



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

OPERADORES NO-EXPANSIVOS Y TRANSPORTE ÓPTIMO: CADENAS DE
MARKOV CON CURVATURA DE RICCI DISCRETA NULA

TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN MATEMÁTICAS
APLICADAS

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL MATEMÁTICO

MARTÍN RAPAPORT BOSCH

PROFESOR GUÍA:
ROBERTO COMINETTI COTTI-COMETTI

MIEMBROS DE LA COMISIÓN:
CLAIRE DELPLANCKE
JOAQUÍN FONTBONA TORRES
JAIME RICARDO SAN MARTÍN ARISTEGUI

Este trabajo ha sido parcialmente financiado por Proyecto Fondecyt 1171501 y CMM
Conicyt PIA AFB170001

SANTIAGO DE CHILE
2019

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE MAGÍSTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN MATEMÁTICAS APLICADAS
POR: MARTÍN RAPAPORT BOSCH
FECHA: 2019
PROF. GUÍA: ROBERTO COMINETTI COTTI-COMETTI

OPERADORES NO-EXPANSIVOS Y TRANSPORTE ÓPTIMO: CADENAS DE
MARKOV CON CURVATURA DE RICCI DISCRETA NULA

En la presente memoria se revisan aspectos relativos al *fenómeno de concentración de la medida*. En particular, se estudia la curvatura de Ricci discreta introducida por Yann Ollivier [26] y su relación con la concentración de la medida.

Nos interesamos al caso en que la curvatura de Ricci discreta es nula. Se define un operador T_m que resulta ser no expansivo con respecto a la distancia de Wasserstein 1 cuando la curvatura de Ricci discreta es nula. En este contexto, se define una iteración de Krasnosel'skii-Mann en el espacios de medidas con primer momento finito como:

$$\mu_{n+1} = (1 - \alpha_{n+1})\mu_n + \alpha_{n+1}T_m\mu_n .$$

con μ_0 una medida arbitraria y $\alpha_n \in (0, 1)$.

Se logra adaptar el Teorema 1.1 del artículo *Rates of Convergence for inexact Krasnosel'skii-Mann iterations in Banach spaces* [19] al Teorema 3.3, que enuncia que si se cumple una condición débil H_0^{ad} , se obtiene la siguiente tasa de convergencia entre μ_n y los iterados $T_m\mu_n$:

$$W_1(\mu_n, T_m\mu_n) \leq p \min\left\{1, \frac{1}{\sqrt{\pi \sum_{k=1}^n a_k(1 - a_k)}}\right\}$$

Se analizan diversos casos en los que la condición H_0^{ad} se satisface. Además, se obtiene un resultado (Teorema 3.6) que establece que la sucesión $\{\mu_n\}_{n \in \mathbb{N}}$ definida mediante las iteraciones de Krasnosel'skii-Mann converge débilmente a una medida invariante para la cadena de Markov.

Nuestro aporte es generalizar resultados de Krasnosel'skii-Mann para espacios normados en espacios métricos con la distancia de Wasserstein 1 y derivar tasas de convergencia para cadenas de Markov con curvatura de Ricci discreta nula.

A mi hermana Matilde

Agradecimientos

A mi hermana Matilde, que día a día nos regala su aire fresco, su ternura, su entusiasmo, su buena onda, su inteligencia y su profundo amor por la vida. Mi mamá repetía siempre hasta el cansancio “los hermanos sean unidos que esa es la ley primera”. Quizás, lo hemos conseguido y espero seguir haciendo camino.

A mi mamá que a lo largo de todos estos años me ha apoyado en infinitas aventuras. Mi mamá ha estado incondicionalmente a mi lado, temperamental (mi papá dice que parece italiana) y a la vez frágil pero que no dudaría un segundo en saltarse al abismo para salvarnos.

A mi papá que ha sido un guardián en nuestra familia, nos aconseja y nos protege como el mejor de los soldados. Si hay un terremoto llegará a la casa antes de que vuelva la señal. Mi mamá ha agregado alguna vez que en las situaciones difíciles conviene “siempre preguntarle que dirección tomar a mi papá”.

A mis abuelos que han sido de siempre cada uno a su manera muy cercanos, infinitamente generosos para conmigo: Pochi, Roberto, Tata, Chechi, Yeye y Yeya.

A la Pepa, mi perra que me saca a pasear todas las noches. Siempre me asombra por que paseamos muchas veces por el mismo camino y siempre me hace ver algo distinto.

A mi amigo Joaquín que es tan cercano como un segundo hermano. Somos tan cercanos que yo me he percatado que a veces adivina lo que estoy pensando. Joaquín es un monumento de rectitud y bondad. A mi amigo Diego otra gran persona, otro ser increíblemente bueno, compañero de tantas batallas (y de las muchas aventuras que espero vendrán).

A mi profesor guía : Roberto Cominetti. Un profesor extraordinario, no pude haber hecho mejor opción que escogerlo a él como profesor guía. Un verdadero y no exagero maestro. No creo ser hoy día tan riguroso como Roberto, pero no me cabe duda que cuando empezamos el viaje yo era muchísimo menos riguroso de lo que soy ahora. Es Roberto quien me ha enseñado aquello y muchísima matemática. Le agradezco que haya revisado cien mil millones de veces los borradores de mi memoria. Le agradezco lo aprendido. Le agradezco también las salidas a tomar café y un cigarrito y las risas también acerca de cualquier cosa. Espero sinceramente que nos volvamos a encontrar en algún camino.

A la comisión integrada por Claire Delplancke, Joaquín Fontbona y Jaime San Martín que han tenido la amabilidad y la disponibilidad de estar presentes para mi defensa. Con Jaime San Martín he compartido menos, sin embargo no olvidaré nunca sus entusiastas clases de

álgebra lineal que nos apasionaron tanto en plan común. Quiero agradecer a Claire y a Joaquín el haberme ayudado en la búsqueda de información para ir a estudiar a Francia. Sin ellos dos eso no se habría podido concretar. De forma más general, quisiera agradecerles el haberlos conocido, son y se nota de lejos, excelentes personas. De esas que hoy en el mundo faltan.

A la comunidad DIM mi más profundo cariño. Quisiera nombrar a dos amigos : a Reidmen y a Pereda. Reidmen es una máquina y es una maquina también de generosidad, de simpatía, de humildad, de excelencia en todos los sentidos. A Pereda, ese agente del mossad o de la cia, otro gigante también, otro tipazo inteligentísimo y a la vez fraterno con cada uno de nosotros. Quisiera también nombrar a Natacha que siempre me ha ayudado con cualquier duda y no sólo académica. Yo creo que todos nosotros en el DIM nos sentiríamos huérfanos sin Natacha.

A los fondos Fondecyt que hacen posible esta memoria.

A todos los que no agradezco pero están presentes en mi vida.

Uno siente (y esto no es falsa modestia) que uno no ha hecho demasiado para haber recibido tanto ¡Muchas gracias !

Tabla de Contenido

Introducción	1
1. Concentración de la medida	4
1.1. Algunos preliminares sobre <i>mixing times</i>	4
1.1.1. <i>Variación total</i> : definición y caracterizaciones	4
1.1.2. Introducción a los <i>mixing times</i>	7
1.1.3. Teoremas de convergencia	9
1.2. Algunos preliminares y técnicas sobre concentración de la medida	12
1.2.1. Concentración de la medida	12
1.2.2. Desigualdades de transporte	14
1.2.3. Desigualdades Log-Sobolev y de Poincaré	20
1.2.4. Relaciones entre distintas desigualdades	23
1.2.5. Concentración de la medida gaussiana vía la desigualdad de Prékopa-Leindler	27
1.2.6. Isoperimetría y concentración de la medida	29
1.2.7. La transformada de Laplace, el <i>largo</i> de un espacio métrico finito y el ejemplo del hipercubo	31
2. La curvatura de Ricci discreta	35
2.1. Introducción a la curvatura de Ricci discreta	35
2.1.1. Definición de curvatura de Ricci discreta	35
2.1.2. Espacios geodésicos	36
2.1.3. Algunos ejemplos	37
2.1.4. El ejemplo de \mathbb{Z}^N y \mathbb{R}^N	37
2.1.5. El ejemplo del hipercubo	38
2.1.6. El ejemplo de los árboles	39
2.1.7. El ejemplo de los grafos completos	42
2.2. La curvatura de Ricci discreta y los <i>mixing times</i>	43
2.2.1. Contracción en el espacio de medidas de probabilidad	43
2.2.2. <i>Mixing times</i> para hipercubo vía la curvatura de Ricci discreta	46
2.2.3. <i>Mixing times</i> para el hipercubo vía el <i>spectral gap</i>	49
2.3. La curvatura de Ricci discreta y el fenómeno de concentración de la medida	50
2.3.1. Algunas definiciones previas al resultado de concentración	50
2.3.2. Teorema de concentración Gaussiana	52
2.3.3. La existencia genuina de un tiempo máximo	58
2.3.4. Concentración exponencial con curvatura no negativa	59

3. Operadores no expansivos en espacios con curvatura de Ricci discreta nula	61
3.1. Operadores T_m y T_{m^α}	61
3.1.1. Propiedad básica de los operadores T_m y T_{m^α}	62
3.2. Tasa de convergencia para iteraciones de Krasnosel'skii-Mann en distancia de Wasserstein 1	63
3.2.1. Introducción	63
3.2.2. Definición de iteración de Krasnosel'skii-Mann en el espacio de medidas $\mathcal{P}_1(\mathcal{X})$	63
3.2.3. Generalización del Teorema 3.1	66
3.3. Convergencia débil de las iteraciones de Krasnosel'skii-Mann	70
3.3.1. Preliminares	70
3.3.2. Convergencia débil de las iteraciones de Krasnosel'skii-Mann	71
 Conclusión	 73
 Bibliografía	 74

Índice de Ilustraciones

1.1. Ilustración de la variación total. (Referencia: Markov Chains and Mixing Times, second edition.)	5
1.2. Síntesis recapitulativa de la sección 1.24: algunas relaciones entre desigualdades funcionales y el fenómeno de concentración de la medida	29
1.3. Ilustración del largo de un espacio métrico finito. (Referencia: Discrete Isoperimetric Inequalities, de Henry Liu.)	33
2.1. Ilustración de la curvatura de Ricci discreta para la cadena de Markov \star en \mathcal{Q}_3 .	39
2.2. Configuración de la arista (x, y) de un árbol en la que el vértice x posee grado 1.	41
2.3. Ilustración de la curvatura de Ricci discreta para un árbol.	42
2.4. Ilustración de la curvatura de Ricci para el grafo completo K_5	43
2.5. Ilustración de la curvatura de Ricci discreta para la cadena de Markov <i>lazy</i> $\star\star$ en \mathcal{Q}_3	47
2.6. Ilustración de la constante de difusión del paseo aleatorio en x	52
3.1. Problema de transporte óptimo recursivo. (Referencia: Sharp convergence rates for averaged nonexpansive maps.)	65

Introducción

El fenómeno de la concentración de la medida establece que, bajo ciertas hipótesis débiles, una función Lipschitz tiende a concentrarse alrededor de su media o mediana. La concentración de la medida es una herramienta poderosa con numerosas aplicaciones en geometría, análisis, teoría de las probabilidades y matemáticas discretas.

Desde el descubrimiento del *fenómeno de concentración de la medida* a finales de los 60', numerosos enfoques han sido relacionados con este fenómeno. Uno de éstos es la *curvatura de Ricci discreta* introducida por Yann Ollivier [26]. Se define mediante el uso del transporte óptimo, en particular mediante la distancia de Wasserstein 1.

Definición:

Sea (\mathcal{X}, d) un espacio métrico y sean $\{m_x\}_{x \in \mathcal{X}}$ medidas de probabilidad en \mathcal{X} .

Sean $x, y \in \mathcal{X}$. Se define $\mathcal{K}(x, y)$ como :

$$\mathcal{K}(x, y) := 1 - \frac{W_1(m_x, m_y)}{d(x, y)} .$$

Se define la curvatura de Ricci discreta como:

$$\mathcal{K} := \inf_{x, y \in \mathcal{X}} \mathcal{K}(x, y) .$$

La curvatura de Ricci discreta es una generalización de la curvatura de Ricci diferencial. En términos de Yann Ollivier, la curvatura de Ricci discreta, expresa que *si la curvatura es positiva las bolas están más cercas que sus centros* (en distancia de Wasserstein 1).

Si la curvatura de Ricci discreta es positiva entonces existe una única medida invariante y ésta satisface el fenómeno de concentración de la medida (Teorema 2.4).

Nos interesamos al caso en que la curvatura de Ricci discreta es nula. Se define un operador T_m de la siguiente manera:

$$\text{Operador } T_m : \mathcal{P}_1(\mathcal{X}) \rightarrow \mathcal{P}_1(\mathcal{X})$$

$$T_m(\mu)(A) := (\mu * m)(A) = \int_{\mathcal{X}} m_x(A) d\mu(x)$$

para todo conjunto boreliano A .

El operador T_m resulta ser no expansivo con respecto a la distancia de Wasserstein 1 cuando la curvatura de Ricci discreta es nula. En este contexto, se define una iteración de Krasnosel'skii-Mann en el espacios de medidas con primer momento finito como:

$$\mu_{n+1} = (1 - \alpha_{n+1})\mu_n + \alpha_{n+1}T_m\mu_n$$

con μ_0 una medida arbitraria y $\alpha_n \in (0, 1)$.

El siguiente Teorema 3.3 enuncia que si se cumple una condición denominada H_0^{ad} , se obtiene la siguiente tasa de convergencia entre μ_n y los iterados $T_m\mu_n$:

Teorema 3.3

Sea $(\mu_n)_{n \in \mathbb{N}}$ una secuencia generada por iteraciones de Krasnosel'skii-Mann y asumamos que la siguiente hipótesis se cumple :

$$\exists p \geq 0 \text{ tal que } W_1(T_m\mu_n, \mu_0) \leq p \quad \forall n \in \mathbb{N}. \quad (H_0^{\text{ad}})$$

Entonces, para todo $n \in \mathbb{N}$ se tiene que:

$$W_1(\mu_n, T_m\mu_n) \leq p \min\left\{1, \frac{1}{\sqrt{\pi \sum_{k=1}^n a_k(1 - a_k)}}\right\}$$

La condición H_0^{ad} se satisface en diversos casos: diámetro de \mathcal{X} acotado (Teorema 3.2), existencia de al menos una medida invariante (Proposición 3.3), existencia de un attracting point (Corolario 3.1), decaimiento rápido (Proposición 3.4), soporte de m_x incluido en un compacto (Corolario 3.2).

Además se logra relacionar la sucesión $\{\mu_n\}_{n \in \mathbb{N}}$ definida mediante las iteraciones de Krasnosel'skii-Mann con la compacidad débil (Teorema 3.6). Específicamente, se demuestra que la sucesión $\{\mu_n\}_{n \in \mathbb{N}}$ converge débil a una medida que es una medida invariante para la cadena de Markov m . El Teorema 3.6 se enuncia de la siguiente forma:

Teorema 3.6

Sea $\mathcal{X} \subset \mathbb{R}^d$ compacto. Sea $\{\mu_n\}_{n \in \mathbb{N}}$ una secuencia de medidas definidas mediante las iteraciones de Krasnosel'skii-Mann. Supongamos $\sup_{x \in \mathcal{X}} \int_{\mathcal{X}} d(y, o) dm_x(y) < \infty$ y que $\int_{\mathcal{X}} d(y, o) d\mu_0(x) < \infty$ para algún $o \in \mathcal{X}$ arbitrario y fijo. Entonces, la sucesión $\{\mu_n\}_{n \in \mathbb{N}}$ converge débil a una medida μ que es invariante para la cadena de Markov m .

La memoria se estructura en tres capítulos. El primer capítulo introduce los *mixing times* y el fenómeno de la concentración de la medida. Además, en este capítulo se revisan distintos métodos para obtener concentración de la medida.

El segundo capítulo trata sobre la curvatura de Ricci discreta y su relación con el fenómeno de concentración de la medida (Teorema 2.4, Teorema 2.5). El tercer capítulo trata sobre los operadores no expansivos en espacios con curvatura de Ricci discreta nula, obteniéndose el resultado citado anteriormente (Teorema 3.3) y explicitando ciertos casos en que se satisface. Además en este último capítulo, se establece una conexión entre las iteraciones de Krasnosel'skii-Mann y la compacidad débil (Teorema 3.6).

Capítulo 1

Concentración de la medida

1.1. Algunos preliminares sobre *mixing times*

Esta primera sección de este capítulo constituye una breve introducción al concepto de *mixing times*. Nos basaremos principalmente en el libro Markov Chains and Mixing Times, second edition de David Levin, Yuval Peres y Elizabeth L. Wilmer [37].

En esta sección, el espacio de estados, que denotamos \mathcal{S} es finito.

1.1.1. *Variación total*: definición y caracterizaciones

Sean μ y ν dos distribuciones de probabilidad en un espacio finito \mathcal{S} . La *variación total* entre μ y ν se define como:

$$\|\mu - \nu\|_{TV} := \max_{A \subset \mathcal{S}} |\mu(A) - \nu(A)|$$

con $\mu(A) = \sum_{x \in A} \mu(x)$.

Proposición 1.1 [37]

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in \mathcal{S}} |\mu(x) - \nu(x)| .$$

Demostración: Sea $B = \{x \in \mathcal{S} : \mu(x) \geq \nu(x)\}$. Sea $A \subset \mathcal{S}$ un conjunto cualquiera. Ahora bien, A es unión disjunta entre $A \cap B$ y $A \cap B^c$. Para todo $x \in A \cap B^c$: $\mu(x) - \nu(x) < 0$. Así:

$$\mu(A) - \nu(A) = \mu(A \cap B) - \nu(A \cap B) + \mu(A \cap B^c) - \nu(A \cap B^c) \leq \mu(A \cap B) - \nu(A \cap B) .$$

Además B es unión disjunta de $A \cap B$ y $B \setminus A$. Para todo $x \in B \setminus A$: $\mu(x) - \nu(x) \geq 0$. Así:

$$\mu(A \cap B) - \nu(A \cap B) \leq \mu(A \cap B) - \nu(A \cap B) + \mu(B \setminus A) - \nu(B \setminus A) = \mu(B) - \nu(B) .$$

Entonces se obtiene la siguiente desigualdad:

$$\mu(A) - \nu(A) \leq \mu(B) - \nu(B) .$$

De manera análoga:

$$\mu(B^c) - \nu(B^c) \leq \mu(A) - \nu(A) .$$

Como $\mu(B) + \mu(B^c) = \nu(B) + \nu(B^c) = 1$, $\mu(B^c) - \nu(B^c) = -(\mu(B) - \nu(B))$. Luego:

$$| \mu(A) - \nu(A) | \leq \mu(B) - \nu(B) .$$

De este modo, $| \mu(A) - \nu(A) |$ está acotada por $\mu(B) - \nu(B)$ y esta cota se alcanza con $A = B$. Entonces:

$$\| \mu - \nu \|_{TV} = \max_{A \subset X} | \mu(A) - \nu(A) | = \mu(B) - \nu(B)$$

$$\| \mu - \nu \|_{TV} = \frac{1}{2} ((\mu(B) - \nu(B)) + (\nu(B^c) - \mu(B^c))) = \frac{1}{2} \sum_{x \in S} | \mu(x) - \nu(x) | .$$



Observación : La demostración anterior también explicita la siguiente caracterización de la variación total.

$$\| \mu - \nu \|_{TV} = \sum_{x \in S, \mu(x) \geq \nu(x)} \mu(x) - \nu(x) .$$

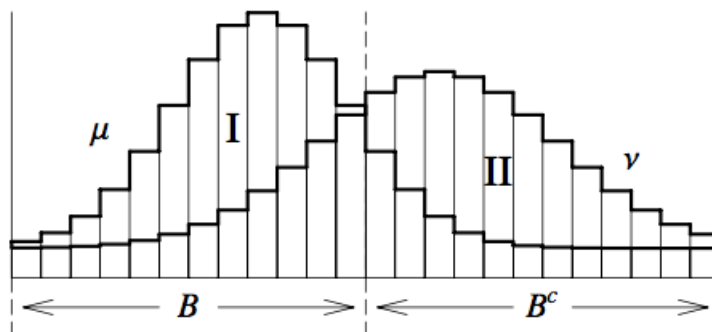


Figura 1.1: Ilustración de la variación total. (Referencia: Markov Chains and Mixing Times, second edition.)

Un *coupling* entre dos medidas de probabilidad μ y ν en \mathcal{S} es un par de variables aleatorias (X, Y) que posee una distribución conjunta q en $\mathcal{S} \times \mathcal{S}$ tal que las distribuciones marginales son μ y ν respectivamente i.e. :

$$\mathbf{P}(X = x) = \sum_{y \in \mathcal{S}} q(x, y) = \mu(x),$$

$$\mathbf{P}(Y = y) = \sum_{x \in \mathcal{S}} q(x, y) = \nu(y).$$

La variación total admite la siguiente caracterización.

Teorema 1.1 [37]

Sean μ y ν dos distribuciones de probabilidad en \mathcal{S} . Entonces:

$$\|\mu - \nu\|_{TV} = \inf\{\mathbf{P}\{X \neq Y\} : (X, Y) \text{ coupling entre } \mu \text{ y } \nu\}$$

y el ínfimo es alcanzado.

Demostración: Para cualquier *coupling* (X, Y) de μ y ν y para cualquier evento $A \subset X$:

$$\begin{aligned} \mu(A) - \nu(A) &= \mathbf{P}\{X \in A\} - \mathbf{P}\{Y \in A\} \\ &= \mathbf{P}\{X \in A, Y \in A\} + \mathbf{P}\{X \in A, Y \notin A\} - \mathbf{P}\{Y \in A\} \\ &\leq \mathbf{P}\{Y \in A\} + \mathbf{P}\{X \in A, Y \notin A\} - \mathbf{P}\{Y \in A\} \\ &= \mathbf{P}\{X \in A, Y \notin A\}. \end{aligned}$$

Dado que el evento $\{X \in A, Y \notin A\}$ está incluido en $\{X \neq Y\}$ se sigue que:

$$\mu(A) - \nu(A) \leq \mathbf{P}\{X \neq Y\}.$$

Se sigue inmediatamente :

$$\|\mu - \nu\|_{TV} \leq \inf\{\mathbf{P}\{X \neq Y\} : (X, Y) \text{ coupling entre } \mu \text{ y } \nu\}.$$

Para demostrar completamente el teorema se construye un *coupling* tal que ínfimo se alcanza. El lector interesado puede consultar la proposición 4.7 del libro Makov Chains and Mixing Times, second edition [37].



1.1.2. Introducción a los *mixing times*

Sea $(X_t)_{t \geq 0}$ una cadena de Markov homogénea en un espacio finito \mathcal{S} , de matriz de transición P , que asumimos irreducible y aperiódica. Entonces, existe una única distribución estacionaria π definida por la ecuación:

$$\pi P = \pi,$$

y la cadena converge a π en el sentido que para todo $(x, y) \in \mathcal{S}^2$,

$$\lim_{t \rightarrow \infty} P^t(x, y) = \pi(y).$$

Una pregunta natural es cuantificar la tasa de convergencia: ¿Dado un número de pasos, cuán cerca está la cadena del equilibrio ?

Para responder esta interrogante consideramos:

$$d(t)_x := \| P^t(x, \cdot) - \pi \|_{TV}.$$

Como se busca un control uniforme sobre todos los puntos, se define el *peor caso*:

$$d(t) := \max_{x \in X} d(t)_x = \max_{x \in X} \| P^t(x, \cdot) - \pi \|_{TV}.$$

Es útil también definir:

$$\bar{d}(t) := \max_{x, y \in X} \| P^t(x, \cdot) - P^t(y, \cdot) \|_{TV}.$$

El *mixing time* se define como:

$$t_{mix}(\varepsilon) := \min\{t : d(t) \leq \varepsilon\} \quad y \quad t_{mix} := t_{mix}\left(\frac{1}{4}\right).$$

Lema 1.1 (La función \bar{d} es submultiplicativa) [37]

$$\bar{d}(s+t) \leq \bar{d}(s)\bar{d}(t).$$

Demostración: Para $x, y \in \mathcal{S}$, sea (X_s, Y_s) un coupling óptimo entre $P^s(x, \cdot)$ y $P^s(y, \cdot)$ cuya existencia está garantizada por el Teorema 1.1 . Entonces:

$$\| P^s(x, \cdot) - P^s(y, \cdot) \|_{TV} = \mathbb{P}\{X_s \neq Y_s\}. \quad (1.1)$$

Así:

$$P^{s+t}(x, w) = \sum_z P^s(x, z) P^t(z, w) = \mathbb{E}(P^t(X_s, w)).$$

Para un conjunto A , sumando sobre $w \in A$:

$$P^{s+t}(x, A) - P^{s+t}(y, A) = \mathbb{E}(P^t(X_s, A) - P^t(Y_s, A)),$$

$$P^{s+t}(x, A) - P^{s+t}(y, A) \leq \mathbb{E}(\bar{d}(t)1_{\{X_s \neq Y_s\}}) = \mathbb{P}\{X_s \neq Y_s\}\bar{d}(t).$$

Por la ecuación (1.1), el lado derecho es menor o igual a $\bar{d}(s)\bar{d}(t)$. ■

El siguiente Lema relaciona d y \bar{d} .

Lema 1.2 [37]

Para $d(t)$ y $\bar{d}(t)$ definidos anteriormente se tiene:

$$d(t) \leq \bar{d}(t) \leq 2d(t).$$

Demostración: Por la desigualdad triangular para la distancia variación total, es inmediato que $\bar{d}(t) \leq 2d(t)$. Para demostrar que $d(t) \leq \bar{d}(t)$, notemos en un primer tiempo que para todo conjunto A , por definición de medida estacionaria se obtiene:

$$\pi(A) = \sum_{y \in X} \pi(y)P^t(y, A).$$

Así:

$$|P^t(x, A) - \pi(A)| = \left| \sum_{y \in X} \pi(y)(P^t(x, A) - P^t(y, A)) \right|$$

$$|P^t(x, A) - \pi(A)| \leq \sum_{y \in X} \pi(y) \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \leq \bar{d}(t)$$

esto último por la desigualdad triangular y la definición de variación total.

Maximizando el lado izquierdo con respecto a x se obtiene :

$$d(t) \leq \bar{d}(t). \quad \blacksquare$$

Observación : Por los dos Lemas anteriores, para c y t enteros positivos:

$$d(ct) \leq \bar{d}(ct) \leq \bar{d}(t)^c \leq 2^c d(t)^c.$$

1.1.3. Teoremas de convergencia

El teorema central de esta sección plantea una condición suficiente para obtener convergencia exponencial de una cadena de Markov. Para ello, es necesario introducir el concepto de aperiodicidad de una cadena de Markov.

Se define $\tau(x)$ como:

$$\tau(x) := \{t \geq 1 : P^t(x, x) > 0\}.$$

Se define el período de x que denotamos $\text{per}(x)$ como m.c.d. $\tau(x)$. Se dice que una cadena de Markov irreducible es aperiódica si $\text{per}(x) = 1 \ \forall x \in \mathcal{S}$.

El siguiente teorema expresa que una cadena de Markov finita, irreducible y aperiódica converge a una tasa exponencial a una distribución estacionaria.

Teorema 1.2 (Convergencia para cadenas de Markov finitas) [37]

Supongamos que P es irreducible y aperiódica, con medida de distribución estacionaria π . Entonces, existen constantes $\alpha \in (0, 1)$ y $C > 0$ tales que:

$$\max_{x \in X} \| P^t(x, \cdot) - \pi \|_{TV} \leq C\alpha^t$$

Para la demostración del Teorema, el siguiente Lema y la siguiente proposición son claves.

Lema 1.3 [37]

Sea $Q \subset \mathbb{Z}^+ \cup \{0\}$ cerrado bajo la operación suma (i.e: $a + b \in Q$ para todo $a, b \in Q$) y m.c.d. $Q = 1$, entonces existe una constante M tal que $a \in Q$ para todo $a \geq M$.

Proposición 1.2 [37]

Si P es aperiódico e irreducible, entonces existe un entero r_0 tal que $P^r(x, y) > 0$ para todo $x, y \in \mathcal{S}$ y $r \geq r_0$.

Demostración: Para todo $x \in \mathcal{S}$, $\tau(x)$ es un conjunto de enteros no negativos cerrado bajo la operación suma. En efecto, si $s, t \in \tau(x)$: $P^s(x, x) > 0$ y $P^t(x, x) > 0$, entonces $P^{s+t}(x, x) \geq P^s(x, x)P^t(x, x) > 0$ y por ende $s + t \in \tau(x)$. Como la cadena es aperiódica, m.c.d. $\tau(x)$ es 1. De este modo, de acuerdo al Lema 2.3, existe $t(x)$ tal que $t \geq t(x)$ implica que $t \in \tau(x)$. Además, como la cadena es irreducible sabemos que para todo $y \in \mathcal{S}$, existe $r = r(x, y)$ tal que $P^r(x, y) > 0$. Luego, para $t \geq t(x) + r$:

$$P^t(x, y) \geq P^{t-r}(x, x)P^r(x, y) > 0.$$

Para $t \geq t'(x) := t(x) + \max_{y \in \mathcal{S}} r(x, y)$, se tiene que $P^t(x, y) > 0$ para todo $y \in \mathcal{S}$. Finalmente, si $t \geq \max_{x \in \mathcal{S}} t'(x)$, entonces $P^t(x, y) > 0$ para todo $x, y \in \mathcal{S}$.

Demostración del Teorema 1.2: Como P es irreducible y aperiódica, por la proposición 1.7, existe un entero r tal que P^r contiene entradas estrictamente positivas. Sea Π la matriz con $|\mathcal{S}|$ filas, tales que cada fila sea el vector π . Para $\delta > 0$ suficientemente pequeño, se tiene :

$$P^r(x, y) \geq \delta\pi(y)$$

para todo $x, y \in \mathcal{S}$. Sea $\theta = 1 - \delta$. La ecuación :

$$P^r = (1 - \theta)\Pi + \theta Q \tag{1.2}$$

define una matriz estocástica Q .

En efecto, todo elemento de Q es no negativo dado que:

$$P^r(x, y) - (1 - \theta)\Pi(x, y) = P^r(x, y) - \delta\pi(y) \geq 0.$$

Además , cada fila de Q suma $\frac{1-(1-\theta)}{\theta} = 1$. Así, Q es una matriz estocástica.

Para todo $n \geq 0$, $Q^n\Pi = \Pi$. En efecto:

$$(Q^n\Pi)(x, y) = \sum_{z \in \mathcal{S}} Q^n(x, z)\Pi(z, y) = \pi(y) \sum_{z \in \mathcal{S}} Q^n(x, z) = \pi(y) = \Pi(x, y).$$

Por otro lado, como π es una medida invariante, para $n \geq 0$:

$$\Pi P^n = \Pi$$

Ahora bien, por inducción se demuestra que:

$$P^{rk} = (1 - \theta^k)\Pi + \theta^k Q^k \tag{1.3}$$

para $k \geq 1$.

Si $k = 1$, lo anterior es cierto por (1.2). Supongamos que (1.3) se satisface para $k = n$.

$$P^{r(n+1)} = P^{rn} P^r = ((1 - \theta^n)\Pi + \theta^n Q^n) P^r.$$

Distribuyendo y expandiendo P^r en el segundo término utilizando (1.2) se obtiene:

$$P^{r(n+1)} = (1 - \theta^n)\Pi P^r + \theta^n Q^n((1 - \theta)\Pi + \theta Q)$$

$$P^{r(n+1)} = (1 - \theta^n)\Pi P^r + (1 - \theta)\theta^n Q^n \Pi + \theta^{n+1} Q^{n+1}.$$

Dado que $\Pi P^r = \Pi$ y $Q^n \Pi = \Pi$ se tiene que :

$$P^{r(n+1)} = (1 - \theta^n)\Pi + (1 - \theta)\theta^n \Pi + \theta^{n+1} Q^{n+1}$$

$$P^{r(n+1)} = (1 - \theta^{n+1})\Pi + \theta^{n+1}Q^{n+1}.$$

Aquello, establece (1.3) para $k = n + 1$ (asumiendo que es cierto para $k = n$) y así es cierto para todo k .

Multiplicando (1.3) por P^j y reordenando los términos se obtiene:

$$P^{r(k+j)} = (1 - \theta^k)\Pi P^j + \theta^k Q^k P^j$$

$$P^{r(k+j)} = (1 - \theta^k)\Pi + \theta^k Q^k P^j$$

$$P^{r(k+j)} - \Pi = \theta^k(Q^k P^j - \Pi).$$

Enfocándose en cualquier fila x de los dos lados de la ecuación se tiene:

$$(P^{r(k+j)} - \Pi)(x, \cdot) = \theta^k(Q^k P^j - \Pi)(x, \cdot).$$

Así:

$$\| P^{r(k+j)} - \pi \|_{TV} = \| P^{r(k+j)}(x, \cdot) - \Pi(x, \cdot) \|_{TV}$$

$$\| P^{r(k+j)} - \pi \|_{TV} = \frac{1}{2} \sum_{y \in \mathcal{S}} | (P^{r(k+j)} - \Pi)(x, y) |$$

$$\| P^{r(k+j)} - \pi \|_{TV} = \frac{\theta^k}{2} \sum_{y \in \mathcal{S}} | (Q^k P^j - \Pi)(x, y) |$$

$$\| P^{r(k+j)} - \pi \|_{TV} = \theta^k \| Q^k P^j(x, \cdot) - \Pi(x, \cdot) \|_{TV} \leq \theta^k.$$

La última desigualdad es cierta pues $Q^k P^j(x, \cdot)$ y $\Pi(x, \cdot)$ son distribuciones y por ende su distancia en variación total está acotada por 1.

Luego, se definen las constantes $\alpha = \theta^{\frac{1}{r}}$ y $C = \alpha^{-r}$. Sea $t \geq 0$ y se define k y j mediante el teorema de la división por r , así $t = rk + j$ con $0 \leq j < r$.

De este modo :

$$\| P^t(x, \cdot) - \pi \|_{TV} \leq \theta^k = \alpha^{rk} \leq \alpha^{rk} \alpha^{j-r} = C \alpha^t$$

El argumento anterior es válido para todo $x \in \mathcal{S}$, entonces para todo $t \geq 0$:

$$\max_{x \in \mathcal{S}} \| P^t(x, \cdot) - \pi \|_{TV} \leq C\alpha^t.$$

■

1.2. Algunos preliminares y técnicas sobre concentración de la medida

La siguiente sección presenta algunos conceptos básicos sobre el fenómeno de concentración de la medida y algunos métodos o técnicas para obtener concentración de la medida. No pretende ser exhaustiva sobre el tema. Para el lector interesado, recomendamos las siguientes referencias: *The Concentration of Measure Phenomenon* (Michel Ledoux) [16], *Concentration Inequalities: A Nonasymptotic Theory of Independence* (Stéphane Boucheron, Gábor Lugosi y Pascal Massart) [35], *Transport Inequalities. A survey* (Nathael Gozlan, Christian Léonard). [24]

En esta sección, \mathcal{X} es un espacio polaco (i.e espacio métrico, completo y separable), equipado con su σ -álgebra de Borel $\beta(\mathcal{X})$, $P(\mathcal{X})$ es el conjunto de medidas de probabilidad de Borel en \mathcal{X} , μ es una medida de probabilidad en \mathcal{X} y $\alpha : \mathbb{R}^+ \rightarrow [0, 1]$ es una función no creciente.

1.2.1. Concentración de la medida

- Se dice que μ satisface la propiedad de concentración de medida con perfil α si para todo $A \subset \mathcal{X}$ con $\mu(A) \geq \frac{1}{2}$ se satisface:

$$\mu(A^r) \geq 1 - \alpha(r) \quad \forall r \geq 0$$

con $A^r = \{x \in \mathcal{X} : d(x, A) \leq r\}$.

Se define de este modo la *función de concentración*:

$$\alpha_\mu(r) := \alpha_{(\mathcal{X}, d, \mu)}(r) = \sup\{1 - \mu(A^r) : \mu(A) \geq \frac{1}{2}\}.$$

- Se dice que μ satisface la propiedad de concentración libre de dimensión con perfil α , si para todo $n \in \mathbb{N}^*$, la medida producto μ^{*n} satisface la desigualdad de concentración con perfil α en \mathcal{X}^n , equipada con la distancia :

$$d_2(x, y) = \sqrt{\sum_{i=1}^n d^2(x_i, y_i)}, \quad x, y \in \mathcal{X}^n.$$

- Se dice que $m_f \in \mathbb{R}$ es una μ -mediana de f si:

$$\mu\{f \leq m_f\} \geq \frac{1}{2} \quad \text{y} \quad \mu\{f \geq m_f\} \geq \frac{1}{2}.$$

Notemos que una mediana no es necesariamente única.

La siguiente proposición clásica, permite describir el fenómeno de concentración de medida en términos equivalentes de las desviaciones de las funciones Lipschitz respecto a una mediana. Usualmente este resultado se interpreta como que *las funciones Lipschitz son esencialmente constantes* cuando existe concentración de la medida para la medida de probabilidad subyacente. En otras palabras, las funciones Lipschitz son, con alta probabilidad, concentradas en la mediana, cuando el perfil de concentración α decrece rápidamente a cero.

Proposición 1.3 [24]

Los siguientes enunciados son equivalentes:

1. La distribución μ satisface la desigualdad de concentración con perfil α .
2. Para toda función 1-Lipschitz f en \mathcal{X} y m_f una mediana de f :

$$\mu(f > m_f + r) \leq \alpha(r), \quad r \geq 0.$$

Demostración:

(1) \implies (2). Sea f una función 1-Lipschitz y definamos $A = \{f \leq m_f\}$. Entonces, dado que f es 1-Lipschitz, se verifica que $A^r \subset \{f \leq m_f + r\}$. Como $\mu(A) \geq \frac{1}{2}$:

$$\mu(\{f \leq m_f + r\}) \geq \mu(A^r) \geq 1 - \alpha(r).$$

Así:

$$\mu(\{f > m_f + r\}) = 1 - \mu(\{f \leq m_f + r\}) \leq \alpha(r)$$

para todo $r \geq 0$.

(2) \implies (1). Para todo $A \subset \mathcal{X}$, la función $f_A : x \rightarrow d(x, A)$ es 1-Lipschitz. Si $\mu(A) \geq \frac{1}{2}$ entonces 0 es una mediana de f_A . Dado que $A^r = \{f_A \leq r\}$:

$$\mu(A^r) = 1 - \mu(\{f_A > r\}) \geq 1 - \alpha(r)$$

para todo $r \geq 0$. ■

Notemos que $-m_f$ es una mediana para $-f$. Si la distribución μ satisface la desigualdad de concentración con perfil α entonces:

$$\mu(\{f > m_f + r\}) \leq \alpha(r) \quad \text{y} \quad \mu(\{f < m_f - r\}) \leq \alpha(r).$$

Combinando ambas desigualdades:

$$\mu(\{|f - m_f| > r\}) = \mu(\{f - m_f > r\}) + \mu(\{f < m_f - r\}) \leq 2\alpha(r) .$$

En particular, aplicando la proposición anterior, si μ satisface la propiedad de concentración de la medida libre de la dimensión con perfil α , entonces, para todo $n \in \mathbb{N}$:

$$\mu^{*n}(\{|f - m_f| > r\}) \leq 2\alpha(r), \quad r \geq 0$$

para toda función $f : \mathcal{X}^n \rightarrow \mathbb{R}$ 1-Lipchitz con respecto a la distancia d_2 .

El siguiente resultado, nos permite describir el fenómeno de concentración de la medida en torno a cualquier constante (en particular, en torno a la media).

Proposición 1.4 [16]

Sea f una función en un espacio de probabilidad de medida μ . Supongamos que existe $a_f \in \mathbb{R}$ y una función α no negativa en \mathbb{R}_+ tal que $\lim_{r \rightarrow \infty} \alpha(r) = 0$, y además:

$$\mu(\{|f - a_f| \geq r\}) \leq \alpha(r)$$

para todo $r > 0$. Entonces:

$$\mu(\{|f - m_f| \geq r + r_0\}) \leq \alpha(r), \quad r > 0,$$

con m_f una mediana de f para μ y con cualquier $r_0 > 0$ tal que $\alpha(r_0) < \frac{1}{2}$.

Si además $\bar{\alpha} = \int_0^\infty \alpha(r) dr < \infty$, entonces f es integrable, $|a_f - \int_{\mathcal{X}} f d\mu| \leq \bar{\alpha}$ y para todo $r > 0$,

$$\mu(\{|f - \int_{\mathcal{X}} f d\mu| \geq r + \bar{\alpha}\}) \leq \alpha(r).$$

1.2.2. Desigualdades de transporte

La técnica de utilizar desigualdades de transporte (también llamadas de información) para demostrar concentración de la medida es un método iniciado por Marton en su artículo sobre concentraciones de Markov [21]. Desde entonces, numerosos autores han generalizado este enfoque.

Sea $\mathcal{P}(\mathcal{X})$ el espacio de medidas de probabilidad. Se definen $\mathcal{P}_1(\mathcal{X})$ y $\mathcal{P}_2(\mathcal{X})$ respectivamente como:

$$\mathcal{P}_1(\mathcal{X}) := \{\mu \in \mathcal{P}(\mathcal{X}) : \int_{\mathcal{X}} d(o, x) d\mu < \infty\}$$

$$\mathcal{P}_2(\mathcal{X}) := \{\mu \in \mathcal{P}(\mathcal{X}) : \int_{\mathcal{X}} d(o, x)^2 d\mu < \infty\}$$

donde $o \in \mathcal{X}$ es un punto arbitrario fijo.

Para un espacio métrico (\mathcal{X}, d) y dos medidas de probabilidad de Borel μ, ν en \mathcal{X} la *distancia de transporte* y la *distancia de transporte cuadrática* se definen respectivamente como:

$$W_1(\mu, \nu) := \inf_{\pi \in \Gamma(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{X}} d(x, y) d\pi(x, y)$$

$$W_2(\mu, \nu) := \sqrt{\inf_{\pi \in \Gamma(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{X}} d(x, y)^2 d\pi(x, y)}$$

con $\Gamma(\mu, \nu) := \{\pi \in P(\mathcal{X} \times \mathcal{X}) \text{ tal que } (\pi_x)_\# \gamma = \mu \text{ y } (\pi_y)_\# \gamma = \nu\}$ y donde π_x y π_y son dos proyecciones de $X \times Y$ en X e Y respectivamente, y donde $\#$ denota el *pushforward* de la medida.

Denotamos $\Gamma_0(\mu, \nu) \subset \Gamma(\mu, \nu)$ como los planes óptimos, donde el mínimo es realizado.

Las cantidades $W_1(\mu, \nu)$ y $W_2(\mu, \nu)$ así definidas son distancias en $\mathcal{P}_1(\mathcal{X})$ y $\mathcal{P}_2(\mathcal{X})$ respectivamente.

Observación : De modo más general, se definen :

$$\mathcal{P}_p(\mathcal{X}) := \{\mu \in \mathcal{P}(\mathcal{X}) : \int_{\mathcal{X}} d(o, x)^p d\mu < \infty\}$$

donde $o \in \mathcal{X}$ es un punto fijo arbitrario.

$$W_p(\mu, \nu) := \inf_{\pi \in \Gamma(\mu, \nu)} \left\{ \int_{\mathcal{X} \times \mathcal{X}} d(x, y)^p d\pi(x, y) \right\}^{\frac{1}{p}} .$$

Las cantidades $W_p(\mu, \nu)$ son distancias en $\mathcal{P}_p(\mathcal{X})$ [7, 18] .

A lo largo de esta sección, en particular para la proposición 1.4, será de gran utilidad la siguiente propiedad denominada *Dualidad de Kantorovich*. Por conveniencia, denotaremos al conjunto de funciones $f : \mathcal{X} \rightarrow \mathbb{R}$ medibles y 1 Lipschitz como $Lip_1(\mathcal{X})$.

Teorema 1.3 (Dualidad de Kantorovich) [17]

$$W_1(\mu_1, \mu_2) = \sup_{f \in Lip_1(\mathcal{X})} \int f d\mu_1 - \int f d\mu_2 .$$

Para toda función $f : \mathcal{X} \rightarrow \mathbb{R}$, medible y no negativa, se define la entropía como:

$$Ent_{\mu}(f) := \begin{cases} \int_{\mathcal{X}} f \log f d\mu - (\int_{\mathcal{X}} f d\mu) \log (\int_{\mathcal{X}} f d\mu), & \text{si } \int_{\mathcal{X}} f \log(1 + f) < \infty \\ +\infty & \text{si no es el caso} \end{cases}$$

Se define la *entropía relativa* o *divergencia de Kullback-Leibler* para cualquier par de medidas de probabilidad ν y μ como:

$$H(\nu | \mu) := \begin{cases} Ent_{\mu}(f) = \int_{\mathcal{X}} f \log f d\mu, & \text{con } f = \frac{d\nu}{d\mu} \text{ si } \nu \ll \mu \\ +\infty, & \text{si no es el caso.} \end{cases}$$

La medida μ satisface una desigualdad de transporte con constante C , abreviada $\mathbf{T}_1(C)$, si:

$$W_1(\mu, \nu) \leq \sqrt{CH(\nu | \mu)} \quad \forall \nu \in \mathcal{P}(\mathcal{X}).$$

De forma análoga, la medida μ satisface una desigualdad de transporte con constante C , abreviada $\mathbf{T}_2(C)$, si:

$$W_2(\mu, \nu) \leq \sqrt{CH(\nu, \mu)} \quad \forall \nu \in \mathcal{P}(X).$$

Tal como lo enuncia el siguiente Teorema, la condición $\mathbf{T}_2(C)$ implica concentración Gaussiana de la medida. Nos restringimos para la demostración a la desigualdad $\mathbf{T}_2(C)$. No obstante, el argumento de Marton también se aplica a la desigualdad $\mathbf{T}_1(C)$.

Teorema 1.4 [21]

Si μ satisface $\mathbf{T}_2(C)$, entonces μ satisface la propiedad de concentración con perfil

$$\alpha(r) = \exp\left(-\frac{1}{C}[r - r_0]_+^2\right), \quad r \geq 0$$

con $r_0 = \sqrt{C \log(2)}$.

Demostración: Sea $A \subset \mathcal{X}$ con $\mu(A) \geq \frac{1}{2}$ y definamos $B = \mathcal{X} \setminus A^r$, $r > 0$. Podemos suponer sin pérdida de generalidad que $\mu(B) > 0$. En efecto, si $\mu(B) = 0$, lo que se desea probar se traduce en :

$$1 - \mu(A^r) \leq \exp\left(-\frac{1}{C}(r - r_0)_+^2\right) \text{ con } \mu(A) \geq \frac{1}{2}$$

$$1 - \exp\left(-\frac{1}{C}(r - r_0)_+^2\right) \leq \mu(A^r) = 1 - \mu(B)$$

lo que se verifica trivialmente si $\mu(B) = 0$. Por ende, supongamos que $\mu(B) > 0$.

Sea $d\mu_A = \frac{1_A}{\mu(A)}d\mu$ y sea $d\mu_B = \frac{1_B}{\mu(B)}d\mu$. Por la desigualdad triangular:

$$W_2(\mu_A, \mu_B) \leq \sqrt{CH(\mu_A | \mu)} + \sqrt{CH(\mu_B | \mu)} .$$

Notemos que:

$$CH(\mu_A | \mu) = -C \log(\mu(A)) .$$

Por hipótesis, $\mu(A) \geq \frac{1}{2}$, lo que equivale a:

$$-\log \mu(A) \leq \log(2) .$$

Luego y utilizando la definición de B :

$$r \leq W_2(\mu_A, \mu_B) \leq \sqrt{C \log(2)} + \sqrt{-C \log(\mu(B))} .$$

Se concluye:

$$\mu(B) \leq \exp\left(\frac{-1}{C}(r - r_0)^2\right) \quad \forall r \geq r_0 = \sqrt{C \log(2)} .$$

■

Observación :

- El ancestro de todas las desigualdades de transporte es la siguiente famosa desigualdad de Csiszar [6], Kullback [15] y Pinsker [29] que compara la variación total con la entropía relativa:

$$\|\mu - \nu\|_{TV}^2 \leq \frac{1}{2} H(\nu, \mu).$$

- La desigualdad $\mathbf{T}_1(C)$ también implica concentración Gaussiana (el argumento de Marton se sigue análogamente). Sin embargo, no se tiene concentración libre de dimensión a diferencia de la desigualdad \mathbf{T}_2 [11] .

La siguiente proposición expresa la existencia de formulación dual equivalente para la desigualdad $\mathbf{T}_1(C)$:

Proposición 1.5 [3]

Sea (\mathcal{X}, d) un espacio métrico, sea $\mu \in \mathcal{P}_1(\mathcal{X})$ y sea $c > 0$. La medida μ verifica la desigualdad de transporte $\mathbf{T}_1(C)$ con $C = 4c$ si y sólo si, para toda función ψ de \mathcal{X} que verifica $\|\psi\|_{Lip} \leq 1$, $\int \psi d\mu = 0$, y para todo t real, se tiene que :

$$\int \exp(t\psi) d\mu \leq \exp(ct^2) .$$

Demostración : Podemos sin pérdida de generalidad restringirnos al caso $t \geq 0$ (cambiando ψ por $-\psi$). La fórmula variacional de la entropía [8] se escribe como :

$$Ent_{\mu}(f) = \sup\left\{ \int fg d\mu, \quad g \text{ tal que } \int e^g d\mu \leq 1 \right\} .$$

Así, para todo $t \geq 0$ y para toda función ψ Lipschitz que verifica $\|\psi\|_{Lip} \leq 1$ y $\int \psi d\mu = 0$,

$$\int \exp(t\psi) d\mu \leq \exp(ct^2)$$

es equivalente a :

$$\int (t\psi - ct^2) f d\mu \leq Ent_{\mu}(f)$$

para toda función f integrable positiva de \mathcal{X} . La implicancia directa es consecuencia directa de la fórmula variacional de la entropía escogiendo $g = t\psi - ct^2$. La recíproca se obtiene escogiendo en la desigualdad anterior $f = \exp(t\psi - ct^2)$. En efecto:

$$\int (t\psi - ct^2) \exp(t\psi - ct^2) d\mu \leq \int \exp(t\psi - ct^2) (t\psi - ct^2) - \int \exp(t\psi - ct^2) \log \left(\int \exp(t\psi - ct^2) \right)$$

$$\int \exp(t\psi - ct^2) \log \left(\int \exp(t\psi - ct^2) \right) \leq 0$$

$$\log \left(\int \exp(t\psi - ct^2) \right) \leq 0$$

Así:

$$\int \exp(t\psi) \leq e^{ct^2} .$$

Notemos que $Ent_{\mu}(\lambda f) = \lambda Ent_{\mu}(f)$.

En efecto:

$$Ent_{\mu}(\lambda f) = \int \lambda f \log(\lambda f) d\mu - \left(\int \lambda f d\mu \right) \log \left(\int \lambda f d\mu \right)$$

$$Ent_{\mu}(\lambda f) = \lambda \int f \log f d\mu + \int \lambda f \log(\lambda) d\mu - \left(\int \lambda f d\mu \right) (\log(\lambda) + \log \left(\int f d\mu \right))$$

$$Ent_{\mu}(\lambda f) = \lambda \int f \log f d\mu - \left(\int \lambda f d\mu \right) \log \left(\int f d\mu \right)$$

$$Ent_{\mu}(\lambda f) = \lambda \left(\int f \log f d\mu - \left(\int f d\mu \right) \log \left(\int f d\mu \right) \right) = \lambda Ent_{\mu}(f) .$$

Luego, supongamos por homogeneidad que f es una densidad de probabilidad con respecto a μ . Dado que la función ψ es de media nula, las siguientes desigualdades son equivalentes para $t > 0$:

$$\int (t\psi - ct^2)f d\mu \leq Ent_{\mu}(f)$$

$$\int (\psi f - \psi) d\mu \leq ct + \frac{1}{t} \int f \log f d\mu .$$

En efecto :

$$\int (t\psi - ct^2)f d\mu \leq Ent_{\mu}(f)$$

\Leftrightarrow

$$\int (t\psi - ct^2)f d\mu \leq \int f \log f d\mu$$

\Leftrightarrow

$$\int t\psi f d\mu \leq \int ct^2 f d\mu + \int f \log f$$

\Leftrightarrow

$$\int \psi f d\mu \leq \int ct f d\mu + \frac{1}{t} \int f \log f d\mu$$

\Leftrightarrow

$$\int (\psi f - \psi) d\mu \leq ct + \frac{1}{t} \int f \log f d\mu .$$

Minimizando en t el término de derecha se obtiene:

$$\int (\psi f - \psi) d\mu \leq \sqrt{4c \int f \log f d\mu}$$

Así, la desigualdad para toda función ψ que verifica $\|\psi\|_{Lip} \leq 1$, $\int \psi d\mu = 0$ y para todo $t \geq 0$,

$$\int \exp(t\psi) d\mu \leq \exp(ct^2)$$

es equivalente a :

$$\int (\psi f - \psi) d\mu \leq \sqrt{4c \int f \log f d\mu}$$

para toda densidad de probabilidad f con respecto a μ . Por el teorema de dualidad de Kantorovich, para toda función ψ que satisface $\|\psi\|_{Lip} \leq 1$ y $\int \psi d\mu = 0$ es equivalente a la desigualdad siguiente:

$$W_1(f d\mu, d\mu) \leq \sqrt{4c Ent_\mu f}$$

$$W_1(\nu, \mu) \leq \sqrt{4c H(\nu | \mu)}$$

lo que es equivalente a $T_1(C)$ y finaliza la demostración. ■

1.2.3. Desigualdades Log-Sobolev y de Poincaré

Sea (\mathcal{X}, d, μ) un espacio métrico de medida. Para una función f localmente Lipschitz, definida en \mathcal{X} a valores en \mathbb{R} , se define el *largo del gradiente* de f en el punto x como:

$$|\nabla f|(x) := \limsup_{y \rightarrow x} \frac{|f(y) - f(x)|}{d(y, x)}$$

Recordemos la definición de la entropía para toda función f , no negativa:

$$Ent_\mu(f) := \begin{cases} \int_{\mathcal{X}} f \log f d\mu - \left(\int_{\mathcal{X}} f d\mu\right) \log \left(\int_{\mathcal{X}} f d\mu\right), & \text{si } \int_{\mathcal{X}} f \log(1+f) < \infty \\ +\infty & \text{si no es el caso} \end{cases}$$

Además, se introducen los siguientes funcionales: la *varianza* y la *energía*.

Para toda función f localmente Lipschitz, se define:

$$Var_\mu(f) := \int_{\mathcal{X}} f^2 d\mu - \left(\int_{\mathcal{X}} f d\mu\right)^2.$$

Para toda función f localmente Lipschitz, se define:

$$\mathcal{E}_\mu(f) = \int_{\mathcal{X}} |\nabla f|^2 d\mu.$$

La medida μ satisface una *desigualdad Log-Sobolev* con constante C , que denotamos **LSI**(C), si para toda función g localmente Lipschitz :

$$Ent_\mu(g^2) \leq 2C\mathcal{E}_\mu(g). \quad (1.4)$$

La medida μ satisface una *desigualdad de Poincaré* con constante C , que denotamos **PI**(C), si para toda función localmente Lipschitz :

$$Var_\mu(f) \leq C\mathcal{E}_\mu(f). \quad (1.5)$$

Teorema 1.5 [16]

1. Si la ecuación (1.4) se satisface, ésta implica *concentración Gaussiana* [16], i.e:

$$\alpha_{(\mathcal{X},d,\mu)}(r) \leq e^{\frac{-r^2}{8C}}, \quad r > 0.$$

2. Si la ecuación (1.5) se satisface, ésta implica *concentración Exponencial* [16], i.e:

$$\alpha_{(\mathcal{X},d,\mu)}(r) \leq e^{\frac{-r}{3\sqrt{C}}}, \quad r > 0.$$

El lector interesado en la segunda propiedad puede leer una demostración en *The Concentration of Measure Phenomenon* [16]. Una de las demostraciones de la primera propiedad es mediante el siguiente argumento atribuido a Herbst.

Demostración : Sea μ una medida que satisface **LSI**(C), i.e :

$$\int g^2 \log(g^2) d\mu - \int g^2 d\mu \log \left(\int g^2 d\mu \right) \leq 2C \int |\nabla g|^2 d\mu$$

para toda función g localmente Lipschitz.

Se escoge $g = \exp(\frac{\lambda f}{2})$, $\lambda > 0$, f 1-Lipschitz. Se define $Z(\lambda) = \int \exp(\lambda f)$.

Remplazando g por la función escogida se obtiene :

$$\lambda \int f \exp(\lambda f) d\mu - \left(\int \exp(\lambda f) d\mu \right) \log \left(\int \exp(\lambda f) d\mu \right) \leq \frac{C\lambda^2}{2} \int |\nabla f|^2 \exp(\lambda f) d\mu .$$

Dado que $|\nabla f| \leq 1$, se tiene que:

$$\lambda Z'(\lambda) - Z(\lambda) \log Z(\lambda) \leq \frac{C\lambda^2}{2} Z(\lambda) .$$

Ahora bien, dividiendo por $\lambda^2 Z(\lambda)$, la expresión deviene:

$$\frac{Z'(\lambda)}{\lambda Z(\lambda)} - \frac{\log Z(\lambda)}{\lambda^2} \leq \frac{C}{2} .$$

Definiendo $K(\lambda) := \frac{1}{\lambda} \log Z(\lambda)$ para $\lambda > 0$, se obtiene:

$$K'(\lambda) \leq \frac{C}{2}$$

Un desarrollo de primer orden asegura la convergencia de $K(\lambda)$ a $\mathbb{E}_\mu(f)$ cuando $\lambda \rightarrow 0$.

Como K es continua en $[0, \lambda]$ y derivable en $]0, \lambda[$:

$$K(\lambda) - K(0) = \int_0^\lambda K'(s) ds \leq \frac{C}{2} \lambda .$$

Así:

$$K(\lambda) \leq \frac{C}{2} \lambda + \mathbb{E}_\mu(f) .$$

Remplazando se obtiene:

$$\frac{1}{\lambda} \log Z(\lambda) \leq \frac{C}{2} \lambda + \mathbb{E}_\mu(f) .$$

Así :

$$Z(\lambda) \leq e^{\frac{C}{2} \lambda^2 + \lambda \mathbb{E}_\mu(f)} .$$

Luego:

$$\mu(f - \mathbb{E}_\mu(f) \geq r) = \mu(e^{\lambda(f - \mathbb{E}_\mu(f))} \geq e^r) .$$

Por la desigualdad de Chebyshev:

$$\mu(f - \mathbb{E}_\mu(f) \geq r) \leq e^{-\lambda r} Z(\lambda) e^{-\mathbb{E}_\mu(f)} .$$

De este modo:

$$\mu(f - \mathbb{E}_\mu(f) \geq r) \leq e^{\frac{C}{2} \lambda^2 - \lambda r} .$$

Optimizando en λ , se obtiene:

$$\mu(f - \mathbb{E}_\mu(f) \geq r) \leq e^{-\frac{r^2}{2C}} .$$

Ahora bien, traduzcamos esta formulación funcional de la concentración de la medida a una formulación geométrica.

Sea $r \geq 0$ y sea $F(x) := \min\{d(x, A), r\}$ que es una función 1-Lipschitz. Se verifica $\mathbb{E}_\mu(F) \leq (1 - \mu(A))r$.

Entonces:

$$1 - \mu(A^r) = \mu(F \geq r) \leq \mu(F \geq \mathbb{E}_\mu(F) + \mu(A)r) \leq e^{-\frac{\mu(A)^2 r^2}{2c}} .$$

En particular, si $\mu(A) \geq \frac{1}{2}$:

$$\mu(A^r) \geq 1 - e^{-\frac{r^2}{8c}} .$$

Se concluye que:

$$\alpha_{(X, d, \mu)} \leq \exp\left(\frac{-r^2}{8C}\right) .$$

■

1.2.4. Relaciones entre distintas desigualdades

Existe una jerarquía natural entre las desigualdades de transporte y las desigualdades Log Sobolev. En efecto, el siguiente Teorema implica que \mathbf{T}_2 está siempre entre \mathbf{PI} y \mathbf{LSI} .

Teorema 1.6 (Otto-Villani [9])

Sea μ una medida de probabilidad absolutamente continua con respecto a la medida de Lebesgue en \mathbb{R}^d :

- Si μ satisface la desigualdad logarítmica de Sobolev $\mathbf{LSI}(C)$ entonces μ satisface $\mathbf{T}_2(C)$
- Si μ satisface $\mathbf{T}_2(C)$ entonces μ satisface la desigualdad de Poincaré $\mathbf{PI}(\frac{C}{2})$

Los dos siguientes teoremas entregan un panorama general de las relaciones entre las desigualdades de transporte y el fenómeno de la concentración de medida.

Teorema 1.7 (Gozlan [10])

Si μ satisface la concentración Gaussiana libre de dimensión con un perfil de la forma:

$$\alpha(r) = \exp\left(-\frac{1}{C}(r - r_0)_+^2\right) \quad r \geq 0$$

entonces μ satisface $\mathbf{T}_2(C)$.

Teorema 1.8 (Gromov y Milman [23], Borovkov y Utev [1])

Sea μ una medida de probabilidad en \mathbb{R}^d . Si μ satisface $\mathbf{PI}(\frac{C}{2})$, entonces μ satisface la concentración de medida libre de dimensión con un perfil exponencial de la forma:

$$\alpha(r) = \exp\left(-\frac{a}{\sqrt{C}}(r - r_0)_+\right)$$

con $r_0 = \sqrt{C}b$ y a , constantes universales numéricas.

Observación :

- Las desigualdades \mathbf{LSI} y \mathbf{T}_2 no son equivalentes. En efecto, Cattiaux y Guillin exhibieron un primer ejemplo de una medida que satisface \mathbf{T}_2 y sin embargo no verifica \mathbf{LSI} [28].
- No obstante la primera observación, existe un refinamiento del Teorema de Otto-Villani. Para enunciarlo, es necesario definir la λ semi-convexidad.

Una función $f : \mathbb{R}^d \rightarrow \mathbb{R}$ es $\lambda \in \mathbb{R}$ semi-convexa si satisface:

$$f((1-t)x_0 + tx_1) \leq (1-t)f(x_0) + tf(x_1) - \frac{\lambda}{2}t(1-t) \|x_0 - x_1\|^2$$

para todo $x_0, x_1 \in \mathbb{R}^n$ y $t \in [0, 1]$. Notemos que cuando $\lambda \geq 0$ la función se dice fuertemente convexa.

Si $f \in \mathcal{C}^1$, λ semi-convexidad es equivalente a :

$$f(y) \geq f(x) + \nabla f(x)(y - x) + \frac{\lambda}{2} \|x - y\|^2 \quad \forall x, y \in \mathbb{R}^n .$$

Si $f \in \mathcal{C}^2$, λ semi-convexidad es equivalente a:

$$\text{Hess} f \geq \lambda .$$

El siguiente Teorema precisa la observación anterior.

Teorema 1.9 [25]

Sea μ una medida de probabilidad en \mathbb{R}^d . Los siguientes enunciados son equivalentes:

- Existe $C > 0$ tal que μ satisface $\mathbf{T}_2(C)$
- Existe $D > 0$ tal que μ satisface la siguiente desigualdad restringida Log-Sobolev, que denotamos $\mathbf{rLSI}(D)$, si : para todo $0 < K < \frac{1}{D}$ y todas las funciones $f : \mathbb{R}^d \rightarrow \mathbb{R}$ que son K semi-convexas, se satisface:

$$\text{Ent}_\mu(e^f) \leq \frac{D}{(1 - DK)^2} \int |\nabla f|^2 e^f d\mu .$$

Además, $D_{opt} \leq C_{opt} \leq e^2 D_{opt}$.

Otra desigualdad que está íntimamente relacionada con el fenómeno de concentración de la medida es la *desigualdad de Prékopa-Leindler*.

Teorema 1.10 (*Desigualdad de Prékopa-Leindler en \mathbb{R}^d*) [2]

Sea $0 < \lambda < 1$ y sean f, g y h funciones no negativas integrables en \mathbb{R}^d que satisfacen:

$$h((1 - \lambda)x + \lambda y) \geq f(x)^{1-\lambda}g(y)^\lambda,$$

para todo $x, y \in \mathbb{R}^d$. Entonces :

$$\int_{\mathbb{R}^d} h(x)dx \geq \left(\int_{\mathbb{R}^d} f(x)dx \right)^{1-\lambda} \left(\int_{\mathbb{R}^d} g(x)dx \right)^\lambda.$$

La *desigualdad de Brunn-Minkowski* puede ser vista como consecuencia directa de la desigualdad de Prékopa-Leindler.

Teorema 1.11 (*Desigualdad de Brunn-Minkowski en \mathbb{R}^d*) [27]

Sean X e Y conjuntos acotados y medibles en \mathbb{R}^d tal que $(1 - \lambda)X + \lambda Y$ sea también medible. Denotemos V_d la medida de Lebesgue en \mathbb{R}^d . Entonces:

$$V_d((1 - \lambda)X + \lambda Y) \geq V_d(X)^{1-\lambda}V_d(Y)^\lambda$$

Demostración: Sea $h = 1_{(1-\lambda)X+\lambda Y}$, $f = 1_X$ y $g = 1_Y$. La hipótesis de Prékopa-Leindler se verifica trivialmente. Así, se concluye inmediatamente:

$$V_d((1-\lambda)X+\lambda Y) = \int_{\mathbb{R}^d} 1_{(1-\lambda)X+\lambda Y}(x)dx \geq \left(\int_{\mathbb{R}^d} 1_X(x)dx \right)^{1-\lambda} \left(\int_{\mathbb{R}^d} 1_Y(x)dx \right)^\lambda = V_d(X)^{1-\lambda}V_d(Y)^\lambda$$

■

Observación : Notemos que la hipótesis que $(1 - \lambda)X + \lambda Y$ sea medible es necesaria. En efecto, un ejemplo de Sierpiński [34] muestra que el conjunto $(1 - \lambda)X + \lambda Y$ no es necesariamente medible cuando X e Y lo son.

Corolario 1.1 (*La desigualdad de Brunn-Minkowski, en su forma estándar*) [27]

Sea $0 < \lambda < 1$ y sean X e Y conjuntos no vacíos medibles y acotados en \mathbb{R}^d , tal que $(1 - \lambda)X + \lambda Y$ sea también medible. Entonces:

$$V_d((1 - \lambda)X + \lambda Y)^{\frac{1}{d}} \geq (1 - \lambda)V_d(X)^{\frac{1}{d}} + \lambda V_d(Y)^{\frac{1}{d}}$$

Demostración: Sea:

$$\lambda' = \frac{V_d(Y)^{\frac{1}{d}}}{V_d(X)^{\frac{1}{d}} + V_d(Y)^{\frac{1}{d}}}$$

Sean $X' = V_n(X)^{-\frac{1}{d}}X$ y $Y' = V_d(Y)^{-\frac{1}{d}}Y$. Entonces:

$$1 - \lambda' = \frac{V_d(X)^{\frac{1}{d}}}{V_d(X)^{\frac{1}{d}} + V_n(Y)^{\frac{1}{d}}}$$

Por homogeneidad de la medida de Lebesgue en \mathbb{R}^d : $V_d(X') = V_d(Y') = 1$.

Luego, por el Teorema de Brunn-Minkowski:

$$V_d((1 - \lambda')X' + \lambda'Y') \geq 1.$$

Notemos :

$$V_d((1 - \lambda')X' + \lambda'Y') = V_d\left(\frac{X + Y}{V_d(X)^{\frac{1}{d}} + V_n(Y)^{\frac{1}{d}}}\right) = \frac{V_d(X + Y)}{(V_d(X)^{\frac{1}{d}} + V_n(Y)^{\frac{1}{d}})^d}$$

Así:

$$V_d(X + Y)^{\frac{1}{d}} \geq V_d(X)^{\frac{1}{d}} + V_n(Y)^{\frac{1}{d}}$$

Se concluye, reemplazando X e Y por $(1 - \lambda)X$ y λY respectivamente. ■

El siguiente Teorema relaciona la convexidad uniforme y el fenómeno de concentración de la medida.

Teorema 1.12 (Gromov-Milman) [22]

Sea K un cuerpo convexo simétrico de \mathbb{R}^d . Supongamos además que K es uniformemente convexo y denotemos por δ su módulo de convexidad. Sea d la distancia asociada a la norma $\|\cdot\|$ inducida por K y sea μ la medida uniforme normalizada de K , es decir, $\mu(A) = \frac{V_d(A)}{V_d(K)}$ para todo $A \subset K$. Entonces, (K, d, μ) satisface el siguiente perfil de concentración :

$$\alpha(\varepsilon) = 2e^{-2n\delta(\varepsilon)}$$

Demostración: Recordemos que si K es uniformemente convexo si existe para todo $\varepsilon > 0$ un $\delta(\varepsilon) > 0$ tal que:

$$\forall x, y \in K, \|x - y\| \geq \varepsilon \implies \left\| \frac{x + y}{2} \right\| \leq 1 - \delta(\varepsilon) \quad (1.6)$$

Sea $A \subset K$ de medida mayor a $\frac{1}{2}$. Sea $B = \{x \in K : d(x, A) > \varepsilon\}$. La ecuación (1.6) implica :

$$\frac{A + B}{2} \subset (1 - \delta(\varepsilon))K.$$

Luego, por Brunn-Minkowski:

$$\mu\left(\frac{A + B}{2}\right) \geq \mu(A)^{\frac{1}{2}}\mu(B)^{\frac{1}{2}}.$$

Así:

$$\mu(B) \leq 2(1 - \delta(\varepsilon))^{2n} \leq 2e^{-2n\delta(\varepsilon)}$$

$$1 - \mu(A_\varepsilon) \leq 2e^{-2n\delta(\varepsilon)} .$$

■

1.2.5. Concentración de la medida gaussiana vía la desigualdad de Prékopa- Leindler

El siguiente Teorema relaciona la desigualdad de Prékopa-Leindler con la concentración de la medida Gaussiana.

Teorema 1.13 [4]

Sea $A \in \mathbb{R}^d$ un conjunto medible y sea γ la medida gaussiana estándar en \mathbb{R}^d . Entonces:

$$\int_{\mathbb{R}^d} e^{\frac{d(x,A)^2}{4}} d\gamma \leq \frac{1}{\gamma(A)} .$$

Por ende, si $\gamma(A) \geq \frac{1}{2}$, entonces:

$$\gamma(A^\varepsilon) \geq 1 - 2e^{-\frac{\varepsilon^2}{4}} .$$

Demostración Se definen u , v y h en \mathbb{R}^d :

$$u(x) := e^{\frac{d(x,A)^2}{4}} \gamma(x)$$

$$v(x) := 1_A(x) \gamma(x)$$

$$h(x) := \gamma(x)$$

con γ la densidad gaussiana.

El enunciado a probar es:

$$\left(\int_{\mathbb{R}^d} e^{\frac{d(x,A)^2}{4}} d\gamma \right) \gamma(A) \leq 1 ,$$

que se traduce inmediatamente en la desigualdad:

$$\left(\int_{\mathbb{R}^d} u\right)\left(\int_{\mathbb{R}^d} v\right) \leq \left(\int_{\mathbb{R}^d} h\right)^2 .$$

Por la desigualdad de Prékopa-Leindler (con $\lambda = \frac{1}{2}$), es suficiente verificar que para todo $x, y \in \mathbb{R}^d$:

$$u(x)v(y) \leq h\left(\frac{x+y}{2}\right)^2 .$$

Es suficiente verificar para $y \in A$, de lo contrario $v(y) = 0$. En este caso, $d(x, A) \leq |x - y|$. Así:

$$(2\pi)^d u(x)v(y) = e^{\frac{d(x,A)^2}{4}} e^{-\frac{x^2}{2}} e^{-\frac{y^2}{2}} \leq \exp\left(\frac{|x-y|^2}{4} - \frac{|x|^2}{2} - \frac{|y|^2}{2}\right) = \exp\left(-\frac{|x+y|^2}{4}\right)$$

Ahora bien :

$$\exp\left(-\frac{|x+y|^2}{4}\right) = \left(\exp\left(-\frac{1}{2} \left|\frac{x+y}{2}\right|^2\right)\right)^2 = \left(\exp\left(-\frac{1}{2} \left|\frac{x+y}{2}\right|^2\right) \frac{1}{(2\pi)^{\frac{d}{2}}}\right)^2 (2\pi)^d = h\left(\frac{x+y}{2}\right)(2\pi)^d .$$

Lo anterior demuestra la primera aseveración.

De este modo, con las hipótesis del teorema:

$$\left(\int_{\mathbb{R}^d} e^{\frac{d(x,A)^2}{4}} d\gamma\right)\gamma(A) \leq 1 .$$

Si nos restringimos a $(A^\varepsilon)^c = \{x : d(x, A) > \varepsilon\}$, se obtiene:

$$\left(\int_{(A^\varepsilon)^c} e^{\frac{d(x,A)^2}{4}} d\gamma\right)\gamma(A) \leq 1 .$$

Luego:

$$\left(\int_{(A^\varepsilon)^c} e^{\frac{\varepsilon^2}{4}} d\gamma\right)\gamma(A) \leq 1 ,$$

$$e^{\frac{\varepsilon^2}{4}} (1 - \gamma(A^\varepsilon))\gamma(A) \leq 1 .$$

Sea $\gamma(A) \geq \frac{1}{2}$, se obtiene:

$$\gamma(A^\varepsilon) \geq 1 - 2e^{-\frac{\varepsilon^2}{4}} .$$

■

El siguiente Teorema relaciona, en ciertos casos, la desigualdad de Prékopa-Leindler con la desigualdad Log-Sobolev.

Teorema 1.14 (Bobkov y Ledoux) [33]

Sea $E = (\mathbb{R}^d, \|\cdot\|)$ un espacio normado finito dimensional y sea $E^* = (\mathbb{R}^d, \|\cdot\|_*)$ su dual. Consideremos una medida de probabilidad μ en E de densidad $\rho e^{-V(x)}$ con V una función convexa definida en un subconjunto convexo y abierto Ω de E . Supongamos además que V es λ semi-convexa, es decir:

$$V(tx + (1-t)y) \leq tV(x) + (1-t)V(y) - \frac{\lambda}{2}t(1-t) \|x - y\|^2 \quad (1.7)$$

Entonces, para toda $f \in C^\infty(\Omega)$:

$$Ent_\mu(f^2) \leq \frac{2}{\lambda} \int \|\nabla f\|_*^2 d\mu .$$

Observación : Si $\mu = \gamma$ entonces $V(x) = \frac{|x|^2}{2}$, y la condición (1.7) se verifica con $\lambda = 2$. Dado que $\|\cdot\|_* = |\cdot|$,

$$Ent_\gamma(f^2) \leq \int |\nabla f|^2 d\gamma .$$

El siguiente esquema recapitula las relaciones entre ciertas desigualdades y el fenómeno de concentración de la medida.

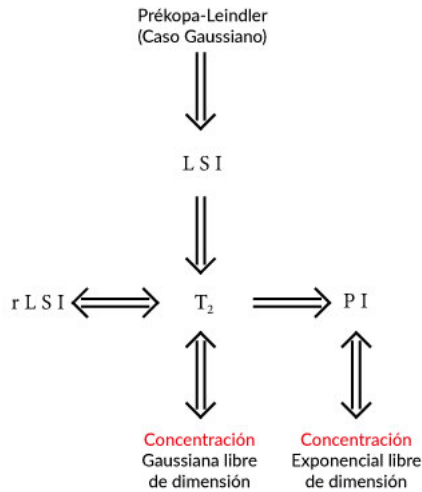


Figura 1.2: Síntesis recapitulativa de la sección 1.24: algunas relaciones entre desigualdades funcionales y el fenómeno de concentración de la medida .

1.2.6. Isoperimetría y concentración de la medida

La isoperimetría clásica relaciona la *medida del borde* de un conjunto con la medida total del conjunto. Sea (\mathcal{X}, d) un espacio métrico con medida de Borel positiva μ , se define la

medida del borde de un conjunto de Borel $A \subset \mathcal{X}$ como:

$$\mu^+(A) := \liminf_{r \rightarrow 0} \frac{1}{r} \mu(A^r \setminus A)$$

La función isoperimétrica $\mathcal{I}_\mu : [0, \mu(\mathcal{X})] \rightarrow \mathbb{R}_+$ se define para cada conjunto de Borel $A \subset X$ con $\mu(A) < \infty$ como el máximo valor que satisface :

$$\mu^+(A) \geq \mathcal{I}_\mu(\mu(A)) \quad (1.8)$$

Todo conjunto que satisface la desigualdad (1.8) con igualdad minimiza la *medida al borde* entre todos los conjuntos A tales que $\mu(A) = \mu(B)$ y se denomina *conjunto extremal*. En otras palabras, si A es un conjunto medible con el mismo volumen que un *conjunto extremal* B , entonces la *medida al borde* de A es mayor o igual a la *medida al borde* de B .

Se puede obtener concentración de la medida a partir de isoperimetría imponiendo *hipótesis débiles* para μ y asumiendo la existencia de una función estrictamente creciente y diferenciable $v : \mathbb{R} \rightarrow [0, \mu(X)]$ que satisface $\mathcal{I}_\mu \geq v' \circ v^{-1}$. Con estas hipótesis, podemos acotar la función de concentración [16]:

$$\alpha(r) \leq 1 - v(v^{-1}(\frac{1}{2}) + r).$$

La función isoperimétrica \mathcal{I}_μ es difícil de calcular en general y es explícitamente conocida en pocos casos.

El ejemplo de la medida Gaussiana:

Sea γ^d la medida Gaussiana estándar en el espacio \mathbb{R}^d , equipado con la distancia estándar Euclideana.

Problema isoperimétrico: ¿Cuáles son los conjuntos extremales para la medida Gaussiana en el espacio \mathbb{R}^d , equipado con la distancia estándar Euclideana ?

Teorema 1.15 [16]

Los semi-espacios son soluciones del problema isoperimétrico Gaussiano. En otras palabras, si $H = \{x \in \mathbb{R}^d : \langle x, u \rangle < a\}$ con $|u| = 1$, $a \in \mathbb{R}$ y si $\gamma^d(A) = \gamma^d(H)$, entonces:

$$(\gamma^d)^+(A) \geq (\gamma^d)^+(H) .$$

Sea $\Phi(t) = (2\pi)^{\frac{1}{2}} \int_{-\infty}^t e^{-\frac{x^2}{2}} dx$, $t \in [-\infty, +\infty]$, la distribución de la medida canónica Gaussiana. Entonces $\gamma^d(H) = \Phi(a)$ y $(\gamma^d)^+(H) = \Phi'(a)$.

Así:

$$(\gamma^d)^+(A) \geq \Phi'(a) = \Phi' \circ \Phi^{-1}(\gamma^d(A)).$$

La desigualdad anterior posee la siguiente formulación integral.

Si $\gamma^d(A) = \gamma^d(H) = \Phi(a)$ entonces, para todo $r \geq 0$:

$$\gamma(A^r) \geq \gamma(H^r) = \Phi(a+r) .$$

En particular, si $\gamma(A) \geq \frac{1}{2} = \Phi(0)$:

$$\gamma(A^r) \geq \Phi(r) \geq 1 - e^{-\frac{r^2}{2}}, \quad r \geq 0.$$

1.2.7. La transformada de Laplace, el *largo* de un espacio métrico finito y el ejemplo del hipercubo

La transformada de Laplace

Sea (\mathcal{X}, d, μ) un espacio métrico de medida. Se define, para todo $\lambda \geq 0$, la *transformada de Laplace* de μ en \mathcal{X} como:

$$Z_{(\mathcal{X}, d, \mu)}(\lambda) = \sup \int e^{\lambda f}$$

donde el supremo actúa sobre todas las funciones f acotadas, con media cero y 1-Lipschitz en (\mathcal{X}, d) .

Proposición 1.6 [16]

Si $Z_{(\mathcal{X}, d, \mu)}(\lambda) \leq e^{\frac{\lambda^2}{2c}}$, para $\lambda \geq 0$, entonces toda función 1-Lipschitz $f : \mathcal{X} \rightarrow \mathbb{R}$ es integrable y para todo $r \geq 0$:

$$\mu(\{f \geq \int f d\mu + r\}) \leq e^{-\frac{cr^2}{2}}.$$

Proposición 1.7 [16]

Si $\text{Diam}(X, d) = D < \infty$, entonces para toda distribución de probabilidad μ en (\mathcal{X}, d) :

$$Z_{(\mathcal{X}, d, \mu)}(\lambda) \leq e^{\frac{D^2 \lambda^2}{2}}, \quad \lambda \geq 0.$$

Una consecuencia de estas proposiciones es el siguiente corolario:

Corolario 1.2 [16]

Sea $P = \mu_1 \otimes \cdots \otimes \mu_n$ un producto de medidas de probabilidad en un producto Cartesiano $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ de espacios métricos (\mathcal{X}_i, d_i) con diámetros finitos D_i , para $i = 1, \dots, n$

equipado con la métrica $d = \sum_{i=1}^N d_i$. Entonces, si f es una función 1-Lipschitz en (\mathcal{X}, d) , para todo $r \geq 0$:

$$P(\{f \geq \int f dP + r\}) \leq e^{\frac{-r^2}{2D^2}}$$

con $D^2 = \sum_{i=1}^n D_i^2$.

Sea \mathcal{Q}_n el hipercubo. Este ejemplo será de particular importancia. En el próximo capítulo, volveremos a demostrar concentración de la medida en este espacio vía la curvatura de Ricci discreta.

El ejemplo del hipercubo:

Sea \mathcal{Q}_n el hipercubo dotado de la medida uniforme y equipado de la distancia de Hamming.

Como aplicación del corolario anterior, consideremos la métrica trivial en cada \mathcal{X}_i , se obtiene así en $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_N$ la *métrica de Hamming*:

$$d(x, y) = \sum_{i=1}^N 1_{\{x_i \neq y_i\}} = \text{Card}\{1 \leq i \leq N; x_i \neq y_i\}$$

Sea $\mathcal{Q}_n = \{0, 1\}^N$ el hipercubo, equipado con la métrica de Hamming, dotado con cualquier medida de probabilidad P . Notemos que $D^2 = \sum_{i=1}^n D_i^2 = N$. De este modo, se obtiene concentración de la medida para el hipercubo:

$$P(\{f \geq \int f dP + r\}) \leq e^{\frac{-r^2}{2N}}$$

para toda función 1-Lipschitz en (\mathcal{Q}_n, d) .

El largo de un espacio métrico finito

Sea \mathcal{X} un conjunto finito. La filtración $\{\mathcal{F}_i\}_{i=0}^m$ es finita y consideramos $\mathcal{F} = \mathcal{F}_m = \mathcal{P}(X)$. Identificamos naturalmente una sub- σ filtración \mathcal{F}_i con una partición $\mathcal{P}_i = \{A_1, \dots, A_q\}$ de X en los átomos de \mathcal{F}_i . Decimos que una partición \mathcal{P} refina una partición \mathcal{P}' si $\forall A \in \mathcal{P}, A \subset B$ para algún $B \in \mathcal{P}'$. Se escribe $\mathcal{P}' \prec \mathcal{P}$. Notemos que podemos identificar una filtración $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_m$ con una partición asociada: $\mathcal{P}_0 \prec \mathcal{P}_1 \prec \dots \prec \mathcal{P}_m$.

Schechtman introdujo la idea de *largo* de un espacio métrico. Sea (\mathcal{X}, d) un espacio métrico finito, equipado con la medida de probabilidad uniforme. El espacio métrico (\mathcal{X}, d) tiene un *largo de a lo más l* si :

- Existen $c_1, \dots, c_m > 0$ con $(\sum_{i=1}^m c_i^2)^{\frac{1}{2}} = l$
- Existe una secuencia de particiones $\mathcal{P}_0 \prec \dots \prec \mathcal{P}_m$ de \mathcal{X} , con \mathcal{P}_0 trivial y \mathcal{P}_m discreta, tal que si $A, B \in \mathcal{P}_i$ con $A \cup B \subset C$ para algún $C \in \mathcal{P}_{i-1}$ entonces $\mathbb{P}(A) = \mathbb{P}(B)$ y existe una biyección $\phi : A \rightarrow B$ con $d(x, \phi(x)) \leq c_i$ para todo $x \in A$.

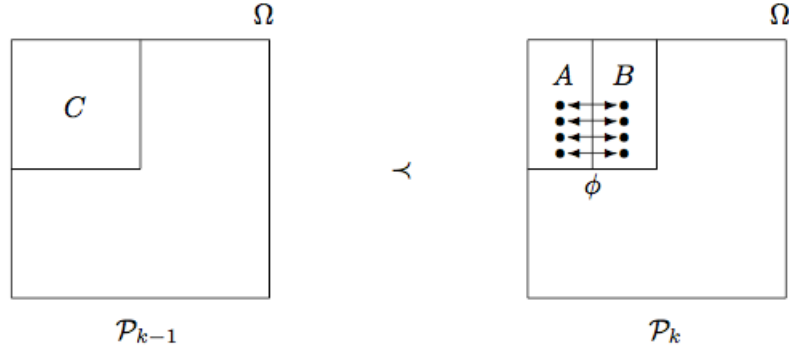


Figura 1.3: Ilustración del largo de un espacio métrico finito. (Referencia: Discrete Isoperimetric Inequalities, de Henry Liu.)

Teorema 1.16 [36]

Sea (\mathcal{X}, d) un espacio métrico finito con largo a lo más l . Sea $f : \mathcal{X} \rightarrow \mathbb{R}$ una función 1-Lipschitz. Entonces, para todo $a \geq 0$:

$$\mathbb{P}(f \geq \mathbb{E}(f) + a) \leq \exp\left(\frac{-a^2}{2l^2}\right)$$

$$\mathbb{P}(f \leq \mathbb{E}(f) - a) \leq \exp\left(\frac{-a^2}{2l^2}\right).$$

Ejemplo (El hipercubo)

El hipercubo \mathcal{Q}_n tiene *largo* a lo más $n^{\frac{1}{2}}$.

Demostración: Consideremos los elementos de \mathcal{Q}_n como secuencias de largo n . Definamos la partición \mathcal{P}_i de \mathcal{Q}_n inducida por la relación de equivalencia \sim_i :

$$x = (x_1, \dots, x_n) \sim_i y = (y_1, \dots, y_n)$$

si y sólo si $x_j = y_j \ \forall j \leq i$.

De esta forma, $\mathcal{P}_0 \prec \dots \prec \mathcal{P}_n$ con \mathcal{P}_0 trivial y \mathcal{P}_n discreta.

Sean $A, B \in \mathcal{P}_i, A \neq B$ y $A \cup B \subset C$ para algún $C \in \mathcal{P}_{i-1}$. Sin pérdida de generalidad, sean:

$$A = \{x \in \{0, 1\}^n : x_j = a_j \text{ para } j < i \text{ y } x_i = 0\}$$

$$B = \{x \in \{0, 1\}^n : x_j = a_j \text{ para } j < i \text{ y } x_i = 1\}$$

para ciertos $a_1 \dots a_{i-1} \in \{0, 1\}$.

Entonces, $|A| = |B|$ y por ende: $\mathbb{P}(A) = \mathbb{P}(B)$. Sea $\phi : A \rightarrow B$ la biyección que *cambia* la i -ésima coordenada x de A , i.e:

$$\phi(x) := \begin{cases} x_j, & \text{para } j \neq i \\ 1 + x_j \pmod{2}, & \text{si } j = i. \end{cases}$$

Entonces ϕ satisface : $d(x, \phi(x)) \leq 1$ para todo $x \in A$. Podemos escoger $c_i = 1$ para todo i y luego \mathcal{Q}_n tiene *largo* a lo más $n^{\frac{1}{2}}$. ■

Observación :

- Se puede verificar que $n^{\frac{1}{2}}$ es la mejor constante.
- Por el Teorema 1.16, se obtiene entonces el fenómeno de *concentración de la medida* para el hipercubo.
- Alternativamente, se puede demostrar la existencia del fenómeno de *concentración de la medida* vía una desigualdad isoperimétrica llamada *la desigualdad de Harper*.

Capítulo 2

La curvatura de Ricci discreta

2.1. Introducción a la curvatura de Ricci discreta

Este capítulo se basa en la curvatura de Ricci discreta introducida en el paper *Ricci curvature of Markov chains on metric spaces* por Yann Ollivier [26].

2.1.1. Definición de curvatura de Ricci discreta

Consideramos una cadena de Markov $(X_n)_{n \in \mathbb{N}}$ en un espacio polaco (\mathcal{X}, d) . El kernel de transición es denotado $(m_x)_{x \in \mathcal{X}}$ con m_x una medida de probabilidad en \mathcal{X} , tal que $m_x(dy)$ es la probabilidad de transición de x a y . El n -ésimo kernel de transición es definido inductivamente:

$$m_x^n(A) := \int_{\mathcal{X}} m_z(A) m_x^{n-1}(dz)$$

para todo conjunto $A \subset \mathcal{X}$ con $m_x^1 = m_x$.

Una medida ν se dice *invariante* para la cadena de Markov m si :

$$\nu(A) = \int_{\mathcal{X}} m_y(A) d\nu(y)$$

para todo $A \subset \mathcal{X}$.

Para toda función $f \in L^2(\mathcal{X}, d, \nu)$ se define *el operador de media* como:

$$Mf(x) := \mathbb{E}_x f = \int_{\mathcal{X}} f(y) m_x(dy).$$

Se verifica que para todo $x \in \mathcal{X}$ y para toda función f acotada en \mathcal{X} :

$$\lim_{k \rightarrow \infty} M^k f(x) = \mathbb{E}_\nu(f) \quad \forall x \in \mathcal{X}$$

con ν la medida invariante.

Del mismo modo que en el capítulo anterior denotamos $\mathcal{P}^1(X)$ el espacio de medidas de probabilidad μ en \mathcal{X} con primer momento finito. Es decir, para algún $o \in X$:

$$\int d(o, x) d\mu(x) < \infty .$$

En $\mathcal{P}^1(\mathcal{X})$, la función W_1 es finita, y es una distancia. En lo que sigue asumiremos que las medidas de probabilidad pertenecen a $\mathcal{P}^1(\mathcal{X})$. Recordemos además la dualidad de Kantorovich:

$$W_1(\mu_1, \mu_2) = \sup_{f \in \text{Lip}_1(\mathcal{X})} \int f d\mu_1 - \int f d\mu_2 .$$

Definición 2.1 Sea (\mathcal{X}, d) un espacio métrico con una cadena de Markov m . Sean $x, y \in \mathcal{X}$ dos puntos distintos. La Curvatura de Ricci discreta se define como:

$$\mathcal{K}(x, y) := 1 - \frac{W_1(m_x, m_y)}{d(x, y)} .$$

Se define \mathcal{K} como:

$$\mathcal{K} = \inf_{x, y \in \mathcal{X}} \mathcal{K}(x, y) .$$

La curvatura de Ricci discreta es una técnica que permite demostrar concentración de la medida. Además, en algunos casos, provee mejores cotas para los mixing times.

2.1.2. Espacios geodésicos

Una idea subyacente a la curvatura es que permite a partir de propiedades locales obtener propiedades globales.

Decimos que (\mathcal{X}, d) es un espacio ε -geodésico si para todo para de puntos $x, y \in \mathcal{X}$ existe un entero n y una secuencia $x = x_0, x_1, \dots, x_n = y$ tal que se satisfacen las dos siguientes condiciones:

$$d(x, y) = \sum_{i=0}^{n-1} d(x_i, x_{i+1})$$

$$d(x_i, x_{i+1}) \leq \varepsilon .$$

Ejemplo En teoría de grafos se denomina distancia entre dos vértices de un grafo al número de vértices mínimo que debe recorrerse para unirlos. Con esta distancia, que denotaremos d_G , un grafo es un espacio 1-geodésico.

Dada la siguiente proposición, a menudo es suficiente calcular la curvatura de Ricci de dos puntos vecinos.

Proposición 2.1 [26]

Si (\mathcal{X}, d) es un espacio ε geodésico y $\mathcal{K}(x, y) \geq \mathcal{K}$ para todo par de puntos con $d(x, y) \leq \varepsilon$, entonces $\mathcal{K}(x, y) \geq \mathcal{K}$ para todo par de puntos $x, y \in \mathcal{X}$.

Demostración:

$$W_1(m_x, m_y) \leq \sum_{i=0}^{n-1} W_1(m_{x_i}, m_{x_{i+1}}) \leq (1 - \mathcal{K}) \sum_{i=0}^{n-1} d(x_i, x_{i+1}) \leq (1 - \mathcal{K})d(x, y).$$



2.1.3. Algunos ejemplos

Los siguientes tres ejemplos corresponden a paseos aleatorios en espacios métricos con curvatura discreta nula, positiva y negativa respectivamente.

2.1.4. El ejemplo de \mathbb{Z}^N y \mathbb{R}^N

Sea m una cadena de Markov en el grafo constituido por el *grid* \mathbb{Z}^N equipado con la métrica del grafo d_G , definida como:

$$m_x(y) := \begin{cases} \frac{1}{2^N}, & \text{si } d_G(x, y) = 1 \\ 0, & \text{si no es el caso.} \end{cases}$$

Entonces, para todo punto $x, y \in \mathbb{Z}^d$, la curvatura de Ricci asociada $\mathcal{K}(x, y)$ es 0.

En efecto, se puede trasladar la medida m_x a la medida m_y por una traslación del vector $y - x$ (que es optimal), así la distancia entre m_x y m_y es exactamente la distancia entre x e y .

Este ejemplo se generaliza al caso de \mathbb{Z}^N o \mathbb{R}^N equipado con cualquier distancia y cualquier paseo aleatorio que sea invariante mediante traslaciones. Por ejemplo, el *tiling* triangular del plano, tiene curvatura cero.

2.1.5. El ejemplo del hipercubo

La distancia de Hamming entre dos secuencias de mismo largo es el número de posiciones en las que éstas dos difieren. Consideremos el hipercubo $\mathcal{Q}_N = \{0, 1\}^N$ equipado con la distancia de Hamming. Definamos la cadena de Markov siguiente:

$$m_x(y) := \begin{cases} \frac{1}{N+1}, & \text{si } y \in B_x(1) \\ 0, & \text{si no es el caso .} \end{cases} \quad (\star)$$

con $B_x(r)$ la bola cerrada de centro x y de radio r .

Examinemos los vecinos $x = (0, 0, \dots, 0), y = (1, 0, 0, \dots, 0)$. Los vecinos de x son los puntos: $x_1 = y = (1, 0, \dots, 0), x_2 = (0, 1, \dots, 0)$, y así sucesivamente $x_N = (0, 0, \dots, 1)$. De manera análoga, los vecinos de y son los puntos: $y_1 = x = (0, 0, \dots, 0), y_2 = (1, 1, 0, \dots, 0)$ y así sucesivamente $y_N = (1, 0, \dots, 1)$.

Notemos que para $2 \leq i \leq N$, $d(x_i, y_i) = 1$ y $d(x_1, y) = d(x, y_1) = 0$.

Calculemos ahora $W_1(m_x, m_y)$ vía el Teorema de la dualidad de Kantorovich. Para ello, notemos que si $f : X \rightarrow \mathbb{R}$:

$$\int f dm_x = \frac{f(x) + f(x_1) + \dots + f(x_N)}{N+1}$$

$$\int f dm_y = \frac{f(y) + f(y_1) + \dots + f(y_N)}{N+1}.$$

El Teorema de dualidad de Kantorovich nos sugiere calcular:

$$\int f dm_x - \int f dm_y = \frac{(f(x) - f(y_1)) + (f(x_1) - f(y)) + (f(x_2) - f(y_2)) + \dots + (f(x_N) - f(y_N))}{N+1}.$$

Dado que $f(x) - f(y_1) = f(x_1) - f(y) = 0$, se obtiene:

$$\int f dm_x - \int f dm_y = \frac{(f(x_2) - f(y_2)) + \dots + (f(x_N) - f(y_N))}{N+1}.$$

Luego, tomando el supremo de todas las funciones f 1- Lipschitz : $\sup_f (f(x_i) - f(y_i)) \leq d(x_i, y_i) = 1$ para $2 \leq i \leq N$.

Escogiendo la función 1-Lipschitz, $f : \mathcal{Q}_N \rightarrow \mathbb{R}$, como:

$$f(x^1, x^2, \dots, x^N) = -x^1$$

se obtiene la igualdad: $\sup_f (f(x_i) - f(y_i)) = d(x_i, y_i) = 1$ para $2 \leq i \leq N$.

Así, $\sup_f (f(x_i) - f(y_i)) = d(x_i, y_i) = 1$ para $2 \leq i \leq N$.

Por ende:

$$W_1(m_x, m_y) = \sup_{f \in Lip_1(Q_N)} \left(\int f dm_x - \int f dm_y \right) = \frac{N-1}{N+1}.$$

con lo cual:

$$\mathcal{K} = \frac{2}{N+1} > 0.$$

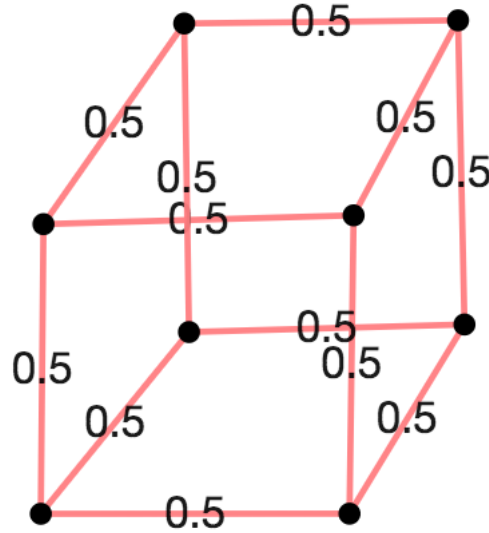


Figura 2.1: Ilustración de la curvatura de Ricci discreta para la cadena de Markov \star en \mathcal{Q}_3 .

2.1.6. El ejemplo de los árboles

Sea $G = (V, E)$ un grafo localmente finito, es decir, un grafo en el que cada nodo tiene grado finito. Sea g_x el grado del vértice x . Utilizamos la métrica del grafo d_G introducida anteriormente. Para vértices vecinos x, y , $d_G(x, y) = 1$ y se denota $x \sim y$. Notemos que un vértice x no es vecino de si mismo. Para todo vértice $x \in V$, consideramos la probabilidad de transición siguiente:

$$m_x(y) = \begin{cases} \frac{1}{g_x}, & \text{si } y \sim x \\ 0, & \text{si no es el caso.} \end{cases}$$

Teorema 2.1 [14]

En un grafo localmente finito $G = (V, E)$ se tiene para todo par de vértices vecinos x e y :

$$\mathcal{K}(x, y) \geq -2\left(1 - \frac{1}{g_x} - \frac{1}{g_y}\right)_+ = \begin{cases} -2 + \frac{2}{g_x} + \frac{2}{g_y}, & \text{si } g_x > 1 \text{ y } g_y > 1, \\ 0 & \text{si no es el caso.} \end{cases}$$

Demostración: Sean x e y tales que $x \sim y$. Así:

$$\mathcal{K}(x, y) = 1 - W_1(m_x, m_y).$$

Utilizando el Teorema de dualidad de Kantorovich:

$$W_1(m_x, m_y) = \sup_{f \in \text{Lip}_1(V)} \left(\frac{1}{g_x} \sum_{z, z \sim x} f(z) - \frac{1}{g_y} \sum_{z', z' \sim y} f(z') \right)$$

$$\begin{aligned} W_1(m_x, m_y) &= \sup_{f \in \text{Lip}_1(V)} \left(\frac{1}{g_x} \sum_{z, z \sim x, z \neq y} (f(z) - f(x)) - \frac{1}{g_y} \sum_{z', z' \sim y, z' \neq x} (f(z') - f(y)) \right) \\ &\quad + \frac{1}{g_x} (f(y) - f(x)) - \frac{1}{g_y} (f(x) - f(y)) + f(x) - f(y) \end{aligned}$$

$$W_1(m_x, m_y) \leq \frac{g_x - 1}{g_x} + \frac{g_y - 1}{g_y} + \left| 1 - \frac{1}{g_x} - \frac{1}{g_y} \right| = 1 + 2\left(1 - \frac{1}{g_x} - \frac{1}{g_y}\right)_+$$

Luego:

$$\mathcal{K}(x, y) \geq -2\left(1 - \frac{1}{g_x} - \frac{1}{g_y}\right)_+.$$

■

Notemos que los árboles alcanzan esta cota inferior.

Proposición 2.2 [14]

Consideremos un árbol $T = (V, E)$. Entonces para todo par de vecinos x, y se tiene:

$$\mathcal{K}(x, y) = -2\left(1 - \frac{1}{g_x} - \frac{1}{g_y}\right)_+.$$

Demostración: Dado el Teorema anterior, nos basta probar que $1 + 2\left(1 - \frac{1}{g_x} - \frac{1}{g_y}\right)_+$ es también una cota inferior para W_1 . Si alguno de los vértices es de grado 1, se verifica fácilmente que $W_1(m_x, m_y) = 1$. En efecto, la configuración del grafo en la arista (x, y) posee la siguiente forma.

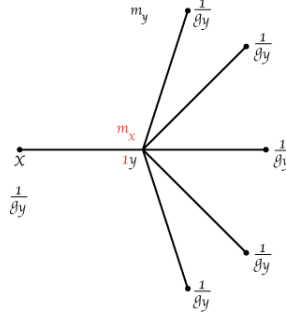


Figura 2.2: Configuración de la arista (x, y) de un árbol en la que el vértice x posee grado 1.

En este caso, es evidente que:

$$W_1(m_x, m_y) = \sum_{z \sim y} \frac{1}{g_y} = \frac{1}{g_y} \sum_{z \sim y} 1 = \frac{g_y}{g_y} = 1.$$

Nos interesamos al caso $1 - \frac{1}{g_x} - \frac{1}{g_y} \geq 0$.

Podemos definir una función 1-Lipschitz en un árbol de este modo:

$$f(z) = \begin{cases} 0, & \text{si } z \sim y, z \neq x \\ 1, & \text{si } z = y \\ 2, & \text{si } z = x \\ 3, & \text{si } z \sim x, z \neq y \end{cases}$$

Dado que en un árbol, el camino que une dos vértices es único, no existen otros caminos entre los vecinos de x y los vecinos de y . Es por ello que la función puede ser extendida como una función 1-Lipschitz a todo el grafo. Por el Teorema de dualidad de Kantorovich:

$$W_1(m_x, m_y) \geq \frac{1}{g_x}(3(g_x - 1) + 1) - \frac{1}{g_y} \cdot 2 = 3 - \frac{2}{g_x} - \frac{2}{g_y}.$$

■

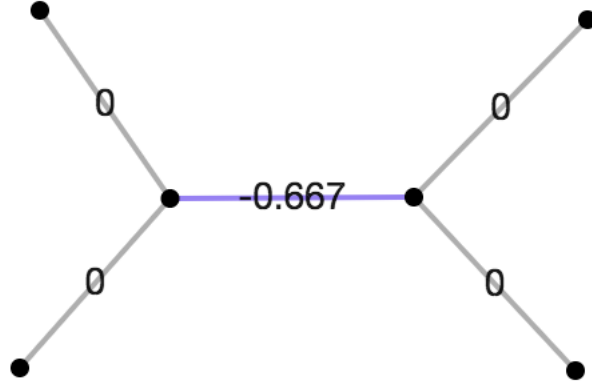


Figura 2.3: Ilustración de la curvatura de Ricci discreta para un árbol.

2.1.7. El ejemplo de los grafos completos

Proposición 2.3 [14]

Consideremos un grafo completo K_n . Consideremos la siguiente cadena de Markov m :

$$m_x(y) := \begin{cases} \frac{1}{n-1}, & \text{si } y \sim x \\ 0, & \text{si } y = x. \end{cases}$$

Entonces para todo par de vértices x, y se tiene:

$$\mathcal{K}(x, y) = \frac{n-2}{n-1}.$$

Demostración: Utilizemos el Teorema de dualidad de Kantorovich. Sean x e y cualesquiera.

$$W_1(m_x, m_y) = \sup_{f \in \text{Lip}(K_n)} \int f dm_x - \int f dm_y$$

$$\begin{aligned} W_1(m_x, m_y) &:= \sup_{f \in \text{Lip}(K_n)} \frac{1}{n-1} \sum_{z \neq x} f(z) - \frac{1}{n-1} \sum_{z \neq y} f(z) \\ &= \sup_{f \in \text{Lip}(K_n)} \frac{1}{n-1} (f(y) - f(x)) \\ &\leq \frac{1}{n-1}. \end{aligned}$$

Escogemos una función f 1-Lipschitz, $f: K_n \rightarrow \mathbb{R}$ como :

$$f(z) := \begin{cases} 1, & \text{si } z = y \\ 0, & \text{si no es el caso.} \end{cases}$$

obteniéndose así la igualdad: $W_1(m_x, m_y) = \frac{1}{n-1}$.

Por ende:

$$\mathcal{K}(x, y) = 1 - \frac{W_1(m_x, m_y)}{d(x, y)} = 1 - W_1(m_x, m_y)$$

$$\mathcal{K}(x, y) = 1 - \frac{1}{n-1} = \frac{n-2}{n-1}.$$

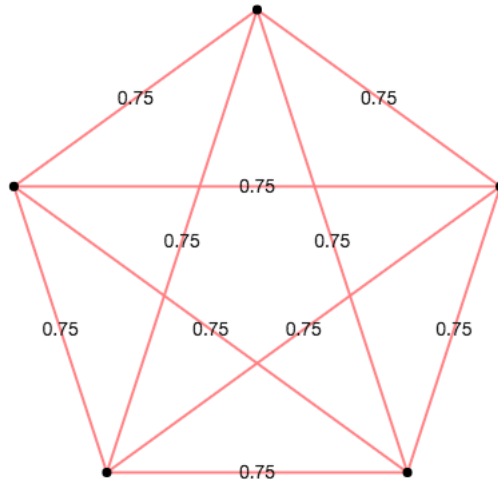


Figura 2.4: Ilustración de la curvatura de Ricci para el grafo completo K_5 .

2.2. La curvatura de Ricci discreta y los *mixing times*

2.2.1. Contracción en el espacio de medidas de probabilidad

Sea μ una medida en \mathcal{X} se define la imagen de μ por el paseo aleatorio como la medida $\mu * m$ dada por:

$$(\mu * m)(A) := \int_{x \in X} m_x(A) d\mu(x) \quad \forall A \in \beta(\mathcal{X}).$$

Teorema 2.2 [26]

Sea (\mathcal{X}, d, m) un espacio métrico con un paseo aleatorio m . Sea $\mathcal{K} \in \mathbb{R}$. Entonces $\mathcal{K}(x, y) \geq \mathcal{K}$ para todo $x, y \in \mathcal{X}$ si y solo si para cualquier par de medidas de probabilidad μ y $\mu' \in \mathcal{P}_1(\mathcal{X})$ se tiene:

$$W_1(\mu * m, \mu' * m) \leq (1 - \mathcal{K})W_1(\mu, \mu').$$

Además, en este caso si $\mu \in \mathcal{P}_1(\mathcal{X})$ entonces $\mu * m \in \mathcal{P}_1(\mathcal{X})$.

Demostración: Supongamos que $\mu \rightarrow \mu * m$ es contractante para la distancia W_1 . Para x e y cualquiera, se define $\mu = \delta_x$ y $\mu' = \delta_y$ como las medidas de Dirac en x e y respectivamente.

Por definición : $\delta_x * m = m_x$ y $\delta_y * m = m_y$, con lo cual:

$$W_1(m_x, m_y) \leq (1 - \mathcal{K})W_1(\delta_x, \delta_y) = (1 - \mathcal{K})d(x, y)$$

de donde resulta $\mathcal{K}(x, y) \geq \mathcal{K}$ para todo $x, y \in \mathcal{X}$.

Para demostrar la recíproca, supongamos $\mathcal{K}(x, y) \geq \mathcal{K}$ para todo $x, y \in \mathcal{X}$. Para cada (x, y) , sea $\xi_{x,y}$ un coupling óptimo entre m_x y m_y que alcanza $W_1(m_x, m_y)$.

Asimismo, sea \mathcal{I} un coupling entre μ y μ' óptimo para $W_1(\mu, \mu')$.

Entonces, $\int_{\mathcal{X} \times \mathcal{X}} d\mathcal{I}(x, y)\xi_{x,y}$ es un coupling entre $\mu * m$ y $\mu' * m$. Así:

$$W_1(\mu * m, \mu' * m) \leq \int_{x,y \in \mathcal{X}} d(x, y)d\left(\int_{x',y' \in \mathcal{X}} d\mathcal{I}(x', y')\xi_{x',y'}\right)(x, y) = \int_{x,y,x',y' \in \mathcal{X}} d\mathcal{I}(x', y')d\xi_{x',y'}(x, y)d(x, y).$$

Notemos que:

$$\int_{x,y \in \mathcal{X}} d\xi_{x',y'}(x, y)d(x, y) = W_1(m_{x'}, m_{y'}) \leq (1 - \mathcal{K}(x', y'))d(x', y') \leq (1 - \mathcal{K})d(x', y').$$

Por el Teorema de Fubini:

$$W_1(\mu * m, \mu' * m) \leq \int_{x',y' \in \mathcal{X}} d\mathcal{I}(x', y')d(x', y')(1 - \mathcal{K}(x', y')) = (1 - \mathcal{K})W_1(\mu, \mu').$$

Para la última parte del Teorema, fijemos algún origen $o \in \mathcal{X}$ y notemos que para toda medida $\mu \in \mathcal{P}_1(\mathcal{X})$ el primer momento de $\mu * m$ es $W_1(\delta_o, \mu * m) \leq W_1(\delta_o, m_o) + W_1(m_o, \mu * m) \leq W_1(\delta_o, m_o) + (1 - \mathcal{K})W_1(\delta_o, \mu)$. Por hipótesis, $W_1(\delta_o, \mu) < \infty$ y $W_1(\delta_o, m_o) < \infty$ ya que m_o por definición tiene primer momento finito. ■

Se define el *salto* $J(x)$ como :

$$J(x) := W_1(\delta_x, m_x).$$

Inductivamente, definamos m^{*n} como:

$$m^{*n} = m^{*(n-1)} * m$$

con $m^1 = m$.

Corolario 2.1 (*W₁ convergencia*) [26]

Supongamos que $\mathcal{K}(x, y) \geq \mathcal{K} > 0$ para todo $x, y \in \mathcal{X}$. Entonces el paseo aleatorio tiene una única medida invariante $\nu \in \mathcal{P}(X)$. Además, para toda medida $\mu \in \mathcal{P}(X)$, la secuencia $\mu * m^{*n}$ tiende exponencialmente rápido a ν en W_1 . Es decir:

$$W_1(\mu * m^{*n}, \nu) \leq (1 - \mathcal{K})^n W_1(\mu, \nu) \quad (2.1)$$

En particular :

$$W_1(m_x^{*n}, \nu) \leq (1 - \mathcal{K})^n \frac{J(x)}{\mathcal{K}}. \quad (2.2)$$

La primera estimación es de gran utilidad para obtener cotas de *mixing times*. Notemos que la segunda estimación depende de la *data* inicial y no requiere conocer la medida invariante ni el diámetro, en particular se aplica a grafos infinitos.

Demostración: Dado que el mapeo $\mu \rightarrow \mu * m$ es una contracción para la distancia W_1 , por el Teorema del punto fijo de Banach existe una única medida invariante ν .

La primera parte es consecuencia directa de la proposición anterior. En efecto: $\nu * m^{*n} = \nu$.

Para la segunda desigualdad basta notar:

$$W_1(\delta_x, \nu) \leq W_1(\delta_x, m_x) + W_1(m_x, \nu)$$

$$W_1(\delta_x, \nu) \leq J(x) + W_1(\delta_x * m, \nu * m).$$

Por la proposición anterior:

$$W_1(\delta_x, \nu) \leq J(x) + (1 - \mathcal{K})W_1(\delta_x, \nu)$$

$$W_1(\delta_x, \nu) \leq \frac{J(x)}{\mathcal{K}}.$$

Escogiendo $\mu = \delta_x$ en (2.1) se obtiene (2.2). ■

2.2.2. Mixing times para hipercubo vía la curvatura de Ricci discreta

Sea $\mathcal{Q}_N = \{0, 1\}^N$ el hipercubo discreto equipado con la distancia de Hamming. Se define la cadena de Markov *lazy* m mediante el siguiente kernel:

$$m_x(z) := \begin{cases} \frac{1}{2}, & \text{si } z = x \\ \frac{1}{2N}, & d_H(z, x) = 1. \end{cases} \quad (**)$$

Sean $x, y \in \mathcal{Q}_N$, entonces $\mathcal{K}(x, y) = \frac{1}{N}$.

En efecto, sin pérdida de generalidad, consideremos $x = (0, \dots, 0)$ e $y = (1, 0, \dots, 0)$. Los vecinos de x son : $x = (0, \dots, 0), x_1 = (1, 0, \dots, 0), x_2 = (0, 1, 0, \dots, 0), \dots, x_N = (0, 0, \dots, 1)$.

Los vecinos de y son: $y = (1, 0, \dots, 0), y_1 = (0, \dots, 0), y_2 = (1, 1, \dots, 0), \dots, y_N = (1, 0, \dots, 0, 1)$.

$$W_1(m_x, m_y) = \sup_{f \in \text{Lip}_1(\mathcal{Q}_N)} \int f dm_x - \int f dm_y .$$

Ahora bien:

$$\int f dm_x - \int f dm_y = \frac{1}{2}f(x) - \frac{1}{2N}f(y_1) + \frac{1}{2N}f(x_1) - \frac{1}{2}f(y) + \frac{1}{2N}((f(x_2) - f(y_2)) + \dots + (f(x_N) - f(y_N))) .$$

Así:

$$\int f dm_x - \int f dm_y = \frac{N-1}{2N}f(x) - \frac{N-1}{2N}f(y) + \frac{1}{2N}((f(x_2) - f(y_2)) + \dots + (f(x_N) - f(y_N))) .$$

Tomando el supremo de funciones 1-Lipschitz, se obtiene:

$$W_1(m_x, m_y) = \frac{N-1}{2N} + \frac{N-1}{2N} \leq \frac{N-1}{N} .$$

Análogamente a la subsección 2.1.5, escogemos la función 1-Lipschitz, $f : \mathcal{Q}_N \rightarrow \mathbb{R}$, como sigue:

$$f(x^1, x^2, \dots, x^N) = -x^1 ,$$

de donde resulta:

$$\mathcal{K} = \frac{1}{N} .$$

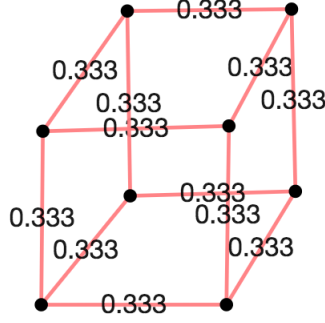


Figura 2.5: Ilustración de la curvatura de Ricci discreta para la cadena de Markov *lazy* $\star\star$ en \mathcal{Q}_3 .

Notemos que por la caracterización del Teorema 1.1 de la variación total:

$$\|\mu - \nu\|_{TV} = \inf_{X \sim \mu, Y \sim \nu} \mathbb{E}(1_{\{X \neq Y\}})$$

Por ende, en el caso de un grafo:

$$\|\mu - \nu\|_{TV} = \inf_{X \sim \mu, Y \sim \nu} \mathbb{E}(1_{\{X \neq Y\}}) \leq \inf_{X \sim \mu, Y \sim \nu} \mathbb{E}(d(x, y)) = W_1(\mu, \nu)$$

En el caso del hipercubo, el diámetro de \mathcal{X} es N . Como aplicación directa del Corolario 2.1 se obtiene:

$$\|m_x^{\star n} - \nu\|_{TV} \leq W_1(m_x^n, \nu) \leq (1 - \mathcal{K})^n W_1(\delta_x, \nu) \leq (1 - \mathcal{K})^n \text{diam}(\mathcal{X}) \quad (2.3)$$

A diferencia de la estimación tradicional de $\mathcal{O}(N^2)$ para *mixing times* utilizando *spectral gap* (ver siguiente subsección), obtenemos con la curvatura de Ricci discreta la estimación $\mathcal{O}(N \log N)$.

En efecto, en virtud de 2.3, para tener $\|m_x^n - \nu\| \leq \varepsilon$ basta:

$$N(1 - \frac{1}{N})^n \leq \varepsilon$$

$$\log N + n \log(1 - \frac{1}{N}) \leq \log \varepsilon$$

$$n \geq \frac{\log \varepsilon - \log N}{\log(1 - \frac{1}{N})} \sim \mathcal{O}(N \log N).$$

La siguiente proposición permite en ciertos casos calcular la curvatura de Ricci .

Proposición 2.4 [26]

Sean $((\mathcal{X}_i, d_i))_{i \in I}$ una familia finita de espacios métricos y supongamos que \mathcal{X}_i está equipado del paseo aleatorio $m^{(i)}$. Sea \mathcal{X} el espacio producto de los espacios \mathcal{X}_i , equipado con la distancia $d := \sum d_i$. Sea (α_i) una familia no negativa de números reales con $\sum \alpha_i = 1$. Consideremos el paseo aleatorio m definido en \mathcal{X} como:

$$m_{(x_1, \dots, x_k)} := \sum \alpha_i \delta_{x_1} \otimes \dots \otimes m_{x_i} \otimes \dots \otimes \delta_{x_k} .$$

Supongamos que para cada i , la curvatura de Ricci de cada $m^{(i)}$ es al menos κ_i . Entonces, la curvatura de Ricci de m es al menos $\inf \alpha_i \kappa_i$.

Para la demostración, recurrimos al siguiente Lema :

Lema 2.1 La distancia $W_1 : \mathcal{P}_1(\mathcal{X}) \times \mathcal{P}_1(\mathcal{X}) \rightarrow \mathbb{R}$ es convexa, i.e. :

$$\begin{aligned} W_1(\alpha(\mu_1, \nu_1) + (1 - \alpha)(\mu_2, \nu_2)) &:= W_1(\alpha\mu_1 + (1 - \alpha)\mu_2, \alpha\nu_1 + (1 - \alpha)\nu_2) \\ &\leq \alpha W_1(\mu_1, \nu_1) + (1 - \alpha)W_1(\mu_2, \nu_2) . \end{aligned}$$

Demostración: Por el Teorema de dualidad de Kantorovich:

$$\begin{aligned} W_1(\alpha(\mu_1, \nu_1) + (1 - \alpha)(\mu_2, \nu_2)) &= \sup_{f \in Lip_1} \int_{\mathcal{X}} f d(\alpha\mu_1 + (1 - \alpha)\mu_2) - \int_{\mathcal{X}} f d(\alpha\nu_1 + (1 - \alpha)\nu_2) \\ &= \sup_{f \in Lip_1} (\alpha \int_{\mathcal{X}} f d\mu_1 + (1 - \alpha) \int_{\mathcal{X}} f d\mu_2) - (\alpha \int_{\mathcal{X}} f d\nu_1 + (1 - \alpha) \int_{\mathcal{X}} f d\nu_2) \\ &\leq \alpha \sup_{f \in Lip_1} \int_{\mathcal{X}} f d\mu_1 - \int_{\mathcal{X}} f d\nu_1 + (1 - \alpha) \sup_{f \in Lip_1} \int_{\mathcal{X}} f d\mu_2 - \int_{\mathcal{X}} f d\nu_2 \\ &= \alpha W_1(\mu_1, \nu_1) + (1 - \alpha)W_1(\mu_2, \nu_2) . \end{aligned}$$

■

Demostración de la Proposición 2.4: Para $x \in \mathcal{X}$ sea $\tilde{m}_x^{(i)}$ el paseo aleatorio dado por $\delta_{x_1} \otimes \dots \otimes m_{x_i} \otimes \dots \otimes \delta_{x_k}$.

Sean $x = (x_i)$ e $y = (y_i)$ dos puntos en \mathcal{X} .

Entonces, por el Lema 2.1 :

$$\begin{aligned}
W_1(m_x, m_y) &\leq \sum \alpha_i W_1(\tilde{m}_x^{(i)}, \tilde{m}_y^{(i)}) \\
&\leq \sum \alpha_i (W_1(m_x^{(i)}, m_y^{(i)}) + \sum_{j \neq i} d_j(x_j, y_j)) \\
&\leq \sum \alpha_i ((1 - \kappa_i) d_i(x_i, y_i) + \sum_{j \neq i} d_j(x_j, y_j)) \\
&= \sum \alpha_i (-\kappa_i d_i(x_i, y_i) + \sum_{j \neq i} d_j(x_j, y_j)) \\
&\leq \sum d_i(x_i, y_i) - \sum \alpha_i \kappa_i d_i(x_i, y_i) \\
&\leq (1 - \inf \alpha_i \kappa_i) \sum d_i(x_i, y_i) \\
&= (1 - \inf \alpha_i \kappa_i) d(x, y) .
\end{aligned}$$

La proposición anterior permite una demostración muy corta que la curvatura de la cadena de markov *lazy* en el hipercubo discreto \mathcal{Q}_N es $\frac{1}{N}$. En efecto, esta cadena m se puede escribir como:

$$m_{(x_1, \dots, x_k)} := \sum_{i=1}^N \frac{1}{N} \delta_{x_1} \otimes \dots \otimes m_{x_i} \otimes \dots \otimes \delta_{x_k} .$$

con m_{x_i} la cadena en $\{0, 1\}$, que envía cada punto a la distribución de equilibrio $(\frac{1}{2}, \frac{1}{2})$ y que posee curvatura igual a 1. Así, $\mathcal{K} \geq \inf \frac{1}{N} \cdot 1 = \frac{1}{N}$.

2.2.3. Mixing times para el hipercubo vía el *spectral gap*

Sea P una matriz de transición.

Se define :

$$\lambda_* := \max\{|\lambda| : \lambda \text{ es un valor propio de } P, \lambda \neq 1\} .$$

La diferencia $\gamma_* := 1 - \lambda_*$ se denomina el *spectral gap absoluto*. Lema 12.1 del libro Markov Chains and Mixing Times [37] implica que si P es aperiódica e irreducible, entonces $\gamma_* > 0$.

Para una matriz de transición reversible P , podemos ordenar de forma decreciente los valores propios:

$$1 = \lambda_1 > \lambda_2 \geq \dots \lambda_{|X|} \geq -1 .$$

El *spectral gap* de una cadena reversible se define como:

$$\gamma := 1 - \lambda_2 .$$

Se verifica que cuando una cadena de Markov es *lazy*, entonces: $\gamma_* = \gamma$.

El *relaxing time* t_{rel} de una cadena de Markov reversible con *spectral gap absoluto* λ_* se define como:

$$t_{rel} := \frac{1}{\gamma_*}.$$

Teorema 2.3 [37]

Sea P una matriz de transición de una cadena de Markov reversible en un espacio \mathcal{X} , sea $\pi_{min} := \min_{x \in \mathcal{X}} \pi(x)$. Entonces:

$$t_{mix}(\varepsilon) \leq \left\lceil \left(t_{rel} \left(\frac{1}{2} \log \left(\frac{1}{\pi_{min}} + \log \left(\frac{1}{2\varepsilon} \right) \right) \right) \right) \right\rceil.$$

Ejemplo (El hipercubo)

Consideremos el *lazy random walk* en el hipercubo n -dimensional $\mathcal{Q}_n = \{-1, 1\}^N$. Para cada conjunto de coordenadas $J \subset \{1, 2, \dots, N\}$:

$$f(x_1, \dots, x_n) = \prod_{j \in J} x_j$$

es un vector propio. El valor propio asociado es :

$$\lambda_J = \frac{\sum_{i=1}^N (1 - 1_{\{i \in J\}})}{N} = \frac{N - |J|}{N}.$$

Por el teorema anterior:

$$t_{mix}(\varepsilon) \leq N(-\log \varepsilon + \log(2^N)) = N^2(\log 2 - N^{-1} \log \varepsilon).$$

Así, $t_{mix} \sim \mathcal{O}(N^2)$. Notemos que esta cota no es óptima como la que obtuvimos con la curvatura de Ricci discreta.

Observación Existen otros métodos como el coupling de cadenas de Markov, o métodos espectrales más refinados que permiten obtener la cota $\mathcal{O}(N \log N)$.

2.3. La curvatura de Ricci discreta y el fenómeno de concentración de la medida

2.3.1. Algunas definiciones previas al resultado de concentración

En lo que sigue consideramos una cadena de Markov m en un espacio métrico (\mathcal{X}, d) .

Se define la *constante de difusión del paseo aleatorio en x* como:

$$\sigma(x) = \left(\frac{1}{2} \int \int d(y, z)^2 dm_x(y) dm_x(z) \right)^{\frac{1}{2}}.$$

Para ν una distribución invariante, se define el *promedio de la constante de difusión*:

$$\sigma := \| \sigma(x) \|_{L^2(\mathcal{X}, \nu)} = \sqrt{\frac{1}{2} \int_{\mathcal{X}} \sigma(x)^2 d\nu(x)}.$$

Sea $\sigma_{\infty}(x) := \frac{1}{2}(\text{diam}(\text{Supp}(m_x)))$ y $\sigma_{\infty} := \sup_{x \in \mathcal{X}} \sigma_{\infty}(x)$.

Se supone $\sigma_{\infty} < \infty$.

Se define n_x la *dimensión local de x* como:

$$n_x := \frac{\sigma(x)^2}{\sup\{\text{Var}_{m_x} f, f : \text{Supp}(m_x) \rightarrow \mathbb{R}, 1\text{-Lipschitz}\}}.$$

Finalmente, denotamos: $n = \inf_x n_x$. Se supone $n > 0$.

Recordemos la definición de $\text{Var}_{\nu} f$:

$$\text{Var}_{\nu} f = \| f - \mathbb{E}f \|_{L^2}^2 = \frac{1}{2} \int \int (f(x) - f(y))^2 d\nu(x) d\nu(y).$$

En general, se define la *dimensión estadística* de un espacio métrico medible como:

$$\text{StatDim}(X, d, \mu) = \frac{1}{2} \frac{\int \int d(x, y)^2 d\mu(x) d\mu(y)}{\sup\{\text{Var}_{\mu} f, f 1\text{-Lipschitz}\}}.$$

Por ende, para cada $x \in \mathcal{X}$ la *dimensión local* de $x \in \mathcal{X}$ es:

$$n_x = \text{StatDim}(X, d, m_x).$$

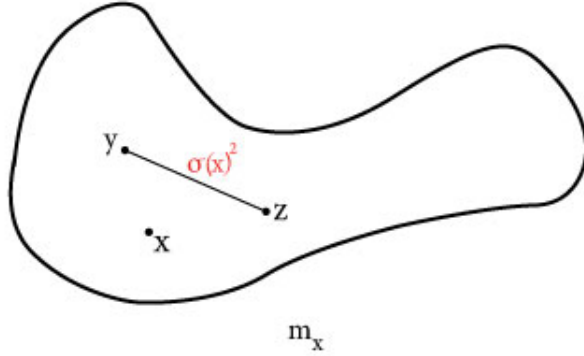


Figura 2.6: Ilustración de la constante de difusión del paseo aleatorio en x .

2.3.2. Teorema de concentración Gaussiana

Los siguientes dos Lemas son esenciales para el teorema de concentración Gaussiana.

Lema 2.2 (*Contracción Lipschitz*) [26]

Sea (X, d, m) un paseo aleatorio en un espacio métrico y sea $k \in \mathbb{R}$. Entonces la curvatura de Ricci discreta de \mathcal{X} es al menos \mathcal{K} si y sólo si para toda función k -Lipschitz $f : X \rightarrow \mathbb{R}$, la función Mf es $k(1 - \mathcal{K})$ Lipschitz.

Demostración: Supongamos que la curvatura de Ricci discreta de \mathcal{X} es al menos \mathcal{K} .

Sea f una función k -Lipschitz, se tiene:

$$Mf(x) - Mf(y) = \int_z f(z) dm_x - \int_z f(z) dm_y$$

$$Mf(x) - Mf(y) \leq k \sup_{g \in \text{Lip}_1(\mathcal{X})} \int_z g d(m_x - m_y) \leq kd(x, y)(1 - \mathcal{K}(x, y)).$$

Recíprocamente, supongamos que f es 1-Lipschitz, de modo tal que Mf es $(1 - \mathcal{K})$ Lipschitz.

Por la dualidad para la distancia de transporte W_1 :

$$W_1(m_x, m_y) = \sup_{f \in \text{Lip}_1(\mathcal{X})} \int f(dm_x - dm_y) = \sup_{f \in \text{Lip}_1(\mathcal{X})} Mf(x) - Mf(y)$$

$$W_1(m_x, m_y) \leq (1 - \mathcal{K})d(x, y).$$

■

Lema 2.3 [26]

Sea $\phi : X \rightarrow \mathbb{R}$ una función α -Lipschitz con $\alpha \leq 1$, y sea $\lambda \leq \frac{1}{3\sigma_\infty}$.

Entonces para todo $x \in \mathcal{X}$ se tiene:

$$(Me^{\lambda\phi})(x) \leq e^{\lambda M\phi(x) + \lambda^2 \alpha^2 \frac{\sigma(x)^2}{n_x}}.$$

Demostración: Para toda función g suficientemente suave y toda variable aleatoria real Y , usando la expansión de Taylor con resto de Lagrange se obtiene:

$$\mathbb{E}g(Y) \leq g(\mathbb{E}(Y)) + \frac{1}{2}(\sup_{x \in \mathcal{X}} g'')(x) \text{Var} Y.$$

Aplicando a $g(x) = e^{\lambda x}$ se obtiene:

$$(Me^{\lambda\phi})(x) = \mathbb{E}_{m_x} e^{\lambda\phi} \leq e^{\lambda M\phi(x)} + \frac{\lambda^2}{2} \left(\sup_{y \in \text{Supp}(m_x)} e^{\lambda\phi(y)} \right) \text{Var}_{m_x} \phi. \quad (2.4)$$

Notando que $\text{diam}(\text{Supp}(m_x)) \leq 2\sigma_\infty$:

$$\phi(y) \leq \phi(z) + 2\sigma_\infty \quad \forall y, z \in \text{Supp}(m_x).$$

Luego integrando con respecto a dm_x se obtiene:

$$\int_{\mathcal{X}} \phi(y) dm_x(z) \leq \int_{\mathcal{X}} \phi(z) dm_x(z) + 2\sigma_\infty \int_{\mathcal{X}} dm_x(z)$$

de donde resulta

$$\phi(y) \leq \mathbb{E}_{m_x}(\phi) + 2\sigma_\infty.$$

Tomando supremo, se obtiene la desigualdad:

$$\sup_{y \in \text{Supp}(m_x)} \phi(y) \leq \mathbb{E}_{m_x}(\phi) + 2\sigma_\infty.$$

De este modo, reemplazando en 2.3 se llega a :

$$(Me^{\lambda\phi})(x) \leq e^{\lambda M\phi(x)} + \frac{\lambda^2}{2} e^{\lambda M\phi(x)+2\lambda\sigma_\infty} Var_{m_x}\phi .$$

Para $\lambda \leq \frac{1}{3\sigma_\infty}$ se tiene $e^{2\lambda\sigma_\infty} \leq 2$, con lo cual la desigualdad anterior deviene:

$$(Me^{\lambda\phi})(x) \leq e^{\lambda M\phi(x)} + \lambda^2 e^{\lambda M\phi(x)} Var_{m_x}\phi$$

o equivalentemente:

$$(Me^{\lambda\phi})(x) \leq e^{\lambda M\phi(x)}(1 + \lambda^2 Var_{m_x}\phi).$$

Ahora bien, por definición de n_x : $Var_{m_x}\phi \leq \alpha^2 \frac{\sigma(x)^2}{n_x}$. Así, se obtiene:

$$(Me^{\lambda\phi})(x) \leq e^{\lambda M\phi(x)}(1 + \lambda^2 \alpha^2 \frac{\sigma(x)^2}{n_x}).$$

Dado que para todo $x \geq 0$, $(1 + x) \leq e^x$, se obtiene el resultado:

$$(Me^{\lambda\phi})(x) \leq e^{\lambda M\phi(x)+\lambda^2 \alpha^2 \frac{\sigma(x)^2}{n_x}} .$$

■

Teorema 2.4 (*Concentración Gaussiana*) [26]

Sea (\mathcal{X}, d, m) un paseo aleatorio en un espacio métrico, con curvatura de Ricci discreta al menos $\mathcal{K} > 0$. Sea ν la única distribución invariante. Definamos:

- $D_x^2 := \frac{\sigma(x)^2}{n_x \mathcal{K}}$
- $D^2 := \mathbb{E}_\nu D_x^2 < \infty$

Supongamos que la función $x \rightarrow D_x^2$ es C -Lipschitz. Sea $t_{max} := \frac{D^2}{\max(\sigma_\infty, \frac{2C}{3})}$. Entonces para toda función f 1-Lipschitz y para todo $t \leq t_{max}$:

$$\nu(\{x, f(x) \geq t + \mathbb{E}_\nu f\}) \leq e^{\frac{-t^2}{6D^2}}$$

y para $t \geq t_{max}$:

$$\nu(\{x, f(x) \geq t + \mathbb{E}_\nu f\}) \leq e^{\frac{-t^2}{6D^2} - \frac{t-t_{max}}{\max(3\sigma_\infty, 2C)}} .$$

Demostración del Teorema: Sea f una función 1-Lipschitz y $\lambda \geq 0$. Se define por inducción:

$$f_0 := f$$

$$f_{k+1}(x) := Mf_k(x) + \lambda \frac{\sigma(x)^2}{n_x} \left(1 - \frac{\mathcal{K}}{2}\right)^{2k}.$$

Supongamos que $\lambda \leq \frac{1}{2C}$.

Por hipótesis:

$h : x \rightarrow \frac{\sigma(x)^2}{n_x \mathcal{K}}$ es C -Lipschitz. Así, $\lambda \frac{\sigma(x)^2}{n_x}$ es $\frac{\mathcal{K}}{2}$ Lipschitz.

Se demuestra fácilmente, por inducción, que f_k es $(1 - \frac{\mathcal{K}}{2})^k$ Lipschitz.

Demostración:

Caso $k=0$:

$$f_1(x) = Mf_0(x) + \frac{\lambda \sigma(x)^2}{n_x} \left(1 - \frac{\mathcal{K}}{2}\right)^0.$$

Por el Lema 2.1 (de contracción Lipschitz): $Mf_0(x)$ es $(1 - \mathcal{K})$ Lipschitz.

Dado que $\lambda \frac{\sigma(x)^2}{n_x}$ es $\frac{\mathcal{K}}{2}$ Lipschitz. Entonces f_1 es $(1 - \frac{\mathcal{K}}{2})$ Lipschitz.

Paso inductivo:

Supongamos por inducción que f_k es $(1 - \frac{\mathcal{K}}{2})^k$ Lipschitz.

$$f_{k+1}(x) - f_{k+1}(y) = Mf_k(x) - Mf_k(y) + \left(1 - \frac{\mathcal{K}}{2}\right)^{2k} \left(\frac{\lambda \sigma(x)^2}{n_x} - \frac{\lambda \sigma(y)^2}{n_y}\right)$$

Por hipótesis de inducción, Mf_k es $(1 - \frac{\mathcal{K}}{2})^k (1 - \mathcal{K})$ Lipschitz. Luego:

$$\begin{aligned} f_{k+1}(x) - f_{k+1}(y) &\leq \left(1 - \frac{\mathcal{K}}{2}\right)^k (1 - \mathcal{K})(x - y) + \left(1 - \frac{\mathcal{K}}{2}\right)^{2k} \frac{\mathcal{K}}{2}(x - y) \\ &= \left(\left(1 - \frac{\mathcal{K}}{2}\right)^k (1 - \mathcal{K}) + \left(1 - \frac{\mathcal{K}}{2}\right)^{2k} \frac{\mathcal{K}}{2}\right)(x - y) \\ &= \left(\left(1 - \frac{\mathcal{K}}{2}\right)^k \left((1 - \mathcal{K}) + \left(1 - \frac{\mathcal{K}}{2}\right)^k \frac{\mathcal{K}}{2}\right)\right)(x - y) \\ &\leq \left(1 - \frac{\mathcal{K}}{2}\right)^{k+1}(x - y). \end{aligned}$$

De este modo, f_{k+1} es $(1 - \frac{\mathcal{K}}{2})^{k+1}$ Lipschitz. ■

Ahora, suponiendo $\lambda \leq \frac{1}{3\sigma_\infty}$ y utilizando el Lema 2.2:

$$(Me^{\lambda f_k})(x) \leq e^{\lambda M f_k(x) + \lambda^2 \frac{\sigma(x)^2}{n_x} (1 - \frac{\mathcal{K}}{2})^{2k}} = e^{\lambda(M f_k(x) + \lambda \frac{\sigma(x)^2}{n_x} (1 - \frac{\mathcal{K}}{2})^{2k})}$$

$$(Me^{\lambda f_k})(x) \leq e^{\lambda f_{k+1}(x)}$$

Aplicando M k -veces por inducción:

$$(M^k e^{\lambda f})(x) \leq e^{\lambda f_k(x)}. \quad (2.5)$$

Definiendo $g = \frac{\sigma(x)^2}{n_x}$, podemos reescribir f_k como:

$$f_k(x) = (M^k f)(x) + \lambda \sum_{i=1}^k (M^{k-i} g)(x) \left(1 - \frac{\mathcal{K}}{2}\right)^{2(i-1)}.$$

Por lo tanto, (utilizando el Teorema de convergencia dominada pues $\frac{\sigma^2(x)}{n(x)} \leq \frac{\sigma_\infty}{n}$):

$$\lim_{k \rightarrow \infty} f_k(x) = \mathbb{E}_\nu f + \lambda \sum_{i=1}^{\infty} \mathbb{E}_\nu g \left(1 - \frac{\mathcal{K}}{2}\right)^{2(i-1)}$$

Ahora bien, dado que $\mathcal{K} \leq 1$

$$\sum_{i=1}^{\infty} \left(1 - \frac{\mathcal{K}}{2}\right)^{2(i-1)} \leq \frac{4}{3\mathcal{K}}$$

$$\lim_{k \rightarrow \infty} f_k(x) \leq \mathbb{E}_\nu f + \lambda \mathbb{E}_\nu g \frac{4}{3\mathcal{K}}. \quad (2.6)$$

Por otro lado:

$$(M^k e^{\lambda f})(x) \rightarrow \mathbb{E}_\nu e^{\lambda f}. \quad (2.7)$$

Por (2.5), (2.6) y (2.7) :

$$\mathbb{E}_\nu e^{\lambda f} \leq \lim_{k \rightarrow \infty} e^{\lambda f_k} \leq e^{\lambda \mathbb{E}_\nu f + \frac{4\lambda^2}{3\mathcal{K}} \mathbb{E}_\nu \frac{\sigma(x)^2}{n_x}} \quad (2.8)$$

Para concluir, utilizamos un argumento de tipo Chebyshev:

$$\nu(\{x : f(x) \geq t + \mathbb{E}_\nu(f)\}) = \nu(\{x : e^{\lambda f(x) - \lambda \mathbb{E}_\nu f} \geq e^{\lambda t}\})$$

$$\nu(\{x : e^{\lambda f(x) - \lambda \mathbb{E}_\nu f} \geq e^{\lambda t}\}) \leq \frac{\mathbb{E}_\nu(e^{\lambda f - \lambda \mathbb{E}_\nu f})}{e^{\lambda t}}$$

$$\nu(\{x : e^{\lambda f(x) - \lambda \mathbb{E}_\nu f} \geq e^{\lambda t}\}) \leq e^{\frac{4\lambda^2}{3} D^2 - \lambda t}.$$

Sea $t \leq t_{\text{máx}} := \min\{\frac{3D^2}{2C}, \frac{D^2}{\sigma_\infty}\}$.

Tomemos $\lambda = \frac{t}{3D^2} \leq \min\{\frac{1}{2C}, \frac{1}{3\sigma_\infty}\}$.

$$\begin{aligned} \nu(\{x : f(x) \geq t + \mathbb{E}_\nu f\}) &\leq e^{\frac{4D^2}{3} \frac{t^2}{9D^4} - \frac{t^2}{3D^2}} \\ &= e^{\frac{t^2}{3D^2} (\frac{4}{9} - 1)} \\ &= e^{\frac{t^2}{6D^2} (\frac{8}{9} - 2)} \\ &\leq e^{\frac{-t^2}{6D^2}}. \end{aligned}$$

Estudieemos ahora el caso $t \geq t_{\text{máx}}$.

Notemos que:

$$\{x : f(x) \geq t + \mathbb{E}_\nu f\} = \{x : f(x) \geq t - t_{\text{máx}} + t_{\text{máx}} + \mathbb{E}_\nu f\}$$

$$\{x : f(x) \geq t + \mathbb{E}_\nu f\} = \{x : e^{\lambda f} \geq e^{\lambda(t - t_{\text{máx}} + t_{\text{máx}} + \mathbb{E}_\nu f)}\}.$$

con lo cual:

$$\nu\{x : e^{\lambda f} \geq e^{\lambda(t - t_{\text{máx}} + t_{\text{máx}} + \mathbb{E}_\nu f)}\} \leq \frac{\mathbb{E}_\nu(e^{\lambda f})}{e^{\lambda t_{\text{máx}}} e^{\lambda(t - t_{\text{máx}})} e^{\lambda \mathbb{E}_\nu f}}.$$

Ahora bien, para $\lambda \leq \frac{1}{\text{máx}\{2C, 2\sigma_\infty\}}$, la desigualdad (2.8) implica:

$$\nu\{x : e^{\lambda f} \geq e^{\lambda(t - t_{\text{máx}} + t_{\text{máx}} + \mathbb{E}_\nu f)}\} \leq \frac{e^{\lambda \mathbb{E}_\nu f} e^{\frac{4\lambda^2 D^2}{3}}}{e^{\lambda(t - t_{\text{máx}})} e^{\lambda t_{\text{máx}}} e^{\lambda \mathbb{E}_\nu f}}.$$

$$\nu\{x : e^{\lambda f} \geq e^{\lambda(t - t_{\text{máx}} + t_{\text{máx}} + \mathbb{E}_\nu f)}\} \leq \frac{e^{-\lambda t_{\text{máx}}} e^{\frac{4\lambda^2}{3} D^2}}{e^{\lambda(t - t_{\text{máx}})}}.$$

Tomando $\lambda = \frac{1}{\text{máx}\{2C, 3\sigma_\infty\}}$ se obtiene:

$$\nu\{x : e^{\lambda f} \geq e^{\lambda(t-t_{\max}+t_{\max}+\mathbb{E}_\nu f)}\} \leq \frac{e^{\frac{-t_{\max}}{\max\{2C, 3\sigma_\infty\}}} e^{\frac{4}{3} \frac{D^2}{\max\{2C, 3\sigma_\infty\}^2}}}{e^{\lambda(t-t_{\max})}}.$$

Sabemos que :

$$t_{\max} = \frac{2D^2}{\max\{2C, 3\sigma_\infty\}}.$$

Luego:

$$\nu\{x : e^{\lambda f} \geq e^{\lambda(t-t_{\max}+t_{\max}+\mathbb{E}_\nu f)}\} \leq \frac{-\frac{2}{3} \frac{D^2}{\max\{2C, 3\sigma_\infty\}^2}}{e^{\lambda(t-t_{\max})}}.$$

Notemos que $\frac{t_{\max}^2}{6D^2} = \frac{2}{3} \frac{D^2}{\max\{2C, 3\sigma_\infty\}^2}$.

Se concluye :

$$\nu\{x : e^{\lambda f} \geq e^{\lambda(t-t_{\max}+t_{\max}+\mathbb{E}_\nu f)}\} \leq e^{\frac{-t_{\max}^2}{6D^2}} e^{-\lambda(t-t_{\max})} \leq e^{\frac{-t_{\max}^2}{6D^2} - \frac{(t-t_{\max})}{\max\{2C, 3\sigma_\infty\}}}.$$

■

2.3.3. La existencia genuina de un tiempo máximo

La existencia de un tiempo t_{max} para un comportamiento Gaussiano es genuina. Hay procesos estocásticos en los que no se puede extender a partir de un cierto tiempo el comportamiento Gaussiano. El siguiente ejemplo ilustra aquello.

Ejemplo Sea $X = \{0, 1\}^N$ equipado con la distancia de Hamming.

Consideremos la siguiente cadena de Markov en \mathcal{X} :

Sea $0 < p < 1$:

- Escojamos 1 bit aleatoriamente entre N bits
- Si es $0 \rightarrow 1$ con probabilidad p
- Si es $1 \rightarrow 0$ con probabilidad $1 - p$

La distribución binomial $\nu(x_1, x_2, \dots, x_N) = \prod p^{x_i} (1-p)^{1-x_i}$ es reversible para esta cadena de Markov.

Además, se puede verificar que la curvatura de Ricci discreta es $\frac{1}{N}$.

Sea k el número de bits de $x \in X$ que son iguales a 1. Entonces k sigue una cadena de Markov en $\{0, 1, \dots, N\}$ cuyas probabilidades de transición son:

- $p_{k, k+1} = p(1 - \frac{k}{N})$

- $p_{k,k-1} = (1-p)\frac{k}{N}$
- $p_{k,k} = p\frac{k}{N} + (1-p)(1 - \frac{k}{N})$

La distribución binomial con parámetros N y p , es decir, $\binom{N}{k}p^k(1-p)^{N-k}$, es reversible para esta cadena de Markov. Además, la curvatura de Ricci de esta cadena de Markov es $\frac{1}{N}$.

Sea $\lambda > 0$ y consideremos $p = \frac{\lambda}{N}$ con $N \rightarrow \infty$.

Es un resultado conocido que esta distribución invariante tiende a la distribución de Poisson $\frac{e^{-\lambda}\lambda^k}{k!}$.

La distribución de Poisson tiene un comportamiento Gaussiano (con varianza λ) en un rango aproximadamente λ alrededor de la esperanza. Posteriormente, decrece como $e^{-k \ln k}$ que no es Gaussiano. Así, en este caso, el rango Gaussiano no puede ser extendido.

2.3.4. Concentración exponencial con curvatura no negativa

Hemos visto que la curvatura de Ricci positiva implica una forma de concentración Gaussiana. La curvatura de Ricci no negativa y una hipótesis adicional de existencia de un *punto de atracción* implica concentración exponencial. El ejemplo básico es el siguiente. Sea \mathbb{N} el conjunto de enteros no negativos equipada con la distancia estándar. Sea $0 < p < 1$ y consideremos la cadena de Markov en \mathbb{N} en la que los enteros se dirigen a la izquierda con probabilidad p y a la derecha con probabilidad $1-p$. Explícitamente:

$$m_k = p\delta_{k-1} + (1-p)\delta_{k+1} \quad \forall k \geq 1$$

$$m_0 = p\delta_0 + (1-p)\delta_1$$

Se verifica inmediatamente que $\mathcal{K}(k, k+1) = 0$ y $\mathcal{K}(0, 1) = p$. Existe una medida invariante si y sólo si $p > \frac{1}{2}$ y se satisface concentración exponencial. Si $p \leq \frac{1}{2}$, no hay distribución invariante, así curvatura no negativa sin hipótesis adicional no implica concentración de la medida. Geométricamente, o es un *punto de atracción* para la cadena de Markov. Se formaliza este fenómeno en el siguiente teorema cuya demostración no exponemos.

Teorema 2.5 [26]

Sea (\mathcal{X}, d, m_x) un espacio métrico con un paseo aleatorio. Supongamos que para algún $o \in X$ y $r > 0$ se tiene:

- $\mathcal{K}(x, y) \geq 0$ para todo $x, y \in X$
- $\forall x \in \mathcal{X}$ con $r \leq d(o, x) < 2r$, se tiene: $W_1(m_x, \delta_o) < d(x, o)$
- \mathcal{X} es r -geodésico

- Existe $s > 0$ tal que toda medida m_x satisface una desigualdad de transformada Gaussiana de tipo Laplace :

$$\mathbb{E}_{m_x} e^{\lambda f} \leq e^{\lambda^2 \frac{s^2}{2}} e^{\lambda \mathbb{E}_{m_x} f}$$

para todo $\lambda > 0$ y para toda función 1-Lipschitz $f : \text{Supp}(m_x) \rightarrow \mathbb{R}$

Sea $\rho = \inf\{d(x, o) - W_1(m_x, \delta_o), r \leq d(o, x) < 2r\}$ y asumamos $\rho > 0$. Entonces, existe una medida invariante para el paseo aleatorio. Además, definiendo $D = \frac{s^2}{\rho}$ y $m = r + 2\frac{s^2}{\rho} + \rho(1 + \frac{J(o)^2}{4s^2})$, para toda medida invariante ν se tiene:

$$\int e^{\frac{d(x,o)}{D}} d\nu(x) \leq (4 + \frac{J(o)^2}{s^2}) e^{\frac{m}{D}}$$

y así para toda función 1-Lipschitz $f : \mathcal{X} \rightarrow \mathbb{R}$ y $t \geq 0$:

$$\mathbb{P}(|f - f(o)| \geq t + m) \leq (8 + 2\frac{J(o)^2}{s^2}) e^{-\frac{t}{D}}.$$

Por ende, se obtiene concentración exponencial con decaimiento característico de distancia $\frac{s^2}{\rho}$.

Observación : En el ejemplo anterior, las hipótesis se satisfacen. Escogemos $o = 0$ y $r = 1$. Podemos escoger $s = 2\sigma_\infty = 2$. Existe un sólo punto x tal que $r \leq d(o, x) < 2r$, que es $x = 1$. Se satisface que $W_1(m_x, \delta_o) = 2(1 - p)$ es estrictamente menor a $d(0, 1) = 1$ si y sólo si $p > \frac{1}{2}$. Así, podemos escoger $\rho = 1 - 2(1 - p) = 2p - 1$. Entonces, se obtiene una concentración exponencial con distancia característica $\frac{4}{2p-1}$.

Capítulo 3

Operadores no expansivos en espacios con curvatura de Ricci discreta nula

3.1. Operadores T_m y T_{m^α}

Sea (\mathcal{X}, d) un espacio métrico polaco, equipado con su σ -álgebra de Borel $\beta(\mathcal{X})$. Denotemos como anteriormente $\mathcal{P}_1(\mathcal{X})$ el espacio de medidas de probabilidad μ sobre \mathcal{X} con primer momento finito, es decir:

$$\int_{\mathcal{X}} d(o, y) d\mu(y) < \infty \text{ para algún } o \in \mathcal{X} .$$

Sobre $\mathcal{P}_1(\mathcal{X})$ consideramos el álgebra estándar de funciones, es decir:

$$(\alpha\mu + \beta\nu)(A) = \alpha\mu(A) + \beta\nu(A) \quad \forall A \in \beta(\mathcal{X}) .$$

Definición 3.1 *Un random walk m en \mathcal{X} es una familia de medidas de probabilidad $m_x(\cdot) \in \mathcal{P}_1(\mathcal{X})$ para cada $x \in \mathcal{X}$, tal que la aplicación $x \rightarrow m_x(A)$ es medible para todo $A \in \beta(\mathcal{X})$.*

Definición 3.2 *Un lazy random walk de lazyness α centrado en x se denota como m_x^α y se define como:*

$$m_x^\alpha := (1 - \alpha)\delta_x + \alpha m_x .$$

Definamos ahora dos operadores que resultarán cruciales en lo que sigue. Para definir adecuadamente los operadores, es necesario que la curvatura \mathcal{K} del *random walk* m sea distinta a $-\infty$. A esta hipótesis la llamaremos (H_1) .

Definición 3.3 *Operador $T_m : \mathcal{P}_1(\mathcal{X}) \rightarrow \mathcal{P}_1(\mathcal{X})$*

$$(\forall A \in \beta(\mathcal{X})) T_m(\mu)(A) := (\mu * m)(A) = \int_{\mathcal{X}} m_x(A) d\mu(x)$$

Operador $T_{m^\alpha} : \mathcal{P}_1(\mathcal{X}) \rightarrow \mathcal{P}_1(\mathcal{X})$

$$(\forall A \in \beta(\mathcal{X})) T_{m^\alpha}(\mu)(A) := (\mu * m^\alpha)(A) = \int_{\mathcal{X}} m_x^\alpha(A) d\mu(x)$$

Recordemos el Teorema 2.3, el cual establece que la cadena de Markov m tiene curvatura de Ricci discreta \mathcal{K} si y solo si para cualquier par de medidas de probabilidad μ y $\mu' \in \mathcal{P}_1(\mathcal{X})$ se tiene:

$$W_1(\mu * m, \mu' * m) \leq (1 - \mathcal{K})W_1(\mu, \mu').$$

Además, en este caso si $\mu \in \mathcal{P}_1(\mathcal{X})$ entonces $\mu * m \in \mathcal{P}_1(\mathcal{X})$.

A priori, $T_m(\mu)$ no pertenece necesariamente a $\mathcal{P}_1(\mathcal{X})$. No obstante, con la hipótesis (H_1) , el Teorema anterior es válido y así $T_m(\mu) \in \mathcal{P}_1(\mathcal{X})$.

Notemos además que en el caso de curvatura de Ricci discreta nula, $T_m(\mu)$ es un operador no expansivo con respecto a la distancia W_1 , es decir :

$$W_1(T_m(\mu), T_m(\mu')) \leq W_1(\mu, \mu').$$

3.1.1. Propiedad básica de los operadores T_m y T_{m^α}

Proposición 3.1 *Se tiene la siguiente relación entre el operador T_m y el operador T_{m^α} :*

$$T_{m^\alpha}(\mu) = (1 - \alpha)\mu + \alpha T_m(\mu).$$

Demostración: Para todo conjunto de Borel A , se tiene que:

$$\begin{aligned} T_{m^\alpha}(\mu)(A) &= \int_{\mathcal{X}} m_x^\alpha(A) d\mu(x) \\ &= \int_{\mathcal{X}} ((1 - \alpha)\delta_x(A) + \alpha m_x(A)) d\mu(x) \\ &= \int_A (1 - \alpha) d\mu + \alpha \int_{\mathcal{X}} m_x(A) d\mu(x) \\ &= (1 - \alpha)\mu(A) + \alpha(\mu * m)(A) \\ &= (1 - \alpha)\mu(A) + \alpha T_m(\mu)(A). \end{aligned}$$

■

3.2. Tasa de convergencia para iteraciones de Krasnosel'skii-Mann en distancia de Wasserstein 1

3.2.1. Introducción

Para un operador $T : \mathcal{X} \rightarrow \mathcal{X}$ en un espacio vectorial normado $(\mathcal{X}, \|\cdot\|)$, la iteración de Krasnosel'skii-Mann se define como:

$$x_{n+1} = (1 - \alpha_{n+1})x_n + \alpha_{n+1}Tx_n$$

con $\alpha_n \in (0, 1)$ para todo $n \in \mathbb{N}$ y $\sum \alpha_k(1 - \alpha_k) = \infty$.

Para un operador $T : \mathcal{C} \rightarrow \mathcal{C}$ un operador no expansivo ie: $\|Tx - Ty\| \leq \|x - y\|$, Cominetti, Soto y Vaisman deducen la siguiente estimación:

$$\|x_n - Tx_n\| \leq \frac{\text{diam}(C)}{\sqrt{\pi \sum_{i=1}^n \alpha_i(1 - \alpha_i)}}$$

La cota fue probada en [31] y conjeturada por Baillon y Bruck [12, 13].

En lo que sigue, se propone extender el resultado al espacio $\mathcal{P}_1(\mathcal{X})$ con la distancia de Wasserstein 1 .

3.2.2. Definición de iteración de Krasnosel'skii-Mann en el espacio de medidas $\mathcal{P}_1(\mathcal{X})$

En lo todo lo que sigue, supondremos que la cadena de Markov m posee curvatura de Ricci discreta nula. En el espacio $\mathcal{P}_1(\mathcal{X})$, la siguiente definición recursiva de μ_k corresponde a una iteración de *Krasnoselskii-Mann*:

$$\mu_{k+1} = (1 - \alpha_{k+1})\mu_k + \alpha_{k+1}T_m\mu_k$$

con T_m no expansivo (es decir $\mathcal{K} = 0$).

Proposición 3.2

$$W_1(\mu_k, \mu_{k+1}) = \alpha_{k+1}W_1(\mu_k, T_m\mu_k) .$$

Demostración:

$$\begin{aligned} W_1(\mu_k, \mu_{k+1}) &= \sup_{f \in \text{Lip}_1(\mathcal{X})} \left(\int_{\mathcal{X}} f d\mu_k - \left(\int_{\mathcal{X}} (1 - \alpha_{k+1})f d\mu_k + \alpha_{k+1} \int_{\mathcal{X}} f d(\mu_k * m) \right) \right) \\ &= \alpha_{k+1} \sup_{f \in \text{Lip}_1(\mathcal{X})} \left(\int_{\mathcal{X}} f d\mu_k - \int_{\mathcal{X}} f d(\mu_k * m) \right) \\ &= \alpha_{k+1}W_1(\mu_k, \mu_k * m) \\ &= \alpha_{k+1}W_1(\mu_k, T_m\mu_k) . \blacksquare \end{aligned}$$

Basándonos en el paper *On the rate of convergence of Krasnoselskii-Mann iterations and their connection with sums of Bernoullis* [31], se puede escribir :

$$\mu_n = \sum_{i=0}^n \pi_i^n T_m \mu_{i-1}$$

$$\mu_m = \sum_{j=0}^m \pi_j^m T_m \mu_{j-1}$$

con $n \geq m \geq 0$ y con $\pi_k^n = \alpha_k \prod_{j=k+1}^n (1 - \alpha_j)$.

Por convención, $\alpha_0 = 1$ y el término $T\mu_{-1}$ se interpreta como μ_0 .

Teorema 3.1 *Se define $c_{m,n}$ recursivamente con $c_{-1,n} = 1$ para todo $n \geq 0$ tal que :*

$$c_{m,n} = \sum_{j=0}^m \sum_{k=m+1}^n \pi_j^m \pi_k^n c_{j-1,k-1} \text{ para } 0 \leq m \leq n.$$

Entonces $W_1(x_m, x_n) \leq c_{m,n}$ para $0 \leq m \leq n$.

Demostración del teorema 3.1:

Sea $m \leq n$, sea $\mu_m = \sum_{i=0}^m \pi_i^m T_m \mu_{i-1}$ y sea $\mu_n = \sum_{j=0}^n \pi_j^n T_m \mu_{j-1}$. Tal como en el paper *Sharp convergence rates for averaged nonexpansive maps* [20], se conectarán las expresiones μ_m y μ_n resolviendo un problema de transporte óptimo entre π^m y π^n .

Sea $\mathcal{N} = \{-1, 0, 1, \dots\}$, definamos $d_{-1,-1} = 0$, $d_{-1,k} = d_{k,-1} = \text{diam}(\mathcal{P}_1(\mathcal{X}))$ para $k \in \mathbb{N}$ y definamos recursivamente $d_{m,n}$ para $m, n \in \mathcal{N}$ resolviendo una secuencia de problemas de transporte óptimo en el grafo bipartito de la figura 3.1. Específicamente, consideremos π_i^m como los suministros con los nodos a la izquierda $i = 0, \dots, m$ y π_j^n como los consumos en los nodos a las derecha $j = 0, \dots, n$, con costos unitarios $d_{i-1,j-1}$ en cada arco (i, j) y definamos:

$$d_{m,n} = \min_{z \in F^{m,n}} \sum_{i=0}^m \sum_{j=0}^n z_{ij} d_{i-1,j-1}$$

donde $F^{m,n}$ como el politopo de planes factibles de transporte z entre π^m y π^n , definido por las siguientes restricciones :

$$z_{ij} \geq 0 \text{ para } i = 0, \dots, m \text{ y } j = 0, \dots, n$$

$$\sum_{j=0}^n z_{ij} = \pi_i^m \text{ para } i = 0, \dots, m$$

$$\sum_{i=0}^m z_{ij} = \pi_j^n \text{ para } j = 0, \dots, n.$$

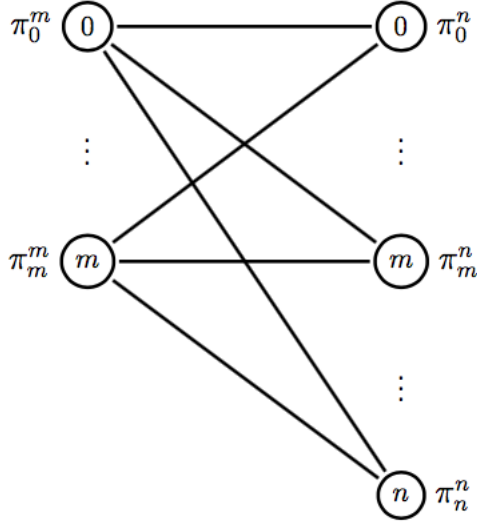


Figura 3.1: Problema de transporte óptimo recursivo. (Referencia: Sharp convergence rates for averaged nonexpansive maps.)

Así, para todo plan de transporte factible $z \in F^{m,n}$:

$$\begin{aligned} W_1(\mu_m, \mu_n) &= W_1\left(\sum_{i=0}^m \sum_{j=0}^n z_{ij} T_m \mu_{i-1}, \sum_{j=0}^n \sum_{i=0}^m z_{ij} T_m \mu_{j-1}\right) \\ &= W_1\left(\sum_{i=0}^m \sum_{j=0}^n z_{ij} (T_m \mu_{i-1}, T_m \mu_{j-1})\right). \end{aligned}$$

Por el Lema 2.1, la función W_1 es convexa:

$$W_1(\mu_m, \mu_n) \leq \sum_{i=0}^m \sum_{j=0}^n z_{ij} W_1(T_m \mu_{i-1}, T_m \mu_{j-1}).$$

Por la no expansividad del operador T :

$$W_1(\mu_m, \mu_n) \leq \sum_{i=0}^m \sum_{j=0}^n z_{ij} W_1(\mu_{i-1}, \mu_{j-1}).$$

Supongamos inductivamente que la cota $W_1(\mu_i, \mu_j) \leq d_{ij}$ se satisface para $i < m$ y $j < n$.

Por la hipótesis de inducción :

$$W_1(\mu_m, \mu_n) \leq \sum_{i=0}^m \sum_{j=0}^n z_{ij} d_{i-1, j-1}$$

Minimizando el lado derecho sobre $z \in F^{m,n}$, se obtiene $W_1(\mu_m, \mu_n) \leq d_{m,n}$, completando la etapa inductiva. Por resultados previos [20] : $d_{m,n} \leq c_{m,n}$, lo que finaliza la demostración. ■

De acuerdo al paper *On the rate of convergence of Krasnosel'skii-Mann iterations and their conection with sums of Bernoullis* [31], el Teorema anterior permite concluir:

Teorema 3.2

$$W_1(\mu_n, T\mu_n) \leq \frac{1}{\sqrt{\pi}} \frac{\text{diam}(\mathcal{P}_1(\mathcal{X}))}{\sqrt{\sum_{i=0}^n \alpha_i(1 - \alpha_i)}}$$

Observación :

- Si $\text{diam}(\mathcal{X})$ es acotado, $\text{diam}(\mathcal{P}_1(\mathcal{X}))$ también lo es. En efecto, si $\text{diam}(\mathcal{X}) = s < \infty$, entonces:

$$\begin{aligned} \text{diam}_{W_1}(\mathcal{P}_1(\mathcal{X})) &= \sup_{\mu, \nu \in \mathcal{P}_1(\mathcal{X})} W_1(\mu, \nu) \\ &= \sup_{\mu, \nu \in \mathcal{P}_1(\mathcal{X})} \inf_{\pi \in \Gamma(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{X}} d(x, y) d\pi(x, y) \leq s. \end{aligned}$$

- El resultado anterior permite establecer una cota con respecto a la distancia de Wasserstein 1 entre la medida μ_n y la medida $\mu_n * m$. En particular, si \mathcal{X} es acotado y $\sum \alpha_k(1 - \alpha_k) = \infty$:

$$\lim_{n \rightarrow \infty} W_1(\mu_n, \mu_n * m) = 0$$

Ejemplo (Grafos finitos)

Los grafos finitos poseen diámetro finito.

3.2.3. Generalización del Teorema 3.1

El siguiente Teorema (adaptado a nuestro contexto de espacio métrico) del paper *Rates of convergence for inexact Krasnosel'skii-Mann iterates in Banach spaces* [19] generaliza el resultado anterior en diversos sentidos.

Teorema 3.3 *Sea $(\mu_n)_{n \in \mathbb{N}}$ una secuencia generada por iteraciones de Krasnosel'skii-Mann y asumamos que la siguiente hipótesis se cumple :*

$$\exists p \geq 0 \text{ tal que } W_1(T_m \mu_n, \mu_0) \leq p \quad \forall n \in \mathbb{N}. \quad (\text{H}_0^{\text{ad}})$$

Entonces, para todo $n \in \mathbb{N}$ se tiene que:

$$W_1(\mu_n, T_m \mu_n) \leq p \min\left\{1, \frac{1}{\sqrt{\pi \sum_{k=1}^n a_k(1-a_k)}}\right\}$$

Observación :

- Si $\mathcal{P}_1(\mathcal{X})$ es acotado, entonces podemos escoger $p = \text{diam}_{W_1}(\mathcal{P}_1(\mathcal{X}))$ y obtener la siguiente cota:

$$W_1(\mu_n, T_m \mu_n) \leq \text{diam}_{W_1}(\mathcal{P}_1(\mathcal{X})) \min\left\{1, \frac{1}{\sqrt{\pi \sum_{k=1}^n a_k(1-a_k)}}\right\}.$$

Con lo cual se recupera el resultado:

$$W_1(\mu_n, T_m \mu_n) \leq \frac{1}{\sqrt{\pi}} \frac{\text{diam}(\mathcal{P}_1(\mathcal{X}))}{\sqrt{\sum_{i=0}^n a_i(1-a_i)}}$$

El Teorema 3.3 permite abordar otros casos en el que $\mathcal{P}_1(\mathcal{X})$ no es acotado.

Notemos que $\text{Fix}(T_m) = \{\nu \in \mathcal{P}_1(\mathcal{X}) : T_m \nu = \nu * m = \nu\}$, es decir $\text{Fix}(T_m)$ corresponde al conjunto de medidas invariantes para la cadena de markov m .

Lema 3.1 Para todo $\bar{\mu} \in \text{Fix}(T_m)$, la sucesión $(\mu_k)_{k \in \mathbb{N}}$ generada por Krasnosel'skii-Mann satisface :

$$W_1(\mu_{k+1}, \bar{\mu}) \leq W_1(\mu_k, \bar{\mu})$$

para todo $k \in \mathbb{N}$.

Demostración: Sea $\bar{\mu} \in \text{Fix}(T_m)$. Entonces:

$$W_1(\mu_{k+1}, \bar{\mu}) = W_1((1 - \alpha_{k+1})\mu_k + \alpha_{k+1}T_m \mu_k, (1 - \alpha_{k+1})\bar{\mu} + \alpha_{k+1}\bar{\mu}).$$

Por el Lema 2.1:

$$\begin{aligned} W_1((1 - \alpha_{k+1})\mu_k + \alpha_{k+1}T_m \mu_k, (1 - \alpha_{k+1})\bar{\mu} + \alpha_{k+1}\bar{\mu}) &\leq (1 - \alpha_{k+1})W_1(\mu_k, \bar{\mu}) + \alpha_{k+1}W_1(T_m \mu_k, \bar{\mu}) \\ &= (1 - \alpha_{k+1})W_1(\mu_k, \bar{\mu}) + \alpha_{k+1}W_1(T_m \mu_k, T_m \bar{\mu}) \\ &\leq (1 - \alpha_{k+1})W_1(\mu_k, \bar{\mu}) + \alpha_{k+1}W_1(\mu_k, \bar{\mu}) \\ &= W_1(\mu_k, \bar{\mu}). \end{aligned}$$

Proposición 3.3 Si $\text{Fix}(T_m) \neq \emptyset$ entonces (H_0^{ad}) se satisface con $p = 2W_1(x_0, \text{Fix}(T_m))$.

Demostración: Sea $\bar{\mu} \in \text{Fix}(T_m)$. Como aplicación directa del Lema 3.1 se obtiene:

$$\begin{aligned}
W_1(T_m\mu_k, \mu_0) &\leq W_1(T_m\mu_k, \bar{\mu}) + W_1(\bar{\mu}, \mu_0) \\
&= W_1(T_m\mu_k, T_m\bar{\mu}) + W_1(\bar{\mu}, \mu_0) \\
&\leq W_1(\mu_k, \bar{\mu}) + W_1(\bar{\mu}, \mu_0) \\
&\leq W_1(\mu_0, \bar{\mu}) + W_1(\bar{\mu}, \mu_0) \\
&= 2W_1(\mu_0, \bar{\mu}) .
\end{aligned}$$

Escogiendo $\bar{\mu}$ ínfimo sobre $\text{Fix}(T_m)$ se concluye. ■

Corolario 3.1 (*Existencia de un Attracting point*)

En el caso de que se cumplan las hipótesis del Teorema 2.5, se cumple (H_0^{ad}) pues el Teorema 2.5 garantiza la existencia de una medida invariante y de este modo $\text{Fix}(T_m) \neq \emptyset$.

Proposición 3.4 (Decaimiento rápido) *Supongamos que para todo $x \in \mathcal{X}$, existe una función $\phi(r)$ tal que $m_x(B(0, r)^c) \leq \phi(r)$ con $\int_{[0, \infty)} \phi(r)dr < \infty$. Entonces (H_0^{ad}) se satisface para una constante \bar{p} .*

Demostración:

$$\begin{aligned}
W_1(\mu_0, T_m\mu) &= \sup_{f \in \text{Lip}_1(\mathcal{X})} \int_{\mathcal{X}} f d\mu_0 - \int_{\mathcal{X}} f d(T_m\mu) \\
&\leq \sup_{f \in \text{Lip}_1(\mathcal{X})} \int_{\mathcal{X}} f d\mu_0 + \sup_{f \in \text{Lip}_1(\mathcal{X})} \int_{\mathcal{X}} f d(T_m\mu) .
\end{aligned}$$

Definamos A y B respectivamente como:

$$\begin{aligned}
A &= \sup_{f \in \text{Lip}_1(\mathcal{X})} \int_{\mathcal{X}} f d\mu_0 \\
B &= \sup_{f \in \text{Lip}_1(\mathcal{X})} \int_{\mathcal{X}} f dT_m\mu(x) .
\end{aligned}$$

Luego :

$$A = \sup_{f \in \text{Lip}_1(\mathcal{X})} \int_{\mathcal{X}} (f(x) - f(o))d\mu_0 + f(o) \leq \int_{\mathcal{X}} d(x, o)d\mu_0 + f(o) =: p_1 < \infty \quad (\mu_0 \in \mathcal{P}_1(\mathcal{X})) .$$

$$B = \sup_{f \in Lip_1(\mathcal{X})} \int_{\mathcal{X}} (f(x) - f(o)) d(T_m \mu)(x) + f(o) \leq \int_{\mathcal{X}} d(x, o) d(T_m \mu)(x) + f(o) .$$

Se define C como sigue:

$$\begin{aligned} C &:= \int_{\mathcal{X}} d(x, o) d(T_m \mu)(x) \\ &= \int_{\mathcal{X}} \int_{[0, \infty)} \mathbf{1}_{\{d(x, o) > r\}} dr dT_m(\mu)(x) \\ &= \int_{[0, \infty)} \int_{\mathcal{X}} \mathbf{1}_{\{d(x, o) > r\}} dT_m(\mu)(x) dr \\ &= \int_{[0, \infty)} T_m(\mu)(B(o, r)^c) dr . \end{aligned}$$

Notemos que:

$$T_m(\mu)(B(o, r)^c) = \int_{\mathcal{X}} m_x(B(o, r)^c) d\mu(x) \leq \phi(r) .$$

Por ende:

$$B \leq \int_{[0, \infty)} \phi(r) dr + f(o) =: p_2 < \infty .$$

Así:

$$W_1(\mu_0, T_m \mu) \leq \bar{p} := p_1 + p_2 .$$

En particular, para todo $n \in \mathbb{N}$:

$$W_1(\mu_0, T_m \mu_n) \leq \bar{p} .$$

■

Corolario 3.2 *Supongamos que para todo $x \in \mathcal{X}$, $\text{supp}(m_x) \subset K$ con K compacto. Entonces, se satisface H_0^{ad} .*

Demostración: Notemos que nos podemos restringir al caso anterior. En efecto, existe R suficientemente grande tal que $\text{supp}(m_x) \subset B(0, R)$. Podemos definir la siguiente función ϕ con decaimiento rápido:

$$\phi(r) = \begin{cases} 0, & \text{para } r > R \\ 1, & \text{para } r \leq R . \end{cases}$$

Notemos que $m_x(B(o, r)^c) \leq \phi(r)$ y ϕ es integrable:

$$\int_0^\infty \phi(r)dr = \int_0^R 1dr = R < \infty .$$

■

3.3. Convergencia débil de las iteraciones de Krasnosel'skii-Mann

3.3.1. Preliminares

Las siguientes definiciones y resultados conocidos serán cruciales en la siguiente sección.

Definición 3.4 *Se dice que una secuencia $\{\mu_n\}_{n \in \mathbb{N}}$ converge débil a μ si:*

$$\int_{\mathcal{X}} \phi(x) d\mu_n \rightarrow \int_{\mathcal{X}} \phi(x) d\mu$$

para toda función ϕ acotada y continua en \mathcal{X} .

El siguiente Teorema establece que con para $\mathcal{X} \subset \mathbb{R}^d$ compacto la convergencia débil es equivalente a la convergencia en la métrica de Wasserstein.

Teorema 3.4 [32]

Sea $\mathcal{X} \subset \mathbb{R}^d$ un compacto. Entonces, una secuencia $\{\mu_n\}_{n \in \mathbb{N}} \subset \mathcal{P}_1(\mathcal{X})$ converge débil a μ si y solo si $W_1(\mu_n, \mu) \rightarrow 0$.

Definición 3.5 *Una secuencia de medidas de probabilidad $\{\mu_n\}_{n \in \mathbb{N}}$ se dice tensa si para todo $\varepsilon > 0$ existe un conjunto K compacto tal que para todo $n \in \mathbb{N}$:*

$$\mu_n(K^c) < \varepsilon$$

Teorema 3.5 (Teorema de Prokhorov) [30]

Toda sucesión de medidas de probabilidad $\{\mu_n\}_{n \in \mathbb{N}}$ tensa posee una subsucesión débil convergente.

3.3.2. Convergencia débil de las iteraciones de Krasnosel'skii-Mann

Definición 3.6 Para una medida μ definida en $\mathcal{X} \subset \mathbb{R}^d$ cerrado y para todo $n \in \mathbb{N}$ se define el momento n -ésimo como:

$$m_n(\mu) := \int_{\mathcal{X}} d(y, o)^n d\mu(x)$$

dato o un punto fijo y arbitrario.

Proposición 3.5 Sea una sucesión $\{\mu_n\}_{n \in \mathbb{N}}$ definida mediante las iteraciones de Krasnosel'skii-Mann. Sea o un punto fijo y arbitrario. Si se cumple que :

$$\sup_{x \in \mathcal{X}} \int_{\mathcal{X}} d(y, o) dm_x(y) < \infty \text{ y } m_1(\mu_0) < \infty ,$$

entonces $\sup_{n \in \mathbb{N}} m_1(\mu_n) < \infty$.

Demostración:

La secuencia $\{\mu_n\}_{n \in \mathbb{N}}$ se define como:

$$\mu_{n+1} = (1 - \alpha_{n+1})\mu_n + \alpha_{n+1}T_m\mu_n$$

Luego:

$$\begin{aligned} \int_{\mathcal{X}} d(y, o) d\mu_{n+1}(x) &= (1 - \alpha_{n+1}) \int_{\mathcal{X}} d(y, o) d\mu_n(x) + \alpha_{n+1} \int_{\mathcal{X}} d(y, o) d(T_m\mu)(x) \\ &= (1 - \alpha_{n+1}) \int_{\mathcal{X}} d(y, o) d\mu_n(x) + \alpha_{n+1} \int_{\mathcal{X}} d(y, o) \int_{\mathcal{X}} dm_x(y) d\mu_n(x) \\ &= (1 - \alpha_{n+1}) \int_{\mathcal{X}} d(y, o) d\mu_n(x) + \alpha_{n+1} \int_{\mathcal{X}} \int_{\mathcal{X}} d(y, o) dm_x(y) d\mu_n(x) . \end{aligned}$$

Sea $C = \sup_{x \in \mathcal{X}} \int_{\mathcal{X}} d(y, o) dm_x(y)$. Entonces:

$$\int_{\mathcal{X}} d(y, o) d\mu_{n+1}(x) \leq (1 - \alpha_{n+1}) \int_{\mathcal{X}} d(y, o) d\mu_n(x) + \alpha_{n+1}C$$

$$m_1(\mu_{n+1}) \leq (1 - \alpha_{n+1})m_1(\mu_n) + \alpha_{n+1}C \leq \max\{C, m_1(\mu_n)\} .$$

Por ende, $m_1(\mu_n) \leq \max\{C, m_1(\mu_0)\} < \infty$ para todo $n \in \mathbb{N}$. ■

Proposición 3.6 Sea $\mu \in \mathcal{P}_1(\mathcal{X})$ una medida invariante para la cadena de Markov m . Entonces:

$$\sup_x \int_{\mathcal{X}} d(y, o) dm_x(y) < \infty .$$

Demostración:

$$\int_{\mathcal{X}} d(y, o) dm_x(y) = \int_{\mathcal{X}} d(y, o) (dm_x(y) - d\mu(y)) + \int_{\mathcal{X}} d(y, o) d\mu(y)$$

$$\int_{\mathcal{X}} d(y, o) dm_x(y) = \int_{\mathcal{X}} d(y, o) d(T_m \delta_x(y) - T_m \mu(y)) + m_1(\mu) .$$

Se define $f(y) := d(y, o)$, función 1-Lipschitz. Luego:

$$\int_{\mathcal{X}} d(y, o) dm_x(y) = Mf(x) - \int_{\mathcal{X}} Mf(y) d\mu(z) + m_1(\mu)$$

$$\int_{\mathcal{X}} d(y, o) dm_x(y) = \int_{\mathcal{X}} (Mf(x) - Mf(y)) d\mu(z) + m_1(\mu) .$$

Por el Lema 2.2 :

$$Mf(x) - Mf(y) \leq d(x, y) .$$

Así:

$$\int_{\mathcal{X}} d(y, o) dm_x(y) \leq \int_{\mathcal{X}} d(x, y) d\mu(z) + m_1(\mu) < \infty .$$

Tomando supremo:

$$\sup_{x \in \mathcal{X}} \int_{\mathcal{X}} d(y, o) dm_x(y) < \infty .$$

■

En el siguiente Teorema nos restringimos al caso $\mathcal{X} \subset \mathbb{R}^d$ compacto.

Teorema 3.6 Sea $\mathcal{X} \subset \mathbb{R}^d$ compacto. Sea $\{\mu_n\}_{n \in \mathbb{N}} \in \mathcal{P}(\mathcal{X})$ una secuencia de medidas definidas mediante las iteraciones de Krasnosel'skii-Mann. Supongamos $\sup_{x \in \mathcal{X}} \int_{\mathcal{X}} d(y, o) dm_x(y) < \infty$ y que $m_1(\mu_0) < \infty$. Entonces, la sucesión $\{\mu_n\}_{n \in \mathbb{N}}$ converge débil a una medida μ que es invariante para la cadena de Markov m .

Demostración: La demostración se basa en 3 etapas.

Primera etapa: La sucesión $\{\mu_n\}_{n \in \mathbb{N}}$ tiene puntos de acumulación.

Por la proposición 3.5, $\sup_n m_1(\mu_n) = C < \infty$ para todo $n \in \mathbb{N}$.

Sea $K := X \cap B(\bar{0}, r)$ que es un compacto en \mathcal{X} . Entonces para todo $n \in \mathbb{N}$:

$$\mu_n(K^c) = \int_{\mathcal{X}} 1_{K^c}(x) d\mu_n(x) \leq \int_{\mathcal{X}} \frac{\|x\|}{R} d\mu_n(x) = \frac{m_1(\mu_n)}{R} \leq \frac{C}{R} .$$

Para todo $\varepsilon > 0$, $\exists R$ tal que $\mu_n(K^c) < \varepsilon$ para todo $n \in \mathbb{N}$. Luego la secuencia $\{\mu_n\}_{n \in \mathbb{N}}$ es tensa. Por ende, por el Teorema de Prokhorov, existe una subsucesión débil convergente, es decir, existe una sucesión $\{\mu_{n_k}\}_{k \in \mathbb{N}}$ tal que $\mu_{n_k} \rightharpoonup \mu$.

Segunda etapa: μ es una medida invariante para la cadena de Markov m .

Por la desigualdad triangular y la no expansividad del operador T_m :

$$\begin{aligned} W_1(\mu, T\mu) &\leq W_1(\mu, \mu_{n_k}) + W_1(\mu_{n_k}, T_m \mu_{n_k}) + W_1(T_m \mu_{n_k}, T\mu) \\ &\leq 2W_1(\mu, \mu_{n_k}) + W_1(\mu_{n_k}, T_m \mu_{n_k}) . \end{aligned}$$

Por el Teorema 3.4 $\lim_{k \rightarrow \infty} W_1(\mu, \mu_{n_k}) = 0$ y por el teorema 3.2 $\lim_{k \rightarrow \infty} W_1(\mu_{n_k}, T_m \mu_{n_k}) = 0$. Luego, $W_1(\mu, T\mu) = 0$ por ende μ es un punto fijo del operador T_m , es decir, μ es una medida invariante para la cadena de Markov m .

Tercera etapa: La sucesión $\{\mu_n\}_{n \in \mathbb{N}}$ converge débil a μ .

Por el Lema 3.1, la secuencia $W_1(\mu_n, \mu)$ es decreciente cualquiera sea $\mu \in \text{Fix}(T_m)$. Por consiguiente:

$$\lim_{n \rightarrow \infty} W_1(\mu_n, \mu) = \lim_k W_1(\mu_{n_k}, \mu) = 0$$

con lo cual $\mu_n \rightharpoonup \mu$.

■

Conclusión

En el trabajo de tesis presente se estableció una tasa de convergencia de iteraciones de Krasnosel'skii-Mann para cadenas de Markov con curvatura de Ricci discreta nula (Teorema 3.3) y diversos casos en los que la tasa de convergencia es válida. Además se logró establecer con ciertas hipótesis la convergencia débil de la sucesión $\{\mu_n\}_{n \in \mathbb{N}}$ definida por Krasnosel'skii-Mann a una medida que es invariante para la cadena de Markov (Teorema 3.6).

Este trabajo ha abierto puertas a diversas interrogantes:

- *¿ Se puede extender estos resultados a distancias de Wasserstein W_p con $p \geq 1$?*
- *¿ Se puede extender estos resultados a la distancia de Wasserstein 1 matricial introducida por Yongxin Chen, Tryphon T.Georgiou, Lipeng Ning y Allen Tannebaum en 2017 [5] ?*
- *¿ Existe una condición adicional débil que permita que las medidas μ_n definidas mediante las iteraciones de Krasnosel'skii-Mann satisfagan el fenómeno de concentración de la medida ?*

Desde una perspectiva personal, se aprendió enormemente sobre concentración de la medida, sobre transporte óptimo y sobre curvatura de Ricci discreta. Fue un inmenso placer aprender tantas cosas en este tiempo.

Bibliografía

- [1] *Sergei A. Utev Aleksandr A. Borovkov. An inequality and a characterization of the normal distribution connected with it. Teor. Veroyatnost i Primenen, 1983.*
- [2] *László Leindler András Prékopa. Logarithmic concave measures with application to stochastic programming. Acta Sci. Math., 1971.*
- [3] *Cécile Ané, Sébastien Blachère, Djalil Chafaï, Pierre Fougères, Ivan Gentil, Florent Malrieu, Cyril Roberto, and Grégory Scheffer. Sur les inégalités de Sobolev logarithmiques. Société mathématique de France, 1998.*
- [4] *Keith Ball. An elementary introduction to modern convex geometry. Mathematical Science Research Institute, 1997.*
- [5] *Yongxin Chen, Tryphon Georgiou, Lipeng Ning, and Allen Tannenbaum. Matricial wasserstein-1 distance. Functional analysis, 2017.*
- [6] *Imre Csiszár. Information- type measures of difference of probability distributions and indirect observations. Studia Sci. Math Hungar, 1967.*
- [7] *Roland Lvovich Dobrushin. Definition of a system of random variables by conditional distributions. Teor. Veroyatnost. i Primenen, 1970.*
- [8] *Michel Ledoux Dominique Bakry, Ivan Gentil. Analysis and Geometry of Markov Diffusion Operators. Springer, 2013.*
- [9] *Cédric Villani Felix Otto. Generalization of an inequality by talagrand and links with the logarithmic sobolev inequality. J. Funct. Anal., 2000.*
- [10] *Nathael Gozlan. A characterization of dimension free concentration in terms of transportation inequalities. Ann. Probab., 2009.*
- [11] *Li-Ming Wu Hacène Djellout, Arnaud Guillin. Transportation cost-information inequalities and applications to dynamical systems and diffusions. Ann. Probab., 2002.*
- [12] *Ronald Bruck Jean Bernard Baillon. Optimal rates of asymptotic regularity for averaged nonexpansive mapping. Proceeding of the Second International Conference on Fixed Point Theory and Applications, 1992.*

- [13] Ronald Bruck Jean Bernard Baillon. *The rate of asymptotic regularity is $\mathcal{O}(1/\sqrt{n})$. Theory and Applications of Nonlinear Operators of Accretive and Monotone Types, 1996.*
- [14] Shiping Liu Jürgen Jost. *Ollivier's ricci curvature, local clustering and curvature dimension inequalities on graphs. Discrete and Computational Geometry, 2013.*
- [15] Salomon Kullback. *A lower bound for discrimination information in terms of variation. IEEE Trans. Information Theory, 1967.*
- [16] Michel Ledoux. *The Concentration of Measure Phenomenon. American Mathematical Society, 2001.*
- [17] G. Rubinstein Leonid Kantorovich. *On a space of completely additive functions. Vestnik Leningradstago Universitata, 1952.*
- [18] C.L. Mallows. *A note on asymptotic joint normality. Ann. Math. Stat., 1970.*
- [19] Matías Pavez-Signé Mario Bravo, Roberto Cominetti. *Rates of convergence for inexact krasnosel'skii-mann iterations in banach spaces. Mathematical Programming, 2019.*
- [20] Roberto Cominetti Mario Bravo. *Sharp convergence rates for averaged nonexpansive maps. Israel Journal of Mathematics, 2018.*
- [21] Katalin Marton. *A simple proof of the blowing-up lemma. IEEE Trans. Information Theory, 1986.*
- [22] Vitali Milman Mijaíl Gromov. *Generalization of the spherical isoperimetric inequality to uniformly convex banach spaces. Compositio Math., 1976.*
- [23] Vitali D. Milman Mikhael Gromov. *A topological application of the isoperimetric inequality. American. J. Math., 1983.*
- [24] Christian Léonard Nathael Gozlan. *Transport inequalities. a survey. Markov Process. Related Fields, 2010.*
- [25] Paul-Marie Samson Nathael Gozlan, Cyril Roberto. *A new characterization of talagrand's transport-entropy inequalities and applications. Ann. Probab., 2011.*
- [26] Yann Ollivier. *Ricci curvature of markov chains on metric spaces. Journal of Functional Analysis, 2008.*
- [27] Robert Osserman. *The isoperimetric inequality. Bull. Amer. Math. Soc., 1978.*
- [28] Arnaud Guillin Patrick Cattiaux. *On quadratic transportation cost inequalities. J. Math. Pures Appl., 2006.*
- [29] Mark Semenovitch Pinsker. *Information and information stability of random variables and processes. Holden-Day, 1964.*

- [30] *Yuri Prokhorov. Convergence of random processes and limit theorems in probability theory.* Theory of Prob. And Appl., 1956.
- [31] *José Vaisman Roberto Cominetti, José Soto. On the rate of convergence of krasnoselskii-mann iterations and their connection with sums of bernoullis.* Israel Journal of Mathematics, 2014.
- [32] *Filippo Santambrogio. Optimal Transport for Applied Mathematicians.* Birkhauser, 2015.
- [33] *Michel Ledoux Sergey Bobkov. From brunn-minkowski to brascamp-lieb and to logarithmic sobolev inequalities.* Geom. funct. anal., 2000.
- [34] *Waclaw Sierpiński. Sur la question de la mesurabilité de la base de m. hamel.* Fund. Math., 1920.
- [35] *Pascal Massart Stéphane Boucheron, Gábor Lugosi. Concentration inequalities: a non-asymptotic theory of independence.* 2016.
- [36] *Gideon Schechtman Vitali Milman. Asymptotic theory of finite dimensional normed space.* Springer-Verlag, 1986.
- [37] *Elizabeth L. Wilmer Yuval Peres, David A. Levin. Markov Chains and Mixing times.* American Mathematical Society, 2017.