UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

EQUIVALENCIA FUERTEMENTE ESTOCÁSTICA DE CADENAS DE MARKOV

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

MATEMÁTICO

P R E S E N T A :

IVÁN IXCÓATL JUÁREZ LÓPEZ

TUTOR
DR. RICARDO GÓMEZ AÍZA

2014

CD. UNIVERSITARIA, D.F.

FACULTAD DE CIENCIAS
UNAM

Índice general

Introducción

El propósito de esta tesis es estudiar el problema de clasificación de cadenas de Markov al verlas como sistemas dinámicos. En probabilidad, una cadena de Markov es un proceso estocástico $(X_n)_{n\in\mathbb{Z}}$ que toma valores en un conjunto finito $\mathcal{A}$ (llamado comúnmente *espacio de estados*), y que cumple con la *propiedad de Markov*, es decir que $\forall\, n \in \mathbb{Z}$, para todo entero $\forall\, k \in \mathbb{N}$ y $\forall\, x_1, \ldots x_k, x_{k+1} \in \mathcal{A}$,

$$\begin{aligned} \mathbb{P}(X_{n+1} = x_{n+1}|X_n = x_n, \cdots X_{n-k} = x_{n-k}) &= \mathbb{P}(X_{n+1} = x_{n+1}|X_n = x_n) \\ &= \mathbb{P}(X_1 = x_{n+1}|X_0 = x_n). \end{aligned} \tag{1}$$

La cadena está determinada por su *matriz de transición de probabilidades* P, la cual es una matrix de $|\mathcal{A}| \times |\mathcal{A}|$ y está definida $\forall\, i, j \in \mathcal{A}$ por

$$P_{i,j} = \mathbb{P}(X_1 = i|X_0 = i).$$

Para nosotros, una cadena de Markov es un sistema dinámico $(X_A, \mathcal{B}, T, \mu_P)$, donde (X_A, T) es el *shift de tipo finito* (shift por vértices) inducido por la matriz de adjacencia $A = P^0$ (donde P^0 denota a la matriz que resulta de P al elevar cada una de sus entradas a la potencia 0), es decir,

$$X_A = \{(x_n)_{n\in\mathbb{Z}} \in \mathcal{A}^{\mathbb{Z}} : A_{x_n,x_{n+1}} = 1 \ \ \forall n \in \mathbb{Z}\},$$

$\mathcal{B}$ es la σ-álgebra de Borel (equipando a X_A con la topología inducida de $\mathcal{A}^{\mathbb{Z}}$ con la topología producto, considerando a $\mathcal{A}$ como un espacio topológico discreto), T es el mapeo *shift* y μ_P es la *medida de Markov* inducida por P.

Para clasificar sistemas dinámicos, la noción de ser "el mismo sistema" es precisamente el concepto de "conjugación" (definida en el capítulo 1 para espacios shift y estudiada en el capítulo 4). En el caso de los shifts de tipo finito, existen dos resultados centrales en términos de conjugaciones. El primero es el "Teorema de Desacomposición", el cual descompone a cualquier conjugación en unidades fundamentales (o *elementales*) llamadas "escisiones" y "amalgamaciones". Por otro lado, dos shifts de tipo finito (de memoria 1) son conjugados si y sólo si sus matrices de adyacencia A y B (sobre $\{0, 1\}$) son *shift-fuertemente equivalentes*. El ser shift-fuertemente equivalente es una relación de equivalencia definida en términos de las matrices de adyacencia de las gráficas dirigidas que inducen a los shifts de tipo finito (de memoria 1). Este resultado es conocido como el "Teorema de Clasificación de Shifts de Tipo Finito", y ambos teoremas los desarrolló R. Williams en [9]. El propósito fundamental de esta tesis es presentar en forma extendida y comprensible la generalización de este teorema de clasificación en el contexto de teoría ergódica, de forma que ahora estarán presentes medidas invariantes μ_P y μ_Q inducidas por matrices de probabilidades de transición P

y Q (sobre $[0,1]$), y no sólo además de las matrices de de adyacencia $A = P^0$ y $B = Q^0$. Esta generalización la realizó por primera por Parry y Williams [8].

La tesis está organizada de la siguiente manera. Los espacios shift los abordaremos en el capítulo 1, los shifts de tipo finito y sus representaciones por medio de gráficas los abordaremos en el capítulo 2, estos dos capítulos son basados en [3]. La parte ocupada de entropía y particiones es revisada en el capítulo 3, auxiliados en su mayoría por [2], y será necesaria para desarrollar el contenido del capítulo 5. Lo que concierne al teorema de descomposición y las prinicipales equivalencias topológicas de los shift's está enfocado en el capítulo 4, nuevamente apoyando principalmente en [3]. El tema principal se trata en el capítulo 5 basado en escencia en [8]. Por último en el apéndice A y B presentamos resultados auxiliares basándonos en [7] , [6] y [4].

Esperamos también que este texto sirva de referencia para más generalizaciones del problema, como por ejemplo desarrollar el concepto de shift equivalente a generalizaciones de shifts de tipo finito, como los shifts de árboles de tipo finito desarrollados por N. Aubrun y M.P. Béal (ver [10]), o los sistemas multidimensionales, usando el formalismo matricial desarrollado por M. Schraudner [11].

Capítulo 1

ESPACIOS SHIFT

1.1. Shifts completos

Definición 1.1.1. Sea $\mathcal{A}$ un conjunto finito al cual llamaremos *alfabeto*, el *$\mathcal{A}$-shift completo* (o *full $\mathcal{A}$-shift*) es el conjunto de todas las sucesiones bi-infinitas de símbolos de $\mathcal{A}$, es decir el producto cartesiano de $\mathcal{A}$ sobre $\mathbb{Z}$, o más precisamente,

$$\mathcal{A}^{\mathbb{Z}} = \{x = (x_i)_{i\in\mathbb{Z}} : x_i \in \mathcal{A}, \forall i \in \mathbb{Z}\}.$$

Cada elemento del $\mathcal{A}$-shift completo será llamado *punto* del $\mathcal{A}$-shift completo.

A lo largo del resto de este capítulo, $\mathcal{A}$ denotará a un alfabeto.

Definición 1.1.2. Un *bloque* (o *palabra*) es una secuencia finita de símbolos de un alfabeto $\mathcal{A}$. La *longitud* de un bloque u es el número de símbolos de $\mathcal{A}$ que contiene y se escribe $|u|$. También conviene incluir entre la palabras o bloques a la secuencia de no símbolos llamado el *bloque vacío* denotado como ϵ, el cual es el único bloque de longitud cero.

Por ejemplo, si $\mathcal{A} = \{D, E, F, G, H\}$ entonces $DEFFDGH$ es un bloque o una palabra sobre el alfabeto $\mathcal{A}$ y tiene longitud 7.

Un bloque que tiene longitud n es llamado *n-bloque*. Al conjunto de todos los n-bloques de $\mathcal{A}$ se escribe como

$$\mathcal{A}^n = \underbrace{\mathcal{A} \times \mathcal{A} \times \cdots \times \mathcal{A}}_{n\text{-veces}}$$

y definimos también

$$\mathcal{A}^* = \bigcup_{n=0}^{\infty} \mathcal{A}^n.$$

Una *subpalabra* (o *subbloque*) de $u = a_1a_2a_3\cdots a_n \in \mathcal{A}^n$ es un elemento $z = a_ia_{i+1}\cdots a_j$ con $1 \leq i \leq n$ e $i \leq j \leq n$. Si $x \in \mathcal{A}^{\mathbb{Z}}$ y $i, j \in \mathbb{Z}$ son tales que $i \leq j$, entonces definimos

$$\begin{aligned} x_{[i,j]} &= x_ix_{i+1}\cdots x_j \\ x_{[i,j)} &= x_ix_{i+1}\cdots x_{j-1} \text{ si } j > i, \\ x_{[i,\infty)} &= x_ix_{i+1}\cdots \end{aligned}$$

Dos bloques $u, v \in \mathcal{A}^*$ pueden ser escritos juntos o *concatenarse*, escribiendo primero u y luego v y formando así un nuevo bloque llamado uv con longitud $|uv| = |u| + |v|$. Hay que notar que esta no es una operación conmutativa, sin embargo se observa fácilmente que uv y vu tienen la misma longitud, y por convención se establece que $\epsilon u = u\epsilon = u$. Si $n \geq 1$ entonces u^n denota la concatenación de u un total de n veces, y escribimos $u^0 = \epsilon$ y de ahí también se sigue la ley de los exponentes $u^n u^m = u^{n+m}$ y se puede escribiremos tambi $u^\infty = \cdots uu.uuu\cdots$.

El subíndice i de un punto $x = (x_i)_{i\in\mathbb{Z}} \in \mathcal{A}^\mathbb{Z}$ se puede pensar como indicador de tiempo (discreto), y esta es una de las principales relaciones que más adelante veremos con este tipo de espacios y cadenas de Markov.

Definición 1.1.3. Se define la *función shift* $T\colon \mathcal{A}^\mathbb{Z} \to \mathcal{A}^\mathbb{Z}$ como $T(x) = (T(x_i))_{i\in\mathbb{Z}} = (x_{i+1})_{i\in\mathbb{Z}}$, es decir,

$$(T(x))_i = x_{i+1} \quad \forall\, i \in \mathbb{Z}.$$

Definición 1.1.4. Se dice que x es un *punto periódico* para T si $T^n(x) = x$ p.a. $n \in \mathbb{N}$, en cuyo caso diremos que x tiene *periodo* n, y también definimos el *periodo mínimo* de x como el número mínimo $n \in \mathbb{N}$ tal que $T^n(x) = x$. Si es el caso que $T(x) = x$, entonces diremos que x es un *punto fijo* (puntos periódicos de periodo mínimo 1).

Observamos que un punto $x \in \mathcal{A}^\mathbb{Z}$ tiene periodo n si y sólo si $x = u^\infty$ para algún bloque $u \in \mathcal{A}^n$.

1.2. Espacios shift

Primero que nada hablaremos de algunos conceptos nuevos surgidos de las siguientes preguntas: ¿qué es que una palabra ocurra en un elemento de un espacio shift?, ¿qué es una colección de bloques prohibidos? y finalmente ¿cuál es el conjunto generado prohibiendo dichos bloques? Se dice que un bloque $w \in \mathcal{A}^n$ ocurre en $x \in \mathcal{A}^\mathbb{Z}$ si existe $i \in \mathbb{Z}$ tal que $x_{[i,i+n-1]} = w$, en cuyo caso diremos que *w ocurre* en x en la posición (o *al tiempo*) i. Luego llamaremos un *conjunto de bloques prohibidos* a todo aquel subconjunto de bloques $\wp \subset \mathcal{A}^*$ que queramos que no ocurran en ningún elemento de cierto subconjunto del full $\mathcal{A}$-shift, el cual será nuestro "subespacio" a estudiar. Denotamos entonces al conjunto generado por la prohibición de $\wp$ como $X_\wp \subset \mathcal{A}^\mathbb{Z}$, es decir,

$$X_\wp = \{x = (x_i)_{i\in\mathbb{Z}} \in \mathcal{A}^\mathbb{Z} : \forall\, w \in \wp,\ w \text{ no ocurre en } x\}.$$

Definición 1.2.1. Sea $X \subseteq \mathcal{A}^\mathbb{Z}$ es un *espacio shift* (o *subshift*) si $X = X_\wp$ p.a. $\wp \subset \mathcal{A}^*$.

1.3. Lenguaje

Definición 1.3.1. Sea $X \subset \mathcal{A}^\mathbb{Z}$ un subconjunto del $\mathcal{A}$-shift completo. Definimos los *bloques de X de longitud* $n \in \mathbb{N}$ como

$$B_n(X) = \{w \in \mathcal{A}^n : w \text{ ocurre en algún elemento } x \in X\}.$$

Definimos también al *lenguaje de X* como $B(X) = \bigcup_{n\in\mathbb{N}} B_n(X)$.

No cualquier subconjunto de $\mathcal{A}^*$ corresponde al lenguaje de un espacio shift. El siguiente resultado caracteriza los lenguajes de los espacios shift.

Proposición 1.3.1. *Sea X un espacio shift, y $\ell = B(X)$ su propio lenguaje.*

1. *Si $w \in \ell$ entonces:*

   a) *Cualquier subbloque de w está en ℓ.*

   b) *Existen dos bloques u, v de longitudes al menos uno y tales que $uwv \in \ell$.*

2. *Los lenguajes de los espacios shift están caracterizados por la propiedad 1, es decir, $\ell = B(X)$ p.a. espacio shift X si y sólo si ℓ cumple 1.*

3. *El lenguaje de los espacios shift determina al espacio shift, lo que es igual que $X = X_{B(X)^c}$ de aquí se puede decir que dos espacios shift son el mismo si tienen el mismo lenguaje.*

Demostración. Primero sea $w \in \ell$. Como w está en el lenguaje de X, existe $x \in X$ tal que w ocurre en x, de ahí que cualquier subbloque de w ocurre en x, y en consecuencia cualquier z subbloque de w cumple que $z \in \ell$, lo cual concluye la prueba de $1a$. Ahora para $1b$, como $w \in \ell$, existe $x \in X$ tal que w ocurre en x, lo que significa que $w = x_{[i,j]}$. Sea $u = x_{[t,i-1]}$ y $v = x_{[j+1,r]}$ con $r > j > i > t$, de forma que $uwv = x_{[t,r]}$ y entonces uwv ocurre en x, lo que significa $uwv \in \ell$, lo cual concluye la prueba de 1. Para 2, la ida se deduce de 1, y para el regreso, sea $X = X_{\ell^c}$. Para demostrar el resultado basta probar que $B(X) = \ell$. Si $w \in B(X)$, entonces existe $x \in X$ tal que w ocurre en x, de lo cual obtenemos que $w \notin \ell^c$, por lo que entonces $w \in \ell$. Ahora para la implicación inversa, si $w \in \ell$, entonces podemos suponer, por $1a$, que $|w| = 1$. Ocupando 1, vemos que existen $u_n, v_n \in \mathcal{A}\,\forall n \in \mathbb{N}$ tales que $u_1 w v_1 \in \ell$ y recursivamente sea $w_n = u_{n-1} w_{n-1} v_{n-1}$, con $w_n \in \mathcal{A}\ \forall\ n \in N$, lo cual es posible por inducción (tomando como base de inducción que $w_2 \in \ell$ y luego suponiendo $w_n \in \ell$, probamos que $w_{n+1} \in \ell$: como las u_n y v_n son exactamente las que pedimos para que $u_n w_n v_n \in \ell$ el resultado es inmediato, por lo tanto $w_n \in \ell\ \ \forall\ n \in \mathbb{N}$). Luego entonces podemos definir a $x \in \mathcal{A}^{\mathbb{Z}}$ de la siguiente manera: $x_0 = w$ y luego, $\forall\ n \in \mathbb{N} \setminus \{0\}$, $x_{-n} = u_n$ y $x_n = v_n$. Ahora veamos que $x \in X$ lo cual bastaría para concluir la prueba debido a que de ser cierto w ocurriría en un elemento de X de lo cual se sigue $w \in B(X)$ con lo cual nos daría la igualdad buscada entre $B(X)$ y ℓ. Para probar que $x \in X$ sólo necesitamos ver que ningún $x_{[i,j]} \in \ell^c$ por la forma en que definimos a x, $x_{[i,j]}$ es subbloque de algún w_n, pero como $w_n \in \ell\ \ \forall\ n \in \mathbb{N}$ por la otra propiedad $1a$, se sigue que $x_{[i,j]} \in \ell\ \ \forall i, j\ \in \mathbb{Z}$, con $i \leq j$, por lo tanto $x_{[i,j]} \notin \ell^c$ lo cual implica que x no tiene bloques prohibidos y por lo tanto $x \in X$ lo cual concluye la prueba de 2.

Ahora para probar 3 sea $x \in X$. Si w ocurre en x, entonces $w \in B(X)$ lo que significa que $w \notin B(X)^c$ lo que nos dice que cualquier bloque que ocurra en x no está en el complemento de $B(X)$ lo que significa que x no tiene bloques prohibidos por $B(X)^c$ lo que quiere decir que $x \in X_{B(X)^c}$. Ahora si ocurre que $x \in X_{B(X)^c}$, entonces x no tiene bloques de $B(X)^c$ lo cual nos dice que $x_{[i,j]} \in B(X)\ \ \forall\, i, j \in \mathbb{Z}$, con $i \leq j$, lo cual nos dice que $x \in X$. □

Corolario 1.3.1. *Sea $X \subset \mathcal{A}^{\mathbb{Z}}$ subconjunto del $\mathcal{A}$-shift completo. Entonces X es un espacio shift si y sólo si siempre que $x \in \mathcal{A}^{\mathbb{Z}}$ y cada $x_{[i,j]} \in B(X)$, entonces $x \in X$.*

Demostración. La primera implicación es consecuencia directa de nuestro resultado anterior. Para el regreso $B(X)$ cumple 1 de la Proposición 1.3.1 puesto que para $1a$ tomamos $w \in B(X)$ y entonces existe $x \in X$ tal que w ocurre en x, es decir, $w = x_{[i,j]}$ para enteros $i \leq j$ y entonces cualquier subbloque z de w también se ve como $z = x_{[l,k]}$. Entonces z ocurre en x y así $z \in B(X)$, y luego, si $w \in B(X)$, entonces existe $y \in X$ tal que w ocurre en $y \in X$, es decir, $w = y_{[i,i+|w|-1]}$. Para concluir sean $f = y_{i-1}$ y $g = y_{i+|w|}$, de forma que $fwg \in B(X)$, lo cual termina la prueba. □

Definición 1.3.2. Un espacio shift es *irreducible* si para cualesquiera dos bloques $u, v \in B(X)$ existe $w \in B(X)$ tal que $uwv \in B(X)$.

1.4. Presentaciones en bloques superiores y potencias de espacios shift

Definición 1.4.1. Sea X un espacio shift sobre $\mathcal{A}$ y sea $R \geq 1$ un entero. Consideremos a $\mathcal{A}_X^{[R]} = B_R(X)$ como un nuevo alfabeto y definamos una función (código) $\beta_R \colon X \to (\mathcal{A}_X^{[R]})^{\mathbb{Z}}$, llamada el *R-ésimo código superior de bloques*, de la siguiente manera:

$$\forall x = (x_i)_{i \in \mathbb{Z}} \in X, \quad \beta_R(x)_i = x_{[i,i+R-1]} \quad \forall\, i \in \mathbb{Z}.$$

Por ejemplo, si $R = 3$, entonces

$$\beta_3(x) = \cdots \begin{bmatrix} x_0 \\ x_{-1} \\ x_{-2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_0 \\ x_{-1} \end{bmatrix} . \begin{bmatrix} x_2 \\ x_1 \\ x_0 \end{bmatrix} \cdots .$$

Definición 1.4.2. Sea X un espacio shift. Entonces el ***shift superior de R-bloques de X*** es la imagen $X^{[R]} = \beta_R(X)$.

Observemos que los símbolos en $\mathcal{A}^R$ que son consecutivos en la imagen bajo β_R de un punto de x tienen la propiedad de que se sobreponen ciertos símbolos de A, es decir, se sobreponen progresivamente. Precisamente, diremos que $u = u_1u_2u_3u_4 \cdots u_s$ y $v = v_1v_2v_3v_4 \cdots v_s$ se *sobreponen progresivamente* si $u_2u_3u_4 \cdots u_s = v_1v_2v_3v_4 \cdots v_{s-1}$.

Proposición 1.4.1. *Los shifts superiores de bloques son también espacios shift.*

Demostración. Sea X un espacio shift y $R > 1$. Ahora como X es espacio shift existe una colección de bloques $\wp \subset \mathcal{A}^*$ tal que $X = X_\wp$. Entonces creamos una nueva colección $\tilde{\wp} \subset \mathcal{A}^*$ remplazando cada bloque u con $|u| < R$ por todos los R-bloques en $\mathcal{A}$ que contengan a u. Lo primero que hay que observar es que $X = X_{\tilde{\wp}}$, pues si $x \in X$, entonces $\forall\, w \in \wp$, w no ocurre en x, y en particular cualquier R-bloque que contenga a w no ocurre en x (de lo contrario también w ocurriría en x), por lo cual $x \in X_{\tilde{\wp}}$. Ahora supongamos que $x \in X_{\tilde{\wp}}$,si existiera alguna $w \in \wp$ que ocurre en x, tendríamos $w = x_{[i,j]}$ para algunas $i, j \in \mathbb{Z}$, $i \leq j$. Si $j - i \geq R - 1$, entonces $w \in \tilde{\wp}$, una contradicción pues w ocurre en $x \in X_{\tilde{\wp}}$. Si $j - i < R - 1$, entonces $x_{[i,j]}x_{[j+1,i+R-1]} = x_{[i,i+R-1]} \in \tilde{\wp}$, lo cual es una contradicción. Concluimos entonces que $x \in X_\wp = X$. Entonces $X = X_\wp = X_{\tilde{\wp}}$.

Ahora, cada bloque en $\tilde{\wp}$ tiene longitud mayor o igual que R. Para cada bloque $w = a_1a_2...a_m \in \tilde{\wp}$ (con $m \geq R$), tenemos

$$w^{[R]} = \begin{bmatrix} a_R \\ \vdots \\ a_2 \\ a_1 \end{bmatrix} \begin{bmatrix} a_{R+1} \\ \vdots \\ a_3 \\ a_2 \end{bmatrix} \cdots \begin{bmatrix} a_m \\ \vdots \\ a_{m-R+2} \\ a_{m-R+1} \end{bmatrix}$$

que es un $m-R+1$-bloque de $(\mathcal{A}^{[R]})^*$. Ahora definimos $\wp_1 \subset (\mathcal{A}^{[R]})^*$ como todos los bloques de la forma $w^{[R]}$ p.a. $w \in \tilde{\wp}$. Si $x \in X^{[R]}$, entonces es de la forma

$$x = \beta_R(z) = \cdots \begin{bmatrix} z_{-3+R-1} \\ z_{-3+R-2} \\ \vdots \\ z_{-3+R-R} \end{bmatrix} \begin{bmatrix} z_{-2+R-1} \\ z_{-2+R-2} \\ \vdots \\ z_{-2+R-R} \end{bmatrix} \begin{bmatrix} z_{-1+R-1} \\ z_{-1+R-2} \\ \vdots \\ z_{-1+R-R} \end{bmatrix} . \begin{bmatrix} z_{0+R-1} \\ z_{0+R-2} \\ \vdots \\ z_{0+R-R} \end{bmatrix} \begin{bmatrix} z_{1+R-1} \\ z_{1+R-2} \\ \vdots \\ z_{1+R-R} \end{bmatrix} \begin{bmatrix} z_{2+R-1} \\ z_{2+R-2} \\ \vdots \\ z_{2+R-R} \end{bmatrix} \cdots$$

para alguna $z \in X$, y ya que ningún bloque de $\tilde{\wp}$ puede ocurrir en z, tampoco bloques de $\wp_1$ pueden ocurrir en x, pues en caso contrario tendríamos que algún $w^{[R]}$ ocurre en x lo que quiere decir que

$$\begin{aligned} w^{[R]} &= \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_R \end{bmatrix} \begin{bmatrix} a_2 \\ a_3 \\ \vdots \\ a_{R+1} \end{bmatrix} \cdots \begin{bmatrix} a_{m-R+1} \\ a_{m-R+2} \\ \vdots \\ a_m \end{bmatrix} \\ &= x_{[i,j]} \\ &= \beta_R(z)_{[i,j]} \\ &= \begin{bmatrix} z_{i+R-1} \\ z_{i+R-2} \\ \vdots \\ z_{i+R-R} \end{bmatrix} \begin{bmatrix} z_{i+1+R-1} \\ z_{i+1+R-2} \\ \vdots \\ z_{i+1+R-R} \end{bmatrix} \cdots \begin{bmatrix} z_{j-1+R-1} \\ z_{j-1+R-2} \\ \vdots \\ z_{j-1+R-R} \end{bmatrix} \begin{bmatrix} z_{j+R-1} \\ z_{j+R-2} \\ \vdots \\ z_{j+R-R} \end{bmatrix} \end{aligned}$$

y lo anterior implica que

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_R \end{bmatrix} = \begin{bmatrix} z_{i+R-1} \\ z_{i+R-2} \\ \vdots \\ z_{i+R-R} \end{bmatrix}, \begin{bmatrix} a_2 \\ a_3 \\ \vdots \\ a_{R+1} \end{bmatrix} = \begin{bmatrix} z_{i+1+R-1} \\ z_{i+1+R-2} \\ \vdots \\ z_{i+1+R-R} \end{bmatrix}, \cdots, \begin{bmatrix} a_{m-R+1} \\ a_{m-R+2} \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} z_{j+R-1} \\ z_{j+R-2} \\ \vdots \\ z_{j+R-R} \end{bmatrix},$$

lo cual nos dice que $w = a_1a_2...a_m = z_{[i,j]}$ y ésto es una contradicción pues en z no ocurren elementos de $\tilde{\wp}$. Entonces $X^{[R]} \subset X_{\wp_1}$ y definimos

$$\wp_2 = \{uv : u, v \in A^{[R]}, u \text{ y } v \text{ no están sobrepuestos progresivamente}\}.$$

Como $X^{[R]}$ satisface la condición de sobreposición, entonces en ninguno de sus elementos pueden ocurrir bloques de $\wp_2$, por lo que tenemos $X^{[R]} \subset X_{\wp_2}$ y también $X^{[R]} \subset X_{\wp_1} \cap X_{\wp_2}$.

Ahora veamos que $X_{\wp_1} \cap X_{\wp_2} = X_{\wp_1 \cup \wp_2}$. Primero tomamos un elemento x en la intersección, de forma que no hay bloques de $\wp_1$ y $\wp_2$ que ocurran en x, lo que quiere decir que $\forall\ w \in \wp_1$ y $\forall\ w \in \wp_2$, w no ocurre en x, y por lo tanto $x \in X_{\wp_1 \cup \wp_2}$. Por otro lado, si $x \in X_{\wp_1 \cup \wp_2}$, entonces ninguún bloque de $\wp_1$ ni de $\wp_2$ ocurre en x, y por lo tanto $X_{\wp_1} \cap X_{\wp_2}$. Entonces a lo que hemos llegado hasta ahora es a que $X^{[R]} \subset X_{\wp_1 \cup \wp_2}$. Ahora si $y \in X_{\wp_1} \cap X_{\wp_2}$, entonces y cumple que sus elementos se sobreponen progresivamente pues los bloques que no cumplen ésto no ocurren en y ya que $y \in X_{\wp_2}$. Definimos entonces x_i como el símbolo de índice mas pequeño del i-ésimo símbolo de

y y notamos que $\beta_R(x) = y$. Ahora, como anteriormente se probó que en y no ocurren los bloques que estén en $\wp_1$, los bloques de $\tilde{\wp}$ no ocurren en x, y por lo tanto $x \in X_{\tilde{\wp}} = X$, lo que nos dice que $y \in \beta_R(X) = X^{[R]}$ por lo tanto $X^{[R]}$ es un espacio shift. $\square$

Dentro de esta demostración se probó un resultado técnico importante que es el siguiente:

Proposición 1.4.2. *Sea $X \subset \mathcal{A}^{\mathbb{Z}}$ un espacio shift y sea $M \in \mathbb{N}$. Entonces existe $\wp \subset B(X)$ tal que $\forall\, w \in \wp$, $|w| \geq M$ y $X = X_{\wp}$.*

La demostración de este resultado es análogo al anterior.

Definición 1.4.3. Sean $X \subset \mathcal{A}^{\mathbb{Z}}$ un espacio shift, $R \geq 2$ un entero y $\mathcal{A}_X^{[R]} = B_R(X)$. Definimos el *R-ésimo código potencia de bloque* como la función $\gamma_R : X \to (\mathcal{A}_X^{[R]})^{\mathbb{Z}}$ definida por

$$\gamma_R(x)_{[i]} = x_{[iR,iR+R-1]}.$$

Por ejemplo, para $R = 3$,

$$\gamma_3(x) = \cdots \begin{bmatrix} x_{-4} \\ x_{-5} \\ x_{-6} \end{bmatrix} \begin{bmatrix} x_{-1} \\ x_{-2} \\ x_{-3} \end{bmatrix} \begin{bmatrix} x_{2} \\ x_{1} \\ x_{0} \end{bmatrix} \cdots .$$

Definición 1.4.4. Sea X un espacio shift y $R \geq 1$ un entero. Entonces el *shift potencia de R-bloques* (o la *presentación de X en potencias de R-bloques*) es la imagen $X^R = \gamma_R(X)$.

La siguiente proposición es equivalente a la Proposición 1.4.1 y la demostración es similar.

Proposición 1.4.3. *Los shifts de potencias de bloques son también espacios shift.*

1.5. Códigos de bloques deslizantes

Los códigos de bloques deslizantes (o simplemente códigos de bloque) son las funciones entre espacios shift que preservan la dinámica del shift y que son continuas, con la topología inducida como subconjunto del producto topológico indexado por $\mathbb{Z}$ del alfabeto como espacio topológico discreto.

Definición 1.5.1. Sea X un espacio shift sobre el alfabeto $\mathcal{A}$ y Y un espacio shift sobre el alfabeto $\mathcal{B}$ y sea

$$\Phi : B_{m+a+1}(X) \longrightarrow \mathcal{B}$$

una funcion de bloques, con $m, a \geq 0$ enteros. Definimos $\phi\colon X \to Y$ para cada $x \in X$ por la regla

$$\phi(x)_i = \Phi(x_{[i-m,i+n]}) \quad \forall\, i \in \mathbb{Z}.$$

A ϕ le llamaremos un *código de bloques deslizante*, o también un *código de* $(m + a + 1)$*-bloques*, de X a Y, con *memoria* m y *anticipación* a, y diremos que está *inducida por la regla local* Φ, en cuyo caso escribiremos

$$\phi = \Phi_{\infty}{}^{[-m,a]}.$$

Cuando la memoria y la anticipación son cero se le llama *1-código bloque* (que por supuesto es una manera mas compacta de decir código de 1 bloque) puesto que la i-ésima entrada del código sólo depende de la i-ésima entrada del dominio.

Proposición 1.5.1. *Sean X y Y dos espacios shift, $\phi\colon X \to Y$ un código de bloques deslizante (Sliding Block Code). Denotaremos la función shift para X como T_X análogo para Y. Entonces*

$$T_Y \circ \phi = \phi \circ T_X.$$

Demostración. Sea ϕ inducida por una regla local $\Phi : B_{m+a+1}(X) \longrightarrow \mathcal{B}$ son memoria m y anticipación a con $\mathcal{B}$ alfabeto de Y. Para $x \in X$ tenemos que

$$\phi(T_X(x))_i = \Phi(T_X(x)_{[i-m,i+a]}) = \Phi((x)_{[i-m+1,i+a+1]})$$

mientras que

$$T_Y(\phi(x))_i = \phi(x)_{i+1} = \Phi((x)_{[i-m+1,i+a+1]}) = \phi(T_X(x))_i$$

por lo tanto $T_Y \circ \phi = \phi \circ T_X$. □

Proposición 1.5.2. *Sean X, Y, W espacios shift y sean $\phi\colon X \to Y$ y $\psi\colon Y \to W$ códigos de bloques. Entonces $\psi \circ \phi$ tambien es un código de bloques.*

Demostración. Sean Φ y Ψ las reglas locales que inducen los respectivos códigos de memoria m y n y de anticipación a y b respectivamente, es decir $\phi(x)_i = \Phi(x_{[i-m,i+a]})$ y $\psi(y)_i = \Psi(y_{[i-n,i+b]})$ para toda $i \in \mathbb{Z}$. Ahora definimos

$$\overline{\psi \circ \phi} : B_{m+a+n+b+1}(X) \longrightarrow B_1(W)$$

para todo $a_1a_2a_3 \ldots a_{m+a+n+b+1} \in B_{m+a+n+b+1}(X)$ por

$$\overline{(\psi \circ \phi)}(a_1a_2a_3 \cdots a_{m+a+n+b+1})$$
$$= \Psi(\Phi(a_1a_2 \cdots a_{m+a+1})\Phi(a_2a_3 \cdots a_{m+a+2}) \cdots \Phi(a_{b+n+1}a_{b+n+2} \cdots a_{m+a+b+n+1})).$$

Ahora sólo falta ver que $(\overline{\psi \circ \phi})^{\infty} = \psi \circ \phi$, para lo cual tenemos que

$$\psi \circ \phi(x)_i = \Psi(\phi(x)_{[i-n,i+b]}) = \Psi(\phi(x)_{i-n}\phi(x)_{i-n+1}\phi(x)_{i-n+2} \cdots \phi(x)_{i+b+1}) =$$
$$= \Psi(\Phi(x_{[i-n-m,i-n+a]})\Phi(x_{[i-n-m+1,i-n+a+1]})\Phi(x_{[i-n-m+2,i-n+a+2]}) \cdots \Phi(x_{[i+b+1,i+b+a+1]})) =$$
$$= \overline{(\psi \circ \phi)}(x_{[i-n-m,i+b+a+1]}),$$

lo que termina la prueba. □

Proposición 1.5.3. *Sean X y Y dos espacios shift. Una función $\phi\colon X \to Y$ es un código de bloques deslizante (Sliding Block Code) si y sólo si $T_Y \circ \phi = \phi \circ T_X$ y existe $R \geq 0$ en los naturales tal que $\phi(x)_0$ es una funcioń de $x_{[-R,R]}$ para toda $x \in X$.*

Demostración. La necesidad proviene de la Definición 1.5.1 y la Proposición 1.5.1. Para la suficiencia, sea R el número entero que hace que $\phi(x)_0$ sea función de $x_{[-R,R]}$. Definimos una función de $(2R+1)$-bloques como sigue: para cada $w \in B_{2R+1}(X)$, existe $x \in X$ tal que w ocurre en x, entonces $w = x_{[i,i+2R]}$, y si $i \neq -R$, tomamos $T^{i+R}(x) \in X$ de forma que podemos suponer que $w = x_{[-R,R]}$. Entonces definimos $\Phi(w) = \phi(x)_0$ puesto que por hipótesis es función de $x_{[-R,R]}$. Así, basta probar que $\phi = \Phi_\infty$. Tenemos que para toda $x \in X$,

$$\phi(x)_i = T_Y^i(\phi(x))_0 = \phi(T_X^i(x))_0 = \Phi(x_{[i-R,i+R]})$$

y por lo tanto $\phi = \Phi_\infty^{[-R,R]}$, lo que demuestra el teorema. □

Definición 1.5.2. Un código de bloques deslizante $\phi\colon X \to Y$ es una *conjugación* de X a Y si ϕ es invertible y la inversa $\phi^{-1}\colon Y \to X$ es un código de bloques. En este caso diremos que X y Y son *conjugados*.

Ya hemos visto ejemplo de conjugaciones, como lo muestra el siguiente resultado.

Proposición 1.5.4. *Sea X un espacio shift sobre $\mathcal{A}$ y sea $R \geq 1$. Entonces las funciones que definen el shift superior y las potencias de bloques $\beta_R\colon X \to X^{[R]}$ y $\gamma_R\colon X \to X^R$ son conjugaciones.*

Demostración. Veremos el argumento únicamente para β_R. Es claro que la regla local $\Phi\colon B_R(X) \to B_1(X)$ definida para toda $w = w_1 \dots w_R \in B_R(X)$ por $\Phi(w) = w_1$ es tal que $\phi = \Phi_\infty\colon X^{[R]} \to X$ es biyectiva y $\phi = \beta_R^{-1}$. □

Proposición 1.5.5. *Sea $\phi\colon X \to Y$ código de bloques deslizante. Si $x \in X$ tiene periodo n bajo T_X, entonces $\phi(x)$ tiene periodo n bajo T_Y, y el periodo mínimo de $\phi(x)$ divide al periodo mínimo de x. En virtud de ésto, el periodo mínimo de un punto se conserva bajo conjugaciones.*

Demostración. Primero, como x tiene periodo n, $T_X^n(x) = x$. Veamos qué ocurre con $\phi(x)$. Tenemos que

$$T_Y^n(\phi(x)) = \phi(T_X^n(x)) = \phi(x)$$

porque el shift y los códigos de bloques conmutan. Entonces también $\phi(x)$ tiene periodo n. Ahora sea m el mínimo entero positivo tal que ${T_Y}^m(\phi(x)) = \phi(x)$. Ocupando el algoritmo de la división vemos que $n = mt + r$ para alguna $r < m$, pero entonces

$$\phi(x) = T_Y^n(\phi(x)) = T_Y^{mt+r}(\phi(x)) = \phi(T_X^{mt+r}(x)) = \phi(T_X^r(x))$$

como m el mínimo periodo de $\phi(x)$, $T_X^r(x) = x$,de lo contrario $\phi(x) = \phi(T_X^r(x)) = T_Y^r(\phi(x))$ pero $r < m$ y m es el periodo mínimo de x, con lo cual $r = 0$, es decir, $m|n$, en particular, el periodo mínimo de $\phi(x)$ divide al periodo mínimo de x. El resultado se sigue entonces de un argumento similar para ϕ^{-1}, pues tendríamos que el periodo mínimo de x divide al periodo mínimo de $\phi(x)$. □

Proposición 1.5.6. *Sea $\phi\colon X \to Y$ un código de bloques deslizante. Entonces existe un shift superior de bloques $\tilde{X}$, una conjugación $\psi\colon X \to \tilde{X}$, y un código de 1-bloque $\tilde{\phi}\colon \tilde{X} \to Y$ tal que $\tilde{\phi} \circ \psi = \phi$.*

Demostración. Supongamos que ϕ es inducida por una función de bloques Φ de memoria m y anticipación a.Definimos entonces $\tilde{X} = X^{[m+a+1]} \subset \mathcal{A}'^{\mathbb{Z}}$, donde $\mathcal{A}' = B_{m+a+1}(X)$, y a la conjugación $\psi\colon X \to \tilde{X}$ por

$$\psi(x) = (x_{[i-m,i+a]})_{i \in \mathbb{Z}} \quad \forall\, x \in X.$$

La Proposición 1.4.1 implica que $\psi(X) = T^{-m} \circ \beta_{m+a+1}(X) = X^{[m+a+1]}$ es un espacio shift. Además ψ^{-1} es un código de 1-bloques, con regla local $\tilde{\Psi}\colon B_1(\tilde{X})\mathcal{A}' \to B_1(X)$ dada por

$$x_{-m} \dots x_a \mapsto x_0.$$

Ahora veamos que si hacemos $\tilde{\phi} = \phi \circ \psi^{-1}$, entonces $\tilde{\phi}$ es un código de 1-bloques. Para toda $x \in \tilde{X}$, tenemos que

$$\tilde{\phi}(x)_i = \phi \circ \psi^{-1}(x)_i = \phi(\psi^{-1}(x))_i = \Phi((\psi^{-1}(x))_{[i-m,i+a]}) = \Phi(\psi(\psi^{-1}(x))_i) = \Phi((x)_i)$$

para toda $i \in \mathbb{Z}$. □

Teorema 1.5.1. *La imagen de un espacio shift bajo un código de bloques deslizante es un espacio shift.*

Demostración. Sean X y Y dos espacios shift y sea $\phi\colon X \to Y$ un código de bloques deslizante. Por la proposición anterior se puede suponer que es un código de 1-bloques. Sea $\Phi\colon B_1(X) \to B_1(Y)$ la función de bloques, o regla local, que induce a ϕ. Sea

$$L = \{\Phi(w) : w \in B(X)\}$$

el lenguaje de la imagen, desde luego al hablar de la imagen de una palabra de longitud mayor a 1 nos referimos a la concatenacion de las imagenes de los símbolos de la palabra. Probaremos que $\phi(X) = X_{L^c}$, lo que terminaría la prueba. Sea $y = \phi(x) \in \phi(X)$, con $x \in X$. Entonces cada bloque de $\phi(x)$ está en L porque todos los bloques que ocurren en x pertenecen a $B(X)$. Por lo tanto $\phi(X) \subset X_{L^c}$ (de lo contrario algún bloque fuera de L tendría que ocurrir en y lo cual es falso por la afirmación hecha).

Para la otra contención, si $y \in X_{L^c}$ entonces en y no ocurren bloques fuera de L (es decir en su complemento), lo que se dice igual que cualquier bloque que ocurra en y está en L, lo que quiere decir que para toda $n \in N$ tenemos que $y_{[-n,n]}$ está en L, y así $y_{[-n,n]} = \Phi(w)$ donde $w \in B_{2n+1}(X)$ ocurre en algún punto $x = x^{(n)} \in X$, o más precisamente, $w = x_{[-n,n]}$. Tenemos entonces una sucesión de puntos $(x^{(n)})_{n=0}^{\infty} \subset X$. Consideremos la coordenada cero de cada uno de los $x^{(n)}$'s. Como el número total de símbolos de X es finito, debe haber una cantidad infinita de números n tales que $x_0^{(n)}$ es el mismo para esas n's, y llamemos a ese conjunto S_0, es decir, $S_0 \subset \mathbb{N}$ es tal que $|S_0| = \infty$ y para toda $n, m \in S_0$, $x_0^{(n)} = x_0^{(m)}$. En forma similar, como $|B_3(X)| < \infty$, existe un conjunto infinito $S_1 \subset S_0$ tal que $x_{[-1,1]}^{(n)}$ es constante para toda $n \in S_1$. Inductivamente, para cada $k \geq 1$ existe un conjunto infinito $S_k \subset S_{k-1}$ tal que $x_{[-k,k]}^{(n)}$ es constante para toda $n \in S_k$. Definimos a $x \in X$ como el elemento que satisface $x_{[-k,k]} = x_{[-k,k]}^{(n)}$ con $n \in S_k$ para toda $k \in \mathbb{N}$. Observemos que x está bien definido porque $S_k \subset S_{k-1}$ para toda $k \in \mathbb{N}$, y como

$$\Phi(x_{[-k,k]}) = \Phi(x_{[-k,k]}^{(n)}) = \phi(x^{(n)})_{[-k,k]} = y_{[-k,k]},$$

entonces $y = \phi(x)$, lo cual demuestra que $X_{L^c} \subset \phi(X)$, lo cual termina la demostración del teorema. □

Teorema 1.5.2. *Sean X y Y dos espacios shift y sea $\phi\colon X \to Y$ un código de bloques deslizante. Si ϕ es biyectivo, entonces naturalmente tiene inversa y dicha inversa es un código de bloques y entonces ϕ es de hecho una conjugación.*

Demostración. Por la Proposición 1.5.6, existe un shift superior de bloques $\tilde{X}$, una conjugación $\psi\colon X \to \tilde{X}$, y un código de 1-bloques $\tilde{\phi}\colon \tilde{X} \to Y$ tal que $\tilde{\phi} \circ \psi = \phi$. Trabajaremos con $\tilde{\phi}$ que es un código de 1-bloques. Como $\phi \circ \psi^{-1} = \tilde{\phi}$, entonces $\tilde{\phi}$ es una composición de dos funciones biyectivas y por lo tanto también es una función biyectiva. Como también es un código de 1-bloques, tenemos que $\tilde{\phi}(x) = (\tilde{\Phi}(x_i))_{i \in Z}$, donde $\tilde{\Phi}$ toma valores en el alfabeto de Y. Por lo tanto $\tilde{\Phi}$ es también una función biyectiva. Sea $\tilde{\Phi}^{-1}$ la inversa de $\tilde{\Phi}$. Bastará que probemos que $\tilde{\phi}^{-1} = \tilde{\Phi}_{\infty}^{-1}$. Tomamos el código de bloques definido por dicha funcion inversa, es decir definimos $\xi\colon Y \to \tilde{X}$ por $\xi(y)_i = \tilde{\Phi}^{-1}(y_i)$ para todas $y \in Y$ e $i \in \mathbb{Z}$. Veamos que $\xi \circ \tilde{\phi}(x) = x$ para toda $x \in X$ y que

$\tilde{\phi} \circ \xi(y) = y$ para toda $y \in Y$, para lo cual veremos en ambos casos que son iguales coordenada a coordenada. Para toda $i \in \mathbb{Z}$, primero tenemos que

$$\xi \circ \tilde{\phi}(x)_i = \xi(\tilde{\phi}(x))_i = \tilde{\Phi}^{-1}(\tilde{\phi}(x)_i) = \tilde{\Phi}^{-1}(\tilde{\Phi}(x_i)) = \tilde{\Phi}^{-1} \circ \tilde{\Phi}(x_i) = id(x_i) = x_i$$

y luego tenemos que

$$\tilde{\phi} \circ \xi(y)_i = \tilde{\phi}(\xi(y))_i = \tilde{\Phi}(\xi(y)_i) = \tilde{\Phi}(\tilde{\Phi}^{-1}(y_i)) = \tilde{\Phi} \circ \tilde{\Phi}^{-1}(y_i) = id(y_i).$$

Entonces $\xi = \tilde{\phi}^{-1}$, y como ξ es un código de bloques, entonces $\tilde{\phi}^{-1}$ es un código de bloques. Como $\tilde{\phi} \circ \psi = \phi$ entonces por la proposición 1.5.2 $\psi^{-1} \circ \tilde{\phi}^{-1} = \phi^{-1}$ es un código de bloques. $\square$

Capítulo 2

SHIFTS DE TIPO FINITO

2.1. Restricciones de tipo finito

Definición 2.1.1. Se define un shift de tipo finito como un shift que es posible caracterizar por una cantidad finita de bloques prohibidos, es decir $X = X_F \subset \mathcal{A}^{\mathbb{Z}}$ con $F \subset \mathcal{A}^*$ finito.

Definición 2.1.2. Un shift de tipo finito tiene *memoria* M si puede ser descrito por una colección de $(M+1)$-bloques prohibidos.

Proposición 2.1.1. *Si X es un shift de tipo finito, entonces existe $M > 0$ tal que X tiene memoria M.*

Demostración. Es una consecuencia directa de la Proposición 1.4.2. □

Proposición 2.1.2. *Un espacio shift X es un shift de tipo finito de memoria M si y sólo si siempre que $uv, vw \in B(X)$ con $|v| \geq M$, $uvw \in B(X)$.*

Demostración. Para la primera implicación supongamos que $X \subset \mathcal{A}^{\mathbb{Z}}$ es un shift de tipo finito de memoria M y sea $F \subset \mathcal{A}^{M+1}$ tal que $X = X_F$. Sean $u, v, w \in B(X)$ tales que $uv, vw \in B(X)$ y $|v| = n \geq M$. Entonces, como X es un espacio shift, existen $x, y \in X$ tales que $x_{[-k,n]} = uv$, $y_{[1,l]} = vw$ con $x_{[1,n]} = y_{[1,n]} = v$. Sea $z = x_{(-\infty,0]}vy_{[n+1,\infty)}$. Ninguún bloque de F ocurre en z porque todos los bloques de F son de longitud $M+1$ y $|v| \geq M+1$, pues si $r \in F$ ocurre en z entonces ocurre en x o en y, una contradicción puesto que $x, y \in X$. Entonces z está en X, por lo que $\forall i, j z_{[i,j]} \in B(X)$, en particular $uvw = x_{[-k,0]}vy_{[n+1,l]} = z_{[-k,l]} \in B(X)$.

Para la segunda implicación supongamos que existe M de tal forma que siempre que $uv, vw \in B(X)$ con $|v| > M$ entonces $uvw \in B(X)$. Sea $F = A^{M+1} - B_{M+1}(X)$, el cual es finito. Probaremos que $X = X_F$ para concluir. Si $x \in X$, entonces los bloques de F no pueden ocurrir en x puesto que los únicos que ocurren en puntos de X de tamaño $M+1$ son los puntos de $B(X)$, lo que significa que $x \in X_F$, y por lo tanto $X \subset X_F$. Ahora para probar la otra contención, sea $x \in X_F$. Entonces $x_{[0,M]}$ y $x_{[1,M+1]}$ son bloques de tamaño $M+1$, y como $x \in X_F$, entonces $x_{[0,M]}, x_{[1,M+1]} \in B_{M+1}(X)$, y además coinciden en un bloque de tamaño mayor o igual que M, lo que por hipóyesses implica que $x_{[0,M+1]} \in B(X)$. Inductivamente veamos que $x_{[0,n]} \in B(X)$ para toda $n \in \mathbb{N}$. Como X es un espacio shift, observamos entonces que para toda $j, k \in \mathbb{Z}$, con $i \leq j$, $x_{[j,k]} \in B(X)$, y por lo tanto $x \in X$ por el Corolario 1.3.1. □

Proposición 2.1.3. *Un espacio shift que es conjugado a un espacio shift de tipo finito también es de tipo finito.*

Demostración. Supongamos que X es un espacio shift que es conjugado a un shift de tipo finito Y. Sea Φ la regla local que induce una conjugación $\phi\colon X \to Y$ y sea Ψ una regla local que induce la correspondiente inversa $\psi\colon Y \to X$. Veremos que X es un espacio shift de tipo finito de memoria M con la caracterización del teorema anterior, viendo que existe $M \in \mathbb{N}$ que cumple que siempre que $uv, vw \in B(X)$ con $|v| \geq M$, $uvw \in B(X)$.

Podemos suponer que ambos códigos tienen la misma memoria y anticipación l (tomando el máximo de la memoria y anticipación de ambos y extendiendo las reglas locales en la forma natural), es decir, tenemos que

$$\Phi\colon B_{l+1}(X) \to B_1(Y), \qquad \phi = \Phi_\infty, \qquad \phi(x)_0 = \Phi(x_{[-l,l]})\ \forall\, x \in X,$$

$$\Psi\colon B_{2l+1}(Y) \to B_1(X), \qquad \psi = \Psi_\infty, \qquad \psi(y)_0 = \Psi(y_{[-l,l]})\ \forall\, y \in Y.$$

La regla local que induce la conjugación $\psi \circ \phi(x) = x$ para toda $x \in X$ es ahora una función de bloques $\overline{\Psi \circ \Phi} : B_{4l+1}(X) \to B_1(X)$. Como Y es un shift de tipo finito, existe $R > 0$ tal que siempre que $uv, vw \in B(Y)$ con $|v| \geq R$, $uvw \in B(Y)$.

Sea $M = R + 4l$ y demostremos que si $uv, vw \in B(X)$ con $|v| \geq M$, entonces $uvw \in B(X)$. Por $1b$ de la Proposición 1.3.1, existen $s, t \in B_{2l}(X)$ tales que $suv, vwt \in B(X)$. Podemos aplicar Φ a estos bloques y obtemos que $\Phi(suv) \in B_{|suv|-2l}(Y)$ y $\Phi(vwt) \in B_{|vwt|-2l}(Y)$, y como $|v| \geq M$, $\Phi(suv)$ y $\Phi(vwt)$ se sobreponen en al menos R símbolos, por lo que $\Phi(suvwt) \in B(Y)$ y de aquí que $uvw \in B(X)$. $\square$

2.2. Gráficas y sus shifts

Los shifts inducidos por gráficas serán de importancia fundamental porque constituirán, módulo una conjugación, al conjunto de shifts de tipo finito.

Comencemos recordando salgo de terminología y notación. Una *gráfica* G consiste de un conjunto de *vértices* $V(G)$ y un conunto de *aristas* $E(G)$ junto con sus *funciones de incidencias inicial* $i\colon E(G) \to V(G)$ y *terminal* $t\colon E(G) \to V(G)$. Si $e \in E(G)$, diremos que e *inicia* en el vértice $i(e)$ y *termina* en $t(e)$. Un *camino* en G es una sucesión de aristas tal que para cualesquiera dos aristas consecutivas del camino, digamos ef, se tiene que $t(e) = i(f)$. Un camino finito $\gamma = e_1 \dots e_n \in E^n$ con $t(e_k) = i(e_{k+1})\ \forall\, k = 1, \dots, n-1$ *inicia* es $v = i(e_1)$ y *termina* en $w = t(e_n)$ y es un (u, v)-camino (o *trayectoria*), y diremos que es de *longitud* n. Asociamos a G la matriz cuadrada $A(G)$ con entradas en $\mathbb{Z}^+$ e indexada por $V(G)$ de forma que para toda $i, j \in V(G)$,

$$A(G)_{i,j} = |\{e \in E(G) \ :\ i(e) = i \text{ y } t(e) = j\}|.$$

Diremos que la gráfica G es *simple* si $A(G)_{i,j} \in \{0, 1\}\ \forall\, i, j \in V(G)$ y que $A(G)$ es la matriz de *adyacencia*.

Definición 2.2.1. Sea G una gráfica con $E = E(G)$ y $A = A(G)$. El *shift de aristas* $X_G = X_A$ es el espacio shift $X_F \subset E^{\mathbb{Z}}$ con alfabeto E y

$$F = \{ef \in E^2 \ :\ t(e) \neq i(f)\},$$

es decir,

$$X_G = X_A = \{e = (e_n)_{n\in\mathbb{Z}} \in E^{\mathbb{Z}} : t(e_n) = i(e_{n+1})\ \ \forall\, n \in \mathbb{Z}\}.$$

Hagamos algunas observaciones.

Observación 2.2.1. *El espacio shift por aristas X_G definido por la gráfica G consiste precisamente de todos los* caminos *doblemente infinitos en la gráfica G.*

Observación 2.2.2. *Sea G una gráfica con matriz de adyacencia $A = A(G)$ entonces el shift de aristas $X_G = X_A$ es un shift de tipo finito de memoria 1.*

Definición 2.2.2. Un vértice *varado* es aquel que no es vértice inicial o no es vértice terminal de alguna arista y una gráfica es *escencial* si no tiene vértices varados.

Observación 2.2.3. *Sea G una gráfica, entonces existe una única subgráfica escencial H de G tal que $X_G = X_H$.*

En virtud de esta última observación, consideraremos de ahora en adelante únicamente gráficas escenciales.

Proposición 2.2.1. *Sea una gráfica G y $A = A(G)$ y sea $m \geq 0$ un entero. Entonces:*

1. *La cantidad de caminos de longitud m de $i \in V(G)$ a $j \in V(G)$ es $(A^m)_{i,j}$.*
2. *El número de caminos cerrados de longitud m es $tr(A^m)$ (donde $tr()$ denota a la función traza, es decir, la suma de los elementos en la diagonal de la matriz), y es el número de puntos en X_G con periodo m.*

Demostración. La primera parte es una demostración propia de teoría de gráficas, es de hecho una simple inducción como sigue: Si $m = 1$ es obviamente verdad que la cantidad de caminos de longitud 1 de i a j es $A_{i,j}$, en el entendido de que la longitud de un camino es el número de aristas que lo conforman. Lo suponemos válido para $m \geq 1$ y con esta hipótesis de inducción lo deducimos para $m + 1$ con la fórmula

$$(A^{m+1})_{i,j} = (A^m A)_{i,j} = \sum_{k\in V(G)} (A^m)_{i,k} A_{k,j}.$$

La afirmación del segundo enunciado es entonces consecuencia directa de su predecesor. □

Definición 2.2.3. Una gráfica G es *fuertemente conexa* si para todo $i, j \in V(G)$, existe un camino en G que inicia en i y termina en j.

Proposición 2.2.2. *Una gráfica G es fuertemente conexa si y sólo si X_G es irreducible.*

Demostración. Supongamos que G es fuertemente conexa y veamos entonces que X_G es irreducible. Sean $u, v \in B(X)$, digamos que $u = u_1 \cdots u_l$ $v = v_1 \cdots v_r$. Entonces $t(u_l)$ y $i(v_1)$ se encuentran en $V(G)$, por lo que existe un $t(u_l), i(v_1)$-camino $D = a_1 \cdots a_r$ con a_i aristas tales que $i(a_1) = t(u_l)$ y $i(v_1) = t(a_r)$. Entonces uDv es un camino en G. Además existen $x, y \in X$ tales que $x_{[1,l]} = u_1 \cdots u_l$ y $y_{[1,r]} = v_1 \cdots v_r$, por lo que $x_{(-\infty,0]} uDv y_{[r+1,\infty)} \in X$ y por lo tanto $uDv \in B(X)$.

Para la segunda implicación tomemos dos vértices $u, v \in V(G)$. La hipótesis de que G es esencial implica que existen $a, b \in E(G)$ tales que $t(a) = u$ e $i(b) = v$, es decir, todo vértice tiene aristas que salen y entran, lo que equivale a que $e \in B_1(X_G)$ para toda $e \in E(G)$. Entonces $a, b \in B_1(X)$ y como X es irreducible, existe $w \in B(X)$ tal que $awb \in B(X)$, y entonces existe un u, v-camino □

Definición 2.2.4. Sea X un espacio shift por aristas definido por la gráfica G con matriz $A = A(G)$. La ***gráfica traspuesta*** G^T es la gráfica que satisface $V(G^T) = V(G)$ y $A(G^T) = A^T$ (donde A^T denota la matriz ***transpuesta*** de A, es decir, $(A^T)_{i,j} = A_{j,i}$ para todo $i, j \in V(G)$), y X_{G^T} es el shift ***transpuesto***.

De cómo está definida G^T se puede ver que esta grfica no es mas que G pero con el sentido de las aristas invertido, lo cual nos dará resultados más intersesantes concernientes al capítulo 4 sección 4.3.

2.3. Representación de shifts de tipo finito con gráficas

Estamos en posición de enunciar un resultado fundamental.

Teorema 2.3.1. *Si X es un shift de tipo finito de memoria M, entonces hay una gráfica G tal que $X^{[M+1]} = X_G$, es decir, X y X_G son conjugados.*

Demostración. Si $M = 0$ entonces X es el shift completo y tomamos a G como un sólo vértice y tomamos una arista por cada símbolo que haya en X y bueno esta gráfica se ve que genera a X y como $M = 0$ $X^{[M+1]} = X$.

Asumimos $M \geq 1$ y entonces definimos el conjunto de vértices de nuestra gráfica como $V(G) = B_M(X)$ y definimos una arista entre $a_1a_2\cdots a_M$ y $b_1b_2\cdots b_M$ si $a_2a_3\cdots a_M = b_1b_2\cdots b_{M-1}$ y $a_1a_2a_3\cdots a_{M-1}b_M \in B(X)$ de lo contrario no hay arista entre ellos. Ahora sólo falta ver que $X_G = X^{[M+1]}$. Sea $x \in X_G$, entonces $x = (a_i)_{i\in\mathbb{Z}}$ donde para cada $i \in \mathbb{Z}$, a_i es una arista de l_i a m_i, donde $l_i, m_i \in B_M(X)$, l_i y m_i se sobreponen y $l_i^1 \dots l_i^M m_i^M \in B_{M+1}(X)$, de forma que a_i puede ser pensado como un bloque de tamaño $M + 1$ en X. La definición de $X^{[M+1]}$ tomando a $X = X_F$ con $F \subset \mathcal{A}^{M+1}$ junto con la definición de G implican entonces que $X^{[M+1]} = X_G$. □

Definición 2.3.1. Sea G una gráfica para $N \geq 2$.Definimos la ***N-ésima gráfica de aristas superior*** $G^{[N]}$ como la gráfica que tiene de vertices a todas trayectorias de longitud $N-1$ en G y 2 trayectorias digamos $e_1e_2\cdots e_{N-1}$,$f_1f_2\cdots f_{N-1}$ tienen una arista entre ellas si $e_2e_3\cdots e_{N-1} = f_1f_2\cdots f_{N-2}$ a dicha arista la llamamos $e_1f_1f_2\cdots f_{N-2}$ o en el caso de aristas (trayectorias de longitud 1) ($N = 2$) tenemos que una empieza donde termina la otra, y por último para $N = 1$ $G^{[1]} = G$.

de aqui es natural preguntar una cosa un tanto obvia que relación hay entre el shift superior de una grafica y el shift de la gráfica superior.

Proposición 2.3.1. *Sea G una gráfica entonces $(X_G)^{[N]} = X_{G^{[N]}}$*

Demostración. Simplemente basta ver que los elementos son de la misma forma, del lado derecho cada coordenada (entrada) de un punto del shift son bloques de tamaño N formados por aristas, cumplen que digamos $e_1e_2\cdots e_{N-1}$,$f_1f_2\cdots f_{N-1}$ tienen una arista entre ellas si $e_2e_3\cdots e_{N-1} = f_1f_2\cdots f_{N-2}$, es decir se sobreponen progresivamente es decir los elementos de dicho shift, son trayectorias de longitud $N + 1$ $e_1f_1f_2\cdots f_{N-1}$, del otro lado tomamos las aristas de la gráfica y vemos el shift superior que son de la misma forma y como en el lado izquierdo $x \in (X_G)^{[N]}$ si y sólo si x_i y x_{i+1} se sobreponen progresivamente es decir $x_i = d_1e_1e_2\cdots e_{N-1}$ $x_{i+1} = e_1f_1f_2\cdots f_{N-1}$ donde las d_i's hacen el papel de la primer bloque de longitud N y analogamente para las e_i's y f_i's de tal forma que esto ocurre si y sólo si cada par de trayectorias de longitud $N + 1$ x_i y x_{i+1} siendo $x_i = e_1e_2\cdots e_N$ y $x_{i+1} = f_1f_2\cdots f_N$ cumplen con $e_2e_3\cdots e_N = f_1f_2\cdots f_{N-1}$ que es lo mismo a que haya una arista en la gráfica superior. □

Definición 2.3.2. Sea B la matriz de adyacencia de una gráfica simple G con r vertices. El *shift de vertices* $\widehat{X}_G = \widehat{X}_B$ es el espacio shift sobre el alfabeto $\mathcal{A} = \{1, 2, 3, ..., r\}$ definido por

$$\widehat{X}_G = \widehat{X}_B = \{x = (x_i)_{i \in \mathbb{Z}} \ : \ B_{x_i x_{i+1}} = 1 \ \forall i \in \mathbb{Z}\}.$$

Proposición 2.3.2. *1. Salvo un cambio de nombre de los símbolos los shift de tipo finito de memoria 1 son los mismos que los shifts de vértices.*

2. *Salvo un cambio de nombre de los símbolos cada shift de aristas es un shift de vértices (de otra gráfica por supuesto).*

3. *Si X es un shift de tipo finito de memoria M, entonces $X^{[M]}$ es un shift de tipo finito de memoria 1, es decir, es un shift de vértices. De hecho hay una gráfica G tal que $X^{[M]} = \widehat{X}_G$ y $X^{[M+1]} = X_G$.*

Demostración. 1. Primero, un shift de vértices definido por una gráfica simple G es un shift de tipo finito de memoria 1 puesto que los bloques prohibidos serían $V(G)^2 \setminus E(G)$, es decir, todas aquellas aristas que no pertenecen a la gráfica. Para ver lo reciproco, si $X \subset A^{\mathbb{Z}}$ es un shift de tipo finito de memoria 1, podemos tomar a la gráfica simple G con $V(G) = B_1(X) = \mathcal{A}$ y $E(G) = V(G)^2 \setminus F$ donde $F \subset \mathcal{A}^2$ es tal que $X = X_F$. Entonces $X = \widehat{X}_G$.

2. Para el segundo punto basta considerar la "gráfica de líneas". Concretamente, dada G una gráfica, su ***gráfica de líneas*** es la gráfica simple $L(G)$ que tiene como vértices a las aristas de G y como aristas al conjunto de 2-bloques de X_G, es decir, $V(L(G)) = E(G)$, y $E(L(G)) = B_2(X_G)$. Es directo entonces probar que X_G es conjugado a $\widehat{X}_{L(G)}$.

3. Para el tercero punto primero nos podemos dar cuenta de que cada $M+1$-bloque en X puede ser visto como un bloque de tamaño dos en $X^{[M]}$, lo que implica que $X^{[M]}$ es un shift de tipo finito de memoria 1, y por lo tanto un shift por vértices (por el inciso 1). Para la segunda afirmación, vemos que G construida como en el Teorema 2.3.1 es simple (porque $M + 1 \geq 1$) y cumple que $X^{[M+1]} = X_G$ y $X^{[M]} = \widehat{X}_G$. □

Definición 2.3.3. Sea $N \geq 1$. Definimos la ***N-ésima gráfica superior de aristas*** $G^{[N]}$ por $V(G^{[N]}) = B_{N-1}(X_G)$ y para cada de vérticese$_1, \dots e_{N-1}, f_1 \dots f_{N-1} \ \in \ V(G^{[N]}$, una arista entre ellos si $e_2, \dots e_{N-1} = f_1 \dots f_{N-2}$.

Definición 2.3.4. Sea $N \geq 1$. Definimos la ***N-ésima potencia*** G^N de la gráfica G como la gráfica con vertices $V(G^N) = V(G)$ y una arista entre $i \in V(G)$ y $j \in V(G)$ por cada camino de longitud N en G de i a j.

Proposición 2.3.3. *Sea una gráfica G con $A = A(G)$. Entonces*

$$A(G^N) = (A_G)^N$$

y además $(X_G)^N = X_{G^N}$.

Demostración. La primera parte es muy sencilla pues de hecho es consecuencia de un resultado de teoria de gráficas como en la Proposición 2.2.1, parte 1. Para la segunda parte, observamos que $(X_G)^N$ y X_{G^N} tienen ambos el mismo alfabeto que son todas las trayectorias de longitud N y ambos cumplen la condición de que $x = x_i \in (X_G)^N$ si y sólo si x_i termina donde x_{i+1} empieza $\forall i \in \mathbb{Z}$ y esto pasa si y sólo si $x = x_i \in X_{G^N}$ lo que concluye con la prueba. □

2.4. División de estados

Definición 2.4.1. Sea G una gráfica y fijémonos ahora para vértice $I \in V(G)$ en el conjunto de aristas que salen del estado I hacia cualquier otro estado, denotado a dicho conjunto por $\mathcal{E}_I$. Ahora tomamos una partición finita de este conjunto, digamos $\mathcal{E}_I^1, \ldots, \mathcal{E}_I^{m(I)}$ y nos fijamos en la particion $\mathcal{P}$ resultante del conjunto total de aristas inducida por cada una de estas particiones pata cada $l \in V(G)$. La *gráfica de division de estados* (de *salida*) $G^{[\mathcal{P}]}$ formada por G usando la partición $\mathcal{P}$ tiene como estados o vértices $I^1, \ldots, I^{m(I)}$ con $I \in V(G)$. Para definir al conjunto de aristas, por cada $e \in A(G)$ con $i(e) = I$ y $t(e) = J$, existe j con $1 \leq j \leq m(i(e))$, de forma que $e \in \mathcal{E}_I^j$, y entonces ponemos una arista e^k de I^j a J^k para cada $k = 1, \ldots, m(J)$.

Podemos dar una definición similar, pero ahora considerando particiones de conjuntos de aristas que llegan a un vértice dado.

Definición 2.4.2. Sea G una gráfica supongamos que $\mathcal{P}$ es una partición de las aristas que resulta de considerar para cada vértice $J \in V(G)$ una partición del conjunto de aristas que llegan al estado J, denotado por $\mathcal{E}^J$, digamos que $\mathcal{E}_1^J, \ldots, \mathcal{E}_{m(J)}^J$. La *gráfica de división de estados* (de *llegada*) es la gráfica $G_{[\mathcal{P}]}$ cuyo conjunto de vértices está formado por $I_1, \ldots, I_{m(I)}$ para cada $I \in V(G)$, y cuyas aristas se construyen a partir de cada arista $e \in A(G)$, con $i(e) = I$ y $t(e) = J$, siendo que existe $j \in \{1, \ldots, m(J)\}$ tal que $e \in \mathcal{E}_j^J$, y entonces ponemos una arista e_i de I_i a J_j para cada $i \in \{1, \ldots, m(I)\}$.

Definición 2.4.3. Una gráfica H es una *escisión* (de *salida*) de una gráfica G, y G es una *amalgamación* (de *salida*) de H, si existe una partición $\mathcal{P}$ de las aristas formada por particiones de cada conjunto de aristas de salida de cada vértice de G de forma tal que $H = G^{[\mathcal{P}]}$. Definimos escisiones y amalgamaciones de *llegada* en forma similar. En cualquier caso, en general diremos que H es una escisión de G y que G es una amalgamación de H.

Teorema 2.4.1. *Si H es una escisión de G, entonces los espacios shifts X_G y X_H son conjugados.*

Un pequeño comentario es importante por hacer es que la escisión siendo un escisión (de salida/ de entrada) entonces a la conjugación le llamaremos código escisión (de salida/ de entrada) o código amalgamación (de salida/ de entrada).

Demostración. Bastará probar un sólo caso, digamos que el de salida, de forma que suponemos que $H = G^{[\mathcal{P}]}$ para $\mathcal{P}$ como en la definición de escisiones y amalgamaciones de salida. Definimos una regla local que es una función de 1-bloques, $\Psi\colon \mathcal{B}_1(X_H) \to \mathcal{B}_1(X_G)$, como $\Psi(e^j) = e$ para toda $e^j \in E(X_H)$ (con $1 \leq j \leq m(i(e))$). Observemos que $e^j f^k \in \mathcal{B}_2(X_H)$ implica que $ef \in \mathcal{B}_2(X_G)$ ya que e^j es una arista que va de I^i a J^k para toda $k \in \{1, \ldots, m(J)\}$ y f^k una arista que sale de J^j para alguna $k \in \{1, \ldots, m(J)\}$. Por lo tanto la imagen de cualquier trayectoria en H bajo Ψ es una trayectoria en G, y siendo así, podemos definir $\psi = \Psi_\infty$ entonces $\psi(X_H) \subseteq X_G$ y con esto tenemos que $\psi\colon X_H \to X_G$ es un código de 1-bloques.

Ahora definimos una regla local que es una función de 2-bloques $\Phi\colon \mathcal{B}_2(X_G) \to \mathcal{B}_1(X_H)$ de la siguiente manera. Si $fe \in \mathcal{B}_2(X_G)$, entonces si $i(e) = I$ y $t(e) = J$, entonces e se encuentra en un único elemento de la partición, digamos en $\mathcal{E}_J^j$ para alguna $j \in \{1, \ldots, m(J)\}$, y entonces definimos $\Phi(fe) = f^j$. Como anteriormente, la imagen de una trayectoria en G bajo Φ es una trayectoria en H. Definimos $\phi = \Phi_\infty$ y tenemos entonces que $\phi(X_G) \subseteq X_H$ y así tenemos un código de bloques $\phi\colon X_G \to X_H$. De hecho Ψ y ϕ son ambas funciones inversas una de otra pues si

$x = \cdots e_{-1}e_0e_1 \cdots \in X_G$, entonces $\phi(x) = \cdots e_{-1}^{j_{-1}}e_0^{j_0}e_1^{j_1}\cdots$ por lo que $\psi(\phi(x)) = x$, y de manera similar si $y = \cdots e_{-1}^{j_{-1}}e_0^{j_0}e_1^{j_1}\cdots$, entonces $\psi(y) = \cdots e_{-1}e_0e_1 \cdots \in X_G$ y como $e_i^{j_i}e_{i+1}^{j_{i+1}} \in \mathcal{B}_2(X_H)$, $e_{i+1} \in \mathcal{E}_{t(e_i)}^{j_i}$, por lo que $\Phi(e_ie_{i+1}) = e_i^{j_i}$ $\forall i \in \mathbb{Z}$ lo cual nos termina por demostrar que $\phi(\psi(y)) = y$ lo que demuestra por fin que X_G y X_H son conjugados. □

Definición 2.4.4. Sea G una gráfica y $H = G^{[\mathcal{P}]}$ una escición de salida usando una partición $\mathcal{P}$. Sean $V = V(G)$ y $W = V(H)$. Definimos la *matriz de división* D como la la matriz de tamaño $|V| \times |W|$ y cuyas entradas están dadas para cada $I \in V$ y $J^k \in W$ por

$$D(I, J^k) = \begin{cases} 1 & \text{si } I = J \\ 0 & \text{en otro cas.} \end{cases}$$

y definimos también la *matriz de aristas* E como la matriz de tamaño $|W| \times |V|$ y cuyas entradas están dadas para cada $I^k \in W$ y $J \in V$ por

$$E(I^k, J) = |\mathcal{E}_I^k \cap \mathcal{E}^J|.$$

Teorema 2.4.2. *Sea G una gráfica y $H = G^{[\mathcal{P}]}$ la gráfica de escisión de estados (de salida) que resulta de G usando la partición $\mathcal{P}$. Sean D la matriz de división y E la matriz de aristas correspondientes. Entonces*

$$A_G = DE \qquad y \qquad A_H = ED. \tag{2.1}$$

Demostración. Usando la definición de ambas matrices vemos que

$$DE(I,J) = \sum_{k \in W} D(I,k)E(k,J) = \sum_{i=1}^{m(I)} D(I,I^i)E(I^i,J) = \sum_{i=1}^{m(I)} \left|\mathcal{E}_I^i \cap \mathcal{E}^J\right| =$$

$$= \left|\bigcup_{i=1}^{m(I)} (\mathcal{E}_I^i \cap \mathcal{E}^J)\right| = \left|\left(\bigcup_{i=1}^{m(I)} \mathcal{E}_I^i\right) \cap \mathcal{E}^J\right| = \left|\mathcal{E}_I \cap \mathcal{E}^J\right| = A_G(I,J),$$

con lo cual se tiene que $DE = A_G$. Para la otra igualdad tenemos

$$ED(I^i, J^j) = \sum_{k \in V} E(I^i, k)D(k, J^j) = E(I^i, J)D(J, J^j) =$$

$$= E(I^i, J) = \left|\mathcal{E}_I^i \cap \mathcal{E}^J\right| = \left|\mathcal{E}_{I^i} \cap \mathcal{E}^{J^j}\right| = A_H(I^i, J^j),$$

lo cual nos permite concluir que $ED = A_H$. □

La matriz de división es importante porque nos proporciona un medio para amalgamar, como lo muestra el Teorema 2.4.3. Primero definamos las matrices de división en abstracto.

Definición 2.4.5. Una *matriz de división* D es una matriz de ceros y unos que en cada columna tiene exactamente un 1 y en cada renglón tiene al menos un 1.

Teorema 2.4.3. *Sea G y H dos gráficas (escenciales). Entonces H es una escisión (de salida) de G si y sólo si existe D una matriz de división y E una matriz de entradas enteras no-negativas de forma que las ecuaciones (2.1) se satisfacen.*

Demostración. Sólo falta probar el regreso por que la ida es el teorema anterior. Partimos entonces de una matriz de división D y una matriz con entradas enteras no-negativas E que satisfacen (2.1). Definimos nuevamente a $V = V(G)$ y $W = V(H)$. Las ecuaciones implican que los tamaños de D y E son $|V| \times |W|$ y $|W| \times |V|$ respectivamente. Hacemos una reetiquetación de los vertices de H de la siguiente forma. Para cada $I \in V$, denotamos por $m(I) \geq 1$ al número de unos en el renglón correspondiente y entonces I^i con $i = 1, \ldots, m(I)$ e $I \in V$ indexan a los vértices de H. Ahora hay que construir una partición. Tenemos que para cada $I \in V$,

$$A_G(I, J) = DE(I, J) = \sum_{k \in W} D(I, k)E(k, J) = \sum_{i=1}^{m(I)} E(I^i, J),$$

de forma que podemos particionar a $\mathcal{E}_I$ en $m(I)$ partes de forma que haya $E(I^i, J)$ aristas en $\mathcal{E}_I^J$ en la parte $i \in \{1, \ldots, m(I)\}$. Si tomamos a $\mathcal{P}$ como la partición de $E(G)$ que resulta al considerar todo $I \in V$, vemos que por definición y por el hecho de que $A_H = ED$, resulta que $H = G^{[\mathcal{P}]}$.

□

Un resultado importante con respecto a escisiones y amalgamaciones de estados es el hecho de que las gráficas superior de aristas puede ser obtenida mediante una sucesión finita de escisiones, lo cual demostraremos más adelante. Primero tenemos el siguiente lema.

Lema 2.4.1. *Sea G una gráfica y $\mathcal{P}$ la partición que resulta de considerar puros singuletes (conjuntos con un solo elemento). Entonces $G^{[\mathcal{P}]}$ es isomorfa a $G^{[2]}$.*

Demostración. Primero notemos que podemos llamar a la partición del conjunto de aristas E_I que salen de I, de la siguiente forma

$$\mathcal{P}^I = \{\{a \in A(G)\} | i(a) = I\}.$$

El conjunto de vértices que se obtiene es entonces

$$V(G^{[\mathcal{P}]}) = \{i(a)^{t(a)} | a \in A(G)\}$$

y el de aristas

$$A(G^{[\mathcal{P}]}) = \left\{ e^K : \begin{array}{ll} (1) & e^K \text{ es la arista que va de } i(a)^{t(a)} \text{ a } i(b)^{t(b)} \\ (2) & i(e) = i(a) \\ (3) & t(e) = t(a) = i(b) \\ (4) & K = t(b) \end{array} \right\}.$$

Es sencillo entonces verificar que las funciones $f\colon V(G^{[\mathcal{P}]}) \to V(G^{[2]})$, $g\colon A(G^{[\mathcal{P}]}) \to A(G^{[2]})$ definidas por la regla de correspondencia

$$f(i(e)^{t(e)}) = e \text{ y } g(e^K) = ef \text{ con } t(f) = K, i(f) = t(e)$$

constituyen un isomorfismo de gráficas. □

Proposición 2.4.1. *Sea G una gráfica, entonces $(G^{[N]})^2 \cong G^{[N+1]}$*

Demostración. Es básicamente el detalle de que las aristas $G^{[N]}$ son exactamente trayectorias de longitud $N+1$ y el isomorfismo está impícito en la definicion de $G^{[M]}$ para cualquier M $\square$

Lema 2.4.2. *Sea G una gráfica. Para cada $N \in \mathbb{N}$ existe una sucesión finita de escisiones (de salida) de G tal que al final obtenemos una gráfica isomorfa a $G^{[N]}$.*

Demostración. Hagamos la prueba por inducción. La base $N = 1$ ya está probada por el lema anterior. Supongamos que con una cantidad finita de escisiones (de salida) de G obtenemos $G^{[N]}$. Tenemos que $\left((G^{[\mathcal{P}_1]})^{\cdots}\right)^{[\mathcal{P}_m]} \cong G^{[N]}$ y por el lemma anterior existe $\mathcal{P}_{m+1}$ partición de las aristas de $G^{[N]}$ tal que ${G^{[N]}}^{[\mathcal{P}_{m+1}]} \cong {G^{[N]}}^{[2]} \cong G^{[N+1]}$, y entonces $\left(((G^{[\mathcal{P}_1]})^{\cdots})^{[\mathcal{P}_m]}\right)^{[\mathcal{P}_{m+1}]} \cong \left(G^{[N]}\right)^{[\mathcal{P}_{m+1}]} \cong \left(G^{[N]}\right)^{[2]} \cong G^{[N+1]}$ con lo cual terminamos la prueba. $\square$

Capítulo 3

ENTROPÍA Y TEORIA ERGÓDICA

3.1. Particiones y funciones de información

Definición 3.1.1. Sea $(X, \mathfrak{A}, \mu)$ un espacio de probabilidad. Una *partición* se define como una colección $(A_i)_{i\in I}$ de subconjuntos de X, donde I es a lo mas numerable y se cumple que

1. $\mu\left(\bigcup_{i\in I} A_i\right) = 1.$

2. Si $i, j \in I$ son distintos, entonces $A_i \cap A_j = \emptyset$.

Una partición es *finita* si todos los elementos de la partición, salvo un número finito de ellos, tienen medida cero.

En nuestro caso particular sólo trataremos con particiones finitas. A los elementos de una partición los llamaremos *celdas*.

SUPOSICIÓIN. Por el resto de este capítulo, a menos que se indique de otra manera, al hablar de particiones o de σ-álgebras estaremos en el entendido que corresponden a un espacio de probabilidad $(X, \mathfrak{A}, \mu)$ dado.

Definición 3.1.2. Sean $\mathcal{P}$ y $\mathcal{Q}$ dos particiones.

1. Diremos que $\mathcal{P}$ es *más fina* que $\mathcal{Q}$ (o también que $\mathcal{P}$ *refina* a $\mathcal{Q}$, o que $\mathcal{P}$ es un *refinamiento* de $\mathcal{Q}$), si cada celda de $\mathcal{P}$ está contenida en una celda de $\mathcal{Q}$, lo cual denotaremos por $\mathcal{P} \succeq \mathcal{Q}$, lo que equivale a que cada elemento de $\mathcal{Q}$ es unión de elementos de $\mathcal{P}$.

2. La *refinación* de $\mathcal{P}$ y $\mathcal{Q}$ es el refinamiento menos fino que refina a $\mathcal{P}$ y a $\mathcal{Q}$ al mismo tiempo (es decir cualquier *refinamiento* de ambas tiene que refinar a su *refinación*) y se denota como $\mathcal{P} \vee \mathcal{Q}$.

Proposición 3.1.1. *La relación $\preceq$ es reflexiva y transitiva.*

Demostración. El hecho que $\preceq$ es reflexiva es claro.

Sean $\mathcal{P} \preceq \mathcal{Q} \preceq \mathcal{R}$ particiones. Si $l \in \mathcal{R}$, entonces existe $s \in \mathcal{Q}$ y así mismo existe $t \in \mathcal{R}$ tales que $s \subset l$ y $l \subset t$, entonces $s \subset t$, lo que concluye la prueba. □

Proposición 3.1.2. *Sean $\mathcal{P}$ y $\mathcal{Q}$ particiones. Entonces $\mathcal{P} = \mathcal{Q}$ si y sólo si $\mathcal{P} \succeq \mathcal{Q}$ y $\mathcal{Q} \succeq \mathcal{P}$.*

Demostración. La ida es directa.

Para el regreso, sea $x \in \mathcal{P}$. Como $\mathcal{P}$ es más fina que $\mathcal{Q}$, existe $y \in \mathcal{Q}$ tal que $x \subset y$, pero también $\mathcal{Q}$ es más fina que $\mathcal{P}$, por lo que existe $w \in \mathcal{P}$ tal que $y \subset w$. Esto implica que $x \subset w$, y como $\mathcal{P}$ es una partición, tenemos que $x = w$, y por las contenciones se deduce que $x = y$, por lo que concluimos que $\mathcal{P} \subset \mathcal{Q}$. La otra contención es análoga y obtenemos $\mathcal{P} = \mathcal{Q}$, lo que concluye la prueba. □

De la definición 3.1.2 se deduce la siguiente proposición directa y sencilla.

Proposición 3.1.3. *$\mathcal{P} \succeq \mathcal{Q}$ si y sólo si $\mathcal{P} \vee \mathcal{Q} = \mathcal{P}$.*

Demostración. Como $\mathcal{P} \succeq \mathcal{Q}$ y $\mathcal{P} \succeq \mathcal{P}$, se sigue directo de la definición de refinación que $\mathcal{P} \succeq \mathcal{P} \vee \mathcal{Q}$, pero como $\mathcal{P} \preceq \mathcal{P} \vee \mathcal{Q}$, por la proposición anterior, concluimos que $\mathcal{P} \vee \mathcal{Q} = \mathcal{P}$.

□

Hagamos notar un hecho consecuencia de los resultados anteriores, la relacion $\preceq$ en la particiones es una relacion de orden.

Proposición 3.1.4. *Sean $\mathcal{P}, \mathcal{Q}, \mathcal{R}, \mathcal{S}$ particiones tales que $\mathcal{P} \preceq \mathcal{R}$ y $\mathcal{Q} \preceq \mathcal{S}$. Entonces*

$$\mathcal{P} \vee \mathcal{Q} \preceq \mathcal{R} \vee \mathcal{S}.$$

Demostración. Tenemos que $\mathcal{R}, \mathcal{S} \preceq \mathcal{R} \vee \mathcal{S}$, entonces por transitividad se obtiene $\mathcal{P}, \mathcal{Q} \preceq \mathcal{R} \vee \mathcal{S}$. La definición 3.1.2 implica entonces que

$$\mathcal{P} \vee \mathcal{Q} \preceq \mathcal{R} \vee \mathcal{S}.$$

□

Proposición 3.1.5. *Sea $\mathcal{P}$ y $\mathcal{Q}$ dos particiones. Entonces*

$$\mathcal{P} \vee \mathcal{Q} = \{x \cap y : x \in \mathcal{P}, y \in \mathcal{Q}\}.$$

Demostración. Primero, como $\mathcal{P}$ y $\mathcal{Q}$ son particiones, $\mathcal{R} = \{x \cap y : x \in \mathcal{P}, y \in \mathcal{Q}\}$ es una partición, puesto que

$$\mu\left(\bigcup_{x \in \mathcal{P}, y \in \mathcal{Q}} x \cap y\right) = \mu\left(\bigcup_{x \in \mathcal{P}} x \cap \bigcup_{y \in \mathcal{Q}} y\right)$$

y como intersección de conjuntos de medida de probabilidad 1 tienen medidad de probabilidad 1,

$$1 = \mu\left(\bigcup_{x \in \mathcal{P}, y \in \mathcal{Q}} x \cap y\right) = \mu\left(\bigcup_{x \in \mathcal{P}} x \cap \bigcup_{y \in \mathcal{Q}} y\right),$$

lo que indica que se satisface la condición 1 de la Definición 3.1.1. Ahora sean $x, y \in \mathcal{P}$ y $v, w \in \mathcal{Q}$ y supongamos que $x \cap v \neq y \cap w$. Entonces no puede ser que los correspondientes elementos x, y coincidan al mismo tiempo que coinciden v, w. Podemos suponer entonces que $x \neq y$, pero entonces $x \cap y = \varnothing$ porque $\mathcal{P}$ es una partición, y entonces $x \cap v \cap y \cap w \subseteq x \cap y = \emptyset$, que quiere decir que se satisface 2 de la Definición 3.1.1. Ahora para ver que son la misma partición, hace falta ver que cada una es más fina que la otra, es decir, que $\mathcal{R} \succeq \mathcal{P} \vee \mathcal{Q}$ y $\mathcal{P} \vee \mathcal{Q} \succeq \mathcal{R}$. Es obvio que

$$\mathcal{R} \succeq \mathcal{P} \quad \text{y} \quad \mathcal{R} \succeq \mathcal{Q}$$

puesto que los elementos del lado izquierdo de nuestras desigualdades son subconjuntos de los elementos del lado derecho. Ésto implica directamente que

$$\{x \cap y : x \in \mathcal{P}, y \in \mathcal{Q}\} \succeq \mathcal{P} \vee \mathcal{Q}$$

puesto que como es más fina que ambas, tiene que ser más fina que $\mathcal{P} \vee \mathcal{Q}$. Ahora para probar lo correspondiente faltante tomamos $\mathcal{S}$ una partición que refina a ambas y veamos que es más fina que $\mathcal{R}$. Primero sea s un elemento de dicha partición $\mathcal{S}$, entonces existe un elemento $f \in \mathcal{P}$ tal que $s \subseteq f$ y también $g \in \mathcal{Q}$ tal que $s \subseteq g$, por lo que entonces $s \subseteq f \cap g$, y con ésto probamos que es más fina. Entonces cualquier partición que refine a $\mathcal{P}$ y a $\mathcal{Q}$ simultaneamente refina a $\mathcal{R}$, en particular

$$\mathcal{P} \vee \mathcal{Q} \succeq \mathcal{R}$$

y por la Proposición 3.1.2 se sigue que

$$\mathcal{P} \vee \mathcal{Q} = \{x \cap y : x \in \mathcal{P}, y \in \mathcal{Q}\}.$$

$\square$

Definición 3.1.3. Sea $(X, \mathfrak{A}, \mu)$ un espacio de probabilidad y $\mathcal{P}$ una partición de dicho espacio. Definimos la *función de información* de dicha partición como la función $I_{\mu,\mathcal{P}} \colon X \to [0, \infty)$ dada para toda $x \in X$ por

$$I_{\mu,\mathcal{P}}(x) = I_{\mathcal{P}}(x) = -\log(\mu(A_x))$$

donde A_x es la única celda de $\mathcal{P}$ que contiene a x. Si $\mathcal{Q}$ es otra partición de X, definimos la *función de información condicional de $\mathcal{P}$ con respecto de* (o *condicionada a*) $\mathcal{Q}$ como la función $I_{\mu,\mathcal{P}|\mathcal{Q}} \colon X \to [0, \infty)$ definida para toda $x \in X$ por

$$I_{\mu,\mathcal{P}|\mathcal{Q}}(x) = I_{\mathcal{P}|\mathcal{Q}}(x) = I_{\mathcal{P}\vee\mathcal{Q}} - I_{\mathcal{Q}}(x) = -\log\left(\frac{\mu(A_x \cap B_x)}{\mu(B_x)}\right),$$

donde A_x y B_x son las únicas celdas respectivamente de $\mathcal{P}$ y de $\mathcal{Q}$ que contienen a x.

3.2. Entropía de particiones

Definición 3.2.1. La *entropía* (o *entropía estática*) *de $\mathcal{P}$ con respecto a μ* se define como la esperanza de su función de información, es decir,

$$H(\mu, \mathcal{P}) = H(\mathcal{P}) = \int_X I_{\mathcal{P}} d\mu.$$

Proposición 3.2.1.

$$H(\mu, \mathcal{P}) = -\sum_{A \in \mathcal{P}} \mu(A) \log(\mu(A)).$$

Demostración. Como $\mathcal{P}$ es una partición, $\sum_{A\in\mathcal{P}} \mathbb{I}_A(x) = 1$ μ-casi-seguramente. Entonces

$$I_{\mathcal{P}}(x) = \left(\sum_{A\in\mathcal{P}} \mathbb{I}_A(x)\right)(-\log(\mu(A_x))) = -\sum_{A\in\mathcal{P}} \mathbb{I}_A(x)\log(\mu(A_x)) = -\sum_{A\in\mathcal{P}} \mathbb{I}_A(x)\log(\mu(A)).$$

A partir de aquí es sencillo cómo concluir: usando la linealidad de la integral y el Teorema de Convergencia Monótona, tenemos que

$$\begin{aligned}
H(\mu,\mathcal{P}) &= \int_X I_{\mathcal{P}} d\mu \\
&= \int_X -\sum_{A\in\mathcal{P}} \mathbb{I}_A(x)\log(\mu(A))d\mu \\
&= -\sum_{A\in\mathcal{P}} \int_X \mathbb{I}_A(x)\log(\mu(A))d\mu \\
&= -\sum_{A\in\mathcal{P}} \log(\mu(A)) \int_X \mathbb{I}_A(x)d\mu \\
&= -\sum_{A\in\mathcal{P}} \mu(A)\log(\mu(A)).
\end{aligned}$$

□

Definición 3.2.2. Definimos la *entropía condicional* de $\mathcal{P}$ dado $\mathcal{Q}$ como

$$H(\mu,\mathcal{P}|\mathcal{Q}) = \int_X I_{\mathcal{P}|\mathcal{Q}} d\mu.$$

Proposición 3.2.2.

$$H(\mu,\mathcal{P}|\mathcal{Q}) = -\sum_{A\in\mathcal{P},B\in\mathcal{Q},A\cap B\neq\emptyset} \mu(A\cap B)\log\left(\frac{\mu(A\cap B)}{\mu(B)}\right).$$

Demostración. Primero,

$$H(\mu,\mathcal{P}|\mathcal{Q}) = \int (I_{\mathcal{P}\vee\mathcal{Q}} - I_{\mathcal{Q}})d\mu,$$

pero por la linealidad de la integral,

$$\int_X (I_{\mathcal{P}\vee\mathcal{Q}} - I_{\mathcal{Q}})d\mu = \int_X I_{\mathcal{P}\vee\mathcal{Q}} d\mu - \int_X I_{\mathcal{Q}} d\mu = H(\mathcal{P}\vee\mathcal{Q}) - H(\mathcal{Q}).$$

Por la Proposición 3.2.1 tenemos entonces que

$$H(\mathcal{P}\vee\mathcal{Q}) - H(\mathcal{Q}) = -\sum_{A\in\mathcal{P}\vee\mathcal{Q}} \mu(A)\log(\mu(A)) - \sum_{B\in\mathcal{Q}} \mu(B)\log(\mu(B)))$$

y así, la Proposición 3.1.5 implica que

$$\begin{aligned}
H(\mathcal{P} \vee \mathcal{Q}) - H(\mathcal{Q}) &= -\left(\sum_{A \in \mathcal{P} \vee \mathcal{Q}} \mu(A)\log(\mu(A)) - \sum_{B \in \mathcal{Q}} \mu(B)\log(\mu(B))\right) \\
&= -\left(\sum_{A \in \mathcal{P}, B \in \mathcal{Q}} \mu(A \cap B)\log(\mu(A \cap B)) - \sum_{B \in \mathcal{Q}} \mu(B)\log(\mu(B))\right) \\
&= -\left(\sum_{A \in \mathcal{P}, B \in \mathcal{Q}} \mu(A \cap B)\log(\mu(A \cap B)) - \sum_{B \in \mathcal{Q}} \mu(B)\log(\mu(\bigcup_{A \in \mathcal{P}} A \cap B))\right) \\
&= -\left(\sum_{A \in \mathcal{P}, B \in \mathcal{Q}} \mu(A \cap B)\log(\mu(A \cap B)) - \sum_{A \in \mathcal{Q}, B \in \mathcal{Q}} \mu(A \cap B)\log(\mu(B))\right) \\
&= -\left(\sum_{A \in \mathcal{P}, B \in \mathcal{Q}} \mu(A \cap B)\log(\mu(A \cap B)) - \mu(A \cap B)\log(\mu(B))\right),
\end{aligned}$$

lo último debido a que $\mathcal{P}$ es una partición. Tenemos entonces que

$$\begin{aligned}
H(\mathcal{P} \vee \mathcal{Q}) - H(\mathcal{Q}) &= -\left(\sum_{A \in \mathcal{P}, B \in \mathcal{Q}} \mu(A \cap B)\log(\mu(A \cap B)) - \mu(A \cap B)\log(\mu(B))\right) \\
&= -\left(\sum_{A \in \mathcal{P}, B \in \mathcal{Q}} \mu(A \cap B)(\log(\mu(A \cap B)) - \log(\mu(B)))\right) \\
&= -\sum_{A \in \mathcal{P}, B \in \mathcal{Q}, A \cap B \neq \emptyset} \mu(A \cap B)\log(\frac{\mu(A \cap B)}{\mu(B)}),
\end{aligned}$$

lo que concluye la prueba. □

3.3. Entropía y σ-álgebras

Vamos a asociar σ-álgebras a sucesiones de particiones, y esta asociación nos va a permitir definir conceptos de entropía para σ-álgebras, incluso junto con particiones. Comencemos con la noción de refinación, la cual se puede extender no solo inductivamente para una cantidad finita de particiones.

Definición 3.3.1. Sea $\mathcal{Q} = (\mathcal{Q}_i)_{i \in \mathbb{N}}$ una sucesión de particiones. La *refinación* de $\mathcal{Q}$ es la σ-álgebra $\mathfrak{Q}$ generada por los elementos de la sucesión, y se denota por

$$\mathfrak{Q} = \bigvee_{i=1}^{\infty} \mathcal{Q}_i.$$

Ahora miremos a la función de información condicional.

Definición 3.3.2. Sean $\mathcal{P}$ una partición y $\mathfrak{F}$ una σ-álgebra. Definimos la *función de información condicional de* $\mathcal{P}$ *dado* $\mathfrak{F}$ como la función $I_{\mu,\mathcal{P}|\mathfrak{F}}\colon X \to [0,\infty)$ definida para cada $x \in X$ por

$$I_{\mu,\mathcal{P}|\mathfrak{F}}(x) = I_{\mathcal{P}|\mathfrak{F}}(x) = -\log(\mathbb{E}_{\mu}[\mathbb{I}_{A_x}|\mathfrak{F}])(x)$$

y la *entropía condicional* de $\mathcal{P}$ dado $\mathfrak{F}$ como

$$H(\mu,\mathcal{P}|\mathfrak{F}) = \int_X I_{\mu,\mathcal{P}|\mathfrak{F}} d\mu.$$

Observación 3.3.1. Es interesante notar que, definiendo

$$\mu_B(A) = \frac{\mu(A\cap B)}{\mu(B)} = \mu(A|B)$$

(lo que se conoce en probabilidad como *medida de probabilidad condicional dado* B), la Proposición 3.2.2 se puede reescribir como

$$H(\mu,\mathcal{P}|\mathcal{Q}) = -\sum_{A\in\mathcal{P},B\in\mathcal{P},A\cap B\neq\emptyset} \mu(B)\mu_B(A)\log(\mu_B(A)).$$

En efecto, multiplicando y dividiendo por $\mu(B)$ y simplificando tenemos

$$\begin{aligned}
H(\mu,\mathcal{P}|\mathcal{Q}) &= -\sum_{A\in\mathcal{P},B\in\mathcal{P},A\cap B\neq\emptyset} \mu(B)\mu_B(A)\log(\mu_B(A))\\
&= -\sum_{B\in\mathcal{Q}}\sum_{A\in\mathcal{P}} \mu(B)\mu_B(A)\log(\mu_B(A))\\
&= \sum_{B\in\mathcal{Q}} \mu(B)(-\sum_{A\in\mathcal{P}} \mu_B(A)\log(\mu_B(A)))\\
&= \sum_{B\in\mathcal{Q}} \mu(B)(H(\mu_B,\mathcal{P}))\\
&= \sum_{B\in\mathcal{Q}} \mu(B)H_B(\mathcal{P}).
\end{aligned}$$

Enuncimos ahora un resultado más, el cual relaciona la definición de entropía condicional para particiones con el de σ-álgebras. De hecho veamos que la Definición 3.2.2 se generaliza con la entropía condicional para particiones.

Proposición 3.3.1.

$$H(\mu,\mathcal{P}|\mathcal{Q}) = H(\mu,\mathcal{P}|\sigma(\mathcal{Q})),$$

donde $\sigma(\mathcal{Q})$ *es la* σ*-álgebra generada por* $\mathcal{Q}$.

Demostración. Primero veamos que

$$\mathbb{E}[\mathbb{I}_A(x)|\mathcal{Q}] = \prod_{B\in\mathcal{Q}} \left(\frac{\mu(A\cap B)}{\mu(B)}\right)^{\mathbb{I}_B(x)}.$$

En el lado derecho vemos una función $\sigma(\mathcal{Q})$-medible puesto que cada B está en $\mathcal{Q}$. Sea N el complemento de la unión de $\mathcal{Q}$ (N tiene medida cero puesto que $\mathcal{Q}$ es partición). Afirmamos que

$$\sigma(\mathcal{Q}) = \left\{ \bigcup_{B\in S} B \cup M \ :\ S \subseteq \mathcal{Q},\ \ M \text{ es } N \text{ o el vacío} \right\}. \tag{3.1}$$

Para ver ésto, primero notemos que $\mathcal{Q} \cup \{N\}$ es una partición algebráica del espacio, puesto que N es el complemento de la unión de los elementos de $\mathcal{Q}$, así que el lado derecho de (3.1) es cerrado bajo complementos (debido a que el complemento de alguna de esas uniones, por ser $\mathcal{Q} \cup \{N\}$ partición algebráica, es unión del resto). Es obvio que es cerrado bajo uniones. El vacío es unión vacía, por lo tanto está ahí, y el total también (por ser partición algebráica). Este conjunto continene a $\mathcal{Q}$ por lo cual es una σ-álgebra que contiene a $\mathcal{Q}$. Así obtenemos

$$\sigma(\mathcal{Q}) \subset \left\{ \bigcup_{B\in S} B \cup M \ :S \subseteq \mathcal{Q}\ \ M \text{ es } N \text{ o el vacío} \right\}.$$

Ahora, cualquier σ-álgebra $\mathfrak{B}$ que contenga a $\mathcal{Q}$ contiene a la unión de $\mathcal{Q}$ y por lo tanto a N, así también contiene a $\mathcal{Q} \cup \{N\}$, por ésto contiene a al vacío y a cualquier unión de $\mathcal{Q} \cup \{N\}$. Luego, a causa de ésto,

$$\mathfrak{B} \supseteq \left\{ \bigcup_{B\in S} B \cup M \ :S \subseteq \mathcal{Q} M \text{ es } N \text{ o el vacío} \right\}$$

y luego

$$\sigma(\mathcal{Q}) \supseteq \left\{ \bigcup_{B\in S} B \cup M \ :S \subseteq \mathcal{Q} M \text{ es } N \text{ o el vacío} \right\}.$$

Regresando a la prueba original, sea $L \in \sigma(\mathcal{Q})$. Como se cumple (3.1), L es la unión de elementos de $\mathcal{Q}$. Tenemos que

$$\begin{aligned}
\mathbb{E}\left[\prod_{B\in\mathcal{Q}} \left(\tfrac{\mu(A\cap B)}{\mu(B)}\right)^{\mathbb{I}_B(x)} \mathbb{I}_L\right] &= \mathbb{E}\left[\prod_{B\in\mathcal{Q}} \left(\tfrac{\mu(A\cap B)}{\mu(B)}\right)^{\mathbb{I}_B(x)} (\sum_{C \text{ es uniendo de } L} \mathbb{I}_C)\right] \\
&= \sum_{C \text{ es uniendo de } L} \mathbb{E}\left[\prod_{B\in\mathcal{Q}} \left(\frac{\mu(A\cap B)}{\mu(B)}\right)^{\mathbb{I}_B(x)} \mathbb{I}_C\right] \\
&= \sum_{C \text{ es uniendo de } L} \mathbb{E}\left[\left(\frac{\mu(A\cap C)}{\mu(C)}\right)^{\mathbb{I}_C(x)} \mathbb{I}_C\right] \\
&= \sum_{C \text{ es uniendo de } L} \frac{\mu(A\cap C)}{\mu(C)}\mu(C),
\end{aligned} \tag{3.2}$$

lo último a causa de que

$$\prod_{B\in\mathcal{Q}}\left(\frac{\mu(A\cap B)}{\mu(B)}\right)^{\mathbb{I}_B(x)}\mathbb{I}_B\mathbb{I}_C$$

sólo toma el valor de $\frac{\mu(A\cap C)}{\mu(C)}$ en el conjunto C o cero en el complemento. Entonces (3.2) es igual a

$$\sum_{C \text{ es uniendo de } L}\mu(A\cap C)=\mu(A\cap(\bigcup_{C \text{ es uniendo de } L}C))=\mu(A\cap L)=\mathbb{E}\left[\mathbb{I}_A(x)\mathbb{I}_L(x)\right].$$

De aquí y del hecho de que $\prod_{B\in\mathcal{Q}}\left(\frac{\mu(A\cap B)}{\mu(B)}\right)^{\mathbb{I}_B(x)}$ es $\sigma(\mathcal{Q})$-medible se sigue que

$$\mathbb{E}[\mathbb{I}_A(x)|\mathcal{Q}]=\prod_{B\in\mathcal{Q}}\left(\frac{\mu(A\cap B)}{\mu(B)}\right)^{\mathbb{I}_B(x)}.$$

Ésto nos dice que

$$\begin{aligned}
I_{\mu,\mathcal{P}|\sigma(\mathcal{Q})}(x) &= I_{\mathcal{P}|\sigma(\mathcal{Q})}(x)\\
&= -\log(\mathbb{E}_\mu[\mathbb{I}_{A_x}|\sigma(\mathcal{Q})])(x)\\
&= -\sum_{A\in\mathcal{P}}\log(\mathbb{E}_\mu[\mathbb{I}_A|\sigma(\mathcal{Q})])\mathbb{I}_A(x)\\
&= -\sum_{A\in\mathcal{P}}\log(\prod_{B\in\mathcal{Q}}\left(\frac{\mu(A\cap B)}{\mu(B)}\right)^{\mathbb{I}_B(x)})\mathbb{I}_A(x)\\
&= -\sum_{A\in\mathcal{P},B\in\mathcal{Q}}\log\left(\frac{\mu(A\cap B)}{\mu(B)}\right)\mathbb{I}_A(x)\mathbb{I}_B(x)\\
&= -\log\left(\frac{\mu(A_x\cap B_x)}{\mu(B_x)}\right)=I_{\mu,\mathcal{P}|\mathcal{Q}}(x)
\end{aligned}$$

con lo cual terminamos la proposición. □

Proposición 3.3.2. *Sean $\mathcal{P},\mathcal{Q}$ y $\mathcal{R}$ tres particiones.*

1. $H(\mathcal{P}|\mathcal{Q}\vee\mathcal{R})=\sum_{C\in\mathcal{R}}\mu(C)H_C(\mathcal{P}|\mathcal{Q})$.

2. $H(\mathcal{P}\vee\mathcal{Q}|\mathcal{R})=H(\mathcal{P}|\mathcal{Q}\vee\mathcal{R})+H(\mathcal{Q}|\mathcal{R})$.

Demostración. Para 1, por la Observación 3.3.1 hecha anteriormente,

$$\begin{aligned} H(\mathcal{P}|\mathcal{Q}\vee\mathcal{R}) &= \sum_{B\cap C\in\mathcal{Q}\vee\mathcal{R}} \mu(B\cap C)H_{B\cap C}(\mathcal{P}) \\ &= \sum_{B\in\mathcal{Q},C\in\mathcal{R}} \frac{\mu(C)}{\mu(C)}\mu(B\cap C)H_{B\cap C}(\mathcal{P}) \\ &= \sum_{B\in\mathcal{Q},C\in\mathcal{R}} \mu(C)\mu_C(B)H_{B\cap C}(\mathcal{P}). \end{aligned} \tag{3.3}$$

Por otro lado, tenemos que

$$\begin{aligned} H_{B\cap C}(\mathcal{P}) &= -\sum_{A\in\mathcal{P}} \mu_{B\cap C}(A)\log\left(\mu_{B\cap C}(A)\right) \\ &= -\sum_{A\in\mathcal{P}} \frac{\mu(B\cap C\cap A)}{\mu(B\cap C)}\log\left(\frac{\mu(B\cap C\cap A)}{B\cap C}\right) \\ &= -\sum_{A\in\mathcal{P}} \frac{\frac{\mu(B\cap C\cap A)}{\mu(C)}}{\frac{\mu(B\cap C)}{\mu(C)}}\log\left(\frac{\frac{\mu(B\cap C\cap A)}{\mu(C)}}{\frac{\mu(B\cap C)}{\mu(C)}}\right) \\ &= -\sum_{A\in\mathcal{P}} \frac{\mu_C(B\cap A)}{\mu_C(B)}\log\left(\frac{\mu_C(B\cap A)}{\mu_C(B)}\right) \\ &= -\sum_{A\in\mathcal{P}} \mu_{C\,B}(A)\log(\mu_{C\,B}(A)) \\ &= H_{C\,B}(\mathcal{P}). \end{aligned} \tag{3.4}$$

Juntando las ecuaciones (3.3) y (3.4) obtenemos

$$\begin{aligned} H(\mathcal{P}|\mathcal{Q}\vee\mathcal{R}) &= \sum_{B\in\mathcal{Q},C\in\mathcal{R}} \mu(C)\mu_C(B)H_{C\,B}(\mathcal{P}) \\ &= \sum_{C\in\mathcal{R}} \mu(C)\sum_{B\in\mathcal{Q}} \mu_C(B)H_{C\,B}(\mathcal{P}) \\ &= \sum_{C\in\mathcal{R}} \mu(C)H(\mu_C,\mathcal{P}|\mathcal{Q}) \\ &= \sum_{C\in\mathcal{R}} \mu(C)H_C(\mathcal{P}|\mathcal{Q}), \end{aligned}$$

donde las dos igualdades finales son resultado de la Observación 3.3.1, lo que concluye el primer resultado.

Para 2, ocupando 1 y la Observación 3.3.1 vemos que

$$\begin{aligned} H(\mathcal{P}|\mathcal{Q}\vee\mathcal{R})+H(\mathcal{Q}|\mathcal{R}) &= \sum_{C\in\mathcal{R}}\mu(C)H_C(\mathcal{P}|\mathcal{Q})+\sum_{C\in\mathcal{R}}\mu(C)H_C(\mathcal{Q}) \\ &= \sum_{C\in\mathcal{R}}\mu(C)(H_C(\mathcal{P}|\mathcal{Q})+H_C(\mathcal{Q})) \\ &= \sum_{C\in\mathcal{R}}\mu(C)H_C(\mathcal{P}\vee\mathcal{Q}) \\ &= H(\mathcal{P}\vee\mathcal{Q}|\mathcal{R}), \end{aligned}$$

donde los pasos finales de nuevo son consecuencia de la Observación 3.3.1, y con ésto se concluye la proposición. □

Definición 3.3.3. Sea $\mathcal{Q}$ una particion y $\mathfrak{B}$ una σ-álgebra. Definimos la *refinación* de $\mathcal{Q}$ y $\mathfrak{B}$ como la σ-álgebra generada por estos dos conjuntos. Como es de esperarse, la refinación se denota de la manera tradicional como $\mathcal{Q}\vee\mathfrak{B}$.

Proposición 3.3.3. *Sean $\mathcal{P}$ y $\mathcal{Q}$ dos particiones y $\mathfrak{B}$ una σ-álgebra. Entonces*

$$H(\mathcal{P}\vee\mathcal{Q}|\mathfrak{B})=H(\mathcal{P}|\mathcal{Q}\vee\mathfrak{B})+H(\mathcal{Q}|\mathfrak{B}).$$

Demostración. La demostración de la afirmación es directa de la misma propiedad para particiones, notando el hecho de que

$$H(\mathcal{P}|\mathfrak{B})=\text{ínf}\left\{H(\mathcal{P}|\mathcal{Q})|\mathcal{Q}\text{ es una partición }\mathfrak{B}\text{-medible}\right\}.$$

□

Lema 3.3.1. *Si $f\colon I\subset\mathbb{R}\to\mathbb{R}$, donde I es un intervalo, f es una función continua y estrictamente concava en I, entonces $\forall\,(t_n)_{n\in\mathbb{N}}$ sucesión de elementos de I y para toda familia de $\alpha_i\in[0,1)$ tal que $\sum_{i=0}^{\infty}\alpha_i=1$,*

$$f\left(\sum_{i=0}^{\infty}\alpha_i t_i\right)\geq\sum_{i=0}^{\infty}\alpha_i f(t_i).$$

Si además $\alpha_i\neq 0$ para un número finito de $i\in\mathbb{N}$ y la sucesión $(t_n)_{n\in\mathbb{N}}$ es no constante sobre el mismo conjunto de índices en donde todos los α_i's $\neq 0$, obtenemos

$$f\left(\sum_{i=0}^{\infty}\alpha_i t_i\right)>\sum_{i=0}^{\infty}\alpha_i f(t_i).$$

Demostración. Probaremos primero la proposición para una cantidad finita de $\alpha_i\neq 0$, es decir para cualquier conjunto finito de $\alpha_0,\ldots,\alpha_n\in[0,1)$ tal que $\sum_{i=0}^{n}\alpha_i=1$, veamos que

$$f\left(\sum_{i=0}^{n}\alpha_i t_i\right)>\sum_{i=0}^{n}\alpha_i f(t_i)\quad\forall\,(t_n)_{n\in\mathbb{N}}\in I.$$

Lo probaremos por inducción. La base de inducción será 2, que es el caso de la definición estándar de una función (estrictamente) concava. Supongamos que se vale si hay n α_i' s $\neq 0$ que cumplen con la proposición, es decir, si $\sum_{i=0}^{n} \alpha_i = 1$, con $\alpha_i \in [0, 1)$, donde la familia de t_i's (para los mismos índices donde $\alpha_i \neq 0$) es una familia no constante, entonces

$$f\left(\sum_{i=0}^{n} \alpha_i t_i\right) > \sum_{i=0}^{n} \alpha_i f(t_i) \quad \forall\, (t_n)_{n\in\mathbb{N}} \in I. \tag{3.5}$$

Probemos la proposición para $n+1$ parámetros α_i's $\neq 0$ bajo las hipótesis, es decir supongamos que $\sum_{i=0}^{n} \alpha_i + \alpha_{n+1} = 1$. Tenemos entonces que

$$\begin{aligned}
f\left(\sum_{i=0}^{n+1} \alpha_i t_i\right) &= f\left(\sum_{i=0}^{n} \alpha_i t_i + \alpha_{n+1} t_{n+1}\right) \\
&= f\left(\frac{\sum_{i=0}^{n} \alpha_i \sum_{i=0}^{n} \alpha_i t_i}{\sum_{i=0}^{n} \alpha_i} + \alpha_{n+1} t_{n+1}\right) \\
&= f\left(\sum_{i=0}^{n} \alpha_i \sum_{i=0}^{n} \frac{\alpha_i}{\sum_{i=0}^{n} \alpha_i} t_i + \alpha_{n+1} t_{n+1}\right) \\
&\geq \sum_{i=0}^{n} \alpha_i f\left(\sum_{i=0}^{n} \frac{\alpha_i}{\sum_{i=0}^{n} \alpha_i} t_i\right) + \alpha_{n+1} f(t_{n+1}),
\end{aligned}$$

donde la última desigualdad es simplemente por concavidad. Ya que existen al menos dos índices i, j tales que $t_i \neq t_j$ y $n \geq 2$, podemos suponer que $i, j \leq n$, por lo que entonces podemos usar la hipótesis de inducción puesto que

$$\sum_{i=0}^{n} \frac{\alpha_i}{\sum_{i=0}^{n} \alpha_i} = 1.$$

Entonces

$$
\begin{aligned}
\sum_{i=0}^{n}\alpha_i f\left(\sum_{i=0}^{n}\frac{\alpha_i}{\sum\limits_{i=0}^{n}\alpha_i}t_i\right)+\alpha_{n+1}f(t_{n+1}) &> \sum_{i=0}^{n}\alpha_i\sum_{i=0}^{n}\frac{\alpha_i}{\sum\limits_{i=0}^{n}\alpha_i}f(t_i)+\alpha_{n+1}f(t_{n+1}) \\
&= \frac{\sum\limits_{i=0}^{n}\alpha_i}{\sum\limits_{i=0}^{n}\alpha_i}\sum_{i=0}^{n}\alpha_i f(t_i)+\alpha_{n+1}f(t_{n+1}) \\
&= \sum_{i=0}^{n+1}\alpha_i f(t_i),
\end{aligned}
$$

lo que prueba el caso donde los α_{n+1}'s distintos de cero son una cantidad finita.

Para pasar ésto al límite, utilicemos la continuidad. Supongamos que hay una cantidad infinita de α_i's distintas de cero. Entonces

$$
\begin{aligned}
\sum_{i=0}^{\infty}\alpha_i f(t_i) &= \lim_{n\to\infty}\sum_{i=0}^{n}\alpha_i f(t_i) &&= \lim_{n\to\infty}\frac{1}{\sum\limits_{i=0}^{n}\alpha_i}\lim_{n\to\infty}\sum_{i=0}^{n}\alpha_i f(t_i) \\
&= \lim_{n\to\infty}\sum_{i=0}^{n}\frac{\alpha_i}{\sum\limits_{i=0}^{n}\alpha_i}f(t_i) &&\leq \lim_{n\to\infty} f\left(\sum_{i=0}^{n}\frac{\alpha_i}{\sum_{i=0}^{n}\alpha_i}t_i\right) \\
&= f\left(\lim_{n\to\infty}\sum_{i=0}^{n}\frac{\alpha_i}{\sum\limits_{i=0}^{n}\alpha_i}t_i\right) &&= f\left(\lim_{n\to\infty}\frac{\sum\limits_{i=0}^{n}\alpha_i t_i}{\sum\limits_{i=0}^{n}\alpha_i}\right) \\
&= f\left(\frac{1}{\sum\limits_{i=0}^{\infty}\alpha_i}\sum_{i=0}^{\infty}\alpha_i t_i\right) &&= f\left(\sum_{i=0}^{\infty}\alpha_i t_i\right),
\end{aligned}
$$

donde las últimas igualdades son consecuencia tanto de lo probado anteriormente para el caso finito como de la continuidad de f. □

Proposición 3.3.4. *Sean $\mathcal{P}$, $\mathcal{Q}$ y $\mathcal{R}$ tres particiones. Entonces*

1. $\mathcal{P} \succeq \mathcal{Q} \Longrightarrow H(\mu, \mathcal{Q}|\mathcal{P}) = 0.$

2. $\mathcal{P} \succeq \mathcal{Q} \Longrightarrow H(\mu, \mathcal{P}|\mathcal{R}) \geq H(\mu, \mathcal{Q}|\mathcal{R}).$

3. $\mathcal{Q} \succeq \mathcal{R} \Longrightarrow H(\mu, \mathcal{P}|\mathcal{Q}) \leq H(\mu, \mathcal{P}|\mathcal{R}).$

Demostración. Para la ida de 1, debido a la Proposición 3.3.2,

$$H(\mathcal{P} \vee \mathcal{Q}) = H(\mathcal{Q}|\mathcal{P}) + H(\mathcal{P}),$$

pero entonces como $\mathcal{P} \succeq \mathcal{Q}$, la Proposición 3.1.3 implica que $\mathcal{P} \vee \mathcal{Q} = \mathcal{P}$, y entonces

$$H(\mathcal{Q}|\mathcal{P}) + H(\mathcal{P}) = H(\mathcal{P} \vee \mathcal{Q}) = H(\mathcal{P}),$$

por lo que $H(\mu, \mathcal{Q}|\mathcal{P}) = 0$.

Ahora para ver 2, ocupando nuevamente la Proposición 3.3.2, vemos que, como $\mathcal{P}$ es más fina que $\mathcal{Q}$, la Proposición 3.1.3 implica que

$$H(\mathcal{P}|\mathcal{R}) = H(\mathcal{P} \vee \mathcal{Q}|\mathcal{R}) = H(\mathcal{P}|\mathcal{Q} \vee \mathcal{R}) + H(\mathcal{Q}|\mathcal{R})$$

y entonces

$$H(\mathcal{P}|\mathcal{R}) - H(\mathcal{Q}|\mathcal{R}) = H(\mathcal{P}|\mathcal{Q} \vee \mathcal{R}) \geq 0,$$

lo cual nos dice que

$$H(\mu, \mathcal{P}|\mathcal{R}) \geq H(\mu, \mathcal{Q}|\mathcal{R}).$$

Para 3, primero lo probamos para la partición que sólo tiene al total. Sea $\mathcal{R}$ dicha partición. Entonces

$$H(\mathcal{P}|\mathcal{R}) = \sum_{B \in \mathcal{R}} \mu(B) H_B(\mathcal{P}) = \mu(X) H_X(\mathcal{P}) = H_X(\mathcal{P}) = H(\mathcal{P})$$

debido a que

$$\mu_X(A) = \frac{\mu(A \cap X)}{\mu(X)} = \mu(A).$$

Ahora probemos que $H(\mathcal{P}|\mathcal{Q}) \leq H(\mathcal{P})$. Por la Proposición 3.2.2 tenemos que

$$\begin{aligned} H(\mathcal{P}|\mathcal{Q}) &= \sum_{A \in \mathcal{P}, B \in \mathcal{Q}, A \cap B \neq \emptyset} -\mu(A \cap B) \log\left(\frac{\mu(A \cap B)}{\mu(B)}\right) \\ &\leq \sum_{A \in \mathcal{P}} (-\mu(A) \log(\mu(A))) \\ &= H(\mathcal{P}) \end{aligned}$$

debido a que como $-t \log t$ es estrictamente cóncava (ésto a raźon de que su segunda derivada es $-1/t$ y, puesto que sólo está definida para $t > 0$, es negativa) y continua. Ocupando el Lema 3.3.1

para $\alpha_B = \mu(B)$ y con $t_B = \dfrac{\mu(A\cap B)}{\mu(B)}$ vemos que

$$\begin{aligned}
-\sum_{B\in\mathcal{Q}} \mu(A\cap B)\log\left(\frac{\mu(A\cap B)}{\mu(B)}\right) &\leq -\sum_{B\in\mathcal{Q}} \mu(A\cap B)\log\left(\sum_{B\in\mathcal{Q}} \mu(A\cap B)\right)\\
&= -\mu\left(\bigcup_{B\in\mathcal{Q}} A\cap B\right)\log\left(\mu\left(\bigcup_{B\in\mathcal{Q}} A\cap B\right)\right)\\
&= -(\mu(A)\log(\mu(A))).
\end{aligned}$$

Con ésto y ocupando las Proposiciones 3.2.2 y 3.1.3 vemos que

$$H(\mathcal{P}|\mathcal{Q}) = H(\mathcal{P}|\mathcal{Q}\vee\mathcal{R}) = \sum_{C\in\mathcal{R}} \mu(C)H_C(\mathcal{P}|\mathcal{Q}) \leq \sum_{C\in\mathcal{R}} \mu(C)H_C(\mathcal{P}) = H(\mathcal{P}|\mathcal{R})$$

lo que termina la prueba. □

Ahora si podemos deducir una proposición en contextos de sucesiones de particiones que se refinan y las σ-álgebras que generan.

Proposición 3.3.5. *Sean $\mathfrak{B}$ una σ-álgebra, $\mathcal{P}$ una partición y $\mathcal{Q}^n$ una sucesión de particiones tales que se refinan sucesivamente (es decir que $\mathcal{Q}^n \preceq \mathcal{Q}^{n+1}$ para toda $n\in\mathbb{N}$) y*

$$\bigvee_{n=1}^{\infty} \mathcal{Q}^n = \mathfrak{B}.$$

Entonces

$$H(\mathcal{P}|\mathfrak{B}) = \lim_{n\to\infty} H(\mathcal{P}|\mathcal{Q}^n).$$

Demostración. Debido a la Proposición 3.3.4, el que los elementos de la sucesión de particiones que se refinen sucesivamente implica que

$$\mathbb{E}[\mathbb{I}_A(x)|\mathcal{Q}^n] \xrightarrow[n\to\infty]{} \mathbb{E}[\mathbb{I}_A(x)|\mathfrak{B}]$$

puesto que $\sigma(\mathcal{Q}^n)$ es una filtración (por el hecho de que $\mathcal{Q}^n$ se refinan) y por la Proposición A.1.3. Obtenemos entonces que

$$\mathcal{I}_{\mathcal{P}|\mathcal{Q}^n} = -\sum_{A\in\mathcal{P}} \log\left(\mathbb{E}[\mathbb{I}_A(x)|\mathcal{Q}^n]\right)\mathbb{I}_A \xrightarrow[n\to\infty]{} -\sum_{A\in\mathcal{P}} \log\left(\mathbb{E}[\mathbb{I}_A(x)|\mathfrak{B}]\right)\mathbb{I}_A = \mathcal{I}_{\mathcal{P}|\mathfrak{B}}.$$

Por la Proposición 3.3.1 tenemos que $\mathcal{I}_{\mathcal{P}|\mathcal{Q}^n}$ es una sucesión no decreciente de funciones no negativas simples que convergen a la función integrable $\mathcal{I}_{\mathcal{P}|\mathfrak{B}}$ y entonces por el Teorema de Convergencia Monótona ([5] pág. 31),

$$\lim_{n\to\infty} H(\mathcal{P}|\mathcal{Q}^n) = \int \mathcal{I}_{\mathcal{P}|\mathfrak{B}}\,d\mu = H(\mathcal{P}|\mathfrak{B}).$$

□

3.4. Entropía dinámica de particiones y particiones de Markov

Definición 3.4.1. Sea $(X, \mathcal{F}, \mu)$ un espacio de probabilidad y $T\colon X \to X$ una función medible con μ una medida T-invariante (es decir $\mu(T^{-1}(A)) = \mu(A)$ siempre que $A \in \mathcal{F}$). Sea $\mathbb{S}$ el semigrupo (o grupo) de iteraciones de T (dependiendo de si T es o no invertible). A la quinteta $(X, \mathcal{F}, \mu, T, \mathbb{S})$ le llamaremos un *sistema dinámico*.

De ahora en adelante trabajaremos con estos espacios, con el pequeño detalle de no mencionar el semigrupo implícito por T.

Lema 3.4.1. *Sea $(a_n)_{n=1}^{\infty}$ una sucesión de números reales con incrementos decrecientes. Entonces*

$$\lim_{n\to\infty} \frac{a_n}{n} = \lim_{n\to\infty} a_n - a_{n-1}.$$

Demostración. Para $a_1 = 0$ tenemos que

$$\lim_{n\to\infty} \frac{a_n}{n} = \lim_{n\to\infty} \frac{\sum_{i=2}^{n} a_i - a_{i-1}}{n} \geq \lim_{n\to\infty} \frac{\sum_{i=1}^{n} a_n - a_{n-1}}{n}$$

(la desigualdad se da porque los incrementos son decrecientes). Luego pues,

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} a_n - a_{n-1}}{n} = \lim_{n\to\infty} \frac{n(a_n - a_{n-1})}{n} = \lim_{n\to\infty} a_n - a_{n-1},$$

lo cual prueba que

$$\lim_{n\to\infty} \frac{a_n}{n} \geq \lim_{n\to\infty} a_n - a_{n-1}.$$

Veamos ahora le desigualdad la recíproca. Como

$$\begin{aligned}
\lim_{n\to\infty} a_n - a_{n-1} &= \lim_{n\to\infty} \frac{n(a_n - a_{n-1})}{n} = \lim_{n\to\infty} \frac{\sum_{i=n}^{2n-1} a_n - a_{n-1}}{n} \\
&\geq \lim_{n\to\infty} \frac{\sum_{i=n}^{2n-1} a_i - a_{i-1}}{n} = \lim_{n\to\infty} \frac{a_{2n-1} - a_{n-1}}{n}
\end{aligned}$$

(los dos últimos pasos anteriores son consecuencia de que a_n tiene incrementos decrecientes y del hecho de tener una suma telescópica), obtenemos que

$$\lim_{n\to\infty} \frac{a_{2n-1} - a_{n-1}}{n} = \lim_{n\to\infty} \frac{a_{2n-1}}{n} - \lim_{n\to\infty} \frac{a_{n-1}}{n} = 2 \lim_{n\to\infty} \frac{a_{2n-1}}{2n-1} - \lim_{n\to\infty} \frac{a_{n-1}}{n-1}$$

debido a que

$$\lim_{n\to\infty} \frac{2n}{2n-1} = 1 \quad \text{y} \quad \lim_{n\to\infty} \frac{n}{n-1} = 1$$

y al multiplicar ese límite con la penúltima igualdad nos resulta lo final. Ahora como toda subsucesión de una sucesión convergente converge al mismo límite que la original,

$$2\lim_{n\to\infty}\frac{a_{2n-1}}{2n-1}-\lim_{n\to\infty}\frac{a_{n-1}}{n-1}=2\lim_{n\to\infty}\frac{a_n}{n}-\lim_{n\to\infty}\frac{a_n}{n}=\lim_{n\to\infty}\frac{a_n}{n},$$

con lo cual concluye se concluye que $\lim_{n\to\infty}\frac{a_n}{n}\leq\lim_{n\to\infty}a_n-a_{n-1}$, lo que termina con la prueba. □

Definición 3.4.2. Sea $\mathcal{P}$ una partición. Para cada $n\in\mathbb{N}$, se define

$$\mathcal{P}^n=\bigvee_{i=0}^{n-1}T^{-i}(\mathcal{P}),$$

Si $\mathbb{D}\subset\mathbb{Z}$ es finito, definimos

$$\mathcal{P}^{\mathbb{D}}=\bigvee_{i\in\mathbb{D}}T^{-i}(\mathcal{P}).$$

Si $\mathbb{D}$ es infinito, definimos a $\mathcal{P}^{\mathbb{D}}$ como la σ-álgebra más pequeña que contiene a

$$(T^{-i}(\mathcal{P}))_{i\in\mathbb{D}},$$

es decir,

$$\mathcal{P}^{\mathbb{D}}=\sigma((T^{-i}(\mathcal{P}))_{i\in\mathbb{D}}).$$

(por ejemplo, $\mathcal{P}^{[0,n-1]}=\mathcal{P}^n$). Escribiremos también $\mathcal{P}^{-n}=\mathcal{P}^{[-n,-1]}$ y de una manera similar, $\mathcal{P}^{[1,\infty)}=\mathcal{P}^+$ y $\mathcal{P}^{(-\infty,-1]}=\mathcal{P}^-$, los cuales se llaman el *n-presente*, el *n-pasado*, el *futuro completo* y el *pasado completo del proceso*.

Claramente, si $\mathbb{D}$ es finito, entoces $\mathcal{P}^{\mathbb{D}}$ es también una partición.

Proposición 3.4.1. *Sean $\mathcal{P}$ una partición y $\mathfrak{B}$ una σ-álgebra T-invariante, es decir $T^{-1}(\mathfrak{B})=\mathfrak{B}$. Entonces $H(\mathcal{P}^n|\mathfrak{B})$ tiene incrementos decrecientes.*

Demostración. Tenemos que

$$\begin{aligned}H(\mathcal{P}^n|\mathfrak{B})-H(\mathcal{P}^{n-1}|\mathfrak{B}) &= H(\bigvee_{i=0}^{n-1}T^{-i}(\mathcal{P})|\mathfrak{B})-H(\mathcal{P}^{n-1}|\mathfrak{B})\\ &= H(T^{-(n-1)}(\mathcal{P})\vee\mathcal{P}^{n-1}|\mathfrak{B})-H(\mathcal{P}^{n-1}|\mathfrak{B})\\ &= H(T^{-(n-1)}(\mathcal{P})|\mathfrak{B}\vee\mathcal{P}^{n-1})\\ &= H(T^{-n+1}(\mathcal{P})|\mathfrak{B}\vee\mathcal{P}^{n-1}),\end{aligned}$$

donde la penúltima igualdad se sigue de la Proposición 3.3.3. De la misma manera se ve que

$$H(\mathcal{P}^{n+1}|\mathfrak{B})-H(\mathcal{P}^n|\mathfrak{B})=H(T^{-n}(\mathcal{P})|\mathfrak{B}\vee\mathcal{P}^n).$$

Ahora nos fijamos en que

$$
\begin{aligned}
H(T^{-n+1}(P)|\mathfrak{B} \vee \mathcal{P}^{n-1}) &= H(T^{-n+1}(P)|\mathfrak{B} \vee (\bigvee_{i=0}^{n-2} T^{-i}(\mathcal{P}))) \\
&= H(T^{-n}(P)|T^{-1}(\mathfrak{B}) \vee (\bigvee_{i=1}^{n-1} T^{-i}(\mathcal{P}))) \\
&= H(T^{-n}(P)|T^{-1}(\mathfrak{B}) \vee (\mathcal{P}^{[1,n-1]})),
\end{aligned}
$$

donde la penúltima igualdad se da debido a que la entropía se calcula en base a las medidas de la partición, a causa de la T-invarianza de la medida y de $\mathfrak{B}$, por lo que pudimos aplicar T^{-1} sin que se afectara la entropía, y como $\mathfrak{B} \bigvee \mathcal{P}^n \succeq T^{-1}(\mathfrak{B}) \bigvee \mathcal{P}^{[1,n-1]}$ concluimos que

$$
\begin{aligned}
H(\mathcal{P}^n|\mathfrak{B}) - H(\mathcal{P}^{n-1}|\mathfrak{B}) &= H(T^{-n}(P)|T^{-1}(\mathfrak{B}) \vee (\mathcal{P}^{[1,n-1]})) \\
&\leq H(T^{-n}(\mathcal{P})|\mathfrak{B} \vee \mathcal{P}^n) \\
&= H(\mathcal{P}^{n+1}|\mathfrak{B}) - H(\mathcal{P}^n|\mathfrak{B}),
\end{aligned}
$$

lo cual concluye con la prueba. □

Definición 3.4.3. Definimos la *entropía dinámica de una partición $\mathcal{P}$ con respecto a una transformacion T y una medida μ* como

$$h(\mu, T, \mathcal{P}) = h(T, \mathcal{P}) = h(\mathcal{P}) = \lim_{n\to\infty} \frac{H(\mathcal{P}^n)}{n}.$$

Ahora si $\mathfrak{B}$ es una σ-álgebra T-invariante y $\mathcal{Q}$ una partición, entonces definimos las siguientes entropías dinámicas:

1. $$h(\mu, T, \mathcal{P}|\mathfrak{B}) = h(T, \mathcal{P}|\mathfrak{B}) = h(\mathcal{P}|\mathfrak{B}) = \lim_{n\to\infty} \frac{H(\mathcal{P}^n|\mathfrak{B})}{n}.$$

2. $$h(\mu, T, \mathcal{P}|\mathcal{Q}) = h(T, \mathcal{P}|\mathcal{Q}) = h(\mathcal{P}|\mathcal{Q}) = \lim_{n\to\infty} \frac{H(\mathcal{P}^n|\mathcal{Q}^n)}{n}.$$

3. $$h(\mu, T, \mathcal{P}|\mathcal{Q} \bigvee \mathfrak{B}) = h(T, \mathcal{P}|\mathcal{Q} \vee \mathfrak{B}) = h(\mathcal{P}|\mathcal{Q} \vee \mathfrak{B}) = \lim_{n\to\infty} \frac{H(\mathcal{P}^n|\mathcal{Q}^n \vee \mathfrak{B})}{n}.$$

Proposición 3.4.2. *Si $h(\mathcal{P}|\mathfrak{B}) < \infty$, entonces $h(\mu, T, \mathcal{P}|\mathfrak{B}) = H(\mu, \mathcal{P}|\mathfrak{B} \bigvee \mathcal{P}^+)$, en particular $h(\mu, T, \mathcal{P}) = H(\mu, \mathcal{P}|\mathcal{P}^+)$ (es el resultado que corresponde a la σ-álgebra trivial).*

Demostración. A causa de que $\mathfrak{B}$ y μ son T-invariantes, tenemos que

$$H(\mathcal{P}^{n+1}|\mathfrak{B}) - H(\mathcal{P}^n|\mathfrak{B}) = H(\mathcal{P}^{n+1}|\mathfrak{B}) - H(T^{-1}(\mathcal{P}^n)|\mathfrak{B}) = H(\mathcal{P}|\mathfrak{B} \vee \mathcal{P}^{[1,n]})$$

ya que con la Proposición 3.3.3 vemos que

$$H(\mathcal{P}^{n+1}|\mathfrak{B}) = H(\mathcal{P}|\mathfrak{B} \vee \mathcal{P}^{[1,n]}) + H(\mathcal{P}^{[1,n]}|\mathfrak{B}) = H(\mathcal{P}|\mathfrak{B} \vee \mathcal{P}^{[1,n]}) + H(T^{-1}(\mathcal{P}^n)|\mathfrak{B})$$

y despejando se obtiene lo deseado. Como $\mathcal{P}^n$ tiene incrementos decrecientes, la Proposición 3.4.1 y luego el Lema 3.4.1 implican que

$$\begin{aligned} h(\mathcal{P}|\mathfrak{B}) &= \lim_{n\to\infty} \frac{H(\mathcal{P}^n|\mathfrak{B})}{n} \\ &= \lim_{n\to\infty} H(\mathcal{P}^{n+1}|\mathfrak{B}) - H(\mathcal{P}^n|\mathfrak{B}) \\ &= \lim_{n\to\infty} H(\mathcal{P}|\mathfrak{B} \vee \mathcal{P}^{[1,n]}) \\ &= H(\mathcal{P}|\mathfrak{B} \vee \mathcal{P}^{+}), \end{aligned}$$

donde la última igualdad es consecuencia de la Proposición 3.3.5, lo cual concluye la prueba. □

A continuación agregaremos un pequeño auxiliar de nuestro trabajo, algo que por el momento no ocuparemos, pero que más adelante servirá para conectar los capítulos 3 y 4.

Definición 3.4.4. Sean X un espacio métrico compacto, $T\colon X \to X$ una función continua y $\mathbb{S}$ el semigrupo o grupo de iteraciones de T. A la terna $(X, T, \mathbb{S})$ le llamaremos un *sistema dinámico topológico*.

Dado un sistema dinámico topológico, podemos considerar la σ-álgebra que generan los conjuntos abiertos, llamada la *σ-álgebra de Borel* (o *σ-álgebra boreliana*). La Proposición A.2.8 garantiza la existencia de una medida de probabilidad T-ergódica y T-invariante, con la cual se obtiene entonces tenemos un sistema dinámico como se definió anteriormente.

Definición 3.4.5. Sea $(X, T, \mathbb{S})$ un sistema dinámico topológico, μ una medida T-invariante y $\mathcal{P} = \{A_1, A_2, \dots\}$ una partición de X. Decimos que $\mathcal{P}$ es una *partición de Markov* si

$$\frac{\mu(A_{x_0} \cap T^{-1}A_{x_1} \cap T^{-2}A_{x_2} \cap T^{-3}A_{x_3} \cdots \cap T^{-n}A_{x_n})}{\mu(A_{x_0} \cap T^{-1}A_{x_1} \cap T^{-2}A_{x_2} \cap T^{-3}A_{x_3} \cdots \cap T^{-(n-1)}A_{x_{n-1}})} = \frac{\mu(A_{x_{n-1}} \cap T^{-1}A_{x_n})}{\mu(A_{x_{n-1}})}$$

para cualquier conjunto finito de índices $(x_i)_{i=0}^n$ (es decir, $x_i \in \{1, 2, \dots\}$) y cualquier familia $(A_{x_i})_{i=0}^n$ de elementos de la partición.

Proposición 3.4.3. *Sea $(X, T, \mathbb{S})$ un sistema dinámico topológico, μ un medida de probabilidad T-invariante y $\mathcal{P}$ una partición de X. Entonces $\mathcal{P}$ es una partición de Markov si y sólo si*

$$\frac{\mu(A_{x_0} \cap T^{-1}A_{x_1} \cap T^{-2}A_{x_2} \cap T^{-3}A_{x_3} \cdots \cap T^{-n}A_{x_n})}{\mu(T^{-1}A_{x_1} \cap T^{-2}A_{x_2} \cap T^{-3}A_{x_3} \cdots \cap T^{-n}A_{x_n})} = \frac{\mu(T^{-1}A_{x_1} \cap A_{x_0})}{\mu(T^{-1}A_{x_1})}$$

para cualquier conjunto finito de índices $(x_i)_{i=0}^n$ y cualquier familia $(A_{x_i})_{i=0}^n$ de elementos de la partición.

Demostración. Tomemos un conjunto finito de índices $(x_i)_{i=0}^n$ y consideramos la familia correspondiente $(A_{x_i})_{i=0}^n$ de elementos de la partición $\mathcal{P}$. Usando la definición de probabilidad condicional

$$\mu(A|B) = \frac{\mu(A \cap B)}{\mu(B)}$$

vemos que, dada una i fija,

$$\begin{aligned}
\mu(T^{-i}A_{x_i}|T^{-(i-1)}A_{x_{i-1}}) &= \frac{\mu(T^{-i}A_{x_i} \cap T^{-(i-1)}A_{x_{i-1}})}{\mu(T^{-(i-1)}A_{x_{i-1}})} \\
&= \frac{\mu(T^{-(i-1)}(T^{-1}A_{x_i} \cap A_{x_{i-1}}))}{\mu(T^{-(i-1)}A_{x_{i-1}})} \\
&= \frac{\mu(T^{-1}A_{x_i} \cap A_{x_{i-1}})}{\mu(A_{x_{i-1}})},
\end{aligned}$$

donde la última igualdad se debe a la T-invarianza de μ. Como la partición es de Markov, tenemos que

$$\frac{\mu(T^{-1}A_{x_i} \cap A_{x_{i-1}})}{\mu(A_{x_{i-1}})} = \frac{\mu(A_{x_0} \cap T^{-1}A_{x_1} \cap \cdots \cap T^{-i}A_{x_i})}{\mu(A_{x_0} \cap T^{-1}A_{x_1} \cap \cdots \cap T^{-(i-1)}A_{x_{i-1}})}$$

y a causa de esto obtenemos que

$$\begin{gathered}
\mu(A_{x_0} \cap T^{-1}A_{x_1} \cap \cdots \cap T^{-i}A_{x_i}) \\
= \frac{\mu(T^{-i}A_{x_i} \cap T^{-(i-1)}A_{x_{i-1}})}{\mu(T^{-(i-1)}A_{x_{i-1}})}\mu(A_{x_0} \cap T^{-1}A_{x_1} \cap \cdots \cap T^{-(i-1)}A_{x_{i-1}}) \\
= \mu(T^{-i}A_{x_i}|T^{-(i-1)}A_{x_{i-1}})\mu(A_{x_0} \cap T^{-1}A_{x_1} \cap \cdots \cap T^{-(i-1)}A_{x_{i-1}}).
\end{gathered}$$

Deducimos entonces que

$$\begin{gathered}
\mu(A_{x_0} \cap T^{-1}A_{x_1} \cap \cdots \cap T^{-l}A_{x_l}) \\
= \mu(T^{-l}A_{x_l}|T^{-(l-1)}A_{x_{l-1}})\mu(T^{-(l-1)}A_{x_{l-1}}|T^{-(l-2)}A_{x_{l-2}}) \cdots \mu(T^{-1}A_{x_1} \cap A_{x_0}) \\
= \mu(T^{-l}A_{x_l}|T^{-(l-1)}A_{x_{l-1}})\mu(T^{-(l-1)}A_{x_{l-1}}|T^{-(l-2)}A_{x_{l-2}}) \cdots \mu(T^{-1}A_{x_1}|A_{x_0})\mu(A_{x_0})
\end{gathered}$$

y que

$$\begin{gathered}
\mu(T^{-1}A_{x_1} \cap T^{-2}A_{x_2} \cap T^{-3}A_{x_3} \cdots \cap T^{-l}A_{x_l}) = \\
= \mu(T^{-l}A_{x_l}|T^{-(l-1)}A_{x_{l-1}})\mu(T^{-(l-1)}A_{x_{l-1}}|T^{-(l-2)}A_{x_{l-2}}) \cdots \mu(T^{-2}A_{x_2} \cap T^{-1}A_{x_1}) \\
= \mu(T^{-l}A_{x_l}|T^{-(l-1)}A_{x_{l-1}})\mu(T^{-(l-1)}A_{x_{l-1}}|T^{-(l-2)}A_{x_{l-2}}) \cdots \mu(T^{-2}A_{x_2}|T^{-1}A_{x_1})\mu(T^{-1}A_{x_1})
\end{gathered}$$

y como consecuencia de ésto se puede observar que

$$\frac{\mu(A_{x_0} \cap T^{-1}A_{x_1} \cap T^{-2}A_{x_2} \cap T^{-3}A_{x_3} \cap \cdots \cap T^{-n}A_{x_n})}{\mu(T^{-1}A_{x_1} \cap T^{-2}A_{x_2} \cap T^{-3}A_{x_3} \cap \cdots \cap T^{-n}A_{x_n})}$$

$$= \frac{\mu(T^{-n}A_{x_n}|T^{-(n-1)}A_{x_{n-1}})\cdots\mu(T^{-2}A_{x_2}|T^{-1}A_{x_1})\mu(T^{-1}A_{x_1}|A_{x_0})\mu(A_{x_0})}{\mu(T^{-n}A_{x_n}|T^{-(n-1)}A_{x_{n-1}})\cdots\mu(T^{-2}A_{x_2}|T^{-1}A_{x_1})\mu(T^{-1}A_{x_1})}$$

$$= \frac{\mu(T^{-1}A_{x_1}|A_{x_0})\mu(A_{x_0})}{\mu(T^{-1}A_{x_1})} = \frac{\mu(T^{-1}A_{x_1}\cap A_{x_0})\mu(A_{x_0})}{\mu(T^{-1}A_{x_1})\mu(A_{x_0})} = \frac{\mu(T^{-1}A_{x_1}\cap A_{x_0})}{\mu(T^{-1}A_{x_1})}$$

con lo cual se concluye la ida de la proposición.

Ahora para el regreso, por hipótesis tenemos que

$$\mu(A_{x_0}\cap T^{-1}A_{x_1}\cap\cdots\cap T^{-i}A_{x_i}) = \frac{\mu(A_{x_0}\cap T^{-1}A_{x_1})}{\mu(T^{-1}A_{x_1})}\mu(T^{-1}A_{x_1}\cap T^{-2}A_{x_2}\cap\cdots\cap T^{-i}A_{x_i}).$$

Aplicando la T-invarianza de μ podemos ver que se cumple un poco más que esta igualdad, es decir, la igualdad se conserva sin importar la familia de subíndices, más precisamente tenemos que

$$\mu(A_{y_0}\cap T^{-1}A_{y_1}\cap\cdots\cap T^{-i}A_{y_i}) = \frac{\mu(A_{y_0}\cap T^{-1}A_{y_1})}{\mu(T^{-1}A_{y_1})}\mu(A_{y_1}\cap T^{-1}A_{y_2}\cap\cdots\cap T^{-(i-1)}A_{y_{i-1}}).$$

Aplicando este paso $n-1$ veces para $\mu(A_{x_0}\cap T^{-1}A_{x_1}\cap\cdots\cap T^{-n}A_{x_n})$ obtenemos que

$$\mu(A_{x_0}\cap T^{-1}A_{x_1}\cap\cdots\cap T^{-n}A_{x_n})$$

$$= \frac{\mu(A_{x_0}\cap T^{-1}A_{x_1})}{\mu(T^{-1}A_{x_1})}\cdots\frac{\mu(A_{x_{n-2}}\cap T^{-1}A_{x_{n-1}})}{\mu(T^{-1}A_{x_{n-1}})}\mu(A_{x_{n-1}}\cap T^{-1}A_{x_n})$$

y de manera similar vemos

$$\mu(A_{x_0}\cap T^{-1}A_{x_1}\cap\cdots\cap T^{-(n-1)}A_{x_{(n-1)}})$$

$$= \frac{\mu(A_{x_0}\cap T^{-1}A_{x_1})}{\mu(T^{-1}A_{x_1})}\cdots\frac{\mu(A_{x_{n-3}}\cap T^{-1}A_{x_{n-2}})}{\mu(T^{-1}A_{x_{n-2}})}\cdot\frac{\mu(A_{x_{n-2}}\cap T^{-1}A_{x_{n-1}})}{\mu(T^{-1}A_{x_{n-1}})}\mu(T^{-1}A_{x_{n-1}}),$$

con lo cual

$$\frac{\mu(A_{x_0}\cap T^{-1}A_{x_1}\cap T^{-2}A_{x_2}\cap T^{-3}A_{x_3}\cdots\cap T^{-n}A_{x_n})}{\mu(A_{x_0}\cap T^{-1}A_{x_1}\cap T^{-2}A_{x_2}\cap T^{-3}A_{x_3}\cdots\cap T^{-(n-1)}A_{x_{n-1}})}$$

$$= \frac{\frac{\mu(A_{x_0}\cap T^{-1}A_{x_1})}{\mu(T^{-1}A_{x_1})}\cdots\frac{\mu(A_{x_{n-2}}\cap T^{-1}A_{x_{n-1}})}{\mu(T^{-1}A_{x_{n-1}})}\mu(A_{x_{n-1}}\cap T^{-1}A_{x_n})}{\frac{\mu(A_{x_0}\cap T^{-1}A_{x_1})}{\mu(T^{-1}A_{x_1})}\cdots\frac{\mu(A_{x_{n-3}}\cap T^{-1}A_{x_{n-2}})}{\mu(T^{-1}A_{x_{n-2}})}\frac{\mu(A_{x_{n-2}}\cap T^{-1}A_{x_{n-1}})}{\mu(T^{-1}A_{x_{n-1}})}\mu(T^{-1}A_{x_{n-1}})}$$

$$= \frac{\mu(A_{x_{n-1}}\cap T^{-1}A_{x_n})}{\mu(T^{-1}A_{x_{n-1}})},$$

por lo tanto $\mathcal{P}$ es una partición de Markov lo cual concluye la prueba. □

Definición 3.4.6. Decimos que dos particiones $\mathcal{P}$ y $\mathcal{Q}$ son independientes si $\forall A\in\mathcal{P}$ y $\forall B\in\mathcal{Q}$ $\mu(A\cap B)=\mu(A)\mu(B)$, o que es lo mismo $\mu(A|B)=\dfrac{\mu(A\cap B)}{\mu(B)}=\mu(A)$.

Lema 3.4.2. *Sean α,β,γ particiones finitas. Entonces:*

1. *α y β son independientes si y sólo si $H(\alpha|\beta) = H(\alpha)$.*
2. *$\forall\, A \in \alpha, B \in \beta, C \in \gamma$,*
$$\frac{\mu(A \cap B \cap C)}{\mu(B \cap C)} = \frac{\mu(A \cap C)}{\mu(C)}$$
si y sólo si $H(\alpha|\beta \vee \gamma) = H(\alpha|\gamma)$.

Demostración. Para probar 1 la ida es inmediata, puesto que si $\mu(A|B) = \mu(A)$ entonces

$$\begin{aligned} H(\alpha|\beta) &= -\sum_{A\in\alpha, B\in\beta} \mu(A \cap B) \log(\mu(A|B)) \\ &= -\sum_{A\in\alpha, B\in\beta} \mu(A \cap B) \log(\mu(A)) \\ &= -\sum_{A\in\alpha} \mu(A) \log(\mu(A)) \\ &= H(\alpha)\,. \end{aligned}$$

Para el regreso, como la función $-t\log(t)$ es estrictamente cóncava, vemos que para cada $A \in \alpha$, ocupando el Lema 3.3.1 para $\alpha_B = \mu(B)$ y con $t_B = \frac{\mu(A\cap B)}{\mu(B)}$, tomando como conjunto de subíndices a β, vemos que

$$\begin{aligned} -\sum_{B\in\beta} \mu(A \cap B) \log\left(\frac{\mu(A\cap B)}{\mu(B)}\right) &\leq -\sum_{B\in\beta} \mu(A \cap B) \log(\sum_{B\in\beta} \mu(A \cap B)) \\ &= -\mu(\bigcup_{B\in\beta} A \cap B) \log(\mu(\bigcup_{B\in\beta} A \cap B)) \qquad (3.6) \\ &= -\mu(A) \log(\mu(A)). \end{aligned}$$

Ahora sí, suponemos que β no es la partición trivial (donde sólo te fijas de un conjunto de medida 1, en cuyo caso la afirmación que queremos demostrar es un caso trivial). Tenemos que $\forall B \in \beta\ \alpha_B = \mu(B) < 1$. Como $H(\alpha|\beta) = H(\alpha)$, tenemos que

$$-\sum_{A\in\alpha, B\in\beta} \mu(A \cap B) \log\left(\frac{\mu(A\cap B)}{\mu(B)}\right) = -\sum_{A\in\alpha} \mu(A) \log(\mu(A)),$$

entonces

$$-\sum_{A\in\alpha}\left(\mu(A)\log(\mu(A)) - \sum_{B\in\beta} \mu(A \cap B) \log\left(\frac{\mu(A\cap B)}{\mu(B)}\right)\right) = 0,$$

de aquí obtenemos, por (3.6), una suma de números no negativos que suman cero, por lo que cada sumando es cero, que es exactamente lo mismo que

$$\mu(A)\log(\mu(A)) - \sum_{B\in\beta} \mu(A \cap B) \log\left(\frac{\mu(A\cap B)}{\mu B)}\right) = 0$$

y así

$$-\mu(A)\log(\mu(A)) = -\sum_{B\in\beta}\mu(A\cap B)\log\left(\frac{\mu(A\cap B)}{\mu(B)}\right).$$

Ahora como sabemos que $-t\log(t)$ es estrictamente cóncava, si $t_B = \frac{\mu(A\cap B)}{\mu(B)}$ no fuera constante en el conjunto de índices donde $\alpha_B \neq 0$, entonces tendríamos la desigualdad estricta, pero esto traería consigo una contradicción por lo cual $t_B = \frac{\mu(A\cap B)}{\mu(B)}$ es constante para toda $B \in \beta$, llamémosle c a dicha constante. Entonces

$$\begin{aligned}
-\mu(A)\log(\mu(A)) &= -\sum_{B\in\beta}\mu(A\cap B)\log\left(\frac{\mu(A\cap B)}{\mu(B)}\right)\\
&= -\sum_{B\in\beta}\mu(A\cap B)\log(c)\\
&= \log(c)(-\sum_{B\in\beta}\mu(A\cap B))\\
&= \log(c)(-\mu(\bigcup_{B\in Q} A\cap B))\\
&= \mu(A)\log(c),
\end{aligned}$$

lo que nos dice que

$$\mu(A) = \log(c) = \log\left(\frac{\mu(A\cap B)}{\mu(B)}\right)$$

y de aquí concluimos que

$$\mu(A) = \left(\frac{\mu(A\cap B)}{\mu(B)}\right).$$

Para 2, la ida es de hecho similar a la ida del primero. Para el regreso, tenemos que $H(\alpha|\beta\vee\gamma) = H(\alpha|\gamma)$ y usando la Proposición 3.3.2 vemos que

$$H(\alpha|\beta\vee\gamma) = \sum_{C\in\gamma}\mu(C)H_C(\alpha|\beta) = H(\alpha|\gamma) = \sum_{C\in\gamma}\mu(C)H_C(\alpha),$$

donde la última igualdad se sigue de la Proposicion 3.2.2. Entonces

$$\sum_{C\in\gamma}\mu(C)(H_C(\alpha|\beta) - H_C(\alpha)) = 0.$$

Si fuera γ la partición con un único elemento de medida 1, el resultado sería obvio. Entonces suponiendo el hecho opuesto, es decir, suponiendo, sin pérdida de generalidad, que $\mu(C) \neq 0\ \forall C \in \gamma$, entonces tenemos $H_C(\alpha|\beta) - H_C(\alpha) = 0$, es decir $H_C(\alpha|\beta) = H_C(\alpha)$, que es el caso anterior para la medida μ_C, por lo que ocurre que α es independiente de β bajo la medida μ_C, y entonces $\mu_C(A|B) = \mu_C(A)$, o lo que es lo mismo,

$$\frac{\mu_C(A\cap B)}{\mu_C(B)} = \mu_C(A),$$

lo cual es equivalente a

$$\frac{\frac{\mu(A\cap B\cap C)}{\mu(C)}}{\frac{\mu(B\cap C)}{\mu(C)}} = \frac{\mu(A\cap C)}{\mu(C)},$$

y por lo tanto

$$\frac{\mu(A\cap B\cap C)}{\mu(B\cap C)} = \frac{\mu(A\cap C)}{\mu(C)},$$

lo que concluye con la prueba. □

Teorema 3.4.1. *Sea $(X, T, \mathbb{S})$ un sistema dinámico topológico y μ una medida T-invariante. Una partición finita $\mathcal{P}$ es de Markov si y sólo si $h(\mu, T, \mathcal{P}) = H(\mathcal{P}|T^{-1}(\mathcal{P}))$.*

Demostración. Basta probar que $\mathcal{P}$ es partición de Markov si y sólo si

$$H(\mathcal{P}|T^{-1}(\mathcal{P})) = H(\mathcal{P}|\mathcal{P}^{+}) = H(\mathcal{P}|\bigvee_{i=1}^{\infty} T^{-i}(\mathcal{P}))$$

(ésto debido a la Proposición 3.4.2). Para la primera implicación vemos que

$$H\left(\mathcal{P}\middle|\bigvee_{i=1}^{\infty} T^{-i}(\mathcal{P})\right) = \lim_{n\to\infty} H\left(\mathcal{P}\middle|\bigvee_{i=1}^{n-1} T^{-i}(\mathcal{P})\right)$$

porque la sucesión de particiones

$$\left(\bigvee_{i=1}^{n-1} T^{-i}(\mathcal{P})\right)_{n\in\mathbb{N}}$$

es refinada y entonces es consecuencia de la Proposición 3.3.5. Luego

$$H\left(\mathcal{P}\middle|\bigvee_{i=1}^{n-1} T^{-i}(\mathcal{P})\right) = -\sum_{\substack{A_{x_0}\in\mathcal{P}\\ B\in\bigvee_{i=1}^{n-1} T^{-i}(\mathcal{P})}} \mu(A_{x_0}\cap B)\log\left(\frac{\mu(A_{x_0}\cap B)}{\mu(B)}\right)$$

$$= -\sum_{\substack{A_{x_0}\in\mathcal{P}\\ A_{x_i}\in\mathcal{P},\ i=1,\ldots,n-1}} \mu\left(A_{x_0}\cap\bigcap_{i=1}^{n-1} T^{-i}(A_{x_i})\right)\log\left(\frac{\mu\left(A_{x_0}\cap\bigcap_{i=1}^{n-1} T^{-i}(A_{x_i})\right)}{\mu\left(\bigcap_{i=1}^{n-1} T^{-i}(A_{x_i})\right)}\right),$$

y siendo $\mathcal{P}$ una partición de Markov, usamos Proposición 3.4.3 y obtenemos que

$$H\left(\mathcal{P}\middle|\bigvee_{i=1}^{n-1} T^{-i}(\mathcal{P})\right)$$

$$= - \sum_{\substack{A_{x_0} \in \mathcal{P} \\ A_{x_i} \in \mathcal{P},\ i=1,\dots,n-1}} \mu\left(A_{x_0} \cap \bigcap_{i=1}^{n-1} T^{-i}(A_{x_i})\right) \log\left(\frac{\mu(A_{x_0} \cap T^{-1}(A_{x_1}))}{\mu(T^{-1}(A_{x_1}))}\right)$$

$$= - \sum_{A_{x_0} \in \mathcal{P}, A_{x_1} \in \mathcal{P}} \mu\left(A_{x_0} \cap T^{-1}(A_{x_1})\right) \log\left(\frac{\mu(A_{x_0} \cap T^{-1}(A_{x_1}))}{\mu(T^{-1}(A_{x_1}))}\right)$$

$$= H\left(\mathcal{P}|T^{-1}(\mathcal{P})\right),$$

lo cual concluye la ida.

Ahora, para demostrar lo que falta, como

$$H\left(\mathcal{P}\left|\bigvee_{i=1}^{n-1} T^{-i}(\mathcal{P})\right.\right)_{n\in\mathbb{N}}$$

es una sucesión decreciente por la Proposición 3.3.4, tenemos que

$$H\left(\mathcal{P}|T^{-1}(\mathcal{P})\right) = h(\mu, T, \mathcal{P}) = \lim_{n\to\infty} H\left(\mathcal{P}\left|\bigvee_{i=1}^{n-1} T^{-i}(\mathcal{P})\right.\right) = \inf_{n\in\mathbb{N}} H\left(\mathcal{P}\left|\bigvee_{i=1}^{n-1} T^{-i}(\mathcal{P})\right.\right).$$

Por otro lado,

$$H\left(\mathcal{P}|T^{-1}(\mathcal{P})\right) \geq H\left(\mathcal{P}\left|\bigvee_{i=1}^{n-1} T^{-i}(\mathcal{P})\right.\right)$$

por la Proposición 3.3.4 (ya que

$$\left(\bigvee_{i=1}^{n-1} T^{-i}(\mathcal{P})\right)_{n\in\mathbb{N}}$$

es una sucesión de particiones que se refina). Con la desigualdad opuesta, concluimos que la sucesión

$$H\left(\mathcal{P}\left|\bigvee_{i=1}^{n-1} T^{-i}(\mathcal{P})\right.\right)_{n\in\mathbb{N}}$$

es constante, lo cual nos dice que $\forall\, n \in \mathbb{N}$,

$$H\left(\mathcal{P}\left|\bigvee_{i=1}^{n-1} T^{-i}(\mathcal{P})\right.\right) = H\left(\mathcal{P}\left|T^{-1}(\mathcal{P})\right.\right)$$

y entonces

$$H\left(\mathcal{P}\left|\bigvee_{i=2}^{n-1} T^{-i}(\mathcal{P}) \bigvee T^{-1}(\mathcal{P})\right.\right) = H\left(\mathcal{P}|T^{-1}(\mathcal{P})\right).$$

Por el Lema 3.4.2,

$$\frac{\mu(A \cap B \cap C)}{\mu(B \cap C)} = \frac{\mu(A \cap C)}{\mu(C)} \quad \forall\, A \in \mathcal{P},\ \forall\, B \in \bigvee_{i=2}^{n-1} T^{-i}(\mathcal{P}),\ \forall\, C \in T^{-1}(\mathcal{P})$$

y por la forma de dichas particiones esto implica que

$$\frac{\mu(A_{x_0} \cap T^{-1}A_{x_1} \cap T^{-2}A_{x_2} \cap T^{-3}A_{x_3} \cdots \cap T^{-n}A_{x_n})}{\mu(T^{-1}A_{x_1} \cap T^{-2}A_{x_2} \cap T^{-3}A_{x_3} \cdots \cap T^{-n}A_{x_n})} = \frac{\mu(T^{-1}A_{x_1} \cap A_{x_0})}{\mu(T^{-1}A_{x_1})}$$

para cualquier conjunto finito de índices $(x_i)_{i=0}^n$ y cualquier familia $(A_{x_i})_{i=0}^n$ de elementos de la partición. Por lo tanto $\mathcal{P}$ es una partición de Markov. □

3.5. Teorema ergódico

En esta sección enunciaremos y demostraremos el Teorema Ergódico de Birkhoff, un resultado de gran importancia en Teoría Ergódica, y bueno, también en el desarrollo de esta tesis. Primero enunciamos una versión simplificada del concepto de ergodicidad.

Definición 3.5.1. Sea $(X, \mathcal{F}, \mu)$ un espacio de probabilidad y $T\colon X \to X$ una transformación medible que preserva la medida μ (i.e. μ es T-invariante). Decimos que T es *ergódica* si para cada $B \in \mathcal{F}$ tal que $T^{-1}(B) = B$ se tiene que $\mu(B) = 0$ o $\mu(B) = 1$ (en cuyo caso diremos que μ es *T-ergódica*)

Proposición 3.5.1. *Sea $(X, \mathcal{F}, \mu)$ un espacio de probabilidad y $T\colon X \to X$ una transformación medible que preserva la medida μ. Entonces T es ergódica respecto de μ si y sólo si para toda $f \in L^1(X, \mathcal{F}, \mu)$ tal que $f = f \circ T$, ocurre que f es constante.*

Demostración. Para la implicación inversa tomamos un $A \in \mathcal{F}$, con $T^{-1}(A) = A$. Entonces $\mathbb{I}_A \circ T = \mathbb{I}_A$, y así, por la hipótesis, $\mathbb{I}_A$ es constante, lo que quiere decir que $\mu(A) = 0$ ó $\mu(A) = 1$.

Ahora para la implicación directa, sea $f \in L^1(X, \mathcal{F}, \mu)$ tal que $f = f \circ T$ y veamos que f es necesariamente constante. Supongamos lo contrario. Entonces existe $\alpha \in \mathbb{R}$ tal que los conjuntos

$$A = \{w \in X : f(w) < \alpha\} \quad \text{y} \quad A^c = \{w \in X : f(w) \geq \alpha\}$$

tienen ambos medida no cero y también tenemos que

$$A = \{w \in X : f(w) < \alpha\} = \{w \in X : f \circ T(w) < \alpha\} = \{w \in X : T(w) \in A\} = T^{-1}(A),$$

lo que directamente nos lleva a una contradicción debido a que $T^{-1}(A) = A$, más sin embargo $\mu(A) > 0$, y como $\mu(A^c) > 0$ entonces $\mu(A) < 1$, pero T es ergódica, por lo tanto f no tiene otra opción que ser constante, lo cual termina esta prueba. □

Podemos ahora enunciar una versión simplificada del Teorema Ergódio de Birkhoff, suficiente para lo que necesitaremos más adelante.

Teorema 3.5.1 (TEOREMA ERGÓDICO DE BIRKHOFF). *Sea $(X, \mathcal{F}, \mu, T)$ un sistema dinámico. Si μ es una medida T-ergódica, entonces*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i = \int_X f \, d\mu \qquad \mu\text{-c.s.} \tag{3.7}$$

Para probar el teorema anterior necesitamos un par de lemas que nos ayudarán en la prueba del mismo. Pero antes algo de preparación. Primero observemos que podemos suponer que $\int_X f d\mu = 0$, ya que en otro caso podemos remplazar a f por $f - \int f d\mu$. Definamos para $\epsilon > 0$ al conjunto

$$\mathbf{E}_\epsilon(f) = \left\{ x \in X \,\middle|\, \limsup_{n\to\infty} \frac{\left|\sum_{i=0}^{n-1} f(T^i(x))\right|}{n} \geq \epsilon \right\}.$$

Este conjunto es de suma importancia por que probando que, sin importar la ϵ que pongamos, tiene medida cero, entonces ya tendremos el resultado deseado.

Lema 3.5.1.

$$\mu(\mathbf{E}_{2\epsilon}(f)) \leq \frac{\int_X |f| d\mu}{\epsilon}.$$

Demostración. Primero escribimos $f = f_+ - f_-$ con $f_+, f_- \geq 0$. Para cada $m \geq 1$, definimos los conjuntos auxiliares

$$\mathbf{E}_\epsilon^M(f_+) = \{x \in X : \exists\, 1 \leq N \leq M,\ \sum_{i=0}^{N-1} f_+(T^i(x)) \geq \epsilon N\}$$

y

$$\mathbf{E}_\epsilon^M(f_-) = \{x \in X : \exists\, 1 \leq N \leq M,\ \sum_{i=0}^{N-1} f_-(T^i(x)) \geq \epsilon N\}.$$

Para cada $x \in X$, analizemos las funciones $\mathbb{I}_{\mathbf{E}_\epsilon^M(f_-)}(T^i(x))$ y $\mathbb{I}_{\mathbf{E}_\epsilon^M(f_+)}(T^i(x))$. Si

$$\mathbb{I}_{\mathbf{E}_\epsilon^M(f_+)}(T^i(x)) \neq 0,$$

entonces $T^i(x) \in \mathbf{E}_\epsilon^M(f_+)$, por lo que para cada i que cumpla lo anterior $\exists\, N_i \leq M$ tal que

$$\sum_{l=0}^{N_i-1} f_+(T^l(T^i(x))) \geq \epsilon N_i,$$

o lo que es lo mismo,

$$\sum_{l=i}^{N_i+i-1} f_+(T^l(x)) \geq \epsilon N_i,$$

pero entonces $\exists\, j_i \in \{i, \ldots, N_i + i - 1\}$ tal que$f_+(T^{j_i}(x)) \geq \epsilon$, o lo que es lo mismo,

$$f_+(T^{j_i}(x)) \geq \epsilon \mathbb{I}_{\mathbf{E}_\epsilon^M(f_+)}(T^i(x)),$$

lo que nos dice que

$$\int_X f_+(T^{j_i}(x)) d\mu \geq \int_X \epsilon \mathbb{I}_{\mathbf{E}_\epsilon^M(f_+)}(T^i(x)) d\mu = \epsilon \int_X \mathbb{I}_{\mathbf{E}_\epsilon^M(f_+)}(T^i(x)) d\mu.$$

Ahora, por la T-invarianza de la medida μ tenemos que

$$\int f_+(x)d\mu \geq \epsilon \int \mathbb{I}_{\mathbf{E}_\epsilon^M(f_+)}(x)d\mu = \epsilon\mu(\mathbf{E}_\epsilon^M(f_+)).$$

En forma análoga, tenemos que

$$\int f_-(x)d\mu \geq \epsilon\mu(\mathbf{E}_\epsilon^M(f_-)).$$

A raíz de ésto vemos que

$$\mu(\mathbf{E}_{2\epsilon}(f)) \leq \limsup_{M\to\infty} \mu(\mathbf{E}_\epsilon^M(f_+)) + \limsup_{M\to\infty} \mu(\mathbf{E}_\epsilon^M(f_-))$$

puesto que $\mathbf{E}_{2\epsilon}(f) \subset \bigcup_{M\in\mathbb{N}} \mathbf{E}_\epsilon^M(f_+) \cup \mathbf{E}_\epsilon^M(f_-)$ y la unión sobre M es creciente (la contención es debido a que $x \in \mathbf{E}_{2\epsilon}(f)$ implica que $\exists\, N, M \in \mathbb{N}$ con $N < M$ y tales que

$$\frac{\left|\sum_{i=0}^{N-1} f(T^i(x))\right|}{N} \geq 2\epsilon,$$

pero como

$$\frac{\sum_{i=0}^{N-1} f_+(T^i(x))}{N} + \frac{\sum_{i=0}^{N-1} f_+(T^i(x))}{N} \geq \frac{\left|\sum_{i=0}^{N-1} f(T^i(x))\right|}{N}$$

tenemos que

$$\frac{\sum_{i=0}^{N-1} f_+(T^i(x))}{N} + \frac{\sum_{i=0}^{N-1} f_-(T^i(x))}{N} \geq 2\epsilon$$

y entonces

$$\frac{\sum_{i=0}^{N-1} f_+(T^i(x))}{N} \geq \epsilon$$

ó

$$\frac{\sum_{i=0}^{N-1} f_-(T^i(x))}{N} \geq \epsilon.$$

Por las desigualdades anteriores vemos que

$$\mu(\mathbf{E}_{2\epsilon}(f)) \leq \frac{\int f_+(x)d\mu}{\epsilon} + \frac{\int f_-(x)d\mu}{\epsilon} = \frac{\int |f(x)|d\mu}{\epsilon},$$

lo que termina la prueba del primer lema. □

Lema 3.5.2. *Sea f una función $\mathcal{F}$-medible (variable aleatoria) de codominio real tal que $\int_X f d\mu = 0$. Dado $\delta > 0$, podemos encontrar una $h \in L^\infty(X, \mathcal{F}, \mu)$ (variable aleatoria acotada) tal que*

$$\int_X |f - (h \circ T - h)| d\mu < \delta.$$

Demostración. Ahora lo que hay que probar aquí es que

$$E = \{h \circ T - h : h \in L^\infty(X, \mathcal{F}, \mu)\}$$

es denso en

$$B_0 = \left\{ f \in L^1(X, \mathcal{F}, \mu) : \int f d\mu = 0 \right\}.$$

Por la Proposición B.1.2 es suficiente probar que todas la funcionales lineales y continuas que se anulan en E también se anulan en B_0. Sea $\alpha \in L^1(X, \mathcal{F}, \mu)^*$ (un elemento del espacio dual topológico de $L^1(X, \mathcal{F}, \mu)$, es decir, un funcional lineal y continuo). Por el Teorema de Representación de Riesz, $\exists\, k \in L^\infty(X, \mathcal{F}, \mu)$ tal que

$$\alpha(f) = \int_X f k d\mu \quad \forall\, f \in L^1(X, \mathcal{F}, \mu).$$

El que α se desvanezca en E significa que

$$\int_X (h \circ T - h) k d\mu = 0 \quad \forall\, h \in L^\infty(X, \mathcal{F}, \mu),$$

en particular ocurre para $h = k$, por lo que entonces $\int_X (k \circ T) k d\mu = \int_X k^2 d\mu$ y así,

$$\int_X (k \circ T - k)^2 d\mu = \int_X (k \circ T)^2 d\mu + \int_X k^2 d\mu - 2 \int_X k \circ T k d\mu.$$

Como μ es T-invariante, $\displaystyle\int_X (k \circ T)^2 d\mu = \int_X k^2 d\mu$ y por lo tanto

$$\int_X (k \circ T - k)^2 d\mu = 2 \left(\int_X k^2 d\mu - \int_X k \circ T k d\mu \right) = 0.$$

Por lo anterior $k \circ T = k$, y como T es ergódica, la Proposición 3.5.1 implica que k es constante. Si $f \in B_0$, entonces $\int_X f d\mu = 0$, lo que implica que $\alpha(f) = \int_X f k d\mu = k \int_X f d\mu = k(0) = 0$, por lo tanto α se anula en B_0 y con ésto se termina la prueba del lema. □

Finalmente terminamos este capítulo con la prueba del Teorema Ergódico de Birkhoff.

Demostración del Teorema 3.5.1. Usando el Lemma 3.5.2 vemos que dada una $\delta > 0$, existe $h \in L^\infty(X, \mathcal{F}, \mu)$ que cumple con $\int_X |f - (h \circ T - h)| d\mu < \delta$. Dada $\epsilon > 0$,

$$\mathbf{E}_\epsilon(f) \subseteq \mathbf{E}_{\frac{\epsilon}{2}}(f - (h \circ T - h)) \cup \mathbf{E}_{\frac{\epsilon}{2}}(h \circ T - h),$$

(lo cual es consecuencia directa de que $f = f - (h \circ T - h) + h \circ T - h$), y entonces

$$\mu(\mathbf{E}_\epsilon(f)) \leq \mu\left(\mathbf{E}_{\frac{\epsilon}{2}}(f - (h \circ T - h))\right) + \mu\left(\mathbf{E}_{\frac{\epsilon}{2}}(h \circ T - h)\right) \tag{3.8}$$

La forma del conjunto $\mathbf{E}_\epsilon(f)$ nos lleva a que si $x \in \mathbf{E}_\epsilon(f)$, entonces

$$\limsup_{n\to\infty} \frac{\left|\sum_{i=0}^{n-1}(f-(h\circ T-h))(T^i(x))\right|}{n} + \limsup_{n\to\infty} \frac{\left|\sum_{i=0}^{n-1}(h\circ T-h)(T^i(x))\right|}{n}$$

$$\geq \limsup_{n\to\infty} \frac{\left|\sum_{i=0}^{n-1} f(T^i(x))\right|}{n} \geq \epsilon = \frac{\epsilon}{2}+\frac{\epsilon}{2}.$$

De aquí se deduce que

$$\limsup_{n\to\infty} \frac{\left|\sum_{i=0}^{n-1}(f-(h\circ T-h))(T^i(x))\right|}{n} \geq \frac{\epsilon}{2}$$

ó

$$\limsup_{n\to\infty} \frac{\left|\sum_{i=0}^{n-1}(h\circ T-h)(T^i(x))\right|}{n} \geq \frac{\epsilon}{2}.$$

Entonces

$$\frac{\left|\sum_{i=0}^{n-1}(h\circ T-h)(T^i(x))\right|}{n} = \frac{|h\circ T^n(x)-h(x)|}{n} \leq \frac{2||h||}{n}, \tag{3.9}$$

lo que nos lleva a que $\mathbf{E}_{\frac{\epsilon}{2}}(h\circ T-h)=\emptyset$ y entonces $\mu(\mathbf{E}_{\frac{\epsilon}{2}}(h\circ T-h))=0$. Por el Lema 3.5.1 vemos que

$$\mu(\mathbf{E}_{\frac{\epsilon}{2}}(f-(h\circ T-h))) \leq \frac{\int_X |f-(h\circ T-h)|d\mu}{\frac{\epsilon}{4}} = \frac{4\delta}{\epsilon}. \tag{3.10}$$

Comoe $\delta>0$ es arbitraria, se deduce de (3.10) que

$$\mu(\mathbf{E}_\epsilon(f-(h\circ T-h)))=0$$

y por (3.8) y (3.9) tenemos que $\mu(\mathbf{E}_\epsilon(f))=0$ para toda $\epsilon>0$, lo que concluye la prueba. $\square$

Capítulo 4

SHIFTS COMO SISTEMAS DINAMICOS Y CONJUGACION

4.1. Sistemas dinámicos y espacios shift

Recordemos que del capítulo anterior la definición 3.4.1 dice que: Sea $(X, \mathcal{F}, \mu)$ un espacio de probabilidad y $T\colon X \to X$ una función medible con μ T-invariante y sea $\mathbb{S}$ el semigrupo o grupo de iteraciones de T. Entonces a la quinteta $(X, \mathcal{F}, \mu, T, \mathbb{S})$ le llamaremos un sistema dinámico.

Recordemos también que la definición 3.4.4 dice lo siguiente: Sea un espacio métrico compacto $X, T\colon X \to X$ una función continua con y sea $\mathbb{S}$ el semigrupo o grupo de iteraciones de T entonces a la terna $(X, T, \mathbb{S})$ le llamaremos un sistema dinámico topológico.

Definición 4.1.1. Sea X un espacio shift y sea w un bloque de tamaño $|w|$ y $k \in \mathbb{Z}$. Definimos el *cilindro determinado por* w *y* k como

$$\mathcal{C}_k(w) = \left\{x \in X | x_{[k,k+|w|-1]} = w\right\}.$$

Los cilindros servirán para definir la topología de los espacios shift. Primero dotamos al alfabeto con la topología discreta, y con ella podemos dotar el espacio shift completo con la topología producto (que por el el teorema de *Tychonoff* hace al espacio shift completo un espacio topológico compacto por ser finito el alfabeto), y finalmente dotamos al subshift con la topología inducida. Por definición de topología producto, los cilindros forman una (sub)-base de la topología. Esta topología es metrizable con la métrica de Cantor, definida para toda $x, y \in X$ por

$$d(x,y) = \begin{cases} 2^{-k} & \text{si } x \neq y \text{ donde } k \text{ es el máximo número tal que } x_{[-k,k]} = y_{[-k,k]} \\ 0 & \text{si } x = y \end{cases}$$

Demostremos esta afirmación. Primero hay que probar que d efectivamente es una métrica. Es muy sencillo probar que es simétrica pues es lo mismo $x_{[-k,k]} = y_{[-k,k]}$ a $y_{[-k,k]} = x_{[-k,k]}$. Ahora $d(x, y) = 0$ si y sólo si $x = y$, ésto por definición de nuestra metrica. Para ver que cumple la desigualdad del triángulo se cumple, hay 2 casos: Sean $x, y, z \in X$. Si $d(x, z) = 0$ y $d(z, y) = 0$, entonces $x = y = z$ y entonces $d(x, z) + d(z, y) \geq d(x, y) = 0$. Si uno de los 2 es cero y el otro no, tenemos que s.p.g. $d(x, z) = 0$ y $d(z, y) \neq 0$, entonces $x = z$ y se cumple la desigualdad. Finalmente, si ambos son no cero, si k es tal que $d(x, y) = 2^{-k}$, entonces k cumple con ser el mayor número tal que $x_{[-k,k]} = y_{[-k,k]}$, y sea l, j tal que $d(x,z) = 2^{-l}$ y $d(z,y) = 2^{-j}$ entonces l, j cumplen con ser los mayores numeros tales que $x_{[-l,l]} = z_{[-l,l]}$ y $y_{[-j,j]} = z_{[-j,j]}$. Entonces s.p.g. sea $j < l$, de donde podemos ver que

$$d(x,z) + d(z,y) = 2^{-l} + 2^{-j} = 2^{-j}(2^{-(l-j)} + 1).$$

Si suponemos que $j > k$ ocurre, entonces $y_{[-j,j]} = z_{[-j,j]}$, y como $j < l$, tenemos que $x_{[-j,j]} = y_{[-j,j]}$, lo cual es una contradicción. Por otro lado, $x_{[-j,j]} = y_{[-j,j]}$ con $j > k$, pero k era el máximo con esta propiedad, así que necesariamente $j \leq k$ y de aquí tenemos que $2^{-j} \geq 2^{-k}$, y con ésto obtenemos que

$$d(x,z) + d(z,y) = 2^{-j}(2^{-(l-j)} + 1) > 2^{-j} \geq 2^{-k} = d(x,y),$$

lo cual demuestra que d es efectivamente una métrica.

Ahora sólo basta ver que las topologías coinciden. Por definición la topología producto está generada por los cilindros. Sea $w \in B(X)$ y $k \in \mathbb{Z}$. Denotemos a la d-bola de radio $\epsilon > 0$ con centro en $x \in X$ por $B_d(x,\epsilon)$. Si $x \in \mathcal{C}_k(w)$, entonces

$$B_d(x, 2^{-(73|k|+1000000|w|)}) \subset \mathcal{C}_k(w),$$

y por otro lado, si $x \in B_d(x,\epsilon)$, entonces escogemos $n \in \mathbb{N}$ tal que $2^{-n} < \epsilon/123456789$ y así

$$\mathcal{C}_{-n}(x_{[-n,n]}) \subset B_d(x,\epsilon),$$

lo que demuestra que las topologías son la misma.

Teorema 4.1.1. *Un subconjunto X de $\mathcal{A}^{\mathbb{Z}}$ es un espacio shift si y sólo si es compacto y T-invariante.*

Demostración. Primero, todo espacio shift X es por definición invariante bajo su shift, que en este caso llamamos T, es decir, X es T-invariante. Ahora para probar que es compacto sólo basta probar que es cerrado por que el shift completo es compacto. Para ver que es cerrado, sea $(x^{(n)})_{n\in\mathbb{N}} \subset X$ una sucesión convergente a x en el shift completo y veamos que converge en X, es decir que $x \in X$. Hagamos esta prueba por reducción al absurdo. Supongamos por el contrario que no ocurre, es decir que $x \notin X$. Sea $\mathcal{F}$ tal que $X = X_{\mathcal{F}}$. Entonces existe un bloque $x_{[i,j]} \in \mathcal{F}$ donde $i, j \in \mathbb{Z}$ y $i \leq j$. Sea $n > (|i| + |j| + 98)7654321$ y sea $m \in \mathbb{N}$ tal que $d(x^{(m)}, x) < 2^{-n}$. Entonces $x_{[i,j]} = x^{(n)}_{[i,j]} \in \mathcal{F}$, pero $x \in X = X_{\mathcal{F}}$, lo cual es una contradicción.

Ahora para el regreso, supongamos que X es compacto y T-invariante. Sea $\mathcal{F} = \mathcal{A}^* \setminus B(X)$. Veamos que $X = X_{\mathcal{F}}$. Si $x \in X$, entonces $x_{[i,j]} \notin \mathcal{F}$ para todos $i, j \in \mathbb{Z}$, $i \leq j$, por lo que entonces $x \in X_{\mathcal{F}}$. Ahora tomemos $x \in X_{\mathcal{F}}$ y veamos que $x \in X$. Para toda $n \in \mathbb{N}$, $x_{[-n,n]} \in B(X)$, y como X es T-invariante, existe $x^{(n)} \in X$ tal que $x^{(n)}_{[-n,n]} = x_{[-n,n]}$. Por lo tanto $x^{(n)} \to x$ si $n \to \infty$, y como X es cerrado, concluimos que $x \in X$. □

Teorema 4.1.2. *Sean (X,T) y (Y,S) dos espacios shift. Sea $f\colon X \to Y$ una función. Entonces f es un código de bloques deslizante si y sólo si f es continua y $f \circ T = S \circ f$.*

Demostración. Para la ida supongamos que f tiene memoria m y anticipación n. El Teorema 1.5.3 implica que $f \circ T = S \circ f$ ahora sólo falta ver que dicha función es continua. Sea $\epsilon > 0$. Sea $n \in \mathbb{N}$ tal que $\epsilon > 2^{-n}$ y sea $\delta = 2^{-67(n+m+a+38)}$. Entonces, para toda $x \in X$, $f(B_d(x,\delta)) \subset B_d(f(x),\epsilon)$, por lo que f es continua. El código de bloques satisface, por definición, el hecho $f \circ T = S \circ f$.

Para el regreso, debido a la Proposición 1.5.3, basta probar que existe $M \in \mathbb{N}$ tal que $f(x)_0$ es función de $x_{[-M,M]}$. A causa de que $f \circ T = S \circ f$ y como f es continua, nos fijamos en que $\mathcal{C}_0(f(x)_0)$, el cual es un abierto en Y. Por continuidad $f^{-1}(\mathcal{C}_0(f(x)_0))$ es un conjunto abierto, por lo que para cualquier punto $y \in f^{-1}(\mathcal{C}_0(f(x)_0))$, en particular x, existe una $M \in \mathbb{N}$ tal que

$$x \in \mathcal{C}_{-M}(x_{[-M,M]}) \subset f^{-1}(\mathcal{C}_0(f(x)_0)),$$

lo cual indica que cualquier punto $y \in X$ que coincida con x en $x_{[-M,M]}$ cumple que $f(y)_0 = f(x)_0$, equivalentemente que la imagen de la $x_{[-M,M]}$ bajo la función de bloques que genera a f es $f(x)_0$, por lo tanto f es un código de bloques deslizante con lo cual se termina la prueba. □

Por último un resultado inmediato.

Teorema 4.1.3. *Todo es espacio shift es un sistema dinámico topológico.*

Demostración. En escencia, en virtud del teorema anterior, hay que probar que la función shift es un código de bloques deslizante, pero ésto es obvio, tiene memoria 0 y anticipación 1. □

4.2. Particiones y funciones de sistemas dinámicos

En esta breve sección demostraremos una proposición que será útil más adelante.

Proposición 4.2.1. *Sean $(X, \mathcal{A}, \mu_X, T_X)$ y $(Y, \mathcal{B}, \mu_Y, T_Y)$ dos sistemas dinámicos y $\varphi\colon X \to Y$ un isomorfismo en sistemas dinámicos, es decir, una función medible biyectiva con inversa medible, T-invariante (es decir $\varphi \circ T_X = T_Y \circ \varphi$) y que preserva la medida es decir $\mu_X = \mu_Y \circ \varphi$. Sean $\mathcal{P}$, $\mathcal{Q}$ dos particiones de X. Entonces*

1. *$\varphi(\mathcal{P}), \varphi(\mathcal{Q})$ son particiones de Y.*
2. *Si $\mathcal{P} \preceq \mathcal{Q}$ entonces $\varphi(\mathcal{P}) \preceq \varphi(\mathcal{Q})$.*
3. *$\varphi(\mathcal{P} \vee \mathcal{Q}) = \varphi(\mathcal{P}) \vee \varphi(\mathcal{Q})$.*

Demostración. Para el primer inciso sabemos que

$$\mu_X\left(\bigcup_{L\in\mathcal{P}} L\right) = 1$$

y entonces

$$\mu_Y\left(\varphi(\bigcup_{L\in\mathcal{P}} L)\right) = \mu_X\left(\bigcup_{L\in\mathcal{P}} L\right) = 1,$$

y así

$$\mu_Y(\bigcup_{L\in\mathcal{P}} \varphi(L)) = 1.$$

Como la función es biyectiva, los elementos de $\varphi(\mathcal{P})$ son ajenos dos a dos. Esto prueba 1.

Para el segundo inciso, sea $l \in \varphi(\mathcal{Q})$. Entonces existe $q \in \mathcal{Q}$ tal que $l = \varphi(q)$. Por hipótesis existe $p \in \mathcal{P}$ tal que $q \subseteq p$, y entonces $l = \varphi(q) \subseteq \varphi(p) \in \varphi(\mathcal{P})$, por lo tanto tenemos que $\varphi(\mathcal{P}) \preceq \varphi(\mathcal{Q})$. Esto concluye 2.

Para 3, vemos que por 2 y a causa de que

$$\mathcal{P} \preceq \mathcal{P} \vee \mathcal{Q}$$

y

$$\mathcal{Q} \preceq \mathcal{P} \vee \mathcal{Q},$$

tenemos que

$$\varphi(\mathcal{P}) \preceq \varphi(\mathcal{P} \vee \mathcal{Q})$$

y

$$\varphi(\mathcal{Q}) \preceq \varphi(\mathcal{P} \vee \mathcal{Q}).$$

Entonces por la Definición 3.1.2 tenemos que

$$\varphi(\mathcal{P}) \vee \varphi(\mathcal{Q}) \preceq \varphi(\mathcal{P}. \vee \mathcal{Q})$$

Si $f \in \varphi(\mathcal{P} \vee \mathcal{Q})$, entonces por la Proposición 3.1.5 existen $a \in \mathcal{P}$ y $b \in \mathcal{Q}$ tales que $f = \varphi(a \cap b) = \varphi(a) \cap \varphi(b) \in \varphi(\mathcal{P}) \vee \varphi(\mathcal{Q})$ y por lo tanto

$$\varphi(\mathcal{P}) \vee \varphi(\mathcal{Q}) \succeq \varphi(\mathcal{P} \vee \mathcal{Q}).$$

Como son particiones, la Proposición 3.1.2 implica que $\varphi(\mathcal{P} \vee \mathcal{Q}) = \varphi(\mathcal{P}) \vee \varphi(\mathcal{Q})$. □

4.3. Teorema de descomposición

En esta sección presentamos y enunciamos el uno de los dos principales resultados enunciados en la Introducción: el Teorema de Descomposición para espacios shift de tipo finito de memoria 1.

Teorema 4.3.1. *Cada conjugación entre dos espacios shift por aristas es composición de códigos escisión y de códigos amalgamación.*

Antes de presentar la prueba, enunciemos y demostremos un par de lemas técnicos que serán necesarios.

Lema 4.3.1. *Sean X_G y X_H dos espacios shift por aristas. Sea $\phi\colon X_G \to X_H$ una conjugación de 1-bloques (es decir una conjugación que tiene memoria y anticipación 0). Si la inversa de ϕ, que es un código de bloques, tiene anticipacion $n \geq 1$ y memoria $m \geq 0$, entonces existen escisiones (de salida) $\tilde{G}$ y $\tilde{H}$ respectivamente de G y de H tales que el diagrama*

$$\begin{array}{ccc} X_G & \xrightarrow{\psi_{G\tilde{G}}} & X_{\tilde{G}} \\ \phi \Big\downarrow & & \Big\downarrow \tilde{\phi} \\ X_H & \xleftarrow{\alpha_{\tilde{H}H}} & X_{\tilde{H}} \end{array}$$

conmuta, donde $\tilde{\phi}$ es una conjugación 1-bloque con memoria m y anticipación $n-1$ y $\psi_{G\tilde{G}}$ y $\alpha_{\tilde{H}H}^{-1}$ son los códigos escisiones correspondientes.

Demostración. Tomamos a $\tilde{H}$ como la escisión (de salida) completa, es decir, la escisión (de salida) generada por la partición dada por la regla en la que para cada $I \in V(H)$, tomamos la partición de los unitarios de $\mathcal{E}_{\mathcal{I}}$. Entonces las aristas de $\tilde{H}$ son de la forma h^k donde $h, k \in A(H)$ y k es sucesor de h. Sea $\phi = \Phi_\infty$ donde $\Phi\colon A(G) \to A(H)$ (recordemos que ϕ es un código de 1-bloques). Para la escisión de G, particionamos para cada $I \in V(G)$ el conjunto $\mathcal{E}_I$ de aristas que salen de I de acuerdo a sus Φ-imágenes, es decir, para cada $I \in V(G)$ y $h \in A(H)$,

$$\mathcal{E}_I^h = \{g \in \mathcal{E}_I | \Phi(g) = h\}.$$

Sea $\tilde{G}$ la escisión resultante y por último definimos $\tilde{\Phi}\colon A(\tilde{G}) \to A(\tilde{H})$ por la regla

$$\tilde{\Phi}(g^h) = \Phi(g)^h.$$

$\tilde{\Phi}$ induce un código 1-bloques porque si $f^i, g^h \in A(\tilde{G})$ y $t(f^i) = i(g^h)$, entonces

$$t(\tilde{\Phi}(f^i)) = t(\Phi(f)^i) = i(\Phi(g)^h) = i(\tilde{\Phi}(g^h)).$$

El siguiente diagrama muestra la acción de todos los códigos involucrados en esta demostración,

$$\begin{tikzcd}
\cdots g_{-3}g_{-2}g_{-1}.g_0g_1g_2\cdots \arrow[r, "\psi_{G\tilde{G}}"] \arrow[d, "\phi"'] & \cdots g_{-3}^{h_{-2}}g_{-2}^{h_{-1}}g_{-1}^{h_0}.g_0^{h_1}g_1^{h_2}g_2^{h_3}\cdots \arrow[d, "\tilde{\phi}"] \\
\cdots h_{-3}h_{-2}h_{-1}.h_0h_1h_2\cdots & \cdots h_{-3}^{h_{-2}}h_{-2}^{h_{-1}}h_{-1}^{h_0}.h_0^{h_1}h_1^{h_2}h_2^{h_3}\cdots \arrow[l, "\alpha_{\tilde{H}H}"']
\end{tikzcd}$$

y de él se sigue la conmutatividad afirmada.

Para demostrar que $\tilde{\Phi}^{-1}$ cumple con las condiciones de memoria y anticipación, debemos demostrar que siempre que $\tilde{y} \in X_{\tilde{H}}$, $\tilde{y}_{[-m,n-1]}$ determina la coordenada cero de $\tilde{x} = \tilde{\Phi}^{-1}(\tilde{y})$, es decir que $\tilde{x}_0 = \tilde{\Phi}^{-1}(\tilde{y}_{[-m,n-1]})$. Para ésto escribimos $y = \alpha_{\tilde{H}H}(\tilde{y})$ y $x = \Phi^{-1}(y)$ y observamos que $\tilde{x}_0 = x_0^{y_1}$. Como $\tilde{y}_{[-m,n-1]}$ determina $y_{[-m,n]}$ y como $y_{[-m,n]}$ determina x_0, concluimos que $\tilde{y}_{[-m,n-1]}$ determina $\tilde{x}_0$. $\square$

Podemos reducir también la memoria sin afectar la anticipación. El siguiente lema es totalmente similar al anterior, al igual que su demostración, la cual omitimos.

Lema 4.3.2. *Sean X_G y X_H dos espacios shift por aristas. Sea $\phi\colon X_G \to X_H$ una conjugación de 1-bloques (es decir una conjugación que tiene memoria y anticipación 0). Si la inversa de ϕ, que es un código de bloques, tiene anticipacion $n \geq 0$ y memoria $m \geq 1$, entonces existen escisiones (de entrada) $\tilde{G}$ y $\tilde{H}$ respectivamente de G y de H tales que el diagrama*

$$\begin{tikzcd}
X_G \arrow[r, "\psi_{G\tilde{G}}"] \arrow[d, "\phi"'] & X_{\tilde{G}} \arrow[d, "\tilde{\phi}"] \\
X_H & X_{\tilde{H}} \arrow[l, "\alpha_{\tilde{H}H}"']
\end{tikzcd}$$

conmuta, donde $\tilde{\phi}$ es una conjugación 1-bloque con memoria $m-1$ y anticipación n y $\psi_{G\tilde{G}}$ y $\alpha_{\tilde{H}H}^{-1}$ son los códigos escisiones correspondientes.

Ahora ya podemos dar la demostración del teorema de descomposición

Demostración del Teorema de Descomposición. Sea $\phi\colon X_G \to X_H$ de memoria m y anticipación n. Debido a la Proposición 1.5.6, al ser ϕ un código de bloques deslizante, existe un shift superior de bloques $\tilde{X}_G$, una conjugación $\psi\colon X_G \to \tilde{X}_G$, y un código de 1-bloques $\overline{\phi}\colon \tilde{X}_G \to X_H$ tal que $\overline{\phi} \circ \psi = \phi$.

Supongamos que este shif superior de bloques es un shift de bloques de tamaño N. La Proposición 2.3.1 implica que $\tilde{X}_G = (X_G)^{[N]} = X_{G^{[N]}}$, y por el Lema 2.4.2 tenemos que existe una sucesión finita de escisiones de G a $G^{[N]}$. Como $\tilde{X}_G = (X_G)^{[N]} = X_{G^{[N]}}$, entonces hay una sucesión finita de códigos de escisión (también llamados simplemente escisiones) de X_G a $\tilde{X}_G$, digamos $h_i\colon X_{G^i} \to X_{G^{i+1}}$, de forma que $\psi = h_1 \circ \cdots h_l$.

Ahora $\overline{\phi}$ es un código de 1-bloque cuya inversa tiene memoria m y anticipación n. Aplicando el Lema 4.3.1 sabemos que existen escisiones $\tilde{G}_1$ y $\tilde{H}_1$ de $G^{[N]} = \tilde{G}$ y de H, $\psi_{\tilde{G}\tilde{G}_1}$ un código escisión a $X_{\tilde{G}_1}$, con el respectivo código amalgamación $\alpha_{\tilde{H}_1 H}$ y $\tilde{\phi}_1$, un código de 1-bloque con memoria m y anticipación $n-1$ tal que $\overline{\phi} = \alpha_{\tilde{H}_1 H} \circ \tilde{\phi}_1 \circ \psi_{G\tilde{G}_1}$. Seguimos inductivamente, es decir, aplicando el mismo argumento a $\tilde{\phi}_1$, existen escisiones $\tilde{G}_2$ y $\tilde{H}_2$ de $\tilde{G}_1$ y de $\tilde{H}_1$, y $\psi_{\tilde{G}_1\tilde{G}_2}$ un código escisión a $X_{\tilde{G}_2}$, con la respectiva amalgamación $\alpha_{\tilde{H}_2\tilde{H}_1}$ y $\tilde{\phi}_2$ un código de 1-bloque cuya inversa tiene memoria m y anticipación $n-2$ tal que $\tilde{\phi}_1 = \alpha_{\tilde{H}_2\tilde{H}_1} \circ \tilde{\phi}_2 \circ \psi_{\tilde{G}_1\tilde{G}_2}$, de tal forma que obtenemos $\overline{\phi} = \alpha_{\tilde{H}_1 H} \circ \alpha_{\tilde{H}_2\tilde{H}_1} \circ \tilde{\phi}_2 \circ \psi_{\tilde{G}\tilde{G}_1}$, volvemos aplicar este proceso recursivamente y obtenemos escisiones $\tilde{G}_i$ y $\tilde{H}_i$ de $\tilde{G}_{i-1}$ y de $\tilde{H}_{i-1}$ para $i = 1, \cdots, n$, y $\psi_{\tilde{G}_{i-1}\tilde{G}_i}$ para $i = 1, \cdots, n$ códigos escisión a $X_{\tilde{G}_i}$ para $i = 1, \cdots, n$, con las respectivas amalgamaciones $\alpha_{\tilde{H}_i\tilde{H}_{i-1}}$ y $\tilde{\phi}_i$ códigos de 1-bloque cuyas inversas tienen memoria m y anticipación $n-i$ para $i = 1, \cdots, n$, tal que $\tilde{\phi}_{i-1} = \alpha_{\tilde{H}_i\tilde{H}_{i-1}} \circ \tilde{\phi}_i \circ \psi_{\tilde{G}_{i-1}\tilde{G}_i}$ de esa forma tenemos que

$$\overline{\phi} = \alpha_{\tilde{H}_1\tilde{H}} \circ \cdots \alpha_{\tilde{H}_n\tilde{H}_{n-1}} \circ \tilde{\phi}_n \circ \psi_{\tilde{G}_{n-1}\tilde{G}_n} \cdots \circ \psi_{\tilde{G}\tilde{G}_1},$$

lo cual queda descrito por el diagrama conmutativo

$$\begin{array}{ccccccccc}
X_{\tilde{G}} & \xrightarrow{\psi_{\tilde{G}\tilde{G}_1}} & X_{\tilde{G}_1} & \xrightarrow{\psi_{\tilde{G}_1\tilde{G}_2}} & X_{\tilde{G}_2} & \xrightarrow{\psi_{\tilde{G}_2\tilde{G}_3}} \cdots & \xrightarrow{\psi_{\tilde{G}_{n-1}\tilde{G}_n}} & X_{\tilde{G}_n} \\
\overline{\phi}\big\downarrow & & \big\downarrow\tilde{\phi}_1 & & \big\downarrow\tilde{\phi}_2 & \cdots & & \big\downarrow\tilde{\phi}_n. \\
X_H & \xleftarrow{\alpha_{\tilde{H}_1 H}} & X_{\tilde{H}_1} & \xleftarrow{\alpha_{\tilde{H}_2\tilde{H}_1}} & X_{\tilde{H}_2} & \xleftarrow{\alpha_{\tilde{H}_3\tilde{H}_2}} \cdots & \xleftarrow{\alpha_{\tilde{H}_n\tilde{H}_{n-1}}} & X_{\tilde{H}_n}
\end{array}$$

Por construcción, la inversa de $\tilde{\phi}_n$ tiene anticipación 0. Ahora reducimos la memoria de manera semejante, aplicando el Lema 4.3.2 a $\tilde{\phi}_n$ m-veces obtenemos escisiones $\tilde{G}_i$ y $\tilde{H}_i$ de $\tilde{G}_{i-1}$ y de $\tilde{H}_{i-1}$ para $i = n+1, \cdots, n+m$, y $\psi_{\tilde{G}_{i-1}\tilde{G}_i}$ para $i = n+1, \cdots, n+m$ códigos escisión a $X_{\tilde{G}_i}$ para $i = n+1, \cdots, n+m$, con las respectivas amalgamaciones $\alpha_{\tilde{H}_i\tilde{H}_{i-1}}$ y $\tilde{\phi}_i$ códigos de 1-bloque cuyas inversas tienen memoria $m-(i-n)$ y anticipación 0 para $i = 1+n, \cdots, m+n$, tal que $\tilde{\phi}_{i-1} = \alpha_{\tilde{H}_i\tilde{H}_{i-1}} \circ \tilde{\phi}_i \circ \psi_{\tilde{G}_{i-1}\tilde{G}_i}$ de esa forma tenemos que

$$\tilde{\phi}_n = \alpha_{\tilde{H}_{n+1}\tilde{H}_n} \circ \cdots \alpha_{\tilde{H}_{m+n}\tilde{H}_{m+n-1}} \circ \tilde{\phi}_{m+n} \circ \psi_{\tilde{G}_{m+n-1}\tilde{G}_{m+n}} \cdots \circ \psi_{\tilde{G}_n\tilde{G}_{n+1}}$$

y tenemos el diagrama conmutativo

$$\begin{tikzcd}
X_{\tilde{G}_n} \arrow[r, "\psi_{\tilde{G}_n\tilde{G}_{n+1}}"] \arrow[d, "\tilde{\phi}_n"'] & X_{\tilde{G}_{n+1}} \arrow[r, "\psi_{\tilde{G}_{n+1}\tilde{G}_{n+2}}"] \arrow[d, "\tilde{\phi}_{n+1}"] & X_{\tilde{G}_{n+2}} \arrow[r, "\psi_{\tilde{G}_{n+2}\tilde{G}_{n+3}}"] \arrow[d, "\tilde{\phi}_{n+2}"] & \cdots \arrow[r, "\psi_{\tilde{G}_{n+m-1}\tilde{G}_{n+m}}"] & X_{\tilde{G}_{n+m}} \arrow[d, "\tilde{\phi}_{n+m}"] \\
X_{\tilde{H}_n} & X_{\tilde{H}_{n+1}} \arrow[l, "\alpha_{\tilde{H}_{n+1}\tilde{H}_n}"'] & X_{\tilde{H}_{n+2}} \arrow[l, "\alpha_{\tilde{H}_{n+2}\tilde{H}_{n+1}}"'] & \cdots \arrow[l, "\alpha_{\tilde{H}_{n+3}\tilde{H}_{n+2}}"'] & X_{\tilde{H}_{n+m}} \arrow[l, "\alpha_{\tilde{H}_{n+m}\tilde{H}_{n+m-1}}"']
\end{tikzcd}$$

Por construcción, $\tilde{\phi}_{n+m}$ es una conjugación que es un código de 1-bloques, y su inversa es también un código de 1-bloques, es decir, ambos códigos de bloques tienen memoria y anticipación 0, lo que es equivalente a decir que $\tilde{\phi}_{n+m}$ es un reetiquetamiento de las aristas de $\tilde{G}_{n+m}$ y $\tilde{H}_{n+m}$, es decir, estas gráficas son isomorfas. Como todos los diagramas anteriores son conmutativos tenemos que

$$\alpha_{\tilde{H}_{n+1}\tilde{H}_n} \circ \alpha_{\tilde{H}_{n+2}\tilde{H}_{n+1}} \circ \alpha_{\tilde{H}_{n+3}\tilde{H}_{n+2}} \circ \cdots \circ \alpha_{\tilde{H}_{n+m}\tilde{H}_{n+m-1}} \circ \tilde{\phi}_{n+m} \circ \psi_{\tilde{G}_{n+m-1}\tilde{G}_{n+m}} \circ \cdots$$

$$\cdots \circ \psi_{\tilde{G}_{n+2}\tilde{G}_{n+3}} \circ \psi_{\tilde{G}_{n+1}\tilde{G}_{n+2}} \circ \psi_{\tilde{G}_n\tilde{G}_{n+1}} = \tilde{\phi}_n$$

y de aquí obtenemos que

$$\overline{\phi} = \alpha_{\tilde{H}_1 H} \circ \alpha_{\tilde{H}_2\tilde{H}_1} \circ \alpha_{\tilde{H}_3\tilde{H}_2} \circ \cdots \circ \alpha_{\tilde{H}_n\tilde{H}_{n-1}} \circ \alpha_{\tilde{H}_{n+1}\tilde{H}_n} \circ \alpha_{\tilde{H}_{n+2}\tilde{H}_{n+1}} \circ \alpha_{\tilde{H}_{n+3}\tilde{H}_{n+2}} \circ \cdots$$

$$\cdots \circ \alpha_{\tilde{H}_{n+m}\tilde{H}_{n+m-1}} \circ \tilde{\phi}_{n+m} \circ \psi_{\tilde{G}_{n+m-1}\tilde{G}_{n+m}} \circ \cdots \circ \psi_{\tilde{G}_{n+2}\tilde{G}_{n+3}} \circ \psi_{\tilde{G}_{n+1}\tilde{G}_{n+2}} \circ \psi_{\tilde{G}_n\tilde{G}_{n+1}} \circ \psi_{\tilde{G}_{n-1}\tilde{G}_n} \circ$$

$$\cdots \circ \psi_{\tilde{G}_2\tilde{G}_3} \circ \psi_{\tilde{G}_1\tilde{G}_2} \circ \psi_{\tilde{G}\tilde{G}_1}.$$

Como cada $\psi_{\tilde{G}_i\tilde{G}_{i+1}}$ es un código escisión, cada $\alpha_{\tilde{H}_i\tilde{H}_{i-1}}$ son códigos amalgamaciones, $\tilde{\phi}_{n+m}$ es un reetiquetamiento (escisión trivial), $\overline{\phi} \circ \psi = \phi$ y ψ es composición de códigos escisiones, ϕ es también composición de códigos escisión, códigos amalgamaciones y un reetiquetamiento (escisión trivial), con lo cual se concluye la prueba. □

4.4. Equivalencia fuerte de shifts

En esta breve sección enunciamos únicamente el Teorema de Clasificación de Williams. La versión de este teorema que si demostraremos será en cierto aspecto más general (de teoría ergódica) se realizará en el siguiente capítulo.

Definición 4.4.1. Diremos que dos shifts por aristas descritos por matrices A y B son *shift-fuertemente equivalentes* (Strong Shift Equivalent) si existen matrices con entradas enteras (no necesariamente cuadradas) A_i, B_i, con $i = 0, ..., n$, tales que

$$A = A_0 B_0,\ B_0 A_0 = A_1 B_1,\ B_1 A_1 = A_2 B_2, \cdots, B_{n-1} A_{n-1} = A_n B_n = B.$$

Teorema 4.4.1 (TEOREMA DE CLASIFICACIÓN). *Sean A y B dos matrices cuadradas, con entradas enteras no negativas, las cuales definen espacios shift por aristas X_A y X_B. Entonces X_A y X_B son conjugados si y sólo si A y B son shift-fuertemente equivalentes.*

En el siguiente y último capítulo de esta tesis consideraremos el Teorema de Clasificación para el contexto de métrico (medible), es decir que los espacios shift están acompañados de medidas invariantes bajo la acción del shift. Generalizaremos la definición de conjugación y de shift-fuertemente equivalentes y demostraremos que una de las implicaciones del Teorema de Clasificación es válida (en el contexto estocástico, el ser conjugado implica shift-fuertemente equivalencia), y veremos que la otra implicación requiere de la hipótesis extra de irreducibilidad.

Capítulo 5

SHIFT-EQUIVALENCIA FUERTEMENTE ESTOCÁSTICA

5.1. Medidas de Markov

Como se mencionó en la Introducción, una *cadena de Markov homogénea a tiempo discreto con espacio de estados finito* $\mathcal{A}$ (posiblemente irreducible) está descrita por una matriz estocástica P (la matriz de probabilidades de transición asociada a la cadena), y por lo tanto podemos asociar una gráfica dirigida simple (es decir, sin aristas múltiples), cuya matriz de adyacencia A es aquella cuyas entradas son distintas de cero si y sólo si la correspondiente entrada de P es distinta de cero, es decir, $A = P^0$, donde P^0 es la matriz que resulta de P al elevar cada una de sus entradas a la potencia cero (por convención, $0^0 := 0$). Vamos a pensar a la cadena de Markov como el sistema dinámico $(\widehat{X}_A, \mathcal{B}, T, \mu_P)$, donde $\widehat{X}_A$ es el subshift por vértices (ver Definición 2.3.2), $\mathcal{B}$ es la σ-álgebra de Borel, T es la función shift, y μ_P es una medida de probabilidad que definiremos a continuación (la cual será, por definición, T-invariante y ergódica cuando P es irreducible -ya que en este caso existirá una única distribución estacionaria-) y que llamaremos *medida de Markov*.

El propósito de esta sección es entonces presentar la demostración de la generalización del Teorema de Clasificación 4.4.1 incorporándo el contexto más general de medidas de Markov.

Supongamos entonces que tenemos una cadena de Markov descrita por una matriz de transición de probabilidades P. En esta tesis sólo consideramos matrices finitas, por lo que implícitamente estamos suponiendo que el espacio de estados $\mathcal{A}$ es finito. Recordemos también que el espacio de estados se puede particionar en "clases de comunicación". En particular (el caso que nos interesa), la cadena es *irreducible* si todo par de estados distintos $i, j \in \mathcal{A}$ se comunican entre sí, es decir, existen $v_1, \ldots, v_n \in \mathcal{A}$ (el número n depende de i y j) tales que $v_1 = i$, $v_n = j$ y $P_{v_1,v_2} \cdots P_{v_{n-1}v_n} > 0$. En este caso, como $|\mathcal{A}| < \infty$, resulta que la cadena es *recurrente positiva*, por lo tanto existe una única *distribución estacionaria* π (es decir, un vector de probabilidad $\pi = (\pi_v)_{v \in \mathcal{A}}$ que es además un vector característico izquierdo de P asociado al valor característico 1 (por ser P estocástica e irreducible, el sistema característico $(\pi, 1)$ es precisamente el *sistema Perron de* P). Vamos a definir la medida de Markov μ_P, primero definiéndola en los conjuntos cilindros (que también son conjuntos borelianos), de la siguiente manera. Para toda $n \geq 1$, si $v_1 \ldots v_n \in B_n(\widehat{X}_A)$ y $k \in \mathbb{Z}$, definimos

$$\mu_P(\mathcal{C}_k(v_1 \ldots v_n)) = \pi_{v_1} \prod_{i=1}^{n-1} P_{v_i, v_{i+1}}$$

(observemos que la medida del cilindro no depende de $k \in \mathbb{Z}$, lo cual resultará en la T-invarianza de μ_P). Aunque hasta ahora μ_P está definida únicamente para conjuntos cilindro, es posible demostrar que se puede extender a toda la σ-álgebra de Borel, usando el Teorema de Extensión de Caratheodory.

Fijenmonos en el shift-completo X con alfabeto $\mathcal{E}$ (donde este último es el espacio de estados de la cadena). Un poco de notación usual ayudara a entender mejor el desarrollo de lo que viene a continuación. Para cada $k \in \mathbb{Z}$ denotamos a la k-ésima proyección $\pi_k \colon X \to \mathcal{E}$ con la siguiente regla de correspondencia:

$$\pi_k(x) = x_k$$

En una primera vista conocemos a los cilindros (también son conjuntos borelianos) y en vista de lo anterior podriamos intentar definir nuestra medida en terminos de ellos como anteriormente se mencionó, como sigue

$$\mu_S(\mathcal{C}_i(x_{n_1} \cdots x_{n_r})) = \pi(x_{n_1}) \prod_{i=1}^{r-1} P_{x_{n_i} x_{n_{i+1}}},$$

pero esta familia de conjuntos lo mas que llega a cumplir (en terminos de definir una medida y que se pueda extender) es que es cerrada bajo intersecciones, pero a continuación tenemos otra familia similar pero que si cumple ciertas propiedades útiles para hacer esto.

Sea

$$\mathcal{S} = \{ \bigcup_{i=1}^{n} \mathcal{C}_i \mid \mathcal{C}_i = \bigcap_{j=1}^{r} \pi_{n_j}^{-1}(x_{n_j}^i) \text{donde } x_{n_j}^i \in \mathcal{E} \ \forall i = 1 \cdots n \text{ para algunos } n_1, \cdots, n_r \in \mathbb{Z} \}$$

Si $\mathcal{C}_i = \bigcap_{j=1}^{r} \pi_{n_j}^{-1}(x_{n_j}^i)$, diremos que $\mathcal{C}_i$ fija las coordenadas $n_1, \cdots, n_r$ con los símbolos $x_{n_j}^i$.

Observación 5.1.1. $\mathcal{S}$ es una semialgebra de X y podemos definir una medida ahi mismo.

Demostración.

1. Es claro que $\emptyset, X \in \mathcal{S}$.

2. Si $A, B \in \mathcal{S}$, con $A = \bigcup_{i=1}^{n} \mathcal{A}_i$ y $B = \bigcup_{i=1}^{n} \mathcal{B}_i$ entonces

$$\begin{aligned} A \cap B &= \textstyle\bigcup_{i=1}^{n} \mathcal{A}_i \cap \bigcup_{i=1}^{m} \mathcal{B}_i \\ &= \textstyle\bigcup_{i=1}^{n} \bigcup_{j=1}^{m} \mathcal{A}_i \cap \mathcal{B}_j = \bigcup_{i=1,j=1}^{n,m} \mathcal{A}_i \cap \mathcal{B}_j. \end{aligned}$$

   Es claro que $\mathcal{A}_i \cap \mathcal{B}_j$ es de esta forma

$$\bigcap_{s=1}^{r} \pi_{n_s}^{-1}(x_{n_s}^{i,j})$$

   donde $\{n_1, \cdots, n_r\}$ es la union de los dos conjuntos de subíndices ocupados para A y B (es decir $n_1^A, \cdots, n_r^A$ y $n_1^B, \cdots, n_r^B$ respectivamente), con lo cual $A \cap B \in \mathcal{S}$

3.

$$A = \bigcup_{i=1}^{n} \bigcap_{r=1}^{s} \pi_{n_r}^{-1}(x_{n_r}^i) \subseteq \bigcup_{i=1}^{n} \bigcap_{r=1}^{s-1} \pi_{n_r}^{-1}(x_{n_r}^i) \subseteq \cdots \subseteq \bigcup_{i=1}^{n} \bigcap_{r=1}^{s-l} \pi_{n_r}^{-1}(x_{n_r}^i) \subseteq \cdots \subseteq \bigcup_{i=1}^{n} \pi_{n_1}^{-1}(x_{n_1}^i) \subseteq X.$$

Sea $A_l = \bigcup_{i=1}^{n} \bigcap_{r=1}^{s-l} \pi_{n_r}^{-1}(x_{n_r}^i)$, ahora veamos que

$$
\begin{aligned}
A_l \backslash A_{l-1} &= \bigcup_{i=1}^{n} \bigcap_{r=1}^{s-l} \pi_{n_r}^{-1}(x_{n_r}^i) \cap \left(\bigcup_{j=1}^{n} \bigcap_{p=1}^{s-l+1} \pi_{n_p}^{-1}(x_{n_p}^j) \right)^c \\
&= \bigcup_{i=1}^{n} \bigcap_{r=1}^{s-l} \pi_{n_r}^{-1}(x_{n_r}^i) \cap \bigcap_{j=1}^{n} \bigcup_{p=1}^{s-l+1} (\pi_{n_p}^{-1}(x_{n_p}^j))^c \\
&= \bigcup_{i=1}^{n} \bigcap_{r=1}^{s-l} \pi_{n_r}^{-1}(x_{n_r}^i) \cap \bigcup_{p=1}^{s-l+1} \bigcap_{j=1}^{n} (\pi_{n_p}^{-1}(x_{n_p}^j))^c \\
&= \bigcup_{i=1}^{n} \bigcap_{r=1}^{s-l} \pi_{n_r}^{-1}(x_{n_r}^i) \cap \bigcup_{p=1}^{s-l+1} \pi_{n_p}^{-1} \left((\bigcup_{j=1}^{n} x_{n_p}^j)^c \right) \quad , \\
&= \bigcup_{i=1}^{n} \bigcap_{r=1}^{s-l} \bigcup_{p=1}^{s-l+1} \pi_{n_r}^{-1}(x_{n_r}^i) \cap \pi_{n_p}^{-1} \left((\bigcup_{j=1}^{n} x_{n_p}^j)^c \right) \\
&= \bigcup_{i=1}^{n} \bigcap_{r=1}^{s-l} \pi_{n_r}^{-1}(x_{n_r}^i) \cap \pi_{n_{s-l+1}}^{-1} \left((\bigcup_{j=1}^{n} x_{n_{s-l+1}}^j)^c \right) \\
&= \bigcup_{z \neq x_{n_{s-l+1}}^j \ \forall j} \ \bigcup_{i=1}^{n} \bigcap_{r=1}^{s-l} \pi_{n_r}^{-1}(x_{n_r}^i) \cap \pi_{n_{s-l+1}}^{-1}(z) \in \mathcal{S}.
\end{aligned}
$$

lo cual demuestra que $\mathcal{S}$ es una semialgebra.

Si $A = \bigcup_{i=1}^{n} \bigcap_{r=1}^{s} \pi_{n_r}^{-1}(x_{n_r}^i)$ (s. p. g. podemos suponer n_r creciente), definimos

$$
\mu_{\mathcal{S}}(A) = \sum_{i=1}^{n} \pi(x_{n_1}^i) \prod_{r=1}^{s} P_{x_{n_r}^i x_{n_{r+1}}^i}^{n_{r+1}-n_r}
$$

si $A \neq \emptyset$, y $\mu_{\mathcal{S}}(\emptyset) = 0$.

Basta probar que $\mu_{\mathcal{S}}$ es σ-aditiva y tendremos que es una medida.

Sea $(A_i)_{i \in \mathbb{N}} \subseteq \mathcal{S}$ tal que $A = \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{S}$, pero $A_i \in \mathcal{S}$ $\forall i \in \mathbb{N}$ entonces $A_i = \bigcup_{j=1}^{n^{A_i}} A_{i,j}$ $\forall i \in \mathbb{N}$ con $A_{i,j} = \bigcap_{r=1}^{s_i} \pi_{n_r^{A_i}}^{-1}(x_{n_r^{A_i}}^{j,i})$ (Nota que $A = \bigcup_{i \in \mathbb{N}} \bigcup_{j=1}^{n^{A_i}} A_{i,j}$), además $A \in \mathcal{S}$ y como consecuencia $A = \bigcup_{t=1}^{n^A} \bigcap_{r=1}^{s} \pi_{n_r^A}^{-1}(x_{n_r^A}^t)$, como $A_i \subseteq A$ debe ocurrir que $\{n_1^A, \cdots, n_r^A\} \subseteq \{n_1^{A_i}, \cdots, n_r^{A_i}\}$ y además que para cada t existe s tal que $x_{n_r^A}^{t,i} = x_{n_r^A}^s$. Para demostrar la σ-aditividad solo hay que hacer notar que para i, j, siendo $A_{i,j} = \bigcap_{r=1}^{m} \pi_{n_r^{A_i}}^{-1}(x_{n_r^{A_i}}^{j,i})$ (es decir $A_{i,j}$ fija unicamente m coordenadas) y como consecuencia el conjunto

$D_{i,j} = \{\, n\colon A_{l,s}$ fija $n \geq m$ coordenadas, ademas coincide con todas las coordenadas que fija $A_{r,p}$
salvo la última, que también fija pero con otro símbolo ,esto si $A_{r,p}$
fija mas o el mismo número de coordenadas que $A_{i,j}$ pero menos que $A_{l,s}$ $\}$

es acotado, nota que el hecho de ser $D_{i,j}$ acotado y que P^h es una matriz de probabilidad nos garantiza la σ-aditividad, esto debido a que como la union está en la semi-algebra entonces todos los uniendos fijan un patron común de coordenadas, todos fijando los mismos estados(símbolos) en cada coordenada.
Veamos que $D_{i,j}$ es acotado, en caso de no serlo $D_{i,j}$ haremos un pequeño algoritmo que nos haria obtener una sucesión que no puede existir.
El algoritmo se sigue asi $A_{i,j}$ fija unicamente m_0 coordenadas, entonces debe existir $A^1_{i,j}$ de los uniendos que forman a A que haga lo mismo, solo que fija m_1 con $m_1 > m_0$ y además fija los mismos $m_0 - 1$ símbolos iniciales, de lo contrario la union completa no resultaría A , hacemos lo mismo para $A^1_{i,j}$ obtenemos $A^2_{i,j}$ que fija m_2 con $m_2 > m_1$ y además fija los mismos $m_1 - 1$ símbolos iniciales, como $D_{i,j}$ no es acotado podemos hacer esto una infinidad de veces , encontrando para cada natural n un $A^n_{i,j}$ que cumple recursivamente que fija m_n con $m_n > m_{n-1}$ y además fija los mismos $m_{n-1} - 1$ símbolos iniciales que fija $A^{n-1}_{i,j}$y el último no , es muy importante hacer notar que al ser ajenos los $A_{i,j}$'s, tenemos la garantía de que no hay una $A_{s,t}$ que haga lo mismo que $A^1_{i,j}$ pero el total de coordenadas fijadas sea el mismo número de coordenadas que $A_{i,j}$.
Sea $z \in A$, definimos a $y = (y_i)$ con $y_{n_r^{m_s}} = x^j_{n_r^{m_s}}$ donde $x^j_{n_r^{m_s}}$ es el primer símbolo que fija $A^s_{i,j}$ distinto de los fijados por $A^{s-1}_{i,j}$, $y_i = z_i \; \forall i \neq n_r^{m_s}, \forall s \in \mathbb{N}$, de aquí que es claro que $y \in A$ pues $z \in A$, pero $y \notin \bigcup_{i\in\mathbb{N}} A_i$ con esto se concluye que ese conjunto $D_{i,j}$ es acotado lo que concluye con la σ-aditividad. □

(Al hablar de estados me refiero a los estados de la cadena que son los símbolos del shift.)
A raíz de esto tenemos una serie de pasos que se hace para llegar a que esta medida se puede extender al algebra generada por nuestra semi-algebra (uniones ajenas finitas de nuestra semi-algebra, la medida se extiende como la suma de las medidas de los uniendos.)

Definición 5.1.1. Sean $\mathcal{A}$ una álgebra y $\mu : \mathcal{A} \longrightarrow \overline{\mathbb{R}}$ una casi medida sobre los reales extendidos.Definimos $\mu^* : \mathcal{P}(X) \longrightarrow \overline{\mathbb{R}}$ como sigue

$$\mu^*(E) = \text{ínf}\left\{\sum_{n=1}^{\infty}\mu(A_n) : E \subset \bigcup_{n=1}^{\infty} A_n, A_n \in \mathcal{A}\; n \in \mathbb{N}\right\}$$

para todo $E \in X$, se prueba que μ^* es una medida exterior en un curso de análisis ó teoría de la medida, de hecho todo esto viene obtenido de [1] en particular el teorema de extensión de caratheodory se ubica en la página 86, y a esta se le conoce como la medida exterior generada por μ.

Ahora por el teorema de extensión de caratheodory, la σ-algebra de caratheodory

$$\mathcal{A}^* = \{E \subset X : \mu^*(B) = \mu^*(B \cap E) + \mu^*(B - E) \text{ para toda } B \subset X\}$$

generada por las algebra y medida antes mencionadas contiene a la σ-algebra de borel de dicha topología y es completa con $\overline{\mu} = \mu^*|_{\mathcal{A}^*}$ como medida completa, la existencia de esta medida nos garantiza que $\mathcal{M}(X, T, \mu_S) \neq \emptyset$, lo que por el teorema A.2.8 podemos asumir que esta es ergódica. Ahora por todos estos resultados podemos obtener un espacio de probabilidad $(X, \mathcal{A}^*, \overline{\mu})$ completo con $\overline{\mu}$ T-invariante y ergódica, además T continua y $\overline{\mu}$-medible, entonces $(X, \mathcal{A}^*, \overline{\mu}, T, \mathbb{Z})$ es un sistema dinámico y su semigrupo de automorfismos es un grupo por que la T es biyectiva,de ahora en adelante tomaremos $\mu_P = \overline{\mu}$.

Observación 5.1.2. La σ-algebra de *caratheodory* $\mathcal{A}^*$, solamente es la compleción $(\overline{B_X(\mathcal{C})})$ de la σ-algebra de borel generada por los cilindros, debido a que la medida es una medida de probabilidad, por tanto finita, por ende σ-finita y $B_X(\mathcal{C})$ como se define en un curso básico de teoria de la medida (una referencia bastante buena es [1] ejercicio 36), además $B_X(\mathcal{C}) = \left\{ \bigcup_{j\in J} \bigcap_{i\in I} \mathcal{C}_{i,j} \mid \mathcal{C}_{i,j} \text{ es un cilindro} \right\}$. Esto es muy sencillo verlo por que el complemento de un cilindro es una unión de cilindros que fijan las coordenadas que el primero fija pero con los símbolos que no ocupa el, el termino de intersección sale por consecuencia de las leyes de Demorgan .

5.2. Teorema de clasificación de cadenas de Markov homogéneas (primera implicación)

En esta sección alcanzamos nuestro objetivo principal de este documento, que es analizar al Teorema de Clasificación 4.4.1 en el contexto métrico.

Recordemos que si P es una matriz estocástica, P^0 es tal que $P^0_{i,j} = 1$ si $P_{i,j} \neq 0$, y en otro caso es cero.

Definición 5.2.1. Sean $(X_n)_{n\in\mathbb{Z}}$ y $(Y_n)_{n\in\mathbb{Z}}$ dos cadenas de Markov irreducibles (finitas), con matrices de probabilidades de transición P y Q respectivamente, y $A = P^0$ y $B = Q^0$. Diremos que $(\widehat{X}_A, T_A, \mu_P)$ y $(\widehat{X}_B, T_B, \mu_Q)$ son *conjugadas* si exite una conjugación $\phi\colon \widehat{X}_A \to \widehat{X}_B$ que preserva la medida, es decir, $\forall\ Z \subset \widehat{X}_B$ medible, $\mu_P(\phi^{-1}(Z)) = \mu_Q(Z)$.

Definición 5.2.2. Diremos que dos cadenas de Markov con matrices de transición P y Q son *estocásticamente shift-fuertemente equivalentes* si, existen matrices estocásticas $U_i, V_i, i = 0, \cdots, n$ tales que

$$P = U_0V_0,\ V_0U_0 = U_1V_1,\ V_1U_1 = U_2V_2, \cdots, V_{n-1}U_{n-1} = U_nV_n = Q \tag{5.1}$$

y

$$P^0 = U_0^0V_0^0,\ V_0^0U_0^0 = U_1^0V_1^0,\ V_1^0U_1^0 = U_2^0V_2^0, \cdots, V_{n-1}^0U_{n-1}^0 = U_n^0V_n^0 = Q^0. \tag{5.2}$$

Nos referiremos a las series de igualdades (5.1) y (5.2) como las *series paralelas*. Presentaremos ahora la demostración de una de las implicaciones del Teorema de Clasificación. La prueba que daremos diferirá del argumento clásico usado en [3], la razón de ésto es que el argumento que presentamos (Parry y Williams) se puede generalizar al caso métrico.

5.2.1. Caso topológico

Teorema 5.2.1. *Sean $(X_n)_{n\in\mathbb{Z}}$ y $(Y_n)_{n\in\mathbb{Z}}$ dos cadenas de Markov irreducibles con espacio de estados finito y matrices de probabilidades de transición P, Q respectivamente. Si $(X_n)_{n\in\mathbb{Z}}$ y $(Y_n)_{n\in\mathbb{Z}}$ son conjugadas, entonces $P^0 = A$, $Q^0 = B$ son shift-fuertemente equivalentes.*

Antes de presentar la demostración de este teorema, daremos algunas definiciones y probaremos unos lemas auxiliares.

Sea k el número de estados de $(X_n)_{n\in\mathbb{Z}}$ y l el de $(Y_n)_{n\in\mathbb{Z}}$. Llamémosle a los diferentes estados $1, ..., k$ y $1, ..., l$. Como las cadenas son conjugadas, existe una conjugación $\varphi\colon X_A \to X_B$, es decir, un homeomorfismo que hace conmutativo el siguiente diagrama

$$\begin{array}{ccc} X_A & \xrightarrow{T_A} & X_A \\ \varphi \Big\downarrow & & \Big\downarrow \varphi. \\ X_B & \xleftarrow{T_B} & X_B \end{array}$$

Sea

$$A_i = \{x \in X_A \mid x_0 = i\}$$

para $i = 1, ..., k$ y

$$B_i = \{y \in X_B \mid y_0 = j\}$$

para $j = 1, ..., l$. Definimos las siguientes particiones $\alpha = (A_i)_{i=1}^k$ y $\beta = (B_i)_{i=1}^l$ (dichas particiones se denominan *naturales*). Los elementos de dichas particiones son imágenes inversas continuas de abiertos y cerrados, por lo que son particiones abiertas y cerradas. Por simplicidad de notación, sea $T = T_A$.

Lema 5.2.1. *Bajo las hipótesis y definiciones anteriores* $\exists\, m \in \mathbb{N}$ *tal que si* $\psi = \varphi \circ T^m$ *y* $\gamma = \psi^{-1}(\beta)$, *entonces*

$$\bigvee_{i=0}^{2m} T^{-i}(\alpha) \succeq \gamma \quad \textit{y también} \quad \bigvee_{i=0}^{2m} T^{-i}(\gamma) \succeq T^{-2m}(\alpha).$$

Demostración. Como B_i es tanto abierto como cerrado y φ es un homeomorfismo (por ser conjugación), entonces $\varphi^{-1}(B_i)$ es abierto y cerrado. Como $\bigcup_{r\in\mathbb{Z}} T^{-r}(\alpha)$ es una subbase de la topología, $\bigvee_{r\in\mathbb{Z}} T^{-r}(\alpha)$ continene una base para la topología generada por $(T^{-r}(\alpha))_{r\in\mathbb{Z}}$ llamada $\overline{\beta}((T^{-i}(\alpha))_{i\in\mathbb{Z}})$. Entonces existe $\mathcal{U}_j^i \in \overline{\beta}((T^{-r}(\alpha))_{r\in\mathbb{Z}})$ donde $j \in I^i$ para I^i conjunto de índices tal que $\varphi^{-1}(B_i) = \bigcup_{j\in I^i} \mathcal{U}_j^i$, pero como $\varphi^{-1}(B_i)$ es cerrado y el espacio total es un espacio shift y por lo tanto compacto. Por ende $\varphi^{-1}(B_i)$ también es compacto, y entonces la cubierta de $\varphi^{-1}(B_i)$ descrita anteriormente tiene una subcubierta finita, es decir, existen $j_1, ..., j_{s^i}$ tales que

$$\varphi^{-1}(B_i) \subseteq \bigcup_{t=1}^{s^i} \mathcal{U}_{j_t}^i.$$

Como $\mathcal{U}_{j_t}^i \in \overline{\beta}((T^{-i}(\alpha))_{i\in\mathbb{Z}})$, para cada j_t existen $w_{j_t}^{1,i}, w_{j_t}^{2,i}, \cdots, w_{j_t}^{h_{j_t},i}$ tal que $\mathcal{U}_{j_t}^i = \bigcap_{q=1}^{h_{j_t}} F_{w_{j_t}^{q,i}}^i$, con $F_{w_{j_t}^{q,i}}^i \in T^{-w_{j_t}^{q,i}}(\alpha)$ para cada q. Con ésto tenemos que

$$\varphi^{-1}(B_i) = \bigcup_{t=1}^{s^i} \bigcap_{q=1}^{h_{j_t}} F_{w_{j_t}^q}^i.$$

Si

$$m_1 = \text{máx}\left\{w_{j_t}^{q,i} | q \in \{1, ..., h_{j_t}\}, t \in \{1, ..., s^i\}, i \in \{1, ..., l\}\right\},$$

entonces

$$\bigcap_{q=1}^{h_{j_t}} F^i_{w^{q,i}_{j_t}} \in \bigvee_{r=-m_1}^{m_1} T^r(\alpha) \quad \forall\, t \in \{1, ..., s^i\}, i \in \{1, ..., l\},$$

lo que demuestra que existe

$$C \in \bigvee_{r=-m_1}^{m_1} T^r(\alpha)$$

tal que

$$C \subseteq \varphi^{-1}(B_i),$$

con lo cual

$$\varphi^{-1}(\beta) \preceq \bigvee_{r=-m_1}^{m_1} T^r(\alpha).$$

Como φ es un homeomorfismo y $\bigcup_{i\in\mathbb{Z}} T_B^{-i}(\beta)$ es una subbase, $\bigcup_{i\in\mathbb{Z}} \varphi^{-1}(T_B^{-i}(\beta))$ también es una subbase. Como φ es conjugación, su inversa también es código de bloques, por tanto es una conjugación debido al Teorema 1.5.2. Así, la inversa conmuta con la función shift, por la misma razón también conmuta con el inverso de la función shift, por lo que obtenemos que $\bigcup_{i\in\mathbb{Z}} T^{-i}(\varphi^{-1}(\beta))$ es subbase de la topología. Debido a lo anterior, $\bigvee_{i\in\mathbb{Z}} T^{-i}(\varphi^{-1}(\beta))$ contiene una base para la topología generada por $(T^{-i}(\varphi^{-1}(\beta)))_{i\in\mathbb{Z}}$ llamada $\overline{\beta}((T^{-r}(\varphi^{-1}(\beta)))_{r\in\mathbb{Z}})$. Como A_i es abierto, existen $\mathcal{V}^i_j \in \overline{\beta}((T^{-r}(\varphi^{-1}(\beta)))_{r\in\mathbb{Z}})$ donde $j \in I^i$ para I^i conjunto de índices tal que $A_i = \bigcup\limits_{j\in I^i} \mathcal{V}^i_j$, pero como A_i es cerrado y el espacio total es un espacio shift, debido al Teorema 4.1.1, el espacio total es compacto y por ende A_i también es compacto, entonces esa cubierta de A_i tiene una subcubierta finita, es decir, existen $j_1, ..., j_{s^i}$ tales que

$$A_i \subseteq \bigcup_{t=1}^{s^i} \mathcal{V}^i_{j_t}.$$

Como $\mathcal{V}^i_{j_t} \in \overline{\beta}((T^{-r}(\varphi^{-1}(\beta)))_{r\in\mathbb{Z}})$ para cada j_t, existen $w^{1,i}_{j_t}, w^{2,i}_{j_t}, \cdots, w^{h_{j_t},i}_{j_t}$ tales que $\mathcal{V}^i_{j_t} = \bigcap\limits_{q=1}^{h_{j_t}} G^i_{w^{q,i}_{j_t}}$ con $G^i_{w^{q,i}_{j_t}} \in T^{-w^{q,i}_{j_t}}(\varphi^{-1}(\beta))$ para cada q. Entonces

$$A_i \subseteq \bigcup_{t=1}^{s^i} \bigcap_{q=1}^{h_{j_t}} G^i_{w^q_{j_t}}.$$

Si

$$m_2 = \max\left\{w^{q,i}_{j_t} | q \in \{1, ..., h_{j_t}\}, t \in \{1, ..., s^i\}, i \in \{1, ..., k\}\right\},$$

entonces

$$\bigcap_{q=1}^{h_{j_t}} G^i_{w^{q,i}_{j_t}} \in \bigvee_{r=-m_2}^{m_2} T^r(\varphi^{-1}(\beta)) \quad \forall t\ \in \{1, ..., s^i\}, i \in \{1, ..., k\},$$

demostrando que existe
$$D \in \bigvee_{r=-m_2}^{m_2} T^r(\varphi^{-1}(\beta))$$
tal que
$$C \subseteq A_i,$$
con lo cual
$$\alpha \preceq \bigvee_{r=-m_2}^{m_2} T^r(\varphi^{-1}(\beta)).$$
Sea $m = \max\{m_1, m_2\}$. Entonces
$$\varphi^{-1}(\beta) \preceq \bigvee_{r=-m}^{m} T^r(\alpha)$$
y
$$\alpha \preceq \bigvee_{r=-m}^{m} T^r(\varphi^{-1}(\beta)).$$
Por la Proposición 4.2.1 tenemos que
$$T^{-m}(\varphi^{-1}(\beta)) \preceq T^{-m}(\bigvee_{r=-m}^{m} T^r(\alpha)) = \bigvee_{r=-m}^{m} T^{r-m}(\alpha) = \bigvee_{r=0}^{2m} T^{-r}(\alpha).$$
Entonces
$$\gamma = (\varphi \circ T^m)^{-1}(\beta) = T^{-m}(\varphi^{-1}(\beta)) \preceq \bigvee_{r=0}^{2m} T^{-r}(\alpha)$$
y con ésto
$$\bigvee_{i=0}^{2m} T^{-i}(\alpha) \succeq \gamma.$$
De la misma manera tenemos que
$$T^{-2m}(\alpha) \preceq T^{-2m}(\bigvee_{r=-m}^{m} T^r(\varphi^{-1}(\beta))) = \bigvee_{r=-m}^{m} T^{r-2m}(\varphi^{-1}(\beta))$$
$$= \bigvee_{r=-m}^{m} T^{r-m}(T^{-m}\varphi^{-1}(\beta)) = \bigvee_{r=0}^{2m} T^{-r}(\gamma)$$
lo que termina con la prueba de este primer lema. □

Definición 5.2.3. Sean ξ y η dos particiones cerradas y abiertas. Diremos que :

1. ξ es *equivalente inferiormente* de η si
$$\xi \preceq \eta \preceq \xi \vee T^{-1}(\xi),$$
denotado como $\xi \xrightarrow[1]{} \eta$.

2. ξ es *equivalente superiormente* de η si
$$T^{-1}(\xi) \preceq \eta \preceq \xi \vee T^{-1}(\xi),$$
denotado como $\xi \xrightarrow{1} \eta$.

3. ξ es *equivalente* por n *pasos inferiormente* de η si existen particiones $\xi_1, \cdots, \xi_{n-1}$ tales que $\xi \xrightarrow[1]{} \xi_1$, $\xi_i \xrightarrow[1]{} \xi_{i+1}$ con $i = 1, ..., n-1$ y $\xi_{n-1} \xrightarrow[1]{} \eta$, denotado como $\xi \xrightarrow[n]{} \eta$.

4. ξ es *equivalente* por n *pasos superiormente* de η si existen particiones $\xi_1, \cdots, \xi_{n-1}$ tales que $\xi \xrightarrow{1} \xi_1$, $\xi_i \xrightarrow{1} \xi_{i+1}$ con $i = 1, ..., n-1$ y $\xi_{n-1} \xrightarrow{1} \eta$, denotado como $\xi \xrightarrow{n} \eta$.

Lema 5.2.2. *Sean A una matriz de entradas enteras no negativas menores o iguales a 1 y X_A su cadena de Markov topológica, es decir el shift de vértices de la gráfica asociada a la matriz, α, γ particiones abiertas y cerradas de X_A tales que*
$$\alpha^{2m} = \bigvee_{i=0}^{2m} T^{-i}(\alpha) \succeq \gamma \quad \textit{y también} \quad \gamma^{2m} = \bigvee_{i=0}^{2m} T^{-i}(\gamma) \succeq T^{-2m}(\alpha).$$

Entonces
$$\alpha \xrightarrow[2m-1]{} \alpha^{2m-1} \xrightarrow[2m]{} \alpha^{2m-1} \vee \gamma^{2m-1} \xrightarrow{2m} \gamma^{4m-1} \xleftarrow[4m-1]{} \gamma.$$

Demostración. Para probar que $\alpha \xrightarrow[2m-1]{} \alpha^{2m-1}$, sea $\xi^n = \bigvee_{i=0}^{n} T^{-i}(\alpha)$. Como T es una conjugación,
$$\begin{aligned}
\xi^i \vee T^{-1}(\xi^i) &= \xi^i \vee T^{-1}(\bigvee_{j=0}^{i} T^{-j}(\alpha)) = \xi^j \vee \bigvee_{j=0}^{i} T^{-(j+1)}(\alpha) \\
&= \xi^i \vee \bigvee_{j=1}^{i+1} T^{-j}(\alpha) = \xi^i \vee \bigvee_{j=1}^{i+1} T^{-j}(\alpha) \\
&= \bigvee_{j=0}^{i} T^{-j}(\alpha) \vee \bigvee_{j=1}^{i+1} T^{-j}(\alpha) = \bigvee_{j=0}^{i+1} T^{-j}(\alpha) \\
&= \xi^{i+1},
\end{aligned}$$
por ende $\xi^i \vee T^{-1}(\xi^i) = \xi^{i+1}$ y entonces
$$\xi^i \preceq \xi^{i+1} = \xi^i \vee T^{-1}(\xi^i) \preceq \xi^i \vee T^{-1}(\xi^i),$$
por lo tanto
$$\xi_i \xrightarrow[1]{} \xi_{i+1}.$$
Análogo a lo anterior vemos que
$$\alpha \preceq \xi_1 = \alpha \vee T^{-1}(\alpha) \preceq \alpha \vee T^{-1}(\alpha),$$

con lo que llegamos a $\alpha \xrightarrow[2m-1]{} \alpha^{2m-1}$. Para demostrar

$$\alpha^{2m-1} \xrightarrow[1]{} \alpha^{2m-1} \vee \gamma$$

usamos la hipótesis de que

$$\gamma \preceq \alpha^{2m}.$$

Debido al Teorema 3.1.3 tenemos que

$$\gamma \vee \alpha^{2m} = \alpha^{2m}$$

y observando que

$$\alpha^{2m-1} \preceq \alpha^{2m-1} \vee \gamma \preceq \alpha^{2m} \vee \gamma = \alpha^{2m} = \alpha^{2m-1} \vee T^{-1}(\alpha^{2m-1})$$

obtenemos

$$\alpha^{2m-1} \xrightarrow[1]{} \alpha^{2m-1} \vee \gamma.$$

De la misma manera, como anteriormente, siendo s una partición, definimos $s^n = \bigvee_{i=0}^{n} T^{-i}(s)$. Similarmente, como se probó antes que $\xi^i \vee T^{-1}(\xi^i) = \xi^{i+1}$, se prueba $(s^i)^1 = s^i \vee T^{-1}(s^i) = s^{i+1}$ (nota que $s^0 = s$). Entonces $\xi^n = \alpha^{2m-1} \vee \gamma^n$, y así

$$\xi^n = \alpha^{2m-1} \vee \gamma^n \preceq \alpha^{2m-1} \vee \gamma^{n+1}$$

a causa de

$$\gamma^n \preceq \gamma^{n+1}.$$

Luego

$$\begin{aligned}
\xi^n \vee T^{-1}(\xi^n) &= \alpha^{2m-1} \vee \gamma^n \vee T^{-1}(\alpha^{2m-1} \vee \gamma^n) \\
&= \alpha^{2m-1} \vee \gamma^n \vee T^{-1}(\alpha^{2m-1}) \vee T^{-1}(\gamma^n) \\
&= \alpha^{2m-1} \vee T^{-1}(\alpha^{2m-1}) \vee \gamma^n \vee T^{-1}(\gamma^n) \\
&= \alpha^{2m} \vee \gamma^{n+1}
\end{aligned}$$

y se sigue que

$$\xi^n = \alpha^{2m-1} \vee \gamma^n \preceq \alpha^{2m-1} \vee \gamma^{n+1} \preceq \alpha^{2m} \vee \gamma^{n+1} = \xi^n \vee T^{-1}(\xi^n),$$

por lo tanto

$$\xi_i \xrightarrow[1]{} \xi_{i+1}$$

para $i = 0, ..., 2m-2$. Tenemos entonces que

$$\alpha^{2m-1} \xrightarrow[2m]{} \alpha^{2m-1} \vee \gamma^{2m-1}.$$

Ahora probaremos que si ζ, η son dos particiones del espacio donde trabajamos, entonces $\zeta \vee \eta \xrightarrow{h} T^{-h}(\zeta) \vee \eta^h$. Hagamos esto por inducción: para $h = 1$ tenemos que

$$T^{-1}(\zeta \vee \eta) = T^{-1}(\zeta) \vee T^{-1}(\eta) \preceq T^{-1}(\zeta) \vee \eta \vee T^{-1}(\eta) = T^{-1}(\zeta) \vee \eta^1$$

$$\preceq \zeta \vee T^{-1}(\zeta) \vee \eta^1 = \zeta \vee \eta \vee T^{-1}(\zeta \vee \eta).$$

Supongamos por hipótesis de inducción que $\zeta \vee \eta \xrightarrow{h} T^{-h}(\zeta) \vee \eta^h$. Ahora, como la base de la inducción vale para cualquier par de particiones, tenemos que funciona para $T^{-h}(\zeta)$ y η^h. Entonces

$$T^{-h}(\zeta) \vee \eta^h \xrightarrow{h} T^{-1}(T^{-h}(\zeta)) \vee (\eta^h)^1 = T^{-(h+1)}(\zeta) \vee \eta^{h+1}$$

y de aquí es directo que

$$\zeta \vee \eta \xrightarrow{h+1} T^{-(h+1)}(\zeta) \vee \eta^{h+1}.$$

Usando lo anterior para α^{2m-1} y γ^{2m-1} tenemos que

$$\alpha^{2m-1} \vee \gamma^{2m-1} \xrightarrow{2m} T^{-2m}(\alpha^{2m-1}) \vee (\gamma^{2m-1})^{2m}$$

$$= T^{-2m}(\bigvee_{i=0}^{2m} T^{-i}(\alpha)) \vee \gamma^{4m-1} = \bigvee_{i=2m}^{4m} T^{-i}(\alpha) \vee \gamma^{4m-1}$$

y ahora ocupando la hipótesis de que

$$\gamma^{2m} = \bigvee_{i=0}^{2m} T^{-i}(\gamma) \succeq T^{-2m}(\alpha)$$

y que T es conjugación, nos da lo siguiente

$$T^{-s}(\gamma^{2m}) = \bigvee_{i=0}^{2m} T^{-(i+s)}(\gamma) \succeq T^{-(2m+s)}(\alpha),$$

es decir

$$T^{-s}(\gamma^{2m}) = \bigvee_{i=s}^{2m+s} T^{-i}(\gamma) \succeq T^{-(2m+s)}(\alpha) \tag{5.3}$$

para cualquier s entero con $0 \leq s \leq 2m-1$. Refinando las particiones en cada lado de (5.3) con $0 \leq s \leq 2m-1$ obtenemos

$$\bigvee_{s\in\{0,\cdots,2m-1\}} \bigvee_{i=s}^{2m+s} T^{-i}(\gamma) \succeq \bigvee_{s\in\{0,\cdots,2m-1\}} T^{-(2m+s)}(\alpha) = \bigvee_{s\in\{2m,\cdots,4m-1\}} T^{-s}(\alpha),$$

pero

$$\gamma^{4m-1} = \bigvee_{s\in\{0,\cdots,4m-1\}} T^{-i}(\gamma) \succeq \bigvee_{s\in\{2m,\cdots,4m-1\}} T^{-s}(\alpha)$$

y con ello se tiene

$$\bigvee_{i=2m}^{4m} T^{-i}(\alpha) \vee \gamma^{4m-1} = \gamma^{4m-1}$$

por lo tanto tenemos que

$$\alpha^{2m-1} \vee \gamma^{2m-1} \xrightarrow{2m} \gamma^{4m-1}.$$

Análogamente a la primera prueba, tenemos que

$$\gamma^{4m-1} \xleftarrow[4m-1]{} \gamma$$

lo que concluye la prueba del lema. □

Definición 5.2.4. Sean ξ y η dos particiones abiertas y cerradas de X. Definimos dos matrices indexadas por $\xi \times \eta$ como sigue:

$$(\xi, \eta)(K, E) = \begin{cases} 1 & \text{si } K \cap E \neq \emptyset \text{ para } K \in \xi, E \in \eta \\ 0 & \text{en otro caso} \end{cases}$$

y

$$(\xi, \eta)_T(K, E) = \begin{cases} 1 & \text{si } K \cap T^{-1}(E) \neq \emptyset \text{ para } K \in \xi, E \in \eta \\ 0 & \text{en otro caso.} \end{cases}$$

Definición 5.2.5. Sean (X, τ, T) un sistema dinamico topológico, μ una medida que lo vuelve sistema dinámico (es decir, T-invariante), y α una partición abierta y cerrada (bajo μ o μ-medible). Decimos que α cumple la *propiedad topológica de Markov* si $\forall\, n \in \mathbb{N}$ dados $A_0, \cdots, A_n \in \alpha$ tales que

$$A_0 \cap T^{-1}(A_1) \neq \emptyset \text{ y } A_1 \cap T^{-1}(A_2) \cap \cdots \cap T^{-(n-1)}(A_n) \neq \emptyset,$$

se tiene que

$$A_0 \cap T^{-1}(A_1) \cap T^{-2}(A_2) \cap \cdots \cap T^{-n}(A_n) \neq \emptyset.$$

Observación 5.2.1. Sean α, β son dos particiones abiertas y cerradas tales que $\alpha \xrightarrow[1]{} \beta$. Entonces

1. $\alpha \vee T^{-1}(\alpha) = \beta \vee T^{-1}(\alpha)$ y $\alpha \vee T^{-1}(\beta) = \beta \vee T^{-1}(\beta)$.

2. $\forall A \cap T^{-1}(C) \in \alpha \vee T^{-1}(\alpha)$ con $A, C \in \alpha$, existe $B \in \beta$ tal que $B \subseteq A$ y

$$A \cap T^{-1}(C) = B \cap T^{-1}(C).$$

3. $\forall\, F \in \beta$, existe $D \in \alpha$ tal que $F \subset D$ y

$$F = \bigcup_{C \in \sigma \subset \alpha} D \cap T^{-1}(C).$$

4. Si además α cumple la propiedad topológica de Markov, entonces β también cumple la propiedad topológica de Markov.

Demostración. Como $\alpha \xrightarrow[1]{} \eta$, $\alpha \vee T^{-1}(\alpha) \succeq \eta \succeq \alpha$, la Proposición 3.1.4 implica que

$$\eta \vee T^{-1}(\alpha) \preceq \alpha \vee T^{-1}(\alpha) \vee T^{-1}(\alpha) = \alpha \vee T^{-1}(\alpha).$$

Como $\alpha \preceq \eta$, obtenemos

$$\alpha \vee T^{-1}(\alpha) \preceq \eta \vee T^{-1}(\alpha),$$

con lo cual, por la Proposición 3.1.2, se sigue

$$\alpha \vee T^{-1}(\alpha) = \eta \vee T^{-1}(\alpha).$$

Ahora como $\alpha \preceq \beta$,

$$\alpha \vee T^{-1}(\beta) \preceq \eta \vee T^{-1}(\beta),$$

y por la misma razón tenemos que $T^{-1}(\alpha) \preceq T^{-1}(\beta)$ y como $\beta \preceq \alpha \vee T^{-1}(\alpha)$ obtenemos

$$\beta \vee T^{-1}(\beta) \preceq \alpha \vee T^{-1}(\alpha) \vee T^{-1}(\beta) = \alpha \vee (T^{-1}(\alpha) \vee T^{-1}(\beta)) = \alpha \vee T^{-1}(\beta),$$

con lo cual concluimos

$$\alpha \vee T^{-1}(\beta) = \beta \vee T^{-1}(\beta).$$

Para 2, sean $A, B \in \alpha$ tales que $A \cap T^{-1}(B) \neq \emptyset$. Entonces

$$A \cap T^{-1}(B) \in \alpha \vee T^{-1}(\alpha) = \eta \vee T^{-1}(\alpha),$$

y como $\alpha \preceq \eta$, podemos ver que

$$A = \bigcup_{E \in \Sigma \subseteq \eta} E$$

y

$$A \cap T^{-1}(B) = \bigcup_{E \in \Sigma \subseteq \eta} (E \cap T^{-1}(B)).$$

Como $A \cap T^{-1}(B) \neq \emptyset$, existe $E \in \Sigma \subseteq \eta$ tal que $E \cap T^{-1}(B) \neq \emptyset$, con ello

$$E \cap T^{-1}(B) \subseteq A \cap T^{-1}(B),$$

pero como ambos son miembros de la misma partición se tiene $E \cap T^{-1}(B) = A \cap T^{-1}(B)$.

Para 3, como $F \in \beta \preceq \alpha \vee T^{-1}(\alpha)$,

$$F = \bigcup_{D_i \in \Sigma_1 H_i \in \Sigma_2, \Sigma_i \subset \alpha} D_i \cap T^{-1}(H_i),$$

pero por 2 tenemos que para cada D_i existe $F_i \in \eta$ con $F_i \subset D_i$ tal que $D_i \cap T^{-1}(H_i) = F_i \cap T^{-1}(H_i)$. Así, $F_i \cap F \neq \emptyset$ con lo que $F = F_i$ para toda i, como consecuencia $D_i \supset F$ y por ello $D_i \cap D_j \neq \emptyset$ para todo par de subíndices i, j. A causa de ésto y del hecho de que α es una partición tenemos que $D_i = D$ o sea que $F \subset D$.

Para el último punto, sin pérdida de generalidad, probémoslo para $n = 1$ (para n's más grandes se hace la prueba análoga). Sean $A, B, C \in \beta$ tales que

$$A \cap T^{-1}(B) \neq \emptyset \text{ y } B \cap T^{-1}(C) \neq \emptyset.$$

Por 2 existe $E, F \in \alpha$ contenidos en A, B respectivamente, tales que

$$A \cap T^{-1}(B) = E \cap T^{-1}(B), B \cap T^{-1}(C) = F \cap T^{-1}(C).$$

Así también existen $C \subseteq G$ y $F \cap T^{-1}(C) \subseteq F \cap T^{-1}(G)$ tales que

$$A \cap T^{-1}(B) \subseteq E \cap T^{-1}(B) \neq \emptyset \text{ y } B \cap T^{-1}(C) \subseteq F \cap T^{-1}(C) \neq \emptyset.$$

De igual manera

$$A \cap T^{-1}(B) \subseteq E \cap T^{-1}(F) \neq \emptyset \text{ y } B \cap T^{-1}(C) \subseteq F \cap T^{-1}(C) \neq \emptyset.$$

Como $\alpha \preceq \beta \preceq \alpha \vee T^{-1}(\alpha)$, obtenemos

$$F \cap T^{-1}(C) = F \cap T^{-1}\left(\bigcup_{M \cap T^{-1}N \subseteq C, M, N \in \alpha} M \cap T^{-1}N \right) \neq \emptyset$$

y entonces existe $\emptyset \neq M \cap T^{-1}N \subseteq C$ de tal manera que

$$F \cap T^{-1}M \cap T^{-2}N = F \cap T^{-1}(M \cap T^{-1}N) \neq \emptyset.$$

Usando la propiedad topológica de Markov en α tenemos que

$$\begin{aligned}
\emptyset &\neq E \cap T^{-1}F \cap T^{-2}M \cap T^{-3}N \\
&\subseteq E \cap T^{-1}F \cap T^{-2}C \\
&= E \cap T^{-1}(F \cap T^{-1}C) \\
&= E \cap T^{-1}(B \cap T^{-1}C) \\
&= (E \cap T^{-1}B) \cap T^{-2}C \\
&= A \cap T^{-1}B \cap T^{-2}C,
\end{aligned}$$

por lo tanto $A \cap T^{-1}B \cap T^{-2}C \neq \emptyset$ y así β tiene la propiedad topológica de Markov. □

Observación 5.2.2. Si en la Observación 5.2.1 suplimos $\alpha \underset{1}{\to} \beta$ por $\alpha \xrightarrow{1} \beta$, se sigue cumpliendo 3.

Demostración. Es claro el resultado del hecho de que α cumple la propiedad topológica de Markov si y sólo si $T(\alpha)$ la cumple. □

Observación 5.2.3. Sea $(X_n)_{\mathbb{Z}}$ una cadena de Markov con matriz de probabilidades de transición P, con $A = P^0$, y $\alpha = \{A_i : i \in \mathbb{N}\}$ la partición natural, es decir,

$$A_i = \{x \in X_A \mid x_0 = i\}.$$

Entonces α tiene la propiedad topológica de Markov.

Demostración. Sean $A_{i_0}, A_{i_1}, A_{i_2}, \cdots, A_n$

$$A_{i_0} \cap T^{-1}(A_{i_1}) \neq \emptyset \text{ y } A_{i_1} \cap \cdots \cap T^{-(n-1)}(A_n) \neq \emptyset$$

entonces existe $x, y \in X_A$ tal que

$$x \in A_{i_0} \cap T^{-1}(A_{i_1}) \text{ y } y \in A_{i_1} \cap \cdots \cap T^{-(n-1)}(A_n)$$

es decir

$$x_0 = i_1 \quad \text{ y } \quad y_{[0,n-1]} = i_1 \cdots i_n,$$

de aquí z con $z_i = x_i$ para $i \leq 0$ y $z_i = y_{i-1}$ para $i > 0$. Como

$$z \in A_{i_0} \cap \cdots \cap T^{-n}(A_n),$$

α tiene la propiedad topológica de Markov. □

Proposición 5.2.1. *Sean α y η dos particiones abiertas y cerradas que cumplen la propiedad topológica de Markov, tales que $\alpha \xrightarrow[1]{} \eta$ ó $\alpha \xrightarrow{1} \eta$. Entonces*

$$(\alpha, \eta)(\eta, \alpha)_T = (\alpha, \alpha)_T \text{ y } (\eta, \alpha)_T(\alpha, \eta) = (\eta, \eta)_T.$$

Demostración. Supongamos que

$$(\alpha, \eta)(\eta, \alpha)_T(A, B) = \sum_{E \in \eta} (\alpha, \eta)(A, E)(\eta, \alpha)_T(E, B) = 1.$$

Ésto ocurre si y sólo si existe $E \in \eta$ tal que

$$(\alpha, \eta)(A, E)(\eta, \alpha)_T(E, B) = 1,$$

si y sólo si $E \cap A \neq \emptyset$ y $E \cap T{-}1(B) \neq \emptyset$. Debido a que η y α son particiones, se tiene que lo anterior es equivalente a que exista $E \in \eta$ tal que $E \subset A$ y $E \cap T{-}1(B) \neq \emptyset$. Por el punto 2 de la Observación 5.2.1, lo anterior es equivalente a que exista $E \in \eta$ tal que $E \subset A$ y $A \cap T^{-1}(B) \neq \emptyset$, y ésto ocurre si y sólo si existe $E \in \eta$ tal que $E \subset A$ y $(\alpha, \alpha)_T(A, B) = 1$. Por lo tanto

$$(\alpha, \eta)(\eta, \alpha)_T = (\alpha, \alpha)_T.$$

Para la otra igualdad,

$$(\eta, \alpha)_T(\alpha, \eta)(E, F) = \sum_{A \in \alpha} (\eta, \alpha)_T(E, A)(\alpha, \eta)(A, F)$$

es el número de A's tal que $(\eta, \alpha)_T(E, A)(\alpha, \eta)(A, F) = 1$, que es lo mismo que el número de A's tales que $(\eta, \alpha)_T(E, A) = 1$ y $(\alpha, \eta)(A, F) = 1$, es decir, el número de A's tales que $E \cap T^{-1}(A) \neq \emptyset$ y $A \cap F \neq \emptyset$, pero como α es partición,

$$\sum_{A \in \alpha} (\eta, \alpha)_T(E, A)(\alpha, \eta)(A, F)$$

es el número de A's tales que $E \cap T^{-1}(A) \neq \emptyset$ y $A \supseteq F$. Sea A el único elemento de α que contiene a F, y supongamos que $E \cap T^{-1}(A) = \emptyset$. Tenemos entonces que $E \cap T^{-1}(F) = \emptyset$ y de ésto concluimos que si

$$(\eta, \alpha)_T(\alpha, \eta)(E, F) = \sum_{A \in \alpha} (\eta, \alpha)_T(E, A)(\alpha, \eta)(A, F) = (\eta, \alpha)_T(E, A) = 0,$$

entonces

$$(\eta, \eta)_T(E, F) = 0.$$

Supongamos que $E \cap T^{-1}(A) \neq \emptyset$ y $F \cap A \neq \emptyset$ (de aquí es inmediato que $F \subseteq A$). Como $E \in \eta \preceq \alpha \vee T^{-1}(\alpha)$,

$$E = \bigcup_{B_i \in \Sigma_1 C_i \in \Sigma_2, \Sigma_i \subset \alpha} B_i \cap T^{-1}(C_i)$$

pero por el punto 2 de la Observación 5.2.1 tenemos que para cada B_i, existe $E_i \in \eta$ con $E_i \subset B_i$ tal que $B_i \cap T^{-1}(C_i) = E_i \cap T^{-1}(C_i)$. A causa de que $E_i \cap E \neq \emptyset$, tenemos que $E = E_i$ para toda i, y con ésto $B_i \supset E$, lo cual nos dice que $B_i \cap B_j \neq \emptyset$ para todo i, j subíndices. A causa de ésto y del hecho de que α es una partición, tenemos que $B_i = B$, es decir, es constante la B_i sobre el subíndice. Ahora como $E \subseteq B$, entonces $E \cap T^{-1}(A) \subseteq B \cap T^{-1}(A)$, y por el hecho de que $\alpha \vee T^{-1}(\alpha)$ es una partición, el punto 1 de la Observación 5.2.1, resulta que los dos elementos en esa contención forman parte de la partición y entonces $E \cap T^{-1}(A) = B \cap T^{-1}(A)$, y de aquí que A es uno de los C_i's.

Continuando con la prueba, si $E \subset B$, entonces $E \cap T^{-1}(F) \subset B \cap T^{-1}(F)$, y como son elementos de la misma partición, tenemos que $E \cap T^{-1}(F) = B \cap T^{-1}(F)$. De igual manera, por el punto 3 de la Observación 5.2.1, existe $D \in \alpha$ tal que $F \subseteq D$ y

$$F = \bigcup_{C \in \sigma \subseteq \alpha} D \cap T^{-1}(C)$$

y entonces $D \cap A \neq \emptyset$, lo que implica que $D = A$. Existe entonces $C \in \alpha$ tal que $F \supseteq A \cap T^{-1}(C) \neq \emptyset$, y como $B \cap T^{-1}(A) \neq \emptyset$, la propiedad topológica de Markov implica que

$$B \cap T^{-1}(A) \cap T^{-2}(C) \neq \emptyset,$$

pero

$$\emptyset \neq B \cap T^{-1}(A) \cap T^{-2}(C) = B \cap T^{-1}(A \cap T^{-1}(C)) \subseteq B \cap T^{-1}(F)$$

por lo que $B \cap T^{-1}(F) \neq \emptyset$. Entonces $E \cap T^{-1}(F) \neq \emptyset$ y concluimos que si

$$(\eta, \alpha)_T(\alpha, \eta)(E, F) = \sum_{A \in \alpha} (\eta, \alpha)_T(E, A)(\alpha, \eta)(A, F) = (\eta, \alpha)_T(E, A) = 1,$$

entonces

$$(\eta, \eta)_T(E, F) = 1,$$

es decir

$$(\eta, \alpha)_T(\alpha, \eta)(E, F) = \sum_{A \in \alpha} (\eta, \alpha)_T(E, A)(\alpha, \eta)(A, F) = (\eta, \alpha)_T(E, A) = 1$$

si y sólo si

$$(\eta,\eta)_T(E,F)=1.$$

Por lo tanto

$$(\eta,\alpha)_T(\alpha,\eta)=(\eta,\eta)_T.$$

Falta ahora el otro caso para concluir la demostración del lema. Supongamos $\alpha \xrightarrow{1} \eta$. Entonces $T^{-1}(\alpha)\preceq\eta\preceq\alpha\vee T^{-1}(\alpha)$. Como T es una conjugación, la Proposición 4.2.1 implica que $\alpha\preceq T(\eta)\preceq\alpha\vee T(\alpha)$. Con ésto y ocupando el resultado anterior para las particiones $\alpha, T(\eta)$ y la conjugación T^{-1}, a causa de que $\alpha \xrightarrow[1]{} T(\eta)_{T^{-1}}$, obtenemos que

$$(\alpha,T(\eta))(T(\eta),\alpha)_{T^{-1}}=(\alpha,\alpha)_{T^{-1}} \text{ y } (T(\eta),\alpha)_{T^{-1}}(\alpha,T(\eta))=(T(\eta),T(\eta))_{T^{-1}}.$$

Tenemos la siguiente serie de igualdades:

$$\begin{aligned}
(\alpha,\alpha)_{T^{-1}}(A,B) &= \begin{cases}1 & \text{si } A\cap T(B)\neq\emptyset\\ 0 & \text{si } A\cap T(B)=\emptyset\end{cases}\\
&= \begin{cases}1 & \text{si } T^{-1}(A)\cap B\neq\emptyset\\ 0 & \text{si } T^{-1}(A)\cap B=\emptyset\end{cases}\\
&= (\alpha,\alpha)_T(B,A)\\
&= (\alpha,\alpha)_T^t(A,B),
\end{aligned}$$

$$\begin{aligned}
(T(\eta),\alpha)_{T^{-1}}(T(E),B) &= \begin{cases}1 & \text{si } T(E)\cap T(B)\neq\emptyset\\ 0 & \text{si } T(E)\cap T(B)=\emptyset\end{cases}\\
&= \begin{cases}1 & \text{si } B\cap E\neq\emptyset\\ 0 & \text{si } B\cap E=\emptyset\end{cases}\\
&= (\eta,\alpha)(B,E)\\
&= (\alpha,\eta)(E,B)
\end{aligned}$$

y

$$\begin{aligned}
(\alpha,T(\eta)) &= \begin{cases}1 & \text{si } A\cap T(E)\neq\emptyset\\ 0 & \text{si } A\cap T(E)=\emptyset\end{cases}\\
&= \begin{cases}1 & \text{si } E\cap T^{-1}(A)\neq\emptyset\\ 0 & \text{si } E\cap T^{-1}(A)=\emptyset\end{cases}\\
&= (\eta,\alpha)_T(E,A),
\end{aligned}$$

$$
\begin{aligned}
(T(\eta),T(\eta))_{T^{-1}}(T(E),T(F)) &= \begin{cases} 1 & \text{si } T(E)\cap T^2(F)\neq\emptyset \\ 0 & \text{si } T(E)\cap T^2(F)=\emptyset \end{cases} \\
&= \begin{cases} 1 & \text{si } T^{-1}(E)\cap F\neq\emptyset \\ 0 & \text{si } T^{-1}(E)\cap F=\emptyset \end{cases} \\
&= (\eta,\eta)_T(F,E) \\
&= (\eta,\eta)^t_T(E,F)
\end{aligned}
$$

Entonces

$$
\begin{aligned}
(\alpha,\alpha)_T(A,B) &= (\alpha,\alpha)^t_{T^{-1}}(A,B) \\
&= (\alpha,\alpha)_{T^{-1}}(B,A) \\
&= (\alpha,T(\eta))(T(\eta),\alpha)_{T^{-1}}(B,A) \\
&= \sum_{E\in\eta}(\alpha,T(\eta))(B,T(E))(T(\eta),\alpha)_{T^{-1}}(T(E),A) \\
&= \sum_{E\in\eta}(T(\eta),\alpha)(T(E),B)(\alpha,T(\eta))_{T^{-1}}(A,T(E)) \\
&= \sum_{E\in\eta}(\eta,\alpha)_T(E,B)(\alpha,\eta)(A,E) \\
&= \sum_{E\in\eta}(\alpha,\eta)(A,E)(\eta,\alpha)_T(E,B) \\
&= (\alpha,\eta)(\eta,\alpha)_T(A,B)
\end{aligned}
$$

y tenemos que

$$(\alpha,\eta)(\eta,\alpha)_T=(\alpha,\alpha)_T,$$

y por otro lado

$$\begin{aligned}
(\eta,\eta)_T(E,F) &= (T(\eta),T(\eta))^t_{T^{-1}}(T(E),T(F)) \\
&= (T(\eta),T(\eta))_{T^{-1}}(T(F),T(E)) \\
&= (T(\eta),\alpha)_{T^{-1}}(\alpha,T(\eta))(T(F),T(E)) \\
&= \sum_{A\in\alpha}(T(\eta),\alpha)_{T^{-1}}(T(F),A)(\alpha,T(\eta))(A,T(E)) \\
&= \sum_{A\in\alpha}(\alpha,T(\eta))(A,T(E))(T(\eta),\alpha)_{T^{-1}}(T(F),A) \\
&= \sum_{A\in\alpha}(\eta,\alpha)_T(E,A)(\eta,\alpha)(A,F) \\
&= (\eta,\alpha)_T(\eta,\alpha)(E,F),
\end{aligned}$$

por lo tanto

$$(\eta,\eta)_T = (\eta,\alpha)_T(\eta,\alpha),$$

lo que concluye la prueba. □

Ahora la prueba del Teorema 5.2.1

Demostración. Para la demostración hace sólo falta ver un par de detalles debido a los lemas ya probados, estos detalles vienen a continuación. Sean $G, H \in \beta$. Entonces

$$\begin{aligned}
(\gamma,\gamma)_T(\varphi^{-1}(G),\varphi^{-1}(H)) &= ((\varphi\circ Tm)^{-1}(\beta),(\varphi\circ Tm)^{-1}(\beta))_T(G,H) \\
&= \begin{cases} 1 & \text{si } (\varphi\circ T^m)^{-1}(G)\cap T^{-1}((\varphi\circ T^m)^{-1}(H)) \neq \emptyset \\ 0 & \text{si } (\varphi\circ T^m)^{-1}(G)\cap T^{-1}((\varphi\circ T^m)^{-1}(H)) = \emptyset \end{cases} \\
&= \begin{cases} 1 & \text{si } (\varphi\circ T^m)^{-1}(G)\cap (\varphi\circ T^{m+1})^{-1}(H) \neq \emptyset \\ 0 & \text{si } (\varphi\circ T^m)^{-1}(G)\cap (\varphi\circ T^{m+1})^{-1}(H) = \emptyset \end{cases} \\
&= \begin{cases} 1 & \text{si } (T_B^m\circ\varphi)^{-1}(G)\cap T^{-1}((T_B^{m+1}\circ\varphi)^{-1}(H)) \neq \emptyset \\ 0 & \text{si } (T_B^m\circ\varphi)^{-1}(G)\cap T^{-1}((T_B^{m+1}\circ\varphi)^{-1}(H)) = \emptyset \end{cases} \\
&= \begin{cases} 1 & \text{si } (T_B^m\circ\varphi)^{-1}(G)\cap (T_B^m\circ\varphi)^{-1}(T_B^{-1}(H)) \neq \emptyset \\ 0 & \text{si } (T_B^m\circ\varphi)^{-1}(G)\cap (T_B^m\circ\varphi)^{-1}(T_B^{-1}(H)) = \emptyset \end{cases} \\
&= \begin{cases} 1 & \text{si } G\cap T_B^{-1}(H) \neq \emptyset \\ 0 & \text{si } G\cap T_B^{-1}(H) = \emptyset \end{cases} \\
&= (\beta,\beta)_{T_B}(G,H)
\end{aligned}$$

y por lo tanto
$$(\beta,\beta)_{T_B} = (\gamma,\gamma)_T.$$
Por otro lado
$$(\alpha,\alpha)_T(A_i,A_j) = 1$$
si y sólo si existe
$$x \in X_A \text{ con } x \in A_i \cap T^{-1}(A_j),$$
si y sólo $1 = A_{x_0,x_1} = A_{i,j}$, y por ende $(\alpha,\alpha)_T = A$. De manera análoga tenemos que $B = (\beta,\beta)_{T_B} = (\gamma,\gamma)_T$, debido a los Lemas 5.2.1, 5.2.2 y la Proposición 5.2.1, tenemos que A y B son matrices shift-fuertemente equivalentes, lo cual concluye la prueba. □

5.2.2. Caso estocástico.

En esta subsección demostramos la versión estocástica del Teorema de Clasificación.

Teorema 5.2.2. *Sean $(X_n)_{n\in\mathbb{Z}}$ y $(Y_n)_{n\in\mathbb{Z}}$ dos cadenas de Markov irreducibles con espacio de estados finito y matrices de transición P y Q respectivamente, siendo $P^0 = A$ y $Q^0 = B$. Si $(X_n)_{n\in\mathbb{Z}}$ y $(Y_n)_{n\in\mathbb{Z}}$ son conjugadas, entonces P y Q son estocásticamente shift-fuertemente equivalentes.*

Nuevamente, posponemos la demostración hasta contar con las herramientas necesarias para llevarla a cabo. Dada P una matriz estocática y $A = P^0$, fijamos la notación $\mu = \mu_P$ y $(X,T) = (X_A,T_A)$.

Definición 5.2.6. Sean ξ,η dos particiones abiertas y cerradas de X. Definimos las siguientes matrices indexadas por los elementos de la particiones, concretamente, para todo $K \in \xi$ y $E \in \eta$,
$$[\xi,\eta](K,E) = \frac{\mu(K\cap E)}{\mu(K)}$$
y
$$[\xi,\eta]_T(K,E) = \frac{\mu(K\cap T^{-1}(E))}{\mu(K)}.$$

Una pequeña observación es importante, que la matrices definidas arriba son matrices estocásticas, ésto debido a que ambas están definidas sobre particiones.

Definición 5.2.7. Definimos un *generador* como una partición $\mathcal{P}$ de X que cumple que para cada $x \neq y \in X$, existe $n \in \mathbb{N}$ tal que x y y están en distinta celda de la partición
$$\bigvee_{i=-n}^{n} T^{-i}(\mathcal{P}).$$

Un *generador de Markov* simplemente es una partición de Markov que es un generador.

Lema 5.2.3. *Sean α y η particiones de Markov abiertas y cerradas tales que $\alpha \xrightarrow[1]{} \eta$ ó $\alpha \xrightarrow{1} \eta$. Entonces*
$$[\alpha,\eta][\eta,\alpha]_T = [\alpha,\alpha]_T \text{ y } [\eta,\alpha]_T[\alpha,\eta] = [\eta,\eta]_T.$$

Demostración. Primero supongamos que $\alpha \underset{1}{\longrightarrow} \eta$. Entonces $\alpha \vee T^{-1}(\alpha) \succeq \eta \succeq \alpha$, y por lo tanto

$$
\begin{aligned}
[\alpha,\eta][\eta,\alpha]_T(A,B) &= \sum_{E\in\eta} \frac{\mu(A\cap T^{-1}(E))}{\mu(A)} \frac{\mu(E\cap B)}{\mu(E)} \\
&= \sum_{E\subset A} \frac{\mu(A\cap E}{\mu(A)} \frac{\mu(E\cap T^{-1}(B))}{\mu(E)} \\
&= \sum_{E\subset A} \frac{\mu(E)}{\mu(A)} \frac{\mu(E\cap T^{-1}(B))}{\mu(E)} \\
&= \sum_{E\subset A} \frac{\mu(E\cap T^{-1}(B))}{\mu(A)} \\
&= \frac{\mu(A\cap T^{-1}(B))}{\mu(A)} \\
&= [\alpha,\alpha]_T(A,B).
\end{aligned}
$$

Para lo otro,

$$
\begin{aligned}
[\eta,\alpha]_T[\alpha,\eta](E,F) &= \sum_{A\in\alpha} [\eta,\alpha]_T(E,A)[\alpha,\eta](A,F) \\
&= \sum_{A\in\alpha} \frac{\mu(E\cap T^{-1}(A))}{\mu(E)} \frac{\mu(A\cap F)}{\mu(A)} \qquad (5.4)\\
&= \frac{\mu(E\cap T^{-1}(A))}{\mu(E)} \frac{\mu(A\cap F)}{\mu(A)},
\end{aligned}
$$

donde A es el único elemento de α que contiene a F. Hagamos una pausa para notar algo. Como A es el único elemento de α que contiene a F, anteriormente probamos que, a raíz de ésto,

$$F = \bigcup_{B\in\Sigma\subset\alpha} A\cap T^{-1}(B)$$

y de aquí vemos que

$$
\begin{aligned}
\mu(E\cap T^{-1}(F)) &= \mu(\bigcup_{B\in\Sigma\subset\alpha} E\cap T^{-1}(A)\cap T^{-2}(B)) \\
&= \sum_{B\in\Sigma\subset\alpha} \mu(E\cap T^{-1}(A)\cap T^{-2}(B)) \\
&= \sum_{B\in\Sigma\subset\alpha} \frac{\mu(E\cap T^{-1}(A)\cap T^{-2}(B))}{\mu(E\cap T^{-1}(A))} \mu(E\cap T^{-1}(A)).
\end{aligned}
$$

Ocupando la propiedad de Markov tenemos que

$$\begin{aligned}
\sum_{B\in\Sigma\subset\alpha}\frac{\mu(E\cap T^{-1}(A)\cap T^{-2}(B))}{\mu(E\cap T^{-1}(A))}\mu(E\cap T^{-1}(A)) &= \sum_{B\in\Sigma\subset\alpha}\frac{\mu(A\cap T^{-1}(B))}{\mu(A)}\mu(E\cap T^{-1}(A)) \\
&= \frac{\mu(E\cap T^{-1}(A))}{\mu(A)}\sum_{B\in\Sigma\subset\alpha}\mu(A\cap T^{-1}(B)) \\
&= \frac{\mu(E\cap T^{-1}(A))}{\mu(A)}\mu(\bigcup_{B\in\Sigma\subset\alpha} A\cap T^{-1}(B)) \\
&= \frac{\mu(E\cap T^{-1}(A))}{\mu(A)}\mu(F),
\end{aligned}$$

con lo cual tenemos que

$$\frac{\mu(E\cap T^{-1}(F))}{\mu(F)}=\frac{\mu(E\cap T^{-1}(A))}{\mu(A)}$$

y así, regresando a la prueba original, de (5.4) concluimos que

$$\begin{aligned}
[\eta,\alpha]_T[\alpha,\eta](E,F) &= \frac{\mu(E\cap T^{-1}(F))}{\mu(F)}\frac{\mu(A\cap F)}{\mu(E)} \\
&= \frac{\mu(E\cap T^{-1}(F))}{\mu(F)}\frac{\mu(F)}{\mu(E)} \\
&= \frac{\mu(E\cap T^{-1}(F))}{\mu(E)} \\
&= [\eta,\eta]_T(E,F)
\end{aligned}$$

y por lo tanto

$$[\eta,\alpha]_T[\alpha,\eta]=[\eta,\eta]_T,$$

lo cual concluye el primer caso.

Supongamos ahora que $\alpha \xrightarrow{1} \eta$. Entonces $\alpha\vee T^{-1}(\alpha)\succeq\eta\succeq T^{-1}(\alpha)$ y hacemos lo mismo que en la prueba sin medida. Obtenemos que $\alpha\preceq T(\eta)\preceq\alpha\vee T(\alpha)$, y con ésto y ocupando el resultado anterior para las particiones α y $T(\eta)$ y la conjugación T^{-1}, a causa de que $\alpha \xrightarrow[1]{} T(\eta)_{T^{-1}}$, obtenemos que

$$[\alpha,T(\eta)][T(\eta),\alpha]_{T^{-1}}=[\alpha,\alpha]_{T^{-1}} \text{ y } [T(\eta),\alpha]_{T^{-1}}[\alpha,T(\eta)]=[T(\eta),T(\eta)]_{T^{-1}}.$$

Ahora observemos que

$$\begin{aligned}
[\alpha,\alpha]^t_{T^{-1}}(A,B) &= [\alpha,\alpha]_{T^{-1}}(B,A) \\
&= \frac{\mu(B\cap T(A))}{\mu(B)} \\
&= \frac{\mu(A\cap T^{-1}(B))}{\mu(B)} \\
&= \frac{\mu(A\cap T^{-1}(B))}{\mu(A)}\frac{\mu(A)}{\mu(B)} \\
&= [\alpha,\alpha]_T(A,B)\frac{\mu(A)}{\mu(B)}
\end{aligned}$$

(la segunda igualdad se debe a que μ es T-invariante) y también

$$\begin{aligned}
[T(\eta),T(\eta)]^t_{T^{-1}}(T(E),T(F)) &= [T(\eta),T(\eta)]_{T^{-1}}(T(F),T(E)) \\
&= \frac{\mu(T(F)\cap T^2(E))}{\mu(T(F))} \\
&= \frac{\mu(E\cap T^{-1}(F))}{\mu(F)} \\
&= \frac{\mu(E\cap T^{-1}(F))}{\mu(E)}\frac{\mu(E)}{\mu(F)} \\
&= [\eta,\eta]_T(E,F)\frac{\mu(E)}{\mu(F)}.
\end{aligned}$$

Ahora

$$\begin{aligned}
[\alpha,\alpha]_T(A,B)\frac{\mu(A)}{\mu(B)} &= [\alpha,\alpha]^t_{T^{-1}}(A,B) \\
&= [\alpha,\alpha]_{T^{-1}}(B,A) = [\alpha,T(\eta)][T(\eta),\alpha]_{T^{-1}}(B,A) \\
&= \sum_{E\in\eta}[\alpha,T(\eta)](B,T(E))[T(\eta),\alpha]_{T^{-1}}(T(E),A) \\
&= \sum_{E\in\eta}\frac{\mu(B\cap T(E))}{\mu(B)}\frac{\mu(T(E)\cap T(A))}{\mu(T(E))} = **
\end{aligned}$$

y como la μ es T-invariante,

$$\begin{aligned}\sum_{E\in\eta}\frac{\mu(E\cap T^{-1}(B))}{\mu(B)}\frac{\mu(E\cap A)}{\mu(E)} &= \sum_{E\in\eta}\frac{\mu(E\cap T^{-1}(B))}{\mu(E)}\frac{\mu(A\cap E)}{\mu(A)}\frac{\mu(A)}{\mu(B)}\\ &= \sum_{E\in\eta}[\alpha,\eta](A,E)[\eta,\alpha]_T(E,B)\frac{\mu(A)}{\mu(B)} \qquad (5.5)\\ &= [\alpha,\eta][\eta,\alpha]_T(A,B)\frac{\mu(A)}{\mu(B)}.\end{aligned}$$

Con ésto tenemos que

$$[\alpha,\alpha]_T(A,B)\frac{\mu(A)}{\mu(B)} = [\alpha,\eta][\eta,\alpha]_T(A,B)\frac{\mu(A)}{\mu(B)}$$

y entonces

$$[\alpha,\alpha]_T(A,B) = [\alpha,\eta][\eta,\alpha]_T(A,B),$$

con lo cual

$$[\alpha,\alpha]_T = [\alpha,\eta][\eta,\alpha]_T.$$

Ahora

$$\begin{aligned}[\eta,\eta]_T(E,F)\frac{\mu(E)}{\mu(F)} &= [T(\eta),T(\eta)]^t_{T^{-1}}(T(E),T(F))\\ &= [T(\eta),T(\eta)]_{T^{-1}}(T(F),T(E))\\ &= [T(\eta),\alpha]_{T^{-1}}[\alpha,T(\eta)](T(F),T(E))\\ &= \sum_{A\in\alpha}[T(\eta),\alpha]_{T^{-1}}(T(F),A)[\alpha,T(\eta)](A,T(E))\\ &= \sum_{A\in\alpha}\frac{\mu(T(F)\cap T(A))}{\mu(T(F))}\frac{\mu(A\cap T(E))}{\mu(A)}\\ &= \sum_{A\in\alpha}\frac{\mu(A\cap F)}{\mu(F)}\frac{\mu(E\cap T^{-1}(A))}{\mu(A)}\\ &= \sum_{A\in\alpha}\frac{\mu(A\cap F)}{\mu(A)}\frac{\mu(E\cap T^{-1}(A))}{\mu(E)}\frac{\mu(E)}{\mu(F)}\\ &= \sum_{A\in\alpha}[\eta,\alpha]_T(E,A)[\alpha,\eta](A,F)\frac{\mu(E)}{\mu(F)}\\ &= [\eta,\alpha]_T[\alpha,\eta](E,F)\frac{\mu(E)}{\mu(F)},\end{aligned}$$

y así obtenemos que

$$[\eta,\eta]_T(E,F)\frac{\mu(E)}{\mu(F)} = [\eta,\alpha]_T[\alpha,\eta](E,F)\frac{\mu(E)}{\mu(F)},$$

por lo que entonces

$$[\eta,\eta]_T(E,F) = [\eta,\alpha]_T[\alpha,\eta](E,F)$$

y así

$$[\eta,\eta]_T = [\eta,\alpha]_T[\alpha,\eta],$$

lo cual concluye la prueba. $\square$

Lema 5.2.4. *Sean α y η dos particiones abiertas y cerradas de X tales que $\alpha \xrightarrow[1]{} \eta$ ó $\alpha \xrightarrow{1} \eta$. Si α es un generador de Markov, entonces η también es un generador de Markov.*

Demostración. Sean $x \neq y \in X$. Como α es un generador de Markov, en particular es un generador, y por ello existe $n \in \mathbb{N}$ tal que x y y pertencen a distintas celdas de

$$\bigvee_{i=-n}^{n} T^{-i}(\alpha).$$

Por hipótesis, $\alpha \preceq \eta$ ó $T^{-1}(\alpha) \preceq \eta$. Para el primer caso tenemos que $T^i(\alpha) \preceq T^i(\eta)$ y para el segundo $T^{i-1}(\alpha) \preceq T^i(\eta)$ $\forall i \in \mathbb{Z}$. Si refinamos estas particiones para $i = -n, ...n+1$, entonces tenemos que para el primer caso

$$\bigvee_{i=-n}^{n+1} T^{-i}(\alpha) \preceq \bigvee_{i=-n}^{n+1} T^{-i}(\eta)$$

y para el segundo

$$\bigvee_{i=-(n+1)}^{n} T^{-i}(\alpha) \preceq \bigvee_{i=-n}^{n+1} T^{-i}(\eta),$$

pero para cualquiera de los dos casos, por transitividad, tenemos que

$$\bigvee_{i=-n}^{n} T^{-i}(\alpha) \preceq \bigvee_{i=-n}^{n+1} T^{-i}(\eta) \preceq \bigvee_{i=-(n+1)}^{n+1} T^{-i}(\eta).$$

Ahora supongamos que $x \in S$ y $y \in R$ para R y S celdas de $\bigvee_{i=-(n+1)}^{n+1} T^{-i}(\eta)$. Así, existen celdas V y W de $\bigvee_{i=-n}^{n} T^{-i}(\alpha)$ tales que $x \in S \subset V$ y $y \in R \subset W$. A causa de ésto, $W \neq V$, y por ello, S y R son conjuntos ajenos y por ende distintos. Entonces x y y pertenecen a distintas celdas de $\bigvee_{i=-(n+1)}^{n+1} T^{-i}(\eta)$, y con ésto obtenemos que η es un generador. Ahora como $\alpha \preceq \eta \preceq \alpha \vee T^{-1}(\alpha)$,

$$T^{-i}(\alpha) \preceq T^{-i}(\eta) \preceq T^{-i}(\alpha) \vee T^{-(i+1)}(\alpha),$$

y de aquí que

$$\bigvee_{i=0}^{n-1} T^{-i}(\alpha) \preceq \bigvee_{i=0}^{n-1} T^{-i}(\eta) \preceq \bigvee_{i=0}^{n-1} T^{-i}(\alpha) \vee T^{-(i+1)}(\alpha) = \bigvee_{i=0}^{n} T^{-i}(\alpha).$$

Entonces, con la notación de el capítulo 3,

$$\alpha^n \preceq \eta^n \preceq \alpha^{n+1},$$

y usando la Proposición 3.3.4 para R la partición trivial obtenemos que

$$H(\alpha^n) \leq H(\eta^n) \leq H(\alpha^{n+1}).$$

Dividiendo entre n y tomando el límite cuando $n \to \infty$, tenemos que $h(T, \eta) = h(T, \alpha)$. Como α es una partición de Markov, obetenemos que

$$\begin{aligned} h(T,\eta) &= h(T,\alpha) \\ &= H(\alpha|T^{-1}(\alpha)) \\ &= H(\alpha \vee T^{-1}(\alpha)|T^{-1}(\alpha)) \\ &\geq H(\eta|T^{-1}(\alpha)) \\ &\geq H(\eta|T^{-1}(\eta)), \end{aligned}$$

donde la últimas dos desigualdades son debido a la Proposición 3.3.4 ya que

$$T^{-1}(\alpha) \preceq T^{-1}(\eta), \eta \preceq \alpha \vee T^{-1}(\alpha).$$

Como

$$\eta^+ = \bigvee_{i=1}^{\infty} \eta^{[1,i]}$$

y $(\eta^{[1,i]})_{i \in \mathbb{N}}$ es una sucesión que se refina,

$$H(\eta|\eta^+) = \lim_{n\to\infty} H(\eta|\eta^{[1,i]}).$$

Como $\eta^{[1,i]} \succeq T^{-1}(\eta)$ para toda i,

$$h(T,\eta) = H(\eta|\eta^+) = \lim_{n\to\infty} H(\eta|\eta^{[1,i]}) \leq H(\eta|T^{-1}(\eta)),$$

con lo cual tenemos que

$$h(T,\eta) \leq H(\eta|T^{-1}(\eta)) \leq h(T,\eta)$$

y por lo tanto

$$h(T,\eta) = H(\eta|T^{-1}(\eta))$$

Por el Teorema 3.4.1 concluimos que η es una partición de Markov y como es un generador, es también un generador de Markov, lo cual termina con la prueba del lema. $\square$

Demostración del Teorema 5.2.2. Sean $E, F \in \beta$. Entonces

$$\begin{aligned}
[\gamma,\gamma]_T((\varphi\circ T^m)^{-1}(E),(\varphi\circ T^m)^{-1}(F)) &= \frac{\mu((\varphi\circ T^m)^{-1}(E)\cap T^{-1}((\varphi\circ T^m)^{-1}(F)))}{\mu((\varphi\circ T^m)^{-1}(E))}\\
&= \frac{\mu((\varphi\circ T^m)^{-1}(E)\cap(\varphi\circ T^{m+1})^{-1}(F))}{\mu((\varphi\circ T^m)^{-1}(E))}\\
&= \frac{\mu(T^{-m}(\varphi^{-1}(E))\cap T^{-(m+1)}((\varphi^{-1}(F))))}{\mu(T^{-m}(\varphi^{-1}(E)))}\\
&= \frac{\mu(T^{-m}(\varphi^{-1}(E)\cap T^{-1}(\varphi^{-1}(F))))}{\mu(T^{-m}(\varphi^{-1}(E)))}\\
&= \frac{\mu(\varphi^{-1}(E)\cap T^{-1}(\varphi^{-1}(F)))}{\mu(\varphi^{-1}(E))}\\
&= \frac{\mu(\varphi^{-1}(E)\cap(\varphi\circ T)^{-1}(F))}{\mu(\varphi^{-1}(E))}\\
&= \frac{\mu(\varphi^{-1}(E)\cap(T_B\circ\varphi)^{-1}(F))}{\mu(\varphi^{-1}(E))}\\
&= \frac{\mu(\varphi^{-1}(E)\cap\varphi^{-1}(T_B^{-1}(F)))}{\mu(\varphi^{-1}(E))}\\
&= \frac{\mu(\varphi^{-1}(E\cap T_B^{-1}(F)))}{\mu(\varphi^{-1}(E))}\frac{\mu_Q(E\cap T_B^{-1}(F))}{\mu_Q(E)}\\
&= [\beta,\beta]_{T_B}(E,F).
\end{aligned}$$

Así,

$$[\gamma,\gamma]_T = [\beta,\beta]_{T_B}$$

y

$$\begin{aligned}
[\alpha,\alpha]_T(A_i,A_j) &= \frac{\mu(A_i\cap T^{-1}(A_j))}{\mu(A_i)}\\
&= \frac{\mu(C_0(ij))}{\mu(A_i)}\\
&= \frac{\mu(X_0=i,X_1=j)}{\mu(X_0=i)}\\
&= \mu(X_0=i,X_1=j|X_0=i)\\
&= P_{i,j},
\end{aligned}$$

por lo que $P=[\alpha,\alpha]_T$ y de manera análoga $Q=[\beta,\beta]_T$.

Veamos que α es una partición de Markov. Si $(x_i)_{i=0}^n$ es un conjunto finito de índices consideramos a la familia de $(A_{x_i})_{i=0}^n$ de elementos de la partición α, entonces

$$\begin{aligned}
\frac{\mu(A_{x_0} \cap T^{-1}A_{x_1} \cap T^{-2}A_{x_2} \cdots \cap T^{-n}A_{x_n})}{\mu(A_{x_0} \cap T^{-1}A_{x_1} \cap T^{-2}A_{x_2} \cdots \cap T^{-(n-1)}A_{x_{n-1}})} &= \frac{\mu(\mathcal{C}_0(x_0x_1\cdots x_n))}{\mu(\mathcal{C}_0(x_0x_1\cdots x_{n-1}))} \\
&= \frac{\mu(X_0 = x_0, X_1 = x_1, \cdots, X_n = x_n)}{\mu(X_0 = x_0, X_1 = x_1, \cdots, X_{n-1} = x_{n-1})} \\
&= \mu(X_n = x_n | X_0 = x_0, X_1 = x_1, \cdots, X_{n-1} = x_{n-1}) \\
&= \mu(X_n = x_n | X_{n-1} = x_{n-1}) \\
&= \frac{\mu(X_n = x_n, X_{n-1} = x_{n-1})}{\mu(X_{n-1} = x_{n-1})} \\
&= \frac{\mu(\mathcal{C}_0(x_{n-1}x_n))}{\mu(A_{x_{n-1}})} \\
&= \frac{\mu(A_{x_{n-1}} \cap T^{-1}(A_{x_n})))}{\mu(A_{x_{n-1}})},
\end{aligned}$$

por lo que efectivamente α es una partición de Markov. Es un generador por que los cilindros centrados cumplen con la propiedad de que cualquier par de elementos de X cumplen con estar en dos cilindros centrados en cero, distintos y del mismo tamaño (para algún tamaño). Ahora, ya que ésto ocurre, tenemos que las hipótesis de los Lemas 5.2.1, 5.2.2, 5.2.3, 5.2.4 y la Proposición 5.2.1 se cumplen, lo que implica que

$$[\alpha, \alpha]_T \quad \text{y} \quad [\gamma, \gamma]_T$$

son estocásticamente shift-fuertemente equivalentes, y por ende P y Q lo son. Por lo tanto $(X_n)_{n\in\mathbb{Z}}$ y $(Y_n)_{n\in\mathbb{Z}}$ son estocásticamente shift-fuertemente equivalentes lo cual concluye con esta prueba y con la sección. □

5.3. Suficiencia de las series paralelas

Observación 5.3.1. Sean P y Q dos matrices estocásticas no triviales (que no tienen renglones o columnas de únicamente de ceros). Supongamos que P y Q son estocásticamente shift-fuertemente equivalentes, y supongamos también que la longitud de dicha equivalencia es 1 (es decir $P = P_1Q_1$ y $Q_1P_1 = Q$). Si P es reducible (es decir existen i, j tal que $\forall n \in \mathbb{N}\ P_{ij}^n = 0$), entonces Q también, y por lo tanto P es irreducible si y sólo si Q es irreducible.

Demostración. Como P es reducible tenemos que existen i, j tal que $\forall n \in \mathbb{N}\ P_{ij}^n = 0$, pero como P y Q son estocásticamente shift-fuertemente equivalentes y dicha condición está dada por 2 matrices P_1 y Q_1, de forma que $P = P_1Q_1$ y $Q_1P_1 = Q$, $PP_1 = P_1Q_1P_1 = P_1Q$. Por lo anterior, $P^{n+1} = PP^n = P_1Q_1P^n = P_1Q^nQ_1$, y con ésto, $\forall n \in \mathbb{N}$,

$$0 = P_{i,j}^{n+1} = (P_1Q^nQ_1)_{ij} = \sum_{t,l} P_{1it}Q_{tl}^nQ_{1lj}.$$

Como P y Q son no triviales, también son no triviales Q_1 y P_1. En efecto, ya que $P = P_1Q_1,\ Q_1P_1 = Q$, tendrías que si existe

1. un renglón en P_1 cero, el mismo renglón tendría que ser cero en P;
2. una columna en P_1 cero, la misma columna tendría que ser cero en Q;
3. un renglón en Q_1 cero, el mismo renglón tendría que ser cero en Q;
4. una columna en Q_1 cero, la misma columna tendría que ser cero en P.

Estos 4 puntos son una contradicción, por lo que entonces existen t y l tales que $P_{1it}Q_{1lj} \neq 0$, lo que indica que el hecho de $P_{1it}Q^n_{tl}Q_{1lj} = 0$ es necesariamente causa de que $Q^n_{tl} = 0$, por lo tanto Q es reducible. □

Teorema 5.3.1. *Sean $(X_n)_{n\in\mathbb{Z}}$ y $(Y_n)_{n\in\mathbb{Z}}$ dos cadenas de Markov irreducibles con espacio de estados finito, sus matrices de probabilidades de transición P y Q respectivamente, siendo $P^0 = A$ y $Q^0 = B$. Si $(X_n)_{n\in\mathbb{Z}}$ y $(Y_n)_{n\in\mathbb{Z}}$ son estocásticamente shift-fuertemente equivalentes, entonces ambas cadenas son isomorfas por bloques (block-isomorfas), es decir, son conjugadas.*

Demostración. Por hipótesis tenemos que existen matrices estocásticas $U_i, V_i, i = 0, ..., n$ tales que

$$P = U_0V_0,\ V_0U_0 = U_1V_1,\ V_1U_1 = U_2V_2, \cdots, V_{n-1}U_{n-1} = U_nV_n = Q$$

y

$$P^0 = U^0_0V^0_0,\ V^0_0U^0_0 = U^0_1V^0_1,\ V^0_1U^0_1 = U^0_2V^0_2, \cdots, V^0_{n-1}U^0_{n-1} = U^0_nV^0_n = Q^0.$$

Basta probar este resultado para $n = 1$. Si suponemos que ocurre para $n = 1$, entonces tomamos en cada producto de arriba la cadena de Markov $(X^i_m)_{m\in\mathbb{Z}}$ asociada a la matriz U_1V_1 (está matriz es no trivial, pues de serlo, por 5.3, Q sería trivial, pero ésto no puede pasar por ser matriz de transición de probabilidades de una cadena de Markov, por ello, debido a la Observación 5.3.1 también es irreducible). Tenemos entonces una serie de conjugaciones que preservan la medida respectiva de un espacio de un sistema dinámico a otro, de forma que la primera conjugación sale de $(X_{P^0}, T_{P^0}, \mathfrak{B}(\tau), \mu_P)$ y la última llega a $(X_{Q^0}, T_{Q^0}, \mathfrak{B}(\tau), \mu_Q)$. Al final, la composición preserva la medida desde el primer espacio hasta el último por transitividad, y composición de conjugaciones es una conjugación, lo cual terminaría el resultado de ser cierta nuestra afirmación.

Ahora veámoslo para $n = 1$. Supongamos que

$$P = UV \text{ y } VU = Q$$

y

$$A = U^0V^0 \text{ y } V^0U^0 = B.$$

Definimos $\varphi\colon X_A \to X_B$ por $\varphi(x) = y$ donde y_n es el único estado tal que

$$A_{x_nx_{n+1}} = U^0_{x_ny_n}V^0_{y_nx_{n+1}}.$$

Entonces sólo hace falta ver que efectivamente φ es el código de bloques definido por

$$\Phi(x_nx_{n+1}) = y_n.$$

Primero, y_n es adyascente a y_{n+1} debido a que

$$B_{y_n y_{n+1}} = V^0_{y_n x_s} U^0_{x_s y_{n+1}} = (U^0_{x_n y_n} V^0_{y_n x_s}) U^0_{x_s y_{n+1}},$$

lo cual nos dice que $s = n+1$ y con ello

$$(U^0_{x_n y_n} V^0_{y_n x_s}) U^0_{x_s y_{n+1}} = (U^0_{x_n y_n} V^0_{y_n x_{n+1}})(U^0_{x_{n+1} y_{n+1}} V^0_{y_{n+1} x_{n+2}}) = (1)(1) = 1.$$

Lo siguiente es definir otra función de manera análoga $\psi : X_B \longrightarrow X_A$, ahora $\psi(y) = x$ donde x_{n+1} es el único estado tal que

$$B_{y_n y_{n+1}} = V^0_{y_n x_{n+1}} U^0_{x_{n+1} y_{n+1}}.$$

Sólo hace falta ver que efectivamente φ es el código de bloques definido por $\Psi(y_n y_{n+1}) = x_{n+1}$. Tenemos que x_{n+1} es adyascente a x_{n+2} debido a que

$$A_{x_{n+1} x_{n+2}} = U^0_{x_{n+1} y_s} V^0_{y_s x_{n+2}} = (V^0_{y_n x_{n+1}} U^0_{x_{n+1} y_s}) V^0_{y_s x_{n+2}},$$

lo cual nos dice que $s = n+1$ y con ello

$$(V^0_{y_n x_{n+1}} U^0_{x_{n+1} y_s}) V^0_{y_s x_{n+2}} = (V^0_{y_n x_{n+1}} U^0_{x_{n+1} y_{n+1}})(V^0_{y_{n+1} x_{n+2}} U^0_{x_{n+2} y_{n+2}}) = (1)(1) = 1.$$

Ahora vemos que

$$\begin{aligned}
(\varphi \circ \psi(y))_i &= \varphi(\psi(y))_i \\
&= \Phi(\psi(y)_i \psi(y)_{i+1}) \\
&= \Phi(\Psi(y_i y_{i+1}) \Psi(y_{i+1} y_{i+2})) \\
&= \Phi(x_{i+1} x_{i+2}) = y_{i+1} = (T_{X_B}(y))_i
\end{aligned}$$

y entonces

$$\varphi \circ \psi(y) = T_{X_B}(y) \text{ y también } \varphi \circ \psi = T_{X_B}.$$

De manera análoga se prueba que

$$\psi \circ \varphi = T_{X_A}$$

y con ésto se ve que las dos efectivamente son conjugaciones puesto que son códigos de bloques biyectivos. Falta ver que esta conjugación preserva la medida, lo cual haremos como sigue. Siendo $\phi(x) = y$,

$$\begin{aligned}
\mu_P(x_0 x_1 \cdots x_{n+1}) &= \pi_P(x_0) P_{x_0 x_1} P_{x_1 x_2} \cdots P_{x_n x_{n+1}} \\
&= \Pi_P(x_0) U_{x_0 y_0} V_{y_0 x_1} U_{x_1 y_1} V_{y_1 x_2} \cdots U_{x_n y_n} V_{y_n x_{n+1}} \\
&= \pi_P(x_0) U_{x_0 y_0} Q_{y_0 y_1} Q_{y_1 y_2} \cdots Q_{y_{n-1} y_n} V_{y_n x_{n+1}} \\
&= \frac{\pi_P(x_0)}{\pi_Q(y_0)} U_{x_0 y_0} V_{y_n x_{n+1}} \mu_Q(y_0 y_1 y_2 \cdots y_n).
\end{aligned}$$

Con esta serie de igualdades, y debido a que tenemos una cantidad finita de estados, podemos tomar el máximo $\frac{\pi_P(x_0)}{\pi_Q(y_0)} U_{x_0 y_0} V_{y_n x_{n+1}}$ sobre todos los estados, lo cual es una cota para el cociente

$$\frac{\mu_P(x_0 x_1 \cdots x_{n+1})}{\mu_Q(y_0 y_1 y_2 \cdots y_n)} = \frac{\mu_P(\mathcal{B})}{\mu_Q(\varphi(\mathcal{B}))}.$$

Por la forma de la σ-álgebra de Borel y por la Observación 5.1.2, podemos afirmar que también hay una cota para

$$\frac{\mu_P(\mathcal{B})}{\mu_Q(\varphi(\mathcal{B}))}$$

con $\mathcal{B} \in B_X(\mathcal{C})$, y como $\mathcal{A}^*$ sólo es la compleción de la σ-álgebra de Borel, también hay una cota para

$$\frac{\mu_P(\mathcal{B})}{\mu_Q(\varphi(\mathcal{B}))}$$

con $\mathcal{B} \in \mathcal{A}^*$. Análogamente podemos encontrar una cota para $\frac{\mu_Q(\varphi(\mathcal{B}))}{\mu_P(\mathcal{B})}$, y de aquí obtenemos que las medidas $\mu_Q \circ \varphi$ y μ_P son equivalentes y ambas son ergódicas y T-invariantes en X_A. Sea $E \in \mathcal{A}^*$. Entonces $\mathbb{I}_E$ es $\mathcal{A}^*$-medible, y usando el Teorema Ergódico de Birkhoff (Teorema 3.5.1), tenemos que

$$\lim_{n\to\infty} \frac{\sum_{i=0}^{n-1} \mathbb{I}_E \circ T^i}{n} = \int \mathbb{I}_E d\mu_P \qquad \mu_P\text{-c.s.} \tag{5.6}$$

y

$$\lim_{n\to\infty} \frac{\sum_{i=0}^{n-1} \mathbb{I}_E \circ T^i}{n} = \int \mathbb{I}_E d\mu_Q \circ \varphi \qquad (\mu_Q \circ \varphi)\text{-c.s.} \tag{5.7}$$

Entonces existe un conjunto N_P tal que $\mu_P(N) = 0$ y tal que

$$\lim_{n\to\infty} \frac{\sum_{i=0}^{n-1} \mathbb{I}_E \circ T^i}{n}(w) = \int \mathbb{I}_E d\mu_P \qquad \forall \qquad w \notin N, \tag{5.8}$$

pero entonces, como son medidas equivalentes, tenemos que $\mu_Q \circ \varphi(N) = 0$. Entonces

$$\lim_{n\to\infty} \frac{\sum_{i=0}^{n-1} \mathbb{I}_E \circ T^i}{n} = \int \mathbb{I}_E d\mu_P \qquad (\mu_Q \circ \varphi)\text{-c.s.} \tag{5.9}$$

y como el límite es único en espacios métricos,

$$\mu_Q \circ \varphi(E) = \int \mathbb{I}_E d\mu_Q \circ \varphi = \int \mathbb{I}_E d\mu_P = \mu_P(E).$$

Por lo tanto, efectivamente la conjugación preserva la medida y por ende las dos cadenas son block-isomorfas (conjugadas)con lo que se concluye el resultado. $\square$

Apéndice A

UN POCO DE TEORIA DE LA MEDIDA

A.1. Integrabilidad uniforme

Definición A.1.1. Ahora decimos que $(X_n)_{n\in\mathbb{Z}}$ un proceso estocástico es u*niformemente integrable* si todas las variables del proceso son acotadas en L^1, es decir,

$$\sup_{n\in\mathbb{N}} \{\mathbb{E}[|X_n|]\} < \infty$$

y además

$$\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|\mathbb{I}_{\{|X_n|>x\}}] \xrightarrow[x\to\infty]{} 0.$$

Teorema A.1.1. *Sea $(X_n)_{n\in\mathbb{Z}}$ un proceso estocástico tal que*

$$X_n \xrightarrow[n\to\infty]{\mathbb{P}\text{-}c.s.} X$$

y además es uniformemente integrable. Entonces

$$X_n \xrightarrow[n\to\infty]{L^1} X.$$

Demostración. Observemos que sólo basta probar que

$$\mathbb{E}[|X_n - X|] = \int |X_n - X| d\mathbb{P} \xrightarrow[n\to\infty]{} 0.$$

Vemos que

$$\begin{aligned}
\mathbb{E}[|X_n - X|] &= \mathbb{E}[|X_n - X|\mathbb{I}_{\{|X_n|>x\}}] + \mathbb{E}[|X_n - X|\mathbb{I}_{\{|X_n|\le x\}}] \\
&\le \sup_{n\in\mathbb{N}}\{\mathbb{E}[|X_n - X|\mathbb{I}_{\{|X_n|>x\}}]\} + \mathbb{E}[|X_n - X|\mathbb{I}_{\{|X_n|\le x\}}],
\end{aligned}$$

y por el Teorema de Convergencia Dominada (revisar [1] pag.61), y a causa de que

$$X_n \xrightarrow[n\to\infty]{\mu\text{-C.S.}} X,$$

al tomar límite conluimos que

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}[|X_n - X|] &\leq \sup_{n\in\mathbb{N}}\{\mathbb{E}[|X_n - X|\mathbb{I}_{\{|X_n|>x\}}]\} + \mathbb{E}[\lim_{n\to\infty}|X_n - X|\mathbb{I}_{\{|X_n|\leq x\}}] \\
&= \sup_{n\in\mathbb{N}}\{\mathbb{E}[|X_n - X|\mathbb{I}_{\{|X_n|>x\}]\}} + \mathbb{E}[0] = \sup_{n\in\mathbb{N}}\{\mathbb{E}[|X_n - X|\mathbb{I}_{\{|X_n|>x\}}]\} \\
&\leq \sup_{n\in\mathbb{N}}\{\mathbb{E}[|X_n|\mathbb{I}_{\{|X_n|>x\}}]\} + \sup_{n\in\mathbb{N}}\{\mathbb{E}[|X|\mathbb{I}_{\{|X_n|>x\}}]\}.
\end{aligned}$$

Tomando ahora el límite de $x \to \infty$, a causa del Teorema de Convergencia Dominada(revisar [1] pag.61), y del hecho de que el proceso es uniformemente integrable, obtenemos que

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}[|X_n - X|] &= \lim_{x\to\infty}\lim_{n\to\infty} \mathbb{E}[|X_n - X|] \\
&\leq \lim_{x\to\infty}\sup_{n\in\mathbb{N}}\{\mathbb{E}[|X_n|\mathbb{I}_{\{|X_n|>x\}}]\} + \lim_{x\to\infty}\sup_{n\in\mathbb{N}}\{\mathbb{E}[|X|\mathbb{I}_{\{|X_n|>x\}}]\} \\
&= \lim_{x\to\infty}\sup_{n\in\mathbb{N}}\{\mathbb{E}[|X_n|\mathbb{I}_{\{|X_n|>x\}}]\} + \sup_{n\in\mathbb{N}}\{\mathbb{E}[|X|\lim_{x\to\infty}\mathbb{I}_{\{|X_n|>x\}}]\} \\
&= 0 + \sup_{n\in\mathbb{N}}\{\mathbb{E}[|X|\mathbb{I}_{\emptyset}]\} \\
&= 0 + 0,
\end{aligned}$$

lo cual concluye la prueba. □

Proposición A.1.1. *Sean $(\mathfrak{F}_n)_{n\in\mathbb{N}}$ una filtración y X una variable aleatoria integrable (esperanza finita). Entonces $(Y_n = [X|\mathfrak{F}_n])_{n\in\mathbb{N}}$ es una martingala con respecto a $(\mathfrak{F}_n)_{n\in\mathbb{N}}$ y además uniformemente integrable .*

Demostración. Para ver que es una martingala, usando la desigualdad de Jensen condicional

$$\mathbb{E}[|\mathbb{E}[X|\mathfrak{F}_n]|] \leq \mathbb{E}[\mathbb{E}[|X|\,|\,\mathfrak{F}_n]] < \infty$$

y la propiedad de medibilidad es obvia,

$$\mathbb{E}[Y_n|\mathfrak{F}_{n-1}] = \mathbb{E}[\mathbb{E}[X|\mathfrak{F}_n]|\mathfrak{F}_{n-1}] = \mathbb{E}[X|\mathfrak{F}_{n-1}] = Y_{n-1}$$

con esto vemos que efectivamente es una martingala.

Para ver que es uniformemente integrable

$$\begin{aligned}
\mathbb{E}[|Y_n|\mathbb{I}_{\{|Y_n|>a\}}] &= \mathbb{E}[|\mathbb{E}[X|\mathfrak{F}_n]|\mathbb{I}_{\{|\mathbb{E}[X|\mathfrak{F}_n]|>a\}}] \\
&\leq \mathbb{E}[\mathbb{E}[|X||\mathfrak{F}_n]\mathbb{I}_{\{\mathbb{E}[|X||\mathfrak{F}_n]>a\}}] \\
&= \mathbb{E}[|X|\mathbb{I}_{\{\mathbb{E}[|X||\mathfrak{F}_n]>a\}}] \\
&= \mathbb{E}[\mathbb{I}_{\{|X|\leq K\}}|X|\mathbb{I}_{\{\mathbb{E}[|X||\mathfrak{F}_n]>a\}}] + \mathbb{E}[\mathbb{I}_{\{|X|>K\}}|X|\mathbb{I}_{\{\mathbb{E}[|X||\mathfrak{F}_n]>a\}}] \\
&\leq K\mathbb{P}(\mathbb{E}[|X||\mathfrak{F}_n] > a) + \mathbb{E}[\mathbb{I}_{\{|X|>K\}}|X|] \\
&\leq K\frac{\mathbb{E}[\mathbb{E}[|X||\mathfrak{F}_n]]}{a} + \mathbb{E}[\mathbb{I}_{\{|X|>K\}}|X|] \\
&\leq \frac{K}{a}\mathbb{E}[|X|] + \mathbb{E}[\mathbb{I}_{\{|X|>K\}}|X|],
\end{aligned}$$

con lo cual obtenemos que

$$\mathbb{E}[|Y_n|\mathbb{I}_{\{|Y_n|>a\}}] \leq \frac{K}{a}\mathbb{E}[|X|] + \mathbb{E}[\mathbb{I}_{\{|X|>K\}}|X|]$$

tomando el límite cuando $a \to \infty$, despues $K \to \infty$ por teorema de convergencia dominada obtenemos que

$$\lim_{a\to\infty} \sup_{n\in\mathbb{N}} \mathbb{E}[|Y_n|\mathbb{I}_{\{|Y_n|>x\}}] \leq 0 + \mathbb{E}[\mathbb{I}_{\{|X|>K\}}|X|]$$

$$\begin{aligned}
\lim_{a\to\infty} \sup_{n\in\mathbb{N}} \mathbb{E}[|Y_n|\mathbb{I}_{\{|Y_n|>a\}}] &= \lim_{K\to\infty} \lim_{a\to\infty} \sup_{n\in\mathbb{N}} \mathbb{E}[|Y_n|\mathbb{I}_{\{|Y_n|>a\}}] \\
&\leq \lim_{K\to\infty} \mathbb{E}[\mathbb{I}_{\{|X|>K\}}|X|] \\
&= \mathbb{E}[\lim_{K\to\infty} \mathbb{I}_{\{|X|>K\}}|X|] \\
&= \mathbb{E}[\mathbb{I}_{\emptyset}|X|] \\
&= 0
\end{aligned}$$

y por lo tanto

$$\lim_{a\to\infty} \sup_{n\in\mathbb{N}} \mathbb{E}[|Y_n|\mathbb{I}_{\{|Y_n|>a\}}]0.$$

Por lo tanto es uniformemente integrable. □

Teorema A.1.2. *(Teorema de convergencia de martingalas) Sea $(X_n)_{n\in\mathbb{N}}$ una submartingala tal que* $\sup_{n\in\mathbb{N}} \mathbb{E}[X_n^+] < \infty$*. Entonces existe X variable aleatoria tal que*

$$X_n \xrightarrow[n\to\infty]{c.s.} X.$$

Teorema A.1.3. *Sean $(\mathfrak{F}_n)_{n\in\mathbb{N}}$ una filtración y X una variable aleatoria integrable(esperanza finita) entonces $(Y_n = \mathbb{E}[X|\mathfrak{F}_n])_{n\in\mathbb{N}}$ converge a $\mathbb{E}[X|\bigvee_{n\in\mathbb{N}}\mathfrak{F}_n]$ donde $\bigvee_{n\in\mathbb{N}}\mathfrak{F}_n = \sigma(\bigcup_{n\in\mathbb{N}}\mathfrak{F}_n)$*

Demostración.

$$\begin{aligned}
\sup_{n\in\mathbb{N}}\mathbb{E}[Y_n^+] &\leq \sup_{n\in\mathbb{N}}\mathbb{E}[|Y_n|] \\
&= \sup_{n\in\mathbb{N}}\mathbb{E}[|[X|\mathfrak{F}_n]|] \\
&\leq \sup_{n\in\mathbb{N}}\mathbb{E}[\mathbb{E}[|X||\mathfrak{F}_n]] \\
&= \sup_{n\in\mathbb{N}}\mathbb{E}[|X|] < \infty
\end{aligned}$$

de aqui que podemos ocupar el teorema anterior y entonces el proceso $(\mathbb{E}[X|\mathfrak{F}_n])_{n\in\mathbb{N}}$ es una martingala uniformemente integrable que converge c.s. a una variable aleatoria integrable, por la proposición A.1.1 converge en L^1, ahora sea $A \in \bigcup_{n\in\mathbb{N}}\mathfrak{F}_n$ tenemos que

$$\mathbb{E}[\lim_{n\to\infty} Y_n\mathbb{I}_A] = \mathbb{E}[\lim_{n\to\infty}\mathbb{E}[X|\mathfrak{F}_n]\mathbb{I}_A] = \lim_{n\to\infty}\mathbb{E}[\mathbb{E}[X|\mathfrak{F}_n]\mathbb{I}_A]$$

entonces existe $N \in \mathbb{N}$ tal que $A \in \mathfrak{F}_N$ y como estamos tratando con una filtración $A \in \mathfrak{F}_n$ $\forall n \geq N$ entonces

$$\begin{aligned}
\lim_{n\to\infty}\mathbb{E}[\mathbb{E}[X|\mathfrak{F}_n]\mathbb{I}_A] &= \lim_{n\to\infty}\mathbb{E}[\mathbb{E}[X\mathbb{I}_A|\mathfrak{F}_n]] \\
&= \lim_{n\to\infty}\mathbb{E}[X\mathbb{I}_A] \\
&= \mathbb{E}[X\mathbb{I}_A].
\end{aligned}$$

Sea

$$\mathcal{H} = \left\{ E \in \bigvee_{n\in\mathbb{N}}\mathfrak{F}_n \mid \mathbb{E}[\lim_{n\to\infty} Y_n\mathbb{I}_A] = \mathbb{E}[X\mathbb{I}_A] \right\}$$

de está definición obtenemos que

$$\bigcup_{n\in\mathbb{N}}\mathfrak{F}_n \subset \mathcal{H}$$

y ahora es fácil ver que $\mathcal{H}$ es una álgebra sea $A \in \mathcal{H}$ entonces

$$\mathbb{E}[\lim_{n\to\infty} Y_n\mathbb{I}_A] = \mathbb{E}[X\mathbb{I}_A]$$

$$\begin{aligned}
\mathbb{E}[\lim_{n\to\infty} Y_n\mathbb{I}_{A^c}] &= \mathbb{E}[\lim_{n\to\infty} Y_n(1-\mathbb{I}_A)] \\
&= \mathbb{E}[\lim_{n\to\infty} Y_n] - \mathbb{E}[\lim_{n\to\infty} Y_n\mathbb{I}_A] \\
&= \mathbb{E}[\lim_{n\to\infty} Y_n\mathbb{I}_\Omega] - \mathbb{E}[X\mathbb{I}_A] \\
&= \mathbb{E}[X\mathbb{I}_\Omega] - \mathbb{E}[X\mathbb{I}_A] \\
&= \mathbb{E}[X] - \mathbb{E}[X\mathbb{I}_A] = \mathbb{E}[X(1-\mathbb{I}_A)] \\
&= \mathbb{E}[X\mathbb{I}_{A^c}]
\end{aligned}$$

y con esto tenemos que $\Omega \backslash A \in \mathcal{H}$, como todas las σ-álgebras tienen a Ω entonces

$$\Omega \in \bigcup_{n\in\mathbb{N}} \mathfrak{F}_n.$$

Sean $A, B \in \bigcup_{n\in\mathbb{N}} \mathfrak{F}_n$ entonces $\exists n, m \in \mathbb{N}$ tal que $A \in \mathfrak{F}_n, B \in \mathfrak{F}_m$,s.p.g. $n \leq m$ lo cual implica que $A, B \in \mathfrak{F}_m$, asi $A \cup B \in \mathfrak{F}_m \subset \bigcup_{n\in\mathbb{N}} \mathfrak{F}_n$ y por ende

$$\bigcup_{n\in\mathbb{N}} \mathfrak{F}_n$$

es una álgebra, ahora veamos que $\mathcal{H}$ es una clase monótona, s.p.g hagamoslo para creciente por lo anterior ahora supongamos que $(A_n)_{n\in\mathbb{N}} \in \mathcal{H}$ una sucesión monotona de conjuntos en $\mathcal{H}$ y sea $A = \bigcup_{n\in\mathbb{N}} A_n$ entonces

$$X\mathbb{I}_{A_n} \xrightarrow[n\to\infty]{} X\mathbb{I}_A,$$

por otro lado

$$X\mathbb{I}_{A_n} \leq X,$$

de igual manera

$$\lim_{m\to\infty} Y_m\mathbb{I}_{A_n} \xrightarrow[n\to\infty]{} \lim_{m\to\infty} Y_m\mathbb{I}_A$$

y además

$$\lim_{m\to\infty} Y_m\mathbb{I}_{A_n} \leq \lim_{m\to\infty} Y_m$$

donde ambas cotas esta en L^1 entonces por el teorema de convergencia dominada (revisar [1] pag.61) tenemos que

$$\mathbb{E}[X\mathbb{I}_A] = \lim_{n\to\infty} \mathbb{E}[X\mathbb{I}_{A_n}] = \lim_{n\to\infty} \mathbb{E}[\lim_{m\to\infty} Y_m\mathbb{I}_{A_n}]$$

$$\mathbb{E}[\lim_{n\to\infty}\lim_{m\to\infty} Y_m\mathbb{I}_{A_n}] = \mathbb{E}[\lim_{m\to\infty} Y_m\mathbb{I}_A]$$

y con esto obtenemos que $A \in \mathcal{H}$ lo cual nos dice que $\mathcal{H}$ es una clase monotona que contiene al álgebra, es decir tenemos que

$$\bigcup_{n\in\mathbb{N}} \mathfrak{F}_n \subseteq \mathcal{H}$$

con lo cual por el lema de clases monotonas (revisar [5] pag.116) tenemos que $\bigvee_{n\in\mathbb{N}} \mathfrak{F}_n \subset \mathcal{H}$ y por lo tanto

$$\mathbb{E}[\lim_{n\to\infty} Y_n\mathbb{I}_A] = \mathbb{E}[X\mathbb{I}_A]$$

$\forall A \in \bigvee_{n\in\mathbb{N}} \mathfrak{F}_n$ por lo tanto

$$\mathbb{E}[X|\mathfrak{F}_n] \xrightarrow[n\to\infty]{c.s.} \mathbb{E}[X|\bigvee_{n\in\mathbb{N}} \mathfrak{F}_n]$$

$\square$

A.2. Medidas de Probabilidad Ergódicas

Para comprender por completo las pruebas a la mayoria de los enunciados que estan a continuación, es conveniente leer antes definiciones básicas topológicas usadas de manera natural en este texto, que se pueden hallar en [6] como normalidad , regularidad (axiomas de separación en general), compacidad y compacidad local. Un resultado básico necesario es el que sigue a continuación.

Teorema A.2.1. *Sea X un espacio topológico compacto Hausdorff, las siguientes son equivalentes :*

1. *X es metrizable.*
2. *X es 2-numerable.*
3. *$\mathcal{C}(X)$ (el espacio de todas las funciones continuas complejo valuadas en X) es separable.*

Definición A.2.1. Sea un espacio topológico compacto X definimos :

$$\mathcal{M}(X) = \{m : m \text{ es una medida de probabilidad en la } \sigma\text{-álgebra de borel de} X\}$$

Definición A.2.2. Definimos la *topología debil** en $\mathcal{M}(X)$ como la topología mas pequeña que hace que cada una de las funciones $\mu \mapsto \int_X f d\mu$ con $f \in \mathcal{C}(X)$ sean continuas. Una base es dada por la colección de todos los conjuntos de la forma

$$\mathcal{V}_\mu(f_1, \cdots f_k; \epsilon) = \left\{ m \in \mathcal{M}(X) \middle| \left| \int f_i dm - \int f_i d\mu \right| < \epsilon, 1 \leq i \leq k \right\}$$

donde $\mu \in \mathcal{M}(X), k \geq 1, f_i \in \mathcal{C}(X)$ y $\epsilon > 0$.

Teorema A.2.2. *Si X es un espacio compacto metrizable entonces el espacio $\mathcal{M}(X)$ es metrizable bajo la topología debil*. Si $(f_i)_{i \in \mathbb{N}} \subseteq \mathcal{C}(X)$ es un conjunto denso entonces*

$$\mathcal{D}(m, \mu) = \sum_{n=1}^{\infty} \frac{|\int f_n dm - \int f_n d\mu|}{2^n ||f_n||}$$

es una metrica para dicho espacio.

La demostración de este teorema se sale un poco del contexto del trabajo, puede revisarse en [7] pag. 148.

Teorema A.2.3. *Sea μ una medida de probabilidad de borel (es decir, una medida en la σ-álgebra de borel) de un espacio métrico. Entonces μ es regular(es decir, $\forall B \in \mathcal{B}(X)$ y $\forall \epsilon > 0 \; \exists$ un abierto U_ϵ y un cerrado C_ϵ con $C_\epsilon \subseteq B \subseteq U_\epsilon$ y $\mu(U_\epsilon \backslash C_\epsilon) < \epsilon$).*

Corolario A.2.1. *Una medida de probabilidad μ de borel de un espacio métrico, para $B \in \mathcal{B}(X)$ se tiene que*

$$\mu(B) = \sup_{C \text{ cerrado}, C \subseteq B} \mu(C) \text{ y } \mu(B) = \sup_{A \text{ abierto}, A \supseteq B} \mu(A)$$

Las demostraciones del corolario y teorema anteriores son resultados clasicos de teoria de la medida, se pueden revisar en [7] pag. 147.

Teorema A.2.4. *Sean μ, ν dos medidas de probabilidad de borel en un espacio metrico X. Si $\int_X f d\nu = \int_X f d\mu \, \forall f \in \mathcal{C}(X)$ entonces $\mu = \nu$.*

Demostración. Por el corolario A.2.1 basta revisar $\mu(C) = \nu(C)$ con C cerrado.Supongamos C cerrado y $\epsilon > 0$.Por la regularidad de μ existe un conjunto abierto U con $C \subseteq U$ con $\mu(U \backslash C) < \epsilon$. Definimos $f\colon X \longrightarrow \mathbb{R}$ por :

$$f(x) = \begin{cases} 0 & \text{si } x \notin U \\ \frac{d(x, X\backslash U)}{d(x, X\backslash U) + d(x, C)} & \text{si } x \in U \end{cases}$$

Esta bien definida pues no tiene denominador 0. A causa de que en cualquier cerrado contenido en $(0,1)$ la $f(x) = \frac{d(x, X\backslash U)}{d(x, X\backslash U) + d(x, C)}$ que es una función continua y asi la imagen inversa de dicho es cerrada, $f = 0$ en $X \backslash U$, $f = 1$ en C entonces f es continua. Así obtenemos

$$\mu(C) \leq \int_X f d\mu = \int_X f d\nu \leq \nu(U) < \nu(C) + \epsilon$$

y como esto se ocurre para ϵ arbitraria positiva, obtenemos

$$\mu(C) \leq \nu(C)$$

y analogamente

$$\nu(C) \leq \mu(C),$$

lo que concluye la prueba. □

Observación A.2.1. Sea X un espacio métrico compacto.

1. $\mu_n \longrightarrow \mu$ en la topología *debil** si y sólo si $\forall\, f \in \mathcal{C}(X) \qquad \int f d\mu_n \longrightarrow \int f d\mu$

2. $\mu_n \mu \in \mathcal{M}(X), n \geq 1$ lo siguientes son equivalentes :

   a) $\mu_n \longrightarrow \mu$

   b) Para cada cerrado F subconjunto de X, $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$ y para cada abierto U de X tenemos que $\liminf_{n\to\infty} \mu_n(U) \geq \mu(U)$

   c) Para cada boreliano B con $\mu(\partial(B)) = 0$ tenemos que $\mu_n(B) \longrightarrow \mu(B)$

Demostración. El primero es obvio del hecho que $\mathcal{M}(X)$ es metrizable y por su forma de la métrica. Como en el trabajo sólo ocupamos que $2a$ implica $2c$, sólo probaremos dicha implicación.

Sea F un cerrado y $k > 0$, definimos $U_k = \{x \in X | d(x, F) < \frac{1}{k}\}$, dichos conjuntos U_k es una familia de abiertos que decrecen a F entonces $\mu(U_k) \longrightarrow \mu(F)$. Debido a que X es compacto también es localmente compacto usando el teorema 6.4 de [6] pag. 238 obtenemos que X es regular y por ser compacto es normal, ahora usando el lema de *Urysonh* (se puede checar en el mismo texto [6] pag.146), existe $f_k \in \mathcal{C}(X)$ tal que $0 \leq f_k(x) \leq 1$, $f_k(x) = 0$ en $X \backslash U_k$ y $f_k(x) = 1$ en F ,entonces

$$\limsup_{n\to\infty} \mu_n(F) \leq \limsup_{n\to\infty} \int f_k d\mu_n = \int f_k d\mu \leq \mu(U_k)$$

entonces

$$\limsup_{n\to\infty} \mu_n(F) \leq \mu(F).$$

Ahora usando este mismo hecho tenemos que

$$\limsup_{n\to\infty} \mu_n(X\backslash U) \leq \mu(X\backslash U)$$

de aquí que

$$\liminf_{n\to\infty} \mu_n(U) \geq \mu(U),$$

de esto tenemos la primera implicación.

De aquí seguimos con suponer que ocurre el segundo punto y que $\mu(\partial(A)) = 0$ entonces

$$\mu(int(A)) = \mu(A) = \mu(\overline{A}).$$

Como

$$\limsup_{n\to\infty} \mu_n(\overline{A}) \leq \mu(\overline{A}) = \mu(A) \text{ y } \liminf_{n\to\infty} \mu_n(int(A)) \geq \mu(intA) = \mu(A)$$

y a razon de que

$$\begin{aligned} \mu(A) &= \mu(intA) \leq \liminf_{n\to\infty} \mu_n(int(A)) \\ &\leq \liminf_{n\to\infty} \mu_n(A) \leq \limsup_{n\to\infty} \mu_n(A) \\ &\leq \limsup_{n\to\infty} \mu_n(\overline{A}) \leq \mu(\overline{A}) \\ &= \mu(A) \end{aligned}$$

de lo cual concluimos que $\mu_n(A) \longrightarrow \mu(A)$ □

Teorema A.2.5. *(Teorema de Representación de Riesz) Sean X un espacio métrico compacto y $J\colon \mathcal{C}(X) \mapsto \mathbb{C}$ una funcional lineal continua tal que J es un operador positivo (es decir si $f \geq 0$ entonces $J(f) \geq 0$) y $J(1) = 1$. Entonces existe $\mu \in \mathcal{M}(X)$ tal que*

$$J(f) = \int_X f d\mu \qquad \forall f \in \mathcal{C}(X)$$

Teorema A.2.6. *Sea X un espacio topológico. Si X es compacto metrizable entonces $\mathcal{M}(X)$ es compacto metrizable con la topología debil*.*

Demostración. Denotaremos $\mu(f) = \int f d\mu$. Sean una sucesion μ_n en $\mathcal{M}(X)$ y $f_1, f_2, f_3, \cdots$ un conjunto denso en $\mathcal{C}(X)$, la sucesion $\mu_n(f_1)$ es acotada en $\mathbb{C}$ entonces tiene una subsucesion convergente $\mu_n^1(f_1)$, de igual manera $\mu_n^1(f_2)$ es acotada entonces tiene una subsucesion convergente $\mu_n^2(f_2)$, aqui hacemos notar que $\mu_n^2(f_1)$ tambien converge, asi sucesivamente construimos μ_n^i una subsucesion convergente de μ_n de tal manera que $\mu_n^i \subseteq \mu_n^{i-1} \subseteq \cdots \subseteq \mu_n^1 \subseteq \mu_n$, además $\mu_n^i(f_j)$ converge para cada $j \leq i$, así tomamos la sucesion diagonal μ_n^n, con esto no es muy dificil ver que $\mu_n^n(f_i)$ converge $\forall i$. Sea $f \in \mathcal{C}(X)$, como $(f_i)_{i\in\mathbb{N}}$ es denso entonces existe $(f_{n_i})_{i\in\mathbb{N}}$ una sucesion que converge a f en $\mathcal{C}(X)$, y asi

$$\mu_n^n(f) = \int f d\mu_n^n = \int \lim_{i\to\infty} f_{n_i} d\mu_n^n = \lim_{i\to\infty} \int f_{n_i} d\mu_n^n$$

la última igualdad es debido a que $f_{n_i} \overset{\text{unif}}{\longrightarrow} f$, de igual manera $\lim_{i\to\infty} \int f_{n_i} d\mu_n^n$ converge, con lo cual $\mu_n^n(f)$ también converge. Sea

$$J(f) = \lim_{n\to\infty} \mu_n^n(f),$$

por propiedades del limite y la integral J es un funcional lineal y como

$$|J(f)| = |\lim_{n\to\infty} \mu_n^n(f)| \leq ||f||,$$

tenemos que es continua y también es positiva entonces por el teorema A.2.5 (Teorema de Representación de Riesz) existe una medida de borel μ en X tal que $J(f) = \int f d\mu$ de aqui $\mu_n^n \longrightarrow \mu$, y lo matrizable sale de el Teorema A.2.2 lo que nos permite concluir que $\mathcal{M}(X)$ es compacto metrizable. $\square$

Definición A.2.3. Sea X un espacio metrico compacto, A un boreliano y $T\colon X \longrightarrow X$ medible denotamos $(\tilde{T}\mu)(A) = \mu(T^{-1}(A))$

Observación A.2.2.

$$\int f d(\tilde{T}\mu) = \int f \circ T d\mu$$

Demostración. Nota que para cualquier indicadora medible se vale la observación de aqui se vale para funciones medibles por el procedimiento estandar, en particular para continuas. $\square$

Corolario A.2.2. *Sean X un espacio métrico compacto, $T\colon X \longrightarrow X$ continua y $\mu \in \mathcal{M}(X)$. Entonces μ es T-invariante si y sólo si*

$$\int f d\mu = \int f \circ T d\mu$$

Demostración. Primero si $\mu \in \mathcal{M}(X)$ y μ es T-invariante tenemos que $\tilde{T}\mu = \mu$, por la observación 5.1.2 obtenemos que

$$\int f d\mu = \int f \circ T d\mu.$$

Para la otra implicación como

$$\int f d\mu = \int f \circ T d\mu \text{ y } \int f d(\tilde{T}\mu) = \int f \circ T d\mu,$$

asi obtenemos que

$$\int f d\mu = \int f d(\tilde{T}\mu)$$

lo cual por el teorema A.2.4 llegamos a que $\mu = \tilde{T}\mu$. $\square$

Definición A.2.4. Sean X un conjunto, $\mathcal{S} \subseteq \wp(X)$ ($\wp(X)$ el conjunto potencia). Decimos que $\mathcal{S}$ es una ***semi-álgebra*** si cumple con lo siguiente:

1. $X, \emptyset \in \mathcal{S}$
2. $A, B \in \mathcal{S}$ entonces $A \cap B \in \mathcal{S}$

3. $A \in \mathcal{S}$ entonces existe $(A_i)_{i=1,\dots,n} \subseteq \mathcal{S}$ tal que $A_1 = A$, $A_i \subseteq A_{i+1}$ y

$$A_{i+1} \backslash A_i \in \mathcal{S} \; \forall i = 1, \dots, n$$

Una medida sobre $\mathcal{S}$ es una medida solamente que es σ-aditiva para el caso de las uniones numerables ajenas que caen en $\mathcal{S}$.

Observación A.2.3. $\mathcal{M}(X)$ es un conjunto convexo.

Esto debido a que $\alpha \in [0,1]$ $\mu, \nu \in \mathcal{M}(X)$ implíca que $\alpha\mu + (1-\alpha)\nu \in \mathcal{M}(X)$

Lema A.2.1. *Los puntos extremos de $\mathcal{M}(X)$ son medidas ergódicas.(Un punto extremo μ es aquel que si $0 < \alpha < 1$ tal que $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$, entonces $\mu_1 = \mu_2$)*

Demostración. Si μ es una medida que no es ergódica entonces existe $B \in \mathcal{B}$ tal que $T^{-1}(B) = B$ con $0 < \mu(B) < 1$.

$$\mu(A) = \mu(A \cap B) + \mu(A \cap X \backslash B) = \mu(B)\frac{A \cap B}{\mu(B)} + \mu(X \backslash B)\frac{\mu(A \cap X \backslash B)}{\mu(X \backslash B)}.$$

Sea $\alpha = \mu(B), \mu_1(A) = \frac{A \cap B}{\mu(B)}$ y $\mu_2(A) = \frac{\mu(A \cap X \backslash B)}{\mu(X \backslash B)}$ entonces

$$\mu = \alpha\mu_1 + (1-\alpha)\mu_2,$$

claramente μ_1 y μ_2 son diferentes lo cual termina la proposición. □

Teorema A.2.7. *Sean X un espacio metrico compacto, $\mathcal{S}$ una semiálgebra que genera a los borelianos cuyos elementos son abiertos y cerrados simultaneamente, $T \colon X \longrightarrow X$ continua, $\mu_{\mathcal{S}}$ una medida sobre $\mathcal{S}$ entonces el conjunto*

$$\mathcal{M}(X, T, \mu_{\mathcal{S}}) = \{\mu \in \mathcal{M}(X) | \mu \text{ es } T\text{-invariante y } \mu|_{\mathcal{S}} = \mu_{\mathcal{S}}\}$$

es compacto.

Demostración. Sólo hace falta ver que dicho conjunto es cerrado. Sea $(\mu_n)_{n \in \mathbb{N}} \subseteq \mathcal{M}(X, T, \mu_{\mathcal{S}})$ con $\mu_n \longrightarrow \mu \in \mathcal{M}(X)$, ocurre que

$$\int f \circ T d\mu = \lim_{n \to \infty} \int f \circ T d\mu_n = \lim_{n \to \infty} \int f d\mu_n = \int f d\mu$$

debido a la observación A.2.1 y a la hipótesis de la sucesión, en consecuencia por el corolario A.2.2 se tiene que μes T-invariante. Ahora para ver la otra condición, tenemos que los elementos de $\mathcal{S}$ son abiertos y cerrados, por ende su frontera es vacía y a causa de que $\mu_n \longrightarrow \mu$, por la observación A.2.1 $\mu_n(B) \longrightarrow \mu(B) \; \forall B \in \mathcal{S}$. Sea $B \in \mathcal{S}$ $\mu(B) = \lim_{n \to \infty} \mu_n(B) = \lim_{n \to \infty} \mu_{\mathcal{S}}(B) = \mu_{\mathcal{S}}(B)$ con lo cual $\mu \in \mathcal{M}(X, T, \mu_{\mathcal{S}})$ lo q concluye la prueba. □

Teorema A.2.8. *Sean X un espacio metrico compacto,$\mathcal{S}$ una semiálgebra que genera a los borelianos cuyos elementos son abiertos y cerrados simultaneamente, $T \colon X \longrightarrow X$ continua, una medida $\mu_{\mathcal{S}}$ sobre $\mathcal{S}$ entonces existe una medida de probabilidad T-ergódica μ tal que $\mu|_{\mathcal{S}} = \mu_{\mathcal{S}}$.*

Demostración. Primero elegimos $(f_k)_{k\in\mathbb{N}}$ un conjunto denso de $\mathcal{C}(X)$ ahora como la función $\mu \mapsto \int f_0 d\mu$ es continua en

$$\mathcal{M}(X,T,\mu_S),$$

debido a que es compacto existe al menos $\nu \in \mathcal{M}(X,T,\mu_S)$ tal que

$$\int f_0 d\nu = \sup_{\mu\in\mathcal{M}(X,T,\mu_S)} \int f_0 d\mu.$$

De nuevo sea

$$\mathcal{N}_0 = \left\{ \nu \in \mathcal{M}(X,T,\mu_S) | \int f_0 d\nu = \sup_{\mu\in\mathcal{M}(X,T,\mu_S)} \int f_0 d\mu \right\}$$

por ser *Haussdorf* $\mathbb{C}$ y $\mu \mapsto \int f_0 d\mu$ una función continua $\mathcal{N}_0$ es cerrado y por tanto compacto, de manera analoga existe al menos $\nu \in \mathcal{M}(X,T,\mu_S)$ tal que

$$\int f_1 d\nu = \sup_{\mu\in\mathcal{N}_0} \int f_1 d\mu.$$

Similarmente

$$\mathcal{N}_1 = \left\{ \nu \in \mathcal{N}_0 | \int f_1 d\mu = \sup_{\mu\in\mathcal{N}_0} \int f_1 d\mu \right\}$$

y siguiendo el mismo razonamiento $\mathcal{N}_1 \subseteq \mathcal{N}_0 \subseteq \mathcal{M}(X,T,\mu_S)$ todos no vacios y cerrados. Definimos de manera inductiva

$$\mathcal{N}_k = \left\{ \nu \in \mathcal{N}_{k-1} | \int f_k d\mu = \sup_{\mu\in\mathcal{N}_{k-1}} \int f_k d\mu \right\}$$

y como en el razonamiento anterior $\mathcal{N}_k \subseteq \mathcal{N}_{k-1} \subseteq \cdots \mathcal{N}_1 \subseteq \mathcal{N}_0 \subseteq \mathcal{M}(X,T,\mu_S)$ donde cada conjunto es compacto y metrizable.Esta sucesion de conjuntos es decreciente anidada y de compactos entonces $\bigcap_{k\in\mathbb{N}} \mathcal{N}_k$ es no vació. Sea $\nu \in \bigcap_{k\in\mathbb{N}} \mathcal{N}_k$ probaremos que este es un punto extremo, supongamos $\nu = \alpha\mu_1 + (1-\alpha)\mu_2$ lo que necesitamos probar es que $\mu_1 = \mu_2$, primero probaremos que $\int f_k d\mu_1 = \int f_k d\mu_2$ para cualquier k. Sea $k = 0$, como $\nu \in \mathcal{N}_0$ observamos que

$$\begin{aligned}
\alpha \int f_0 d\mu_1 + (1-\alpha) \int f_0 d\mu_2 &= \int f_0 d\nu \\
&= \sup\nolimits_{\mu\in\mathcal{M}(X,T,\mu_S)} \int f_0 d\mu \\
&\geq \alpha \int f_0 d\mu_i + (1-\alpha) \int f_0 d\mu_i \text{ para } i = 1,2,
\end{aligned}$$

esto nos dice directamente que $\int f_0 d\mu_j \geq \int f_0 d\mu_i$ con $i,j = 1,2$ con lo cual $\int f_0 d\mu_1 = \int f_0 d\mu_2$.Ahora supongamos que pasa para k, como $\nu \in \mathcal{N}_{k+1}$ tenemos que

$$\begin{aligned}
\alpha \int f_{k+1} d\mu_1 + (1-\alpha) \int f_{k+1} d\mu_2 &= \int f_{k+1} d\nu \\
&= \sup\nolimits_{\mu\in\mathcal{N}_k} \int f_{k+1} d\mu \\
&\geq \alpha \int f_{k+1} d\mu_i + (1-\alpha) \int f_{k+1} d\mu_i,
\end{aligned}$$

analogamente $\int f_{k+1} d\mu_j \geq \int f_{k+1} d\mu_i$ y finalmente que $\int f_{k+1} d\mu_1 = \int f_{k+1} d\mu_2$. Sabemos que cualquier función en $\mathcal{C}(X)$ es limite uniforme de f_{n_i}'s, entonces también se vale $\int f d\mu_1 = \int f d\mu_2$ $\forall f \in \mathcal{C}(X)$, por el teorema A.2.4 tenemos que $\mu_1 = \mu_2$ y por el lema A.2.1 tenemos que ν es T-ergódica. $\square$

Apéndice B

ALGO DE ANALISIS MATEMÁTICO

B.1. Teorema de Hahn-Banach

Teorema B.1.1. *(Teorema de Hahn-Banach) Sean $(X, \| \ \|)$ un espacio de Banach, $Y \subset X$ un subespacio, $f : (Y, \| \ \|) \longrightarrow \mathbb{F}$ una funcional lineal continua. Entonces existe $\tilde{f}$ una extension de f, hacia X lineal y continua, además $\|f\| = \|\tilde{f}\|$(es decir existe $\tilde{f} : (X, \| \ \|) \longrightarrow \mathbb{F}$ una funcional lineal y continua con $\|f\| = \|\tilde{f}\|$ y $\tilde{f}|_Y = f$).*

Corolario B.1.1. *Sean $(X, \| \ \|)$ un espacio de Banach, $Y \subset X$ un subespacio cerrado propio, $x_0 \in X \backslash Y$ y $d(x, Y) = d > 0$. Entonces existe $f \in X^*$ tal que $\|f\| = 1$, $f(Y) = \{0\}$ y $f(x_0) = d$*

Corolario B.1.2. *Sean $(X^*, \| \cdot \|)$ un espacio de Banach y $Y^* \subseteq W^* \subseteq X^*$ con W^* cerrado (de Banach). Entonces Y^* es denso en W^* si y sólo si para toda $f \in W^{**}$ que se anule en Y^* tenemos que también se anula en W^* (e.d. $f(y) = 0$, $\forall y \in Y^*$ entonces $f(w) = 0$, $\forall w \in W^*$)*

Demostración. Para la ida supongamos Y^* denso en W^*, $f \in W^{**}$ con $f(y) = 0 \, \forall y \in Y^*$. Como Y^* es denso en W^*, sea $\epsilon > 0$ y $w \in W^*$ entonces existe $y \in Y^*$ tal que $\|y - w\| < \epsilon$, como f es continua existe $K \geq 0$ tal que

$$
\begin{aligned}
|f(w)| &= |f(w-y) + f(y)| \\
&= |f(w-y)| \\
&= K\|w-y\| \\
&< \epsilon K + 0 \\
&= \epsilon K
\end{aligned}
$$

y como epsilon es arbitraria concluimos que $f(w) = 0$, $\forall w \in W^*$.

Para el regreso aplicamos el corolario del teorema de Hahn-Banach $\overline{Y^*} \subseteq W^*$, supongamos que Y^* no es denso en W^*, es decir $\overline{Y^*} \neq W^*$, así existe $x_0 \in W^* \backslash \overline{Y^*}$, a causa del corolario anterior existe $f \in W^{**}$ tal que $f(\overline{Y^*}) = \{0\}$ y $f(x_0) = d$, entonces f se anula en $\overline{Y^*}$ en particular en Y^*, pero no se anula en todo W^* pues no se anula en x_0 con lo cual llegamos a una contradicción, por lo tanto $\overline{Y^*} = W^*$ con lo cual tenemos que Y^* es denso en W^*. □

Bibliografía

[1] GRAVINSKY S., GUILLERMO, *Teoría de la medida*, D.R. Universidad Nacional Autónoma de México, Facultad de Ciencias, México D.F., 1a edición, 2009.

[2] DOWNAROWICZ, TOMASZ, *Entropy in Dynamical Systems*, Cambridge University Press, Cambridge, 1a edición, 2011.

[3] LIND, DOUGLAS AND MARCUS, BRIAN, *An introduction to symbolic dynamics and coding*, Cambridge University Press, 1a edición, 1995.

[4] POLLICOTT, MARK AND YURY, MICHIKO, *Dynamical Systems and Ergodic Theory*, Cambridge University Press, 1a edición, 1998, 179pp.

[5] G.BARTLE, ROBERT, *The Elements of Integration*, Jonh Wiley and Sons, Inc., United States of America, 1a edición, 1966, 129pp.

[6] DUGUNDJI, JAMES, *Topology*, Allyn and Bacon Inc., United States of America, 1a edición, 1966, 447pp.

[7] WALTERS, PETER, *An Introduction to Ergodic Theory*, Springer-Verlag New York inc., United States of America, 1a edición, 1943, 252pp.

[8] PARRY, WILLIAM AND WILLIAMS, ROBERT, *Block coding and zeta functions of finite Markov chains*, Mathematics Institue University of Warwick, United States of America, 1976,13pp.

[9] WILLIAMS, R.F., *Classification of subshifts of finite type*, Annals of Math., 1974,381pp.

[10] AUBRUN, NATHALIE AND BÉAL, MARIE-PIERRE, *Tree-shifts of finite type.*, Theoretical Computer Science, Volume 459, 25pp., 2012.

[11] SCHRAUDNER, MICHAEL, *A matrix formalism for conjugacies of higher-dimensional shifts of finite type*, Centro de Modelamiento Matemático de la Universidad de Chile, Colloq. Math. (2008), 515 pp.