MATEMÁTICA UNIVERSITÁRIA
nº29 – dezembro 2000 – pp. 45–74

Bilhares. A Hipótese Ergódica de Boltzmann

R. Markarian

Muchos nombres de uruguayos están directamente vinculados a mi reencuentro con la matemática luego de la dictadura que asolara nuestro país. Jacob dió un impulso en el momento preciso y fuerza intelectual y moral a mi avance en ese sentido.
Los matemáticos uruguayos apreciamos enormemente sus esfuerzos y logros en la consolidación de ciencia de calidad en el continente.

1 Introdução

A Hipótese Ergódica – formulada há mais de cem anos pelo físico alemão Ludwig Boltzmann – é parte de um modelo mecânico para explicar as propriedades dos gases. O estudo das dinâmicas desordenadas no bilhar está diretamente relacionado com esta hipótese.

Este trabalho se propõe a expor de forma simples essa relação, começando com uma descrição dos bilhares planos. A seguir são dadas algumas razões pelas quais é interessante o estudo da dinâmica dos bilhares. Na seção 3 são analisados com mais detalhes os bilhares planos e são dados exemplos relacionados aos seus comportamentos.

Nas seções 4 a 7 segue-se a evolução cronológica que vai da formulação física à formulação matemática da Hipótese de Boltzmann. Na seção 4 é dada a formulação física da Hipótese e na seção 5 se colocam elementos de sua formulação matemática. A seção 7 contém um resumo da formulação do modelo dos gases através dos bilhares.

Nas seções 5 a 8 tenta-se mostrar como a formulação matemática da Hipótese se relaciona com o desenvolvimento das áreas da matemática

conhecidas na atualidade pelos nomes Teoria Ergódica, Transformações Contínuas com Singularidades e, em particular, Teoria dos Bilhares. A dificuldade na colocação dos problemas aumenta a partir da seção 7. Na última seção resumimos e comentamos os principais resultados exatos recentes vinculados à formulação matemática da Hipótese de Boltzmann.

2 A beleza dos bilhares

Comecemos dando a definição de bilhares planos para que assim tenhamos um objeto matemático ao qual possamos nos referir.

Todo fenômeno que evolui no tempo pode ser considerado como um **sistema dinâmico**. Diremos que um tal sistema é determinístico se, conhecido o estado do sistema em um tempo inicial, as regras da evolução futura (e passada) são conhecidas com precisão. Um **bilhar plano** é o sistema dinâmico que descreve o movimento de uma partícula pontual em um conjunto compacto conexo $\mathbf{Q} \subset \mathbb{R}^2$ (ou no toro $\mathbf{T}^2$), cuja fronteira é a união de um número finito de curvas regulares, por exemplo C^3. Dentro de $\mathbf{Q}$ o movimento é uniforme (velocidade constante) e a reflexão na fronteira $\partial\mathbf{Q}$ é elástica (ângulo de saída igual ao de entrada). Lembro que o toro $\mathbf{T}^2$ no espaço $\mathbb{R}^3$ – a superfície exterior de uma rosquinha de padaria – pode ser obtido a partir de um quadrado de papel, colando (identificando) os lados paralelos. Na figura 1 está representado o movimento de uma partícula em $\mathbf{T}^2$ com um obstáculo circular: quando a partícula chega ao bordo do quadrado, ou podemos fazê-la tabelar com choque elástico (como se fosse um bilhar quadrado com obstáculo circular), ou, como está feito na figura, a fazemos reaparecer no lado oposto. Esta última representação corresponde ao movimento no toro onde o único obstáculo é o círculo.

Como o movimento é uniforme dentro de $\mathbf{Q}$, o sistema do bilhar se torna determinado por seus sucessivos choques com $\partial\mathbf{Q}$ e, no lugar do fluxo a tempo contínuo, podemos considerar a transformação que a cada ponto da fronteira e a cada vetor de saída faz corresponder os do próximo choque (se este estiver bem definido). Assim, a cada ponto $(q_0, v_0) \in \partial\mathbf{Q} \times \mathcal{S}_1$ corresponde um ponto (q_1, v_1) no mesmo espaço produto ($\mathcal{S}_1$ é a circunferência de raio um). Ver figura 2.

Os arcos da fronteira são ditos **dispersores** se são convexos quando vistos de dentro da mesa de bilhar (como a circunferência da figura 1),

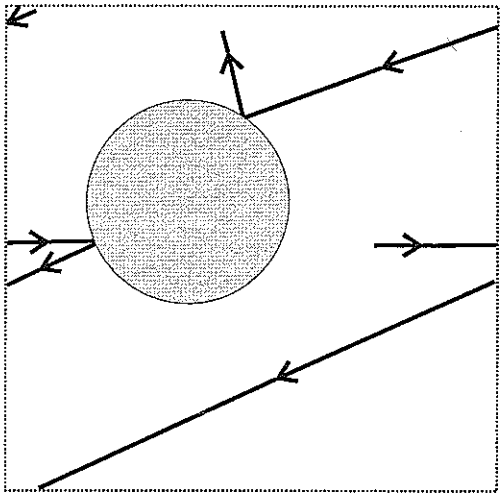

Figura 1: Bilhar no toro.

são ditos **focalizadores** se são côncavos quando vistos de dentro da mesa de bilhar (como uma circunferência vista de dentro) e são ditos neutros se são segmentos de reta. Chamamos de dispersores e focalizadores porque um feixe de trajetórias paralelas é "dispersado" ou "focalizado", respectivamente, ao se chocar com cada um deles. Por favor pegue lápis e papel e faça um desenho com arcos de cada um desses tipos de feixes paralelos incidindo sobre eles.

Em geral os bilhares são sistemas dinâmicos que correspondem ao movimento uniforme de um ponto material sobre uma variedade riemanniana, com choque elástico na fronteira. Portanto são fluxos geodésicos em variedades com bordos. As características específicas dos bilhares aparecem quando o papel da fronteira (curvatura, posição relativa, etc.) é muito mais importante que o da variedade subjacente.

Nos últimos trinta anos a Teoria dos Bilhares tem se desenvolvido principalmente devido às seguintes razões:

1. Algumas classes de bilhares apresentam um forte comportamento caótico e podem ser considerados entre os melhores exemplos de caos determinístico;
2. Muitos exemplos interessantes de sistemas dinâmicos de origem física

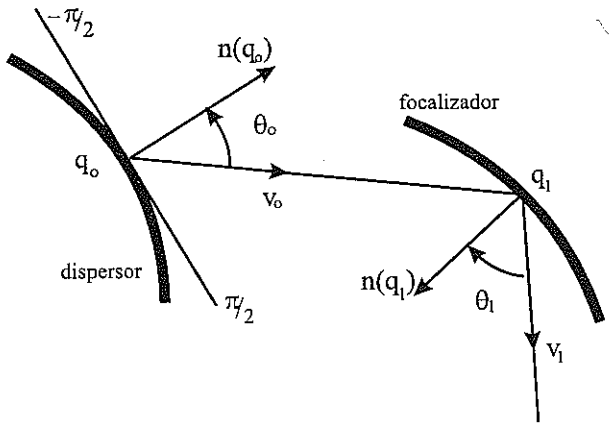

Figura 2: Espaço de configuração do bilhar. Se $x_0 = (q_0, v_0)$, então $t(x_0) = \text{distância}\,(q_0, q_1)$.

(especialmente aqueles em que a interação entre partículas envolve choques elásticos) podem ser reduzidos a bilhares. Alguns destes exemplos serão estudados a partir da seção 4;
3. Importantes problemas na teoria do caos quântico envolvem uma análise profunda dos bilhares clássicos;
4. O estudo dos bilhares sugere muitos problemas bonitos e interessantes em geometria e probabilidade.

A seguir nos referiremos a alguns destes problemas. Apresentaremos os resultados para bilhares planos, mas estes são válidos, com as devidas correções, em bilhares de qualquer dimensão.

Consideremos dois arcos de curvas dispersoras que se cruzam em um ponto V, formando um vértice da mesa de bilhar. Existem trajetórias que vão direto ao vértice. Qualquer outra trajetória que entra no "ângulo" do vértice V começará a sair em algum momento: o número de tabelas perto do vértice é finito. Mais ainda, se as curvas dispersoras não são tangentes, o número de tabelas na região do vértice tem uma cota superior finita. Ou seja, dado um número pequeno ε, existe um número C (que depende do ângulo entre ambas as curvas, de suas curvaturas e de ε) tal que toda trajetória que passa a uma distância menor

que ε do vértice, acabará saindo após no máximo C tabelas. Se as curvas são tangentes, em geral existem trajetórias que se mantêm próximas do vértice por tantas tabelas quantas queiramos.

É também interessante calcular, para qualquer bilhar, a média dos comprimentos t dos segmentos de trajetória entre duas tabelas. Ou seja, tomamos todos os possíveis pontos de partida e, com algum critério razoável, efetuamos a média dos comprimentos até a próxima tabela. O resultado desta **trajetória livre média** (mean free pass), que será definida rigorosamente mais tarde, é

$$\tau_m = \pi \frac{\text{área da mesa de bilhar}}{\text{comprimento do bordo do bilhar}} .$$

3 Bilhares planos

Fixemos uma origem na fronteira a partir da qual mediremos seu comprimento; sejam $\partial \mathbf{Q}_i$ as componentes regulares da fronteira (que se cortam nos "vértices", formando os "ângulos" do bilhar) e $n(q)$ o versor normal interior a $\partial \mathbf{Q}_i$ em $q \in \partial \mathbf{Q}_i$. Então, o **espaço de fase** onde está definido o sistema dinâmico pode ser parametrizado pelo comprimento de arco s e pelo ângulo θ de saída entre $n(q)$ e o versor v de saída. Para simplificar a compreensão, suponhamos inicialmente que a mesa não tenha obstáculos em seu interior e que o comprimento total de $\partial \mathbf{Q}$ seja L. Então o espaço de fase está contido no retângulo $[0,\ L] \times [-\pi/2,\ \pi/2]$ ou, mais exatamente, no cilindro $\mathcal{R}$ que obtemos se identificamos de maneira natural as posições $s = 0$ com $s = L$. É neste espaço que representaremos o movimento de uma partícula material se chocando contra os bordos da **mesa de bilhar**. É claro que o que se representa são os choques e que os **trechos de trajetórias** (segmentos entre um choque e outro) só serão visíveis no **espaço de configurações**, isto é, na própria mesa de bilhar.

Seja S a **transformação do bilhar** tal que $S(s_0, \theta_0) = (s_1, \theta_1)$, onde s_0 e s_1 são as coordenadas dos pontos q_0 e q_1 de saída e chegada (respectivamente) na fronteira e θ_0 e θ_1 são os ângulos de saída das trajetórias em q_0 e q_1 respectivamente. A função S (ou sua inversa) não está bem definida se q_1 ou q_0 está em um vértice da fronteira e é descontínua nos pontos em que a trajetória tangencia o bordo. Observa-se que a derivada de S não é limitada nestes pontos de tangência.

Com efeito, se $\tilde{x}_1 = (\tilde{q}_1, \tilde{v}_1) = T(\tilde{x}_0)$ está definida para $\tilde{x}_0 = (\tilde{q}_0, \tilde{v}_0)$ então para todo $x_0 = (q_0, v_0)$ em uma pequena vizinhança de $\tilde{x}_0$ a matriz derivada nas coordenadas (s, θ) é dada por (ver [6]):

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

onde $K_i = K(x_i)$, $i \in \mathbb{N}$ são as curvaturas de ∂Q em q_i e t_0 é a distância entre q_0 e q_1, ou seja, o comprimento do trecho de trajetória entre x_0 e x_1. Portanto os coeficientes desta matriz tendem a infinito quando a imagem de (q_0, v_0) por S tende a tangenciar ∂Q (isto é, quando θ_1 se aproxima de $\pm\pi/2$).

Os segmentos verticais determinados pelos vértices da fronteira, os pontos de tangência e as imagens e pré-imagens por S de todos estes segmentos se denominam **singularidades** de S e nós as representamos por $\mathcal{D}$.

Com essas restrições a transformação S é diferenciável, com inversa diferenciável (difeomorfismo C^2) no conjunto $\mathcal{M}$ que é obtido quando retiramos $\mathcal{D}$ de $\mathcal{R}$.

Além disso, essa transformação preserva a medida $d\mu = c\cos\theta ds d\theta$ ($c = 1/2L$ é uma constante de normalização). Isto significa que se $\mu(A) = \int_A c\cos\theta ds d\theta$, então $\mu(A) = \mu(S^{-1}A)$ para todo conjunto boreliano A: a μ-medida dos conjuntos borelianos do espaço de fase não varia quando os "movemos" por S. É com respeito a esta medida padrão que fazemos todas as médias e estudos probabilísticos dos bilhares. Em particular, a trajetória livre média, cuja definição foi dada no fim da seção anterior, é calculada da seguinte maneira: $\tau_m = \int_{\mathcal{M}} t(x) d\mu(x)$ onde $t(x)$ é o comprimento da trajetória entre o ponto $x = (s, \theta)$ e Sx (entre uma tabela e a seguinte).

Todas essas características da transformação do bilhar em regiões limitadas, conexas e cujos bordos são formados por finitas curvas C^3 no plano, são válidas também – com as devidas adaptações – para os bilhares em $\mathbb{R}^d$ ou em $\mathbb{T}^d$, o toro de dimensão d para qualquer $d \geq 2$.

Vejamos alguns exemplos de bilhares planos.

Consideremos em primeiro lugar a mesa de bilhar cujo bordo é uma circunferência. Sabe-se que se uma partícula sai com ângulo θ_0, todos os choques ocorrerão com este mesmo ângulo e portanto a representação

dos pontos desta trajetória no espaço de fase estará restrita a um segmento de reta horizontal. Os trechos de trajetória entre um choque e outro terão todos mesmo comprimento, e serão tangentes a uma mesma circunferência. O leitor interessado pode tentar descobrir quando haverá trajetórias periódicas (trajetória que volta a passar no ponto inicial) e quando a trajetória cobre densamente o segmento de reta horizontal do espaço de fase.

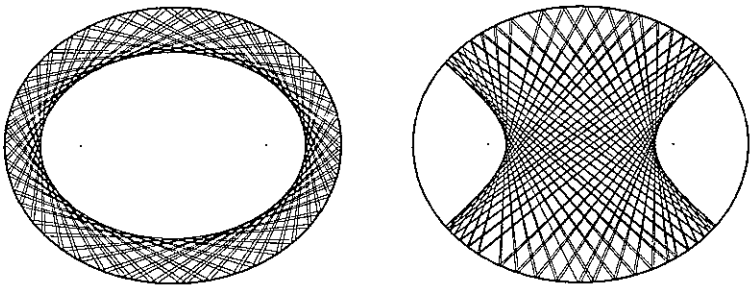

Figura 3: Cáusticas elípticas e hiperbólicas

Se a mesa de bilhar tem como bordo uma elipse de semieixo vertical com comprimento um e semieixo horizontal com comprimento $a > 1$, a situação se complica um pouco e haverá essencialmente dois tipos de trajetórias. Se um trecho de trajetória corta o eixo maior entre um foco e o vértice mais próximo, esta propriedade continuará se verificando para os outros trechos de trajetórias que cortem esse eixo. Além disso, todos os trechos de trajetórias serão tangentes a uma elipse interior. Diz-se que esta elipse, que tem os mesmos focos que a do bordo do bilhar - confocal - é uma **cáustica** do bilhar (observe que se $a = 1$ a elipse é uma circunferência e cada uma dessas cáusticas também). Ver figura 3.

Se um trecho de uma trajetória corta o eixo maior entre os dois focos, todos os outros trechos desta trajetória também o farão, e as retas que contêm estes trechos serão tangentes (em pontos dentro ou fora da mesa elíptica) a uma cáustica hiperbólica.

Uma pergunta surge naturalmente: o que acontece com as tajetórias que passam pelos focos? A resposta é ainda mais simples que nos casos anteriores: logo após se chocar com a elipse a partícula se dirigirá ao outro foco e assim sucessivamente. Estas trajetórias que passam pelos

focos "separam" os comportamentos anteriores (com cáusticas elípticas e hiperbólicas) e são muito importantes no estudo da dinâmica deste bilhar e de outros relacionados a este. Todos os resultados sobre cáusticas confocais e trajetórias que passam por focos são consequências do chamado Teorema de Poncelet da Geometria Projetiva.

Se representarmos algumas trajetórias de cada tipo no espaço de fase veremos que as mais parecidas com as da circunferência são as do primeiro tipo (cáusticas elípticas) e que as do segundo tipo estão contidas em curvas em torno de dois pontos do espaço de fase que correspondem à trajetória periódica que se move ao longo do eixo menor. As trajetórias que passam pelos dois focos são representadas sobre duas curvas ("separatrizes" dos dois movimentos principais) que unem os dois pontos periódicos das trajetórias que estão sobre o eixo maior. Ver figura 4.

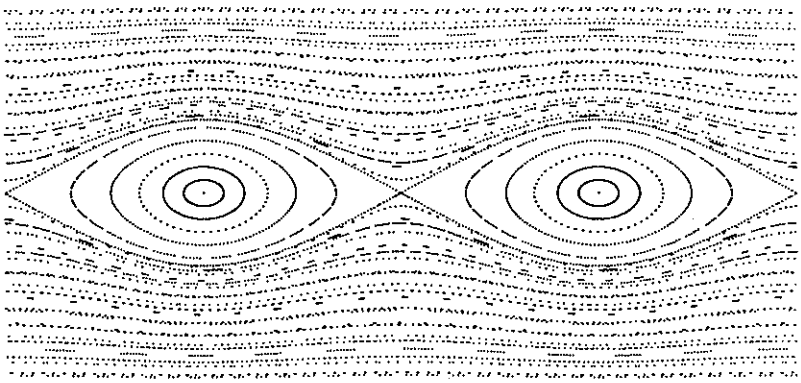

Figura 4: Espaço de fase do bilhar elíptico

No bilhar elíptico o espaço de fase não tem singularidades "visíveis"; não há vértices do bilhar porque a fronteira é uma única curva regular, e as trajetórias tangentes não existem dentro do bilhar, porque estão restritas aos pontos $\theta = \pm\pi/2$ do espaço de fase. Isto quer dizer que o conjunto $\mathcal{D}$ das singularidades é formado por apenas dois segmentos e que o espaço $\mathcal{M}$ é um cilindro com altura π e sem os bordos superior e inferior.

Agora modifiquemos um pouco a situação cortando a elipse ao longo

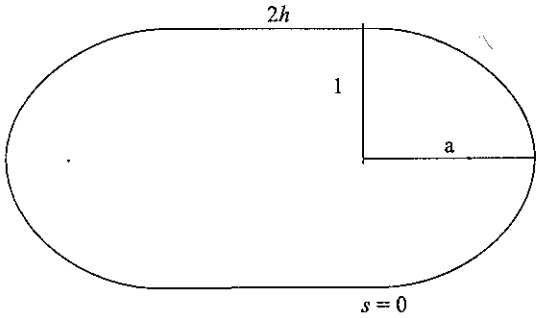

Figura 5: O estádio elíptico

do eixo menor, separando estas duas metades e colando as pontas livres com dois segmentos de comprimento $2h$. Teremos o *estádio elíptico* da figura 5. Se $a = 1$ as curvas dos extremos são semicircunferências e o estádio é dito de Bunimovich, por ser este o nome do primeiro matemático que o estudou em detalhe.

De qualquer forma, novamente as trajetórias tangentes não aparecem dentro do espaço de fase do bilhar, só que agora temos quatro "vértices" nos pontos onde as pontas da elipse cortada tocam os segmentos de reta. Nestes pontos os bordos têm derivada contínua, mas a curvatura (que está diretamente relacionada com a segunda derivada) é descontínua. Então, se o comprimento da meia elipse é m, o comprimento total do bordo é $L = 2m + 4h$ e se começamos a medir o comprimento de arco a partir de um dos (antigos) vértices do semieixo menor da elipse, o espaço de fase será como na figura 6. Nela representamos ainda o conjunto dos pontos cujas imagens por S estão no ponto $s = 0$ (o ângulo θ é medido a partir da normal interior, em sentido horário). O aspecto de cada curva varia segundo o valor de a (comprimento do semieixo maior). Aqui representamos o caso em que a é muito próximo ou igual a 1. O leitor interessado pode tentar ver como essa curva se modifica quando a aumenta.

Essa curva e outras muito semelhantes a ela, que são as pré-imagens e imagens por S das tangências e são obtidas por diversas simetrias, formam parte do conjunto $\mathcal{D}$ das singularidades. Então o conjunto $\mathcal{M}$, que originalmente parecia ser um simples cilindro, ficou dividido em várias

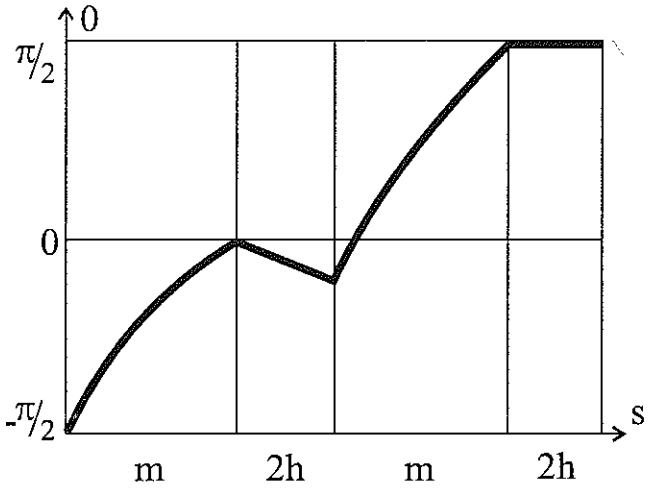

Figura 6: Espaço de fase do estádio elíptico. São indicadas as curvas de pontos cujas imagens estão no vértice inferior do cixo menor da semielipse da direita.

partes cujos bordos são essas curvas. A maneira como elas se cortam e sua distribuição no cilindro original são assuntos que se encontram entre os estudos mais avançados e interessantes dos bilhares.

No caso do estádio elíptico as trajetórias são muito mais complicadas que no caso da elipse. Há infinitas trajetórias com período dois entre os lados paralelos; há trajetórias que passam de uma meia elipse para a outra, mudando o sentido de sua circulação. O comportamento geral depende muito dos valores de a e h. Isto será visto com um pouco mais de detalhe na seção 7. Na figura 7 estão representadas 150000 iterações de uma única trajetória para o bilhar com $a = 1.24, h = 1.04$. No trabalho [8] são estudadas muitas propriedades do estádio elíptico.

4 Hipótese de Boltzmann

Em 1964 Werner Heisenberg afirmou que "Um físico teórico se sente melhor se não há objetos matemáticos rigorosamente definidos por trás de suas considerações.". Seguramente neste momento tinha em mente

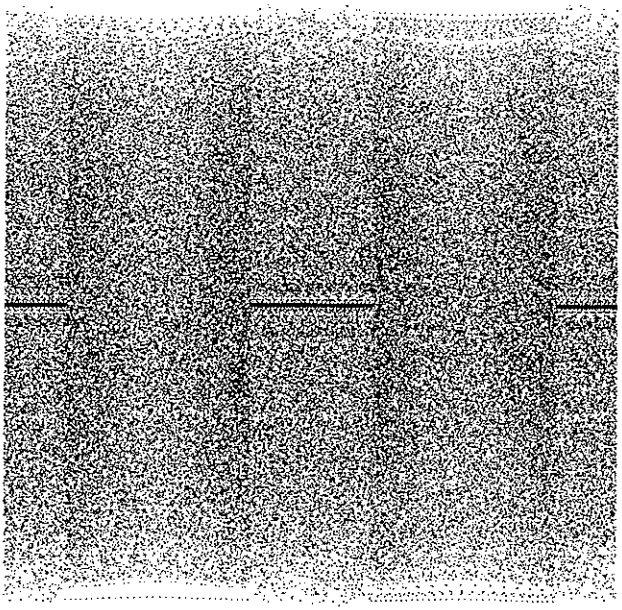

Figura 7: Espaço de fase do estadio elíptico. 150000 iterações de uma única condição inicial.

os primeiros anos da Mecânica Quântica, mas a frase pode igualmente ser aplicada à obra de Ludwig Boltzmann. E não apenas à sua Hipótese Ergódica, sobre a qual versa parte substancial do resto deste artigo, como também a outras áreas por ele estudadas.

Vale também recordar que, motivado pelas equações de Boltzmann, David Hilbert incluiu em sua célebre coletânea de 23 problemas (apresentados no Congresso Internacional de Matemática que teve lugar em Paris, em 1900) o denominado "Tratamento Matemático dos Axiomas da Física". Sobre este escreveu: "é muito desejável que a discussão dos fundamentos da Mecânica seja tomada também por matemáticos. Assim o trabalho de Boltzmann sobre os princípios da Mecânica sugere o problema do desenvolvimento matemático dos processos limites ali meramente indicados, que levam da visão atomística às leis do movimento contínuo".

Entre 1870 e 1884 Boltzmann usou várias formas da Hipótese Ergó-

dica. [1] Uma formulação avançada seria:

Hipótese Ergódica de Boltzmann. *Para grandes sistemas de partículas interagindo em equilíbrio, as médias temporais estão próximas das médias espaciais.*

Acho que é o caso de fazer alguns comentários sobre essa formulação. Por "grandes sistemas de partículas interagindo em equilíbrio" entende-se um sistema com muitas partículas que não recebe influências externas; por exemplo, um número muito grande de partículas que se chocam com as paredes completamente rígidas de uma caixa e entre si, sem a incidência de outros fatores. As "médias temporais" são as médias dos valores observados (medições) de uma função numérica, com o passar tempo; este medido em alguma unidade que corresponda à escala do problema. Se o modelo matemático do fenômeno é descrito por uma equação diferencial, sua solução será dada por um fluxo e a evolução do tempo será contínua; mas se o modelo é dado por uma transformação (uma função) a evolução do tempo não é contínua, é discreta. Neste artigo trabalharemos com este último tipo de modelo porque o tempo será medido pelo número de vezes que se aplica a transformação estudada (S^n indica a aplicação sucessiva da transformação S, n vezes, n aqui é o "tempo"). Então somamos os valores medidos no decorrer de um longo tempo T, e dividimos por T (que é um número inteiro: $T \in \mathbb{Z}$). As "médias espaciais" consistem de médias (integrais) de medições (ou registros) simultâneas em todos os pontos com respeito a uma medida com um sentido físico no espaço de fase. As funções medidas podem ser, por exemplo, a temperatura ou a pressão.

Lembremos que o conceito abstrato de medida generaliza os de área e volume, e se caracteriza pela aditividade: a medida da união de conjuntos disjuntos é a soma das medidas de cada um deles. Mais precisamente uma **medida de probabilidade** μ sobre uma σ-álgebra $\mathcal{A}$ de subconjuntos do espaço M no qual trabalhamos é uma função $\mu : \mathcal{A} \to [0,1]$ tal que $\mu(M) = 1$, $\mu(\emptyset) = 0$ e $\mu(\cup A_i) = \sum \mu(A_i)$ se $A_i \cap A_j = \emptyset$ para $i \neq j$. Diremos que os conjuntos de $\mathcal{A}$ são **mensuráveis** e que uma propriedade se verifica μ-quase todo ponto, se o conjunto dos pontos que a verificam tem medida total (um em nosso caso).

[1] A palavra ergódica, usada neste contexto, provém do grego *ergon* (trabalho) e *odos* (caminho, passo).

Observação: os físicos entendem as médias espaciais como médias de equilíbrio (microcanônico), ou seja, com respeito à medida de Liouville na subvariedade do espaço de fase determinada pelos invariantes triviais do movimento. É desta medida que falamos no comentário anterior.

Introduzidos esses conceitos matemáticos, podemos fazer uma apresentação mais precisa da Hipótese Ergódica. Ela estabelece que se f é uma medição – uma função no espaço de fase do sistema – e o tamanho do sistema – o número N de partículas – tende a infinito, então

$$\frac{1}{T}\sum_{n=0}^{T-1} f(S^n x) \to \int_M f(x)d\mu(x) \tag{2}$$

onde μ é a medida de equilíbrio e $S^n x$ é a evolução temporal do ponto x no espaço de fase.

Vemos imediatamente que f e μ também dependem de N, e portanto, para uma formulação matemática rigorosa deveríamos especificar o significado da convergência em (2). Vamos em busca desse significado.

5 Encontrando um objeto e um problema matemáticos (de Boltzmann a Von Neumann: de 1870 a 1931)

Levou muito tempo para que o objeto matemático da Hipótese Ergódica fosse encontrado. Só em 1929 Koopmam começou a investigar grupos de transformações que preservam medidas, ou seja, transformações que levam cada conjunto em outro que mede o mesmo. Esses progressos foram precedidos pelos êxitos da Teoria da Medida, a qual, em outro sentido, permitiria a Kolmogorov em 1933 estabelecer os fundamentos axiomáticos da Teoria da Probabilidade.

Mais precisamente, seja M um espaço abstrato, o espaço de fase do sistema, e μ uma medida de probabilidade sobre uma σ-álgebra de M. A dinâmica é dada pela aplicação sucessiva de uma transformação S (ou sua inversa S^{-1}), que preserva medida. Isto significa que para todo conjunto mensurável $A \subset M$, e para todo $n \in \mathbb{Z}$, $\mu(S^n A) = \mu(A)$.

Finalmente, seja $f : M \to \mathbb{R}$ uma função (o observável) razoavelmente regular, por exemplo, pertencendo ao espaço das funções quadrado integráveis: $\int f^2(x)d\mu(x) < \infty$; neste caso diz-se que $f \in L_2(\mu)$.

Dessa forma o objeto matemático – (M, S, μ), com as funções f – está definido.

No mesmo ano (1929) Von Neumann provou o primeiro teorema ergódico, o qual estabelece que a média temporal de uma função f ao longo da trajetória que passa por um ponto x tem limite $\tilde{f}(x)$:

Teorema Ergódico. *Quando* $T \to \infty$,

$$\frac{1}{T}\sum_{n=0}^{T-1} f(S^n x) \to \tilde{f}(x) \ , \quad \textit{sendo a convergência no sentido } L_2(\mu).$$

Em 1931 Birkhoff e Khincin separadamente provaram a mesma fórmula, mas com convergência para quase todo x no sentido da medida μ, ou seja, que essa convergência se dá para todos os pontos de M exceto aqueles em um conjunto com μ-medida nula. A função limite $\tilde{f}(x)$, que normalmente é chamada de média de Birkhoff, é tal que sua integral (sobre todo o espaço) tem mesmo valor que a de f : $\int_M \tilde{f}(x)d\mu(x) = \int_M f(x)d\mu(x)$.

Dado que o termo que depende de T é exatamente a média temporal citada na Hipótese de Boltzmann, esta passou a ter um significado preciso logo depois de formulado o Teorema Ergódico: nos grandes sistemas de partículas interagindo em equilíbrio, as médias temporais $\tilde{f}(x)$ não dependem de x. Observe que, como μ é uma medida de probabilidade, se $\tilde{f}(x)$ for constante, quando integrarmos com respeito a μ obteremos o mesmo valor $\tilde{f}(x)$ e, pela observação final do parágrafo anterior, resulta $\tilde{f}(x) = \int_M \tilde{f}(x)d\mu(x) = \int_M f(x)d\mu(x)$.

A propriedade de que em um sistema as médias de Birkhoff não dependem do ponto inicial x foi considerada muito importante desde o começo, a tal ponto que os sistemas (M, S, μ) que verificam esta propriedade receberam o nome de ergódicos.

Definição: Ergodicidade.O sistema (M, S, μ) é ergódico se para toda função razoavelmente regular f, por exemplo, para toda $f \in L_2$, temos $\tilde{f}(x)$ constante para μ-quase todo ponto $x \in M$.

Esta definição é equivalente a cada uma das três seguintes afirmações:

Se f é razoavelmente regular e constante ao longo das órbitas ($f \circ S^n = f$, para todo $n \in \mathbb{Z}$), então f é constante para μ-quase todo ponto $x \in M$.

Se um conjunto $A \subset M$ é invariante por S^n ($S^n A = A$), então A é quase todo o espaço M ou tem medida nula ($\mu(A) = 0$ ou 1).

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

Resumindo: temos um modelo matemático (grupo de transformações que preservam medidas), a noção de ergodicidade e, finalmente, o problema de estabelecer a ergodicidade de um sistema que pode ser interessante do ponto de vista da mecânica.

Esses progressos deram origem a uma nova área da Matemática, a **Teoria Ergódica**, e a várias subáreas. Uma dessas estuda generalizações dos teoremas ergódicos; outra, formas mais fortes de estocasticidade (mixing, sistemas de Bernoulli, decaimento de correlação; ver, por exemplo, [4], Ch 2); uma área especial – particularmente importante para este trabalho – estuda a ergodicidade de certos sistemas, dentre os que surgem da Mecânica; outra, os problemas de isomorfismos e classificação dos sistemas dinâmicos, etc.

Devemos esclarecer que os sistemas de partículas cujos movimentos são descritos pela transformação S da seção 4 correspondem à classe mais geral dos chamados **fluxos hamiltonianos em variedades simpléticas**. Acreditava-se que poderíamos aplicar a Hipótese de Boltzmann para tais fluxos. Mas existe um exemplo célebre de Michel Herman que mostra a existência de uma quantidade muito grande de tais fluxos que não são ergódicos (ver [5]). Esse exemplo foi apresentado no final dos anos 80, mas já anteriormente, as tentativas de provas para a Hipótese Ergódica se restringiam aos modelos de bolas com choques elásticos.

6 Provando os primeiros teoremas relevantes (de Von Neumann a Sinai de 1931 a 1970)

Os métodos para estabelecer a ergodicidade de sistemas mecânicos vieram de outras áreas da Matemática, em particular da Teoria de Sistemas Dinâmicos. Em 1938-39 Hedlund e Hopf encontraram um método para demonstrar a ergodicidade de fluxos geodésicos em variedades compactas de curvatura negativa (trata-se de movimentos em inércia sobre, por exemplo, superfícies que resultam de girar uma hipérbole equilátera em torno de uma assíntota). A principal descoberta foi o, assim chamado,

comportamento hiperbólico dos sistemas dinâmicos, que implicava a ergodicidade de tais sistemas.

Hiperbolicidade significa a existência de curvas transversais sobre as quais o sistema dinâmico atua expandindo (variedade instável) ou contraindo (estável). Se a variedade não é uma superfície, ou seja, se tem dimensão maior que dois, as curvas são subvariedades cuja soma das dimensões é igual a dimensão da variedade original. Hiperbolicidade implica instabilidade para todas as órbitas: trajetórias que começam arbitrariamente próximas umas das outras se separam no futuro ou no passado; chamamos esta última propriedade de **sensibilidade com respeito às condições iniciais**.

O exemplo mais simples de sistema hiperbólico é o famoso "cat" de Arnold (o nome não vem do gato que era usado para representá-lo, mas de "Continuous Automorphisms of the Torus"). Consideramos a aplicação T_A do toro bidimensional $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ em si próprio, definida pela matriz $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. Observamos que a transformação expande em uma direção e contrai em outra, sendo estas direções as correspondentes aos vetores próprios com autovalores respectivamente $\lambda_u > 1$ e $\lambda_s < 1$ da matriz A.

Em 1942, pouco depois dos resultados fundamentais de Hedlund e Hopf, o físico russo N. S. Krylov mostrou que a instabilidade observada nos fluxos geodésicos também existia nos sistemas de bolas elásticas que modelam os gases. Esta descoberta e o avanço das idéias de Hedlund e Hopf na Teoria dos Sistemas Dinâmicos Hiperbólicos justificaram a versão de Sinai da Hipótese Ergódica de Boltzmann formulada em 1963 para o caso particular das bolas elásticas.

Hipótese Ergódica de Boltzmann-Sinai [11]. O sistema de N bolas elásticas sobre $\mathbb{T}^2$ ou $\mathbb{T}^3$ é ergódico para todo $N \geq 2$.

Como estes sistemas mecânicos têm muitas quantidades que se conservam ao longo das órbitas, a interpretação que devemos dar à conjectura é que teremos hiperbolicidade nas (componentes conexas das) subvariedades do espaço de fase definidas pelos invariantes do movimento. A primeira diferença conceitual que tem esta conjectura quando comparada com a formulação original de Boltzmann é que aqui não se faz nenhuma suposição sobre o número N de bolas. A ergodicidade (e outras propriedades estatísticas mais fortes) é prevista para qualquer

$N \geq 2$! De acordo com as idéias dos fundadores da Mecânica Estatística o estado de um sistema de partículas é aleatório porque qualquer sistema na natureza contém um número enorme destas: a aleatoriedade surge do grande número de graus de liberdade. Na abordagem de Sinai o número de partículas não tem nenhuma importância: um sistema de duas partículas elásticas se chocando apresenta comportamento estocástico.

Uma segunda característica importante no argumento de Sinai é que o conhecimento local de sistemas hiperbólicos leva a uma descrição global. A macrodinâmica surge da microdinâmica.

Em 1963 Sinai pensou ter provado o resultado. Escreveu: "Among the chief results of the present article should be counted the theorem of ergodicity of systems of hard pellets in a rectangular box". Mas só em 1970 [12] pode provar sua conjectura para o caso $N = 2$: discos movendo-se no toro $\mathbb{T}^2$.

7 O argumento de Sinai. Bilhares (1970)

Comecemos com um procedimento comum em Física e Matemática. Em vez de tratar o problema com N bolas consideremos uma só partícula em um espaço de fase com dimensão mais alta. Mais concretamente, suponhamos que o sistema é composto por $N \geq 2$ bolas, cada uma com massa unitária e raio $r > 0$, movendo-se em $\mathbb{T}^\nu$, o toro ν-dimensional ($\nu \geq 2$). No espaço de fase, a i-ésima bola é representada por $(q_i, v_i) \in \mathbb{T}^\nu \times \mathbb{R}^\nu$, onde q_i indica a posição do centro desta bola e v_i sua velocidade.

O espaço de configuração $\tilde{\mathbf{Q}}$ das N bolas em $\mathbb{T}^\nu$ é um subconjunto do toro de dimensão $N \cdot \nu$ pois, uma vez que as bolas são maciças e, portanto, seus centros devem estar a uma distância maior ou igual a duas vezes seus raios r, devemos retirar de $\mathbb{T}^{N\cdot\nu}$ os seguintes $\binom{N}{2}$ obstáculos cilíndricos:

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

$1 \leq i < j \leq N$. A energia $H = \frac{1}{2}\sum_1^N v_i^2$ e o momento total $P = \sum_1^N v_i$ são as integrais primeiras do movimento. Então, sem perda de generalidade, podemos supor que $H = \frac{1}{2}$, $P = 0$ e mais ainda, que também a soma das componentes espaciais $B = \sum_1^N q_i = 0$. Estas

condições são muito naturais do ponto de vista físico; algumas estão relacionadas com o estado de equilíbrio em que estamos considerando o fenômeno. Para estes valores de H, P e B, o espaço de fase e se reduz a $M := \mathbf{Q} \times \mathcal{S}_{d-1}$ onde

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

onde $d := \dim \mathbf{Q} = N \cdot \nu - \nu$ e $\mathcal{S}_k$ denomina a esfera unitária k-dimensional. A dimensão de $\mathbf{Q}$ é $N \cdot \nu - \nu$ porque a condição $\sum q_i = 0$ impõe ν restrições (uma para cada coordenada) aos pontos de $\tilde{\mathbf{Q}}$. A dimensão do espaço dos vetores é $N \cdot \nu - \nu$ porque $\sum v_i = 0$ e tais vetores estão na esfera unitária $(N \cdot \nu - \nu - 1)$-dimensional porque seu módulo (ao quadrado) $\sum v_i^2$ é constante. É fácil ver que a dinâmica das N bolas, determinada por seu movimento uniforme com colisões elásticas, e o fluxo do bilhar $\{S^t : t \in \mathbb{R}\}$ sobre $\mathbf{Q}$ com reflexões na fronteira $\partial\mathbf{Q}$ são isomorfos e conservam a medida de Liouville $d\mu = const \cdot dq \cdot dv$ (para detalhes ver [1], Ch 6).

Para explicar de maneira mais simples as principais características desses sistemas, nos restringiremos ao caso do bilhar com $N = 2$ e $\nu = 2$, ou seja, consideraremos duas bolas (discos) no toro $\mathbf{T}^2$. Então $\partial\mathbf{Q}$ é a união de um número finito de curvas regulares e os vetores v estão na circunferência $\mathcal{S}_1$.

Podemos dar mais um passo e supor que um dos discos está fixo e que só o centro do outro se move. Chegamos assim ao bilhar que analisamos na primeira seção, figura 1. Mais ainda, para não nos restringirmos ao toro, podemos considerar mesas de bilhar no plano que satisfazem as principais propriedades dinâmicas daquele bilhar de $\mathbf{T}^2$. Os bilhares planos que surgem do modelo com bolas duras para $N = 2$ e $\nu = 2$ são como os da figura 8. Sua ergodicidade foi provada por Sinai em 1970 e são chamados de **bilhares dispersores de Sinai** (ver a definição de arcos dispersores na seção 2).

É claro que podemos considerar bilhares planos cujas fronteiras não são geradas por condições do modelo das bolas duras e perguntar como devem ser as fronteiras para que a transformação do bilhar seja ergódica.

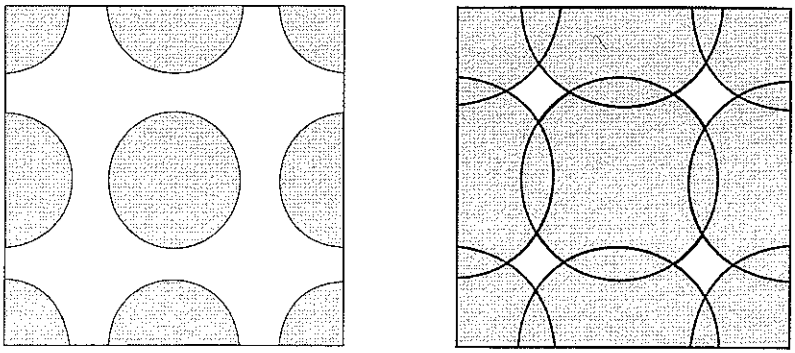

Figura 8: Bilhares dispersores. (a) Quadrado de lado 1 e raio $0 < r < 1/4$. (b) Quadrado de lado 1 e raio $1/4 \leq r < 1/2$

Em particular é interessante saber como "montar" uma mesa de bilhar para que seu sistema dinâmico seja ergódico. Está demonstrado que o bilhar no estádio de Bunimovich (duas semicircunferências unidas por segmentos paralelos), na cardióide, e em muitos estádios elípticos (duas semielipses cortadas ao longo do eixo menor, unidas por segmentos paralelos) são ergódicos e satisfazem outras propriedades estocásticas. No caso do estádio elíptico, se $a < \sqrt{2}$, para ter um bilhar ergódico a separação entre as duas metades deve crescer tendendo a infinito à medida que a cresce tendendo a $\sqrt{2}$. Ver [8]. Para o bilhar da figura 5, se $a \geq \sqrt{2}$ não é possível construir estádios elípticos ergódicos.

8 Como se prova a ergodicidade de bilhares

Desde já devemos esclarecer que apesar de usarmos o termo ergodicidade para os sistemas dinâmicos com o tipo de desordem definido na seção 5, na realidade nesses casos temos provadas propriedades estocásticas muito mais fortes.

De uma forma ou de outra, a prova da ergodicidade de sistemas dinâmicos diferenciáveis passa pela prova de algum tipo de hiperbolicidade (ver seção 6). O procedimento padrão para isso é o estudo dos expoentes de Liapunov da transformação e a aplicação da chamada Teoria de Pesin, que permite construir as curvas expansoras e contratoras das

quais falamos anteriormente. Para evitar maiores complicações técnicas vamos supor que trabalhamos no espaço euclidiano $\mathbf{R}^d, d \geq 2$. Quem desejar um modelo ainda mais intuitivo pode pensar no plano.

Se p é um ponto fixo de um difeomorfismo S (S e sua inversa são diferenciáveis, $S(p) = p$) definido em um conjunto de $\mathbb{R}^d$ e queremos analisar o comportamento de S em uma vizinhança de p, como primeira aproximação podemos estudar sua parte linear, ou seja, a derivada (diferencial) $(S)'_p : \mathbb{R}^d \to \mathbb{R}^d$. Se α_i são as raízes reais do polinômio característico e v_i seus respectivos vetores próprios, teremos que $\lim_{n\to\pm\infty} 1/n \log \|(S^n)'_p v_i\| = \log |\alpha_i|$. Para o caso de autovalores complexos (que aparecem aos pares conjugados α_j, $\bar{\alpha}_j$), esta igualdade se verifica para v_j contido em subespaços de dimensão igual ao dobro da multiplicidade de α_j. Observe que isso não ocorre se em lugar do autovetor v_i tomamos qualquer vetor v. Chamamos de *expoentes de Liapunov* a estes números $\log |\alpha_i| = \lambda_i$. É fácil ver que a transformação S expande os subespaços correspondentes aos autovetores com expoente de Liapunov $\lambda_i > 0$ e contrai os subespaços correspondentes aos autovetores com expoente de Liapunov $\lambda_i < 0$. Em qualquer caso, temos que desconsiderar o vetor nulo nesses subespaços.

Se p não é um ponto fixo e existe uma medida μ invariante por S, um célebre teorema de Oseledets estabelece a existência do limite que define os expoentes de Liapunov para μ quase todo ponto p. O teorema de Oseledets pode ser aplicado a transformações S definidas em conjuntos muito gerais. Para estendê-lo ao caso dos bilhares com os quais trabalhamos daremos um enunciado no contexto mais geral de **transformações contínuas com singularidades**. Para evitar a introdução de notações que complicam a compreensão do que é realmente importante, nos restringiremos a trabalhar em $\mathbb{R}^d$. Para uma abordagem mais detalhada deste tema consultar [6] e [7].

Seja M a união de um número finito de regiões em $\mathbb{R}^d$, coladas ao longo de "superfícies" cuja união está contida no conjunto das **singularidades** $\mathcal{D}$. Este conjunto, por sua vez, é a união de um conjunto finito de variedades fechadas de dimensão menor que d (se $d = 2$, $\mathcal{D}$ é a união de um número finito de curvas e pontos). O conjunto $P = M \setminus \mathcal{D}$ é aberto. Seja μ uma medida definida sobre os borelianos de M, tal que $\mu(M) = 1$ (medida de probabilidade). Em primeiro lugar daremos uma versão bastante precisa do teorema de Oseledets e a seguir uma

interpretação mais geométrica.

Seja $S : P \to M$ uma transformação com inversa, ambas suficientemente diferenciáveis (difeomorfismo de classe $C^r, r \geq 1$, para o teorema de Oseledets e $r \geq 2$ para o resto da teoria), de P em sua imagem, que preserva a medida μ e que verifica algumas outras condições de crescimento. Seja H o conjunto dos pontos que têm infinitos iterados tanto para o futuro como para o passado: $H = \cap_{n=-\infty}^{\infty} S^n P$; segue que $\mu(H) = 1$. Então o teorema de Oseledets diz que para quase todo ponto $p \in H$ existe $m(p)$ números reais $\lambda_i(p) : \lambda_1(p), \lambda_2(p), \cdots, \lambda_{m(p)}(p)$ e subespaços $E_1(p), E_2(p), \cdots, E_{m(p)}(p)$, tais que

$$\lim_{n \to \pm\infty} \frac{1}{n} \log \|(S^n)'_p v_i\| = \lambda_i(p) \qquad \textit{para todo vetor não nulo } v_i \in E_i(p).$$

$E_i(p)$ é o **subespaço próprio do expoentes de Liapunov** $\lambda_i(p)$. Estes números podem variar com p, mas são invariantes ao longo de sua órbita: $\lambda_i(S^n(p)) = \lambda_i(p)$. O teorema de Oseledets assegura que para quase todo ponto (pontos **regulares**) do espaço onde atua uma transformação que preserva medida, o comportamento dos pontos próximos de sua órbita pode ser aproximado levando em conta as contrações ou expansões em direções conhecidas; expansões ou contrações respectivamente para λ_i maior ou menor que zero.

Garantida a existência destes subespaços e dos expoentes de Liapunov que assintoticamente se comportam como os subespaços próprios e os autovalores de uma matriz, surge a seguinte pergunta: como deduzir das propriedades desses números as propriedades da função S, em particular as propriedades relacionadas com a hiperbolicidade?

A resposta para esta pergunta é dada pela chamada **Teoria de Pesin**, que trata, em primeiro lugar, da construção de **variedades invariantes** expansivas e contrativas nos pontos em que não há expoente de Liapunov nulo.

Não seremos muito rigorosos no enunciado dos teoremas de Pesin para não complicar a história. Definimos a **região de Pesin**, Σ, como o conjunto dos pontos regulares p para os quais todos os expoentes de Liapunov são diferentes de zero ($\lambda_i(p) \neq 0$ para todo $1 \leq i \leq m(p)$). Então para cada ponto $p \in \Sigma$ existem variedades invariantes estáveis ($W^s(p)$) e instáveis ($W^u(p)$) cujos pontos se aproximam ou se afastam respectivamente entre si quando iteramos por S. A soma das dimensões

destas variedades é igual à dimensão de todo o espaço, e cada uma delas é tangente ao hiperplano (subespaço) gerado pelos vetores com expoentes de Liapunov negativos ($W^s(p)$) e positivos ($W^u(p)$). Além disso, Σ pode ser decomposta em um conjunto enumerável de subconjuntos invariantes, em cada um dos quais S é ergódica.

Se a região de Pesin tem medida total ($\mu(\Sigma) = 1$) diz-se que S é **(não uniformemente) hiperbólica** ou que o sistema (M, S, μ) tem comportamento caótico. Usamos a expressão não uniformemente porque os ângulos entre os subespaços estáveis e instáveis, assim como os $\lambda(p)$, podem se aproximar muito de zero, gerando uma série de dificuldades no controle das propriedade básicas da hiperbolicidade.

Faremos uma reinterpretação do significado do teorema quando a região de Pesin tem medida um. Se os expoentes de Liapunov são não nulos em quase todo ponto de H (ou de P ou M) então em quase todo ponto temos variedades instável e estável local (cujos pontos, quando iterados por S^n se aproximam ou se afastam exponencialmente com $n \to \infty$), tais que a soma de suas dimensões é d. Mais ainda, a segunda parte do teorema estabelece que existe um conjunto enumerável Σ_i de **componentes ergódicas**, ou seja, conjuntos de medida positiva que não possuem subconjuntos S-invariantes com medida maior que zero e menor que um: S é caótica dentro de cada Σ_i.

Dessa forma o problema de provar a ergodicidade de todo o sistema se transforma em provar: 1) que os expoentes de Liapunov são não nulos μ-quase todo ponto e, portanto, 2) que a componente ergódica é única.

Existem duas maneiras tradicionais de atacar o primeiro problema: mostrar que

A) em quase todo o ponto $x \in N$ existem **cones** $C(x)$ invariantes por S' : $S'_x(C(x)) \subset C(S(x))$ (um cone $C(x) \subset \mathbf{R}^2$, por exemplo, é um ângulo com vértice em x, ver figura 9); ou

B) em quase todo ponto $x \in N$ pode ser definida uma **forma quadrática** não degenerada Q_x crescente ao longo das órbitas de S : $Q_{S(x)}(S'_x v) - Q_x v > 0$ (*) se v é um vetor qualquer com origem em x. Uma forma quadrática é determinada por uma matriz A_x simétrica sendo $Q_x : \mathbf{R}^d \to \mathbf{R}$, definida por $Q_x v = v^t A v$; se A não tem autovalores nulos, Q é não degenerada.

A) e B) são essencialmente equivalentes; em particular observamos

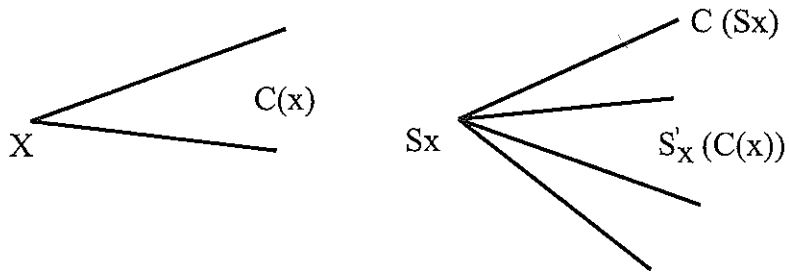

Figura 9: Cone invariante $C(x)$.

que o conjunto $\{v : Q_x v > 0\}$ constitui um cone $C(x)$. Se uma das condições é verificada, o sistema dinâmico (M, S, μ) tem expoentes de Liapunov não nulos μ-quase todo ponto.

As condições para dar uma resposta afirmativa ao segundo problema são mais complexas mas tentaremos explicá-las no contexto da hiperbolicidade (não uniforme) provada pelo método B), e supondo que $d = 2$. Se além de B), o sistema (M, S, μ) com $M \subset \mathbb{R}^2$ verifica as condições a seguir então é ergódico:

B1) as órbitas com n pontos que começam e terminam no conjunto das singularidades $\mathcal{D}$ constituem um conjunto finito;

B2) as tangente a $\mathcal{D}$ estão nos cones de vetores onde a forma quadrática Q é positiva ou negativa;

B3) para quase todo ponto de $\mathcal{D}$, suas órbitas entram de maneira recorrente em regiões onde Q cresce uniformemente. Isto significa que na expressão (*) a diferença é maior que uma constante positiva δ;

B4) a medida dos pontos cujas variedades invariantes (estável e instável) intersectam $\mathcal{D}$ e têm comprimento maior que ϵ, é menor que ϵ;

B5) existe um $x \in H$ tal que sua órbita $\{S^n(x)\}_{n\in\mathbb{Z}}$ passa por todas as componentes conexas por arcos de N.

Estas condições asseguram, por exemplo, que as variedades instáveis locais $W^u(x)$, ao evoluir por S, vão se particionando por seus cortes com $\mathcal{D}$ (lembramos que em $\mathcal{D}$ a função S não está definida), mas ainda assim, os trechos cortados são suficientemente longos: **a hiperbolicidade prevalece sobre o fracionamento**. O fato de serem as variedades invariantes suficientemente longas implica que podemos unir dois pontos de uma vizinhança de quase todo ponto através de um zigue-zague por

arcos estáveis e instáveis; ver figura 10.

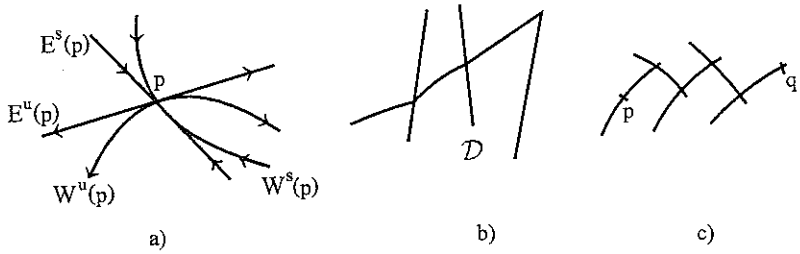

Figura 10: a) Subespaços estáveis e instáveis $[E^{s,u}(p)]$ e suas variedades locais $W^{s,u}(p)$; b) Partições de $S^n(W^u(p))$ por $\mathcal{D}$; c) Zig-zag de variedades estáveis e instáveis unindo p, q.

Os sistemas do tipo dos bilhares satisfazem as hipóteses desses teoremas. A formulação de Sinai da Hipótese de Boltzmann a coloca naturalmente no contexto dos bilhares em espaços de dimensão alta. Lembramos que de acordo com o indicado na seção 7, a dimensão do espaço é $2N\nu - 2\nu - 1$ e o número de cilindros é igual a $N(N-1)/2$ onde N é o número de bolas e ν é a dimensão do espaço onde se movem as bolas.

9 Provando a Hipótese de Boltzmann-Sinai

A principal dificuldade que se coloca com o modelo dos gases, em sua formulação via bilhares, é que se o obstáculo é apenas convexo, e não estreitamente convexo, (cilindros, por exemplo, ver figura 11), não é simples provar a existência de expoentes de Liapunov diferentes de zero.

Krylov e Sinai já haviam observado que enquanto um bilhar com obstáculos estritamente convexos se comporta como um sistema dinâmico "bem" hiperbólico, quando os obstáculos não são estritamente convexos, existe apenas hiperbolicidade parcial. Os obstáculos das figuras 1, 8 e 12 são estritamente convexos (o bilhar é **dispersor**), enquanto que os da figura 13 são apenas convexos (o bilhar é **semidispersor**).

Geometricamente, o mecanismo de geração de hiperbolicidade pode ser ilustrado com imagens inspiradas na ótica. Consideremos, como na figura 12, um obstáculo estritamente convexo que visualizamos como um espelho. Consideremos também uma frente de ondas plana (que

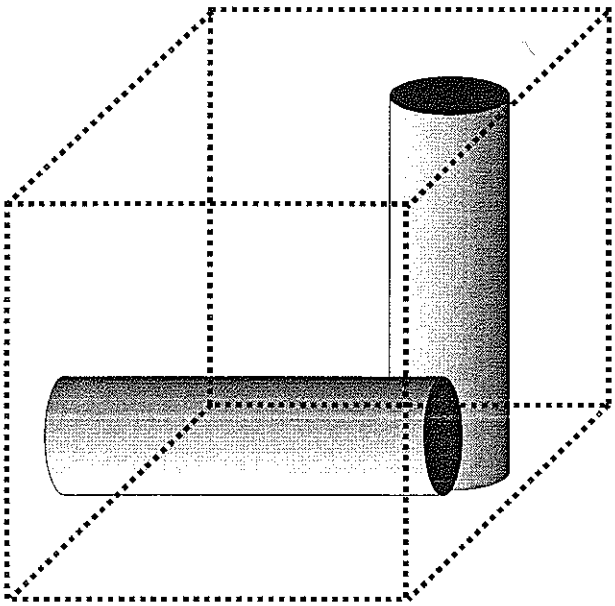

Figura 11: Obstáculos cilíndricos que aparecem na formulação bilharística do modelo dos gases.

corresponde a dar um ponto $x = (Q, V)$ no espaço de fase M, considerar no espaço de configurações – onde realmente ocorre o movimento – um hiperplano Γ por Q e, em cada ponto desse hiperplano, tomar vetores paralelos a V). Então, logo que a frente de ondas alcança o espelho, ela se torna estritamente convexa enquanto que as distâncias lineares medidas a partir de Γ são uniformemente expandidas. Este é exatamente o mecanismo que gera a hiperbolicidade uniforme nos bilhares. Neste caso as dificuldades vêm, como já foi em parte observado, da existência de trajetórias tangentes que fazem com que o sistema perca diferenciabilidade.

Se os obstáculos não são estritamente convexos aparecem situações como a da figura 13 em que a imagem do hiperplano Γ não é curvada em todas as direções: as trajetórias que se chocam com uma diretriz do cilindro (as retas paralelas ao eixo, na figura) saem paralelas, são "neutras", e portanto não se separam nem aumentam as distâncias me-

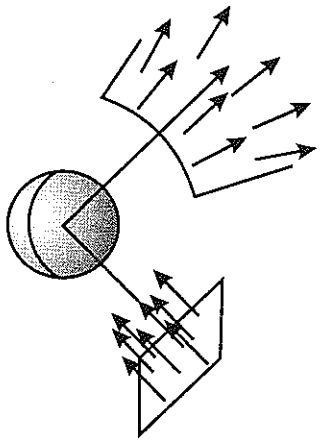

Figura 12: Obstáculo estritamente convexo: um feixe paralelo é completamente dispersado depois do choque.

didas a partir de Γ. Esta direção neutra pode desaparecer devido a muitos choques com outros cilindros. O fato de que a hiperbolicidade global (a expansão em todos os sentidos) seja obtida através de muitos choques com obstáculos, levou naturalmente à introdução do conceito de **suficiência** de trajetórias, cuja descrição intuitiva damos a seguir.

Seja $S^{[a,b]}x$ um segmento finito de trajetória que não passa por singularidades, ou seja, tal que $S^j x \notin \mathcal{D}$ para $j \in [a,b]$. Seja $S^a x = (Q,V) \in P$ e consideremos o hiperplano $\tilde{\Gamma}(S^a x) := \{(Q + dQ, V) : dQ \in \mathbb{R}^d$, pequeno, perpendicular a $V\}$. Dizemos que um segmento de trajetória $S^{[a,b]}x$ é **suficiente** se $\pi(S^b\tilde{\Gamma})$ é estritamente convexo (ver figura 14). Um ponto do espaço de fase $x \in P$ é suficiente se sua trajetória é suficiente, ou seja, se contém um segmento de trajetória suficiente.

Se o segmento de trajetória não é suficiente, a curvatura de $\pi(S^b\tilde{\Gamma})$ em $\pi(S^b x)$ necessariamente se anula em alguma direção, formando o assim chamado **subespaço neutro**. Não é difícil ver que na vizinhança de um ponto suficiente os expoentes de Liapunov relevantes são não nulos.

Essas observações levam ao não trivial

Teorema Fundamental dos Bilhares Semidispersores (Sinai-Chernov [14]). Se em um bilhar semidispersor são satisfeitas as con-

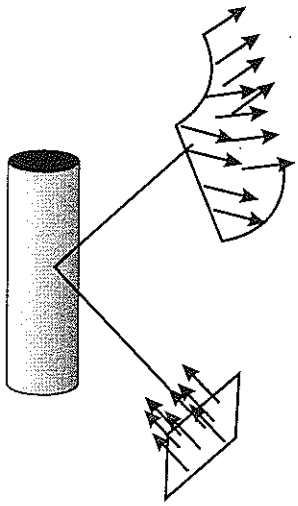

Figura 13: Obstáculo não estritamente convexo: existem direções em que vetores do feixe paralelo incidente não são dispersados.

dições geométricas e de hiperbolicidade, relacionadas com as singularidades (ver B1) - B4) na seção anterior), então todo ponto $x \in P$ suficiente tem uma vizinhança pertencente a uma componente ergódica do sistema (ver comentários depois dos Teoremas de Pesin).

A partir desse teorema Sinai e Chernov provaram em 1987 que o sistema formado por $N = 2$ bolas no ν-toro é um sistema ergódico.

Os resultados posteriores foram obtidos pelos matemáticos húngaros Krámli, Simányi e Szász, os quais entre 1990 e 1992 estenderam os resultados para três e quatro bolas no ν-toro, com $\nu \geq 2$ (ver [2], [3]).

A obtenção de resultados muito dependentes do número de bolas e da velocidade do processo, se deve às dificuldades para driblar o problema da suficiência ao aumentarmos a dimensão do espaço em que se trabalha e o número de cilindros.

Essas dificuldades foram resolvidas de maneira parcial, muito recentemente, por Simányi e Szász, introduzindo condições que permitem analisar o quão pequeno é o conjunto não suficiente (deve ser formado por uma união de um conjunto enumerável de variedades com codimensão pelo menos dois). Essa pequenez relativa permite obter uma

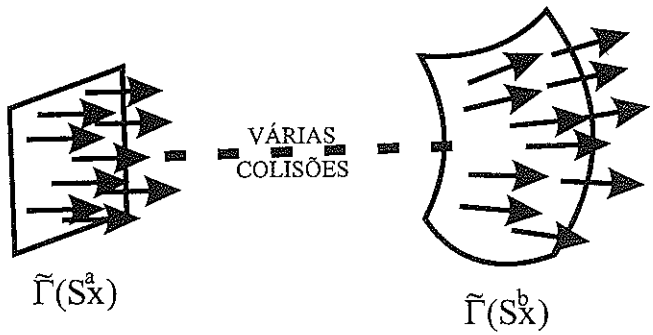

Figura 14: Trajetórias suficientes. Depois de muitas colisões todos os vetores de um feixe paralelo são dispersados.

rica estrutura de colisões entre as bolas, assegurando as condições às quais se refere o teorema fundamental.

Ainda assim não conseguiram provar o caráter ergódico do sistema de bolas elásticas, mas apenas a não anulação dos expoentes de Liapunov em quase todo ponto. Isso permite aplicar a segunda parte dos teoremas de Pesin, e deduzir que existe um conjunto enumerável de componentes ergódicas. Os autores comentam: "Our methods so far do not give the expected global ergodicity.".

Como novidade, estudam o movimento de bolas de massas distintas $\vec{m} = (m_1, m_2, \cdots, m_N) \in \mathbb{R}^N$ (ao contrário de todos os trabalhos anteriores e da hipótese original, onde as massas são iguais), mas o teorema resultante vale para $\vec{m}$ fora de um conjunto pequeno de $\mathbb{R}^N$ ($\vec{m}$ não pertence a uma união enumerável de subvariedades analíticas próprias).

Seja $\mathbb{R}^N_+$ o conjunto das N-uplas de números reais positivos. Simányi e Szász provaram o seguinte

Teorema [9]. *Se $N \geq 2$, $\nu \geq 2$, $\vec{m} = (m_1, m_2, \cdots, m_N) \in \mathbb{R}^N_+$ são as massas das bolas e seus raios são iguais a $r = R_0(N, \nu)$ então o sistema $(M,\ S,\ \mu)_{\vec{m}}$ tem todos os seus expoentes de Liapunov relevantes não nulos se $\vec{m}$ está fora de um conjunto "pequeno" de $\mathbb{R}^N_+$.*

Este resultado foi melhorado em um trabalho mais recente de Simányi. Os últimos progressos na direção de provar a ergodicidade do modelo de Boltzmann-Sinai têm levado à utilização de ferramentas refinadas da geometria algébrica. Chernov, Simányi e Szász (e seus alunos)

continuam trabalhando nesses temas.

CONCLUSÃO. Hoje em dia não é clara a relevância **para a Física** da Hipótese Ergódica e dos resultados matemáticos a seu respeito. Entretanto a elucidação de seu significado e sua apresentação rigorosa vêm tendo um impacto muito grande tanto na Física como na Matemática. Contribuíram de maneira fundamental para o desenvolvimento da Mecânica Estatística e da Teoria Ergódica de Sistemas Dinâmicos e, de maneira geral, para o estudo formalizado dos movimentos caóticos. Neste sentido são de grande transcendência os resultados que foram expostos, sem demasiado rigor, neste trabalho.

Agradecimentos. Com Dómokos Szász mantive interessantes conversações sobre estes temas. Agradeço-lhe por me haver permitido usar partes de seu artigo [15] neste trabalho. Sylvie Oliffson e Sônia Pinto colaboraram com a confecção de várias figuras e Michelle Dysman realizou uma atenta tradução crítica para o português.

Referências

[1] I. P. Cornfeld, S. V. Fomin & Ya. G. Sinai, *Ergodic Theory*, Springer Verlag, 1982.

[2] A. Krámli, N. Simányi & D. Szász, *The K–Property of Three Billiard Balls*, Annals of Mathematics, **133** (1991), 37–72.

[3] A. Krámli, N. Simányi & D. Szász, *The K–Property of Four Billiard Balls*, Commun. Math. Phys., **144** (1992), 107-142.

[4] R. Mañé, *Introducao à Teoria Ergódica*, IMPA, Rio de Janeiro, 1986.

[5] R. Mañé, *Global Variational Methods in Conservative Dynamics*, IMPA, Rio de Janeiro, 1991.

[6] 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.

[7] R. Markarian, *New ergodic billiards: exact results*, Nonlinearity, **6** (1993), 819–841.

[8] R. Markarian, S. Oliffson & S. Pinto, *Chaotic properties of the elliptical stadium*, Commun Math. Phys., **174** (1996), 661–679.

[9] N. Simányi & D. Szász, *Hard ball systems are completely hyperbolic*, Ann. of Math. **149** (1999), 35–96.

[10] N. Simányi, *Ergodicity of hard spheres in a box*, Ergodic Theory Dynam. Systems **19** (1999), 741–766.

[11] 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

[12] Ya.G. Sinai, *Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards*, Russ. Math. Surv. **25** (1970), 137–189.

[13] 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.

[14] 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.

[15] D. Szász, *Boltzmann's ergodic hypothesis, a conjecture for centuries?*, Studia Sci. Math. Hung. **31** (1996), 299–322.

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