

# Billares. Modelos con dinámicas caóticas

Roberto Markarian

## Resumen

Se exponen en forma sencilla las relaciones del modelo de los gases de Boltzmann y su Hipótesis Ergódica con la moderna Teoría Ergódica. Se da una descripción de la formulación matemática de la Teoría de Billares y sus consecuencias en el modelo de las bolas duras.

## 1 Introducción

En los cursos elementales de física nos han enseñado que un gas en un recipiente cerrado puede concebirse como muchísimas bolas pequeñas (moléculas) que se mueven y chocan entre sí. Si no hay intercambio de energía con el exterior esta situación puede modelarse como un sistema conservativo (hamiltoniano). Contrariamente a lo que uno esperaría de esa descripción tan elemental, el diagrama de fase, que incluye las posiciones y velocidades de todas las bolas, dista mucho de ser simple. Tanto el análisis teórico como las simulaciones computacionales cada vez más afinadas, detectan que los sistemas hamiltonianos típicos manifiestan una extraordinaria riqueza dinámica donde coexisten hipersuperficies (toros) invariantes (comportamientos integrables a los que se refiere la llamada teoría KAM) y una o varias regiones caóticas (no integrables).

Ante esta complejidad el físico y el matemático deben ser modestos y contentarse con el estudio de modelos simples en que sólo uno de los comportamientos sucede: o bien el espacio de fase está completamente laminado por toros invariantes o bien todo el espacio es “fácilmente” divisible en una cantidad finita de conjuntos invariantes de medida no nula (componentes ergódicas).

El modelo más simple de sistema hamiltoniano en que diversas manifestaciones de la no integrabilidad son demostrables lo constituye el **sistema de las bolas duras** (hard ball systems): una colección de bolas de billar sin rotación que interactúan elásticamente entre sí. Este sistema y algunos relacionados con él, constituyen modelos auténticamente físicos en que varias leyes de la mecánica estadística pueden ser probadas rigurosamente.

Como se verá en la Sección 7 de este trabajo el sistema de las bolas duras es isomorfo a un sistema de billar con una sola partícula evolucionando en un espacio de dimensión muy grande. Ella se mueve libremente (movimiento uniforme, inercial) con choques elásticos (ángulo de entrada igual al de salida) contra obstáculos fijos. La Hipótesis Ergódica – formulada hace más de cien años por el físico alemán Ludwig Boltzmann – se relaciona directamente con estos modelos mecánicos para explicar las propiedades de los gases.

En la Sección 2 se dan algunas razones por las cuales es interesante el estudio de la dinámica de los billares. En la sección 3 serán analizados con más detalle los billares planos y se darán ejemplos sobre sus comportamientos. En las Secciones 4 a 7 se sigue la evolución cronológica que va de la formulación física a la formulación matemática de la Hipótesis de Boltzmann y se intenta mostrar como la formulación matemática de la Hipótesis se relaciona con el desarrollo de las áreas de la matemática conocidas en la actualidad por los nombres Teoría Ergódica, Transformaciones Continuas con Singularidades y, en particular, Teoría de Billares. En la Sección 4 se da la formulación física de la Hipótesis y en la 5 se presentan los elementos de su formulación matemática. En la Sección 7 se describen el modelo de los gases a través de los billares y otros modelos físicos relacionados con él. La dificultad en la presentación aumenta a partir de la Sección 7. En la última Sección resumimos y comentamos los principales resultados exactos recientes vinculados a la formulación matemática de la Hipótesis de Boltzmann.

## 2 La Belleza de los Billares

Comenzamos dando la definición de billares planos para así tener un objeto matemático al cual referirnos.

Todo fenómeno que evoluciona en el tiempo puede ser considerado como un **sistema dinámico**. Diremos que un tal sistema es determinístico si las reglas de evolución para el futuro (y/o para el pasado) son conocidas con precisión o sea, si conocido el estado del sistema en un momento inicial, se conoce el estado en cualquier tiempo futuro (y/o pasado). Un **billar plano** es el sistema dinámico que describe el movimiento de una partícula puntual en un conjunto compacto conexo  $Q \subset \mathbb{R}^2$  (o en el toro  $\mathbb{T}^2$ ), cuya frontera es la unión de un número finito de curvas regulares, por ejemplo  $C^3$ . Dentro de  $Q$  el movimiento es uniforme (velocidad constante) y la reflexión en la frontera  $\partial Q$  es elástica (ángulo de salida igual al de entrada).

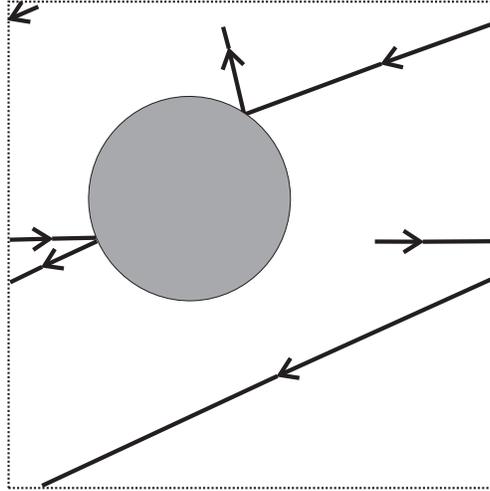


Figura 1: Billar en el toro.

En la Figura 1 está representado el movimiento de una partícula en  $\mathbb{T}^2$  con un obstáculo circular: cuando la partícula llega al borde del cuadrado, podemos hacerla rebotar con choque elástico (como si fuese un billar cuadrado con obstáculo circular) o, como está hecho en la figura, la hacemos reaparecer en el lado opuesto. Esta última representación corresponde al movimiento en el toro donde el único obstáculo es el círculo.

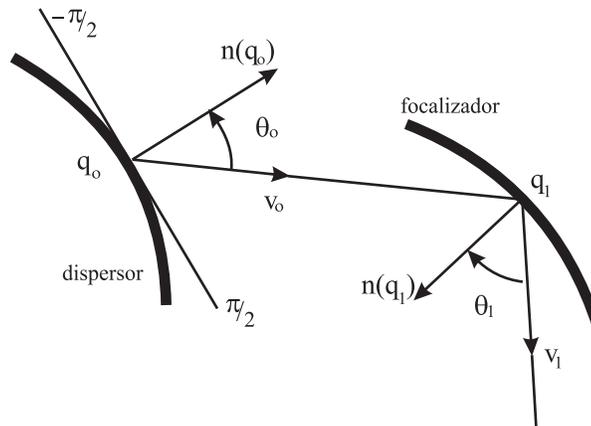


Figura 2: Espacio de configuración del billar. Si  $x_0 = (q_0, v_0)$ , entonces  $t(x_0) = \text{distancia}(q_0, q_1)$ .

Como el movimiento es uniforme dentro de  $\mathbf{Q}$ , el sistema del billar queda determinado por sus sucesivos choques con  $\partial\mathbf{Q}$  y, en lugar del flujo a tiempo continuo, podemos considerar la transformación que a cada punto de la frontera y a cada vector de salida les hace corresponder los del próximo choque (si estuviera bien definido). Así, a cada punto  $(q_0, v_0) \in \partial\mathbf{Q} \times \mathcal{S}_1$  corresponde un punto  $(q_1, v_1)$  en el mismo espacio producto (aquí,  $\mathcal{S}_1$  es media circunferencia  $[-\pi/2, \pi/2]$  de radio uno). Ver Figura 2.

Los arcos de frontera son llamados **dispersores** si son convexos vistos desde dentro de la mesa de billar (como la circunferencia de la figura 1), son llamados **focalizadores** si son cóncavos vistos desde dentro de la mesa (como una circunferencia vista desde dentro) y son llamados **neutros** si son segmentos de recta. Los llamamos dispersores y focalizadores porque un haz de trayectorias paralelas es “dispersado” o “focalizado”, respectivamente, al chocar con cada uno de ellos. Por favor tome un lápiz y papel y haga un dibujo con haces paralelos incidiendo sobre arcos de cada uno de esos tipos.

En general los billares son sistemas dinámicos que corresponden al movimiento uniforme de un punto material sobre una variedad riemanniana, con choque elástico en la frontera. Por tanto son flujos geodésicos en variedades con bordes. Las características específicas de los billares aparecen cuando el papel de la frontera (curvatura, posición relativa, etc.) es mucho más importante que el de la variedad subyacente.

En los últimos treinta años la Teoría de Billares se ha desarrollado principalmente debido a las siguientes razones:

1. Algunas clases de billares presentan un fuerte comportamiento caótico y pueden ser considerados entre los mejores ejemplos de caos determinístico;
2. Muchos ejemplos interesantes de sistemas dinámicos de origen físico (especialmente aquellos en que la interacción entre partículas envuelve choques elásticos) pueden ser reducidos a billares. Algunos de estos ejemplos serán estudiados a partir de la Sección 4;
3. Importantes problemas en la teoría de caos cuántico involucran un análisis profundo de los billares clásicos;
4. El estudio de los billares sugiere muchos problemas bonitos e interesantes en geometría y probabilidad.

A continuación nos referiremos a algunos de estos problemas. Presentaremos los resultados para billares planos, aunque son válidos, con las debidas adaptaciones, para billares en cualquier dimensión.

Consideremos dos arcos de curvas dispersoras que se cruzan en un punto  $V$ , formando un vértice de la mesa de billar. Existen trayectorias que van directamente al vértice. Cualquier trayectoria que entra en el “ángulo” del vértice  $V$ , y no va directo hacia él, comenzará a salir en algún momento: el número de choques cerca del vértice es finito. Más aún, si las curvas dispersoras no son tangentes, el número de choques cerca del vértice tiene una cota superior finita. O sea, dado un número pequeño  $\varepsilon$ , existe un número  $C$  (que depende del ángulo entre ambas curvas, de sus curvaturas y de  $\varepsilon$ ) tal que toda trayectoria que pasa a una distancia menor que  $\varepsilon$  del vértice, acabará saliendo después de a lo máximo  $C$  choques. En la Sección 7 daremos una fórmula explícita para un caso particular de esta situación. Si las curvas son tangentes, en general existen trayectorias que se mantienen próximas al vértice por la cantidad de choques que se desee. En este caso el número de choques es finito pero no hay una cota superior  $C$ .

Es también interesante calcular, para cualquier billar, la distancia media de las longitudes  $t$  de los segmentos de trayectoria entre dos choques. O sea, tomamos todos los posibles puntos de partida y, con algún criterio estadístico razonable, efectuamos la media de las longitudes hasta el próximo choque. El resultado de esta **trayectoria libre media** (mean free pass), que será definida rigurosamente más adelante, es

$$\tau_m = \pi \frac{\text{área de la mesa de billar}}{\text{longitud del borde del billar}} .$$

### 3 Billares planos

Fijemos un origen en la frontera a partir del cual mediremos su longitud de arco; sean  $\partial\mathbf{Q}_i$  las componentes regulares de la frontera (que se cortan en los “vértices”, formando los “ángulos” del billar) y  $n(q)$  el versor normal interior a  $\partial\mathbf{Q}_i$  en  $q \in \partial\mathbf{Q}_i$ . Entonces, el **espacio de fase** donde está definido el sistema dinámico puede ser parametrizado por la longitud de arco  $s$  y por el ángulo  $\theta$  entre el versor  $v$  de salida y  $n(q)$ . Para simplificar la comprensión, supongamos inicialmente que la mesa no tiene obstáculos en su interior y que la longitud total de  $\partial\mathbf{Q}$  sea  $L$ . Entonces el espacio de fase estará contenido en un cilindro  $\mathcal{R}$  que se obtiene identificando de manera natural las posiciones  $s = 0$  con  $s = L$ . Es en este espacio que representaremos el movimiento de una partícula material chocando contra los bordes de la **mesa de billar**. Es claro que lo que se representa son los choques y que los **trozos de trayectorias** (segmentos entre un choque y otro) sólo serán visibles en el **espacio de configuraciones**, esto es, en la propia mesa del billar.

Sea  $S$  la **transformación del billar** tal que  $S(s_0, \theta_0) = (s_1, \theta_1)$ , donde  $s_0$  y  $s_1$  son las coordenadas de los puntos  $q_0$  y  $q_1$  de salida y llegada (respectivamente) en la frontera y  $\theta_0, \theta_1$  son los ángulos de salida de las trayectorias en  $q_0$  y  $q_1$  respectivamente. La función  $S$  (o su inversa) no está bien definida si  $q_1$  o  $q_0$  está en un vértice de la frontera y es discontinua en los puntos en que la trayectoria es tangente al borde. Obsérvese que la derivada de  $S$  no está acotada en estos puntos de tangencia.

En efecto, si  $\tilde{x}_1 = (\tilde{q}_1, \tilde{v}_1) = T(\tilde{x}_0)$  está definida para  $\tilde{x}_0 = (\tilde{q}_0, \tilde{v}_0)$  entonces para todo  $x_0 = (q_0, v_0)$  en un pequeño entorno de  $\tilde{x}_0$  la matriz derivada en las coordenadas  $(s, \theta)$  está dada por (ver [9], [2]):

$$DT(x_0) = - \begin{pmatrix} \frac{t_0 K_0 + \cos \theta_0}{\cos \theta_1} & \frac{t_0}{\cos \theta_1} \\ K_1 \frac{t_0 K_0 + \cos \theta_0}{\cos \theta_1} + K_0 & \frac{K_1 t_0}{\cos \theta_1} + 1 \end{pmatrix}, \quad (1)$$

donde  $K_i = K(x_i)$ ,  $i \in \mathbf{N}$  son las curvaturas de  $\partial\mathbf{Q}$  en  $q_i$  y  $t_0$  es la distancia entre  $q_0$  y  $q_1$ , o sea la longitud del trozo de trayectoria entre  $x_0$  y  $x_1$ . En esa fórmula la curvatura de los arcos dispersores es positiva y la de los arcos focalizadores negativa. Obsérvese que los coeficientes de esta matriz tienden a infinito cuando la imagen de  $(q_0, v_0)$  por  $S$  tiende a ser tangente a  $\partial\mathbf{Q}$  (o sea, cuando  $\theta_1$  se aproxima a  $\pm\pi/2$ ).

Los segmentos verticales determinados por los vértices de la frontera, los puntos de tangencia y las imágenes y preimágenes por  $S$  de todos estos segmentos se denominan **singularidades** de  $S$  y las representamos por  $D$ .

Con esas restricciones la transformación  $S$  es diferenciable, con inversa diferenciable (difeomorfismo  $C^2$ ) en el conjunto  $M$  que se obtiene retirando  $D$  de  $\mathcal{R}$ .

Además, esa transformación preserva la medida  $d\mu = c \cos \theta ds d\theta$  ( $c = 1/(2L)$  es una constante de normalización). Esto significa que si  $\mu(A) = \int_A c \cos \theta ds d\theta$ , entonces  $\mu(A) = \mu(S^{-1}A)$  para todo conjunto boreliano  $A$ : la  $\mu$ -medida de los conjuntos borelianos del espacio de fase no varía cuando los “movemos” por  $S$ . Es con respecto a esta medida estándar que hacemos todas las medias y estudios probabilísticos de los billares. En particular, la trayectoria libre media, cuya definición fue dada al final de la Sección anterior, es calculada de la siguiente manera:  $\tau_m = \int_M t(x) d\mu(x)$  donde  $t(x)$  es la longitud de la trayectoria entre los puntos  $x = (s, \theta)$  y  $Sx$  (esto es, entre un choque y el siguiente).

Todas esas características de la transformación de billar en regiones acotadas, conexas, cuyos bordes están formados por un número finito de curvas  $C^3$  en el plano, son válidas también – con las debidas adaptaciones – para los billares en  $\mathbf{R}^d$  o en  $\mathbf{T}^d$ , el toro de dimensión  $d$  para cualquier  $d \geq 2$ .

Veamos algunos ejemplos de billares planos.

**Círculos.** Consideremos en primer lugar la mesa de billar cuyo borde es una circunferencia. Se sabe que para una partícula que sale con ángulo  $\theta_0$ , todos los choques ocurrirán con ese mismo ángulo y por tanto la representación de los puntos de esa trayectoria en el espacio de fase estará contenida en una recta horizontal. Los segmentos de trayectoria entre un choque y otro tendrán todos la misma longitud, y serán tangentes a una misma circunferencia. El lector interesado puede proponerse como ejercicio

descubrir cuáles trayectorias son periódicas y cuándo una trayectoria cubre densamente un segmento de recta horizontal.

Veamos estos y otros problemas con un enfoque más formal. Sea  $\mathbf{Q}$  el disco de radio uno. La superficie  $M = \partial\mathbf{Q} \times [-\pi/2, \pi/2]$  (con coordenadas  $s, \theta$ ) es un cilindro cuya base es una circunferencia de longitud  $2\pi$  y altura  $\pi$ .

Para ver cómo actúa la transformación  $S : M \rightarrow M$ , sea  $x = (s, \theta) \in M$ . La siguiente reflexión ocurrirá en el punto  $Sx = (s + \pi - 2\theta, \theta)$ , y la  $n$ -ésima reflexión  $S^n x = (s + n(\pi - 2\theta), \theta)$ . En esta fórmula se supone que  $s$  es una coordenada cíclica  $s$ , i.e.  $s + n(\pi - 2\theta)$  se toma módulo  $2\pi$ . Dos conclusiones surgen de inmediato.

(i) El ángulo de reflexión  $\theta$  es mantenido por  $S$ , i.e. toda curva  $\theta = \text{constante}$  en  $M$  es invariante por  $S$ . Cada una de esas curvas es un círculo paralelo a la base del cilindro  $M$ .

(ii) La transformación  $S$  restringida a cualquier curva  $\theta = \text{constante}$  la hace rotar por un ángulo constante,  $\pi - 2\theta$ . Por tanto,  $S$  actúa como una rotación en el círculo sobre cada curva invariante.

También observamos que la distancia entre dos choques  $t(x) = 2 \cos \theta$  es constante sobre cada círculo  $\theta = \text{constante}$ . Todos los trozos entre dos choques de una trayectoria que comienza con ángulo dado  $\theta$  son tangentes a un círculo de radio  $R = \sin \theta$  concéntrico con la mesa de billar. Ver Figura 3.

Si el ángulo de rotación  $\pi - 2\theta$  es un *múltiplo racional* de  $\pi$ , i.e. si  $\frac{\pi - 2\theta}{2\pi} = \frac{m}{n}$  con  $m, n \in \mathbb{Z}$ , entonces la transformación  $S$  sobre el círculo  $\theta = \text{constante}$  es periódica con período  $n$ . Más aún, cada punto sobre este círculo es periódico con el mismo período.

Si el ángulo de rotación  $\pi - 2\theta$  es *irracional*, i.e. si  $\theta/\pi$  es un número irracional, entonces la rotación en el círculo es ergódica y cada trayectoria es densa y uniformemente distribuida en el círculo. En ese caso los trozos de cualquier trayectoria en el espacio de configuraciones llenan densamente el anillo de la mesa de billar con radio interior  $R = \sin \theta$  (ver Figura 3).

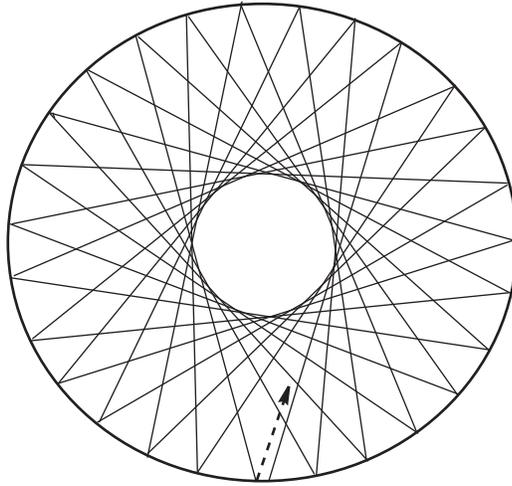


Figura 3: Trayectoria en el billar circular.

Se ve claramente en la Figura 3 que los trozos de trayectoria se ven más densos cerca del círculo interior. Si nuestra trayectoria de billar fueran pasajes de un rayo laser y el borde de la mesa de billar  $Q$  fuera un espejo perfecto, habría “mucho calor” allí, cerca del círculo interior. Por esta razón el círculo interior es llamado una **cáustica** (lo cual significa “que se quema”). En general, una cáustica para un billar es una curva tal que si un trozo de trayectoria es tangente a ella, entonces todo otro trozo de la misma trayectoria es tangente a la cáustica.

Dado que la superficie  $M$  está laminada por curvas invariantes  $\theta = \text{const}$ , la transformación  $S$  tiene

conjuntos invariantes de medida no total (se dice que no es ergódica) porque cualquier conjunto que es unión de curvas  $S$ -invariantes es  $S$ -invariante. Otra manera de ver la no ergodicidad es encontrar una función invariante no constante. En este caso, la función  $F : M \rightarrow \mathbb{R}$  definida por  $F(r, \theta) = \theta$  es una función suave no constante invariante por  $S$ , i.e. satisface  $F(Sx) = F(x)$  para todo  $x \in M$ .

**Definición.** Si un sistema dinámico suave  $S : M \rightarrow M$  sobre una variedad  $M$  admite una función suave no constante invariante por  $S$ , entonces  $F$  es llamada una **integral primera** y se dice que  $S$  es **integrable**.

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

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

El billar en el círculo es completamente integrable. Su derivada y la de su potencia  $n$ -ésima para todo  $n \in \mathbb{Z}$ , son

$$DS(x) = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \quad DS^n(x) = \begin{pmatrix} 1 & -2n \\ 0 & 1 \end{pmatrix} \quad (2)$$

Esto sale de (1), con  $\theta_1 = \theta$ ,  $\tau = 2 \cos \theta$  y  $K = -1$ . En ningún sentido se puede decir que el billar en el círculo es caótico. En particular (ver Sección 8) las fórmulas anteriores permiten probar que ambos exponentes de Liapunov son cero en cualquier punto  $x \in M$  y, como consecuencia, que la entropía de  $S$  es nula.

Obsérvese, sin embargo, que vectores tangentes típicos crecen bajo la acción de  $DS^n$  cuando  $n \rightarrow \infty$ , pero sólo linealmente en  $n$  (y no exponencialmente). Estas transformaciones, donde la separación de trayectorias cercanas son típicamente lineales, se dice que exhiben comportamiento **parabólico**, como contrapuesto al hiperbólico que será expuesto en la Sección 6.

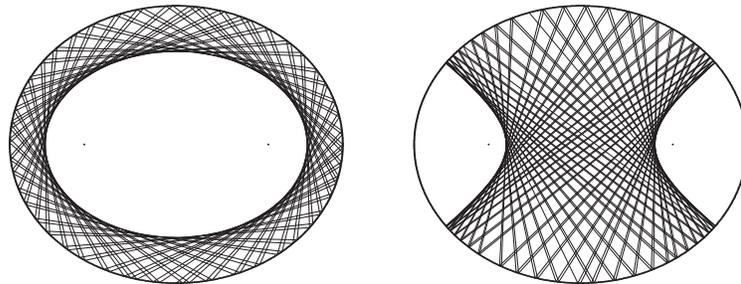


Figura 4: Cáusticas elípticas e hiperbólicas

**Elipses.** Si la mesa de billar tiene como borde una elipse de semieje vertical con longitud uno y semieje horizontal con longitud  $a > 1$ , la situación se complica un poco y habrá esencialmente dos tipos de trayectorias.

Si un trozo de trayectoria corta al eje menor entre un foco y el vértice más próximo, esta propiedad continuará verificándose para los otros trozos de trayectorias que cortan ese eje. Además, todos los trozos de trayectoria serán tangentes a una elipse interior con los mismos focos que la frontera. De acuerdo con la definición dada anteriormente esta elipse confocal es una cáustica del billar. Ver Figura 4. Obsérvese que si  $a = 1$  la elipse es una circunferencia y cada una de esas cáusticas también.

Si un trozo de trayectoria corta al eje menor entre los dos focos, todos los otros trozos también lo harán, y las rectas que contienen esos trozos son tangentes (en puntos dentro o fuera de la mesa elíptica) a una cáustica hiperbólica. Este tipo de cáusticas no se dan en la circunferencia.

Una pregunta surge naturalmente: ¿qué sucede con las trayectorias que pasan por los focos? La respuesta es aún más simple que en los casos anteriores: después de chocar con la elipse la partícula se dirigirá al otro foco y así sucesivamente. Estas trayectorias que pasan por los focos “separan” los comportamientos anteriores (con cáusticas elípticas e hiperbólicas) y son muy importantes en el estudio de la dinámica de este billar y de otros relacionados con él. Todos los resultados sobre cáusticas confocales y trayectorias que pasan por focos son consecuencias del llamado Teorema de Poncelet de la Geometría Proyectiva.

Hay dos órbitas periódicas distinguidas, una sobre el eje mayor y otra sobre el eje menor de la elipse.

El espacio de fase es el cilindro  $M$  con base la elipse y altura  $\pi$ . Si representamos algunas trayectorias de cada tipo en el espacio de fase veremos que las más parecidas con las de la circunferencia son las del primer tipo (cáusticas elípticas); en realidad éstas aparecen como deformaciones de las rectas invariantes  $\theta = \text{const}$  del billar circular. Las trayectorias del segundo tipo están contenidas en curvas cerradas alrededor de dos puntos del espacio de fase que corresponden a las trayectorias periódicas que se mueven a lo largo del eje menor (vertical). Estas trayectorias llenan las dos regiones en  $M$  limitadas por la curva con la forma del símbolo  $\infty$  en la Figura 5. Las trayectorias que pasan por los dos focos están representadas sobre dos curvas (“separatrices” de los dos movimientos principales) que unen los dos puntos periódicos de las trayectorias que están sobre el eje mayor. Estas curvas cerradas **separatrices** en el cilindro  $M$  son las que forman el símbolo  $\infty$  en Figura 5.

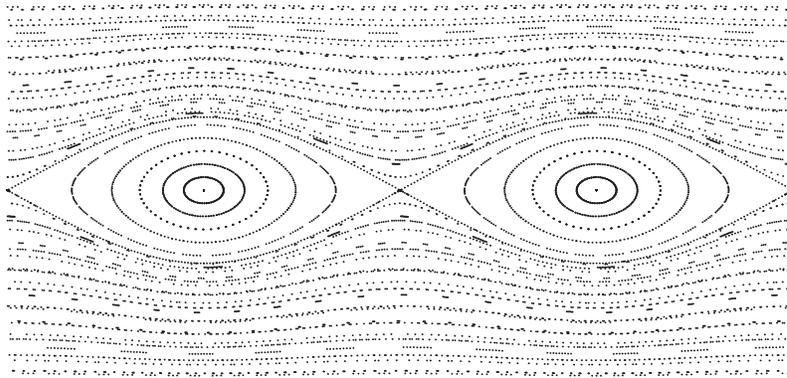


Figura 5: Espacio de fase  $M$  del billar elíptico

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.

Sobre cada curva invariante la transformación  $S$  (o  $S^2$ ) es conjugada a una rotación del círculo con algún ángulo (llamado el **número de rotación**). Este número cambia continua y monótonamente con la curva invariante. La curva especial con forma de símbolo  $\infty$  que separa las cáusticas hiperbólicas y elípticas, la separatriz, es invariante por  $S$ , pero no rota. Por el contrario, toda trayectoria comenzando en la separatriz converge (tanto en el futuro como en el pasado) a la órbita 2-periódica que está sobre el eje mayor.

La fórmula para la derivada de  $S$  en las coordenadas  $s, \theta$  es complicada y no nos dará mucha información a nuestros actuales efectos. Pero se pueden elegir unas coordenadas curvilíneas especiales en  $M$  (no válidas sobre la separatriz) con un eje a lo largo de las curvas invariantes, en que la matriz  $DS$  es triangular y tiene unos en la diagonal. Y podremos sacar las mismas conclusiones que sacamos de la fórmula (2): los exponentes de Liapunov y las entropías son nulas; para cualquier estándar la

transformación  $S$  no es caótica.

El comportamiento dinámico de la transformación  $S$ , dejando de lado la separatriz y las órbitas periódicas distinguidas, es parabólico, como en el caso del billar circular. Pero ahora tenemos otras clases de fenómenos. La trayectoria periódica sobre el eje mayor atrae o repele a los puntos de la separatriz (se dice que la variedad estable o inestable de los puntos periódicos están sobre la separatriz, y que estos puntos son hiperbólicos (Sección 6)). La otra órbita periódica distinguida, sobre el eje menor no es ni hiperbólica ni parabólica, es elíptica.

**Definición.** Sea  $x$  un punto periódico, i.e.  $S^n x = x$  para algún  $n \geq 1$ . Si la derivada  $DS^n(x)$  tiene un valor propio complejo (no real) en el círculo unidad, entonces se dice que el punto  $x$  es **elíptico**.

En este caso, cuando  $\dim M = 2$ , entonces la transformación derivada  $DS^n$  es una rotación con algún ángulo. Un entorno de  $x$  contiene frecuentemente muchas curvas cerradas  $S^n$ -invariantes (son curvas como elipses, de allí viene el nombre de *elíptico*). Estos hechos están relacionados con la teoría KAM mencionada anteriormente. Nuestra trayectoria periódica sobre el eje menor es elíptica y hay un gran entorno completamente foliado por óvalos invariantes por  $S^2$ . Por tanto, el billar dentro de una elipse tiene una combinación de estructuras parabólicas, hiperbólicas y elípticas.

**Estadios.** En el billar elíptico el espacio de fase no tiene singularidades “visibles”; no hay vértices del billar porque la frontera es una única curva regular, y las trayectorias tangentes no existen dentro del billar, porque están restringidas a los puntos  $\theta = \pm\pi/2$  del espacio de fase. Esto quiere decir que el conjunto  $D$  de las singularidades está formado por apenas dos segmentos y que el espacio  $M$  es un cilindro con altura  $\pi$  y sin los bordes superior e inferior.

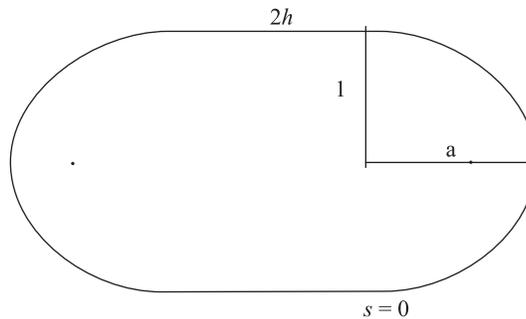


Figura 6: El estadio elíptico

Ahora modifiquemos un poco la situación cortando la elipse a lo largo del eje menor, separando estas dos mitades y pegando las puntas libres con dos segmentos de longitud  $2h$ . Tendremos el *estadio elíptico* de la Figura 6. Si  $a = 1$  las curvas de los extremos son semicircunferencias y el estadio se llama de Bunimovich, por ser éste el nombre del primer matemático que lo estudió en detalle.

Nuevamente, las trayectorias tangentes no aparecen dentro del espacio de fase del billar. Pero ahora tenemos cuatro vértices en los puntos donde los extremos de las semielipses se tocan con los segmentos de recta. En esos puntos la frontera tiene derivada primera continua, pero su curvatura (que está directamente relacionada con la segunda derivada) es discontinua. Entonces, si la longitud de media elipse es  $m$ , la longitud total de la frontera es  $L = 2m + 4h$ , y si comenzamos a medir la longitud de arco a partir de uno de los (antiguos) vértices del semieje menor de la elipse, el espacio de fase será como en la Figura 7. En ella hemos representado también el conjunto de los puntos cuyas imágenes por  $S$  están en los puntos con  $s = 0$  (recuérdese que el ángulo  $\theta$  es medido a partir del vector de salida hacia la normal interior). El aspecto de cada curva varía de acuerdo con el valor  $a$  (longitud del semieje mayor). Aquí representamos el caso en que  $a$  es muy próximo a 1. El lector interesado puede intentar ver como esa curva se modifica cuando  $a$  aumenta.

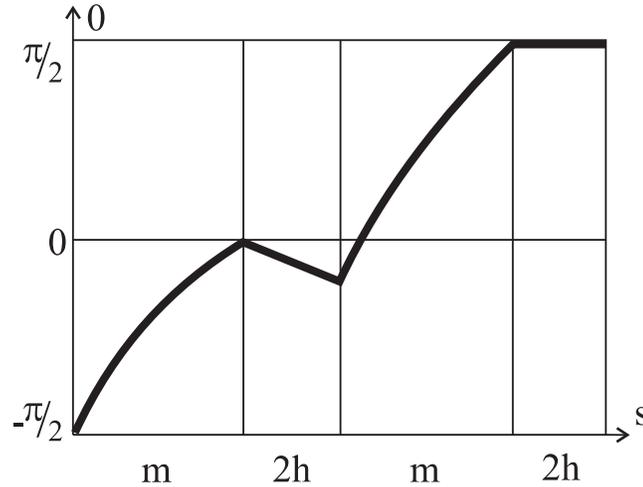


Figura 7: Espacio de fase del estadio elíptico. Se indican las curvas de los puntos cuyas imágenes están en el vértice inferior del eje menor de la semielipse de la derecha.

Esa curva y las que son preimágenes o imágenes por  $S$  de los “vértices”  $s = m, m + 2h, 2m + 2h$  son obtenidas por diversas simetrías a partir de la dibujada. Ellas forman parte del conjunto  $D$  de las singularidades. Entonces el conjunto  $M$ , que originalmente parecía ser un simple cilindro, quedó dividido en varias partes cuyos bordes son esas curvas. El estudio de cómo ellas se cortan y se distribuyen en el cilindro original es fundamental para entender el comportamiento estadístico de la dinámica de los billares.

En el caso del estadio elíptico las trayectorias son mucho más complicadas que en el caso de la elipse. Hay infinitas trayectorias con período dos entre los lados paralelos; hay trayectorias que pasan de una media elipse a la otra, cambiando el sentido de su circulación. El comportamiento general depende mucho de los valores de  $a$  y  $h$ . Esto será visto con un poco más de detalle en la sección 7. Por ejemplo, la representación de una simulación de 150.000 iteraciones de una única trayectoria del billar con  $a = 1.24, h = 1.04$  parece distribuirse casi uniformemente en el espacio de fase. En el trabajo [11] se estudian muchas propiedades del estadio elíptico.

## 4 La Hipótesis de Boltzmann

En 1964 Werner Heisenberg afirmó que “Un físico teórico se siente mejor si no hay objetos matemáticos rigurosamente definidos por detrás de sus consideraciones.”. Seguramente en ese momento tenía en mente los primeros años de la Mecánica Cuántica, pero esa frase puede igualmente ser aplicada a la obra de Ludwig Boltzmann. Y no sólo a su Hipótesis Ergódica, de la que trata la parte sustancial de lo que resta de este artículo, sino a otros temas estudiados por él.

Corresponde también recordar que, motivado por las ecuaciones de Boltzmann, David Hilbert incluyó en su célebre colección de 23 problemas (presentada en el Congreso Internacional de Matemática que tuvo lugar en París, en 1900) el denominado “Tratamiento Matemático de los Axiomas de la Física”. Sobre esto escribió: “es muy deseable que la discusión de los fundamentos de la Mecánica sea tomada también por los matemáticos. Así el trabajo de Boltzmann sobre los principios de la Mecánica sugiere el problema del desarrollo matemático de los procesos límite allí meramente indicados, que llevan de la visión atomística a las leyes del movimiento continuo”.

Entre 1870 y 1884 Boltzmann usó varias formas de la Hipótesis Ergódica.<sup>1</sup> Una formulación avanzada sería:

**Hipótesis Ergódica de Boltzmann.** *Para grandes sistemas de partículas interactuando en equilibrio, las medias temporales están próximas de las medias espaciales.*

Creo que viene al caso hacer algunos comentarios sobre esa formulación. Por “grandes sistemas de partículas interactuando en equilibrio” se entiende un sistema con muchas partículas que no reciben influencias externas; por ejemplo, un número muy grande de partículas que chocan con las paredes completamente rígidas de una caja y entre si, sin incidencia de otros factores. Las “medias temporales” son los promedios de los valores observados (mediciones) de una función numérica, al pasar el tiempo; éste se mide en alguna unidad que corresponda a la escala del problema. Si el modelo matemático del fenómeno es descrito por una ecuación diferencial, sus soluciones estarán dadas por un flujo y la evolución será a tiempo continuo; pero si el modelo está dado por una transformación (un mapa, una función) la evolución del tiempo no es continua, es discreta. En este artículo seguiremos trabajando con este último tipo de modelo porque el tiempo será medido por el número de veces que se aplica la transformación estudiada.  $S^n$  indicará la aplicación sucesiva de la transformación  $S$ ,  $n$  veces; por tanto aquí  $n$  es el “tiempo”. Entonces sumamos los valores medidos al transcurrir un largo tiempo  $T$ , y dividimos por  $T$  (que es un número entero:  $T \in \mathbb{Z}$ ). Las “medias espaciales” consisten de medias (integrales) de mediciones (o registros) simultáneos en todos los puntos con respecto a una medida con sentido físico en el espacio de fase. Las funciones medidas pueden ser, por ejemplo, la temperatura o la presión.

Observación: los físicos entienden las medias espaciales como promedio de equilibrio (microcanónico), o sea, con respecto a la medida de Liouville en la subvariedad del espacio de fase determinada por los invariantes triviales del movimiento. A esta medida nos referimos en el comentario anterior.

Introducidos esos conceptos matemáticos, podemos hacer una presentación más precisa de la Hipótesis Ergódica. Ella establece que si  $f$  es una medición – una función en el espacio de fase del sistema – y el tamaño del sistema – el número  $N$  de partículas – tiende a infinito, entonces

$$\frac{1}{T} \sum_{n=0}^{T-1} f(S^n x) \rightarrow \int_M f(x) d\mu(x) \quad (3)$$

donde  $\mu$  es una medida de equilibrio y  $S^n x$  es la evolución temporal del punto  $x$  en el espacio de fase.

Vemos inmediatamente que  $f$  y  $\mu$  también dependen de  $N$ , y por tanto una formulación matemática rigurosa debería especificar el significado de convergencia en (3). Vamos en busca de ese significado.

## 5 Encontrando un objeto y un problema matemáticos (de Boltzmann a Von Neumann: de 1870 a 1931)

Llevó mucho tiempo para que el objeto matemático de la Hipótesis Ergódica fuese encontrado. Sólo en 1929 Koopman comenzó a investigar grupos de transformaciones que preservan medidas, o sea, transformaciones que llevan conjuntos en otros que miden lo mismo. Esos progresos fueron precedidos por los éxitos de la Teoría de la Medida, la cual, en otro sentido, permitiría a Kolmogorov en 1933 establecer los fundamentos axiomáticos de la Teoría de la Probabilidad.

Más precisamente, sea  $M$  un espacio abstracto, el espacio de fase del sistema, y  $\mu$  una medida de probabilidad sobre una  $\sigma$ -álgebra de  $M$ . La dinámica está dada por la aplicación sucesiva de una transformación  $S$  (o su inversa  $S^{-1}$ ), que preserva medida. Esto significa que para todo conjunto del  $\sigma$ -álgebra,  $A \subset M$ , y para todo  $n \in \mathbb{Z}$ ,  $\mu(S^n A) = \mu(A)$ .

Finalmente, sea  $f : M \rightarrow \mathbb{R}$  una función (u observable) razonablemente regular, por ejemplo, perteneciente al espacio de las funciones de cuadrado integrable:  $\int f^2(x) d\mu(x) < \infty$ ; en este caso se dice que  $f \in L_2(\mu)$ .

---

<sup>1</sup>La palabra ergódica, usada en este contexto, proviene del griego *ergon* (trabajo) y *odos* (camino, paso).

De esa forma el objeto matemático  $(M, S, \mu)$ , con las funciones  $f$  – está definido.

En el mismo año (1929) Von Neumann probó el primer teorema ergódico, el cual establece que la media temporal de una función  $f$  a lo largo de una trayectoria que pasa por un punto  $x$  tiene límite  $\tilde{f}(x)$ :

**Teorema Ergódico.** Cuando  $T \rightarrow \infty$ ,

$$\frac{1}{T} \sum_{n=0}^{T-1} f(S^n x) \rightarrow \tilde{f}(x), \text{ siendo la convergencia en el sentido } L_2(\mu).$$

En 1931 Birkhoff y Khincin separadamente probaron la misma fórmula, pero con convergencia para casi todo  $x$  en el sentido de la medida  $\mu$ , o sea, que esa convergencia se da para todos los puntos de  $M$  excepto los que están en un conjunto con  $\mu$ -medida nula. La función límite  $\tilde{f}(x)$ , que normalmente es llamada media de Birkhoff, es tal que su integral (sobre todo el espacio) tiene el mismo valor que la de  $f$ :  $\int_M \tilde{f}(x) d\mu(x) = \int_M f(x) d\mu(x)$ .

Dado que el término que depende de  $T$  es exactamente el promedio temporal que aparece en la Hipótesis de Boltzmann, esta pasó a tener un significado preciso inmediatamente después de formulado el Teorema Ergódico: en grandes sistemas de partículas interactuando en equilibrio, las medias temporales  $\tilde{f}(x)$  no dependen de  $x$ . Observe que, como  $\mu$  es una medida de probabilidad, si  $\tilde{f}(x)$  fuera constante, cuando integramos con respecto a  $\mu$  obtendremos el mismo valor  $\tilde{f}(x)$  y, por la observación final del párrafo anterior, resulta  $\tilde{f}(x) = \int_M \tilde{f}(x) d\mu(x) = \int_M f(x) d\mu(x)$ .

La propiedad de que en un sistema las medias de Birkhoff no dependan del punto inicial  $x$  fue considerada muy importante desde el comienzo, a tal punto que los sistemas  $(M, S, \mu)$  que verifican esta propiedad reciben el nombre de ergódicos.

**Definición: Ergodicidad.** El sistema  $(M, S, \mu)$  es ergódico si para toda función razonablemente regular  $f$ , por ejemplo, para toda  $f \in L_2$ , se satisface  $\tilde{f}(x)$  constante para  $\mu$ -casi todo punto  $x \in M$ .

Esta definición es equivalente a cada una de las siguientes tres afirmaciones:

# Si  $f$  es razonablemente regular y constante a lo largo de las órbitas ( $f \circ S^n = f$ , para todo  $n \in \mathbb{Z}$ ), entonces  $f$  es constante para  $\mu$ -casi todo punto  $x \in M$ .

# Si un conjunto  $A \subset M$  es invariante por  $S^n$  ( $S^n A = A$ ), entonces  $A$  es o bien casi todo el espacio  $M$  o tiene medida nula ( $\mu(A) = 0$  o  $1$ ).

#  $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{n=0}^{T-1} \mu(S^{-n} A \cap B) = \mu(A)\mu(B)$ , para todo par de conjuntos medibles  $A, B \subset M$ .

Resumiendo: tenemos un modelo matemático (grupo de transformaciones que preservan medidas) y la nueva noción de ergodicidad. Se plantea el problema de establecer si un sistema interesante desde el punto de vista físico es ergódico o no.

Estos progresos dieron origen a una nueva área de la Matemática, la **Teoría Ergódica**, y a varias subareas. Una de ellas estudia generalizaciones de los teoremas ergódicos; otra, formas más fuertes de estocasticidad (mixing, sistemas de Bernoulli, decaimiento de correlaciones; ver, por ejemplo, [6], Cap. 2); una área especial – particularmente importante para este trabajo – estudia la ergodicidad de ciertos sistemas que aparecen en la Mecánica; otra, los problemas de isomorfismos y clasificación de los sistemas dinámicos, etc.

Los sistemas de partículas cuyos movimientos son descritos por la transformación  $S$  de la Sección 4 corresponden a una clase más general de los llamados **flujos hamiltonianos en variedades simplécticas**. Se creía que se podría aplicar la Hipótesis de Boltzmann a tales flujos. Pero existe un ejemplo célebre de Michel Herman que muestra la existencia de una cantidad muy grande de tales flujos que no son ergódicos (ver [7]). Ese ejemplo fue presentado al final de los años 80, pero ya anteriormente, las tentativas de prueba de la Hipótesis Ergódica se restringían a los modelos de bolas con choques elásticos.

## 6 Probando los primeros teoremas relevantes (de Von Neumann a Sinai de 1931 a 1970)

Los métodos para establecer la ergodicidad de sistemas mecánicos vinieron de otras áreas de la Matemática, en particular de la Teoría de Sistemas Dinámicos. En 1938-39 Hedlund y Hopf encontraron un método para demostrar la ergodicidad de flujos geodésicos en variedades compactas de curvatura negativa (se trata de movimientos inerciales sobre, por ejemplo, superficies que resultan de girar una hipérbola equilátera en torno de una asíntota). El principal descubrimiento fue el, así llamado, **comportamiento hiperbólico de los sistemas dinámicos**, que implicaba la ergodicidad de tales sistemas.

Hiperbolicidad significa la existencia de curvas transversales sobre las que el sistema dinámico actúa expandiendo (variedad inestable) o contrayendo (estable). Si la variedad no es una superficie, o sea, si tiene dimensión mayor que dos, las curvas son subvariedades cuya suma de dimensiones es igual a la dimensión de la variedad original. Hiperbolicidad implica inestabilidad para todas las órbitas: trayectorias que comienzan arbitrariamente próximas unas de las otras se separan en el futuro o en el pasado: llamamos a esta última propiedad **sensibilidad con respecto a las condiciones iniciales**.

El ejemplo mas simple de sistema hiperbólico es el famoso “cat” de Arnold (el nombre no proviene del gato que era usado para representarlo, sino de “Continuous Automorphisms of the Torus”). Consideramos la aplicación  $S_A$  del toro bidimensional  $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$  en si mismo, definida por la matriz  $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ . Observamos que la transformación expande en una dirección y contrae en otra, siendo estas direcciones las correspondientes a los vectores propios con autovalores respectivamente  $\lambda_u > 1$  y  $\lambda_s < 1$  de la matriz  $A$ .

En 1942, poco después de los resultados fundamentales de Hedlund y Hopf, el físico ruso N. S. Krylov mostró que la inestabilidad observada en los flujos geodésicos también existían en los sistemas de bolas elásticas que modelan los gases. Este descubrimiento y el avance de las ideas de Hedlund y Hopf en la Teoría de Sistemas Dinámicos Hiperbólicos fueron la base para la versión de Sinai de la Hipótesis Ergódica de Boltzmann formulada en 1963 para el caso particular de bolas elásticas.

**Hipótesis Ergódica de Boltzmann-Sinai** [14]. El sistema de  $N$  bolas elásticas sobre  $\mathbb{T}^2$  o  $\mathbb{T}^3$  es ergódico para todo  $N \geq 2$ .

Como estos sistemas mecánicos tienen muchas cantidades que se conservan a lo largo de las órbitas, la interpretación que debemos dar a la conjetura es que tendremos hiperbolicidad en las (componentes conexas de las) subvariedades del espacio de fase definidas por los invariantes del movimiento. La primera diferencia conceptual que tiene esta conjetura comparada con la formulación original de Boltzmann es que aquí no se hace ninguna suposición sobre el número  $N$  de bolas. La ergodicidad (y otras propiedades estadísticas más fuertes) se verificaría para cualquier  $N \geq 2$ . De acuerdo con las ideas de los fundadores de la Mecánica Estadística el estado de un sistema de partículas es aleatorio porque cualquier sistema en la naturaleza contiene un número enorme de ellas: la aleatoriedad surge del gran número de grados de libertad. En el planteamiento de Sinai el número de partículas no tiene ninguna importancia: un sistema de dos partículas elásticas chocándose presenta comportamiento estocástico.

Una segunda característica importante del argumento de Sinai es que el conocimiento local de sistemas hiperbólicos lleva a una descripción global. La macrodinámica surge de la microdinámica.

En 1963 Sinai pensó haber probado el resultado. Escribió: “Entre los principales resultados de este artículo debería contarse el teorema de la ergodicidad de los sistemas de las bolitas (pellets) duras en una caja rectangular”. Pero sólo en 1970 [15] pudo probar su conjetura para el caso  $N = 2$ : discos moviéndose en el toro  $\mathbb{T}^2$ .

Los intentos de probar la conjetura de Boltzmann-Sinai se extienden a lo largo de 40 años. A esta altura parece que el problema está casi resuelto. Ver última Sección.

## 7 El argumento de Sinai. Billares (1970)

Considérese el modelo 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, suponga que todos los discos (bolas) tienen el mismo radio  $r$  y la misma masa  $m$ . Cada bola (disco) se mueve libremente, i.e. con velocidad constante, hasta que choca con otra bola. Cuando dos bolas (discos) chocan, cambian sus velocidades de acuerdo con las leyes de los choques elásticos.

Estas leyes significan lo siguiente. Supongamos que dos bolas chocan. Llamemos  $q_1$  y  $q_2$  a sus centros y por  $v_1$  y  $v_2$  sus vectores velocidad en el momento de la colisión. Sea  $L$  la línea que une los centros  $q_1$  y  $q_2$ . 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 la componente perpendicular a  $L$ . Entonces las nuevas velocidades, salientes, de las bolas son

$$v_1^{\text{nue}} = v_1^\perp + v_2^0 \quad \text{y} \quad v_2^{\text{nue}} = v_2^\perp + v_1^0$$

En otras palabras, las bolas intercambian las componentes de la velocidad paralelas a la línea de los centros  $L$  y mantienen las componentes ortogonales. Observamos que las leyes de los choques elásticos implican la conservación de la energía cinética total  $\sum m \|v_i\|^2 / 2$  y el momento total  $\sum m v_i$  del sistema de las  $N$  bolas (discos). También observamos que la colisión de dos bolas duras con centros  $q_1$  y  $q_2$  sólo puede ocurrir si  $\text{dist}(q_1, q_2) = 2r$ , i.e.  $\|q_1 - q_2\|^2 = (2r)^2$ .

El sistema de las  $N$  bolas (discos) moviéndose en el espacio abierto (o en el plano) sin paredes no es muy interesante desde el punto de vista dinámico. Como es intuitivamente claro, el número total de choques entre pelotas es siempre finito, y después del último choque las pelotas se van para siempre libremente. Más aún, el número de choques entre las  $N$  pelotas en el espacio abierto está uniformemente acotado por una constante que sólo depende de  $N$ .<sup>2</sup>

Consideremos ahora  $N$  bolas o discos encerrados en un dominio acotado  $R$ , llamado el recipiente. Las bolas (discos) chocan elásticamente entre ellas y con las paredes del recipiente. Precisamente, si una bola con centro  $q$  choca una pared en el punto  $w \in \partial R$ , entonces descomponemos el vector velocidad como  $v = v^0 + v^\perp$ , donde  $v^0$  es la componente paralela a la línea pasando a través de  $q$  y  $w$ , y  $v^\perp$  es perpendicular a esa línea. La nueva velocidad, saliente, es  $v^{\text{nue}} = v^\perp - v^0$ . Obsérvese que la energía cinética total del sistema es preservada, pero *no* su momento total.

**Modelo de Sinai de las bolas duras.** Ahora reducimos el sistema de  $N$  bolas duras a un billar. Supongamos que el recipiente  $R$  es un toro, por lo que no hay paredes (i.e.,  $\partial R = \emptyset$ ). Entonces las bolas sólo chocan entre ellas y también el momento total se conserva.

Más concretamente, el sistema está compuesto por  $N \geq 2$  bolas, cada una con masa unitaria y radio  $r > 0$ , moviéndose en  $\mathbb{T}^d$ , el toro  $d$ -dimensional ( $d \geq 2$ , siempre se puede suponer  $d = 2$  –discos– o  $d = 3$  –bolas). En el espacio de fase, la  $i$ -ésima bola es representada por  $(q_i, v_i) \in \mathbb{T}^d \times \mathbb{R}^d$ , donde  $q_i = (q_i^1, q_i^2, \dots, q_i^d)$  indica la posición del centro de esta bola y  $v_i$  su velocidad.

El espacio de configuraciones  $\tilde{\mathbf{Q}}$  de las  $N$  bolas en  $\mathbb{T}^d$  es un subconjunto del toro de dimensión  $N \cdot d$  pues, dado que las bolas son macizas, sus centros deben estar a una distancia mayor o igual a dos veces

<sup>2</sup>Esto último fue probado sólo recientemente:

**Teorema [1].** *El sistema de  $N$  bolas duras con masas y radios arbitrarios en el espacio abierto  $\mathbb{R}^d$  puede tener a lo sumo  $C < \infty$  choques.*

$$C = \left( 32 \sqrt{\frac{m_{\max}}{m_{\min}}} \frac{r_{\max}}{r_{\min}} N^{\frac{3}{2}} \right)^{N^2}$$

donde  $m_{\max}$  y  $m_{\min}$  son las masas máximas y mínimas y  $r_{\max}$  y  $r_{\min}$  son los radios máximos y mínimos de las bolas, respectivamente.

Es destacable que este resultado admite una prueba geométrica relativamente elemental. Obsérvese que si los radios y masas son iguales  $C$  sólo depende de  $N$ .

sus radios, y por tanto debemos retirar de  $\mathbb{T}^{N \cdot d}$  los siguientes  $\binom{N}{2}$  obstáculos cilíndricos:

$$C_{i,j} = \left\{ Q = (q_1, \dots, q_N) \in \mathbb{T}^{N \cdot d} : |q_i - q_j| < 2r \right\},$$

$1 \leq i < j \leq N$ . Como ya fue observado, la energía  $H = \frac{1}{2} \sum_1^N v_i^2$  y el momento total  $P = \sum_1^N v_i$  son integrales primeras del movimiento. Entonces, sin pérdida de generalidad, podemos suponer que  $H = \frac{1}{2}$ ,  $P = 0$  y, también, que la suma de las componentes espaciales  $B = \sum_1^N q_i = 0$ . Estas condiciones son muy naturales desde el punto de vista físico; algunas están relacionadas con el estado de equilibrio en que estamos considerando el fenómeno. Para estos valores de  $H, P$  y  $B$ , el espacio de fase se reduce a  $M := \mathbf{Q} \times S_{a-1}$  donde

$$\begin{aligned} \mathbf{Q} &:= \left\{ Q \in \tilde{\mathbf{Q}} : \sum_1^N q_i = 0 \right\} \\ &= \left\{ Q \in \mathbb{T}^{N \cdot d} \setminus \bigcup_{1 \leq i < j \leq N} C_{i,j} : \sum_1^N q_i = 0 \right\} \end{aligned}$$

Aquí  $a := \dim \mathbf{Q} = N \cdot d - d$  y  $S_k$  denomina a la (media) esfera unitaria  $k$ -dimensional. La dimensión de  $\mathbf{Q}$  es  $N \cdot d - d$  porque la condición  $\sum q_i = 0$  impone  $d$  restricciones (una para cada coordenada) a los puntos de  $\tilde{\mathbf{Q}}$ . La dimensión del espacio de los vectores es  $N \cdot d - d$  porque  $\sum v_i = 0$  y tales vectores están en la esfera unitaria  $(N \cdot d - d - 1)$ -dimensional porque sus módulos (al cuadrado)  $\sum v_i^2$  son constantes. Es fácil ver que la dinámica de las  $N$  bolas, determinada por su movimiento uniforme con colisiones elásticas y el flujo del billar  $\{S^t : t \in \mathbb{R}\}$  sobre  $\mathbf{Q}$  con reflexiones en la frontera  $\partial \mathbf{Q}$  son isomorfos y conservan la medida de Liouville  $d\mu = \text{const} \cdot dq \cdot dv$  (para detalles ver [3], Cap. 6).

El conjunto  $D$  de las singularidades contiene todas las intersecciones de las superficies cilíndricas  $\partial C_{ij}$ . Estas intersecciones corresponden a colisiones simultáneas de tres o más bolas. El proceso posterior a tales choques múltiples no está definido. La regla general es, por tanto, ignorar las trayectorias del billar que chocan en  $D$ .

La formulación de Sinai de la Hipótesis de Boltzmann la coloca naturalmente en el contexto de billares en espacios de dimensión alta. Recordemos, para resumir, que el espacio es de dimensión  $2Nd - 2d - 1$  y que el número de cilindros es  $N(N-1)/2$  donde  $N$  es el número de bolas y  $d$  es la dimensión del espacio donde se mueven las bolas.

**Gas de Lorentz.** Este es otro modelo muy popular en física. Imagínese una partícula puntual moviéndose entre dos bolas rígidas fijas, en el espacio  $\mathbb{R}^3$ . Las bolas tienen el mismo radio  $r$ . Pueden ser colocadas aleatoriamente o formar una estructura del tipo cristalina (por ejemplo, sus centros pueden colocarse en los vértices del reticulado –lattice–  $\mathbb{Z}^3$ ). Las bolas no se mueven, solamente una partícula puntual se mueve entre ellas y choca elásticamente. Las bolas actúan como obstáculos.

Este modelo fue introducido por H. Lorentz en 1905 en el estudio de electrones en metales. La partícula móvil representa un electrón, y las bolas juegan el papel de las moléculas del metal. Cuando las bolas están colocadas en los vértices de un reticulado regular, se dice que el gas de Lorentz es periódico.

En lugar de bolas, podemos colocar otros cuerpos idénticos en los vértices del reticulado, y requerir a la partícula móvil que choque con los obstáculos elásticamente. Los cuerpos pueden ser cúbicos o cuerpos poliédricos más generales. Uno también puede definir un gas de Lorentz en un plano; en este caso obstáculos bidimensionales son colocados en los vértices del reticulado  $\mathbb{Z}^2$ .

Reducimos ahora el gas de Lorentz a un billar. De hecho él es “casi” un billar, excepto que el dominio por el que puede andar la partícula no es acotado y tiene volumen infinito. Esto se soluciona fácilmente porque desde que los obstáculos están colocados en los vértices del reticulado, su estructura es periódica y uno puede encontrar un **dominio fundamental**  $H$ , o sea un domino cuya traslación paralela cubra todo el espacio. El dominio  $H$  contiene sólo unos pocos (a veces sólo uno) obstáculos, y todos los otros

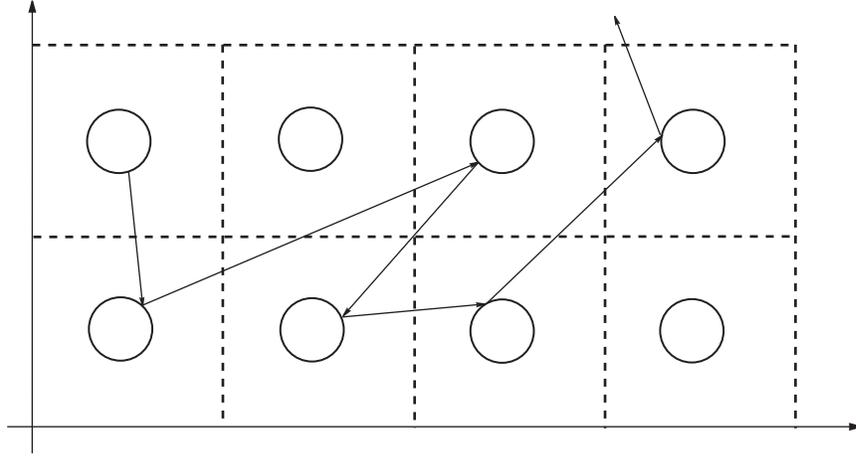


Figura 8: Gas de Lorentz periódico.

se obtienen por la traslación paralela de aquéllos en  $H$ . La Figura 8 muestra un gas periódico de Lorentz en un plano, donde los obstáculos son discos fijos y el dominio fundamental es un cuadrado conteniendo un disco.

El movimiento de la partícula puede ser proyectado al dominio fundamental  $H$  y tendremos un sistema de billar en  $H$  con condiciones de frontera periódicas, eso es un billar en el toro  $\mathbb{T}^d$ ,  $d = 2, 3$ . La mesa del billar  $\mathbf{Q}$  se obtiene sacando uno o varios obstáculos del toro. En el caso del gas de Lorentz de la Figura 8 estamos nuevamente ante el billar de la Figura 1.

**Billares de Sinai.** Para explicar de manera más simple las principales características de los sistemas que hemos venido describiendo, nos restringiremos al modelo de las bolas duras con  $N = 2$  y  $d = 2$ , o sea, consideraremos dos bolas (discos) en el toro  $\mathbb{T}^2$ . Entonces  $\partial\mathbf{Q}$  es la unión de un número finito de curvas regulares y los vectores  $v$  están en la (media) circunferencia  $\mathcal{S}_1$ .

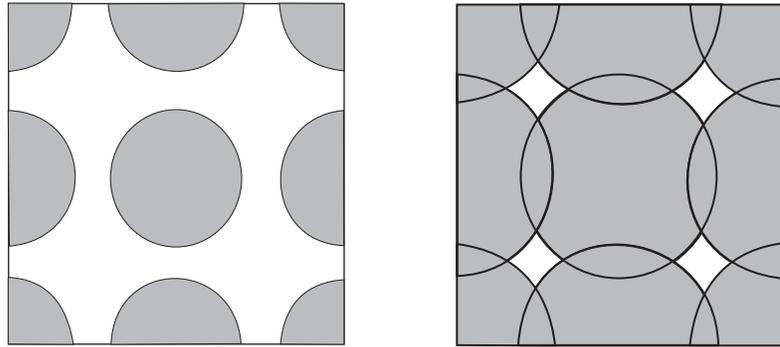


Figura 9: Billares dispersores. (a) Cuadrado de lado 1 y radio  $0 < r < 1/4$ . (b) Cuadrado de lado 1 y radio  $1/4 \leq r < 1/2$

Podemos dar un paso más y suponer que uno de los discos está fijo y que sólo el centro del otro se mueve. Llegamos así al billar que analizamos en la Sección 1, Figura 1. Más aún, para no restringirnos al toro, podemos considerar mesas de billar en el plano que satisfacen las principales propiedades dinámicas

de aquel billar en  $\mathbb{T}^2$ . Los billares planos que surgen del modelo de las bolas duras para  $N = 2$  y  $d = 2$  son como los de la Figura 9. Su ergodicidad fue probada por Sinai en 1970 y son llamados **billares dispersores de Sinai** (ver la definición de arcos dispersores en la Sección 2).

Es claro que podemos considerar billares planos cuyas fronteras no son generadas por condiciones derivadas de modelos físicos y preguntarnos cómo deben ser las fronteras para que la correspondiente transformación de billar sea ergódica. En particular es interesante saber cómo “montar” una mesa de billar para que su sistema dinámico sea ergódico. Está demostrado que el billar en el estadio de Bunimovich (dos semicircunferencias unidas por segmentos paralelos), en la cardioide y en muchos estadios elípticos son ergódicos y satisfacen otras propiedades estocásticas. En el caso del estadio elíptico descrito en la Sección 3, si  $a < \sqrt{2}$ , para tener un billar ergódico la separación entre las dos mitades debe crecer tendiendo a infinito a medida que  $a$  crece tendiendo a  $\sqrt{2}$ . Ver [11]. Para el billar de la Figura 6, si  $a \geq \sqrt{2}$  no es posible construir estadios elípticos ergódicos.

## 8 Cómo se prueba la Ergodicidad de los Billares

Desde ya debemos aclarar que si bien seguiremos usando el término ergódico para los sistemas dinámicos con el tipo de desorden (caos) definido en la Sección 5, en realidad en casi todos esos casos se han probado propiedades estadísticas mucho más fuertes.

De una forma u otra, la prueba de la ergodicidad de los sistemas dinámicos diferenciables pasa por la prueba de algún tipo de hiperbolicidad (ver Sección 6). El procedimiento padrón es el estudio de los exponentes de Liapunov de la transformación y la aplicación de la llamada Teoría de Pesin, que permite construir las curvas expansoras y contractoras a las que nos referimos anteriormente. Para evitar mayores complicaciones técnicas vamos a suponer que trabajamos en el espacio euclidiano  $\mathbf{R}^d$ ,  $d \geq 2$ . Quien quiera un modelo aún más intuitivo puede pensar en el plano.

Si  $p$  es un punto fijo de un difeomorfismo  $S$  ( $S$  y su inversa tienen derivadas continuas,  $S(p) = p$ ) definido en un conjunto de  $\mathbf{R}^d$  y queremos analizar el comportamiento de  $S$  en un entorno de  $p$ , como primera aproximación podemos estudiar su parte lineal, o sea, la derivada (diferencial)  $(S)'_p : \mathbf{R}^d \rightarrow \mathbf{R}^d$ . Si  $\alpha_i$  son las raíces reales del polinomio característico y  $v_i$  sus respectivos vectores propios, tendremos que  $\lim_{n \rightarrow \pm\infty} 1/n \log \|(S^n)'_p v_i\| = \log |\alpha_i|$ . Para el caso de autovalores complejos (que aparecen de a pares conjugados  $\alpha_j, \bar{\alpha}_j$ ), esta igualdad se verifica para  $v_j$  contenido en subespacios de dimensión igual al doble de la multiplicidad de  $\alpha_j$ . Observe que eso no ocurre si en lugar del autovector  $v_i$  tomamos cualquier vector  $v$ . Llamamos **exponentes de Liapunov** a estos números  $\log |\alpha_i| = \lambda_i$ . Es fácil ver que la parte lineal de la transformación  $S$  (su diferencial) expande en los subespacios correspondientes a los autovectores con exponente de Liapunov  $\lambda_i > 0$  y contrae en los subespacios correspondientes a los autovectores con exponente de Liapunov  $\lambda_i < 0$ . En cualquier caso, no debemos considerar el vector nulo en esos subespacios.

Si existe una medida  $\mu$  invariante por  $S$ , un célebre teorema de Oseledets establece la existencia del límite que define a los exponentes de Liapunov para  $\mu$ -casi todo punto  $p$ , independientemente de si éste es periódico o no. El teorema de Oseledets puede ser aplicado a transformaciones  $S$  definidas en conjuntos mucho más generales. Para extenderlo al caso de los billares con los que trabajamos daremos un enunciado en el contexto más general de **transformaciones continuas con singularidades**. Para evitar la introducción de notaciones que complican la comprensión de lo que es realmente importante, nos restringiremos a trabajar en  $\mathbf{R}^d$ . Para un enfoque más detallado y general se puede consultar [9], [10], [2].

Sea  $M$  la unión de un número finito de regiones en  $\mathbf{R}^d$ , pegadas a lo largo de “hipersuperficies” cuya unión está contenida en el conjunto de las **singularidades**  $D$ . Este conjunto, a su vez, es la unión de un conjunto finito de variedades cerradas de dimensión menor que  $d$  (si  $d = 2$ ,  $D$  es la unión de un número finito de curvas y puntos). El conjunto  $P = M \setminus D$  es abierto. Sea  $\mu$  una medida de probabilidad definida sobre los borelianos de  $M$ .

Sea  $S : P \rightarrow M$  una transformación con inversa, ambas suficientemente diferenciables (difeomorfismo

de clase  $C^r$ ,  $r \geq 1$ , para el teorema de Oseledets, y  $r \geq 2$  para el resto de la teoría), de  $P$  en su imagen, que preserva la medida  $\mu$  y que verifica algunas otras condiciones de crecimiento de la derivada cerca de las singularidades. Sea  $H$  el conjunto de los puntos que tienen infinitos iterados tanto para el futuro como para el pasado:  $H = \bigcap_{n=-\infty}^{\infty} S^n P$ ; por tanto  $\mu(H) = 1$ . Entonces, el **teorema de Oseledets** dice que para casi todo punto  $p \in H$  existen  $m(p)$  números reales  $\lambda_i(p) : \lambda_1(p), \lambda_2(p), \dots, \lambda_{m(p)}(p)$  y subespacios  $E_1(p), E_2(p), \dots, E_{m(p)}(p)$ , tales que

$$\lim_{n \rightarrow \pm\infty} \frac{1}{n} \log \|(S^n)'_p v_i\| = \lambda_i(p) \quad \text{para todo vector no nulo } v_i \in E_i(p).$$

$E_i(p)$  es el **subespacio propio del exponente de Liapunov**  $\lambda_i(p)$ . Estos números pueden variar con  $p$ , pero son invariantes a lo largo de su órbita:  $\lambda_i(S^n(p)) = \lambda_i(p)$ . El teorema de Oseledets asegura que para casi todo punto (puntos **regulares**) del espacio donde actúa una transformación que preserva medida, el comportamiento de los puntos próximos a su órbita puede ser aproximado teniendo en cuenta el comportamiento en direcciones conocidas, que serán contracciones o expansiones según  $\lambda_i$  sea menor o mayor que zero.

Estos subespacios y exponentes de Liapunov se comportan asintóticamente como los subespacios propios y el logaritmo del valor absoluto de los autovalores de una matriz. ¿Cómo deducir de propiedades de esos números propiedades de la función  $S$ , en particular propiedades relacionadas con la hiperbolicidad?

La respuesta a esta pregunta es dada por la llamada **Teoría de Pesin**, que trata, en primer lugar, de la construcción de **variedades invariantes** expansivas y contractivas en puntos en que hay exponente de Liapunov no nulos.

No seremos muy rigurosos en los enunciados de los teoremas de Pesin, ni los presentaremos con su máxima generalidad, para no complicar la exposición. Definimos la **región de Pesin**,  $\Sigma$ , como el conjunto de los puntos regulares  $p$  para los que todos los exponentes de Liapunov son diferentes de zero ( $\lambda_i(p) \neq 0$  para todo  $1 \leq i \leq m(p)$ ). Entonces para cada punto  $p \in \Sigma$  existen variedades invariantes **estables** ( $W^s(p)$ ) e **inestables** ( $W^u(p)$ ) cuyos puntos se aproximan o se apartan respectivamente entre sí cuando iteramos por  $S$ . La suma de las dimensiones de estas variedades es igual a la dimensión de todo el espacio, y cada una de ellas es tangente al hiperplano (subespacio) generado por los vectores con exponentes de Liapunov negativos ( $W^s(p)$ ) o positivos ( $W^u(p)$ ). Además,  $\Sigma$  puede ser descompuesto en un conjunto numerable de subconjuntos invariantes, en cada uno de los cuales  $S$  es ergódica.

Si la región de Pesin tiene medida total ( $\mu(\Sigma) = 1$ ) se dice que  $S$  es (**no uniformemente**) **hiperbólica** o que el sistema  $(M, S, \mu)$  tiene **comportamiento caótico**. Usamos la expresión no uniformemente porque los ángulos entre los subespacios estables e inestables, así como los  $\lambda(p)$ , puede aproximarse mucho a cero, generando una serie de dificultades en el control de las propiedades básicas de la hiperbolicidad.

Entonces, si los exponentes de Liapunov son no nulos en casi todo punto de  $H$  (o de  $P$  o  $M$ ) entonces en casi todo punto tenemos variedades inestable y estable local (cuyos puntos cuando iterados por  $S^n$  se aproximan o se alejan exponencialmente cuando  $n \rightarrow \infty$ ); la suma de sus dimensiones es  $d$ . Más aún, la segunda parte del teorema establece que existe un conjunto numerable  $\Sigma_i$  de **componentes ergódicas**, o sea, conjuntos de medida positiva que no poseen subconjuntos  $S$ -invariantes con medida mayor que cero y menor que uno:  $S$  es caótica dentro de cada  $\Sigma_i$ .

De esta manera el problema de probar la ergodicidad de todo el sistema se transforma en probar: 1) que los exponentes de Liapunov son no nulos  $\mu$ -casi todo punto y, por tanto, 2) que la componente ergódica es única.

Existen dos maneras tradicionales de atacar el primer problema: mostrar que

**A)** en casi todo punto  $x \in N$  existen **conos**  $C(x)$  invariantes por  $S' : S'_x(C(x)) \subset C(S(x))$  (un cono  $C(x) \subset \mathbf{R}^2$ , por ejemplo, es un ángulo con vértice en  $x$ , ver Figura 9); o

**B)** en casi todo punto  $x \in N$  puede ser definida una **forma cuadrática** no degenerada  $Q_x$  creciente a lo largo de las órbitas de  $S : Q_{S(x)}(S'_x v) - Q_x v > 0$  (\*) si  $v$  es cualquier vector con origen en  $x$ . Una forma

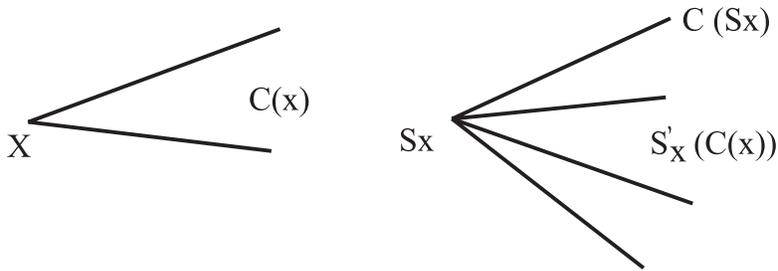


Figura 10: Cono invariante  $C(x)$ .

cuadrática es determinada por una matriz  $A_x$  simétrica siendo  $Q_x : \mathbf{R}^d \rightarrow \mathbf{R}$ , definida por  $Q_x v = v^t A v$ ; si  $A$  no tiene autovalores nulos, se dice que  $Q$  es no degenerada.

A) e B) son esencialmente equivalentes; en particular observamos que el conjunto  $\{v : Q_x v > 0\}$  constituye un cono  $C(x)$ . Si una de las condiciones es verificada, el sistema dinámico  $(M, S, \mu)$  tiene exponentes de Liapunov no nulos  $\mu$ -casi todo punto.

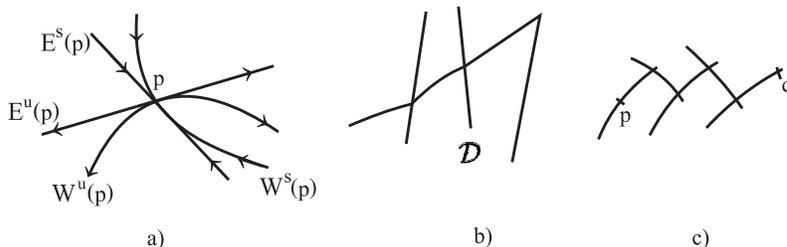


Figura 11: a) Subespacios estables e inestables  $[E^{s,u}(p)]$  y sus variedades locales  $W^{s,u}(p)$ ; b) Particiones de  $S^n(W^u(p))$  por  $D$ ; c) Zig-zag de variedades estables e inestables uniendo  $p, q$ .

Las condiciones para dar pruebas del segundo problema (unicidad de la componente ergódica) son más complejas. Intentaremos explicarlas en el contexto de la hiperbolicidad (no uniforme) probada por el método B), y suponiendo que  $d = 2$ . Si además de B), el sistema  $(M, S, \mu)$  con  $M \subset \mathbf{R}^2$  verifica las condiciones siguientes, entonces el sistema es ergódico (ver [8] y [10]):

B1) para cada  $n \in \mathbf{N}$ , las órbitas con  $n$  iteraciones que comienzan y terminan en el conjunto de las singularidades  $D$  constituyen un conjunto finito;

B2) las tangentes a  $D$  están en los conos de vectores donde la forma cuadrática  $Q$  es estrictamente positiva o negativa;

B3) para casi todo punto de  $D$ , sus órbitas entran de manera recurrente en regiones donde  $Q$  crece uniformemente. Esto significa que en la expresión (\*) la diferencia es mayor que una constante positiva  $\delta$ ;

B4) la medida de los puntos cuyas variedades invariantes (estable e inestable) cortan  $D$  y tienen longitud mayor que  $\epsilon$ , es menor que  $\text{cte} \cdot \epsilon$ ;

B5) existe un  $x \in H$  tal que su órbita  $\{S^n(x)\}_{n \in \mathbf{Z}}$  pasa por todas las componentes conexas por arcos de  $N$ .

Las variedades inestables locales  $W^u(x)$ , al evolucionar por  $S$ , se van partiendo al cortarse con  $D$  (recordamos que en  $D$  la función  $S$  o su inversa no está definida). Estas condiciones aseguran que aún así, los trechos cortados se hacen suficientemente largos al continuar iterándolos: **la hiperbolicidad prevalece sobre el fraccionamiento**. El hecho de que las variedades invariantes sean suficientemente

largas implica que podemos unir dos puntos de un entorno de casi todo punto con un zig-zag de arcos estables e inestables; ver Figura 11.

## 9 Probando la Hipótesis de Boltzmann-Sinai

La principal dificultad con el modelo de los gases, en su formulación billarística, es que si el obstáculo es no estrictamente convexo (cilindros, por ejemplo, ver Figura 12), no es simple probar la existencia de exponentes de Liapunov diferentes de cero; y es aún más difícil probar la unicidad de la componente ergódica.

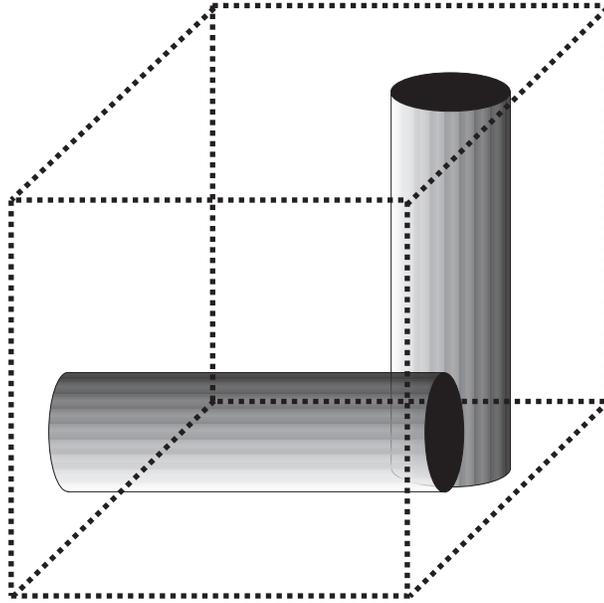


Figura 12: Obstáculos cilíndricos que aparecen en la formulación billarística del modelo de los gases.

Krylov y Sinai ya habían observado que mientras un billar con obstáculos estrictamente convexos se comportan como un sistema dinámico “bien” hiperbólico, cuando los obstáculos no son estrictamente convexos, existe apenas hiperbolicidad parcial. Los obstáculos de las Figuras 1, 8, 9 y 13 son estrictamente convexos (el billar es **dispersor**), mientras que los de las Figuras 12 y 14 son apenas convexos (el billar es **semidispersor**).

Geoméricamente, el mecanismo de generación de hiperbolicidad puede ser ilustrado con imágenes inspiradas en la óptica, como ya sugiriéramos en la Sección 2 al definir componentes focalizadoras y dispersoras de la frontera. Consideremos, como en la Figura 13, un obstáculo estrictamente convexo que oficia de espejo. Consideremos también un frente de onda plano (que corresponde a dar un punto  $x = (Q, V)$  en el espacio de fase  $M$ , considerar en el espacio de configuraciones – donde realmente ocurre el movimiento – un hiperplano  $\Gamma$  por  $Q$  y, en cada punto de ese hiperplano, tomar vectores paralelos a  $V$ ). Entonces, luego que el frente de ondas alcance el espejo, el se vuelve estrictamente convexo y las distancias lineales medidas a partir de  $\Gamma$  son uniformemente expandidas. Éste es exactamente el mecanismo que genera la hiperbolicidad uniforme en los billares. En este caso las dificultades vienen, como ya fue en parte observado, de la existencia de trayectorias tangentes que hacen que el sistema pierda diferenciabilidad.

Si los obstáculos no son estrictamente convexos aparecen situaciones como las de la Figura 14 en el hiperplano  $\Gamma$  no es curvada en todas las direcciones: las trayectorias que chocan con una directriz del

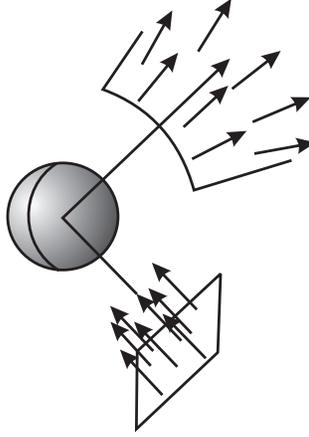


Figura 13: Obstáculo estrictamente convexo: un haz paralelo es completamente dispersado después del choque.

cilindro (las rectas paralelas al eje, en la figura) salen paralelas, son “neutras”, y por tanto no se separan ni aumentan las distancias medidas a partir de  $\Gamma$ . Esta dirección neutra puede desaparecer debido a otros choques con otros cilindros. El hecho de que la hiperbolicidad global (la expansión en todos los sentidos) sea obtenida a través de muchos choques con obstáculos, lleva naturalmente a la introducción del concepto de *suficiencia* de las trayectorias, cuya descripción intuitiva damos a continuación.

Sea  $S^{[a,b]}x$  un segmento finito de trayectoria que no pasa por singularidades, o sea, tal que  $S^j x \notin D$  para  $j \in [a, b]$ . Sea  $S^a x = (Q, V) \in P$  y consideremos el hiperplano  $\Gamma(S^a x) := \{(Q + dQ, V) : dQ \in \mathbb{R}^d, \text{pequeño, perpendicular a } V\}$ . Decimos que un segmento de trayectoria  $S^{[a,b]}x$  es **suficiente** si  $\pi(S^b \tilde{\Gamma})$  es estrictamente convexo. Un punto del espacio de fase  $x \in P$  es suficiente si su trayectoria es suficiente, o sea, si contiene un segmento de trayectoria suficiente.

Si el segmento de trayectoria no es suficiente, la curvatura de  $\pi(S^b \tilde{\Gamma})$  en  $\pi(S^b x)$  necesariamente se anula en alguna dirección, formando el así llamado **subespacio neutro**. No es difícil ver que en el entorno de un punto suficiente los exponentes de Liapunov relevantes son no nulos.

Esas observaciones llevan al no trivial

**Teorema Fundamental de los Billares Semidispersores (Sinai- Chernov [17]).** *Si en un billar semidispersor se satisfacen las condiciones geométricas y de hiperbolicidad relacionadas con las singularidades (ver B1) - B4) en la Sección anterior), entonces todo punto  $x \in P$  suficiente tiene un entorno perteneciente a una componente ergódica del sistema (ver comentarios después de los Teoremas de Pesin).*

A partir de ese teorema Sinai y Chernov probaron en 1987 que el sistema formado por  $N = 2$  bolas en el  $d$ -toro es un sistema ergódico.

Los resultados posteriores fueron obtenidos por los matemáticos húngaros Krámli, Simányi y Szász, quienes entre 1990 y 1992 extendieron los resultados para tres y cuatro bolas en el  $d$ -toro, con  $d \geq 2$  (ver [4], [5]).

La obtención de resultados muy dependientes del número de bolas se debe a las dificultades para evitar el problema de la suficiencia al aumentar la dimensión del espacio en que se trabaja y el número de cilindros.

Esas dificultades fueron resueltas de manera parcial, muy recientemente por Simányi y Szász, introduciendo condiciones que permiten analizar cuan pequeño es el conjunto no suficiente (debe ser formado por la unión de un conjunto numerable de variedades de codimensión mayor o igual a dos). Esa pequeñez relativa permite obtener una rica estructura de colisiones entre las bolas, asegurando las condiciones a que se refiere el teorema fundamental.

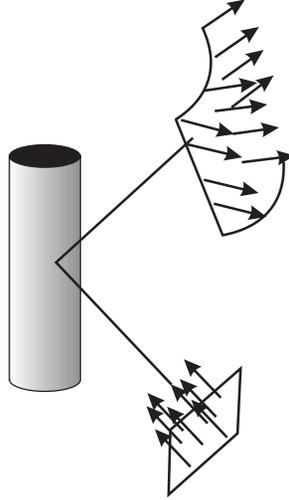


Figura 14: Obstáculo no estrictamente convexo: existen direcciones en que vectores del haz paralelo incidente no son dispersados.

Aún así no consiguieron probar el carácter ergódico del sistema de bolas elásticas, sino sólo la anulación de los exponentes de Liapunov en casi todo punto. Eso permite aplicar la segunda parte de los teoremas de Pesin, y deducir que existe un conjunto numerable de componentes ergódicas. Los autores comentan: “Our methods so far do not give the expected global ergodicity.”.

Como novedad, estudian el movimiento de bolas de masas distintas  $\vec{m} = (m_1, m_2, \dots, m_N) \in \mathbb{R}_+^N$ , al contrario de todos los trabajos anteriores y de la hipótesis original, donde las masas son iguales. Pero el teorema resultante vale para  $\vec{m}$  fuera de un conjunto pequeño de  $\mathbb{R}_+^N$  ( $\vec{m}$  no pertenece a una unión numerable de subvariedades analíticas propias).

Simányi y Szász probaron el siguiente

**Teorema [12].** Si  $N \geq 2$ ,  $d \geq 2$ ,  $\vec{m} = (m_1, m_2, \dots, m_N) \in \mathbb{R}_+^N$  son las masas de las bolas y sus radios son iguales a  $r = R_0(N, d)$  entonces el sistema  $(M, S, \mu)_{\vec{m}}$  tiene todos sus exponentes de Liapunov relevantes no nulos si  $\vec{m}$  está fuera de un conjunto “pequeño” de  $\mathbb{R}_+^N$ .

Este resultado fue mejorado en trabajos más recientes de Simányi, quien probó que todo sistema de bolas duras en un toro de cualquier dimensión es hiperbólico y que en el toro de dimensión 2, casi todos estos sistemas son ergódicos.

Los últimos progresos en la dirección de probar la ergodicidad del modelo de Boltzmann-Sinai ha llevado a la utilización de herramientas refinadas de geometría algebraica. Chernov, Simányi y Szász (y sus alumnos) continúan trabajando en esos temas. Recientemente han aparecido nuevas complicaciones porque se ha probado que la curvatura de las variedades invariantes puede no ser acotada como hasta entonces se había aceptado.

## Conclusión

Hoy en día no es clara la relevancia **para la Física** de la Hipótesis Ergódica y de los resultados matemáticos al respecto. Sin embargo, la discusión de su contenido y su presentación rigurosa tuvo y tiene un impacto muy grande tanto en la Física como en la Matemática. Contribuyeron de manera fundamental para el desarrollo de la Mecánica Estadística y de la Teoría Ergódica de los Sistemas Dinámicos y, de manera general, para el estudio formalizado de los movimientos caóticos.

**Agradecimientos.** Con Dómokos Szász y Nikolai Chernov mantuve interesantes conversaciones sobre estos temas. Varias partes de estas notas están inspiradas en el artículo [18]. Sylvie Oliffson, Sônia

Pinto y Nikolai Chernov colaboraron con la confección de varias de las figuras.

## Referencias

- [1] D. Burago, S. Ferleger & A. Kononenko, *Uniform estimates on the number of collisions in semi-dispersing billiards*, Ann. of Math. **147** (1998), 695–708.
- [2] N Chernov & R. Markarian, *Introduction to the ergodic theory of chaotic billiards*, IMCA, Lima, 2001.
- [3] I. P. Cornfeld, S. V. Fomin & Ya. G. Sinai, *Ergodic Theory*, Springer, Berlín, 1982.
- [4] A. Krámli, N. Simányi & D. Szász, *The K-Property of Three Billiard Balls*, Annals of Mathematics, **133** (1991), 37–72.
- [5] A. Krámli, N. Simányi & D. Szász, *The K-Property of Four Billiard Balls*, Commun. Math. Phys., **144** (1992), 107-142.
- [6] R. Mañé, *Introducao à Teoria Ergódica*, IMPA, Rio de Janeiro, 1986.
- [7] R. Mañé, *Global Variational Methods in Conservative Dynamics*, IMPA, Rio de Janeiro, 1991.
- [8] R. Markarian, *Billiards with Pesin region of measure one*, Comm. Math. Phys. **118** (1988), 87–97.
- [9] R. Markarian, *Introduction to the ergodic theory of plane billiards* en Dynamical Systems. Santiago de Chile 1990, R. Bamón, R. Labarca, J. Lewowicz, and J. Palis, editors, Longman, 1993, 327-439.
- [10] R. Markarian, *New ergodic billiards: exact results*, Nonlinearity, **6** (1993), 819–841.
- [11] R. Markarian, S. Oliffson & S. Pinto, *Chaotic properties of the elliptical stadium*, Commun Math. Phys., **174** (1996), 661–679.
- [12] N. Simányi & D. Szász, *Hard ball systems are completely hyperbolic*, Ann. of Math. **149** (1999), 35–96.
- [13] N. Simányi, *Ergodicity of hard spheres in a box*, Ergodic Theory Dynam. Systems **19** (1999), 741–766.
- [14] 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
- [15] Ya.G. Sinai, *Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards*, Russ. Math. Surv. **25** (1970), 137–189.
- [16] Ya.G. Sinai, *Development of Krylov's ideas*, Afterwards to N.S. Krylov, *Works on the foundations of statistical physics*, Princeton Univ. Press, 1979, 239–281.
- [17] 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.
- [18] D. Szász, *Boltzmann's ergodic hypothesis, a conjecture for centuries?*, Studia Sci. Math. Hung. **31** (1996), 299–322.
- [19] D. Szász, Editor *Hard Ball Systems and the Lorentz gas*, Springer, Berlin, 2000.

Instituto de Matemática y Estadística “Prof. Ing. Rafael Laguardia”  
IMERL, Facultad de Ingeniería, Universidad de la República  
C.C. Nro. 30, Montevideo, Uruguay  
c-electrónico: roma@fing.edu.uy; Fax: (598-2)-711 5446