



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS
MATEMÁTICAS Y DE LA
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA

DISTANCIA DE HAUSDORFF-FERMAT: PROPUESTA Y
APLICACIONES

TESIS
QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN CIENCIAS

PRESENTA:
BRANDON SAÚL GÓMEZ BRAVO

TUTOR PRINCIPAL
DR. SERGIO IVÁN LÓPEZ ORTEGA
FACULTAD DE CIENCIAS

CIUDAD DE MÉXICO, MAYO 2024



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Resumen

En los últimos años, el uso de algoritmos de Machine Learning ha experimentado un crecimiento significativo, impulsado por su diversidad de aplicaciones, lo que ha planteado nuevos desafíos. En particular, en el ámbito del análisis de datos multidimensionales, el objetivo central radica en transformar la representación de conjuntos de datos caracterizados por una alta dimensionalidad, reduciendo la cantidad de variables. Este proceso tiene como meta mejorar el rendimiento de los modelos predictivos. Sin embargo, surge el problema de la dimensionalidad. En este contexto, la distancia Euclideana, ampliamente utilizada, muestra sus limitaciones, ya que en espacios de alta dimensión, los puntos de un conjunto de datos tienden a estar a distancias similares, lo que dificulta la diferenciación entre ellos. Además, los enfoques tradicionales tienden a fracasar cuando el conjunto de datos está inmerso en estructuras no lineales. Diversos estudios previos han abordado este desafío desde la perspectiva de considerar tanto la densidad como la geometría de los datos, intentando equilibrar dos objetivos aparentemente opuestos: reducir la dimensionalidad sin perder información significativa.

En este contexto, este trabajo abarca desde los fundamentos de la noción de distancia hasta la aplicación de distancias innovadoras, como la distancia de Hausdorff-Fermat, en el análisis de datos complejos en estructuras no lineales.

En el primer capítulo, sentamos las bases al introducir la noción de distancia y profundizar en la conocida Distancia de Hausdorff. Esta distancia sirve para cuantificar la similitud entre conjuntos. Esto ha tenido una gran cantidad de aplicaciones en la época moderna, desde la detección de caras hasta segmentación de imágenes. Este capítulo establece el escenario para comprender cómo las distancias pueden ser poderosas herramientas en la resolución de problemas reales.

En el segundo capítulo se presenta la Distancia Power-Weighted Shortest-Path y algunos teoremas que aseguran convergencia en ciertos contextos. Además, introducimos la Distancia de Fermat, que se convierte en un componente fundamental en nuestro enfoque. Ambas distancias tienen como objetivo principal alcanzar un equilibrio que permita capturar tanto la densidad de los datos como la estructura

de su soporte, superando las limitaciones de las distancias tradicionales.

El tercer capítulo mostramos aplicaciones prácticas de la Distancia de Fermat, con un énfasis especial en algoritmos de clusterización. Posteriormente definimos la Hausdorff-Fermat, una variante que incorpora la densidad de los datos. A través de ejemplos simples pero reveladores, exploramos el potencial de esta distancia en la segmentación de imágenes, especialmente en el campo de la medicina.

Agradezco al proyecto PAPIIT-DGAPA IN102822 por el apoyo en infraestructura y condiciones de trabajo que permitieron realizar esta tesis.

Índice general

1. Distancia de Hausdorff	1
1.1. Introducción	1
1.2. Distancia de Hausdorff	3
1.2.1. Aplicaciones	6
2. Distancias de trayectoria óptima	8
2.1. Motivación	8
2.2. Distancia PWSP	10
2.3. Constante del tiempo de percolación	13
2.4. Resultados de convergencia local	14
2.5. Resultados en variedades	26
2.6. Distancia de Fermat	34
2.6.1. Complejidad	36
3. Aplicaciones	38
3.1. Clusterización	38
3.2. Distancia de Hausdorff-Fermat	39
4. Conclusiones	47
Bibliografía	48

Distancia de Hausdorff

La noción de distancia es algo que surge naturalmente desde los animales que tienen la intuición de cercanía o lejanía para conseguir alimentos y/o modificar su lugar de estancia dependiendo del clima de una región. En la humanidad, el concepto de distancia ha sido de vital importancia para transporte de recursos. Para dar un ejemplo simple, los vuelos siguen *trayectorias geodésicas* para ir de un poblado a otro que no son necesariamente una línea recta.

En las Matemáticas se ha axiomatizado el concepto de distancia. Esto ayuda a tener un formalismo para obtener resultados interesantes al respecto. En este capítulo presentamos este concepto matemático y algunas consecuencias básicas. También daremos la definición de la *distancia de Hausdorff* y algunas de sus aplicaciones. A lo largo de este capítulo seguiremos las exposiciones del libro [1].

1.1. Introducción

Definición 1.1 (Distancia) Una función $d : X \times X \rightarrow \mathbb{R}^+$ es considerada una *distancia o métrica* en un conjunto X si para cualesquiera dos elementos $p, q \in X$ se cumple lo siguiente:

1. $d(p, q) > 0$ si $p \neq q$; y $d(p, p) = 0$.
2. $d(p, q) = d(q, p)$.
3. $d(p, q) \leq d(p, r) + d(r, q)$ para cualquier $r \in X$.

Además, el conjunto X que cumple con lo anterior es considerado *espacio métrico*.

Por ejemplo, una distancia posible en el conjunto \mathbb{R} es el valor absoluto de la diferencia entre dos elementos, es decir,

$$d(p, q) = |p - q| \quad \forall p, q \in \mathbb{R}.$$

También es posible definir una distancia entre nodos de una gráfica, por ejemplo la conocida *distancia del taxi* [2]. En este caso, el espacio X es la gráfica, y la distancia entre dos nodos se define como el mínimo número de segmentos que tienen que ser recorridos para ir de un nodo a otro.

Es posible definir una distancia en diferentes espacios, como la distancia entre gráficas, entre matrices, entre funciones o entre conjuntos. Profundizaremos en este último caso más adelante.

Una de las distancias más importantes en el espacio \mathbb{R}^k es la *distancia Euclidiana*.

Definición 1.2 (Distancia Euclidiana) Sean $p, q \in \mathbb{R}^k$, con $p = (p_1, \dots, p_k)$ y $q = (q_1, \dots, q_k)$, la distancia Euclidiana está definida como

$$d(p, q) = \sqrt{\sum_{i=1}^k (p_i - q_i)^2}. \quad (1.1)$$

A pesar de que cumple las condiciones de ser una distancia, no siempre es la mejor manera de medir, porque medir depende del problema y del espacio con el que se esté tratando. Por ejemplo, podemos pensar en la superficie de la tierra siendo un subconjunto de \mathbb{R}^3 . Cuando queremos medir la distancia entre dos ciudades distintas, no ocupamos la distancia Euclidiana, ya que estaríamos midiendo con una línea recta, es decir, como si pasáramos por un túnel a través de la tierra. Lo que se hace realmente es medir el camino más corto entre las ciudades con la restricción que todos los puntos de la trayectoria se encuentren sobre la superficie de la tierra. Entonces, podemos definir la distancia entre dos puntos p, q sobre la superficie S como

$$d_S(p, q) = \min_{\gamma \subset S} \text{longitud}(\gamma), \quad (1.2)$$

donde γ es cualquier trayectoria que inicia en p y termina q , tal que $\gamma \subset S$. De aquí es importante observar que todo subconjunto $Y \subset X$ de un espacio métrico X , es por sí mismo un espacio métrico, pues la función d restringida al subconjunto $Y \times Y$ sigue satisfaciendo las propiedades requeridas.

1.2. Distancia de Hausdorff

Una cuestión que surge es cómo definir una distancia entre conjuntos, es decir, cuantificar la similitud entre dos conjuntos. Una de las soluciones más conocidas es la *distancia de Hausdorff* [3]. Primero daremos unas definiciones que nos ayudarán a entender mejor los conceptos de esta sección.

Definición 1.3 Consideremos un espacio métrico (X, d) . Sea $p \in X$ y $r > 0$. La bola abierta centrada en p con radio r es definida por $B_p(r) = \{x \in X : d(x, p) < r\}$.

Definición 1.4 Un conjunto $E \subseteq X$ es *acotado* en (X, d) si existen $p \in X$ y $M > 0$ tales que $E \subseteq B_p(M)$.

Definición 1.5 Una sucesión $\{x_n\}$ es una *sucesión de Cauchy* si para cada $\varepsilon > 0$ existe un entero positivo N tal que $d(x_n, x_m) < \varepsilon$ para todo $m, n \geq N$.

Definición 1.6 Un espacio métrico (X, d) es un *espacio métrico completo* si toda sucesión de Cauchy en (X, d) converge a un punto en X .

Definición 1.7 Un conjunto $E \subseteq X$ es *compacto* en (X, d) si cada sucesión en E tiene una subsucesión que converge a un punto en E .

Consideremos un espacio métrico completo (X, d) . Y consideremos $\mathcal{M}(X)$ el conjunto de todos los subconjuntos compactos no vacíos de X . Consideremos un elemento $A \in \mathcal{M}(X)$ y definimos la r -expansión de A , denotada por $A^{(r)}$, como

$$A^{(r)} = \bigcup_{x \in A} B_x(r) = \bigcup_{x \in A} \{y \in X : d(x, y) \leq r\}. \quad (1.3)$$

Definición 1.8 (Distancia de Hausdorff) Sea (X, d) un espacio métrico completo. Definimos la distancia de Hausdorff entre $A, B \in \mathcal{M}(X)$ como

$$d_H(A, B) = \inf\{r > 0 : B \subset A^{(r)} \text{ y } A \subset B^{(r)}\}, \quad (1.4)$$

con la convención de que el ínfimo del conjunto vacío es ∞ .

Entonces, la distancia de Hausdorff es el menor valor de r para el cual la r -expansión de A contiene a B y la r -expansión de B contiene a A .

Una definición equivalente puede ser construida. Consideremos un elemento $p \in X$ y $A \in \mathcal{M}(X)$, definimos la distancia de un punto p a un conjunto A por

$$d(p, A) = \inf_{y \in A} d(p, y). \quad (1.5)$$

Definición 1.9 La distancia de Hausdorff está definida como

$$d_H(A, B) = \max \left\{ \sup_{p \in A} d(p, B), \sup_{q \in B} d(q, A) \right\}. \quad (1.6)$$

Entonces, la distancia de Hausdorff es la distancia más alejada de cualquier punto de B al conjunto A o la distancia de cualquier punto de A al conjunto B ; la que sea mayor.

Proposición 1.1 Las definiciones 1.8 y 1.9 son equivalentes.

Demostración. Para fines prácticos denotemos a la distancia de la Definición 1.8 como d_1 y a la de la Definición 1.9 como d_2 .

Sea $r = d_2(A, B)$. Por definición se tiene que $A \subset B^{(r)}$, entonces $d(a, B) \leq r$ para toda $a \in A$, por lo que $\sup_{a \in A} d(a, B) \leq r$. De igual manera se tiene que $B \subset A^{(r)}$, entonces $d(b, A) \leq r$ para toda $b \in B$, por lo tanto $\sup_{b \in B} d(b, A) \leq r$. Entonces,

$$d_1(A, B) = \max \left\{ \sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A) \right\} \leq r = d_2(A, B).$$

Ahora, consideremos $d_1(A, B) = r$, entonces $d(a, B) \leq r$ para toda $a \in A$, por lo que $A \subset B^{(r)}$. También, por definición, tenemos que $d(b, A) \leq r$ para toda $b \in B$, entonces $B \subset A^{(r)}$. Así obtenemos que

$$d_2(A, B) \leq d_1(A, B).$$

Notemos que $d_2(A, B) = \infty$ es equivalente a que para toda $r \geq 0$ ocurre que $A \not\subset B^{(r)}$ o $B \not\subset A^{(r)}$ esto ocurre si y solamente si existe $a \in A$ tal que $d(a, B) \geq r$ o $b \in B$ tal que $d(b, A) \geq r$. A su vez, esto es equivalente a que $d_1(A, B) \geq r$ y como es válido para toda $r \geq 0$ entonces $d_1(A, B) = \infty$. Se concluye que $d_1 = d_2$. ■

Proposición 1.2 Sea (X, d) un espacio métrico completo. Entonces $d_H(\cdot)$ induce una distancia en el conjunto $\mathcal{M}(X)$.

Demostración. A continuación mostramos que se cumplen las tres propiedades de la Definición 1.1.

1. Por definición y considerando que $d(\cdot)$ es una función no negativa, es claro que $d_H(A, B) \geq 0$.

Por otro lado, si $A = B$ entonces $d(p, B) = 0 = d(q, A)$ para toda $p \in A$ y $q \in B$, entonces $d_H(A, B) = 0$. Ahora, supongamos que $d_H(A, B) = 0$ lo cual implica que $\sup_{p \in A} d(p, B) = 0$, entonces, $d(p, B) = 0$ para todo $p \in A$. Por lo que $A \subseteq B$. De manera similar obtenemos que $B \subseteq A$. Por lo tanto $A = B$.

2. La simetría se cumple por definición,

$$d_H(A, B) = \max \left\{ \sup_{p \in A} d(p, B), \sup_{q \in B} d(q, A) \right\} = d_H(B, A).$$

3. Consideremos $A, B, C \in \mathcal{M}(X)$. Ya que A, B, C son compactos podemos garantizar que para cada $a \in A$ existe $c_a \in C$ tal que $d(a, C) = d(a, c_a)$. Entonces,

$$\begin{aligned} d(a, B) &= \inf \{ d(a, b) : b \in B \} \leq \inf \{ d(a, c_a) + d(c_a, b) : b \in B \} \\ &= d(a, c_a) + \inf \{ d(c_a, b) : b \in B \} = d(a, c_a) + d(c_a, B) \\ &\leq d(a, C) + d(c_a, B). \end{aligned}$$

Entonces,

$$\sup_{a \in A} d(a, B) \leq \sup_{a \in A} d(a, C) + \sup_{c \in C} d(c, B) \quad (1.7)$$

De manera análoga a la anterior, podemos probar que

$$\sup_{b \in B} d(b, A) \leq \sup_{b \in B} d(b, C) + \sup_{c \in C} d(c, A) \quad (1.8)$$

Con los resultados de las ecuaciones (1.7) y (1.8), obtenemos que

$$\begin{aligned} d_H(A, B) &= \text{máx} \left\{ \sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A) \right\} \\ &\leq \text{máx} \left\{ \sup_{a \in A} d(a, C), \sup_{c \in C} d(c, A) \right\} + \text{máx} \left\{ \sup_{b \in B} d(b, C), \sup_{c \in C} d(c, B) \right\} \\ &= d_H(A, C) + d_H(C, B). \end{aligned}$$

Se concluye que $d_H(A, B) \leq d_H(A, C) + d_H(C, B)$.

Por lo tanto, $d_H(\cdot)$ induce una distancia en $\mathcal{M}(X)$. ■

Cabe mencionar que comúnmente la literatura se refiere a la distancia de Hausdorff cuando $d(\cdot)$ es la norma Euclídeana. Sin embargo, como hemos visto, puede definirse para cualquier espacio métrico completo (X, d) .

1.2.1. Aplicaciones

Como hemos mencionado antes, la distancia de Hausdorff sirve para cuantificar la similitud entre conjuntos. Esto ha tenido una gran cantidad de aplicaciones en la época moderna. Desde de la década de los noventa múltiples trabajos se han enfocado en implementar la distancia de Hausdorff para comparar imágenes bidimensionales y tridimensionales. Sin embargo consideran ciertas restricciones a la imagen, como la identificación de un objeto sólo si éste está escalado [4, 5].

A pesar de que teóricamente la implementación es directa, de manera computacional puede llegar a causar inconvenientes debido a la cantidad de información que se tiene que evaluar. Es por ello que diferentes variaciones de la distancia de Hausdorff han sido implementadas con el fin de ser más eficientes numéricamente. Su principal uso ha sido en la detección de caras [6–8]. También ha tenido gran importancia en el ámbito de la medicina, donde es utilizada para segmentación de imágenes con el objetivo de aislar el área de interés para visualización o un posterior análisis, por ejemplo, en la detección de zonas cancerígenas [9].

En la siguiente figura se muestra un ejemplo del uso de la distancia de Hausdorff para encontrar un figura objetivo en una imagen.

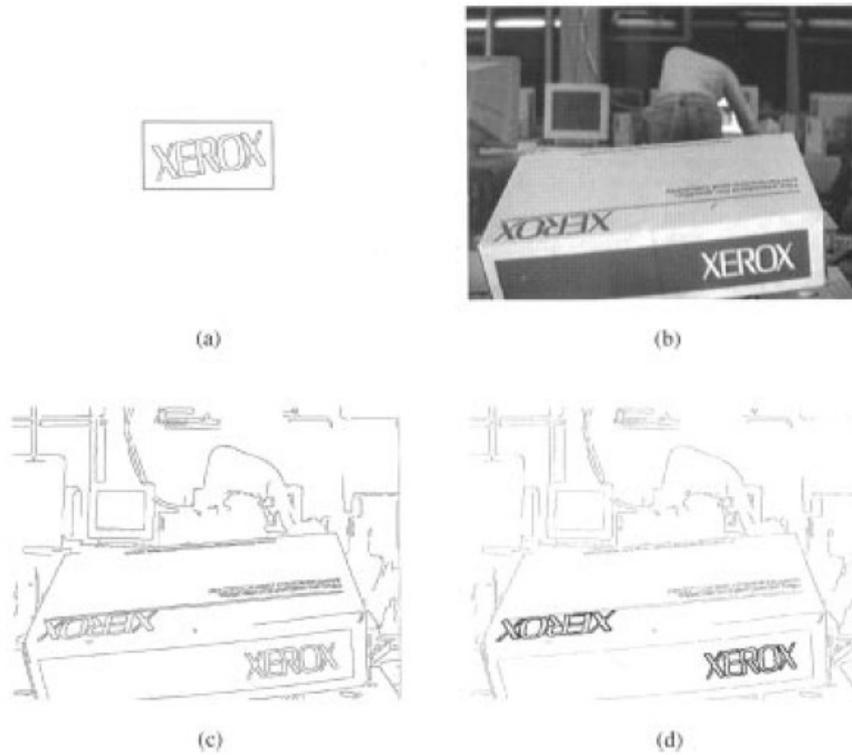


Figura 1.1: En el panel (a) se muestra el objeto a buscar en la imagen mostrada en el panel (b). En el panel (c) se muestra la imagen tras un procesamiento para resaltar los bordes encontrados. En el panel (d) se muestra en que parte de la imagen se encontró el objeto a buscar. Imagen tomada de [10].

Actualmente, las distintas variaciones y mejoras computacionales han permitido obtener una cuantificación robusta ante errores en la posición de los puntos de la imagen, así como ante la presencia de outliers y puntos faltantes [10].

Distancias de trayectoria óptima

El problema de la trayectoria más corta entre dos puntos es un problema muy interesante, tanto teóricamente como en aplicaciones, ya que surge naturalmente en problemas de optimización. En este capítulo definiremos una distancia en un conjunto de puntos aleatorios $Q_n = \{q_1, \dots, q_n\}$ muestreados de una densidad f con soporte en una variedad Riemanniana $\mathcal{M} \subseteq \mathbb{R}^d$, con la intención de que la definición de la distancia nos dé un balance que logre capturar la estructura geométrica de \mathcal{M} y la densidad f .

En este capítulo presentaremos la distancia PWSP (Power-Weighted Shortest-Path) y la distancia de Fermat. Empezaremos mostrando pruebas de convergencia para el caso donde el soporte de los datos es \mathbb{R}^d . Posteriormente iremos extendiendo los resultados al caso de variedades. Por último, daremos una breve descripción de la complejidad computacional de las distancias. Este aspecto nos brindará una comprensión más profunda de las implicaciones prácticas y los posibles desafíos en la implementación y aplicaciones de estas distancias en entornos reales.

2.1. Motivación

En los últimos años el uso de algoritmos de Machine Learning ha aumentado debido a su amplia variedad de aplicaciones y con esto nuevos retos han surgido. En el caso de análisis de datos multidimensionales se trata de ocupar la mayor cantidad de variables para hacer modelos más complejos que tengan un mejor desempeño, por ejemplo, en algoritmos de clusterización o en calibración de modelos estadísticos. Sin embargo, surge el problema de la dimensionalidad que consiste en que la distancia Euclideana no es buena opción debido a que en espacios de alta dimensión los puntos de un conjunto grande de datos están a una distancia similar [11]. Existen enfoques clásicos como análisis de componentes principales (PCA) o

escalamiento multidimensional (MDS) que dan buenos resultados cuando los datos se encuentran en un subespacio lineal de un espacio de dimensión más alta [12]. Esto hace posible cambiar la representación del conjunto de datos, que se encuentra en términos de un gran número de variables, a una descripción que requiera menos variables. Sin embargo, estos métodos sólo funcionan cuando el subespacio es aproximadamente lineal. Cuando el conjunto de datos está contenido en estructuras no lineales, los métodos de PCA y MDS fallan. Por ejemplo consideremos el conjunto de imágenes de un mismo objeto tomadas de diferentes ángulos bajo la misma condición de iluminación, como se muestra en la Fig. 2.1. Si una imagen puede representarse como una malla de píxeles de $n \times n$ en escalas de grises, donde cada píxel guarda la intensidad de luminosidad, entonces, cada posible imagen es un punto en \mathbb{R}^{n^2} . Sin embargo, el espacio de todas las imágenes del mismo objeto tiene una estructura no lineal, esto causa que ambos métodos fallen [12].

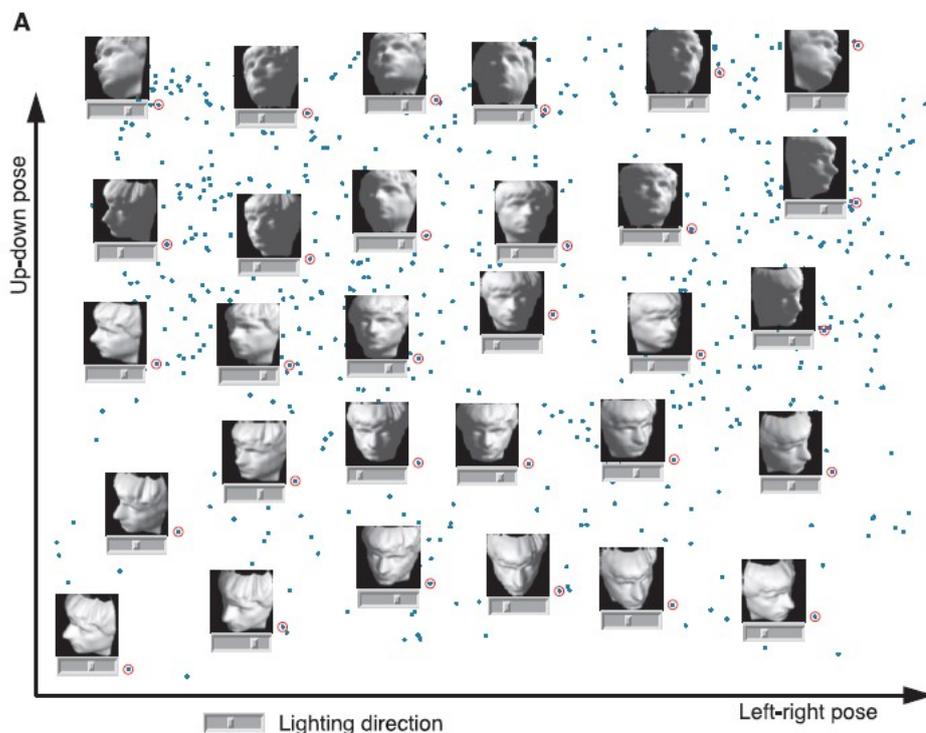


Figura 2.1: Un ejemplo de reducción de dimensión. Los datos corresponden a 698 imágenes, de 64×64 píxeles, de una cara desde diferentes ángulos. Imagen tomada de [12].

Existen diferentes métodos donde cada dato multidimensional es mapeado a un espacio bidimensional o tridimensional, de manera que sea más fácil ver algunas similitudes entre los datos [12–15]. Sin embargo, algunas veces estos métodos

fracasan al tratar de preservar la estructura local de los clusters y la geometría global de los datos [16]. Otro posible camino para resolver el problema consiste en elegir una distancia de manera adecuada. Diferentes trabajos se han basado en considerar la densidad y la geometría de los datos, tratando de equilibrar dos objetivos opuestos [17–22]. Estos últimos trabajos mencionados son en los que se basa la definición de distancia que se estudia en este trabajo.

Notación	Definición
\mathcal{X}_n	$\mathcal{X}_n = \{X_1, \dots, X_n\} \subset \mathbb{R}^d$, conjunto finito de datos.
(\mathcal{M}, g)	Variedad Riemaniana con tensor métrico asociado
$D_\alpha(x, y)$	Distancia PWSP, véase ecuación (2.1).
$\hat{d}_\alpha(x, y; \mathcal{X}_n)$	Distancia muestral PWSP dado \mathcal{X}_n , véase ecuación (2.3).
$\mathcal{T}_{x,y}$	Trayectoria asociada a $\hat{d}_\alpha(x, y; \mathcal{X}_n)$.
μ	Constante de tiempo de percolación
\mathcal{H}_λ	Proceso Poisson Homogéneo de intensidad λ .
$B(z; R)$	Bola de radio R respecto a la distancia especificada.
(q_i, q_j)	Trayectoria geodésica entre q_i y q_j .
(q_1, q_2, \dots, q_n)	Unión de segmentos de trayectorias geodésicas consecutivas.
$\overline{(q_i, q_j)}$	Recta entre q_i y q_j .
$\overline{(q_1, q_2, \dots, q_n)}$	Unión de segmentos de trayectorias rectas consecutivas.

Tabla 2.1: Notaciones usadas en este trabajo.

2.2. Distancia PWSP

A continuación daremos algunas notaciones y supuestos que serán usados a lo largo de este capítulo. Consideremos \mathcal{M} una variedad Riemaniana suave compacta d -dimensional sin frontera con tensor métrico g para $d > 1$. También, consideremos el espacio de probabilidad $(\mathcal{M}, \mathcal{B}(\mathcal{M}), \mathbb{P})$ donde \mathbb{P} es la ley de la distribución sobre los conjuntos de Borel $\mathcal{B}(\mathcal{M})$ del espacio \mathcal{M} . Asumamos que la ley de distribución tiene densidad de probabilidad f . Consideremos $\mathcal{X}_n = \{X_1, \dots, X_n\}$ una colección de variables aleatorias (v.a.) independientes idénticamente distribuidas (i.i.d.) sobre \mathcal{M} con función de densidad f . Por último, consideraremos $\alpha \geq 1$.

Definición 2.1 (Distancia PWSP) Sea f como anteriormente descrita, $x, y \in \mathcal{M}$. Definimos la distancia PWSP (Power-Weighted Shortest-Path) por sus siglas en ingles de x a y como

$$D_\alpha(x, y) = \inf_{\gamma} \int_0^1 f(\gamma_t)^{(1-\alpha)/d} \sqrt{g(\gamma'_t, \gamma'_t)} dt, \quad (2.1)$$

donde el ínfimo es tomado sobre todas las curvas contenidas en \mathcal{M} , $\gamma : [0, 1] \rightarrow \mathcal{M}$, que empiezan en $\gamma_0 = x$ y terminan en $\gamma_1 = y$. Cuando una curva alcanza el ínfimo, la denominamos *curva D_α -geodésica*.

Notemos que cuando $\alpha = 1$ el factor de la densidad desaparece y se obtiene la longitud de la curva geodésica clásica entre dos puntos. De cierta manera, el parámetro α es un equilibrio entre las trayectorias que pasan por regiones de alta densidad y entre las trayectorias que no se “desvían” tanto de la curva geodésica clásica. Además, podemos asegurar que el ínfimo se alcanza debido a que \mathcal{M} es compacto.

Para el caso $0 < \alpha < 1$, la trayectoria prefiere pasar por regiones de baja densidad incluso si eso implica grandes deformaciones, dando lugar a resultados contraintuitivos. Lo cual es opuesto a lo que se requiere en aplicaciones, principalmente de Machine Learning [23]. Cuando $\alpha \rightarrow \infty$, la trayectoria óptima es aquella que tenga el mínimo sobre el segmento más largo de las trayectorias. Este caso ha sido utilizado y ha demostrado tener gran precisión en algoritmos de clusterización, principalmente para mezcla de modelos no paramétricos [24].

Ahora, supongamos que no conocemos la función de densidad f y sólo tenemos una muestra de n observaciones. Una forma de estimar la ecuación (2.1) es de la siguiente forma.

Definición 2.2 (Distancia muestral PWSP) Sea un conjunto no vacío $\mathcal{X} \subset \mathcal{M}$. Para $x, y \in \mathcal{M}$, definimos la *distancia muestral PWSP* como

$$\hat{d}_\alpha(x, y; \mathcal{X}) = \inf \left\{ \sum_{i=1}^{k-1} D_1(a_i, a_{i+1})^\alpha : a_i \in \mathcal{X} \cup \{x, y\}, \right. \\ \left. (a_1, \dots, a_k) \text{ una trayectoria de } x \text{ a } y \right\}. \quad (2.2)$$

Notamos que la distancia muestral PWSP busca la trayectoria de menor costo que pase por los puntos $a_i \in \mathcal{X}_n$, (véase Fig. 2.2), donde el costo de cada segmento es la distancia geodésica entre a_i y a_{i+1} elevada a la α -potencia. En este caso se está ignorando de manera directa la influencia de la función de densidad, sin embargo, su relevancia está implícita en la muestra \mathcal{X}_n . Denotamos por $\mathcal{T}_{x,y}$ a la trayectoria entre x y y que alcanza el valor $\hat{d}_\alpha(x, y; \mathcal{X}_n)$.

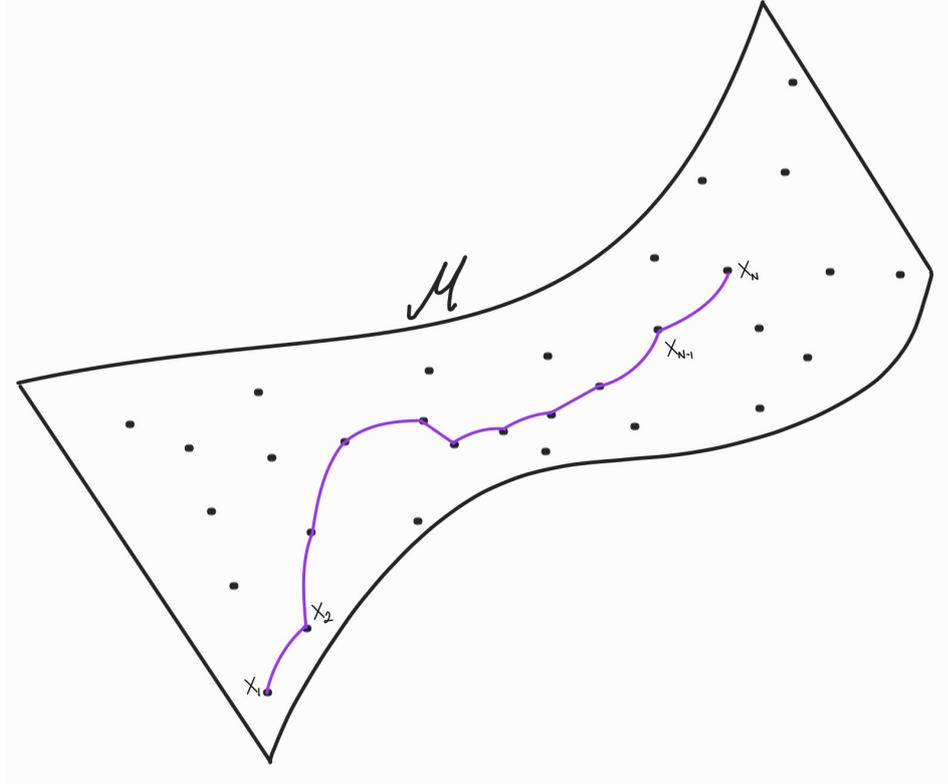


Figura 2.2: Representación esquemática de la trayectoria más corta, respecto a \hat{d}_α , entre x_1 y x_N . Como se muestra, la trayectoria entre x_i y x_{i+1} no es una línea recta necesariamente, sino la geodésica entre x_i y x_{i+1} que está determinada por la estructura \mathcal{M} .

Proposición 2.1 La distancia muestral PWSP es una distancia en \mathcal{X}_n .

Demostración. Probemos que la distancia muestral PWSP cumple las propiedades de la Def. 1.1.

1. Dados $x, y \in \mathcal{X}_n \subset \mathcal{M}$, $x \neq y$, existe una trayectoria $(q_{k_1}, \dots, q_{k_l})$, con $q_{k_i} \neq q_{k_j}$, $q_{k_1} = x$, $q_{k_l} = y$, tal que $\hat{d}_\alpha(x, y; \mathcal{X}) = \sum_{i=1}^{l-1} D_\alpha(q_{k_i}, q_{k_{i+1}}) > 0$. Por lo

tanto, $\hat{d}_\alpha(x, y; \mathcal{X}) > 0$.

Por otro lado, consideremos x, y tales que $\hat{d}_\alpha(x, y; \mathcal{X}) = 0$. Esto es equivalente a que exista una trayectoria $(q_{k_1}, \dots, q_{k_l})$ tal que $D_\alpha(q_{k_i}, q_{k_{i+1}}) = 0$ para todo i , que a su vez es equivalente a $q_{k_i} = q_{k_{i+1}}$. Por lo que $x = y$.

2. La igualdad $\hat{d}_\alpha(x, y; \mathcal{X}) = \hat{d}_\alpha(y, x; \mathcal{X})$ se cumple trivialmente debido a la simetría de D_α .
3. Consideremos $\hat{d}_\alpha(x, y; \mathcal{X})$ y $\hat{d}_\alpha(y, z; \mathcal{X})$ y sus respectivas trayectorias $(q_{k_1}, \dots, q_{k_l})$ y $(q_{j_1}, \dots, q_{j_m})$. Ya que $x = q_{k_1}, y = q_{k_l} = q_{j_1}, z = q_{j_m}$, existe una trayectoria $(q_{k_1}, \dots, q_{k_n}, q_{j_1}, \dots, q_{j_m})$ de x a z de longitud $\hat{d}_\alpha(x, y; \mathcal{X}) + \hat{d}_\alpha(y, z; \mathcal{X})$. Por lo tanto, $\hat{d}_\alpha(x, z; \mathcal{X}) \leq \hat{d}_\alpha(x, y; \mathcal{X}) + \hat{d}_\alpha(y, z; \mathcal{X})$.

Así, mostramos que cumple con la definición de distancia. ■

Notemos que dado \mathcal{X}_n , $\hat{d}_\alpha(\cdot)$ no cumple con la definición de distancia en \mathcal{M} . Por ejemplo, consideremos $\mathcal{M} = \mathbb{R}$, $\mathcal{X}_n = \{0, 1\}$ y $\alpha \geq 2$. Tomemos el punto $z = 0.5$. Veamos que $\hat{d}_\alpha(0, 1; \mathcal{X}_n) = 1$ y $\hat{d}_\alpha(0, 0.5; \mathcal{X}_n) + \hat{d}_\alpha(0.5, 1; \mathcal{X}_n) = 2(0.5)^\alpha < 1$. Entonces, la desigualdad del triángulo no se cumple. En general, no se cumple para todo punto $z \in (0, 1)$. Por lo tanto, no cumple con la definición de distancias para todo \mathbb{R} .

2.3. Constante del tiempo de percolación

A continuación enunciaremos un teorema de convergencia cuando \mathcal{X}_n es un *Proceso Poisson homogéneo* y $\mathcal{M} = \mathbb{R}^d$, cuya prueba puede consultarse en [25, 26].

Denotemos por \mathcal{H}_λ a un proceso homogéneo Poisson en \mathbb{R}^d de intensidad constante $\lambda > 0$. Sea $\hat{e}_1 = (1, 0, \dots, 0) \in \mathbb{R}^d$ un vector unitario. Por la invarianza traslacional y rotacional del proceso homogéneo Poisson, tenemos que para $x, y \in \mathbb{R}^d$ la distribución de $\hat{d}_\alpha(x, y; \mathcal{H}_\lambda)$ es igual a la distribución $\hat{d}_\alpha(0, t\hat{e}_1; \mathcal{H}_\lambda)$ con $t = |x - y|$ [27].

Teorema 2.1 Sea \mathcal{H}_1 un proceso homogéneo Poisson en \mathbb{R}^d con intensidad $\lambda = 1$.

Entonces,

$$\lim_{t \rightarrow \infty} \mathbb{E} \left[\frac{\hat{d}_\alpha(0, t\hat{e}_1; \mathcal{H}_1)}{t} \right] = \mu, \quad (2.3)$$

donde $\mu = \mu(d, \alpha)$ es una constante positiva que depende de sólo de d y α .

Teorema 2.2 Bajo las mismas condiciones del teorema anterior,

$$\lim_{t \rightarrow \infty} \frac{\hat{d}_\alpha(0, t\hat{e}_1; \mathcal{H}_1)}{t} = \mu, \quad \text{casi seguramente.} \quad (2.4)$$

Los teoremas anteriores pueden ser extendidos al caso donde el proceso Poisson está definido en un conjunto abierto convexo $S \subset \mathbb{R}^d$, aplicando un reescalamiento [28].

Teorema 2.3 Sea $S \subset \mathbb{R}^d$ un conjunto abierto convexo y \mathcal{H}_n un proceso Poisson con intensidad n dividida por el volumen de S . Entonces existe $\mu = \mu(d, \alpha)$ tal que para toda $x, y \in S$, se tiene que

$$\lim_{n \rightarrow \infty} \frac{\hat{d}_\alpha(x, y; \mathcal{H}_n)}{n^{(1-\alpha)/d}|x-y|} = \mu \quad \text{casi seguramente.} \quad (2.5)$$

Más aún, dados $b > 0$ y $\varepsilon > 0$, existen constantes positivas θ_1, β tales que si $|x - y| > b$, entonces

$$\mathbb{P} \left(\left| \frac{\hat{d}_\alpha(x, y; \mathcal{H}_n)}{n^{(1-\alpha)/d}|x-y|} - \mu \right| \geq \varepsilon \right) \leq \exp(-\theta_1 n^\beta), \quad (2.6)$$

para toda $n \geq 1$.

La constante μ es conocida como *constante del tiempo de percolación de primer pasaje* [29, 30]. En tal contexto, se tiene una malla de nodos en donde cada arista entre los nodos tiene asociada una variable aleatoria que representa el tiempo que toma ir de un nodo a otro. Entonces, μt caracteriza el tiempo asintótico más rápido posible que toma ir de un nodo a otro cuando están a distancia t . En nuestro caso, el peso entre los nodos está determinado por la distancia. Sin embargo, el esquema de estudio tiene cierta analogía.

2.4. Resultados de convergencia local

En esta sección mostraremos algunos resultados para el caso $\mathcal{M} = \mathbb{R}^d$, que después se extenderán para el caso de variedades Riemannianas. Notemos que para

este caso, $D_1(x, y) = |x - y|$ donde $|\cdot|$ es la norma Euclidea. A lo largo de esta sección, ocuparemos la notación

$$B(z; R) = \{x \in \mathbb{R}^d : |x - y| < R\}, \quad \beta = \frac{1}{d + 2\alpha}. \quad (2.7)$$

Empezaremos acotando la probabilidad de la distancia mínima entre los elementos de \mathcal{X}_n .

Lema 2.1 Sea \mathcal{H}_λ un proceso Poisson en un conjunto conexo $S \subset \mathbb{R}^d$ y \mathcal{X}_n una muestra. Sea (X_1, \dots, X_{k_i}) una trayectoria minimizante. Dado $\delta > 0$, existe una constante positiva $\theta_1 > 0$ tal que

$$\mathbb{P} \left(\max_{j < k_i} |X_{j+1} - X_j| > \delta \right) \leq \exp(-\theta_1 n).$$

Demostración. Definamos la función $h(X_i, X_j; \cdot) : S \rightarrow \mathbb{R}$,

$$h(X_j, X_{j+1}; u) = |X_j - u|^\alpha + |X_{j+1} - u|^\alpha - |X_j - X_{j+1}|^\alpha, \quad (2.8)$$

y consideremos $\theta(X_j, X_{j+1}) = \{u \in S : h(X_j, X_{j+1}; u) < 0\}$, es decir, $\theta(X_j, X_{j+1})$ es el conjunto de puntos u tal que que la trayectoria (X_j, u, X_{j+1}) es más corta que (X_j, X_{j+1}) , (véase Fig. 2.3). Entonces, existe una constante $\theta'_1 = \theta'_1(d, \alpha) > 0$ tal que existe un conjunto \mathcal{C} de volumen $\theta'_1 |X_j - X_{j+1}|^d$ contenido en $\theta(X_j, X_{j+1})$. La probabilidad que no haya un punto contenido en \mathcal{C} está acotada por $\exp(-cn)$ con $c > 0$. Si $\max_{j < k_i} |X_{j+1} - X_j| > \delta$, entonces

$$\mathbb{P} \left(\max_{j < k_i} |X_{j+1} - X_j| > \delta \right) \leq \exp(-\theta_1 n).$$

■

A continuación enunciamos una variación del lema anterior cuya demostración es similar y se dejará de lado.

Lema 2.2 Sea $z \in \mathbb{R}^d$ y $R > 0$. Supongamos que \mathcal{X}_n es una muestra en \mathbb{R}^d con función de densidad f y que $f_m = \inf\{f(u) : u \in B(z; R)\}$ es estrictamente positivo. Sea $b > 0$. Para cada par $1 \leq i < j \leq n$, definamos el evento $H_n(i, j)$ como la intersección de los siguientes eventos:

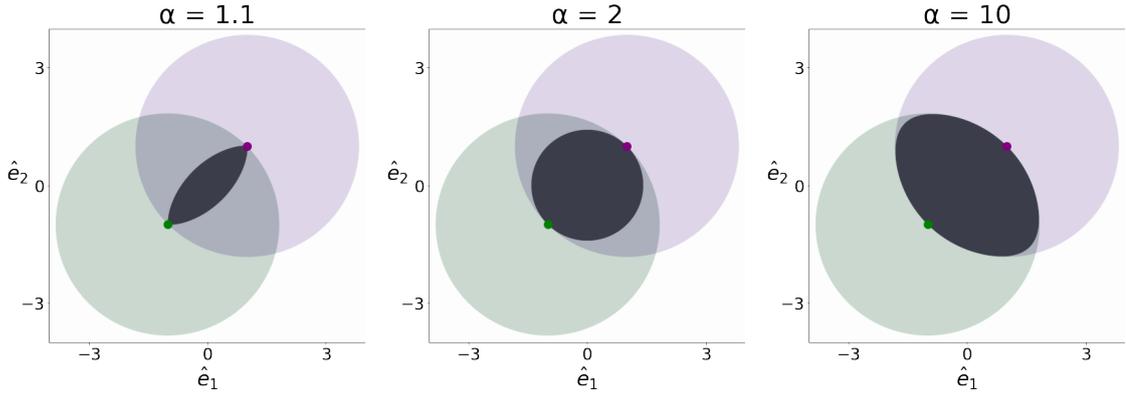


Figura 2.3: Ejemplo bidimensional donde $X_1 = (-1, -1)$ (punto color verde) y $X_2 = (1, 1)$ (punto color púrpura), donde la región $\theta(X_1, X_2)$ está sombreada de color gris. Para valores de α cercanos a 1, la región es “delgada”. Conforme α aumenta, la región se va ensanchando hasta alcanzar la intersección de los círculos centrados en X_1, X_2 , respectivamente, con radio $|X_1 - X_2|$.

- G.1: X_i y X_j están en $B(z; R)$,
- G.2: $|X_i - X_j| > b^\beta (nf_m)^{(\beta-1)/d}$,
- G.3: la trayectoria más cercana de X_i a X_j sobre \mathcal{X}_n respecto a \hat{d}_α no contiene a algún otro punto de \mathcal{X}_n .

Sea $F_n = \bigcap_{i,j} H_n(i, j)^c$. Entonces, existe una constante $\theta_1 > 0$ tal que

$$1 - \mathbb{P}(F_n) \leq \exp(-\theta_1 n^\beta). \quad (2.9)$$

Notemos que la condición de la cota inferior, $|X_i - X_j| > b^\beta (nf_m)^{(\beta-1)/d}$, decrece en función de n . Entonces, la expresión $1 - \mathbb{P}(F_n) \leq \exp(-\theta_1 n^\beta)$ nos dice que conforme n crece, la probabilidad de que haya un segmento de trayectoria “largo” también decrece exponencialmente.

A continuación enunciaremos dos lemas cuya demostración puede consultarse en [21, 25], donde se ocupa la teoría y resultados del Proceso Poisson [27]. Daremos dichos resultados por verdaderos para demostraciones posteriores. Además, en todo este trabajo, μ denotará a la constante del tiempo de percolación de primer pasaje.

Lema 2.3 Sean $z \in \mathbb{R}^d$, $R_2 > R_1 > 0$. Suponga que la función de probabilidad f es uniforme en $B(z; R_2)$, pero no necesariamente en el resto del soporte. Sean $\varepsilon > 0$

y $b \in (0, 2R_1)$. Entonces, existe $n_0 > 0$ tal que para toda $n \geq n_0$ y $x, y \in B(z; R_1)$ con $|x - y| \geq b$, se tiene que

$$\left| \frac{\mathbb{E} \left[\hat{d}_\alpha(x, y; \mathcal{X}_n \cap B(z; R_2)) \right]}{(nf(z))^{(1-\alpha)/d} |x - y|} - \mu \right| < \varepsilon. \quad (2.10)$$

Lema 2.4 Sean $z \in \mathbb{R}^d$, $R_2 > R_1 > 0$. Suponga que la función de probabilidad f es uniforme en $B(z; R_2)$, pero no necesariamente en el resto del soporte. Sean $\varepsilon > 0$ y $b \in (0, 2R_1)$. Entonces, existe una constante $\theta_3 > 0$ tal que para n suficientemente grande y para toda $x, y \in B(z; R_1)$ con $|x - y| \geq b$, se tiene que

$$\mathbb{P} \left(\left| \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n \cap B(z; R_2))}{(nf(z))^{(1-\alpha)/d} |x - y|} - \mu \right| > \varepsilon \right) \leq \exp(-\theta_3 n^\alpha). \quad (2.11)$$

Los Lemas 2.3 y 2.4 nos indican convergencia en media y en probabilidad, respectivamente. Lo que nos indica es que si tenemos una región donde la probabilidad es uniforme, podemos asegurar que la distancia PWSP, entre cualquiera dos puntos x, y , escalada por un factor $(nf(z))^{(1-\alpha)/d} |x - y|$ converge a una constante μ que sólo depende de la dimensión de la región y α . De cierta manera, dichos resultados son una extensión de los Teoremas 2.1, 2.2 y 2.3.

A continuación enunciamos un lema que nos da una cota para la diferencia de las distancias entre elementos contenidos en una bola de radio R . Dicha cota será usada más adelante.

Lema 2.5 Sea $z \in \mathbb{R}^d$, $R > 0$ y $b \in (0, R/2)$. Suponga que la función de probabilidad f es uniforme en $B(z; R)$, pero no necesariamente en el resto del soporte. Sea $\varepsilon > 0$ lo suficientemente pequeña para que $(\mu + \varepsilon)/2 < (\mu - \varepsilon)5/8$, es decir, $9\varepsilon < \mu$. Sea $v, u, x, y \in B(z; R)$. Si los siguientes eventos ocurren,

- A.1: el evento F_n del Lema 2.2,
- A.2: $|x - u|$ y $|y - v|$ es menor que $b^\beta (nf(z))^{(\beta-1)/d}$,

▪ A.3: $|x - y| \geq b$,

▪ A.4:

$$\frac{\hat{d}_\alpha(u, v; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}|u - v|} \leq \mu + \varepsilon, \quad (2.12)$$

entonces, existe $n_0 > 0$ tal que

$$\left| \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}|x - y|} - \frac{\hat{d}_\alpha(u, v; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}|u - v|} \right| < \varepsilon/2. \quad (2.13)$$

Demostración.

Por la desigualdad del triángulo podemos acotar el lado izquierdo de la ecuación (2.13) por

$$\begin{aligned} & \left| \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}|x - y|} - \frac{\hat{d}_\alpha(u, v; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}|x - y|} \right| \\ & + \left| \frac{\hat{d}_\alpha(u, v; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}|x - y|} - \frac{\hat{d}_\alpha(u, v; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}|u - v|} \right|. \end{aligned} \quad (2.14)$$

Por A.1 tenemos que cada segmento en la trayectoria más corta es de longitud a lo más $b^\beta(nf(z))^{(\beta-1)/d}$. Aunado a la condición A.2, tenemos que

$$|\hat{d}_\alpha(x, y; \mathcal{X}_n \cap B(z; R)) - \hat{d}_\alpha(u, v; \mathcal{X}_n \cap B(z; R))| \leq 2(b^\beta)^\alpha(nf(z))^{(\beta-1)\alpha/d}. \quad (2.15)$$

Y usando A.3 tenemos que

$$\left| \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}|x - y|} - \frac{\hat{d}_\alpha(u, v; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}|x - y|} \right| \leq \frac{2b^{\alpha\beta}(nf(z))^{\alpha\beta-1/d}}{b}. \quad (2.16)$$

Ya que $(\alpha\beta - 1) < 0$, existe n_1 tal que la ecuación anterior es acotada por $\varepsilon/4$ para toda $n \geq n_1$.

Por otro lado, de A.2 tenemos que $||u - v| - |x - y|| < 2b^\beta(nf(z))^{(\beta-1)/d}$, y

junto con A.4, el segundo término de (2.14) es acotado por

$$\begin{aligned}
& \left| \frac{\hat{d}_\alpha(u, v; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}|x-y|} - \frac{\hat{d}_\alpha(u, v; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}|u-v|} \right| \\
& \leq \frac{\hat{d}_\alpha(u, v; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}|u-v|} \left| \frac{|u-v| - |x-y|}{|x-y|} \right| \\
& \leq (\mu + \varepsilon) \frac{2b^\beta (nf(z))^{(\beta-1)/d}}{b}.
\end{aligned} \tag{2.17}$$

Ya que $(\beta - 1) < 0$, existe n_2 tal que la ecuación de arriba puede ser acotada por $\varepsilon/4$ para toda $n \geq n_2$.

En conclusión, la ecuación (2.13) se cumple tomando $n_0 = \max\{n_1, n_2\}$. ■

A continuación enunciaremos un lema, el cual asegura que con muy alta probabilidad para n suficientemente grande, la trayectoria entre dos elementos $x, y \in B(z; R/4)$ no se sale de $B(z; R)$.

Lema 2.6 Sea $z \in \mathbb{R}^d$, $R > 0$ y $b \in (0, R/2)$. Suponga que la función de probabilidad f es uniforme en $B(z; R)$, pero no necesariamente en el resto del soporte. Definamos los eventos:

- $E_n(\varepsilon)$ al evento tal que para toda $x \in B(z; R/4)$ y $u \notin B(z; R)$,

$$\frac{\hat{d}_\alpha(x, u; \mathcal{X}_n)}{(nf(z))^{(1-\alpha)/d}} > (\mu - \varepsilon) \frac{5}{8} R. \tag{2.18}$$

- $E'_n(\varepsilon)$ al evento tal que para toda $x, y \in B(z; R/4)$ con $|x - y| \geq b$ se tiene que

$$\hat{d}_\alpha(x, y; \mathcal{X}_n) = \hat{d}_\alpha(x, y; \mathcal{X}_n \cap B(z; R)) \tag{2.19}$$

y

$$\left| \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{(nf(z))^{(1-\alpha)/d}|x-y|} - \mu \right| \leq \varepsilon. \tag{2.20}$$

2. DISTANCIAS DE TRAYECTORIA ÓPTIMA

Sea $\varepsilon > 0$ lo suficientemente pequeño para que $(\mu + \varepsilon)/2 < (\mu - \varepsilon)5/8$, es decir, $9\varepsilon < \mu$. Entonces, existe $\theta_4 > 0$ tal que

$$1 - \mathbb{P}(E_n(\varepsilon) \cap E'_n(\varepsilon)) \leq \exp(-\theta_4 n^\beta). \quad (2.21)$$

Demostración.

Definamos $\xi_n = b^\beta (nf(z))^{(\beta-1)/d}$. Para un conjunto de puntos $\{w_i\}_{i=1}^m \in \mathbb{R}^d$, sea $\{B(w_i; \xi_n) : w_i \in B(z; R/4), 1 \leq i \leq m\}$ una cubierta abierta finita de $B(z; R/4)$ con m del orden de n , $m = O(n)$. De igual manera, para un conjunto de puntos $\{v_k\}_{k=1}^l \in \mathbb{R}^d$, sea $\{B(v_k; \xi_n) : v_k \in B(z; 7R/8), 1 \leq k \leq l\}$ una cubierta abierta finita de la frontera de $B(z; 7R/8)$ con $l = O(n)$.

Definamos los supuestos:

- B.1: el evento F_n , del Lema 2.2, ocurre.
- B.2: para toda w_i, v_k se tiene que

$$\left| \frac{\hat{d}_\alpha(w_i, v_k; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d} |w_i - v_k|} - \mu \right| \leq \varepsilon/2. \quad (2.22)$$

- B.3:

$$\left| \frac{\hat{d}_\alpha(w_i, w_j; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d} |w_i - w_j|} - \mu \right| \leq \varepsilon/2. \quad (2.23)$$

Afirmamos que el evento $E_n(\varepsilon)$, ecuación (2.18), ocurre bajo los supuestos B.1 y B.2.

Sea $r > 0$. Denotemos por $\hat{d}_\alpha(x; \mathcal{X}_n; r)$ a la distancia mínima, respecto a \hat{d}_α , de x a la frontera de $B(z; r)$, es decir,

$$\hat{d}_\alpha(x; \mathcal{X}_n; r) = \min_{|z-y|=r} \hat{d}_\alpha(x, y; \mathcal{X}_n). \quad (2.24)$$

Notemos que para $x \in B(z; R/4)$ y $u \notin B(z; R)$ se tiene que $\hat{d}_\alpha(x, u; \mathcal{X}_n) \geq \hat{d}_\alpha(x; \mathcal{X}_n; 7R/8)$. Entonces, para asegurar que se cumple la ecuación (2.18) es suficiente con mostrar que

$$\frac{\hat{d}_\alpha(x; \mathcal{X}_n; 7R/8)}{(nf(z))^{(1-\alpha)/d}} \geq (\mu - \varepsilon)5R/8, \quad (2.25)$$

Además, notemos que $\hat{d}_\alpha(x; \mathcal{X}_n; 7R/8) = \hat{d}_\alpha(x; \mathcal{X}_n \cap B(z; R); 7R/8)$, ya que si la trayectoria más corta a la frontera saliera de $B(z; R)$, la trayectoria ya habría pasado por la frontera, lo cual es una contradicción.

Para toda $x \in B(z; R/4)$, existe w_i tal que $|x - w_i| < \xi_n$, y para toda q en la frontera de $B(z; 7R/8)$, existe v_k tal que $|q - v_k| < \xi_n$. Notemos que se cumplen las hipótesis del Lema 2.5, entonces tenemos que

$$-\frac{\varepsilon}{4} + \frac{\hat{d}_\alpha(w_i, w_k; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}|w_i - v_k|} \leq \frac{\hat{d}_\alpha(x, q; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}|x - q|}. \quad (2.26)$$

Junto con el supuesto B.2, obtenemos que para n suficientemente grande

$$\frac{\hat{d}_\alpha(x, q; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}|x - q|} \geq \mu - \varepsilon, \quad (2.27)$$

para toda $x \in B(z; R/4)$ y para toda q que satisfaga $|z - q| = 7R/8$. Así, teniendo en cuenta que $|x - q| \geq 5R/8$, afirmamos que el evento $E_n(\varepsilon)$, ecuación (2.18), ocurre bajo los supuestos B.1 y B.2.

Ahora, mostraremos por contradicción que el evento $E'_n(\varepsilon)$ se cumple bajo las condiciones B.1, B.2 y B.3.

Sea $x, y \in B(z; R/4)$ y $|x - y| \geq b$. Entonces existen w_i, w_j tales que $|w_i - x| < \xi_n$ y $|w_j - y| < \xi_n$. Supongamos que la ecuación (2.19) no se cumple, es decir, $\hat{d}_\alpha(x, y; \mathcal{X}_n) \neq \hat{d}_\alpha(x, y; \mathcal{X}_n \cap B(z; R))$. Esto implica que la trayectoria entre x, y ha cruzado la frontera de $B(z; 7R/8)$, se sigue que $\hat{d}_\alpha(x; \mathcal{X}_n; 7R/8) \leq \hat{d}_\alpha(x, y; \mathcal{X}_n)$. Por la ecuación (2.25) se tendría que

$$(\mu - \varepsilon)5R/8 \leq \frac{\hat{d}_\alpha(x; \mathcal{X}_n; 7R/8)}{(nf(z))^{(1-\alpha)/d}} \leq \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{(nf(z))^{(1-\alpha)/d}}. \quad (2.28)$$

Pero por el otro lado, por Lema 2.5, tenemos que

$$\frac{\hat{d}_\alpha(x, y; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}|x - y|} \leq \frac{\hat{d}_\alpha(w_i, w_j; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}|w_i - w_j|} + \varepsilon/2, \quad (2.29)$$

y por B.3 tenemos que

$$\frac{\hat{d}_\alpha(w_i, w_j; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}|w_i - w_j|} \leq \mu + \varepsilon/2. \quad (2.30)$$

Además, teniendo en cuenta que $|x - y| < R/2$, da a lugar

$$\frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{(nf(z))^{(1-\alpha)/d}} \leq \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}} \leq (\mu + \varepsilon)R/2. \quad (2.31)$$

Sin embargo, recordemos que ε fue elegido lo suficientemente pequeño para asegurar que $(\mu + \varepsilon)/2 < (\mu - \varepsilon)5/8$. Entonces, por las ecuaciones (2.28) y (2.31) obtenemos una contradicción. Concluimos que las condiciones B.1, B.2 y B.3 implican la ecuación (2.19).

Ahora, consideremos

$$\begin{aligned} & \left| \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{(nf(z))^{(1-\alpha)/d}|x - y|} - \mu \right| \\ & \leq \left| \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{(nf(z))^{(1-\alpha)/d}|x - y|} - \frac{\hat{d}_\alpha(w_i, w_j; \mathcal{X}_n)}{(nf(z))^{(1-\alpha)/d}|w_i - w_j|} \right| \\ & \quad + \left| \frac{\hat{d}_\alpha(w_i, w_j; \mathcal{X}_n)}{(nf(z))^{(1-\alpha)/d}|w_i - w_j|} - \mu \right| \\ & \leq \varepsilon/2 + \varepsilon/2 = \varepsilon, \end{aligned} \quad (2.32)$$

donde hemos ocupado la condición de la ecuación (2.19), el Lema 2.5 para acotar el primer término y la condición B.3 para el segundo término. Así obtenemos que también se cumple la ecuación (2.20).

Por lo tanto, las condiciones B.1, B.2 y B.3 implican que se cumplen los eventos $E_n(\varepsilon)$ y $E'_n(\varepsilon)$. Entonces,

$$\begin{aligned} 1 - \mathbb{P}(E_n(\varepsilon) \cap E'_n(\varepsilon)) & \leq (1 - \mathbb{P}(F_b)) \\ & \quad + \sum_{w_i, v_k} \mathbb{P} \left(\left| \frac{\hat{d}_\alpha(w_i, v_k; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}|w_i - v_k|} - \mu \right| > \varepsilon/2 \right) \\ & \quad + \sum_{w_i, w_j} \mathbb{P} \left(\left| \frac{\hat{d}_\alpha(w_i, w_j; \mathcal{X}_n \cap B(z; R))}{(nf(z))^{(1-\alpha)/d}|w_i - w_j|} - \mu \right| > \varepsilon/2 \right). \end{aligned} \quad (2.33)$$

El primer término puede ser acotado por el Lema 2.2. Ambas sumas son de orden polinomial y cada término puede ser acotado por el Lema 2.4. De esta manera, obtenemos que

$$1 - \mathbb{P}(E_n(\varepsilon) \cap E'_n(\varepsilon)) \leq \exp(-\theta_4 n^\beta). \quad (2.34)$$

Los resultados anteriores dan paso a un lema donde la condición de uniformidad en $B(z; R)$ es relajada. Este resultado será usado para el caso en donde el soporte es una variedad. ■

Lema 2.7 Suponga que f es continua en $z \in \mathbb{R}^d$ y $f(z) > 0$. Sea $\varepsilon > 0$ lo suficientemente pequeña para que se cumpla que $(\mu + \varepsilon)/2 < (\mu - \varepsilon)5/8$, es decir, $9\varepsilon < \mu$. Para $R > 0$ y $b \in (0, R/2)$ definimos $H_n = H_n(z, R, \varepsilon, b)$ al evento que cumple

$$\left| \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{(nf(z))^{(1-\alpha)/d}|x-y|} - \mu \right| \leq \varepsilon \quad (2.35)$$

para toda $x, y \in B(z; R/4)$ con $|x - y| \geq b$, y simultáneamente

$$\frac{\hat{d}_\alpha(x, u; \mathcal{X}_n)}{(nf(z))^{(1-\alpha)/d}} \geq (\mu - \varepsilon)\frac{5}{8}R \quad (2.36)$$

para toda $x \in B(z; R/4)$ y $u \notin B(z; R)$. Entonces existe $R = R(z) > 0$ tal que para toda $b \in (0, R/2)$ se cumple que

$$1 - \mathbb{P}(H_n(z, R, \varepsilon, b)) \leq \exp(-\theta_5 n^\alpha), \quad (2.37)$$

con $\theta_5 > 0$ para n suficientemente grande.

Demostración.

Sean

$$f_m = \inf_{q \in B(z; R)} f(q) \quad \text{y} \quad f_M = \sup_{q \in B(z; R)} f(q). \quad (2.38)$$

Ya que f es continua en z podemos elegir a R lo suficientemente pequeña para que f_m y f_M sean valores cercanos a $f(z)$ para asegurar que se cumplen las siguientes desigualdades

$$\left(\mu + \frac{\varepsilon}{2}\right) \left(\frac{f(z)}{f_m}\right)^{(\alpha-1)/d} \leq \mu + \varepsilon, \quad (2.39)$$

$$\left(\mu - \frac{\varepsilon}{2}\right) \left(\frac{f(z)}{f_M}\right)^{(\alpha-1)/d} \geq \mu - \varepsilon, \quad (2.40)$$

$$f_M V_d R^d \leq 1, \quad (2.41)$$

donde V_d denota el volumen de una bola unitaria d -dimensional.

Para cada $X_i \in \mathcal{X}_n$, sea Y_i un punto arbitrario fuera de $B(z; R)$, y sea Z_i una v.a. Bernoulli con probabilidad de éxito $f_m/f(X_i)$. Definamos una nueva v.a.

$$X_i^m = \begin{cases} X_i & \text{if } X_i \notin B(z; R), \\ X_i Z_i + Y_i(1 - Z_i) & \text{if } X_i \in B(z; R). \end{cases} \quad (2.42)$$

Así $\mathcal{X}_n^m = \{X_1^m, \dots, X_n^m\}$ es una muestra que restringida a $B(z; R)$ es uniforme con intensidad f_m .

Por otro lado, para $X_i \in \mathcal{X}_n$, definimos otra nueva v.a. X_i^M como sigue. Sea $\sigma = \int (f_M - f(u)) du$ donde la integral es sobre $B(z; R)$. Por (2.41), $0 \leq \sigma \leq 1$. Sea \tilde{Y}_i un punto aleatorio dentro de $B(z; R)$ con función de densidad $(f_M - f(u))/\sigma$ para $u \in B(z; R)$, y sea \tilde{Z}_i una v.a. Bernoulli con $\mathbb{P}(\tilde{Z}_i = 1) = 1 - \sigma$. Definamos

$$X_i^M = \begin{cases} X_i & \text{si } X_i \in B(z; R), \\ X_i \tilde{Z}_i + \tilde{Y}_i(1 - \tilde{Z}_i) & \text{si } X_i \notin B(z; R). \end{cases} \quad (2.43)$$

Así $\mathcal{X}_n^M = \{X_1^M, \dots, X_n^M\}$ es una muestra que restringida a $B(z; R)$ es uniforme con intensidad f_M .

Definamos los eventos

- C.1: los eventos $E_n(\varepsilon/2)$ y $E'_n(\varepsilon/2)$ ocurren para \mathcal{X}_n^m ,
- C.2: los eventos $E_n(\varepsilon/2)$ y $E'_n(\varepsilon/2)$ ocurren para \mathcal{X}_n^M .

Consideremos que C.1 ocurre. Sean $x, y \in B(z; R/4)$ con $|x - y| \geq b$. Ya que $(\mathcal{X}_n^m \cap B(z; R)) \subset (\mathcal{X}_n \cap B(z; R))$, tenemos que

$$\hat{d}_\alpha(x, y; \mathcal{X}_n) \leq \hat{d}_\alpha(x, y; \mathcal{X}_n \cap B(z; R)) \leq \hat{d}_\alpha(x, y; \mathcal{X}_n^m \cap B(z; R)). \quad (2.44)$$

Entonces, por (2.19), tenemos que $\hat{d}_\alpha(x, y; \mathcal{X}_n) \leq \hat{d}_\alpha(x, y; \mathcal{X}_n^m)$, lo que nos lleva a

$$\frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{(nf(z))^{(1-\alpha)/d}|x-y|} \leq \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n^m)}{(nf_m)^{(1-\alpha)/d}|x-y|} \left(\frac{f(z)}{f_m} \right)^{\frac{\alpha-1}{d}}. \quad (2.45)$$

Por (2.20) y (2.39) tenemos que

$$\frac{\hat{d}_\alpha(x, y; \mathcal{X}_n^m)}{(nf_m)^{(1-\alpha)/d}|x-y|} \leq \mu + \varepsilon/2, \quad \left(\frac{f(z)}{f_m}\right)^{\frac{\alpha-1}{d}} \leq \frac{\mu + \varepsilon}{\mu + \varepsilon/2}. \quad (2.46)$$

Entonces, de (2.45) obtenemos

$$\frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{(nf(z))^{(1-\alpha)/d}|x-y|} \leq \mu + \varepsilon. \quad (2.47)$$

Lo que nos da la cota superior de la ecuación (2.35).

Ahora, supongamos que el evento C.2 ocurre. Sean $x \in B(z; R/4)$ y

$$v = \arg \min_{y:|z-y|=R} \hat{d}_\alpha(x, y; \mathcal{X}_n). \quad (2.48)$$

Entonces, $\hat{d}_\alpha(x, v; \mathcal{X}_n) = \hat{d}_\alpha(x, v; \mathcal{X}_n \cap B(z; R))$. De otra manera la trayectoria más cercana de x a v pasaría a través de otro punto en la frontera de $B(z; R)$, y esto contradice la elección de v . Ya que $(\mathcal{X}_n \cap B(z; R)) \subset (\mathcal{X}_n^M \cap B(z; R))$,

$$\hat{d}_\alpha(x, v; \mathcal{X}_n) = \hat{d}_\alpha(x, v; \mathcal{X}_n \cap B(z; R)) \geq \hat{d}_\alpha(x, v; \mathcal{X}_n^M \cap B(z; R)) = \hat{d}_\alpha(x, v; \mathcal{X}_n^M), \quad (2.49)$$

entonces

$$\frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{(nf(z))^{(1-\alpha)/d}} \geq \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n^M)}{(nf_M)^{(1-\alpha)/d}} \left(\frac{f(z)}{f_M}\right)^{\frac{\alpha-1}{d}}. \quad (2.50)$$

Debido a que $v \notin B(z; R)$, se sigue de la ecuación (2.18) que

$$\frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{(nf(z))^{(1-\alpha)/d}} \geq (\mu - \varepsilon/2) \frac{5}{8} R \left(\frac{f(z)}{f_M}\right)^{\frac{\alpha-1}{d}} \quad (2.51)$$

y por (2.40)

$$\frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{(nf(z))^{(1-\alpha)/d}} \geq (\mu - \varepsilon) \frac{5}{8} R. \quad (2.52)$$

Ahora, si $u \notin B(z; R)$, la trayectoria $\hat{d}_\alpha(x, u; \mathcal{X}_n)$ cruza la frontera de $B(z; R)$ en algún punto u' , entonces $\hat{d}_\alpha(x, u; \mathcal{X}_n) \geq d_\alpha(x, u'; \mathcal{X}_n) \geq \hat{d}_\alpha(x, v; \mathcal{X}_n)$ por la minimalidad de v y recordemos que $x \in B(z; R/4)$. Lo que implica la ecuación

(2.36) bajo la suposición de que el evento C.2 ocurre.

Ahora, mostraremos que la ecuación (2.35) se cumple bajo C.1 y C.2. Sean $x, y \in B(z; R/4)$. Entonces $\hat{d}_\alpha(x, y; \mathcal{X}_n) = d_\alpha(x, y; \mathcal{X}_n \cap B(z; R))$. De otra manera, si $\hat{d}_\alpha(x, y; \mathcal{X}_n)$ alcanza algún punto $u \notin B(z; R)$, entonces $\hat{d}_\alpha(x, u; \mathcal{X}_n) \leq \hat{d}_\alpha(x, y; \mathcal{X}_n)$, debido a que $|x - y| < R/2$, por la ecuación (2.47) tendríamos que

$$\frac{\hat{d}_\alpha(x, u; \mathcal{X}_n)}{R/2(nf(z))^{(1-\alpha)/d}} \leq \frac{\hat{d}_\alpha(x, u; \mathcal{X}_n)}{R/2(nf(z))^{(1-\alpha)/d}} \leq \frac{\hat{d}_\alpha(x, u; \mathcal{X}_n)}{|x - y|(nf(z))^{(1-\alpha)/d}} \leq \mu + \varepsilon, \quad (2.53)$$

y por la condición $(\mu + \varepsilon)/2 < (\mu - \varepsilon)5/8$, se tendría que

$$\frac{\hat{d}_\alpha(x, u; \mathcal{X}_n)}{R/2(nf(z))^{(1-\alpha)/d}} < (\mu - \varepsilon)\frac{5}{8}R, \quad (2.54)$$

lo cual entra en contradicción con la ecuación (2.36).

Ahora, podemos repetir un argumento similar usado para obtener (2.47) y mostrar que

$$\frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{(nf(z))^{(1-\alpha)/d}|x - y|} \geq \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n^M)}{(nf_M)^{(1-\alpha)/d}|x - y|} \left(\frac{f(z)}{f_M} \right)^{\frac{\alpha-1}{d}} \geq (\mu - \varepsilon). \quad (2.55)$$

Entonces, con (2.47) y (2.55), obtenemos (2.35). Así, mostramos que la ocurrencia de los eventos C.1 y C.2 asegura las condiciones necesarias enunciadas en el Lema 2.7. Las muestras \mathcal{X}_n^m y \mathcal{X}_n^M son independientes entre sí. Aplicando Lema 2.6 a cada una de las muestras, obtenemos que

$$1 - \mathbb{P}(H_n(z, R, \varepsilon, b)) \leq \exp(-\theta_5 n^\alpha), \quad (2.56)$$

con $\theta_5 > 0$ para n suficientemente grande, lo cual establece el Lema 2.7. ■

2.5. Resultados en variedades

En esta sección extendemos los resultados de la sección pasada al caso donde el soporte de la función es una variedad Riemanniana suave compacta. Para $z \in \mathcal{M}$

y $R > 0$ consideramos

$$B(z; R) = \{u \in \mathcal{M} : D_1(z, u) < R\}, \quad (2.57)$$

donde $D_1(\cdot)$ es la distancia geodésica descrita en la ecuación (2.1).

Lema 2.8 Sea (\mathcal{M}, g_1) una variedad Riemanniana con tensor métrico g_1 . Sean $z \in \mathcal{M}$ y $\varepsilon > 0$. Suponga que $f(z) > 0$ y que f es continua en z . Para $R > 0$ y $b \in (0, 2R)$, denotamos por $J_n(B(z; R), \varepsilon, b)$ al evento que cumple

- D.1: Si la trayectoria más corta pasa a través de $B(z; R)$ entonces los segmentos de trayectoria en $B(z; R)$, ponderados por D_1 , miden a lo más $b^\beta (nf(z))^{(\beta-1)/d}$.
- D.2: para cualesquiera $x, y \in B(z; R)$ con $D_1(x, y) \geq b$ se cumple que

$$\left| \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{n^{(1-\alpha)/d} D_\alpha(x, y)} - \mu \right| \leq \varepsilon \quad (2.58)$$

Entonces, para n suficientemente grande, existe $R = R(z) > 0$ tal que para toda $b \in (0, 2R)$ existe $\theta_6 > 0$ tal que

$$1 - \mathbb{P}(J_n(B(z; R), \varepsilon, b)) \leq \exp(-\theta_6 n^\beta). \quad (2.59)$$

Demostración.

Mostraremos el lema, mostrando que la probabilidad de que ocurra cada una de las dos condiciones anteriores está acotada de manera exponencial como en la ecuación (2.59). Acotar la probabilidad de la primera condición es muy similar al argumento dado en la prueba del Lema 2.2. La diferencia es que en vez de tomar la distancia Euclideana, tomamos la distancia $D_1(\cdot)$.

Para acotar la probabilidad de la segunda condición elijamos $\delta > 0$ lo suficientemente pequeña para asegurar que

$$2(1 + \delta)^\alpha \leq \frac{5}{2}(1 - \delta)^\alpha, \quad (2.60)$$

$$\left(\frac{1+\delta}{1-\delta}\right)^\alpha (\mu + \varepsilon/2) \leq \mu + \varepsilon, \quad (2.61)$$

$$\left(\frac{1-\delta}{1+\delta}\right)^\alpha (\mu - \varepsilon/2) \geq \mu - \varepsilon. \quad (2.62)$$

Definamos $U = B(z; 4R)$. Ya que f es continua, podemos elegir $R > 0$ lo suficientemente pequeña para asegurar que existe una transformación $\phi : U \subset \mathcal{M} \rightarrow V \subset \mathbb{R}^d$ tal que el evento $H_n(\phi(z), 4R, \varepsilon/2, b/(1+\delta))$, definido por las ecuaciones (2.35) y (2.36), satisface la ecuación (2.37) y además satisface que

$$(1-\delta)^d \sup_U f \leq f(z) = f(\phi(z)) \leq (1+\delta)^d \inf_U f, \quad (2.63)$$

$$1-\delta \leq \frac{D_1(u, v)}{|\phi(u) - \phi(v)|} \leq 1+\delta, \quad (2.64)$$

para todo $u, v \in U$, con $u \neq v$ [31]. Afirmamos que $H_n(\phi(z), 4R, \varepsilon/2, b/(1+\delta))$ implica D.2.

Sean $x, y \in B(z; R) \subset U$. Por la desigualdad del triángulo tendremos que $D_1(x, y) \leq D_1(x, z) + D_1(z, y) \leq 2R$ por lo que la curva D_1 -geodésica está contenida en U . Entonces, tendremos que

$$D_\alpha(x, y) \leq D_1(x, y) \left(\inf_U f\right)^{(1-\alpha)/d}. \quad (2.65)$$

Además, notemos que la curva D_α -geodésica está contenida en U por lo que

$$D_\alpha(x, y) \geq D_1(x, y) \left(\sup_U f\right)^{(1-\alpha)/d}. \quad (2.66)$$

Supongamos ahora que la curva D_α -geodésica no está contenida en U . Entonces, la distancia $D_\alpha(x, y)$ es al menos $3R \left(\sup_U f\right)^{(1-\alpha)/d}$ ya que $x, y \in B(z; R)$ y $U = B(z; 4R)$. Sin embargo, recordemos que $D_1(x, y) \leq 2R$ y por la ecuación (2.63), tenemos que

$$D_\alpha(x, y) \leq D_1(x, y) \left(\inf_U f\right)^{(1-\alpha)/d} \leq 2R \left(\frac{1+\delta}{1-\delta}\right)^{\alpha-1} \left(\sup_U f\right)^{(1-\alpha)/d}, \quad (2.67)$$

y usando el resultado de la ecuación (2.60) obtenemos que

$$D_\alpha(x, y) \leq \frac{5}{2}R \left(\sup_U f \right)^{(1-\alpha)/d}, \quad (2.68)$$

lo cual entra en contradicción con la suposición de que la curva D_α -geodésica se sale de U .

Ahora mostraremos que $\hat{d}_\alpha(x, y; \mathcal{X}_n) = \hat{d}_\alpha(x, y; \mathcal{X}_n \cap U)$, es decir, la trayectoria más corta entre $x, y \in B(z; R)$ está contenida en U cuando $H_n(\phi(z), 4R, \varepsilon/2, b/(a+\delta))$ ocurre. Para este argumento, denotemos por \hat{d}_α a la distancia en U y por \hat{d}'_α a la distancia en V .

Por contradicción, supongamos que la trayectoria $\hat{d}_\alpha(x, y; \mathcal{X}_n)$ sale de U . Entonces la trayectoria correspondiente en V empieza en $\phi(x)$ y sale de V , entonces por la ecuación (2.36) tenemos que

$$\frac{\hat{d}'_\alpha(\phi(x), \phi(y); \phi(\mathcal{X}_n))}{(nf(z))^{(1-\alpha)/d}} \geq (\mu - \varepsilon)5R/2, \quad (2.69)$$

y usando el resultado la condición (2.64) tenemos que

$$\frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{(nf(z))^{(1-\alpha)/d}} \geq (\mu - \varepsilon)(1 - \delta)^\alpha 5R/2. \quad (2.70)$$

Por otro lado, recordemos que $\phi(x), \phi(y) \in B(\phi(z); R) \subset V$, lo que implica $|\phi(x) - \phi(y)| < 2R$. Entonces, por la ecuación (2.35) tenemos que

$$\frac{\hat{d}'_\alpha(\phi(x), \phi(y); \phi(\mathcal{X}_n \cap U))}{(nf(\phi(z)))^{(1-\alpha)/d}} \leq (\mu + \varepsilon)2R, \quad (2.71)$$

y usando la ecuación (2.64) obtenemos

$$\frac{\hat{d}_\alpha(x, y; \mathcal{X}_n \cap U)}{(nf(z))^{(1-\alpha)/d}} \leq (\mu + \varepsilon)(1 + \delta)^\alpha 2R. \quad (2.72)$$

Entonces, por las ecuaciones (2.70) y (2.72) tenemos que

$$(\mu - \varepsilon)(1 - \delta)^\alpha 5R/2 \leq \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{(nf(z))^{(1-\alpha)/d}} < \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n \cap U)}{(nf(z))^{(1-\alpha)/d}} \leq (\mu + \varepsilon)(1 + \delta)^\alpha 2R. \quad (2.73)$$

Sin embargo, la ecuación anterior entra en contradicción con la condición impuesta en la ecuación (2.60). Por lo tanto, concluimos que $\hat{d}_\alpha(x, y; \mathcal{X}_n) = \hat{d}_\alpha(x, y; \mathcal{X}_n \cap U)$ para toda $x, y \in B(z; R)$.

Ahora, queremos mostrar $H_n(\phi(z), 4R, \varepsilon/2, b/(1 + \delta))$ implica D.3, es decir, que se cumple la ecuación (2.58). Veamos que por los resultados de las ecuaciones (2.65), (2.64), (2.63) tenemos que

$$\begin{aligned} \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{n^{(1-\alpha)/d} D_\alpha(x, y)} &\geq \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{(n \sup_U f)^{(1-\alpha)/d} D_1(x, y)} \\ &\geq \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{(1 + \delta)^\alpha |\phi(x) - \phi(y)| (nf(\phi(z)))^{(1-\alpha)/d}}. \end{aligned} \quad (2.74)$$

Utilizamos el resultado de la ecuación (2.64) para relacionar la distancia $\hat{d}_\alpha(x, y; \mathcal{X}_n \cap U)$ en U y $\hat{d}_\alpha(\phi(x), \phi(y); \phi(\mathcal{X}_n \cap U))$ en V . Además, por la definición de H_n , ecuación (2.35), y la condición (2.62) obtenemos

$$\begin{aligned} \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{(1 + \delta)^\alpha |\phi(x) - \phi(y)| (nf(\phi(z)))^{(1-\alpha)/d}} &\geq \left(\frac{1 - \delta}{1 + \delta} \right)^\alpha \frac{\hat{d}_\alpha(\phi(x), \phi(y); \phi(\mathcal{X}_n \cap U))}{(nf(\phi(z)))^{(1-\alpha)/d} |\phi(x) - \phi(y)|} \\ &\geq \left(\frac{1 - \delta}{1 + \delta} \right)^\alpha (\mu - \varepsilon/2) \geq \mu - \varepsilon. \end{aligned} \quad (2.75)$$

Por lo tanto, $\hat{d}_\alpha(x, y; \mathcal{X}_n)/n^{(1-\alpha)/d} D_\alpha(x, y) \geq \mu - \varepsilon$. Para la cota superior el procedimiento es análogo. De esta manera, aseguramos que H_n implica D.2 y por lo tanto se cumple la cota de la ecuación (2.59). ■

A continuación mostraremos que si el resultado del Lema 2.8 se cumple localmente para cada punto en \mathcal{M} , entonces se tiene una convergencia global de la longitud de la curva.

Teorema 2.4 Suponga que \mathcal{M} es compacto, y f es una función continua con

$\inf_{\mathcal{M}} f > 0$. Sea $b > 0$, $x, y \in \mathcal{M}$ tal que $D_1(x, y) \geq b$, y $\varepsilon > 0$. Entonces, existen constantes $\mu = \mu(d, \alpha) > 0$ y $\theta_0 > 0$ que satisfacen lo siguiente

$$\mathbb{P} \left(\sup_{x, y} \left| \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{n^{(1-\alpha)/d} D_\alpha(x, y)} - \mu \right| \right) \leq \exp(-\theta_0 n^{1/(d+2\alpha)}), \quad (2.76)$$

para n suficientemente grande, donde el supremo es tomado respecto a los elementos de \mathcal{M} con $D_1(x, y) \geq b$.

Demostración.

Para cada $w \in \mathcal{M}$ puede ser asociada una $R(w) > 0$ y $V(w) = B(w; 3R(w))$ tal que el evento $J_n(V(w), \varepsilon/2, b)$, definido en el Lema 2.8, cumple la ecuación (2.59).

Por compacidad, existe una colección finita $\{w_i \in \mathcal{M}\}_{i=1}^m$ tal que $\bigcup_{i=1}^m B(w_i; R(w_i))$ es una cubierta finita de \mathcal{M} . Definamos $R_i = R(w_i)$, $U_i = B(w_i; R_i)$ y $V_i = V(w_i)$, $i = 1, \dots, m$.

Consideremos $x, y \in \mathcal{M}$ y la trayectoria asociada a $\hat{d}_\alpha(x, y; \mathcal{X}_n)$ la cual denotamos por $\mathcal{T}_{x,y}$. Recordemos que $\mathcal{T}_{x,y}$ es una unión de segmentos de D_1 -geodésicas. Sin pérdida de generalidad, consideremos $x \in U_1$ y definamos $z_1 = x$. Si la trayectoria sale de V_1 , entonces un punto $z_2 \in \mathcal{T}_{x,y}$ puede ser elegido tal que $z_2 \notin U_1$ y $D_1(z_1, z_2) \geq R_1$. De igual manera, si la trayectoria sale de V_2 , existe $z_3 \in \mathcal{T}_{x,y}$ tal que $z_3 \notin U_2$ y $D_1(z_2, z_3) > R_2$. Repetimos el mismo proceso hasta que la trayectoria termine en $y \in V_k$, $y = z_{k+1}$. Entonces la trayectoria (z_1, \dots, z_k) cumple que el segmento $(z_i, z_{i+1}) \in V_i$ y $D_1(z_i, z_{i+1}) > R_i$ para toda $i = 1, \dots, k-1$. El valor de k es finito pues estamos considerando una cubierta finita por lo que $k \leq m$. Además, notemos que los puntos z_i 's no necesariamente pertenecen al conjunto \mathcal{X}_n .

Definamos $R^* = \min\{R_1, \dots, R_k\}$. Ya que estamos suponiendo que el evento $J_n(V_i, \varepsilon/2, b)$ ocurre, podemos ocupar el resultado de la ecuación (2.58),

$$(\mu - \varepsilon) D_\alpha(z_i, z_{i+1}) \leq \frac{\hat{d}_\alpha(z_i, z_{i+1}; \mathcal{X}_n)}{n^{(1-\alpha)/d}} \quad i = 1, \dots, k-1, \quad (2.77)$$

de manera que

$$\begin{aligned}
 (\mu - \varepsilon)D_\alpha(x, y) &\leq (\mu - \varepsilon) \sum_{i=1}^{k-1} D_\alpha(z_i, z_{i+1}) \\
 &\leq \sum_{i=1}^{k-1} \frac{\hat{d}_\alpha(z_i, z_{i+1}; \mathcal{X}_n)}{n^{(1-\alpha)/d}} \\
 &\leq \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{n^{(1-\alpha)/d}},
 \end{aligned} \tag{2.78}$$

donde hemos ocupado que D_α cumple la desigualdad del triángulo, el resultado de la ecuación (2.77), y que $a^\alpha + b^\alpha \leq (a + b)^\alpha$ para la primera, segunda y tercera desigualdad, respectivamente. Los eventos $J_n(V_i, \varepsilon/2, b)$ satisfacen (2.59) para toda $i = 1, \dots, k$, por lo tanto

$$1 - \mathbb{P} \left(\inf_{x, y} \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{n^{(1-\alpha)/d} D_\alpha(x, y)} - \mu \geq -\varepsilon \right) \leq k \exp(-\theta_6 n^\beta). \tag{2.79}$$

Ahora calculemos la probabilidad de la cota inferior. Recordemos los puntos z_i no pertenecen a \mathcal{X}_n y estamos considerando que $z_i \in \mathcal{T}_{x, y}$. Por lo que en general $\sum_{i=1}^k \hat{d}_\alpha(z_i, z_{i+1}; \mathcal{X}_n) \leq \hat{d}_\alpha(x, y; \mathcal{X}_n)$. Pero notemos que a partir de los caminos más cortos de z_{i-1} a z_i y de z_i a z_{i+1} se puede crear un camino entre z_{i-1} a z_{i+1} quitando el punto z_i . Esto puede ser aplicado múltiples veces y seguir obteniendo una trayectoria de x a y . Ya que el Lema (2.8) aplica en cada V_i , la D_1 -distancia de z_i a z_{i+1} es a lo más de $b^\beta (n \inf f)^{(\beta-1)/d}$. De manera que cada “recorte” de la trayectoria puede tener un costo adicional de a lo más $(2b^\beta)^\alpha (n \inf f)^{(\beta-1)\alpha/d}$. De manera que

$$\hat{d}_\alpha(x, y; \mathcal{X}_n) \leq \sum_{i=1}^k \left(\hat{d}_\alpha(z_i, z_{i+1}; \mathcal{X}_n) \right) + k(2b^\beta)^\alpha (n \inf f)^{(\beta-1)\alpha/d}. \tag{2.80}$$

Si el evento $J_n(V_i, \varepsilon/2, \min\{R^*, b\})$ sucede para cada V_i entonces

$$n^{(\alpha-1)/d} \hat{d}_\alpha(x, y; \mathcal{X}_n) \leq D_\alpha(x, y)(\mu + \varepsilon/2) + k(2b^\beta)^\alpha n^{(\alpha-1)/d} (n \inf f)^{(\beta-1)\alpha/d}. \tag{2.81}$$

Para n suficientemente grande, tenemos que

$$\frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{n^{(1-\alpha)/d} D_\alpha(x, y)} - \mu \leq \varepsilon. \quad (2.82)$$

De manera que el resultado del teorema se cumple al aplicar el Lema 2.8 a cada V_i . ■

A continuación mostramos un resultado análogo al Teo. 2.3 en el cual considerábamos que la muestra provenía de un proceso Poisson Homogéneo. En el caso de una variedad tenemos que considerar $D_\alpha(x, y)$ en vez de $|x - y|$. Sin embargo, en ambos obtenemos convergencia casi segura.

Teorema 2.5 Suponga que \mathcal{M} es completo y que f es continua con $f(z) > 0$ para toda $z \in \mathcal{M}$. Sea $x, y \in \mathcal{M}$. Entonces,

$$\lim_{n \rightarrow \infty} \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{n^{(1-\alpha)/d} D_\alpha(x, y)} = \mu \quad \text{casi seguramente.} \quad (2.83)$$

Demostración.

Consideremos el conjunto

$$\mathcal{A} = \{u \in \mathcal{M} : (\mu - \varepsilon) D_\alpha(x, u) \leq (\mu - \varepsilon) D_\alpha(x, y)\}. \quad (2.84)$$

De igual manera, definamos $R_i = R(w_i)$, $U_i = B(w_i; R_i)$ y $V_i = V(w_i)$, $i = 1, \dots, m$. Existe una colección finita $\{w_i \in \mathcal{A}\}_{i=1}^m$ tal que $\bigcup_{i=1}^m B(w_i; R_i)$ es una cubierta finita de \mathcal{A} y los eventos $J_n(V_i, \varepsilon/2, b)$, definido en el Lema 2.8, cumple la ecuación (2.59). Por los mismos argumentos del Teorema anterior, para n suficientemente grande tenemos que

$$n^{(\alpha-1)/d} \hat{d}_\alpha(x, y; \mathcal{X}_n) \leq (\mu + \varepsilon) D_\alpha(x, y), \quad (2.85)$$

y si la trayectoria asociada $\mathcal{T}_{x,y}$ a $\hat{d}_\alpha(x, y; \mathcal{X}_n)$ no sale de \mathcal{A} , entonces

$$n^{(\alpha-1)/d} \hat{d}_\alpha(x, y; \mathcal{X}_n) \geq (\mu - \varepsilon) D_\alpha(x, y). \quad (2.86)$$

A continuación argumentaremos por que la trayectoria $\mathcal{T}_{x,y}$ no sale de \mathcal{A} , por

contradicción, sea $u = \arg \min_{u'} \hat{d}_\alpha(x, u'; \mathcal{X}_n)$ donde el mínimo es tomado sobre la frontera de \mathcal{A} . Ya que la trayectoria se sale de \mathcal{A} , tenemos que $\hat{d}_\alpha(x, y; \mathcal{X}_n) > \hat{d}_\alpha(x, u; \mathcal{X}_n)$. Además, la trayectoria $\mathcal{T}_{x,u}$ está contenida en \mathcal{A} , entonces cumple con la ecuación (2.86) y con la igualdad $(\mu - \varepsilon)D_\alpha(x, u) = (\mu + \varepsilon)D_\alpha(x, y)$. Considerando la ecuación (2.85), tenemos que

$$\begin{aligned}
 (\mu + \varepsilon)D_\alpha(x, y) &\geq \frac{\hat{d}_\alpha(x, y; \mathcal{X}_n)}{n^{(1-\alpha)/d}} \\
 &> \frac{\hat{d}_\alpha(x, u; \mathcal{X}_n)}{n^{(1-\alpha)/d}} \\
 &\geq (\mu - \varepsilon)D_\alpha(x, u) \\
 &= (\mu + \varepsilon)D_\alpha(x, y).
 \end{aligned} \tag{2.87}$$

Lo cual es una contradicción.

Por lo tanto, usando el Lema 2.8 y el Lema de Borel-Cantelli, obtenemos el resultado de la ecuación (2.83). ■

2.6. Distancia de Fermat

Como mencionamos al principio del capítulo, diferentes trabajos se han basado en considerar la densidad y la geometría de los datos. En esta sección detallaremos los resultados de [28, 32], en donde las distancias estudiadas son muy similares a las presentadas en este capítulo salvo pequeños detalles. Estos trabajos, proponen la *distancia de Fermat* y la *distancia muestral de Fermat*, las cuales denotaremos por D'_α y \hat{d}'_α , respectivamente. En el caso donde $\mathcal{M} = \mathbb{R}^d$, la distancia de Fermat es la misma que la usada en este trabajo.

El nombre de la “distancia de Fermat” surge del principio de Fermat en óptica, el cual dice que la luz siempre elige la trayectoria que le permite llegar más rápido entre dos puntos que no necesariamente es una línea recta. En el contexto de la clusterización de datos, la distancia de Fermat busca la rutas más cortas para conectar puntos, tratando de pasar por zonas de mayor densidad.

Definición 2.3 (Distancia muestral de Fermat) Sean $\alpha \geq 1$, \mathcal{X} un conjunto

no vacío y $x, y \in \mathbb{R}^d$, definimos la *distancia muestral de Fermat* como

$$\hat{d}'_{\alpha}(x, y; \mathcal{X}) = \inf \left\{ \sum_{i=1}^{k-1} |a_{i+1} - a_i|^{\alpha} : a_i \in \mathcal{X} \cup \{x, y\}, \right. \\ \left. (a_1, \dots, a_k) \text{ una trayectoria de } x \text{ a } y \right\}. \quad (2.88)$$

La diferencia es en el peso que le damos a los segmentos de trayectoria. En este trabajo consideramos el peso $D_1(a_i, a_{i+1})$ y en la caso de la Definición 2.3, el peso asignado es la norma Euclideana $|a_{i+1} - a_i|^{\alpha}$. Sin embargo, se obtienen resultados similares de convergencia, siguiendo pruebas análogas.

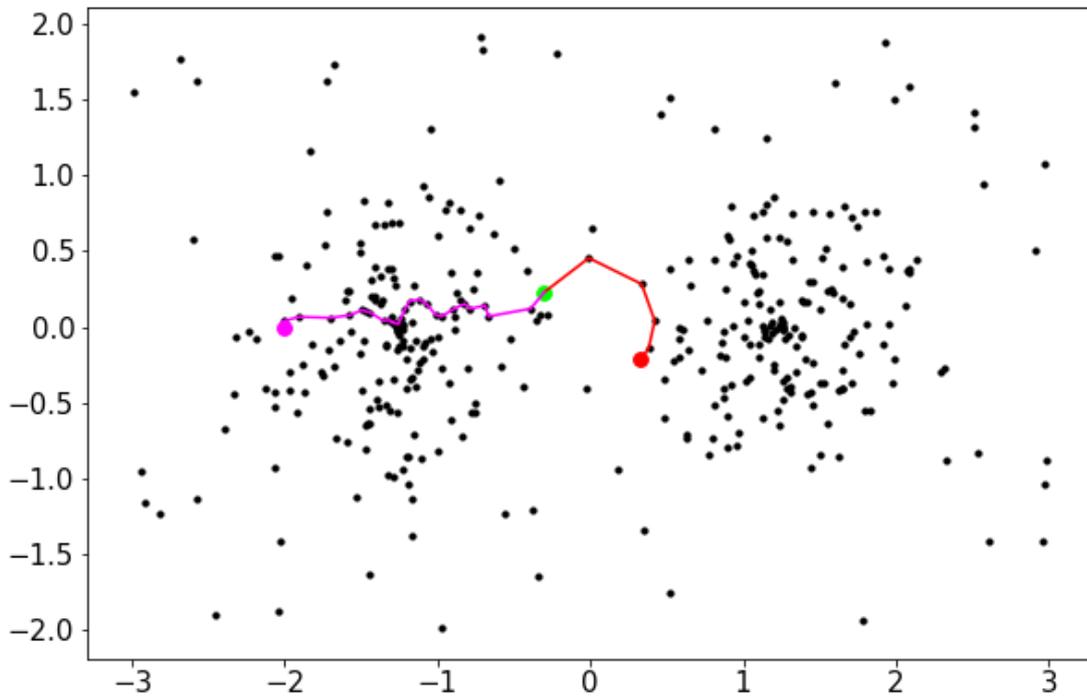


Figura 2.4: Ejemplo de la distancia de Fermat. Tomemos de referencia los puntos $p_1 = (-.3, .23)$ (verde), $p_2 = (-2, 0)$ (morado) y $p_3 = (.33, -.21)$ (rojo). Respecto a la distancia Euclideana, p_1 está más cerca de p_3 que de p_2 . Sin embargo, si consideramos la distancia de Fermat con $\alpha = 2$, tenemos que $\hat{d}_{\alpha}(p_1, p_3) = .39$ (trayectoria roja), mientras que $\hat{d}_{\alpha}(p_1, p_2) = .25$ (trayectoria morada). Esto se debe a que la región entre p_1 y p_3 es muy poco densa, mientras que para ir a p_2 se tiene lo contrario.

Teorema 2.6 Sea $\mathcal{M} \subset \mathbb{R}^d$ una variedad compacta. Sean $f : \mathcal{M} \rightarrow [m_f, M_f]$ una función continua de densidad y \mathcal{X}_n una muestra de f . Dados $x, y \in \mathcal{M}$ y $\varepsilon > 0$, existen constantes μ, c_1, c_2, n_0 tales que

$$\mathbb{P} \left(\left| \frac{\hat{d}'_\alpha(x, y; \mathcal{X}_n)}{n^{(1-\alpha)/d} D'_\alpha(x, y)} - \mu \right| > \varepsilon \right) \leq e^{-c_1 n^{c_2}}, \quad (2.89)$$

para toda $n \geq n_0$. Además,

$$\lim_{n \rightarrow \infty} \frac{\hat{d}'_\alpha(x, y; \mathcal{X}_n)}{n^{(1-\alpha)/d} D'_\alpha(x, y)} = \mu \quad \text{casi seguramente.} \quad (2.90)$$

2.6.1. Complejidad

Un aspecto importante a tomar en cuenta es la complejidad temporal del algoritmo la cual definimos como la cantidad de tiempo que le toma al algoritmo finalizar en función de la cantidad de datos. Es una de las posibilidades para definir la complejidad computacional, y será a la que nos referiremos en este trabajo.

En el caso discreto calcular la distancia muestral de Fermat es equivalente a encontrar el camino más corto entre dos nodos de una gráfica pesada donde el peso de la arista entre dos nodos es justo la distancia Euclideana elevada a la potencia α . Este problema ha sido extensamente estudiado y existen algoritmos como Floyd-Warshall o Dijkstra cuya complejidad es $\mathcal{O}(n^3)$ y $\mathcal{O}(n^2)$ [33], lo cual podría no ser deseable cuando se tiene una cantidad grande de datos. Una forma de simplificar este problema es restringir la búsqueda a trayectorias en donde cada punto q_{i+1} de la trayectoria satisface ser uno de los k -vecinos más cercanos de q_i , considerando $k \approx \log(n)$ [28]. Además, con alta probabilidad la distancia muestral de Fermat coincide con la trayectoria restringida. Sin embargo, es importante tener en cuenta que el k elegido asegure que la gráfica generada sea conexa. Ya que si la densidad es muy bajas en algunas zonas y elegimos un k pequeño, puede que se generen subgráficas disconexas, lo cual llevaría a resultados erróneos.

Definición 2.4 Sean $\alpha \geq 1$, \mathcal{X} un conjunto no vacío y $x, y \in \mathbb{R}^d$. Dado $1 \leq k \leq n$ con $k \in \mathbb{N}$, $q \in \mathcal{X}$, definamos al conjunto de los k -vecinos más cercanos de q por la

recursión:

$$q^{(1)} = \arg \min_{q' \in \mathcal{X} \setminus \{q\}} |q' - q| \quad q^{(k)} = \arg \min_{q' \in \mathcal{X} \setminus \{q, q^{(1)}, \dots, q^{(k-1)}\}} |q' - q|. \quad (2.91)$$

Entonces, $q^{(1)}$ es el vecino más cercano a q , $q^{(2)}$ es el segundo vecino más cercano a q y así sucesivamente. Notemos que debido a la recursión los empates también son considerados. En caso de que haya empate con el $q^{(k)}$ consideramos sólo uno. Denotamos al conjunto de los k -vecinos más cercanos de q como $\mathcal{N}_k(q) = \{q^{(1)}, \dots, q^{(k)}\}$.

Definición 2.5 (Distancia muestral de Fermat restringida) Para $x, y \in \mathcal{X}$, $\alpha \geq 1$ y $k \in \mathbb{N}$, definimos la distancia muestral de Fermat restringida, a los k vecinos más cercanos, por

$$\hat{d}_\alpha^{k'}(x, y; \mathcal{X}) = \min \left\{ \sum_{i=1}^{k-1} |q_{i+1} - q_i|^\alpha : q_1 = x, q_k = y, q_{i+1} \in \mathcal{N}_k(q_i) \right\}. \quad (2.92)$$

Proposición 2.2 Consideremos de nuevo las condiciones del enunciado del Teorema 2.6. Dada $\varepsilon > 0$, existen constantes positivas c_4, c_5 tales que si $k > c_4 \log(n/\varepsilon) + c_5$ ocurre que

$$\mathbb{P} \left(\hat{d}_\alpha^{k'}(x, y; \mathcal{X}) = \hat{d}_\alpha'(x, y; \mathcal{X}) \right) > 1 - \varepsilon. \quad (2.93)$$

La demostración de la proposición anterior se encuentra en [28]. De esta manera, la forma de encontrar la trayectoria restringida más corta entre cualquiera dos puntos tiene una complejidad de $\mathcal{O}(n^2 \log^2(n))$ y con probabilidad $1 - \varepsilon$ la trayectoria restringida coincide con la trayectoria sin restricción.

Aplicaciones

En este capítulo, exploraremos algunas de las aplicaciones de la distancia de Fermat, y su impacto en el campo de los algoritmos. Investigaremos cómo esta distancia influye en la creación de grupos coherentes y discutiremos posibles mejoras comparativas con métodos existentes. Además, consideraremos una innovadora propuesta: la *distancia de Hausdorff-Fermat*. Esta nueva distancia combina los principios de la distancia de Fermat con la robustez de la distancia de Hausdorff, y discutiremos algunas posibles aplicaciones.

3.1. Clusterización

Como hemos mencionado al principio del Capítulo 2, los enfoques clásicos como análisis de componentes principales (PCA) o escalamiento multidimensional (MDS) dan buenos resultados cuando los datos se encuentran en un subespacio lineal de un espacio de dimensión más alta [12]. La elección del algoritmo adecuado y la distancia correcta son esenciales para obtener resultados precisos. Un ejemplo de ello, es el *algoritmo K-means* [34]. Una componente fundamental en el funcionamiento de algoritmos de clusterización es la definición de la distancia utilizada para medir la similitud entre puntos. A continuación presentaremos un ejemplo ilustrativo en el que la distancia Euclideana falla mientras que la distancia de Fermat destaca en la eficacia para el algoritmo *K-means*.

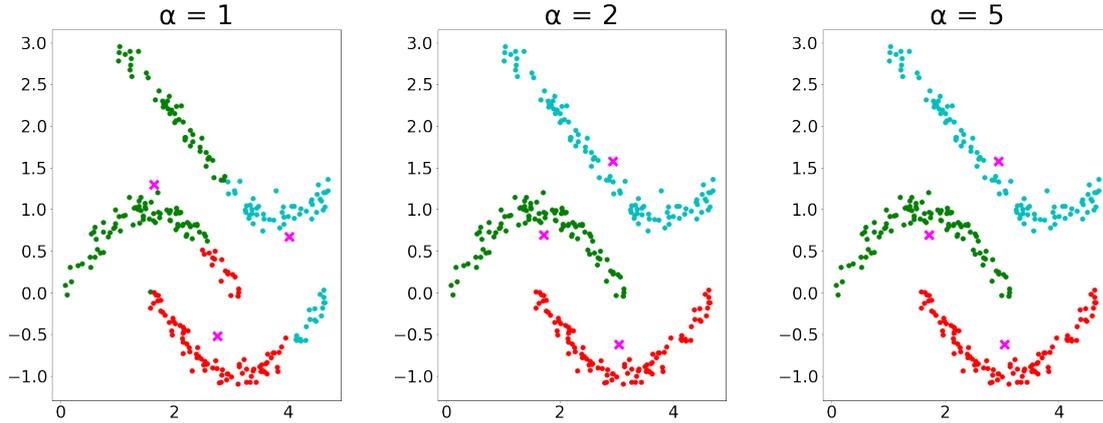


Figura 3.1: En la gráfica se muestra un ejemplo de clusterización con el algoritmo k-means. En el primer panel se ilustra los clusters obtenidos con la distancia Euclídeana. En el segundo y tercer panel se muestra los clusters identificados para $\alpha = 2, 5$, respectivamente. Debido a que las regiones de los clusters no son lineales, se obtienen resultados poco óptimos con la distancia Euclídeana. Este ejemplo destaca la ventaja de ocupar la distancia de Fermat en la clusterización de datos con estructuras geométricas complejas.

3.2. Distancia de Hausdorff-Fermat

Como vimos en el primer capítulo, la distancia de Hausdorff definida en 1.8 funciona con cualquier distancia inducida. La más común y como normalmente se le conoce a la distancia es la distancia Euclídeana. Entonces, podremos ocupar la distancia de Fermat para inducir la distancia de Hausdorff a lo cual llamaremos la *distancia de Hausdorff-Fermat (HF)* $\hat{d}_{\text{HF},\alpha}(\cdot)$.

Definimos la distancia de Fermat de un punto p a un conjunto A como

$$\hat{d}'_{\alpha}(p, A) = \inf_{y \in A} \hat{d}'_{\alpha}(x, y). \quad (3.1)$$

Definición 3.1 La distancia de Hausdorff-Fermat está definida como

$$\hat{d}_{\text{HF},\alpha}(A, B) = \max \left\{ \sup_{p \in A} \hat{d}'_{\alpha}(p, B), \sup_{q \in B} \hat{d}'_{\alpha}(q, A) \right\}. \quad (3.2)$$

Por la Proposición 1.2, podemos asegurar que la distancia de HF cumple con ser distancia.

3. APLICACIONES

A continuación presentamos un ejemplo de la distancia de Hausdorff-Fermat. Recordemos que para $\alpha = 1$ la distancia de Hausdorff-Fermat se reduce a la distancia de Hausdorff clásica con la distancia Euclídeana. Para mostrar la diferencia, presentamos tres imágenes con diferentes valores de parámetro $\alpha = 1, 2, 5$. Consideramos tres conjuntos distintos a los cuales denotamos por P_1 (puntos color verde), P_2 (puntos color rojo) y P_3 (puntos color cian). A continuación mostramos un resumen de las distancias obtenidas y gráfico de dichos conjuntos.

	$\alpha = 1$	$\alpha = 2$	$\alpha = 5$
$\hat{d}_{\text{HF},\alpha}(P_1, P_2)$	1.6511	0.576	0.0133
$\hat{d}_{\text{HF},\alpha}(P_2, P_3)$	3.1331	1.0764	0.0414
$\hat{d}_{\text{HF},\alpha}(P_3, P_1)$	2.5657	0.5598	0.0105

Tabla 3.1: Tabla de la distancia Hausdorff-Fermat entre los conjuntos P_1 , P_2 y P_3 .

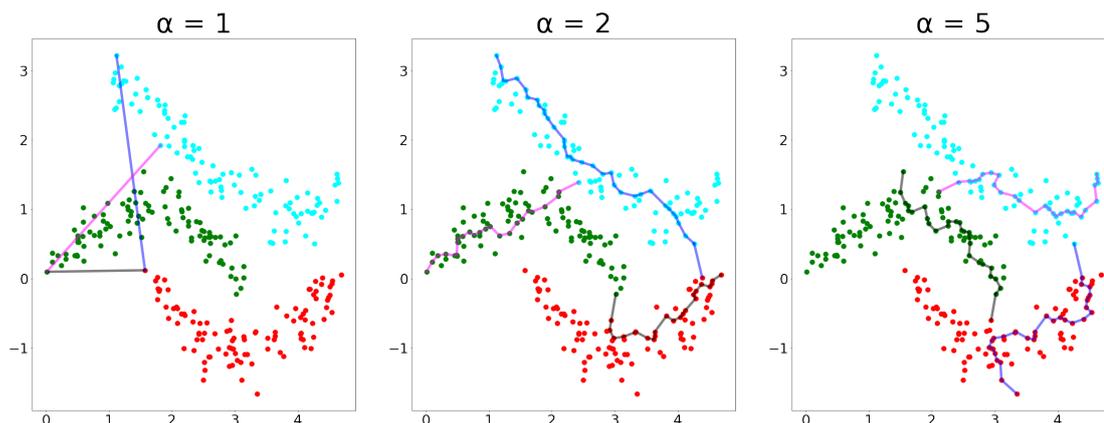


Figura 3.2: En la gráfica se muestran tres conjuntos distintos: P_1 (puntos color verde), P_2 (puntos color rojo) y P_3 (puntos color cian). También se muestran las trayectorias asociadas a las distancias de HF: $\hat{d}_{\text{HF},\alpha}(P_1, P_2)$ (color gris), $\hat{d}_{\text{HF},\alpha}(P_2, P_3)$ (color azul), y $\hat{d}_{\text{HF},\alpha}(P_3, P_1)$ (color magenta).

Recordemos que las variaciones en la densidad están implícitamente dentro de la distancia HF. Entonces, es de esperarse que para conjuntos provenientes de una densidad uniforme no se obtenga información relevante. A continuación mostramos la distancia HF para dos conjuntos, P_4 y P_5 , muestreados de una densidad uniforme con el mismo soporte y la gráfica de dichos datos. Como es de esperarse realmente no hay alguna diferencia significativa ni en la distancia ni en la trayectoria que corresponde a dicha distancia.

	$\alpha = 1$	$\alpha = 2$	$\alpha = 5$
$\hat{d}_{\text{HF},\alpha}(P_4, P_5)$	0.43	0.16	0.01

Tabla 3.2: Tabla de la distancia Hausdorff-Fermat entre los conjuntos P_4 y P_5 .

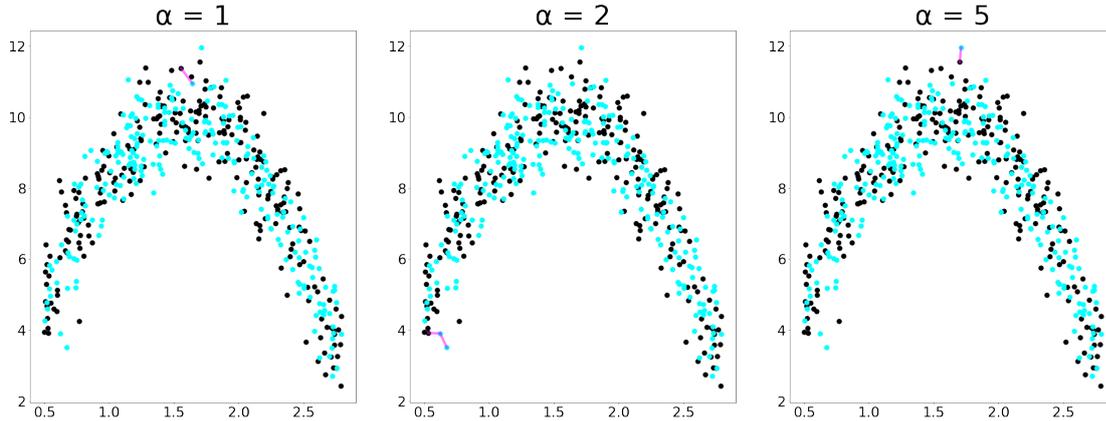


Figura 3.3: En la gráfica se muestran dos conjuntos distintos: P_4 (puntos color negro) y P_5 (puntos color cian). También se muestran las trayectorias asociadas a las distancias de HF (color magenta).

Donde esperamos que tenga una aplicación tangible es en la segmentación de imágenes. Una tarea crucial en el diagnóstico y tratamiento de enfermedades es la segmentación de tumores en imágenes médicas como resonancias magnéticas (RM) o tomografías computarizadas (TC). La distancia de Hausdorff es una herramienta fundamental utilizada en varios algoritmos para lograr esta segmentación precisa. Uno de los enfoques comunes es el uso de la distancia de Hausdorff como medida de similitud entre las regiones tumorales predichas por el algoritmo y las regiones tumorales reales en la imagen original [9, 35]. A continuación, se presenta el procedimiento de manera general de cómo se usa la distancia de Hausdorff en la segmentación de tumores:

1. Preprocesamiento de imágenes: se comienza con la imagen médica que contiene el tumor a segmentar, como una RM o una TC. Se aplican técnicas de preprocesamiento para mejorar el contraste, reducir el ruido y realzar las características relevantes.
2. Umbralización inicial: se aplica un umbral inicial para separar la región de interés, que en este caso es el tumor, del fondo y otros tejidos circundantes.

3. Algoritmo de segmentación: se aplica un algoritmo de segmentación para identificar la región aproximada del tumor en la imagen. Algunos algoritmos comunes utilizados en este paso son la segmentación basada en crecimiento de regiones [36] y métodos de contorno activo [37], entre otros.
4. Generación de máscara: a partir de la región estimada del tumor, se crea una máscara binaria que representa la ubicación aproximada del tumor en la imagen.
5. Evaluación con distancia de Hausdorff: se compara la máscara binaria generada con una máscara de referencia que contiene la ubicación real del tumor en la imagen original. La distancia de Hausdorff se calcula entre los bordes de las dos máscaras para medir la discrepancia entre las regiones segmentadas y las reales.
6. Ajuste y refinamiento: basándose en la distancia de Hausdorff calculada, el algoritmo ajusta y refina la segmentación. Puede expandir o contraer la región estimada del tumor para que se ajuste mejor a la ubicación real del tumor.
7. Iteraciones y convergencia: los pasos 3 a 6 se pueden repetir iterativamente para lograr una segmentación cada vez más precisa. El algoritmo puede converger cuando la distancia de Hausdorff entre las regiones segmentadas y reales es suficientemente baja.

En resumen, la distancia de Hausdorff se utiliza como una medida de similitud para guiar y ajustar la segmentación de tumores en imágenes médicas. Esto ayuda a mejorar la precisión de los algoritmos de aprendizaje. Sin embargo, en el caso clásico de la distancia de Hausdorff no se toma en cuenta la densidad del conjunto.

En la Fig. 3.4 se muestran dos casos distintos de conjuntos. En las gráficas del primer panel mostramos los conjuntos P_6 (puntos negros) y P_7 (puntos azules). En las gráficas del segundo panel mostramos los conjuntos P_8 (puntos negros) y P_9 (puntos azules). En ambos casos, el par de conjuntos están distribuidos en el mismo soporte, pero con densidades diferentes. En el primer caso, los puntos P_6 tienden a estar mayormente distribuidos en el extremo derecho, mientras que hay más puntos del conjunto P_7 en el extremo izquierdo. En el segundo caso, si consideramos la parametrización clásica del círculo con θ refiriéndose al ángulo, notamos que el conjunto P_8 tiene mayor densidad para valores de $\theta \in [0, \pi)$ mientras que el P_9 para los valores $\theta \in [\pi, 2\pi)$. Sin embargo, en ambos casos notamos que tenemos puntos distribuidos sobre toda la “dona”.

En los dos pares de conjuntos mencionados anteriormente calculamos la distancia HF para los valores $\alpha = 1, 2$ y 5 presentados en la Tab. 3.3. Recordemos

que $\alpha = 1$ corresponde al caso clásico de Hausdorff. En ambos casos obtenemos una distancia alta cuando se considera $\alpha = 1$ y se va reduciendo para $\alpha = 2, 5$.

Tomando como ejemplo $\hat{d}_{\text{HF},\alpha}(P_6, P_7)$ (primer panel de la Fig. 3.4). Para todos los casos, los puntos asociados a la distancia HF son: el punto azul más extremo a la derecha p con el punto negro q más extremo a la derecha inferior. La trayectoria se muestra en color magenta. En el caso $\alpha = 1$ vemos que se obtiene un valor de 2.1, mientras que para $\alpha = 2, 5$ se obtiene 1.2 y 1.0, respectivamente. Esto se debe que al considerar la densidad, es decir, considerar que hay una cantidad significativa de puntos entre p y q lo cual reduce la distancia HF. Con este en mente, creemos que la distancia HF puede ser ocupada en segmentación de imágenes cuando la densidad no es tan relevante y cuando se tiene cierta tolerancia a la densidad que en el caso de imágenes se traduciría a intensidad de los píxeles. Por ejemplo, si quisiéramos segmentar el área de un tumor cancerígeno, quisiéramos detectar toda el área posible. En ese caso no importaría tanto la densidad, solo nos interesa si hemos pasado cierto umbral para considerar esa área como afectada.

	$\alpha = 1$	$\alpha = 2$	$\alpha = 5$
$\hat{d}_{\text{HF},\alpha}(P_6, P_7)$	2.1571	1.2277	1.0281
$\hat{d}_{\text{HF},\alpha}(P_8, P_9)$	0.7417	0.2314	0.0097

Tabla 3.3: Valores de la distancia HF entre los conjuntos P_6 y P_7 , P_8 y P_9 .

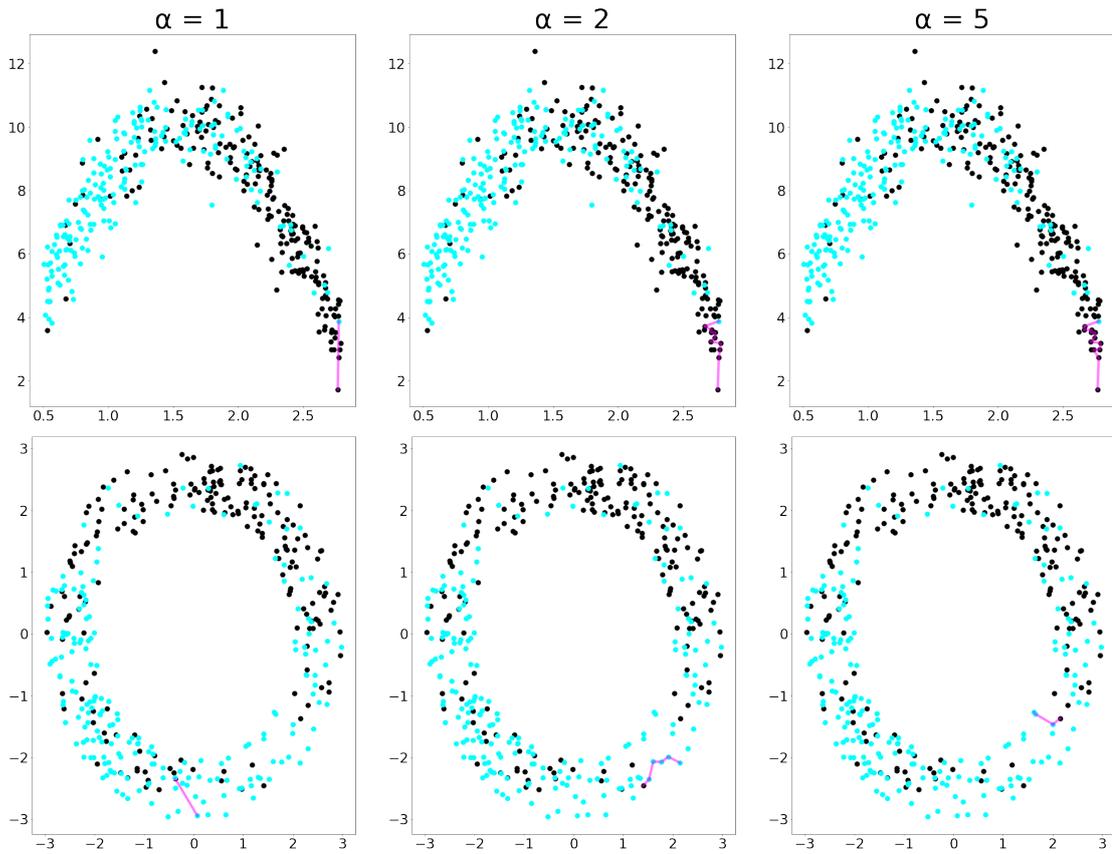


Figura 3.4: En los paneles del primer renglón se muestra P_6 (puntos color negro), P_7 (puntos color cian) y sus trayectorias asociadas a la distancias HF (color magenta). En los paneles del segundo renglón tenemos lo mismo pero para los conjuntos P_8 y P_9 .

Otra forma en la cual creemos que pueda servir la distancia HF es en dar un valor cualitativo de si un conjunto ha sido “desprendido” de otro. De igual manera, lo hacemos pensando en un caso médico. En caso de tener una región con múltiples áreas cancerígenas, un aspecto importante es dar una medida que cuantifique si un área ha sido “desprendida” de otra. En este caso, se tendría un indicador de posible metástasis. En este caso, a menor distancia le asignamos más posibilidad de desprendimiento.

En la Fig. 3.5 consideramos los conjuntos P_{10} (puntos negros) y P_{11} (puntos cian) y en la Tab 3.4 se muestran los valores obtenidos para $\alpha = 1, 2, 5$. Debido a la forma del soporte de P_{10} hace que la distancia de Hausdorff considere la distancia entre los extremos superiores de ambos conjuntos, generando un valor alto, lo cual indicaría un sentido de lejanía. Sin embargo, cuando se considera $\alpha \geq 2$, notamos que entra en juego el funcionamiento de la densidad. Esto se puede notar en el

segundo panel y tercer panel, donde lo trayectoria pasa primero por la parte más densa de P_{10} hasta que encuentra una zona la longitud del “salto” de un conjunto a otro no sea tan grande.

	$\alpha = 1$	$\alpha = 2$	$\alpha = 5$
$\hat{d}_{\text{HF},\alpha}(P_{10}, P_{11})$	2.9363	0.8387	0.07501

Tabla 3.4: Valores de la distancia HF entre los conjuntos P_{10} y P_{11} .

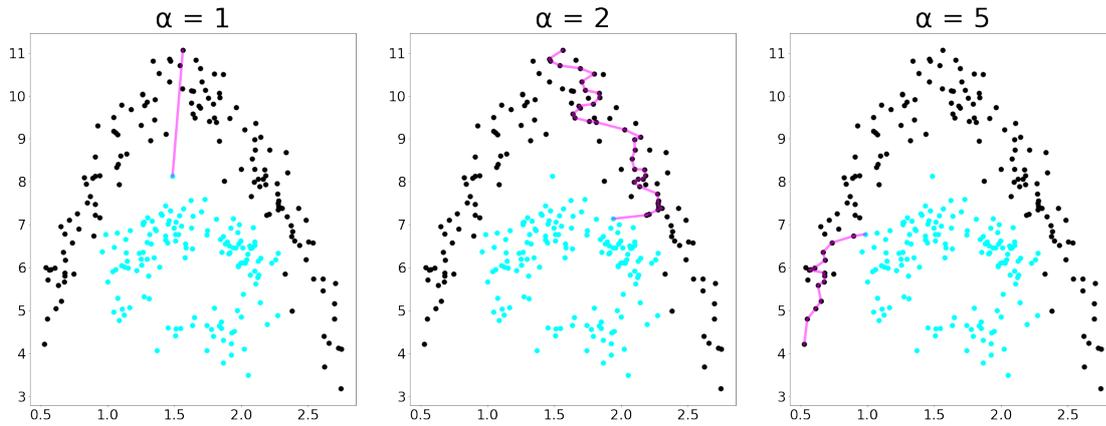


Figura 3.5: En la gráfica se muestran dos conjuntos distintos: P_{10} (puntos color negro) y P_{11} (puntos color cian). También se muestran las trayectorias asociadas a las distancias de HF (color magenta).

Por último, el hecho de que la distancia HF tome en cuenta la densidad de puntos, hace que sirva de mejor manera para medir la similitud entre muestras aleatorias de funciones de densidad, particularmente en densidades de colas pesadas. En la Fig. 3.6 se muestran dos conjuntos de puntos que marginalmente en el eje horizontal se distribuyen como una t-Student. Recordemos que la función de densidad está dada por

$$f(x; \nu, \mu, \sigma) = \frac{\Gamma(\frac{\nu+1}{2})}{\sigma\sqrt{\pi\nu}\Gamma(\nu/2)} \left(1 + \frac{(x - \mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}},$$

donde ν son los grados de libertad, σ el parámetro de escalamiento, μ el parámetro de locación. El conjunto P_{12} (puntos negros) ha sido con $\mu = 2.8$ y el conjunto P_{13} (puntos cian) con $\mu = 4.2$. En ambos casos se considero $\sigma = 0.5$ y $\nu = 15$.

3. APLICACIONES

El problema de ocupar el distancia de Hausdorff como medida de similitud para variables aleatorias es que tiende a ocupar los puntos más extremos. En el caso de densidades de colas pesadas, da grandes problemas ya que tiende dar valores muy grandes. Sin embargo, con la distancia HF, no necesariamente se mide la distancia entre los extremos ya que es ahí donde el papel de la densidad entra en juego. A mayor cantidad de puntos “en medio” de los extremos, menor el valor de la distancia HF, lo cual indicaría un nivel de similitud mayor.

	$\alpha = 1$	$\alpha = 2$	$\alpha = 5$
$\hat{d}_{\text{HF},\alpha}(P_{12}, P_{13})$	2.1837	1.1531	0.7168

Tabla 3.5: Valores de la distancia HF entre los conjuntos P_{12} y P_{13} .

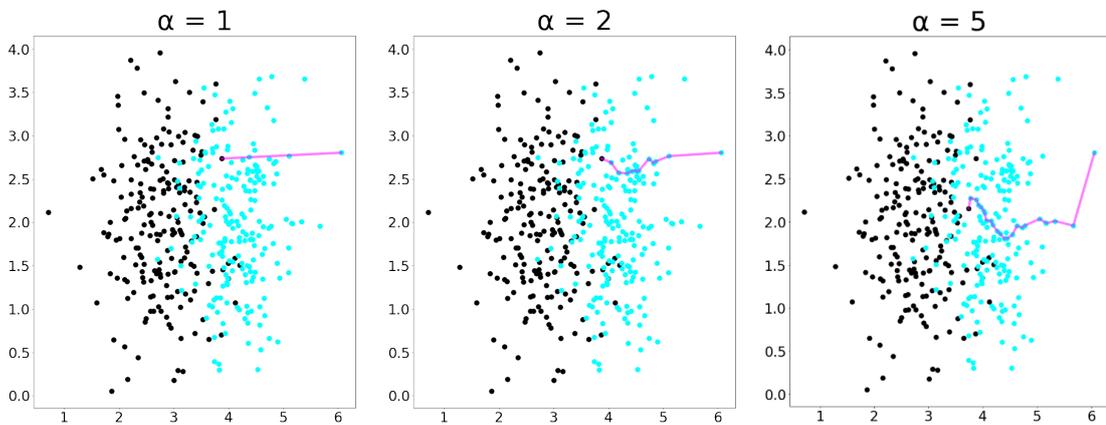


Figura 3.6: En la gráfica se muestran dos conjuntos distintos: P_{12} (puntos color negro) y P_{13} (puntos color cian). También se muestran las trayectorias asociadas a las distancias HF (color magenta). Ambos conjuntos se distribuyen marginalmente en el eje horizontal con una distribución t-Student.

Conclusiones

En este trabajo hemos explorado y analizado los conceptos fundamentales en torno a la distancia de Hausdorff, distancia PWSP y la distancia de Fermat, así como su aplicación en el ámbito computacional. Los resultados obtenidos por distintos autores y presentados en este trabajo han enriquecido la comprensión de estas distancias y han planteado posibles nuevas perspectivas en el campo de la análisis de datos.

En el primer capítulo, se ha desarrollado una introducción a la distancia de Hausdorff y se han discutido las aplicaciones computacionales que esta distancia ofrece. Al comprender su significado y utilidad en la comparación de conjuntos y la evaluación de similitudes, se ha sentado una base sólida para explorar su potencial en problemas del mundo real.

En el segundo capítulo, nos enfocamos en demostrar resultados de convergencia que dan una amplia perspectiva sobre el comportamiento de la distancia PWSP y la distancia de Fermat. Estos resultados revelan que a medida que aumentamos la cantidad de muestras o datos, la distancia muestral PWSP converge a la distancia PWSP salvo una constante, donde dicha constante es la constante del tiempo de percolación. Este resultado es de gran interés para la aplicación de análisis de datos, principalmente en la medida de similitud entre objetos.

En el tercer capítulo introdujimos un nuevo concepto al proponer la distancia de Hausdorff-Fermat. Aunque se presentaron ejemplos simplificados para ilustrar su potencial de aplicación, estos casos sencillos sirven como cimiento para futuras exploraciones más complejas. La distancia de Hausdorff-Fermat, al considerar la densidad de conjuntos, resulta ser más robusta en términos de las densidades de los datos, es decir, menos dependiente de datos atípicos. Por ello, tiene el potencial de aportar nueva perspectiva en la segmentación de imágenes y otros campos.

En resumen, el presente trabajo pretende mejorar la comprensión de la distancia de Hausdorff y Fermat, y hacer una propuesta nueva mediante la distancia de Hausdorff-Fermat, inspirada dentro del contexto de análisis de datos.

Bibliografía

- [1] W. Rudin. *Principles of mathematical analysis*. McGraw-Hill New York, 1964. [1](#)
- [2] E. F. Krause. *Taxicab Geometry: An adventure in Non-Euclidean Geometry*. Dover Publications, 1986. [2](#)
- [3] J. Banás y A. Martín. Some properties of the hausdorff distance in metric spaces. *Bull. Austral. Math. Soc.*, **42**(3):511–516, 1990. [3](#)
- [4] G. A. Klanderman y W. J. Rucklidge D. P. Huttenlocher. Comparing images using the hausdorff distance. *IEEE*, 15(9):850–863, 1993. [6](#)
- [5] D. Santa-cruz y T. Ebrahimi N. Aspert. Mesh: Measuring errors between surfaces using the hausdorff distance. *Proceedings of the IEEE International Conference in Multimedia and Expo (ICME)*, 1:705–708, 2002. [6](#)
- [6] B. Takács. Comparing face images using the modified hausdorff distance. *Pattern Recognition*, 31(12):1873–1881, 1998. [6](#)
- [7] Y. Gao. Efficiently comparing face images using a modified hausdorff ditance. *IEE Pro. Vis. Image Signal Process*, 150(6):346–350, 2003.
- [8] M. Dubuisson y A. Jain. A modified hausdorff distance for object matching. *Proceedings of 12th International Conference on Pattern Recognition*, 2:566–568, 1994. [6](#)
- [9] D. Karimi y S. E. Salcudean. Reducing the hausdorff ditance in medical image segmentation with convolutional neural networks. *IEEE Transactions on Medical Imaging*, 39(2):499–513, 2020. [6](#), [41](#)
- [10] W. J. Rucklidge. Efficiently locating objects using the hausdorff ditance. *International Journal of Computer Vision*, 24(3):251–270, 1997. [7](#)

-
- [11] A. Hinneburg y D. A. Keim C. C. Aggarwal. On the surprising behavior of distance metrics in high dimensional space. *Database Theory, ICDT 2001*, pages 420–434, 2001. [8](#)
- [12] V. De Silva y J. C. Langford J. B. Tenenbaum. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2008. [9](#), [38](#)
- [13] L. van der Maaten y G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(11):2579–2605, 2008.
- [14] M. Belkin y P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.
- [15] M. R. Coifman y S. Lafon. Diffusion maps. *Applies and Computational Harmonic Analysis*, 21(1):5–30, 2006. [9](#)
- [16] A. Little y Y. Xie A. Manousidaki. Clustering and visualization of single-cell rna-seq data using path metrics. 2021. [10](#)
- [17] L. L. Saul y M. I. Jordan. A variational principle for model-based interpolation. *NIPS*, pages 267–273, 1997. [10](#)
- [18] A. Jaffe y B. Nadler A. Moscovich. Minimax-optimal semi-supervised regression on unknown manifolds. *AISTATS*, pages 849–856, 2017.
- [19] D. McKenzie y S. Damelin. Power weighted shortest paths for clustering euclidean data. 2019.
- [20] S. T. Roweis y L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2008.
- [21] S. B. Damelin y A. O. Hero S. J. Hwang. Shortest path through random points. *Ann. Appl. Probab.*, **26**(5):2791–2823, 2016. [16](#)
- [22] D. McKenzie y J. M. Murphy A. Little. Balancing geometry and density: path distances on high-dimensional data. 2021. [10](#)
- [23] M. Alamgir y U. V. Luxburg. Shortest path distance in random k-nearest neighbor graphs. *ICML*, pages 1251–1258, 2012. [11](#)
- [24] M. Maggioni y J. M. Murphy A. Little. Path-based spectral clustering: Guarantees, robustness to outliers, and fast algorithms. *Journal of the Machine Learning Research*, **21**(6):1–66, 2020. [11](#)
- [25] C. D. Howard y C. M. Newman. Euclidean models of first-passage percolation. *Probability Theory and Related Fields*, **108**(2):153–170, 1997. [13](#), [16](#)
-

- [26] C. D. Howard y C. M. Newman. Geodesics and spanning trees for euclidean first-passage percolation. *Ann. Appl. Probab.*, **29**(2):577–623, 2001. [13](#)
- [27] J. F. C. Kingman. *Poisson Processes*. Oxford University Press, 1993. [13](#), [16](#)
- [28] M. Jonckheere y F. Sapienza P. Groisman. Nonhomogeneous euclidean first-passage percolation and distance learning. *Bernoulli*, **28**(1):255–276, 2022. [14](#), [34](#), [36](#), [37](#)
- [29] H. Kesten. On the time constant and path length of first-passage percolation. *Advances in Applied Probability*, 12(4):848–863, 1980. [14](#)
- [30] J. T. Cox. The time constant of first-passage percolation on the square lattice. *Advances in Applied Probability*, 12(4):864–879, 1980. [14](#)
- [31] B. O’Neill. *Semi-Riemannian geometry*. Academic Press, 1983. [28](#)
- [32] P. Groisman y M. Jonckheere F. Sapienza. Weighted geodesic distance following fermat’s principle. *International Conference on Learning Representation*, 2018. [34](#)
- [33] D. B. Johnson. Efficient algorithms for shortest paths in sparse networks. *Journal of the ACM*, **21**(1):1–13, 1977. [36](#)
- [34] S. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982. [38](#)
- [35] N. B. Puhan y G. Panda R. Panda. New binary hausdorff symmetry measure based seeded region growing for retinal vessel segmentation. *Biocybern Biomed Eng*, 36(1):119–129, 2016. [41](#)
- [36] L. P. Suresh y S. H. K. Veni S. Angelina. Image segmentation based on genetic algorithm for region growth and region merging. *ICCEET*, pages 970–974, 2012. [42](#)
- [37] D. J. Williams y M. Shah. A fast algorithm for active contours and curvature estimation. *CVGIP: Image understanding*, 55(1):14–26, 1992. [42](#)