

XVIII ESCUELA VENEZOLANA DE MATEMÁTICA

Introducción al estudio de billares caóticos

Nikolai Chernov ¹ y Roberto Markarian ²

¹N. Chernov, Department of Mathematics University of Alabama at Birmingham, Birmingham, AL 35294, USA.
C-electrónico: chernov@math.uab.edu

²R. Markarian, Instituto de Matemática y Estadística “Prof. Ing. Rafael Laguardia”, Facultad de Ingeniería,
Universidad de la República, C.C. 30, Montevideo, Uruguay. C-electrónico: roma@fing.edu.uy

Índice general

I. Transformaciones que preservan medida. Entropía	7
I.1. Definición. Ejemplos	7
I.2. Teorema de Recurrencia de Poincaré	11
I.3. El Problema de la Equivalencia	12
I.4. Entropía	13
II. Ergodicidad	17
II.1. Teorema Ergódico de Birkhoff-Khinchin	17
II.2. Ergodicidad	21
II.3. Jerarquía Ergódica	24
II.4. Propiedades Estadísticas de Sistemas Dinámicos	25
III. Exponentes de Liapunov. Teoría de Pesin	29
III.1. Exponentes de Liapunov	29
III.2. Teorema de Oseledec	32
III.3. Teoría de Pesin. Hiperbolicidad no uniforme	33
III.4. Condiciones Suficientes para la Hiperbolicidad	39
III.5. Condiciones Suficientes para la Ergodicidad	41
IV. Billares. Propiedades generales. Ejemplos.	47
IV.1. Billares Planos. Generalidades	47
IV.2. Cuadrados y Círculos.	50
IV.3. Curvas Convexas y Estadios.	54
IV.4. Medidas. Teorema de Oseledec	58
IV.5. Billares en Dimensiones Mayores. Modelos	61
IV.6. Otras Propiedades Geométricas	65
V. Billares. Hiperbolicidad	71
V.1. Frentes de Onda y Conos.	71
V.2. Billares Dispersores.	76
V.3. Billares de Bunimovich	81
V.4. Otros Billares Hiperbólicos.	83
V.5. Observaciones Finales y Perspectives	90
A. Teoría de la medida. Ejemplos de ergodicidad	93
A.1. Medidas	93
A.2. Transformaciones medibles	95
A.3. Integración	96
A.4. Ergodicidad de los Shifts y de Traslaciones y Transformaciones Lineales en el Toro	97
Bibliography	101

Introducción

Estas notas en gran parte son traducción de una obra anterior de los mismos autores [CM03]. Esta versión contiene diversas correcciones, actualizaciones, algunos agregados y varias supresiones y presentaciones más breves.

Se ha reducido en forma importante la parte dedicada a los aspectos generales de Teoría Ergódica, que ahora constituyen los capítulos I, II y el Apéndice, y antes eran más de la mitad del libro. Se han eliminado -por razones de espacio- varias demostraciones, en particular la del Teorema de Birkhoff - Khincin. Se ha eliminado también la prueba de la ergodicidad de un tipo particular de billares (stadium elíptico)

Se dan en estas notas versiones más precisas y actualizadas de uno de los teoremas sobre condiciones suficientes para la ergodicidad local (Sección III.5), y se ha ampliado con ejemplos, figuras y comentarios actualizados el anterior capítulo sobre Billares, que ahora son dos.

En las últimas tres décadas del siglo XX, los billares caóticos se transformaron en un área muy activa de investigación en mecánica estadística y disciplinas relacionadas. Comenzó con el trabajo seminal de Ya. G. Sinai de 1970 [Si70], en que se desarrolló el aparato matemático para el estudio de hiperbolicidad y propiedades ergódicas de una clase muy amplia de billares planos, y se dió una fórmula exacta para su entropía. La teoría de Sinai llevó a una explosión de trabajos en revistas de matemática y física sobre diversas clases de billares en el plano y en espacios de cualquier dimensión. El progreso fundamental de la teoría llevó a la solución, en algún sentido, de la clásica hipótesis de Boltzmann (enunciada por 1880) sobre la ergodicidad del modelo de bolas duras de los gases (ver Sección IV.5). El avance en el estudio de los billares también penetró en la mecánica estadística del no equilibrio y otras ciencias.

El propósito de estas notas es introducir al lector al estudio actualizado de los billares caóticos y habilitarlo para avanzar más rápido en el estudio de los trabajos originales. Por ello, entre otras razones, la bibliografía es amplia y variada. Está dirigido a estudiantes de posgrado e investigadores, tanto físicos como matemáticos. Conocimientos básicos de cálculo, geometría Riemanniana y probabilidad son necesarios. En las notas se incluye un panorama extenso de los elementos necesarios de teoría de medida.

Los Capítulos I y II y el actual Apéndice deben mucho a unas notas de Ricardo Mañé para un curso en el ICTP, Trieste. 1988. Él me había amablemente autorizado a incluirlas en mi versión original del libro, que es una monografía publicada en 1993.

Agradezco: los aportes y correcciones de diversos alumnos de cursos sobre estos temas dictados en el último lustro, a Nelson Möller por haber hecho algunas de las Figuras, a los organizadores de la XVIII Escuela Venezolana de Matemática por darme la oportunidad de dictar este curso, en especial a Carlos Di Prisco por apoyos de todo tipo.

Montevideo, junio 2005, Roberto Markarian

Capítulo I

Transformaciones que preservan medida. Entropía

I.1. Definición. Ejemplos

Si (X, \mathcal{O}, μ) y (Y, \mathcal{S}, ν) son espacios de medida, decimos que la transformación $T : X \rightarrow Y$ *preserva medida* si

$$B \in \mathcal{S} \Rightarrow T^{-1}(B) \in \mathcal{O} \text{ y } \mu(T^{-1}(B)) = \nu(B).$$

Si T va de un espacio de medida (X, \mathcal{O}, μ) en si mismo, suele decirse que μ *es preservada por T* o que T *preserva μ* .

Un método útil y simple para verificar que una transformación preserva medida es el siguiente:

Proposición I.1.1. Sean (X, \mathcal{O}, μ) y (Y, \mathcal{S}, ν) espacios de probabilidad y $\mathcal{S}_0 \subset \mathcal{S}$ un álgebra que genera \mathcal{S} . Si $T^{-1}(B) \in \mathcal{O}$ y $\mu(T^{-1}(B)) = \nu(B)$ para todo $B \in \mathcal{S}_0$, entonces T preserva medida.

Demostración. Sea $\hat{\mathcal{S}}$ una familia de conjuntos $B \in \mathcal{S}$ tales que $T^{-1}(B) \in \mathcal{O}$ y $\mu(T^{-1}(B)) = \nu(B)$. Es sencillo ver que $\hat{\mathcal{S}}$ es un σ -álgebra que contiene a \mathcal{S}_0 . Entonces $\hat{\mathcal{S}} \supset \mathcal{S}$. Por tanto $B \in \mathcal{S}$ implica $T^{-1}(B) \in \mathcal{O}$ y $\mu(T^{-1}(B)) = \nu(B)$. \square

Ejemplos:

1. **Traslaciones en \mathbb{T}^n .** Para cada $k = (k_1, \dots, k_n) \in \mathbb{T}^n$ se define la traslación $L_k : \mathbb{T}^n \rightarrow \mathbb{T}^n$ por

$$L_k(x_1, \dots, x_n) = (k_1 x_1, \dots, k_n x_n)$$

Proposición I.1.2. La probabilidad de Lebesgue en \mathbb{T}^n es invariante bajo todas las traslaciones. Más aún, es la única probabilidad sobre el σ -álgebra de Borel de \mathbb{T}^n con esa propiedad.

Demostración. Sea L_k una traslación. Si $Q \subset \mathbb{T}^n$ es un cubo, es claro que

$$\lambda(L_k^{-1}(Q)) = \text{Vol } L_k^{-1}(Q) = \text{Vol } Q = \lambda(Q).$$

Entonces, aplicando la Proposición I.1.1 al álgebra de las uniones disjuntas de cubos, se obtiene que λ es L_k -invariante. Inversamente, suponga que λ' es otra probabilidad sobre el σ -álgebra de Borel de \mathbb{T}^n invariante bajo todas las traslaciones. Sea $\{Q_1, \dots, Q_m\}$ una colección finita de cubos disjuntos cubriendo \mathbb{T}^n , y que cada cubo es traslación de un mismo cubo. Entonces $\lambda'(Q_1) = \dots = \lambda'(Q_m)$ y $\lambda(Q_1) = \dots = \lambda(Q_m)$. Por tanto, si $\lambda'(Q_i) \neq \lambda(Q_i)$ para algún i , digamos $\lambda'(Q_i) > \lambda(Q_i)$, esto implicará de inmediato que $\lambda'(Q_j) > \lambda(Q_j)$ para todo $1 \leq j \leq m$. esto implicaría que $1 = \lambda'(\mathbb{T}^n) = \sum_j \lambda'(Q_j) > \sum_j \lambda(Q_j) = \lambda(\mathbb{T}^n) = 1$, una contradicción. Por tanto, $\lambda(Q_i) = \lambda'(Q_i)$ para todo i . Esto muestra que λ y λ' coinciden sobre cubos pertenecientes a este tipo de partición. Pero es simple ver que todo cubo $Q \subset \mathbb{T}^n$ es una unión (disjunta) de cubos pertenecientes a esa partición. Entonces λ y

λ' coinciden sobre la familia de todos los cubos, y por lo tanto coinciden sobre el álgebra de las uniones disjuntas de cubos. Así $\lambda = \lambda'$ sobre el σ -álgebra de Borel de \mathbb{T}^n . \square

2. Transformaciones lineales de \mathbb{T}^n . Defina la transformación $\pi : \mathbb{R}^n \rightarrow \mathbb{T}^n = S^1 \times \cdots \times S^1$ por

$$\pi(t_1, \dots, t_n) = (e^{2\pi i t_1}, \dots, e^{2\pi i t_n}).$$

Claramente $\pi(x) = \pi(y) \iff x - y \in \mathbb{Z}^n$ y $\pi(\mathbb{Z}^n) = \{(1, \dots, 1)\}$. Dado un isomorfismo lineal $\hat{T} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ tal que $\hat{T}(\mathbb{Z}^n) \subset \mathbb{Z}^n$, o, equivalentemente, cuya matriz en la base canónica tiene componentes enteras, existe $T : \mathbb{T}^n \rightarrow \mathbb{T}^n$ tal que $\pi \circ \hat{T} = T \circ \pi$. La transformación T se define tomando para cada $p \in \mathbb{T}^n$, un punto $x \in \pi^{-1}(p)$ y estableciendo

$$T(p) = \pi[\hat{T}(x)].$$

Esta definición es consistente porque si $y \in \pi^{-1}(p)$ es otro punto, con $x - y \in \mathbb{Z}^n$ entonces $\hat{T}(x - y) \in \mathbb{Z}^n$, y así $\pi[\hat{T}(y)] = \pi[\hat{T}(x)]$. Transformaciones construídas como la T de arriba son llamadas *transformaciones lineales del toro \mathbb{T}^n* , y \hat{T} es llamado el *levantamiento lineal de T al espacio de (re)cubrimiento \mathbb{R}^n* .

Proposición I.1.3. *La probabilidad de Lebesgue sobre \mathbb{T}^n es invariante bajo toda transformación lineal de \mathbb{T}^n .*

Demostración. Sea $T : \mathbb{T}^n \rightarrow \mathbb{T}^n$ una transformación lineal. Defina una probabilidad ν sobre la σ -álgebra de Borel de \mathbb{T}^n estableciendo que $\nu(A) = \lambda(T^{-1}(A))$. Si $L_k : \mathbb{T}^n \rightarrow \mathbb{T}^n$ es una traslación y $Tx = k$, tenemos, para todo conjunto de Borel:

$$\nu(L_k^{-1}(A)) = \lambda(T^{-1}L_k^{-1}(A)) = \lambda(L_x^{-1}T^{-1}(A)) = \lambda(T^{-1}(A)) = \nu(A)$$

Así ν es invariante bajo todas las traslaciones, por lo que $\lambda = \nu$. Esto prueba que λ es invariante bajo T . \square

3. Shifts de Bernoulli. Sea X un espacio métrico compacto.¹ Llamemos $B(X)$ al espacio de las sucesiones dobles $\theta : \mathbb{Z} \rightarrow X$ con la métrica

$$d(\alpha, \beta) = \sum_{n=-\infty}^{\infty} \frac{1}{k^{|n|}} d_0(\alpha(n), \beta(n))$$

donde $k > 1$ es una constante y d_0 la métrica de X . Observe que en esta métrica, una sucesión de sucesiones $\{\theta_n\} \in B(X)$ converge a una sucesión $\theta \in B(X)$ si y sólo si converge cada una de sus componentes, o sea, si $\lim_{n \rightarrow \infty} \theta_n(j) = \theta(j)$ para todo $j \in \mathbb{Z}$. Así la convergencia es independiente de la constante $k > 1$ usada para definir la métrica d . El *shift (transformación de corrimiento) $\sigma : B(X) \rightarrow B(X)$* es definido por

$$(\sigma\theta)(n) = \theta(n + 1)$$

Claramente σ es un homeomorfismo. Cuando X es un conjunto finito, $X = \{1, \dots, m\}$, denominamos $B(\{1, \dots, m\})$ simplemente por $B(m)$.

Dados conjuntos de Borel A_0, \dots, A_m en X y $j \in \mathbb{Z}$ definimos un *cilindro* $C(j, A_0, \dots, A_m)$ por

$$C(j, A_0, \dots, A_m) = \{\theta \in B(X) \mid \theta(j + i) \in A_i, \quad 0 \leq i \leq m.\}$$

Uniones finitas de cilindros disjuntos forman un álgebra que genera el σ -álgebra de Borel de $B(X)$. Más aún, dada una medida de probabilidad μ_0 sobre el σ -álgebra de Borel de X , existe una única probabilidad μ sobre el σ -álgebra de Borel de $B(X)$ (llamada la *medida producto asociada con μ_0*) tal que para todo cilindro:

$$\mu(C(j, A_0, \dots, A_m)) = \prod_{i=0}^m \mu_0(A_i) \tag{I.1.1}$$

¹Obsérvese que a los efectos de la pura teoría ergódica, no importa trabajar sobre un espacio topológico, por lo que alcanzaría con tener la probabilidad μ_0 indicada un poco más adelante definida sobre un σ -álgebra de X .

La existencia y unicidad de μ puede ser deducida del Teorema A.1.3 siguiendo una construcción similar a la de la medida de Lebesgue sobre \mathbb{R}^n o \mathbb{T}^n . Más aún, μ es invariante bajo σ . Esto se deduce de $\mu(\sigma^{-1}(C)) = \mu(C)$ para cada cilindro C , como se verifica usando la fórmula de arriba y del hecho de que las uniones disjuntas de cilindros forman un álgebra que genera el σ -álgebra de Borel de $B(X)$. Denote por $B_\mu(X)$ al espacio $B(X)$ con la probabilidad μ . El shift $\sigma : B_\mu(X) \rightarrow B_\mu(X)$ es llamado *shift de Bernoulli*.

Cuando X es un conjunto finito, $X = \{1, \dots, m\}$, la probabilidad μ_0 es determinada por los números $p_i = \mu_0(\{i\})$ y en este caso $B_\mu(\{1, \dots, m\})$ es denotada por $B(p_1, \dots, p_m)$.

De manera similar definimos $B^+(X)$, el espacio de las sucesiones $\theta : \mathbb{Z}^+ \rightarrow X$ con la correspondiente métrica. El *shift unilateral* $\sigma : B^+(X) \rightarrow B^+(X)$ se define también por la fórmula $(\sigma\theta)(n) = \theta(n+1)$, pero ahora $n \geq 0$. La transformación σ sobre $B^+(X)$ es solamente una transformación sobreyectiva continua y no un homeomorfismo. Si μ_0 es una probabilidad sobre el σ -álgebra de Borel de X , una medida producto μ sobre la σ -álgebra de Borel de $B^+(X)$ es definida por la misma fórmula para los cilindros $C(j, A_0, \dots, A_m)$ con $j \geq 0$. Otra vez μ es σ -invariante. Entonces $B_\mu^+(X)$ y $B^+(p_1, \dots, p_m)$ se definen como el caso previo.

4. Difeomorfismos que preservan volumen. Sean U y V conjuntos abiertos de \mathbb{R}^n . Decimos que un difeomorfismo $f : U \rightarrow V$ *preserva volumen* si $|\det f'(x)| = 1$ para todo $x \in U$. Entonces

$$\lambda(f^{-1}(A)) = \lambda(A)$$

para todo subconjunto de Borel $A \subset V$. Para probar esta propiedad primero observe que si $\mu^*(\cdot)$ denota la medida exterior de Lebesgue

$$\lambda(A) = \int \chi_A(x) dx$$

para todo abierto $A \subset \mathbb{R}^n$ con $\mu^*(\partial A) = 0$. Aquí la integral es de Riemann y χ_A es la función característica (indicatriz) de A (o sea $\chi_A(x) = 0$ si $x \notin A$ y $\chi_A(x) = 1$ si $x \in A$, ver también Ej. I.2.7). Pero $\mu^*(\partial A) = 0$ implica $\mu^*(\partial f^{-1}(A)) = 0$. Por tanto

$$\lambda(f^{-1}(A)) = \int \chi_{f^{-1}(A)}(x) dx.$$

Observe que $\chi_{f^{-1}(A)} = \chi_A \circ f$. Por tanto

$$\begin{aligned} \lambda(f^{-1}(A)) &= \int \chi_{f^{-1}(A)}(x) dx = \int (\chi_A \circ f) dx \\ &= \int (\chi_A \circ f) |\det f'| dx = \int \chi_A(x) dx = \lambda(A) \end{aligned}$$

donde la cuarta identidad se obtiene por cambio de variables. Esto muestra que la fórmula deseada se satisface cuando A es un conjunto abierto con $\mu^*(\partial A) = 0$; en particular, cuando A es un cubo. Luego tomamos un cubrimiento de V por cubos disjuntos Q_1, Q_2, \dots y definimos el σ -álgebra $\hat{\mathcal{O}}$ de todos los conjuntos de Borel $A \subset V$ tales que $\lambda(f^{-1}(A \cap Q_i)) = \lambda(A \cap Q_i)$ para todo i . Este es un σ -álgebra que contiene el subálgebra de las uniones disjuntas de cubos. Por tanto $\hat{\mathcal{O}}$ es el σ -álgebra de Borel y la fórmula está probada.

5. Medidas invariantes de funciones continuas en espacios compactos. Sea X un espacio métrico compacto y \mathcal{O} su σ -álgebra de Borel. Se prueba que el conjunto $m(X)$ de las probabilidades $\mu : \mathcal{O} \rightarrow [0, 1]$ admite una única topología metrizable tal que $\mu_n \rightarrow \mu$ si y sólo si

$$\int \varphi d\mu_n \rightarrow \int \varphi d\mu$$

para cada función continua $\varphi : X \rightarrow \mathbb{R}$. Con esta topología $m(X)$ se transforma en un espacio compacto, y usando sus propiedades se muestra que toda función continua $T : X \rightarrow X$ tiene por lo menos una probabilidad T -invariante. Si T no es continua puede suceder que no haya probabilidad T -invariante; vea

el Ejercicio I.6. Pruebas de este teorema se encuentran en todos los libros generales de Teoría ergódica, por ejemplo [CFS82] o [Mn83].

Si T es continua la situación más frecuente es que tenga infinitas probabilidades invariantes. Cuando T tiene sólo un número finito de probabilidades invariantes es claro que tiene sólo una (porque si tuviera dos, μ_1 y μ_2 , entonces todas las combinaciones lineales $\lambda\mu_1 + (1-\lambda)\mu_2$, $0 \leq \lambda \leq 1$, serían también probabilidades invariantes); en este caso, como veremos más abajo T tiene propiedades ergódicas muy fuertes.

Dada una función continua T , denote por $m_T(X)$ al conjunto de las probabilidades T -invariantes.

Teorema I.1.4 (Bogolyubov-Krylov, [BK37]). $m_T(X)$ es no vacío.

Como ya dijimos, puede suceder que una transformación tenga exactamente una medida invariante. En este caso se dice que es *únicamente ergódica*². La clase de transformaciones únicamente ergódicas es muy restringida, aún en el caso de las transformaciones continuas sobre espacios compactos que estamos estudiando en esta Sección. Los ejemplos estándar son ciertas traslaciones en el toro \mathbb{T}^n .

Teorema I.1.5. Sea f una traslación en \mathbb{T}^n . Entonces, las siguientes propiedades son equivalentes:

- (a) Si $e = (1, \dots, 1)$, la órbita $\{f^m(e) | m \in \mathbb{Z}\}$ es densa en \mathbb{T}^n
- (b) $\{f^m(x) | m \in \mathbb{Z}\}$ es densa en \mathbb{T}^n para todo $x \in \mathbb{T}^n$
- (c) f es únicamente ergódica.

EJERCICIO I.1. Sea X un espacio métrico compacto y $T : X \rightarrow X$ una función continua. Las siguientes propiedades son equivalentes:

- (a) T es únicamente ergódica;
- (b) Para cada $f \in C^0(X)$ el límite $\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{j=0}^n f(T^j(x))$ existe para cada $x \in X$ y no depende de x .
- (c) Para cada $f \in C^0(X)$ la sucesión de funciones $\frac{1}{n+1} \sum_{j=0}^n f \circ T^j$ converge uniformemente a una constante.

EJERCICIO I.2. Una transformación de intercambio de intervalo S^1 es una transformación $T : S^1 \rightarrow S^1$ tal que existe una familia finita de intervalos abiertos disjuntos J_1, \dots, J_n , cuyas clausuras cubren S^1 y tales que $T|_{J_i} : J_i \rightarrow T(J_i)$ es una isometría y los intervalos $T(J_i)$ son disjuntos, también. Probar que T preserva la probabilidad de Lebesgue.

EJERCICIO I.3. Sea $J_i = (a_i, b_i)$, $i \geq 1$ una familia numerable de intervalos disjuntos abiertos en $(0, 1)$ tal que

$$\sum_{i=1}^{\infty} (b_i - a_i) = 1$$

Sea $T : [0, 1] \rightarrow [0, 1]$ una transformación satisfaciendo:

$$T(x) = \sigma_i \frac{x - a_i}{b_i - a_i} + \frac{1 - \sigma_i}{2}$$

cuando $x \in J_i$. Aquí σ_i puede ser $+1$ o -1 . Probar que T preserva la probabilidad de Lebesgue.

EJERCICIO I.4. Probar que el conjunto de los puntos periódicos del shift $\sigma : B(X) \rightarrow B(X)$ es denso en $B(X)$.

EJERCICIO I.5. * Considere $B(m)$ con la métrica

$$d(\alpha, \beta) = \sum_n \frac{1}{k^{|n|}} |\alpha(n) - \beta(n)|$$

donde $k > 1$. Denote por $B_r(\theta)$ la bola cerrada de radio r centrada en θ y ponga

$$S_N(\theta) = \{\alpha \in B(m) | \alpha(n) = \theta(n) \quad \text{for } |n| \leq N\}$$

- (a) Probar que $S_N(\theta) \subset B_r(\theta)$ si

$$\frac{1}{k^N} \cdot \frac{2m}{k-1} \leq r$$

- (b) Probar que $S_N(\theta) \supset B_r(\theta)$ si

$$N \leq \frac{\log(1/r)}{\log k}$$

- (c) Probar, usando (a) y (b), que si μ es la medida producto asociada con la probabilidad uniforme μ_0 sobre $\{1, \dots, m\}$ dada por $\mu_0(\{i\}) = 1/m$, entonces existe $C > 0$ tal que

$$C^{-1}r^\delta \leq \mu(B_r(\theta)) \leq Cr^\delta \quad \text{con } \delta = \frac{2 \log m}{\log k}.$$

Deducir que $2 \log m / \log k$ es igual a la dimensión de Hausdorff de $B(m)$.

EJERCICIO I.6. Considere la transformación $T : [0, 1] \rightarrow [0, 1]$ definida por $T(x) = x/2$ para $0 < x \leq 1$ y $T(0) = 1$. Probar que T es medible y no es T -invariante para ninguna probabilidad sobre el σ -álgebra de Borel de $[0, 1]$.

²Esta denominación es coherente con lo explicado al final de la Sección II.2.

I.2. Teorema de Recurrencia de Poincaré

Trataremos primero la versión probabilística del Teorema de Recurrencia de Poincaré, que no hace referencia alguna a conceptos topológicos.

Teorema I.2.1. *Sea T una transformación que preserva medida en el espacio de probabilidad (X, \mathcal{O}, μ) . Dada $A \in \mathcal{O}$, sea A_0 el conjunto de puntos $x \in A$ tal que $T^n(x) \in A$ para una cantidad infinita de $n \geq 0$. Entonces A_0 pertenece a \mathcal{O} , y $\mu(A_0) = \mu(A)$.*

Demostración. Sea $C_n := \{x \in A \mid T^j(x) \notin A \text{ para todo } j \geq n\}$. Es claro que

$$A_0 = A \setminus \bigcup_{n=1}^{\infty} C_n$$

Entonces el teorema quedará probado si mostramos que $C_n \in \mathcal{O}$ y $\mu(C_n) = 0$ para cada $n \geq 1$. Observe que

$$C_n = A \setminus \bigcup_{j \geq n} T^{-j}(A).$$

lo cual muestra que $C_n \in \mathcal{O}$, e implica que

$$C_n \subset \bigcup_{j \geq 0} T^{-j}(A) \setminus \bigcup_{j \geq n} T^{-j}(A)$$

Pero desde que

$$\bigcup_{j \geq n} T^{-j}(A) = T^{-n}(\bigcup_{j \geq 0} T^{-j}(A)) \implies \mu(\bigcup_{j \geq n} T^{-j}(A)) = \mu(\bigcup_{j \geq 0} T^{-j}(A)).$$

Esto implica $\mu(C_n) = 0$. □

Para enunciar la versión topológica del teorema de recurrencia, necesitamos la noción de conjunto ω -límite de un punto. Sea X un espacio topológico y $T : X \rightarrow X$ una transformación. Definimos al *conjunto ω -límite de un punto $x \in X$* ($\omega(x)$) como el conjunto de puntos $y \in X$ tales que para cualquier entorno U de y , sucede que $T^n(x) \in U$ para infinitos valores de n . Si X es un espacio métrico, esto es equivalente a

$$\liminf_{n \rightarrow \infty} \text{dist}(T^n(x), y) = 0$$

Teorema I.2.2. *Sea X un espacio métrico separable³ y $T : X \rightarrow X$ una transformación Borel-medible. Sea μ una probabilidad T -invariante sobre el σ -álgebra de Borel de X . Entonces $\mu(\{x : x \notin \omega(x)\}) = 0$. En otras palabras, casi todo punto es recurrente.*

Demostración. Sea $\{U_n\}_{n=0}^{\infty}$ una base de conjuntos abiertos tales que

$$\lim_{n \rightarrow \infty} \text{diam } U_n = 0 \quad \text{y} \quad \bigcup_{n \geq m} U_n = X$$

para todo $m \geq 0$. Sea $\tilde{U}_n := \{x \in U_n \mid T^j(x) \in U_n \text{ para infinitos valores positivos de } j\}$. Por el teorema precedente

$$\mu(U_n \setminus \tilde{U}_n) = 0. \quad \text{Ponga } \tilde{X} := \bigcap_{m=0}^{\infty} \bigcup_{n \geq m} \tilde{U}_n$$

Se deduce que

$$\begin{aligned} \mu(X \setminus \tilde{X}) &= \mu(\bigcup_{m=0}^{\infty} (X \setminus \bigcup_{n \geq m} \tilde{U}_n)) \\ &= \mu(\bigcup_{m=0}^{\infty} (\bigcup_{n \geq m} U_n \setminus \bigcup_{n \geq m} \tilde{U}_n)) \\ &\leq \mu(\bigcup_{m=0}^{\infty} \bigcup_{n \geq m} (U_n \setminus \tilde{U}_n)) = 0 \end{aligned}$$

Por tanto sólo tenemos que probar que $x \in \tilde{X}$ implica $x \in \omega(x)$. Sea $r > 0$. Elija m tal que $\text{diam } U_n \leq r/3$ si $n \geq m$. Puesto que $x \in \tilde{X}$, se tiene que $x \in \bigcup_{n \geq m} \tilde{U}_n$. Por tanto existe $n \geq m$ tal que $x \in \tilde{U}_n$. Dado que $\text{diam } U_n \leq r/3$, se sigue que $U_n \subset B_r(x)$, lo cual implica que $T^j(x) \in B_r(x)$ si $T^j(x) \in U_n$. Pero desde que $x \in \tilde{U}_n$, $T^j(x) \in U_n$ para una cantidad infinita de j , lo cual muestra que $x \in \omega(x)$. □

³Un espacio topológico es *separable* si tiene un conjunto denso numerable

EJERCICIO I.7. Sea T una transformación que preserva medida en el espacio de probabilidad (X, \mathcal{O}, μ) . Dado $A \in \mathcal{O}$, defina \tilde{A} como el conjunto de puntos $x \in A$ tales que $T^n(x) \in A$ para una cantidad infinita de $n > 0$. Defina $N : \tilde{A} \rightarrow \mathbb{Z}$ por $N(x) = m$ si $m > 0$, $T^m(x) \in A$ y $T^n(x) \notin A$ para $0 < n < m$. La función $N(x)$ es llamada *tiempo de primer retorno*.

- (a) Probar que $\tilde{A} \in \mathcal{O}$, $\mu(A \setminus \tilde{A}) = 0$, y $N : \tilde{A} \rightarrow \mathbb{Z}$ es una función medible;
- (b) Defina $\tilde{T} : \tilde{A} \rightarrow X$ por $\tilde{T}(x) = T^{N(x)}(x)$. Probar que $\tilde{T}(\tilde{A}) \subset \tilde{A}$ y que si T es invertible (existe una transformación S que preserva medida en (X, \mathcal{O}, μ) tal que $TS(x) = ST(x) = x$ para μ -casi todo x), entonces $\tilde{T} : \tilde{A} \rightarrow \tilde{A}$ deja invariante la medida $\mu|_{\tilde{A}}$. Sugerencia: Ponga $\tilde{A}_n = N^{-1}(\{n\})$; probar que $\mu(\tilde{T}(C)) = \mu(C)$ para todo $C \in \mathcal{O}$ tal que $C \subset \tilde{A}_n$.

I.3. El Problema de la Equivalencia

La noción natural de equivalencia entre dos transformaciones que preservan medida está dada por la siguiente definición.

Definición 1. Decimos que dos transformaciones que preservan medida T_i , $i = 1, 2$, de dos espacios de medida $(X_i, \mathcal{O}_i, \mu_i)$, $i = 1, 2$, respectivamente, son equivalentes si existe una transformación que preserva medida F que lleva $(X_1, \mathcal{O}_1(\text{mod } 0), \mu_1)$ en $(X_2, \mathcal{O}_2(\text{mod } 0), \mu_2)$ satisfaciendo

- (a) F es invertible, o sea que existe una transformación medible $G : X_2 \rightarrow X_1$ tal que $GF(x) = x$ para c.t. $x \in X_1$ y $FG(y) = y$ para c.t. $y \in X_2$.
- (b) F preserva medida, o sea $\mu_1(F^{-1}(A)) = \mu_2(A) \pmod{0}$ para todo $A \subset \mathcal{O}_2$.
- (c) $T_2F = FT_1$ para c.t. $x \in X_1$.

Observe que (a)-(b) implican que G preserva medida de $(X_2, \mathcal{O}_2, \mu_2)$ en $(X_1, \mathcal{O}_1, \mu_1)$ y, por (c), $GT_2 = T_1G$ casi todo punto.

Por tanto, la equivalencia es simétrica. Claramente es transitiva y reflexiva, por lo que es una verdadera relación de equivalencia.

Uno de los objetivos de la teoría ergódica es clasificar a las transformaciones que preservan medida, módulo esta relación de equivalencia. Uno de los métodos para este análisis consiste en asociar con una transformación que preserva medida $T : (X, \mathcal{O}, \mu) \rightarrow (Y, \mathcal{S}, \nu)$ un operador lineal $U_T : \mathcal{L}^2(Y) \rightarrow \mathcal{L}^2(X)$ definido por

$$U_T f = f \circ T.$$

Como T preserva medida entonces U_T es un operador unitario: $\langle U_T f, U_T g \rangle = \langle f, g \rangle$ para cada $f, g \in \mathcal{L}^2(Y)$ ($\langle \cdot, \cdot \rangle$ es el producto interno en \mathcal{L}^2).

Definición 2. Sean $(X_i, \mathcal{O}_i, \mu_i)$, $i = 1, 2$, espacios de medida $T_i : X_i \rightarrow X_i$ transformaciones que preservan medida, con operadores lineales asociados U_{T_i} . Decimos que T_1 y T_2 son espectralmente equivalentes si existe un isometría invertible

$$L : \mathcal{L}^2(X_2) \rightarrow \mathcal{L}^2(X_1)$$

tal que $LU_{T_2} = U_{T_1}L$.

Si T_1 and T_2 son equivalentes, son espectralmente equivalentes puesto que la transformación $F : X_1 \rightarrow X_2$ dada en la definición de equivalencia da lugar a una isometría $U_F : \mathcal{L}^2(X_2) \rightarrow \mathcal{L}^2(X_1)$ que satisface la condición $U_F U_{T_2} = U_{T_1} U_F$ (alcanza con tomar $U_F f = f \circ F$).

En general, sin embargo, las transformaciones espectralmente equivalentes no son necesariamente equivalentes; por ejemplo todos los shifts de Bernoulli son espectralmente equivalentes, pero Kolmogorov probó en 1958 que no todos son equivalentes. Hizo esto asociando a cada transformación que preserva medida $T : X \rightarrow X$ un número real $h(T) \in [0, +\infty]$, llamado la entropía de T que es invariante bajo equivalencias (significando que todas las transformaciones equivalentes tienen la misma entropía). Como se verá en la Sección siguiente, la entropía de un shift de Bernoulli $B(p_1, \dots, p_n)$ es igual a $-\sum_{i=1}^n p_i \log p_i$. Entonces, en particular, los shifts $B(1/2, 1/2)$, $B(1/3, 1/3, 1/3)$, y $B(1/4, 1/4, 1/2)$ tienen diferentes entropías y no pueden ser equivalentes.

En el caso de los shifts de Bernoulli, el recíproco también se cumple: dos shifts de Bernoulli con la misma entropía son equivalentes. Este importante resultado fue probado por Ornstein en 1970. Entonces tenemos el siguiente

Teorema I.3.1. *Dos shifts de Bernoulli son equivalentes si y sólo si tienen la misma entropía.*

EJERCICIO I.8. Probar que si los shifts de Bernoulli $B_\nu^+(p_1, \dots, p_m)$ y $B_\mu^+(q_1, \dots, q_\ell)$ son equivalentes, entonces $m = \ell$ y $\{p_1, \dots, p_m\}$ pueden ser indexados de tal manera que $p_i = q_i$ para todo $1 \leq i \leq m = \ell$. Sugerencia: Ponga $A_i = \{\theta \in B^+(p_1, \dots, p_m) \mid \theta(0) = i\}$ y $C_j = \{\theta \in B^+(q_1, \dots, q_\ell) \mid \theta(0) = j\}$. (a) Pruebe que el conjunto A_i tiene las siguientes propiedades: $\sigma|_{A_i}$ es una biyección entre A_i y $B^+(p_1, \dots, p_m)$, y $\nu(\sigma(S)) = p_i^{-1}\nu(S)$ para todos los subconjuntos de Borel $S \subset A_i$. (b) Luego, si $T : B_\nu^+(p_1, \dots, p_m) \rightarrow B_\mu^+(q_1, \dots, q_\ell)$ realiza la equivalencia entre los dos shifts, muestre que $T(A_i)$ contiene un subconjunto de Borel D tal que $\sigma|_D$ es una biyección entre D y un conjunto de medida total de $B^+(q_1, \dots, q_\ell)$, y $\mu(\sigma(S)) = p_i^{-1}\mu(S)$ para todo $S \subset D$. (c) Deduzca que existe j tal que $q_j = p_i$ y $T(A_i) = C_j \pmod{0}$. Asuma primero que los q_j son todos distintos.

EJERCICIO I.9. Sea $f : [0, 1] \rightarrow [0, 1]$ la transformación dada en el Ejercicio I.3. Probar que f es equivalente a $\sigma : B^+(p_1, p_2, \dots) \rightarrow B^+(p_1, p_2, \dots)$ con $p_i = b_i - a_i$. Sugerencia: Defina la función $T : [0, 1] \rightarrow B^+(p_1, p_2, \dots)$ estableciendo $(Tx)(n) = j$ si $f^{n-1}(x) \in (a_j, b_j)$.

I.4. Entropía

La entropía, como se vió en la Sección anterior, es una importante característica numérica de una transformación que preserva medida. La noción de entropía tiene una larga historia; proviene de ramas de la física y otras ciencias. Originalmente, R. Klausius introdujo la entropía en 1864 para describir la transformación de energía calórica en energía cinética. Actualmente la entropía en física es considerada como una medida de la complejidad de sistemas con gran número de componentes y configuraciones posibles.

En 1948, C. Shannon introdujo la entropía en la teoría de la información como una característica cuantitativa de la incertidumbre (o información) de sucesos aleatorios. Sea (X, \mathcal{A}, μ) un espacio de probabilidad. Queremos medir la cantidad de información agregada por el conocimiento de que un suceso $A \in \mathcal{A}$ haya ocurrido. Claramente tal medida debería ser una función, i , de la probabilidad $\mu(A)$, y

- ser no negativa ($i : [0, 1] \rightarrow [0, +\infty)$),
- ser cero si el suceso tiene probabilidad uno ($i(1) = 0$),
- crecer cuando $\mu(A)$ decrece, y
- satisfacer la siguiente relación de independencia: si $A, B \in \mathcal{A}$ son sucesos independientes ($\mu(A \cap B) = \mu(A)\mu(B)$), la información agregada por saber que ambos A y B ocurrieron ($A \cap B$ ocurre) es igual a la suma de la información dada por el conocimiento de que A y B han ocurrido.

Se puede probar que hay una única función que satisface todas esas condiciones: $i(p) = -k \log p$, donde $k > 0$ es una constante; ver el Ejercicio I.10. Si $k = 1/\log 2$, la unidad de información es llamada “bit” (se usa en teoría de la información).

Sean ahora A_1, \dots, A_n sucesos aleatorios de los cuales uno y sólo uno puede realmente ocurrir (son, por lo tanto mutuamente excluyentes). Sean $p_i = \mu(A_i)$, $1 \leq i \leq n$, sus probabilidades y, obviamente, $p_1 + \dots + p_n = 1$. Entonces, la cantidad de información agregada por conocer cual de esos sucesos realmente ocurre es la media (valor esperado) de la información definida por:

$$h(p_1, \dots, p_n) = -p_1 \log p_1 - \dots - p_n \log p_n \quad (\text{I.4.1})$$

Esta cantidad es llamada la *entropía* de la distribución de probabilidad $\{p_1, \dots, p_n\}$. Observamos que $h \geq 0$. Si se fija n , entonces h toma su máximo $h = \log n$ con la distribución uniforme $p_1 = \dots = p_n = 1/n$; ver Ejercicio I.11. En este caso el número de sucesos, n , puede ser expresado por $n = e^h$.

¿Cuál es el significado de h en el caso general, cuando n es muy grande y la distribución es muy ‘no uniforme, por lo que $h \ll \log n$? La respuesta es dada por el clásico teorema de Shannon-McMillan-Breiman de teoría de la información, que describiremos sin mucho rigor. Dice que podemos dividir el conjunto de los sucesos $\{A_1, \dots, A_n\}$ en dos grupos: sucesos esenciales, cuyas probabilidades son suficientemente grandes, y sucesos despreciables (los otros), de modo que (i) los sucesos despreciables pueden ser eliminados de la expresión (I.4.1) sin pérdida significativa del valor de h , y (ii) el número de sucesos esenciales es aproximadamente

$$n_{\text{ess}} \approx e^h \quad (\text{I.4.2})$$

Este es el significado básico de la entropía en teoría de la información: e^h da, ‘esencialmente’, el número de sucesos posibles.

En 1958, basado en las ideas de Shannon, A.Ñ. Kolmogorov introdujo la entropía en la teoría de sistemas dinámicos. Su versión de la entropía, desarrollada junto con Ya. G. Sinai, es ahora llamada *entropía de teoría de la medida* (*measure-theoretic entropy*) o *entropía de Kolmogorov-Sinai*. Las pruebas de los resultados que siguen se encuentran en, por ejemplo, [CFS82], Ch. 10; [Mn83], Cap 4.

Sea $T : X \rightarrow X$ una transformación que preserva la probabilidad μ . Vamos a medir la complejidad de la transformación T^N cuando $N \rightarrow \infty$. ¿Qué es esta complejidad⁴? Sea $X = A_1 \cup \dots \cup A_n$ una partición de X en subconjuntos medibles disjuntos, $A_i \cap A_j = \emptyset$, $i \neq j$. Si etiquetamos cada A_i por i , entonces el punto x del espacio de fase puede ser codificado por la etiqueta i si $x \in A_i$. La complejidad de la partición $\xi = \{A_1, \dots, A_n\}$ es medida por su entropía definida por la fórmula de Shannon:

$$h(\xi) = -\mu(A_1) \log \mu(A_1) - \dots - \mu(A_n) \log \mu(A_n)$$

Ahora la transformación T entra en la contrucción. Dado un punto $x \in A_{i_0}$ (con la etiqueta i_0), el punto $T^k x$ para un $k \geq 1$ puede ser etiquetado por i_k si $T^k x \in A_{i_k}$. Entonces, la órbita $\{x, Tx, \dots, T^N x\}$ puede ser etiquetada por una cuerda $\{i_0, i_1, \dots, i_N\}$. Uno puede fácilmente ver que esta cuerda etiquetará de hecho a todos los puntos en el conjunto $A_{i_0} \cap T^{-1}A_{i_1} \cap \dots \cap T^{-N}A_{i_N}$. Todos estos conjuntos (con un N fijo) crean una partición de X que denominaremos ξ_N . Su complejidad es una vez más medida por la entropía $h(\xi_N)$ definida por la misma fórmula de Shannon.

Para entender como luce la partición ξ_N , para cada $j \geq 1$ considere una partición de X dada por $T^{-j}\xi = \{T^{-j}A_1, \dots, T^{-j}A_n\}$. Debido a la invariancia de μ por T tenemos $h(T^{-j}\xi) = h(\xi)$ para todo $j \geq 1$. Ahora ξ_N es obtenida tomando todas las posibles intersecciones de elementos de $T^{-j}\xi$, $0 \leq j \leq N$. Esta construcción es llamada *producto de particiones* y denotada por $\bigvee_{j=0}^N T^{-j}\xi$. Observe que si $N \geq M$, cada conjunto en ξ_N está contenido en ξ_M y cada elemento en ξ_M es la union de elementos de ξ_N ; por lo que ξ_N es un *refinamiento* de ξ_M .

La partición ξ_N consiste de n^{N+1} conjuntos, llamados *elementos* o *átomos* (algunos de ellos pueden ser vacíos, en este caso ponemos 0 en lugar de $0 \log 0$). Por tanto, su máxima entropía no puede exceder $h(\xi_N) \leq \log n^{N+1} = (N+1) \log n$, o sea que $h(\xi_N)$ crece a lo sumo linealmente en el tiempo N . De hecho, esta estimación es más bien optimista, puesto que para muchas transformaciones la mayoría de los elementos de ξ_N serán vacíos, y $h(\xi_N)$ crecerá lentamente, si es que lo hace; ver Ejercicio I.12. Pero para muchas transformaciones hay crecimiento lineal de $h(\xi_N)$; ver Ejercicio I.13.

La cantidad

$$h(T, \xi) = \lim_{N \rightarrow \infty} \frac{1}{N+1} h(\xi_N)$$

es llamada la *entropía de la transformación T con respecto a la partición ξ* . Este límite siempre existe y es igual a $\inf_{N \geq 1} (N+1)^{-1} h(\xi_N)$, porque la sucesión $(N+1)^{-1} h(\xi_N)$ decrece monótonamente con N (omitimos la demostración de esto que es la parte más complicada de la definición de entropía). En particular, para cada N tenemos $h(T, \xi) \leq (N+1)^{-1} h(\xi_N) \leq h(\xi)$.

El valor de $h(T, \xi)$ admite la siguiente interpretación en términos de teoría de la información. Queremos codificar la información acerca del estado del sistema (X, T, μ) en cada iteración futura j usando la partición ξ . El itinerario $\{i_0, i_1, \dots, i_N\}$ definido arriba para un punto dado, x , codifica un átomo de la partición ξ_N que contiene a x . Como $h(\xi_N)$ es la cantidad (esperada) de información que se tiene por conocer el itinerario del punto x durante las primeras $N+1$ iteraciones, el valor $h(\xi_N)/(N+1)$ es la velocidad media de transmisión de datos que se requiere para codificar la evolución del sistema. Por tanto, $h(T, \xi)$ es la velocidad asintótica de transmisión de datos cuando $N \rightarrow \infty$. Más específicamente, la cantidad de bits transmitidos por unidad de tiempo es $h(T, \xi)/\log 2$. Si esta cantidad es positiva, el sistema es suficientemente complejo de modo que la descripción de su evolución durante el tiempo N requiere una cantidad de memoria de computadora que debe crecer linealmente con N . Ninguna cantidad finita de información sobre el sistema puede describir su evolución en todo el futuro. Prácticamente, esto significa que el futuro no puede ser predecido, o sea que el sistema es *determinístico* pero *impredicible*.

⁴Hay otras medidas de la complejidad, algunas introducidas por el mismo Kolmogorov

Como se sigue de las estimaciones anteriores, $h(T, \xi) \leq h(\xi)$, o sea que $h(T, \xi)$ es siempre finita, pero puede no ser acotada como función de ξ . La cantidad

$$h_\mu(T) = \sup\{h(T, \xi) : \xi \text{ cualquier partición medible de } X\}$$

es llamada la *entropía de teoría de la medida (measure-theoretic entropy)*. Esta es la *entropía de Kolmogorov-Sinai*.

Enfatizamos el significado de la entropía. Por la fórmula (I.4.2), la entropía $h(T, \xi)$ mide la tasa exponencial de crecimiento del número de elementos esenciales en la partición ξ_N cuando $N \rightarrow \infty$ (esos elementos que hacen una contribución esencial a la cantidad $h(\xi_N)$). Uno puede pensar entonces a $h(T)$ como la tasa de crecimiento exponencial de la complejidad de T^N cuando $N \rightarrow \infty$.

Existe un teorema de Kolmogorov y Sinai, muy útil prácticamente, que simplifica los cálculos de la entropía. Necesitamos algunas definiciones. Si $\{\mathcal{P}_n\}_{n \geq 1}$ es una sucesión de particiones, denotamos por $\bigvee_{n \geq 1} \mathcal{P}_n$ al menor σ -álgebra que contiene a todos los átomos de todas esas particiones. En otras palabras, es el σ -álgebra generado por $\bigcup_{n \geq 1} \mathcal{P}_n$. Si T es una transformación medible en el espacio de medida (X, \mathcal{A}, μ) , decimos que \mathcal{P} es una *partición generadora* si satisface una de las siguientes condiciones:

- i) $\bigvee_{j=0}^{\infty} T^{-j} \mathcal{P} = \mathcal{A} \pmod{0}$, o
- ii) $\bigvee_{j=0}^{\infty} T^{-j} \mathcal{P} \neq \mathcal{A}$, pero T es invertible con inversa medible y $\bigvee_{j=-\infty}^{\infty} T^{-j} \mathcal{P} = \mathcal{A}$

Entonces el Teorema de Kolmogorov-Sinai establece que si \mathcal{P} es una partición generadora con $h(\mathcal{P}) < \infty$, entonces

$$h_\mu(T) = h(T, \mathcal{P}).$$

EJERCICIO I.10. Probar que la única función $i : [0, 1] \rightarrow [0, +\infty)$ que satisface (a)-(d) al comienzo de esta Sección es $i(p) = -k \log p$, con una constante $k > 0$. Observar que la condición (d) significa $i(pq) = i(p) + i(q)$ para todo $0 \leq p, q \leq 1$. Sugerencia: Cambie la variable $x = -\log p$, entonces la función $f(x) = i(e^{-x})$ es creciente y satisface $f(0) = 0$ y $f(x+y) = f(x) + f(y)$ para todo $x, y > 0$. Denote $f(1) = a$ y muestre que $f(m/n) = am/n$ para todo $m, n \in \mathbb{N}$. Entonces use la monotonía de f para probar que $f(x) = ax$ para todo $x > 0$.

EJERCICIO I.11. Mostrar que $h(p_1, \dots, p_n) \leq \log n$ y la igualdad se satisface si y sólo si $p_1 = \dots = p_n = 1/n$. Sugerencia: observe que la función $g(x) = -x \log x - (a-x) \log(a-x)$, con $a > 0$ dado, toma su máximo en $x = a/2$.

EJERCICIO I.12. Sea $T : S^1 \rightarrow S^1$ una rotación del círculo S^1 con ángulo fijo $\theta > 0$. Mostrar que $h(T, \xi) = 0$ para toda partición finita de S^1 en arcos. Sugerencia: Sea n el número de arcos. Verifique que ξ_N es una partición de S^1 en no más de $n(N+1)$ arcos.

EJERCICIO I.13. Considere el shift de Bernoulli $B_\mu(p_1, \dots, p_m)$ y una partición ξ en conjuntos $A_i = \{\theta \in B(m) : \theta(0) = i\}$, para $1 \leq i \leq m$. Mostrar que

$$h(\sigma, \xi) = h(p_1, \dots, p_m) = -p_1 \log p_1 - \dots - p_m \log p_m$$

Sugerencia: Verique por inducción que $h(\xi_N) = Nh(p_1, \dots, p_m)$ para todo $N \geq 1$.

Capítulo II

Ergodicidad

II.1. Teorema Ergódico de Birkhoff-Khinchin

El Teorema de Birkhoff-Khinchin se refiere a la distribución de las órbitas de una transformación T que preserva medida en un espacio de probabilidad (X, \mathcal{O}, μ) . Para estudiar cómo una órbita $\{x, T(x), T^2(x), \dots\}$ se distribuye asintóticamente en X introducimos el *tiempo de estadía* de x en un conjunto $A \in \mathcal{O}$ por

$$\tau(x, A) = \lim_{n \rightarrow \infty} \frac{1}{n} \#\{0 \leq m < n \mid T^m(x) \in A\}.$$

Obsérvese que el tiempo de estadía es el límite del promedio de las estadías en A de $T^m(x)$ hasta el iterado $(n-1)$ -ésimo. El Teorema de Birkhoff establece que ese límite existe para casi todo x , que $\tau(x, A)$ es una función integrable de x cuya integral es $\int_X \tau(x, A) d\mu(x) = \mu(A)$. Más aún, puede verificarse fácilmente que, como función de x , τ es T -invariante, o sea $\tau(x, A) = \tau(T(x), A)$ c.t. $x \in X$. Esto motiva la siguiente definición: decimos que T es ergódica si toda función T -invariante es constante c.t.p. Entonces τ debe ser constante y su integral es igual a su valor c.t.p., y para c.t. x tenemos: $\tau(x, A) = \mu(A)$. Esta es una conclusión muy importante y coloca el problema de desarrollar métodos para decidir cuando una transformación T es ergódica. La Ergodicidad es una propiedad muy fuerte y muchas transformaciones importantes, como las que provienen de la Mecánica Hamiltoniana, frecuentemente no son ergódicas. Por otro lado, como veremos en este Capítulo, hay muchas clases de transformaciones ergódicas. Más aún, cuando X es un espacio métrico compacto y $T: X \rightarrow X$ es continua, siempre existe una probabilidad T -invariante sobre el σ -álgebra de Borel de X respecto del cual T es ergódica. Estas probabilidades respecto de las que T es ergódica son muy importantes en el análisis de la dinámica de T .

Un ejemplo interesante de transformación ergódica es $T: [0, 1] \rightarrow [0, 1]$ dada por $T(x) = 10x - [10x]$ (a veces representada por $\{10x\}$, donde $\{\cdot\}$ indica la parte fraccionaria de un número). Esta transformación preserva la probabilidad de Lebesgue λ en $[0, 1]$ y es ergódica (por una prueba vea Ch. III.1. en [Mn83]). Una consecuencia directa de su ergodicidad es el siguiente importante resultado de teoría de números. Escriba $x \in [0, 1]$ en representación decimal $x = 0.a_0a_1a_2\dots$ y sea $N_n(x, j)$ el número de veces que el dígito $0 \leq j \leq 9$ aparece en la cuerda $[a_0 \dots a_{n-1}]$. Entonces para c.t. $x \in [0, 1]$

$$\lim_{n \rightarrow \infty} \frac{1}{n} N_n(x, j) = \frac{1}{10}$$

Esto es una consecuencia de la ergodicidad de T porque $a_m = j$ si y sólo si $T^m(x) \in [j/10, (j+1)/10)$. Entonces, para c.t. x :

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} N_n(x, j) &= \tau(x, [j/10, (j+1)/10)) \\ &= \lambda([j/10, (j+1)/10)) = 1/10. \end{aligned}$$

Más sutil es una propiedad similar de fracciones continuas. Cada número irracional $x \in (0, 1)$ puede ser reescrito de manera única como una fracción continua

$$x = \frac{1}{a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\ddots}}}}$$

donde a_0, a_1, \dots son enteros positivos. Sea $P_n(x, k)$ el número de veces que k aparece entre los números a_0, \dots, a_{n-1} . Entonces, para c.t. $x \in (0, 1)$

$$\lim_{n \rightarrow \infty} \frac{1}{n} P_n(x, k) = \frac{1}{\log 2} \log \left(1 + \frac{1}{k(k+2)} \right). \quad (\text{II.1.1})$$

La prueba de esta propiedad requiere primero transformar $P_n(x, k)$ en un tiempo de estadía. Esto se hace con la ayuda de la *transformación de Gauss* $T : [0, 1] \rightarrow [0, 1]$ y la medida μ en $[0, 1]$, estudiadas por Gauss en 1799 y definidas por

$$T(x) = \begin{cases} \frac{1}{x} - \left[\frac{1}{x} \right] & \text{si } x \neq 0 \\ 0 & \text{si } x = 0 \end{cases}, \quad \mu(A) = \frac{1}{\log 2} \int_A \frac{dx}{1+x}$$

para todo conjunto de Borel A . La probabilidad μ es T -invariante. Más aún, se puede probar que T es ergódica con respecto a la probabilidad μ . De hecho ella pertenece a un clase bien estudiada de transformaciones del intervalo llamadas *transformaciones de Markov*. [Mn83], III.1. Ver Ejercicio II.1.

Dada una transformación T que preserva medida en el espacio de probabilidad (X, \mathcal{O}, μ) decimos que una función $f : X \rightarrow \mathbb{R}$ es *invariante* (o T -invariante) si $f(T(x)) = f(x)$ para c.t. $x \in X$.

Teorema II.1.1 (Birkhoff-Khinchin). *Sea (X, \mathcal{O}, μ) un espacio de probabilidad y $T : X \rightarrow X$ una transformación que preserva medida. Si $f : X \rightarrow \mathbb{R}$ es una función integrable, el límite*

$$\tilde{f}(x) := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) \quad (\text{II.1.2})$$

existe para casi todo $x \in X$, y la función \tilde{f} es T -invariante, integrable, $\|\tilde{f}\|_1 \leq \|f\|_1$,

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

Además, si $f \in \mathcal{L}^p(X, \mathcal{O}, \mu)$ para algún $1 \leq p < \infty$, entonces $\tilde{f} \in \mathcal{L}^p(X, \mathcal{O}, \mu)$, y

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

Pruebas de este importante teorema se pueden encontrar en [CFS82], [Mn83]. La función \tilde{f} es llamada la *media temporal* de f . Los tiempos de estadía son medias temporales de funciones características porque

$$\#\{0 \leq j < n : T^j(x) \in A\} = \sum_{j=0}^{n-1} \chi_A(T^j(x)),$$

por lo que

$$\tau(x, A) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_A(T^j(x)) = \tilde{\chi}_A(x)$$

Entonces

$$\int \tau(x, A) d\mu(x) = \int \tilde{\chi}_A d\mu = \int \chi_A d\mu = \mu(A).$$

Corolario II.1.2. Si T es invertible y $f \in \mathcal{L}^1(X, \mathcal{O}, \mu)$, entonces $\tilde{f}(x) = \tilde{f}^-(x)$ para c.t.x $x \in X$, donde

$$\tilde{f}^-(x) := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^{-j}(x)).$$

$$\text{Entonces } \tilde{f}(x) = \lim_{n \rightarrow \infty} \frac{1}{2n+1} \sum_{j=-n}^n f(T^j(x)).$$

Demostración. Si $f_n \rightarrow f$ en \mathcal{L}^1 , la desigualdad

$$\int |f_n - f| d\mu = \int |f_n - f|^\sim d\mu \geq \int |\tilde{f}_n - \tilde{f}| d\mu$$

muestra que $\tilde{f}_n \rightarrow \tilde{f}$ in \mathcal{L}^1 , y lo mismo es cierto para \tilde{f}_n^- y \tilde{f}^- . Sobre un espacio de probabilidad X , las funciones de cuadrado integrable $f \in \mathcal{L}^2(X)$ son densas en $\mathcal{L}^1(X)$ (en realidad, cualquier función en \mathcal{L}^1 puede ser aproximada por funciones simples). Por lo tanto, es suficiente probar el Corolario para $f \in \mathcal{L}^2$.

Sea $F \subset \mathcal{L}^2$ el subespacio de las funciones T -invariantes. Es un subespacio cerrado \mathcal{L}^2 . Si $\pi : \mathcal{L}^2 \rightarrow F$ es la proyección ortogonal, la igualdad $\tilde{f} = \tilde{f}^-$ seguirá de $\pi f = \tilde{f}$ para todo $f \in \mathcal{L}^2$, puesto que F es también el espacio de las funciones T^{-1} -invariantes (tales que $g \circ T^{-1} = g$ c.t.p.). Pero, para cualquier $g \in F$

$$\int (f - \tilde{f})g d\mu = \int (fg - \tilde{f}g)^\sim d\mu = \int (fg)^\sim d\mu - \int (\tilde{f}g)^\sim d\mu = 0$$

porque $(fg)^\sim = \tilde{f}g = (\tilde{f}g)^\sim$, como consecuencia de que \tilde{f} y g son T -invariantes. \square

OBSERVACIÓN II.1.3. Sea T^t un grupo continuo de automorfismos sobre un espacio de probabilidad (X, \mathcal{O}, μ) . Esto significa que $t \in \mathbb{R}$, T^t es invertible, $T^0 = \text{Id}$, $T^{t+s}(x) = T^t(T^s(x))$, y $\mu(T^t(A)) = \mu(A)$ para todo $t, s \in \mathbb{R}$, $x \in X$, $A \in \mathcal{O}$. Decimos que T^t es un *flujo (medible)* si para toda función medible $g : X \rightarrow \mathbb{R}$, $g \circ T^t$ es medible en $X \times \mathbb{R}$.

La correspondiente versión del Teorema de Birkhoff-Khinchin para flujos expresa que para cada $f \in \mathcal{L}^1(X, \mathcal{O}, \mu)$

$$\begin{aligned} \tilde{f}(x) &= \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t f(T^s(x)) ds = \\ &= \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t f(T^{-s}(x)) ds = \lim_{t \rightarrow \infty} \frac{1}{2t} \int_{-t}^t f(T^s(x)) ds \end{aligned}$$

donde los límites existen para casi todo $x \in X$, y $\tilde{f} \in \mathcal{L}^1(X, \mathcal{O}, \mu)$ con $\int \tilde{f} d\mu = \int f d\mu$.

Corolario II.1.4. Para cada $A, B \in \mathcal{O}$ existe el límite

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j}(A) \cap B).$$

Demostración. Tenemos

$$\mu(T^{-n}(A) \cap B) = \int_X \chi_{T^{-n}(A)} \chi_B d\mu = \int_X (\chi_A \circ T^n) \chi_B d\mu$$

Como $\chi_A \in \mathcal{L}^2(X)$, podemos aplicar el Teorema de Birkhoff-Khinchin para concluir que la sucesión

$$\frac{1}{n} \sum_{j=0}^{n-1} \chi_A \circ T^j$$

converge en \mathcal{L}^2 . Entonces

$$\frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j}(A) \cap B) = \int_X \left(\frac{1}{n} \sum_{j=0}^{n-1} \chi_A \circ T^j \right) \chi_B d\mu$$

también converge. □

EJERCICIO II.1. En relación con la fórmula II.1.1:

- i) Mostrar que la probabilidad de Lebesgue **no** es T -invariante.
- ii) Probar que la medida de Gauss es invariante. Sugerencia: $\mu(A) = \mu(T^{-1}(A))$ cuando A es un intervalo; luego use el hecho de que la uniones finitas de intervalos forman un subálgebra que genera el σ -álgebra de Borel.
- iii) Probar que

$$a_n = k \iff T^n(x) \in \left(\frac{1}{k+1}, \frac{1}{k} \right].$$

Sugerencia: Observe que

$$T(x) = \frac{1}{x} - \left[\frac{1}{x} \right] = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\ddots}}} \quad \text{y} \quad T^n(x) = \frac{1}{a_n + \frac{1}{a_{n+1} + \frac{1}{\ddots}}}$$

por lo que $a_n = [1/T^n(x)]$.

- iv) Admitiendo que T es ergódica respecto a la probabilidad μ y observando que $\lim_{n \rightarrow \infty} \frac{1}{n} P_n(x, k) = \tau(x, (1/(k+1), 1/k])$, deducir (II.1.1). Desde que la medida μ es equivalente a la de Lebesgue, concluimos que la propiedad anterior se verifica para c.t. x con respecto a la medida de Lebesgue.

EJERCICIO II.2. Admita que el límite II.1.2 existe.

- i) De

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

- ii) Probar que \tilde{f} es T -invariante.

EJERCICIO II.3. Sea (X, \mathcal{O}, μ) un espacio de probabilidad, $T : X \rightarrow X$ una transformación que preserve medida y $\{F_n\}_{n \geq 0}$ una sucesión dominada de funciones en $\mathcal{L}^1(X, \mathcal{O}, \mu)$ que converge casi todo punto en $F \in \mathcal{L}^1(X, \mathcal{O}, \mu)$. Probar que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} F_j(T^j(x)) = \tilde{F}(x) \quad \text{c.t.p.}$$

EJERCICIO II.4. Sean (X, \mathcal{O}, μ) un espacio de probabilidad y $T : X \rightarrow X$ una transformación que preserve medida.

- (a) Probar que para todo $A \in \mathcal{O}$ con medida positiva existe un conjunto $A_0 \in \mathcal{O}$ contenido en A , de medida positiva tal que para cada $x \in A_0$, tenemos $\tau(x, A) \geq \mu(A)$. Sugerencia: Poner $A_1 := \{x : \tau(x, A) \geq \mu(A)\}$, y probar que $\mu(A_1) > 0$ y $A_1 = \cup_{n \geq 0} T^{-n}(A_1 \cap A)$.
- (b) Probar que $\tau(x, A) > 0$ para casi todo $x \in A$.
- (c) Sea X un espacio métrico separable y \mathcal{O} el σ -álgebra de Borel con una probabilidad T -invariante. Probar que para casi todo $x \in X$ tenemos $\tau_U(x) > 0$ para cada entorno U de x .

EJERCICIO II.5. Sea (X, \mathcal{O}, μ) un espacio de probabilidad y $T : X \rightarrow X$ una transformación que preserve medida.

- (a) Probar que si $C : X \rightarrow (0, +\infty)$ es medible, entonces $\liminf_{n \rightarrow \infty} \frac{1}{n} C(T^n(x)) = 0$ c.t.p. Sugerencia: Si esa propiedad no se cumple, existe $A \in \mathcal{O}$, $K_1 > 0$ y $K_2 > 0$ tal que

$$\liminf_{n \rightarrow \infty} \frac{1}{n} C(T^n(x)) \geq K_1, \quad C(x) \leq K_2 \quad \text{para todo } x \in A$$

y $\mu(A) > 0$. Use el Teorema de Recurrencia de Poincaré (Teorema I.2.1) para obtener una contradicción entre esas desigualdades.

- (b) Probar que si $C : X \rightarrow (0, +\infty)$ es medible y $C \circ T - C$ es integrable, entonces

$$\lim_{n \rightarrow \infty} \frac{1}{n} C(T^n(x)) = 0 \quad \text{c.t.p.}$$

Sugerencia: Escriba

$$\frac{1}{n} C(T^n(x)) = \frac{1}{n} \sum_{j=0}^{n-1} (C \circ T - C)(T^j(x)) + \frac{1}{n} C(x)$$

y aplique el Teorema de Birkhoff-Khinchin.

II.2. Ergodicidad

Considere una transformación que preserva medida T de un espacio de probabilidad (X, \mathcal{O}, μ) . Recuerde que un conjunto $A \in \mathcal{O}$ es T -invariante si $T^{-1}(A) = A$.

Definición 3. T es ergódica si cada conjunto T -invariante tiene medida 0 or 1.

Los shifts de Bernoulli y ciertas traslaciones y transformaciones lineales en el toro son ergódicas. Ver Secciones I.1 para las definiciones y A.4 para las demostraciones

Proposición II.2.1. Las siguientes propiedades son equivalentes:

- (1) T es ergódica;
- (2) Si $f \in \mathcal{L}^1(X)$ es T -invariante, entonces f es constante casi todo punto;
- (3) Si $f \in \mathcal{L}^p(X)$ es T -invariante, entonces f es constante casi todo punto;
- (4) Para cada $A, B \in \mathcal{O}$ tenemos

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=0}^{n-1} \mu(T^{-m}(A) \cap B) = \mu(A)\mu(B)$$

- (5) Para cada $f \in \mathcal{L}^1(X)$ tenemos $\tilde{f}(x) = \int_X f d\mu$ casi todo punto.

Demostración. (3) \Rightarrow (1). Si $A \in \mathcal{O}$ es T -invariante, su función característica χ_A es T -invariante y pertenece a $\mathcal{L}^p(X)$. Entonces χ_A es constante casi todo punto, esto es $\mu(A) = 0$ ó 1.

(1) \Rightarrow (2). Si $f \in \mathcal{L}^1(X)$ es T -invariante, el conjunto $A_c := \{x : f(x) \leq c\}$ es invariante para cada c . Como T es ergódica, esto significa que $\mu(A_c) = 0$ ó 1 para cada c . dejamos al lector mostrar que esto implica que f es constante casi todo punto.

(2) \Rightarrow (5). Como \tilde{f} pertenece a $\mathcal{L}^1(X)$ y es T -invariante, debe ser una constante. De $\int_X \tilde{f} d\mu = \int_X f d\mu$ se deduce (5).

(5) \Rightarrow (4). Por el Teorema de Birkhoff-Khinchin,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=0}^{n-1} \chi_A(T^j(x)) = \tilde{\chi}_A = \int_X \chi_A d\mu = \mu(A)$$

casi todo punto. Por el Teorema de Convergencia Dominada A.3.3,

$$\begin{aligned} \mu(A)\mu(B) &= \int_X \left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_A(T^j(x)) \right) \chi_B d\mu = \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \int_X \sum_{j=0}^{n-1} \chi_A(T^j(x)) \chi_B d\mu = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j}(A) \cap B) \end{aligned}$$

(4) \Rightarrow (1). Si A es T -invariante, aplicamos (4) a los conjuntos A y A^c . Entonces

$$\mu(A)\mu(A^c) = \lim_{n \rightarrow \infty} \sum_{j=0}^{n-1} \mu(T^{-j}(A) \cap A^c) = 0$$

por lo que $\mu(A) = 0$ ó 1. □

De hecho, la condición (5) sólo necesita satisfacerse en un conjunto denso de $\mathcal{L}^1(X)$ para implicar las otras cuatro:

Proposición II.2.2. Si existe un conjunto denso $F \subset \mathcal{L}^1(X)$ tal que $\tilde{f}(x) = \int_X f d\mu$ c.t.p. para cada $f \in F$, entonces T es ergódica.

Demostración. Como la sucesión

$$\frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j$$

converge a \tilde{f} en $\mathcal{L}^1(X)$, es suficiente verificar que para cada $f \in \mathcal{L}^1(X)$

$$\lim_{n \rightarrow \infty} \left\| \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j - \int_X f d\mu \right\|_1 = 0$$

Tome $\varepsilon > 0$, y elija $g \in F$ tal que $\|g - f\|_1 \leq \varepsilon/3$. Sea n_0 tal que $n \geq n_0$ implica

$$\left\| \frac{1}{n} \sum_{j=0}^{n-1} g \circ T^j - \int_X g d\mu \right\|_1 \leq \varepsilon/3$$

Entonces, para todo $n \geq n_0$,

$$\begin{aligned} \left\| \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j - \int_X f d\mu \right\|_1 &\leq \frac{1}{n} \left\| \sum_{j=0}^{n-1} f \circ T^j - \sum_{j=0}^{n-1} g \circ T^j \right\|_1 \\ &+ \left\| \frac{1}{n} \sum_{j=0}^{n-1} g \circ T^j - \int_X g d\mu \right\|_1 + \left| \int_X g d\mu - \int_X f d\mu \right|. \end{aligned}$$

Como

$$\|f \circ T^j - g \circ T^j\|_1 = \|(f - g) \circ T^j\|_1 = \|f - g\|_1$$

$$\text{y } \left| \int_X g d\mu - \int_X f d\mu \right| \leq \|g - f\|_1,$$

se concluye que

$$\left\| \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j - \int_X f d\mu \right\|_1 \leq \|g - f\|_1 + \varepsilon/3 + \|g - f\|_1 \leq \varepsilon$$

La Proposición está probada. \square

Otra caracterización de la ergodicidad puede darse en términos del tiempo de estadía $\tau(x, A)$ de la trayectoria de un punto x en el conjunto A , que existe por el Teorema de Birkhoff-Khinchin.

Proposición II.2.3. T es ergódica si y sólo si $\tau(x, A) = \mu(A)$ para casi todo $x \in X$ y todo $A \in \mathcal{O}$.

Demostración. Si T es ergódica entonces

$$\tau(x, A) = \tilde{\chi}_A(x) = \int_X \chi_A d\mu = \mu(A) \quad \text{c.t.p.}$$

Inversamente, sea $A \in \mathcal{O}$ un conjunto T -invariante. Si $\mu(A) > 0$, como $\tau(x, A) = 1$ para $x \in A$, se deduce que $\mu(A) = 1$. \square

El siguiente “teorema de unicidad” se satisface para transformaciones ergódicas:

Proposición II.2.4. Si T es ergódica con respecto a μ y $\mu_1 : \mathcal{O} \rightarrow [0, 1]$ es otra probabilidad T -invariante, las siguientes condiciones son equivalentes:

(a) $\mu_1 = \mu$;

(b) $\mu_1 \ll \mu$;

(c) No existe conjunto T -invariante $A \in \mathcal{O}$ tal que $\mu(A) = 0$ y $\mu_1(A) \neq 0$.

Demostración. (b) \Rightarrow (a). Si $\mu_1 \ll \mu$ y ambas, μ_1 y μ , son invariantes, la derivada de Radon-Nikodym $d\mu_1/d\mu$ es una función invariante. Como T es ergódica, $d\mu_1/d\mu$ es constante c.t.p. y, por tanto $\mu_1 = \mu$.

(c) \Rightarrow (b). Suponga que (b) no se cumple y tome $A_0 \in \mathcal{O}$ tal que $\mu(A_0) = 0$ y $\mu_1(A_0) \neq 0$. El conjunto $A := \cup_{n \in \mathbb{Z}} T^n(A_0)$ contradice (c).

(a) \Rightarrow (c). Trivial. \square

Si X es un conjunto y \mathcal{O} es un σ -álgebra en X , el conjunto de todas las medidas con signo sobre \mathcal{O} tiene una estructura de espacio vectorial trivial. Si $T : X \rightarrow X$ es una transformación medible, el conjunto $m_T(X, \mathcal{O})$ de las probabilidades T -invariantes sobre \mathcal{O} es un conjunto convexo de ese espacio vectorial. La siguiente proposición caracteriza a las transformaciones ergódicas con respecto a T . Observar que la expresión: “ μ is ergódica con respecto a T ” significa que $\mu \in m_T(X, \mathcal{O})$ y T es una transformación ergódica de (X, \mathcal{O}, μ) . Obsérvese también que esta proposición implica que si existe una única medida invariante para una transformación medible en un espacio de probabilidad, entonces la transformación es ergódica. Por ello es que si hay una única medida invariante la transformación es también únicamente ergódica.

Proposición II.2.5. La medida $\mu \in m_T(X, \mathcal{O})$ es ergódica si y sólo si μ es un punto extremal¹ of $m_T(X, \mathcal{O})$.

Demostración. Suponga que $\mu \in m_T(X, \mathcal{O})$ es ergódica y $\mu = \lambda\mu_1 + (1 - \lambda)\mu_2$ para algunas $\mu_1, \mu_2 \in m_T(X, \mathcal{O})$ y $0 < \lambda < 1$. Tenemos $\mu_1 \ll \mu$ puesto que $\lambda \neq 0$, entonces la Proposición II.2.4 implica que $\mu_1 = \mu$. Similarmente, tenemos $\mu_2 = \mu$, lo cual muestra que $\mu_1 = \mu_2$, por lo que μ es extremal.

Ahora, si $\mu \in m_T(X, \mathcal{O})$ no es ergódica, existe un conjunto T -invariante $A_0 \in \mathcal{O}$ satisfaciendo $0 < \mu(A_0) < 1$. Definimos medidas $\mu_i \in m_T(X, \mathcal{O})$, $i = 1, 2$, por

$$\mu_1(A) = \frac{1}{\mu(A_0)}\mu(A \cap A_0) \quad \text{y} \quad \mu_2(A) = \frac{1}{\mu(A_0^c)}\mu(A \cap A_0^c)$$

para todo $A \in \mathcal{O}$. Entonces podemos escribir $\mu = \mu(A_0)\mu_1 + \mu(A_0^c)\mu_2$ lo cual muestra que μ no es extremal. \square

EJERCICIO II.6. Sean X un conjunto, \mathcal{O} un σ -álgebra sobre X y $T : X \rightarrow X$ una transformación medible. Si $\mu_i \in m_T(X, \mathcal{O})$, $i = 1, \dots, n$ son medidas ergódicas y $\mu_i \not\ll \mu_j$ para $i \neq j$, probar que existen conjuntos disjuntos $A_i \in \mathcal{O}$, $i = 1, \dots, n$, tales que $\cup_{i=1}^n A_i = X$ y $\mu_j(A_i) = \delta_{ij}$.

EJERCICIO II.7. Sean T una transformación continua de un espacio métrico compacto X y $\mu \in m_T(X)$ una medida ergódica.

(a) Probar que existe un conjunto de medida total $A \subset X$ tal que para cada $x \in A$ tenemos:

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

para toda función continua $f : X \rightarrow \mathbb{C}$.

(b) Probar que para casi todo $x \in X$ tenemos $\tau(x, A) = \mu(A)$ para todo $A \in \mathcal{O}$ tal que $\mu(\partial A) = 0$.

(c) Probar que para casi todo $a \in X$ existe un conjunto numerable $S \subset (0, +\infty)$, tal que para $r \notin S$ tenemos $\tau(a, B_r(a)) = \mu(B_r(a))$

EJERCICIO II.8. Sean (X, \mathcal{O}, μ) un espacio de probabilidad y $T : X \rightarrow X$ una transformación que preserva medida. Probar que si $f \in \mathcal{L}^2(X)$ satisface

$$\sum_{n=0}^{\infty} \left| \langle U_T^n f, f \rangle - \left(\int_X f d\mu \right)^2 \right| < +\infty \quad \text{entonces} \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) = \int_X f d\mu$$

para casi todo $x \in X$. Sugerencia: Use la desigualdad de Chebyshev:

$$\mu(S_\varepsilon) \leq \frac{\sigma^2(f)}{\varepsilon^2}$$

donde $\sigma^2(f) = \langle f, f \rangle - \left(\int f d\mu \right)^2$ y $S_\varepsilon = \{x \in X : |f(x) - \int f d\mu| > \varepsilon\}$.

¹Un punto extremal x de un conjunto convexo C es uno que no puede ser presentado como $x = \lambda y + (1 - \lambda)z$ donde $y, z \in C$ son puntos distintos y $0 < \lambda < 1$.

II.3. Jerarquía Ergódica

En la Sección A.1 probamos que una transformación T que preserva medida de un espacio de probabilidad (X, \mathcal{O}, μ) es ergódica si y sólo si para cada $A, B \in \mathcal{O}$

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{j=0}^{m-1} \mu(T^{-j}(A) \cap B) = \mu(A)\mu(B)$$

En este caso, si el límite de $\mu(T^{-j}(A) \cap B)$ cuando $j \rightarrow \infty$ existe, su valor debe ser igual a $\mu(A)\mu(B)$.

Definición 4. Un endomorfismo T que preserva medida de un espacio de probabilidad (X, \mathcal{O}, μ) se denomina *mixing* (mezcladora²) si para cualquier par $A, B \in \mathcal{O}$

$$\lim_{n \rightarrow \infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B).$$

Una descripción gráfica del clásico libro de Arnold y Avez [AA67] explica cómo actúa una transformación *mixing*. Suponga que una coctelera M , $\mu(M) = 1$ es llenada con 85% de pisco y 15% de jugo de limón. Sea A la parte de la coctelera originalmente ocupada por el jugo y B cualquier parte del recipiente. Sea $T: M \rightarrow M$ la transformación que consiste en cada movimiento abrupto de la coctelera al actuar el barman (quien está sacudiendo reiteradamente el recipiente). Entonces, después de n sacudidas la fracción de jugo en la parte B será $\mu(T^n(A) \cap B)/\mu(B)$. Mientras el barman siga sacudiendo ($n \rightarrow \infty$), la fracción de jugo en cualquier parte B se aproxima a $\mu(A) = 15\%$, o sea que el jugo de limón se habrá distribuido uniformemente en la mezcla.

Observamos que la definición que hemos dado para una transformación *mixing map* es buena tanto para el caso invertible como para el no invertible (endomorfismos).

Proposición II.3.1. Toda transformación *mixing* es ergódica.

Demostración. Sea A cualquier conjunto medible T -invariante, entonces $T^{-n}(A) = A$ y $\mu(A \cap B) = \lim_{n \rightarrow \infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B)$. En particular, para $A = B$ tenemos $\mu(A) = \mu^2(A)$. Esto significa $\mu(A) = 0$ ó 1 ; por tanto T es ergódica. \square

Observamos que no todas las transformaciones ergódicas son *mixing*; ver Ejercicio II.9. Por tanto, *mixing* es una propiedad más fuerte que la ergodicidad.

Definición 5. Si X es un espacio topológico, una transformación $T: X \rightarrow X$ es topológicamente *mixing* si para cualquier par de conjuntos abiertos $U, V \subset X$ existe $N \in \mathbb{N}$ tal que $T^{-n}(U) \cap V \neq \emptyset$ para todo $n \geq N$.

Proposición II.3.2. Si X es un espacio topológico, \mathcal{O} es el σ -álgebra de Borel y μ una medida de probabilidad positiva sobre abiertos, entonces, si $T: X \rightarrow X$ es *mixing*, también es topológicamente *mixing*.

Demostración. Como para cualquier par de abiertos U, V tenemos $\lim_{n \rightarrow \infty} \mu(T^{-n}(U) \cap V) = \mu(U)\mu(V) \neq 0$ resulta que, para todo $n \geq N$, $\mu(T^{-n}(U) \cap V) \neq 0$, por lo que $T^{-n}(U) \cap V \neq \emptyset$. \square

Definición 6. Una transformación invertible T que preserva medida en un espacio de probabilidad (X, \mathcal{O}, μ) es *K-mixing* o de Kolmogorov si existe una sub- σ -álgebra $\mathcal{O}_0 \subset \mathcal{O}$ tal que

- 1) $\mathcal{O}_0 \subset T\mathcal{O}_0$,
- 2) $\bigvee_{n=0}^{\infty} T^n \mathcal{O}_0 = \mathcal{O} \pmod{0}$ ³ y
- 3) $\bigcap_{n=0}^{\infty} T^{-n} \mathcal{O}_0 = \mathcal{N} = \{X, \emptyset\} \pmod{0}$.

Definición 7. Un automorfismo T que preserva medida de un espacio de probabilidad (X, \mathcal{O}, μ) es Bernoulli (o satisface la propiedad Bernoulli) si es equivalente a un *shift de Bernoulli*. Ver Ejemplo 3 en la Sección I.1 y Sección I.3 (el *shift* puede estar definido sobre un espacio de probabilidad, sin estructura topológica –ver llamada al pie de página en el Ejemplo 3 de la Sección I.1).

²Usaremos la expresión en inglés aunque no haya sido la norma en estas notas.

³Si \mathcal{A}_n es una familia de subconjuntos de X , entonces $\bigvee_n \mathcal{A}_n$ es el σ -álgebra generado por $\bigcup_n \mathcal{A}_n$.

Se puede probar que todo automorfismo K-mixing es mixing, pero no vice versa. Todos estos resultados son discutidos en el libro por Mañé [Mn83], Sections II.8 and II.11. Se prueba que hay una llamada *jerarquía ergódica*:

$$\text{Bernoulli} \Rightarrow (\text{K - mixing}) \Rightarrow \text{Mixing} \Rightarrow \text{Ergódica} \quad (\text{II.3.1})$$

Cada propiedad en esa lista está referida a transformaciones que preservan medidas de un espacio de probabilidad (X, \mathcal{O}, μ) . Cada una de ellas implica la siguiente, la cual, por tanto, es más débil que la anterior.

Notamos que todas las implicaciones en (II.3.1) son en un sentido solamente; ninguna de ellas puede ser dirigida en el sentido opuesto. Esto significa que la ergodicidad no implica mixing (como fue observado arriba), mixing no implica K-mixing, etc.: existen ejemplos de la falsedad de cada implicación inversa.

Por tanto la propiedad Bernoulli es el más alto grado de caos en términos de la teoría ergódica, e implica todas las otras. Además, todas las transformaciones Bernoulli con la misma entropía son equivalentes desde el punto de vista de la teoría de la medida.

EJERCICIO II.9. Sea λ la probabilidad de Lebesgue sobre el círculo unitario S^1 . Probar:

- (a) Las rotaciones irracionales de S^1 son ergódicas.
- (b) Las rotaciones de S^1 nunca son mixing. Sugerencia: Sea, por ejemplo R_ω , la rotación de S^1 de ángulo $0 < \omega < 1/2$. Tome dos arcos A, B de longitud $\pi\omega$. Observe que si $R_\omega^{-n}A \cap B \neq \emptyset$, entonces $R_\omega^{-n-1}A \cap B = \emptyset$.

II.4. Propiedades Estadísticas de Sistemas Dinámicos

Para aquellos familiarizados con la Teoría de Probabilidad presentamos aquí una discusión adicional que introduce algunos conceptos relacionados con la velocidad de mezclado y de acercamiento de las medias de Birkhoff a su límite. Lamentablemente no desarrollaremos las técnicas que permiten estudiar estos conceptos en ejemplos concretos, pero la buena comprensión de estos conceptos permitirá entender las grandes motivaciones de toda esta teoría. Y los senderos de su desarrollo actual.

En el lenguaje de la teoría de probabilidad, un shift de Bernoulli representa una sucesión de variables aleatorias independientes. Esto se deduce de inmediato de la fórmula producto (I.1.1) en el Capítulo I. Por esta razón la propiedad Bernoulli es vista como una propiedad *estadística* de un sistema dinámico. Establece una equivalencia entre un sistema dinámico y una sucesión puramente aleatoria de sucesos independientes: un modelo canónico de la teoría de probabilidad.

Esta es una observación muy interesante e importante. Un sistema dinámico $T : X \rightarrow X$ es, por naturaleza, completamente determinístico. Esto significa que si se tiene un punto $x \in X$, su futuro $\{T^n x\}$, $n \geq 1$, está unívocamente determinado y puede ser calculado con precisión. Cuando la transformación T es invertible, el pasado $\{T^n x\}$, $n \leq -1$, está también determinado y es computable. Uno puede mirar esto de la siguiente manera: conociendo el *estado actual* de un sistema dinámico (dado por $x \in X$), puede determinarse su futuro y, frecuentemente, su pasado. Este es el significado preciso que damos a la palabra “determinístico”.

Por otro lado, en una sucesión de sucesos independientes el resultado de cualquier ensayo no da ninguna pista de cuales serán (o fueron) los resultados de otros ensayos. Así, el conocimiento del estado actual no nos dice nada sobre el futuro o el pasado, el resultado de cualquier ensayo será completamente aleatorio e impredecible⁴. ¿Una paradoja? En algún sentido, lo es, y existen discusiones relevantes en física en que la teoría de sistemas dinámicos tiene la mayor parte de sus aplicaciones. No desarrollamos estos aspectos aquí. Ver, por ejemplo, [Lb99a, Lb99b].

Sin embargo haremos algunos otros comentarios. Mientras que la propiedad Bernoulli es una manifestación de aleatoriedad o caos absoluto, extrañamente, tiene poca relevancia en las aplicaciones físicas directas. ¿Por qué? Porque la equivalencia entre un sistema dinámico y un shift de Bernoulli es, normalmente, dada por una transformación meramente medible con una estructura muy complicada, no suave (diferenciable) ni aún continua. En física, por otro lado, las leyes de los movimientos son normalmente especificadas por ecuaciones diferenciales (como las ecuaciones de Newton o Hamiltonianas), y todas las funciones interesantes (como la temperatura, la energía, la presión) son también suaves. Por tanto, sólo las propiedades de los sistemas dinámicos expresadas por transformaciones y funciones suaves son relevantes en física.

⁴Observamos que, a pesar de este hecho, el comportamiento a largo plazo de los sucesos independientes puede ser bastante bien descrito por las leyes de la teoría de probabilidad.

Por estas razones, suponiendo que X es una variedad, $T : X \rightarrow X$ una transformación suave que preserva una probabilidad μ y $f : X \rightarrow \mathbb{R}$ una función suave, uno puede caracterizar al sistema de una manera significativa físicamente como sigue. Considere

$$S_n = f + f \circ T + f \circ T^2 + \dots + f \circ T^{n-1} \quad (\text{II.4.1})$$

Los cocientes S_n/n son llamados *media temporal* de la función f . Adoptando la notación física, denotamos por $\langle \cdot \rangle$ al valor esperado de la función con respecto a μ , o sea $\langle f \rangle = \int_X f d\mu$. La integral $\langle f \rangle$ es también llamada *media espacial* de f .

Ahora, el Teorema Ergódico de Birkhoff establece que si T es ergódica, entonces S_n/n converge casi todo punto a $\langle f \rangle$ as $n \rightarrow \infty$. En lenguaje físico, esto significa que **las medias temporales convergen a las medias espaciales**. En teoría de probabilidad, este resultado es llamado la *ley fuerte de los grandes números*.

Una importante característica de los sistemas dinámicos es la *función de correlación temporal*

$$C_f(n) = \langle f \cdot (f \circ T^n) \rangle - \langle f \rangle^2. \quad (\text{II.4.2})$$

Si la transformación T es mixing, uno puede mostrar que $C_f(n) \rightarrow 0$ cuando $n \rightarrow \infty$ (Ver Ejercicio II.10), o sea, como dicen los físicos, la *correlación decae*. La velocidad asintótica de convergencia $C_f(n) \rightarrow 0$ caracteriza la “velocidad de mezclado” del sistema. Ver [Vi97, Ba00, CY00]

Por otro lado, decimos que f satisface el *teorema del límite central* si

$$\lim_{n \rightarrow \infty} \mu \left\{ x: \frac{S_n(x) - n\langle f \rangle}{\sqrt{n}} < z \right\} = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^z e^{-\frac{s^2}{2\sigma^2}} ds \quad (\text{II.4.3})$$

para toda $-\infty < z < \infty$. Aquí $\sigma = \sigma_f \geq 0$ es una constante relacionada con la función de correlación:

$$\sigma_f^2 = C_f(0) + 2 \sum_{n=1}^{\infty} C_f(n). \quad (\text{II.4.4})$$

La ecuación (II.4.3) es equivalente a la convergencia en distribución de $(S_n - n\langle f \rangle)/\sqrt{n}$ a la variable aleatoria normal $N(0, \sigma_f^2)$. Destacamos que el teorema del límite central es considerablemente más refinado que el Teorema Ergódico de Birkhoff-Khincin; nos dice que la distribución de las desviaciones de las medias temporales S_n/n respecto de su valor límite $\langle f \rangle$, cuando son “escaladas” por $1/\sqrt{n}$, es asintóticamente Gaussiana. Este resultado no exige ergodicidad, y se refiere a la manera como las medias de Birkhoff tienden a su límite, en las condiciones que comentamos a continuación.

Es claro de (II.4.4) que el teorema del límite central sólo es válido si $\sum_n |C_f(n)| < \infty$. En realidad, la prueba de este teorema requiere una convergencia aún más rápida de $C_f(n)$ a cero. Por estas (y otras) razones la velocidad de esta convergencia es vista como una importante característica estadística del sistema. Dos tipos principales de convergencia son: la *exponencial*, cuando $|C_f(n)| < \text{const} \cdot e^{-an}$, $a > 0$, y la *polinomial*, cuando $|C_f(n)| < \text{const} \cdot n^{-b}$, $b > 0$. Los sistemas con decaimiento exponencial son los “más caóticos”, poseen muchas características necesarias para las aplicaciones en física estadística. Los sistemas con decaimiento polinomial son vistos como de compartamiento intermedio (“intermitente”) entre “regular” y “caótico”, y esas características son muy sensibles al valor exacto de la potencia $b > 0$ y otros factores que pueden limitar sus aplicaciones en física estadística.

Es también interesante observar que si uno disminuye los requerimientos de suavidad de la función f en (II.4.2), uno pierde totalmente el control sobre el decaimiento de correlaciones. En todos los sistemas dinámicos mixing conocidos, la convergencia $C_f(n) \rightarrow 0$ tiene velocidad arbitrariamente baja para funciones integrables genéricas, y aún para funciones continuas genéricas (D. Volný, 1990). Ver, por ejemplo, [CY00]. Por lo tanto, la suavidad de f es esencial.

Por otra parte, muy sorprendentemente, para muchos sistemas dinámicos interesantes uno puede probar el teorema del límite central, y obtener buenas estimaciones para el decaimiento de correlación de funciones suaves f . Esto abre las puertas a una interacción cercana entre la teoría de sistemas dinámicos y la mecánica estadística, que es actualmente un área de investigación muy activa.

EJERCICIO II.10. Sea $T : X \rightarrow X$ una transformación mixing que preserva una probabilidad μ . Probar que para todas $f, g \in L^2(X)$

$$\int_X f(T^n x)g(x) d\mu \rightarrow \int_X f(x) d\mu \cdot \int_X g(x) d\mu$$

Sugerencia: Primero considere funciones simples f, g , y después aproxime funciones L^2 por funciones simples.

EJERCICIO II.11. Bajo las condiciones del Teorema de Recurrencia de Poincaré, suponga que $f : X \rightarrow \mathbb{C}$ es una función medible tal que $f(x) \neq 0$ c.t.p. Mostrar que la sucesión $S_n(x)$ definida por (II.4.1) diverge c.t.p. Sugerencia: considere los conjuntos $A_n = \{x \in X : |f(x)| > 1/n\}$.

Capítulo III

Exponentes de Liapunov. Teoría de Pesin

III.1. Exponentes de Liapunov

Sea p un punto fijo de un difeomorfismo $f : A \rightarrow A$, $A \subset \mathbb{R}^d$. Queremos estudiar el comportamiento de f^n ($n \in \mathbb{Z}$) en un entorno de p . Como primera aproximación a f consideramos su parte lineal, o sea su derivada $f'_p : \mathbb{R}^d \rightarrow \mathbb{R}^d$.

Sean $\alpha_1, \alpha_2, \dots, \alpha_r, \alpha_{r+1}, \bar{\alpha}_{r+1}, \dots, \alpha_s, \bar{\alpha}_s$ todas las raíces distintas del polinomio característico de f'_p . En esa sucesión denominamos $\alpha_1, \dots, \alpha_r$ todas las raíces reales y por $\alpha_j, \bar{\alpha}_j$, $r+1 \leq j \leq s$, todos los pares conjugados de raíces complejas. Sean también \tilde{m}_j , $1 \leq j \leq s$, las respectivas multiplicidades. El teorema de Jordan (forma canónica real) establece que todas las raíces (*valores propios -eigenvalues- de f'_p*) están asociados a subespacios propios generalizados f'_p -invariantes \tilde{E}_j , $1 \leq j \leq s$, cuyas respectivas dimensiones son \tilde{m}_j para $1 \leq j \leq r$ y $2\tilde{m}_j$ for $r < j \leq s$ (en el último caso \tilde{E}_j está asociado al par $\alpha_j, \bar{\alpha}_j$). Más aún, $\mathbb{R}^d = \bigoplus_{j=1}^s \tilde{E}_j$. Observe que dado que f es un difeomorfismo, entonces $\det f'_p \neq 0$, por lo que tenemos $\alpha_j \neq 0$ para todo j .

Si v_i es un vector propio de \tilde{E}_i , entonces $(f^n)'_p v_i = \alpha_i^n v_i$, por lo que

$$\log \|(f^n)'_p v_i\| = n \log |\alpha_i| + \log \|v_i\|$$

para todo $n \in \mathbb{Z}$. Esta relación no es cierta para *cualquier* vector $v_i \in \tilde{E}_i$, pero sí es cierto que

$$\lim_{n \rightarrow \pm\infty} \frac{1}{n} \log \|(f^n)'_p v_i\| = \log |\alpha_i| \quad (\text{III.1.1})$$

para todo $\vec{0} \neq v_i \in \tilde{E}_i$, ver Ejercicio III.1. Esto muestra que cuando $|\alpha_i| > 1$, el vector $(f^n)'_p v_i$ crece exponencialmente cuando $n \rightarrow \infty$ y disminuye exponencialmente cuando $n \rightarrow -\infty$. Si $|\alpha_i| < 1$, sucede lo mismo, con las evoluciones temporales inversas. Si $|\alpha_i| = 1$, no hay crecimiento o decrecimiento exponencial, pero podría haber crecimiento o contracción más lentos (por ejemplo, lineal en n) del vector $(f^n)'_p v_i$.

La ecuación (III.1.1) sugiere estudiar no los autovalores α_i , sino los logaritmos de sus módulos,

$$\lambda_i = \log |\alpha_i|$$

que son llamados los *exponentes de Liapunov* de la transformación f en el punto fijo p . Observamos que algunos autovalores distintos $\alpha_i \neq \alpha_j$ corresponden al mismo exponente de Liapunov si $|\alpha_i| = |\alpha_j|$. En este caso cada vector no nulo de la suma directa de los correspondientes subespacios $\tilde{E}_i \oplus \tilde{E}_j$ satisfacen (III.1.1). Entonces, tenemos una decomposición de \mathbb{R}^d como suma directa de subespacios $E_1 \oplus \dots \oplus E_{m(p)}$ tal que si $\vec{0} \neq v_i \in E_i$, entonces

$$\lim_{n \rightarrow \pm\infty} \frac{1}{n} \log \|(f^n)'_p v_i\| = \lambda_i(p) \quad (\text{III.1.2})$$

donde $\lambda_i(p)$ es el exponente de Liapunov asociado a E_i .

Más aún, si se supone que no hay exponentes de Liapunov nulos en el punto fijo p (diremos que tal punto es *hiperbólico*), podemos sumar todos los subespacios con exponentes de Liapunov negativos y todos los subespacios con exponentes de Liapunov positivos para obtener, respectivamente, subespacios E^s y E^u , tales que

$$(E1) \quad \mathbb{R}^d = E^s \oplus E^u,$$

$$(E2) \quad f'_p(E^s) = E^s \text{ y } f'_p(E^u) = E^u$$

(E3) existen $\lambda > 0$ y $n_0 \geq 1$ tales que para todo $|n| \geq n_0$

$$\frac{1}{n} \log \|(f^n)'_p v\| \leq -\lambda \quad v \in E^s, \|v\| = 1 \text{ y } \frac{1}{n} \log \|(f^n)'_p v\| \geq \lambda \quad v \in E^u, \|v\| = 1.$$

Esto significa que los vectores de E^s son contraídos por las iteraciones hacia adelante (futuro) de f'_p y expandidos por las iteraciones hacia atrás (pasado) de f'_p . Es al revés para los vectores de E^u . Observamos que podría suceder que $E^u = \{\vec{0}\}$ o $E^s = \{\vec{0}\}$. Estaremos primariamente interesados en el estudio de los puntos hiperbólicos.

El Teorema de Großman-Hartman asegura que en un punto fijo hiperbólico el comportamiento de f es similar al de su parte lineal. Precisamente, a los subespacios (f'_p) -invariantes E^s, E^u les corresponden subvariedades f -invariantes $W^s(p), W^u(p) \subset A$ (diferenciablemente immersas en A) tales que

$$(W1) \quad \mathcal{T}_p W^s(p) = E^s \text{ y } \mathcal{T}_p W^u(p) = E^u,$$

(W2) existe un entorno $U(p) \subset A$ tal que $f(W^s(p) \cap U(p)) \subset W^s(p)$ and $f(W^u(p) \cap U(p)) \subset W^u(p)$ y

(W3) se cumple

$$\lim_{n \rightarrow \infty} f^n(y) = p \quad y \in W^s(p) \text{ y } \lim_{n \rightarrow \infty} f^{-n}(y) = p \quad y \in W^u(p).$$

Se encuentran pruebas de este Teorema en [PM78, KH95]. La idea principal es representar f en un sistema local de coordenadas de p , asociado con E^s y E^u , y luego contruir $W^s(p)$ como el grafo de una función en aquellas coordenadas con sucesivas iteraciones por f . El resultado para W^u sale sustituyendo f por f^{-1} .

El resultado de arriba fácilmente se extiende a cualquier difeomorfismo $f : A \rightarrow A$ de un conjunto abierto $A \subset M$ de una variedad Riemanniana M , en lugar de $A \subset \mathbb{R}^d$. Una estructura Riemanniana en M es necesaria para que la norma $\|\cdot\|$ esté bien definida. Las pruebas son esencialmente las mismas que en el caso de \mathbb{R}^d . De ahora en adelante supondremos que f está definida en un subconjunto abierto de una variedad Riemanniana.

Es necesario para nuestro estudio ulterior especificar la velocidad de convergencia en (W3). Denotamos $\text{dist}(\cdot, \cdot)$ a la distancia sobre la variedad Riemanniana M . Sea $\lambda > 0$ tal que ningún exponente de Liapunov está en el intervalo $(-\lambda, \lambda)$. Entonces podemos especificar (W3) como

(W3') se cumple

$$\text{dist}(f^n(y), p) \leq C e^{-\lambda n}, \quad y \in W^s(p) \text{ y } \text{dist}(f^{-n}(y), p) \leq C e^{-\lambda n}, \quad y \in W^u(p)$$

para todo $n \geq 0$ y alguna constante $C > 0$.

Si p no es un punto fijo, pero periódico, con período k , todos los resultados se aplican a la transformación $f^k : A \rightarrow A$.

Si p no es un punto periódico podemos aún definir los exponentes de Liapunov de una manera similar:

Definición 8. Sea la transformación f^n diferenciable en $p \in M$ para todo $n \in \mathbb{Z}$. Suponga que el espacio tangente $\mathcal{T}_p M$ se descompone en suma directa de subespacios $E_1 \oplus \cdots \oplus E_m(p)$ tales que

$$\lim_{n \rightarrow \pm\infty} \frac{1}{n} \log \|(f^n)'_p v_i\| = \lambda_i(p), \quad \text{para } \vec{0} \neq v_i \in E_i. \quad (\text{III.1.3})$$

Los valores $\lambda_i(p)$ son llamados exponentes de Liapunov en p , siendo $\dim E_i$, sus respectivas multiplicidades.

Observamos que la existencia del límite (III.1.3) no está garantizado para ningún punto $p \in A$; volveremos a este asunto en la próxima Sección. Por ahora, diremos que p tiene todos los exponentes de Liapunov si los límites de arriba existen.

Si un punto p tiene todos sus exponentes de Liapunov y ninguno es cero, decimos que el punto p es un *punto hiperbólico*. Para un punto hiperbólico $p \in M$, se cumple $\mathcal{T}_p M = E_p^s \oplus E_p^u$, donde

$$E_p^s = \oplus_{\lambda_i(p) < 0} E_i \quad y \quad E_p^u = \oplus_{\lambda_i(p) > 0} E_i \quad (\text{III.1.4})$$

Todos los resultados de arriba se aplican a los puntos (no periódicos) hiperbólicos, incluyendo la existencia de subvariedades invariantes $W^s(p)$ y $W^u(p)$, pero la propiedad (W3') debe ser modificada correspondientemente:

(W3'') para algún $C > 0$ y todo $n \geq 0$ se cumole:

$$\begin{aligned} \text{dist}(f^n(y), f^n(p)) &\leq C e^{-\lambda n} \quad y \in W^s(p) \quad y \\ \text{dist}(f^{-n}(y), f^{-n}(p)) &\leq C e^{-\lambda n} \quad y \in W^u(p). \end{aligned}$$

Analicemos las propiedades anteriores con más cuidado. Ellas significan que las órbitas para el futuro de los puntos de $W^s(p)$ se están acercando entre ellas (convergen) con velocidad exponencial. Por esta razón $W^s(p)$ es llamada la *variedad estable*. La denominación proviene de las ecuaciones diferenciales, donde la convergencia de soluciones es interpretada como estabilidad: se quiere destacar que órbitas que han comenzado cercanas se mantienen cercanas: son estables en el sentido usual de la palabra. Las órbitas para el futuro de puntos de $W^u(p)$ se van separando (divergen) con velocidad exponencial, y por esta razón $W^u(p)$ es llamada la *variedad inestable*. Observe, sin embargo, que las órbitas hacia el pasado de $W^u(p)$ convergen, y las órbitas hacia el pasado de $W^s(p)$ divergen.

Vemos que las órbitas de todos los puntos cerca de un punto hiperbólico p son muy inestables: ellas divergen (se van separando) con velocidad exponencial sea en el futuro o en el pasado. Si $\dim W^s(p) \neq 0$ y $\dim W^u(p) \neq 0$, entonces cualquier punto y cercano a p , que no está exactamente en $W^u(p)$ o en $W^s(p)$, tiene su trayectoria de y se separándose de la de p tanto en el futuro como en el pasado! Esta separación exponencial de las trayectorias es la fuente principal de inestabilidad, turbulencia, mezclado – todo eso que llamamos *caos*.

Los estudios de sistemas dinámicos con puntos hiperbólicos comenzaron hace mucho tiempo. Alrededor de 1900 J. Hadamard probó la hiperbolicidad de los flujos geodésicos sobre variedades con curvatura negativa constante. Por los años 30 del siglo XX J. G. Hedlund and E. Hopf estudiaron las propiedades ergódicas de tales flujos.

Basado en los estudios de flujos geodésicos, por los años 60 D. Anosov (y S. Smale, de manera diferente) introdujo clases generales de difeomorfismos con puntos hiperbólicos. Daremos la definición de los que ahora llamamos *difeomorfismos de Anosov* (a los que Anosov mismo denominó “difeomorfismos con una condición”, o C-difeomorfismos, la letra “C” aparece por “condición”).

Definición 9. . Un difeomorfismo $f : M \rightarrow M$ de una variedad compacta Riemanniana es de Anosov si existen constantes $K > 0$, $\lambda > 0$ tales que para cada punto $p \in M$ existe una descomposición $\mathcal{T}_p M = E_p^s \oplus E_p^u$ con las propiedades

(A1) E_p^s y E_p^u son f' -invariantes, o sea $f'_p(E_p^s) = E_{f(p)}^s$ y $f'_p(E_p^u) = E_{f(p)}^u$,

(A2) para todo $n \geq 0$

$$\begin{aligned} \|(f^n)'_p v\| &\leq K e^{-n\lambda} \|v\| \quad v \in E_p^s \\ \|(f^{-n})'_p v\| &\leq K e^{-n\lambda} \|v\| \quad v \in E_p^u \end{aligned}$$

(A3) los espacios E_p^s y E_p^u dependen continuamente de p .

La condición (A3) se incluye por tradición, en realidad se deduce de las otras dos.

Observe que las condiciones (A1) y (A2) de Anosov son en realidad más débiles que los requerimientos de la definición de exponentes de Liapunov: los exponentes de Liapunov no tienen porque existir en todos los puntos para un difeomorfismo de Anosov. Sin embargo, las condiciones (A1) y (A2) captan la esencia de la hiperbolicidad, en particular en todo $p \in M$ existen variedades estable e inestable $W^s(p)$ y $W^u(p)$ que satisfacen (W3''), y sus hiperplanos tangentes son E_p^s y E_p^u , respectivamente. Por estas razones, los puntos que satisfacen las condiciones (A1) y (A2) son frecuentemente llamados *hiperbólicos* (aún cuando los exponentes de Liapunov técnicamente no existan). Esta puede parecer una terminología confusa pero, en lo esencial, será clarificada en la próxima Sección.

OBSERVACIÓN III.1.1. La constante K en la condición (A2) juega un rol auxiliar. Siempre se pueda cambiar la métrica Riemanniana en M de modo que la condición (A2) se satisfaga con $K = 1$, o sea, que la constante K puede ser eliminada. En este caso la contracción de los vectores estables $v \in E_p^s$ y la expansión de los vectores inestables $v \in E_p^u$ por $(f^n)'_p$ será monótona en n . Tal métrica, con $K = 1$, es llamada *métrica adaptada* o *métrica de Liapunov*.

EJERCICIO III.1. Verificar la fórmula (III.1.1). Sugerencia: Suponga, por simplicidad, que $\dim \tilde{E}_i = 2$. Hay dos casos. 1) Si α_i es una raíz real de multiplicidad 2, entonces f' restringida a \tilde{E}_i está dada por la matriz de Jordan $J = \begin{pmatrix} \alpha_i & 1 \\ 0 & \alpha_i \end{pmatrix}$. Verifique que $J^n = \begin{pmatrix} \alpha_i^n & n\alpha_i^{n-1} \\ 0 & \alpha_i^n \end{pmatrix}$ para todo $n \in \mathbb{Z}$ y entonces deduzca (III.1.1). 2) Si $\alpha_i = a + bi$ es una raíz compleja, $b \neq 0$, entonces la correspondiente forma canónica de Jordan es $J = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$. Verificar que $J^n = |\alpha_i|^n \begin{pmatrix} \cos n\varphi & \sin n\varphi \\ -\sin n\varphi & \cos n\varphi \end{pmatrix}$ para algún $\varphi \in [0, 2\pi)$ y todo $n \in \mathbb{Z}$, y luego deduzca (III.1.1).

III.2. Teorema de Oseledec

Ahora estudiaremos la existencia de exponentes de Liapunov, o sea la existencia del límite (III.1.3). El siguiente ejemplo muestra que puede haber muchos puntos donde los exponentes de Liapunov no existen.

Ejemplo. Sea $f : S^1 \rightarrow S^1$ un difeomorfismo del círculo dado por $f(x) = x + \frac{1}{3\pi} \sin 2\pi x$, donde $0 \leq x < 1$ es una coordenada cíclica en S^1 . Tenemos dos puntos fijos: $x_0 = 0$ y $x_1 = 1/2$, ambos con exponentes de Liapunov: $\lambda(x_0) = \log f'(x_0) = \log(5/3) > 0$ and $\lambda(x_1) = \log f'(x_1) = \log(1/3) < 0$. Entonces el punto x_0 es inestable (un *repulsor*); x_1 es un punto estable (un *atractor*). Para cualquier punto $p \in (0, 1/2)$ tenemos $f^n(p) \rightarrow 1/2$ y $f^{-n}(p) \rightarrow 0$ cuando $n \rightarrow \infty$. Entonces, por la regla de la cadena, para cualquier vector no nulo $v \in \mathcal{T}_p(S^1)$ tenemos

$$\log(5/3) = \lim_{n \rightarrow -\infty} \frac{1}{n} \log \|(f^n)'_p v\| \neq \lim_{n \rightarrow \infty} \frac{1}{n} \log \|(f^n)'_p v\| = \log(1/3)$$

Esto muestra que el límite en (III.1.3) no existe. La misma conclusión se obtiene para cualquier $p \in (1/2, 1)$.

Por tanto vemos que los exponentes de Liapunov sólo existen en los dos puntos fijos, x_0 y x_1 : los exponentes de Liapunov son “muy raros” en cualquier sentido razonable: por simple conteo, topológicamente, con respecto a la medida de Lebesgue ... Pero cuidado! ¿Porqué tomar en cuenta la medida de Lebesgue, si ella no es invariante por f ? El punto de vista de la teoría de medida desarrollado en el capítulo II nos lleva a estudiar f con la ayuda de medidas invariantes. No es difícil encontrar todas las medidas invariantes de f (ver Ejercicio III.2). Cualquier medida invariante está soportado en los puntos fijos $\{x_0, x_1\}$. Por tanto, con respecto a cualquier medida invariante, los exponentes de Liapunov existe cada todo punto!

Es de destacar que lo antes indicado es muy general, y ese es precisamente el contenido del Teorema Multiplicativo de Oseledec¹ que por brevedad denominaremos Teorema de Oseledec. La versión de este Teorema que enunciaremos más abajo es suficiente para todas las transformaciones suaves a trozos que estudiaremos en estas notas, incluyendo los billares en los Capítulos IV y V, pero hay una versión más general que involucra *cociclos*, que no trataremos aquí.

Primero, introducimos una clase general de transformaciones suaves con singularidades.

Definición 10. Sea M una unión finita de variedades Riemannianas M_1, M_2, \dots, M_s (quizás con fronteras y ángulos), todas de la misma dimensión $d \geq 2$, pegadas a lo largo de un número finito de subvariedades C^1 de codimensión positiva. Estas subvariedades están contenidas en G , la unión de un número finito de subvariedades compactas C^1 de codimensión positiva en M_1, \dots, M_s . Entonces $V = M \setminus G$ es un conjunto abierto denso en M . Por último, sea $N \subset V$ un subconjunto abierto y $f : N \rightarrow V$ un difeomorfismo C^r , $r \geq 1$, de N en su imagen, o sea una inmersión difeomórfica de N en V . f es llamada una transformación suave con singularidades. La transformación inversa f^{-1} está bien definida en un abierto $f(N)$. Por tanto, f^{-1} es también una transformación suave con singularidades. Denotamos $H = \bigcap_{n=-\infty}^{\infty} f^n N$ al conjunto de los puntos que tienen bien definidos todos sus iterados (pasados y futuros) por f . El conjunto $S := M \setminus N$ donde f no está definida es llamado conjunto de singularidades².

¹El nombre se pronuncia *Oseledets*.

²Este nombre será extendido luego a las singularidades de f y f^{-1} .

Teorema III.2.1 (Oseledec, [Os68]). *Suponga que $f : N \rightarrow V$ preserva la probabilidad de Borel μ en M y $\mu(H) = 1$. Si*

$$\int_M \log^+ \|(f)_p'\| d\mu(p) < \infty \quad \text{y} \quad \int_M \log^+ \|(f^{-1})_p'\| d\mu(p) < \infty$$

donde $\log^+ s = \max\{\log s, 0\}$, entonces existe un conjunto f -invariante $E \subset H$, $\mu(E) = 1$, tal que para cada punto $p \in E$ todos los exponentes de Liapunov existen.

Se suele llamar puntos regulares (de Oseledec) a los puntos del conjunto E . Entonces, los exponentes de Liapunov existen casi todo punto (con respecto a cualquier medida invariante). Puesto que es estándar en teoría ergódica ignorar a los conjuntos de medida cero, podemos simplificar las cosas pensando que los “exponentes de Liapunov siempre existen”. Esto explica porqué los puntos satisfaciendo las condiciones de Anosov (A1) and (A2) en la Sección anterior se dice que son *hiperbólicos* - la existencia de los exponentes de Liapunov (casi todo punto) es garantizada por el teorema de Oseledec.

Para $p \in \Gamma$ denotamos por $\lambda_1(p) < \lambda_2(p) < \dots < \lambda_{m(p)}(p)$ los distintos exponentes de Liapunov y por $E_1(p), \dots, E_{m(p)}(p)$ los correspondientes subespacios en el espacio tangente $\mathcal{T}_p M$. Para cualquier número real $\kappa \in \mathbb{R}$ y $p \in \Gamma$ denote

$$E_{p,\kappa}^- = \oplus_{\lambda_i(p) \leq \kappa} E_i(p), \quad E_{p,\kappa}^+ = \oplus_{\lambda_i(p) > \kappa} E_i(p)$$

(esta es una generalización de los E_p^s y E_p^u de la sección previa).

OBSERVACIÓN III.2.2. El Teorema de Oseledec también incluye el siguiente hecho que enunciamos separadamente. Denótese por $\gamma_\kappa(p)$ al ángulo entre los espacios $E_{p,\kappa}^-$ y $E_{p,\kappa}^+$. Entonces

$$\lim_{n \rightarrow \pm\infty} \frac{1}{n} \log \gamma_\kappa(f^n(p)) = 0 \tag{III.2.1}$$

o sea, el ángulo $\gamma_\kappa(f^n(p))$ cambia lentamente con n (más lentamente que cualquier función exponencial).

Sea $p \in N$ un punto periódico de período k y μ la probabilidad atómica invariante soportada por el conjunto finito $\{p, f(p), \dots, f^{k-1}(p)\}$, o sea $\mu(f^i(p)) = 1/k$ para $0 \leq i < k$. Entonces las hipótesis del Teorema de Oseledec, son satisfechas y todos los exponentes de Liapunov existen en los puntos $p, f(p), \dots, f^{k-1}(p)$. Así, este teorema incluye los resultados obtenidos en la Sección previa para puntos periódicos, como caso particular.

La prueba del Teorema de Oseledec es muy difícil, y la omitimos.

La existencia de los exponentes de Liapunov lleva a la siguiente pregunta: ‘se pueden construir variedades estables e inestables como hicimos para puntos fijos en la Sección previa? En la siguiente Sección daremos una respuesta afirmativa en un contexto levemente diferente.

EJERCICIO III.2. En el ejemplo $f : S^1 \rightarrow S^1$ discutido al comienzo de esta Sección, encontrar todos las probabilidades f -invariantes en S^1 . Sugerencia: considere un pequeño intervalo $I = (a, b) \subset (0, 1/2)$ tal que $f(I) \cap I = \emptyset$ y muestre que $\mu(I) = 0$ para toda probabilidad invariante μ .

EJERCICIO III.3. En el contexto del Oseledec, probar que $m(p)$ y $\lambda_1(p), \dots, \lambda_{m(p)}(p)$ son funciones f -invariantes y $f_p'(E_i(p)) = E_i(f(p))$ para todo $1 \leq i \leq m(p)$. Esto tiene una importante consecuencia: si la medida μ es ergódica, entonces todos los exponentes de Liapunov son constantes casi todo punto.

III.3. Teoría de Pesin. Hiperbolicidad no uniforme

Ahora volvemos a la noción de hiperbolicidad introducida en Sección III.1. Recuerde que el punto p es hiperbólico si tiene todos sus exponentes de Liapunov y ninguno de ellos es igual a cero. Además, acordamos en llamar a un punto hiperbólico si satisface las condiciones de Anosov (A1) y (A2) dadas al final de la Sección III.1. Dijimos que mientras que esas condiciones técnicamente no implican la existencia de los exponentes de Liapunov, ellas garantizan que de existir son distintos de cero (y valen por lo menos λ en valor absoluto!).

Ahora es importante observar que las condiciones (A1) y (A2) son en realidad más fuertes que la mera hiperbolicidad, porque la constante K es la misma para todo $p \in M$. Ciertamente, si sólo asumimos que el

punto p es hiperbólico y ninguno de los exponentes de Liapunov está en el intervalo $(-\lambda, \lambda)$, entonces sólo podemos deducir (ver Ejercicio III.4) que para cualquier $\varepsilon > 0$ existe un $K(p, \varepsilon) > 0$ tal que parfa todo $n \geq 1$

$$\|(f^n)'_p v\| \leq K(p, \varepsilon) e^{-n(\lambda-\varepsilon)} \|v\| \quad v \in E_p^s \quad (\text{III.3.1})$$

$$\|(f^{-n})'_p v\| \leq K(p, \varepsilon) e^{-n(\lambda-\varepsilon)} \|v\| \quad v \in E_p^u \quad (\text{III.3.2})$$

donde E_p^s y E_p^u están definidos por (III.1.4).

Es importante observar que la constante $K(p, \varepsilon)$ depende, generalmente, del punto p . En otras palabras, es la *uniformidad* del factor K en las condiciones (A1)-(A2) lo que distingue a los difeomorfismos de Anosov de las transformaciones hiperbólicas más generales. Por ello, las transformaciones que satisfacen las condiciones (A1)-(A2) con constante K uniforme para todos los puntos, son llamadas *uniformemente hiperbólicas*. Las transformaciones hiperbólicas que sólo satisfacen las condiciones (III.3.1)-(III.3.2) son llamadas *no uniformemente hiperbólicas*.

Observamos que si $K(p, \varepsilon)$ es grande, entonces el efecto de contracción de vectores estables y de expansión de vectores inestables sólo puede ser observado para n suficientemente grande. Por tanto, la contracción y la expansión son sólo asintóticas y, para períodos de tiempo arbitrariamente largo puede que nada de eso ocurra. Más aún, los vectores estables pueden temporariamente expandir bajo la acción de $(f^n)'_p$. Todo esto es muy característico de los sistemas no uniformemente hiperbólicos.

Un estudio sistemático de los difeomorfismos no uniformemente hiperbólicos con medida invariante absolutamente continua fue hecho por Ya. Pesin a mediados de los setenta [Pe76, Pe77]. En los ochenta, la teoría de Pesin fue extendida a transformaciones no uniformemente hiperbólicas más generales por F. Ledrappier, A. Katok, D. Ruelle, y L.-S. Young, entre otros. Seguimos la exposición de A. Katok y J.-M. Strelcyn [KS86] que cubre una clase muy amplia de transformaciones no uniformemente hiperbólicas con singularidades. Esta clase incluye modelos de billares físicamente importantes, que estudiaremos separadamente más adelante.

Sea f una transformación suave con singularidades y M, G, V, N, S y H como en la Definición 10 de la Sección anterior. Sea μ una medida f -invariante satisfaciendo las condiciones del teorema de Oseledec. Adicionalmente requerimos que la clase de suavidad de f sea por lo menos $r \geq 2$.

Sea $d(x, S) = \inf\{\text{dist}(x, y) : y \in S\}$ la distancia de $x \in N$ a S . Para $x \in N$, denotamos por $\exp_x: \mathcal{T}_x N \rightarrow N$ la *transformación exponencial* (definida por $\exp_x(v) = \gamma(x, v, 1)$, donde $\gamma(x, v, t)$ es la geodésica en N definida por $\gamma(x, v, 0) = x$ y $\dot{\gamma}(x, v, 0) = v$). Sea $R(x, N)$ el *radio de inyectividad* de $\exp_x: \mathcal{T}_x N \rightarrow N$, o sea $R(x, N) = \sup\{r: \exp_x \text{ es uno a uno en la bola } B_r(0) \subset \mathcal{T}_x N\}$. \exp_x está definida y es inyectiva en la bola $B_{r(x, N)}(0) \subset \mathcal{T}_x N$ donde $r(x, N) = \min\{R(x, N), d(x, S)\}$.

Definimos $f_{ox} = \exp_{f(x)}^{-1} \circ f \circ \exp_x$. Esta es la manera de definir f en coordenadas lineales cerca de un punto x . Esta transformación está bien definida en un entorno de $0 \in \mathcal{T}_x N$.

Introducimos ahora dos hipótesis adicionales hechas por Katok and Strelcyn [KS86]:

(KS1) Existen constantes $C_1 > 0$ y $a > 0$ tales que para todo $\varepsilon > 0$ la μ -medida de los ε -entornos de S satisfacen

$$\mu(U_\varepsilon(S)) \leq C_1 \varepsilon^a$$

o sea, la medida μ no crece demasiado cerca del conjunto de singularidades S .

(KS2) Existen constantes $C_2 > 0$ y $b > 0$ tales que para todo $x \in N$ y $v \in \mathcal{T}_x N$, $\|v\| \leq r(x, N)$ tenemos

$$\|f''_{ox}(v)\| \leq C_2 d(\exp_x(v), S)^{-b}$$

o sea, la segunda derivada f''_{ox} no crece demasiado rápido cerca de S .

El teorema de Oseledec implica que μ -casi todo $x \in H$ tiene todos sus exponentes de Liapunov $\lambda_1(x) < \dots < \lambda_{m(x)}(x)$. El conjunto de todos los puntos hiperbólicos de N es frecuentemente llamado la *región de Pesin* f :

$$\Sigma(f) = \{x \in H : \lambda_i(x) \neq 0, \text{ for every } i = 1, \dots, m(x)\}$$

Observamos que la región de Pesin $\Sigma(f)$ es invariante por f . En cada punto $x \in \Sigma(f)$ tenemos los subespacios E_x^s and E_x^u definidos por (III.1.4); sean

$$\lambda^+(x) = \min\{\lambda_i(x) > 0\} \quad \text{y} \quad \lambda^-(x) = \max\{\lambda_i(x) < 0\}$$

los más pequeños (en valor absoluto) exponentes positivos y negativos, respectivamente.

El siguiente teorema es, de hecho, una generalización de nuestras condiciones anteriores (III.3.1), (III.3.2); ver también Ejercicio III.4.

Teorema III.3.1. *Dado $\varepsilon > 0$, existe una función medible $C(x, \varepsilon)$ en $\Sigma(f)$ tal que para todo $x \in \Sigma(f)$, $n \geq 1$ y $m \in \mathbb{Z}$*

$$\|(f^n)'_{f^m x} v\| \leq C(x, \varepsilon) e^{\lambda^-(x)n + \varepsilon n + \varepsilon |m|} \|v\| \quad v \in E_x^s$$

$$\|(f^n)'_{f^m x} v\| \geq C^{-1}(x, \varepsilon) e^{\lambda^+(x)n - \varepsilon n - \varepsilon |m|} \|v\| \quad v \in E_x^u$$

y el ángulo $\gamma(f^m x)$ entre $E_{f^m x}^s$ y $E_{f^m x}^u$ satisface $\gamma(f^m x) \geq C^{-1}(x, \varepsilon) e^{-\varepsilon |m|}$

Este teorema permite usar el mismo argumento del Teorema de Groβman-Hartman y construir variedades estables e inestables:

Teorema III.3.2. *Para μ -casi todo $x \in \Sigma(f)$ existe un $\delta(x) \in (0, r(x, N))$ tal que para todo $\varepsilon > 0$ pequeño el conjunto*

$$W^s(x) = \{y \in \exp_x B_{\delta(x)}(0) : \limsup_{n \rightarrow \infty} \frac{1}{n} \log \text{dist}(f^n(x), f^n(y)) \leq \lambda^-(x) + \varepsilon\}$$

es una variedad (estable) C^r diferenciable. De manera similar,

$$W^u(x) = \{y \in \exp_x B_{\delta(x)}(0) : \liminf_{n \rightarrow -\infty} \frac{1}{n} \log \text{dist}(f^n(x), f^n(y)) \geq \lambda^+(x) - \varepsilon\}$$

es una variedad (inestable) C^r diferenciable. También tenemos

$$\mathcal{T}_x W^s(x) = E_x^s \quad \text{y} \quad \mathcal{T}_x W^u(x) = E_x^u$$

Observamos que las variedades estables e inestables en x son transversales, o sea, se intersectan solamente en el punto x y el ángulo entre ellas es positivo. También observamos que $W^s(x)$ and $W^u(x)$ existen casi todo punto, pero no necesariamente en todos los puntos $\Sigma(f)$. Su existencia es una cuestión muy importante, por lo que daremos algunos detalles más. Estas variedades dejan de existir para los $x \in \Sigma(f)$ cuyas trayectorias se acercan demasiado rápido al conjunto de singularidades S . Más específicamente

$$d(f^{n_k} x, S) < \beta^{n_k} \quad \text{para algún} \quad \beta < e^{-\lambda_m(x)}$$

y para alguna sucesión infinita $n_k \rightarrow \infty$. Entonces $W^s(x)$ no puede existir. Si existiera, sería contraída demasiado lentamente al iterar por f y por tanto acabaría finalmente en S : para algún $n_k > 0$ tendríamos $f^{n_k}(W^s(x)) \cap S \neq \emptyset$, lo cual es imposible por la definición de $W^s(x)$ dada en el teorema anterior.

Afortunadamente nuestra hipótesis (KS1) garantiza que para casi todo punto $x \in N$ existen $c(x) > 0$ y $\kappa > 0$ tales que $d(f^n(x), S) > c(x) |n|^{-\kappa}$ para todo $0 \neq n \in \mathbb{Z}$ (ver Ejercicio III.5). Este hecho puede ser usado para deducir la existencia de $W^s(x)$ y $W^u(x)$ e incluso estimar su tamaño, omitimos otros detalles (ver [CM05], Chapter 6).

Las variedades estables e inestables pueden ser eficientemente usadas en el estudio de las propiedades ergódicas de f . Sea $B \subset H$ un conjunto f -invariante tal que $\mu(B) > 0$. Entonces la transformación $f_B := f|_B$ (la restricción de f a B) preserva la probabilidad μ_B , que se obtiene condicionando la medida μ a B . Recordamos que $f_B: B \rightarrow B$ es ergódica si y sólo si cada $g \in \mathcal{L}^1(B)$ que es f_B -invariante es constante casi todo punto de B ; cf. Sección II.2.

Un procedimiento clásico de construir un conjunto B en el cual f_B es ergódica (con respecto a la medida μ_B) usa las variedades estables e inestables por el método de E. Hopf. Tome cualquier punto $x \in \Sigma(f)$ y sus variedades estable e inestable $W^s(x)$, $W^u(x)$. Sea g una función f -invariante y suponga, por simplicidad, que es continua en M (las funciones integrables pueden ser aproximadas por continuas, pero este paso es puramente técnico, y lo omitimos). Entonces uno puede mostrar que g es constante en $W^u(x) \cup W^s(x)$, ver Ejercicio III.6. Puesto que esto se aplica a cualquier otro punto del conjunto $\Sigma(f)$ podemos proceder de la siguiente manera. Comenzamos fijando $x \in \Sigma(f)$ y construimos el conjunto B_1 (una primera aproximación

a B) como la unión de todas las variedades inestables $W^u(y)$, $y \in W^s(x)$, y todas las variedades estables $W^s(z)$, $z \in W^u(x)$. La función g es constante en B_1 .

Entonces definimos conjuntos B_n , $n \geq 2$, recursivamente por $B_n = \cup\{W^u(y) \cup W^s(y) : y \in B_{n-1}\}$. Uno puede mostrar, por inducción en n , que la función g es constante en B_n para cada n . Ponemos ahora $B_\infty = \cup_n B_n$. La función g es constante en todo el conjunto B_∞ . Por tanto es suficiente poner

$$B = \cup_{n=-\infty}^{\infty} f^n(B_\infty) \quad (\text{III.3.3})$$

y la transformación $f_B : B \rightarrow B$ será ergódica.

Describamos ahora al conjunto B_∞ . Uno puede fácilmente verificar que B_∞ consiste de todos los puntos $y \in \Sigma(f)$ para los cuales existe una sucesión finita $x = z_0, z_1, \dots, z_{k-1}, z_k = y$ con la propiedad que para todo $0 \leq i \leq k-1$ o bien $W^s(z_i) \cap W^u(z_{i+1}) \neq \emptyset$, o bien $W^u(z_i) \cap W^s(z_{i+1}) \neq \emptyset$ (claramente, todo $z_i \in B_\infty$). En otras palabras, para cada $y \in B_\infty$ hay una cadena de variedades estables e inestables que une al punto y con el punto original x . Tal cadena es normalmente llamada una *cadena de Hopf* o un *zig-zag*. Ahora podemos decir que el conjunto B_∞ es la unión de todas las cadenas de Hopf (o zig-zags) comenzando en x .

Veamos ahora cuan grande es B_∞ is. Las variedades estables e inestables, $W^s(y)$ y $W^u(y)$, son transversales entre ellas en cualquier $y \in \Sigma(f)$ (de hecho el ángulo no sólo es positivo sino alejado de cero en un entorno de y). Observe también que $\dim W^s(y) + \dim W^u(y) = \dim M$ por la hiperbolicidad. Suponga por simplicidad que todas las variedades estables e inestables $W^s(y)$, $W^u(y)$, en un entorno de x son suficientemente largas, digamos que $\text{dist}(y, \partial W^s(y)) \geq c$ y $\text{dist}(y, \partial W^u(y)) \geq c$ para alguna constante $c > 0$ y todo y cercano a x . Entonces uno puede fácilmente probar que el conjunto B_∞ contienen un entorno abierto de x en la región de Pesin $\Sigma(f)$. Este hecho es llamado *ergodicidad local* (o, a veces teorema de ergodicidad local). Volveremos a la ergodicidad local en las próximas Secciones.

En sistemas con singularidades, sin embargo, las variedades estables e inestables pueden ser arbitrariamente cortas. Esto sucede por las mismas razones que a veces dejan de existir, como hemos mostrado antes. Por tanto, B_∞ puede no cubrir ningún abierto de x en $\Sigma(f)$, o sea, pueden haber algunas islas diminutas no cubiertas tan cerca como se quiera de x . Pero es posible aún probar que B_∞ tiene μ -medida positiva. El conjunto B definido por (III.3.3) es llamado una *componente ergódica* de f . El siguiente teorema, dado sin demostración, resume nuestras discusiones:

Teorema III.3.3 (Pesin [Pe77]). *Sea $\mu(\Sigma(f)) > 0$. Existen conjuntos $\Sigma_i \subset \Sigma(f)$, $i = 0, 1, 2, \dots, J$ ($J \leq +\infty$) tales que*

(i) $\Sigma_i \cap \Sigma_j = \emptyset$ para $i \neq j$ y $\cup_i \Sigma_i = \Sigma(f)$;

(ii) $\mu(\Sigma_0) = 0$ y $\mu(\Sigma_i) > 0$ para $i > 0$;

(iii) $f(\Sigma_i) = \Sigma_i$ para $i \geq 0$;

(iv) $f|_{\Sigma_i}$ es ergódica con respecto a μ_{Σ_i} para $i > 0$.

Más aún, para todo $i > 0$ tenemos $\Sigma_i = \Sigma_{i,1} \cup \dots \cup \Sigma_{i,J(i)}$ para algún $1 \leq J(i) < \infty$ tal que

(v) $\Sigma_{i,j} \cap \Sigma_{i,k} = \emptyset$ para $j \neq k$;

(vi) $f(\Sigma_{i,j}) = \Sigma_{i,j+1}$ para $1 \leq j < J(i)$ y $f(\Sigma_{i,J(i)}) = \Sigma_{i,1}$;

(vii) la transformación $f^{J(i)}$ restringida a $\Sigma_{i,j}$ es K -mixing para cada $1 \leq j \leq J(i)$.

De acuerdo a la tradición la partición de $\Sigma(f)$ en los conjuntos $\Sigma_{i,j}$ es llamada la *descomposición espectral* de f . El teorema anterior es llamado *Teorema de descomposición espectral*. Este nombre viene de que está relacionado con los autovalores líderes (aquellos cuyos valores absolutos son iguales a uno), y sus multiplicidades, del operador unitario U_f asociado con f en $\Sigma(f)$, como se definió en la Sección I.3.

Definición 11. *Si $\mu(\Sigma(f)) = 1$, esto es, si la región de Pesin tiene medida total en N , decimos que f es no uniformemente hiperbólica, o que tiene comportamiento caótico.*

Ahora, enunciaremos otros dos importantes teoremas de la teoría de Pesin. Ellos se refieren a la entropía y a los puntos periódicos de f . Sus pruebas, aún a nivel heurístico, están más allá del objetivo de estas notas.

En la Sección I.4 vimos que la entropía es una característica numérica de los sistemas dinámicos caóticos. Representa la tasa exponencial de crecimiento de la complejidad de f^n cuando $n \rightarrow \infty$. También sabemos que los exponentes de Liapunov positivos caracterizan la tasa exponencial de separación de trayectorias por f^n cuando $n \rightarrow \infty$. Por tanto no es sorprendente que la entropía $h_\mu(f)$ esté directamente relacionada con los exponentes de Liapunov positivos de f , como se expresa a continuación:

Teorema III.3.4 (Fórmula de Pesin). *Suponga que la medida μ es absolutamente continua con respecto a la de medida de Lebesgue en N , y suponga la condición extra (KS3) que se enuncia a continuación. Entonces la entropía de f con respecto a la medida μ es*

$$h_\mu(f) = \int_H \sum_{\lambda_i(x) > 0} \lambda_i(x) \cdot \dim E_i(x) d\mu \quad (\text{III.3.4})$$

(la entropía es igual a la media de la suma de todos los exponentes de Liapunov positivos, contando sus multiplicidades).

La condición extra en ese teorema es:

(KS3) Existen $C_3 > 0$ y $c > 0$ tales que para todo $x \in N$ tenemos $\|f'_x\| < C_3 d(x, S)^{-c}$.

Esta condición fortalece un poco las hipótesis del Teorema de Oseledec, pero no mucho.

OBSERVACIÓN III.3.5. De hecho, el Teorema III.3.4 es la versión original de la fórmula de Pesin. Más tarde (III.3.4) fue extendida a medidas μ que no son absolutamente continuas en N pero cuyas medidas condicionales sobre las variedades inestables son absolutamente continuas. Lo último significa: si se toma una partición medible ξ de N tal que cada átomo (elemento) $A \in \xi$ es un subconjunto de una variedad inestable, o sea $A \subset W^u(x)$ para algún $x \in N$, entonces la medida condicional $\mu|_\xi$ es absolutamente continua con respecto a la medida de Lebesgue sobre las variedades inestables. Tales medidas son ahora llamadas *medidas de Sinai-Ruelle-Bowen*³ (o *medidas SRB*).

La fórmula de Pesin (III.3.4) implica que si $\mu(\Sigma(f)) > 0$, entonces $h_\mu(f) > 0$ para medidas absolutamente continuas (y SRB). Katok motró que este hecho es en realidad más general:

Proposición III.3.6 (Katok [Ka80]). *Suponga que la medida μ no tiene átomos ($\mu(\{x\}) = 0$ para todo punto $x \in N$). Si $\mu(\Sigma(f)) > 0$ y la restricción de f a la región de Pesin $\Sigma(f)$ es ergódica, entonces $h_\mu(f) > 0$.*

Otro teorema en la teoría de Pesin involucra a los puntos periódicos de f . Sea

$$P_n(f) = \#\{x \in N : f^n(x) = x\}$$

el número de puntos periódicos de período n (el número de puntos fijos de f^n). También denotamos por $\text{supp } \mu$ el *soporte de la medida* μ (el conjunto cerrado minimal que tiene medida total), definido por $\text{supp } \mu = \bigcap \{E : \mu(E) = 1 \text{ y } E \text{ es cerrado}\}$.

Teorema III.3.7 (Katok [Ka80]). *Si $\mu(\Sigma(f)) = 1$, entonces*

(a) *Los puntos periódicos son densos en el conjunto $\text{supp } \mu$;*

(b) $\limsup_{n \rightarrow \infty} \frac{\log P_n(f)}{n} \geq h_\mu(f)$

El Teorema de Katok III.3.7 implica que el número de puntos periódicos $P_n(f)$ crece por lo menos exponencialmente en n , y que el exponente es por lo menos la entropía de f respecto de cualquier medida invariante. Esto lleva naturalmente a plantearse el problema del máximo de esas entropías para todas las medidas invariantes. Para los difeomorfismos de Anosov definidos en la Sección III.1 un resultado preciso en este sentido se satisface:

Teorema III.3.8 (Bowen [Bo70]). *Si $f : M \rightarrow M$ es un difeomorfismo de Anosov en una variedad compacta M , entonces*

$$\lim_{n \rightarrow \infty} \frac{\log P_n(f)}{n} = h(f) := \sup_{\mu} h_\mu(f)$$

donde el supremo es tomado sobre todas las medidas f -invariantes. La cantidad $h(f)$ es independiente de la medida y llamada entropía topológica de f .

Uno podría preguntarse qué hace a los puntos periódicos tan interesantes en la teoría de los sistemas dinámicos suaves, porque en la mayoría de los casos hay sólo una cantidad numerable de puntos periódicos y su medida es nula. Por tanto, desde el punto de vista de la teoría de medida parecen no significar nada.

³Estas medidas fueron primero estudiadas por Sinai [?], Ruelle [Ru76] y Bowen [Bo75] en los setenta.

Sin embargo, los puntos periódicos juegan un papel importante en física y en los trabajos experimentales, y hay muchas otras razones por las que son tan relevantes.

Primero, las órbitas periódicas son muy simples – están completamente especificadas por un número finito de iteraciones de la transformación – lo cual hace realizables con gran exactitud los cálculos numéricos de sus exponentes de Liapunov y otras características. Segundo, ellos son relativamente fáciles de hallar en ejemplos físicamente interesantes – dado n , un programa de computadora puede hallar todos los puntos periódicos de período n . Tercero, por la parte (a) del Teorema de Katok III.3.7, cualquier órbita finita $\{f^i(x)\}$, $1 \leq i \leq n$, puede ser arbitrariamente bien aproximada por una órbita periódica (para cualquier $\varepsilon > 0$ existe un punto periódico y tal que $\text{dist}(f^i(x), f^i(y)) < \varepsilon$ para todo $1 \leq i \leq n$). La última propiedad de hecho permite aproximar cualquier medida invariante μ por medidas soportadas en órbitas periódicas. Tales aproximaciones fueron halladas por Bowen para los difeomorfismos de Anosov y más tarde han sido usadas bastante eficientemente en estudios físicos y experimentales de otros sistemas caóticos.

Para resumir, los puntos periódicos permiten estudiar muchas propiedades delicadas de los sistemas caóticos suaves. Dijo H. Poincaré in 1892: “Lo que hace a estas soluciones periódicas tan preciosas es que son, para decirlo así, la única brecha a través de la cual podemos intentar penetrar en un área que hasta ahora hemos juzgado inaccesible”. Ver [Poin], Vol 1 , #36.

Finalmente discutimos las relaciones entre las tres propiedades más importantes en la teoría de los sistemas dinámicos caóticos suaves.:

1. **Hiperbolicidad** (no anulación de todos los exponentes de Liapunov o $\mu(\Sigma(f)) = 1$);
2. **Entropía positiva**: $h_\mu(f) > 0$;
3. **Ergodicidad** (y su versión más fuerte - **mixing**).

Cada una de estas propiedades indica de alguna manera *caos* en el sistema dinámico. Aunque no hay una definición aceptada unánimemente del término *caos*, cada una de las propiedades anteriores puede ser (y a veces lo es) vista como una posible definición de caos.

Sin embargo, estas propiedades no son equivalentes y, de hecho, ninguna de ellas implica lógicamente a ninguna de las otras (sin embargo, bajo ciertas condiciones, la hiperbolicidad implica que la entropía es positiva, como lo indica la Fórmula de Pesin III.3.4 y la Proposición de Katok III.3.6). Pero, más importante para nosotros es entender que estas tres propiedades caracterizan diferentes aspectos del comportamiento caótico de los sistemas dinámicos. La entropía positiva es la característica más *local* – aún un ínfimo conjunto f -invariante $E \subset M$ con pequeña medida positiva, $\mu(E) > 0$, puede hacer que $h_\mu(f) > 0$, aunque en el resto del espacio $M \setminus E$ la transformación f pueda no ser caótica en sentido alguno. La hiperbolicidad es también una condición *local* (caracteriza la separación de trayectorias *cercanas*), pero por lo menos debe satisfacerse casi todo punto, desde que se requiere $\mu(\Sigma(f)) = 1$. Aún así, no hay garantía de que una transformación hiperbólica f sea ergódica o mixing – de acuerdo a la descomposición espectral (Teorema III.3.3), la región de Pesin $\Sigma(f)$ puede ser descompuesta en una cantidad numerable de subregiones que no interactúan entre sí!

La ergodicidad es la propiedad más *global* de las tres. Sin embargo, la ergodicidad sola no implica hiperbolicidad, o caos. Por ejemplo, la rotación del círculo $T : S^1 \rightarrow S^1$ con ángulo irracional $\phi > 0$ es ergódica; ver Ejercicio II.9 a, pero no es auténticamente caótica – su único exponente de Liapunov es obviamente nulo y por tanto también lo es su entropía (ver Ejercicio I.12).

Una observación final. La pérdida de ergodicidad en sistemas con entropía positiva y grandes regiones de Pesin es un fenómeno bastante común en sistemas Hamiltonianos. Fue ampliamente estudiado por Kolmogorov, Arnold y Moser, cuya teoría es ahora llamada *teoría KAM*. Ellos descubrieron que para un sistema Hamiltoniano típico $f : M \rightarrow M$, hay regiones f -invariantes $D \subset M$ de medida positiva donde la dinámica es muy estable. Esto significa que todos los exponentes de Liapunov son nulos en D y cada región D contiene un subconjunto de medida positiva, que es la unión de curvas o toros f -invariantes (estos son llamados *toros KAM*) – una característica que hace a la dinámica dentro de D casi integrable. Tales regiones estables ocurren alrededor de puntos periódicos $p \in M$, $f^n(p) = p$, cuando el espectro de la derivada $(f^n)'_p : T_p M \rightarrow T_p M$ está en el círculo unidad (hay algunas otras condiciones técnicas que omitimos aquí). Estos puntos periódicos son llamados *elípticos* (como opuestos a hiperbólicos), y las regiones estables alrededor de ellos son llamadas *islas elípticas* o *islas de estabilidad* (desarrollaremos un poco estas ideas en las Secciones IV.2, IV.3 donde se encontrarán además definiciones más precisas).

En un sistema Hamiltoniano típico, las islas de estabilidad coexisten con una región de Pesin $\Sigma(f)$ grande, de medida positiva. La región de Pesin es a veces llamada *mar caótico*. En ese mar, uno puede encontrar pequeñas islas estables separadas de $\Sigma(f)$ por algunas curvas f -invariantes. Tal es la figura típica que muestran muchos experimentos numéricos. El fenómeno de coexistencia de comportamiento caótico y estable en sistemas dinámicos no es aún muy bien comprendido, y es una de las cuestiones más intrigantes de la moderna físico-matemática.

EJERCICIO III.4. Deducir las condiciones (III.3.1)-(III.3.2) de la definición de exponentes de Liapunov (III.1.3). Sugerencia: primero muestre que para cualquier vector no nulo $v \in E_p^s$ (resp., $v \in E_p^u$), existe un $K(p, \varepsilon)$ que satisface (III.3.1)-(III.3.2), pero depende de v . Luego tome una base ortonormal e_1, \dots, e_k en E_p^s (resp., E_p^u), pruebe (III.3.1)-(III.3.2) con la misma constante $K(p, \varepsilon)$ para todos los vectores e_1, \dots, e_k y use la desigualdad triangular para deducir (III.3.1)-(III.3.2) para cualquier vector v .

EJERCICIO III.5. Sea $H_\kappa = \{x \in H : d(f^n(x), S) > c(x)|n|^{-\kappa} \text{ para algún } c(x) > 0 \text{ y todo } 0 \neq n \in \mathbb{Z}\}$. Mostrar que $\mu(H_\kappa) = 1$ para algún $\kappa > 0$. Sugerencia: Observe que el conjunto $\tilde{H}_{\kappa, n} := \{d(f^n(x), S) < |n|^{-\kappa}\}$ coincide con $f^{-n}(U_{|n|^{-\kappa}}(S))$, por tanto $\mu(\tilde{H}_{\kappa, n}) = \mu(U_{|n|^{-\kappa}}(S))$. Ahora ponga $\kappa = 2/a$ y use (KS1) para probar que $\sum_n \mu(\tilde{H}_{\kappa, n}) < +\infty$. Observe que H_κ consiste de puntos que pertenecen sólo a una cantidad finita de conjuntos $\tilde{H}_{\kappa, n}$.

EJERCICIO III.6. Sea g una función continua f -invariante en M . Mostrar que g es constante sobre cada variedad estable e inestable. Sugerencia: Como M es compacta, la función g es uniformemente continua. Entonces use el hecho que $f^n(W^s(x))$ y $f^{-n}(W^u(x))$ se encogen cuando $n \rightarrow \infty$.

III.4. Condiciones Suficientes para la Hiperbolicidad

En esta Sección continuamos estudiando transformaciones suaves con singularidades y usando la misma notación M, G, V, N, S, H, μ, f de las secciones anteriores. Ver Definición 10.

Describiremos condiciones suficientes bajo las cuales se da la hiperbolicidad. Estas condiciones –y las de la Sección siguiente– se han aplicado exitosamente en el estudio de muchos modelos interesantes para la física. En los Capítulos siguientes presentaremos nuestro principal ejemplo: los billares.

Formas cuadráticas. Una forma cuadrática B en \mathbb{R}^d es una función $B : \mathbb{R}^d \rightarrow \mathbb{R}$ tal que $B(u) = Q(u, u)$, donde Q es una función bilineal simétrica en $\mathbb{R}^d \times \mathbb{R}^d$. Equivalentemente, $B : \mathbb{R}^d \rightarrow \mathbb{R}$ es una forma cuadrática si existe una matriz simétrica A tal que $B(u) = u^t A u$ para $u \in \mathbb{R}^d$; aquí u^t es la (fila) traspuesta de un vector columna u . Una *forma cuadrática* B en M es una función $B : TM \rightarrow \mathbb{R}$ tal que su restricción B_x a $T_x M$ en cada punto $x \in M$ es una forma cuadrática en el sentido usual.

Decimos que una forma cuadrática B es *no degenerada* en x si $B_x(u) \neq 0$ para todo vector no nulo $u \in T_x M$ (equivalentemente, $\det A \neq 0$ para la correspondiente matriz simétrica A). Decimos que B es *positiva* (no negativa) si en cada punto x la forma B_x es positiva definida (semidefinida positiva), o sea $B_x(u) > 0$ (respectivamente, $B_x(u) \geq 0$) para todo $0 \neq u \in T_x M$. Sea $K_B(x)$ el número de valores propios positivos de la matriz que define a la form B_x (la dimensión máxima de un subespacio de $T_x M$ en el que la forma es positiva).

Si $f : M \rightarrow M$ es un difeomorfismo, sea $f^\# B$ (*pull back de B por f*) la forma cuadrática $(f^\# B)_x u = B_{f(x)}(f'_x u)$. Uno puede fácilmente verificar que $f^\# B$ es una forma cuadrática, y que $f^\# B$ es no degenerada en x si y sólo si B es no degenerada en $f(x)$. Observamos que $P = f^\# B - B$ es también una forma cuadrática.

Para una forma cuadrática B definida en la órbita de un punto x , ponemos

$$S_x := \{u \in T_x M : B((f^n)'_x u) < 0, \quad n \geq 0\}$$

$$U_x := \{u \in T_x M : B((f^n)'_x u) > 0, \quad n \leq 0\}$$

Teorema III.4.1 (Markarian [Ma94]). *Sea $B : TM \rightarrow \mathbb{R}$ una forma cuadrática no degenerada tal que*

- (i) B_x depende mediblemente de x ,
- (ii) Para cada $x \in H$ tenemos $0 < K_B(x) < \dim M$ y $K_B(f^n(x)) = K_B(x)$ para todo $n \in \mathbb{Z}$.
- (iii) $P_x = (f^\# B - B)_x$ es positiva para cada $x \in H$;

Entonces $\mu(\Sigma(f)) = 1$ (la transformación f es no uniformemente hiperbólica. Más aún, para cada $x \in \Sigma(f)$ we tenemos $S_x = E_x^s$ y $U_x = E_x^u$.

Damos un bosquejo de la prueba para el caso $\dim M = 2$ en el Ejercicio III.7.

OBSERVACIÓN III.4.2. La condición (iii) puede ser relajada: es suficiente requerir que $P \geq 0$ y P positiva eventualmente, o sea que para casi todo $x \in M$ existe $k = k(x) \in \mathbb{N}$ tal que $B((f^{k+1})'u) - B((f^k)'u) > 0$ para todo vector no nulo $u \in \mathcal{T}_x M$.

Notamos que el requerimiento $K_B(f^n(x)) = K_B(x)$ en la hipótesis (ii) se satisface automáticamente en dos casos importantes:

- (a) M tiene dimensión 2. En este caso $K_B(x) = 1$ para todo $x \in H$;
- (b) B es continua en M , y M es conexa. En este caso $K_B(x)$ es una función continua a valores enteros en M , por tanto es constante.

Estos Teoremas son versiones ergódicas del Teorema 2.1 en [Lew80] donde se probó que la existencia de una forma cuadrática continua no degenerada B satisfaciendo (iii) es equivalente a que f es Anosov. Si la forma cuadrática es continua en M , los espacios E_x^u y E_x^s dependen continuamente de x . En este caso el Teorema fue probado en [Ma88].

Obsérvese que el recíproco también se cumple:

Teorema III.4.3 (Markarian [Ma94]). *Si $\mu(\Sigma(f)) = 1$, entonces existe una forma cuadrática $B : \mathcal{T}M \rightarrow \mathbb{R}$ tal que las condiciones (i) - (iii) del Teorema III.4.1 son satisfechas.*

Conos. Sean $L \subset \mathbb{R}^d$ un subespacio, y $\alpha > 0$. Denote por L^\perp el complemento ortogonal de L . El conjunto

$$C = \{u + v : u \in L, v \in L^\perp, \|u\| \geq \alpha\|v\|\}$$

es denominado un *cono* y L su *eje*. En el caso $d = 3$ y $\dim L = 1$, es la unión de dos conos circulares simétricos con eje común L . En el caso $d = 2$ y $\dim L = 1$ el cono C es la unión de dos sectores simétricos, por lo que en este caso C también es llamado un *sector*.

Sea C un cono en \mathbb{R}^d . Note que el complemento $\mathbb{R}^d \setminus C$ es un conjunto cuya clausura es también un cono (llamado el *cono complementario*, C^-), cuyo eje es L^\perp . El conjunto

$$\partial C = \{u + v : u \in L, v \in L^\perp, \|u\| = \alpha\|v\|\}$$

es la frontera común de C y C^- .

Sea ahora un cono $C(y)$ definido en $\mathcal{T}_y M$ para casi todo $y \in M$. A la colección $\{C(y)\}$ la llamamos un *campo de conos* en M . El campo de conos es *invariante* por f si

$$f'_y(C(y)) \subset C(f(y)) \tag{III.4.1}$$

para casi todo $y \in M$. Es *eventualmente estrictamente invariante* por f si para casi todo $y \in M$ existe $n = n(y)$ tal que

$$(f^n)'_y(C(y)) \subset \text{int } C(f^n(y)) \tag{III.4.2}$$

De manera similar, se define la invariancia y la eventual invariancia estricta por f^{-1} , reemplazando f' por $(f^{-1})'$ en (III.4.1) y $(f^n)'$ por $(f^{-n})'$ en (III.4.2).

Decimos que un campo de conos $\{C^u(x)\}$ es *inestable* si es invariante y eventualmente estrictamente invariante por f y para casi todo $x \in M$

$$\|f'_x(u)\| > \|u\| \quad \text{para todo } u \in \text{int } C^u(x) \tag{III.4.3}$$

Decimos que el campo de conos $\{C^s(x)\}$ es *estable* si es invariante y eventualmente estrictamente invariante por f^{-1} y para casi todo $x \in M$ $\|(f^{-1})'_x(u)\| > \|u\|$ for all $u \in \text{int } C^s(x)$.

Para un campo de conos $\{C^s\}$ definido en la órbita de un punto x , ponemos

$$S_x := \bigcap_{n \geq 0} (f^{-n})'_x [C^s(f^n x)]$$

Para un campo de conos inestable $\{C^u\}$ definido sobre la órbita de un punto x , ponemos

$$U_x := \bigcap_{n \geq 0} (f^n)'_x [C^u(f^{-n} x)]$$

Teorema III.4.4. Sea $\{C^u(x)\}$ un campo de conos inestable y $\{C^s(x)\}$ un campo de conos estable, ambos dependiendo mediblemente de x . Sea $\dim L^u(x) + \dim L^s(x) = \dim M$ para casi todo x , donde $L^u(x)$ y $L^s(x)$ son los ejes de $C^u(x)$ y $C^s(x)$, respectivamente. Entonces $\mu(\Sigma(f)) = 1$. Más aún, para casi todo $x \in \Sigma(f)$ tenemos $S_x = E_x^s$ y $U_x = E_x^u$.

La prueba de este Teorema es bosquejada en el Ejercicio III.8.

Podemos relajar el requerimiento de la eventual estricta invariancia del campo de conos si, en su lugar, asumimos la expansión (III.4.3) para todos los vectores $u \in C^u(x)$ y la expansión hacia atrás para todos los vectores $u \in C^s(x)$.

El Teorema III.4.4 y sus varias versiones han sido usadas en el estudio de modelos físicos desde el comienzo de los setenta. Un toque elegante de esta técnica fue dado por M. Wojtkowski en [W85, W86]:

Teorema III.4.5 (Wojtkowski [W85]). Sea $\dim M = 2$ y la medida invariante μ absolutamente continua con respecto a la medida de Lebesgue en M . Si existe un campo de conos $C(x)$ invariante y eventualmente estrictamente invariante en M que depende mediblemente en x , entonces $\mu(\Sigma(f)) = 1$.

En este Teorema no se hace ningún requerimiento sobre la expansión (o contracción) de los vectores del cono $C(x)$! La expansión de los vectores $u \in C(x)$ por $(f^n)_x'$ para $n > 0$ grande se deduce de la invariancia eventual estricta invariancia del campo de conos y de la continuidad absoluta de la medida invariante. Wojtkowski y Liverani [LW95] también generalizaron el Teorema III.4.5 a $\dim M > 2$, pero una estructura especial adicional – una forma simpléctica f -invariante – es necesaria. Omitimos la consideración de este caso.

EJERCICIO III.7. Bosquejo de la prueba del Teorema III.4.1 en dimensión $d = 2$. Sugerencias:

(a) Sea $x \in H$ y ponga $x_n = f^n(x)$ para $n \in \mathbb{Z}$. Para cada $n > 0$, tome un vector $w_n \in T_{x_n}M$ tal que $B_{x_n}(w_n) < 0$. Ponga $v_n = (f^{-n})_{x_n}' w_n$. Como $P > 0$, resulta que $B_x(v_n) < 0$. Entonces la sucesión de vectores normalizados $v_n/\|v_n\|$ tiene una subsucesión que converge a un vector unitario $v \in T_x M$. Ahora se puede verificar que $v \in S_x$;

(b) De manera similar construimos un vector unitario $u \in U_x$. Por tanto, ambos subespacios U_x y S_x existen, y por tanto deben ser de dimensión uno;

(c) Dado $a > 0$, denote por $D_a \subset M$ al conjunto de puntos $y \in M$ tales que $a B_y(w) \leq \min\{\|w\|^2, P_y(w)\}$ para todo $w \in T_y M$. Si a es suficientemente pequeño, entonces $\mu(D_a) > 0$, y $\lim_{a \rightarrow 0} \mu(D_a) = 1$;

(d) Por el Teorema Ergódico de Birkhoff-Khinchin aplicado a la función indicatriz I_{D_a} (ver Ejercicio II.4b), para casi todo punto $x \in D_a$ tenemos $\lim_n \tau(x, D_a, n)/n = \tau(x, D_a) > 0$, donde $\tau(x, D_a, n) = \#\{0 \leq j < n : f^j(x) \in D_a\}$;

(e) Como P es positivo tenemos que $B((f^n)_x' u) \geq (1+a)^{\tau_n} B(u)$ para $u \in U_x$, donde $\tau_n = \tau(x, D_a, n)$. Entonces, $\liminf_n n^{-1} \log \|(f^n)_x' u\| \geq \log(1+a)\tau(x, D_a)/2 > 0$. Por lo tanto, x tiene un exponente de Liapunov positivo y $E_x^u \supset U_x$ casi todo punto en D_a , por tanto en M ;

(f) De manera similar, $E_x^s \supset S_x$ casi todo punto.

EJERCICIO III.8. Siguiendo las líneas del Ejercicio anterior, pruebe el Teorema III.4.4. Sugerencia: debido a la eventual invariancia estricta del campo de conos $\{C^u(x)\}$, para casi todo $x \in M$ existe un $k = k(x)$ tal que $(f^k)_{f^{-k}x}'(C^u(f^{-k}x)) \subset \text{int } C^u(x)$; por lo tanto, hay un $a = a(x) > 0$ tal que para cada $u \in (f^k)_{f^{-k}x}'(C^u(f^{-k}x))$ tenemos $\|f_x'(u)\| \geq (1+a)\|u\|$.

III.5. Condiciones Suficientes para la Ergodicidad

Ahora suponemos que f es hiperbólica ($\mu(\Sigma(f)) = 1$), y describiremos condiciones suficientes bajo las cuales f es ergódica. Como se vió en el Teorema III.3.3, esto es generalmente falso –la región de Pesin $\Sigma(f)$ se descompone en un número finito o numerable de subregiones Σ_i en las cuales la restricción de f es ergódica. Los conjuntos Σ_i son llamados *componentes ergódicas* de f . Generalmente, pueden ser bastante complicadas y disconexas.

Bajo ciertas condiciones adicionales, uno puede probar que f es ergódica, o sea que hay exactamente una componente ergódica $\Sigma_1 \subset \Sigma(f)$ y $\mu(\Sigma_1) = 1$. Esto será hecho en dos etapas. Primero daremos condiciones que son suficientes para probar un “teorema ergódico local” (esta es la parte más difícil) y después agregaremos otras condiciones para probar la ergodicidad global. Presentaremos dos versiones del teorema ergódico local. Sus pruebas están mucho más allá de los propósitos de estas notas, pero ambas siguen las ideas de los trabajos de Sinai [Si70], Sinai and Chernov [SC87], mejoradas por Krámli, Simányi and Szász [KSS90]

Sea M contenida en una variedad Riemannian de dimensión $d \geq 2$, dividida en una cantidad finita de dominios compactos, de dos maneras:

$$M = B_1^+ \cup \dots \cup B_r^+ = B_1^- \cup \dots \cup B_r^-$$

de modo que $\text{int } B_i^\pm$ es conexo y denso en B_i^\pm , y que las fronteras ∂B_i^\pm están formadas por una cantidad finita de subvariedades de dimensión $d - 1$ que se superponen en a lo sumo sus fronteras. Estas fronteras son una unión finita de subvariedades compactas de dimensión $d - 2$. La transformación f es definida separadamente sobre cada dominio B_i^\pm , $1 \leq i \leq r$, de modo que f es un difeomorfismo C^r ($r \geq 2$) del interior de B_i^+ sobre el interior de B_i^- y un homeomorfismo de B_i^+ sobre B_i^- . Entonces tenemos $N = \cup_i \text{int } B_i^\pm$, $f(N) = \cup_i \text{int } B_i^\mp$ y en cada componente conexa de N f puede ser extendida continuamente a la frontera ∂N , y en cada componente conexa de $f(N)$ la transformación inversa f^{-1} puede ser extendida continuamente a la frontera $\partial f(N)$. Como las fronteras de B_i^\pm pueden tener puntos comunes, la función f puede ser multivaluada en ∂B_i^\pm .

Suponemos que μ es una probabilidad f -invariante absolutamente continua respecto de la medida Riemanniana de M , con densidad acotada.

Denotemos $S^+ = \partial N$ el conjunto de singularidades de f y $S^- = \partial f(N)$ el conjunto de singularidades de f^{-1} . Para $n \geq 1$ ponemos $S_n^+ = \cup_{i=0}^{n-1} f^{-i}(S^+)$, que es el conjunto de singularidades de f^n , y de manera similar $S_n^- = \cup_{i=0}^{n-1} f^i(S^-)$, el conjunto de singularidades de f^{-n} . Observar que todos estos conjuntos son variedades compactas de dimensión $d - 1$.

Transformaciones como la de arriba son llamadas *suaves de a trozos*. Compare con M, f , etc. en la Sección III.2. Veremos más ejemplos de este tipo en el Capítulo siguiente.

Por razones técnicas nos restringimos a $\dim M = 2$. Por tanto, los conjuntos S^\pm, S_n^\pm son de dimensión uno, son curvas suaves compactas. Entonces podemos concebir B_i^\pm como superficies de dimensión dos (quizás con fronteras y vértices).

Ergodicidad Local para transformaciones hiperbólicas suaves de a trozos (primera versión).

Comenzamos dando la versión del teorema ergódico local presentada en [C93]. Hemos supuesto que f es hiperbólica ($\mu(\Sigma(f)) = 1$). Por tanto, en casi todo punto $x \in M$ los subespacios estable e inestable E_x^s y E_x^u existen (ambos tienen dimensión uno).

Por simplicidad, suponemos que los vectores estables e inestables son expandidos de modo monótono:

$$\|f'_x(u)\| \geq \|u\|, \quad u \in E_x^u \quad \text{y} \quad \|(f^{-1})'_x(u)\| \geq \|u\|, \quad u \in E_x^s \quad (\text{III.5.1})$$

Frecuentemente es posible definir la métrica Riemanniana de M de modo que se satisfaga (III.5.1) (tal métrica se llama adaptada o de Liapunov, cf. Observación III.1.1 al final de la Sección III.1). En algunos casos es suficiente una pseudo-métrica que satisfaga (III.5.1). Mostraremos esto en el Capítulo IV. Por ahora, suponemos que una métrica adaptada existe. En la segunda versión usaremos explícitamente una pseudo-métrica que es una métrica si se restringen los vectores a ciertos conos.

Definición 12. *Un punto $x \in N$ es u-esencial si para todo $A > 1$ existe un $n \geq 1$ y un entorno U del punto x tal que para todo $y \in U$*

$$\|(f^n)'_y(u)\| > A\|u\|, \quad u \in E_y^u$$

(siempre que el espacio E_y^u exista). De manera similar, se definen puntos s-esenciales (reemplazando f^n por f^{-n} y E_y^u por E_y^s).

Definición 13. *Un punto $x \in N$ es SC-suficiente si existen $A > 1$ y dos enteros $n < m$, tales que está definido $f^n(x)$ y hay un entorno U del punto $f^n(x)$ tal que para todo $y \in U$ tenemos*

$$\|(f^{m-n})'_y(u)\| > A\|u\|, \quad u \in E_y^u \quad \text{y} \quad \|(f^{m-n})'_y(v)\| < A^{-1}\|v\|, \quad v \in E_y^s$$

(siempre que los espacios E_y^u, E_y^s existan).

Ahora enunciamos cinco condiciones suficientes para probar la ergodicidad:

Propiedad 1 (“singularidades dobles”). Para cualquier $n \geq 1$ la intersección $S_n^+ \cap S_n^-$ es un conjunto finito o numerable.

Propiedad 2 (“delgadez de los entornos de las singularidades”). Para cualquier $\delta > 0$ sea U_δ el δ -entorno del conjunto $S^+ \cup S^-$. Entonces $\mu(U_\delta) \leq \text{const} \cdot \delta$.

Propiedad 3 (“continuidad”). Las familias de subespacios estables e inestables E_x^s y E_x^u son continuas en sus dominios ⁴. Más aún los espacios $\lim_{y \rightarrow x} E_y^u$ y $\lim_{y \rightarrow x} E_y^s$ son siempre transversales entre si en cada punto suficiente x , aún si E_x^u or E_x^s no existen.

Propiedad 4 (“ansatz”). Casi todo punto de S^+ (con la longitud de Lebesgue) es u-esencial, y casi todo punto de S^- es s-esencial.

Propiedad 5 (“transversalidad”). En casi todo punto $x \in S^+$ los subespacios $E^u(x)$ están definidos y son transversales a S^+ , y en casi todo punto $x \in S^-$ los subespacios $E^s(x)$ están definidos y son transversales a S^- .

Teorema III.5.1 (Teorema Ergódico Local 1). *En las condiciones de a-rrriba, todo punto suficiente x tiene un entorno abierto U que pertenece a una componente ergódica (hay una componente $\Sigma_i \subset \Sigma(f)$ tal que $\mu(U \setminus \Sigma_i) = 0$).*

La primera versión de este teorema fue obtenida por Sinai [Si70] al estudiar los billares matemáticos. Luego fue mejorada por Sinai y Chernov [SC87]. Más tarde fue generalizada por Krámlí, Simányi y Szász [KSS90], Liverani y Wojtkowski [LW95], y Markarian [Ma93b] a otras clases de sistemas hiperbólicos de origen físico. En su forma más general fue enunciado por Chernov [C93]. ⁵

Ergodicidad Local para transformaciones hiperbólicas suaves de a trozos (segunda versión). Damos ahora la versión del teorema probada por Liverani y Wojtkowski [LW95]. Sigue las líneas del Teorema III.4.5.

Sean $X_1(z), X_2(z)$ vectores no nulos en los lados del cono $C(z) = \{u \in \mathcal{T}_z M : u = aX_1(z) + bX_2(z), ab \geq 0\}$. Para cada cono $C(z)$, hay una forma cuadrática asociada dada por $Q_z(u) = A(X_1(z), X_2(z))ab$, donde $A(\cdot, \cdot)$ es la forma de volumen generada por la métrica α (ver Condición 4) que se esté utilizando.⁶ La cantidad

$$\sigma(D_z f) = \inf_{u \in \text{int}(C(z))} \sqrt{\frac{Q_{fz}(D_z f u)}{Q_z(u)}}.$$

mide la cantidad de expansión por $D_z f$. De manera similar definimos σ para f^{-1} reemplazando $C(z)$ por su cono complementario.

Definición 14. *Un punto $z \in M$ es llamado LW-suficiente si existe un $n > 0$ tal que $z \notin S_n^+ \cap S_n^-$ y $\sigma(D_z f^n) > 3$ o $\sigma(D_z f^{-n}) > 3$.*

Para aplicar el Teorema Ergódico Local necesitamos que se satisfagan las siguientes condiciones:

1. (Monotonía) El campo de conos $\{C(z)\}$ es eventualmente estrictamente f -invariante, cada caja B_i^+ se puede descomponer en una cantidad finita de sub-cajas tal que la restricción de C al interior de cada sub-caja de B_i^+ es continua.⁷
2. (Alineamiento) El espacio tangente a S^- en cualquier punto de $z \in S^-$ está estrictamente contenido en $C(z)$, y el espacio tangente a $z \in S^+$ está estrictamente contenido en el cono complementario $C'(z)$.
3. (Regularidad) Para todo $n \geq 1$, los conjuntos S_n^+ y S_n^- son regulares (uniones finitas de curvas cerradas suaves que sólo se intersectan en sus extremos).
4. (No contracción) Sea α una métrica Riemanniana en M que puede degenerar sobre ∂M ; las normas y longitudes generadas por α son $\|\cdot\|', \mathcal{L}'$, respectivamente. Existe una constante $\rho > 0$ tal que para todo $n \geq 1$ y todo $z \in M \setminus S_n^+$

$$\|D_z f^n u\|' \geq \rho \|u\|'$$

para todo $u \in C(z)$.

⁴Esa condición es automáticamente satisfecha si la dependencia de la forma cuadrática o del campo de conos en $x \in H$ es continua.

⁵[C93] sólo contiene bosquejos de las pruebas; una exposición completa de este teorema se puede encontrar en [KSS90]

⁶Dado que las medidas utilizadas son absolutamente continuas respecto de la Euclídiana en \mathbb{R}^2 , lo normal será que $A(X_1(z), X_2(z)) = (\text{Área Euclídiana del paralelogramo determinado por } X_1(z), X_2(z)) \times (\text{Valor medio de la densidad de la medida respecto de la medida Euclídiana})$

⁷En realidad, la cantidad puede ser numerable, pero en este caso, la definición de σ se debe hacer de modo algo más complicada; este caso no está incluido en la presentación de [LW95]. Ver [DMM05].

5. (Sinai-Chernov Ansatz) Sea μ_S la longitud Riemanniana en $S^- \cup S^+$. Para μ_S -c.t $z \in S^-(S^+)$

$$\lim_{n \rightarrow +\infty(-\infty)} \sigma(D_z f^n) = +\infty.$$

6. (Estimaciones de volumen) Existe una constante positiva C , tal que para cada $Y \subset S^-(S^+)$ cerrado respecto de la topología relativa de $S^-(S^+)$

$$\limsup_{\delta \rightarrow 0^+} \frac{\nu(Y_{u(s)}^\delta)}{\delta} \leq C \cdot (\text{longitud euclidiana de } Y),$$

donde $Y_{u(s)}^\delta = \{z \in M : \text{existen } \gamma \subset W^{u(s)}(z) \text{ tales que } \gamma \cap Y \neq \emptyset, \mathcal{L}'(Y) < \delta\}$.

Teorema III.5.2 (Teorema Ergódico Local 2). *Si las Condiciones 1-6 son satisfechas, entonces todo punto suficiente en una caja B_i^+ de M tiene un entorno en B_i^+ que pertenece (mod 0) a una componente ergódica de f .*

Como $\{C\}$ es eventualmente estrictamente f -invariante, el conjunto de puntos suficientes de M tiene μ -medida total [LW95, §7F]. Esta versión generalizada, con respecto a la presentación de [LW95], está probada en [DMM05].

La idea de la prueba de ambas versiones está esencialmente descrita en la Section III.3: dado un punto suficiente x , uno debe conectarlo con casi todo punto $y \in U$ por una cadena de Hopf. En otras palabras, la unión de las variedades estables e inestables U debe estar en una componente conexa por arcos de medida total en U .

Sin embargo, uno debe recordar que las variedades estables e inestables (y, por tanto, las cadenas de Hopf) pueden ser arbitrariamente cortas en M debido a las singularidades de f . De hecho, la unión de todos los conjuntos de singularidades $\cup_n S_n^\pm$ es denso en M , por lo que las variedades estables e inestables se pueden quebrar o terminar abruptamente en cualquier parte. Este es el problema más serio en la prueba de los Teoremas III.5.1, III.5.2, que fue resuelto por primera vez por Sinai en el contexto de billares. Él llamó a su versión de estos teoremas *Teorema Fundamental en la Teoría de Billares*. Este Teorema de Ergodicidad Local sigue siendo hoy uno de los resultados más avanzados y difíciles en la teoría de de las transformaciones hiperbólicas con singularidades.

Ahora procedemos a la prueba de la ergodicidad de f . El Teorema de Ergodicidad Local muestra que las componentes ergódicas Σ_i consisten de bolas abiertas alrededor de los puntos suficientes, pero nada dicen de sus tamaños. Ahora agregamos la siguiente hipótesis:

Propiedad (“abundancia de puntos suficientes”). El conjunto $R \subset M$ de los puntos suficientes tiene medida total y existe una trayectoria finita $\{f^j x_0\}, 0 \leq j \leq n$ que pasa por todas las componentes conexas por arco de R .

Teorema III.5.3 (Teorema Ergódico Global). *Si se cumplen todas la condiciones de arriba ⁸ f is ergódica.*

Demostración. Supongamos primero que x, y están en una componente conexa de R . Por la Propiedad 6, existe una curva continua, $\gamma \subset R$ que une x e y . Por el Teorema Ergódico Local, para cada punto $z \in \gamma$ existe un entorno $U(z)$ que pertenece a Σ_p (que aún puede depender de z). Como γ es compacta, puede ser cubierta por un número finito de tales entornos: existen z_1, \dots, z_k tales que $\gamma \subset \cup_{i=1}^k U(z_i)$. Los entornos $U(z_i)$ se sobreponen por lo que ellos pertenecen a una componente ergódica Σ_p , la cual por tanto contiene a x and y . Por tanto cada componente conexa de R forma una componente ergódica de f . Además, si dos puntos $x, y \in R$ están en distintas componentes conexas por arco, sean x_1, x_2 los puntos de la trayectoria referida en la Propiedad 6 que pertenecen a cada una de ellas, entonces $f^j(U(x_1))$ intersecta a $U(x_2)$, para algún j , por lo que ambos entornos están en la misma componente ergódica y todo R está en la misma componente ergódica. Como R tiene medida total, Σ_p es la única componente ergódica y $\mu(\Sigma_p) = 1$. \square

⁸En cualquiera de las dos versiones del teorema de Ergodicidad local.

En la práctica, uno tiene que verificar las propiedades 1-6 para asegurar la ergodicidad de una transformación hiperbólica. Esto puede no ser muy fácil, pero fue hecho exitosamente para muchos sistemas físicos interesantes.

Finalmente observamos que como consecuencia de la teoría de Pesin, extendida a los sistemas suaves con singularidades por Katok y Strelcyn, las componentes ergódicas Σ_1 pueden ser partidas en una cantidad finita de conjuntos medibles disjuntos $\Sigma_{1,j}$, $1 \leq j \leq J$ tales que $f(\Sigma_{1,j}) = \Sigma_{1,j+1}$ para $1 \leq j < J$ y $f(\Sigma_{1,J}) = \Sigma_{1,1}$ y f^J restringida a $\Sigma_{1,j}$ es K - mixing para cada $1 \leq j \leq J$. Ver Teorema III.3.3.

Por tanto, si f^n es ergódica para cada $n \geq 1$, entonces f es K-mixing, y como consecuencia de un resultado general de Chernov y Haskell [CH96] (u Ornstein y Weiss [OW98]) es Bernoulli; ver Sección II.3.

Capítulo IV

Billares. Propiedades generales. Ejemplos.

Aquí llegamos al principal objetivo de estas notas. Estudiaremos una clase de sistemas dinámicos llamados billares. Ellos ilustran todos los conceptos y fenómenos discutidos en los capítulos anteriores: medidas invariantes, ergodicidad y mixing, exponentes de Liapunov y singularidades. El estudio de los billares matemáticos es especialmente importante porque tienen cantidad de aplicaciones en física.

La popularidad de los billares entre los matemáticos y los físicos es también debida a la relativa simplicidad de las reglas que definen la dinámica de los billares. Como expresara A. Katok, los billares son un patio de recreo para los investigadores. Aquí uno puede jugar toda clase de juegos y todos pueden divertirse ...

IV.1. Billares Planos. Generalidades

Imagine una partícula (una masa puntual) moviéndose libremente (sin fricción) en una mesa y chocando sobre los bordes de la mesa elásticamente. La mesa es sencillamente una región acotada del plano.

Con cierta amplitud, este modelo recuerda el popular juego de billares (o pool) donde unas pocas bolas son golpeadas con un taco de modo de hacer carambolas (choques sucesivos de varias bolas luego de un sólo golpe) o mandar algunas de ellas en unos agujeros con redes ubicados sobre los bordes. Pero nuestro modelo luce mucho más simple – tenemos una sola bola, la bola es un punto “sin dimensión”, se mueve sin fricción, y la mesa no tiene agujeros, sólo bordes¹. Por otro lado la forma de la mesa no es necesariamente rectangular² Puede ser más bien arbitraria, y esto es lo que hace al modelo realmente interesante – la forma de la mesa es la que determina el carácter de la dinámica y todas sus propiedades.

Mesa del billar. Sea $Q \subset \mathbb{R}^2$ un conjunto abierto, acotado, conexo, que llamaremos *mesa del billar*. Su frontera $\Gamma = \partial Q$ es la unión finita de curvas compactas suaves (C^k , $k \geq 3$):

$$\partial Q = \Gamma = \Gamma_1 \cup \dots \cup \Gamma_s \tag{IV.1.1}$$

que son disjuntas pero pueden tener extremos comunes. Por ejemplo, si Q es un polígono, entonces Γ_i son sus lados.

Convención y definición. Para fijar los signos de la curvatura K de la frontera, establecemos que $K(r) \geq 0$ si ∂Q es cóncava (convexa para dentro) en r (como en la Figura IV.2), $K(r) \leq 0$ si ∂Q es cóncava para adentro. Una componente suave de la frontera $\Gamma_i \subset \partial Q$ es *dispersora*, si su curvatura es positiva, *focalizadora* si su

¹Los billares matemáticos con agujeros son mucho más complicados; ver por ejemplo [CM97b, CMT00]. Hay pocos estudios matemáticos de modelos “más cercanos” a los billares “reales”; ver [BG93]. Existe un viejo (1835) y entretenido libro de Coriolis [Cor35] en que se estudia el juego de billar. En su Prefacio se dice que “las personas que tengan conocimientos de mecánica racional, como los alumnos de la *École Polytechnique*, verán con interés la explicación de todos los efectos singulares que se observa en el movimiento de las bolas.” “Debo a la gentileza de M. Mingaud –un célebre jugador– haber podido asegurarme, viéndolo jugar, que las fórmulas y construcciones que se deducen dan resultados conformes a la experiencia.”

²En los billares poligonales se estudian otras propiedades y en general con métodos diferentes. Ver [Gu86, Gu96]

curvatura es negativa y *neutra* si su curvatura es cero. Si la curvatura de Γ_i cambia de signo, dividimos la curva en partes menores cuyas curvaturas tienen signo constante. Por tanto, suponemos que la frontera ∂Q consiste de componentes de esos tres tipos – dispersoras, focalizadoras and neutras. Denote $\Gamma_+, \Gamma_-, \Gamma_0$ la unión de los arcos dispersores, focalizadores y de las líneas rectas de Γ , respectivamente. Es muy importante observar que si bien la descripción y algunos de los resultados generales sobre billares son válidos en las condiciones arriba indicadas, la mayoría de los resultados sobre propiedades ergódicas valen con componentes dispersoras y focalizadoras de curvaturas estrictamente mayores y menores que cero

El conjunto $\Gamma^* = \partial\Gamma_1 \cup \dots \cup \partial\Gamma_s$ de los extremos de nuestras curvas frontera jugará un papel muy importante. Sea Γ^* el conjunto de los vértices de la mesa Q .

La partícula móvil tiene posición $q \in Q$ y vector de velocidad $v \in \mathbb{R}^2$, que dependen del tiempo. Si $q \in Q$, entonces la partícula se mueve libremente:

$$\dot{q} = v \quad \text{y} \quad \dot{v} = 0 \tag{IV.1.2}$$

donde los puntos indican las derivadas con respecto al tiempo. Podemos suponer que la partícula tiene masa uno, por lo que su momento es $p = v$, y las ecuaciones de arriba son Hamiltonianas, con función Hamiltoniana $H(p, q) = \|p\|^2/2$. Cuando $q \in \partial Q$, la velocidad v de la partícula cambia discontinuamente, de acuerdo con la clásica regla *el ángulo de incidencia es igual al de reflexión*. Por tanto, el nuevo vector (de salida) v_+ está relacionado con el viejo (entrante) v_- por

$$v_+ = v_- - 2\langle v_-, n(q) \rangle n(q) \tag{IV.1.3}$$

Aquí $\langle \cdot, \cdot \rangle$ es el producto escalar y $n(q)$ es el vector unitario de la normal entrante a la frontera ∂Q en el punto q .

Para usar la regla de reflexión (IV.1.3), necesitamos la normal entrante $n(q)$, por lo que la regla no se puede usar en puntos de $q \in \Gamma^*$, donde ese vector deja de existir (o no está unívocamente determinado). Para ser preciso uno debe definir $n(q)$ con continuidad en los puntos de Γ^* , pero esto daría más de una normal $n(q)$ en todos los vértices donde dos curvas diferentes $\Gamma_i, \Gamma_j \subset \Gamma$ se encuentran con ángulo distinto de π , por lo que la dinámica estaría definida múltiplemente. Y esto no es bueno. Más aún, adoptamos la convención estándar de que la reflexión no está definida en ningún $q \in \Gamma^*$. Por tanto, cuando la partícula choca en un punto de Γ^* (llamado *puntos singulares de la frontera*), ella se para y la trayectoria deja de existir. Obsérvese que este criterio general, elimina trayectorias donde la evolución es continua -cuando dos arcos de la frontera se cortan y tienen la misma tangente interior (un arco de círculo y su tangente, por ejemplo) o cuando dos segmentos de recta se cortan perpendicularmente (ver Ejercicio IV.1). La razón de esta convención es que muchos resultados necesitan que la transformación sea $C^{1+\varepsilon}$ (ver Capítulo anterior) $C^{1+\varepsilon}$ y para esto se necesita que la curva sea $C^{2+\varepsilon}$ (ver Lemma IV.1.1)

Ejemplo. Defina el dominio Q en coordenadas polares ρ, θ por $Q = \{0 < \rho < 1, \theta \neq 0\}$. Este es el círculo unitario con un segmento radial eliminado. En este caso $\partial Q = \{\rho = 1\} \cup \{\theta = 0\}$ consiste de dos curvas suaves, y Γ^* consiste de dos puntos: el origen $\rho = 0$ y el punto $(\rho = 1, \theta = 0)$. Observe que la línea $\theta = 0$ es una “frontera interior” – las trayectorias del billar la chocan de los dos lados.

Hay una razón más por la que la dinámica puede no estar definida. Ésta es bastante peculiar y poco intuitiva. Si la partícula experimenta infinitas reflexiones en un intervalo finito de tiempo, entonces su movimiento no puede ser definido más allá del punto donde las reflexiones se acumulan. Por extraño que parezca, esto puede suceder, sea cerca de un vértice de la mesa o aún donde ∂Q es enteramente suave (Halpern, 1977). Discutiremos los efectos de este extraño fenómeno más adelante.

Espacio de fase. Ahora construimos el espacio de fase de nuestro sistema. La dinámica (IV.1.2)-(IV.1.3) preserva la norma $\|v\|$, por lo que podemos establecer $\|v\| = 1$. Entonces el *espacio de fase* (del flujo) será $\mathcal{M} = \bar{Q} \times S^1$ donde \bar{Q} es la clausura de la mesa de billar Q , y S^1 es el círculo unitario de todas los vectores velocidad ³. Para cada $q \in \partial Q$ los puntos (q, v_-) y (q, v_+) relacionados por (IV.1.3) son identificados

³En la literatura, una mesa de billar Q es a veces definida como la clausura de un conjunto abierto. Hay un pequeño problema con esa definición puesto que no permitiría que ∂Q tuviera líneas “interiores” como la que hemos visto en el ejemplo de arriba.

4. El grupo a un parámetro de transformaciones (flujo) en \mathcal{M} es denotado por Φ^t , donde $t \in \mathbb{R}$ es el tiempo (y la longitud recorrida puesto que la norma de la velocidad es uno).

Es estándar en teoría ergódica reducir el estudio de flujos a transformaciones construyendo una sección transversal. Para el flujo Φ^t , una hipersuperficie natural en \mathcal{M} puede ser construida muy naturalmente con la ayuda de la frontera de la mesa Q . Sea

$$M = \{x = (q, v) \in \mathcal{M}: q \in \partial Q, \langle v, n(q) \rangle \geq 0\}. \quad (\text{IV.1.4})$$

Esta es una subvariedad de dimensión dos en \mathcal{M} . Consiste de todas las posibles velocidades salientes resultantes de las reflexiones en ∂Q . Claramente, cualquier trayectoria del flujo Φ^t cruza la superficie M cada vez que se refleja en ∂Q . Esto define la *transformación de primer retorno* $T: M \rightarrow M$ por $Tx = \Phi^{\tau(x)+0}x$ donde

$$\tau(x) = \min\{t > 0 : \Phi^{t+0}x \in M\} \quad (\text{IV.1.5})$$

La transformación T es frecuentemente llamada *transformación de billar -billiard map-* o *transformación de colisión -collision map*. M es denominado *espacio de fase de la transformación de billar* o *espacio de colisión*⁵.

Denote $\pi : M \rightarrow \Gamma$ la proyección canónica de M sobre $\Gamma = \partial Q$, y por ρ la distancia Euclidiana en \mathbb{R}^2 .

El conjunto M es una unión finita de variedades compactas suaves M_i (con frontera) pegadas a lo largo de sus fronteras (remarcamos que M_i incluye al conjunto de puntos que están saliendo de vértices y apuntan hacia adentro de Q). Está dividido en tres subconjuntos, correspondientes a las partes dispersoras, focalizadoras o neutras de la frontera: $M = M_+ \cup M_- \cup M_0$ con

$$M_+ = \bigcup_{i:\Gamma_i \subset \Gamma_+} M_i, \quad M_- = \bigcup_{i:\Gamma_i \subset \Gamma_-} M_i, \quad M_0 = \bigcup_{i:\Gamma_i \subset \Gamma_0} M_i.$$

Introducimos coordenadas r, ϕ en M , donde r es el parámetro longitud de arco en ∂Q y $\phi \in [-\pi/2, \pi/2]$ es el ángulo entre v y $n(q)$. Estas coordenadas están orientadas como lo indica la Figura IV.1. El punto de referencia en ∂Q (donde $r = 0$) puede ser elegido arbitrariamente (si ∂Q tuviera más de una componente conexa habría que tomar orígenes de la longitud de arco en cada una de ellas).

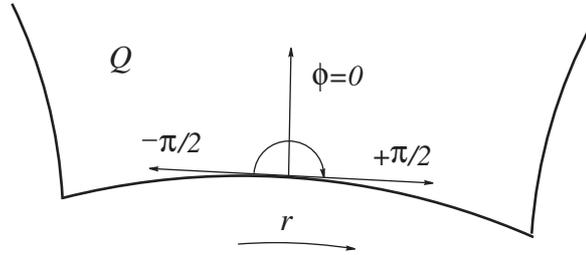


Figura IV.1: Orientación de las coordenadas r y ϕ .

Si la frontera ∂Q es una curva cerrada (este es el caso cuando Q es un polígono, un disco o, para referirnos a algo más complicado -e interesante-, la *cardioide* definida en coordenadas polares ρ, θ por $\rho = 1 - \cos \theta, 0 \leq \theta \leq 2\pi$), entonces M es, topológicamente, un cilindro $M = \partial Q \times [-\pi/2, \pi/2]$, y r es una coordenada cíclica. Si ∂Q consiste de varias curvas cerradas (esto sucede si Q no es simplemente conexa), entonces M es la unión de varios cilindros. En todos los casos $\partial M = \{|\phi| = \pi/2\}$. Sea

$$S_0 = \{|\phi| = \pi/2\} \cup \{(q, v): q \in \Gamma^*\}$$

Este es el conjunto donde las transformación de billar T tiene singularidades:

⁴Sobre las “fronteras interiores” la identificación deben ser hecha separadamente de cada lado de la línea.

⁵Observe que si la mesa fuera abierta habrían trayectorias en que el tiempo va a infinito sin choques. En estas notas no estudiaremos estos casos aunque tienen gran importancia en física y han sido bien estudiados cuando las fronteras son dispersoras (gas de Lorenz, etc).

Lema IV.1.1. Si la frontera ∂Q consiste de curvas suaves C^k entonces la transformación T es C^{k-1} en cada punto de $x \in M \setminus S_0$ tal que $Tx \in M \setminus S_0$.

Demostración. Sea $x = (r, \phi) \in M$ y $Tx = (r_1, \phi_1)$. Denote, como en (IV.1.5), $\tau = \tau(x)$ el tiempo de retorno (longitud de la trayectoria) entre x y Tx , y por $K = K(r)$ la curvatura de la frontera ∂Q en r (por lo que el ángulo entre los vectores normales $n(r)$ y $n(r + dr)$ es $K dr + o(dr)$). De manera similar, ponemos $K_1 = K(r_1)$. Recuerde las convenciones sobre los signos de las curvaturas.

Un detallado (pero elemental) análisis geométrico, ilustrado en la Figura IV.2, da la derivada de T :

$$D_x T = -\frac{1}{\cos \phi_1} \begin{pmatrix} \tau K + \cos \phi & \tau \\ \tau K K_1 + K \cos \phi_1 + K_1 \cos \phi & \tau K_1 + \cos \phi_1 \end{pmatrix} \quad (\text{IV.1.6})$$

Observar que, puesto que $\cos \phi_1 \neq 0$ y $\cos \phi \neq 0$, esta matriz está definida y es no singular. De igual manera, como la primer derivada de T involucra a la curvatura K de ∂Q (relacionada con la segunda derivada de Γ_i), entonces la suavidad de T es solamente C^{k-1} . \square

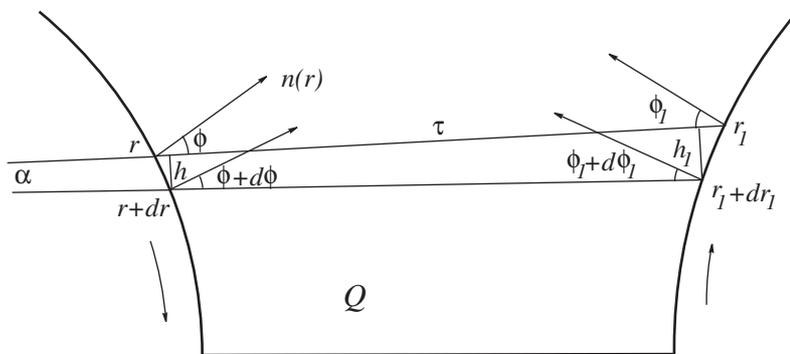


Figura IV.2: El cálculo de DT : observe que, infinitesimalmente, $h = dr \cos \phi$, $h_1 = -dr_1 \cos \phi_1$, $\alpha = d\phi + K dr$ (donde α es el ángulo entre las líneas de r a r_1 y de $r + dr$ a $r_1 + dr_1$), $h_1 = h + \tau \alpha$ y $d\phi_1 = K_1 dr_1 - \alpha$.

OBSERVACIÓN IV.1.2. Es fácil calcular que $|\det DT(x)| = \cos \phi / \cos \phi_1$.

EJERCICIO IV.1. Probar que si dos segmentos de recta se cortan perpendicularmente, el flujo del billar se puede extender continuamente para las trayectorias que llegan al vértice.

IV.2. Cuadrados y Círculos.

En el estudio de billares, aún ejemplos simples pueden exhibir propiedades dinámicas sorprendentemente ricas. Comenzamos con dos ejemplos particularmente simples, concebidos para un lector completamente principiante en el estudio de billares. Esta sección debe ser leído -más aún que cualquier texto de matemática- haciendo dibujos y cálculos simples.

Billar en un cuadrado Sea el cuadrado unitario $Q = \{(x, y): 0 \leq x, y \leq 1\}$ con las leyes de movimiento habituales en un billar.

Si la partícula llega a un vértice del cuadrado Q , la ley de reflexión (IV.1.3) no se puede aplicar por haber dos normales n en el vértice. Por tanto la partícula se detiene y la trayectoria termina. Estos casos singulares los estudiaremos un poco después; ahora concentrémonos en las trayectorias *regulares* que no van a los vértices.

Si $v_t = (u_t, w_t)$ denota el vector velocidad de la partícula móvil en el tiempo t (en las coordenadas x, y). Si choca el lado vertical de Q en el tiempo t , entonces u_t cambia de signo ($u_{t+0} = -u_{t-0}$) y w_t se mantiene incambiada. Si la partícula choca con el lado horizontal de Q , entonces w_t cambia de signo ($w_{t+0} = -w_{t-0}$)

y u_t se mantiene incambiada. Por tanto, $u_t = (-1)^m u_0$ y $w_t = (-1)^n w_0$, donde m y n denotan el número de colisiones en los lados verticales y horizontales, respectivamente, en el intervalo de tiempo $(0, t)$.

Ahora haremos uso de un procedimiento muy usado en billares poligonales: en lugar de reflejar la trayectoria en un lado del cuadrado Q la continuamos como si el *cuadrado hubiera sido reflejado* por ese lado; entonces la trayectoria continuará en línea recta a través de la imagen reflejada de Q . Reiterando este procedimiento tendremos que la trayectoria continúa en línea recta a través de las múltiples imágenes reflejadas del cuadrado. Ver Figura IV.3. Esta construcción es llamada el *desdoblamiento -unfolding-* de la trayectoria de billar. Para obtener la trayectoria original (verdadera) hay que *plegar -fold-* la cadena de copias adyacentes de Q y volver a Q misma.

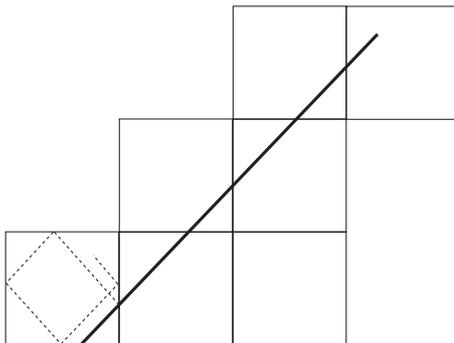


Figura IV.3: Desdoblamiento de la trayectoria en el cuadrado. La trayectoria “real” es la del primer cuadrado de la izquierda

Denotamos las copias de Q por

$$Q_{m,n} = \{(x, y): m \leq x \leq m + 1, n \leq y \leq n + 1\} \quad (\text{IV.2.1})$$

Los cuadrados $Q_{m,n}$ con $m, n \in \mathbb{Z}$ *embaldosan* (cubren con baldosas) *-tile-*, el plano entero \mathbb{R}^2 . Cualquier trayectoria de billar se desdobra en líneas rectas del plano, y cualquier recta del plano (que no pase por puntos con coordenadas ambas enteras) se pliega en una trayectoria del billar. Una trayectoria pasa por un vértice si y sólo si la correspondiente recta pasa por un punto de coordenadas enteras del plano.

La estructura de bloques $Q_{m,n}$ con las correspondientes reglas de plegado es claramente periódica, y el cuadrado 2×2 , $Q_2 = \{(x, y): 0 \leq x, y \leq 2\}$, juega el papel de un dominio fundamental –el plano entero es cubierto por traslaciones paralelas de Q_2 . Por tanto la proyección estándar de \mathbb{R}^2 sobre Q_2 transforma trayectorias desdobladas en trayectorias rectas del toro \mathbb{T}^2 (que se obtiene por identificaciones de los lados opuestos del cuadrado Q_2). El billar en el cuadrado unitario Q se reduce así al flujo geodésico en el toro plano \mathbb{T}^2 .

El flujo geodésico en el toro es uno de los ejemplos estándar en teoría ergódica. [KH95, PM78]. Las traslaciones en el toro, estudiadas en las Secciones I.1 A.4, corresponden a la transformación por el flujo geodésico, con tiempo fijo (Cuál?). Sus principales propiedades son:

- (a) si la trayectoria tiene pendiente racional, $dy/dx \in \mathbb{Q}$, entonces es periódica (va a lo largo de una geodésica cerrada);
- (b) si la trayectoria tiene pendiente irracional $dy/dx \notin \mathbb{Q}$, entonces es densa (su clausura es todo el toro).

Esto se transforma en estas alternativas para la trayectoria de billar en el cuadrado unitario Q :

Corolario IV.2.1. *Si $w_0/u_0 \in \mathbb{Q}$, entonces la correspondiente trayectoria de billar en Q es periódica. Si $w_0/u_0 \notin \mathbb{Q}$, es densa.*

El espacio de fase del flujo de billar en el cuadrado unitario Q es la variedad de dimensión tres $\mathcal{M} = Q \times S^1$; el flujo de billar Φ^t está definido para todo tiempo $-\infty < t < \infty$ en las trayectorias regulares. En las trayectorias excepcionales (aquellas que pasan en algún momento por un vértice), el flujo sólo se define hasta

que llega al vértice. El conjunto de las trayectorias excepcionales forma una unión numerable de superficies en \mathcal{M} .

Como ya sabemos (ver Ejercicio IV.1) en este caso podemos desembarazarnos de estas singularidades, extendiendo el flujo por continuidad, de modo que si una trayectoria pasa por un vértice ella sea devuelta y haga el recorrido inverso.

Esta extensión define el flujo de billar Φ^t en todo el espacio de fase \mathcal{M} y lo hace continuo en todo punto. Observamos, sin embargo, que tales extensiones bonitas, son posibles raramente, como veremos al estudiar billares con fronteras algo más complicadas.

Círculos. Sea ahora una mesa de billar circular: sea Q un disco de radio uno y centro en el origen. La superficie $M = \partial Q \times [-\pi/2, \pi/2]$ (con las coordenadas r, ϕ) es un cilindro cuya base es un círculo de longitud 2π y altura π .

Denote por $q_t = (x_t, y_t)$ las coordenadas de la partícula móvil en el tiempo t y por $v_t = (u_t, w_t)$ su vector velocidad. Entonces su posición y velocidad en el tiempo $t + s$ es

$$x_{t+s} = x_t + u_t s, \quad u_{t+s} = u_t; \quad y_{t+s} = y_t + w_t s, \quad w_{t+s} = w_t, \quad (\text{IV.2.2})$$

mientras la partícula se mantenga dentro de Q , o sea sin tocar su frontera ∂Q .

Cuando la partícula colide con ∂Q , su vector velocidad es reflejado por la línea tangente a ∂Q en el punto de colisión. Ver Figura IV.4

Después de la reflexión la partícula reasume el movimiento libre (IV.2.2) dentro del disco Q , hasta la siguiente colisión en la frontera ∂Q . Entonces rebota nuevamente, y así sucesivamente. El movimiento continuá infinitamente, tanto en el futuro como en el pasado.

Por ejemplo, si la partícula se mueve en un diámetro del disco, su vector velocidad se transformará en el opuesto cada vez que la partícula choque, y ella quedará moviéndose a lo largo del diámetro. Otros ejemplos de movimientos periódicos pueden ser fácilmente encontrados por el lector, analizando los polígonos regulares inscritos en el círculo. Para ver como actúa la transformación $T : M \rightarrow M$, considere $x = (r, \phi) \in M$. La siguiente reflexión ocurre en el punto $Tx = (r + \pi - 2\phi, \phi)$, y la n -ésima reflexión en

$$T^n x = (r + n(\pi - 2\phi), \phi)$$

En esta fórmula r es tomada como una coordenada cíclica: $r + n(\pi - 2\phi)$ es tomada módulo 2π . Varias conclusiones surgen de inmediato.

(i) El ángulo de reflexión ϕ es mantenido por T , por lo que toda curva $\phi = \text{const}$ en M es invariante por T . Cada una de esas curvas es un círculo paralelo a la base del cilindro M .

(ii) La transformación T restringida a cualquier curva $\phi = \text{const}$ la hace rotar un ángulo constante, $\pi - 2\phi$. Por tanto, T actúa como una rotación del círculo sobre cada curva invariante. Más aún, el ángulo de rotación cambia continuamente de círculo en círculo, disminuyendo de 2π en el círculo inferior ($\{\phi = -\pi/2\}$) a 0 en el superior ($\{\phi = \pi/2\}$), por lo que de hecho los círculos inferior y superior son dejados fijos por T . Por tanto el cilindro M es "retorcido" -*twisted*- hacia arriba por T .

(iii) La función tiempo de retorno $\tau(x) = 2 \cos \phi$ (o sea la distancia entre dos choques) es constante en cada círculo invariante $\phi = \text{const}$, y todos los segmentos de trayectoria comenzando con un ángulo dado ϕ son tangentes a un círculo de radio $R = \sin \phi$ concéntrico con la mesa de billar; ver Figura IV.4.

Si el ángulo de rotación $\pi - 2\phi$ es un *múltiplo racional* de π ($\frac{\pi - 2\phi}{2\pi} = \frac{m}{n}$ con $m, n \in \mathbb{Z}$), entonces la transformación T en el círculo $\phi = \text{const}$ es periódica con período n . Más aún, cada punto en este círculo es periódico con el mismo período.

Si el ángulo de rotación $\pi - 2\phi$ es *irracional*, (ϕ/π es un número irracional) entonces la rotación es ergódica respecto de la medida de Lebesgue (Ejercicio IV.11) y *únicamente ergódica*, lo cual significa que la medida invariante es única (ver Teorema I.1.5 en Sección I.1). Por tanto cada trayectoria es densa y uniformemente distribuida en el círculo ⁶. En este caso los segmentos de trayectoria cubren densamente un anillo de la mesa del billar con radio interior $R = \sin \phi$ (ver Figura IV.4).

Cada trozo de trayectoria es tangente a un círculo interior. Ver Ejercicio IV.6. En la Figura IV.4 se ve claramente que los segmentos de trayectoria se ven más densos cerca del círculo interior del anillo. En otras

⁶Una sucesión de puntos $x_n \in \mathcal{C}$ de un círculo \mathcal{C} está *uniformemente distribuida* si para cada intervalo $I \subset \mathcal{C}$ se tiene $\lim_{N \rightarrow \infty} \#\{n: 0 < n < N, x_n \in I\}/N = \text{length}(I)/\text{length}(\mathcal{C})$.

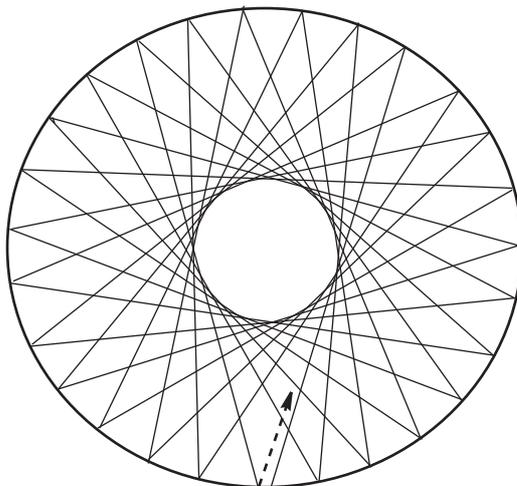


Figura IV.4: Una trayectoria no periódica en el círculo.

palabras, los segmentos se focalizan en el círculo interior. Si nuestra trayectoria fuera el pasaje de un rayo laser y la frontera de la mesa Q fuera un espejo perfecto, allí estaría “muy caliente”. Por esta razón el círculo interior es llamado una *cáustica* (de la palabra “quemar”, en griego). En general, una cáustica para un billar es una curva tal que si un trozo de trayectoria es tangente a ella, entonces la recta que contiene todo otro trozo de la misma trayectoria es tangente a la cáustica.

Como la superficie M está foliada por curvas invariantes $\phi = \text{const}$, T no es ergódica. Cualquier conjunto medible es unión de curvas T -invariantes. Otra manera de ver la no ergodicidad es encontrar funciones invariantes no constantes. Una tal función es $F : M \rightarrow \mathbb{R}$ definida por $F(r, \phi) = \phi$ que es suave T -invariante porque $F(Tx) = F(x)$ para todo $x \in M$.

Definición 15. Si un sistema dinámico $T : M \rightarrow M$ admite una función suave no constante invariante por T , entonces F es llamada una primer integral y T se dice integrable.

Si $T : M \rightarrow M$ es integrable, toda superficie de nivel $S_c = \{F(x) = c\}$ es T -invariante, o sea, M puede ser foliada por hipersuperficies. Si $\dim M = d$ y $T : M \rightarrow M$ admite $d - 1$ integrales primeras F_1, \dots, F_{d-1} , entonces M puede ser foliada por subvariedades T -invariantes de dimensión uno $\{F_1(x) = c_1, \dots, F_{d-1}(x) = c_{d-1}\}$, donde $c_1, \dots, c_{d-1} \in \mathbb{R}$.

Definición 16. Si M puede ser foliada por subvariedades T -invariantes de dimensión uno (curvas), entonces se dice que T es completamente integrable.

Por tanto el billar en un círculo es completamente integrable.

Las derivadas de T y T^n para todo $n \in \mathbb{Z}$, se obtienen tomando $\phi_1 = \phi$, $\tau = 2 \cos \phi$ y $K = -1$ en (IV.1.6):

$$DT(x) = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}, \quad DT^n(x) = \begin{pmatrix} 1 & -2n \\ 0 & 1 \end{pmatrix}. \quad (\text{IV.2.3})$$

Esto implica, de inmediato que ambos exponentes de Liapunov son cero en cada punto $x \in M$. Como consecuencia, la entropía también es cero. Por lo que, para todos los estándares T no es caótica.

Observe, sin embargo, que un vector tangente típico crece por DT^n cuando $n \rightarrow \infty$, pero sólo linealmente en n (y no exponencialmente). Las transformaciones en que la separación de trayectorias cercanas es típicamente lineal, se dice que tienen comportamiento *parabólico* (como opuesto a hiperbólico, de acuerdo a la definición vista en el Capítulo anterior, o elíptico, como veremos más adelante).

Modificación del billar circular. Denote Q_+ al semicírculo superior $x^2 + y^2 \leq 1$, $y \geq 0$, y sea una partícula puntual que se mueve dentro de Q_+ con choques en ∂Q_+ ; si la partícula choca ∂Q_+ en $(1, 0)$ o $(-1, 0)$, reiteramos los razonamientos hechos con los vértices del cuadrado estudiado anteriormente.

Este modelo se puede reducir al círculo unitario Q por *plegamientos*, como ya vimos en el caso del cuadrado. Sea $Q_- = Q \setminus Q_+$ la imagen especular de Q_+ a través del eje Ox . Cuando la partícula choca con el eje la trayectoria es reflejada, pero nosotros consideraremos su continuación (imagen especular) debajo del eje. Esta continuación evolucionará en Q_- simétricamente a la trayectoria real en Q_+ hasta que vuelva a chocar al eje. Entonces ambas trayectorias “se juntarán” y moverán juntas en Q_+ por un tiempo, hasta el siguiente choque en el eje (una irá por Q_- y la otra por Q_+), etc.

Es importante observar que la segunda trayectoria (la imaginaria) nunca es reflejada en el eje, ella sólo lo cruza, por lo que evoluciona como una trayectoria en el disco Q completo. Entonces las propiedades del billar en Q_+ se deducen de aquellas ya discutidas para todo el disco Q . Este tipo de reducción es común en el estudio de billares.

EJERCICIO IV.2. En el billar en un cuadrado, mostrar que si $u_0 \neq 0$ y $w_0 \neq 0$ y la partícula nunca va a un vértice, entonces se dan las cuatro combinaciones, $(\pm u_0, \pm w_0)$, al considerar su trayectoria infinita.

EJERCICIO IV.3. Mostrar que si m y n son pares, entonces el procedimiento de plegado del desdoblamiento del cuadrado (Figura IV.3), lleva $Q_{m,n}$ nuevamente en $Q = Q_{0,0}$ por traslaciones $x \mapsto x - m$, $y \mapsto y - n$, y por tanto preserva la orientación de x e y . Si m es impar, entonces la orientación de x es invertida (con más precisión, $x \mapsto m + 1 - x$). Si n es impar, entonces la orientación de y es invertida ($y \mapsto n + 1 - y$). Estas reglas no dependen de la trayectoria particular que haya sido desdoblada.

EJERCICIO IV.4. Probar que la trayectoria de billar en el rectángulo \mathcal{R} de lados a y b es periódica si y sólo si $(aw_0)/(bu_0) \in \mathbb{Q}$; caso contrario es densa. Sugerencia: transforme el rectángulo en un cuadrado unitario, reescalando la coordenadas: $(x, y) \mapsto (x/a, y/b)$. Muestre que las trayectorias en \mathcal{R} se transformarán en las de Q .

EJERCICIO IV.5. Extender el resultado anterior a billares en estos polígonos: triángulo equilátero, triángulo rectángulo isósceles, triángulo rectángulo con ángulo agudo $\pi/6$, y exágono regular. Qué es común en estos polígonos?

EJERCICIO IV.6. Probar que todo segmento entre dos colisiones sucesivas en el círculo de radio 1 es tangente al círculo menor $S_\psi = \{x^2 + y^2 = \cos^2 \psi\}$. Mostrar que si $\phi/2\pi$ es irracional, la trayectoria llena densamente el anillo entre ∂Q y S_ψ (ver Figura IV.4). Mostrar que una trayectoria no puede ser densa en todo el disco Q .

EJERCICIO IV.7. Probar que en el círculo, $T: M \rightarrow M$ preserva toda medida definida por $d\mu = f(r, \phi) dr d\phi$ con $f(r, \phi) = f(\phi)$.

EJERCICIO IV.8. Mostrar que si se cambia la velocidad de la partícula de la siguiente manera $\|v\|_{\text{new}} = c\|v\|_{\text{old}}$ con constante $c > 0$, entonces la trayectoria se mantiene incambiada y se produce un simple reescalamiento del tiempo: $t_t^{\text{nuevo}} = a_{ct}^{\text{viejo}}$ and $v_t^{\text{nuevo}} = v_{ct}^{\text{viejo}}$ para todo $t \in \mathbb{R}$.

EJERCICIO IV.9. Probar que las trayectorias periódicas en el medio círculo Q_+ corresponden a las trayectorias periódicas en Q . Observar, sin embargo, que el período puede diferir.

EJERCICIO IV.10. Investigar el movimiento del billar en un cuarto del círculo unitario $x^2 + y^2 \leq 1$, $x \geq 0$, $y \geq 0$.

EJERCICIO IV.11. Demostrar que la rotación en círculo invariante definida por las trayectorias del billar circular con ángulo irracional con π son ergódicas. Sugerencia: Usar el Teorema A.4.3

EJERCICIO IV.12. Sea Q un anillo – un dominio limitado por dos círculos concéntricos. Describa el espacio de fase M del billar en Q . Observe que $T: M \rightarrow M$ no es continua.

IV.3. Curvas Convexas y Estadios.

Estudiaremos en esta sección más ejemplos. Un poco más complicados que los de la sección anterior.

Elipses. Considere ahora el billar en el interior de la elipse

$$\frac{x^2}{a^2} + y^2 = 1 \tag{IV.3.1}$$

con algún $a > 1$. La superficie M es un cilindro con base (IV.3.1) y altura π .

Este billar tiene muchas propiedades geométricas bonitas. Por las pruebas, que en general son elementales, pero rebuscadas, recomendamos [CFS82, Ta95, Bi27a]. Se puede también usar métodos de la geometría proyectiva (el llamado Teorema de Poncelet -Ejercicio IV.13) para deducir muchas de esas propiedades, que enunciaremos a continuación.

Los focos F_1 y F_2 de la elipse (IV.3.1), están en el eje Ox . Si una trayectoria de billar pasa por uno de los focos, luego de reflejarse en ∂Q pasará por el otro foco, y así sucesivamente (ver Ejercicio IV.13). Las

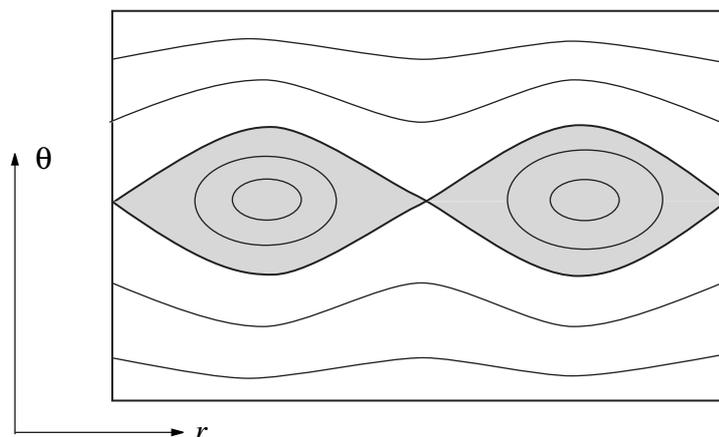


Figura IV.5: La superficie M para el billar elíptico.

trayectorias que pasan por los focos son especiales. Ellas forman una curva cerrada en la superficie M (la que tiene forma ∞ en la Figura IV.5).

Si la trayectoria cruza el segmento F_1F_2 que une los focos, entonces luego de reflejarse en ∂Q ella cruza ese segmento nuevamente. Tales trayectorias llenan dos dominios en M acotados por la curva con forma ∞ (marcada en gris en la Figura IV.5).

Las demás trayectorias nunca cruzan el segmento F_1F_2 . Cada vez que esas trayectorias cruzan el eje mayor (el de las x), el punto de intersección está fuera del segmento F_1F_2 (alternadamente, a izquierda y derecha de ese segmento). Tales trayectorias llenan dos dominios en M (las áreas blancas arriba y abajo de la curva con forma ∞ en la Figura IV.5). Ver Ejercicio IV.14.

Hay dos órbitas periódicas distinguidas -una sobre el eje mayor (representada por los puntos de intersección de las dos ramas de la curva con forma ∞) y otra sobre el eje menor (representada por los centros de las dos partes grises de la curva con forma ∞).

Cada trayectoria que no pasa por los focos y distinta de la órbita periódica en el eje menor tiene una cáustica (Figura IV.6). Para las trayectorias que cruzan el segmento F_1F_2 las cáusticas son hipérbolas confocales con la elipse (IV.3.1). Para las otras las cáusticas son elipses confocales e interiores a la del billar (IV.3.1) (estas cáusticas elípticas generalizan los círculos concéntricos del billar circular estudiado en la Sección anterior).

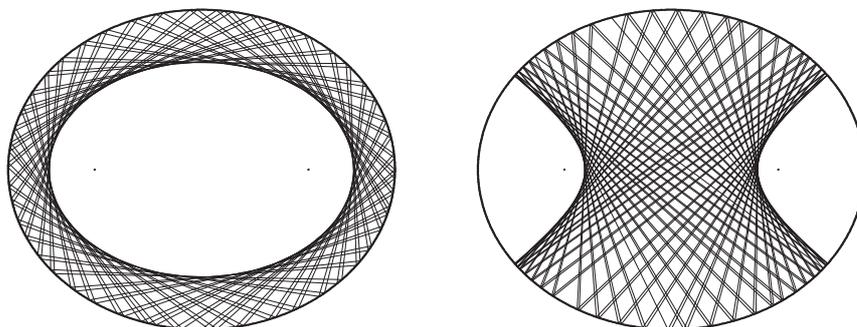


Figura IV.6: Trayectorias con cáustica elíptica e hiperbólica en el billar de la elipse.

Todas las trayectorias tangentes a una cáustica elíptica están en una curva de M que es invariante por T (“ondas horizontales” en el área blanca de la Figura IV.5). Estas curvas invariantes aparecen como deformaciones de las curvas invariantes $\phi = \text{const}$ del billar circular. Todas las trayectorias tangentes a las

cáusticas hiperbólicas están en dos curvas de M , cada una dentro de una mitad de la curva con forma ∞ , ambas invariantes por T^2 . Por tanto, la superficie M está completamente foliada por curvas invariantes. En este sentido el billar elíptico es similar al circular -ambos son completamente integrables. Más abajo, estudiamos las ecuaciones de las curvas invariantes.

Sobre cada curva la transformación T (o T^2) es conjugada a una rotación del círculo (cuyo ángulo se llama *número de rotación*). Este número cambia continua y monótonamente con las curvas invariantes (ver [MOP96]). La curva especial con forma ∞ que separa las regiones con diferentes tipos de cáusticas (hiperbólicas y elípticas) es llamada una *separatriz*. También es invariante por T , pero no rota: cada trayectoria comenzando en la separatriz converge (tanto en el futuro como en el pasado) a una de las órbitas 2-periódica del eje mayor.

La fórmula para la derivada de T en las coordenadas r, ϕ es complicada y no nos diría mucho. Pero uno puede elegir en M (excluyendo a la separatriz) un sistema de coordenadas curvilíneas especial con un eje a lo largo de una curva invariante, en el que la matriz DT será triangular y tendrá unos en la diagonal. Entonces uno puede fácilmente deducir que los exponentes de Liapunov son cero (dado que estos no dependen del sistema de coordenadas ni de la métrica). Entonces la entropía también es nula y nuevamente, por todos los estándares, la transformación T no es caótica.

El comportamiento de T , fuera de las separatrices y las órbitas periódicas distinguidas, es parabólico, como en el caso del billar circular. La trayectoria periódica del eje mayor es hiperbólica, y sus variedades estable e inestable están en la separatriz; ver Ejercicio IV.15. La órbita periódica del eje menor no es hiperbólica ni parabólica.

Definición 17. . Sea x un punto periódico ($T^n x = x$ para algún $n \geq 1$). Si la derivada $DT^n(x)$ tiene sólo valores propios simples complejos (no reales) en el círculo unitario, entonces el punto x es llamado elíptico.

Si $\dim M = 2$, entonces la derivada de DT^n es una rotación con algún ángulo. En este caso un entorno de x frecuentemente tiene curvas cerradas T^n -invariantes (óvalos semejantes a elipses, de allí el término *elíptico*). Este hecho está relacionado con la Teoría KAM mencionada al final de la Sección III.3. Las trayectorias periódicas del eje menor son elípticas (Ejercicio IV.15) y sus entornos están completamente foliados por óvalos T^2 -invariantes.

Por tanto, el billar dentro de una elipse despliega una combinación de estructuras parabólica, hiperbólicas y elípticas.

La elipse puede también ser parametrizada por la coordenada φ que es el ángulo del eje horizontal $y = 0$ con la tangente orientada a la elipse en el punto (x, y) . La ecuación de nuestra elipse y su curvatura, parametrizada por φ están dadas por

$$y(\varphi) = \pm \frac{1}{\sqrt{1 + a^2 \tan^2 \varphi}}, \quad x(\varphi) = -a^2 y(\varphi) \tan \varphi,$$

$$K(\varphi) = \frac{(1 + a^2 \tan^2 \varphi)^{1/2} (1 + \tan^2 \varphi + a^2 \tan^2 \varphi + a^2 \tan^4 \varphi)}{-a^2 (1 + \tan^2 \varphi)^{5/2}}.$$

La relación entre r y φ es $K(r)dr = -d\varphi$ de donde obtenemos de inmediato que $d\phi/d\varphi = -K(r)^{-1}d\phi/dr$.

El espacio de fase de nuestro billar elíptico puede ser parametrizado por dos conjuntos de coordenadas (r, ϕ) y (φ, ϕ) . Usando el segundo conjunto, definimos la función

$$G(\varphi, \phi) = \frac{\sin^2 \phi - \epsilon^2 \cos^2 \varphi}{1 - \epsilon^2 \cos^2 \varphi} \quad [\epsilon = \frac{\sqrt{a^2 - 1}}{a} \text{ es la excentricidad de la elipse}], \quad (\text{IV.3.2})$$

es una integral primera de la transformación de billar en la elipse: G es constante (≤ 1) a lo largo de las órbitas de T (Ejercicio IV.17). El espacio de fase del billar elíptico está foliado por las curvas de nivel de G .

Para diversas aplicaciones es importante calcular la pendiente de las tangentes a estas curvas de nivel. Daremos el resultado en las coordenadas (r, ϕ) . La pendiente de la curva de nivel $G = G(z)$ en $z = (r, \phi)$ está dada por

$$p(z) := \frac{d\phi}{dr} = \frac{K(r)\epsilon^2 \sin 2\varphi}{\sin 2\phi} (1 - G). \quad (\text{IV.3.3})$$

Las curvas de nivel en el espacio de fase se asocian (como ya fue observado) con las *cáusticas*. Por tanto podemos dividir las curvas de nivel de G en dos clases: elípticas e hiperbólicas.

Definición 18. Denote por \mathcal{E} el conjunto de M que consiste de las curvas de nivel elípticas y por \mathcal{H} , el de las hiperbólicas. Los puntos de \mathcal{E} tienen trayectorias con cáusticas elípticas, y los puntos en \mathcal{H} tienen trayectorias con cáusticas hiperbólicas.

Tenemos $0 < G < 1$ en \mathcal{E} , y $1 - a^2 \leq G < 0$ en \mathcal{H} . En las trayectorias por los focos, cuya unión forma la conexión silla del espacio de fase, tenemos $G = 0$.

Mesas convexas (suaves). Uno se puede preguntar si hay billares completamente integrables distintos de los círculos y las elipses. La cuestión es conocida como la

Conjetura de Birkhoff. Los únicos billares completamente integrables son los círculos y elipses.

Sorprendentemente, esta conjetura -que tiene más de 70 años- está aún abierta. Aunque la mayoría de los matemáticos creemos que es cierta.

Por otro lado, los billares en mesas con forma oval, a pesar de no ser completamente integrables, tienen muchas características comunes con los billares en las elipses. En 1973 Lazutkin [La73] probó: si Q es un dominio estrictamente convexo (la curvatura de la frontera nunca se anula) con la frontera suficientemente regular, entonces existe un conjunto de medida positiva $N \subset M$ que es foliado por curvas invariantes. El conjunto N se acumula cerca de la frontera ∂M . Todas las trayectorias que comienzan en N tienen cáusticas, que son curvas convexas cerradas en el interior de Q . Esta región es todo el espacio de fase en el caso del círculo y el conjunto \mathcal{E} de la Definición 18 en el caso de la elipse.

Naturalmente que el billar no puede ser ergódico porque $\nu(N) > 0$. Los exponentes de Liapunov de puntos $x \in N$ son cero. Sin embargo, fuera de N la dinámica puede ser completamente diferente.

Originalmente Lazutkin demandó 553 derivadas continuas para su teorema, pero más tarde R. Douady [Dy82] probó que 6 eran suficientes (se conjetura que la curva podría ser C^4). Esta mejora es consecuencia de avances en el Teorema KAM. Existen ejemplos que muestran que por lo menos dos derivadas son necesarias para la existencia de cáusticas ⁷. También, J. Mather [Mt82] probó que aún un único punto con curvatura nula en un mesa convexa con frontera C^2 puede obstruir la existencia de cáusticas cerca de ∂Q .

Stadium. Ahora modificaremos un poco la mesa de billar circular. Tome dos semicírculos de radio uno, juntos, y sepárelos en dirección perpendicular al diámetro común, una distancia 2ε . Una las dos mitades separadas por segmentos paralelos. El dominio resultante es descrito por

$$Q = \{|y| < 1, |x| < \varepsilon + \sqrt{1 - y^2}\}$$

Este es un dominio convexo con frontera C^1 (pero no C^2). Recuerda a un estadio griego o romano, un *stadium*.

L. Bunimovich fue el primero en introducir el estudio de los estadios en los setenta [Bu74, Bu79]. Descubrió este hecho impactante: el billar en el stadium tiene exponentes de Liapunov no nulos. Su entropía es positiva. Más aún es ergódico, mixing e isomorfo a un shift de Bernoulli. Entonces uno puede perturbar apenas el billar circular para obtener un sistema dinámico completamente caótico.

Modificaciones similares se pueden hacer en las elipses. Se corta la elipse a lo largo del eje menor, se separan las dos mitades en la dirección del eje mayor, y se juntan las dos partes con dos segmentos paralelos de longitud $2h$. Se obtiene un *stadium elíptico*. Es hiperbólico y ergódico para $h > h_{\min} > 0$, donde h_{\min} depende de la excentricidad de la elipse [MOP96, DMM03].

Hay otra manera de obtener billares caóticos por modificaciones de círculos y elipses. Tome un círculo y corta una rebanado a lo largo de una cuerda (que puede ser arbitrariamente pequeña) El billar en tales círculos “truncados” es hiperbólico y ergódico, como probó Bunimovich en [Bu74]. O tome la elipse (IV.3.1) y corte dos rebanadas simétricas a lo largo de las líneas cut two symmetric slices along the lines $y = \pm(1 - h)$, $0 < h < 1$. El billar en esa elipse truncada es también hiperbólico y ergódico [DelM01].

Todos los billares en los stadiums son convexas con frontera C^1 (pero no C^2). La transformación $T : M \rightarrow M$ es continua pero no siempre diferenciable. Hay puntos donde DT no existe (puntos singulares). Estos son puntos $x \in M$ tales que $Tx \in \Gamma^*$. En la Figura IV.7 se muestra el espacio de fase del stadium elíptico con la parametrización de la curva comenzando en el extremo inferior de la semielipse de la derecha.

⁷Billares convexas con fronteras C^1 fueron estudiados por Hubacher [Hu87]. Sus derivadas segundas existen y son continuas excepto en un conjunto finito de puntos donde sus límites laterales existen, pero la función curvatura es discontinua. Se prueba que no existen cáusticas en un entorno de la frontera, pero se observa que pueden existir lejos de ella.

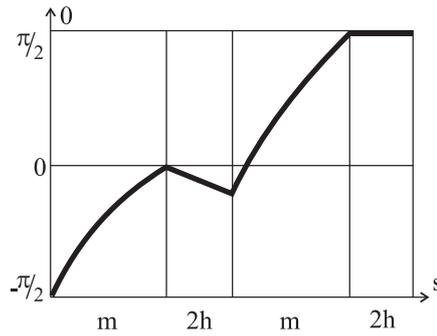


Figura IV.7: Espacio de fase de la transformación de billar en el stadium elíptico. m es la longitud de media elipse. La curva gruesa indica el conjunto de puntos que van a uno de los puntos de contacto de una de las media elipses y uno de los segmentos paralelos de la frontera.

Se ha dibujado una de las 4 curvas del conjunto de singularidades S^+ (ver Sección III.5). Parecería que las singularidades de T siempre existen en los billares caóticos.

EJERCICIO IV.13. Sea Q una elipse con focos F_1 and F_2 . Sea $A \in \partial Q$ y L la tangente a ∂Q en A . Probar que los segmentos AF_1 y AF_2 forman ángulos iguales con la línea L (este es conocido en geometría proyectiva como Teorema de Poncelet). Sugerencia: Use la propiedad característica de la elipse: $|AF_1| + |AF_2|$ no depende de A .

EJERCICIO IV.14. En el espacio M del billar elíptico marcar 5 rebotes sucesivos de distintos tipos de trayectorias. En particular distinguir las trayectorias que avanzan en sentido horario de las que lo hacen en sentido antihorario, entre las que no cruzan el segmento F_1F_2 .

EJERCICIO IV.15. (a) Sea x un punto 2-periódico de la transformación de billar T . Observe que entonces $\cos \phi = \cos \phi_1 = 1$ en la fórmula (IV.1.6). Muestre que los valores propios de DT^2 son complejos (no reales) si y sólo si

$$\tau K K_1 + K + K_1 < 0 \quad \text{y} \quad (\tau K + 1)(\tau K_1 + 1) > 0$$

con la notación de (IV.1.6).

(b) Probar que la órbita periódica en el eje menor de la elipse es elíptica.

(c) Probar que la órbita periódica en el eje mayor de la elipse es hiperbólica. Describir con precisión sus variedades estable e inestable.

EJERCICIO IV.16. Probar que la función $G(\varphi, \phi)$ definida en (IV.3.2) es una integral primera del billar elíptico.

EJERCICIO IV.17. En la Figura IV.7, dibuje las otras tres curvas de puntos que tienen imagen en los puntos de contacto de una media elipse y un segmento de recta.

IV.4. Medidas. Teorema de Oseledec

Medida Invariante. Las ecuaciones de Hamilton (IV.1.2) preservan la medida de Liouville $dq dv$ en el espacio de fase \mathcal{M} , donde dq y dv son las medidas uniformes en Q y S^1 , respectivamente. Uno puede comprobar directamente (lo omitimos) que esta medida es también invariante en las reflexiones (IV.1.3). Como la mesa Q es compacta, la medida $dq dv$ en \mathcal{M} es finita y puede ser normalizada; tendremos así una probabilidad

$$d\mu = c_\mu dq dv \quad \text{con} \quad c_\mu = (2\pi|Q|)^{-1};$$

$|Q|$ es el área del dominio Q .

La medida invariante μ induce una medida T -invariante en M . Para hallarla, introducimos un sistema especial de coordenadas en el espacio \mathcal{M} . Para cada punto $x = (q, v) \in \mathcal{M}$ sea $s(x) = -\max\{t < 0: \Phi^{t+0}x \in M\}$ y $y = (q', v') = \Phi^{-s(x)+0}x$. Claramente, $s(x)$ es el tiempo transcurrido desde la reflexión anterior, $q' \in \partial Q$ es el punto de reflexión, y v' es el vector de velocidad saliente. Observamos que $v' = v$, y el punto q' y el valor de s pueden ser determinados unívocamente resolviendo la ecuación $q - sv \in \partial Q$. Sean r, ϕ las coordenadas del punto $(q', v') \in M$. Ahora la variedad \mathcal{M} puede ser parametrizada por las tres nuevas coordenadas r, ϕ, s .

Un corto razonamiento geométrico (ilustrado en la Figura IV.8) permite verificar que el volumen infinitesimal en \mathcal{M} puede ser representado por $dq dv = \cos \phi dr d\phi ds$. Entonces la medida μ en \mathcal{M} satisface

$$d\mu = c_\mu \cos \phi dr d\phi ds \tag{IV.4.1}$$

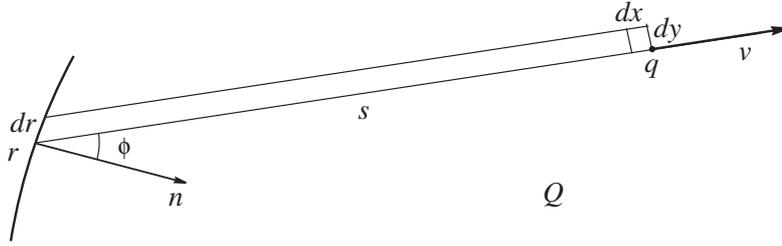


Figura IV.8: Las coordenadas r, ϕ, s .

Como el flujo Φ^t tiene velocidad constante (igual a uno), entonces la medida $\cos \phi dr d\phi$ es invariante por la transformación $T : M \rightarrow M$. La normalizamos para obtener la probabilidad

$$d\nu = c_\nu \cos \phi dr d\phi \quad \text{con} \quad c_\nu = (2|\partial Q|)^{-1};$$

$|\partial Q|$ es la longitud total de la frontera ∂Q (perímetro de Q). La invariancia de la medida por T también se deduce de la Observación IV.1.2 luego del Lemma IV.1.1, como se puede ver haciendo cambio de variable al integrar directamente sobre cualquier conjunto de Borel $U \subset M$:

$$\nu(T(U)) = c_\nu \int_{T(U)} \cos \phi_1 dr_1 d\phi_1 = c_\nu \int_U \cos \phi_1 \frac{\cos \phi}{\cos \phi_1} dr d\phi = \nu(U).$$

Lema IV.4.1. *Existe un conjunto $M' \subset M$ de medida total ($\nu(M') = 1$), en el que la transformación T^n está definida para todo $n \in \mathbb{Z}$. de igual manera, hay un subconjunto $\mathcal{M}' \subset \mathcal{M}$ de medida total, en el que el flujo Φ^t está definido para todo tiempo $t \in \mathbb{R}$.*

Demostración. T no está definida en $x \in M$ sólo si la siguiente reflexión ocurre en algún $q \in \Gamma^*$. Tales puntos x pertenecen a una unión finita o numerable de curvas suaves en M . Por lo tanto, los puntos $x \in M$ cuyas imágenes para adelante o para atrás por T alguna vez pasan al conjunto singular Γ^* están en una unión numerable de curvas suaves en M , por lo que su medida total es cero. Esto prueba la primer parte del lema.

El flujo Φ^t no está definido en $x \in \mathcal{M}$ si la trayectoria por x o bien (i) pasa por un punto singular $q \in \Gamma^*$ o bien (ii) experimenta infinitas reflexiones en tiempo finito. El conjunto de puntos de tipo (i) tiene medida cero por la primera parte. Los del tipo (ii) también tienen medida nula como consecuencia de Teorema de Poincaré. Ver Ejercicio IV.18. \square

Billares en un toro. Una interesante modificación del modelo de billar se obtiene tomando una mesa Q en el toro bidimensional \mathbb{T}^2 , en lugar de en plano \mathbb{R}^2 . Por ejemplo, sea $D \subset \mathbb{T}^2$ un pequeño disco y $Q = \mathbb{T}^2 \setminus D$. Tal sistema puede ser pensado como un billar en el cuadrado unitario con condiciones periódicas de frontera como se ve en la Figura IV.9. En este caso el tiempo de retorno $\tau(x)$, como función en M , es no acotado. Sin embargo, la transformación $T : M \rightarrow M$ sigue estando bien definida: si $x \in M$ el punto $Tx \in M$ siempre existe; ver Ejercicio IV.19.

Involución. El sistema dinámico $(\mathcal{M}', \mu, \Phi^t)$ tiene una interesante propiedad llamada *involución*: para cada $x = (q, v) \in \mathcal{M}'$ el punto $I(x) = (q, -v) \in \mathcal{M}'$ satisface

$$\Phi^t(I(x)) = I(\Phi^{-t}x)$$

para todo $t \in \mathbb{R}$. Por tanto, la involución $I : \mathcal{M}' \rightarrow \mathcal{M}'$ anticonmuta con la dinámica Φ^t , lo cual puede ser escrito $\Phi^t \circ I = I \circ \Phi^{-t}$. Observe que I también preserva la medida μ .

La transformación $T : M' \rightarrow M'$ también admite una involución, I_1 , definida por $(r, \phi) \mapsto (r, -\phi)$. Ella anticonmuta $T : T^k \circ I_1 = I_1 \circ T^{-k}$ para todo $k \in \mathbb{Z}$. También $I_1 : M \rightarrow M$ preserva la medida ν . Suele representarse I_1 usando sólo el signo $-$ y entonces la anticonmutación se escribe, para $x \in M$: $T^k(-x) = -T^{-k}x$.

Exponentes de Liapunov. Ahora probamos que el Teorema de Oseledec III.2.1 se aplica a la transformación de billar T .

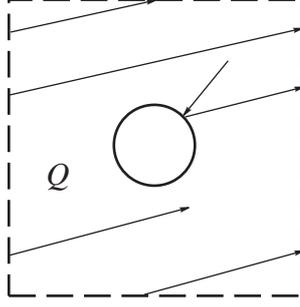


Figura IV.9: Un billar en un toro.

Teorema IV.4.2. *Si el valor absoluto de la curvatura de ∂Q está uniformemente acotado, entonces el Teorema de Oseledec se aplica a la transformación de billar T .*

Demostración. Necesitamos verificar que las funciones $\log^+ \|DT(x)\|$ y $\log^+ \|DT^{-1}(x)\|$ tienen integrales finitas en M . Por la propiedad de involución, es suficiente hacerlo para $DT(x)$. Se deduce de (IV.1.6) que $\|DT(x)\| \leq C/\cos \phi_1$, donde $C > 0$ es una constante. Entonces

$$\begin{aligned} c_\nu \int_M \log^+ \|DT(x)\| d\nu &\leq c_\nu \int_M |\log C + \log \cos \phi_1| \cos \phi d\phi dr \\ &\leq |\log C| + c_\nu \int_M |\log \cos \phi| \cos \phi d\phi dr \end{aligned}$$

donde hemos usado la invariancia de la medida ν . Finalmente,

$$\begin{aligned} \int_M |\log \cos \phi| \cos \phi d\phi dr &= |\partial Q| \int_{-\pi/2}^{\pi/2} |\log \cos \phi| \cos \phi d\phi \\ &= |\partial Q| (2 - \log 4) < \infty \end{aligned}$$

El teorema está probado. □

OBSERVACIÓN IV.4.3. La condición en la curvatura de ∂Q es necesaria. Katok y Strelcyn [KS86] dan un ejemplo de una mesa de billar simplemente conexa cuya frontera es C^∞ en todo punto, excepto uno, en el que la curvatura se va para infinito, para la cual $\int_M \log^+ \|DT(x)\| d\nu = \infty$.

El Teorema de arriba asegura la existencia de dos exponentes de Liapunov, $\lambda_1(x)$ y $\lambda_2(x)$, y los correspondientes subespacios invariantes, $E_1(x)$ y $E_2(x)$, en casi todo punto $x \in M'$. Observamos que si $\lambda_1(x) \neq \lambda_2(x)$, entonces el ángulo $\lambda(x)$ entre las líneas $E_1(x)$ y $E_2(x)$ está definido. Por la Observación III.2.2,

$$\lim_{n \rightarrow \pm\infty} \frac{1}{n} \log \gamma(T^n x) = 0 \tag{IV.4.2}$$

Corolario IV.4.4. $\lambda_1(x) + \lambda_2(x) = 0$ para casi todo $x \in M$.

Demostración. Sea un paralelogramo en $\mathcal{T}_x M$ con lados s_1, s_2 paralelos a $E_1(x)$ y $E_2(x)$, respectivamente. Entonces su área es $s_1 s_2 \sin \gamma(x)$. Su imagen por DT^n es un paralelogramo similar en $\mathcal{T}_{T^n x} M$ con lados $s_1(n), s_2(n)$. Por tanto,

$$\det(DT^n(x)) = \frac{\sin \gamma(T^n x)}{\sin \gamma(x)} \frac{s_1(n)}{s_1} \frac{s_2(n)}{s_2} \tag{IV.4.3}$$

También observamos que

$$\lim_{n \rightarrow \pm\infty} \frac{1}{n} \log \frac{s_i(n)}{s_i} = \lambda_i(x) \tag{IV.4.4}$$

para $i = 1, 2$. Considere el punto $T^n x = (r_n, \phi_n)$. Por la Observación IV.1.2, luego del Lema IV.1.1, $\det(DT^n(x)) = \cos \phi / \cos \phi_n$. Sea $A_e = \{x : \cos \varphi > e\}$; el Teorema de Recurrencia de Poincaré implica que $\mu(\cup_{e>0} A_e) = 1$, por lo que $\mu(\{x : \cos \varphi_n(x) \rightarrow 0\}) = 0$. Por tanto, $\det(DT^n(x))$ está acotado entre 0 e infinito la mayor parte del tiempo. Ahora, tomando logaritmos en (IV.4.3), dividiendo por n , pasando al límite con $n \rightarrow \pm\infty$, y usando (IV.4.2) y (IV.4.4) queda probado el corolario. \square

Por tanto, casi todo punto $x \in M$ o bien tiene ambos exponentes de Liapunov nulos o bien es hiperbólico.

Un modelo mecánico. Considere un sistema de dos partículas materiales de masas m_1 y m_2 en el intervalo unitario $0 \leq x \leq 1$. Las partículas se mueven libremente y chocan elásticamente entre ellas y con las “paredes” en $x = 0$ y $x = 1$. Sean $x_1 \leq x_2$ las posiciones de las partículas y u_1, u_2 sus velocidades. Suponga que $x_1 \leq x_2$. Cuando una partícula choca con una pared, simplemente cambia su velocidad. Cuando las dos partículas chocan entre ellas preservan el momento total $m_1 u_1 + m_2 u_2$ y la energía cinética total $(m_1 u_1^2 + m_2 u_2^2)/2$.

Ahora introducimos nuevas variables:

$$q_i = x_i \sqrt{m_i} \quad y \quad v_i = \dot{q}_i = u_i \sqrt{m_i} \quad (\text{IV.4.5})$$

para $i = 1, 2$. El estado del sistema es descrito por un punto $q = (q_1, q_2) \in \mathbb{R}^2$ (llamado *punto de configuración*) y su vector velocidad $v = (v_1, v_2)$. El conjunto de todos los puntos de configuración (llamado *espacio de configuración*) es el triángulo rectángulo

$$Q = \{q = (q_1, q_2) : 0 \leq q_1 / \sqrt{m_1} \leq q_2 / \sqrt{m_2} \leq 1\}.$$

Se verifica fácilmente (ver Ejercicio IV.20) que la trayectoria del sistema en Q es gobernada por las reglas del billar (IV.1.2) y (IV.1.3). Por lo tanto, el estudio de este modelo mecánico se reduce al estudio de la dinámica del billar en el triángulo Q . Veremos más ejemplos de estas reducciones en la próxima Sección.

EJERCICIO IV.18. Mostrar que para μ -casi todo punto $x \in M$ y todo $T > 0$ el “segmento” $\Phi^t x$, $0 < t < T$, contiene una cantidad finita de reflexiones ∂Q . Sugerencia: Es suficiente mostrar la misma propiedad para ν -casi todo punto $x \in M$, lo cual se puede hacer con la ayuda del Teorema de Recurrencia de Poincaré aplicado a la transformación $T : M \rightarrow M$.

EJERCICIO IV.19. Sea $Q \subset \mathbb{T}^2$ cualquier mesa de billar en el toro. Probar que para todo $x \in M$ su semiórbita $\Phi^t x$, $t > 0$, necesariamente corta ∂Q , por lo que Tx existe (aunque pertenezca al conjunto singular S_0). Sugerencia: Si la semiórbita no cruzara ∂Q , sería o bien periódica o bien densa en el toro, see [PM78] or [KH95]. Ver Teorema A.4.3

EJERCICIO IV.20. Verificar que la dinámica del modelo mecánico anterior corresponde a un billar en el triángulo Q . Sugerencia: La única parte no trivial es mostrar que las colisiones entre las partículas corresponden a reflexiones especulares en la hipotenusa del triángulo Q . Esta es la razón por la que los factores $\sqrt{m_i}$, $i = 1, 2$, fueron introducidos en (IV.4.5).

IV.5. Billares en Dimensiones Mayores. Modelos

Esta Sección es más avanzada que las anteriores: se recomienda andar más despacio. Incluso puede saltarse y seguir con la siguiente, pero luego volver, porque muchas de las motivaciones de los primeros estudios de teoría ergódica se reflejan en su contenido. Ver [?].

Sea Q una región abierta acotada en el espacio \mathbb{R}^d o en el toro d -dimensional \mathbb{T}^d . Suponga que la frontera $\partial Q = \Gamma$ es o bien una subvariedad suave C^k ($k \geq 3$), $(d-1)$ -dimensional o consiste de un número finito de subvariedades suaves C^k , $(d-1)$ -dimensionales $\Gamma_1, \dots, \Gamma_s$ con frontera, de modo que vale (IV.1.1). Por ejemplo, si Q es un poliedro de dimensión d , entonces ∂Q consiste de caras $(d-1)$ -dimensionales cuyas fronteras son aristas $(d-2)$ -dimensionales. El conjunto $\Gamma^* = \partial\Gamma_1 \cup \dots \cup \partial\Gamma_s$ es nuevamente llamado el conjunto singular de ∂Q .

La partícula del billar se mueve libremente dentro de Q de acuerdo con las ecuaciones (IV.1.2) y se refleja elásticamente en la frontera ∂Q de acuerdo con la ecuación (IV.1.3), a menos que $q \in \Gamma^*$, en cuyo caso la reflexión no es definida. Estas ecuaciones preservan la norma de $\|v\|$, y fijamos $\|v\| = 1$. El dominio Q es llamado ahora *espacio de configuración* del sistema del billar.

Muchos resultados de las secciones anteriores se extienden a los billares multidimensionales con las modificaciones obvias. repetimos esos resultados brevemente, enfatizando las diferencias con los billares planos.

El espacio de fase del flujo es $\mathcal{M} = \bar{Q} \times S^{d-1}$, donde \bar{Q} es otra vez la clausura del dominio del billar Q , y S^{d-1} es la esfera unitaria $(d-1)$ -dimensional de todos los vectores velocidad. Para cada $q \in \partial Q$ se identifican los puntos (q, v_-) y (q, v_+) relacionados por (IV.1.3).

La dinámica del billar induce un flujo Φ^t en el espacio \mathcal{M} que tiene una medida invariante $d\mu = c_\mu dq dv$, donde dq y dv son las medidas de Lebesgue en Q y S^{d-1} , respectivamente, y c_μ es el factor de normalización. Es fácil ver que

$$c_\mu = (|Q| \cdot |S^{d-1}|)^{-1} \quad (\text{IV.5.1})$$

donde $|Q|$ es el volumen d -dimensional del dominio Q y $|S^{d-1}|$ es el volumen $(d-1)$ -dimensional de la esfera S^{d-1} .

La medida μ puede ser representada en un nuevo sistema de coordenadas de una manera similar a (IV.4.1):

$$d\mu = c_\mu \langle v, n(q) \rangle dr dv ds \quad (\text{IV.5.2})$$

donde r y s fueron mostrados en la Figura IV.8 de la Sección IV.1. La coordenada ϕ en (IV.4.1) es reemplazada por $v \in S^{d-1}$, y $\cos \phi$ por $\langle v, n(q) \rangle$ (aunque por ser $v, n(q)$ versores (de norma 1) ese producto interno también representa el coseno del ángulo respectivo).

La sección M del espacio de fase \mathcal{M} sigue siendo definida por (IV.1.4). Observe que $\dim M = 2d - 2$. La transformación del billar $T : M \rightarrow M$ y la función tiempo de retorno $\tau(x)$ en M se definen de manera natural, de la misma manera que en la Sección IV.1. Si las componentes suaves de la frontera ∂Q son de clase C^k , entonces T es C^{k-1} en todos los puntos $x \in M \setminus S_0$ tales que $Tx \in M \setminus S_0$. Aquí

$$S_0 = \{(q, v) \in M : \langle v, n(q) \rangle = 0\} \cup \{(q, v) \in M : q \in \Gamma^*\}$$

La transformación de billar $T : M \rightarrow M$ preserva la medida

$$dv = c_\nu \langle v, n(r) \rangle dr dv \quad \text{con} \quad c_\nu = (|\partial Q| \cdot |B^{d-1}|)^{-1} \quad (\text{IV.5.3})$$

donde dr y dv son las medidas de Lebesgue en ∂Q and S^{d-1} , respectivamente, c_ν es el factor de normalización (que se calcula por integración directa en M -lo dejamos como ejercicio), $|\partial Q|$ es el volumen $(d-1)$ -dimensional de ∂Q y $|B^{d-1}|$ es el volumen $(d-1)$ -dimensional de la bola unitaria $B^{d-1} \subset \mathbb{R}^{d-1}$.

Existe un subconjunto $\mathcal{M}' \subset \mathcal{M}$ de μ -medida total en el que el flujo Φ^t está definido para todos los tiempos $t \in \mathbb{R}$. De igual manera, hay un subconjunto $M' \subset M$ de ν -medida total en que T^k está definida para todo $k \in \mathbb{Z}$.

La involución $I : \mathcal{M}' \rightarrow \mathcal{M}'$ es definida como antes por $I(q, v) = (q, -v)$; satisface la relación $(\Phi^t \circ I = I \circ \Phi^{-t})$ y preserva la medida μ .

También $T : M' \rightarrow M'$ admite una involución, I_1 , definida por $(r, v_-) \mapsto (r, v_+)$ en la notación de (IV.1.3). La involución I_1 anticonmuta con T : $T^k \circ I_1 = I_1 \circ T^{-k}$ para todo $k \in \mathbb{Z}$, y preserva la medida ν .

Si los valores absolutos de todas las curvaturas seccionales de ∂Q están uniformemente acotadas, entonces el Teorema de Osedelec se aplica a T y asegura la existencia de $2d - 2$ exponentes de Liapunov $\{\lambda_1(x), \dots, \lambda_{2d-2}(x)\}$ en casi todo $x \in M$. Como en Corolario IV.4.4, su suma se anula:

$$\lambda_1(x) + \dots + \lambda_{2d-2}(x) = 0 \quad (\text{IV.5.4})$$

OBSERVACIÓN IV.5.1. Un resultado mucho más fuerte que (IV.5.4) se puede obtener por medio de geometría simpléctica. Todo billar es un sistema Hamiltoniano y preserva una forma simpléctica natural; por detalles, ver [LW95]. Se deduce que si ordenamos los exponentes de Liapunov, $\lambda_1(x) \geq \dots \geq \lambda_{2d-2}(x)$, entonces $\lambda_i(x) + \lambda_{2d-1-i}(x) = 0$ para todo $1 \leq i \leq 2d - 2$ y casi todo $x \in M$. Esto es llamado *simetría de los exponentes de Liapunov*.

Un modelo mecánico. Esta es una generalización del sistema de dos partículas de la Sección IV.4. Considere n partículas puntuales de masas m_1, \dots, m_n en el intervalo unitario $0 \leq x \leq 1$. Las partículas se mueven libremente y chocan elásticamente entre si y con las “paredes” en $x = 0$ y $x = 1$. Sean $x_1 \leq \dots \leq x_n$ las posiciones de las partículas y u_1, \dots, u_n sus velocidades.

Introducimos nuevas variables:

$$q_i = x_i \sqrt{m_i} \quad \text{y} \quad v_i = \dot{q}_i = u_i \sqrt{m_i}$$

para $i = 1, \dots, n$. Entonces el estado del sistema es descripto por un punto $q = (q_1, \dots, q_n) \in \mathbb{R}^n$ (llamado *punto de configuración*) y su vector velocidad $v = (v_1, \dots, v_n)$. El *espacio de configuración* es una pirámide recta en \mathbb{R}^n :

$$Q = \{q : 0 \leq q_1 / \sqrt{m_1} \leq \dots \leq q_n / \sqrt{m_n} \leq 1\}$$

Como en la Sección IV.4, uno puede verificar que la trayectoria del sistema en Q es gobernada por las reglas del billar (IV.1.2) y (IV.1.3). Por lo tanto, el estudio del modelo mecánico de n partículas en el intervalo puede ser reducido al estudio de la dinámica del billar en la pirámide Q . Ver [CFS82], [?].

Esta reducción tiene interesantes consecuencias. Por ejemplo, recuerde que la medida invariante $d\mu$ es el producto de las medidas de Lebesgue tanto en la pirámide Q como en la esfera $(n-1)$ -dimensional S_v^{n-1} de los vectores velocidad. Entonces para cada partícula i la distribución de su velocidad v_i es una medida marginal de la distribución uniforme en la esfera S_v^{n-1} . La norma del vector velocidad v determina

$$v_1^2 + \cdots + v_n^2 = Kn, \quad K = \text{const} \quad (\text{IV.5.5})$$

Entonces la distribución de cada v_i tiene un límite cuando $n \rightarrow \infty$. Este límite es una medida Gaussiana con media cero y variancia K : $v_i \sim N(0, K)$ cuando $n \rightarrow \infty$. La distribución normal de vectores velocidad es conocida como *distribución de Maxwell* en mecánica estadística y es característica para moléculas de gases y fluidos.

(IV.5.5) puede ser traducido en las variables originales q_i, u_i , y se transforma en una condición para la energía cinética total:

$$\frac{m_1 u_1^2}{2} + \cdots + \frac{m_n u_n^2}{2} = \frac{Kn}{2}.$$

$K/2$ es la energía cinética media por partícula. Ahora como v_i^2 tiene la misma distribución de probabilidad para todo i , también lo hace la energía cinética $m_i u_i^2/2$ de cada partícula. Este es un hecho importante: cada partícula (independientemente de su masa) tiene la misma distribución (y, naturalmente, el mismo valor medio, igual a $K/2$) para su energía cinética. Este hecho es conocido en mecánica estadística como *equipartición de energía* – la energía total en un sistema de múltiples partículas está igualmente dividida entre las partículas.

Gases de discos y bolas duras. El sistema de partículas moviéndose en el intervalo discutido arriba es un modelo muy simplificado de un gas. Considere un modelo más real de n discos moviéndose en el plano, o n bolas moviéndose en el espacio. Estos dos modelos son muy similares, por lo que los discutiremos en paralelo. Por simplicidad, supongamos que todos los discos (bolas) tienen el mismo radio r y la misma masa m . Esta simplificación no introduce ninguna restricción, sólo la hacemos para facilitar la presentación analítica de algunas relaciones y fórmulas (ver la siguiente llamada a pie de página).

Cada disco (bola) se mueve libremente, o sea con velocidad constante, hasta que choca con otro disco (bola). Cuando dos chocan, cambian su velocidad de acuerdo con las leyes de la colisión elástica. Éstas se manifiestan así. Suponga que dos bolas chocan. Denote por q_1 y q_2 sus centros y por v_1 y v_2 sus vectores velocidad en el momento de la colisión. Sea L la línea entre sus centros. Descomponemos $v_i = v_i^0 + v_i^\perp$ para $i = 1, 2$, donde v_i^0 es la componente del vector v_i paralelo a L y v_i^\perp es perpendicular a L . Las nuevas velocidades, salientes, de las bolas, son:

$$v_1^{\text{nueva}} = v_1^\perp + v_2^0 \quad \text{y} \quad v_2^{\text{nueva}} = v_2^\perp + v_1^0$$

En otras palabras, las bolas intercambian las componentes de las velocidades paralelas a la línea de los centros y mantienen sus componentes ortogonales. Observamos que estas leyes implican la preservación de la energía cinética total $\sum m \|v_i\|^2/2$ y del momento total $\sum m v_i$ del sistema de las n bolas (discos). También observamos que una colisión de dos bolas duras con centros q_1 and q_2 puede sólo ocurrir si $\text{dist}(q_1, q_2) = 2r$, o sea $\|q_1 - q_2\|^2 = (2r)^2$.

El sistema de n bolas (discos) moviéndose en el espacio abierto (o en el plano) sin paredes no es dinámicamente muy interesante. Es intuitivamente claro que el número de choques es siempre finito, y que luego de la última colisión las bolas se irán libremente para siempre. Más aún, el número de colisiones entre n bolas en el espacio abierto está uniformemente acotado por una constante que sólo depende de n . Este último resultado fue probado muy recientemente – en 1998 – por Burago, Ferleger and Kononenko [BFK98b]; ver, más abajo, Corolario ??.

Consideremos n bolas o discos moviéndose en un dominio acotado cerrado R , llamado el contenedor (o reservatorio). Las bolas (discos) chocan elásticamente entre ellas y con las paredes del contenedor. Si la bola con centro q toca a la pared en el punto $w \in \partial R$, entonces descomponemos su vector velocidad como $v = v^0 + v^\perp$, donde v^0 es una componente paralela a la línea pasando por q y w , y v^\perp es perpendicular a esa

línea. La nueva velocidad, saliente, de la bola es $v^{\text{nueva}} = v^\perp - v^0$. Note que esta regla preserva la energía cinética total del sistema, pero **no preserva** el momento total.

Ahora reducimos el sistema de n bolas duras en el contenedor R a un billar. Denotamos $q_i = (q_i^1, q_i^2, q_i^3)$ el centro de la i ésima bola y $v_i = (v_i^1, v_i^2, v_i^3)$ su vector velocidad, $1 \leq i \leq n$. Para discos en el plano tenemos dos coordenadas en lugar de tres. Todo el sistema puede ser descrito por el punto de configuración

$$q = (q_1^1, q_1^2, q_1^3, q_2^1, \dots, q_n^2, q_n^3) \in \mathbb{R}^{3n}$$

y su vector velocidad

$$v = (v_1^1, v_1^2, v_1^3, v_2^1, \dots, v_n^2, v_n^3) \in \mathbb{R}^{3n} \quad (\text{IV.5.6})$$

(para sistemas de discos hay que reemplazar $3n$ por $2n$).

Observe que $q \in R^n$. No toda la región R^n (R es el contenedor) es alcanzable por el punto de configuración q : las bolas no pueden sobreponerse. Esto requiere la exclusión de las configuraciones que satisfacen

$$(q_i^1 - q_j^1)^2 + (q_i^2 - q_j^2)^2 + (q_i^3 - q_j^3)^2 < (2r)^2 \quad (\text{IV.5.7})$$

para $1 \leq i < j \leq n$ (aquí r es el radio de las bolas). La desigualdad (IV.5.7) determina cilindros esféricos en \mathbb{R}^{3n} , que denotamos C_{ij} . Para el modelo de los discos duros en el plano, obtenemos cilindros circulares C_{ij} en \mathbb{R}^{2n} . Los cilindros C_{ij} , $1 \leq i < j \leq n$, contienen todas las configuraciones prohibidas de las bolas (discos), por lo que deben ser sacados del espacio accesible. Como consecuencia obtenemos un dominio menor $Q = R^n \setminus \cup_{i \neq j} C_{ij}$; este dominio Q es llamado *espacio de configuración* del sistema.

Ahora se puede inspeccionar directamente (se lo dejamos al lector; es más bien tedioso) que la trayectoria del punto de configuración q en Q es gobernada por las reglas del billar (IV.1.2) y (IV.1.3). Las reflexiones especulares en la superficie de los cilindros C_{ij} corresponden a las colisiones entre las bolas i, j . Por tanto, el estudio del modelo mecánico de las n bolas o discos es reducido al estudio de la dinámica de billar en el dominio Q . La conservación de la energía cinética total $\sum_i m \|v_i\|^2 / 2$ es equivalente a la preservación de la norma $\|v\|$ del vector velocidad. (IV.5.6).

El conjunto de singularidad Γ^* contiene todas las intersecciones de las superficies cilíndricas ∂C_{ij} entre sí. Tales intersecciones corresponden a las colisiones de tres o más bolas. El resultado de tales colisiones múltiples no es definido. Nuestra regla general es ignorar las trayectorias que pasan por Γ^* .

El gas de las bolas duras es un modelo clásico en física estadística. Su estudio se remonta a L. Boltzmann en el siglo XIX. Muchas leyes de la física habían sido enunciadas para gases de bolas duras, y luego verificadas experimentalmente para otros gases. Boltzmann fue el primero en enunciar la célebre *hipótesis ergódica*. Él supuso que los gases de bolas duras son, en general, ergódicos y usó esta suposición para justificar las leyes de la mecánica estadística (en un nivel “heurístico”). Desde entonces, permanece como un gran desafío para físicos y matemáticos probar esta hipótesis, así como hacer uso de la ergodicidad para construir los fundamentos matemáticos de la mecánica estadística.

Al comienzo de los 60, Ya. Sinai estudió una versión específica del modelo de Boltzmann –el gas de n bolas (o discos) duras en el toro \mathbb{T}^d , $d \geq 2$. En este caso el contenedor R es un toro, por lo que no hay paredes ($\partial R = \emptyset$). Por tanto, las bolas sólo chocan entre ellas, por lo que además de la energía cinética total, se conserva el momento total. Sinai conjeturó que si uno establece que el momento total es cero y fija el centro de masa, entonces el sistema reducido resultante debería ser ergódico.

Intentos por probar la conjetura de Boltzmann-Sinai se extendieron por 40 años, y tuvieron una colorida y a veces dramática historia descrita en [Sz96]. Ver también [Sz00]. Importantes avances se han hecho con trabajos recientes de N. Simányi and D. Szász [SS99, Sm99]. Pero esta solución está mucho más allá de los objetivos de estas notas.

Gas de Lorentz. He aquí otro modelo popular en física. Imagine una partícula puntual moviéndose entre bolas rígidas fijas en el espacio \mathbb{R}^3 . Las bolas son del mismo radio r . Pueden ser colocadas o bien aleatoriamente o bien formar una estructura regular del tipo de un cristal (por ejemplo sus centros pueden estar colocados en los sitios de la red \mathbb{Z}^3). Las bolas no se mueven, sólo la partícula puntual se mueve entre ellas y choca con las bolas elásticamente. Las bolas actúan como obstáculos.

Este modelo fue introducido por H. Lorentz en 1905 [Lo05] para el estudio de electrones en metales. La partícula puntual móvil representaba a los electrones y las bolas jugaban el papel de las moléculas de metal.

Cuando las bolas son colocadas en los sitios de una malla regular, el gas de Lorentz se dice periódico (como opuesto a aleatorio). Aquí sólo trataremos gases periódicos de Lorentz.

En lugar de bolas, podemos colocar otros objetos idénticos en los sitios de la malla, y requerir que la partícula choque las fronteras de los obstáculos elásticamente. Los cuerpos pueden tener formas cúbicas o poliédricas más generales. Varias modificaciones de este modelo son estudiadas en física [?]. Uno puede también definir el gas de Lorentz en el plano; en este caso los obstáculos bidimensionales están colocados en los sitios de una malla bidimensional, por ejemplo, \mathbb{Z}^2 .

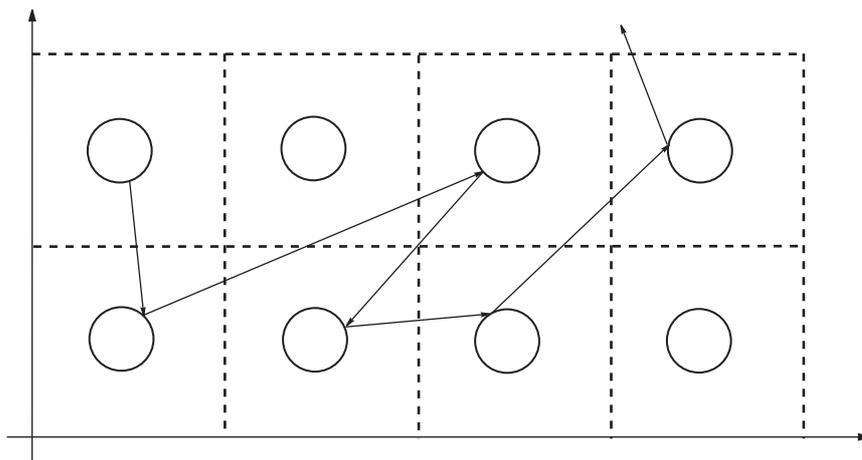


Figura IV.10: Un gas periódico de Lorentz.

Ahora reducimos el gas de Lorentz a un billar. En realidad es “casi” un billar, porque el dominio accesible por la partícula móvil no es acotado y tiene volumen infinito. Esto puede ser evitado fácilmente. Como los obstáculos están colocados en los sitios de una malla, su estructura es periódica, y uno puede encontrar un *dominio fundamental* D , cuyas traslaciones paralelas pueden cubrir todo el espacio. El dominio D contiene unos pocos (a veces uno) obstáculos, y todos los otros obstáculos son obtenidos por traslaciones paralelas de aquellos en D . La Figura IV.10 muestra un gas periódico de Lorentz en el plano, en el que discos fijos forman una matriz periódica, y el dominio fundamental es un cuadrado conteniendo un disco.

El movimiento de la partícula puede ser proyectado al dominio fundamental D , y obtener un sistema de billar en D con condiciones de frontera periódicas, o sea un billar en un toro \mathbb{T}^d , $d = 2, 3$. El dominio del billar Q se obtiene removiendo uno o varios obstáculos del toro. Tal ejemplo fue ya mostrado en la Figura IV.9 de la Sección IV.1.

IV.6. Otras Propiedades Geométricas

En esta Sección estudiaremos dos propiedades de billares que tienen naturaleza más geométrica que dinámica, aunque la primera involucra a la medida invariante.

Paso libre medio. Recuerde que la función $\tau(x)$ en M es el tiempo de retorno del flujo del billar, que define a la transformación $T : M \rightarrow M$, o el tiempo entre dos reflexiones en los puntos x y Tx . Como la velocidad de la partícula tiene módulo uno, $\tau(x)$ también es igual a la distancia en Q entre esos puntos de reflexión, que es llamado *paso libre* de la trayectoria del billar. Queremos estimar el *paso libre medio*, o sea el valor asintótico

$$\bar{\tau}(x) = \lim_{n \rightarrow \infty} \frac{\tau(x) + \tau(Tx) + \dots + \tau(T^{n-1}x)}{n} \tag{IV.6.1}$$

por el Teorema Ergódico de Birkhoff-Khinchin, el valor $\bar{\tau}(x)$ existe casi todo punto en M y su valor promedio

es

$$\bar{\tau} := \int_M \bar{\tau}(x) d\nu(x) = \int_M \tau(x) d\nu(x). \quad (\text{IV.6.2})$$

Si la transformación de billar T es ergódica, entonces, la función $\bar{\tau}(x)$ es constante casi todo punto e igual a $\bar{\tau}$.

El cálculo de $\bar{\tau}$, en término de características del dominio Q , es extraordinariamente simple. Escribimos $\tau(x) = \int_0^{\tau(x)} ds$ donde s es el parámetro introducido en (IV.5.2), entonces

$$\bar{\tau} = \int_M \tau(x) d\nu(x) = c_\nu \int_M \int_0^{\tau(x)} \langle v, n(r) \rangle ds dr d\phi$$

Entonces usamos la identidad (IV.5.2) y obtenemos

$$\bar{\tau} = \frac{c_\nu}{c_\mu} \int_{\mathcal{M}} d\mu = \frac{c_\nu}{c_\mu} \quad (\text{IV.6.3})$$

Usando ahora (IV.5.1) y (IV.5.3) completamos el cálculo:

$$\bar{\tau} = \frac{|Q| \cdot |S^{d-1}|}{|\partial Q| \cdot |B^{d-1}|} \quad (\text{IV.6.4})$$

Observamos que el paso libre medio entre colisiones sólo depende del volumen del dominio Q y del área de su frontera, pero no de su forma.

En particular, para billares planos $d = 2$, tenemos $|S^1| = 2\pi$, $|B^1| = 2$, y la fórmula (IV.6.4) se vuelve muy simple: $\bar{\tau} = \frac{\pi|Q|}{|\partial Q|}$.

La fórmula (IV.6.4) es bien conocida en probabilidad geométrica y geometría integral. Su versión plana es frecuentemente llamada *fórmula de Santaló*, porque apareció por primera vez en el libro [Sa76].

Por ejemplo, considere otra vez una mesa de billar Q en el toro unitario \mathbb{T}^2 donde un pequeño disco D de radio r es removido, como se ve en la Figura IV.9. Claramente, para pequeño r la partícula del billar puede moverse libremente por largo tiempo sin colisiones con el disco D , y la función $\tau(x)$ puede tomar valores arbitrariamente grandes. La fórmula de Santaló (??) da su valor medio:

$$\bar{\tau} = \frac{\pi(1 - \pi r^2)}{2\pi r} = \frac{1 - \pi r^2}{2r}$$

o sea que el paso libre medio es asintóticamente igual a $\frac{1}{2r}$ cuando $r \rightarrow 0$. Veremos más adelante que T es ergódica en ese ejemplo, por lo que $\bar{\tau}(x) = \bar{\tau}$ casi todo punto.

Como un sistema mucho más complicado, considere un sistema de N bolas duras en el toro d -dimensional. Se reduce a un billar en el dominio multidimensional Q cuyas fronteras consisten de cilindros (recorda: $\dim Q = Nd$). El paso libre medio entre colisiones puede ser estimado por (IV.6.4). Esto requiere calcular el volumen de Q y el área de la frontera. Esta es una tarea difícil pero realizable, que fue hecha en [C97]. La expresión final coincide con la clásica fórmula de Boltzmann para el tiempo de inercolisión media usado en física estadística durante décadas. El lector interesado puede recurrir a [C97] y a las referencias allí citadas.

Estimaciones del número de reflexiones. Consideraemos el siguiente problema: dado un trozo de trayectoria de billar de longitud L , cuántas reflexiones en ∂Q pueden haber en ese trozo? En particular, puede el número de reflexiones, n , ser infinito? Mostramos en la Sección IV.1 que $n < \infty$ con probabilidad uno. Pero desde un punto de vista geométrico, a uno le gustaría saber si n puede ser alguna vez infinito, y cuan grande puede ser n . Estas preguntas también surgen en el estudio de propiedades ergódicas de billares.

Comenzamos con un caso simple: una trayectoria de billar moviéndose entre dos líneas, l_1 y l_2 , que se intersecatan en un punto A con ángulo $\alpha > 0$, ver Figura IV.11. Llamamos Q al dominio (infinito) limitado por esas líneas. Cuando la trayectoria corta una de esas líneas, digamos l_1 , se refleja, pero su imagen especular a través de l_1 continuará moviéndose derecho del otro lado de l_1 . Seguiremos la imagen especular, en lugar de la trayectoria misma. Continuará moviéndose en un dominio Q_1 que es la imagen especular de Q a través de la línea l_1 . Cuando nuestra trayectoria corta a la otra línea l_2 , su imagen especular también corta el otro

lado del dominio Q_1 , etc. Estaremos reflejando los dominios Q_i a través de sus lados y siguiendo la línea recta de las imágenes especulares de nuestra trayectoria. Esto lucirá como un cuarto de espejos en un parque de entretenimientos, con muchos reflexiones en distintos espejos. Este método de reflejar la mesa de billar Q a través de sus lados e ir siguiendo las imágenes de la trayectoria es el *desdoblamiento* – *unfolding*, ya introducido en la Sección IV.2.

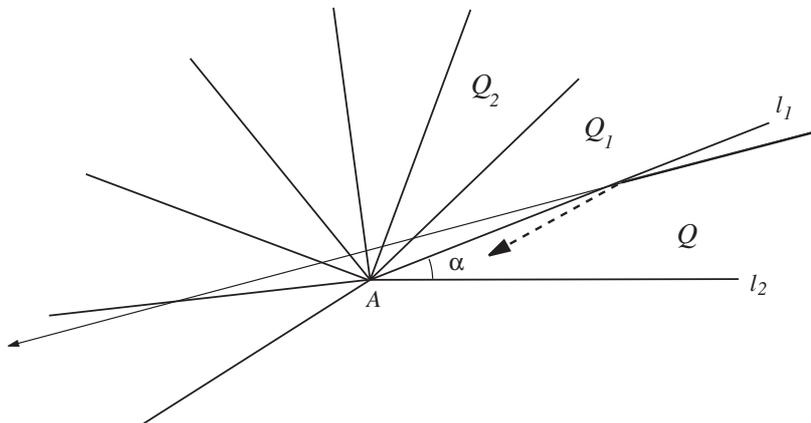


Figura IV.11: El desdoblamiento –unfolding– de una trayectoria de billar.

En la Figura IV.11 queda claro que el ángulo total en las imágenes desdobladas Q_1, Q_2, \dots , entre la primera y la última reflexión no puede exceder π , por lo que

$$n < \frac{\pi}{\alpha} + 1$$

Esta simple estimación da una cota superior para el número de reflexiones en Q . Notamos que esta cota es uniforme: es la misma para todas las trayectorias de billar en Q .

La estimación anterior tiene una versión multidimensional. Suponga que varios hiperplanos en \mathbb{R}^d , $d \geq 2$, se intersectan en un punto A , de modo que forman un “ángulo poliédrico” con vértice en A . Ya. Sinai probó en 1978 [Si78a] que el número de reflexiones de cualquier trayectoria de billar dentro de tal ángulo está uniformemente acotada, la cota depende sólo de la configuración de los hiperplanos.

Es más difícil obtener estimaciones del número de reflexiones en billares con fronteras curvilíneas. Aquí tenemos dos casos distintos. Uno es un dominio Q con fronteras convexas, como un disco en un plano o una bola en \mathbb{R}^d . Cerca de una frontera convexa ∂Q , un corto trozo de trayectoria puede experimentar un cantidad arbitraria de reflexiones. Esto sucede cuando el vector velocidad v en un punto de reflexión $q \in \partial Q$ es casi tangente a la frontera ∂Q . Tal trayectoria simplemente “resbalará” a lo largo de ∂Q experimentando muchas colisiones “rozantes” con ∂Q en rápida sucesión.

Incluso es posible construir mesas de billar convexas $Q \subset \mathbb{R}^2$ en que una corta trayectoria experimente un número infinito de reflexiones que se acumulan en un punto de ∂Q en que la curvatura se anula. Tales ejemplos “anómalos” fueron encontrados por Halpern [Hn77].

Por lo contrario, cuando la frontera de Q es cóncava (convexa hacia afuera), el número de reflexiones puede ser acotado. Para imaginarse una mesa de billar con frontera cóncava, tome un polígono y arquée un poco cada lado hacia adentro. O recuerde la mesa en un toro en que se ha sacado un disco; ver Figura IV.9.

Es claro que cerca de una pieza suave cóncava de ∂Q cualquier trayectoria corta puede tener una sola reflexión. Considere ahora un vértice A donde se encuentran dos (o más piezas cóncavas de ∂Q). Estimaciones del número de reflexiones cerca de tales vértices han sido muy estudiadas por Galperin [Ga81], Vasserstein [Va79] y otros. aquí presentamos la estimación más general obtenida por Burago, Ferleger & Kononenko en 1998 [BFK98b].

Para definir un vértice en \mathbb{R}^d , uno puede considerar una cantidad finita de subconjuntos cerrados convexos

$B_i \subset \mathbb{R}^d$, $i = 1, \dots, n$, cuyas fronteras son hipersuperficies C^1 y definir un dominio de billar

$$Q = \mathbb{R}^d \setminus (\cup_{i=1}^n B_i) \quad (\text{IV.6.5})$$

Es suficiente suponer que $B := \bar{Q} \cap (\cap_{i=1}^n B_i) \neq \emptyset$ y considerar trayectorias de billar cortas cerca de B (el conjunto B juega el papel de un vértice).

Para cualesquiera dos puntos $X, Y \in Q$ denotamos $T(X, Y)$ el trozo de trayectoria de billar que comienza en X y termina en Y (si existe una), y por $|T(X, Y)|$ su longitud. El siguiente lema compara $|T(x, y)|$ con la distancia de X e Y al “fondo del vértice” – el conjunto B .

Lema IV.6.1 (Lema de Comparación). *Para cada $X, Y \in Q$ y cada $A \in B$*

$$|XA| + |AY| \geq |T(X, Y)|$$

La desigualdad es estricta si una de las reflexiones ocurre en una parte estrictamente cóncava de la frontera de Q .

Demostración. Denote $X_i \in \partial Q$, $i = 1, \dots, k$, los puntos de reflexión de la trayectoria $T(X, Y)$ y ponga $X_0 = X$, $X_{k+1} = Y$. Considere la superficie triangulizada por los triángulos $AX_i X_{i+1}$, $i = 0, \dots, k$. Sea $A'X'_0 X'_1 \cdots X'_{k+1} \subset \mathbb{R}^2$ el *desdoblamiento -unfolding-* (o *desarrollo*) de esta superficie al poner esos triángulos en un plano con lados adyacentes, ver Figura IV.12. En el Ejercicio IV.21, se prueba que la curva $\gamma = X_0 \cdots X_{k+1}$ es convexa: para cualquier $i = 0, \dots, k$ ella queda de un lado de la línea $X_i X_{i+1}$ (opuesta al punto A'). La convexidad de γ implica $|XA| + |AY| = |X'_0 A'| + |A' X'_{k+1}| \geq \sum_{i=0}^k |X'_i X'_{i+1}| = \sum_{i=0}^k |X_i X_{i+1}| = |T(X, Y)|$. \square

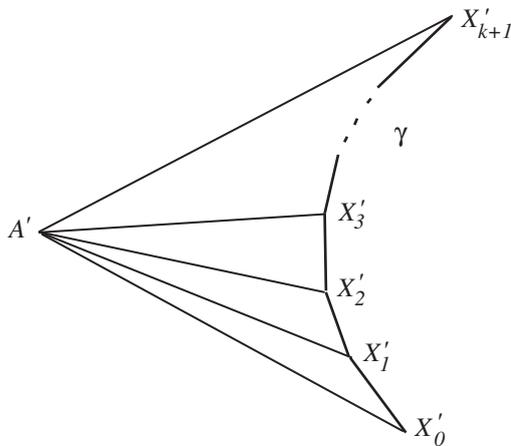


Figura IV.12: El desarrollo de una superficie triangulizada del plano.

Teorema IV.6.2 ([Ga81], [Va79]). *Para cualquier trayectoria de billar de longitud finita en Q el número de reflexiones es finito.*

Demostración. Suponga lo opuesto – una trayectoria comenzando en $X \in Q$ tiene infinitas reflexiones que se acumulan en un punto $A \in B$ (si $A \notin B$, podemos remover algún B_i de nuestra construcción). Sean X_1, X_2, \dots los puntos de reflexión, y $X_i \rightarrow A$ cuando $i \rightarrow \infty$. La longitud del segmento $X_1 A$ es menor que la longitud $|T(X_1, A)|$ de la trayectoria entera entre X_1 y A . Por lo tanto podemos encontrar X_k suficientemente cerca de A de modo que $|X_1 A| + |AX_k| < |T(X_1, X_k)|$, lo cual contradice el Lema de Comparación. \square

Una cota uniforme en el número de reflexiones requiere algunas condiciones más en el dominio del billar. En efecto, si el vértice A de una mesa de billar plano $Q \subset \mathbb{R}^2$ con frontera cóncava forma una cúspide,

formada por dos curvas cóncavas tangentes en A , entonces una corta trayectoria de billar puede tener una cantidad arbitraria de reflexiones cerca de A ; ver Ejercicio IV.22. Por tanto, es necesario algún tipo de transversalidad de las B_i 's en su intersección B . tal condición fue encontrada en [BFK98b]:

Definición 19. *Un dominio de billar Q dado por (IV.6.5) es no degenerado en el subconjunto $U \subset \mathbb{R}^d$ con constante $C > 0$ si para cada $I \subset \{1, \dots, n\}$ y para cualquier $y \in (U \cap Q) \setminus (\cap_{i \in I} B_i)$*

$$\max_{k \in I} \frac{\text{dist}(y, B_k)}{\text{dist}(y, \cap_{i \in I} B_i)} \geq C$$

siempre que $\cap_{i \in I} B_i$ sea no vacío.

Groseramente hablando, esto significa que si un punto está d -cerca de toda pared de I , entonces está d/C -cerca de su intersección.

Teorema IV.6.3 ([BFK98b]). *Sea un billar Q semidispersor no degenerado en un dominio abierto $U \subset \mathbb{R}^d$. Entonces, para cualquier punto $x \in U$ existe un número $M_x < \infty$ y un entorno menor U_x de x tales que toda trayectoria de billar entrando en U_x lo deja, haciendo no más de M_x colisiones con la frontera ∂Q .*

El sistema de N bolas duras de masa y radio arbitrarios en un espacio abierto \mathbb{R}^d puede experimentar no más de un cierto número $M < \infty$ de colisiones. Aquí

$$M = \left(32 \sqrt{\frac{m_{\text{máx}}}{m_{\text{mín}}}} \frac{r_{\text{máx}}}{r_{\text{mín}}} N^{\frac{3}{2}} \right)^{N^2}$$

donde $m_{\text{máx}}$ y $m_{\text{mín}}$ son las masas máximo y mínimo, y $r_{\text{máx}}$ y $r_{\text{mín}}$ son los radios máximo y mínimo de las bolas, respectivamente.

Es importante observar que todos los resultados anteriores admiten pruebas geoméricamente bastante elementales; ver [BFK98b].

EJERCICIO IV.21. Probar que la curva $\gamma = X'_0 \cdots X'_{k+1}$ definida en la prueba del Lema de Comparación es convexa. Sugerencia: Observe la relación entre el “desarrollo” de la línea tangente en X_i y la línea de desdoblamiento $A'X'_i$ y use el hecho de que el ángulo de reflexión es igual al ángulo de incidencia.

EJERCICIO IV.22. Sea $Q \subset \mathbb{R}^2$ una mesa de billar con un vértice A en el que se tocan dos componentes cóncavas $\gamma_1, \gamma_2 \subset \partial Q$ de la frontera, formando una cúspide (las curvas se encuentran con ángulo cero). Mostrar que una trayectoria de billar corta puede experimentar una cantidad arbitraria de reflexiones en ∂Q cerca del punto A . Sugerencia: Considere una línea tangente común L a γ_1 y γ_2 en A y comience la trayectoria en $L \cap Q$ con vector velocidad inicial apuntando casi, pero no exactamente, hacia A . Muestre que en cada reflexión la dirección del vector velocidad va cambiando una cantidad arbitrariamente pequeña, por lo que lleva una cantidad arbitrariamente grande de reflexiones hacer que trayectoria se de vuelta y comience a alejarse de A .

Capítulo V

Billares. Hiperbolicidad

Los billares caóticos están caracterizados en primer lugar por la no anulación de sus exponentes de Liapunov, y luego por la entropía positiva, la ergodicidad, el mixing y la propiedad Benoulli. Nos restringiremos a billares planos, o sea $d = 2$; como ya fue observado en la Sección IV.5, la teoría de billares multidimensionales está menos desarrollada.

V.1. Frentes de Onda y Conos.

Recuerde que los exponentes de Liapunov de una transformación $T : M \rightarrow M$ son propiedades de vectores tangentes $u \in \mathcal{T}M$. Estos admiten una representación geométrica explícita en los sistemas de billares. En cualquier variedad, un vector tangente $u \in \mathcal{T}_x M$ en un punto $x \in M$ puede ser representado por una curva infinitesimal $\gamma \subset M$ que pasa por x con dirección u . En un billar, las trayectorias de los puntos $y \in \gamma$, al dejar la frontera ∂Q y entrar en el dominio Q , definen una familia a un parámetro de líneas orientadas, que llamaremos un *haz de rayos* (*frente de onda*).

Por otro lado, un cono $C \in \mathbb{R}^2$ (ver definición en Sección III.4) puede ser concebido como un subconjunto cerrado y conexo del espacio proyectivo \mathbb{P}^1 . Dada una coordenada proyectiva α de \mathbb{P}^1 , podemos identificar C con un intervalo cerrado de valores de α .

Coordenadas relacionadas geoméricamente con familias de trayectorias de billar pueden ser particularmente útiles y hacer que las pruebas de muchos resultados de la teoría de billares resulten más transparentes. La coordenada que usaremos será el tiempo de focalización de familias infinitesimales de trayectorias. En esta Sección, después de analizar frentes de onda empezando por los referidos a billares dispersores (siguiendo de alguna manera la evolución histórica del concepto), estudiaremos en general la evolución de los frentes de onda, daremos una definición precisa de tiempo de focalización de un vector y discutiremos algunas de sus propiedades.

Frentes de onda. Muchas veces es más conveniente trabajar con haces de rayos en el dominio Q que con curvas γ en M . Tomemos una sección ortogonal de ese haz, que pasa por el punto $x = (q, v)$. La llamamos Σ , ver Figura V.1. Es una curva Q que intersecta cada rayo del haz perpendicularmente. Los vectores velocidad de los puntos de la curva son entonces perpendiculares a ella. Por tanto, Σ es una curva suave equipada con una familia de vectores normales apuntando en la dirección del movimiento. Por ello llamamos a Σ un *frente de onda*, término que se pidió prestado de la física.

La curvatura del frente Σ juega un papel crucial en nuestro análisis. El signo de la curvatura se elige de acuerdo a la siguiente regla. Si el frente Σ es *divergente*, como se muestra en la Figura V.1, entonces su curvatura es positiva. Si el frente es *convergente*, su curvatura es negativa. Si el haz es de líneas paralelas (el frente Σ es perpendicular a ellas), entonces la curvatura es nula, y diremos que el frente es *neutro*.

La curvatura de la frontera ∂Q jugará un papel igualmente importante. Recordar nuestra convención en Sección IV.1: la curvatura de la frontera es positiva si es cóncava (convexa para dentro, dispersora) y negativa si es convexa.

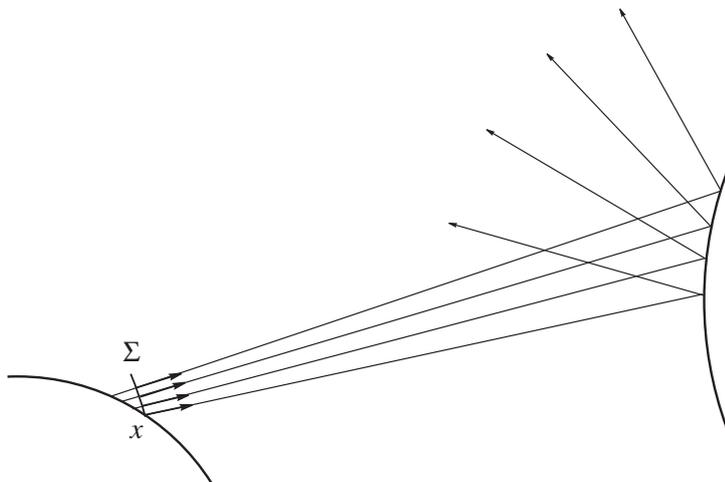


Figura V.1: Un frente de onda divergente en un billar de Sinai.

Definición 20. Si todas las componentes de la frontera ∂Q son dispersoras, se dice que el billar es dispersor. Si ∂Q consiste de curvas dispersoras y neutras, se dice que el billar es semidispersor.

Los billares dispersores son también llamados *billares de Sinai*, porque fueron introducidos por Ya. Sinai en 1970 y así comenzó el estudio matemático de los billares caóticos.

Es geoméricamente evidente que cuando un haz neutro (paralelo) de rayos llega a una componente dispersora, los rayos salientes (luego de la reflexión) forman un haz dispersor. Es también fácil ver que en las siguientes reflexiones en componentes dispersoras el haz se “abrirá” aún más, y sus rayos rápidamente divergerán entre ellos; ver Figura V.1. Esta es la principal razón para el crecimiento exponencial de los frentes de onda en los billares dispersores, que se traduce en la existencia de exponentes de Liapunov positivos. Ya sabemos (Sección IV.1) que la suma de los dos exponentes de Liapunov de la transformación $T : M \rightarrow M$ es cero, por lo que el otro será automáticamente negativo.

Resulta interesante que uno puede construir explícitamente el frente de onda $\Sigma^-(x)$ correspondiente al exponente de Liapunov negativo en $x \in M$. Debe converger exponencialmente, o sea achicarse a medida que el tiempo avanza. La existencia de tal frente recuerda la idea de *estabilidad* en la teoría de ecuaciones diferenciales. Recordar que una trayectoria (solución) de una ecuación diferencial, es estable si todas las trayectorias cercanas no se separan demasiado de ella en el futuro.

Comenzamos la construcción de $\Sigma^-(x)$. Sea $x = (q, v) \in \mathcal{M}$. Para un valor grande de $t > 0$ considere el punto $(q_t, v_t) = \Phi^t(q, v)$, o sea una imagen distante de (q, v) . De vuelta el tiempo por medio de la involución (definida en la Sección IV.4): tome $(q_t, -v_t) = I(q_t, v_t)$. Considere un frente de onda neutro Σ_t^+ que pasa por $(q_t, -v_t)$, o sea un haz de rayos paralelos por q_t que se mueven en la dirección de $-v_t$. Entonces, por la principal propiedad de la involución, la imagen $\Phi^t(\Sigma_t^+)$ será un frente de onda conteniendo el punto $(q, -v) = I(q, v)$. Su involución $\Sigma_t^- = I(\Phi^t(\Sigma_t^+))$ contendrá el punto original (q, v) .

En los billares de Sinai los frentes de onda paralelos se hacen divergentes y crecen exponencialmente rápido con el tiempo. Por tanto, el frente $\Phi^t(\Sigma_t^+)$ será exponencialmente (en t) más largo que Σ_t^+ (recordar que todos nuestros frentes son, de cualquier manera, infinitesimalmente pequeños). Por tanto, el frente $\Phi^t(\Sigma_t^-)$ será *exponencialmente más corto* que el frente Σ_t^- . Esto nos da un frente de onda que partiendo del punto dado (q, v) se achica exponencialmente durante el intervalo de tiempo $(0, t)$. Falta tomar el límite para $t \rightarrow \infty$ y obtener el frente $\Sigma^-(x) := \lim_{t \rightarrow \infty} \Sigma_t^-$ que se achica exponencialmente rápido para todo tiempo futuro. Esta es nuestra familia estable de trayectorias que corresponde al exponente de Liapunov negativo en (q, v) . Note que los frentes Σ_t^- y el frente límite $\Sigma^-(x)$ son convergentes (tienen curvatura negativa).

Además de las familias estables, representadas por los frentes convergentes, necesitaremos sus contrapartes en “marcha atrás”. Dado el punto $x = (q, v) \in \mathcal{M}$, sea $\Sigma^-(I(x))$ la familia estable construída para el punto $I(q, v) = (q, -v)$. Entonces el frente de onda $\Sigma^+(x) := I(\Sigma^-(I(x)))$ será divergente (tendrá curvatura

positiva). Crecerá exponencialmente rápido en el futuro y contraerá en el pasado. Por tanto su exponente de Liapunov será positivo.

Ahora deduciremos las ecuaciones exactas que describen la dinámica de los frentes de onda de cualquier billar.

Sea Σ_0 un frente de onda o sea, un arco C^1 equipado con vectores normales apuntando en la dirección del movimiento. Sea $(q, v) \in \Sigma_0$ uno de sus elementos. Sea χ_0 la curvatura del frente en el punto q cuyo signo se establece por las reglas ya vistas (positiva si es divergente). Considere el frente $\Sigma_t = \Phi^t(\Sigma_0)$, $t > 0$, que contiene al punto $(q_t, v_t) = \Phi^t(q, v)$. Calcularemos la curvatura χ_t del frente Σ_t en el punto q_t .

Primero suponga que t es suficientemente pequeño, de modo que en el intervalo $(0, t)$ no hay reflexiones en ∂Q . Entonces Φ^t evoluciona en movimiento libre por lo que la ecuación es muy simple:

$$\chi_t = \frac{1}{t + 1/\chi_0} \quad (\text{V.1.1})$$

Es bien conocida en óptica geométrica y se puede deducir directamente, prestando atención a las reglas de los signos de las curvaturas (es suficiente aproximar Σ_0 por un arco de círculo y recordar que $|\chi_0|$ es el inverso del radio de ese círculo).

Cuando el frente de onda rebota en la frontera ∂Q , su curvatura cambia instantáneamente. Entonces la curvatura de la frontera se “combina” con la curvatura del frente mismo. La siguiente es una de las leyes básicas de la óptica geométrica conocida como *ecuación del espejo -mirror equation-* o *ley de reflexión* (V.1.9). Sean χ_- y χ_+ , las curvaturas del frente antes y después de la reflexión, respectivamente, K_1 la curvatura de la frontera en el punto de reflexión (cuyo signo está dado por las convenciones ya aludidas), y ϕ_1 el ángulo de reflexión (ver Sección IV.1). Entonces la ecuación del espejo es

$$\chi_+ = \chi_- + \frac{2K_1}{\cos \phi_1} \quad (\text{V.1.2})$$

La prueba se deja como ejercicio (V.1). Es bastante complicado, pero recomendamos hacer algunos casos: ayudará a “ver qué es lo que sucede” en los choques.

Combinando las ecuaciones (V.1.1) y (V.1.2), se calcula la curvatura en cualquier tiempo t .

Fraciones continuas. A efectos de describir los frentes estable e inestable tal como lo hemos bosquejado anteriormente, es mejor expresar χ_0 en función de χ_t , $t > 0$. es lo que haremos a continuación. Sea χ_1 la curvatura del frente de onda inmediatamente después del primer reflexión en ∂Q . Entonces

$$\chi_0 = -\frac{1}{\tau + \frac{1}{\frac{2K_1}{\cos \phi_1} - \chi_1}}, \quad (\text{V.1.3})$$

donde τ es el momento de la primera reflexión. Luego, el valor de χ_1 puede ser expresado como función de la curvatura luego de la segunda reflexión, etc. Procediendo de esa manera en las siguientes reflexiones iremos agregando fracciones a nuestra fracción principal de la ecuación (V.1.3), extendiéndola hacia abajo. La fracción límite será infinita. Tales expresiones son llamadas *fracciones continuas*. Su valor es definido como el límite (si existe) de la sucesión de fracciones finitas, que resultan de truncar esa fracción en el paso n -ésimo. Esta fracción truncada se denomina *reducida n -ésima de la fracción continua*. Si $x = (q, v) \in M$, nuestra fracción continua (estable) es

$$\kappa^s(x) = -\frac{1}{b_1(x) + \frac{1}{b_2(x) + \frac{1}{b_3(x) + \frac{1}{b_4(x) + \dots}}}} \quad (\text{V.1.4})$$

$$b_{2k}(x) = \frac{2K(T^k x)}{\cos \phi(T^k x)}, \quad b_{2k+1}(x) = \tau(T^k x), \quad k \in \mathbb{N}. \quad (\text{V.1.5})$$

Aquí $K(T^k x)$ es la curvatura de la frontera en el punto $T^k x$ de la k -ésima reflexión, $\phi(T^k x)$ es el ángulo de esa reflexión y $\tau(T^k x)$ es el paso libre entre la k -ésima y la $(k+1)$ -ésima reflexión.

La fracción en la ecuación (V.1.4) tiene una estructura alternada –los componentes impares son τ 's y los pares son fracciones $2K/\cos\phi$. Esto refleja la alternancia natural de los pasos libres con las reflexiones a lo largo de la trayectoria, y se manifiesta en la estructura diferente de las reducidas pares e impares de la fracción continua.

Observamos que si uno trunca la fracción infinita (V.1.4) en un término par, agregando cero, el valor de la fracción truncada es la curvatura del frente del onda Σ_t^- antes construida, donde t es el momento de la reflexión.

Sea ahora

$$\kappa^u(x) = a_0(x) + \frac{1}{a_1(x) + \frac{1}{a_2(x) + \frac{1}{a_3(x) + \frac{1}{a_4(x) + \dots}}}} \quad (\text{V.1.6})$$

para $(x, v) \in M$, donde

$$a_{2k}(x) = \frac{2K(T^{-k}(x))}{\cos\phi(T^{-k}(x))}, \quad a_{2k-1}(x) = \tau(T^{-k}(x)), \quad k \geq 0. \quad (\text{V.1.7})$$

La fracción continua (V.1.6) es la curvatura del frente de onda inestable en $x = (q, v)$. Puede ser deducido de manera similar a como lo hicimos con (V.1.4) trabajando hacia el pasado (iteraciones negativas de T). El término extra $a_0(x)$ aparece debido a que (q, v) ya se reflejó en el primer punto de la trayectoria, por lo que al ir para atrás la reflexión en este punto implica la aplicación de la ecuación del espejo (V.1.2). La convergencia de la fracción (V.1.6) depende de los mismos factores.

En realidad, la convergencia de las fracciones continuas (V.1.4) , (V.1.6) no implica hiperbolicidad ni existencia de los exponentes de Liapunov. Por ejemplo, esas fracciones convergen para el billar circular (ver Ejercicio V.2) y para el billar elíptico, (sus valores no son nulos) pero los exponentes de Liapunov se anulan. L. Bunimovich probó que esas fracciones convergen para clases muy amplias de billares caóticos y no caóticos. Ver Sección V.4.

De hecho no hay ejemplos de billares cuyas fracciones dejen de converger en conjuntos de medida positiva. Por otro lado, para todos los billares caóticos las fracciones (V.1.4) y (V.1.6) representan la curvatura de los frentes estables, relacionados con los exponentes de Liapunov negativo y positivo, respectivamente.

Tiempos de focalización. Ahora introducimos, como coordenadas proyectivas, los tiempos de focalización de familias infinitesimales de trayectorias .

Una *variación* $\eta(\alpha)$ es una familia suave a un parámetro de rectas en \mathbb{R}^2 . Sea $u \in \mathcal{T}_z M, z \in M$ y $\xi : (-\varepsilon, \varepsilon) \rightarrow M$ una curva con $\xi(0) = z$ y $\xi'(0) = u$. Asociamos con u a la variación

$$\eta_+(\alpha) = q(\alpha) + t v(\alpha), \quad t \in \mathbb{R}$$

donde $\xi(\alpha) = (q(\alpha), v(\alpha))$. Decimos que $\eta_+(\alpha)$ *focaliza* si existe $f^+ = \lim_{\alpha \rightarrow 0} t(\alpha)$, donde $t(\alpha)$ está determinada por la intersección de las rectas $\eta_+(\alpha)$ y $\eta_+(0) = q(0) + t v(0)$. Ese valor, de existir es el *tiempo de focalización* (para adelante). También consideramos la variación

$$\eta_-(\alpha) = q(\alpha) + t \tilde{v}(\alpha), \quad t \in \mathbb{R}$$

donde el vector $\tilde{v}(\alpha)$ es obtenido reflejando $v(\alpha)$ en $q(\alpha) \in \partial Q$. Definimos el *tiempo de focalización* f^- de la variación $\eta_-(\alpha)$ de la misma manera que antes. A pesar de que para cada vector $u \in \mathcal{T}_z M$, podemos construir infinitas variaciones distintas $\eta_+(\alpha)$ ($\eta_-(\alpha)$), si una de ellas se focaliza, entonces todas las otras lo hacen, y sus tiempos de focalización son los mismos. Una definición alternativa para los tiempos de focalización es la siguiente: $\eta_+(\alpha)$ se focaliza si la componente perpendicular u de $\partial\eta_+/\partial\alpha|_{\alpha=0}$ es igual a cero, para algún $t = f^+ \in \mathbb{R}$. Decimos que el vector u se focaliza hacia el futuro (pasado) -también se usa hacia adelante (atrás)- si la variación $\eta_+(\alpha)$ ($\eta_-(\alpha)$) focaliza.

Sea $z = (r, \phi) \in M_i$ para algún $1 \leq i \leq n$ y $X_r = \partial/\partial r$, $X_\theta = \partial/\partial\theta$. Dado un vector $u = u_r \partial/\partial r + u_\phi \partial/\partial\phi \in \mathcal{T}M$ con $u_r, u_\phi \in \mathbb{R}$, su *tiempo de focalización para adelante y para atrás* $f^+(u), f^-(u)$ están dados por (see [W86])

$$f^\pm(u) = \begin{cases} \frac{-\cos\phi}{K(r) \pm m(u)} & \text{if } u_r \neq 0 \\ 0 & \text{if } u_r = 0 \end{cases} \quad (\text{V.1.8})$$

donde $m(u) = u_\phi/u_r$ es la pendiente de u . Decimos que u es *divergente* (*convergente*) si $f^+(u)$ es negativo (positivo), y u es *plano* si $f^+(u) = \infty$.

Lema V.1.1 (Ley de Reflexión). Sea $z = (r, \phi) \in M \setminus S_0$ tal que $Tz = (r_1, \phi_1) \in M \setminus S_0$ ¹. Para cada $u \in \mathcal{T}_z M$, sea $f_0 = f^+(u)$ y $f_1 = f^+(D_z T u)$. Entonces la ecuación del espejo (V.1.2) es equivalente a

$$\frac{1}{f_1} + \frac{1}{\tau(z) - f_0} = \frac{-2K(r_1)}{\cos \phi_1}. \quad (\text{V.1.9})$$

La curvatura $\chi = -1/f$ cambia de signo de $-$ a $+$ cuando focaliza. Si un frente de onda convergente arriva a una componente dispersora ($K(r_1) > 0$) seguirá o bien “menos convergente” (su tiempo de focalización será más corto de lo que lo sería sin reflexión), o bien divergente.

En condiciones adecuadas, el tiempo de focalización f tiene una importante propiedad de ordenamiento. Esta se utilizará en la prueba del Lemma V.1.3 que será crucial en la prueba de la hiperbolicidad de transformaciones de billar.

Lema V.1.2 (Propiedad de Ordenamiento de f). Sean $z \in M$ y $u, w \in \mathcal{T}_z M$. Suponga que $D_z T$ está definida, $0 < f^+(z, w) < \tau(z)$ y $0 < f^+(Tz, D_z T w)$. Entonces

$$0 \leq f^+(z, u) \leq f^+(z, w) \Rightarrow 0 < f^+(Tz, D_z T u) \leq f^+(Tz, D_z T w).$$

La consecuencia también es cierta si se reemplazan las desigualdades por desigualdades estrictas.

Demostración. Por la fórmula (V.1.9), tenemos que $1/f^+(D_z T w) < 1/f^+(D_z T u)$. □

Se puede ver inmediatamente de su definición que f^+ es una coordenada proyectiva en $\mathcal{T}_z M$. Un cono $C(z) \subset \mathcal{T}M$ es entonces identificado con un intervalo cerrado de f^+ , $C(z) = \{u \in \mathcal{T}_z M : f_1 \leq f^+(u) \leq f_2\}$ para algunos números reales f_1, f_2 tales que $-\infty \leq f_1 < f_2 \leq +\infty$.

Definición 21. Sea $C(z)$ un campo de conos en M_- y $C'(z)$ el campo de conos complementarios de $C(z)$. Para todo $z \in M_-$, defina

$$f^+(z) = \sup_{u \in C(z)} f^+(z, u), \quad f^-(z) = \sup_{u \in C'(z)} f^-(z, u).$$

El siguiente resultado da un criterio simple que ayuda a verificar cuando un campo de conos en M_- es invariante. Fue probado por primera vez en [Do91].

Lema V.1.3 (Lema de Focalización). Sea $z \in M_- \setminus S^+$ tal que Tz está definida y $Tz \in M_-$. Suponga que $0 \leq f^+(z, u) \leq \tau(z)$ para cada $u \in C(z)$ y $0 < f^-(Tz) \leq \tau(z)$. Entonces

$$0 < f^+(z) + f^-(Tz) \leq \tau(z) \Rightarrow D_z T C(z) \subseteq C(Tz),$$

y si reemplazamos la desigualdad en el lado izquierdo por una desigualdad estricta, el término de la derecha debe ser reemplazado por $D_z T C(z) \subset \text{int}(C(Tz))$ (invariancia estricta).

Demostración. Sea $w_1 \in C(Tz)$ tal que $f^+(Tz, w_1) = f^-(Tz)$, lo cual significa que w_1 está en uno de los lados de $C(Tz)$. Si $w = D_{Tz} T^{-1} w_1$, entonces $f^+(z, w) = \tau - f^-(Tz)$. Si $u \in C(z)$, entonces $f^+(z, u) \leq f^+(z)$. Por hipótesis tenemos $f^+ \leq (<) \tau - f^-(Tz) = f^+(z, w)$ por lo que $f^+(z, u) \leq (<) f^+(z, w)$. La prueba se completa aplicando el Lema V.1.2.

Ahora enunciamos una versión del Lema de Focalización para trayectorias que no chocan en otras componentes focalizadoras entre medio de dos pasadas por componentes focalizadoras.

Sean $\tilde{\gamma}_1, \tilde{\gamma}_2$ arcos focalizadores de ∂Q (no son necesariamente diferentes). Suponga que la trayectoria de un punto z deja $\tilde{\gamma}_1$ y vuelve por primera vez a la parte estrictamente convexa de ∂Q al cabo de $n \geq 1$ iteraciones por T , al punto $z' = (q', v')$, $q' \in \tilde{\gamma}_2$. Más precisamente, considere una órbita $\{z, Tz, \dots, T^n z\}$, $n \geq 1$, tal que $\pi(z) \in \tilde{\gamma}_1$, $\pi(T^k z) \notin M_-$, $k = 1, \dots, n-1$ y $\pi(T^n z) = z' \in \tilde{\gamma}_2$. Denotamos $\lambda(z)$ la suma de las longitudes de los segmentos de trayectoria entre z y z' , i.e., $\lambda(z) = \sum_{i=0}^{n-1} \tau_i$.

Lema V.1.4. Para $z \in M_-$ defina, como se vió antes, $z' = T^n z$. Sea $u \in C(z)$, $u' = D_z T^n u$, y $w \in C_{z'}$. Entonces

$$0 < f^+(z) + f^-(z') \leq \lambda(z) \Rightarrow m(u') > m(w).$$

¹Más adelante -Sección V.4- definiremos el conjunto S_{-1} de modo que estos puntos z serán de $M \setminus S_{-1}$

Demostración. Se deduce de la prueba original del Lemma 1.6 en [Do91] y de las observaciones inmediatamente posteriores a la formula (V.1.9). \square

EJERCICIO V.1. Probar la ecuación (V.1.2). Sugerencia: fije el signo de K y χ_- ; por ejemplo, suponga primero que $K > 0$ y $\chi_- > 0$ (frente divergente y frontera dispersora), y luego analice los demás casos.

EJERCICIO V.2. Probar que para el billar en un círculo de radio R la expresión (V.1.6) converge a $-1/\cos\phi$. (Recuerde que los exponentes de Liapunov se anulan!) Sugerencia: $a_{2k+1}(x) = \frac{-2}{R\cos\phi}$, $a_{2k}(x) = 2R\cos\phi$.

EJERCICIO V.3. Probar la convergencia de la fracción continua (V.1.4) para billares dispersores. Sugerencia: Considere las fracciones truncadas en componentes pares e impares, llámelas A_{2n} y A_{2n+1} , respectivamente. Pruebe que $A_{2n+2} < A_{2n+4} < A_{2n+3} < A_{2n+1}$ para cada $n \geq 0$. También observe que la suma de los elementos impares es infinito (la convergencia de fracciones continuas con elementos positivos es conocida como Teorema de Seidel-Stern theorem).

V.2. Billares Dispersores.

Las definiciones de billares dispersores (de Sinai) y semidispersores fueron dadas en la Sección anterior. También concluimos que en los dispersores los frentes de onda paralelos se hacen divergentes y crecen exponencialmente con el tiempo.

Para los billares de Sinai, las componentes de la frontera son dispersoras, por tanto $K(T^k x) > 0$. Como $\tau(T^k x) > 0$ y $\cos\phi(T^k x) > 0$ en todos los casos, todos los términos de nuestra fracción $\kappa^s(x)$ son positivos. Esto simplifica la verificación de la convergencia de (V.1.4). Esto fue observado en primer lugar por Sinai en [Si70]; ver Ejercicio V.3.

El valor límite de la fracción en (V.1.4) es por tanto positivo, por lo que $\kappa^s(x)$ es negativo. Entonces, vemos una vez más que el frente de onda estable es convergente. Para billares dispersores la fracción (V.1.6) es siempre convergente y $\kappa^u(x)$ es positiva, o sea que el frente inestable es divergente.

Teorema V.2.1. *En todo billar dispersor, las fracciones (V.1.4) y (V.1.6) convergen en cada punto $x \in \mathcal{M}'$. El valor $\kappa^s(x) < 0$ da la curvatura el frente de onda estable. El valor $\kappa^u(x) > 0$ da la curvatura del frente de onda inestable.*

En los billares semidispersores, la curvatura de las componentes neutras de la frontera es cero, por lo que algunos términos impares en (V.1.4) y (V.1.6) se anulan (ver también el Ejercicio V.5). Esto no afecta la convergencia, por lo que $\kappa^s(x)$ y $\kappa^u(x)$ aún existen en cada punto $x \in M'$. Sin embargo podrían ser igual a cero. Si $\kappa^s(x) = 0$ o $\kappa^u(x) = 0$, ellas dejan de representar frentes de onda estables o inestables. En ese caso los exponentes de Liapunov en x o bien se anulan o bien no existen. En particular cuando ∂Q consiste sólo de componentes neutras, o sea cuando Q es un polígono, $\kappa^s(x) = \kappa^u(x) = 0$ en cada $x \in M'$, y ambos exponentes de Liapunov se anulan casi todo punto.

Ecuaciones de los frentes de onda y su transversalidad. Deducimos ahora las ecuaciones de los subespacios tangentes correspondientes a los exponentes de Liapunov positivo y negativo. Son subespacios de $\mathcal{T}M$ que llamamos *inestable* (E^u) y *estable* (E^s), respectivamente.

Sean $x = (q, v) \in M$ un punto de coordenadas (r, ϕ) y $u = (dr, d\phi) \in \mathcal{T}_x M$ un vector tangente en x . dr y $d\phi$ son cantidades infinitesimales. Como ya fue explicado, todo vector tangente $u \neq 0$ puede ser representado por una curva en M , por ejemplo $(r + s dr, \phi + s d\phi)$, donde $0 < s < \varepsilon$ es un parámetro pequeño. Esta curva define un haz de trayectorias salientes, cuya sección es un frente de onda Σ que pasa por q . La curvatura de este frente en q , llamada χ , se calcula así

$$\chi = \frac{1}{\cos\phi} \left(\frac{d\phi}{dr} + K(r) \right) \quad (\text{V.2.1})$$

Esto surge de un análisis infinitesimal como el que usamos en el cálculo de la derivada DT de la transformación de billar (Lema IV.1.1); lo dejamos como ejercicio (V.4). Observe que el frente Σ no está unívocamente determinado por el vector u , pero su curvatura χ está completamente determinada por u . Además, χ sólo depende de la dirección de u , y no de su longitud.

De (V.2.1) se deduce que

$$\frac{d\phi}{dr} = -K(r) + \chi \cos\phi, \quad (\text{V.2.2})$$

por lo que los subespacios inestable E^u (definido por el vector $(dr^u, d\phi^u)$) y estable E^s satisfacen esa relación sustituyendo χ por κ^u, κ^s definidos en (V.1.6), (V.1.4).

Supongamos que las curvaturas de las componentes dispersoras y focalizadoras de ∂Q están acotadas por encima y por debajo:

$$K_{\text{máx}} = \max_r |K(r)| < \infty, \quad K_{\text{mín}} = \min_r |K(r)| > 0.$$

donde los mínimos se toman sobre las componentes no neutras de la frontera.

En los billares de Sinai $\kappa^s < 0$, $\kappa^u \geq 2K(r)/\cos\phi$ y $K(r) > 0$, por lo que

$$\frac{d\phi^s}{dr^s} < -K(r) \leq -K_{\text{mín}} < 0, \quad \frac{d\phi^u}{dr^u} > K(r) \geq K_{\text{mín}} > 0 \quad (\text{V.2.3})$$

Entonces, las direcciones inestables son siempre positivas, o crecientes (en las coordenadas r, ϕ), y las direcciones estables son negativas, o decrecientes. Más aún, ambas están separadas de la dirección horizontal $d\phi = 0$.

Podrían, sin embargo, no estar separadas de la dirección vertical porque las derivadas en (V.2.3) pueden ser arbitrariamente grandes. Pero en el caso típico ellas están acotadas, dado que fácilmente puede ver $|\kappa^s(x)| < 1/\tau(x)$ y $|\kappa^u(x) - a_0(x)| < 1/\tau(x)$. Por tanto,

$$\left| \frac{d\phi^a}{dr^a} \right| \leq \max_r |K(r)| + \frac{1}{\min_x \tau(x)}, \quad a = u, s.$$

Nuestro análisis muestra que las acotaciones superior e inferior de $\tau(x)$ están involucradas en el estudio de las direcciones estables e inestables de los billares de Sinai. Tenemos los siguientes resultados simples:

- (i) si una mesa de billar de Sinai tiene *vértices* (intersecciones entre las componentes suaves de ∂Q con ángulo $< \pi$), entonces $\min_x \tau(x) = 0$; en caso contrario $\tau_{\text{mín}} = \min_x \tau(x) > 0$;
- (ii) para cualquier mesa de billar de Sinai $Q \subset \mathbb{R}^2$ se tiene $\tau_{\text{máx}} = \max_x \tau(x) < \infty$;
- (iii) para billares de Sinai en el toro, $Q \subset \mathbb{T}^2$, el valor $\tau_{\text{máx}}$ puede ser o bien finito o bien infinito. En el primer caso decimos que Q tiene *horizonte finito*.

Veremos ahora que las variedades estable e inestable pueden aproximarse a la dirección vertical $dr = 0$ cuando $\tau(x) \approx 0$, lo cual sólo ocurre, en los billares de Sinai, cerca de los vértices. Si no hay vértices, $\tau_{\text{mín}} > 0$ y las derivadas de (V.2.3) están acotadas por encima. En este caso los subespacios estables e inestables, E^s y E^u son uniformemente transversales – el ángulo entre ellos está separado de cero. A este respecto, los billares de Sinai son similares a los difeomorfismos de Anosov.

Decimos que una mesa de billar de Sinai con vértices es una mesa *propia* si todos sus vértices tienen ángulo positivo, o sea si los lados de ∂Q se intersectan transversalmente. Originalmente, Sinai y su escuela sólo estudiaron mesas de billar propias. Las mesas impropias, donde algunos vértices son *cúspides* (los lados forman ángulo cero) tienen propiedades algo diferentes. En particular la estimación del número de reflexiones estudiada con la Figura IV.5 de la Sección IV.6, no es válida. Ver Ejercicio IV.22.

En las mesas de billar propias con vértices, las direcciones estables e inestables, E^s y E^u , no se pueden aproximar simultáneamente a la dirección vertical $dr = 0$. Hay un argumento geométrico explicado en [BSC91] que previene esta anomalía ². Por tanto, el ángulo entre E^s y E^u permanece separado de cero para todos los billares propios de Sinai! La transversalidad de E^u y E^s juega un importante papel en el estudio de las propiedades ergódicas de los billares caóticos.

Tasas de expansión y contracción de los frentes de onda. Como fue explicado anteriormente, en los billares de Sinai los frentes divergentes crecen constantemente al avanzar. Por otra lado la derivada DT (ver Lema IV.1.1) no expande todos los vectores $u = (dr^u, d\phi^u)$, como veremos de aquí a poco. Este hecho sugiere que la norma euclidiana estándar $|u| = [(dr)^2 + (d\phi)^2]^{1/2}$ de vectores tangentes no sea tan conveniente para el estudio de frentes divergentes de billares de Sinai ³. En realidad, parece más interesante medir las longitudes de los vectores tangentes $u = (dr, d\phi)$ por el ancho de la sección ortogonal Σ asociada al frente de onda. Este ancho es denominado *norma p* y representado por $|u|_p$. Fácilmente se ve que $|u|_p = \cos\phi |dr|$.

²Ya hemos visto que el número de choque cerca de un vértice con ángulo interno distinto de cero, tiene una cota superior uniforme; ver Sección IV.5

³La elección de la norma adecuada es un problema complicado en los distintos tipos de billares. Se puede ver un estudio de este asunto en la Sección 2.1 de [BCST02].

Estrictamente hablando, esta no es una norma puesto que $|u| = 0$ para todo vector vertical $(0, d\phi)$. Es, en realidad, una *pseudonorma*. Pero ya hemos visto que los vectores estables e inestables nunca son verticales, por lo que siempre tendrán p-norma positiva. Esto agrega justificaciones para usar la p-norma en lugar de la Euclidiana.

La p-norma de un vector tangente $u \in \mathcal{T}_x M$ cambia DT por la simple fórmula

$$\frac{|DT(u)|_p}{|u|_p} = \left| 1 + \frac{\tau(x)}{\cos \phi} \left(\frac{d\phi}{dr} + K(r) \right) \right| = |1 + \tau(x)\chi| \quad (\text{V.2.4})$$

donde χ es la curvatura del correspondiente frente de onda en x .

En los billares de Sinai, los vectores inestables u satisfacen (V.2.3), por lo que la cantidad (V.2.4) es siempre más grande que uno. En otras palabras, los vectores inestables siempre son expandidos por T en la p-norma, tal como esperábamos. Más aún, si no hay vértices en la mesa, entonces $\tau(x) \geq \tau_{\min} > 0$, y la cantidad (V.2.4) tiene una cota inferior $\Lambda > 1$, i.e.

$$|DT(u)|_p \geq \Lambda |u|_p, \quad \Lambda > 1 \quad (\text{V.2.5})$$

para todos los vectores inestables. Esto demuestra la hiperbolicidad uniforme, en el sentido del Capítulo III, otra característica común de los billares de Sinai y los difeomorfismos de Anosov.

Las mesas de billar de Sinai propias con vértices son también uniformemente hiperbólicas, luego de unas modificaciones menores. En realidad, para estos billares el número de reflexiones cerca de un vértice es uniformemente acotado, ver Sección IV.5. Por tanto hay una constante $m \geq 1$ tal que ninguna trayectoria puede tener más de m reflexiones cerca de cada vértice. En este caso, la transformación T^m es uniformemente hiperbólica, o sea (V.2.5) se cumple con T reemplazado por T^m .

Por otro lado, el factor de expansión en la p-métrica es no acotado. Si x está cerca de S_0 , donde $\cos \phi \approx 0$, entonces la cantidad (V.2.4) puede ser arbitrariamente grande. Esto constituye una importante diferencia entre los billares de Sinai y las transformaciones de Anosov. Es también fácil de ver que en los billares de Sinai con horizonte finito, $\cos \phi \approx 0$ es la única razón para que la cantidad (V.2.4) pueda acercarse a infinito. En otras palabras, tenemos

$$c_1 / \cos \phi \leq |DT(u)|_p / |u|_p \leq c_2 / \cos \phi \quad (\text{V.2.6})$$

con algunas constantes $c_1, c_2 > 0$.

En el estudio de las propiedades estadísticas de los billares de Sinai (su decaimiento de correlación y los teoremas límite, ver Secciones II.4 y V.5), se necesita un control muy preciso del factor de expansión $|DT(u)|_p / |u|_p$. Las estimaciones de arriba muestran que tal control es difícil en la vecindad del conjunto $\partial M = \{\cos \phi = 0\} = \{|\phi| = \pi/2\}$. Para mejorar ese control uno puede partir la superficie M en una cantidad numerable de fajas por las líneas $|\phi| = \pi/2 - a_k$, con algún $a_k \rightarrow 0$ cuando $k \rightarrow \infty$. Por ejemplo, es costumbre poner $a_k = 1/k^2$. Entonces el área crítica cerca de las líneas $|\phi| = \pi/2$ es dividida en fajas estrechas donde el factor de expansión $|DT(u)|_p / |u|_p$ puede ser controlado más fácilmente. Estas regiones son llamadas *fajas de homogeneidad* y fueron introducidas por primera vez en [BSC91].

Consideremos ahora la evolución de los vectores tangentes en la norma estándar Euclideana $|u|$. La relación entre las dos normas es

$$J(u) = \frac{|u|_p}{|u|} = \cos \phi \left[1 + \left(\frac{d\phi}{dr} \right)^2 \right]^{-1/2}. \quad (\text{V.2.7})$$

$$\text{Entonces} \quad \frac{|DT(u)|}{|u|} = \frac{|DT(u)|_p J(u)}{|u|_p J(DT(u))} \quad (\text{V.2.8})$$

No es difícil ver que los vectores inestables u pueden no ser expandidos por DT en la norma Euclidiana, si $J(u)$ es pequeño. Cuando $\cos \phi \approx 0$ y $\tau(x)$ son pequeños, podemos tener $|DT(u)|/|u| \approx 0$, por lo que el vector inestable u puede ser reducido por un factor arbitrariamente grande! Este es un efecto totalmente indeseable, lo cual justifica una vez más nuestro interés por la p-norma.

Conos invariantes. Hemos descrito los subespacios estable e inestable, E^s y E^u , en cada punto $x \in M'$ de billares de Sinai. Esto fue hecho siguiendo la líneas de los estudios originales de Sinai sobre billares

dispersores. Recientemente, sin embargo, las técnicas de conos invariantes se hicieron populares en los estudios de dinámica hiperbólica. Ahora construiremos conos invariantes para los billares de Sinai.

Sea un punto $x = (q, v) \in M$ de coordenadas (r, ϕ) . Cada vector tangente $u = (dr, d\phi)$ en x puede ser representado por un frente de onda Σ pasando por x , con curvatura χ .

Definición 22. *Un cono estable en x consiste de los vectores con $\chi < 0$: los frentes de onda Σ convergentes. El cono inestable es el conjunto de vectores u cuyos frentes de onda Σ fueron divergentes antes de la reflexión en x .*

Por tanto, el cono inestable consiste de vectores u tales que $\chi_- > 0$, por lo que $\chi - 2K(r)/\cos\phi > 0$, de acuerdo a la Ecuación del Espejo (V.1.2).

Por la ecuación (V.2.1), los conos estable e inestable son descritos por $\mathcal{C}^s(x) = \{u : -\infty < d\phi/dr \leq$
Por tanto, en las coordenadas r, ϕ , el cono inestable está en el primer y tercer cuadrantes (direcciones crecientes) y el cono estable está en el segundo y cuarto cuadrantes (direcciones decrecientes). Por esta razón, a veces en los estudios de billares (por ejemplo [BSC91]), las direcciones inestables son llamadas crecientes y las estables, decrecientes.

Se deduce de inmediato de las observaciones previas que la familia de conos $\mathcal{C}^u(x)$ y $\mathcal{C}^s(x)$ son invariantes:

$$DT(\mathcal{C}^u(x)) \subset \mathcal{C}^u(Tx) \quad DT^{-1}(\mathcal{C}^s(x)) \subset \mathcal{C}^s(T^{-1}x).$$

En realidad, son estrictamente invariantes en los billares de Sinai.

En lugar de $\mathcal{C}^u(x)$ y $\mathcal{C}^s(x)$ se puede fijar un $m \geq 1$ y considerar la familia de conos $DT^m(\mathcal{C}^u(x))$ y $(DT^{-m}(\mathcal{C}^s(x)))$. Estas familias también serán invariantes, y para m grande los conos pueden ser arbitrariamente delgados. Y, naturalmente, el ángulo entre esos conos será positivo para cualquier $m \geq 1$.

Variedades estable e inestable. En billares dispersores, la transformación $T : M \rightarrow M$ es uniformemente hiperbólica en el sentido del Capítulo III, pero no suave. Esta última característica establece una gran diferencia entre los billares y los difeomorfismos de Anosov. T es discontinua en puntos $x \in M$ tales que o bien $x \in S_0$ o $Tx \in S_0$; ver Lema IV.1.1. Es costumbre llamar a este conjunto $S_{-1} = S_0 \cup T^{-1}S_0$.

Recordar que $M = \partial Q \times [-\pi/2, \pi/2]$ es una variedad de dimensión dos que consiste de piezas suaves $M_k := \Gamma_k \times [-\pi/2, \pi/2]$ para $1 \leq k \leq s$; ver (IV.1.1). Es fácil ver que $\partial M = \cup_{k=1}^s \partial M_k = S_0$, por lo que S_0 es la frontera natural de la variedad M .

El conjunto $T^{-1}S_0$ tiene una estructura más complicada. Sin entrar en detalles, afirmamos que $T^{-1}S_0$ es una unión de curvas suaves en M ; ver Figura IV.7. Para billares en el plano el número de curvas es finito, pero para billares en un toro, como el gas de Lorentz (ver IV.5), el número de curvas puede ser infinito.

OBSERVACIÓN V.2.2. . Todas las curvas en $T^{-1}S_0$ son decrecientes pues están dadas por ecuaciones $\phi = f(r)$ con alguna $f'(r) < 0$. Corresponden a frentes convergentes. Para ver esto, dibuje una mesa de billar dispersora Q cuya frontera contiene un vértice y una familia de rayos comenzando en ∂Q y convergiendo a un vértice.

De modo similar, se observa que T^n , $n > 0$, tiene singularidades en el conjunto $S_{-n} := S_0 \cup T^{-1}S_0 \cup \dots \cup T^{-n}S_0$ que es también la unión de curvas decrecientes en M (recordar que la clausura de este conjunto fue llamado S_n^+ en la Sección III.5). La transformación T^{-n} para $n > 0$ tiene singularidades en el conjunto S_n definido de manera similar y consistente de curvas crecientes.

En la Sección III.3 introdujimos la clase general de las transformaciones suaves con singularidades. La transformación de billar $T : M \rightarrow M$ siempre pertenece a esa clase, en particular satisface los requerimientos técnicos de Katok y Strelcyn (KS1), (KS2) and (KS3). En realidad, esta clase fue introducida por estos autores en [KS86] con la finalidad principal de incluir a las transformaciones de billar. La verificación de los requerimientos técnicos no es tarea sencilla; está hecha para clases muy generales de billares en la Parte V de ese libro fundamental.

Entonces, del Teorema III.3.2 se deduce que las variedades estables e inestables de T existen en ν -casi todo punto $x \in \Sigma(T)$. Para los billares dispersores, $\nu(\Sigma(T)) = 1$, por lo que las variedades estable e inestable existen casi todo punto:

Teorema V.2.3. *En billares dispersores, para ν -casi todo punto $x \in M$ existen las variedades estable $W^s(x)$ e inestable $W^u(x)$ por x .*

Estas variedades son curvas de dimensión uno en M . La curva $W^s(x)$ corresponde al frente de onda convergente que permanece convergente todo el tiempo. La curva $W^u(x)$ corresponde a un frente de onda divergente que permanece divergente para todos los tiempos.

Cada variedad estable $W^s(x)$ e inestable $W^u(x)$ es una curva compacta suave de longitud finita. Su longitud está determinada por cuan cerca están las imágenes de x del conjunto de singularidad S_0 , ver la discusión después del Teorema III.3.2. Es también cierto que las imágenes de los puntos extremos de $W^s(x)$ y $W^u(x)$ pasan por el conjunto de singularidad S_0 .

Entropía. Otros teoremas enunciados en la Sección III.3 se aplican a los billares. En particular, la fórmula de Pesin (Teorema III.3.4) expresa la entropía de la transformación de billar T :

$$h_\nu(T) = \int_M \lambda_+(x) d\nu(x) \quad (\text{V.2.9})$$

donde $\lambda_+(x)$ es el (único) exponente de Liapounov positivo de T en x .

Esta fórmula tiene un valor teórico fundamental, pero en la práctica no siempre es conveniente, puesto que el exponente de Liapounov está definido por (III.1.3) en el Capítulo III como el límite de una expresión que involucra a todos los iterados de T .

Hay una extraordinaria simplificación de la ecuación (V.2.9). Para cada punto $x \in M$ sea $D^u(x) = \|(DT)_x(v)\|/\|v\|$, $v \in E_x^u$ el factor de expansión de los vectores inestables $v \in E_x^u$. Entonces por la regla de la cadena, la definición de (III.1.3) puede ser reescrita como

$$\lambda_+(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \log D^u(T^i x) \quad (\text{V.2.10})$$

Combinando (V.2.9), (V.2.10) y el Teorema Ergódico de Birkhoff-Khincin, da

$$h_\nu(T) = \int_M \log D^u(x) d\nu(x) \quad (\text{V.2.11})$$

La integral no depende de la métrica usada (ver Ejercicio V.6) por lo que podemos definir $D^u(x)$ en la p -norma por (V.2.4). Esto lleva al importante resultado

Teorema V.2.4. *En los billares dispersores, la entropía está dada por*

$$h_\nu(T) = \int_M \log |1 + \tau(x)\kappa^u(x)| d\nu(x) \quad (\text{V.2.12})$$

Este Teorema puede ser generalizado a cualquier sistema de billar en cualquier dimensión. Ver [C97].

En los sistemas dinámicos, uno puede también definir la entropía de un flujo Φ^t . Se define como la entropía de la transformación Φ^1 (que se obtiene tomando $t = 1$; la transformación Φ^1 es llamada *transformación de tiempo uno* de Φ^t), entonces $h_\mu(\{\Phi^t\}) := h_\mu(\Phi^1)$. Esta definición está basada en el hecho estándar de que para cualquier real $t \in \mathbb{R}$ tenemos $h_\mu(\Phi^t) = |t|h_\mu(\Phi^1)$, por lo que $h(\Phi^t)$ es suficiente para determinar $h(\Phi^t)$ para todo t .

Hay una fórmula relativamente simple de Abramov [Ab59] que relaciona la entropía del flujo Φ^t con la de T :

$$h_\mu(\{\Phi^t\}) = h_\nu(T)/\bar{\tau}$$

donde $\bar{\tau}$ es el tiempo libre medio, ver (IV.6.2) y (IV.6.3). Por tanto, obtenemos

Teorema V.2.5. *En billares dispersores,*

$$h_\mu(\{\Phi^t\}) = h_\nu(T)c_\mu c_\nu^{-1} = h_\nu(T) |\partial Q| (\pi|Q|)^{-1} \quad (\text{V.2.13})$$

Las fórmulas para la entropía fueron esencialmente obtenidas por Sinai antes de 1970 [Si70]. Fueron generalizadas por él mismo, unos años después (1978). Por algunas exposiciones más recientes y extensiones a otras clases de billares, ver [CM92] y [C97].

Se pueden emplear las ecuaciones (V.2.12) y (V.2.13) para calcular estimaciones de la entropía de muchos modelos físicos interesantes, como los gases de Lorentz y los sistemas de bolas duras (Sección IV.5). Algunos

resultados fueron obtenidos en [C97, CM92]. En particular, para el gas de Lorentz d -dimensional ($d \geq 2$) con un único obstáculo esférico de pequeño radio r en el toro unitario, las entropías están dadas por

$$h_\nu(T) = -d(d-1) \log r + O(1)$$

$$h_\mu(\{\Phi^t\}) = -d(d-1) |B^{d-1}| r^{d-1} \log r + O(r^{d-1}) \text{ cuando } r \rightarrow 0.$$

Observamos que la entropía de T crece a infinito, mientras que la de Φ^t se anula. Lo mismo sucede con los exponentes de Liapunov de T y Φ^t , respectivamente. Impulsamos al lector a encontrar un explicación cualitativa de porqué estos exponentes se comportan de manera tan diferente cuando $r \rightarrow 0$.

EJERCICIO V.4. Probar la ecuación (V.2.1). Sugerencia: la curva Σ puede ser aproximada por un arco de círculo. Entonces es parametrizada por la longitud de arco s . El ángulo entre los vectores normales $N(s)$ y $N(s+ds)$ mide $\chi(s) ds + o(ds)$ donde $\chi(s)$ es la curvatura en el punto s .

EJERCICIO V.5. En los billares semidispersores, algunos elementos de la fracción continua (V.1.4) son cero. Mostrar que se puede reescribir (V.1.4) saltando los choques en componentes neutras y agregando en su lugar la suma de las distancias de segmentos de trayectorias alguno de cuyos extremos está en componentes neutras. Sugerencia: Si $T^k x$ está en una componente neutra, entonces $b_{2k} = 0$ y se puede saltar la k -ésima reflexión agregando el paso libre $\tau(T^{k-1}x) + \tau(T^k x)$.

EJERCICIO V.6. Complete los detalles de la demostración del Teorema V.2.4. En particular, verifique que la integral en (V.2.11) no depende de la métrica. Sugerencia: use la fórmula (V.2.8) y la invariancia de la medida ν .

V.3. Billares de Bunimovich

Esta sección y la siguiente están dedicadas al estudio de billares cóticos planos cuyas fronteras contienen curvas focalizadoras. Usaremos la siguiente notación (ver Sección IV.1):

$$\mathcal{M}_+ = \bigcup_{\Gamma_i \in \Gamma_+} \mathcal{M}_i, \quad \mathcal{M}_- = \bigcup_{\Gamma_i \in \Gamma_-} \mathcal{M}_i, \quad \mathcal{M}_0 = \bigcup_{\Gamma_i \in \Gamma_0} \mathcal{M}_i. \quad (\text{V.3.1})$$

Desde que Sinai probó antes de 1970 el carácter caótico de los billares dispersores, ellos se mantuvieron, por muchos años, como la única clase de billares caóticos. Más aún, en 1973 Lazutkin probó que los billares en una mesa genérica convexa tenía cústicas, ver Sección. IV.3. Parecía, por tanto, que la caoticidad en billares podía ser producida por fronteras cóncavas (dispersoras). O, por lo menos, que si existían billares con fronteras focalizadoras, ellos deberían tener fronteras muy especiales, porque en los dominios convexos simples – círculos y elipses – los billares son completamente integrable (lo cual es casi lo opuesto al caos).

Resultó, por tanto una gran sorpresa cuando a mediados de los setenta un alumno de Sinai, L. Bunimovich, construyó billares caóticos cuyas fronteras contenían curvas focalizadores, y estas consistían de simples arcos circulares!

Hay una significativa diferencia entre las teorías de billares dispersores y focalizadores. Para los últimos, el estudio de la hiperbolicidad está muy desarrollado -una amplia clase de ellos se sabe que son hiperbólicos. Pero propiedades más avanzadas son prácticamente desconocidas. Sólo para la clase muy limitada de los billares de Bunimovich se han probado propiedades más técnicas sobre regularidad de las variedades invariantes y las medidas inducidas en ellas, otras propiedades estadísticas. Las pruebas de hiperbolicidad, ergodicidad y otras propiedades de la jerarquía ergódica (ver Sección II.3) se han realizado por diferentes métodos. En general, ellos son menos “billarísticos” y por ello mismo más generales y aplicables a otras clases de sistemas dinámicos caóticos.

Mesas de billar de Bunimovich. La construcción de Bunimovich está basada en el fenómeno de *desfocalización*, que describiremos más abajo. Recuérdese que en los billares de Sinai, los frentes de onda neutros (paralelos) se hacen divergentes después de chocar con la frontera, y durante el tiempo libre entre colisiones, ellos continúan agrandándose. el factor de crecimiento es

$$\Lambda = |1 + \tau\chi| \quad (\text{V.3.2})$$

donde τ es el tiempo entre colisiones y $\chi > 0$ es la curvatura del frente al comienzo de cada corrida libre; ver (V.2.4).

Considere ahora una mesa de billar con frontera focalizadora. Un frente de onda paralelo que choca con la frontera, se hará convergente, o sea $\chi < 0$ (esto sale de inmediato de la ecuación del espejo V.1.2). Pero

puede focalizarse (converger a un punto dentro de Q) y continuar como un frente de onda divergente! Esta transformación de un frente convergente en un divergente es lo que se llama *desfocalización*. Entonces nuestro frente seguirá viajando creciendo de tamaño. Cuando llegue a la frontera ∂Q y se refleje nuevamente, su tamaño podrá ser aún más grande que el tamaño del frente original, ver Figura V.2. El factor de crecimiento es aún dado por (V.3.2). Es verdad que $\chi < 0$, por lo que Λ no tiene porque ser más grande que uno. No obstante, tenemos $\Lambda > 1$ si establecemos la condición adicional $\tau\chi < -2$, o sea cuando $1/|\chi| < \tau/2$ (*). Es claro que un frente de onda convergente cuya curvatura es $\chi < 0$ se focalizará (convergerá a un punto) en el tiempo $1/|\chi|$. Entonces, la condición (*) significa que el frente de onda se focaliza antes de llegar al punto medio entre colisiones. Y por tanto $\Lambda > 1$, por lo que nuestro frente de onda crecerá antes de que ocurra la siguiente colisión. Esta es la condición clave par la hiperbolicidad de T .

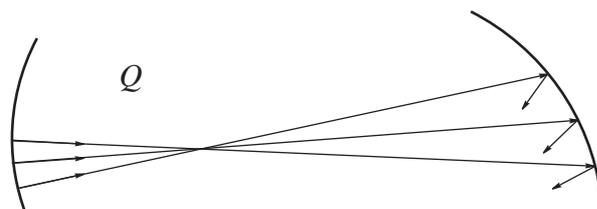


Figura V.2: Desfocalización de un frente convergente.

¿Podemos asegurar la condición (*) al construir mesas de billar con fronteras focalizadoras? Si $\Gamma_i \subset \partial Q$ es una componente focalizadora de la frontera, no podemos colocar otras componentes de ∂Q demasiado cerca de Γ_i . Los frentes de onda que dejan Γ_i necesitan moverse libremente y desfocalizarse antes de que choquen a ∂Q nuevamente, y necesitan suficiente espacio para volver a crecer. Esto explica el requerimiento $K_i \subset Q$ en la siguiente definición de Bunimovich.

Definición 23. Sea Q una mesa de billar cuya frontera consiste de componentes dispersoras, neutras y focalizadoras. Cada componente focalizadora Γ_i es un arco de círculo, K_i , de radio R_i . Si que $K_i \subset Q$, decimos que la mesa es un billar de Bunimovich.

Destacamos que la última hipótesis implica que el ángulo interior de dos componentes adyacentes de frontera, una de ellas focalizadora, debe ser mayor o igual a π .

Un vector tangente $u \in T_x M$ en el punto $x \in M$ representado por el frente de onda Σ es llamado *inestable* si

- (i) Σ es divergente si x pertenece a una componente dispersora o neutra de ∂Q ;
- (ii) Σ es convergente, y su curvatura satisface $\chi < -(R_i \cos \phi)^{-1}$, (**) si x pertenece a una componente focalizadora $\Gamma_i \subset \partial Q$.

Como normalmente, llamaremos C_x^u al conjunto de los vectores inestables. Expresiones de los vectores inestables, en las coordenadas $(dr, d\phi)$ se dan en el Ejercicio V.7. Uno puede fácilmente ver que (**) is equivalente a (*). El siguiente es el teorema clave de Bunimovich.

Teorema V.3.1 ([Bu74], [Bu79]). Sea Q la mesa de billar de la definición de arriba. Si u es un vector inestable, también lo es $DT(u)$.

Demostración. Ver Ejercicio V.8. □

Por lo tanto, los vectores tangentes inestables definen un campo de conos $C^u(x)$ invariant por DT . Esto indica que los métodos de la Sección III.4 se aplican y parece deducirse la hiperbolicidad. Sin embargo, el campo de conos invariante no implica aún la hiperbolicidad. En realidad, el teorema anterior se satisface aún en billares circulares, los cuales, ya lo sabemos, no son hiperbólicos. Por tanto, podemos tener un campo de conos invariante y no tener hiperbolicidad. Los métodos de la Sección III.4 requieren que el campo de conos sea estrictamente invariante, por lo menos eventualmente. En los billares de Bunimovich arriba definidos, la invariancia estricta de los conos, esto es, la propiedad $DT(\bar{C}^u(x)) \subset \text{int } C^u(Tx)$ se satisface si:

- (a) el punto x pertenece a una componente dispersora de la frontera, o

(b) el punto x pertenece a una componente focalizadora Γ_i , pero su imagen Tx pertenece a cualquier otra componente. Que estas condiciones son suficientes para la invariancia del campo de conos se prueba en el Ejercicio V.9.

Mostraremos ahora que casi toda trayectoria de una mesa de billar de Bunimovich tiene puntos de los tipos (a) o (b). Si así sucede, tendremos hiperbolicidad. Un simple análisis geométrico muestra que podrían haber dos tipos de trayectorias que no tienen tales puntos:

A. Las trayectorias que sólo se reflejan en la misma componente focalizadora Γ_i .

B. Las trayectorias que sólo se reflejan en las componentes neutras de ∂Q .

Es fácil ver que las del tipo A, son trayectorias periódicas y su medida total es nula (recuerde que las Γ_i son arcos de círculos, no círculos enteros). Ver Ejercicio V.10. Respecto de las de tipo B, en cada ejemplo particular, es fácil probar que tales trayectorias forman un conjunto de medida nula. Por ejemplo, billares que tienen sólo una (o ninguna) componente neutra, o dos componentes neutras paralelas, etc. Sin embargo, en general es un problema abierto (y desafiante) probar que en general las trayectorias de tipo B forman un conjunto de medida nula. Estos son problemas de billares poligonales que no tratamos en estas notas (por un panorama general, ver [Gu86, Gu96], resultados parciales en [CT98, DMM05]). El siguiente es el teorema probado por Bunimovich:

Teorema V.3.2 ([Bu74], [Bu79]). *Sea Q una mesa de billar que satisface la Definición 23 en que las trayectorias de tipo B forman un conjunto de medida ν -nula. Entonces la transformación de billar T es hiperbólica.*

En la Sección siguiente mostraremos que las fracciones continuas (V.1.4) y (V.1.6) son convergentes para estos billares de Bunimovich y que, por tanto las direcciones estables e inestables están dadas por (V.2.2), sustituyendo χ por $\kappa^s(x)$, $\kappa^u(x)$, respectivamente.

El “stadium” mencionado al final de la Sección IV.3 es, probablemente, el más famoso de los billares de Bunimovich. se pueden construir otros ejemplos bonitos, como la “mesa flor” (unión de varios círculos de los cuales uno, el mayor, está en el centro, y los otros están colocados alrededor del central, y se sobrepone con él.

Si se corta un círculo y se corta a lo largo de una cuerda, sacando la parte menor, da otro ejemplo de billar hiperbólico, aunque éste no satisface los requerimientos de la Definición 23. Pero esto se puede solucionar, a efectos de la prueba tomando la unión de la mesa original y de su imagen especular por la cuerda. Esta unión es un billar de Bunimovich cuya hiperbolicidad (y ergodicidad, etc.) implica la de la mesa original, más chica.

EJERCICIO V.7. Para el campo de conos definido en esta Sección, probar que

a) $x \in M_0$, $C_x^u = \{(dr, d\phi) \in T_x M : d\phi/dr \geq 0\}$;

b) $x \in M_-$, $C_x^u = \{(dr, d\phi) \in T_x M : d\phi/dr \leq 0\}$. Sugerencia: Use (V.2.2).

EJERCICIO V.8. Probar el Teorema V.3.1. Sugerencia: Pruebe primero que la longitud entre dos choques sucesivos en la misma circunferencia C_i es $2R_i \cos \phi$. Considere los diferentes casos dependiendo de los tipos de componentes de dos choques sucesivos. Use la ecuación (V.1.3).

EJERCICIO V.9. Verificar la invariancia estricta de los campos de cono en los dos casos observados luego del enunciado del Teorema V.3.1. Sugerencia: Por simple inspección más detallada en al prueba del Ejercicio V.8.

EJERCICIO V.10. Dado un arco de círculo de ángulo al centro $\Theta > \pi$ calcular el número de rebotes máximos necesarios para salir de él, de una trayectoria que entra con ángulo ϕ . Observar que el asunto es particularmente interesante cuando ϕ es cercano a 0 (trayectorias diametrales) y a $\pi/2$ (trayectorias rasantes, en este caso alcanza que $\Theta > 0$).

V.4. Otros Billares Hiperbólicos.

Desde el descubrimiento del mecanismo de desfocalización por Bunimovich muchos matemáticos intentaron construir billares caóticos con otras clases de curvas focalizadoras. Esto no era sencillo puesto que el círculo es una figura muy “rígida” cuyas propiedades jugaban un papel fundamental en los cálculos de Bunimovich.

Recién a mediados de los ochenta M. Wojtkowski [W86] y R. Markarian [Ma88] construyeron nuevas clases de billares hiperbólicos con fronteras focalizadoras distintas de arcos de círculos. Más tarde V. Donnay

[Do91] y L. Bunimovich [Bu92] generalizaron estas construcciones y presentaron una visión unificada de lo que el último llamó arcos absolutamente focalizadores.

Nuestra presentación de estos temas, combina la maquinaria analítica de formas cuadráticas (Teorema III.4.1) con argumentos geométicos referidos a la evolución de frentes de onda (Sección V.2).

Sea $\Gamma \subset \mathbb{R}^2$ una curva C^r , focalizadora (al suponerla parte de la frontera de una mesa de billar Q), con curvatura estrictamente negativa y puntos extremos A_1 y A_2 . Suponga que su curvatura tiene signo constante y nunca se anula. Sea $\alpha(\Gamma)$ la variación del ángulo de la tangente a Γ al moverse de A_1 a A_2 a largo de Γ .

Definición 24. Una curva Γ es un arco absolutamente focalizador si

(AF1) $\alpha(\Gamma) \leq \pi$, y

(AF2) cualquier frente de onda infinitesimal paralelo que choca en Γ desde Q se focaliza entre dos choques sucesivos, y luego del último choque, en Γ .

La importancia del requerimiento (AF2) fue demostrada arriba –el frente de onda crece después de pasar el punto de focalización, que es una de las claves para la eventual expansión.

Formas cuadráticas. Desarrollamos ahora un método general que permite estudiar la hiperbolicidad de prácticamente cualquier billar. Muchos estudios recientes de billares comienzan con teoremas sobre hiperbolicidad (no anulación de exponentes de Liapunov). Todos esos teoremas están probados, sea construyendo conos invariantes o usando formas cuadráticas crecientes. Como se vió en el Capítulo III ambos métodos son esencialmente equivalentes. Usaremos aquí el segundo método en un contexto bastante general de billares planos.

Sean $x = (q, v) \in M$ un punto con coordenadas (r, ϕ) y $(dr, d\phi)$ un vector tangente. Introducimos un nuevo sistema de coordenadas en el espacio tangente $\mathcal{T}_x M$, que está directamente relacionado con la α y h usadas en la Figura IV.2 de la Sección IV.1. Las nuevas coordenadas (U, V) se define con la ayuda de la proyección ortogonal de $(dr, d\phi)$ sobre el subespacio perpendicular a v :

$$U = \cos \phi dr \quad V = Kdr + d\phi.$$

U es la restricción en M de un campo de Jacobi transversal a lo largo de la trayectoria de billar, y V es su derivada. Esta es la razón por la que normalmente U is denotad por J y V por J' . La siguiente igualdad relaciona esas nuevas coordenadas con los tiempos de focalización: si $u = (U(u), V(u))$ entonces $f^+(u) = -U/V$.

OBSERVACIÓN V.4.1. Considere la métrica Riemanniana ρ dada por $U^2 + V^2$ sobre el fibrado tangente de M . Puede ser probado que la forma simpléctica $\omega = \cos \phi dr \wedge d\phi$ se transforma en $U \wedge V$ en las coordenadas (U, V) , por lo que la forma de volumen inducida por la métrica ρ coincide con la forma simpléctica ω ; en este sentido, estas coordenadas son “naturales” para los billares.

Ahora, si $D_x T(U, V) = (U_1, V_1) \in \mathcal{T}_{T_x M}$ es la imagen de nuestro vector (U, V) , con $U_1 = \cos \phi_1 dr_1$ y $V_1 = K_1 dr_1 + d\phi_1$, entonces un cálculo simple da

$$U = -(\tau K_1 + \cos \phi_1) dr_1 + \tau d\phi_1, \quad V = K_1 dr_1 - d\phi_1$$

donde todos los símbolos tienen su sentido usual.

La evolución de (U, V) a lo largo de un segmento de trayectoria de longitud τ entre dos colisiones consecutivas, y en la reflexión, están dadas, respectivamente, por

$$\begin{pmatrix} 1 & \tau \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} -1 & 0 \\ -\frac{2}{d} & -1 \end{pmatrix}. \quad (\text{V.4.1})$$

donde $d = \cos(\phi)/K(r)$. Entonces, la derivada de T en $x = (r, \phi)$ con respecto a las coordenadas (U, V) es dada por

$$D_x T = - \begin{pmatrix} 1 & \tau \\ \frac{2}{d} & \frac{\tau}{d} + 1 \end{pmatrix}. \quad (\text{V.4.2})$$

Definimos ahora una forma cuadrática medible no degenerada

$$B_x(U, V) = aU^2 + 2bUV + cV^2 \quad (\text{V.4.3})$$

donde a, b, c son funciones medibles de x (en realidad serán diferenciables de a trozos) y $ac - b^2 \neq 0$. Los coeficientes a, b, c satisfacen los requerimientos (iii) del Teorema III.4.1, o sea, la forma cuadrática será creciente, $P_x(U, V) \geq 0$, donde

$$\begin{aligned} P_x(U, V) &= (T^\# B - B)_x(U, V) \\ &= a_1 U_1^2 + 2b_1 U_1 V_1 + c_1 V_1^1 - (aU^2 + 2bUV + cV^2) \end{aligned}$$

Denote $\phi' = d\phi_1/dr_1$, entonces

$$\begin{aligned} P_x(U, V) &= (dr_1)^2 [(a_1 - a) \cos^2 \phi_1 + 2(b_1 - b)(K_1 + \phi')(\cos \phi_1) \\ &\quad + (c_1 - c)(K_1 + \phi')^2 + a(2\tau \cos \phi_1 (K_1 + \phi') - \tau^2 (K_1 - \phi')^2) \\ &\quad + 2b(-2K_1 \cos \phi_1 + \tau(K_1 - \phi')^2) + c4K_1 \phi'] \end{aligned}$$

Por tanto, el signo de P depende del siguiente polinomio de grado dos en ϕ' :

$$\begin{aligned} &(c_1 - c - a\tau^2 + 2b\tau)(\phi')^2 + \\ &+ 2[(b_1 - b) \cos \phi_1 + K_1(c_1 - c) + a\tau \cos \phi_1 + a\tau^2 K_1 - 2b\tau K_1 + 2cK_1] \phi' + \\ &+ (a_1 - a) \cos^2 \phi_1 + 2(b_1 - b)K_1 \cos \phi_1 + (c_1 - c)K_1^2 - \\ &- 2a\tau K_1 \cos \phi_1 - a\tau^2 K_1^2 + 4bK_1 \cos \phi_1 + 2b\tau K_1^2 \end{aligned} \quad (\text{V.4.4})$$

Comenzamos fijando $b = b_1 = 1$ para tener un control simple de la no degeneración de B y simplificar nuestros cálculos. Entonces $P > 0$ para todo ϕ' si y sólo si

$$c_1 - c - a\tau^2 + 2\tau > 0, \quad y \quad (\text{V.4.5})$$

$$\begin{aligned} &(c - 2\tau + a\tau^2)[4c_1 - 2E + (a_1 - a)E^2/4] \\ &+ c_1 E [2 - 2a\tau - (a_1 - a)E/4] + a^2 \tau^2 E^2/4 < 0 \end{aligned} \quad (\text{V.4.6})$$

si $K_1 \neq 0$, donde $E = -2K_1^{-1} \cos \phi$. Si la componente de q_1 es focalizadora ($K_1 < 0$), entonces $E = L_1 = 2R_1 \cos \phi_1$ es el tiempo que la trayectoria llegando a (o saliendo de) q_1 , se queda en el círculo de curvatura de ∂Q en q_1 . En el caso $K_1 = 0$ la condición (V.4.6) se transforma en

$$(a_1 - a)(c - c_1 - 2\tau) + a a_1 \tau^2 < 0 \quad (\text{V.4.7})$$

Observamos que el método de las formas cuadráticas es perfectamente equivalente al de los conos descripto antes, si $b^2 - ac = 1$ en (V.4.3). En los casos que analizaremos a continuación, esta condición será satisfecha.

OBSERVACIÓN V.4.2. Esta presentación da una nueva prueba de: un billar semidispersor es hiperbólico si el conjunto de los puntos cuyas trayectorias sólo chocan componentes neutras de la frontera tiene medida nula. En realidad, si tomamos $a \equiv c \equiv 0$, entonces (V.4.6) es satisfecho para fronteras dispersoras ($K > 0$). Sobre componentes neutras consideramos la misma forma cuadrática. Entonces $P_x \geq 0$ (en realidad es positiva en $T_x M$, con la excepción de una dirección). También, P se hace estrictamente positiva siempre que la trayectoria choque una componente dispersora.

Estudiaremos ahora condiciones **locales** a ser satisfechas por un arco focalizador Γ de modo que pueda hacer parte de la frontera de un billar hiperbólico. Las siguientes observaciones heurísticas son muy importantes. Analizamos el comportamiento de las expresiones (V.4.5) y (V.4.6) durante una serie de reflexiones sucesivas en Γ , con $\phi \cong \pm\pi/2$. Tenemos que $L \cong \tau \cong 0$, y a, c son continuas en un entorno de los puntos $(q_1, \pm\pi/2)$. Entonces, si $c \gg \tau$, el discriminante Δ de (V.4.4) satisface: $\Delta \cong 4cc_1 \cong 4c_1^2$. Como debemos tener $\Delta < 0$, se deduce que $\lim_{\phi \rightarrow \pm\pi/2} c = 0$. Si $c \ll \tau$, tenemos $\Delta \cong 4\tau E \cong 4\tau^2$. En conclusión, debe satisfacerse $c \cong \tau$, lo cual justifica la elección $c = \tau$ o $c = L = 2R \cos \phi$, que corresponde al segmento de trayectoria dentro del círculo de curvatura.

Si $a \equiv 0$, $c = L$, y $c_1 = L_1$, un cálculo inmediato en (V.4.6) da

$$L + L_1 < 2\tau. \quad (\text{V.4.8})$$

Wojtkowski [W86] probó que las curvas que satisfacen estas condición son absolutamente focalizadoras. La prueba está basada en que los frentes de onda entrantes en ∂Q con el mismo ángulo, se focalizan a distancia $L/2$ en la trayectoria que sale de $x = (q, v)$. Ver Ejercicio V.11.a. Si la componente de la frontera es C^4 , la condición (V.4.8) es equivalente a

$$\frac{d^2 R}{ds^2} < 0 \quad (\text{V.4.9})$$

donde s es la longitud de arco de la curva focalizadora y $R = -1/K$ es el radio de su curvatura [W86]. Ver Ejercicio V.11.b.

Si $a \equiv 0$, $c = \tau$, y $c_1 = \tau_1$, obtenemos la condición

$$L_1(\tau + \tau_1) < 2\tau\tau_1. \quad (\text{V.4.10})$$

Curvas que satisfacen (V.4.10) también son absolutamente focalizadoras. Ver Teorema V.4.5. La condición de arriba es “localmente equivalente” ([Ma88], ver Ejercicio V.13) a

$$\frac{d^2 R^{1/3}}{ds^2} > 0 \quad (\text{V.4.11})$$

Por localmente equivalente significamos que ambas condiciones son equivalentes para τ y τ_1 suficientemente cortos. Esta equivalencia es más débil que la establecida entre, (V.4.8) y (V.4.9).

Diremos que una curva suave puede ser parte de un billar hiperbólico si es posible definir una forma cuadrática creciente para choque sucesivos en esa curva (esta definición puede ser también expresada en términos de campos de conos invariantes). Ahora bosquejamos la prueba de este resultado sorprendente, que fue obtenido por Donnay [Do91], para curvas C^6 , y por Markarian, independientemente, para curvas C^4 . El teorema define *arcos cortos focalizadores*:

Teorema V.4.3 ([Ma94]). *Cualquier arco focalizador C^4 suficientemente corto puede ser una componente regular de la frontera de un billar hiperbólico.*

Demostración. Sea $q_1 + r(A)e^{iA}q_1'$ una parametrización de una curva C^k , ($k \geq 4$) en coordenadas polares, con eje polar formado por la tangente orientada en q_1 donde A es el ángulo polar y $r(0) = 0$. Se toma el desarrollo de Taylor de $r(A)$ en $r(A) = \dot{r}A + \frac{\ddot{r}}{2}A^2 + \dots$ de modo que $\dot{r} = \dot{r}(0) = -2/K_1 > 0$.

Si los coeficientes a y c son diferenciables con respecto a s, B (aquí B es el ángulo entre la tangente orientada y la dirección de la trayectoria), entonces ellos también pueden ser desarrollados en series de Taylor:

$$\begin{aligned} c(s, B) &= C_1s + C_2B + C_3s^2 + C_4sB + C_5B^2 \\ &\quad + C_6s^3 + C_7s^2B + C_8sB^2 + C_9B^3 + \dots \\ a(s, B) &= A_0 + A_1s + A_2B + A_3s^2 + \dots \end{aligned}$$

En el punto q_1 se tiene $s = 0$ y $B = A$ por lo que $c_1 = c(0, A) = C_2A + C_5A^2 + C_9A^3 + \dots$. En el punto q tenemos $s = s(-A)$ y $c = c(q, s(-A)) = (-C_1 + 1)\dot{r}A + (C_1\frac{\ddot{r}}{2} + \frac{\ddot{r}}{2} + \dot{r}^2C_3 - \dot{r}C_4 + C_5)A^2$. Como $c \approx \tau = r(A)$, en primera aproximación, resulta $C_1 = 0$. Además:

$$\begin{aligned} a_1 &= A_0 + A_2A + A_5A^3 + \dots, \\ a &= A_0 + (-\dot{r}A_1 + A_2)A + [\frac{\ddot{r}}{2}A_1 + \frac{\ddot{r}}{2}A_2 + \dot{r}^2A_3 - \dot{r}A_4 + A_5]A^2 + \dots \end{aligned}$$

Si sustituimos estas expresiones y el desarrollo de τ y L en (V.4.5) y (V.4.6), obtenemos que (V.4.5) es inmediatamente verificada y que el término de la izquierda en (V.4.6) es $\zeta A^3 + \sigma A^4 + \dots$ con

$$\zeta = 3\dot{r}\ddot{r} + 2\dot{r}^3C_3 - 2\dot{r}^2C_4$$

$$\ddot{r}^2 - \frac{4}{3}\dot{r}(\dot{r} + \ddot{r})2\dot{r}C_5 + 4C_5^2 + 2\dot{r}^2(-\dot{r}C_3 - \dot{r}^2C_6 + \dot{r}C_7 - C_8) +$$

$$2\dot{r}^2A_0C_5 - \dot{r}^2\ddot{r}A_0 + \frac{\dot{r}^4}{4}(A_0^2 - 2A_1) + \frac{2C_5}{\dot{r}}\zeta. \quad (\text{V.4.12})$$

Dado cualquier arco C^4 pequeño (esto es, dado $\dot{r} > 0$ en nuestras coordenadas polares), podemos elegir C_i, A_i convenientemente, de modo que $\zeta = 0$ y $\sigma < 0$. Entonces (V.4.6) serán satisfechas. Por ejemplo, podemos elegir C_6 de modo que $-2\dot{r}^4C_6$ domine todos los otros (se pueden consultar todos los detalles en [Ma94]). \square

Hiperbolicidad de mesas de billar más generales. Hemos descrito clases de arcos focalizadores que *podrían* ser usados en la construcción de billares hiperbólicos. Esto no significa, naturalmente, que cualquier mesa de billar con esos arcos en la frontera, sea hiperbólica. Ahora explicamos cómo unir diversos tipos de arcos en la frontera de modo de construir una mesa de billar hiperbólica. Algunas reglas simples deben ser respetadas. Debemos verificar condiciones en los ángulos interiores de las componentes adyacentes, distancias entre distintas componentes, etc. El siguiente teorema muestra como construir billares planos con comportamiento hiperbólico cuyas componentes de frontera sean de cualquier tipo con la única condición que las componentes focalizadoras satisfagan (V.4.8) o (V.4.10). Observamos que el caso (iii) del Teorema incluye a todos los billares de Bunimovich.

Teorema V.4.4. *Se pueden construir billares hiperbólicos de las siguientes maneras:*

- (i) *las componentes de las fronteras pueden ser cualquier curva C^3 , siempre que las curvas focalizadoras satisfagan (V.4.8) (para curvas C^4 , (V.4.8) puede ser reemplazada por (V.4.9)). Los círculos de semicurvatura de cada punto de cada componente focalizador no pueden contener puntos de círculos de semicurvatura de otras componentes focalizadoras. Componentes focalizadoras adjuntas deben formar ángulo interior mayor que π . Componentes focalizadoras y dispersoras adyacentes deben formar ángulo interior no menor que $\pi/2$. Componentes focalizadoras y neutras adyacentes deben formar ángulos interiores mayores que $\pi/2$.*
- (ii) *las componentes de las fronteras pueden ser cualquier curva C^3 , siempre que las curvas focalizadoras satisfagan (V.4.10). Los círculos de curvatura de las componentes focalizadoras no deben contener puntos de otras componentes de la frontera. Las condiciones sobre los arcos adyacentes son las mismas que en (i).*
- (iii) *Los arcos de la frontera son como en (i) o (ii), pero las igualdades son admitidas en (V.4.8), (V.4.10). Casi toda trayectoria debe pasar por lo menos por dos componentes distintas de la frontera.*

Demostración. Probaremos (ii). La prueba de (i) se hace con ideas análogas. Definiremos formas cuadráticas convenientes con $a = 0, b = 1$ en todos los casos, y estableceremos $c = \tau$ si la reflexión ocurre en una componente focalizadora y $c = 0$ en los demás casos.

Se tienen varios casos. Si q y q_1 están en componentes dispersoras o focalizadoras, el argumento de la Observación V.4.2 se aplica. Si q y q_1 están en la misma componente focalizadora, se satisface la condición (V.4.10) y la forma cuadrática es creciente. Si q y q_1 pertenecen a diferentes componentes focalizadoras y q_2 no está en la intersección de una componente neutra con el círculo de curvatura en q_1 , entonces $\tau, \tau_1 > L_1$. Sea $\tau \geq \tau_1$, entonces $L_1(\tau + \tau_1) < \tau_1(\tau + \tau) = 2\tau\tau_1$, y la condición (V.4.10) es satisfecha. Si q, q_1 están en componentes focalizadoras y q_2 pertenece a una componente neutra adyacente, observamos que $\tau_1 + \tau_2 > L_1$ porque q_3 no está en el círculo de curvatura de q_1 . Más aún, si q_2 está en una componente neutra, entonces $D_{x_2}T \cdot D_{x_1}T$ actúa como si el punto x_2 fuera saltado, y los tiempos entre colisiones son sumados; ver Ejercicio V.5. Finalmente, si q_1 es una componente neutra contenida en el círculo de curvatura en q_2 , el análisis es similar al anterior, tomando en cuenta que $\tau + \tau_1 > L_2$.

La parte (iii) se prueba usando las mismas formas cuadráticas, las cuales en este caso crecerán eventualmente (ver Observación III.4.2 después del Teorema III.4.1). \square

El teorema anterior es válido si en ambas partes (i) and (ii) también admitimos arcos focalizadores cortos, tal como fueron definidos en el Teorema V.4.3. Damos un bosquejo de la prueba (que el lector interesado puede completar), en el caso que las curvas focalizadoras puedan ser arcos focalizadores cortos C , o satisfagan la condición (V.4.8). Definimos formas cuadráticas con $a = 0$:

(i) Si $q_1, q_2 \in C$ y $q_3 \notin C$ definimos $c_2 = L_2$. Entonces (V.4.5) se satisface de inmediato porque el primer término, en primera aproximación, es igual a $2\dot{r}A$. (V.4.6) se satisface si en (V.4.12) tomamos $\zeta = 0, C_5 = 0 = A_0 = A_1$, y C_6 es suficientemente grande.

(ii) Si $q_1 \in C$ y $q_2 \notin C$, consideramos inicialmente sólo la condición $c_2 \geq 0$. En el primer término de (V.4.5) tenemos $c_2 - L_1 + 2\tau_1 > 2\tau_1 - L_1 > 0$ si $2\tau_1 > L_1$; esto significa que la componente donde está q_2 debe estar fuera del círculo de semicurvatura de cualquier punto de C . Si $K_2 = 0$, el primer término de (V.4.7) se hace cero. Entonces para que la forma B sea creciente a lo largo de las trayectorias, ellas deben eventualmente pasar por componentes no neutras.

Si $K_2 \neq 0$, el primer término de (V.4.6) es $(L_1 - 2\tau_1)(4c_2 - 2E) + 2c_2E$.

Si $K_2 > 0$, entonces $E < 0$ y es suficiente, nuevamente, considerar $2\tau_1 > L_1$.

Si $K_2 < 0$, definimos $c_2 = L_2$ y la expresión anterior se transforma en $2L_2(L_1 + L_2 - 2\tau_1)$ la cual es

negativa si $L_1 + L_2 < 2\tau_1$. Esto significa que el círculo de semicurvatura de las componentes focalizadoras no se intersectan entre ellas.

(iii) Si $q_1, q_3 \notin C$, $q_2 \in C$, sea $c_1 \geq 0$, $c_2 = L_2$. El primer término de (V.4.5) es $L_2 - c_1 + 2\tau_1 > 2\tau_1 - c_1$; por lo que (V.4.5) se satisface si $2\tau_1 \geq c_1$.

El primer término en (V.4.6) es $2L_2(c_1 + L_2 - 2\tau_1)$. Si la componente a la que pertenece q_1 es dispersora o neutra, definimos $c_1 = 0$; si es focalizadora, definimos $c_1 = L_1$. Entonces (V.4.6) se satisface si los arcos satisfacen las condiciones que aparecen en la parte (ii).

(iv) Si $q_1 \notin C$ y $q_2, q_3 \in C$, la verificación de (V.4.5) no necesita de nuevas condiciones y, puesto que $c_2 \simeq L_2$, el primer término de (V.4.6) es $(c_1 - 2\tau_1)(4c_2 - 2L_2) + 2c_2L_2 \simeq 2c_2(c_1 + L_2 - 2\tau_1)$, el cual es negativo en las mismas condiciones halladas en (iii).

(v) Si un segmento de trayectoria está entre dos componentes que no son arcos focalizadores cortos, la forma cuadrática se define con $c = L$ si q está en una componente focalizadora y con $c = 0$ en los demás casos.

Argumentos similares funcionan si las componentes focalizadoras satisfacen (V.4.10). Ver [Ma94].

Conos invariantes y fracciones continuas. Fue observado en la Sección III.4 que una forma cuadrática no degenerada B automáticamente define un campo de conos en $\mathcal{T}M$ consistente en los vectores cuyos valores por B son no negativos. Si la forma cuadrática es creciente por DT , estos conos son invariantes por T . Dada una forma cuadrática no degenerada B definimos

$$C^u(x) = \{v \in \mathcal{T}_x M : B_x v \geq 0\} \quad (\text{V.4.13})$$

Este campo de conos satisface la condición de invariancia $D_x T(C^u(x)) \subset C^u(Tx)$.

Los campos de conos pueden ser definidos sin formas cuadráticas. En realidad es a veces más fácil construir un campo de conos invariantes usando propiedades geométricas u “ópticas” del sistema. Este fue el caso de los billares dispersores y de Bunimovich en las dos Secciones anteriores.

Esas construcciones fueron usadas con éxito para probar hiperbolicidad y ergodicidad de billares. En todos los casos alguna versión de los Teoremas III.4.4 y III.4.5 del capítulo III fueron usadas. Estos métodos están muy bien explicados en trabajos de Wojtkowski [W86], Donnay [Do91] y Bunimovich [Bu92]. Ellos incluyen detallados estudios de propiedades de desfocalización de frentes de ondas chocando en fronteras focalizadoras de la mesa de billar.

La hiperbolicidad de billares con fronteras focalizadoras satisfaciendo la condición (V.4.8) fue probado en primer lugar usando esas técnicas geométricas de conos. La hiperbolicidad de algunas otras mesas de billar convexas (que nos satisfacen ninguna de las condiciones estudiadas previamente) fue probada usando otras técnicas de conos. Ver ejemplos, al final de esta Sección.

Una pregunta natural aparece con los billares hiperbólicos con fronteras focalizadoras. ¿Es posible encontrar expresiones para las direcciones estables e inestables, similares a las fracciones continuas de la Sección V.2? El siguiente teorema da una respuesta afirmativa.

Teorema V.4.5. *Considere mesas de billar satisfaciendo las condiciones del Teorema V.4.4. Para ν -casi todo punto las fracciones continuas (V.1.4) y (V.1.6) convergen y dan las direcciones de E_x^s, E_x^u de acuerdo con la fórmula (V.2.1).*

Demostración. Comenzamos recordando un interesante teorema sobre la convergencia de fracciones continuas cuyos elementos pueden ser positivos o negativos ([Bu92], Teoremas 10 y 11). Una fracción continua

$$a_1(x) + \frac{1}{a_2(x) + \frac{1}{a_3(x) + \frac{1}{a_4(x) + \frac{1}{a_5(x) + \dots}}}} \quad (\text{V.4.14})$$

es convergente si las siguientes dos condiciones se satisfacen:

- (a) todos los elementos impares son positivos y su suma es infinita,
- (b) para cualesquiera dos elementos negativos $a_{2k'-1}, a_{2k''+1}$, $k' < k''$, entre los cuales no hay otros elementos negativos, tenemos

$$a_{2k'} \geq |a_{2k'-1}|^{-1}(2 + \delta_{k'}) + |a_{2k''+1}|^{-1} \left(2 - \frac{\delta_{k''}}{1 + \delta_{k''}} \right) \quad (\text{V.4.15})$$

para algunos $\delta_l \geq -1$, $l = 1, 2, \dots$.

Probaremos el Teorema V.4.5 en el caso inestable. Observamos que las condiciones (V.4.8) y (V.4.10) son similares en términos de las componentes de $k^u(x)$ porque pueden ser reescritas, respectivamente, como

$$a_{2k} \geq |a_{2k-1}|^{-1} + 2|a_{2k+1}|^{-1} \quad (W)$$

$$|a_{2k+1}| \geq 2a_{2k}^{-1} + 2a_{2k+2}^{-1}, \quad k = 1, 2, \dots \quad (M)$$

Entonces, si la convergencia se prueba suponiendo, por ejemplo, (V.4.8), la convergencia suponiendo (V.4.10) se deducirá automáticamente porque la última corresponde a un corrimiento de una posición en la fracción continua. La condición (W) para dos choques consecutivos en la misma componente focalizadora es exactamente (V.4.15) para $k' = k''$, con $\delta_{k'} = 0$. Si $T^{-k'+1}x$ y $T^{-k''}x$ son choques sucesivos en componentes focalizadoras (no adyacentes), entonces las condiciones sobre componentes no adyacentes en el Teorema V.4.4 implican directamente (V.4.15) con $\delta_{k'} = \delta_{k''} = 0$. Las fracciones (V.1.4) y (V.1.6) dan las direcciones de los vectores estables e inestables en virtud de nuestra manera de construir los frentes de onda estables e inestables. \square

Anotamos que los billares hiperbólicos con componentes focalizadoras de la frontera tienen entropía positiva, que puede ser calculada por las mismas fórmulas que en el caso dispersor. En particular, (V.2.9), (V.2.12) y (V.2.13) son válidas. Ver [CM92] y [C97] por detalles sobre estas cuestiones.

Por último, probamos

Lema V.4.6. *Las curvas que satisfacen (M) son arcos absolutamente focalizadores.*

Demostración. Observe que la convergencia de $k^u(x)$ implica la convergencia de

$$k^-(x) = \frac{1}{\tau(T^{-1}x) + \frac{1}{\frac{2K(T^{-1}x)}{\cos \phi(T^{-1}x)} + \dots}} \quad (V.4.16)$$

que es la curvatura de la variedad inestable antes de la reflexión en x (cf. la ecuación del espejo (V.1.2)). Como las curvas que satisfacen (W) son arcos absolutamente focalizadores –esto fue observado inmediatamente después de la fórmula (V.4.8)– el signo de $k_W(x)$, o sea el signo del primer término de $k_W(x)$, es negativo, por lo que el signo de k_W^- , o sea el signo del segundo término en k_W , es positivo. Entonces el signo de k_M^- , o sea el signo del primer término de k_M^- , es positivo y, finalmente, $k_M(x)$ (considerado como una parte de $k_M(Tx)$) tiene el signo de su tercer término, o sea negativo. Como estas relaciones son satisfechas en cada reflexión en una componente focalizadora, el punto de focalización del frente de onda inestable ($k_M(x) < 0$) está antes de la siguiente reflexión ($k_M^-(Tx) > 0$). Si esto es así para los frentes divergentes (antes de la reflexión), lo es también para los frentes paralelos. \square

Ejemplos:

1. Curvas que satisfacen la condición $d^2R/ds^2 < 0$. En este grupo están la epicicloide, la hipocicloide, la cicloide y, en particular, la cardioide (una curva cerrada cuya ecuación en coordenadas polares es $r(t) = 1 + \cos t$, $-\pi \leq t \leq \pi$). El arc de elipse $x = a \cos t$, $y = b \sin t$, $b^2 > a^2$, con $-\pi/4 < t < \pi/4$ también satisface esta condición. Ver [W86].

2. Curvas que satisfacen la condición $L_2(\tau_1 + \tau_2) < 2\tau_1\tau_2$. En la elipse recién mencionada, el arco definido por $\sin^2 t > b^2/(b^2 + a^2)$. Curiosamente, este arco es disjunto del anterior. Ver [Ma88].

3. Observamos que las condiciones en las partes (i) y (ii) del Teorema V.4.4 son C^3 abiertas. Por tanto, las componentes no neutras de la frontera de tales billares pueden ser perturbadas en la topología C^3 , manteniendo la hiperbolicidad de los billares resultantes. Si se aceptan las igualdades – como en (iii) del mismo Teorema – entonces las perturbaciones pueden destruir el comportamiento hiperbólico de las nuevas mesas de billar. Este es el caso cuando arcos de círculos son partes de la frontera. Si arcos de círculo son perturbados, las condiciones (V.4.8) y (V.4.10) pueden fallar, y la hiperbolicidad podría no sobrevivir. Sin embargo, en [Ma94] se probó que arcos de círculo menores que medio círculo pueden ser C^4 perturbados manteniendo el comportamiento hiperbólico del billar resultante. Un resultado similar para perturbaciones C^6 fue obtenido en [Do91].

4. Donnay [Do91] probó que arcos absolutamente focalizadores pueden ser parte de la frontera de billares hiperbólicos si son juntados con rectas suficientemente largas, formando una región convexa. Su prueba está basada en un estudio sutil de las propiedades de los frentes de onda convergentes en billares integrables. Se prueba que en la elipse ya mencionada, su mitad $y \geq 0$ es un arco absolutamente focalizador si y sólo si $b/a < \sqrt{2}$. En [MOP96] se dan cotas inferiores para la separación entre las componenes focalizadoras.

5. Es importante distinguir entre hiperbolicidad y ergodicidad (o mixing). No son equivalentes. Ver Ejercicio V.14.

EJERCICIO V.11. (a) Probar que las componentes focalizadoras de la frontera que satisfacen la condición (V.4.8) son arcos absolutamente focalizadores. Sugerencia: Considere frentes de onda paralelos y frentes que llegan en esa curva con ángulo de incidencia cosnante. Compare las posiciones relativas de sus puntos de focalización en la trayectoria saliente.

(b) Probar que las condiciones (V.4.8) y (V.4.9) en curvas Γ focalizadoras C^4 son equivalentes. Sugerencia [W86]: Sean q_0, q_1 puntos distintos de Γ , elija coordenadas cartesianas (x, y) en el plano de modo que el eje Ox pase por q_0 y q_1 . Suponga que q_0 es el origen, q_1 tiene abscisa positiva, y el arco q_0q_1 de Γ está debajo del eje de Ox , o sea que tiene ordenadas no positivas. Si $\Gamma(s) = (x(s), y(s))$ es parametrizada por la longitud de arco, sea $\phi(s)$ el ángulo del eje Ox con la tangente a γ en s . Note que el radio de curvatura de γ es $R = ds/d\phi$. Sea $\tau = |q_0q_1|$. Entonces $L = \int_a^b \frac{dx}{ds} ds = R(b) \sin \phi(b) - R(a) \sin \phi(a) - \int_a^b \sin \phi \frac{dR}{ds} ds$ y $2\tau - L - L_1 = \int_a^b y(s) \frac{d^2R}{ds^2} ds$.

EJERCICIO V.12. Probar que las condiciones (V.4.10) y (V.4.11) de un arco Γ focalizador C^4 son “localmente equivalentes”. Sugerencia [Ma88]: Sea $A = \pi/2 - \phi$ y $\gamma(A) = \rho(A)\gamma'(0)e^{iA}$, $\rho(0) = 0$ la expresión de la frontera en el entorno de un punto de colisión $\gamma(0)$. Escriba el desarrollo de Taylor de τ_1, τ para pequeños valores de A . La expresión de la curvatura en coordenadas polares es $K(A) = [\rho^2 + 2\rho'^2 - \rho\rho'']/[(\rho^2 + \rho'^2)^{3/2}]$. Entonces pruebe que (V.4.10) $\iff 3K_1 \rho''(0)^2 < 4[\rho'(0) + \rho'''(0)] \iff 3KK'' < 2K'^2 \iff$ (V.4.11).

EJERCICIO V.13. Complete los detalles en las pruebas de los Teoremas V.4.3 y V.4.4.

EJERCICIO V.14. Considere una elipse, córtela a lo largo del eje mayor. Aparte ambas mitades en la dirección del eje menor. Junte ambas mitades con líneas paralelas al eje menor. Agregue dos “paredes” paralelas uniendo las posiciones de los focos. Pruebe que la mesa de billar resultante es hiperbólica, pero tiene por lo menos tres componentes ergódicas diferentes. Sugerencia: ver [W86], Appendix C.

EJERCICIO V.15. Calcule el producto de las matrices $D_{x_1}T \cdot D_xT$, donde $x_1 = Tx$ y x_1 pertenecen a componentes neutras de la frontera.

V.5. Observaciones Finales y Perspectives

Los dos últimos capítulos introducen al lector en la moderna teoría matemática de billares caóticos. Son una introducción técnica y munuciosa, después de la cual el lector puede profundizar en los los principales trabajos de investigación en el área. Pero son sólo una introducción. Cubre los resultados básicos; como fue indicado en varios casos, muchos resultados fundamentales están más allá de los propósitos y el alcance de estas notas.

Las más importantes dificultades aparecen en el estudio de las propiedades ergódicas de los billares caóticos. En la Sección III.5 dimos el plan general de la prueba de la ergodicidad para sistemas hiperbólicos con singularidades. Se aplica a los billares dispersores y otros billares hiperbólicos, pero la verificación de cada una de las condiciones involucra un trabajo muy grande y difícil. Hay varios trabajos exponiendo pruebas de esos resultados, que fueron citados en esa Sección. Ellos se refieren a los mismo métodos con diferentes enfoques. En realidad las diferencias principales, imposibles de explicar aquí, se dan ya en los trabajos claves de Sinai [Si70] y de Sinai y Chernov [SC87]. Un resultado muy global sobre la ergodicidad (y la propiedad Bernoulli) de billares planos está probado en el reciente [DMM05]

Mientras que la maquinaria para la prueba de la ergodicidad es bien entendida en dimensión dos (o sea $\dim M = 2$), la situación es mucho menos clara en dimensiones más altas. La geometría de los billares multidimensionales es muy complicada. La estructura de las singularidades y el comportamiento de la transformación de billar T es aún más complicada. Ver, por ejemplo [BCST02], [BCST03].

Uno de los modelos más interesantes es el célebre gas de las bolas duras, para el cual Boltzmann conjeturó, en términos bastante oscuros y ya hace más de 100 años, la ergodicidad. Hemos tratado este modelo en la Sección IV.5 y se puede encontrar una descripción bastante sencilla de la historia del asunto en [Ma02] y [Sz96]. Sinai especificó la conjetura de Boltzmann en 1963 [Si63] afirmando que el sistema de N bolas duras en el toro (en espacios de cualquier dimensión $d \geq 2$) es ergódico. Él mismo obtuvo una prueba para $N = d = 2$ (dos discos duros en el toro de dimensión dos) en 1970 [Si70], reduciendo el sistema a un gas de Lorentz en

el plano. Este es el trabajo seminal de la teoría de billares (y sistemas con singularidades) caóticos. El caso $N = 3$ mostró ser mucho más complicado, y fue resuelto en 1991 por Krámli, Simányi y Szász [KSS91]. La conjetura para cualquier N sigue abierta. Sobre el estado de este tema a principios de siglo el libro [Sz00], compilado por D. Szász, es la mejor referencia.

A pesar de las dificultades, es de esperar que la mayoría de los billares descritos en nuestras Secciones IV.5, V.2, V.4 sean, en realidad, ergódicos. En este sentido, pues, ellos deberían ser similares a los difeomorfismos de Anosov.

Resulta que los billares caóticos también tienen propiedades ergódicas más avanzadas, de acuerdo a la jerarquía vista en la Sección II.3. Como ya lo observara Sinai en su trabajo de 1970 [Si70], la ergodicidad de los billares dispersores implica casi automáticamente las propiedades mixing y K-mixing. Más tarde, Gallavotti y Ornstein [GO74] probaron la propiedad Bernoulli, y después Chernov and Haskell [CH96], Ornstein y Weiss [OW98] (independientemente) probaron que sistemas con singularidades hiperbólicos y K-mixing son siempre Bernoulli, o sea isomorfos a un shift de Bernoulli. Esta es la propiedad más alta en la caoticidad desde el punto de vista de la teoría de la medida, poniendo a los billares en el mismo nivel que los sistemas uniformemente hiperbólicos.

Para fines más prácticos (aplicaciones en física), la propiedad de Bernoulli es, sin embargo, de poca ayuda. Como explicamos en la Sección II.4, las estadísticas de las iteraciones de las funciones suaves juegan el papel fundamental. Especialmente, la tasa de decaimiento de correlaciones y el Teorema del Límite Central (TLC) son los asuntos más importantes. Para los difeomorfismos de Anosov, el TLC fue probado al principio de los setenta, y al mismo tiempo se estableció la tasa exponencial de decaimiento de correlaciones, que es aceptada como la más rápida posible. Esto fue hecho por Sinai, Ruelle and Bowen [?, Ru76, Ru78, Bo75]; el elemento clave fue la construcción de las llamadas particiones de Markov.

Desde 1980, un proyecto mayor fue extender las técnicas de Sinai-Ruelle-Bowen a los billares caóticos. Este trabajo continúa, se ha exployado por más de 20 años y nuevas técnicas se han introducido (ver, por ejemplo [Ba00]). Primero, las particiones de Markov fueron construídas [BS80, BSC90], después fue probado el TLC [BS81, BSC91]. Sobre el decaimiento decorrelaciones, primero se probó [BS81, BSC91] un resultado más débil (una cota superior subexponencial). Al final de los noventa se probó por diferentes aproximaciones que el decaimiento de los billares dispersores es exponencial [Y98, C99a]. Al final se hizo claro que los billares dispersores se comportan a los efectos de las aplicaciones físicas como los difeomorfismos de Anosov. Esto fue resumido por Gallavotti y Cohen en su “Axioma C” [GC95].

Debe ser observado que todas las propiedades estadísticas antes referidas han sido obtenidas sólo para billares dispersores en dimensión dos. Para otros billares caóticos, y especialmente en dimensiones altas, el asunto está ámpliamente abierto. Más aún, la tasa de decaimiento de correlaciones para billares no uniformemente hiperbólicos, como el stadium y el gas de las bolas duras, no puede ser exponencial. Ver Sección II.4. Resultados recientes sobre decaimiento polinomial en billares, se encuentran en [Ma04], [CZ05].

Faltan estudiar otras propiedades finas de los billares caóticos. Por ejemplo, en [St89, C91b] se han estimado cálculos asintóticos de las órbitas periódicas, pero no se ha obtenido aún nada semejante al Teorema III.3.8.

Recientemente, se han hecho populares estudios de modificaciones más “realistas” de los billares. Uno reemplaza las reflexiones elásticas en la frontera rígida ∂Q por un potencial repulsivo, o por diversas formas de choques no elásticos sobre la frontera rígida o le agrega fuerzas externas que actúan sobre la partícula dentro de Q , o en partes de la mesa. Más aún se pueden establecer diversos tipos de salidas aleatorias y estudiar la dinámica resultante.

Una corriente firme de nuevos resultados interesantes en estas direcciones indica que la teoría de billares caóticos no se restringe sólo a los billares clásicos. Y que juega un papel creciente en diversos campos de la moderna físico-matemática, incluyendo la mecánica cuántica.

Apéndice A

Teoría de la medida. Ejemplos de ergodicidad

A.1. Medidas

La familia \mathcal{O} de subconjuntos del conjunto X es llamada un *álgebra* si

$$X \in \mathcal{O}; \quad A \in \mathcal{O} \Rightarrow A^c \in \mathcal{O}; \quad A, B \in \mathcal{O} \Rightarrow A \cup B \in \mathcal{O}$$

Fácilmente se deduce que

$$\begin{aligned} A, B \in \mathcal{O} &\Rightarrow A \cap B = (A^c \cup B^c)^c \in \mathcal{O} \\ A, B \in \mathcal{O} &\Rightarrow A \setminus B = A \cap B^c \in \mathcal{O} \end{aligned}$$

La familia \mathcal{O} es llamada una σ -álgebra si es un álgebra y satisface

$$A_i \in \mathcal{O}, i = 1, 2, \dots \Rightarrow \cup_i A_i \in \mathcal{O}.$$

Si \mathcal{O}_0 es una familia de subconjuntos de X , la σ -álgebra generada por \mathcal{O}_0 es la más pequeña σ -álgebra que contiene \mathcal{O}_0 : ésta es la σ -álgebra \mathcal{O} tal que $\mathcal{O} \supset \mathcal{O}_0$ y $\mathcal{O} \subset \mathcal{O}_1$ para toda σ -álgebra \mathcal{O}_1 que contiene \mathcal{O}_0 .

Si \mathcal{O} es un álgebra de subconjuntos X , decimos que una función $\mu : \mathcal{O} \rightarrow [0, \infty]$ es una *medida sobre \mathcal{O}* si $\mu(\emptyset) = 0$ y para toda colección finita o numerable de conjuntos disjuntos $A_i \in \mathcal{O}$, $i = 1, 2, \dots$, $\cup_i A_i \in \mathcal{O}$, tenemos

$$\mu(\cup_i A_i) = \sum_i \mu(A_i)$$

Una medida es σ -finita si X puede ser descompuesto en una unión numerable de conjuntos con medida finita.

Cabe preguntarse si se puede extender siempre una medida sobre un álgebra a una σ -álgebra más grande. La respuesta está contenida en el siguiente Teorema de Extensión cuya prueba puede encontrada en [Ha50] or [F96].

Teorema A.1.1 (Hahn-Kolmogorov). *Sea \mathcal{O}_0 un álgebra de conjuntos de X y ν una medida definida sobre \mathcal{O}_0 . Entonces, existe una σ -álgebra \mathcal{O} , y una medida μ on \mathcal{O} tal que $\mathcal{O}_0 \subset \mathcal{O}$ y $\nu(A) = \mu(A)$ para cada $A \in \mathcal{O}_0$ (μ es una extensión ν de \mathcal{O}). Si ν es σ -finita, la extensión sobre la σ -álgebra generada por \mathcal{O}_0 es única.*

Si X es un espacio topológico, definimos la σ -álgebra de Borel de X como la σ -álgebra generada por los conjuntos cerrados de X . Puesto que los complementos de los abiertos son cerrados, coincide con la σ -álgebra generada por los subconjuntos abiertos de X . Los conjuntos en el σ -álgebra de Borel son llamados *conjuntos de Borel*.

El ejemplo estándar de una σ -álgebra de Borel con una medida es la medida de Lebesgue en el espacio euclidiano \mathbb{R}^n . Un *cubo* en \mathbb{R}^n es un conjunto Q de la forma $Q = J_1 \times \dots \times J_n$ donde J_1, \dots, J_n son intervalos acotados en \mathbb{R} . Su volumen $\text{Vol}(Q)$ está definido por el producto

$$\text{Vol}(Q) = |J_1| \cdot |J_2| \cdot \dots \cdot |J_n|$$

En el caso del toro \mathbb{T}^n se trabaja de manera similar. El toro \mathbb{T}^n es el producto cartesiano $S^1 \times \dots \times S^1$, donde S^1 es el círculo unitario $S^1 = \{z \in \mathbb{C} : |z| = 1\}$. Un cubo en \mathbb{T}^n es un conjunto Q de la forma $Q = J_1 \times \dots \times J_n$ donde J_1, \dots, J_n son intervalos en S^1 . El volumen $\text{Vol}(Q)$ de un cubo $Q \subset \mathbb{T}^n$ se define por

$$\text{Vol}(Q) = \prod_{i=1}^n \frac{|J_i|}{2\pi}.$$

Teorema A.1.2. *Existe una única medida (de Lebesgue) λ sobre el σ -álgebra de \mathbb{R}^n (\mathbb{T}^n) tal que $\lambda(Q) = \text{Vol}(Q)$ para todo cubo $Q \subset \mathbb{R}^n$ (\mathbb{T}^n). En el caso del toro esta medida es de probabilidad.*

La unicidad se demuestra en el Ejercicio A.1.

Para probar la existencia de la medida de Lebesgue en \mathbb{R}^n se introduce la noción de medida exterior. Sea X un espacio métrico y denótese $\mathcal{P}(X)$ la familia de todos sus subconjuntos. Una función $\mu^* : \mathcal{P}(X) \rightarrow [0, +\infty]$ es llamada una *medida exterior* en X si satisface las siguientes propiedades:

- (a) $\mu^*(\emptyset) = 0$
- (b) $A_i \subset X, i = 1, 2, \dots \Rightarrow \mu^*(\cup_i A_i) \leq \sum_i \mu^*(A_i)$
- (c) Si $A \subset X, B \subset X$ y $\text{dist}(A, B) > 0$, entonces $\mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$.

Aquí $\text{dist}(A, B) \stackrel{\text{def}}{=} \inf\{\text{dist}(x, y) : x \in A, y \in B\}$.

Teorema A.1.3. *Si X es un espacio métrico separable y μ^* es una medida exterior en X , entonces μ^* es una medida sobre la σ -álgebra de Borel de X .*

Por una prueba de este teorema ver [Ha50].

Se define la medida exterior de Lebesgue μ^* en \mathbb{R}^n por $\mu^*(S) = \inf \sum_i \text{Vol}(Q_i)$ donde el ínfimo es tomado sobre todo cubrimiento finito o numerable de S por cubos. $\{Q_i\}$. Es fácil probar que es una medida exterior. Por el Teorema A.1.3, μ^* es una medida sobre la σ -álgebra de Borel de \mathbb{R}^n . Entonces sólo falta probar que $\mu^*(Q) = \text{Vol}(Q)$ cuando Q es un cubo y luego tomar λ como la restricción de μ^* a la σ -álgebra de Borel. Para la verificación que $\mu^*(Q) = \lambda(Q)$ para cubos, ver Ejercicio A.2.

Definición 25. *Un espacio de medida es una terna (X, \mathcal{O}, μ) , donde X es un conjunto, \mathcal{O} una σ -álgebra de subconjuntos de X , y $\mu : \mathcal{O} \rightarrow [0, \infty]$ una medida. Si $\mu(X) = 1$, decimos que (X, \mathcal{O}, μ) es un espacio de probabilidad y μ una medida de probabilidad, o una probabilidad*

Definición 26. *Sea (X, \mathcal{O}, μ) un espacio de medida. Un conjunto $A \subset X$ es llamado un conjunto de medida nula (o un conjunto nulo) si existe $\hat{A} \in \mathcal{O}, A \subset \hat{A}$ y $\mu(\hat{A}) = 0$. Dos conjuntos $A_1, A_2 \subset X$ son equivalentes (mod 0) si su diferencia simétrica $A_1 \Delta A_2 = (A_1 \setminus A_2) \cup (A_2 \setminus A_1)$ es un conjunto de medida nula. Si \mathcal{S} es una familia de conjuntos de X decimos que $A \in \mathcal{S}$ (mod 0) si existe $A_1 \in \mathcal{S}$, que es equivalente (mod 0) a A . Decimos que $\mathcal{S}_1 = \mathcal{S}_2$ (mod 0) si para cada $A_1 \in \mathcal{S}_1$ y $A_2 \in \mathcal{S}_2$ tenemos $A_1 \in \mathcal{S}_2$ (mod 0) y $A_2 \in \mathcal{S}_1$ (mod 0). Finalmente, \mathcal{S} (mod 0)-genera \mathcal{O} si $\mathcal{O} = \bar{\mathcal{S}}$ (mod 0) donde $\bar{\mathcal{S}}$ es la σ -álgebra generada por \mathcal{S} .*

Definición 27. *Una propiedad en un espacio de medida se satisface casi todo punto (c.t.p.) si el conjunto de puntos donde puede no satisfacerse es de medida nula. O sea si el conjunto donde se satisface es todo el espacio (mod 0).*

Teorema A.1.4. *Sea (X, \mathcal{O}, μ) un espacio de medida y $\mathcal{O}_0 \subset \mathcal{O}$ un álgebra que (mod 0)-genera \mathcal{O} . Entonces para cada $A \in \mathcal{O}$ y $\varepsilon > 0$ existe $A_0 \in \mathcal{O}_0$ tal que $\mu(A \Delta A_0) \leq \varepsilon$. Cuando (X, \mathcal{O}, μ) es un espacio de probabilidad, el recíproco también es cierto.*

Teorema A.1.5 (Lema de Borel-Cantelli). *Sea (X, \mathcal{A}, μ) un espacio de probabilidad, y $\limsup_n B_n = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} B_n$, el conjunto de puntos que pertenecen a infinitos $B_n \subset \mathcal{A}$.*

(i) *Si $\sum \mu(A_n) < \infty$, entonces $\mu(\limsup_n A_n) = 0$; o sea que casi todo punto $x \in X$ pertenece a un número finito de B_n .*

(ii) *Si $\sum \mu(A_n) = \infty$, y los B_n son independientes, entonces $\mu(\limsup_n A_n) = 1$; o sea que casi todo punto $x \in X$ pertenece a un número infinito de B_n .*

Teorema A.1.6. *Si μ es una probabilidad sobre el σ -álgebra de Borel de un espacio métrico separable¹, entonces para cada conjunto de Borel A y cada $\varepsilon > 0$ existe un conjunto cerrado $A_0 \subset A$ y un conjunto abierto $A_1 \supset A$ tales que $\mu(A \setminus A_0) \leq \varepsilon$ y $\mu(A_1 \setminus A) \leq \varepsilon$.*

Teorema A.1.7. *Si X es un espacio métrico completo separable y μ es una probabilidad sobre su σ -álgebra de Borel, entonces para cada conjunto de Borel A y cada $\varepsilon > 0$ existe un conjunto compacto $K \subset A$ tal que $\mu(A \setminus K) \leq \varepsilon$.*

EJERCICIO A.1. La medida λ del Teorema A.1.2 es única. Sugerencia: Si λ' fuera otra tal que $\lambda'(Q) = \text{Vol}(Q)$ para cada cubo Q . Cubra \mathbb{R}^n por cubos disjuntos $Q_1, Q_2, \dots \in \mathcal{O}_0$, y pruebe que los conjuntos de Borel A tales que $\lambda(A \cap Q_i) = \lambda'(A \cap Q_i)$ para todo i forman un σ -álgebra que contiene todos los cubos y todos los abiertos. Por tanto incluye el σ -álgebra de Borel y $\lambda(A) = \lambda'(A)$ para todo conjunto de Borel.

¹Un espacio topológico es *separable* si tiene un conjunto denso numerable

EJERCICIO A.2. (a) Si Q, Q_1, \dots, Q_m son cubos y $Q \subset \cup_{i=1}^m Q_i$ probar que $\text{Vol}(Q) \leq \sum_{i=1}^m \text{Vol}(Q_i)$.

(b) Si Q y $Q_i, i = 1, 2, \dots$ son cubos y $Q \subset \cup_i Q_i$, probar que $\text{Vol}(Q) \leq \sum_i \text{Vol}(Q_i)$. Sugerencia: Tome un cubo compacto $Q_0 \subset Q$ con $\text{Vol}(Q_0) \geq \text{Vol}(Q) - \varepsilon$ y cubos abiertos $\hat{Q}_i \supset Q_i$ con $\text{Vol}(\hat{Q}_i) \leq \text{Vol}(Q_i) + \varepsilon/2^i$. Observe que Q_0 es cubierto por una colección finita $\hat{Q}_{i_1}, \dots, \hat{Q}_{i_m}$ y aplique (a).

EJERCICIO A.3. Sea (X, \mathcal{O}, μ) un espacio de probabilidad y $A_i \in \mathcal{O}, i = 1, 2, \dots$. Probar que

$$\mu\left(\bigcap_n \bigcup_{i \geq n} A_i\right) = \lim_{n \rightarrow +\infty} \mu\left(\bigcup_{i \geq n} A_i\right) \geq \limsup_{n \rightarrow +\infty} \mu(A_n)$$

EJERCICIO A.4. Sea (X, \mathcal{O}, μ) un espacio de probabilidad. Suponga que $A \in \mathcal{O}$ y que una sucesión $A_i \in \mathcal{O}, i = 1, 2, \dots$ satisface $\sum_i \mu(A \Delta A_i) < +\infty$. Probar que para casi todo $x \in A$ existe n tal que $x \in A_i$ for all $i \geq n$. Sugerencia: El conjunto de puntos x in A tal que $x \notin A_n$ para una cantidad infinita de valores de n coincide con $\bigcap_{n=1}^{\infty} \bigcup_{i \geq n} (A \setminus A_i)$ y $\mu\left(\bigcap_{n=1}^{\infty} \bigcup_{i \geq n} (A \setminus A_i)\right) \leq \sum_{i \geq n} \mu(A \setminus A_i)$ para todo n .

EJERCICIO A.5. Probar que si X es un espacio métrico completo y μ es una probabilidad sobre su σ -álgebra de Borel, entonces todo conjunto de Borel de X es la unión de conjuntos de Borel N con $\mu(N) = 0$ y una cantidad numerable de conjuntos compactos disjuntos.

EJERCICIO A.6. Si $A \subset \mathbb{R}^n$ es un conjunto abierto con $\lambda^*(\partial A) = 0$, entonces la función característica f_A de A (definida por $f_A(x) = 1$ si $x \in A$ y por $f_A(x) = 0$ si $x \notin A$) es integrable en el sentido de Riemann y $\lambda(A) = \int f_A(x) dx$

A.2. Transformaciones medibles

Sean X, Y dos conjuntos y sean \mathcal{O} y \mathcal{S} σ -álgebras de conjuntos de X e Y , respectivamente. Una transformación $f: X \rightarrow Y$ es *medible con respecto a \mathcal{O} y \mathcal{S}* si $f^{-1}(B) \in \mathcal{O}$ para cada $B \in \mathcal{S}$. Si X e Y son espacios topológicos, decimos que $f: X \rightarrow Y$ es medible si es medible respecto a las σ -álgebras de Borel de X e Y .

Proposición A.2.1. *Si X e Y son espacios topológicos y $f: X \rightarrow Y$ es una transformación tal que $f^{-1}(A)$ es un subconjunto de Borel X para cada abierto de Y , entonces f es medible.*

Corolario A.2.2. *Si X e Y son espacios topológicos, cada función continua $f: X \rightarrow Y$ es medible.*

Si X es un conjunto e Y es un espacio topológico decimos que una sucesión de funciones $f_n: X \rightarrow Y, n \geq 1$, converge (puntualmente) a la función $f: X \rightarrow Y$ si $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ para cada $x \in X$.

Corolario A.2.3. *Si X es un conjunto con un σ -álgebra \mathcal{O} , Y un espacio métrico con su σ -álgebra de Borel, y $f_n: X \rightarrow Y, n \geq 1$, a sucesión de funciones medibles que converge puntualmente a una función $f: X \rightarrow Y$, entonces f es medible*

Teorema A.2.4 (Teorema de Egorov). *Sea (X, \mathcal{O}, μ) un espacio de probabilidad y Y un espacio métrico. Sea $f_n: X \rightarrow Y$ una sucesión de funciones medibles que converge casi todo punto a una función $f: X \rightarrow Y$. Entonces para cada $\varepsilon > 0$ existe un conjunto $A \in \mathcal{O}$ con $\mu(A) \geq 1 - \varepsilon$ tal que $f_n|_A$ converge uniformemente a $f|_A$.*

Un espacio topológico X es *separable* si existe un conjunto numerable denso en X . Si un espacio topológico tiene base numerable es separable.

Teorema A.2.5 (Teorema de Lusin). *Sea (X, \mathcal{O}, μ) un espacio de probabilidad con X métrico separable, Y un espacio métrico separable y $f: X \rightarrow Y$ una transformación medible. Entonces para cada $\varepsilon > 0$ existe un conjunto compacto $K \subset X$ tal que $\mu(K) > 1 - \varepsilon$ y la transformación $f|_K$ es continua.*

EJERCICIO A.7. Sean X un espacio topológico y $f_\alpha: X \rightarrow \mathbb{R}, \alpha \in \mathbb{N}$, una familia numerable de transformaciones. Defina la función $\sup_\alpha f_\alpha: X \rightarrow \mathbb{R}$ by $(\sup_\alpha f_\alpha)(x) = \sup_\alpha f_\alpha(x)$. Probar que

$$\left(\sup_\alpha f_\alpha\right)^{-1}((a, b]) = \bigcap_\alpha f_\alpha^{-1}((-\infty, b]) \cap \left(\bigcup_\alpha f_\alpha^{-1}((a, +\infty))\right)$$

y que $\sup_\alpha f_\alpha$ es medible.

EJERCICIO A.8. Sea (X, \mathcal{O}, μ) un espacio de probabilidad, Y un espacio métrico separable. Una función $f: X \rightarrow Y$ es *simple* si es medible y $f(X)$ es un conjunto finito. Probar que toda función medible es el límite de una sucesión de funciones simples. Sugerencia: Mostrar que para cada $n \geq 1$ existe un cubrimiento de Y por conjuntos de Borel disjuntos $P_1^{(n)}, P_2^{(n)}, \dots$ con $\text{diam } P_i^{(n)} \leq 1/n$ para todo i . Para cada n tome N_n such that: $\mu(f^{-1}(\cup_{i > N_n} P_i^{(n)})) \leq 1/4^n$. Tome un punto $a \in Y$ y puntos arbitrarios $p_i^{(n)} \in P_i^{(n)}$. Defina $f_n: X \rightarrow Y$ by $f_n(x) = p_i^{(n)}$ if $x \in f^{-1}(P_i^{(n)}), 1 \leq i \leq N_n$, and $f_n(x) = a$ si $x \in f^{-1}(\cup_{i > N_n} P_i^{(n)})$

EJERCICIO A.9. Sean (X, \mathcal{O}, μ) un espacio de probabilidad, $\mathcal{O}_0 \subset \mathcal{O}$ una subálgebra que genera \mathcal{O} y $\mu': \mathcal{O} \rightarrow [0, 1]$ una probabilidad. Si $\mu|_{\mathcal{O}_0} = \mu'|_{\mathcal{O}_0}$, entonces $\mu = \mu'$. Sugerencia: Probar que la familia \mathcal{O}' de conjuntos $A \in \mathcal{O}$ tales que $\mu(A) = \mu'(A)$ es a σ -álgebra que contiene \mathcal{O}_0 .

A.3. Integración

Sea (X, \mathcal{O}, μ) un espacio de medida. Dado un conjunto $A \in \mathcal{O}$, su función característica χ_A se define por $\chi_A(x) = 0$ si $x \notin A$ y $\chi_A(x) = 1$ si $x \in A$. Una función $f : X \rightarrow \mathbb{C}$ es *simple* si puede ser escrita como $f = \sum_{i=1}^n \lambda_i \chi_{A_i}$, donde A_1, \dots, A_n están en \mathcal{O} y $\mu(A_i) < +\infty$ cuando $\lambda_i \neq 0$. La integral de f se define por

$$\int_X f d\mu = \sum_{\lambda_i \neq 0} \lambda_i \mu(A_i)$$

si la serie converge absolutamente.

Decimos que $f : X \rightarrow \mathbb{C}$ es *integrable* si existe una sucesión de funciones simples $f_n : X \rightarrow \mathbb{C}$ tal que $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ for a.e. x y

$$\lim_{n, m \rightarrow \infty} \int_X |f_n - f_m| d\mu = 0 \quad (\text{I.3.1})$$

La *integral* de f es definida por

$$\int_X f d\mu = \lim_{n \rightarrow \infty} \int_X f_n d\mu \quad (\text{I.3.2})$$

La existencia del límite de la derecha se deduce de (I.3.1), porque ella implica

$$\limsup_{n, m \rightarrow \infty} \left| \int_X f_n d\mu - \int_X f_m d\mu \right| \leq \lim_{n, m \rightarrow \infty} \int_X |f_n - f_m| d\mu = 0$$

lo cual muestra que la sucesión en (I.3.2) es de Cauchy. Es más difícil pero necesario, probar que el límite en (I.3.2) es independiente de la sucesión $\{f_n\}$. esto está claramente garantido por la propiedad: si $g_n : X \rightarrow \mathbb{C}$, $n \geq 1$, es una sucesión de funciones simples tal que $\lim_{n \rightarrow \infty} g_n(x) = 0$ c.t., y $\lim_{n, m \rightarrow \infty} \int_X |g_n - g_m| d\mu = 0$ entonces $\lim_{n \rightarrow \infty} \int_X g_n d\mu = 0$. El propósito de esta sección es brindar un compendio de las propiedades básicas de la integración. Las pruebas de los resultados se pueden encontrar, por ejemplo, en [Rd66, Ha50] o [F96].

Las funciones integrables no tienen porque ser medibles, pero se deduce de la definición que cualquier función integrable f coincide con una función medible sobre un conjunto $X_0 \subset X$ con $\mu(X_0^c) = 0$. Por tanto, $f^{-1}(A) \in \mathcal{O} \pmod{0}$, si $A \subset \mathbb{C}$ es un conjunto de Borel. Es también claro que dadas dos funciones $f_i : X \rightarrow \mathbb{C}$, $i = 1, 2$, que coinciden c.t.p., f_1 es integrable si y sólo si lo es f_2 , y en este caso sus integrales son iguales. Dada $f : X \rightarrow \mathbb{C}$, decimos que es integrable sobre $A \subset X$ si $f \cdot \chi_A$ es integrable, y ponemos

$$\int_A f d\mu = \int_X f \cdot \chi_A d\mu.$$

Es fácil ver que si f es integrable, entonces lo es en cada $A \in \mathcal{O}$. Más aún, observe que f es integrable si y sólo si $|f|$ lo es.

Cuando se trata con aplicaciones, un frecuente problema importante es decidir si una función $f : X \rightarrow \mathbb{C}$, que es el límite c.t.p. de una sucesión de funciones integrables $f_n : X \rightarrow \mathbb{C}$, es integrable y si su integral es el límite de las integrales de f_n cuando $n \rightarrow \infty$. Obviamente, la respuesta afirmativa requiere hipótesis adicionales. Los siguientes tres teoremas son herramientas fundamentales para manejarse con este problema.

Teorema A.3.1 (Lema de Fatou). *Sea $f_n : X \rightarrow \mathbb{R}$ una sucesión de funciones integrables positivas tales que*

$$\liminf_{n \rightarrow \infty} \int_X f_n d\mu < +\infty$$

y que convergen c.t.p. $f : X \rightarrow \mathbb{R}$. Entonces f es integrable y

$$\liminf_{n \rightarrow \infty} \int_X f_n d\mu \geq \int_X f d\mu$$

Teorema A.3.2 (Teorema de la Convergencia Monótona). *Sea $f_n : X \rightarrow \mathbb{R}$ una sucesión de funciones integrables tal que para c.t. x la sucesión $\{f_n(x)\}$ es monótonamente creciente y*

$$\sup_n \int_X f_n d\mu < +\infty$$

Entonces la función $f(x) = \lim_{n \rightarrow \infty} f_n(x)$ es integrable y

$$\lim_{n \rightarrow \infty} \int_X f_n d\mu = \int_X f d\mu$$

Teorema A.3.3 (Teorema de la Convergencia Dominada). Sea $f_n : X \rightarrow \mathbb{C}$ una sucesión de funciones integrables dominada por una función integrable $f : X \rightarrow \mathbb{R}$, o sea $|f_n(x)| \leq f(x)$ para todo n y c.t. x . Entonces, si la sucesión $f_n(x)$, converge para c.t. x , la función límite $f(x) = \lim_{n \rightarrow \infty} f_n(x)$ satisface

$$\int f d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu$$

Dado $p \geq 1$, denote por $\mathcal{L}^p(X, \mathcal{O}, \mu)$ el espacio de las funciones $f : X \rightarrow \mathbb{C}$ tal que $|f|^p$ es integrable.

Teorema A.3.4. $\mathcal{L}^p(X, \mathcal{O}, \mu)$ es un espacio vectorial de funciones sobre \mathbb{C} y es un espacio de Banach con la norma

$$\|f\|_p = \left(\int |f|^p d\mu \right)^{1/p}$$

De hecho, $\|\cdot\|_p$ no es exactamente una norma porque $\|f\|_p = 0$ sólo implica $f(x) = 0$ para c.t. x . Este problema de la “norma” se resuelve considerando, cuando necesario, $\mathcal{L}^p(X, \mathcal{O}, \mu)$ como el espacio cociente de los espacios de funciones f tales que $|f|^p$ es integrable sobre el espacio de funciones que son cero c.t.p. Recordamos que si $\mu(X) < \infty$ y $0 < p < q < \infty$ entonces $\mathcal{L}^q \subset \mathcal{L}^p$ y $\|f\|_p \leq \|f\|_q (\mu(X))^{1/p-1/q}$ para toda $f \in \mathcal{L}^q$.

Una función $f : X \rightarrow \mathbb{C}$ es \mathcal{L}^∞ si es medible y existe $K \geq 0$ tal que $|f(x)| \leq K$ para c.t. x . $\|f\|_\infty$ denota el ínfimo de las constantes K satisfaciendo esa propiedad. Denote por $\mathcal{L}^\infty(X, \mathcal{O}, \mu)$ al conjunto de las funciones \mathcal{L}^∞ . Claramente es un espacio vectorial de funciones, y en su espacio cociente sobre el espacio de las funciones que son cero c.t.p., $\|\cdot\|_\infty$ es una norma de Banach.

Sea X un conjunto y \mathcal{O} una σ -álgebra de subconjuntos de X . Si $\mu : \mathcal{O} \rightarrow [0, +\infty]$ y $\nu : \mathcal{O} \rightarrow [0, +\infty]$ son medidas, decimos que μ es *absolutamente continua* con respecto ν , y escribimos $\mu \ll \nu$, si $\nu(A) = 0$ implica $\mu(A) = 0$.

Teorema A.3.5 (Radon-Nikodym). sea (X, \mathcal{O}, μ) un espacio de medida y $\nu : \mathcal{O} \rightarrow [0, +\infty)$ una medida satisfaciendo $\mu \ll \nu$. Si (X, \mathcal{O}, μ) es σ -finita, entonces existe una función ν -integrable $f : X \rightarrow \mathbb{R}^+$ tal que para cada $A \in \mathcal{O}$

$$\mu(A) = \int_A f d\nu.$$

Más aún, la función $g : X \rightarrow \mathbb{C}$ está en $\mathcal{L}^1(X, \mathcal{O}, \mu)$ si sólo si $gf \in \mathcal{L}^1(X, \mathcal{O}, \nu)$, y entonces

$$\int_X g d\mu = \int_X fg d\nu.$$

Está claro que la función f es esencialmente única: si $f_1 : X \rightarrow \mathbb{R}$ es otra función satisfaciendo los requerimientos del Teorema, entonces $f_1 = f$ c.t.p. La función f es llamada la *derivada de Radon-Nikodym* de μ con respecto a ν y denotada por $d\mu/d\nu$. Cuando $\mu \ll \nu$ y $\nu \ll \mu$, decimos que μ es *equivalente* a ν . En este caso $d\mu/d\nu = (d\nu/d\mu)^{-1}$. a.e.p.

EJERCICIO A.10. Si T preserva la medida μ y f es una función integrable, probar que $\int f \circ T d\mu = \int f d\mu$.

A.4. Ergodicidad de los Shifts y de Traslaciones y Transformaciones Lineales en el Toro

En el Ejemplo 3 de la Sección I.1 se introdujeron los shifts de Bernoulli.

Proposición A.4.1. *Los shifts de Bernoulli son ergódicos.*

Demostración. La prueba usa el siguiente hecho que debe ser verificado por el lector: Si $A_i \subset B(X)$, $i = 1, 2$, son cilindros, existe $m_0 > 0$ tal que para todo $m > m_0$

$$\mu(\sigma^{-m}(A_1) \cap A_2) = \mu(A_1)\mu(A_2) \tag{I.4.1}$$

Sea ahora $A \subset B(X)$ σ -invariante. Dado que el σ -álgebra of $B(X)$ es generada por los cilindros, existe, para cualquier $\varepsilon > 0$, una unión finita A_0 de cilindros disjuntos tal que

$$\mu(A_0 \Delta A) \leq \varepsilon.$$

se ve fácilmente que la propiedad (I.4.1) se satisface para uniones finitas de cilindros. Por tanto, para algún $m \geq 0$,

$$\begin{aligned}\mu(\sigma^{-m}(A_0^c) \cap A_0) &= \mu(A_0)\mu(A_0^c) \\ \mu(\sigma^{-m}(A_0) \cap A_0^c) &= \mu(A_0)\mu(A_0^c)\end{aligned}$$

Se deduce que

$$\begin{aligned}\mu(\sigma^{-m}(A_0) \Delta A_0) &\leq \mu(\sigma^{-m}(A) \Delta \sigma^{-m}(A_0)) + \mu(\sigma^{-m}(A) \Delta A) + \mu(A_0 \Delta A) \\ &= 2\mu(A \Delta A_0) \leq 2\varepsilon\end{aligned}\tag{I.4.2}$$

Por otro lado,

$$\begin{aligned}\mu(\sigma^{-m}(A_0) \Delta A_0) &= \mu(\sigma^{-m}(A_0) \cap A_0^c) + \mu(\sigma^{-m}(A_0^c) \cap A_0) \\ &= 2\mu(A_0)\mu(A_0^c) = 2\mu(A_0)(1 - \mu(A_0))\end{aligned}\tag{I.4.3}$$

De (I.4.2) y (I.4.3) obtenemos

$$2\mu(A_0)(1 - \mu(A_0)) \leq 2\varepsilon$$

y, dado que ε es arbitrario, $\mu(A)(1 - \mu(A)) = 0$ o sea, $\mu(A) = 0$ ó 1 . \square

En el Capítulo I (Sección I.1) introdujimos las traslaciones y transformaciones lineales en el toro $\mathbb{T}^n = S^1 \times \cdots \times S^1$. En el caso de las transformaciones lineales, los autovalores de T son los autovalores de \tilde{T} .

Sea λ la probabilidad de Lebesgue sobre \mathbb{T}^n (la cual, como ya se probó, es invariante tanto por las traslaciones como por las transformaciones lineales) y denote $\mathcal{L}^2(\mathbb{T}^n)$ espacio de las funciones a valores complejos \mathbb{T}^n que son \mathcal{L}^2 con respecto a λ , con la estructura habitual de espacio de Hilbert.

Teorema A.4.2. *Una transformación lineal $T : \mathbb{T}^n \rightarrow \mathbb{T}^n$ es ergódica si y sólo si ninguno de sus valores propios es raíz de la unidad.*

Teorema A.4.3. *Si $x \in \mathbb{R}^n$, la traslación $L_{\pi(x)} : \mathbb{T}^n \rightarrow \mathbb{T}^n$ es ergódica si y sólo si $\langle k, x \rangle \notin \mathbb{Z}$ para cualquier $k \in \mathbb{Z}^n$, $k \neq 0$.*

Las pruebas utilizan bases ortonormales de $\mathcal{L}^2(\mathbb{T}^n)$ obtenidas de la base de Fourier $\{e^{2\pi i \langle k, x \rangle} | k \in \mathbb{Z}^n\}$ de $\mathcal{L}^2([0, 1] \times \cdots \times [0, 1])$. La descripción formal de esta base y sus propiedades básicas son el motivo del próximo lema.

Lema A.4.4. *Existe una base ortonormal $\{\phi_k | k \in \mathbb{Z}^n\}$ de $\mathcal{L}^2(\mathbb{T}^n)$ tal que*

(a) $\phi_0 = 1$;

(b) *Para cada transformación lineal $T : \mathbb{T}^n \rightarrow \mathbb{T}^n$ se cumple (\tilde{T}^* es la adjunta de \tilde{T}):*

$$\phi_k \circ T = \phi_{\tilde{T}^*(k)}.$$

(c) *Para cada $x \in \mathbb{R}^n$ la traslación $L_{\pi(x)}$ satisface*

$$\phi_k \circ L_{\pi(x)} = e^{i \langle k, 2\pi x \rangle} \phi_k$$

Demostración. Para cada $\psi : \mathbb{R}^n \rightarrow \mathbb{C}$ satisfaciendo $\psi(x) = \psi(x + m)$ para todo $x \in \mathbb{R}^n$, $m \in \mathbb{Z}^n$, existe una única $\check{\psi} : \mathbb{T}^n \rightarrow \mathbb{C}$ tal que $\check{\psi} \circ \pi = \psi$. Sean $\psi_k := \mathbb{R}^n \rightarrow \mathbb{C}$ definida por $\psi_k(x) = e^{i \langle k, 2\pi x \rangle}$ y $\phi_k : \mathbb{T}^n \rightarrow \mathbb{C}$ por $\phi_k = \check{\psi}_k$. Entonces

$$\phi_k \circ T \circ \pi = \check{\psi}_k \circ T \circ \pi = \check{\psi}_k \circ \pi \circ \tilde{T} = \psi_k \circ \tilde{T} = \psi_{\tilde{T}^*(k)},$$

lo cual prueba la parte (b). Ahora, para $x \in \mathbb{R}^n$,

$$\phi_k \circ L_{\pi(x)} \circ \pi = \phi_k \circ \pi \circ L_x = \psi_k \circ L_x = e^{i \langle k, 2\pi x \rangle} \psi_k,$$

dando (c). Queda por probar que $\{\phi_k, k \in \mathbb{Z}^n\}$ es una base ortonormal de $\mathcal{L}^2(\mathbb{T}^n)$. Puesto que $\pi : [0, 1] \times \cdots \times [0, 1] \rightarrow \mathbb{T}^n$ es una transformación que preserva medida, tenemos $\langle \psi_1, \psi_2 \rangle = \langle \psi_1 \circ \pi, \psi_2 \circ \pi \rangle$ para todo $\psi_1, \psi_2 \in \mathcal{L}^2(\mathbb{T}^n)$; aquí los productos internos son tomados en $\mathcal{L}^2(\mathbb{T}^n)$ y $\mathcal{L}^2([0, 1] \times \cdots \times [0, 1])$, respectivamente. Entonces, para cada $\psi \in \mathcal{L}^2(\mathbb{T}^n)$ que satisface $\langle \psi, \phi_k \rangle = 0$ para todo $k \in \mathbb{Z}^n$, tenemos $\langle \psi \circ \pi, \phi_k \circ \pi \rangle = \langle \psi \circ \pi, \psi_k \rangle = 0$ para todo $k \in \mathbb{Z}^n$. Esto muestra que $\psi \circ \pi = 0$, porque $\{\psi_k, k \in \mathbb{Z}^n\}$ es la base ortonormal de Fourier $\mathcal{L}^2([0, 1] \times \cdots \times [0, 1])$. Puesto que π es sobreyectiva, obtenemos $\psi = 0$. \square

Demostración del Teorema A.4.2. Como T preserva medida tenemos

$$\langle \psi_1 \circ T, \psi_2 \circ T \rangle = \langle \psi_1, \psi_2 \rangle$$

para toda $\psi_1, \psi_2 \in \mathcal{L}^2(\mathbb{T}^n)$. Entonces, para $\phi \in \mathcal{L}^2(\mathbb{T}^n)$ tal que $\phi \circ T = \phi$, obtenemos

$$\langle \phi \circ T, \phi_{\tilde{T}^{*i}(k)} \rangle = \langle \phi \circ T, \psi_k \circ T \rangle = \langle \phi, \phi_k \rangle.$$

Entonces, para cada $k \in \mathbb{Z}^n$ y $i \geq 0$

$$\langle \phi \circ T, \phi_{\tilde{T}^{*i}(k)} \rangle = \langle \phi, \phi_k \rangle. \quad (\text{I.4.4})$$

Hay dos posibilidades para la sucesión $\{\tilde{T}^{*i}(k)\}$ – o bien todos sus elementos son distintos, o bien, para algún i , tenemos $\tilde{T}^{*i}(k) = k$. Si $k \neq 0$ el segundo caso significa que 1 es un autovalor de \tilde{T}^{*i} , o sea de \tilde{T}^i , y así T tiene una raíz de la unidad como autovalor. Por tanto, si ningún autovalor de T es una raíz de la unidad, sólo podemos tener el primer caso, y

$$\|\phi\|^2 = \sum_{k \in \mathbb{Z}^n} |\langle \phi, \phi_k \rangle|^2 \geq \sum_{i \in \mathbb{Z}} |\langle \phi, \phi_{\tilde{T}^{*i}(k)} \rangle|^2$$

En la última suma todos los términos son iguales, o sea cero porque la suma es finita. Por tanto, para todo $k \neq 0$ tenemos $\langle \phi, \phi_k \rangle = 0$, lo cual muestra que ϕ es c.t.p. igual a $\langle \phi, \phi_0 \rangle \phi_0$, una constante, y T es ergódica.

Inversamente, suponga que uno de los autovalores de T es raíz de la unidad. Entonces, para algún $i \geq 0$, tenemos $\det(\tilde{T}^{*i} - I) = 0$. Como los componentes de la matriz $\tilde{T}^{*i} - I$ son enteros, existe algún $0 \neq k \in \mathbb{Z}^n$ tal que $\tilde{T}^{*i}(k) = k$. Definiendo

$$\phi := \sum_{j=0}^{i-1} \phi_{\tilde{T}^{*j}}^{*j}(k) \implies \phi \circ T = \sum_{j=0}^{i-1} \phi_{\tilde{T}^{*j}(k)} \circ T = \sum_{j=0}^{i-1} \phi_{\tilde{T}^{*j+1}(k)} = \phi$$

lo cual muestra que T no es ergódica. □

Demostración del Teorema A.4.3. Sean $\phi \in \mathcal{L}^2(\mathbb{T}^n)$ y $\phi \circ L_{\pi(x)} = \phi$. Entonces

$$\langle \phi, \phi_k \rangle = \langle \phi \circ L_{\pi(x)}, \phi_k \rangle = \langle \phi, \phi_k \circ L_{\pi(-x)} \rangle = e^{-i\langle k, 2\pi x \rangle} \langle \phi, \phi_k \rangle.$$

Si $\langle k, x \rangle \notin \mathbb{Z}$ para todo $0 \neq k \in \mathbb{Z}^n$, tenemos $\langle \phi, \phi_k \rangle = 0$ para todo $0 \neq k \in \mathbb{Z}^n$, o sea que ϕ es una constante. Por otro lado, si existe $0 \neq k_0 \in \mathbb{Z}^n$ tal que $\langle k_0, x \rangle \in \mathbb{Z}$, entonces la función ϕ_{k_0} satisface $\phi_{k_0} \circ L_{\pi(x)} = \phi_{k_0}$. □

Bibliografía

- [Ab59] L. M. Abramov, *On the entropy of a flow* (in Russian), Dokl. Akad. Nauk SSSR **128**, 873–875 (1959).
- [An67] D. V. Anosov, *Geodesic flows on closed Riemannian manifolds with negative curvature*, Proc. Steklov Inst. Math. **90** (1967).
- [AS67] D.V. Anosov & Ya.G. Sinai, *Some smooth ergodic systems*, Russ. Math. Surveys **22** (1967), 103–167.
- [AA67] V. I. Arnold & A. Avez, *Problèmes ergodiques de la Mécanique Classique*. Gauthier-Villars, Paris (1967).
- [ACG96] R. Artuso, G. Casati & I. Guarneri, *Numerical Experiments on Billiards*, J. Statist. Phys. **83** (1996), 145–166.
- [Ba00] V. Baladi, *Positive transfer operators and decay of correlations*. World Scientific, River Edge, N. Y. (2000).
- [BCST02] P. Bálint, N. Chernov, N.; D. Szász & I. P. Tóth, *Multi-dimensional semi-dispersing billiards: Singularities and the Fundamental theorem*, Ann. Henri Poincaré **3** (2002), 451-482.
- [BCST03] P. Bálint, N. Chernov, D. Szász, & I. P. Tth, *Geometry of multi-dimensional dispersing billiards* Astrisque **286** (2003), 119-150.
- [BS78] G. Benettin, & J. M. Strelcyn, *Numerical experiments on the free motion of a point mass moving in a plane convex regime: Stochastic transition and entropy*, Physical Review A **17** (1978), 773–785.
- [Bi27a] G. D. Birkhoff, *Dynamical systems*, A.M.S. Colloquium Publications, New York (1927).
- [BK37] Bogoliouboff, Nicolas & Kryloff, Nicolas: *La thorie gnrale de la mesure dans son application l'étude des systmes dynamiques de la mcanique non linéaire*, Ann. of Math. (2) **38** (1937), 65–113
- [Bo71] L. Boltzmann, *Einige allgemene Sätze über das Wärmegleichgewicht*, Wien. Ber. **63** (1871) 679-711.
- [Bo70] R. Bowen, *Topological entropy and Axiom A*. In *Global Analysis (Proc. Symp. Pure Math., Vol XIV*, A.M.S., Providence, R.I. (1970), 23–41.
- [Bo75] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lect. Notes Math. **470**, Springer, Berlin (1975).
- [BG93] D. S. Broomhead & E. Gutkin, *The dynamics of billiards with no-slip collisions*, Phys. D **67** (1993), 188–197.
- [Bu74] L. A. Bunimovich, *On billiards close to dispersing*, Math. USSR. Sb. **23** (1974), 45–67.
- [Bu79] L. A. Bunimovich, *On ergodic properties of nowhere dispersing billiards*, Commun. Math. Phys. **65** (1979), 295–312.
- [Bu90] L. A. Bunimovich, *A Theorem on Ergodicity of Two-Dimensional Hyperbolic Billiards*, Commun. Math. Phys **130** (1990), 599–621.
- [Bu92] L. A. Bunimovich, *On absolutely focusing mirrors*, In *Ergodic Theory and related topics, III (Güstrow, 1990)*. Edited by U. Krengel et al., Lecture Notes in Math. **1514**, Springer, Berlin (1992), 62–82.
- [BR97] L. A. Bunimovich & J. Rehacek, *Nowhere dispersing 3D billiards with non-vanishing Lyapunov exponents*, Comm. Math. Phys. **189** (1997), 729–757.
- [BR98b] L. A. Bunimovich & J. Rehacek, *On the ergodicity of many-dimensional focusing billiards* Ann. Inst. H. Poincaré Phys. Théor. **68** (1998), 421–448.
- [BS73] L. A. Bunimovich, & Ya. G. Sinai, *On a fundamental theorem in the theory of dispersing billiards*, Math. USSR Sbornik **19** (1973), 407-423.
- [BS80] L. A. Bunimovich & Ya. G. Sinai, *Markov partitions for dispersed billiards*, Commun. Math. Phys. **73** (1980), 247–280.
- [BS81] L. A. Bunimovich & Ya. G. Sinai, *Statistical properties of Lorentz gas with periodic configuration of scatterers*, Commun. Math. Phys. **78** (1981), 479–497.
- [BSC90] L. A. Bunimovich, Ya. G. Sinai & N. I. Chernov, *Markov partitions for two-dimensional billiards*, Russ. Math. Surv. **45** (1990), 105–152.
- [BSC91] L. A. Bunimovich, Ya. G. Sinai & N. I. Chernov, *Statistical properties of two-dimensional hyperbolic billiards*, Russ. Math. Surv. **46** (1991), 47–106.
- [BFK98a] D. Burago, S. Ferleger & A. Kononenko, *A geometric approach to semi-dispersing billiards*, Ergodic Theory & Dynam. Systems **18** (1998), 303–319.

- [BFK98b] D. Burago, S. Ferleger & A. Kononenko, *Uniform estimates on the number of collisions in semi-dispersing billiards*, Ann. of Math. **147** (1998), 695–708.
- [CMOP] E. Canale, R. Markarian, S. K. Oliffson & S. Pinto, *A lower bound for chaos on the elliptical stadium*, Phys. D **115** (1998), 189–202.
- [C91a] N. I. Chernov, *A new proof of Sinai’s Formula of entropy of hyperbolic billiards. Its application to Lorentz gas ans stadium*, Funct. An. and its Appl. **25** (1991), 204–219.
- [C91b] N. I. Chernov, *Topological entropy and periodic points of two-dimensional hyperbolic billiards*, Functional Anal. Appl. **25** (1991), 39–45.
- [C93] N. I. Chernov, *On local Ergodicity in Hyperbolic Systems with Singularities*, Functional Anal. Appl. **27** (1993), 51–54.
- [C94] N.I. Chernov, *Statistical properties of the periodic Lorentz gas. Multidimensional case*, J. Stat. Phys. **74** (1994), 11–53.
- [C97] N. Chernov, *Entropy, Lyapunov exponents, and mean free path for billiards*, J. Statist. Phys. **88** (1997), 1–29.
- [C99a] N.I. Chernov, *Decay of correlations and dispersing billiards*, J. Statist. Phys. **94** (1999), 513–556.
- [CH96] N.I. Chernov & C. Haskell, *Nonuniformly hyperbolic K-systems are Bernoulli*, Ergodic Theory and Dynamical Systems, **16** (1996), 19–44.
- [CM92] N. I. Chernov & R. Markarian, *Entropy of non-uniformly hyperbolic plane billiards*, Bol. Soc. Bras. Mat. **23** (1992) 121–135.
- [CM97b] N. Chernov & R. Markarian, *Anosov maps with rectangular holes. Nonergodic cases*, Bol. Soc. Bras. Mat. **28** (1997), 315–342.
- [CM03] N. I. Chernov & R. Markarian, *Introduction to the ergodic theory of chaotic billiards*, Second edition - revised and enlarged, IMPA, Rio de Janeiro (2003).
- [CM05] N. I. Chernov & R. Markarian, *Theory of chaotic billiards*, in preparation (2005).
- [CMT00] N. Chernov, R. Markarian & S. Troubetzkoy *Invariant measurues for Anosov maps with small holes*, Ergod Th & Dynam. Sys **20** (2000), 1007–1044.
- [CT98] N. I. Chernov & S. Troubetzkoy, *Ergodicity of billiards in polygons with pockets*, Nonlinearity **11** (1998), 1095–1102.
- [CY00] N. I. Chernov & L.-S. Young, *Decay of correlations of Lorentz gases and hard balls*. In *Hard ball systems and Lorentz gas*, D. Szász, editor, Springer, Berlin (2000), 89–120.
- [CZ05] N. I. Chernov & H.-K. Zhang, *Billiards with polynomial mixing rates*, Nonlinearity (2005), –.
- [Cor35] G. Coriolis, *Théorie Mathématique des effets du jeu de billard*. Carilian - Gœury, Libraire-éditeur, Paris (1835).
- [CFS82] I. P. Cornfeld, S. V. Fomin & Ya. G. Sinai, *Ergodic Theory*. Springer, Berlin (1982).
- [DDL00] M. Degli Esposti, G. Del Magno & M. Lenci, *Escape orbits and ergodicity in infinite step billiards*, Nonlinearity **13** (2000), 1275–1292.
- [DelM01] G. Del Magno, *Ergodicity of a class of truncated elliptical biliard*, Nonlinearity **14** (2001), 1761–1786.
- [DMM03] G. Del Magno & R. Markarian, *Bernoulli Elliptical Stadia*, Commun. Math. Phys. **233** (2003), 211–230.
- [DMM05] G. Del Magno & R. Markarian, *On the Bernoulli property of planar hyperbolic billiards*. Sent for publication (2005)
- [Do91] V. Donnay, *Using Integrability to Produce chaos: Billiards with Positive Entropy*, Commun. Math. Phys. **141** (1991), 225–257.
- [DoL91] V. Donnay & C. Liverani, *Potentials on the two-torus for which the Hamiltonian flow is ergodic*, Comm. Math. Phys. **135** (1991), 267–302.
- [Dy82] R. Douady, *Applications du théorème des tores invariants*. These. Université Paris VII (1982).
- [F96] P. J. Fernandez, *Medida e integracao*. IMPA, Rio de Janeiro (1996).
- [Ga74] G. Gallavotti, *Lectures on the billiards*. In *Dynamical systems, theory and applications (Rencontres, Battelle Res. Inst., Seattle., Wash., 1974)*, Lecture Notes in Physics **38**, Springer, Berlin(1975), 236–295.
- [GO74] G. Gallavotti & D. S. Ornstein, *Billiards and Bernoulli Schemes*, Commun. Math. Phys. **38** (1974), 83–101.
- [GC95] G. Gallavotti & E. D. G. Cohen, *Dynamical ensembles in stationary states*, J. Stat. Phys. **80** (1995), 931–970.
- [Ga81] Galperin, G. A. , *On systems of locally interacting and repelling particles moving in space*, Trans. Moscow Math. Soc. **43** (1981), 159–214.
- [Gu86] E. Gutkin, *Billiards in polygons*, Physica D **19** (1986), 311–333.
- [Gu96] E. Gutkin, *Billiards in polygons: survey of recent results*, J. Statist. Phys. **83** (1996), 7–26.
- [Ha50] P. R. Halmos, *Measure Theory*. D. van Nostrand, New York (1950); Springer, New York (1974) .
- [Hn77] B. Halpern, *Strange billiard tables*, Trans. Am. Math. Soc. **232** (1977), 297–305.
- [HDM87] A. Hayli, Th. Dumont, J. Moulin-Ollagnier, & J. M. Strelcyn, *Quelques résultats nouveaux sur les billiards de Robnik*, J. Phys. A: Math. Gen. **20** (1987), 3237–3249.

- [HW83] M. Hénon & J. Wisdom, *The Benettin Strelcyn oval billiard revisited*, Physica D **8** (1983), 157–169.
- [He83] M. - R. Herman, *Sur les courbes invariantes par les difféomorphismes de l'anneau, Vol. I. Complété par un appendice au chapitre I de Albert Fathu*, Asterisque **103-104** (1983), 3–221.
- [Hu87] A. Hubacher, *Instability of the boundary on the billiard ball problem*, Commun. Math. Phys. **108** (1987), 483–488.
- [Ka80] A. Katok, *Lyapunov exponents, entropy and periodic orbits for diffeomorphisms*, Publ. Math. I.H.E.S. **51** (1980), 137–173.
- [KH95] A. Katok & B. Hasselblatt, *Introduction to the modern theory of dynamical systems*. Cambridge Univ. Press (1995).
- [KS86] A. Katok & J.-M. Strelcyn, with the collaboration of F. Ledrappier and F. Przytycki, *Invariant manifolds, entropy and billiards; smooth maps with singularities*, Lect. Notes Math., **1222**, Springer, New York (1986).
- [KMS86] S. Kerckhoff, H. Masur & J. Smillie, *Ergodicity of billiard flows and quadratic differentials*, Ann. of Math. **124** (1986), 293–311.
- [KMOP95] J. Koiller, R. Markarian, S. Oliffson & S. Pinto, *Time-dependent billiards*, Nonlinearity **8** (1995), 983–1003.
- [Ko54] A. N. Kolmogorov, *On the Conservation of Conditionally Periodic Motions under Small Perturbations of the Hamiltonian*, Dokl. Akad. Nauk SSSR, **98** (1954), 527–530
- [KT91] V. V. Kozlov & D. V. Treshchëv, *Billiards. A genetic introduction to the dynamics of systems with impacts*. Translations of Mathematical Monographs, **89** American Mathematical Society, Providence, RI (1991).
- [KSS90] A. Krámli, N. Simányi & D. Szász, *A “Transversal” Fundamental Theorem for Semi-Dispersing Billiards*, Commun. Math. Phys., **129** (1990), 535–560
- [KSS91] A. Krámli, N. Simányi & D. Szász, *The K-Property of Three Billiard Balls*, Annals of Mathematics, **133** (1991), 37–72
- [KSS92] A. Krámli, N. Simányi & D. Szász, *The K-Property of Four Billiard Balls*, Commun. Math. Phys. **144** (1992), 107–142.
- [Kr79] N. S. Krylov, *Works on the foundation of statistical physics*. Princeton U. Press, Princeton, N. J. (1979).
- [La73] V. F. Lazutkin, *On the existence of caustics for the billiard ball problem in a convex domain*, Math. USSR Izv. **7** (1973), 185–215.
- [Lb99a] J. Lebowitz, *Microscopic origins of irreversible macroscopic behavior*, Physica A **263** (1999), 516–527.
- [Lb99b] J. Lebowitz, *A century of statistichal mechanics: a selective review of two central issues*, Reviews of Modern Physics **71** (1999), 346–357.
- [Lew80] J. Lewowicz, *Lyapunov functions and topological stability*, J. Differ. Equations **38** (1980), 192–209.
- [LW95] C. Liverani & M. Wojtkowski, *Ergodicity in Hamiltonian Systems* Dynamics reported, Dynam. Report. Expositions Dynam. Systems (N.S.) **4**, Springer, Berlin (1995), 130–202. [Sunny Stony Brook Preprint # 1992/16]
- [LM96] A. Lopes & R. Markarian, *Open billiards: Cantor sets, invariant and conditionally invariant probabilities*, SIAM J. of Applied Math. **56** (1996), 651–680.
- [Lo05] H. A. Lorentz, *The motion of electrons in metallic bodies*, Proc. Amst. Acad. **7** (1905), 438–453.
- [Mc83] J. Machta, *Power Law Decay of Correlations in a Billiard Problem*, J. Statist. Phys. **32** (1983), 555–564.
- [Mn83] R. Mañe, *Introducao à Teoria Ergódica*. IMPA, Rio de Janeiro (1983). English Ed. *Ergodic Theory and Differentiable Dynamics*, Springer, Berlin, Heidelberg, New York (1987).
- [Ma88] R. Markarian, *Billiards with Pesin region of measure one*, Comm. Math. Phys. **118** (1988), 87–97.
- [Ma93a] R. Markarian, *New ergodic billiards: exact results*, Nonlinearity **6** (1993), 819–841.
- [Ma93b] R. Markarian, *The fundamental theorem of Sinai-Chernov for dynamical systems with singularities*. In *Dynamical Systems. Santiago de Chile 1990*, R. Bamón, R. Labarca, J. Lewowicz, and J. Palis, editors, Longman, Harlow (1993), 131–158.
- [Ma94] R. Markarian, *Non-uniformly hyperbolic billiards*, Ann. Fac. Sci. Toulouse Math. **3** (1994), 223–257.
- [Ma95] R. Markarian, *Statistical properties of dynamical systems with singularities*, J. Statist. Phys. **80** (1995), 1207–1239.
- [Ma02] R. Markarian, *Modelos con dinámicas caóticas*, Boletín Asoc. Mat. Ven. **9** (2002), 109 – 140.
- [Ma04] R. Markarian, *Billiards with Polynomial Decay of correlations*, Erg. Th. & Dynam. Syst. **24** (2004), 177 – 197.
- [MOP96] R. Markarian, S. Oliffson & S. Pinto, *Chaotic properties of the elliptical stadium*, Commun Math. Phys., **174** (1996), 661–679.
- [Mt82] J.N. Mather, *Glancing billiards*, Erg. Th. & Dynam. Syst. **2** (1982), 397–403.
- [Mo91] T. Morita, *The symbolic representation of billiards without boundary condition*, Trans. Amer. Math. Soc. **325** (1991), 819–828.
- [OP99] S. Oliffson & S. Pinto, *Bounded gain of energy on the breathing circle billiard*, Nonlinearity **12** (1999), 1363–1371.
- [Or70] D. S. Ornstein, *Bernoulli shifts with the same entropy are isomorphic*, Advances in Math. **4** (1970), 337–352.

- [OW98] D. S. Ornstein & B. Weiss, *On the Bernoulli nature of systems with some hyperbolic structure*, Erg. Th. & Dynam. Syst. **18** (1998), 441–456.
- [Os68] V. I. Oseledec, *A multiplicative ergodic theorem*, Trans. Moscow Math. Soc., **19** (1968), 197–231.
- [PM78] J. Palis & W. de Melo, *Introducao aos Sistemas Dinâmicos*. IMPA, Rio de Janeiro (1978). English Ed. *Geometric Theory of Dynamical Systems*, Springer, New York - Berlin (1982).
- [Pe76] Ya. B. Pesin, *Families of invariant manifolds corresponding to non zero Characteristic exponent* Math. USSR Izvestija **10** (1976), 1261–1305.
- [Pe77] Ya. B. Pesin, *Characteristic Lyapunov exponents and smooth ergodic theory*, Russ. Math. Surv. **32** (1977), 55–114.
- [PS88] V. M. Petkov & L. Stojanov, *On the number of periodic reflecting rays in generic domains*, Ergodic Theory & Dynamical Systems **8** (1988), 81–91.
- [Poin] H. Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste* (1892-1899). 3 Volumes. Dover Publications, Inc., New York, N.Y., (1957).
- [Re95] J. Reháček, *On the ergodicity of dispersing billiards*, Random Comput. Dynam. **3** (1995), 35–55.
- [Ro83] M. Robnik, *Classical dynamics of a family of billiards with analytic boundaries*, J. Phys. A; Math. Gen. **16** (1983), 3971–3986.
- [Rd66] W. Rudin, *Real and Complex Analysis*. Mc Graw-Hill, New York (1966).
- [Ru76] D. Ruelle, *A measure associated with Axiom-A attractors*, Amer. J. Math. **98** (1976), 619–654.
- [Ru78] D. Ruelle, *Thermodynamic Formalism*, Addison-Wesley, Reading, Mass. (1978).
- [Sa76] L. A. Santaló, *Integral Geometry and geometric probability*. Addison Wesley, Reading, Mass. (1976).
- [Sar02] O. M. Sarig, *Subexponential decay of correlations*, Invent. Math. **150** (2002), 629–653.
- [Sm99] N. Simányi, *Ergodicity of hard spheres in a box*, Ergodic Theory & Dynam. Systems **19** (1999), 741–766.
- [SS99] N. Simányi & D. Szász, *Hard ball systems are completely hyperbolic*, Ann. of Math. **149** (1999), 35–96.
- [Si63] Ya. G. Sinai, *On the Foundation of the Ergodic Hypothesis for a Dynamical System of Statistical Mechanics*, Dokl. Akad. Nauk SSSR, **153** (1963), 1261–1264.
- [Si70] Ya.G. Sinai, *Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards*, Russ. Math. Surv. **25** (1970), 137–189.
- [Si78a] Ya. G. Sinai, *Billiard trajectories in a polyhedral angle*, Russ. Math. Surveys **33** (1978), 229–230.
- [Si79] Ya.G. Sinai, *Development of Krylov's ideas*, Afterwards to N. S. Krylov, *Works on the foundations of statistical physics*. Princeton Univ. Press, Princeton, N. J. (1979), 239–281.
- [Si91a] Ya. G. Sinai, *Finite-Dimensional Randomness*, Russ. Math. Surveys **46** (1991), 177–190.
- [Si91b] Ya.G. Sinai, *Hyperbolic billiards*. In *Proceedings of the International Congress of Mathematicians. Kyoto 1990* Math. Soc. Japan, Tokyo (1991), 249–260.
- [SC82] Ya.G. Sinai & N.I. Chernov, *Entropy of a gas of hard spheres with respect to the group of space-time translations*, Trudy Sem. Petrovsk. **8**, (1982), 218–238. In Russian. In *Dynamical Systems*, Ed. by Ya. Sinai, Adv. Series in Nonlinear Dynamics, **1** [4]
- [SC87] Ya.G. Sinai & N.I. Chernov, *Ergodic properties of some systems of 2-dimensional discs and 3-dimensional spheres*, Russ. Math. Surv. **42** (1987), 181–207.
- [St89] L. Stojanov, *An Estimate from above of the number of periodic orbits for semi-dispersed billiards*, Commun. Math. Phys. **124** (1989), 217–227.
- [St99] L. Stoyanov, *Exponential instability for a class of dispersing billiards*, Ergodic Theory & Dynam. Systems **19** (1999), 201–226.
- [Sz96] D. Szász, *Boltzmann's ergodic hypothesis, a conjecture for centuries?*, Studia Sci. Math. Hung. **31** (1996), 299–322.
- [Sz00] D. Szász, *Hard ball systems and the Lorentz gas*. Edited by D. Szász. Springer, Berlin (2000).
- [Ta95] S. Tabachnikov, *Billiards*. Panor. Synth. No. 1, SMF, Paris (1995).
- [Va79] L. N. Vaserstein, *On Systems of particles with finite - range and/or repulsive interactions*, Commun. Math. Phys. **69** (1979), 31–56.
- [Vi97] M. Viana, *Stochastic dynamics of deterministic systems*. IMPA, Rio de Janeiro (1997).
- [VCG83] F. Vivaldi, G. Casati & I. Guarneri, *Origin of long-time tails in strongly chaotic systems*, Phys. Rev. Letters **51** (1983), 727–730.
- [Wa82] P. Walters, *An Introduction to Ergodic Theory*. Springer, Berlin, Heidelberg, New York (1982).
- [W85] M. Wojtkowski, *Invariant families of cones and Lyapunov exponents*, Ergod. The. & Dynam. Sys. **5** (1985), 145–161.
- [W86] M. Wojtkowski, *Principles for the design of billiards with nonvanishing Lyapunov exponents*, Commun. Math. Phys. **105** (1986), 391–414.
- [Y98] L.-S. Young, *Statistical properties of dynamical systems with some hyperbolicity*, Annals of Math. **147** (1998), 585–650.
- [Y99b] L.-S. Young, *Recurrence times and rates of mixing*, Israel J. of Mathematics **110** (1999), 153–188.