X \subseteq \mathbb{R}^n$  y una función continua  $f: X \rightarrow \mathbb{R}^d$ .
2. Para una cubierta  $\mathcal{U}$  de  $\text{im}(f)$  calculemos conglomerados de  $f^{-1}[U] \subseteq Y$  con  $U \in \mathcal{U}$  (cada conglomerado juega el papel de la componente conexa).
3. El nervio que se obtiene de este conjunto de conglomerados es una aproximación topológica de  $X$ .

Sin embargo, desde mi punto de vista podemos obtener una motivación del algoritmo MAPPER más fundamental. Para justificar esto, cuestionaremos la heurística de MAPPER mencionada en el párrafo anterior.

Si  $X$ , el conjunto del cual una nube de datos  $Y$  fue extraída, es compacto, sabemos que sus componentes conexas son acotadas. Así, las componentes conexas de  $f^{-1}[U]$  resultan ser subespacios de las componentes de  $X$ . Así, podemos encontrar encerrar a las componentes conexas de  $f^{-1}[U]$  en una bola abierta con centro en algún elemento de  $Y$ . Es verosímil pensar que tales bolas son lo suficientemente grandes como para formar una cubierta de  $X$ . Por el Corolario 1, sabríamos con absoluta certeza que el nervio de esta cubierta es homotópicamente equivalente a  $X$ , no solo que es una imagen continua de él. No es muy difícil pensar a los conglomerados como bolas abiertas. Creo que este argumento, nos ofrece una mejor motivación del algoritmo MAPPER como herramienta para aproximar  $X$ .

Hay varios problemas con el algoritmo MAPPER.

- ¿Cómo afecta la función  $f$  al nervio que se obtiene?
- ¿La forma en la que se obtengan los conglomerados altera el nervio?

- ¿Podemos garantizar estadísticamente que con una muestra diferente los nervios serán “similares”?
- Si dos funciones son cercanas bajo alguna métrica útil, ¿los nervios serán “cercanos”?

No profundizaremos en ninguna de estas preguntas pues son temas de investigación actual. Sin embargo, referimos a [5, 4] para introducirse en ellos.

# Capítulo 3

## Complejos de Čech y Vietoris-Rips

Ahora hablaremos de los complejos de Čech y Vietoris-Rips. Ambos complejos se definen para cualquier espacio topológico y sirven para visualizar datos. A diferencia del complejo que se encuentra con MAPPER que requiere muchas elecciones, los complejos que definiremos aquí sólo dependen de una sola elección y, como veremos más adelante, diferentes elecciones se pueden combinar para estudiar homológicamente nuestros datos. Excelentes referencias para este capítulo son [8, Capítulo III] y [7, Capítulo 10].

### 3.1. Complejo de Čech

Empecemos directamente con la definición.

**Definición 27.** Sean  $X \subseteq \mathbb{R}^d$  un subespacio topológico,  $Y \subseteq X$  finito y  $r > 0$ . El *complejo de Čech*,  $\text{Cech}(r, Y)$  (o  $\text{Cech}_r(Y)$ ) es el nervio de la cubierta cerrada  $\{\bar{B}_r(x) \mid x \in Y\}$ .

**Observación 5.** En la definición anterior, no somos específicos acerca de el espacio que es cubierto con  $\{\bar{B}_r(x) \mid x \in Y\}$ . En general deseamos que sea una cubierta de  $X$  pues, siguiendo con las ideas del capítulo anterior, si  $Y$  es una nube de datos extraída de  $X$ , entonces nuestro interés es describir a  $X$ . Sin embargo, esto no es posible en general. Las condiciones sobre  $X$ ,  $r$  y  $Y$  que nos permiten garantizar una descripción homotópica de  $X$  a través de  $\text{Cech}(r, Y)$  están fuera del alcance de este documento, pero referimos a [5, Section 4.2] para una breve introducción al tema.

Observemos que este nervio se obtiene de una cubierta de cerrados convexos. Aunque es posible adaptar el Lema del nervio (Teorema 2) para este tipo de cubiertas, aquí será suficiente el siguiente ejercicio.

**Ejercicio 32.** Usa el Corolario 1 para mostrar que  $\text{Cech}_r(Y)$  es homotópicamente equivalente a  $\bigcup_{x \in Y} \bar{B}_r(x)$ . Pista: primero muestra que si  $r < s$ , entonces  $\text{Cech}_r(Y) \subseteq \text{Cech}_s(Y)$ .

Ahora nos enfrentamos a un problema. ¿Cómo sabemos qué conjuntos forman un simplejo en  $\text{Cech}_r(Y)$ ? Es decir, es claro que un conjunto de bolas de radio  $r$  tiene intersección no vacía si y sólo si sus centros se encuentran en una bola de radio  $r$  y centro en la intersección. Pero ¿cómo determinamos esto con un algoritmo? Para responder esto usaremos el teorema de Helly. Primero necesitamos recordar un resultado sobre conjuntos convexos.

**Teorema 4.** Si  $A$  y  $B$  son conjuntos convexos en  $\mathbb{R}^d$  tales que  $\text{int}(A) \neq \emptyset$  y  $\text{int}(A) \cap \text{int}(B) = \emptyset$ , entonces existe un hiperplano  $H$  que *separa estrictamente* a  $\bar{A}$  y  $\bar{B}$ ; es decir, pertenecen a diferentes semiespacios abiertos determinados por  $H$ .

El teorema anterior es en realidad un corolario del Teorema del hiperplano separador. Hay muchas referencias al respecto pero recomendamos [10, Sección 3.6].

**Teorema 5.** Sea  $F$  una familia de convexos cerrados en  $\mathbb{R}^d$ . La intersección de cualesquiera  $d + 1$  conjuntos de  $F$  es no vacía si y sólo si  $\bigcap F$  es no vacía.

La siguiente prueba es básicamente la que se encuentra en [8, Capítulo Complexes]

*Prueba del teorema de Helly.* Es claro que si  $\bigcap F \neq \emptyset$ , entonces la intersección de cualesquiera  $d + 1$  conjuntos de  $F$  es no vacía.

La prueba de la conversa es por inducción doble sobre  $n = \#F$  y  $n \geq d$ . Para  $d = 1$ , la inducción sobre  $n$  se deja como ejercicio. También es claro que el enunciado es verdadero para  $n = d + 1$ .

Supongamos que, para  $d > 1$  tenemos una familia  $F$  de  $n$  convexos cerrados de  $\mathbb{R}^d$  tal que la intersección de cualesquiera  $d + 1$  conjuntos de  $F$  es no vacía,  $\bigcap F \neq \emptyset$ . Supongamos, además, que tanto  $n$  como  $d$  son mínimas con respecto a la existencia de  $F$ .

Por la elección de  $n$  y  $d$ , sabemos que si  $F = \{X_1, \dots, X_n\}$ , entonces  $Y_n = \bigcap_{i < n} X_i \neq \emptyset$ . Por otro lado,  $Y_n$  es convexo y cerrado. Como por hipótesis  $Y_n \cap X_n = \emptyset$ , el Teorema 4 nos asegura que existe un hiperplano  $H$  que separa estrictamente a  $Y_n$  y  $X_n$ .

Consideremos  $Z_i = X_i \cap H$  con  $i < n$  y sea  $F'$  el conjunto cuyos elementos son precisamente estos conjuntos. Observemos que  $Z_i$  es cerrado y convexo de  $H \cong \mathbb{R}^{d-1}$ . Por otro lado, elegimos  $F$  de tal manera que la intersección de cualesquiera  $d$  de los  $X_i$ , con  $i < n$ , tiene intersección no vacía



con  $X_n$ . Esto implica que si  $i_k < d$ , entonces  $\bigcap_{k \leq d} X_{i_k}$  contiene puntos en ambos semiespacios determinados por  $H$  (pues tiene intersección no vacía con  $X_n$  y en  $Y_n$ ). Como consecuencia la intersección de cualesquiera  $d$  elementos de  $F'$  es no vacía. Dada la elección de  $n$  y  $d$  concluimos que  $\bigcap F' \neq \emptyset$ .

Para terminar, notemos que

$$\bigcap F' = \bigcap_{i < n} X_i \cap H = Y_n \cap H.$$

Lo anterior es imposible pues  $H$  separa estrictamente a  $Y_n$  y  $X_n$ . □

**Ejercicio 33.** Demuestra la base de la inducción sobre  $n$  para  $d = 1$  que se omitió en el teorema anterior. Para ser precisos, muestra que si una familia de intervalos cerrados es tal que dos a dos no son ajenos, entonces la intersección de todos ellos es un intervalo cerrado.

Volvamos a la pregunta ¿Cómo sabemos qué conjuntos forman un simplejo en  $\text{Cech}_r(Y)$ ? Ya dijimos que es claro que  $\sigma \subseteq Y$  induce un simplejo en  $\text{Cech}_r(Y)$  si y sólo si los elementos de  $\sigma$  pertenecen a una bola de radio  $r$ . Gracias al teorema de Helly podemos contestar a la pregunta de arriba y, después dar un algoritmo muy eficiente al respecto.

**Teorema 6** (Teorema de Jung). Sea  $\sigma \subseteq \mathbb{R}^d$ . Cualesquiera  $d + 1$  puntos de  $\sigma$  pertenecen a una bola de radio  $r$  si y solo si  $\sigma$  misma está contenida en una bola de radio  $r$ .

**Ejercicio 34.** Muestra el Teorema 6.

Así, determinar si  $\sigma \subseteq Y$  induce un simplejo en  $\text{Cech}_r(Y)$  es equivalente a determinar si  $\sigma$  está contenida en una bola de radio  $r$ .

**Definición 28.** Sea  $\sigma \subseteq \mathbb{R}^d$ . La *minibola* de  $\sigma$  es la bola cerrada mínima que contiene a  $\sigma$ .

**Ejercicio 35.** La minibola de  $\sigma \subseteq Y$  es  $r$  si y sólo si  $\sigma$  induce un simplejo en  $\text{Cech}_r(Y)$ .

Es claro que la minibola de  $\sigma$  está determinada por un subconjunto  $\nu \subseteq \sigma$  que se encuentra en la frontera. Por ello, el siguiente algoritmo busca una bola cerrada que contenga a  $\nu$  en la frontera y  $\tau = \sigma \setminus \nu$  en el interior. Si tal bola existe, esa es la minibola de  $\sigma$ ; si no, entonces aumenta el número de puntos que se esperan en la frontera uno a la vez. Así, para encontrar la minibola de  $\sigma$  basta con llamar  $\text{Miniball}(\sigma, \emptyset)$

Para un análisis de complejidad de este algoritmo, referimos a [8].

**Algoritmo 1:** Algoritmo para calcular la minibola mínima de un conjunto de puntos.

**Entrada:**  $\tau, \nu$  ajenos

**Salida:**  $B$  una bola cerrada que contiene a  $\tau \cup \nu$ .

```

1 Función MiniBall( $\tau, \nu$ ):
2   Si  $\tau = \emptyset$  entonces
3     Calcular directamente la minibola  $B$  de  $\nu$ 
4   finSi
5   Sino
6     Elegir un punto aleatorio  $u \in \tau$ 
7      $B \leftarrow \text{MiniBall}(\tau - \{u\}, \nu)$ 
8     Si  $u \notin B$  entonces
9        $B \leftarrow \text{MiniBall}(\tau - \{u\}, \nu \cup \{u\})$ 
10    finSi
11  finSi
12  Retornar  $B$ 

```

### 3.2. Complejo de Vietoris-Rips

Otro complejo simplicial que se puede asociar a una nube de datos es el complejo de Vietoris-Rips. Como vimos antes, determinar los simplejos de  $\text{Cech}_r(Y)$  requiere encontrar las minibolas de los subconjuntos de  $Y$ . En contraste, para determinar los complejos del complejo de Vietoris-Rips, sólo necesitamos verificar que dos a dos se encuentran en una bola de radio  $r$ .

**Definición 29.** Sean  $X \subseteq \mathbb{R}^d$  un subespacio topológico,  $Y \subseteq X$  finito y  $r > 0$ . El *complejo de Vietoris-Rips*,  $\text{VRips}(r, Y)$  (o  $\text{VRips}_r(Y)$ ) es el complejo simplicial generado por los conjuntos  $\sigma \subseteq Y$  tales que  $d(x, y) \leq 2r$  para cualesquiera  $x, y \in \sigma$ .

**Ejercicio 36.** Muestra que  $\text{Skel}_1(\text{VRips}_r(Y)) = \text{Skel}_1(\text{Cech}_r(Y))$

**Ejercicio 37.** Muestra que  $\text{Cech}_r(Y)$  es un subcomplejo de  $\text{VRips}_r(Y)$

Aunque el ejercicio anterior es muy sencillo, el siguiente resultado sí requiere un argumento más complejo [8].

**Lema 1.** Para cada  $r$  el complejo de Vietoris-Rips  $\text{VRips}_r(Y)$  es un subcomplejo de  $\text{Cech}_{\sqrt{2}r}(Y)$

De estas desigualdades, tenemos certeza de que podemos aproximar el complejo de Čech por el de Vietoris-Rips. Esto es ventajoso pues para determinar el último basta con calcular  $\text{Skel}_1(\text{Cech}_r(Y))$  y luego determinar el complejo de *clanes* de esa gráfica (véase [13]).

Otra cosa que vale la pena notar es que ambos complejos forman cadenas de subcomplejos simpliciales de  $\mathcal{P}(Y)$ . Esto es muy importante y regresaremos a considerarlo cuando estudiemos homología persistente.

Para terminar recomendamos [8, 7] para conocer otras estructuras que se pueden asociar a nubes de datos.

# Capítulo 4

## Homología persistente

Este capítulo no pretende ser una introducción a Homología simplicial, sino que buscamos proveer las herramientas teóricas necesarias para poder estudiar aplicaciones de homología persistente. Por ello, recomendamos complementar la lectura de este capítulo con un curso de topología algebraica. En particular recomendamos [16] como referencia básica para topología algebraica. Otras referencias que pueden ser útiles para acompañar este capítulo son [8, Capítulos IV y VII] y [18]. El primer texto es excelente para estudiar persistencia, pues los autores trabajan ampliamente en el tema. El segundo texto tiene la misma meta que este capítulo.

### 4.1. Complejos de cadenas sobre $\mathbb{Z}_2$

Dado que sólo queremos estudiar homología de tal manera que podamos hacer aplicaciones pronto, no vamos a estudiar homología en general sino que estudiaremos homología simplicial con coeficientes en  $\mathbb{Z}_2$ . Empezamos con complejos de cadenas sobre  $\mathbb{Z}_2$ .

**Definición 30.** Un *complejo de cadenas sobre  $\mathbb{Z}_2$*  es una colección de  $\mathbb{Z}_2$ -espacios vectoriales  $\mathcal{C} = (C_n)_{n \in \mathbb{Z}}$  junto con una familia de morfismos de grupos  $\partial_n: C_n \rightarrow C_{n-1}$  tales que  $\partial_n \circ \partial_{n+1} = 0$ . Los elementos de  $C_n$  se llaman *n-cadenas* y los morfismos  $\partial_n$  se llaman *mapas frontera*.

Observemos que la definición anterior no depende en realidad de  $\mathbb{Z}_2$ , así que en general se puede definir un complejo de cadenas sobre cualquier campo y más generalmente sobre cualquier categoría abeliana. Sin embargo, hay que notar que los mapas frontera son morfismos de grupos.

**Ejercicio 38.** Demuestra que  $\text{im}(\partial_{n+1}) \leq \ker(\partial_n)$

**Definición 31.** El grupo de  $n$ -ciclos de  $\mathcal{C}$  es  $Z_n(\mathcal{C}) = \ker(\partial_n)$ ; el grupo de  $n$ -fronteras de  $\mathcal{C}$  es  $B_n(\mathcal{C}) = \text{im}(\partial_{n+1})$ .

Del Ejercicio 38 (en realidad de la Definición 30) sabemos que:  $B_n(\mathcal{C}) \leq Z_n(\mathcal{C})$ .

**Definición 32.** El  $n$ -ésimo grupo de homología de  $\mathcal{C}$  es  $H_n(\mathcal{C}) = Z_n(\mathcal{C})/B_n(\mathcal{C})$ .

Para profundizar en algebra homológica, recomendamos [17].

## 4.2. Homología simplicial con coeficientes en $\mathbb{Z}_2$

Ahora, vamos a construir los grupos de homología con coeficientes en  $\mathbb{Z}_2$  para un complejo simplicial.

Sea  $K$  un complejo simplicial. Definimos  $\mathcal{C}_n(K)$  como el  $\mathbb{Z}_2$ -espacio vectorial con base  $K_n$ . Para cada  $n$  sea  $\partial_n: \mathcal{C}_n(K) \rightarrow \mathcal{C}_{n-1}(K)$  la función definida como

$$\partial_n(\sigma) = \sum_{\tau \text{ es frontera de } \sigma} \tau.$$

**Definición 33.** El  $n$ -ésimo grupo de homología con coeficientes en  $\mathbb{Z}_2$  de  $K$  es el  $n$ -ésimo grupo de homología de  $\mathcal{C}(K)$ . Si escribimos  $Z_n(K) = Z_n(\mathcal{C}(K))$  y  $B_n(K) = B_n(\mathcal{C}(K))$ , entonces  $H_n(K) = Z_n(K)/B_n(K)$

**Ejercicio 39.** Prueba que el  $n$ -ésimo grupo de homología de un complejo simplicial está bien definido; es decir, demuestra que los mapas frontera de  $\mathcal{C}(K)$  satisfacen  $\partial_n \circ \partial_{n+1} = 0$ .

**Ejemplo 3** (La homología del 2-simplejo). Supongamos que tenemos el dos simplejo  $K = \mathcal{P}(\{x, y, z\})$ .

Entonces

$$\mathcal{C}_n(K) = \begin{cases} \{0, \emptyset\} \cong \mathbb{Z}_2 & \text{si } n = -1 \\ \{0, x, y, z, x+y, x+z, y+z, x+y+z\} = \mathbb{Z}_2^{\{x,y,z\}} & \text{si } n = 0 \\ \{0, xy, yz, zx, xz+yz, xy+zy, yx+xz, xy+yz+xz\} = \mathbb{Z}_2^{\{xy,yz,xz\}} & \text{si } n = 1 \\ \{0, xyz\} \cong \mathbb{Z}_2 & \text{si } n = 2 \\ 0 & \text{en otro caso.} \end{cases} \quad (4.1)$$

Así,

$$Z_n(K) = \begin{cases} 0 & \text{si } n = -1 \\ \{0, x + y, x + z, y + z\} & \text{si } n = 0 \\ \{0, xy + yz + xz\} & \text{si } n = 1 \\ 0 & \text{si } n = 2 \\ 0 & \text{en otro caso.} \end{cases} \quad (4.2)$$

Además

$$B_n(K) = \begin{cases} \{0, \emptyset\} & \text{si } n = -1 \\ \{0, x + y, x + z, y + z\} & \text{si } n = 0 \\ \{0, xy + yz + xz\} = \mathbb{Z}_2^{\{xy, yz, xz\}} & \text{si } n = 1 \\ \{0, xyz\} \cong \mathbb{Z}_2 & \text{si } n = 2 \\ 0 & \text{en otro caso.} \end{cases} \quad (4.3)$$

En consecuencia  $H_n(K) = 0$  para todo  $n$ . Esto refleja el hecho de que  $K$  es contraíble.

**Ejercicio 40.** Calcula los grupos de homología de  $\partial\sigma$  donde  $\sigma$  es un 3-simplejo.

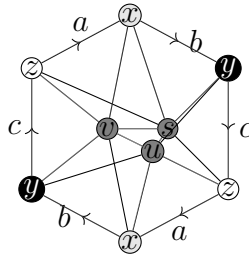


Figura 4.1: El plano proyectivo.

**Ejemplo 4** (La homología del plano proyectivo). content

# Bibliografía

- [1] ALBERTO, L. Pseudospheres: combinatorics, topology and distributed systems. *Journal of Applied and Computational Topology* 8 (2024), 1023–1052.
- [2] BAUER, U., KERBER, M., ROLL, F., AND ROLLE, A. A unified view on the functorial nerve theorem and its variations. *Expositiones Mathematicae* 41, 4 (2023), 125503.
- [3] BORSUK, K. On the imbedding of systems of compacta in simplicial complexes. *Fundamenta Mathematicae* 35, 1 (1948), 217–234.
- [4] CARRIÈRE, M., MICHEL, B., AND OUDOT, S. Statistical analysis and parameter selection for mapper. *Journal of Machine Learning Research* 19, 12 (2018), 1–39.
- [5] CHAZAL, F., AND MICHEL, B. An introduction to topological data analysis: Fundamental and practical aspects for data scientists. *Frontiers in Artificial Intelligence* 4 (2021).
- [6] DEY, T. K., MÉMOLI, F., AND WANG, Y. Multiscale mapper: topological summarization via codomain covers. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms* (USA, 2016), SODA '16, Society for Industrial and Applied Mathematics, p. 997–1013.
- [7] EDELSBRUNNER, H. *A Short Course in Computational Geometry and Topology*. Springer Publishing Company, Incorporated, 2014.
- [8] EDELSBRUNNER, H., AND HARER, J. *Computational Topology: An Introduction*. Applied Mathematics. American Mathematical Society, 2010.
- [9] KOZLOV, D. N. *Combinatorial Algebraic Topology*, vol. 21 of *Algorithms and computation in mathematics*. Springer, Berlin, Heidelberg, 2008.

- [10] LEONARD, I., AND LEWIS, J. *Geometry of Convex Sets*. Wiley, 2015.
- [11] MAY, J. *Simplicial Objects in Algebraic Topology*. Chicago Lectures in Mathematics. University of Chicago Press, 1967.
- [12] MCCORD, M. C. Homotopy type comparison of a space with complexes associated with its open covers. *Proceedings of the American Mathematical Society* 18, 4 (1967), 705–708.
- [13] MESHULAM, R. The clique complex and hypergraph matching. *Combinatorica* 21, 1 (2001), 89–94.
- [14] MESHULAM, R. Domination numbers and homology. *Journal of Combinatorial Theory, Series A* 102, 2 (2003), 321–330.
- [15] MEUNIER, F., AND MONTEJANO, L. Different versions of the nerve theorem and colourful simplices. *Journal of Combinatorial Theory, Series A* 169 (2020), 105125.
- [16] MUNKRES, J. R. *Elements of algebraic topology*. Addison-wesley, 1984.
- [17] ROTMAN, J. J. *An Introduction to Algebraic Topology*. Graduate Texts in Mathematics: 119. Springer New York, 1988.
- [18] SCOVILLE, N. A. *Discrete Morse Theory*, vol. 90 of *Student mathematical library*. American Mathematical Society, 2019.
- [19] SINGH, G., MEMOLI, F., AND CARLSSON, G. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. In *Eurographics Symposium on Point-Based Graphics* (2007), M. Botsch, R. Pajarola, B. Chen, and M. Zwicker, Eds., The Eurographics Association.
- [20] TOM DIECK, T. *Algebraic topology*. European Mathematical Society, Germany, 2008.
- [21] WEIL, A. Sur les théorèmes de de rham. *Commentarii mathematici Helvetici* 26 (1952), 119–145.