UNIVERSIDAD NACIONAL DE LA PLATA

FACULTAD DE CIENCIAS EXACTAS

DEPARTAMENTO DE MATEMÁTICA

**Trabajo de tesis doctoral: Fórmulas ergódicas para
medias geométricas y curvas minimales**

Tesista: Eduardo Mario Ghiglioni

Director: Jorge Abel Antezana

Codirector: Demetrio Stojanoff

Año: 2018

A Mechi

Agradecimientos

Muchas personas forman parte de este logro y por eso quiero agradecerles:

A mi sostén, mi compañera de vida, la que estuvo presente minuto a minuto en cada pequeño avance “Mechi Mechi”. Sobre todo por llevarme adelante de la mano cuando la meta aún estaba lejos.

A mis directores Jorge y Demetrio. Ellos me motivaron a estudiar los problemas de mi tesis y me aconsejaron en cada paso. Pero además de guiarme en mi camino como investigador, me motivaron a crecer como docente en la facultad, dándome su apoyo para transmitir los conocimientos que aprendí durante la licenciatura de la mejor manera. Y sin olvidar el viaje a Corea que fue una oportunidad única para mi y no hubiera logrado sin sus incentivos.

Quiero agradecer a mis padres porque siempre apoyaron cada decisión que tomé. Desde chico entendieron que la matemática era mi pasión y siempre me acompañaron.

A mi hermano y mi cuñada que con quienes intercambiamos experiencias facultativas y me aconsejaron a lo largo de estos años.

También a mis amigos de toda la vida Juan, Dani y Pinto quienes me acompañan en todo momento. Como también a mi ahijadito Chesco y a mi comadre Flor.

A todos mis amigos de la facultad con quienes compartí momentos únicos. En especial a Lu “La capitana” que día a día me demuestra su gran amistad. A Nico que fue mi compañero de estudio de toda la carrera y siempre está a mi lado, brindo con él con sus cervezas artesanales. A Blas por su predisposición e hizo a mi pedido unos carteles de primera. A Juli Lavié que me acompañó también durante la carrera y es una gran amiga. A Gisel “Souto” con quién formamos un lindo lazo. A Juli Ferrario, hemos compartido oficina y ahora cuida de Mechi en la ofi. Además a Mimi, Noe y Mati, Ceci, Adri y Pancho, Elona, Ale, Belén y Gisela, que son un sostén en todo momento.

A todos mis familiares de sangre y de afecto. En especial a mi suegra Elsa. Ellos siempre están presentes y de alguna manera me ayudaron a que esto sea posible.

A los chicos de bibliotekita (Jorge, Demetrio, Francisco, Pedro, Edu y Mariano) que en muchas charlas me compartieron sus experiencias de doctorado y me sumaron al grupo de investigación.

A los bibliotekids (Noe, Lupe, Pablin, Majo, Seba y Clau), cuyo nombre nos pusimos por ser dirigidos por alguien del grupo de bibliotekita. Ellos son mis compañeros incondicionales de oficina tanto en la UNLP como en el IAM.

A todos los que forman parte del IAM (investigadores, becarios y personal de apoyo) quienes durante estos años me han demostrado mucho afecto.

A todos los guitarristas del IFA quienes comparten mi alegría y mi pasión por la música. Con ellos tenemos una amistad única hace muchos años.

Al jurado (Gabriel, Alejandro y “El Solo”) por el tiempo dedicado en este rol. Por las

sugerencias y comentarios que ayudaron a mejorar notablemente este trabajo.

A todos los profesores, JTPs, ayudantes, alumnos que tuve a lo largo de este período. Por enseñarme e instruirme a lo largo de la carrera y también por formar un lazo de amistad.

Como instituciones agradezco al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) por el financiamiento otorgado para la realización de esta tesis doctoral, al Instituto Argentino de Matemática, a la facultad de Ciencias Exactas y al Departamento de Matemática de la Universidad Nacional de La Plata.

Por último quiero agradecer a quienes nombramos Imanuel y Agustina, a pesar que no pudimos conocer sus rostros los llevamos en nuestro corazón y van a estar a nuestro lado por siempre. Con Mechi los amamos.

A todos ellos estoy eternamente agradecido.

¡¡¡Muchas gracias!!!

Índice general

1. Introducción	1
1.1. Estructura de la tesis	13
2. Preliminares del Capítulo 3	15
2.1. Espacios de Hadamard	15
2.1.1. Ejemplos de espacios de Hadamard	17
2.2. Baricentros	18
2.3. Teoría ergódica	19
2.4. Teorema ergódico de Birkhoff y la ley de los grandes números	23
2.5. Sistemas de Kronecker	25
2.5.1. Sistemas de Kronecker y medidas de Haar.	28
3. Teorema Ergódico en espacios de Hadamard en términos de la media inductiva	30
3.1. Motivación	30
3.2. Una variante del Teorema ergódico	38
3.2.1. La media inductiva	39
3.3. Fórmula ergódica asociada a la media inductiva en el caso continuo	41
3.3.1. El marco de trabajo y notación básica	41
3.3.2. Demostración del caso continuo	42
3.4. Fórmula ergódica asociada a la media inductiva en el caso L^1	47
3.4.1. Dos lemas de aproximación	47
3.4.2. Buena aproximación por funciones continuas (Mollifiers)	48
3.4.3. Demostración del caso L^1	55
3.5. Independencia del punto de comienzo	56
3.6. Otra forma de convergencia	56
4. Estructura de las curvas minimales en $\mathcal{H}(n)$, $\mathcal{P}(n)$ y $\mathcal{U}(n)$ con respecto a la métrica de Finsler bi-invariante inducida por la norma traza	59
4.1. Motivación	59
4.2. Estructura de las curvas minimales para la norma traza	60
4.2.1. Estudio de curvas minimales en el espacio de matrices Hermitianas dotada de la norma traza	61
4.2.2. El espacio de matrices positivas	69
4.2.3. El espacio de matrices unitarias	73
4.3. Geometría de los puntos medios en $\mathcal{P}(n)$ y $\mathcal{H}(n)$	75

5. Estructura de las curvas minimales en $\mathcal{U}(n)$ con respecto a la métrica de Finsler bi-invariante inducida por la norma espectral	78
5.1. El grupo de un parámetro	78
5.2. El grupo de matrices unitarias	80
5.3. Estructura de las curvas minimales en $\mathcal{U}(n)$	81
5.3.1. Demostración del Teorema 5.3.1	82
5.4. Geometría de los puntos medios	85
5.5. La Grassmaniana	88
6. Apéndice	92
6.1. Análisis matricial	92
6.2. Geometría de las matrices positivas	96
6.3. Teorema de Lusin	99
6.4. Centro de masa	100

1 | Introducción

El espacio de matrices $n \times n$ complejas, $\mathcal{M}(n)$, es un espacio de Hilbert con el producto interno $\langle A, B \rangle = \text{tr}(A^* B)$ y la norma asociada $\|A\|_2 = (\text{tr}(A^* A))^{1/2}$. Esta norma es conocida como norma Frobenius o Hilbert-Schmidt. El conjunto de matrices Hermitianas $\mathcal{H}(n)$ es un subespacio vectorial real de $\mathcal{M}(n)$. Como el subconjunto $\mathcal{P}(n)$ de matrices (estrictamente) positivas es un subconjunto abierto en $\mathcal{H}(n)$, resulta ser una variedad diferenciable. En particular, el espacio tangente a $\mathcal{P}(n)$ en un punto A , es el espacio $T_A \mathcal{P}(n) = \mathcal{H}(n)$. El producto interno en $\mathcal{H}(n)$ nos permite definir una métrica Riemanniana en la variedad $\mathcal{P}(n)$, la cual en un punto A , está dada por

$$ds = \|A^{-1/2} dA A^{-1/2}\|_2 = [\text{tr}(A^{-1} dA)^2]^{1/2}.$$

En otras palabras, dada una curva suave a trozos $\alpha : [a, b] \rightarrow \mathcal{P}(n)$, su longitud se define como

$$L(\alpha) = \int_a^b \|\alpha^{-1/2}(t) \dot{\alpha}(t) \alpha^{-1/2}(t)\|_2 dt.$$

A partir de esto, como es usual, la distancia rectificable entre $A, B \in \mathcal{P}(n)$ se define por medio de

$$\delta(A, B) = \inf \{L(\alpha) \text{ donde } \alpha : [a, b] \rightarrow \mathcal{P}(n) \text{ une } A \text{ con } B \text{ en } \mathcal{P}(n)\}.$$

Más aún, la distancia Riemanniana δ puede ser calculada por

$$\delta(A, B) = \|\log(A^{-1/2} B A^{-1/2})\|_2.$$

El ínfimo es en realidad un mínimo, y la geodésica que conecta dos matrices positivas A y B tiene la siguiente expresión simple:

$$\gamma_{AB}(t) = A^{1/2} (A^{-1/2} B A^{-1/2})^t A^{1/2}. \quad (1.0.1)$$

Es usual en análisis matricial usar la notación $A \#_t B$ en vez de $\gamma_{AB}(t)$. El punto medio $A \#_{1/2} B$ se conoce como **media geométrica** o **baricentro** entre A y B , y también admite la siguiente caracterización variacional

$$A \#_{\frac{1}{2}} B = \operatorname{argmin}_{C \in \mathcal{P}(n)} \left(\delta^2(A, C) + \delta^2(B, C) \right).$$

No hay razón para restringirnos a sólo dos matrices. La noción de **media geométrica** se puede generalizar para más de dos matrices de manera natural

$$\Lambda(A_1, \dots, A_n) := \operatorname{argmin}_{C \in \mathcal{P}(n)} \left(\sum_{j=1}^n \delta^2(A_j, C) \right).$$

La solución del problema de minimización de cuadrados mínimos, existe y es única, por las propiedades de convexidad de la distancia $\delta(\cdot, \cdot)$.

La media geométrica aparece de manera natural en varios problemas aplicados. Por ejemplo, en el estudio de señales de radar. En este problema, cada señal es detectada por más de un sensor. La información de cada sensor se codifica en un núcleo de covarianza, el cual es semidefinido positivo. Estos núcleos tienen que ser promediados para obtener el resultado final. Resulta que la mejor manera de promediar estos núcleos no es la media aritmética estándar, si no, la media geométrica que introducimos previamente (ver [19] y las referencias allí para más detalles). Otra aplicación típica de la media geométrica es con problemas relacionados con métodos de optimización del tipo Newton (ver [15],[46]).

El inconveniente usual al lidiar con la media geométrica es que, cuando trabajamos con tres o más matrices, no existe en general una fórmula cerrada para calcularla. Por esta razón, hubo una extensa investigación con el objetivo de encontrar buenas maneras de aproximar la media geométrica de tres o más matrices ([14], [35], [42]). Un avance importante se logró al intentar demostrar la monotonía de la media geométrica. En otras palabras, se desconocía si $\Lambda(A_1, \dots, A_n) \leq \Lambda(B_1, \dots, B_n)$, para $A_i \leq B_i$, donde $1 \leq i \leq n$. La resolución de este problema condujo a la introducción de las llamadas **medias inductivas**.

Para motivar la definición de las medias inductivas, notar que dada una sucesión $\{a_n\}_{n \in \mathbb{N}}$ de números complejos tenemos la siguiente relación

$$\begin{aligned} \frac{a_1 + a_2 + a_3}{3} &= \frac{2}{3} \left(\frac{a_1 + a_2}{2} \right) + \frac{1}{3} a_3 \\ &\vdots \\ \frac{a_1 + \dots + a_n}{n} &= \frac{n-1}{n} \left(\frac{a_1 + \dots + a_{n-1}}{n-1} \right) + \frac{1}{n} a_n. \end{aligned}$$

Sea $\gamma_{a,b}(t) = (1-t)a + tb$, y por un momento usemos la notación $a \#_t b = \gamma_{a,b}(t)$. Entonces

$$\begin{aligned} \frac{a_1 + a_2 + a_3}{3} &= (a_1 \#_{\frac{1}{2}} a_2) \#_{\frac{1}{3}} a_3 \\ \frac{a_1 + a_2 + a_3 + a_4}{4} &= ((a_1 \#_{\frac{1}{2}} a_2) \#_{\frac{1}{3}} a_3) \#_{\frac{1}{4}} a_4 \end{aligned}$$

y así sucesivamente. En el espacio Euclídeo los segmentos son las geodésicas. En nuestro contexto, podemos reemplazar los segmentos por las geodésicas asociadas a la estructura Riemanniana. Esta es la idea que nos lleva a la definición de la media inductiva. Dada una sucesión de matrices estrictamente positivas $\{A_n\}_{n \in \mathbb{N}}$, las **medias inductivas** se definen de la siguiente manera:

$$\begin{aligned} S_1(A) &= A_1 \\ S_n(A) &= S_{n-1}(A) \#_{\frac{1}{n}} A_n \quad (n \geq 2). \end{aligned}$$

La **propiedad de monotonía** primero fue probada por Lawson y Lim (ver [35], Corolario 5.1 y Proposición 5.2), basándose en los resultados de Sturm sobre estadísticas en espacio de curvatura no positiva (ver [56], Proposición 6.6). Ellos probaron que si la sucesión está

formada por la elección aleatoria e independiente de elementos del conjunto $\{A_1, \dots, A_n\}$, entonces la media inductiva converge casi seguro a $\Lambda(A_1, \dots, A_n)$. Es decir,

$$\lim_{k \rightarrow \infty} S_k(A_{\omega(1)}, \dots, A_{\omega(k)}) = \Lambda(A_1, \dots, A_n) \quad \text{c.t.p. } \omega \in \{1, \dots, n\}^{\mathbb{N}}.$$

La relación con la monotonía de $\Lambda(\cdot)$ yace en que la media geométrica se obtiene de una sucesión binaria de pasos, cada uno de los cuales es monótono. Es decir,

$$X_1 \leq X_2, Y_1 \leq Y_2 \quad \text{implica} \quad X_1 \#_t X_2 \leq Y_1 \#_t Y_2 \quad \text{para todo } t \in [0, 1]. \quad (1.0.2)$$

De esta manera, la media inductiva resulta monótona, y esta propiedad se preserva en el límite. Observar que el resultado de Lawson y Lim puede ser visto como una “ley fuerte de los grandes números”, pues como mencionamos en la motivación de la media inductiva, la traducción de este resultado al caso Euclídeo es la convergencia en casi todo punto de los promedios a su esperanza. Bhatia y Karandikar [14] simplificaron la prueba de la monotonía de la media geométrica notando que una “ley débil de los grandes números” es suficiente.

A partir de estos resultados el problema de la monotonía estaba resuelto. Notar que además la media inductiva es computable utilizando la conocida expresión de la geodésica (1.0.1). Sin embargo, mientras que el enfoque probabilístico es muy interesante, también es verdad que la sucesión es desconocida. De esta manera, para encontrar una mejor aproximación, todo decanta en encontrar una sucesión concreta que converja a la media geométrica. El candidato para este propósito es la sucesión que simplemente se repite de manera cíclica en $\{A_0, \dots, A_{m-1}\}$. Consideremos m matrices positivas $A_0, \dots, A_{m-1}$, y definamos la función $F : \mathbb{Z}_m \rightarrow \mathcal{P}(n)$ por $F(\bar{k}) = A_k$, donde $\mathbb{Z}_m$ denota el grupo abeliano de enteros módulo m . Si definimos la sucesión periódica $A = \{F(\bar{n})\}_{n \in \mathbb{N}}$, entonces Holbrook [29] probó

$$\lim_{k \rightarrow \infty} S_k(A) = \Lambda(A_0, \dots, A_{m-1}).$$

Ejemplo. Supongamos que tenemos 3 matrices estrictamente positivas A_1, A_2 y A_3 . Recordamos que en este caso particular:

$$\Lambda(A_1, A_2, A_3) = \operatorname{argmin}_{C \in \mathcal{P}(n)} \left(\delta^2(A_1, C) + \delta^2(A_2, C) + \delta^2(A_3, C) \right).$$

Luego Holbrook probó que

$$A_1 \#_{\frac{1}{2}} A_2 \#_{\frac{1}{3}} A_3 \#_{\frac{1}{4}} A_1 \#_{\frac{1}{5}} A_2 \#_{\frac{1}{6}} A_3 \#_{\frac{1}{7}} A_1 \#_{\frac{1}{8}} A_2 \dots \rightarrow \Lambda(A_1, A_2, A_3)$$

Figura 1.1: Ejemplo de la convergencia de la media inductiva en el caso Euclídeo.

Fórmulas ergódicas para medias geométricas

Sea $(G, +)$ un grupo topológico compacto abeliano, dotado con una medida de Haar m , y sea $\tau : G \rightarrow G$ una transformación ergódica. El Teorema ergódico clásico de Birkhoff nos dice que, dada $f \in L^1(G)$, entonces para m -casi todo $\omega \in G$

$$\frac{1}{n} \sum_{k=0}^{n-1} f(\tau^k(\omega)) \xrightarrow{n \rightarrow \infty} \int_G f(\omega) dm(\omega). \quad (1.0.3)$$

Luego, el Teorema de Holbrook puede ser visto como un Teorema ergódico para funciones definidas en $\mathbb{Z}_p$ que toman valores en $\mathcal{P}(n)$. En efecto, podemos pensar que la media inductiva juega el papel de los promedios del lado izquierdo y la media geométrica $\Lambda(A_0, \dots, A_{d-1})$ juega el papel de la integral de f con respecto a la medida de Haar en $\mathbb{Z}_d$.

Esta interpretación del resultado de Holbrook nos sugiere que podemos obtener una aproximación computable de la media geométrica de varias matrices en término de promedios ergódicos. Uno de los objetivos de esta tesis es probar tal Teorema ergódico en término de la media inductiva en un contexto más general. Más precisamente, consideremos un sistema dinámico (G, τ) , donde G es un grupo compacto y τ es ergódico con respecto a la medida de Haar m . Observar que el grupo tiene que ser abeliano porque las órbitas $\{\tau^n g\}_{n \in \mathbb{N}}$ son densas en G .

En este contexto, dada una función $F : G \rightarrow \mathcal{P}(n)$, decimos que $F \in L^p(G, \mathcal{P}(n))$ si

$$\int_G \delta^p(F(g), B) dm(g) < \infty,$$

donde B es cualquier matriz positiva. Por la desigualdad triangular, la definición no depende de la elección de B . Siguiendo el trabajo de Sturm [56], el **baricentro** de $F \in L^1(G, \mathcal{P}(n))$ se define como

$$\beta_F := \operatorname{argmin}_{C \in \mathcal{P}(n)} \int_G \delta^2(F(g), C) - \delta^2(F(g), B) dm(g).$$

En otras palabras, β_F denota al baricentro del *pushforward measure* F_*m . En este caso es posible probar la existencia y unicidad de dicho elemento por las propiedades de continuidad

y convexidad uniforme de la función,

$$C \mapsto \int_G \delta^2(F(g), C) - \delta^2(F(g), B) dm(g).$$

Como antes, esta definición no depende de B . Notar que si $F \in L^2(G, \mathcal{P}(n))$ entonces

$$\beta_F := \operatorname{argmin}_{C \in \mathcal{P}(n)} \int_G \delta^2(F(g), C) dm(g).$$

En esta tesis probamos el siguiente resultado:

Teorema. Dada $A \in L^1(G, \mathcal{P}(n))$, para casi todo $g \in G$

$$\lim_{n \rightarrow \infty} S_n(A(g), A(\tau(g)), \dots, A(\tau^{n-1}(g))) = \beta_A. \quad (1.0.4)$$

Observación. Consideremos como grupo topológico compacto al toro $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$, con la estructura de grupo usual. A continuación vamos a hacer una analogía entre el resultado de Holbrook y el teorema antes mencionado para este grupo particular. Supongamos que tomamos $A \in L^1(\mathbb{T}, \mathcal{P}(n))$ fija:

1. Si tomamos una rotación racional $\theta = 2\pi/3$ en el toro entonces

$$\begin{aligned} A(\theta(g)) &= A_1; \\ A(\theta^2(g)) &= A_2; \\ A(\theta^3(g)) &= A_3; \\ A(\theta^4(g)) &= A_1; \\ A(\theta^5(g)) &= A_2; \\ &\vdots \end{aligned}$$

para ciertas matrices positivas A_1, A_2, A_3 . Por el Teorema de Holbrook

$$\lim_{n \rightarrow \infty} S_n(A(g), A(\theta(g)), \dots, A(\theta^{n-1}(g))) = \Lambda(A_1, A_2, A_3) = \operatorname{argmin}_{C \in \mathcal{P}(n)} \sum_{j=1}^3 \delta^2(A_j, C).$$

2. El caso complementario es tomar una rotación irracional τ . Las rotaciones irracionales son las únicas rotaciones ergódicas en el toro. Por lo cual, el teorema anterior prueba

$$\lim_{n \rightarrow \infty} S_n(A(g), A(\tau(g)), \dots, A(\tau^{n-1}(g))) = \beta_A = \operatorname{argmin}_{C \in \mathcal{P}(n)} \int_{\mathbb{T}} \delta^2(A(g), C) dm(g).$$

Con esta observación podríamos decir de manera figurativa que, en el toro, *el Teorema de Holbrook es el caso racional mientras que (1.0.4) corresponde al caso irracional*.

Figura 1.2: Rotación racional.
Ejemplo de una rotación racional en el toro
con un ángulo igual a $\frac{2\pi}{3}$.

Figura 1.3: Rotación irracional.
Ejemplo de una rotación con un ángulo
irracional en el toro.

Más aún, probamos este resultado no sólo para funciones que toman valores en $\mathcal{P}(n)$, si no también, para los espacios de Hadamard (M, δ) . Recordamos que un espacio métrico completo (M, δ) se llama **espacio de Hadamard** si la distancia satisface la ley del semiparalelogramo. Es decir, para todo $x, y \in M$ existe $m \in M$ tal que

$$\delta^2(m, z) \leq \frac{1}{2}\delta^2(x, z) + \frac{1}{2}\delta^2(y, z) - \frac{1}{4}\delta^2(x, y).$$

La desigualdad del semiparalelogramo implica que el punto medio es único. La existencia y unicidad de puntos medios dan lugar a una **única geodésica** $\gamma_{a,b} : [0, 1] \rightarrow M$ que conecta dos puntos cualesquiera a y b . En efecto, primero se define $\gamma_{a,b}(1/2)$ como el punto medio entre a y b . Luego, usando un argumento inductivo, se define la geodésica para todos los números diádicos racionales en $[0, 1]$. Finalmente por completitud, se extiende para todo $t \in [0, 1]$. Continuaremos usando la notación $a \#_t b$ en vez de $\gamma_{a,b}(t)$.

Como en el caso de matrices estrictamente positivas, definimos la media inductiva en espacios de Hadamard de la siguiente manera:

Definición. (Media inductiva). Sea (M, δ) un espacio de Hadamard. Dado $a \in M^{\mathbb{N}}$ se define

$$\begin{aligned} S_1(a) &= a_1 \\ S_n(a) &= S_{n-1}(a) \#_{\frac{1}{n}} a_n \quad (n \geq 2). \end{aligned}$$

Tanto $L^1(G, M)$ como β_A , al igual que las medias inductivas, se definen mutatis mutandis. En este contexto el enunciado del teorema resulta ser:

Teorema. Dada $A \in L^1(G, M)$, para casi todo $g \in G$,

$$\lim_{n \rightarrow \infty} S_n(A(g), A(\tau(g)), \dots, A(\tau^{n-1}(g))) = \beta_A. \quad (1.0.5)$$

Con respecto a la media inductiva: la diferencia principal entre (1.0.4) y (1.0.5) es que para los espacios de Hadamard, en general, no existe una fórmula para la geodésica que une

dos puntos. En otras palabras, al tratar el problema con más generalidad se puede perder la computabilidad de la media inductiva. Sin embargo, en muchos casos particulares de gran interés como el caso de las matrices positivas, operadores positivos en espacios de Hilbert, entre otros, existe una fórmula cerrada para las geodésicas que unen dos puntos.

Cabe mencionar que en [56], Sturm desarrolló una teoría de baricentros de medidas de probabilidad para espacios de Hadamard (ver Subsección 2.2 para algunas definiciones y resultados básicos). Dotado con este baricentro, los espacios de Hadamard juegan un papel importante en la teoría de integración (variables aleatorias, esperanza y varianza), ley de los grandes números, teoría ergódica, desigualdad de Jensen (ver [17], [27], [38], [48], y [56]), generalizaciones estocásticas de retracciones Lipschitz y problemas de extensión de funciones Lipschitz o Hölder (ver [39], [45], y [50]) y teoría de transporte óptimo en variedades Riemannianas (ver [52], y [53]), etc.

Mollifiers en espacios de Hadamard

La generalización de $\mathcal{P}(n)$, o incluso de variedades Riemannianas con curvatura no positiva, a espacios de Hadamard no es directa. Uno de los motivos es que los espacios de Hadamard, en general, no tienen un espacio vectorial finito dimensional subyacente, como en el caso de variedades. Esto hace que algunos pasos de las demostraciones sean mucho más complicados, y nos llevan a una nueva definición de mollifiers que sólo usa la estructura métrica (ver subsección 3.4.2). Estos resultados originales son interesantes por sí mismos y generalizan unos resultados probados por Karcher en [31].

Mejoras que obtuvimos con respecto a resultados previos

En [6] Austin probó un Teorema ergódico general en espacios de Hadamard, el cual en nuestro contexto nos dice que, dada $A \in L^2(G, M)$, para casi todo $g \in G$ tenemos

$$\lim_{n \rightarrow \infty} \Lambda(A(g), A(\tau(g)), \dots, A(\tau^{n-1}(g))) = \beta_A.$$

Luego, en [48] Navas extendió el resultado de Austin a funciones en $A \in L^1(G, M)$, y que toma valores en espacios métricos más generales (espacios de Busemann). En este resultado, las medias aritméticas discretas del Teorema de Birkhoff son reemplazadas por la media geométricas (conjunta) de la n -upla

$$(A(g), A(\tau(g)), \dots, A(\tau^{n-1}(g))).$$

Por otro lado, la integral del Teorema de Birkhoff se reemplaza por el baricentro β_A . Observar que en la sucesión

$$A(g), \Lambda(A(g), A(\tau(g))), \Lambda(A(g), A(\tau(g), A(\tau^2(g)))), \dots$$

del segundo término en adelante, no hay computabilidad en general, incluso en el caso de matrices positivas. En esta dirección, las medias inductivas son más simples y proveen en muchos ejemplos importantes una aproximación computable al baricentro. Esta es la ventaja de nuestro resultado, pero hay un precio que pagar. Por un lado, necesitamos buen control de la convexidad de la métrica. Por esta razón trabajamos en espacios de Hadamard, como en el caso del resultado de Austin [6]. Por otro lado, las medias inductivas son simplemente promedios de puntos. Esta simplicidad es buena desde el punto de vista computacional, pero confinan nuestro resultado a $\mathbb{Z}$ -acciones. Los resultados mencionados de Austin y Navas también son válidos para acciones más generales.

Curvas minimales

El Teorema ergódico que obtuvimos está basado, en parte, en que existe una única geodésica $\gamma_{a,b}$ que une dos puntos, lo cual nos permite dar una buena definición de la media inductiva. Es interesante entonces preguntarse si al cambiar la métrica seguimos teniendo esta propiedad y podemos generalizar aún más el Teorema ergódico.

Si consideramos como en un comienzo el espacio de matrices (estrictamente) positivas $\mathcal{P}(n)$ entonces: Dada una curva suave a trozos $\alpha : [a, b] \rightarrow \mathcal{P}(n)$, podemos definir su longitud de curva por

$$L_{\|\cdot\|_p}(\alpha) = \int_a^b \left\| \alpha^{-1/2}(t) \dot{\alpha}(t) \alpha^{-1/2}(t) \right\|_p dt,$$

donde $\|\cdot\|_p$ es la norma Schatten p definida como

$$\begin{aligned} \|A\|_p &= \text{tr}(|A|^p)^{1/p} \quad \text{si } 1 \leq p < \infty \\ \|A\|_{sp} &= \sup_{x \in \mathbb{C}^n \setminus \{0\}} \frac{\|Ax\|}{\|x\|}. \end{aligned}$$

La distancia rectificable entre $A, B \in \mathcal{P}(n)$ se define como antes

$$\delta_{\|\cdot\|_p}(A, B) = \inf \left\{ L_{\|\cdot\|_p}(\alpha) \text{ donde } \alpha : [a, b] \rightarrow \mathcal{P}(n) \text{ une } A \text{ con } B \text{ en } \mathcal{P}(n) \right\}.$$

Es conocido que para las normas Schatten p , donde $1 < p < \infty$, el ínfimo se alcanza en una única curva que une A con B y tiene la expresión antes mencionada:

$$A \#_t B = A^{1/2} (A^{-1/2} B A^{-1/2})^t A^{1/2}.$$

Existen infinitos puntos medios (y geodésicas minimales) entre un cierto par de matrices definidas positivas con respecto a la métrica de Finsler asociada a la norma espectral $\|\cdot\|_\infty$, que coincide con la métrica de Thompson,

$$d_\infty(A, B) = \left\| \log(A^{-1/2} B A^{-1/2}) \right\|_\infty = \max \left\{ \log \lambda_1(A^{-1/2} B A^{-1/2}), \log \lambda_1(A^{-1/2} B^{-1} A^{-1/2}) \right\}.$$

Más aún, están caracterizadas las **curvas minimales** con respecto a la **métrica de Thompson**. Recordamos la construcción de Nussbaum para curvas minimales con respecto a la métrica de Thompson.

Teorema ([49]). Sean $A, B \in \mathcal{P}(n)$ y sea $\lambda_k = \lambda_k(A^{-1}B)$, $k = 1, \dots, n$.

i) Si $\lambda_n^{-1} \leq \lambda_1$, entonces para todo $\alpha \in [\lambda_1^{-1}, \lambda_n]$,

$$f_\alpha(t) = \begin{cases} \frac{\lambda_1^t - \alpha^t}{\lambda_1 - \alpha} B + \frac{\lambda_1 \alpha^t - \alpha \lambda_1^t}{\lambda_1 - \alpha} A & \lambda_1 \neq \lambda_n \\ \lambda_1^t A & \lambda_1 = \lambda_n \end{cases}$$

es una curva minimal de A a B con respecto a la métrica de Thompson.

ii) Si $\lambda_1 \leq \lambda_n^{-1}$, entonces para todo $\alpha \in [\lambda_1, \lambda_n^{-1}]$,

$$f_\alpha(t) = \begin{cases} \frac{\alpha^t - \lambda_n^t}{\alpha - \lambda_n} B + \frac{\alpha \lambda_n^t - \lambda_n \alpha^t}{\alpha - \lambda_n} A & \lambda_1 \neq \lambda_n \\ \lambda_1^t A & \lambda_1 = \lambda_n \end{cases}$$

es una curva minimal de A a B con respecto a la métrica de Thompson.

Un subconjunto $\Omega \subset \mathcal{P}(n)$ es **geodésicamente convexo** si $A \#_t B \in \Omega$ para todo $t \in [0, 1]$, siempre que $A, B \in \Omega$. Se puede definir el conjunto de los d_Φ -puntos medios t -pesados entre A y B de la siguiente manera

$$\mathfrak{M}(t; A, B) := \{X \in \mathcal{P}(n) : d_\Phi(X, A) = td_\Phi(A, B), d_\Phi(B, X) = (1 - t)d_\Phi(A, B)\}.$$

En [40] Lim demostró que este conjunto es compacto y geodésicamente convexo para la métrica de Thompson. A partir del estudio de este conjunto también obtuvo el siguiente criterio para la unicidad de curvas minimales con respecto a la métrica de Thompson.

Teorema ([40]). *Sean $A, B \in \mathcal{P}(n)$. Entonces existe una única d_∞ -geodésica minimal si, y sólo si, existe un único punto medio entre A y B si, y sólo si, $\sigma(A^{-1}B) \subset \{a, a^{-1}\}$ para algún $a > 0$.*

Luego, la construcción de Nussbaum nos llevó a pensar en una caracterización para las curvas minimales en el espacio de matrices (estrictamente) positivas en el caso $p = 1$. Para estudiar propiedades de las curvas minimales en el espacio de matrices positivas, nos enfocamos en el estudio del espacio de matrices Hermitianas $\mathcal{H}(n)$, porque toda matriz positiva $P \in \mathcal{P}(n)$ se puede escribir de la forma $P = e^A$, donde $A \in \mathcal{H}(n)$.

A continuación desarrollamos este nuevo problema: Dadas dos matrices $A, B \in \mathcal{H}(n)$ y una curva $\alpha : [0, 1] \rightarrow \mathcal{H}(n)$ tal que $\alpha(0) = A, \alpha(1) = B$, podemos medir la longitud de esta curva de la siguiente manera:

$$L(\alpha) = \int_0^1 \|\dot{\alpha}(t)\|_1 dt, \quad (1.0.6)$$

donde $\|\cdot\|_1$ es la norma traza. La distancia rectificable inducida, denotada $d_1(\cdot, \cdot)$, se calcula como antes, como el ínfimo de las longitudes de todas las curvas suaves a trozos que unen dos puntos dados:

$$d_1(A, B) = \inf \{L(\alpha) \text{ donde } \alpha : [a, b] \rightarrow \mathcal{H}(n) \text{ une } A \text{ con } B \text{ en } \mathcal{H}(n)\}.$$

En este contexto es posible probar que el segmento

$$\alpha(t) = (1 - t)A + tB,$$

es una **curva minimal** que une la matriz A con la matriz B . Es decir,

$$L(\alpha) = d_1(A, B).$$

Sin embargo es fácil construir ejemplos para probar que no es la única.

Ejemplo. Sean

$$Id = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} \frac{\pi}{2} & 0 \\ 0 & \frac{\pi}{4} \end{pmatrix}.$$

Entonces la curva

$$\alpha(t) = \begin{pmatrix} \frac{\pi}{2}t & 0 \\ 0 & \frac{\pi}{2}t^2 \end{pmatrix},$$

une Id con X y es minimal.

Teorema. Sea $X : [0, 1] \rightarrow \mathcal{H}(n)$ una curva minimal. Supongamos también que $X(0) = 0$, y que $D = X(1) = \text{diag}(d_1, \dots, d_n)$, es una matriz diagonal tal que las entradas d_j están ordenadas de manera decreciente. Entonces la curva minimal que une 0 con D tiene la siguiente estructura:

$$X(t) = \begin{pmatrix} P(t) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & N(t) \end{pmatrix} \begin{matrix} \mathcal{S}_1 \\ \mathcal{S}_2 \\ \mathcal{S}_3 \end{matrix},$$

$$\mathcal{S}_1 = \text{span}\{e_j : d_j > 0\}, \mathcal{S}_2 = \text{span}\{e_j : d_j = 0\} \quad \text{y} \quad \mathcal{S}_3 = \text{span}\{e_j : d_j < 0\}.$$

La implicación que es más fácil de demostrar es que las curvas que tienen esa estructura son curvas minimales. La otra implicación es posible lograrla utilizando el operador pinching y reduciendo, de alguna manera, el problema a caracterizar las curvas minimales en matrices 2×2 . Además es interesante que las condiciones que se requieren, son las extensiones naturales de las condiciones que se necesitan para caracterizar las curvas minimales en $\mathbb{R}^n$. Recordamos que si $\alpha : [0, 1] \rightarrow \mathbb{R}^n$ una curva minimal que une el vector $O = (0, \dots, 0)$ con $V = (v_1, \dots, v_n)$ en $\mathbb{R}^n$. Son equivalentes:

- La curva $\alpha : [0, 1] \rightarrow \mathbb{R}^n$ es minimal.
- Las curvas $\alpha_j : [0, 1] \rightarrow \mathbb{R}$ que unen 0 con v_j son minimales.
- Las curvas $\alpha_j : [0, 1] \rightarrow \mathbb{R}$ que unen 0 con v_j son tales que $\dot{\alpha}_j(t) > 0$, si v_j es estrictamente positivo, o $\dot{\alpha}_j(t) < 0$, si v_j es estrictamente negativo, para todo $t \in [0, 1]$. Si α_j une 0 con 0 entonces es la curva idénticamente nula.

Este resultado se puede probar pensando que la distancia entre dos puntos de $\mathbb{R}^n$ es la suma de las diferencias (absolutas) de sus coordenadas. Esta geometría es conocida como *la métrica del taxista*. Podemos pensar que para llegar de un punto a otro lo importante es, por ejemplo, cuantas cuadras nos movemos hacia la derecha y cuantas hacia arriba.

Por otro lado, las matrices inversibles actúan en $\mathcal{P}(n)$ de manera natural por conjugación. Dicha acción es transitiva e induce una estructura métrica dada por

$$ds = \|A^{-1/2}dA A^{-1/2}\|_1.$$

Es decir, la longitud de una curva $\gamma : [0, 1] \rightarrow \mathcal{P}(n)$ se define como

$$L(\gamma) = \int_0^1 \|\gamma^{-1/2}(t) \dot{\gamma}(t) \gamma^{-1/2}(t)\|_1 dt.$$

Con esta estructura métrica, la acción de los inversibles resulta isométrica. Como consecuencia de este cambio, el espacio $\mathcal{P}(n)$ resulta un espacio de curvatura no positiva (NPC). Las curvas minimales también cambian, al igual que en el caso de $\mathcal{H}(n)$, se ve que en general puede existir más de una. En esta tesis daremos una caracterización de dichas curvas minimales, la cual pasa por relacionar las **curvas minimales** en el espacio de **matrices positivas**, con las de $\mathcal{H}(n)$, utilizando la función exponencial.

Teorema. Sea $\alpha : [0; 1] \rightarrow \mathcal{P}(n)$ una curva que une I con e^X , donde X es una matriz Hermitiana diagonal. En otras palabras, $\alpha(t) = e^{X(t)}$ donde la curva $X : [0, 1] \rightarrow \mathcal{H}(n)$ es tal que: $X(0) = 0$, $X(1) = X$ y X es una matriz Hermitiana diagonal. La curva $X(t)$ es minimal si, y sólo si, $\alpha(t)$ es una curva minimal.

Asimismo, mostraremos como la misma idea se puede utilizar en el grupo de Lie de las matrices unitarias $U \in \mathcal{U}(n)$, donde la longitud de una curva en $\mathcal{U}$ se mide como en (1.0.6). No obstante, debido a que la curvatura de este espacio es no negativa, si bien el método permite construir familias de curvas minimales, no permite probar que dichas familias son todas. En ambos casos, las técnicas utilizadas provienen de la teoría espectral de matrices normales, y en particular, de ciertas desigualdades de matrices.

Para completar el estudio que realizamos para $p = 1$ y los trabajos existentes para matrices positivas donde $p = \infty$, surge el siguiente problema: Si consideramos $\mathcal{U}(n) \subset L(\mathcal{H})$ como subvariedad, resulta natural, considerar la estructura asociada a la norma de operadores. La métrica de Finsler bi-invariante dada por la norma espectral es la siguiente,

$$\|X\|_U = \|U^* X\|_{sp} = \|X\|_{sp},$$

para todo X tangente a un operador unitario U . Con respecto a esta métrica de Finsler, la longitud de la curva $\alpha : [0, 1] \rightarrow \mathcal{U}(n)$, se puede calcular de la siguiente manera

$$L(\alpha) = \int_0^1 \|\dot{\alpha}(t)\|_{sp} dt.$$

La distancia rectificable inducida, denotada por $d_\infty(\cdot, \cdot)$, se calcula como en el caso Riemanniano, como el ínfimo de las longitudes de todas las curvas suaves a trozos que unen dos puntos dados:

$$d_\infty(U, V) = \inf \{L(\alpha) \text{ donde } \alpha : [a, b] \rightarrow \mathcal{U}(n) \text{ une } U \text{ con } V \text{ en } \mathcal{U}(n)\}.$$

Sean $U, V \in \mathcal{U}(n)$ donde $V = Ue^{iX}$ con $X \in \mathcal{H}(n)$ y $\|X\|_{sp} \leq \pi$. Entonces la curva $\delta(t) = Ue^{itX}$ es más corta que cualquier otra curva suave a trozos γ en $\mathcal{U}(n)$ que una U con V . Es decir,

$$L(\delta) = d_\infty(U, V).$$

Sin embargo es fácil construir ejemplos para probar que no es la única.

Ejemplo. Sean

$$U = e^{iX} = \begin{pmatrix} e^{i\frac{\pi}{2}} & 0 \\ 0 & e^{i\frac{\pi}{4}} \end{pmatrix}, \quad X = \begin{pmatrix} \frac{\pi}{2} & 0 \\ 0 & \frac{\pi}{4} \end{pmatrix}.$$

Entonces la curva

$$\alpha(t) = \begin{pmatrix} e^{it\frac{\pi}{2}} & 0 \\ 0 & \tilde{\alpha}(t) \end{pmatrix}, \quad \tilde{\alpha}(t) = \begin{cases} e^{i\frac{\pi}{2}t} & t \in [0, 1/4] \\ e^{i\frac{\pi}{2}(\frac{1}{2}-t)} & t \in [1/4, 1/2] \\ e^{i\frac{\pi}{2}(t-\frac{1}{2})} & t \in [1/2, 1] \end{cases}$$

une Id con U y es minimal.

Con este problema en vista, obtuvimos la siguiente estructura para las **curvas minimales** que unen dos **matrices unitarias con respecto a la norma espectral**.

Teorema. Dado $U \in \mathcal{U}(n)$, sea $X \in \mathcal{H}(n)$ tal que $U = e^{iX}$ y $\|X\|_{sp} \leq \pi$. Entonces:

- a) Si $\sigma(|X|) \subseteq [0, \pi)$ y tiene más de un elemento, entonces una curva minimal que une I con U tiene la siguiente estructura:

$$\alpha(t) = \begin{pmatrix} e^{itX_S} & 0 \\ 0 & \alpha_1(t) \end{pmatrix} \begin{matrix} \mathcal{S} \\ \mathcal{S}^\perp \end{matrix},$$

donde $\mathcal{S} = \ker(\|X\|_{sp}I - |X|)$, $X_S = X|_{\mathcal{S}} \in \mathcal{H}(\mathcal{S})$, y $\alpha_1 : [0, 1] \rightarrow \mathcal{U}(\mathcal{S}^\perp)$ es cualquier curva que une $I_{\mathcal{S}^\perp}$ con $U_{\mathcal{S}^\perp} = U|_{\mathcal{S}^\perp}$ tal que $\|\dot{\alpha}_1\|_{sp} \leq \|X\|_{sp}$.

- b) Si $\pi \in \sigma(|X|)$, entonces una curva minimal que une I con U tiene la siguiente estructura:

$$\alpha(t) = \begin{pmatrix} e^{itY_S} & 0 \\ 0 & \alpha_1(t) \end{pmatrix} \begin{matrix} \mathcal{S} \\ \mathcal{S}^\perp \end{matrix},$$

donde $\mathcal{S} = \ker(\pi I - |X|)$, $Y_S \in \mathcal{H}(\mathcal{S})$ y satisface que $\sigma(Y_S) \subseteq \{\pi, -\pi\}$, y $\alpha_1 : [0, 1] \rightarrow \mathcal{U}(\mathcal{S}^\perp)$ es cualquier curva que une $I_{\mathcal{S}^\perp}$ con $U_{\mathcal{S}^\perp} = U|_{\mathcal{S}^\perp}$ tal que $\|\dot{\alpha}_1\|_{sp} \leq \pi$.

Por otro lado, esto lleva a una caracterización de aquellos pares de matrices unitarias conectadas por una única curva minimal.

Corolario. La geodésica es la única curva minimal que une $U, V \in \mathcal{U}(n)$ si, y sólo si, el espectro de U^*V está contenido en $\{e^{i\theta}, e^{-i\theta}\}$.

Este resultado es análogo al que obtuvo Y. Lim en [40] para matrices positivas con la métrica de Thompson. A su vez estudiamos la geometría de los puntos medios obteniendo el siguiente resultado.

Proposición. Dados $U, V \in \mathcal{U}(n)$ tal que $d_\infty(U, V) < \pi/2$, y $t \in (0, 1)$, el conjunto

$$\mathcal{M}_t(U, V) = \{W \in \mathcal{U}(n) : d_\infty(U, W) = t d_\infty(U, V) \text{ y } d_\infty(W, V) = (1-t) d_\infty(U, V)\}$$

es geodésicamente convexo.

Una vez obtenidos estos resultados pasamos al estudio de la Grassmanniana. La **Grassmanniana**, que denotaremos por $\mathcal{G}_n$, es el conjunto de subespacios de $\mathbb{C}^n$, que puede ser identificado con el conjunto de proyecciones ortogonales en $M_n(\mathbb{C})$. Si consideramos en $M_n(\mathbb{C})$ la topología definida por cualquiera de sus normas equivalentes, el espacio Grassmanniano dotado con esta topología, es un conjunto compacto. Sin embargo, no es conexo. En efecto, es suficiente considerar la traza, la cual es una función continua definida en todo el espacio $M_n(\mathbb{C})$, y restringida a $\mathcal{G}_n$ toma sólo valores naturales. En particular, esto prueba que las componentes conexas de $\mathcal{G}_n$ son subconjuntos $\mathcal{G}_{m,n}$, definidos como:

$$\mathcal{G}_{m,n} := \{P \in \mathcal{G}_n : \text{tr}(P) = m\}.$$

Cada una de estas componentes es una subvariedad de $M_n(\mathbb{C})$. Para una norma simétrica dada $\|\cdot\|$ en $\mathcal{M}_n(\mathbb{C})$, el espacio Grassmanniano lleva la estructura de Finsler dada por

$$\|X\|_P = \|X\|_{sp} \quad \text{para} \quad X \in T_P \mathcal{G}_n.$$

Con esta estructura, la componente Grassmanniana $\{UPU^* : U \in \mathcal{U}(n)\}$ es isométrica (módulo un factor 2) a la órbita $\{US_P U^* : U \in \mathcal{U}(n)\}$ de la simetría $S_P = 2P - I$. Una cuenta sencilla muestra que, si $X = XP + PX$, entonces

$$e^{iX} S_P = S_P e^{-iX}.$$

Esta simple observación nos permite usar los resultados previos obtenidos en el grupo unitario y aplicarlos a la Grassmanniana. En este caso, obtuvimos los siguientes resultados.

Teorema. Sean $P, Q \in \mathcal{G}_n$ con $\|P - Q\|_{sp} < 1$. Sea $X \in \mathcal{H}(n)$ tal que $X = PX + XP \in T_P \mathcal{G}_n$ y $Q = e^{iX} P e^{-iX}$. Entonces $\|X\|_{sp} < \pi/2$ por lo cual $\sigma(|X|) \subseteq [0, \pi/2)$. Si $\sigma(|X|)$ tiene más de un elemento, entonces una curva minimal que une P con Q tiene la siguiente estructura:

$$\delta(t) = \begin{pmatrix} \gamma_X(t) & 0 \\ 0 & \alpha_1(t) \end{pmatrix} \begin{pmatrix} \mathcal{S}_X \\ \mathcal{S}_X^\perp \end{pmatrix},$$

donde $\alpha_1 : [0, 1] \rightarrow \mathcal{G}(\mathcal{S}_X^\perp)$ es cualquier curva que une, $P|_{\mathcal{S}_X^\perp}$ con $Q|_{\mathcal{S}_X^\perp}$, tal que $\|\dot{\alpha}_1\|_{sp} \leq \|X\|_{sp}$.

Corolario. Dados $P, Q \in \mathcal{G}_n$ con $\|P - Q\|_{sp} < 1$, existe una única curva minimal que los une si, y sólo si, $\sigma(S_Q S_P) \subseteq \{e^{i\theta}, e^{-i\theta}\}$ para algún $|\theta| < \pi/2$.

1.1 Estructura de la tesis

La tesis se divide como sigue:

- En el capítulo 2 damos las herramientas que necesitamos para el capítulo 3. Definimos los espacios de Hadamard, damos ejemplos de los mismos y algunos resultados preliminares. Luego definimos el baricentro. A continuación pasamos a dar algunas nociones básicas de Teoría ergódica. Particularmente nos interesa el Teorema de Birkhoff y su relación con la ley de los grandes números. Por último pasamos a sistemas de Kronecker. Estos resultados son necesarios para el desarrollo del capítulo 3. No obstante, si el lector conoce estos temas, puede obviar este capítulo y pasar al siguiente.

- En el capítulo 3 damos la demostración del Teorema ergódico para espacios de Hadamard. Comenzamos el capítulo con una motivación del mismo. La relevancia de esta motivación yace en el hecho que utilizamos algunos lemas relacionados con la misma. Los resultados originales se encuentran en las secciones 3.3 y 3.4, donde demostramos el Teorema ergódico para el caso continuo y L^1 respectivamente. A su vez, utilizamos una nueva noción de mollifiers en espacios de Hadamard, la cual es interesante por sí misma y había sido probada por Karcher para variedades Riemannianas [31].
- En el capítulo 4 estudiamos la estructura de las curvas minimales en el caso de la norma traza. Comenzamos el capítulo motivando este problema, a partir de un estudio análogo realizado por Y. Lim en [40], en el espacio de matrices positivas. Los resultados originales se encuentran en la sección 4.2 y los resultados más importantes obtenidos son: la caracterización de las curvas minimales para matrices Hermitianas y positivas.
- En el capítulo 5 estudiamos la estructura de las curvas minimales para matrices unitarias medidas con la norma espectral. Los resultados originales se encuentran en las secciones 5.3, 5.4 y 5.5. En la primera, estudiamos dicha estructura. En la segunda, la geometría de los puntos medios y en la tercera, la Grassmaniana. .
- En el capítulo 6 incluimos resultados que utilizamos del análisis matricial. Entre estos se encuentran resultados de mayorización, normas unitariamente invariantes u otros. A su vez se pueden encontrar resultados que utilizamos sobre la geometría de las matrices positivas. También agregamos algunos resultados que son conocidos, pero tal vez no en toda su generalidad (para espacios métricos), como el Teorema de Lusin.

2 | Preliminares del Capítulo 3

2.1 Espacios de Hadamard

En esta sección resumimos algunos hechos básicos sobre espacios de Hadamard, también llamados espacios globales CAT(0) o de curvatura no positiva. Esta área de estudio empezó con el trabajo de Alexandrov [1] y Reshetnyak [55]. Hoy en día existe una gran bibliografía sobre este tema [7], [8], [18], [30] y otros.

Definición. Un espacio métrico completo (M, δ) se llama **espacio de Hadamard** si la distancia satisface la ley del semiparalelogramo, es decir, para todo $x, y \in M$ existe $m \in M$ tal que

$$\delta^2(m, z) \leq \frac{1}{2}\delta^2(x, z) + \frac{1}{2}\delta^2(y, z) - \frac{1}{4}\delta^2(x, y) \quad (2.1.1)$$

para todo $z \in M$. El punto m se llama **punto medio** entre x e y .

Unicidad y propiedades del punto medio: Tomando $z = x$ y $z = y$ en la desigualdad (2.1.1), es fácil ver que,

$$\delta(x, m) = \delta(m, y) = \frac{1}{2}\delta(x, y).$$

En efecto, tomando $z = x$, obtenemos

$$\delta(m, x) \leq \frac{1}{2}\delta(x, y).$$

Análogamente, tomando $z = y$, obtenemos

$$\delta(m, y) \leq \frac{1}{2}\delta(x, y).$$

Pero como

$$\delta(x, y) \leq \delta(x, m) + \delta(y, m) \leq \delta(x, y),$$

deducimos las igualdades antes mencionadas. Más aún, la desigualdad del semiparalelogramo implica que el punto medio es único. En efecto, si m, m' son puntos medios entre x e y ,

$$\begin{aligned} \delta^2(m, m') &\leq \frac{1}{2}\delta^2(x, m') + \frac{1}{2}\delta^2(y, m') - \frac{1}{4}\delta^2(x, y) \\ &\leq \frac{1}{2} \left[\frac{1}{2}\delta^2(x, x) + \frac{1}{2}\delta^2(y, x) - \frac{1}{4}\delta^2(x, y) \right] + \\ &\quad + \frac{1}{2} \left[\frac{1}{2}\delta^2(x, y) + \frac{1}{2}\delta^2(y, y) - \frac{1}{4}\delta^2(x, y) \right] - \frac{1}{4}\delta^2(x, y) = 0. \end{aligned}$$

La existencia y unicidad de puntos medios dan lugar a una **única geodésica** $\gamma_{a,b} : [0, 1] \rightarrow M$ que conecta dos puntos cualesquiera a y b . Para probarlo, primero se define $\gamma_{a,b}(1/2)$ como el punto medio entre a y b . Luego, usando un argumento inductivo, se define la geodésica para todos los números diádicos racionales en $[0, 1]$. Finalmente por completitud, se extiende para todo $t \in [0, 1]$. De ahora en adelante vamos a usar la notación $a \#_t b$ en vez de $\gamma_{a,b}(t)$.

La desigualdad (2.1.1) se extiende también para puntos arbitrarios en geodésicas.

Proposición 2.1.1. *Sea (M, δ) un espacio de Hadamard. Entonces para todo $t \in [0, 1]$ y $x, y, z \in M$,*

$$\delta^2(x \#_t y, z) \leq (1-t)\delta^2(x, z) + t\delta^2(y, z) - t(1-t)\delta^2(x, y). \quad (2.1.2)$$

Demostración. El razonamiento es análogo al anterior: Si $t = 1/2$ la desigualdad que obtenemos es la del semiparalelogramo. Luego se prueba inductivamente para todos los números diádicos racionales en $[0, 1]$. Finalmente por continuidad, se extiende para todo $t \in [0, 1]$. ■

Una consecuencia de este resultado es:

Corolario 2.1.2. *Dados cuatro puntos $a, a', b, b' \in M$ sea*

$$f(t) = \delta(a \#_t a', b \#_t b').$$

Entonces f es convexa en $[0, 1]$; es decir,

$$\delta(a \#_t a', b \#_t b') \leq (1-t)\delta(a, b) + t\delta(a', b'). \quad (2.1.3)$$

Demostración. Como f es continua es suficiente probar que es convexa en $t = 1/2$. Sea $m_1 = a \#_{1/2} a'$, $m_2 = b \#_{1/2} b'$ y $m = a \#_{1/2} b'$. Por la desigualdad del semiparalelogramo (2.1.1): $\delta(m_1, m) \leq \delta(a', b')/2$ y $\delta(m, m_2) \leq \delta(a, b)/2$. Luego

$$\delta(m_1, m_2) \leq \delta(m_1, m) + \delta(m, m_2) \leq \frac{1}{2} (\delta(a, b) + \delta(a', b')).$$

Concluimos esta sección con el llamado **Teorema de cuádruple comparación de Reshetnyak**.

Teorema 2.1.3. *Sea (M, δ) un espacio de Hadamard. Para todo $x_1, x_2, x_3, x_4 \in M$,*

$$\delta^2(x_1, x_3) + \delta^2(x_2, x_4) \leq \delta^2(x_2, x_3) + \delta^2(x_1, x_4) + 2\delta(x_1, x_2)\delta(x_3, x_4). \quad (2.1.4)$$

$$e^2 + f^2 \leq b^2 + d^2 + 2ac.$$

La demostración original de Reshetnyak [55] está basada en el hecho que el cuádruple puede ser *embebido* en un espacio Euclídeo de dimensión dos, de manera que, las longitudes de los laterales se preserva y las diagonales se expanden. Otra demostración que evita el *embedding* se encuentra en [30].

2.1.1 Ejemplos de espacios de Hadamard

Acabamos de definir los espacios de Hadamard como los espacios métricos completos que poseen como propiedad la ley del semiparalelogramo. Puede resultar poco intuitivo darnos cuenta qué espacios cumplen esta propiedad. Sin embargo, hay varios espacios que son muy utilizados que son espacios de Hadamard. En esta subsección vamos a mencionar algunos de ellos. Los ejemplos más conocidos de espacios de Hadamard son las variedades, los espacios de Hilbert y los espacios de *árbol*. También pueden ser construidos a partir de un espacio de Hadamard dado, como subconjuntos, imágenes, productos o espacios L^2 .

Comenzamos entonces con los ejemplos más utilizados:

Proposición 2.1.4. (*Variedades*). Sea (M, d) una variedad Riemanniana y sea d su distancia Riemanniana. Luego, (M, d) es un espacio de Hadamard si, y sólo si, es un espacio completo, simplemente conexo y de curvatura seccional no positiva.

Proposición 2.1.5. *Todo espacio de Hilbert es un espacio de Hadamard.*

También podemos encontrar un ejemplo en la teoría de grafos. Un **árbol** es un grafo no dirigido en el que dos vértices están conectados por exactamente un camino. En otras palabras, cualquier grafo conectado acíclico es un árbol.

Proposición 2.1.6. *Todo árbol es un espacio de Hadamard.*

Un ejemplo, particularmente importante de árbol, son las *telarañas*.

Ejemplo. (*Telarañas*). Sea K un conjunto arbitrario y para todo $i \in K$, sea

$$N_i := \{(i, r) : r \in \mathbb{R}^+\},$$

la copia de $\mathbb{R}^+$ (dotada con la métrica usual). Definimos la *telarañas* sobre K o K -*telarañas* (M, d) a partir del pegado de todos estos espacios juntos $N_i, i \in K$, en el origen. Es decir,

$$M = \{(i, r) : i \in K, r \in \mathbb{R}^+\} / \sim$$

donde $(i, 0) \sim (j, 0)$ para todo i, j . La distancia que consideramos es

$$d((i, r), (j, s)) = \begin{cases} |r - s| & \text{si } i = j \\ |r| + |s| & \text{en caso contrario} \end{cases}.$$

Los rayos n_i pueden ser considerados como subconjuntos cerrados de M . Dos rayos N_i y N_j con $i \neq j$ se intersecan en el origen $o := (i, 0) = (j, 0)$ de N .

Los siguientes ejemplos se construyen a partir de un espacio de Hadamard.

Lema 2.1.7. (*Subconjuntos*). *Un subconjunto $N \subset M$ de un espacio de Hadamard M es un espacio de Hadamard si, y sólo si, es cerrado y convexo.*

Lema 2.1.8. (*Imágenes*). *Sea $\varphi : M \rightarrow M'$ una isometría entre los espacios (M, d) y (M', d') . Luego, (M, d) es un espacio de Hadamard si, y sólo si, (M', d') es un espacio de Hadamard.*

Lema 2.1.9. (*Productos*). *El producto directo de espacios métricos (M_i, d_i) , $i = 1, \dots, k$ es el espacio métrico (M, d) definido por*

$$M = \prod_{i=1}^k M_i, \quad d(x, y) = \left(\sum_{i=1}^k d_i^2(x_i, y_i) \right)^{1/2}.$$

Este es un espacio de Hadamard si todos sus factores lo son.

Lema 2.1.10. (*Espacios L^2*). *Para cada espacio medible $(\Omega, \mathcal{M}, m)$, cada espacio métrico (M, d) y cada función fuertemente medible $h : \Omega \rightarrow M$, el espacio $L^2(\Omega, M, h)$ de todas las funciones fuertemente medibles $f : \Omega \rightarrow M$ con $d_2(f, h) < \infty$, es un espacio métrico completo con la métrica*

$$d_2(f, g) = \left(\int_{\Omega} d(f(x), g(x))^2 m(dx) \right)^{1/2}.$$

Si (M, d) es un espacio de Hadamard entonces el espacio $L^2(\Omega, M, h)$ también lo es.

Proposición 2.1.11. (*Límites*). *El límite de Gromov-Hausdorff (M, d) de una sucesión convergente de espacios de Hadamard (M_k, d_k) , $k \in \mathbb{N}$, es un espacio de Hadamard.*

2.2 Baricentros

Sea (M, δ) un espacio de Hadamard, y sea $\mathcal{B}(M)$ la σ -álgebra de Borel (es decir, la σ -álgebra más chica que contiene los conjuntos abiertos). Denotemos por $\mathcal{P}(M)$ al conjunto de todas las medidas de probabilidad en $\mathcal{B}(M)$ con soporte separable. Para $1 \leq \theta < \infty$, sea $\mathcal{P}^\theta(M)$, el conjunto de $\mu \in \mathcal{P}(M)$ tales que

$$\int \delta^\theta(x, y) d\mu(y) < \infty,$$

para algún (y por lo tanto para todo) $x \in M$. Por medio de $\mathcal{P}^\infty(M)$ denotamos al conjunto de todas las medidas en $\mathcal{P}(M)$ con soporte acotado.

Proposición 2.2.1. Sea (M, δ) un espacio de Hadamard y fijemos $y \in M$. Para cada $\mu \in \mathcal{P}^1(M)$ existe un único punto $\beta_\mu \in M$ que minimiza la función continua uniformemente convexa

$$z \mapsto \int_M [\delta^2(z, x) - \delta^2(y, x)] d\mu(x).$$

Este punto es independiente de y .

Utilizaremos la notación del trabajo de Sturm [56]: llamaremos **baricentro** al punto β_μ de μ . Notar que si $\mu \in \mathcal{P}^2(M)$ entonces β_μ coincide con la definición usual de baricentro de Cartan:

$$\operatorname{argmin}_{z \in M} \int_M \delta^2(z, x) d\mu(x).$$

La siguiente desigualdad que satisfacen los baricentros resulta de mucha utilidad.

Proposición 2.2.2 (Desigualdad de la varianza). Sea (M, δ) un espacio de Hadamard. Para toda medida de probabilidad $\mu \in \mathcal{P}^1(M)$ y para todo $z \in M$:

$$\delta^2(z, \beta_\mu) \leq \int_M [\delta^2(z, x) - \delta^2(\beta_\mu, x)] d\mu(x). \quad (2.2.1)$$

Sea (Ω, P) un espacio de probabilidad arbitrario y sea $F : \Omega \rightarrow M$ una función medible. Esta función define una medida de probabilidad $F_*P \in \mathcal{P}(M)$ dada por

$$F_*P(A) := P(F^{-1}(A)) = P(\{\omega \in \Omega : F(\omega) \in A\}) \quad (\forall A \in \mathcal{B}(M)),$$

la cual se llama *pushforward measure* de P por la función F . En lenguaje probabilístico, el *pushforward measure* F_*P se llama *función de distribución* de F .

Dado $1 \leq \theta \leq \infty$, decimos que $F \in L^\theta(\Omega, M)$ si $F_*P \in \mathcal{P}^\theta(M)$. En otras palabras, para $1 \leq \theta < \infty$, decimos que $F \in L^\theta(\Omega, M)$ si para algún (y por lo tanto para todo) $y \in M$ se tiene que

$$\int_\Omega \delta^\theta(F(\omega), y) dP(\omega) < \infty. \quad (2.2.2)$$

Por otro lado, decimos que $F \in L^\infty(\Omega, M)$ si para algún (y por lo tanto para todo) $y \in M$ la función $\omega \mapsto \delta(F(\omega), y)$ está esencialmente acotada.

2.3 Teoría ergódica

El siguiente enunciado: “Si dos sistemas están en equilibrio térmico con un tercer sistema, entonces están en equilibrio térmico entre sí”, es una conocida ley de la termodinámica. Desde un punto de vista microscópico, tal vez es sorprendente que cualquier sistema aproxime un punto de equilibrio, ya que microscópicamente, no hay estados estables, y por lo tanto no hay equilibrio. Sin embargo, por su naturaleza, las leyes de la termodinámica no son estados absolutos en un sistema en un tiempo fijo. Esto sugiere que se estudie el siguiente hecho: Sea Γ un estado de fases. Para todo t (tiempo) existe una función $T_t : \Gamma \rightarrow \Gamma$, donde $T_t(x)$ es un estado que resulta de tomar el estado x en el tiempo t_0 y esperar al tiempo $t_0 + t$. En

mecánica clásica, los observables de un sistema como la energía o el momento angular, son funciones en un espacio de fases. Esto sugiere estudiar

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(T_t(x)) dt.$$

Los resultados clásicos de la teoría ergódica estudian este fenómeno.

Para motivar la siguiente definición, sea s fijo y supongamos que χ_F es la función característica de un conjunto medible F . Entonces

$$\frac{1}{T} \int_0^T \chi_{s^{-1}F}(T_t(w)) dt = \frac{1}{T} \int_0^T \chi_F(T_s T_t(w)) dt.$$

Si el límite cuando $T \rightarrow \infty$ existe, entonces $\mu(T_s^{-1}F) = \mu(F)$.

Definición. Sean $(X, \mathcal{A}, \mu)$ un espacio de medida y $T : X \rightarrow X$. Decimos que T preserva μ o que μ es **T-invariante** si para todo $A \in \mathcal{A}$, $T^{-1}(A) \in \mathcal{A}$ y $\mu(A) = \mu(T^{-1}(A))$.

El siguiente teorema nos da una condición suficiente para la existencia de una medida T -invariante.

Teorema 2.3.1. (Krylov-Bogoliubov). Sean X un espacio métrico compacto, $T : X \rightarrow X$ una función continua y $\mathcal{B}$ la σ -álgebra de Borel. Entonces existe una medida μ de probabilidad T -invariante.

Uno de los teoremas clásicos de la teoría ergódica es el **Teorema de Birkhoff**. Este teorema permite sustituir promedios temporales del sistema, por un promedio espacial sobre una región del espacio de las fases. Más precisamente:

Teorema 2.3.2. (Birkhoff). Sean $(X, \mathcal{A}, \mu)$ un espacio de probabilidad y $T : X \rightarrow X$ una transformación que preserva medida. Sea $f \in L^1(X)$, entonces

$$\left\{ \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) \right\},$$

converge casi en todo punto a $\tilde{f} \in L^1(X)$ y

$$\int_X f d\mu = \int_X \tilde{f} d\mu.$$

Demostración. Para $\alpha \in \mathbb{R}$, sean

$$E_\alpha^+(f) = \{x : f^+(x) > \alpha\}, \quad E_\alpha^-(f) = \{x : f^-(x) < \alpha\}.$$

Luego

$$\begin{aligned} T^{-1}(E_\alpha^+(f)) &= E_\alpha^+(f); \\ T^{-1}(E_\alpha^-(f)) &= E_\alpha^-(f); \\ E_\alpha^+(f) &= E_0^+(f - \alpha); \\ E_\alpha^-(f) &= E_{-\alpha}^+(-f). \end{aligned}$$

Afirmamos que si $A \subset E_\alpha^+(f)$ y $T^{-1}A = A$, entonces

$$\int_A f d\mu \geq \alpha \mu(A).$$

En efecto

$$\int_A f d\mu = \int_A (f - \alpha) d\mu + \alpha \mu(A).$$

Como $A \subset E_\alpha^+(f) = E_0^+(f - \alpha) \subset E(f - \alpha)$, entonces

$$\int_A (f - \alpha) d\mu \geq 0.$$

Análogamente, si $A \subset E_\beta^-(f)$ y $T^{-1}A = A$, entonces

$$\int_A f d\mu \leq \beta \mu(A).$$

Por lo tanto, tomando

$$A = E_\alpha^+(f) \cap E_\beta^-(f),$$

tenemos

$$\alpha \mu(A) \leq \int_A f d\mu \leq \beta \mu(A),$$

y así $\mu(E_\alpha^+(f) \cap E_\beta^-(f)) = 0$.

Tomando una sucesión $\{\alpha_n\}$ en $\mathbb{R}$ tenemos

$$\{x : f^+(x) > f^-(x)\} = \bigcup_{\alpha_n > \alpha_m} E_{\alpha_n}^+(f) \cap E_{\alpha_m}^-(f),$$

lo que implica $\mu(\{x : f^+(x) > f^-(x)\}) = 0$.

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} |f(T^j(x))| = g,$$

en casi todo punto y

$$|\tilde{f}| \leq g.$$

Como T preserva μ ,

$$\int_X \frac{1}{n} \sum_{j=0}^{n-1} |f \circ T^j| d\mu = \frac{1}{n} \sum_{j=0}^{n-1} \int_X |f \circ T^j| d\mu = \int_X |f| d\mu.$$

Por el Lema de Fatou

$$\int_X g d\mu \leq \int_X |f| d\mu,$$

y así

$$\int_X |\tilde{f}| d\mu \leq \int_X |f| d\mu.$$

Si $f \in L^\infty(X)$,

$$\left\| \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j \right\|_\infty \leq \|f\|_\infty.$$

Así

$$\|\tilde{f}\|_\infty \leq \|f\|_\infty \quad \text{y} \quad \left\| \tilde{f} - \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j \right\|_\infty \leq 2\|f\|_\infty.$$

Por convergencia dominada

$$\lim_{n \rightarrow \infty} \int_X \left| \tilde{f} - \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j \right| d\mu = \int_X \lim_{n \rightarrow \infty} \left| \tilde{f} - \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j \right| d\mu = 0.$$

En el caso general, dado $\varepsilon > 0$, existe $f_0 \in L^1(X)$ tal que

$$\int_X |f - f_0| d\mu \leq \varepsilon,$$

y tomamos N tal que si $n \geq N$,

$$\int_X \left| f_0 - \frac{1}{n} \sum_{j=0}^{n-1} f_0 \circ T^j \right| d\mu \leq \varepsilon.$$

Entonces

$$\begin{aligned} \int_X \left| \tilde{f} - \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j \right| d\mu &\leq \int_X |\tilde{f} - \tilde{f}_0| d\mu + \int_X \left| \tilde{f}_0 - \frac{1}{n} \sum_{j=0}^{n-1} f_0 \circ T^j \right| d\mu \\ &\quad + \int_X \left| \frac{1}{n} \sum_{j=0}^{n-1} (f - f_0) \circ T^j \right| d\mu. \end{aligned}$$

Pero

$$\int_X |\tilde{f} - \tilde{f}_0| d\mu = \int_X |\widetilde{(f - f_0)}| d\mu \leq \int_X |f - f_0| d\mu \leq \varepsilon,$$

y

$$\int_X \left| \frac{1}{n} \sum_{j=0}^{n-1} (f - f_0) \circ T^j \right| d\mu \leq \int_X |f - f_0| d\mu \leq \varepsilon.$$

Luego,

$$\int_X \left| \tilde{f} - \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j \right| d\mu \leq 3\varepsilon,$$

si $n \geq N$. Luego,

$$\int_X \tilde{f} d\mu = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \int_X f \circ T^j d\mu = \int_X f d\mu.$$

■

Es útil a veces expresar la ergodicidad en términos de la medida.

Definición. Sean $(X, \mathcal{A}, \mu)$ un espacio de probabilidad y $T : X \rightarrow X$ una transformación que preserva medida. Decimos que T es μ -ergódica o que μ es una medida **T-ergódica**, si para $B \in \mathcal{A}$, $T^{-1}(B) = B$ implica $\mu(B) = 0$ ó 1 .

Lema 2.3.3. Son equivalentes

1. T es μ -ergódica,
2. $f \circ T = f$, $f \in L^1(X, \mathcal{A}, \mu)$ entonces f es constante,
3. $f \circ T = f$, $f \in L^2(X, \mathcal{A}, \mu)$ entonces f es constante.

Una consecuencia inmediata del Teorema de Birkhoff y las equivalencias dadas en el Lema anterior 2.3.3 es la siguiente.

Corolario 2.3.4. Si T es μ -ergódica entonces

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) = \int f d\mu, \quad (2.3.1)$$

para toda $f \in L^1(X, \mathcal{A}, \mu)$.

2.4 Teorema ergódico de Birkhoff y la ley de los grandes números

Primero necesitamos recordar diferentes maneras en las cuales una sucesión de variables aleatorias puede converger. Sea $\{Y_n\}$ una sucesión de variables aleatorias e Y una variable aleatoria a la cual queremos que la sucesión *converja*.

Definición. La noción más fuerte de convergencia es **casí seguro**. Decimos que

$$Y_n \xrightarrow{a.s.} Y$$

si

$$\mathbb{P}(Y_n \rightarrow Y) = 1.$$

Si Ω es un espacio de probabilidad en el cual las variables aleatorias están definidas y ν es la medida de probabilidad definida por $\mathbb{P}$, entonces esta condición puede ser reescrita como

$$\nu\{\omega \in \Omega \mid Y_n(\omega) \rightarrow Y(\omega)\} = 1.$$

Definición. Una noción más débil de convergencia es la **convergencia en probabilidad**. Decimos que

$$Y_n \xrightarrow{p} Y$$

si

$$\mathbb{P}(|Y_n - Y| \geq \epsilon) \rightarrow 0 \text{ para todo } \epsilon > 0.$$

En términos de Ω y ν , esta condición es

$$\nu\{\omega \in \Omega \mid |Y_n(\omega) - Y(\omega)| \geq \epsilon\} \rightarrow 0.$$

Definición. Una noción aún más débil de convergencia es la **convergencia en distribución**. Decimos que

$$Y_n \xrightarrow{d} Y$$

si, escribiendo $F_n, F: \mathbb{R} \rightarrow [0, 1]$ para la función de distribución acumulada de Y_n e Y respectivamente, tenemos

$$F_n(t) \rightarrow F(t)$$

para todo t donde $F(t)$ es continua.

Ley débil de los grandes números: Dada una sucesión de variables aleatorias X_n , consideremos las sumas

$$S_n = X_1 + X_2 + \cdots + X_n.$$

Entonces $\frac{S_n}{n}$ es el promedio de las primeras n observaciones. Supongamos que la sucesión X_n es independiente e idénticamente distribuida (IID) y que X_n es integrable. Es decir, $\mathbb{E}(|X_n|) < \infty$. Luego, en particular, la esperanza $\mu = \mathbb{E}(X_n)$ es finita. La ley débil de grandes números dice que $\frac{S_n}{n}$ converge en probabilidad a la función constante μ . Como la distribución límite es constante, es suficiente probar convergencia en distribución. Este hecho nos lleva a una prueba muy conocida de la ley débil usando funciones características.

Si una variable aleatoria Y es absolutamente continua, es decir, si tiene función de densidad f , entonces su función característica φ_Y es la trasformada de Fourier de f . Más generalmente, la función característica de Y es

$$\varphi_Y(t) = \mathbb{E}(e^{itY}).$$

Las funciones características están relacionadas a la convergencia en distribución por el Teorema de continuidad de Lévy, el cual dice (entre otras cosas) que $Y_n \xrightarrow{d} Y$ si, y sólo si, $\varphi_{Y_n}(t) \rightarrow \varphi_Y(t)$ para todo $t \in \mathbb{R}$. En particular, para probar la ley débil de grandes números es suficiente demostrar que las funciones características de $\frac{1}{n}S_n$ convergen a la función $e^{it\mu}$.

Sea φ la función característica de X_n . Observar que cada X_n tiene la misma función característica porque están idénticamente distribuidas. Sea φ_n la función característica de $\frac{1}{n}S_n$. Luego

$$\varphi_n(t) = \mathbb{E}(e^{\frac{it}{n}(X_1 + \cdots + X_n)}).$$

Como las variables X_n son independientes, tenemos

$$\varphi_n(t) = \prod_{j=1}^n \mathbb{E}(e^{\frac{it}{n}X_j}) = \varphi\left(\frac{t}{n}\right)^n. \quad (2.4.1)$$

Por el Teorema de Taylor y por linealidad de la esperanza, tenemos para $t \approx 0$ que

$$\varphi(t) = \mathbb{E}(e^{itX_j}) = \mathbb{E}(1 + itX_j + o(t^2)) = 1 + it\mu + o(t),$$

y junto con (2.4.1) obtenemos

$$\varphi_n(t) = \left(1 + \frac{it\mu}{n} + o(t/n)\right)^n \rightarrow e^{it\mu},$$

lo que completa la prueba.

Teorema ergódico y ley fuerte de los grandes números: la ley fuerte de los grandes números establece que no sólo $\frac{1}{n}S_n$ converge a μ en probabilidad, también converge casi seguro. En este caso no incluiremos la demostración clásica, pero vamos a ver que la ley fuerte de grandes números se puede ver como un caso especial del Teorema ergódico de Birkhoff (en particular del Corolario 2.3.4). Sea $X_n : \Omega \rightarrow \mathbb{R}$ una sucesión de variables aleatorias IID, y definamos la función $\pi : \Omega \rightarrow \mathbb{R}^{\mathbb{N}}$ por

$$\pi(\omega) = (X_1(\omega), X_2(\omega), \dots).$$

Sea ν la medida de probabilidad en Ω que determina $\mathbb{P}$, y sea $\mu = \pi_*\nu = \nu \circ \pi^{-1}$ la medida de probabilidad correspondiente en $\mathbb{R}^{\mathbb{N}}$.

Como las variables X_n son independientes, μ tiene la forma $\mu = \nu_1 \times \nu_2 \times \dots$, y como son idénticamente distribuidas todas las distribuciones marginales ν_j son las mismas, entonces $\mu = \nu^{\mathbb{N}}$ para alguna distribución de probabilidad ν en $\mathbb{R}$.

La medida μ es invariante y ergódica con respecto a las dinámicas en $\mathbb{R}^{\mathbb{N}}$ dadas por la función *shift* $f(x_1, x_2, x_3, \dots) = (x_2, x_3, x_4, \dots)$. Escribiendo $x = (x_1, x_2, x_3, \dots) \in \mathbb{R}^{\mathbb{N}}$ y tomando $\varphi(x) = x_1$, vemos que para $x = \pi(\omega)$ tenemos

$$\begin{aligned}\pi(\omega) &= (X_1(\omega), X_2(\omega), \dots); \\ f^n(\pi(\omega)) &= (X_{n+1}(\omega), X_{n+2}(\omega), \dots); \\ \varphi(f^n(\pi(\omega))) &= X_{n+1}(\omega).\end{aligned}$$

Luego

$$\begin{aligned}\frac{X_1(\omega) + \dots + X_n(\omega)}{n} &= \frac{\varphi(f^0(\pi(\omega))) + \varphi(f^1(\pi(\omega))) + \dots + \varphi(f^{n-1}(\pi(\omega)))}{n} \\ &= \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(\pi(\omega))).\end{aligned}$$

En particular, la convergencia en (2.3.1) implica la ley de los grandes números.

2.5 Sistemas de Kronecker

Un sistema dinámico es un sistema en el que una función describe el comportamiento de un punto dependiente del tiempo (de manera continua o discreta) en un espacio geométrico. Por ejemplo: una partícula o conjunto de partículas cuyo estado varía con el tiempo, y por lo tanto, obedece a ecuaciones diferenciales que implican derivadas. Con el fin de hacer una predicción sobre el comportamiento futuro del sistema, se realiza una solución analítica de tales ecuaciones. Esta sección está basada en [57] (capítulo 2). Un **sistema dinámico** puede definirse formalmente, como una transformación que preserva la medida de una σ -álgebra.

Definición. Un *sistema topológico dinámico* $(X, T) = (X, \mathcal{F}, T)$ es un espacio topológico, compacto y metrizable (y por lo tanto Hausdorff) tal que T es un isomorfismo topológico (es decir, un homeomorfismo).

En particular estamos interesados en dos tipos de sistemas.

Definición. (Sistemas equicontínuos e isométricos). Sea $(X, \mathcal{F}, T)$ un sistema topológico dinámico.

- Decimos que un sistema es **isométrico** si existe una métrica d en X tal que las funciones shift $T^n : X \rightarrow X$ son todas isometrías. Por lo tanto, $d(T^n x, T^n y) = d(x, y)$ para todo $n \in \mathbb{N}$, $x, y \in X$. Notar que una vez que T es una isometría, entonces todas las potencias T^n son isometrías.
- Decimos que un sistema es **equicontínuo** si existe una métrica d en X tal que las funciones shift $T^n : X \rightarrow X$ forman una familia uniformemente equicontinua. Por lo tanto, para todo $\varepsilon > 0$ existe $\delta > 0$ tal que $d(T^n x, T^n y) \leq \varepsilon$ siempre que n, x, y son tales que $d(x, y) \leq \delta$. Notar que como X es compacto, equicontinuidad y uniformemente equicontinua son conceptos equivalentes.

Proposición 2.5.1. Todo sistema topológico dinámico es isométrico si, y sólo si, es equicontínuo.

Demostración. Si el sistema es isométrico entonces es equicontínuo, se deduce directamente de la definición. Por otro lado, si T^n es una familia equicontinua con respecto a la métrica d , considerar la métrica

$$\tilde{d}(x, y) = \sup_{n \in \mathbb{N}} d(T^n x, T^n y).$$

■

En vista de este resultado podemos mejorar la estructura topológica (equicontinuidad) utilizando la estructura geométrica (isometrías). También podemos mejorar la estructura geométrica de isometrías utilizando la estructura algebraica de acciones de grupos compactos abelianos.

Definición. (Sistemas de Kronecker). Un sistema topológico dinámico $(X, \mathcal{F}, T)$ es un **sistema de Kronecker** si es isomorfo a un sistema de la forma $(K, \mathcal{K}, S)$ donde $(K, +, \mathcal{K})$ es un grupo topológico abeliano, compacto y metrizable, y $S : x \mapsto x + \alpha$ es una rotación del grupo para algún $\alpha \in K$.

Observar que **un sistema de Kronecker es equicontínuo** (y por lo tanto isométrico). En efecto, la compacidad del grupo topológico K nos asegura que la rotación del grupo $g : x \mapsto x + g$ es uniformemente equicontinua cuando varía $g \in K$. Como los shifts $T^n : x \mapsto x + n\alpha$ son todas rotaciones del grupo, obtenemos el resultado. Por otro lado, no todo sistema isométrico o equicontínuo es un sistema de Kronecker. Sin embargo, contienen un sistema de Kronecker.

Definición. Un sistema dinámico **minimal** es un sistema (X, T) que no tiene subsistemas propios (Y, S) . Un sistema topológico dinámico minimal es un sistema $(X, \mathcal{F}, T)$ que no tiene subsistemas propios $(Y, \mathcal{G}, S)$.

Dado un punto x de un sistema (X, f) , denotamos por,

$$\text{Orb}_f = \{x, f(x), f_2(x), \dots\},$$

la órbita de f y $\omega_f(x)$ denota su conjunto ω -límite, es decir, el conjunto de puntos límites de la sucesión $x, f(x), f_2(x), \dots$

Proposición 2.5.2. *Las siguientes condiciones son equivalentes:*

1. (X, f) es minimal,
2. Toda órbita es densa en X ,
3. $\omega_f(x) = X$ para todo $x \in X$.

Una función minimal f es necesariamente suryectiva si X es un conjunto compacto Hausdorff.

Proposición 2.5.3. *Todo sistema $(X, \mathcal{F}, T)$ minimal equicontinuo (o isométrico) es un sistema de Kronecker. Es decir, es isomorfo a una rotación del grupo abeliano $(K, \mathcal{K}, x \mapsto x + \alpha)$. Más aún, la órbita $\{n\alpha : n \in \mathbb{Z}\}$ es densa en K .*

Demostración. Por la Proposición 2.5.1 podemos suponer que el sistema es isométrico, por lo tanto, podemos encontrar una métrica d tal que las funciones shift T^n son isometrías. Hemos visto que T^n pertenece al espacio $C(X \rightarrow X)$ de funciones continuas de X en sí mismo (dotado de la topología uniforme). Sea $G \subset C(X \rightarrow X)$ la clausura de las funciones $\{T^n : n \in \mathbb{Z}\}$. Es fácil ver que G es un grupo topológico, cerrado y metrizable de isometrías en $C(X \rightarrow X)$; por el Teorema de Arzelà-Ascoli G es compacto. Además, como T^n y T^m conmutan para todo n, m , G es abeliano.

Sea $x \in X$, entonces la imagen $\{f(x) : x \in G\}$ de G bajo la evaluación $f \mapsto f(x)$ es un subconjunto compacto, no vacío, invariante de X , y por lo tanto, es igual a todo X por minimalidad. Si definimos el estabilizador $\Gamma := \{f : f(x) = x\}$, entonces Γ es un subgrupo cerrado (por lo tanto compacto) del grupo abeliano G . Como $X = \{f(x) : f \in G\}$, hay una biyección continua $f\Gamma \mapsto f(x)$ del grupo cociente $K := G/\Gamma$ (dotado de la topología cociente) en X . Como ambos espacios son compactos Hausdorff, la función es un homeomorfismo. Luego, esta función es un isomorfismo de sistemas topológicos dinámicos entre el sistema de Kronecker K y X . Como K es un grupo topológico compacto metrizable (por la distancia Hausdorff) obtenemos el resultado. Notar que la densidad de $\{n\alpha : n \in \mathbb{Z}\}$ en K es directa por la construcción. ■

Proposición 2.5.4. *Todo sistema isométrico $(X, \mathcal{F}, T)$ se puede partir en una unión disjunta de sistemas isométricos minimales.*

Demostración. Como los sistemas minimales son automáticamente disjuntos, es suficiente probar que todo punto $x \in X$ está contenido en un sistema dinámico minimal, o equivalentemente, que la clausura de la órbita $\overline{T^{\mathbb{Z}}x}$ es minimal. Si esto no sucede, entonces existe $y \in \overline{T^{\mathbb{Z}}x}$ tal que x no pertenece a la clausura de la órbita de y . Pero por definición de la clausura de la órbita, podemos encontrar una sucesión n_j tal que $T^{n_j}x$ converge a y . Por propiedad de las isometrías, esto implica que $T^{-n_j}y$ converge a x . Luego, x está en la clausura de la órbita de y , lo cual es una contradicción. ■

2.5.1 Sistemas de Kronecker y medidas de Haar.

En esta parte vamos a considerar aquellos sistemas que preservan medida $(X, \mathcal{X}, \mu, T)$, que son compactos o casi periódicos. Estos sistemas son análogos a los sistemas equicontinuos o isométricos en dinámicas topológicas y, con estos sistemas, vamos a poder caracterizar algebraicamente los sistemas que preservan medida (más precisamente los ergódicos) como sistemas de Kronecker.

Definición. Sea $(X, \mathcal{X}, \mu, T)$ un sistema que preserva medida. Una función $f \in (X, \mathcal{X}, \mu)$ es **casi periódica** si la clausura de la órbita $\{T^n f : n \in \mathbb{Z}\}$ es compacta en $L^2(X, \mathcal{X}, \mu)$.

Definición. Un sistema que preservan medida $(X, \mathcal{X}, \mu, T)$ es **compacto** si toda función en $L^2(X, \mathcal{X}, \mu)$ es casi periódica.

Una fuente de sistemas compactos proviene de sistemas de Kronecker. Tales sistemas son topológicos, y vamos a necesitar dotarlos con una medida canónica, que es la medida de Haar.

Sea G un grupo topológico compacto metrizable (no necesariamente abeliano). Sin un ambiente, no podemos definir la convolución $f * g$ de dos funciones continuas $f, g \in C(G)$. Sin embargo, podemos definir la convolución $\mu * f$ de una medida de Borel finita μ en G y una función continua $f \in C(G)$ como la función

$$\mu * f(x) := \int_G f(y^{-1}x) d\mu(y),$$

la cual (por la continuidad uniforme de f) es otra función continua. Análogamente definimos

$$f * \mu(x) := \int_G f(xy^{-1}) d\mu(y).$$

Decimos que un número c es una media a izquierda (resp. media a derecha) de una función $f \in C(G)$ si existe una medida de probabilidad μ tal que $\mu * f$ (resp. $f * \mu$) es igual a la constante c . Para grupos compactos metrizables G , esta media c , está bien definida.

Lema 2.5.5. (*Existencia y unicidad*). Sea G un grupo topológico compacto metrizable, y sea $f \in C(G)$. Entonces existe una única constante c que es media a derecha e izquierda de f .

La función $f \mapsto c$ que toma una función continua y nos devuelve su media es un funcional lineal acotado no negativo en $C(G)$ que preserva constantes, y por lo tanto, por el Teorema de representación de Riesz está dada por una única medida de probabilidad μ . Como la convolución a izquierda y derecha conmutan, esta medida es invariante a derecha e izquierda. Recíprocamente, dada cualquier medida μ de probabilidad Boreliana, se puede probar que $f * \mu = \mu * f = c$. Luego, tenemos el siguiente teorema:

Teorema 2.5.6. (*Existencia y unicidad de medidas de Haar*). Si G es un grupo topológico compacto metrizable, entonces existe una única medida de probabilidad Boreliana μ en G que es invariante a izquierda y derecha.

Como caso particular obtenemos el siguiente corolario.

Corolario 2.5.7. Todo sistema topológico de Kronecker $(K, x \mapsto x + \alpha)$ se puede convertir canónicamente en un sistema dinámico que preservan medida, el cual es compacto.

La demostración de este resultado está basada en los mismos argumentos que veremos en el siguiente ejemplo.

Ejemplo. Consideremos el toro $\mathbb{T}$ y el sistema shift $(\mathbb{T}, x \mapsto x + \alpha)$. Entonces toda función $f \in L^2(\mathbb{T})$ es casi periódica, porque la clausura de la órbita pertenece al conjunto $\{f(\cdot + \theta) : \theta \in \mathbb{T}\}$, que es la imagen continua del toro $\mathbb{T}$, y por lo tanto, es compacta.

Teorema 2.5.8. *(Clasificación de sistemas ergódicos compactos)*

1. *Todo sistema ergódico compacto es equivalente a un sistema de Kronecker (abeliano).*
2. *Todo sistema compacto se puede desintegrar en sistemas de Kronecker ergódicos.*

3 | Teorema Ergódico en espacios de Hadamard en términos de la media inductiva

3.1 Motivación

El problema que vamos a estudiar en este capítulo surge de manera natural en el estudio de la media geométrica y sus propiedades. Sin embargo es necesario recapitular algunos resultados recientes, particularmente en el espacio de matrices positivas, para entender mejor el problema. Parte de la motivación fue mencionada en la introducción de la tesis. En esta sección vamos a desarrollar con más profundidad ciertas cuestiones.

El espacio de matrices $n \times n$ complejas, $\mathcal{M}(n)$, es un espacio de Hilbert con el producto interno $\langle A, B \rangle = \text{tr}(A^*B)$ y la norma asociada $\|A\|_2 = (\text{tr}(A^*A))^{1/2}$. Esta norma es conocida como norma Frobenius o Hilbert-Schmidt. El conjunto de matrices Hermitianas $\mathcal{H}(n)$ es un subespacio vectorial real de $\mathcal{M}(n)$. Como el subconjunto $\mathcal{P}(n)$ de matrices (estrictamente) positivas es un subconjunto abierto en $\mathcal{H}(n)$, resulta ser una variedad diferenciable. En particular el espacio tangente a $\mathcal{P}(n)$ en un punto A , es el espacio $T_A\mathcal{P}(n) = \mathcal{H}(n)$. El producto interno en $\mathcal{H}(n)$ nos permite definir una métrica Riemanniana en la variedad $\mathcal{P}(n)$, la cual en un punto A está dada por,

$$ds = \|A^{-1/2}dAA^{-1/2}\|_2 = [\text{tr}(A^{-1}dA)^2]^{1/2}.$$

En otras palabras, dada una curva suave a trozos $\alpha : [a, b] \rightarrow \mathcal{P}(n)$, su longitud se define como

$$L(\alpha) = \int_a^b \|\alpha^{-1/2}(t)\dot{\alpha}(t)\alpha^{-1/2}(t)\|_2 dt.$$

A partir de esto, como es usual, la distancia rectificable entre $A, B \in \mathcal{P}(n)$ se define por medio de

$$\delta(A, B) = \inf \{L(\alpha) \text{ donde } \alpha : [a, b] \rightarrow \mathcal{P}(n) \text{ une } A \text{ con } B \text{ en } \mathcal{P}(n)\}.$$

Más aún, la distancia Riemanniana δ puede ser calculada por

$$\delta(A, B) = \|\log(A^{-1/2}BA^{-1/2})\|_2.$$

El ínfimo es en realidad un mínimo, y la geodésica que conecta dos matrices positivas A y B tiene la siguiente expresión simple:

$$\gamma_{AB}(t) = A^{1/2}(A^{-1/2}BA^{-1/2})^t A^{1/2}. \quad (3.1.1)$$

Es usual en análisis matricial usar la notación $A \#_t B$ en vez de $\gamma_{AB}(t)$. El punto medio $A \#_{1/2} B$ se conoce como **media geométrica** o **baricentro** entre A y B , y también admite la siguiente caracterización variacional

$$A \#_{\frac{1}{2}} B = \operatorname{argmin}_{C \in \mathcal{P}(n)} \left(\delta^2(A, C) + \delta^2(B, C) \right).$$

No hay razón para restringirnos a sólo dos matrices. La noción de **media geométrica** se puede generalizar para más de dos matrices de manera obvia

$$\Lambda(A_1, \dots, A_n) := \operatorname{argmin}_{C \in \mathcal{P}(n)} \left(\sum_{j=1}^n \delta^2(A_j, C) \right). \quad (3.1.2)$$

La solución del problema de minimización de cuadrados mínimos, existe y es única, por las propiedades de convexidad de la distancia $\delta(\cdot, \cdot)$. Esta media recibe muchos nombres diferentes dependiendo del contexto en el que se use. Algunos de ellos son: media de cuadrados mínimos, centro de gravedad, media de Frechet, media de Cartan, centro Riemanniano de masa, media geométrica Riemanniana, o más frecuentemente media de Karcher para el caso de matrices positivas.

La media geométrica aparece de manera natural en varios problemas aplicados. Por ejemplo en el estudio de señales de radar. En este problema, cada señal es detectada por más de un sensor. La información de cada sensor se codifica en un núcleo de covarianza, el cual es semidefinido positivo. Estos núcleos tienen que ser promediados para obtener el resultado final. Resulta que la mejor manera de promediar estos núcleos no es la media aritmética estándar, si no, la media geométrica que introducimos previamente (ver [19] y las referencias allí para más detalles). Otra aplicación típica de la media geométrica es con problemas relacionados con métodos de optimización del tipo Newton (ver [15],[46]).

El inconveniente usual al lidiar con la media geométrica es que, cuando trabajamos con tres o más matrices, no existe en general una fórmula cerrada para calcularla. Por esta razón, hubo una extensa investigación con el objetivo de encontrar buenas maneras de aproximar la media geométrica de tres o más matrices ([14], [35], [42]). Un avance importante se logró al intentar demostrar la monotonía de la media geométrica. En otras palabras, se desconocía si $\Lambda(A_1, \dots, A_n) \leq \Lambda(B_1, \dots, B_n)$, para $A_i \leq B_i$, donde $1 \leq i \leq n$. La resolución de este problema condujo a la introducción de las llamadas **medias inductivas**.

Para motivar la definición de las medias inductivas, notar que, dada una sucesión $\{a_n\}_{n \in \mathbb{N}}$ de números complejos tenemos la siguiente relación

$$\begin{aligned} \frac{a_1 + a_2 + a_3}{3} &= \frac{2}{3} \left(\frac{a_1 + a_2}{2} \right) + \frac{1}{3} a_3 \\ &\vdots \\ \frac{a_1 + \dots + a_n}{n} &= \frac{n-1}{n} \left(\frac{a_1 + \dots + a_{n-1}}{n-1} \right) + \frac{1}{n} a_n. \end{aligned}$$

Sea $\gamma_{a,b}(t) = (1-t)a + tb$, y por un momento usemos la notación $a \#_t b = \gamma_{a,b}(t)$. Entonces

$$\begin{aligned} \frac{a_1 + a_2 + a_3}{3} &= (a_1 \#_{\frac{1}{2}} a_2) \#_{\frac{1}{3}} a_3 \\ \frac{a_1 + a_2 + a_3 + a_4}{4} &= ((a_1 \#_{\frac{1}{2}} a_2) \#_{\frac{1}{3}} a_3) \#_{\frac{1}{4}} a_4 \end{aligned}$$

y así sucesivamente. En el espacio Euclídeo los segmentos son las geodésicas. En nuestro contexto, podemos reemplazar los segmentos por las geodésicas asociadas a la estructura Riemanniana. Esta es la idea que nos lleva a la definición de la media inductiva. Dada una sucesión de matrices estrictamente positivas $\{A_n\}_{n \in \mathbb{N}}$, las **medias inductivas** se definen de la siguiente manera:

$$\begin{aligned} S_1(A) &= A_1 \\ S_n(A) &= S_{n-1}(A) \#_{\frac{1}{n}} A_n \quad (n \geq 2). \end{aligned}$$

La **propiedad de monotonía** primero fue probada por Lawson y Lim (ver [35], Corolario 5.1 y Proposición 5.2), basándose en los resultados de Sturm sobre estadísticas en espacio de curvatura no positiva. El resultado en el que se basaron es el siguiente:

Proposición 3.1.1. ([56] *Ley de los grandes números*). Sea (M, d) un espacio de Hadamard con un baricentro contractivo

$$\beta : \mathcal{P}^1(M) \rightarrow M,$$

y fijemos $p \in \mathcal{P}^\infty(M)$. Más aún, sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad y $(X_i)_{i \in \mathbb{N}}$ una sucesión independiente de funciones medibles $X_i : \Omega \rightarrow M$ con distribución idéntica $\mathbb{P}_{X_i} = p$. Definimos el valor medio del baricentro $s_n : \Omega \rightarrow M$ como

$$s_n(\omega) := \beta \left(\frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)} \right).$$

Entonces para $\mathbb{P}$ -casi todo $\omega \in \Omega$,

$$s_n(\omega) \xrightarrow{n \rightarrow \infty} \beta(p).$$

Demostración. Ver [56], Proposición 6.6. ■

Lawson y Lim probaron que si la sucesión está formada por la elección aleatoria e independiente de elementos del conjunto $\{A_1, \dots, A_n\}$, entonces la media inductiva converge casi seguro a $\Lambda(A_1, \dots, A_n)$. Es decir,

$$\lim_{k \rightarrow \infty} S_k(A_{\omega(1)}, \dots, A_{\omega(k)}) = \Lambda(A_1, \dots, A_n) \quad \text{c.t.p. } \omega \in \{1, \dots, n\}^{\mathbb{N}}.$$

La relación con la monotonía de $\Lambda(\cdot)$ yace en el hecho que, la media geométrica se obtiene de una sucesión binaria de pasos, cada uno de los cuales es monótono. Es decir,

$$X_1 \leq X_2, Y_1 \leq Y_2 \quad \text{implica} \quad X_1 \#_t X_2 \leq Y_1 \#_t Y_2 \quad \text{para todo } t \in [0, 1]. \quad (3.1.3)$$

De esta manera la media inductiva resulta monótona y esta propiedad se preserva en el límite. Observar que el resultado de Lawson y Lim puede ser visto como una “ley fuerte de los grandes números”, pues como mencionamos en la motivación de la media inductiva, la traducción de este resultado al caso Euclídeo es la convergencia en casi todo punto de los promedios a su esperanza. Bhatia y Karandikar [14] simplificaron la prueba de la monotonía de la media geométrica notando que una “ley débil de los grandes números” es suficiente.

Con el fin de profundizar en la conexión que existe entre los resultados que obtuvimos y la “ley débil de los grandes números” obtenida por Bhatia y Karandikar, vamos a dar la demostración de la monotonidad de la media geométrica que se encuentra en [14].

Ley débil de los grandes números: Sean $A_1, \dots, A_m$ elementos de $\mathcal{P}(n)$, sea $G = \Lambda(A)$ la media geométrica de estas matrices, y

$$\alpha := \sum_{j=1}^m \frac{1}{m} \delta_2^2(G, A_j). \quad (3.1.4)$$

Para $n \geq 1$, sea $\mathcal{J}_n$ el conjunto de todas la n -uplas ordenadas $(j_1, \dots, j_n)$, con $j_k \in \{1, \dots, m\}$. Este conjunto tiene m^n elementos. Para cada elemento de este conjunto definimos una *media*, $\mathcal{M}_n(j_1, \dots, j_n)$, de las matrices dadas $A_1, \dots, A_m$, a partir del siguiente proceso inductivo:

$$\begin{aligned} \mathcal{M}_1(j) &= A_j, \text{ para todo } j \in \mathcal{J}_1 \\ \mathcal{M}_{n+1}(j_1, \dots, j_n, k) &= \mathcal{M}_n(j_1, \dots, j_n) \#_{\frac{1}{n+1}} A_k, \end{aligned}$$

para todo $(j_1, \dots, j_n) \in \mathcal{J}_n$ y $k \in \mathcal{J}_1$. En otras palabras

$$\mathcal{M}_{n+1}(j_1, \dots, j_n, k) = (\dots (((A_{j_1} \#_{\frac{1}{2}} A_{j_2}) \#_{\frac{1}{3}} A_{j_3}) \#_{\frac{1}{4}} A_{j_4}) \dots) \#_{\frac{1}{n+1}} A_k,$$

con $(j_1, \dots, j_n) \in \mathcal{J}_n$ y $k \in \mathcal{J}_1$. La *clave* de la demostración es la siguiente estimación:

Teorema 3.1.2. *Para todo $n \in \mathbb{N}$, tenemos*

$$\frac{1}{m^n} \sum_{(j_1, \dots, j_n) \in \mathcal{J}_n} \delta_2^2(G, \mathcal{M}_n(j_1, \dots, j_n)) \leq \frac{1}{n} \alpha. \quad (3.1.5)$$

Idea de la demostración. La prueba es por inducción. El caso $n = 1$ es trivial ya que tenemos α en ambos lados. Supongamos que es verdadero para n , y vamos a probar que es verdadero para $n + 1$. El argumento clave es la desigualdad del semiparalelogramo (2.1.1), que nos permite reducir $\mathcal{M}_{n+1}$ a $\mathcal{M}_n$ y luego aplicar la hipótesis inductiva, y la desigualdad de la varianza (2.2.1) que nos permite estimar el término que queremos. ■

Ahora damos la prueba para la monotonía [14]:

Teorema 3.1.3. *Sean $A_1, \dots, A_m$ y $A'_1, \dots, A'_m$ matrices definidas positivas tales que $A_j \leq A'_j$ para $1 \leq j \leq m$. Entonces*

$$G(A_1, \dots, A_m) \leq G(A'_1, \dots, A'_m). \quad (3.1.6)$$

Demostración. Denotemos por G y G' los dos lados de (3.1.6) y $\alpha, \alpha', \mathcal{M}, \mathcal{M}'$ las sumas correspondientes a las m -uplas y sus medias. Para todo $\varepsilon > 0$, existe n , tal que

$$\frac{\alpha}{n} < \frac{\varepsilon^2}{4}, \quad \frac{\alpha'}{n} < \frac{\varepsilon^2}{4}.$$

Sea $\mathcal{I}_n$ el subconjunto de $\mathcal{J}_n$, tal que $\delta_2(G, \mathcal{M}(j_1, \dots, j_n)) > \varepsilon$. Por (3.1.5) tenemos

$$\#\mathcal{I}_n \leq \frac{1}{4} \#\mathcal{J}_n.$$

Análogamente, sucede lo mismo para $\mathcal{I}'_n$. Esto prueba que la intersección de ambos conjuntos

$$\mathcal{C}_1 = \{(j_1, \dots, j_n) \in \mathcal{J}_n : \delta_2(G, \mathcal{M}(j_1, \dots, j_n)) < \varepsilon\},$$

y

$$\mathcal{C}_2 = \{(j_1, \dots, j_n) \in \mathcal{J}_n : \delta_2(G', \mathcal{M}'(j_1, \dots, j_n)) < \varepsilon\}$$

contiene al menos un elemento. De hecho, al menos contiene $\frac{m^n}{2}$ elementos.

Figura 3.1: Cada bloque contiene al menos $\frac{m^n}{4}$ elementos.

Sea $(j_1^*, \dots, j_m^*) \in \mathcal{C}_1 \cap \mathcal{C}_2$. Usando (3.1.3) vemos que

$$\mathcal{M}_n(j_1^*, \dots, j_m^*) \leq \mathcal{M}'_n(j_1^*, \dots, j_m^*).$$

Para $\varepsilon = \frac{1}{k}$, elegimos n y $(j_1^*, \dots, j_m^*)$ como describimos antes. Luego

$$\begin{aligned} \delta_2(G, \mathcal{M}_n(j_1^*, \dots, j_m^*)) &\leq \frac{1}{k} \\ \delta_2(G', \mathcal{M}'_n(j_1^*, \dots, j_m^*)) &\leq \frac{1}{k} \\ \mathcal{M}_n(j_1^*, \dots, j_m^*) &\leq \mathcal{M}'_n(j_1^*, \dots, j_m^*). \end{aligned}$$

De donde concluimos que $G \leq G'$ tomando límite. ■

Mientras que el enfoque probabilístico es muy interesante, también es verdad que la sucesión es desconocida. Eso llevó a Bhatia a preguntarse si una elección más natural, que es la sucesión que simplemente se repite de manera cíclica en $\{A_1, \dots, A_m\}$, converge a la media geométrica. La respuesta es afirmativa y fue demostrada por Holbrook. Más aún, la convergencia es independiente del punto de comienzo. Este último fenómeno es consecuencia de la curvatura del espacio como veremos más adelante (sección 3.5).

El resultado de Holbrook [29] es el siguiente: consideremos d matrices positivas $A_0, \dots, A_{d-1}$, y definamos la función $F : \mathbb{Z}_d \rightarrow \mathcal{P}(n)$ por $F(\bar{k}) = A_k$, donde $\mathbb{Z}_d$ denota el grupo abeliano de enteros módulo d . Si definimos la sucesión periódica $A = \{F(\bar{n})\}_{n \in \mathbb{N}}$, entonces Holbrook probó que

$$\lim_{n \rightarrow \infty} S_n(A) = \Lambda(A_0, \dots, A_{d-1}).$$

Ejemplo. Supongamos que tenemos 3 matrices estrictamente positivas A_1, A_2 y A_3 . Recordamos que en este caso particular:

$$\Lambda(A_1, A_2, A_3) = \operatorname{argmin}_{C \in \mathcal{P}(n)} \left(\delta^2(A_1, C) + \delta^2(A_2, C) + \delta^2(A_3, C) \right).$$

Luego, Holbrook probó que

$$A_1 \#_{\frac{1}{2}} A_2 \#_{\frac{1}{3}} A_3 \#_{\frac{1}{4}} A_1 \#_{\frac{1}{5}} A_2 \#_{\frac{1}{6}} A_3 \#_{\frac{1}{7}} A_1 \#_{\frac{1}{7}} \dots \rightarrow \Lambda(A_1, A_2, A_3)$$

Figura 3.2: Ejemplo de la convergencia de la media inductiva en el caso Euclídeo.

Fórmulas ergódicas para medias geométricas

Sea $(G, +)$ un grupo topológico compacto, dotado con una medida de Haar m , y sea $\tau : G \rightarrow G$ una transformación ergódica. El teorema ergódico clásico de Birkhoff nos dice que, dada $f \in L^1(G)$, entonces para m -casi todo $\omega \in G$

$$\frac{1}{n} \sum_{k=0}^{n-1} f(\tau^k(\omega)) \xrightarrow{n \rightarrow \infty} \int_G f(\omega) dm(\omega). \quad (3.1.7)$$

Luego, el Teorema de Holbrook puede ser visto como un Teorema ergódico para funciones definidas en $\mathbb{Z}_p$ que toman valores en $\mathcal{P}(n)$. En efecto, pensar que la media inductiva juega el papel de los promedios del lado izquierdo y la media geométrica $\Lambda(A_0, \dots, A_{d-1})$ juega el papel de la integral de f con respecto a la medida de Haar en $\mathbb{Z}_d$.

Esta interpretación del resultado de Holbrook nos sugiere que podemos obtener una aproximación computable de la media geométrica de varias matrices en término de promedios ergódicos. Uno de los objetivos de esta tesis es probar tal Teorema ergódico en término de la media inductiva en un contexto más general. Más precisamente, consideremos un sistema dinámico (G, τ) , donde G es un grupo compacto y τ es ergódico con respecto a la medida de Haar m . Observar que el grupo tiene que ser abeliano porque las órbitas $\{\tau^n g\}_{n \in \mathbb{N}}$ son densas en G .

En este contexto, dada una función $F : G \rightarrow \mathcal{P}(n)$, decimos que $F \in L^p(G, \mathcal{P}(n))$ si

$$\int_G \delta^p(F(g), B) dm(g) < \infty,$$

donde B es cualquier matriz positiva. Por la desigualdad triangular, la definición no depende de la elección de B . Siguiendo el trabajo de Sturm [56], el **baricentro** de $F \in L^1(G, \mathcal{P}(n))$ se define como

$$\beta_F := \operatorname{argmin}_{C \in \mathcal{P}(n)} \int_G \delta^2(F(g), C) - \delta^2(F(g), B) dm(g).$$

En otras palabras, β_F denota al baricentro del *pushforward measure* F_*m . En este caso es posible probar la existencia y unicidad de dicho elemento por las propiedades de continuidad y convexidad uniforme de la función,

$$C \mapsto \int_G \delta^2(F(g), C) - \delta^2(F(g), B) dm(g).$$

Notar que si $F \in L^2(G, \mathcal{P}(n))$ entonces

$$\beta_F := \operatorname{argmin}_{C \in \mathcal{P}(n)} \int_G \delta^2(F(g), C) dm(g).$$

Como antes, esta definición no depende de B . En esta tesis probamos el siguiente resultado:

Teorema. Dada $A \in L^1(G, \mathcal{P}(n))$, para casi todo $g \in G$

$$\lim_{n \rightarrow \infty} S_n(A(g), A(\tau(g)), \dots, A(\tau^{n-1}(g))) = \beta_A. \quad (3.1.8)$$

Más aún, probamos este resultado no sólo para funciones que toman valores en $\mathcal{P}(n)$, si no también, para los espacios de Hadamard M . Recordamos que un espacio métrico completo (M, δ) se llama **espacio de Hadamard** si la distancia satisface la ley del semiparalelogramo. Es decir, para todo $x, y \in M$ existe $m \in M$ tal que

$$\delta^2(m, z) \leq \frac{1}{2}\delta^2(x, z) + \frac{1}{2}\delta^2(y, z) - \frac{1}{4}\delta^2(x, y).$$

Existencia y unicidad de geodésicas: La desigualdad del semiparalelogramo implica que el punto medio es único. En efecto, tomando $z = x$ y $z = y$ en la desigualdad (2.1.1), es fácil ver que,

$$\delta(x, m) = \delta(m, y) = \frac{1}{2}\delta(x, y).$$

En efecto, tomando $z = x$, obtenemos

$$\delta(m, x) \leq \frac{1}{2}\delta(x, y).$$

Análogamente, tomando $z = y$, obtenemos

$$\delta(m, y) \leq \frac{1}{2}\delta(x, y).$$

Pero como

$$\delta(x, y) \leq \delta(x, m) + \delta(y, m) \leq \delta(x, y),$$

deducimos las igualdades antes mencionadas. Más aún, la desigualdad del semiparalelogramo implica que el punto medio es único. En efecto, si m, m' son puntos medios entre x e y ,

$$\begin{aligned}\delta^2(m, m') &\leq \frac{1}{2}\delta^2(x, m') + \frac{1}{2}\delta^2(y, m') - \frac{1}{4}\delta^2(x, y) \\ &\leq \frac{1}{2} \left[\frac{1}{2}\delta^2(x, x) + \frac{1}{2}\delta^2(y, x) - \frac{1}{4}\delta^2(x, y) \right] + \\ &\quad + \frac{1}{2} \left[\frac{1}{2}\delta^2(x, y) + \frac{1}{2}\delta^2(y, y) - \frac{1}{4}\delta^2(x, y) \right] - \frac{1}{4}\delta^2(x, y) = 0.\end{aligned}$$

La existencia y unicidad de puntos medios dan lugar a una **única geodésica** $\gamma_{a,b} : [0, 1] \rightarrow M$ que conecta dos puntos cualesquiera a y b . Para probarlo, primero se define $\gamma_{a,b}(1/2)$ como el punto medio entre a y b . Luego, usando un argumento inductivo, se define la geodésica para todos los números diádicos racionales en $[0, 1]$. Finalmente por completitud, se extiende para todo $t \in [0, 1]$. De ahora en adelante vamos a usar la notación $a \#_t b$ en vez de $\gamma_{a,b}(t)$.

Como en el caso de matrices estrictamente positivas, definimos la media inductiva en espacios de Hadamard de la siguiente manera:

Definición. (*Media inductiva*). Sea (M, δ) un espacio de Hadamard. Dado $a \in M^{\mathbb{N}}$ se define

$$\begin{aligned}S_1(a) &= a_1 \\ S_n(a) &= S_{n-1}(a) \#_{\frac{1}{n}} a_n \quad (n \geq 2).\end{aligned}$$

Tanto $L^1(G, M)$ como β_A , al igual que las medias inductivas, se definen mutatis mutandis. En este contexto el enunciado del teorema resulta ser:

Teorema. Dada $A \in L^1(G, M)$, para casi todo $g \in G$,

$$\lim_{n \rightarrow \infty} S_n(A(g), A(\tau(g)), \dots, A(\tau^{n-1}(g))) = \beta_A. \quad (3.1.9)$$

Con respecto a la media inductiva, la diferencia principal entre (3.1.8) y (3.1.9) es que para los espacios de Hadamard, en general, no existe una fórmula para la geodésica que une dos puntos. En otras palabras, al tratar el problema con más generalidad se puede perder la computabilidad de la media inductiva. Sin embargo, en muchos casos particulares de gran interés como el caso de las matrices positivas, operadores positivos en espacios de Hilbert, entre otros, existe una fórmula cerrada para las geodésicas que unen dos puntos.

Cabe mencionar que en [56], Sturm desarrolló una teoría de baricentros de medidas de probabilidad para espacios de Hadamard (ver Subsección 2.2 para algunas definiciones y resultados básicos). Dotado con este baricentro, los espacios de Hadamard juegan un papel importante en la teoría de integración (variables aleatorias, esperanza y varianza), ley de los grandes números, teoría ergódica, desigualdad de Jensen (ver [17], [27], [38], [48], y [56]), generalizaciones estocásticas de retracciones Lipschitz y problemas de extensión de funciones Lipschitz o Hölder (ver [39], [45], y [50]) y teoría de transporte óptimo en variedades Riemannianas (ver [52], y [53]), etc.

Mollifiers en espacios de Hadamard

La generalización de $\mathcal{P}(n)$, o incluso de variedades Riemannianas con curvatura no positiva, a espacios de Hadamard no es directa. Uno de los motivos es que los espacios de Hadamard, en general, no tienen un espacio vectorial finito dimensional subyacente, como en el caso de variedades. Esto hace que algunos pasos de las demostraciones sean mucho más complicados, y nos llevan a una nueva definición de mollifiers que sólo usan la estructura métrica (ver subsección 3.4.2). Estos resultados originales son interesantes por sí mismos y generalizan unos resultados probados por Karcher en [31].

3.2 Una variante del Teorema ergódico

Antes de encarar un problema general del tipo (3.1.8) pensemos nuevamente en el conjunto de matrices $\mathcal{P}(n)$ pero, esta vez, consideremos que las matrices $A_1, \dots, A_m$ son mutuamente conmutativas. Recordar que en este caso la media geométrica satisface $G(A_1, \dots, A_m) = (A_1 \dots A_m)^{1/m}$. Por otro lado, una consecuencia del Teorema ergódico de Birkhoff tomando $f = \chi_B$ es:

Proposición 3.2.1. *Sea*

$$\tau_B^{(m)} := \frac{\# \{0 \leq j \leq m-1 : T^j(x) \in B\}}{m},$$

donde $T : X \rightarrow X$ es ergódica. Entonces

$$\tau_B^{(m)} \xrightarrow{m \rightarrow \infty} \mu(B) \quad \text{casi todo } x \in X.$$

Esto nos dice que en el caso conmutativo es necesario y suficiente que, a medida que n crece, la cantidad de veces que aparece cada matriz tienda a ser la misma, para tener

$$S_n(A) \rightarrow \Lambda(A_1, \dots, A_m).$$

Una pregunta previa que podríamos plantearnos es si permutamos (infinitas) matrices en $S_n(A)$ del resultado de Holbrook podemos probar la convergencia a $\Lambda(A_1, \dots, A_m)$.

El hecho de poder permutar matrices en la demostración de Holbrook no es directa, pero hay otro resultado que nos puede ayudar en esta dirección, y en nuestro problema ergódico. Existe una versión general del Teorema de Holbrook en espacios de Hadamard donde los n puntos tienen asignados diferentes pesos. Para ser más precisos, sea (M, δ) un espacio métrico completo. El centro de masa $\Lambda(\omega; \mathbf{a})$ de la n -upla $\mathbf{a} = (a_1, \dots, a_n) \in M^n$ con pesos $\omega = (\omega_1, \dots, \omega_n) \in \Delta_n$ está definido como el único minimizador de la función

$$\Lambda(\omega; \mathbf{a}) = \operatorname{argmin}_{x \in M} \sum_{i=1}^n \omega_i \delta^2(x, a_i).$$

En [43], Lim Y Pálfi probaron el siguiente resultado.

Teorema 3.2.2. *Sea $\omega = (\omega_1, \dots, \omega_n) \in \Delta_n$ y $\mathbf{a} = (a_1, \dots, a_n) \in M^n$. Sea*

$$\begin{aligned} S_1(\mathbf{a}) &= a_1 \\ S_{k+1}(\mathbf{a}) &= S_k(\mathbf{a}) \#_{s_{k+1}} a_{k+1}, \end{aligned}$$

donde $s_k = \omega_{\bar{k}}/l(k)$ con $l(k) = \sum_{i=1}^k \omega_{\bar{i}}$. Entonces

$$\delta^2(\Lambda(\omega; \mathbf{a}), S_k(a)) \leq \frac{1}{l(k)} \left[3\Delta(\mathbf{a})^2 + \sum_{i=1}^n \omega_i \delta^2(\Lambda(\omega; \mathbf{a}), a_i) \right],$$

para todo $k \in \mathbb{N}$, es decir,

$$\lim_{k \rightarrow \infty} S_k(a) = \Lambda(\omega; \mathbf{a}),$$

ya que $l(k) \rightarrow \infty$ cuando $k \rightarrow \infty$.

Demostración. Ver [43] Teorema 3.4. ■

La prueba del resultado de Lim y Pálfi está basada en las desigualdades propias del espacio de Hadamard. Pero lo que nos interesa es que este nuevo resultado también nos dice que

$$\delta^2(\Lambda(\omega; \mathbf{a}), S_k(a)) \leq \frac{4\Delta(\mathbf{a})^2}{l(k)},$$

donde

$$\Delta(\mathbf{a}) := \sup_{i,j \in \{1, \dots, m\}} \delta(a_i, a_j).$$

Ahora tenemos una cota superior que depende solamente del diámetro del conjunto. Esto nos hace suponer que es posible modificar la demostración para poder realizar permutaciones. De hecho, a partir de modificaciones de sus resultados, obtenemos los primeros resultados en dirección al Teorema ergódico que buscamos (para el caso de funciones continuas).

3.2.1 La media inductiva

Como en el caso de matrices estrictamente positivas (que consideramos en la introducción), definimos la media inductiva en espacios de Hadamard de la siguiente manera:

Definición. (*Media inductiva*). Sea (M, δ) un espacio de Hadamard. Dado $a \in M^{\mathbb{N}}$ se define

$$\begin{aligned} S_1(a) &= a_1 \\ S_n(a) &= S_{n-1}(a) \#_{\frac{1}{n}} a_n \quad (n \geq 2). \end{aligned}$$

Observación. Dado $a \in M^n$ y $1 \leq k \leq n$, por medio de $S_k(a)$ vamos a denotar la media k -inductiva de cualquier sucesión cuyos primeros n términos son los elementos del vector a .

Ejemplo. Supongamos que M pertenece a $\mathbb{C}^n$ con la distancia Euclídea usual. Como las geodésicas en este caso son los segmentos, si tomamos una sucesión $\{a_n\}_{n \in \mathbb{N}}$ en $\mathbb{C}^n$, entonces

$$\begin{aligned} S_1(a) &= a_1, \\ S_2(a) &= \frac{a_1 + a_2}{2}, \\ S_3(a) &= \frac{2}{3} \left(\frac{a_1 + a_2}{2} \right) + \frac{1}{3} a_3 = \frac{a_1 + a_2 + a_3}{3}, \end{aligned}$$

y así sucesivamente. Luego, en este caso la media inductiva coincide con la media aritmética. ▲

De ahora en más, sea (M, δ) un espacio de Hadamard. Como una consecuencia de (2.1.3), obtenemos directamente este resultado de contractividad de la media inductiva.

Corolario 3.2.3. (*Propiedad de contracción*) *Dados $a, b \in M^{\mathbb{N}}$, tenemos que*

$$\delta(S_n(a), S_n(b)) \leq \frac{1}{n} \sum_{i=1}^n \delta(a_i, b_i). \quad (3.2.1)$$

El siguiente lema se deduce de (2.1.2), y es un caso particular de la desigualdad con pesos considerada por Lim y Pálfi en [43].

Lema 3.2.4. *Dados $a \in M^{\mathbb{N}}$ y $z \in M$, para todo $k, m \in \mathbb{N}$*

$$\begin{aligned} \delta^2(S_{k+m}(a), z) &\leq \frac{k}{k+m} \delta^2(S_k(a), z) + \frac{1}{k+m} \sum_{j=0}^{m-1} \delta^2(a_{k+j+1}, z) \\ &\quad - \frac{k}{(k+m)^2} \sum_{j=0}^{m-1} \delta^2(S_{k+j}(a), a_{k+j+1}). \end{aligned}$$

Demostración. Por la desigualdad (2.1.2), aplicada a $S_{n+1}(a) = S_n(a) \#_{\frac{1}{n+1}} (a_{n+1})$, obtenemos

$$(n+1) \delta^2(S_{n+1}(a), z) - n \delta^2(S_n(a), z) \leq \delta^2(a_{n+1}, z) - \frac{n}{(n+1)} \delta^2(S_n(a), a_{n+1}).$$

Sumando estas desigualdades de $n = k$, a $n = k + m - 1$, obtenemos que la diferencia

$$(k+m) \delta^2(S_{k+m}(a), z) - k \delta^2(S_k(a), z),$$

obtenida de la suma telescópica del lado izquierdo, es menor o igual que

$$\sum_{j=0}^{m-1} \left(\delta^2(a_{k+j+1}, z) - \frac{k+j}{(k+j+1)} \delta^2(S_{k+j}(a), a_{k+j+1}) \right).$$

Finalmente, Usando que $\frac{k+j}{k+j+1} \geq \frac{k}{k+m}$ para todo $j \in \{0, \dots, m-1\}$, esta suma está acotada por arriba por

$$\sum_{j=0}^{m-1} \left(\delta^2(a_{k+j+1}, z) - \frac{k}{(k+m)} \delta^2(S_{k+j}(a), a_{k+j+1}) \right),$$

lo que completa la demostración. ■

Dada una sucesión $a \in M^{\mathbb{N}}$, sea $\Delta(a)$ el diámetro de su imagen, es decir

$$\Delta(a) := \sup_{n, m \in \mathbb{N}} \delta(a_n, a_m).$$

Notar que, también por (2.1.2), $\delta(S_n(a), a_k) \leq \Delta(a)$ para todo $n, k \in \mathbb{N}$.

Lema 3.2.5. Dado $a \in M^{\mathbb{N}}$ tal que $\Delta(a) < \infty$, entonces para todo $k, m \in \mathbb{N}$ se tiene que

$$\frac{1}{m} \sum_{j=0}^{m-1} \delta^2(S_k(a), a_{k+j+1}) \leq \tilde{R}_{m,k} + \frac{1}{m} \sum_{j=0}^{m-1} \delta^2(S_{k+j}(a), a_{k+j+1}).$$

$$\text{donde } \tilde{R}_{m,k} = \left(\frac{m^2}{(k+1)^2} + 2 \frac{m}{k+1} \right) \Delta^2(a).$$

Demostración. Notar que por (3.2.1), y para todo k ,

$$\delta(S_{k+j}(a), S_{k+j+1}(a)) \leq \frac{1}{k+j+1} \Delta(a).$$

Por lo tanto

$$\begin{aligned} \delta(S_k(a), a_{k+j+1}) &\leq \delta(S_k(a), S_{k+j}(a)) + \delta(S_{k+j}(a), a_{k+j+1}) \\ &\leq \sum_{h=1}^j \frac{1}{k+h} \Delta(a) + \delta(S_{k+j}(a), a_{k+j+1}) \\ &\leq \frac{m}{k+1} \Delta(a) + \delta(S_{k+j}(a), a_{k+j+1}). \end{aligned}$$

Luego, para todo $j \leq m$,

$$\delta^2(S_k(a), a_{k+j+1}) \leq \left(\frac{m^2}{(k+1)^2} + 2 \frac{m}{k+1} \right) \Delta^2(a) + \delta^2(S_{k+j}(a), a_{k+j+1}),$$

donde usamos que $\delta(S_{k+j}(a), a_{k+j+1}) \leq \Delta(a)$ para todo $k, j \in \mathbb{N}$. Sumando estas desigualdades y dividiendo por m , obtenemos el resultado. ■

3.3 Fórmula ergódica asociada a la media inductiva en el caso continuo

3.3.1 El marco de trabajo y notación básica

Sea $(G, +)$ un grupo topológico metrizable, compacto y abeliano. En este grupo fijamos una medida de Haar m , y tomamos un automorfismo ergódico $\tau(g) = g + h_0$ para algún $h_0 \in G$. A una métrica invariante por **desplazamientos** (shift) del grupo G la denotamos por d_G (para la existencia de tal métrica ver [47], sección 1.22). En esta sección vamos a trabajar con un sistema dinámico (G, τ) .

Observación. Todo sistema de grupo dinámico (G, τ) como el descrito antes es equivalente a un **sistema de Kronecker** (ver Teorema 2.5.8). Es conocido que todo sistema dinámico isométrico (o equicontinuo) contiene un sistema de Kronecker, pues todo sistema dinámico minimal es un sistema Kronecker (Ver Proposición 2.5.3 y Proposición 2.5.4).

Por otro lado, vamos a fijar un espacio de Hadamard (M, δ) . Dada una función $A : G \rightarrow M$, definimos $a^\tau : G \rightarrow M^\mathbb{N}$, por

$$a^\tau(x) := \{a_n^\tau(x)\}_{n \in \mathbb{N}} \quad \text{donde} \quad a_j^\tau(x) = A(\tau^j(x)). \quad (3.3.1)$$

Finalmente, dada una función $F \in L^1(G, M)$, por medio de β_F vamos a denotar el baricentro del *pushforward measure* F_*m . En otras palabras, dado cualquier punto $y \in M$,

$$\beta_F = \operatorname{argmin}_{z \in M} \int_G [\delta^2(z, F(g)) - \delta^2(y, F(g))] dm(g).$$

Recordar que si $F \in L^2(G, M)$ entonces,

$$\beta_F = \operatorname{argmin}_{z \in M} \int_G \delta^2(z, F(g)) dm(g).$$

Luego, si $F \in L^2(G, M)$, entonces β_F coincide con el baricentro de Cartan correspondiente a la distribución F .

3.3.2 Demostración del caso continuo

En esta sección vamos a probar la fórmula ergódica para funciones continuas. Más precisamente vamos a probar el siguiente resultado:

Teorema 3.3.1. *Sean (M, δ) un espacio de Hadamard, $(G, +)$ un grupo topológico metrizable, compacto y abeliano, y $A : G \rightarrow M$ una función continua. Entonces*

$$\lim_{n \rightarrow \infty} S_n(a^\tau(g)) = \beta_A, \quad (3.3.2)$$

uniformemente en $g \in G$.

Con este objetivo, primero vamos a probar algunos resultados técnicos, que vamos a combinar al final de la sección para probar el Teorema 3.3.1.

Lema 3.3.2. *Sea $A : G \rightarrow M$ una función continua, y sea K un subconjunto compacto de M . Para todo $n \in \mathbb{N}$, definimos $F_n : G \times K \rightarrow \mathbb{R}$ por*

$$F_n(g, x) = \frac{1}{n} \sum_{j=0}^{n-1} \delta^2(a_j^\tau(g), x).$$

Entonces, la familia $\{F_n\}_{n \in \mathbb{N}}$ es equicontinua.

Demostración. Por la desigualdad triangular, la función $y \mapsto \delta^2(A(\cdot), y)$ es continua de (K, δ) , en el conjunto de funciones continuas reales definidas en G considerando la norma uniforme. Como K es compacto, la familia $\{\delta^2(A(\cdot), x)\}_{x \in K}$ es (uniformemente) equicontinua. Por lo tanto, dado $\varepsilon > 0$, existe $\delta > 0$ tal que si

$$d_G(g_1, g_2) < \delta \quad \text{entonces} \quad |\delta^2(A(g_1), x) - \delta^2(A(g_2), x)| < \frac{\varepsilon}{2},$$

para todo $x \in K$. Como τ es una isometría y $d_G(g_1, g_2) < \delta$, obtenemos

$$|F_n(g_1, x) - F_n(g_2, x)| = \left| \frac{1}{n} \sum_{j=0}^{n-1} \delta^2(a_j^\tau(g_1), x) - \delta^2(a_j^\tau(g_2), x) \right| < \frac{\varepsilon}{2}.$$

Sea Δ el diámetro del conjunto $(\text{Imagen}(A) \times K)$ en M^2 . Como ambos conjuntos son compactos, $\Delta < \infty$. Entonces, tomemos (g_1, x_1) y (g_2, x_2) tales que $d_G(g_1, g_2) < \delta$ y $\delta(x_1, x_2) < \frac{\varepsilon}{4\Delta}$. Luego

$$\begin{aligned} |F_n(g_1, x_1) - F_n(g_2, x_2)| &\leq |F_n(g_1, x_1) - F_n(g_1, x_2)| + |F_n(g_1, x_2) - F_n(g_2, x_2)| \\ &\leq \frac{2\Delta}{n} \sum_{k=0}^{n-1} \delta(x_1, x_2) + \frac{\varepsilon}{2} < \varepsilon. \end{aligned}$$

■

Ahora, como consecuencia de los Teoremas de Arzelà-Ascoli y Birkhoff obtenemos.

Proposición 3.3.3. *Sea $A : G \rightarrow M$ una función continua, y K un conjunto compacto de M . Entonces*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \delta^2(a_j^\tau(g), x) = \int_G \delta^2(A(\gamma), x) dm(\gamma),$$

donde la convergencia es uniforme en $(g, x) \in G \times K$.

Convención: De ahora en adelante, hasta finalizar la subsección, vamos a fijar una función continua $A : G \rightarrow M$.

Sea

$$\alpha := \min_{x \in M} \int_G \delta^2(A(g), x) dm(g),$$

y β_A el punto donde se alcanza el mínimo, es decir, β_A es el baricentro del *pushforward* por A de la medida de Haar en G . Entonces obtenemos la siguiente cota superior.

Lema 3.3.4. *Para todo $\varepsilon > 0$, existe $m_0 \in \mathbb{N}$ tal que, para todo $m \geq m_0$ y todo $k \in \mathbb{N}$,*

$$\begin{aligned} \delta^2(S_{k+m}(a^\tau(g)), \beta_A) &\leq \frac{k}{k+m} \delta^2(S_k(a^\tau(g)), \beta_A) + \frac{m}{k+m} (\alpha + \varepsilon) \\ &\quad - \frac{km}{(k+m)^2} \left(\frac{1}{m} \sum_{j=0}^{m-1} \delta^2(S_{k+j}(a^\tau(g)), a_{k+j+1}^\tau(g)) \right). \end{aligned}$$

Demostración. Para todo $\varepsilon > 0$, existe $m_0 \in \mathbb{N}$ tal que, para todo $m \geq m_0$,

$$\left| \frac{1}{m} \sum_{j=0}^{m-1} \delta^2(a_{k+j+1}^\tau(g), \beta_A) - \alpha \right| < \varepsilon.$$

Notar que m_0 es independiente de k por la Proposición 3.3.3. Ahora, por el Lema 3.2.4

$$\begin{aligned}
\delta^2(S_{k+m}(a^\tau(g)), \beta_A) &\leq \frac{k}{k+m} \delta^2(S_k(a^\tau(g)), \beta_A) + \frac{1}{k+m} \sum_{j=0}^{m-1} \delta^2(a_{k+j+1}^\tau(g), \beta_A) \\
&\quad - \frac{k}{(k+m)^2} \sum_{j=0}^{m-1} \delta^2(S_{k+j}(a^\tau(g)), a_{k+j+1}^\tau(g)) \\
&= \frac{k}{k+m} \delta^2(S_k(a^\tau(g)), \beta_A) + \frac{m}{k+m} \left(\frac{1}{m} \sum_{j=0}^{m-1} \delta^2(a_{k+j+1}^\tau(g), \beta_A) \right) \\
&\quad - \frac{km}{(k+m)^2} \left(\frac{1}{m} \sum_{j=0}^{m-1} \delta^2(S_{k+j}(a^\tau(g)), a_{k+j+1}^\tau(g)) \right) \\
&\leq \frac{k}{k+m} \delta^2(S_k(a^\tau(g)), \beta_A) + \frac{m}{k+m} (\alpha + \varepsilon) \\
&\quad - \frac{km}{(k+m)^2} \left(\frac{1}{m} \sum_{j=0}^{m-1} \delta^2(S_{k+j}(a^\tau(g)), a_{k+j+1}^\tau(g)) \right).
\end{aligned}$$

■

Recordar que, dada una sucesión $a \in M^{\mathbb{N}}$, $\Delta(a)$ denota el diámetro de su imagen. Es decir,

$$\Delta(a) := \sup_{n,m \in \mathbb{N}} \delta(a_n, a_m).$$

Como $A : G \rightarrow M$ es continua, notar que

$$C_a := \sup_{g \in G} \Delta(a^\tau(g)) < \infty.$$

Lema 3.3.5. *Para todo $\varepsilon > 0$, existe $m_0 \in \mathbb{N}$ tal que, para todo $m \geq m_0$ y todo $k \in \mathbb{N}$,*

$$\delta^2(S_k(a^\tau(g)), \beta_A) - \varepsilon + \alpha - R_{m,k} \leq \frac{1}{m} \sum_{j=0}^{m-1} \delta^2(S_{k+j}(a^\tau(g)), a_{k+j+1}^\tau(g)),$$

$$\text{donde } R_{m,k} = \left(\frac{m^2}{(k+1)^2} + 2 \frac{m}{k+1} \right) C_a^2.$$

Demostración. Considerar el conjunto compacto

$$K := \overline{cc \{S_k(a^\tau(x)) : k \in \mathbb{N}\}},$$

donde la cápsula convexa se toma en el sentido geodésico. Para todo $\varepsilon > 0$, existe $m_0 \in \mathbb{N}$ tal que, para todo $m \geq m_0$, por la desigualdad de la varianza (Proposición 2.2.2) y la Proposición 3.3.3, se tiene que

$$\begin{aligned}
\delta^2(S_k(a^\tau(x)), \beta_A) &\leq \int_G \delta^2(S_k(a^\tau(g)), A(\gamma)) dm(\gamma) - \alpha \\
&\leq \varepsilon + \frac{1}{m} \sum_{j=0}^{m-1} \delta^2(S_{k+j}(a^\tau(g)), a_{k+j+1}^\tau(g)) - \alpha.
\end{aligned}$$

Finalmente, por el Lema 3.2.5

$$\delta^2(S_k(a^\tau(x)), \beta_A) \leq \varepsilon + \frac{1}{m} \sum_{j=0}^{m-1} \delta^2(S_{k+j}(a^\tau(x)), a_{k+j+1}^\tau(x)) + R_{m,k} - \alpha,$$

$$\text{donde } R_{m,k} = \left(\frac{m^2}{(k+1)^2} + 2 \frac{m}{k+1} \right) C_a^2. \quad \blacksquare$$

Lema 3.3.6. Dado $\varepsilon > 0$, existe $m_0 \geq 1$ tal que, para todo $\ell \in \mathbb{N}$

$$\delta^2(S_{\ell m_0}(a^\tau(g)), \beta_A) \leq \frac{L}{\ell} + \varepsilon,$$

uniformemente en $g \in G$, donde $L = \alpha + 3C_a^2$.

Demostración. Fijemos $\varepsilon > 0$. Por los Lemas 3.3.4 y 3.3.5, existe $m_0 \geq 1$ tal que, para todo $k \in \mathbb{N}$,

$$\begin{aligned} \delta^2(S_{k+m_0}(a^\tau(g)), \beta_A) &\leq \frac{k}{k+m_0} \delta^2(S_k(a^\tau(g)), \beta_A) + \frac{m_0}{k+m_0} (\alpha + \varepsilon) \\ &\quad - \frac{km_0}{(k+m_0)^2} \left(\frac{1}{m_0} \sum_{j=0}^{m_0-1} \delta^2(S_{k+j}(a^\tau(x)), a_{k+j+1}^\tau(x)) \right), \end{aligned}$$

y

$$\frac{1}{m_0} \sum_{j=0}^{m_0-1} \delta^2(S_{k+j}(a^\tau(g)), a_{k+j+1}^\tau(g)) \geq \delta^2(S_k(a^\tau(g)), \beta_A) - \varepsilon + \alpha - R_{m_0,k}.$$

Luego, combinando estas dos desigualdades obtenemos

$$\begin{aligned} \delta^2(S_{k+m_0}(a^\tau(g)), \beta_A) &\leq \frac{k}{k+m_0} \delta^2(S_k(a^\tau(g)), \beta_A) + \frac{m_0}{k+m_0} (\alpha + \varepsilon) \\ &\quad - \frac{km_0}{(k+m_0)^2} (\delta^2(S_k(a^\tau(g)), \beta_A) - \varepsilon + \alpha - R_{m_0,k}). \end{aligned}$$

Consideremos ahora el caso particular cuando $k = \ell m_0$. Como $R_{m_0, \ell m_0} \leq \frac{3}{\ell} C_a^2$ tenemos

$$\begin{aligned} \delta^2(S_{(\ell+1)m_0}(a^\tau(g)), \beta_A) &\leq \frac{\ell}{\ell+1} \delta^2(S_{\ell m_0}(a^\tau(g)), \beta_A) + \frac{1}{\ell+1} (\alpha + \varepsilon) - \\ &\quad \frac{\ell}{(\ell+1)^2} (\delta^2(S_{\ell m_0}(a^\tau(g)), \beta_A) - \varepsilon + \alpha - R_{m_0, \ell m_0}) \\ &\leq \frac{\ell^2 \delta^2(S_{\ell m_0}(a^\tau(g)), \beta_A) + (2\ell+1)\varepsilon + \alpha + 3C_a^2}{(\ell+1)^2}. \end{aligned} \quad (3.3.3)$$

Usando la desigualdad recursivamente, el resultado se obtiene por inducción en ℓ . De hecho, si $\ell = 1$ entonces

$$\delta^2(S_{m_0}(a^\tau(g)), \beta_A) \leq C_a^2 \leq L.$$

Por otro lado, si suponemos que el resultado es verdadero para algún $\ell \geq 1$, es decir,

$$\delta^2(S_{\ell m_0}(a^\tau(x)), g) \leq \frac{L}{\ell} + \varepsilon.$$

Combinando esta desigualdad con (3.3.3) tenemos que

$$\delta^2(S_{(\ell+1)m_0}(a^\tau(g)), \beta_A) \leq \frac{\ell L + \ell^2 \varepsilon + (2\ell + 1)\varepsilon + \alpha + 3C_a^2}{(\ell + 1)^2} = \frac{L}{\ell + 1} + \varepsilon.$$

■

Ahora estamos listos para probar la fórmula ergódica para funciones continuas.

Demostración del Teorema. 3.3.1. Dado $\varepsilon > 0$, por el Lema 3.3.6, existe $m_0 \in \mathbb{N}$ tal que,

$$\delta^2(S_{\ell m_0}(a^\tau(g)), \beta_A) \leq \frac{L}{\ell} + \frac{\varepsilon^2}{8},$$

para todo $\ell \in \mathbb{N}$. Tomemos $\ell_0 \in \mathbb{N}$ tal que, para todo $\ell \geq \ell_0$,

$$\delta^2(S_{\ell m_0}(a^\tau(g)), \beta_A) \leq \frac{\varepsilon^2}{4}. \quad (3.3.4)$$

Sea $n = \ell m_0 + d$ tal que $\ell \geq \ell_0$ y $d \in \{1, \dots, m_0 - 1\}$. Notar que

$$S_{\ell m_0}(a^\tau(g)) = S_{\ell m_0 + d}((a_1^\tau(g), \dots, a_{\ell m_0}^\tau(g), \underbrace{S_{\ell m_0}(a^\tau(g)), \dots, S_{\ell m_0}(a^\tau(g))}_{d \text{ veces} })).$$

Entonces, usando el Corolario 3.2.3 con las sucesiones

$$(a_1^\tau(g), \dots, a_{\ell m_0}^\tau(g), \underbrace{S_{\ell m_0}(a^\tau(g)), \dots, S_{\ell m_0}(a^\tau(g))}_{d \text{ veces} })$$

y

$$(a_1^\tau(g), \dots, a_{\ell m_0}^\tau(g), a_{\ell m_0 + 1}^\tau(g), \dots, a_{\ell m_0 + d}^\tau(g)),$$

obtenemos

$$\delta(S_{\ell m_0}(a^\tau(g)), S_{\ell m_0 + d}(a^\tau(g))) \leq \frac{1}{\ell m_0 + d} \sum_{j=1}^d \delta(S_{\ell m_0}(a^\tau(g)), a_{\ell m_0 + j}^\tau(g)).$$

Como $\delta(S_{\ell m_0}(a^\tau(g)), a_{\ell m_0 + j}^\tau(g)) \leq C_a$ para todo $j \in \{1, \dots, m_0 - 1\}$, finalmente tenemos que

$$\delta(S_{\ell m_0}(a^\tau(g)), S_{\ell m_0 + d}(a^\tau(g))) \leq \frac{d}{\ell m_0 + d} C_a \leq \frac{1}{\ell} C_a \xrightarrow{k \rightarrow \infty} 0.$$

Combinando esto con (3.3.4) obtenemos que para un n suficientemente grande,

$$\delta(S_n(a^\tau(g)), \beta_A) < \varepsilon.$$

■

3.4 Fórmula ergódica asociada a la media inductiva en el caso L^1

Como mencionamos antes, queremos extender el Teorema 3.3.1 a funciones en $L^1(G, M)$. En esta sección vamos a probar el siguiente Teorema ergódico para funciones en $L^1(G, M)$, en término de la media inductiva, el cual es uno de los resultados principales de esta tesis.

Teorema 3.4.1. *Sean (M, δ) un espacio de Hadamard y $(G, +)$ un grupo topológico metrizable, compacto y abeliano. Dada $A \in L^1(G, M)$, para casi todo $g \in G$*

$$\lim_{n \rightarrow \infty} S_n(a^\tau(g)) = \beta_A. \quad (3.4.1)$$

Observación. Sea M el cuerpo de números complejos con la distancia usual. Como observamos en el ejemplo de la sección 3.2.1, la media inductiva $S_n(a^\tau(g))$ se convierte en la media aritmética usual

$$\frac{1}{n} \sum_{k=0}^{n-1} A(\tau^k(g)).$$

Por otro lado, el baricentro β_A es la integral de A con respecto a la medida de Haar m . Luego, la ecuación (3.4.1) se traduce en

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} A(\tau^k(g)) = \int_G A(g) dm(g),$$

en casi todo punto $g \in \mathbb{C}$, el cual es el Teorema clásico de Birkhoff. ▲

Es natural pensar entonces, en aproximar una función de $L^1(G, M)$, por funciones continuas. Para ello necesitamos definir *mollifiers* en espacios de Hadamard.

3.4.1 Dos lemas de aproximación

La estrategia de la demostración consiste en construir una buena aproximación por funciones continuas, y obtener el resultado en L^1 como consecuencia del teorema para funciones continuas (Teorema 3.3.1). Las preguntas que surgen son: ¿Qué significa una buena aproximación? ¿Qué debemos requerir a la aproximación para obtener el caso L^1 como límite del caso continuo? Los siguientes lemas contienen las claves para responder estas dos preguntas.

Lema 3.4.2. *Sea $(\Omega, \mathcal{B}, P)$ un espacio de probabilidad, y $A, B \in L^1(\Omega, M)$. Si*

$$\begin{aligned} \beta_A &= \operatorname{argmin}_{z \in M} \int_{\Omega} [\delta^2(A(\omega), z) - \delta^2(A(\omega), y)] dP(\omega), \\ \beta_B &= \operatorname{argmin}_{z \in M} \int_{\Omega} [\delta^2(B(\omega), z) - \delta^2(B(\omega), y)] dP(\omega), \end{aligned}$$

entonces

$$\delta(\beta_A, \beta_B) \leq \int_{\Omega} \delta(A(\omega), B(\omega)) dP(\omega). \quad (3.4.2)$$

Observación. Recordar que la definición de β_A (resp. β_B) no depende de la elección de $y \in M$.

Demostración. Por la desigualdad de la varianza (Proposición 2.2.2) obtenemos

$$\begin{aligned}\delta^2(\beta_A, \beta_B) &\leq \int_{\Omega} \delta^2(\beta_A, B(\omega)) - \delta^2(\beta_B, B(\omega)) dP(\omega), \\ \delta^2(\beta_A, \beta_B) &\leq \int_{\Omega} \delta^2(\beta_B, A(\omega)) - \delta^2(\beta_A, A(\omega)) dP(\omega),\end{aligned}$$

y la combinación de estas dos desigualdades nos llevan a

$$\begin{aligned}2\delta^2(\beta_A, \beta_B) &\leq \int_{\Omega} \delta^2(\beta_A, B(\omega)) + \delta^2(\beta_B, A(\omega)) \\ &\quad - \delta^2(\beta_B, B(\omega)) - \delta^2(\beta_A, A(\omega)) dP(\omega).\end{aligned}$$

Finalmente, usando el Teorema de cuádruple comparación de Reshetnyak (Teorema 2.1.3) obtenemos

$$2\delta^2(\beta_A, \beta_B) \leq 2\delta(\beta_A, \beta_B) \int_{\Omega} \delta(A(\omega), B(\omega)) dP(\omega),$$

lo cual es, después de una simplificación algebraica, el resultado deseado. ■

Lema 3.4.3. Sean $A, B \in L^1(G, M)$. Dado $\varepsilon > 0$, para casi todo $g \in G$ existe n_0 , que puede depender de g , tal que

$$\delta(S_n(a^\tau(g)), S_n(b^\tau(g))) \leq \varepsilon + \int_G \delta(A(g), B(g)) dm(g), \quad (3.4.3)$$

siempre que $n \geq n_0$.

Demostración. Por el Corolario 3.2.3

$$\delta(S_n(a^\tau(g)), S_n(b^\tau(g))) \leq \frac{1}{n} \sum_{k=0}^{n-1} \delta(a_k^\tau(g), b_k^\tau(g)) = \frac{1}{n} \sum_{k=0}^{n-1} \delta(A(\tau^k(g)), B(\tau^k(g))),$$

por lo tanto, el lema se deduce del Teorema ergódico de Birkhoff. ■

3.4.2 Buena aproximación por funciones continuas (Mollifiers)

Los dos lemas anteriores nos dan indicio de que necesitamos algún tipo de buena aproximación en L^1 . Más precisamente, dada $A \in L^1(G, M)$ y $\varepsilon > 0$, buscamos una función continua $A_\varepsilon : G \rightarrow M$ tal que

$$\int_G \delta(A(g), A_\varepsilon(g)) dm(g) < \varepsilon.$$

En algunos casos existe un espacio vectorial finito dimensional subyacente. Este es el caso, por ejemplo, cuando M es el conjunto de matrices estrictamente positivas, o más general, cuando M es una variedad Riemanniana de curvatura no positiva. En este caso las funciones A_ε se pueden construir usando *Mollifiers*. Esta idea fue usada por Karcher en [31]. En el caso general, podemos usar una idea similar.

Mollifiers en variedades Riemannianas

En esta subsección veremos como es posible definir los Mollifiers para variedades Riemannianas. Dicha idea fue utilizada por Karcher en [31]. Recordemos primero como se definen los mollifiers en $\mathbb{R}^n$. Las siguientes definiciones y resultados son clásicos, cuyas demostraciones se encuentran en varios libros, por lo cual, no incluiremos las mismas. Solamente las mencionamos, para hacer una analogía con los mollifiers en variedades Riemannianas.

Definición. Sea $\eta \in C^\infty(\mathbb{R}^n)$ definida por,

$$\eta(x) := \begin{cases} C e^{\frac{1}{|x|^2-1}} & \text{si } |x| < 1 \\ 0 & \text{si } |x| \geq 1 \end{cases}$$

donde la constante C se elige de modo que,

$$\int_{\mathbb{R}^n} \eta dx = 1.$$

Figura 3.3: Ejemplo de una función $\eta \in C^\infty(\mathbb{R})$.

Para todo $\varepsilon > 0$, sea

$$\eta_\varepsilon(x) = \frac{1}{\varepsilon^n} \eta\left(\frac{x}{\varepsilon}\right).$$

Se suele llamar a η el *mollifier* estándar. Las funciones η_ε cumplen:

1. $\eta_\varepsilon \in C^\infty$,
2. $\int_{\mathbb{R}^n} \eta_\varepsilon(x) dx = 1$,
3. $\text{spt}(\eta_\varepsilon) \subset B(0, \varepsilon)$.

Esta función se utiliza para suavizar a una función f dada.

Definición. Si $f : U \rightarrow \mathbb{R}$ es localmente integrable, definimos su Mollificación o aproximación de la identidad por

$$f^\varepsilon(x) := \eta_\varepsilon * f(x),$$

donde $x \in U_\varepsilon = \{x \in U : \text{dist}(x, \partial U) > \varepsilon\}$.

Teorema 3.4.4. (*Propiedades de los Mollifiers*)

1. $f^\varepsilon \in C^\infty(U_\varepsilon)$.
2. $f^\varepsilon \rightarrow f$ cuando $\varepsilon \rightarrow 0$ ctp.
3. Si $f \in C(U)$, entonces $f^\varepsilon \rightarrow f$ uniformemente en subconjuntos compactos de U .
4. Si $1 \leq p < \infty$ y $f \in L^p_{loc}$, entonces $f^\varepsilon \rightarrow f$ en L^p_{loc} .

Ahora si pasamos al contexto de variedades Riemannianas. Sea $\varphi : \mathbb{R} \rightarrow [0, 1]$ una función C^∞ tal que

$$\begin{aligned}\varphi(r) &= 1 & \text{para } r \leq \varepsilon; \\ \varphi(r) &= 0 & \text{para } r \geq 1 - \varepsilon; \\ \varphi'(r) &\leq 0.\end{aligned}$$

Figura 3.4: Ejemplo de dicha función.

Sea (M, δ) una variedad Riemanniana e identificamos la bola centrada en $m \in M$, de radio ρ , $B_\rho(m) \subset M$ vía $\exp_m^{-1}$ con la bola Euclídea $D_\rho(0) \subset T_m M$. Por cuestiones de conveniencia se usa, en la definición de mollifier en variedades, el volumen Euclídeo $d_m x$ (porque las cancelaciones son más obvias que usando el volumen Riemanniano no lineal $B_\rho(m)$).

Notación básica: Sean $M, \hat{M}$ dos variedades Riemannianas tales que un centro de masa puede ser definido en las bolas $B_{\hat{\rho}} \subset \hat{M}$. Supongamos también que $f : M \rightarrow \hat{M}$ es una función medible y existe un número ρ_0 tal que, para toda bola $B_{\rho_0} \subset M$, existe una bola $B_{\hat{\rho}} \subset \hat{M}$ con $f(B_{\rho_0}) \subset B_{\hat{\rho}}$.

Definición. Definimos la función $\Phi : M \times M \rightarrow \mathbb{R}$ por

$$\Phi_\rho(m, y) = \varphi\left(\frac{1}{\rho}d(m, y)\right) \left(\int_{B_\rho(m)} \varphi\left(\frac{1}{\rho}d(m, y)\right) d_m x \right)^{-1}.$$

Es necesario definir el siguiente espacio de probabilidad para suavizar a f en m :

$$A_{\rho, m} = (M, \Phi_\rho(m, y) d_m x).$$

Como fuera de la bola $B_\rho(m)$ la medida $\Phi_\rho(m, y) d_m x$ es cero, podemos considerar la función $f : M \rightarrow \hat{M}$ como una función medible

$$f_{\rho, m} = f : A_{\rho, m} \rightarrow B_{\hat{\rho}}.$$

Hasta ahora las definiciones son obvias, los problemas son causados por el espacio de llegada $\hat{M}$ el cual no es lineal. El centro de masa hace posible la siguiente generalización de los *mollifier* estándar:

Definición. Dado $\rho > 0$, sea $f_\rho : M \rightarrow \hat{M}$, definida por

$$f_\rho(m) = \operatorname{argmin}_{\hat{m} \in B_{\hat{\rho}}} \frac{1}{2} \int_{A_{\rho,m}} \delta(\hat{m}, f(a))^2 da.$$

Notar que el mínimo siempre existe y es único por las propiedades de convexidad de la función, siempre que trabajemos dentro de bolas convexas. Dentro de tales bolas, es necesario elegir un subconjunto compacto y una función de distribución de masa total 1. Razón por la cual, es necesario definir el espacio de probabilidad $A_{\rho,m}$. Este espacio de probabilidad nos otorga la función de distribución da de masa total 1 en $A_{\rho,m}$, que nos garantiza la existencia y unicidad del minimizador (Ver apéndice sección 6.4). Por otro lado los mollifiers son funciones suaves y deben aproximar a nuestra función original. De esto trata el siguiente resultado.

Teorema 3.4.5.

1. La función $f_\rho : M \rightarrow \hat{M}$ es diferenciable, y si M y $\hat{M}$ son espacios C^∞ entonces f_ρ es una función C^∞ .
2. Si f es continua en m entonces

$$\lim_{\rho \rightarrow 0} f_\rho(m) = f(m).$$

3. Si f es C^1 entonces

$$\lim_{\rho \rightarrow 0} df_\rho = df.$$

Por último concluimos esta subsección con un resultado de continuidad. Una pregunta natural es: dadas dos funciones, si estas se aproximan entre sí ¿También lo hacen sus centros de masa? Un resultado de este estilo nos sería de utilidad y también fue anticipado por Karcher en su trabajo.

Proposición 3.4.6. Si $f_1, f_2 : A \rightarrow B_\rho$ son distribuciones de masa y $\delta \leq K \leq \Delta$ son cotas de la curvatura, entonces

$$d(C_{f_1}, C_{f_2}) \leq (1 + \operatorname{const.}(\delta, \Delta)(2\rho)^2) \int_A d(f_1(a), f_2(a)) da.$$

Demostración. Ver [31], Corolario 1.6, pág. 512. ■

Observar que ya dimos una demostración más sencilla y general (para espacio de Hadamard) de esta proposición en la parte anterior (ver Lema 3.4.2).

Mollifiers en espacios de Hadamard

Para el caso general podemos usar la siguiente idea: Dado $\eta > 0$, sea U_η un entorno de la identidad de G tal que $m(U_\eta) < \eta$, cuyo diámetro es menor que η . Fijemos $y \in M$, y definimos

$$A_\eta(g_0) = \operatorname{argmin}_{z \in M} \int_{U_\eta} [\delta^2(z, A(g + g_0)) - \delta^2(y, A(g + g_0))] dm(g). \quad (3.4.4)$$

Equivalentemente, $A_\eta(g_0)$ es el baricentro del *pushforward* por A de la medida de Haar restringida a $g_0 + U_\eta$.

Figura 3.5: Ejemplo de los mollifiers en espacios de Hadamard. Con naranja representamos un entorno de la identidad el cual trasladamos, y luego minimizamos la función en la nueva región naranja.

Esta definición usa la idea de *mollifier*, reemplazando la media aritmética por los promedios inducidos por los baricentros. Vamos a probar que, como en el caso usual de *mollifiers*, estas funciones continuas nos dan una buena aproximación a funciones de L^1 (Teorema 3.4.10). Con este objetivo, primero probamos el siguiente lema.

Lema 3.4.7. Sea $A \in L^p(G, M)$ donde $1 \leq p < \infty$. Definimos la función $\varphi : G \rightarrow [0, +\infty)$ por

$$\varphi(h) = \int_G \delta^p(A(g), A(g + h)) dm(g).$$

Luego, esta función es continua.

Demostración. Fijado $z_0 \in M$, definamos la medida ν en los conjuntos Borelianos de G por

$$\nu(B) := \int_B \delta^p(A(g), z_0) dm(g).$$

Por definición, ν es absolutamente continua con respecto a la medida de Haar m . En consecuencia, dado $\varepsilon > 0$, existe $\eta > 0$, tal que, siempre que el conjunto Boreliano B satisfaga

$$\int_B dm(g) < \eta,$$

se tiene que

$$\nu(B) = \int_B \delta^p(A(g), z_0) dm(g) < \frac{\varepsilon}{2^{p+2}}, \quad (3.4.5)$$

Por el Teorema de Lusin ([26, Teorema 7.5.2] o Teorema 6.3.1 del apéndice), existe un conjunto compacto $C_\eta \subset G$ tal que $m(C_\eta) \geq 1 - \eta/2$, y la restricción de A a C_η es (uniformemente) continua.

Como m es una medida de Haar, es suficiente probar la continuidad de φ en la identidad. Con este objetivo, tomemos un entorno de la identidad U tal que, siempre que $g_1, g_2 \in C_\eta$ satisfaga $g_1 - g_2 \in U$, se tiene que

$$\delta^p(A(g_1), A(g_2)) \leq \frac{\varepsilon}{2}.$$

Dado $h \in U$, definimos $\Omega := C_\eta \cap (C_\eta + h)$, y $\Omega^c := G \setminus \Omega$. Entonces

$$\begin{aligned} \int_G \delta^p(A(g), A(g+h)) dm(g) &= \int_\Omega + \int_{\Omega^c} \delta^p(A(g), A(g+h)) dm(g) \\ &\leq \frac{\varepsilon}{2} + \int_{\Omega^c} \delta^p(A(g), A(g+h)) dm(g) \\ &\leq \frac{\varepsilon}{2} + \int_{\Omega^c} [\delta(A(g), z_0) + \delta(A(g+h), z_0)]^p dm(\omega) \\ &= \frac{\varepsilon}{2} + 2^{p+1} \int_{\Omega^c} \delta^p(A(g), z_0) dm(g), \end{aligned}$$

donde en la última igualdad usamos que m es *shift* invariante (invariante a desplazamientos). Como $|\Omega^c| < \delta$ obtenemos que

$$\int_G \delta^p(A(g), A(g+h)) dm(g) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

■

Corolario 3.4.8. *Para todo $\eta > 0$, las funciones A_η son continuas.*

Demostración. Por el Lema 3.4.2

$$\begin{aligned} \delta(A_\eta(h_1), A_\eta(h_2)) &\leq \frac{1}{m(U_\eta)} \int_{U_\eta} \delta(A(g+h_1), A(g+h_2)) dm(g) \\ &\leq \frac{1}{m(U_\eta)} \int_G \delta(A(g+h_1), A(g+h_2)) dm(g) \\ &= \frac{1}{m(U_\eta)} \int_G \delta[A(g), A(g+(h_2-h_1))] dm(g). \end{aligned}$$

Luego, la continuidad de A_η es una consecuencia de la continuidad de φ en la identidad. ■

La función $A \mapsto A_\varepsilon$ tiene la siguiente propiedad de continuidad que nos será útil.

Lema 3.4.9. *Sea $A, B \in L^1(G, M)$, y $\eta > 0$. Para todo $\varepsilon > 0$, existe $\rho > 0$ tal que si*

$$\int_G \delta(A(g), B(g)) dm(g) \leq \rho,$$

entonces las funciones continuas correspondientes A_η y B_η satisfacen que

$$\max_{g \in G} \delta(A_\eta(g), B_\eta(g)) \leq \varepsilon.$$

Demostración. Dado $\varepsilon > 0$, tomemos $\rho = m(U_\eta)\varepsilon$. Entonces, por el Lema 3.4.2

$$\begin{aligned}\delta(A_\eta(g), B_\eta(g)) &\leq \frac{1}{|U_\eta|} \int_{U_\eta} \delta(A(g+h), B(g+h)) dm(h) \\ &\leq \frac{1}{|U_\eta|} \int_G \delta(A(h), B(h)) dm(h) \leq \varepsilon,\end{aligned}$$

para todo $g \in G$. ■

Llegamos al resultado principal de la sección.

Proposición 3.4.10. *Dada una función continua $A \in L^1(G, M)$, si A_η son las funciones continuas definidas por (3.4.4) entonces*

$$\lim_{\eta \rightarrow 0^+} \int_G \delta(A(g), A_\eta(g)) dm(g) = 0.$$

Demostración. Primero, supongamos que $A \in L^2(G, M)$. En este caso, por la desigualdad de la varianza, se tiene que

$$\delta^2(A(g), A_\eta(g)) \leq \frac{1}{|U_\eta|} \int_{U_\eta} \delta^2(A(g), A(g+h)) dm(h).$$

Entonces, usando el Teorema de Fubini

$$\begin{aligned}\int_G \delta^2(A(g), A_\eta(g)) dm(g) &\leq \frac{1}{|U_\eta|} \int_{U_\eta} \int_G \delta^2(A(g), A(g+h)) dm(g) dm(h) \\ &= \frac{1}{|U_\eta|} \int_{U_\eta} \varphi(h) dm(h).\end{aligned}$$

Por el Lema 3.4.7, la función φ es continua. En consecuencia, si e denota la identidad de G ,

$$\lim_{\eta \rightarrow 0^+} \frac{1}{|U_\eta|} \int_{U_\eta} \varphi(h) dm(h) = \varphi(e) = 0.$$

Esto prueba el resultado para las funciones en $L^2(G, M)$ ya que por la desigualdad de Jensen tenemos

$$\int_G \delta(A(g), A_\eta(g)) dm(g) \leq \left(\int_G \delta^2(A(g), A_\eta(g)) dm(g) \right)^{1/2}.$$

Ahora, consideremos una $A \in L^1(G, M)$. Fijado $z_0 \in M$, para cada número natural N , definimos el truncamiento

$$A^{(N)}(g) := \begin{cases} A(g) & \text{if } \delta(A(g), z_0) < N \\ z_0 & \text{if } \delta(A(g), z_0) \geq N \end{cases}.$$

Para todo N tenemos $A^{(N)} \in L^1(G, M) \cap L^\infty(G, M)$, y luego, también pertenece a $L^2(G, M)$. Por otro lado, como la función definida en G por $g \mapsto \delta(A(g), z_0)$ es integrable, tenemos

$$\int_G \delta(A(g), A^{(N)}(g)) dm(g) = \int_{\{g: \delta(A(g), z_0) \geq N\}} \delta(A(g), z_0) dm(g) \xrightarrow[N \rightarrow \infty]{} 0. \quad (3.4.6)$$

Entonces, si A_η y $A_\eta^{(N)}$ son las funciones continuas asociadas a A y $A^{(N)}$ respectivamente,

$$\begin{aligned} \int_G \delta(A(g), A_\eta(g)) \, dm(g) &\leq \int_G \delta(A(g), A^{(N)}(g)) \, dm(g) \\ &+ \int_G \delta(A^{(N)}(g), A_\eta^{(N)}(g)) \, dm(g) \\ &+ \int_G \delta(A_\eta^{(N)}(g), A_\eta(g)) \, dm(g). \end{aligned}$$

Notar que cada término del lado derecho tiende a cero: el primero por (3.4.6), el segundo por el caso L^2 hecho en la primer parte de la demostración y el último por el Lema 3.4.9. ■

3.4.3 Demostración del caso L^1

Ahora si podemos dar una demostración del Teorema ergódico para funciones en L^1 .

Prueba del Teorema 3.4.1

Sea $\varepsilon > 0$. Para todo $k \in \mathbb{N}$ sea A_k una función continua tal que

$$\int_G \delta(A(g), A_k(g)) \, dm(g) \leq \frac{1}{k}.$$

Por el Lema 3.4.3, podemos tomar un conjunto de medida nula $N \subseteq G$, tal que si tomamos $g \in G \setminus N$ y $k \in \mathbb{N}$, existe n_0 , que puede depender de g y k , tal que

$$\delta(S_n(a^\tau(g)), S_n(a_{(k)}^\tau(g))) \leq \frac{\varepsilon}{4} + \int_G \delta(A(g), A_k(g)) \, dm(g),$$

siempre que $n \geq n_0$. En esta expresión, $a_{(k)}^\tau$ es la sucesión definida en términos de A_k y τ como en (3.3.1). Sea $g \in G \setminus N$ fijo. Tomando k tal que $1/k < \varepsilon/4$, tenemos que

$$\delta(S_n(a^\tau(g)), S_n(a_{(k)}^\tau(g))) \leq \frac{\varepsilon}{2},$$

para todo $n \geq n_0$. Por el Lema 3.4.2, se tiene que $\delta(\beta_A, \beta_{A_k}) \leq \frac{\varepsilon}{4}$ donde

$$\begin{aligned} \beta_A &= \operatorname{argmin}_{z \in M} \int_G [\delta^2(A(g), z) - \delta^2(A(g), y)] \, dm(g), \\ \beta_{A_k} &= \operatorname{argmin}_{z \in M} \int_G [\delta^2(A_k(g), z) - \delta^2(A_k(g), y)] \, dm(g). \end{aligned}$$

Finalmente, por el Teorema 3.3.1, existe $n_1 \geq 1$ tal que, para todo $n \geq n_1$,

$$\delta(S_n(a_{(k)}^\tau(g), \beta_{A_k}) \leq \frac{\varepsilon}{4}.$$

Combinando estas desigualdades tenemos

$$\begin{aligned} \delta(S_n(a^\tau(g)), \beta_A) &\leq \delta(S_n(a^\tau(g)), S_n(a_{(k)}^\tau(g))) \\ &+ \delta(S_n(a_{(k)}^\tau(g), \beta_{A_k}) + \delta(\beta_{A_k}, \beta_A) \leq \varepsilon, \end{aligned}$$

lo que completa la prueba. ■

3.5 Independencia del punto de comienzo

Como mencionamos en la motivación, Holbrook prueba que la convergencia de la media inductiva es independiente del punto de comienzo. Aquí vamos a presentar una prueba de este fenómeno, que podría no ser más que una simple observación pues, este resultado se extiende a espacios de Hadamard, como consecuencia directa de la contractividad de la media inductiva (Corolario 3.2.3). Sin embargo vamos a demostrar la independencia del punto de comienzo a falta de una referencia bibliográfica.

Con el fin de no introducir notación nueva notar que:

1. Dado $x \in M$, entonces $x \#_t x = x$ para todo $t \in [0, 1]$.
2. Utilizando la notación de las secciones anteriores, entendemos que la convergencia es independiente del punto de comienzo si para todo $x \in M$,

$$x \#_{\frac{1}{m}} a_m^\tau(g) \#_{\frac{1}{m+1}} a_{m+1}^\tau(g) \#_{\frac{1}{m+2}} a_{m+2}^\tau(g) \#_{\frac{1}{m+3}} a_{m+3}^\tau(g) \cdots \xrightarrow{m \rightarrow \infty} \beta_A.$$

3. Si queremos comparar la sucesión del ítem anterior con la que conocíamos, podemos completarla utilizando el ítem uno, es decir, podemos comparar

$$\underbrace{x \#_{\frac{1}{2}} x \#_{\frac{1}{3}} \cdots x \#_{\frac{1}{m}} a_m^\tau(g)}_{m - 2 \text{ veces}} \#_{\frac{1}{m+1}} a_{m+1}^\tau(g) \#_{\frac{1}{m+2}} a_{m+2}^\tau(g) \#_{\frac{1}{m+3}} a_{m+3}^\tau(g),$$

con

$$a_1^\tau(g) \#_{\frac{1}{2}} a_2^\tau(g) \#_{\frac{1}{3}} a_3^\tau(g) \#_{\frac{1}{4}} \cdots \#_{\frac{1}{m}} a_m^\tau(g) \#_{\frac{1}{m+1}} a_{m+1}^\tau(g) \#_{\frac{1}{m+2}} a_{m+2}^\tau(g) \#_{\frac{1}{m+3}} a_{m+3}^\tau(g).$$

Cuya distancia está acotada por,

$$\frac{1}{n} \sum_{i=1}^m \delta(x, a_i^\tau(g)) \xrightarrow{n \rightarrow \infty} 0,$$

como consecuencia del Corolario 3.2.3.

3.6 Otra forma de convergencia

En las secciones anteriores trabajamos en la demostración del siguiente teorema.

Teorema. Dada $A \in L^1(G, M)$, para casi todo $g \in G$

$$\lim_{n \rightarrow \infty} S_n(a^\tau(g)) = \beta_A.$$

Observar que este teorema nos da la convergencia en casi todo punto para funciones $A \in L^1(G, M)$. El siguiente teorema nos dice que, además de la convergencia en casi todo punto, tenemos un tipo de convergencia en L^p para funciones en $L^p(G, M)$. Más precisamente, terminaremos este capítulo demostrando el siguiente resultado.

Teorema 3.6.1. Sean $1 \leq p < \infty$ y $A \in L^p(G, M)$. Entonces

$$\lim_{n \rightarrow \infty} \int_G \delta^p(S_n(a^\tau(g)), \beta_A) dm(g) = 0. \quad (3.6.1)$$

Demostración. Definamos la siguiente medida en los conjuntos Borelianos de G :

$$\nu(B) := \int_B \delta^p(A(g), \beta_A) dm(g).$$

Por definición, ν es absolutamente continua con respecto a la medida de Haar m . En consecuencia, dado $\varepsilon > 0$, existe $\eta > 0$, tal que, siempre que el conjunto de Borel B satisfaga

$$\int_B dm(g) < \eta,$$

se tiene que

$$\nu(B) = \int_B \delta^p(A(g), \beta_A) dm(g) < \frac{\varepsilon}{2}. \quad (3.6.2)$$

Por el Teorema de Egoroff, como

$$\delta^p(S_n(a^\tau(g)), \beta_A) \xrightarrow{n \rightarrow \infty} 0$$

en casi todo punto en un espacio de medida finita, entonces existe un conjunto $C_\eta \subset G$ con $m(C_\eta) < \eta$, tal que

$$\delta^p(S_n(a^\tau(g)), \beta_A) \xrightarrow{n \rightarrow \infty} 0 \quad (3.6.3)$$

uniformemente en $G \setminus C_\eta$.

Por otro lado,

$$\delta(S_n(a^\tau(g)), \beta_A) \leq \frac{1}{n} \sum_{k=0}^{n-1} \delta(a_k^\tau(g), \beta_A) = \frac{1}{n} \sum_{k=0}^{n-1} \delta(A(\tau^k(g)), \beta_A),$$

Luego, por la desigualdad de Jensen,

$$\delta^p(S_n(a^\tau(g)), \beta_A) \leq \frac{1}{n} \sum_{k=0}^{n-1} \delta^p(A(\tau^k(g)), \beta_A).$$

Tomando integral sobre C_η en la desigualdad anterior obtenemos

$$\begin{aligned} \int_G \delta^p(S_n(a^\tau(g)), \beta_A) dm(g) &= \int_{G \setminus C_\eta} \delta^p(S_n(a^\tau(g)), \beta_A) dm(g) + \int_{C_\eta} \delta^p(S_n(a^\tau(g)), \beta_A) dm(g) \\ &\leq \int_{G \setminus C_\eta} \delta^p(S_n(a^\tau(g)), \beta_A) dm(g) + \int_{C_\eta} \frac{1}{n} \sum_{k=0}^{n-1} \delta^p(A(\tau^k(g)), \beta_A) dm(g) \\ &= \int_{G \setminus C_\eta} \delta^p(S_n(a^\tau(g)), \beta_A) dm(g) + \sum_{k=0}^{n-1} \frac{1}{n} \int_{C_\eta} \delta^p(A(\tau^k(g)), \beta_A) dm(g). \end{aligned}$$

Ahora,

1. Para el ε dado en un comienzo, existe $n_0 \in \mathbb{N}$, tal que para todo $n \geq n_0$,

$$\int_{G \setminus C_\eta} \delta^p(S_n(a^\tau(g)), \beta_A) dm(g) < \frac{\varepsilon}{2},$$

como consecuencia de (3.6.3).

2. Como

$$\int_{C_\eta} dm(g) < \eta,$$

entonces por (3.6.2)

$$\nu(C_\eta) = \int_{C_\eta} \delta^p(A(g), \beta_A) dm(g) < \frac{\varepsilon}{2}.$$

Luego, para todo $n \in \mathbb{N}$,

$$\sum_{k=0}^{n-1} \frac{1}{n} \int_{C_\eta} \delta^p(A(\tau^k(g)), \beta_A) dm(g) < \frac{\varepsilon}{2}.$$

Finalmente, combinando estas dos cotas, dado ε , existe $n_0 \in \mathbb{N}$, tal que para todo $n \geq n_0$,

$$\int_G \delta^p(S_n(a^\tau(g)), \beta_A) dm(g) < \varepsilon,$$

lo que termina la prueba. ■

4 | Estructura de las curvas minimales en $\mathcal{H}(n)$, $\mathcal{P}(n)$ y $\mathcal{U}(n)$ con respecto a la métrica de Finsler bi-invariante inducida por la norma traza

4.1 Motivación

El cono convexo abierto de matrices (estrictamente) positivas $\mathcal{P}(n)$ del espacio $\mathcal{H}(n)$ de matrices Hermitianas, tiene una rica estructura geométrica, heredada de las normas Gauge simétricas (equivalentemente, normas unitariamente invariantes - ver capítulo 6 apéndice). Para una norma Gauge simétrica Φ en $\mathbb{R}^n$, la norma en $\mathcal{H}(n)$ definida por

$$\|A\|_{\Phi} := \Phi(\lambda_1(A), \lambda_2(A), \dots, \lambda_n(A)),$$

donde $\lambda_j(A)$ son los autovalores (ordenados decrecientemente) de $A \in \mathcal{H}(n)$, induce una métrica de Finsler $\mathcal{G}l(n)$ -invariante en $\mathcal{P}(n)$ y la distancia correspondiente entre dos matrices definidas positivas A y B es

$$d_{\Phi}(A, B) = \|\log(A^{-1/2}BA^{-1/2})\|_{\Phi}.$$

Una de las propiedades comunes de esta geometría es que la curva $t \mapsto A \#_t B$ actúa como una geodésica minimal de A a B para d_{Φ} , donde

$$A \#_t B = A^{1/2}(A^{-1/2}BA^{-1/2})^t A^{1/2}.$$

Una consecuencia inmediata es que, $A \#_{1/2} B$, es el punto medio entre A y B para d_{Φ} . Ver por ejemplo [11], [12] para más detalles.

La norma Frobenius $\|\cdot\|_2$ da lugar a la estructura Riemanniana

$$\langle X, Y \rangle_A = \text{Tr}(A^{-1}XA^{-1}Y),$$

donde $A \in \mathcal{P}(n)$, $X, Y \in T_A(\mathcal{P}(n)) = \mathcal{H}(n)$. En otras palabras, $\mathcal{P}(n)$ se convierte en una variedad de Cartan-Hadamard, simplemente conexa, con curvatura seccional no positiva (ver [11], [12], [32]). En este caso, hay una única geodésica métrica minimal entre A y B , que coincide con la media geométrica $A \#_t B$.

En general, el punto medio y la geodésica minimal entre dos puntos de $\mathcal{P}(n)$ no es única bajo d_{Φ} . Las normas Schatten p tiene esta propiedad de unicidad para $1 < p < \infty$, pero no

para $p = 1$ o $p = \infty$. Por ejemplo, hay infinitos (geodésicas mínimas) puntos medios entre un cierto par de matrices definidas positivas con respecto a la métrica de Finsler heredada por la norma espectral $\|\cdot\|_{sp}$, que coincide con la métrica de Thompson,

$$d_{\infty}(A, B) = \|\log(A^{-1/2}BA^{-1/2})\|_{\infty} = \max \{ \log \lambda_1(A^{-1/2}BA^{-1/2}), \log \lambda_1(A^{-1/2}B^{-1}A^{-1/2}) \}.$$

Recordamos la construcción de Nussbaum para curvas minimales con respecto a la métrica de Thompson.

Teorema 4.1.1 ([49]). Sean $A, B \in \mathcal{P}(n)$ y sea $\lambda_k = \lambda_k(A^{-1}B)$, $k = 1, \dots, n$.

i) Si $\lambda_n^{-1} \leq \lambda_1$, entonces para todo $\alpha \in [\lambda_1^{-1}, \lambda_n]$,

$$f_{\alpha}(t) = \begin{cases} \frac{\lambda_1^t - \alpha^t}{\lambda_1 - \alpha} B + \frac{\lambda_1 \alpha^t - \alpha \lambda_1^t}{\lambda_1 - \alpha} A & \lambda_1 \neq \lambda_n \\ \lambda_1^t A & \lambda_1 = \lambda_n \end{cases}$$

es una curva minimal de A a B con respecto a la métrica de Thompson.

ii) Si $\lambda_1 \leq \lambda_n^{-1}$, entonces para todo $\alpha \in [\lambda_1, \lambda_n^{-1}]$,

$$f_{\alpha}(t) = \begin{cases} \frac{\alpha^t - \lambda_n^t}{\alpha - \lambda_n} B + \frac{\alpha \lambda_n^t - \lambda_n \alpha^t}{\alpha - \lambda_n} A & \lambda_1 \neq \lambda_n \\ \lambda_1^t A & \lambda_1 = \lambda_n \end{cases}$$

es una curva minimal de A a B con respecto a la métrica de Thompson.

Este problema tienen sus análogos cuando consideramos $p = 1$. En este capítulo vamos a estudiar el caso $p = 1$. Es decir, nos preguntamos: Dadas $A, B \in \mathcal{P}(n)$ ¿Podemos caracterizar las curvas minimales que los unen con respecto a la métrica heredada por la norma traza? Vamos a responder esta pregunta a lo largo del capítulo en el Corolario 4.2.10 y el Teorema 4.2.11.

4.2 Estructura de las curvas minimales para la norma traza

El objetivo de esta sección es estudiar la estructura de las curvas minimales que unen dos matrices medidas con la norma traza. Vamos a dar condiciones necesarias y suficientes para que esto suceda. Los espacios que nos interesan son: el espacio de matrices positivas $\mathcal{P}(n)$, estudiado por Lim [40] para la norma espectral y el espacio de matrices unitarias $\mathcal{U}(n)$.

Para estudiar propiedades de las curvas minimales en estos espacios, nos enfocamos en el estudio del espacio de matrices Hermitianas $\mathcal{H}(n)$, porque toda matriz positiva $P \in \mathcal{P}(n)$ se puede escribir de la forma $P = e^A$, donde $A \in \mathcal{H}(n)$ y toda matriz unitaria $U \in \mathcal{U}(n)$ se puede escribir de la forma $U = e^{iB}$ donde $B \in \mathcal{H}(n)$.

4.2.1 Estudio de curvas minimales en el espacio de matrices Hermitianas dotada de la norma traza

Convención: En esta subsección vamos a suponer lo siguiente:

1. Las curvas son suaves, es decir, curvas C^1 a trozos.
2. Sea $\mathcal{H}(n)$ el conjunto de matrices Hermitianas $n \times n$.

Dada una curva suave a trozos $\alpha : [a, b] \rightarrow \mathcal{H}(n)$, podemos definir la longitud de curva dada por

$$L(\alpha) = \int_a^b \|\dot{\alpha}(t)\|_1 dt.$$

La norma que consideramos $\|\cdot\|_1$ es la norma traza de estas matrices.

La distancia rectificable entre $A, B \in \mathcal{H}(n)$ está dada por

$$d_1(A, B) = \inf \{L(\alpha) \text{ donde } \alpha : [a, b] \rightarrow \mathcal{H}(n) \text{ une } A \text{ con } B \text{ en } \mathcal{H}(n)\}.$$

Una curva es **minimal** si su longitud es igual a la distancia de los puntos que une. Es decir,

$$L(\alpha) = d_1(A, B).$$

En esta sección vamos a dar una caracterización de las curvas minimales medidas con la norma traza en el espacio de matrices Hermitianas. Con este objetivo, nos va a ser útil recordar un resultado que seguramente le es conocido: ¿Cuáles son las curvas minimales en $\mathbb{R}^n$ medidas con la norma traza?

Proposición 4.2.1. *Sea $\gamma : [0, 1] \rightarrow \mathbb{R}^n$ una curva minimal que une el vector $O = (0, \dots, 0)$ con $V = (v_1, \dots, v_n)$ en $\mathbb{R}^n$. Son equivalentes:*

1. *La curva $\gamma : [0, 1] \rightarrow \mathbb{R}^n$ es minimal.*
2. *Las curvas $\gamma_j : [0, 1] \rightarrow \mathbb{R}$ que unen 0 con v_j son minimales.*
3. *Las curvas $\gamma_j : [0, 1] \rightarrow \mathbb{R}$ que unen 0 con v_j son tales que $\dot{\gamma}_j(t) > 0$, si v_j es estrictamente positivo, o $\dot{\gamma}_j(t) < 0$, si v_j es estrictamente negativo, para todo $t \in [0, 1]$. Si la curva $\gamma_j : [0, 1] \rightarrow \mathbb{R}$ une 0 con 0 debe ser la curva idénticamente nula.*

Ahora volvemos al espacio de matrices Hermitianas para acercarnos a nuestro resultado. El siguiente resultado también le debe ser conocido y nos dice que el segmento que une dos matrices Hermitianas tiene longitud menor o igual que la longitud de cualquier curva poligonal que los une. En particular,

$$d_1(A, B) = \|A - B\|_1.$$

Vamos a dar una demostración simple de este resultado.

Lema 4.2.2. *Los segmentos son curvas minimales en $\mathcal{H}(n)$ con respecto a la norma traza.*

Demostración. Como toda curva puede ser aproximada por poligonales, es suficiente probar que el segmento que une las matrices 0 con $A \in \mathcal{H}(n)$ es más corto que toda curva poligonal que los une. Esto se deduce inductivamente y el argumento se puede ver en la siguiente imagen.

Sea $\gamma(t) = tA$ el segmento que une 0 con A y tomemos cualquier $B \in \mathcal{H}(n)$. Entonces la poligonal

$$\gamma_1(t) = \begin{cases} 2tB & t \in [0, 1/2] \\ (2-2t)B + (2t-1)A & t \in [1/2, 1] \end{cases}$$

satisface que

$$L(\gamma_1) = \int_0^1 \|\dot{\gamma}_1(t)\|_1 dt = \|B\|_1 + \|A - B\|_1 \geq \|A\|_1 = L(\gamma).$$

■

Ahora vamos a dar una condición necesaria para que una curva en el espacio de matrices Hermitianas sea minimal. Esta primera condición nos dice que las entradas de la diagonal tienen que ser curvas minimales. Antes de demostrar este hecho, vamos a recordar la definición del operador Pinching.

Definición. Sea $\mathcal{P} = (P_1, \dots, P_k) \in \mathcal{M}_n(\mathbb{C})^k$ un sistema de proyectores en $\mathcal{M}_n(\mathbb{C})$. Esto significa que las entradas P_i son proyecciones mutuamente ortogonales tales que

$$\sum_{i=1}^k P_i = I.$$

El **operador Pinching** asociado, $\mathcal{C}_{\mathcal{P}} : \mathcal{M}_n(\mathbb{C}) \rightarrow \mathcal{M}_n(\mathbb{C})$, está dado por

$$\mathcal{C}_{\mathcal{P}}(A) = \sum_{j=1}^k P_j A P_j, \quad \text{para todo } A \in \mathcal{M}_n(\mathbb{C}).$$

Una propiedad conocida de este operador es que, si $\|\cdot\|$ es una norma unitariamente invariante, entonces

$$\|\mathcal{C}(A)\| \leq \|A\|, \quad \text{para todo } A \in \mathcal{M}_n(\mathbb{C}). \quad (4.2.1)$$

En realidad, con respecto a la norma traza, vamos a probar un resultado más general, que de algún modo nos dice que, el operador Pinching preserva la minimalidad de curvas.

Observación. Como la norma $\|\cdot\|_1$ es unitariamente invariante, sin pérdida de generalidad y en busca de simplicidad, podemos suponer que una de las matrices es la nula y la otra matriz Hermitiana es diagonal. En efecto, como

$$d_1(A, B) = d_1(0, B - A),$$

y $B - A \in \mathcal{H}(n)$, existe una matriz unitaria U que la diagonaliza. ▲

Lema 4.2.3. Sea $X : [0; 1] \rightarrow \mathcal{H}(n)$ una curva minimal. Supongamos además que $X(0) = 0$. Sea $D = X(1) \in \mathcal{H}(n)$ y $\mathcal{P}$ un sistema de proyectores, todos ellos tales que conmutan con D . Entonces la curva “Pinchada” $\mathcal{C}_{\mathcal{P}}(X(t))$ es también una curva **minimal** que une 0 con $D \in \mathcal{H}(n)$.

Demostración. Es una consecuencia inmediata de (4.2.1) y el hecho que

$$\frac{d}{dt} \mathcal{C}_{\mathcal{P}}(X(t)) = \mathcal{C}_{\mathcal{P}}(\dot{X}(t)) \xrightarrow{(4.2.1)} \left\| \mathcal{C}_{\mathcal{P}}(\dot{X}(t)) \right\|_1 \leq \left\| \dot{X}(t) \right\|_1 \quad \text{para } t \in [0, 1]. \quad (4.2.2)$$

Luego $L(\mathcal{C}_{\mathcal{P}}(X(t))) \leq L(X(t)) = \|D\|_1$. Por otro lado, $\mathcal{C}_{\mathcal{P}}(X(1)) = \mathcal{C}_{\mathcal{P}}(D) = D$ por la hipótesis de conmutatividad. ■

Un caso particular es el siguiente: sea $\mathcal{P}_0$ un sistema de proyectores asociados a la base canónica de $\mathbb{C}^n$, tal que, la imagen de $\mathcal{C}_{\mathcal{P}_0}$ es el conjunto de matrices diagonales usual.

Corolario 4.2.4. Sea $X : [0; 1] \rightarrow \mathcal{H}(n)$ una curva como en el Lema 4.2.3. Supongamos que $D = X(1) = \text{diag}(d_1, \dots, d_n)$ es una matriz diagonal. Entonces para todo $t \in [0, 1]$,

$$\mathcal{C}_{\mathcal{P}_0}(X(t)) = \text{diag}(x_1(t)d_1, \dots, x_n(t)d_n),$$

donde $x_i : [0, 1] \rightarrow \mathbb{R}$ son tales que $x_i(0) = 0, x_i(1) = 1$ y $\dot{x}_i(t) > 0$ si d_i es positivo, o $\dot{x}_i(t) < 0$ si d_i es negativo. En otras palabras, las curvas en la diagonal de la matriz, son minimales también.

Ya casi estamos listos para caracterizar las curvas minimales en el espacio de matrices Hermitianas, pero antes, vamos a probar como se caracterizan en el caso de matrices 2×2 . Las condiciones que obtuvimos son las generalizaciones naturales de las condiciones que se tienen en la Proposición 4.2.1 (inciso 3). Este resultado va a ser importante para generalizarlo a matrices $n \times n$.

Lema 4.2.5. Sea $X : [0; 1] \rightarrow \mathcal{H}(2)$ una curva minimal que une 0 con $D \in \mathcal{H}(2)$,

$$D = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}.$$

1. Si $\alpha, \beta \geq 0$, entonces $\dot{X}(t) \geq 0$.
2. Si $\alpha, \beta < 0$, entonces $\dot{X}(t) \leq 0$.
3. Si $\alpha > 0, \beta < 0$, entonces $X(t)$ es una matriz diagonal.

Demostración.

1. Observar que, por el Lema 4.2.3, la curva diagonal

$$\mathcal{C}(X(t)) = \begin{pmatrix} x_{11}(t) & 0 \\ 0 & x_{22}(t) \end{pmatrix}$$

es minimal. Usando la ecuación (4.2.2), obtenemos que

$$\operatorname{tr} |\mathcal{C}(\dot{X}(t))| = \operatorname{tr} |\dot{X}(t)|, \quad (4.2.3)$$

ya que la curva es suave y $L(\mathcal{C}(X(t))) = L(X(t))$. Por otro lado, por el Corolario 4.2.4 sabemos que $\dot{x}_{11}(t) \geq 0$ y $\dot{x}_{22}(t) \geq 0 \implies \mathcal{C}(\dot{X}(t)) \geq 0$. Si combinamos

$$\operatorname{tr} |\mathcal{C}(\dot{X}(t))| = \operatorname{tr} \mathcal{C}(\dot{X}(t)) = \operatorname{tr} \dot{X}(t) \leq \operatorname{tr} |\dot{X}(t)|, \quad (4.2.4)$$

con (4.2.3), tenemos $\operatorname{tr} \dot{X}(t) = \operatorname{tr} |\dot{X}(t)|$. Lo que implica que $\dot{X}(t) = |\dot{X}(t)| \geq 0$.

2. Se deduce del caso anterior tomando $Y(t) = -X(t)$.

3. Sea

$$X(t) = \begin{pmatrix} a(t) & b(t) \\ b(t) & c(t) \end{pmatrix},$$

y sean $\lambda_1(t), \lambda_2(t)$ los autovalores de $\dot{X}(t)$ ordenamos de manera decreciente. Como

$$(\dot{a}(t), \dot{c}(t)) \prec (\lambda_1(t), \lambda_2(t)),$$

y $\dot{a}(t) \geq 0, \dot{c}(t) \leq 0$ (por el Corolario 4.2.4), entonces

$$\lambda_2(t) \leq \dot{c}(t) \leq 0 \leq \dot{a}(t) \leq \lambda_1(t) \quad \text{y} \quad |\dot{c}(t)| + |\dot{a}(t)| \leq |\lambda_1(t)| + |\lambda_2(t)| \quad (4.2.5)$$

para todo $t \in [0, 1]$. Por otro lado, usando el Lema 4.2.3, tenemos que

$$\begin{aligned} \int_0^1 |\dot{c}(t)| + |\dot{a}(t)| dt &= \int_0^1 \operatorname{tr}(|\mathcal{C}(\dot{X}(t))|) dt = \int_0^1 \operatorname{tr}(|\dot{X}(t)|) dt = \\ &= \int_0^1 \|\dot{X}(t)\|_1 dt = \int_0^1 |\lambda_1(t)| + |\lambda_2(t)| dt. \end{aligned}$$

Como la curva es suave, por la ecuación (4.2.5),

$$|\dot{c}(t)| + |\dot{a}(t)| = |\lambda_1(t)| + |\lambda_2(t)| \implies \dot{a}(t) = \lambda_1(t), \quad \dot{c}(t) = \lambda_2(t).$$

Aplicando la norma Frobenius a la matriz $\dot{X}(t)$, concluimos que $\dot{b}(t) = 0$ para todo $t \in [0, 1]$. Como $b(1) = 0$, concluimos que $b(t) = 0$ para todo $t \in [0, 1]$.

La implicación recíproca, es decir, si vale alguna de las condiciones entonces la curva es minimal, se demuestra de la siguiente manera:

1. Supongamos que $\alpha, \beta \geq 0$ y $\dot{X}(t) \geq 0$. Luego $\operatorname{tr} \dot{X}(t) = \operatorname{tr} |\dot{X}(t)|$, por lo cual

$$\begin{aligned} L(X(t)) &= \int_0^1 \operatorname{tr} |\dot{X}(t)| dt = \int_0^1 \operatorname{tr} \dot{X}(t) dt = \int_0^1 \dot{x}_{11}(t) + \dot{x}_{22}(t) dt \\ &= x_{11}(1) - x_{11}(0) + x_{22}(1) - x_{22}(0) = \alpha - 0 + \beta - 0 = \|D\|_1. \end{aligned}$$

2. Supongamos que $\alpha, \beta < 0$ y $\dot{X}(t) \leq 0$. Nuevamente, se deduce del caso anterior tomando $Y(t) = -X(t)$.
3. Supongamos que $\alpha > 0, \beta < 0$ y que $X(t)$ es una matriz diagonal. Luego

$$\begin{aligned} L(X(t)) &= \int_0^1 \text{tr} |\dot{X}(t)| dt = \int_0^1 |\dot{x}_{11}(t)| + |\dot{x}_{22}(t)| dt \stackrel{\text{Cor 4.2.4}}{=} \int_0^1 \dot{x}_{11}(t) - \dot{x}_{22}(t) dt \\ &= (x_{11}(1) - x_{11}(0)) - (x_{22}(1) - x_{22}(0)) = (\alpha - 0) - (\beta - 0) = |\alpha| + |\beta| = \|D\|_1. \end{aligned}$$

■

Observación. La igualdad

$$\text{tr} \dot{X}(t) = \text{tr} |\dot{X}(t)|,$$

es una de las claves en la caracterización, pues juega un papel fundamental, en ambas implicaciones de los dos primeros items. ▲

Todas las curvas minimales tienen que tener la siguiente estructura:

Teorema 4.2.6. *Sea $X : [0; 1] \rightarrow \mathcal{H}(n)$ una curva minimal. Supongamos también que $X(0) = 0$, y que $D = X(1) = \text{diag}(d_1, \dots, d_n)$, es una matriz diagonal tal que las entradas d_j están ordenadas de manera decreciente. Entonces la curva minimal que une 0 con D tiene la siguiente estructura:*

$$X(t) = \begin{pmatrix} P(t) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & N(t) \end{pmatrix} \begin{matrix} \mathcal{S}_1 \\ \mathcal{S}_2 \\ \mathcal{S}_3 \end{matrix},$$

donde

$$\mathcal{S}_1 = \text{span} \{e_j : d_j > 0\}, \mathcal{S}_2 = \text{span} \{e_j : d_j = 0\} \quad \text{y} \quad \mathcal{S}_3 = \text{span} \{e_j : d_j < 0\}.$$

Más aún, $P : [0, 1] \rightarrow \mathcal{H}(\mathcal{S}_1)$ es una curva minimal que une $0_{\mathcal{S}_1}$ con $D_{\mathcal{S}_1}$, tal que $\dot{P}(t) \geq 0$, para todo $t \in [0, 1]$, y $N : [0, 1] \rightarrow \mathcal{H}(\mathcal{S}_3)$ es una curva minimal que une $0_{\mathcal{S}_3}$ con $D_{\mathcal{S}_3}$, tal que $\dot{N}(t) \leq 0$, para todo $t \in [0, 1]$.

Demostración de un ejemplo 5×5 : Antes de dar la demostración de este teorema vamos demostrar un ejemplo concreto, donde se van a ver las técnicas utilizadas, y facilitarán la lectura del caso general. Supongamos que $X(t)$ es una curva minimal que une la matriz nula con una matriz diagonal tal que $d_1 \geq d_2 > 0$, $d_3 = 0$, $d_5 \leq d_4 < 0$.

$$\begin{pmatrix} x_{11}(t) & x_{12}(t) & x_{13}(t) & x_{14}(t) & x_{15}(t) \\ x_{21}(t) & x_{22}(t) & x_{23}(t) & x_{24}(t) & x_{25}(t) \\ x_{31}(t) & x_{32}(t) & x_{33}(t) & x_{34}(t) & x_{35}(t) \\ x_{41}(t) & x_{42}(t) & x_{43}(t) & x_{44}(t) & x_{45}(t) \\ x_{51}(t) & x_{52}(t) & x_{53}(t) & x_{54}(t) & x_{55}(t) \end{pmatrix}$$

Observar que podemos utilizar el operador Pinching para probar que los elementos de la diagonal deben ser curvas minimales (Corolario 4.2.4).

$$\begin{pmatrix} x_{11}(t) & x_{12}(t) & x_{13}(t) & x_{14}(t) & x_{15}(t) \\ x_{21}(t) & x_{22}(t) & x_{23}(t) & x_{24}(t) & x_{25}(t) \\ x_{31}(t) & x_{32}(t) & 0 & x_{34}(t) & x_{35}(t) \\ x_{41}(t) & x_{42}(t) & x_{43}(t) & x_{44}(t) & x_{45}(t) \\ x_{51}(t) & x_{52}(t) & x_{53}(t) & x_{54}(t) & x_{55}(t) \end{pmatrix}$$

Como la entrada $x_{33}(t)$ es una curva minimal que une 0 con 0 debe ser constantemente igual a 0. Este el primer paso, poner ceros en las entradas que unen un bloque nulo con un bloque nulo.

$$\begin{pmatrix} x_{11}(t) & x_{12}(t) & x_{13}(t) & x_{14}(t) & x_{15}(t) \\ x_{21}(t) & x_{22}(t) & x_{23}(t) & x_{24}(t) & x_{25}(t) \\ x_{31}(t) & x_{32}(t) & 0 & x_{34}(t) & x_{35}(t) \\ x_{41}(t) & x_{42}(t) & x_{43}(t) & x_{44}(t) & x_{45}(t) \\ x_{51}(t) & x_{52}(t) & x_{53}(t) & x_{54}(t) & x_{55}(t) \end{pmatrix}$$

Ahora queremos tener ceros en las entradas $x_{13}(t)$ y $x_{31}(t)$ para lo cual vamos a utilizar las entradas marcadas en rojo.

$$\begin{pmatrix} x_{11}(t) & x_{12}(t) & x_{13}(t) & x_{14}(t) & x_{15}(t) \\ x_{21}(t) & x_{22}(t) & x_{23}(t) & x_{24}(t) & x_{25}(t) \\ x_{31}(t) & x_{32}(t) & 0 & x_{34}(t) & x_{35}(t) \\ x_{41}(t) & x_{42}(t) & x_{43}(t) & x_{44}(t) & x_{45}(t) \\ x_{51}(t) & x_{52}(t) & x_{53}(t) & x_{54}(t) & x_{55}(t) \end{pmatrix}$$

Utilizando nuevamente el operador Pinching (el cual nos da una curva minimal por el Lema 4.2.3) en las entradas marcadas con color

$$\begin{pmatrix} x_{11}(t) & x_{13}(t) & 0 & 0 & 0 \\ x_{31}(t) & 0 & 0 & 0 & 0 \\ 0 & 0 & x_{22}(t) & x_{24}(t) & x_{25}(t) \\ 0 & 0 & x_{42}(t) & x_{44}(t) & x_{45}(t) \\ 0 & 0 & x_{52}(t) & x_{54}(t) & x_{55}(t) \end{pmatrix}$$

y una matriz de permutación, la cual no afecta la longitud de curva (por ser una matriz unitaria y la norma traza unitariamente invariante), obtenemos una matriz con esta estructura.

$$\begin{pmatrix} x_{11}(t) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & x_{22}(t) & x_{24}(t) & x_{25}(t) \\ 0 & 0 & x_{42}(t) & x_{44}(t) & x_{45}(t) \\ 0 & 0 & x_{52}(t) & x_{54}(t) & x_{55}(t) \end{pmatrix}$$

Utilizando la caracterización para matrices 2×2 en el primer bloque, como la curva derivada de ese bloque debe ser mayor o igual a cero para todo t , los elementos codiagonales deben ser nulos.

$$\begin{pmatrix} x_{11}(t) & x_{12}(t) & 0 & x_{14}(t) & x_{15}(t) \\ x_{21}(t) & x_{22}(t) & x_{23}(t) & x_{24}(t) & x_{25}(t) \\ 0 & x_{32}(t) & 0 & x_{34}(t) & x_{35}(t) \\ x_{41}(t) & x_{42}(t) & x_{43}(t) & x_{44}(t) & x_{45}(t) \\ x_{51}(t) & x_{52}(t) & x_{53}(t) & x_{54}(t) & x_{55}(t) \end{pmatrix}$$

Como resultado obtenemos esta matriz.

$$\begin{pmatrix} x_{11}(t) & x_{12}(t) & 0 & x_{14}(t) & x_{15}(t) \\ x_{21}(t) & x_{22}(t) & 0 & x_{24}(t) & x_{25}(t) \\ 0 & 0 & 0 & x_{34}(t) & x_{35}(t) \\ x_{41}(t) & x_{42}(t) & x_{43}(t) & x_{44}(t) & x_{45}(t) \\ x_{51}(t) & x_{52}(t) & x_{53}(t) & x_{54}(t) & x_{55}(t) \end{pmatrix}$$

Con un razonamiento análogo podemos poner ceros en las entradas $x_{23}(t)$ y $x_{32}(t)$. De esta manera concluimos el segundo paso de la demostración.

$$\begin{pmatrix} x_{11}(t) & x_{12}(t) & 0 & x_{14}(t) & x_{15}(t) \\ x_{21}(t) & x_{22}(t) & 0 & x_{24}(t) & x_{25}(t) \\ 0 & 0 & 0 & 0 & x_{35}(t) \\ x_{41}(t) & x_{42}(t) & 0 & x_{44}(t) & x_{45}(t) \\ x_{51}(t) & x_{52}(t) & x_{53}(t) & x_{54}(t) & x_{55}(t) \end{pmatrix}$$

Análogamente podemos poner ceros en las entradas $x_{34}(t)$ y $x_{43}(t)$, esta vez utilizando la parte b) del Lema 2 $\times$ 2.

$$\begin{pmatrix} x_{11}(t) & x_{12}(t) & 0 & x_{14}(t) & x_{15}(t) \\ x_{21}(t) & x_{22}(t) & 0 & x_{24}(t) & x_{25}(t) \\ 0 & 0 & 0 & 0 & 0 \\ x_{41}(t) & x_{42}(t) & 0 & x_{44}(t) & x_{45}(t) \\ x_{51}(t) & x_{52}(t) & 0 & x_{54}(t) & x_{55}(t) \end{pmatrix}$$

Como también podemos poner ceros en las entradas $x_{35}(t)$ y $x_{53}(t)$. Finaliza de esta manera el tercer paso de la demostración.

$$\begin{pmatrix} \textcolor{red}{x}_{11}(t) & x_{12}(t) & 0 & x_{14}(t) & \textcolor{blue}{0} \\ x_{21}(t) & x_{22}(t) & 0 & x_{24}(t) & x_{25}(t) \\ 0 & 0 & 0 & 0 & 0 \\ x_{41}(t) & x_{42}(t) & 0 & x_{44}(t) & x_{45}(t) \\ \textcolor{blue}{0} & x_{52}(t) & 0 & x_{54}(t) & \textcolor{red}{x}_{55}(t) \end{pmatrix}$$

Usando el operador Pinching y la parte c) del Lema 2 $\times$ 2 podemos poner ceros en las entradas $x_{15}(t)$ y $x_{51}(t)$.

$$\begin{pmatrix} \textcolor{red}{x}_{11}(t) & x_{12}(t) & 0 & \textcolor{blue}{0} & 0 \\ x_{21}(t) & x_{22}(t) & 0 & x_{24}(t) & x_{25}(t) \\ 0 & 0 & 0 & 0 & 0 \\ \textcolor{blue}{0} & x_{42}(t) & 0 & \textcolor{red}{x}_{44}(t) & x_{45}(t) \\ 0 & x_{52}(t) & 0 & x_{54}(t) & x_{55}(t) \end{pmatrix}$$

Usando el mismo ítem podemos poner ceros en las entradas $x_{14}(t)$ y $x_{41}(t)$.

$$\begin{pmatrix} x_{11}(t) & x_{12}(t) & 0 & 0 & 0 \\ x_{21}(t) & \textcolor{red}{x}_{22}(t) & 0 & \textcolor{blue}{0} & x_{25}(t) \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \textcolor{blue}{0} & 0 & \textcolor{red}{x}_{44}(t) & x_{45}(t) \\ 0 & x_{52}(t) & 0 & x_{54}(t) & x_{55}(t) \end{pmatrix}$$

Como en las entradas $x_{24}(t)$ y $x_{42}(t)$.

$$\begin{pmatrix} x_{11}(t) & x_{12}(t) & 0 & 0 & 0 \\ x_{21}(t) & \textcolor{red}{x}_{22}(t) & 0 & 0 & \textcolor{blue}{0} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{44}(t) & x_{45}(t) \\ 0 & \textcolor{blue}{0} & 0 & x_{54}(t) & \textcolor{red}{x}_{55}(t) \end{pmatrix}$$

Por último en las entradas $x_{25}(t)$ y $x_{52}(t)$.

$$\begin{pmatrix} x_{11}(t) & x_{12}(t) & 0 & 0 & 0 \\ x_{21}(t) & x_{22}(t) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{44}(t) & x_{45}(t) \\ 0 & 0 & 0 & x_{54}(t) & x_{55}(t) \end{pmatrix}$$

Con este razonamiento podemos ver que la curva minimal tiene esta forma, donde $\dot{P}(t) \geq 0$ y $\dot{N}(t) \leq 0$ como consecuencia del Lema 2 $\times$ 2.

Demostración. (Teorema 4.2.6) Es fácil probar que si $X(t)$ tiene la forma dada en el enunciado, entonces

$$L(X(t)) = \int_0^1 \|\dot{X}(t)\|_1 dt = \int_0^1 \|\dot{N}(t)\|_1 dt + \int_0^1 \|\dot{P}(t)\|_1 dt = \sum_{j=1}^n |d_j|,$$

por lo que es una curva minimal. Para probar la otra implicación, la clave es el Lema 4.2.5. Sea

$$X(t) = \begin{pmatrix} X_{11}(t) & X_{12}(t) & X_{13}(t) \\ X_{21}(t) & X_{22}(t) & X_{23}(t) \\ X_{31}(t) & X_{32}(t) & X_{33}(t) \end{pmatrix} \begin{matrix} \mathcal{S}_1 \\ \mathcal{S}_2 \\ \mathcal{S}_3 \end{matrix},$$

entonces

1. *Primer paso:* $\dot{X}_{11}(t) \geq 0$, $X_{22}(t) = 0$ y $\dot{X}_{33}(t) \leq 0$. En efecto, si

$$\mathcal{C}(X(t)) = \sum_{j=1}^3 P_j X(t) P_j,$$

donde P_j es la proyección sobre $\mathcal{S}_j$, por el Lema 4.2.3, el bloque $X_{22}(t) = 0$ (porque tiene que ser una curva minimal que une 0 con 0). Una consecuencia del Lema 4.2.5(1)

es que $\dot{X}_{11}(t) \geq 0$, porque es una curva minimal que une $0_{\mathcal{S}_1}$ con $D_{\mathcal{S}_1}$ (las entradas positivas). Es natural renombrarlo como $X_{11}(t) = P(t)$. Análogamente, $\dot{X}_{33}(t) \leq 0$, por lo cual es natural renombrarlo como $X_{33}(t) = N(t)$.

2. *Segundo paso:* $X_{12}(t) = X_{21}(t) = 0$. Tomemos un operador Pinching apropiado de la siguiente manera: Elegimos una entrada diagonal del bloque matricial $P(t)$, y lo llamamos $p_{11}(t)$. Esta entrada está asociada a un autovalor $d_i > 0$. Ahora elegimos una entrada diagonal del bloque matricial 0. Esta entrada está asociada a un autovalor $d_j = 0$. Sean $\mathcal{M}_i = \text{span}\{e_i\}$, $\mathcal{M}_j = \text{span}\{e_j\}$ y P la proyección sobre $\mathcal{M}_i \oplus \mathcal{M}_j$. Esta proyección junto con la proyección ortogonal sobre $(\mathcal{M}_i \oplus \mathcal{M}_j)^\perp$, va a producir el operador Pinching deseado (el cual conmuta con D porque es una matriz diagonal). Usando una matriz de permutación, podemos tener en el primer bloque, la siguiente matriz:

$$Y(t) = \begin{pmatrix} p_{11}(t) & y_{12}(t) \\ y_{21}(t) & 0 \end{pmatrix} \begin{matrix} \mathcal{M}_i \\ \mathcal{M}_j \end{matrix},$$

la cual es minimal y une 0 con $\text{diag}(d_i, 0)$ en $\mathcal{H}(2)$, por el Lema 4.2.3. Pero $\dot{Y}(t) \geq 0$ por el Lema 4.2.5(1), entonces $\dot{y}_{12}(t) = \dot{y}_{21}(t) = 0$, y luego $y_{12}(t) = y_{21}(t) = 0$. Usando este truco podemos poner un cero en cada entrada de $X_{12}(t)$, y cada entrada de $X_{21}(t)$.

3. *Tercer paso:* $X_{23}(t) = X_{32}(t) = 0$. Se deduce del caso anterior como antes, tomar

$$Y(t) = - \begin{pmatrix} 0 & y_{23}(t) \\ y_{32}(t) & n_{22}(t) \end{pmatrix} \begin{matrix} \mathcal{M}_i \\ \mathcal{M}_j \end{matrix}.$$

4. *Cuarto paso:* $X_{13}(t) = X_{31}(t) = 0$. Se deduce de una restricción similar a la anterior, usando en este caso el Lema 4.2.5(3). Eligiendo $e_i \in \mathcal{S}_1$, y $e_j \in \mathcal{S}_3$, podemos obtener un bloque matricial de la forma

$$Y(t) = \begin{pmatrix} p(t) & y_{12}(t) \\ y_{21}(t) & n(t) \end{pmatrix} \begin{matrix} \mathcal{M}_i \\ \mathcal{M}_j \end{matrix}, \quad \text{con } p(t) > 0 \quad \text{y} \quad n(t) < 0,$$

el cual es minimal por el Lema 4.2.3. En este caso $y_{12}(t) = y_{21}(t) = 0$, por el Lema 4.2.5(3). De esta manera podemos poner en cero en cada entrada $X_{13}(t)$, y cada entrada de $X_{31}(t)$. ■

Observación. Usando las hipótesis y la notación del Teorema 4.2.6, no sólo $\dot{P}(t) \geq 0$ y $\dot{N}(t) \leq 0$, si no también $P(t) \geq 0$ y $N(t) \leq 0$: para $t \in [0, 1]$ fijo, sea $X(t) := Y_t$. Entonces

$$\begin{aligned} \|D\|_1 - \|Y_t\|_1 &= \sum_{j=1}^n |d_j| - \left| \lambda_j^\downarrow(Y_t) \right| \leq \sum_{j=1}^n \left| d_j - \lambda_j^\downarrow(Y_t) \right| \leq \\ &\leq \sum_{j=1}^n |\lambda_j(D - Y_t)| = \|D - Y_t\|_1, \end{aligned}$$

usando que $(d_1, \dots, d_n) = \lambda^\downarrow(D)$ y el Teorema de Lidskii, el cual asegura la fórmula de mayorización,

$$\lambda^\downarrow(D) - \lambda^\downarrow(Y_t) \prec \lambda^\downarrow(D - Y_t)$$

(como también la mayorización débil de sus módulos). Como la curva es minimal, la igualdad vale, y

$$|d_j| - \left| \lambda_j^\downarrow(Y_t) \right| = \left| d_j - \lambda_j^\downarrow(Y_t) \right|,$$

es decir, $\lambda_j(Y_t)$ y d_j , deben tener el mismo signo para todo $t \in [0, 1]$. ▲

Una consecuencia de esta caracterización, es la siguiente condición necesaria, que dice que también las curvas que unen a los autovalores, tienen que ser minimales.

Corolario 4.2.7. *Sea $X : [0; 1] \rightarrow \mathcal{H}(n)$ una curva minimal que une 0 con $A \in \mathcal{H}(n)$. Entonces las curvas $\lambda_j(X(t))$, que unen 0 con $\lambda_j(A)$, son curvas minimales (en $\mathbb{R}$).*

Demostración. Por la conjugación unitaria (que preserva todos los autovalores), podemos suponer que, $A = D = \text{diag}(d_1, \dots, d_n)$ es una matriz diagonal como en el Teorema 4.2.6. Como $X(t)$ es una curva minimal, tiene que tener la estructura dada en el Teorema 4.2.6, por lo que los autovalores positivos corresponden a $P(t)$, y los negativos a $N(t)$. Supongamos que $\dim(\mathcal{S}_1) = k$, y fijemos $j \leq k$.

Notar que $L(\lambda_j(P(t))) \geq d_j$, porque la curva une 0 con d_j , y $L(t d_j) = L(p_{jj}(t)) = d_j$, las cuales son minimales por el Corolario 4.2.4. Por otro lado, como $d(P(t)) \prec \lambda(P(t))$,

$$\sum_{j=1}^k p_{jj}(t) = \sum_{j=1}^k \lambda_j(P(t)).$$

Si tomamos derivada y luego integramos, tenemos

$$\sum_{j=1}^k d_j = \sum_{j=1}^k L(\lambda_j(P(t))),$$

es decir, $L(\lambda_j(P(t))) = d_j$, por lo que son minimales. Con cambios menores, se puede probar que, $L(\lambda_j(N(t))) = -d_j$ si $e_j \in \mathcal{S}_3$. ■

4.2.2 El espacio de matrices positivas

Recordar que el espacio de matrices positivas es un espacio de curvatura no positiva por lo cual la función exponencial incrementa distancias en general. Es natural pensar que dada una curva minimal en el espacio de matrices positivas, la curva que se obtiene en el espacio de matrices Hermitianas a partir de tomar logaritmo a dicha curva, es también minimal. Lo que no es natural pensar es que si tomamos una curva minimal en el espacio de matrices Hermitianas obtenemos una curva minimal en el espacio de matrices positivas. Esto último lo podemos lograr gracias a la caracterización que tenemos para curvas minimales en el espacio de matrices Hermitianas.

Convención: En esta subsección vamos a suponer lo siguiente:

1. Las curvas son suaves, es decir, curvas C^1 a trozos.
2. Sea $\mathcal{P}(n)$ el conjunto de matrices positivas $n \times n$.

Dada una curva suave $\alpha : [a, b] \rightarrow \mathcal{P}(n)$, se define la longitud de curva dada por

$$L(\alpha) = \int_a^b \left\| \alpha^{-1/2}(t) \dot{\alpha}(t) \alpha^{-1/2}(t) \right\|_1 dt.$$

La norma que consideramos $\|\cdot\|_1$ es la norma traza de estas matrices.

La distancia rectificable entre $A, B \in \mathcal{P}(n)$ está dada por

$$d_1(A, B) = \inf \{L(\alpha) \text{ donde } \alpha : [a, b] \rightarrow \mathcal{P}(n) \text{ une } A \text{ con } B \text{ en } \mathcal{P}(n)\}.$$

Una curva es **minimal** si su longitud es igual a la distancia de los puntos que une. Es decir,

$$L(\alpha) = d_1(A, B).$$

La fórmula usada para definir la longitud de curva es natural en el siguiente sentido: Observar que el espacio $\mathcal{P}(n)$ es un espacio homogéneo, lo cual se deduce del hecho que, toda matriz $P \in \mathcal{P}(n)$, se puede descomponer como

$$P = GG^* = (U|G|)(U|G|)^* = U|G|^2U^* = |G^*|^2 \quad (4.2.6)$$

donde $G \in \mathcal{G}l(n)$, $|G|, |G^*| \in \mathcal{P}(n)$, $U \in SO(n)$. Las descomposiciones polares a izquierda y derecha $G = U|G| = |G^*|U$ son únicas. Las igualdades matriciales en (4.2.6) nos dan la geometría cociente del cono $\mathcal{P}(n)$. En otras palabras,

$$\mathcal{P}(n) \cong \mathcal{G}l(n)/SO(n), \quad (4.2.7)$$

donde $SO(n) = \{G \in \mathcal{P}(n) : GG^* = I_n\}$, es decir, el espacio de matrices de rotación $n \times n$. En vista de (4.2.6), $\mathcal{G}l(n)$ tiene una acción transitiva en $\mathcal{P}(n)$ vía congruencias

$$P \mapsto GPG^*,$$

donde $G \in \mathcal{G}l(n)$. Toda matriz $P \in \mathcal{P}(n)$ es llevada a la matriz identidad eligiendo $G = P^{-1/2} \in \mathcal{G}l(n)$. Análogamente, todo vector simétrico X tangente a la identidad $I \in \mathcal{P}(n)$, se puede llevar a un vector tangente $A^{1/2}XA^{1/2} \in T_A\mathcal{P}(n)$. En la identidad I , la métrica $\mathcal{G}l(n)$ -invariante en $\mathcal{P}(n)$ está definida como el producto escalar usual

$$\langle X_1, X_2 \rangle_I = Tr(X_1X_2^*) = Tr(X_1X_2).$$

La invariancia de las métricas entonces implica

$$\langle A^{1/2}X_1A^{1/2}, A^{1/2}X_2A^{1/2} \rangle_A = Tr(X_1X_2),$$

la cual sólo se puede satisfacer con la definición

$$\langle D_1, D_2 \rangle_A = Tr(D_1A^{-1}D_2A^{-1}) \quad (4.2.8)$$

para todo $A \in \mathcal{P}(n)$, $D_1, D_2 \in \mathcal{H}(n)$. Entonces existe una métrica $\mathcal{G}l(n)$ -invariante en la variedad

$$\mathcal{P}(n) \cong \mathcal{G}l(n)/SO(n).$$

La invariancia de la métrica (4.2.8) tiene implicaciones directas en la expresión de las geodésicas y la métrica Riemanniana. Luego, dado cualquier intervalo $[a, b] \subset \mathbb{R}$ y una curva suave $\alpha : [a, b] \rightarrow \mathcal{P}(n)$, la longitud de la curva está definida como

$$L(\alpha) = \int_a^b \|\alpha^{-1/2}(t)\dot{\alpha}(t)\alpha^{-1/2}(t)\|_1 dt.$$

donde $\|\cdot\|_1$ es la norma traza de estas matrices.

Recordemos la siguiente fórmula para la derivada (atribuida a varias personas Duhamel, Dyson, Feynman, y Schwingerof) de la función exponencial:

$$De^X(Y) = \int_0^1 e^{tX} Y e^{(1-t)X} dt.$$

También recordamos una demostración simple de la generalizada *infinitesimal exponential metric increasing property* (IEMI). Para probar nuestro resultado vamos a usar el siguiente teorema.

Teorema 4.2.8. *Para la norma traza tenemos*

$$\|A^{1/2}XB^{1/2}\|_1 \leq \left\| \int_0^1 A^t X B^{1-t} dt \right\|_1.$$

Idea de la demostración. Supongamos que $A = B$ y que A es diagonal con entradas $\lambda_1, \dots, \lambda_n$ en su diagonal. La matriz $A^{1/2}XA^{1/2}$ se obtiene de $\int_0^1 A^t X B^{1-t} dt$ de multiplicarla entrada a entrada con la matriz X cuyas entradas son

$$y_{ij} = \frac{\lambda_i^{1/2} \lambda_j^{1/2} (\log(\lambda_i) - \log(\lambda_j))}{\lambda_i - \lambda_j}.$$

Esta matriz es congruente con la matriz de entradas

$$z_{ij} = \frac{\log(\lambda_i) - \log(\lambda_j)}{\lambda_i - \lambda_j}.$$

Notar que esta matriz es positiva, por lo cual, obtenemos el resultado para $A = B$. Por otro lado, es suficiente probar este caso para obtener el caso general. ■

Proposición 4.2.9. *(IEMI generalizada) Para todo H y K en $\mathcal{H}(n)$ tenemos*

$$\|e^{-H/2}De^H(K)e^{-H/2}\|_1 \geq \|K\|_1,$$

para toda norma unitariamente invariante.

Demostración. Como $K = e^{H/2}(e^{-H/2}Ke^{-H/2})e^{H/2}$,

$$\begin{aligned} \|K\|_1 &= \|e^{H/2}(e^{-H/2}Ke^{-H/2})e^{H/2}\|_1 \leq \left\| \int_0^1 e^{tH}(e^{-H/2}Ke^{-H/2})e^{(1-t)H} dt \right\|_1 = \\ &= \left\| e^{-H/2} \left(\int_0^1 e^{tH} K e^{(1-t)H} dt \right) e^{-H/2} \right\|_1 = \|e^{-H/2}De^H(K)e^{-H/2}\|_1. \end{aligned}$$

■

Una consecuencia inmediata de la IEMI generalizada para el caso particular de la norma traza es:

Corolario 4.2.10. *Sea $\alpha : [0; 1] \rightarrow \mathcal{P}(n)$ una curva (suave) que une I con e^X , donde X es una matriz Hermitiana diagonal. En otras palabras, $\alpha(t) = e^{X(t)}$ donde la curva $X : [0, 1] \rightarrow \mathcal{H}(n)$ es tal que: $X(0) = 0$, $X(1) = X$ y X es una matriz Hermitiana diagonal. Luego, si $\alpha(t)$ es una curva minimal entonces $X(t)$ es una curva minimal.*

Demostración. Por regla de la cadena $\dot{\alpha}(t) = De^{X(t)}(\dot{X}(t))$. Entonces la desigualdad

$$L(\alpha) \geq \int_0^1 \|\dot{X}(t)\|_1 dt,$$

se deduce de la definición de

$$L(\alpha) = \int_0^1 \|\alpha^{-1/2}(t)\dot{\alpha}(t)\alpha^{-1/2}(t)\|_1 dt,$$

y la IEMI generalizada. ■

Lo que es interesante es que funciona de la otra manera también: si $X(t)$ es una curva minimal entonces $\alpha(t)$ es una curva minimal. La demostración está basada en la caracterización de la sección previa pero primero necesitamos algunos resultados técnicos.

Sea $I \subset \mathbb{R}$ un intervalo abierto y sea $C^1(I)$ el espacio de funciones reales continuamente diferenciables en I . Para todo $A \in \mathcal{H}_n(I)$ la derivada $Df(A)$ en A es una función lineal de $\mathcal{H}_n$ en sí mismo. Existe una fórmula para la derivada

$$Df(A)(B) = f^{[1]}(A) \circ B, \quad (4.2.9)$$

donde el producto es el de Hadamard (entrada a entrada) y $f^{[1]}$ es la función en $I \times I$ definida como

$$\begin{aligned} f^{[1]}(\lambda, \mu) &= \frac{f(\lambda) - f(\mu)}{\lambda - \mu} & \lambda \neq \mu \\ f^{[1]}(\lambda, \lambda) &= f'(\lambda). \end{aligned}$$

Esto se llama **primer diferencia dividida** de f . Para $A \in \mathcal{H}_n(I)$, $f^{[1]}(A)$ es la matriz $n \times n$

$$f^{[1]}(A) = \frac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j},$$

donde $\lambda_1, \dots, \lambda_n$ son los autovalores de A .

Teorema 4.2.11. *Sea $\alpha : [0; 1] \rightarrow \mathcal{P}(n)$ una curva (suave) que une I con e^X , donde X es una matriz Hermitiana diagonal. En otras palabras, $\alpha(t) = e^{X(t)}$ donde la curva $X : [0, 1] \rightarrow \mathcal{H}(n)$ es tal que: $X(0) = 0$, $X(1) = X$ y X es una matriz Hermitiana diagonal. Luego, si $X(t)$ es una curva minimal entonces $\alpha(t)$ es una curva minimal.*

Demostración. Por regla de la cadena $\dot{\alpha}(t) = De^{X(t)}(\dot{X}(t)) = f^{[1]}(X(t)) \circ \dot{X}(t)$. Ahora elegimos una base ortonormal en la cual $X(t) = \text{diag}(\lambda_1(t), \dots, \lambda_n(t))$. Por la fórmula (4.2.9)

$$De^{X(t)}(\dot{X}(t)) = \exp^{[1]}(X(t)) \circ \dot{X}(t) = \frac{e^{\lambda_i(t)} - e^{\lambda_j(t)}}{\lambda_i(t) - \lambda_j(t)} \dot{X}_{ij}(t).$$

En particular, las entradas diagonales son $e^{\lambda_i(t)} \dot{X}_{ii}(t)$. Luego las entradas diagonales de

$$L = \alpha^{-1/2}(t) \dot{\alpha}(t) \alpha^{-1/2}(t) = e^{-X(t)/2} De^{X(t)}(\dot{X}(t)) e^{-X(t)/2},$$

son

$$L_{ii} = e^{-\lambda_i(t)/2} (e^{\lambda_i(t)} (\dot{X}_{ii}(t))) e^{-\lambda_i(t)/2} = \dot{X}_{ii}(t).$$

Análogamente, si $\lambda_i = \lambda_j$, entonces $L_{ij} = \dot{X}_{ij}(t)$. Si $\lambda_i \neq \lambda_j$ entonces,

$$L_{ij} = \frac{\sinh(\lambda_i(t) - \lambda_j(t))/2}{(\lambda_i(t) - \lambda_j(t))/2} \dot{X}_{ij}(t).$$

Notar que $\dot{P}(t), f^{[1]}(P(t)) \geq 0$ luego $De^{P(t)}(\dot{P}(t)) \geq 0$ por el Teorema de Hadamard. Finalmente

$$\begin{aligned} L(\alpha) &= \int_0^1 \left\| \alpha^{-1/2}(t) \dot{\alpha}(t) \alpha^{-1/2}(t) \right\|_1 dt \\ &= \int_0^1 \left\| e^{-X(t)/2} De^{X(t)}(\dot{X}(t)) e^{-X(t)/2} \right\|_1 dt = \int_0^1 \text{Tr} \left| e^{-X(t)/2} De^{X(t)}(\dot{X}(t)) e^{-X(t)/2} \right| dt \\ &= \int_0^1 \text{Tr} \left| e^{-P(t)/2} De^{P(t)}(\dot{P}(t)) e^{-P(t)/2} \right| dt + \text{Tr} \left| e^{-N(t)/2} De^{N(t)}(\dot{N}(t)) e^{-N(t)/2} \right| dt \\ &= \int_0^1 \sum_{j=1}^n \left| \dot{X}_{jj}(t) \right| dt = \sum_{j=1}^n |\lambda_j| = \|X\|_1. \end{aligned}$$

■

Otra condición necesaria para que una curva sea minimal en este espacio es que las curvas que unen los autovalores sean minimales también.

Corolario 4.2.12. Sea $\alpha : [0; 1] \rightarrow \mathcal{P}(n)$ una curva minimal que une I con $P = e^X \in \mathcal{P}(n)$, donde $X \in \mathcal{H}(n)$. Entonces las curvas que unen $\lambda_j(I)$ con $\lambda_j(P)$, los autovalores respectivos, son curvas minimales (en $\mathbb{R}$).

Demostración. Notar que $\alpha(t)$ es una curva minimal si, y sólo si, $X(t)$ es una curva minimal. Ahora si $X(t)$ es una curva minimal entonces las curvas que unen los autovalores $\lambda_j(t)$ son minimales por el Corolario 4.2.7. Luego las curvas $e^{\lambda_j(t)}$, que son las curvas que unen los autovalores de $\alpha(t)$, son minimales también.

■

4.2.3 El espacio de matrices unitarias

Ahora recordamos que el espacio de matrices unitarias es un espacio de curvatura no negativa, entonces la función exponencial decrece distancias en general. Es natural pensar que si tomamos la exponencial de una curva minimal en el espacio de matrices antihermitianas vamos a obtener una curva minimal en el espacio de matrices unitarias. Podemos usar la caracterización de curvas minimales en el espacio de matrices Hermitianas para construir curvas minimales en el espacio de matrices unitarias.

Convención: En esta subsección vamos a suponer lo siguiente:

1. Las curvas son suaves, es decir, curvas C^1 a trozos.
2. Sea $\mathcal{U}(n)$ el conjunto de matrices unitarias $n \times n$.

Dada una curva suave $\alpha : [a, b] \rightarrow \mathcal{U}(n)$, se define la longitud de curva dada por

$$L(\alpha) = \int_a^b \|\dot{\alpha}(t)\|_1 dt.$$

La norma que consideramos $\|\cdot\|_1$ es la norma traza de estas matrices.

La distancia rectificable entre $U, V \in \mathcal{U}(n)$ está dada por

$$d_1(U, V) = \inf \{L(\alpha) \text{ donde } \alpha : [a, b] \rightarrow \mathcal{U}(n) \text{ une } U \text{ con } V \text{ en } \mathcal{U}(n)\}.$$

Una curva es **minimal** si su longitud es igual a la distancia de los puntos que une. Es decir,

$$L(\alpha) = d_1(U, V).$$

En [33] Larotonda demostró un teorema de las características del siguiente enunciado en un contexto mucho más general. Aquí presentamos una demostración simple para nuestro caso particular de matrices unitarias.

Teorema 4.2.13. *Sea $\alpha : [0; 1] \rightarrow \mathcal{U}(n)$ una curva (suave) que une I con e^{iX} , donde X es una matriz Hermitiana diagonal tal que $\|X\|_{sp} < \pi$. En otras palabras, $\alpha(t) = e^{iX(t)}$ donde $X : [0, 1] \rightarrow \mathcal{H}(n)$ es tal que: $X(0) = 0$, $X(1) = X$ y X es una matriz Hermitiana diagonal tal que $\|X\|_{sp} < \pi$. Luego, si $X(t)$ es una curva minimal entonces $\alpha(t)$ es una curva minimal.*

Demostración. Si $X(t)$ es una curva minimal, como

$$\begin{aligned} L(\alpha) &= \int_0^1 \|\dot{\alpha}(t)\|_1 dt = \int_0^1 \left\| \int_0^1 e^{isX(t)} \dot{X}(t) e^{i(1-s)X(t)} ds \right\|_1 dt \leq \\ &\leq \int_0^1 \int_0^1 \left\| e^{isX(t)} \dot{X}(t) e^{i(1-s)X(t)} \right\|_1 ds dt = \int_0^1 \left\| \dot{X}(t) \right\|_1 dt = L(X(t)), \end{aligned}$$

entonces $\alpha(t)$ es una curva minimal. ■

Otra condición necesaria para que una curva sea minimal en este espacio es que las curvas que unen los autovalores sean minimales también.

Proposición 4.2.14. *Sea $\alpha : [0, 1] \rightarrow \mathcal{U}(n)$ una curva (suave) minimal para la norma traza que une I con $U = e^{iX}$, donde $\|X\|_{sp} < \pi/2$. Sea $Y : [0, 1] \rightarrow \mathcal{H}(n)$ la única curva suave tal que $\alpha(t) = e^{iY(t)}$ y $\|Y(t)\|_{sp} < \pi/2$ para todo $t \in [0, 1]$. Supongamos que las curvas*

$$\lambda_j(\alpha(t)) \stackrel{\text{def}}{=} e^{i\lambda_j(Y(t))}$$

para $t \in [0, 1]$ son suaves a trozos. Entonces

1. Las curvas $\lambda_j(Y(t))$ son monótonas, luego, son curvas minimales en $\mathbb{R}$ para la norma traza.
2. Las curvas $\lambda_j(\alpha(t))$ son minimales en $\mathbb{C}$ para la norma traza.

Demostración. Sea $t \in [0, 1)$ fijo y notemos por $Y_t := Y(t)$. Como la curva α es minimal y $\|Y_t\|_{sp} < \pi/2$,

$$\begin{aligned}
\|X\|_1 &= d_1(I, e^{iX}) = d_1(I, e^{iY_t}) + d_1(e^{iY_t}, e^{iX}) \\
&= \|Y_t\|_1 + d_1(I, e^{-iY_t} e^{iX}) = \|Y_t\|_1 + d_1(I, e^{i(U_t X U_t^* - V_t Y_t V_t^*)}) \\
&= \|Y_t\|_1 + \|U_t X U_t^* - V_t Y_t V_t^*\|_1 \\
&\geq \|Y_t\|_1 + \sum_{j=1}^n \left| \lambda_j^\downarrow(U_t X U_t^*) - \lambda_j^\downarrow(V_t Y_t V_t^*) \right| \\
&= \|Y_t\|_1 + \sum_{j=1}^n \left| \lambda_j^\downarrow(X) - \lambda_j^\downarrow(Y_t) \right| \\
&\geq \|Y_t\|_1 + \sum_{j=1}^n \left| \lambda_j^\downarrow(X) \right| - \left| \lambda_j^\downarrow(Y_t) \right| = \|X\|_1.
\end{aligned}$$

La primer desigualdad se deduce del Teorema de Lidskii y la convexidad que surge de tomar módulo. En conclusión,

$$\left| \lambda_j^\downarrow(X) - \lambda_j^\downarrow(Y_t) \right| = \left| \lambda_j^\downarrow(X) \right| - \left| \lambda_j^\downarrow(Y_t) \right| \quad \text{para todo } 1 \leq j \leq n.$$

Usando la misma cuenta reemplazando X por Y_s para todo $s \in (t, 1]$ (por la minimalidad de α y el hecho que $\|Y_s\|_{sp} < \pi/2$, se tiene que $d(I, \alpha(s)) = \|Y_s\|_1$), se deduce que las funciones

$$t \mapsto \left| \lambda_j^\downarrow(Y_t) \right|$$

son crecientes y que $\lambda_j(X), \lambda_j(Y_t) \geq 0$ (tienen el mismo signo) para todo $1 \leq j \leq n$ y todo $t \in [0, 1]$. Luego las funciones

$$t \mapsto \lambda_j^\downarrow(Y_t)$$

son monótonas y minimales en $\mathbb{R}$ y también las curvas $\lambda_j(\alpha(t)) = e^{i\lambda_j^\downarrow(Y_t)}$ son minimales en $\mathbb{C}$ para la norma traza (la segunda parte se deduce de usar el Teorema 4.2.13 con $n = 1$). ■

4.3 Geometría de los puntos medios en $\mathcal{P}(n)$ y $\mathcal{H}(n)$

Sean $P \in \mathcal{P}(n)$ y $t \in (0, 1)$. En esta subsección vamos a estudiar el conjunto de puntos medios:

$$\mathcal{M}_t(I, P) = \{W \in \mathcal{P}(n) : d_{\|\cdot\|}(I, W) = t d_{\|\cdot\|}(I, P) \text{ y } d_{\|\cdot\|}(W, P) = (1 - t) d_{\|\cdot\|}(I, P)\}.$$

Recordamos que, para una norma unitariamente invariante, la longitud de curva se define por medio de

$$L_{\|\cdot\|}(\alpha) = \int_a^b \|\alpha^{-1/2}(t)\dot{\alpha}(t)\alpha^{-1/2}(t)\| dt.$$

La distancia rectificable asociada se define como,

$$d_{\|\cdot\|}(A, B) = \inf \{L_{\|\cdot\|}(\alpha) \text{ donde } \alpha : [a, b] \rightarrow \mathcal{P}(n) \text{ une } A \text{ con } B \text{ en } \mathcal{P}(n)\}, \quad (4.3.1)$$

la cual es una métrica en $\mathcal{P}(n)$ invariante por congruencias. El conjunto de puntos medios correspondiente a la métrica de Thompson, es geodésicamente convexo (ver [40]). A continuación probaremos que este conjunto es geodésicamente convexo para toda norma unitariamente invariante. La demostración pasa por probar que la función

$$s \mapsto d_{\|\cdot\|}(I, \gamma(s)),$$

donde $\gamma(s)$ es la geodésica que une dos matrices positivas, es convexa. Luego, usaremos esta propiedad para probar el resultado que buscamos.

Teorema 4.3.1. *Si $\gamma(s) = A^{1/2}(A^{-1/2}BA^{-1/2})^sA^{1/2}$ denota la geodésica que une $A, B \in \mathcal{P}(n)$, entonces*

$$s \mapsto d_{\|\cdot\|}(I, \gamma(s))$$

es convexa.

Demostración. En efecto,

$$\begin{aligned} d_{\|\cdot\|}(I, \gamma(s)) &\leq d_{\|\cdot\|}(I, A^{1-s}) + d_{\|\cdot\|}(A^{1-s}, \gamma(s)) = d_{\|\cdot\|}(I, A^{1-s}) + d_{\|\cdot\|}(A^{-s}, (A^{-1/2}BA^{-1/2})^s) \leq \\ &\leq (1-s)d_{\|\cdot\|}(I, A) + sd_{\|\cdot\|}(A^{-1}, A^{-1/2}BA^{-1/2}) = (1-s)d_{\|\cdot\|}(I, A) + sd_{\|\cdot\|}(I, B). \end{aligned}$$

■

Proposición 4.3.2. *Sean $P \in \mathcal{P}(n)$ y $t \in (0, 1)$. Entonces el conjunto*

$$\mathcal{M}_t(I, P) = \{W \in \mathcal{P}(n) : d_{\|\cdot\|}(I, W) = t d_{\|\cdot\|}(I, P) \text{ y } d_{\|\cdot\|}(W, P) = (1-t) d_{\|\cdot\|}(I, P)\}$$

es geodésicamente convexo.

Demostración. Observar que si $\gamma(s)$ es la geodésica que une $W_0, W_1 \in \mathcal{M}_t(U, V)$, es decir, $W_0 = \gamma(0), W_1 = \gamma(1)$ y $\gamma(s) = W_0^{1/2}(W_0^{-1/2}W_1W_0^{-1/2})^sW_0^{1/2}$, entonces

$$\begin{aligned} d_{\|\cdot\|}(I, \gamma(s)) &\leq (1-s)d_{\|\cdot\|}(I, W_0) + sd_{\|\cdot\|}(I, W_1) \\ &= (1-s)td_{\|\cdot\|}(I, P) + std_{\|\cdot\|}(I, P) = td_{\|\cdot\|}(I, P); \\ d_{\|\cdot\|}(P, \gamma(s)) &\leq (1-s)d_{\|\cdot\|}(P, W_0) + sd_{\|\cdot\|}(P, W_1) \\ &= (1-s)(1-t)d_{\|\cdot\|}(I, P) + s(1-t)d_{\|\cdot\|}(I, P) = (1-t)d_{\|\cdot\|}(I, P) \end{aligned}$$

Por lo tanto

$$d_{\|\cdot\|}(I, P) \leq d_{\|\cdot\|}(I, \gamma(s)) + d_{\|\cdot\|}(P, \gamma(s)) \leq d_{\|\cdot\|}(I, P).$$

De lo cual deducimos que,

$$\begin{aligned} d_{\|\cdot\|}(I, \gamma(s)) &= td_{\|\cdot\|}(I, P) \\ d_{\|\cdot\|}(P, \gamma(s)) &= (1-t)d_{\|\cdot\|}(I, P). \end{aligned}$$

Luego $\gamma(s) \in \mathcal{M}_t(I, P)$ para todo $s \in [0, 1]$.

■

Utilizando las mismas ideas vamos a probar por completitud que, en $\mathcal{H}(n)$,

$$\mathcal{M}_t(0, X) = \{W \in \mathcal{H}(n) : d_{\|\cdot\|}(0, W) = t d_{\|\cdot\|}(0, X) \text{ and } d_{\|\cdot\|}(W, X) = (1 - t) d_{\|\cdot\|}(0, X)\},$$

es geodésicamente convexo. Este caso puede ser menos interesante, o más sencillo de probar, pues las geodésicas son segmentos.

Teorema 4.3.3. *Si $\gamma(s) = (1 - s)A + sB$ denota la geodésica que une $A, B \in \mathcal{H}(n)$, entonces*

$$s \mapsto d_{\|\cdot\|}(0, \gamma(s))$$

es convexa.

Demostración. En efecto,

$$\begin{aligned} d_{\|\cdot\|}(0, \gamma(s)) &\leq d_{\|\cdot\|}(0, (1 - s)A) + d_{\|\cdot\|}((1 - s)A, \gamma(s)) = (1 - s)d_{\|\cdot\|}(0, A) + d_{\|\cdot\|}(0, sB) = \\ &= (1 - s)d_{\|\cdot\|}(0, A) + sd_{\|\cdot\|}(0, B). \end{aligned}$$

■

Proposición 4.3.4. *Sean $X \in \mathcal{H}(n)$ y $t \in (0, 1)$. Entonces el conjunto*

$$\mathcal{M}_t(0, X) = \{W \in \mathcal{H}(n) : d_{\|\cdot\|}(0, W) = t d_{\|\cdot\|}(0, X) \text{ and } d_{\|\cdot\|}(W, X) = (1 - t) d_{\|\cdot\|}(0, X)\}$$

es geodésicamente convexo.

Demostración. Haciendo los cambios necesarios, la prueba es igual a la dada en la Proposición 4.3.2. ■

Observación. Es importante notar que si la norma $\|\cdot\|$ es estrictamente convexa entonces existe una única geodésica que une dos puntos dados. En consecuencia, el conjunto de puntos medios consta de un solo elemento, razón por la cual no tiene sentido preguntarse sobre la convexidad geodésica.

5 | Estructura de las curvas minimales en $\mathcal{U}(n)$ con respecto a la métrica de Finsler bi-invariante inducida por la norma espectral

5.1 El grupo de un parámetro

Si consideramos $\mathcal{U}(n) \subset L(\mathcal{H})$ como subvariedad, resulta natural, considerar la estructura asociada a la norma de operadores. La métrica de Finsler bi-invariante dada por la norma espectral es la siguiente,

$$\|X\|_U = \|U^*X\|_{sp} = \|X\|_{sp},$$

para todo X tangente a un operador unitario U . Con respecto a esta métrica de Finsler, la longitud de la curva $\alpha : [0, 1] \rightarrow \mathcal{U}(n)$, se puede calcular de la siguiente manera

$$L(\alpha) = \int_0^1 \|\dot{\alpha}(t)\|_{sp} dt.$$

La distancia rectificable inducida, denotada por $d_\infty(\cdot, \cdot)$, se calcula como en el caso Riemanniano, como el ínfimo de las longitudes de todas las curvas suaves a trozos que unen dos puntos dados.

$$d_\infty(U, V) = \inf \{L(\alpha) \text{ donde } \alpha : [a, b] \rightarrow \mathcal{U}(n) \text{ une } U \text{ con } V \text{ en } \mathcal{U}(n)\}.$$

Sean $U, V \in \mathcal{U}(n)$ donde $V = Ue^{iX}$ con $X \in \mathcal{H}(n)$ y $\|X\|_{sp} \leq \pi$. Entonces la curva $\delta(t) = Ue^{itX}$ es más corta que cualquier otra curva suave a trozos γ en $\mathcal{U}(n)$ que una U con V . Es decir

$$L(\delta) = d_\infty(U, V).$$

En efecto, la idea de la demostración utilizada en [5], involucra los siguientes resultados.

Teorema. (Thompson [60]) *Dados $X, Y \in \mathcal{H}(n)$, existen matrices unitarias U, V tales que*

$$e^{iX}e^{iY} = e^{i(UXU^* + VYV^*)}.$$

Corolario. *Sean $X, Y, Z \in \mathcal{H}(n)$ tales que $\|Z\|_{sp} \leq \pi$, $e^{iX}e^{iY} = e^{iZ}$. Entonces existen matrices unitarias U, V tales que*

$$|Z| \leq |UXU^* + VYV^*|.$$

Teorema. (Thompson [[58], [59]]). Dados $X, Y \in M_n(\mathbb{C})$, existen matrices unitarias U, V tales que

$$|X + Y| \leq U |X| U^* + V |Y| V^*.$$

El teorema demostrado en [5] prueba la existencia de curvas cortas con respecto a la métrica de Finsler para toda norma unitariamente invariante en el espacio de matrices unitarias. Esta generalidad nos va a resultar conveniente en la motivación del siguiente problema.

Teorema: Sean $U, V \in \mathcal{U}(n)$ y $V = Ue^{iX}$, con $\|X\|_{sp} \leq \pi$, $X \in \mathcal{H}(n)$. Entonces, la curva $\delta(t) = Ue^{itX}$ es más corta o igual que cualquier otra curva suave a trozos γ que una U con V en $\mathcal{U}(n)$. En particular, $d(U, V) = \|X\|$.

Idea de la demostración: Sin perder generalidad podemos suponer que $U = Id$. Primero se reduce a estudiar una curva poligonal vía un argumento de aproximación estándar.

Luego, vía un argumento inductivo, se reduce a estudiar la desigualdad triangular.

Sabemos que:

- $e^{iX} = e^{iY}e^{iZ}$;
- $\|X\|_{sp} \leq \pi$;

y queremos probar que: $\|X\| \leq \|Y\| + \|Z\|$. Y esto se deduce de:

- Por el Teorema de Thompson $e^{iX} = e^{iY}e^{iZ} = e^{i(UYU^* + VZV^*)}$, $U, V \in \mathcal{U}(n)$;
- Como $\|X\|_{sp} \leq \pi$ se tiene que $\|X\| \leq \|UYU^* + VZV^*\|$.

Sin embargo es fácil construir ejemplos para probar que no es la única si la norma no es estrictamente convexa.

Ejemplo. Sean

$$U = e^{iX} = \begin{pmatrix} e^{i\frac{\pi}{2}} & 0 \\ 0 & e^{i\frac{\pi}{4}} \end{pmatrix}, \quad X = \begin{pmatrix} \frac{\pi}{2} & 0 \\ 0 & \frac{\pi}{4} \end{pmatrix}.$$

Entonces la curva

$$\alpha(t) = \begin{pmatrix} e^{it\frac{\pi}{2}} & 0 \\ 0 & \tilde{\alpha}(t) \end{pmatrix}, \quad \tilde{\alpha}(t) = \begin{cases} e^{i\frac{\pi}{2}t} & t \in [0, 1/4] \\ e^{i\frac{\pi}{2}(\frac{1}{2}-t)} & t \in [1/4, 1/2] \\ e^{i\frac{\pi}{2}(t-\frac{1}{2})} & t \in [1/2, 1] \end{cases}$$

une Id con U y es minimal para la distancia d_∞ de la norma $\|\cdot\|_{sp}$.

Con este problema en vista y los resultados obtenidos en la sección anterior nos preguntamos

- Sean $U, V \in \mathcal{U}(n)$ ¿Podemos caracterizar las curvas minimales que los unen con respecto a la métrica heredada por la norma espectral?
- Definiendo el conjunto de puntos medios entre U y V ¿Es posible probar que bajo ciertas condiciones este conjunto es geodésicamente convexo?
- ¿Se puede caracterizar la unicidad de curvas minimales en términos del espectro de U y V ?

Vamos a responder estas preguntas a lo largo del capítulo en los siguientes resultados: Teorema 5.3.1, Corolario 5.3.2 y Proposición 5.4.3, respectivamente.

5.2 El grupo de matrices unitarias

El grupo de matrices $n \times n$ complejas unitarias $\mathcal{U}(n)$, como todo grupo de Lie, tiene una conexión canónica libre de torsión sobre campos vectoriales invariantes a izquierda X, Y definida como $\nabla_X Y = \frac{1}{2}[X, Y]$. La geodésica asociada a esta conexión son los grupos de un parámetro $t \mapsto Ue^{tZ}$ (donde U es una matriz unitaria y Z una matriz anti-Hermitiana). Podemos introducir una métrica Riemanniana en el grupo unitario de manera estándar

$$\langle X, Y \rangle_g = \text{Tr}(U^* X (U^* Y)^*) = \text{Tr}(XY^*),$$

para $U^* X, U^* Y$ en el álgebra de Lie del grupo, es decir, para $U^* X, U^* Y$ matrices anti-Hermitianas. Es conocido que la conexión mencionada, es la conexión de Levi-Civita de la métrica g inducida por la traza, y que las geodésicas son curvas minimales siempre que el espectro de Z esté contenido en $(-\pi, \pi)$.

Ahora consideremos la métrica de Finsler bi-invariante dada por la norma espectral,

$$\|X\|_U = \|U^* X\|_{sp} = \|X\|_{sp},$$

para todo X tangente a un operador unitario U . Con respecto a esta métrica de Finsler, la longitud de la curva $\alpha : [0, 1] \rightarrow \mathcal{U}(n)$, se puede calcular de la siguiente manera

$$L(\alpha) = \int_0^1 \|\dot{\alpha}(t)\|_{sp} dt.$$

La distancia rectificable inducida, denotada por $d_\infty(\cdot, \cdot)$, se calcula como en el caso Riemanniano, como el ínfimo de las longitudes de todas las curvas suaves a trozos que unen dos puntos dados.

Con respecto a esta estructura métrica, las geodésicas de la conexión son curvas minimales (ver [4]). Sin embargo, como la norma espectral no es estrictamente convexa, en la mayoría de los casos los grupos de un parámetro no son las únicas curvas minimales que conectan dos puntos dados. Luego, una pregunta natural es: si es posible o no, caracterizar todas las curvas minimales que unen dos puntos en $\mathcal{U}(n)$. El objetivo principal de este capítulo es dar tal caracterización. Por otro lado, esto lleva a una caracterización de aquellos pares de matrices unitarias conectadas por una única curva minimal. Más precisamente, la geodésica es la única curva minimal que une $U, V \in \mathcal{U}(n)$ si, y sólo si, el espectro de U^*V está contenido en $\{e^{i\theta}, e^{-i\theta}\}$ para algún $\theta \in [0, \pi)$. Un resultado análogo, pero en el conjunto de matrices estrictamente positivas, lo obtuvo Yongdo Lim en [40]. A su vez, la descripción de todas las curvas minimales que unen dos puntos nos lleva a considerar el conjunto de puntos medios

$$\mathcal{M}_{1/2}(U, V) = \{W \in \mathcal{U}(n) : d_\infty(U, W) = d_\infty(W, V) = \frac{1}{2} d_\infty(U, V)\},$$

o más generalmente para $t \in (0, 1)$

$$\mathcal{M}_t(U, V) = \{W \in \mathcal{U}(n) : d_\infty(U, W) = t d_\infty(U, V) \text{ and } d_\infty(W, V) = (1 - t) d_\infty(U, V)\}.$$

vamos a probar que este conjunto es geodésicamente convexo siempre que $d_\infty(U, V) < \pi/2$.

5.3 Estructura de las curvas minimales en $\mathcal{U}(n)$

En esta sección, vamos a estudiar la estructura de las curvas minimales que unen dos matrices unitarias U y V dadas, donde la minimalidad es con respecto a la estructura métrica de Finsler dada por la norma espectral.

Notación: Por $\mathcal{M}_n(\mathbb{C})$ vamos a denotar al álgebra de matrices complejas $n \times n$. A veces vamos a usar $\mathcal{GL}(\mathcal{S})$ (resp. $\mathcal{U}(\mathcal{S})$, $\mathcal{H}(\mathcal{S})$) para indicar que los operadores actúan en algún subespacio específico $\mathcal{S}$. Si $T \in \mathcal{M}_n(\mathbb{C})$, entonces $|T|$ indica el módulo de T , es decir, $|T| = \sqrt{T^*T}$, y $\text{tr}(T)$ denota la traza de T . Por medio de $\sigma(T)$ denotamos al conjunto de autovalores de T , mientras que $\rho(T)$ denota el radio espectral de T . Dada $A \in \mathcal{H}(n)$, $\lambda_1(A) \geq \dots \geq \lambda_n(A)$ denota los autovalores de A ordenados de manera no creciente y contados con multiplicidad. Análogamente, dada una matriz arbitraria $T \in \mathcal{M}_n(\mathbb{C})$, $s_1(T) \geq \dots \geq s_n(T)$ denota los valores singulares de T (también contados con multiplicidad), es decir, los autovalores de $|T|$. Finalmente, dadas $A, B \in \mathcal{H}(n)$, por medio de $A \leq B$ denotamos que A es menor o igual que B con respecto al orden de Löwner.

Convención: de ahora en adelante, vamos a suponer que las curvas están parametrizadas por el intervalo $[0, 1]$, y de tal manera que $t \mapsto \|\dot{\alpha}(t)\|$ es constante. Esto nos va a evitar problemas de notación o pérdida de unicidad por posibles reparametrizaciones.

Recordar que, en el caso de matrices Hermitianas, la norma espectral coincide con el radio espectral. Por lo tanto, *hablando mal y pronto*, podemos cambiar una matriz Hermitiana dentro del autoespacio correspondiente a los “autovalores más chicos” y la matriz todavía tendría la misma norma espectral. Esta simple observación, nos da una estrategia fácil para construir curvas minimales, por perturbaciones del grupo de un parámetro.

El siguiente resultado nos dice que precisamente estas son todas las curvas minimales. En busca de simplicidad, y sin pérdida de generalidad (pues la norma espectral es unitariamente invariante), vamos a suponer que una de las matrices es la identidad.

Teorema 5.3.1. *Dado $U \in \mathcal{U}(n)$, sea $X \in \mathcal{H}(n)$ tal que $U = e^{iX}$ y $\|X\|_{sp} \leq \pi$. Entonces:*

- a) *Si $\sigma(|X|) \subseteq [0, \pi)$ y tiene más de un elemento, entonces una curva minimal que une I con U tiene la siguiente estructura:*

$$\alpha(t) = \begin{pmatrix} e^{itX_S} & 0 \\ 0 & \alpha_1(t) \end{pmatrix} \begin{matrix} \mathcal{S} \\ \mathcal{S}^\perp \end{matrix},$$

donde $\mathcal{S} = \ker(\|X\|_{sp}I - |X|)$, $X_S = X|_{\mathcal{S}} \in \mathcal{H}(\mathcal{S})$, y $\alpha_1 : [0, 1] \rightarrow \mathcal{U}(\mathcal{S}^\perp)$ es cualquier curva que une $I_{\mathcal{S}^\perp}$ con $U_{\mathcal{S}^\perp} = U|_{\mathcal{S}^\perp}$ tal que $\|\dot{\alpha}_1\|_{sp} \leq \|X\|_{sp}$.

- b) *Si $\pi \in \sigma(|X|)$, entonces una curva minimal que une I con U tiene la siguiente estructura:*

$$\alpha(t) = \begin{pmatrix} e^{itY_S} & 0 \\ 0 & \alpha_1(t) \end{pmatrix} \begin{matrix} \mathcal{S} \\ \mathcal{S}^\perp \end{matrix},$$

donde $\mathcal{S} = \ker(\pi I - |X|)$, $Y_S \in \mathcal{H}(\mathcal{S})$ y satisface que $\sigma(Y_S) \subseteq \{\pi, -\pi\}$, y $\alpha_1 : [0, 1] \rightarrow \mathcal{U}(\mathcal{S}^\perp)$ es cualquier curva que une $I_{\mathcal{S}^\perp}$ con $U_{\mathcal{S}^\perp} = U|_{\mathcal{S}^\perp}$ tal que $\|\dot{\alpha}_1\|_{sp} \leq \pi$.

Una consecuencia directa del primer ítem de este resultado es el siguiente corolario sobre unicidad de curvas minimales:

Corolario 5.3.2. *Dados $U, V \in \mathcal{U}(n)$, existe una única curva minimal que los une si, y sólo si, $\sigma(U^*V) \subseteq \{e^{i\theta}, e^{-i\theta}\}$ para algún $|\theta| < \pi$.*

Otra consecuencia de este teorema es el siguiente resultado:

Corolario 5.3.3. *Dada una curva minimal $\alpha : [0, 1] \rightarrow \mathcal{U}(n)$, también es minimal entre dos puntos cualesquiera de su traza, y*

$$d_\infty(U, \alpha(r)) = rd_\infty(U, V).$$

5.3.1 Demostración del Teorema 5.3.1

Para comenzar, vamos a probar el siguiente lema, el cual es una pequeña modificación de un hermoso truco usado por Porta y Recht en [54] (ver también [5]):

Lema 5.3.4. *Sea $\alpha : [0, 1] \rightarrow \mathcal{U}(n)$ una curva minimal (suave) que une I con $U = e^{iX}$, donde $\|X\|_{sp} < \pi$. Si ξ es un autovector unitario de X correspondiente a un autovalor λ tal que $\rho(X) = |\lambda|$. Entonces*

$$\alpha(t) = \begin{pmatrix} e^{it\lambda} & 0 \\ 0 & \alpha_1(t) \end{pmatrix} \begin{matrix} \mathcal{S} \\ \mathcal{S}^\perp \end{matrix},$$

donde $\mathcal{S} = \text{span}\{\xi\}$, y $\alpha_1 : [0, 1] \rightarrow \mathcal{U}(\mathcal{S}^\perp)$ es cualquier curva que une $I_{\mathcal{S}^\perp}$ con $U_{\mathcal{S}^\perp} = U|_{\mathcal{S}^\perp}$ tal que $\|\dot{\alpha}_1\|_{sp} \leq \|X\|_{sp}$.

Demostración. Sea $\gamma : [0, 1] \rightarrow \mathcal{U}(n)$ la curva definida por $\gamma(t) = e^{itX}$, y sea S^{2n-1} la esfera:

$$S^{2n-1} = \{\eta \in \mathbb{C}^n : \|\eta\|_2 = 1\}.$$

Definimos las curvas $\tilde{\gamma} : [0, 1] \rightarrow S^{2n-1}$ y $\tilde{\alpha} : [0, 1] \rightarrow S^{2n-1}$ de la siguiente manera:

$$\tilde{\gamma}(t) = \gamma(t)\xi \quad \text{y} \quad \tilde{\alpha}(t) = \alpha(t)\xi.$$

Un cálculo sencillo nos prueba que $L(\tilde{\gamma}) = L(\tilde{\alpha})$, donde la longitud aquí se computa con respecto a la estructura Riemanniana de la esfera. Como $\|X\|_{sp} < \pi$, $\tilde{\gamma}$ es la única geodésica en S^{2n-1} que une ξ con $U\xi = e^{iX}\xi$. Por lo tanto, $\tilde{\gamma} = \tilde{\alpha}$. En particular, ξ es un autovector de $\alpha(t)$ asociado a $e^{it\lambda}$ para todo $t \in [0, 1]$. El resto del enunciado es una consecuencia de la descomposición por bloques de α , inducida por la descomposición del espacio $\mathbb{C}^n = \mathcal{S} \oplus \mathcal{S}^\perp$. ■

Iterando este lema obtenemos la primera parte del Teorema 5.3.1.

Corolario 5.3.5. Sea $\alpha : [0, 1] \rightarrow \mathcal{U}(n)$ una curva minimal (suave) que une I con $U = e^{iX}$, donde $\|X\|_{sp} < \pi$. Entonces

$$\alpha(t) = \begin{pmatrix} e^{itX_S} & 0 \\ 0 & \alpha_1(t) \end{pmatrix} \begin{matrix} \mathcal{S} \\ \mathcal{S}^\perp \end{matrix},$$

donde $\mathcal{S} = \ker(\|X\|_{sp}I - |X|)$, $X_S = X|_{\mathcal{S}}$, y $\alpha_1 : [0, 1] \rightarrow \mathcal{U}(\mathcal{S}^\perp)$ es una curva que une $I_{\mathcal{S}^\perp}$ con $U_{\mathcal{S}^\perp} = U|_{\mathcal{S}^\perp}$ tal que $\|\dot{\alpha}_1\|_{sp} \leq \|X\|_{sp}$.

Para la segunda parte, vamos a usar un argumento por aproximación.

Lema 5.3.6. Sea $\alpha : [0, 1] \rightarrow \mathcal{U}(n)$ una curva minimal (suave) que une I con $U = e^{iX}$, donde $\|X\|_{sp} = \pi$. Entonces existe $Y \in \mathcal{H}(n)$, con $\|Y\|_{sp} = \pi$, tal que $U = e^{iY}$ y un autovector unitario ξ de Y correspondiente a un autovalor λ con $|\lambda| = \pi$, tal que

$$\alpha(t) = \begin{pmatrix} e^{it\lambda} & 0 \\ 0 & \alpha_1(t) \end{pmatrix} \begin{matrix} \mathcal{S} \\ \mathcal{S}^\perp \end{matrix},$$

donde $\mathcal{S} = \text{span}\{\xi\}$, y $\alpha_1 : [0, 1] \rightarrow \mathcal{U}(\mathcal{S}^\perp)$ es cualquier curva que une $I_{\mathcal{S}^\perp}$ y $U_{\mathcal{S}^\perp} = U|_{\mathcal{S}^\perp}$ tal que $\|\dot{\alpha}_1\|_{sp} \leq \|X\|_{sp}$.

Demostración. Para todo $m \in \mathbb{N}$ vamos a considerar la curva

$$\alpha_m(t) = \alpha\left(\left(\frac{m}{m+1}\right)t\right), \quad \text{para todo } t \in [0, 1].$$

En otras palabras, $\alpha_m(t)$ es la parte de la curva $\alpha(t)$, que une I con $\alpha_m(1) = e^{iX_m}$ para algún $X_m \in \mathcal{H}(n)$ con $\|X_m\|_{sp} < \pi$. Como $\alpha(t)$ tiene velocidad constante, se tiene que

$$\|X_m\|_{sp} = d_\infty(I, e^{iX_m}) = \frac{m}{m+1}L(\alpha) < \pi.$$

Luego, $\alpha_m(t)$ es una curva minimal que une I con e^{iX_m} . Más aún,

$$e^{iX_m} = \alpha\left(\frac{m}{m+1}\right) \xrightarrow{m \rightarrow \infty} \alpha(1) = e^{iX}.$$

Para $\{X_m\}_{m \in \mathbb{N}}$ existe una subsucesión convergente, que denotamos por X_m con algún abuso de notación, es decir,

$$X_m \xrightarrow{m \rightarrow \infty} Y \in \mathcal{H}(n) \implies U = e^{iX} = e^{iY} . \quad (5.3.1)$$

Denotamos por $U_m = e^{iX_m}$ para $m \in \mathbb{N}$, y

$$\eta := \min \{ |\lambda_j(U) - \lambda_k(U)| : \text{donde } \lambda_j(U) \neq \lambda_k(U) \} > 0 .$$

Indexemos por $j = 0, \dots, r$ a los distintos autovalores de U . Para todo $j = 0, \dots, r$, consideramos el disco D_j , con centro $\lambda_j(U)$ y radio $\eta/3$. Por (5.3.1) y el hecho que $\sigma(U) \subseteq \bigcup_j D_j$, existe un $N_1 \geq 1$ tal que

$$\sigma(U_m) \subseteq \bigcup_j D_j \quad \text{para todo } m \geq N_1 .$$

Observar que como $\|X\|_{sp} = \pi$ entonces $-1 \in \lambda(U)$. Sea $D_0 = \overline{D}(-1, \eta/3)$ y tomemos una función continua $f : \mathbb{T} \rightarrow [0, 1]$ tal que, $f|_{D_0} = 1$, y $f|_{D_k} = 0$ para $k \neq 0$, $k = 1, \dots, r$. Entonces $f(U_m) \xrightarrow{m \rightarrow \infty} f(U)$ por el Teorema espectral. En otras palabras,

$$P_{\mathcal{T}_m} \xrightarrow{m \rightarrow \infty} P_{\mathcal{T}} , \quad (5.3.2)$$

donde $P_{\mathcal{T}} = f(U)$ es la proyección sobre el subespacio $\mathcal{T} = \ker(U + I)$ y $P_{\mathcal{T}_m} = f(U_m)$ es la proyección sobre el subespacio

$$\mathcal{T}_m = \text{span} \{ v \in \mathbb{C}^n : U_m v = \lambda_k(U_m) v \quad \text{para algún } \lambda_k(U_m) \in D_0 \} .$$

Para todo $m \in \mathbb{N}$, sea $\lambda_m \in \sigma(X_m)$, tal que $|\lambda_m| = \|X_m\|$. Como $\lambda_m \in [-\pi, \pi]$ existe una subsucesión convergente, que denotamos por λ_m , con algún abuso de notación nuevamente, es decir,

$$\lambda_m \xrightarrow{m \rightarrow \infty} \lambda ,$$

donde $|\lambda| = \pi$. Como $\|X_m\| < \pi$ podemos aplicar el Lema 5.3.4 a la curva $\alpha_m(t)$. Para todo $m \in \mathbb{N}$, sea $\xi_m \in \ker(X_m - \lambda_m I)$ un vector unitario, y denotamos por $\mathcal{S}_m = \text{span}\{\xi_m\}$. Entonces

$$\alpha_m(t) = \begin{pmatrix} e^{it\lambda_m} & 0 \\ 0 & \alpha_{m,1}(t) \end{pmatrix} \begin{pmatrix} \mathcal{S}_m \\ \mathcal{S}_m^\perp \end{pmatrix} ,$$

donde $\alpha_{m,1} : [0, 1] \rightarrow \mathcal{U}(\mathcal{S}_m^\perp)$, es una curva que une $I_{\mathcal{S}_m^\perp}$ con $U_{m, \mathcal{S}_m^\perp}$, tal que $\|\dot{\alpha}_{m,1}\|_{sp} \leq \|X_m\|_{sp}$. Como $\xi_m \in \mathcal{S}_m \subseteq \mathcal{T}_m$ y $P_{\mathcal{T}_m} \xrightarrow{m \rightarrow \infty} P_{\mathcal{T}}$, Concluimos que existe un vector unitario $\xi \in \mathcal{T}$, y una subsucesión convergente tal que

$$\xi_{j_m} \xrightarrow{m \rightarrow \infty} \xi \implies P_{\mathcal{S}_{j_m}} \xrightarrow{m \rightarrow \infty} P_{\mathcal{S}} ,$$

donde $\mathcal{S} = \text{span}\{\xi\}$. Como $X_{j_m} \xrightarrow{m \rightarrow \infty} Y$ y $\lambda_{j_m} \xrightarrow{m \rightarrow \infty} \lambda$, luego $Y\xi = \lambda\xi$. Entonces

$$\langle e^{it\lambda_{j_m}} \xi_{j_m}, \xi_{j_m} \rangle \xrightarrow{m \rightarrow \infty} \langle e^{it\lambda} \xi, \xi \rangle \quad \text{para todo } t \in [0, 1] .$$

Como también $\alpha_{j_m}(t) \xrightarrow{m \rightarrow \infty} \alpha(t)$ para todo $t \in [0, 1]$, deducimos que

$$\alpha_{j_m}(t) = \begin{pmatrix} e^{it\lambda_{j_m}} & 0 \\ 0 & \alpha_{j_m,1}(t) \end{pmatrix} \begin{pmatrix} \mathcal{S}_{j_m} \\ \mathcal{S}_{j_m}^\perp \end{pmatrix} \xrightarrow{m \rightarrow \infty} \alpha(t) = \begin{pmatrix} e^{it\lambda} & 0 \\ 0 & \alpha_1(t) \end{pmatrix} \begin{pmatrix} \mathcal{S} \\ \mathcal{S}^\perp \end{pmatrix},$$

donde $\alpha_1 = P_{\mathcal{S}^\perp} \alpha P_{\mathcal{S}^\perp}|_{\mathcal{S}^\perp} : [0, 1] \rightarrow \mathcal{U}(\mathcal{S}^\perp)$ es una curva que une $I_{\mathcal{S}^\perp}$ con $U_{\mathcal{S}^\perp} = U|_{\mathcal{S}^\perp}$ tal que $\|\dot{\alpha}_1\|_{sp} \leq \|\dot{\alpha}\|_{sp} = \|X\|_{sp}$. ■

Para completar la demostración del Teorema 5.3.1 parte b) vamos a iterar este lema y, en cada iteración, vamos a obtener una nueva Y que puede ser diferente de X . La razón de esto es que no hay unicidad en las geodésicas que unen puntos antipodales en la esfera.

Demostración del Teorema 5.3.1. La primera parte ya fue probada en el Corolario 5.3.5. La segunda parte es una consecuencia del Lema 5.3.6: Sea $\alpha : [0, 1] \rightarrow \mathcal{U}(n)$ una curva minimal (suave) que une I con $U = e^{iX}$, donde $\|X\|_{sp} = \pi$. Por el Lema 5.3.6 existe Y_1 con $\|Y_1\|_{sp} = \pi$, y un autovector unitario ξ_1 de Y_1 correspondiente a un autovalor λ_1 , tal que $|\lambda_1| = \pi$, y

$$\alpha(t) = \begin{pmatrix} e^{it\lambda_1} & 0 \\ 0 & \alpha_1(t) \end{pmatrix} \begin{pmatrix} \mathcal{S}_1 \\ \mathcal{S}_1^\perp \end{pmatrix},$$

donde $\mathcal{S}_1 = \text{span}\{\xi_1\}$, y $\alpha_1 : [0, 1] \rightarrow \mathcal{U}(\mathcal{S}_1^\perp)$ es cualquier curva que une $I_{\mathcal{S}_1^\perp}$ con $U_{\mathcal{S}_1^\perp} = U|_{\mathcal{S}_1^\perp}$ tal que $\|\dot{\alpha}_1\|_{sp} \leq \|X\|_{sp}$.

Si $\alpha_1(1) = e^{iX_1}$ es tal que $\|X_1\|_{sp} = \pi$, aplicamos el Lema 5.3.6 a la curva $\alpha_1(t)$: existe Y_2 con $\|Y_2\|_{sp} = \pi$ y un autovector unitario ξ_2 de Y_2 correspondiente a un autovalor λ_2 tal que $|\lambda_2| = \pi$, tal que,

$$\alpha_1(t) = \begin{pmatrix} e^{it\lambda_2} & 0 \\ 0 & \alpha_2(t) \end{pmatrix} \begin{pmatrix} \mathcal{S}_2 \\ \mathcal{S}_2^\perp \end{pmatrix},$$

donde $\mathcal{S}_2 = \text{span}\{\xi_2\}$, y $\alpha_2 : [0, 1] \rightarrow \mathcal{U}(\mathcal{S}_2^\perp)$ es cualquier curva que une $I_{\mathcal{S}_2^\perp}$ con $U_{\mathcal{S}_2^\perp} = U|_{\mathcal{S}_2^\perp}$ tal que $\|\dot{\alpha}_2\|_{sp} \leq \|X\|_{sp}$.

Si $\alpha_2(1) = e^{iX_2}$ es tal que $\|X_2\|_{sp} = \pi$, continuamos iterando el lema. En algún punto vamos a terminar porque estamos en dimensión finita. De esta manera construimos $Y_{\mathcal{S}}$.

Es claro que si una curva tiene la estructura dada en el enunciado, entonces es minimal, lo cual se prueba a partir de un cálculo sencillo. ■

5.4 Geometría de los puntos medios

Dados $U, V \in \mathcal{U}(n)$ y $t \in (0, 1)$, en esta sección vamos a estudiar el conjunto de puntos intermedios:

$$\mathcal{M}_t(U, V) = \{W \in \mathcal{U}(n) : d_\infty(U, W) = t d_\infty(U, V) \text{ y } d_\infty(W, V) = (1 - t) d_\infty(U, V)\}.$$

En el caso de matrices positivas, el conjunto correspondiente de puntos intermedios, fue estudiado y se probó que es geodésicamente convexo (ver [40]). No podemos esperar que un

resultado similar con toda generalidad sea válido en nuestro contexto. De hecho, por ejemplo, en el caso unidimensional, el conjunto de puntos intermedios puede ser disconexo. En el caso que la dimensión sea mayor a uno, hay también un caso particular, donde la convexidad geodésica no es cierta:

Ejemplo. Sean $U = I$ y $V = -I$, y consideremos el conjunto de puntos intermedios $\mathcal{M}_t(I, -I)$ para algún $t \in (0, 1/2)$ (para $t \in (1/2, 1)$ es análogo). Entonces

$$W^+ = \begin{pmatrix} e^{it\pi} & 0 \\ 0 & e^{it\pi} \end{pmatrix} \quad y \quad W^- = \begin{pmatrix} e^{-it\pi} & 0 \\ 0 & e^{-it\pi} \end{pmatrix}.$$

pertenece a $\mathcal{M}_t(I, -I)$. Sin embargo, no es difícil ver que I es el punto medio de la geodésica (y única curva minimal por el Corolario 5.3.2) que une W^+ con W^- . $\blacktriangle$

Primero vamos a dar una caracterización del conjunto $\mathcal{M}_t(U, V)$.

Lema 5.4.1. *Dados $U, V \in \mathcal{U}(n)$ y $t \in (0, 1)$, entonces $M_t(U, V)$ se puede caracterizar como el siguiente conjunto*

$$\{W \in \mathcal{U}(n) : \text{existe } \gamma : [0, 1] \rightarrow \mathcal{U}(n) \text{ una curva entre } U \text{ y } V \text{ con } L(\gamma) = d_\infty(U, V) : \gamma(t) = W\}.$$

Demostración. Si existe una curva minimal γ tal que $\gamma(t) = W$ entonces, por el Corolario 5.3.3,

$$\begin{aligned} d_\infty(U, W) &= d_\infty(\gamma(0), \gamma(t)) = td_\infty(U, V) \\ d_\infty(W, V) &= d_\infty(\gamma(t), \gamma(1)) = (1 - t)d_\infty(U, V). \end{aligned}$$

Esto prueba una inclusión. Recíprocamente, si $W \in M_t(U, V)$, tomemos una curva minimal β_1 que une U con W y β_2 una curva minimal que une W con V . Entonces

$$\begin{aligned} L(\gamma) &= L(\beta_1) + L(\beta_2) \\ &= d_\infty(U, W) + d_\infty(W, V) \\ &= td_\infty(U, V) + (1 - t)d_\infty(U, V) = d_\infty(U, V). \end{aligned}$$

Si $\gamma = \beta_1 * \beta_2$ es la concatenación de estas dos curvas parametrizadas con velocidad constante en $[0, 1]$, entonces $W = \gamma(t)$. $\blacksquare$

A continuación vamos a probar que si $W_0, W_1 \in M_t(U, V)$ entonces $\beta(s) \in M_t(U, V)$, para todo $s \in [0, 1]$, donde $\beta(s)$ es una geodésica que une W_0, W_1 , con la condición que $d_\infty(U, V) < \pi/2$. La idea de la demostración es principalmente la misma que aquella usada en la Sección 4.3. Con este propósito vamos a usar el siguiente teorema.

Teorema 5.4.2 ([3]). *Si $d_\infty(W, \gamma(s)) < \pi/2$ para todo $s \in [0, 1]$, donde $\gamma(s)$ es una geodésica, entonces la función*

$$s \mapsto d_\infty(W, \gamma(s))$$

es convexa.

Veamos que β verifica las hipótesis del teorema:

Si $d_\infty(U, V) < \pi/2$, dados $W_0, W_1 \in M_t(U, V)$,

$$\begin{aligned} d_\infty(U, \beta(s)) &\leq d_\infty(U, W_j) + \frac{d_\infty(W_0, W_1)}{2} \\ &\leq d_\infty(U, W_j) + \frac{d_\infty(W_0, V)}{2} + \frac{d_\infty(W_1, V)}{2} \\ &= td_\infty(U, V) + \frac{(1-t)d_\infty(U, V)}{2} + \frac{(1-t)d_\infty(U, V)}{2} = d_\infty(U, V) < \frac{\pi}{2} \end{aligned}$$

tomando un $j = 0, 1$ conveniente.

Proposición 5.4.3. *Dados $U, V \in \mathcal{U}(n)$ tal que $d_\infty(U, V) < \pi/2$, y $t \in (0, 1)$, el conjunto*

$$\mathcal{M}_t(U, V) = \{W \in \mathcal{U}(n) : d_\infty(U, W) = td_\infty(U, V) \text{ y } d_\infty(W, V) = (1-t)d_\infty(U, V)\}$$

es geodésicamente convexo.

Demostración. Notar que, si $\beta(s)$ es la geodésica que une $W_0, W_1 \in \mathcal{M}_t(U, V)$, es decir, $W_0 = \beta(0), W_1 = \beta(1)$, entonces

$$\begin{aligned} d_\infty(U, \beta(s)) &\leq sd_\infty(U, W_1) + (1-s)d_\infty(U, W_0) \\ &= std_\infty(U, V) + (1-s)td_\infty(U, V) = td_\infty(U, V); \\ d_\infty(V, \beta(s)) &\leq sd_\infty(V, W_1) + (1-s)d_\infty(V, W_0) \\ &= s(1-t)d_\infty(U, V) + (1-s)(1-t)d_\infty(U, V) = (1-t)d_\infty(U, V) \end{aligned}$$

Luego

$$d_\infty(U, V) \leq d_\infty(U, \beta(s)) + d_\infty(V, \beta(s)) \leq d_\infty(U, V).$$

Es decir,

$$\begin{aligned} d_\infty(U, \beta(s)) &= td_\infty(U, V) \\ d_\infty(V, \beta(s)) &= (1-t)d_\infty(U, V). \end{aligned}$$

Entonces $\beta(s) \in M_t(U, V)$.

■

5.5 La Grassmaniana

La Grassmaniana $\mathcal{G}_n$ es el conjunto de subespacios de $\mathbb{C}^n$, que puede ser identificado con el conjunto de proyecciones ortogonales en $M_n(\mathbb{C})$. Si consideramos en $M_n(\mathbb{C})$ la topología definida por cualquiera de sus normas equivalentes, el espacio Grassmaniano dotado con esta topología, es un conjunto compacto. Sin embargo, no es conexo. En efecto, es suficiente considerar la traza, la cual es una función continua definida en todo el espacio $M_n(\mathbb{C})$, y restringida a $\mathcal{G}_n$ toma sólo valores naturales. En particular, esto prueba que las componentes conexas de $\mathcal{G}_n$ son subconjuntos $\mathcal{G}_{m,n}$, definidos como:

$$\mathcal{G}_{m,n} := \{P \in \mathcal{G}_n : \text{tr}(P) = m\}.$$

Cada una de estas componentes es una subvariedad de $M_n(\mathbb{C})$ [[61], p. 129], y las componentes conexas están dadas, por la órbita unitaria de una proyección dada P , tal que $\text{tr}(P) = \text{rk } P = m$:

$$\mathcal{G}_{m,n} = \{UPU^* : U \in \mathcal{U}(n)\}.$$

El espacio tangente en un punto $P \in \mathcal{G}_{m,n}$ se puede identificar con el subespacio de matrices Hermitianas P -codiagonales, es decir,

$$T_P\mathcal{G}_n = \{X \in \mathcal{H}(n) : X = PX + XP\}.$$

Denotamos por $\mathcal{S} = R(P)$. Luego cada $X \in T_P\mathcal{G}_n$ tiene una descomposición en bloques

$$X = \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix} \begin{matrix} \mathcal{S} \\ \mathcal{S}^\perp \end{matrix} \quad \text{para algún} \quad A \in L(\mathcal{S}^\perp, \mathcal{S}).$$

En particular notar que $T_P\mathcal{G}_n$ tiene un complemento natural N_P , que es el espacio de matrices Hermitianas que conmutan con P , es decir, las matrices P -diagonales Hermitianas. La descomposición en matrices diagonales y codiagonales define un **fibrado normal**, y nos lleva a una derivada covariante

$$\nabla_V \Gamma(P) = \Pi_{T_P||N_P} \frac{d}{dt} \Gamma(\alpha(t))|_{t=0},$$

donde Γ es un campo vectorial a lo largo de la curva $\alpha : (-\varepsilon, \varepsilon) \rightarrow \mathcal{G}_{m,n}$ que satisface $\alpha(0) = P$ y $\dot{\alpha}(0) = V$. Luego, tenemos una noción de paralelismo, y las geodésicas en este sentido, se describen en el siguiente teorema:

Teorema 5.5.1. (Ver Porta-Recht [54], también Davis-Kahan [25] o Halmos [28])

- La única geodésica en P con dirección $X \in T_P\mathcal{G}_n$ es:

$$\gamma(t) = e^{itX} P e^{-itX}.$$

- Si $P, Q \in \mathcal{G}_n$ son tales que $\|P - Q\|_{sp} < 1$, entonces existe un único $X \in T_P\mathcal{G}_n$ con $\|X\| < \pi/2$ tal que

$$X = PX + XP \in T_P\mathcal{G}_n \quad \text{y} \quad Q = e^{iX} P e^{-iX}. \quad (5.5.1)$$

Por lo tanto, pueden ser unidos por la geodésica $\gamma(t) = e^{itX} P e^{-itX}$.

La métrica de Finsler en la Grassmanniana: Para una norma simétrica dada $\|\cdot\|$ en $\mathcal{M}_n(\mathbb{C})$, el espacio Grassmanniano lleva la estructura de Finsler dada por

$$\|X\|_P = \|X\|_{sp} \quad \text{para} \quad X \in T_P \mathcal{G}_n .$$

Con esta estructura, la componente Grassmanniana $\{UPU^* : U \in \mathcal{U}(n)\}$ es isométrica (módulo un factor 2) a la órbita $\{US_P U^* : U \in \mathcal{U}(n)\}$ de la simetría $S_P = 2P - I$. Una cuenta sencilla muestra que, si $X = XP + PX$, entonces

$$e^{iX} S_P = S_P e^{-iX} .$$

Esta simple observación nos permite usar los resultados previos obtenidos en el grupo unitario y aplicarlos a la Grassmanniana.

Como el grupo unitario actúa de manera transitiva en estas componentes vía $U \cdot P = UPU^*$, ellas también son espacios homogéneos de $\mathcal{U}(n)$. Se las puede distinguir de otras subvariedades homogéneas de $\mathcal{U}(n)$, porque la función

$$P \mapsto S_P = 2P - 1$$

embebe a ellas en $\mathcal{U}(n)$, y la función S es dos veces una isometría. Las imágenes S_P son simetrías, es decir, matrices que satisfacen $S_P^* = S_P = S_P^{-1}$.

Notación 5.5.2. Sean $P \in \mathcal{G}_{m,n}$, $X \in T_P \mathcal{G}_n$ y $A \in L(\mathcal{S}^\perp, \mathcal{S})$ tales que

$$X = \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix} \begin{pmatrix} \mathcal{S} \\ \mathcal{S}^\perp \end{pmatrix} . \quad \text{Entonces} \quad |X| = \begin{pmatrix} |A^*| & 0 \\ 0 & |A| \end{pmatrix} \begin{pmatrix} \mathcal{S} \\ \mathcal{S}^\perp \end{pmatrix} . \quad (5.5.2)$$

1. Es conocido que, si $\sigma(A) = (\sigma_1, \dots, \sigma_n)$ son los valores singulares de A , entonces

$$\lambda(X) = (\sigma_1, \dots, \sigma_n, -\sigma_n, \dots, -\sigma_1) \in (\mathbb{R}^{2n})^\downarrow ,$$

y los valores singulares de X aparecen en pares.

2. Vamos a llamar $P_X \in \mathcal{M}_n(\mathbb{C})$ a la proyección ortogonal sobre el subespacio

$$S_X \stackrel{\text{def}}{=} \ker(\|X\|_{sp} I - |X|) \stackrel{(5.5.2)}{=} \Omega_1 \oplus \Omega_2 , \quad (5.5.3)$$

donde

$$\Omega_1 = \left\{ v \in \mathcal{S} : |A^*| v = \|A\|_{sp} v \right\} \quad \text{y} \quad \Omega_2 = \left\{ w \in \mathcal{S}^\perp : |A| w = \|A\|_{sp} w \right\} .$$

3. Vamos a llamar $\gamma_X : \mathbb{R} \rightarrow L(S_X)$ a la curva dada por

$$\gamma_X(t) = P_X \left(e^{itX} P e^{-itX} \right) P_X \Big|_{S_X} \quad \text{para} \quad t \in \mathbb{R} .$$

Por la ecuación (5.5.3), $P P_X = P_X P$ y $X P_X = P_X X$. Entonces también

$$P_X e^{itX} P e^{-itX} = e^{itX} P e^{-itX} P_X ,$$

y $\gamma_X(t) \in \mathcal{G}(S_X)$ para todo $t \in \mathbb{R}$ (como también la compresión a $S_X^\perp$). ▲

En el siguiente resultado vamos a usar la notación 5.5.2 dada anteriormente.

Teorema 5.5.3. Sean $P, Q \in \mathcal{G}_n$ con $\|P - Q\|_{sp} < 1$. Sea $X \in \mathcal{H}(n)$ como en la ecuación (5.5.1). Entonces $\|X\|_{sp} < \pi/2$ por lo cual $\sigma(|X|) \subseteq [0, \pi/2)$. Si $\sigma(|X|)$ tiene más de un elemento, entonces una curva minimal que une P con Q tiene la siguiente estructura:

$$\delta(t) = \begin{pmatrix} \gamma_X(t) & 0 \\ 0 & \alpha_1(t) \end{pmatrix} \begin{pmatrix} \mathcal{S}_X \\ \mathcal{S}_X^\perp \end{pmatrix},$$

donde $\alpha_1 : [0, 1] \rightarrow \mathcal{G}(\mathcal{S}_X^\perp)$ es cualquier curva que une, $P|_{\mathcal{S}_X^\perp}$ con $Q|_{\mathcal{S}_X^\perp}$, tal que $\|\dot{\alpha}_1\|_{sp} \leq \|X\|_{sp}$.

Demostración. Tomemos $\delta(t)$ una curva minimal que une P con Q en $\mathcal{G}_n$ (en este caso $L(\delta) = \|X\|_{sp}$). Notar que si $\gamma(t) = e^{itX} P e^{-itX}$ entonces

$$S_{\gamma(t)} = 2\gamma(t) - 1 = e^{itX} S_P e^{-itX} = e^{2itX} S_P = S_P e^{-2itX},$$

y esta curva es la geodésica que une S_P con S_Q y tiene longitud $2\|X\|_{sp}$. Ahora consideremos la curva $\alpha(t) = S_{\delta(t)}$ que une S_P con S_Q en $\mathcal{U}(n)$ y, notar que esta es una curva minimal, ya que

$$L(\alpha) = \int_0^1 \|\dot{\alpha}(t)\|_{sp} = \int_0^1 \|2\dot{\delta}(t)\|_{sp} = 2\|X\|_{sp}.$$

Entonces por el Teorema 5.3.1, como $\sigma(|X|) \subseteq [0, \pi/2)$ y tiene más de un elemento, la curva minimal que une S_P con S_Q tiene la siguiente estructura:

$$\alpha(t) = \begin{pmatrix} S_P e^{-2itX_S} & 0 \\ 0 & \alpha_1(t) \end{pmatrix} \begin{pmatrix} \mathcal{S} \\ \mathcal{S}^\perp \end{pmatrix},$$

donde $\mathcal{S} = \ker(\|X\|_{sp} I - |X|)$, $X_S = X|_{\mathcal{S}}$, y $\alpha_1 : [0, 1] \rightarrow \mathcal{U}(\mathcal{S}^\perp)$ es cualquier curva que une $S_P|_{\mathcal{S}^\perp}$ con $S_Q|_{\mathcal{S}^\perp}$, tal que $\|\dot{\alpha}_1\|_{sp} \leq 2\|X\|_{sp}$. Es fácil ver que en realidad $\mathcal{S} = \mathcal{S}_X$.

Como la función $\Phi : \mathcal{G}_n \rightarrow \mathcal{U}(n)$ dada por $\Phi(P) = S_P = 2P - I$ es dos veces una isometría y satisface que $\Phi(P)^2 = I$ para todo $P \in \mathcal{G}_n$, se obtiene la estructura dada. ■

Como antes, una consecuencia directa de este resultado, es el siguiente corolario sobre unicidad de curvas minimales:

Corolario 5.5.4. Dados $P, Q \in \mathcal{G}_n$ con $\|P - Q\| < 1$, existe una única curva minimal que los une si, y sólo si, $\sigma(S_Q S_P) \subseteq \{e^{i\theta}, e^{-i\theta}\}$ para algún $|\theta| < \pi/2$.

Observación. El corolario anterior nos dice que hay algunos **casos donde la unicidad es imposible**. Como mencionamos antes, los valores singulares de X están dados en pares, para todo autovalor real positivo hay un autovalor real negativo, además de los posibles autovalores nulos. Si la dimensión de n es impar entonces cero también es un autovalor y la condición $\sigma(S_Q S_P) = \{e^{i\theta}, e^{-i\theta}\}$ no se puede satisfacer. Si la dimensión de n es par, es una condición necesaria para la unicidad, que $\dim(R(p)) = \dim(N(P))$.

Proposición 5.5.5. Dados $P, Q \in \mathcal{G}_n$ tales que $\|P - Q\| < 1/\sqrt{2}$, y $t \in (0, 1)$, el conjunto

$$\mathcal{M}_t(P, Q) = \{R \in \mathcal{U}(n) : d_\infty(P, R) = t d_\infty(P, Q) \text{ y } d_\infty(R, Q) = (1 - t) d_\infty(P, Q)\}$$

es geodésicamente convexo.

Demostración. Por la Proposición 5.4.3, si $d(S_P, S_Q) < \pi/2$, entonces el conjunto $M_t(S_P, S_Q)$ es geodésicamente convexo. Como $d(S_P, S_Q) < \pi/2$ es equivalente a $\|S_P - S_Q\| < \sqrt{2}$, entonces

$$\|P - Q\| = \frac{1}{2} \|2P - 2Q\| = \frac{1}{2} \|(2P - 1) - (2Q - 1)\| = \frac{1}{2} \|S_P - S_Q\| < \frac{1}{2}\sqrt{2},$$

implica que $M_t(P, Q)$ es geodésicamente convexo. ■

6 | Apéndice

6.1 Análisis matricial

En este capítulo vamos a recordar algunas definiciones, resultados y demostraciones (o ideas de demostración) que utilizamos a lo largo del texto sobre matrices. El objetivo de esta sección es hacer que la tesis sea de alguna manera autocontenida, o que el lector tenga las herramientas para buscar en la bibliografía, los trabajos pertinentes.

En esta sección estudiaremos las relaciones de mayorización entre vectores reales asociados a matrices. Particularmente vectores de autovalores, de valores singulares y diagonales de matrices. A su vez estudiaremos la relación entre funciones Gauge simétricas y normas unitariamente invariantes.

Si bien la mayorización aparece como una forma de comparar vectores de $\mathbb{R}^n$, cuando se la piensa en vectores de autovalores o de valores singulares, se percibe rápidamente que es una noción intrínsecamente relacionada con la teoría de matrices.

Definición. Decimos que x está **mayorizada** por y , y lo simbolizamos como $x \prec y$, si

$$\sum_{j=1}^k x_j^\downarrow \leq \sum_{j=1}^k y_j^\downarrow,$$

para todo $1 \leq k \leq n$, y

$$\sum_{j=1}^n x_j^\downarrow = \sum_{j=1}^n y_j^\downarrow.$$

Si solamente se cumple la primer condición decimos que está mayorizada débilmente y lo notamos $x \prec_w y$.

El primer resultado que mencionamos es debido a Schur y relaciona la diagonal de una matriz con sus autovalores.

Teorema 6.1.1. (Teorema de mayorización de Schur) Sea $A \in \mathcal{H}(n)$. Utilizando la notación $d(A) = (a_{11}, \dots, a_{nn}) \in \mathbb{R}^n$, se tiene que

$$d(A) \prec \lambda(A).$$

Idea de la demostración. Para demostrar que $d(A) \prec \lambda(A)$ es suficiente probar que $d(A) = B\lambda(A)$, para cierta matriz B doble estocástica. Como $A \in \mathcal{H}(n)$, si $D = \text{diag}(\lambda(A))$,

existe $U \in \mathcal{U}(n)$ tal que $A = U^*DU$. Mediante cuentas elementales de matrices, se puede verificar que cada entrada de A tiene la expresión:

$$a_{ij} = \sum_{k=1}^n \bar{u}_{ki} \lambda_k u_{kj},$$

donde $i, j \in \mathbb{I}_n$. En particular

$$a_{ii} = \sum_{k=1}^n \lambda_k |u_{ki}|^2.$$

Consideremos ahora la matriz $B = (|u_{ji}|^2)_{ij}$ que, por ser U unitaria, es doble estocástica. Además, un cálculo sencillo, muestra que $d(A) = B\lambda(A)$. ■

A continuación recordamos la noción de función Gauge simétrica.

Definición. Una función $\Phi : \mathbb{R}^n \rightarrow R_+$ es una **función Gauge simétrica** si

1. Φ es una norma.
2. $\Phi(Px) = \Phi(x)$ para todo $x \in \mathbb{R}^n$ y P permutación.
3. $\Phi(\varepsilon_1 x_1, \dots, \varepsilon_n x_n) = \Phi(x_1, \dots, x_n)$ si $\varepsilon_j = \pm 1$.
4. $\Phi(1, 0, \dots, 0) = 1$.

Ejemplo. Si las coordenadas de x están ordenadas de manera que $|x_1| \geq |x_2| \geq \dots \geq |x_n|$, entonces para todo $k = 1, \dots, n$, la función

$$\Phi_{(k)}(x) = \sum_{j=1}^k |x_j|,$$

es una función Gauge simétrica. Vamos a usar la notación $\|x\|_{(k)}$ para esta función.

Proposición 6.1.2. Para todo $k = 1, \dots, n$,

$$\Phi_{(k)}(x) = \min \{ \Phi_{(n)}(u) + k\Phi_{(1)}(v) : x = u + v \}.$$

Idea de la demostración. Podemos suponer, sin pérdida de generalidad, que $x \in \mathbb{R}_+^n$. Si $x = u + v$,

$$\Phi_{(k)}(x) \leq \Phi_{(k)}(u) + \Phi_{(k)}(v) \leq \Phi_{(n)}(u) + k\Phi_{(1)}(v).$$

Si elegimos

$$\begin{aligned} u &= (x_1^\downarrow - x_k^\downarrow, x_2^\downarrow - x_k^\downarrow, \dots, x_k^\downarrow - x_k^\downarrow, 0, \dots, 0); \\ v &= (x_k^\downarrow, \dots, x_k^\downarrow, x_{k+1}^\downarrow, x_n^\downarrow), \end{aligned}$$

entonces

$$\begin{aligned} u + v &= x^\downarrow; \\ \Phi_{(k)}(u) &= \Phi_{(k)}(x) + kx_k^\downarrow; \\ \Phi_{(1)}(v) &= x_k^\downarrow, \end{aligned}$$

y obtenemos el resultado. ■

Definición. Una norma es **unitariamente invariante** en el espacio de matrices $M_n(\mathbb{C})$, y la notamos por $|||\cdot|||$, si para todo $U, V \in \mathcal{U}(n)$,

$$|||UAV||| = |||A|||.$$

Estas normas son muy útiles en análisis. Vamos a normalizar tales normas de modo tal que tomen el valor 1 en la matriz $\text{diag}(1, 0, \dots, 0)$. Hay una conexión, que provee los valores singulares, entre estas normas y las funciones Gauge simétricas.

Teorema 6.1.3. Dada una función Gauge simétrica Φ en $\mathbb{R}^n$, definimos una función en $M_n(\mathbb{C})$ dada por

$$|||A|||_{\Phi} = \Phi(s(A)).$$

Entonces, esto define una norma unitariamente invariante en $M_n(\mathbb{C})$. Recíprocamente, toda norma unitariamente invariante $|||\cdot|||$ en $M_n(\mathbb{C})$, define una función en $\mathbb{R}^n$ por

$$\Phi_{|||\cdot|||(x)} = |||\text{diag}(x)|||, \quad (6.1.1)$$

donde $\text{diag}(x)$ es la matriz diagonal con entradas $x_1, x_2, \dots, x_n$ en su diagonal. Esto define una función Gauge simétrica en $\mathbb{R}^n$.

Idea de la demostración. Como $s(UAV) = s(A)$ para todo U, V unitarios, $|||\cdot|||_{\Phi}$ es unitariamente invariante. Vamos a probar que obedece la desigualdad triangular, pues, las otras condiciones que debe cumplir una norma son fáciles de probar. Para esto, recordar la mayorización

$$s(A+B) \prec_w s(A) + s(B),$$

para todo $A, B \in M_n(\mathbb{C})$. Por último usar el hecho que Φ es monótona. Recíprocamente, notar que (6.1.1) nos da una norma en $\mathbb{R}^n$. Como las matrices diagonales de la forma $\text{diag}(e^{i\theta_1}, \dots, e^{i\theta_n})$ y las matrices de permutación son todas unitarias, esta norma es absolutamente invariante e invariante por permutaciones, y por lo tanto es una función Gauge simétrica. ■

Dos clases de normas unitariamente invariantes especialmente importantes son:

1. las **normas Schatten** p definidas como

$$\|A\|_p = \Phi_p(s(A)) = \left(\sum_{j=1}^n (s_j(A))^p \right)^{1/p},$$

donde $1 \leq p < \infty$, y

$$\|A\|_{\infty} = \Phi_{\infty}(s(A)) = s_1(A) = \|A\|_{sp}.$$

2. Las **normas Ky- Fan** k , definidas como

$$\|A\|_{(k)} = \sum_{j=1}^k s_j(A),$$

donde $1 \leq k \leq n$.

Proposición 6.1.4. *Para todo $k = 1, \dots, n$,*

$$\|A\|_{(k)} = \min \left\{ \|B\|_{(n)} + k \|C\|_{sp} : A = B + C \right\}.$$

Demostración. Si $A = B + C$, entonces

$$\|A\|_{(k)} \leq \|B\|_{(k)} + \|C\|_{(k)} \leq \|B\|_{(k)} + k \|C\|_{sp}.$$

Sea $s(A) = (s_1, \dots, s_n)$ y elegimos matrices unitarias U, V tales que

$$A = U[\text{diag}(s_1, \dots, s_n)]V.$$

Sean

$$\begin{aligned} B &= U[\text{diag}(s_1 - s_k, s_2 - s_k, \dots, s_k - s_k, 0, \dots, 0)]V; \\ C &= U[\text{diag}(s_k, \dots, s_k, s_{k+1}, \dots, s_n)]. \end{aligned}$$

Luego

$$A = B + C,$$

por lo tanto

$$\|B\|_{(n)} = \sum_{j=1}^k s_j - ks_k = \|A\|_{(k)} - ks_k.$$

Por otro lado

$$\|C\|_{sp} = s_k.$$

En consecuencia

$$\|A\|_{(k)} = \|B\|_{(n)} + k \|C\|_{sp}.$$

■

Teorema 6.1.5. *Para matrices $A, B \in M_n(\mathbb{C})$, tenemos la mayorización*

$$|s(A) - s(B)| \prec_w s(A - B).$$

Idea de la demostración. Elegimos cualquier índice $k = 1, 2, \dots, n$ y lo fijamos. Por la Proposición 6.1.4, existen $X, Y \in M_n(\mathbb{C})$ tales que

$$A - B = X + Y,$$

y

$$\|A - B\|_{(k)} = \|X\|_{(n)} + k \|Y\|_{sp}.$$

Definimos

$$\begin{aligned} \alpha &= s(X + B) - s(B); \\ \beta &= s(A) - s(X + B). \end{aligned}$$

Luego

$$s(A) - s(B) = \alpha + \beta.$$

Por lo tanto, por la Proposición 6.1.2, por

$$\Phi_{(1)}(s(A) - s(B)) \leq \Phi_{(1)}(s(A - B)),$$

y por

$$\sum_{j=1}^n \left| \lambda_j^\downarrow(A) - \lambda_j^\downarrow(B) \right| \leq \|A - B\|_{(n)},$$

tenemos

$$\begin{aligned} \Phi_{(k)}(s(A) - s(B)) &\leq \Phi_{(n)}(\alpha) + k\Phi_{(1)}(\beta) \\ &= \Phi_{(n)}(s(X + B) - s(B)) + k\Phi_{(1)}(s(A) - s(X + B)) \\ &\leq \|X\|_{(n)} + k\|A - (X + B)\|_{sp} \\ &= \|X\|_{(n)} + k\|Y\|_{sp} \\ &= \|A - B\|_{(k)} \end{aligned}$$

■

Teorema 6.1.6. Sean A, B matrices Hermitianas y sea Φ una función Gauge simétrica en $\mathbb{R}^n$. Entonces

$$\Phi(\lambda^\downarrow(A) - \lambda^\downarrow(B)) \leq \Phi(\lambda(A - B)) \leq \Phi(\lambda^\downarrow(A) - \lambda^\uparrow(B)).$$

Demostración. Ver, por ejemplo, Teorema III.4.4 [10].

■

Proposición 6.1.7. Sea $t \mapsto A(t)$ una función continua de un intervalo I al espacio de matrices $n \times n$. Entonces existen funciones continuas, $\lambda_1(t), \dots, \lambda_n(t)$ que, para todo $t \in I$, son los autovalores de $A(t)$.

Demostración. Ver, por ejemplo, Corolario VI.I.6 [10].

■

6.2 Geometría de las matrices positivas

Dado un intervalo $[a, b] \subset \mathbb{R}$ y una curva suave a trozos $\alpha : [a, b] \rightarrow \mathcal{P}(n)$, podemos definir la longitud de curva por

$$L(\alpha) = \int_a^b \left\| \alpha^{-1/2}(t) \dot{\alpha}(t) \alpha^{-1/2}(t) \right\|_2 dt. \quad (6.2.1)$$

La norma que consideramos es la norma Frobenius de estas matrices, es decir, $\|A\|_2 = (tr(A^*A))^{1/2}$. La distancia rectificable entre $A, B \in \mathcal{P}(n)$ está dada por

$$\delta_2(A, B) = \inf \{L(\alpha) \text{ donde } \alpha : [a, b] \rightarrow \mathcal{P}(n) \text{ une } A \text{ con } B \text{ en } \mathcal{P}(n)\}. \quad (6.2.2)$$

En esta sección vamos a ver que el ínfimo se alcanza en una única curva que une A con B . Esta curva se llama la *geodésica* de A a B .

Proposición 6.2.1. Sean A y B matrices que conmutan en $\mathcal{P}(n)$. Entonces la función exponencial envía el segmento que une $\log(A)$ con $\log(B)$ en $\mathcal{H}(n)$ en la geodésica que une A con B en $\mathcal{P}(n)$. En este caso,

$$\delta_2(A, B) = \|\log(A) - \log(B)\|_2$$

Demostración. Es suficiente probar que la curva,

$$\gamma(t) = \exp((1-t)\log(A) + t\log(B)),$$

donde $t \in [0, 1]$, es la única curva minimal que une A con B en el espacio $(\mathcal{P}(n), \delta_2)$. Como A conmuta con B entonces $\gamma(t) = A^{1-t}B^t$ y $\gamma'(t) = (\log(A) - \log(B))\gamma(t)$. Luego

$$L(\gamma) = \int_0^1 \|\log(A) - \log(B)\|_2 dt = \|\log(A) - \log(B)\|_2.$$

La EMI (Proposición 6.2.5) nos dice que ninguna curva puede ser más corta que esta. Entonces la curva γ bajo esta consideración es una curva minimal. Supongamos ahora que $\tilde{\gamma}$ es otra curva que une A con B de igual longitud que γ . Entonces $\tilde{H}(t) = \log(\tilde{\gamma}(t))$ es una curva que une $\log(A)$ con $\log(B)$ en $\mathcal{H}(n)$, y nuevamente por la EMI, esta curva tiene longitud $\|\log(A) - \log(B)\|_2$. Pero en el espacio Euclídeo los segmentos rectos son las únicas curvas minimales que unen dos puntos. Entonces $\tilde{H}(t)$ es una reparametrización de $(1-t)\log(A) + t\log(B)$. ■

Teorema 6.2.2. Sean A y B dos elementos de $\mathcal{P}(n)$. Entonces existe una única geodésica $A \#_t B$ que une A con B . Esta geodésica tiene la parametrización

$$A \#_t B = A^{1/2}(A^{-1/2}BA^{-1/2})^t A^{1/2}, \quad 0 \leq t \leq 1,$$

la cual es natural en el sentido que

$$\delta_2(A, A \#_t B) = t\delta_2(A, B),$$

para todo $t \in [0, 1]$. Más aún, tenemos que

$$\delta_2(A, B) = \|\log(A^{-1/2}BA^{-1/2})\|_2.$$

Demostración. Las matrices I y $A^{-1/2}BA^{-1/2}$ conmutan. Luego la geodésica que los une tiene la parametrización

$$I \#_t A^{-1/2}BA^{-1/2} = (A^{-1/2}BA^{-1/2})^t.$$

Aplicando la isometría $\Gamma_{A^{1/2}}$ (donde $\Gamma_X(Y) = X^*YX$) obtenemos la curva

$$\Gamma_{A^{1/2}}(I \#_t A^{-1/2}BA^{-1/2}) = A^{1/2}(A^{-1/2}BA^{-1/2})^t A^{1/2}.$$

Como $\Gamma_{A^{1/2}}$ es una isometría, la curva es una geodésica que une A con B . Usando la Proposición 6.2.1 vemos que

$$\begin{aligned} \delta_2(A, B) &= \delta_2(I, A^{-1/2}BA^{-1/2}) \\ &= \|\log(I) - \log(A^{-1/2}BA^{-1/2})\|_2 \\ &= \|\log(A^{-1/2}BA^{-1/2})\|_2. \end{aligned}$$
■

Por último veamos una demostración sencilla de este hecho que utilizamos en la motivación del capítulo 3.

Proposición 6.2.3. Sean A y B dos elementos de $\mathcal{P}(n)$, y sea

$$f(X) = \delta_2^2(A, X) + \delta_2^2(B, X).$$

Entonces la función f es estrictamente convexa en $\mathcal{P}(n)$, y tiene un único mínimo en el punto $X_0 = A \#_{1/2} B$.

Demostración. La convexidad de la función f se prueba a partir de la convexidad de la función

$$g(X) = \delta_2^2(A, X).$$

Es decir,

$$g(X_1 \#_t X_2) \leq (1-t)g(X_1) + tg(X_2),$$

lo cual se prueba primero para $t = 1/2$, luego para los números diádicos y por último se completa al intervalo $[0, 1]$. Por la desigualdad del semiparalelogramo

$$\delta_2(A \#_{1/2} B, X) \leq \frac{1}{2}f(X) - \frac{1}{4}\delta_2^2(A, B) = \frac{1}{2}f(X) - \frac{1}{4}f(A \#_{1/2} B).$$

Luego

$$f(A \#_{1/2} B) \leq f(X) - 2\delta_2(A \#_{1/2} B, X).$$

Esto prueba que f tiene un único mínimo en $X_0 = A \#_{1/2} B$. ■

Algunas de las demostraciones anteriores (Capítulo 4) se pueden generalizar de $\|\cdot\|_2$ a todas las normas Schatten p , $\|\cdot\|_p$ e incluso a las normas unitariamente invariantes $|||\cdot|||$.

Teorema 6.2.4. Para toda norma unitariamente invariante y toda $A, B \in \mathcal{P}(n)$ tenemos

$$|||A^{1/2}XB^{1/2}||| \leq |||\int_0^1 A^tXB^{1-t}dt|||.$$

Idea de la demostración. Supongamos que $A = B$ y que A es diagonal con entradas $\lambda_1, \dots, \lambda_n$ en su diagonal. La matriz $A^{1/2}XA^{1/2}$ se obtiene de $\int_0^1 A^tXB^{1-t}dt$ de multiplicarla entrada a entrada con la matriz Y cuyas entradas son

$$y_{ij} = \frac{\lambda_i^{1/2}\lambda_j^{1/2}(\log(\lambda_i) - \log(\lambda_j))}{\lambda_i - \lambda_j}.$$

Esta matriz es congruente con la matriz de entradas

$$z_{ij} = \frac{\log(\lambda_i) - \log(\lambda_j)}{\lambda_i - \lambda_j}.$$

Notar que esta matriz es positiva, por lo cual, obtenemos el resultado para $A = B$. Por otro lado, es suficiente probar este caso para obtener el caso general. ■

Proposición 6.2.5. (IEMI generalizada) Para todo H y K en $\mathcal{H}(n)$ tenemos

$$|||e^{-H/2}De^H(K)e^{-H/2}||| \geq |||K|||,$$

para toda norma unitariamente invariante.

Demostración. Como $K = e^{H/2}(e^{-H/2}Ke^{-H/2})e^{H/2}$,

$$\begin{aligned} |||K||| &= |||e^{H/2}(e^{-H/2}Ke^{-H/2})e^{H/2}||| \leq \left\| \int_0^1 e^{tH}(e^{-H/2}Ke^{-H/2})e^{(1-t)H} dt \right\| = \\ &= \left\| e^{-H/2} \left(\int_0^1 e^{tH}Ke^{(1-t)H} dt \right) e^{-H/2} \right\| = |||e^{-H/2}De^H(K)e^{-H/2}|||. \end{aligned}$$

■

Observar que como caso particular tenemos la Proposición 4.2.9. En la definición (6.2.1) podemos reemplazar la norma 2 por una norma unitariamente invariante, es decir,

$$L_{|||\cdot|||}(\alpha) = \int_a^b |||\alpha^{-1/2}(t)\dot{\alpha}(t)\alpha^{-1/2}(t)||| dt.$$

El análogo de (6.2.2),

$$\delta_{|||\cdot|||}(A, B) = \inf \{ L_{|||\cdot|||}(\alpha) \text{ donde } \alpha : [a, b] \rightarrow \mathcal{P}(n) \text{ une } A \text{ con } B \text{ en } \mathcal{P}(n) \}, \quad (6.2.3)$$

es una métrica en $\mathcal{P}(n)$ invariante por congruencias. La IEMI generalizada lleva a una EMI generalizada.

Proposición 6.2.6. Para todo $A, B \in \mathcal{P}(n)$,

$$\delta_{|||\cdot|||}(A, B) \geq |||\log(A) - \log(B)|||.$$

Equivalentemente, para toda $H, K \in \mathcal{H}(n)$,

$$\delta_{|||\cdot|||}(e^H, e^K) \geq |||H - K|||.$$

Muchas normas unitariamente invariantes tiene la propiedad que, en la métrica inducida por $\mathcal{H}(n)$, el segmento recto es la única geodésica que une dos puntos dados. Si la norma $|||\cdot|||$ tiene esta propiedad, entonces la métrica $\delta_{|||\cdot|||}$ en $\mathcal{P}(n)$ los hereda. Las normas Schatten p tiene esta propiedad para $1 < p < \infty$, pero no para $p = 1$ o $p = \infty$ (como ya hemos visto a lo largo del trabajo). Con esta observación, las demostraciones de esta sección se pueden probar en un contexto más general. En particular, tenemos que

$$\delta_{|||\cdot|||}(A, B) = |||\log(A^{-1/2}BA^{-1/2})|||.$$

La media geométrica $A \# B$ es equidistante de A a B en todas las métricas $\delta_{|||\cdot|||}$. Para ciertas métricas, tales como las correspondientes a las normas Schatten p para $1 < p < \infty$, este es el único punto medio entre A y B .

6.3 Teorema de Lusin

El próximo teorema es una extensión a dominios más generales del Teorema de Lusin para funciones a valores reales de valores reales.

Teorema 6.3.1. (*Lusin*) Sea (X, τ) un espacio topológico y μ una medida de Borel regular, cerrada y finita en X . Sea (S, d) un espacio métrico separable y sea f una función de Borel medible de X a S . Entonces para todo $\varepsilon > 0$, existe un conjunto cerrado $F \subset X$ tal que $\mu(X \setminus F) < \varepsilon$ y la restricción de f a F es continua.

Demostración. Sea $\{s_n\}_{n \geq 1}$ un conjunto denso contable de S . Para $m = 1, 2, \dots$, y todo $x \in X$, sea $f_m(x) = s_n$ para el menor n tal que

$$d(f(x), s_n) < \frac{1}{m}.$$

Entonces f_m es medible y está definida para todo X . Para todo m , sea $n(m)$ suficientemente grande tal que,

$$\mu\{x : d(f(x), s_n) \geq 1/m \text{ para todo } n \leq n(m)\} \leq \frac{1}{2^m}.$$

Por regularidad cerrada, para $n = 1, \dots, n(m)$, tomemos un conjunto cerrado $F_{mn} \subset f_m^{-1}\{s_n\}$ con

$$\mu(f_m^{-1}\{s_n\} \setminus F_{mn}) < \frac{1}{2^m n(m)}.$$

Para todo m fijo, los conjuntos F_{mn} son disjuntos para diferentes valores de n . Sea

$$F_m = \bigcup_{n=1}^{n(m)} F_{mn}.$$

Entonces f_m es continua en F_m . Por la elección de $n(m)$ y F_{mn} ,

$$\mu(X \setminus F_m) < \frac{2}{2^m}.$$

Como

$$d(f_m, f) < \frac{1}{m},$$

en todo punto, claramente $f_m \rightarrow f$ uniformemente. Para $r = 1, 2, \dots$, sea

$$H_r := \bigcap_{m=r}^{\infty} F_m.$$

Entonces H_r es un conjunto cerrado y $\mu(X \setminus H_r) \leq 4/2r$. Tomemos un r suficientemente grande tal que $4/2r < \varepsilon$. Entonces f restringido a H_r es continua, ya que es el límite uniforme de funciones continuas f_m en $H_r \subset F_m$ para $m \geq r$, entonces podemos llamar $F = H_r$ y completamos la prueba. ■

6.4 Centro de masa

Recordar que en el espacio Euclídeo, el centro de masa de un conjunto finito $\{x_i\}_{i=1}^n$, se define como

$$x = \frac{1}{n}(x_1 + \dots + x_n).$$

Esta igualdad es equivalente a preguntarse cuando

$$\overrightarrow{xx_1} + \dots + \overrightarrow{xx_n} = 0.$$

Más importante es la propiedad (descubierta por Appolonius) la cual nos dice que x , es de hecho, el (único) minimizador de la función

$$f : y \mapsto \sum_{i=1}^n d(y, x_i)^2.$$

Esta igualdad se extiende sin problemas a variedades Riemannianas siempre que nos quedemos dentro de bolas convexas. Dentro de dichas bolas, elegimos un subconjunto compacto $A \subset M$, y una distribución de masa da en A de masa total 1.

Proposición 6.4.1. (*Cartan*) *La función*

$$f : m \in M \mapsto \frac{1}{2} \int_A d(m, a)^2 da,$$

*es estrictamente convexa y alcanza un único mínimo en un punto llamado **centro de masa** de A para la distribución da . Más aún, este punto esta caracterizado por ser el único cero del campo vectorial gradiente¹*

$$\nabla f(x) = \int_A \exp_x^{-1}(a) da.$$

Demostración. Ver por ejemplo [9], Proposición 11, pág. 264. ■

¹El campo vectorial gradiente ∇f se define como el campo vectorial dual al diferencial df (el cual es una 1-forma) en el sentido que

$$g(\nabla f, v) = df(v),$$

para todo vector tangente v .

Bibliografía

- [1] Alexandrov, A. D.: A theorem on triangles in a metric space and some applications, Trudy Math. Inst. Steklov 38 (1951), 5-23.
- [2] Ando T., Li C. K., Mathias R.: Geometric means, Linear Algebra Appl. 385 (2004), 305-334.
- [3] Andruchow, E., Larotonda, G.: The rectifiable distance in the unitary Fredholm group. Polish Acad. Sciences Inst. Mathematics. (2010) vol. 196 p. 151-181
- [4] Andruchow, E., Larotonda, G., Recht, L.: Finsler geometry and actions of the p -Schatten unitary groups. Trans. Amer. Math. Soc. 62 (2010), 319-344.
- [5] Antezana J., Larotonda G.: Varela A.: Optimal paths for symmetric actions in the unitary group, Comm. Math. Phys. 328 (2014), no. 2, 481–497.
- [6] Austin T.: A CAT(0) valued pointwise ergodic theorem, J. Topol. Anal. 3 (2011) 145-152.
- [7] Ballmann W.: Lectures on spaces of nonpositive curvature. DMV Seminar Band 25, Birkhäuser Verlag, Basel.
- [8] Ballmann W., Gromov M., Schroeder V.: Manifolds of nonpositive curvature. Progress in Mathematics 61. Birkhäuser Boston Inc., Boston, MA
- [9] Berger M.: A Panoramic View of Riemannian Geometry, Springer, 2003.
- [10] Bhatia R.: Matrix analysis. Volume 169 of Graduate texts in mathematics. Springer, 1997.
- [11] Bhatia R.: On the exponential metric increasing property, Linear Algebra Appl. 375 (2003) 211–220.
- [12] Bhatia R.: Positive Definite Matrices, Princeton Series in Applied Mathematics, Princeton University Press, 2007.
- [13] Bhatia R.: Holbrook J., Noncommutative geometric means, Math. Intelligencer 28 (2006), no. 1, 32–39. MR MR2202893 (2007g:47023)
- [14] Bhatia R., Karandikar R.: Monotonicity of the matrix geometric mean. Math. Ann. 353(4)(2012) 1453-1467.

- [15] Bini D., Iannazzo B.: Computing the Karcher mean of symmetric positive definite matrices, *Linear Algebra Appl.* 438 (2013) 1700-1710.
- [16] Bini D., Meini B. and Poloni F.: an effective matrix geometric mean satisfying the Ando-Li-Mathias properties, *Math. Comp.* 79 (2010), 437-452.
- [17] Bochi J., Navas A.: A geometric path from zero Lyapunov exponents to rotation cocycles, *Ergodic Theory Dynam. Systems* 35 (2015) 374-402.
- [18] Bridson M.R., Haefliger A.: Metric spaces of non-positive curvature, *Grundlehren der Mathematischen Wissenschaften*, 319. Springer-Verlag, Berlin.
- [19] Barbaresco F.: Interactions between symmetric cone and information geometries: Bruhat-Tits and Siegel spaces models for higher resolution autoregressive Doppler imagery, *Emerging Trends in Visual Computing, Lecture Notes in Computer Science* 5416 (2009) 124-163.
- [20] Conde, C. Tesis doctoral: Desigualdades Geométricas e Interpolación de Operadores p -Schatten (2008).
- [21] Corach, G.; Maestripieri, A. L.: Differential and metrical structure of positive operators. *Positivity* 3 (1999), no. 4, 297–315.
- [22] Corach, G., Porta, H., Recht, L.: A geometric interpretation of Segal's inequality $\|e^{X+Y}\| \leq \|e^{X/2}e^Y e^{X/2}\|$. *Proc. Amer. Math. Soc.* 115 (1992), no. 1, 229–231.
- [23] Corach G., Porta H. and Recht L., Geodesics and operator means in the space of positive operators, *Int. J. Math.*, 4 (1993)
- [24] Corach G., Porta H. and Recht L., The geometry of the space of selfadjoint invertible elements in a C^* -algebra. *Integral Equations Operator Theory* 16 (1993), no. 3, 333–359.
- [25] Davis, C., Kahan, W.M.: The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* 7, 1-46 (1970).
- [26] Dudley R. M.: Real analysis and probability, *Cambridge Studies in Advanced Mathematics*, 74. Cambridge University Press, Cambridge.
- [27] Es-Sahib A., Heinich H.: Barycentre canonique pour un espace metrique a courbure negative, in: *Seminaire de Probabilites*, in: *Lecture Notes in Math.*, vol. 1709, Springer, Berlin, 1999.
- [28] Halmos, P.R.: Two subspaces. *Trans. Am. Math. Soc.* 144, 381-389 (1969).
- [29] Holbrook J.: No dice: a deterministic approach to the Cartan centroid, *J. Ramanujan Math. Soc.* 27 (2012) 509-521.
- [30] Jost J.: Nonpositive Curvature: Geometric and Analytic Aspects, *Lectures in Mathematics ETH Zurich*, Birkhauser, 1997.
- [31] Karcher H.: Riemannian center of mass and mollifier smoothing, *Comm. Pure Appl. Math.* 30 (1977) 509-541.

- [32] Lang S.: Fundamentals of Differential Geometry, Graduate Texts in Mathematics, Springer-Verlag, 1999.
- [33] Larotonda G.: The metric geometry of infinite dimensional Lie groups and their homogeneous spaces preprint arXiv:1805.02631
- [34] Lawson J., Lim Y.: A general framework for extending means to higher orders, Colloq. Math. 113 (2008) 191-221.
- [35] Lawson J., Lim Y.: Monotonic properties of the least squares mean, Math. Ann. 351 (2011) 267-279.
- [36] Lawson J., Lim Y.: Weighted means and Karcher equations of positive operators, Proc. Natl. Acad. Sci. USA 110 (2013) 15626-15632.
- [37] Lawson J., Lim Y.: Karcher means and Karcher equations of positive definite operators, Trans. Amer. Math. Soc. Ser. B1 (2014) 1-22.
- [38] Lawson J., Lim Y.: Contractive barycentric maps, J. Operator Theory 77 (2017) 87-107.
- [39] Lee J., Naor A.: Extending Lipschitz functions via random metric partitions, Invent. Math. 160 (2005) 59-95.
- [40] Lim Y.: Geometry of midpoint sets for Thompson's metric, Linear Algebra Appl. 439 (2013), 211-227.
- [41] Lim Y.: Riemannian distances between Geometric means. SIAM J. Matrix Anal. Appl. 34 (2013), 932-945.
- [42] Lim Y., Pálfi M.: Matrix power mean and the Karcher mean, J. Funct. Anal. 262 (2012) 1498-1514.
- [43] Lim Y., Pálfi M.: Weighted deterministic walks and no dice approach for the least squares mean on Hadamard spaces, Bull. Lond. Math. Soc. 46 (2014) 561-570.
- [44] Lim Y., Pálfi M.: Approximations to the Karcher mean on Hadamard spaces via geometric power means, Forum Math. 27 (2015) 2609-2635.
- [45] Mendel M., Naor A.: Spectral calculus and Lipschitz extension for barycentric metric spaces, Anal. Geom. Metr. Spaces 1 (2013) 163-199.
- [46] Moakher M., Zerai M.: The Riemannian geometry of the space of positive-definite matrices and its application to the regularization of positive-definite matrix-valued data, J. Math. Imaging Vision 40 (2011) 171-187.
- [47] Montgomery D., Zippin L.: Topological transformation groups. Interscience Publishers (1955) Geometry, Algebraic - 282 hojas.
- [48] Navas A.: An L^1 ergodic theorem with values in a non-positively curved space via a canonical barycenter map, Ergodic Theory Dynam. Systems 33 (2013) 609-623.

- [49] Nussbaum R.D.: Finsler structures for the part metric and Hilbert's projective metric and applications to ordinary differential equations, *Differ. Integral Equ.* 7 (1994) 1649–1707.
- [50] Ohta S.: Extending Lipschitz and Hölder maps between metric spaces, *Positivity* 13 (2009) 407–425.
- [51] Pálfi M.: Means in metric spaces and the center of mass, *J. Math. Anal. Appl.* 381 (2011) 383–391.
- [52] Pass B.: Uniqueness and Monge solutions in the multimarginal optimal transportation problem, *SIAM J. Math. Anal.* 43 (2011) 2758–2775.
- [53] Pass B.: Optimal transportation with infinitely many marginals, *J. Funct. Anal.* 264 (2013) 947–963.
- [54] H. Porta, L. Recht.: Minimality of geodesics in Grassmann manifolds, *Proc. Amer. Math. Soc.* 100 (1987), 464–466.
- [55] Reshetnyak Y. G.: (1968): Inextensible mappings in a space of curvature no greater than K . *Sib. Math. Jour.* 9, 683–689
- [56] Sturm K.-T.: Probability measures on metric spaces of nonpositive curvature, in: P. Auscher, et al. (Eds.), *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces*, in: *Contemp. Math.*, vol. 338, Amer. Math. Soc. (AMS), Providence, 2003.
- [57] Tao T.: Poincaré's legacies: pages from year two of a mathematical blog (Volume I, Volume II) American Mathematical Society.
- [58] Thompson, R.C.: Convex and concave functions of singular values of matrix sums. *Pacific J. Math.* 66(1), 285–290 (1976)
- [59] Thompson, R.C.: Matrix type metric inequalities. *Linear Multilinear Algebra* 5(4), 303–319 (1977/78)
- [60] Thompson, R.C.: Proof of a conjectured exponential formula. *Linear Multilinear Algebra* 19(2), 187–197 (1986)
- [61] Warner F. W.: *Foundations of differentiable manifolds and Lie groups*, Graduate Texts in Mathematics 94, Springer-Verlag, New York-Berlin, 1983.