Escuola Doctoral en Probabilidades y Sistémas Dinâmicos

PUC - Santiago, 09 a 19 de Octobre de 2018

Conteúdo

Capítulo 1

Medidas Invariantes e Recorrência

A Teoria Ergódica estuda o comportamento de sistemas dinâmicos relativamente a medidas que permanecem invariantes sob a ação da dinâmica. Mais precisamente, busca-se descrever as propriedades que são válidas para quase toda a trajetória do sistema, relativamente à medida invariante. Começaremos, na Seção 1.1, por definir estas noções de sistema dinâmico e de medida invariante.

As raízes da teoria remontam à primeira metade do século 19. De fato, em 1838 o matemático francês Joseph Liouville observou que todo sistema da Mecânica Newtoniana (com conservação da energia) admite uma medida invariante natural no seu espaço de configurações. Além disso, em 1845 o grande matemático e físico alemão Carl Friedrich Gauss observou que uma certa transformação no intervalo que tem um papel importante na Teoria dos Números admite uma medida invariante que é equivalente à medida de Lebesgue. Estes são dois dos exemplos de aplicação da Teoria Ergódica que apresentaremos na Seção 1.3. Muitos outros surgirão ao longo deste livro.

O primeiro resultado importante foi devido ao grande matemático francês Henri Poincaré, ao final do século 19. Ele estava especialmente interessado no movimento dos corpos celestes, tais como planetas e cometas, o qual é descrito por certas equações diferenciais que resultam da Lei da Gravitação de Newton. A partir da observação de Liouville, Poincaré mostrou que para quase todo estado inicial do sistema, ou seja, quase todo valor das posições e velocidades iniciais, a solução da equação diferencial regressa arbitrariamente perto desse estado inicial, a menos que vá para infinito. Mais ainda, ele apontou que essa propriedade de *recorrência* não é exclusiva dos sistemas da Mecânica Celeste: ela vale sempre que o sistema admite uma medida invariante. Este será o tema da Seção 1.2.

Ele reaparecerá na Seção 1.4 num contexto mais elaborado: consideramos um número finito de sistemas dinâmicos que comutam entre si e buscamos retornos *simultâneos* das órbitas de todos esses sistemas à vizinhança do estado inicial.

Este tipo de resultado tem importantes aplicações em Combinatória e Teoria dos Números, como veremos mais tarde.

A ideia de recorrência também está por trás das construções que são conhecidas como *transformações induzidas*. A ideia básica é fixar um subconjunto do domínio com medida positiva e considerar o primeiro retorno a esse conjunto. Frequentemente, essa transformação de primeiro retorno é mais fácil de analisar e, por outro lado, ela pode ser usada para entender o comportamento da transformação original.

1.1 Medidas invariantes

Seja $(M, \mathcal{B}, \mu)$ um espaço de medida e seja $f : M \to M$ uma transformação mensurável. Dizemos que a medida μ é *invariante* por f se

$$\mu(E) = \mu(f^{-1}(E)) \quad \text{para todo conjunto mensurável } E \subset M. \tag{1.1.1}$$

Nesse caso também dizemos que f *preserva* μ. Note que a definição (1.1.1) faz sentido, uma vez que a pré-imagem de um conjunto mensurável por uma transformação mensurável ainda é um conjunto mensurável. Heuristicamente, ela significa que a probabilidade de um ponto estar num dado conjunto é igual à probabilidade de que a sua imagem esteja nesse conjunto.

É possível, e conveniente, estender esta definição a outros tipos de sistemas dinâmicos além das transformações. Estamos especialmente interessados em *fluxos*, ou seja, famílias de transformações $f^t : M \to M$, onde $t \in \mathbb{R}$, satisfazendo as seguintes condições:

$$f^0 = \mathrm{id} \quad \text{e} \quad f^{s+t} = f^s \circ f^t \text{ para todo } s, t \in \mathbb{R}. \tag{1.1.2}$$

Isto também implica que toda a transformação f^t é invertível e a sua inversa é f^{-t}. Fluxos aparecem naturalmente associados a equações diferenciais do tipo $\gamma'(t) = X(\gamma(t))$, onde X é um campo de vetores, do seguinte modo: sob condições adequadas sobre X, para cada ponto x existe uma única solução $t \mapsto \gamma_x(t)$ da equação que satisfaz $\gamma_x(0) = x$; então $f^t(x) = \gamma_x(t)$ define um fluxo no domínio M da equação diferencial.

Dizemos que uma medida μ é *invariante* pelo fluxo $(f^t)_t$ se ela é invariante por cada uma das transformações f^t, ou seja, se

$$\mu(E) = \mu(f^{-t}(E)) \quad \text{para todo mensurável } E \subset M \text{ e todo } t \in \mathbb{R}. \tag{1.1.3}$$

Proposição 1.1.1. *Sejam $f : M \to M$ uma transformação mensurável e μ uma medida em M. Então f preserva μ se, e somente se,*

$$\int \phi \, d\mu = \int \phi \circ f \, d\mu \tag{1.1.4}$$

para toda função μ-integrável $\phi : M \to \mathbb{R}$.

Demonstração. Suponhamos que a medida μ é invariante. Vamos mostrar que a relação (1.1.4) é válida para classes de funções sucessivamente mais amplas. Inicialmente, observe que por hipótese $\mu(B) = \mu(f^{-1}(B))$ para todo conjunto mensurável B. Como,

$$\int \mathcal{X}_B \, d\mu = \mu(B) \quad \text{e} \quad \mu(f^{-1}(B)) = \int (\mathcal{X}_B \circ f) \, d\mu,$$

isto mostra que (1.1.4) é válida para as funções características. Então, por linearidade da integral, (1.1.4) é válida para funções simples. Em seguida, vamos usar um argumento de aproximação para concluir que (1.1.4) vale para toda função integrável. Dada qualquer função integrável $\phi : M \to \mathbb{R}$, considere uma sequência $(s_n)_n$ de funções simples convergindo para ϕ e tal que $|s_n| \le |\phi|$ para todo n. Tal sequência existe, pela Proposição A.1.33. Então, usando o teorema da convergência dominada (Teorema A.2.11) duas vezes:

$$\int \phi \, d\mu = \lim_n \int s_n \, d\mu = \lim_n \int (s_n \circ f) \, d\mu = \int (\phi \circ f) \, d\mu.$$

Isto mostra que (1.1.4) vale para toda função integrável se μ é invariante. A recíproca também segue imediatamente dos argumentos que apresentamos. $\square$

1.2 Teorema de recorrência de Poincaré

Vamos estudar duas versões do teorema de Poincaré. A primeira (Seção 1.2.1) está formulada no contexto de espaços de medida (finita). O teorema de Kač, que provaremos na Seção 1.2.2 complementa este resultado de forma quantitativa. A segunda versão do teorema de recorrência (Seção 1.2.3) supõe que o ambiente é um espaço topológico com certas propriedades adicionais. Também provaremos uma terceira versão do teorema de recorrência, devida a Birkhoff, cuja formulação é puramente topológica.

1.2.1 Versão mensurável

O nosso primeiro resultado afirma que, dada qualquer medida invariante *finita*, quase todo ponto de qualquer conjunto mensurável E regressa a E um número infinito de vezes:

Teorema 1.2.1 (Recorrência de Poincaré)**.** *Seja $f : M \to M$ uma transformação mensurável e seja μ uma medida finita invariante por f. Seja $E \subset M$ qualquer conjunto mensurável com $\mu(E) > 0$. Então, para μ-quase todo ponto $x \in E$ existem infinitos valores de n para os quais $f^n(x)$ também está em E.*

Demonstração. Representemos por E_0 o conjunto dos pontos $x \in E$ que nunca regressam a E. Inicialmente, vamos provar que E_0 tem medida nula. Para isso, começamos por observar que as suas pré-imagens $f^{-n}(E_0)$ são disjuntas duas-a-duas. De fato, suponhamos que existem $m > n \ge 1$ tais que $f^{-m}(E_0)$ intersecta

$f^{-n}(E_0)$. Seja x um ponto na interseção e seja $y = f^n(x)$. Então $y \in E_0$ e $f^{m-n}(y) = f^m(x) \in E_0$, que está contido em E. Isto quer dizer que y volta pelo menos uma vez a E, o que contradiz a definição de E_0. Esta contradição, prova que as pré-imagens são disjuntas duas-a-duas, como afirmamos.

Observando que $\mu(f^{-n}(E_0)) = \mu(E_0)$ para todo $n \geq 1$, porque μ é invariante, concluímos que

$$\mu\big(\bigcup_{n=1}^{\infty} f^{-n}(E_0)\big) = \sum_{n=1}^{\infty} \mu(f^{-n}(E_0)) = \sum_{n=1}^{\infty} \mu(E_0).$$

Como supomos que a medida é finita, a expressão do lado esquerdo é finita. Por outro lado, à direita temos uma soma de infinitos termos, todos iguais. O único jeito desta soma ser finita é que as parcelas sejam nulas. Portanto, devemos ter $\mu(E_0) = 0$, tal como foi afirmado.

Agora, denotemos por F o conjunto dos pontos $x \in E$ que regressam a E apenas um número finito de vezes. Como consequência direta da definição, temos que todo ponto $x \in F$ tem algum iterado $f^k(x)$ em E_0. Ou seja,

$$F \subset \bigcup_{k=0}^{\infty} f^{-k}(E_0).$$

Como $\mu(E_0) = 0$ e μ é invariante, temos:

$$\mu(F) \leq \mu\big(\bigcup_{k=0}^{\infty} f^{-k}(E_0)\big) \leq \sum_{k=0}^{\infty} \mu\big(f^{-k}(E_0)\big) = \sum_{k=0}^{\infty} \mu(E_0) = 0.$$

Portanto, $\mu(F) = 0$ como queríamos provar. $\square$

Observe que o Teorema 1.2.1 implica um resultado análogo para sistemas com tempo contínuo. De fato, suponha que μ é uma medida invariante finita de um fluxo $(f^t)_t$. Segue imediatamente da definição que μ é invariante pela respectiva transformação f^1, chamada *tempo* 1 do fluxo. Aplicando o Teorema 1.2.1 à transformação tempo 1, concluímos que, dado qualquer conjunto $E \subset M$ com medida positiva, para quase todo $x \in E$ existem tempos $t_j \to +\infty$ tais que $f^{t_j}(x) \in E$. Valem observações análogas para as outras versões do teorema de recorrência, que apresentaremos posteriormente. Por outro lado, o teorema que apresentamos a seguir é específico de sistemas com tempo discreto.

1.2.2 Teorema de Kač

Seja $f : M \to M$ uma transformação mensurável e seja μ uma medida finita invariante por f. Seja $E \subset M$ qualquer conjunto mensurável com $\mu(E) > 0$. Considere a função *tempo de primeiro retorno* $\rho_E : E \to \mathbb{N} \cup \{\infty\}$ definida da seguinte forma:

$$\rho_E(x) = \min\{n \geq 1 : f^n(x) \in E\} \tag{1.2.1}$$

sempre que o conjunto do lado direito for não vazio, isto é, se x tiver algum iterado em E; caso contrário, $\rho_E(x) = \infty$. De acordo com o Teorema 1.2.1, a segunda alternativa só ocorre para um conjunto de pontos com medida nula.

O resultado que vamos apresentar a seguir mostra que esta função é integrável e exibe o valor da sua integral. Para o enunciado precisamos da seguinte notação:

$$E_0 = \{x \in E : f^n(x) \notin E \text{ para todo } n \geq 1\} \quad \text{e}$$
$$E_0^* = \{x \in M : f^n(x) \notin E \text{ para todo } n \geq 0\}.$$

Ou seja, E_0 é o conjunto dos pontos de E que nunca regressam a E e E_0^* é o conjunto dos pontos de M que nunca entram em E. Note que $\mu(E_0) = 0$, pelo teorema de recorrência de Poincaré.

Teorema 1.2.2 (Kač)**.** *Seja $f : M \to M$, μ uma medida invariante finita e E um subconjunto com medida positiva. Então a função ρ_E é integrável e*

$$\int_E \rho_E \, d\mu = \mu(M) - \mu(E_0^*).$$

Demonstração. Para cada $n \geq 1$, defina

$$E_n = \{x \in E : f(x) \notin E, \ldots, f^{n-1}(x) \notin E, \text{ mas } f^n(x) \in E\} \quad \text{e}$$
$$E_n^* = \{x \in M : x \notin E, f(x) \notin E, \ldots, f^{n-1}(x) \notin E, \text{ mas } f^n(x) \in E\}.$$

Ou seja, E_n é o conjunto dos pontos de E que retornam a E pela primeira vez exatamente no momento n,

$$E_n = \{x \in E : \rho_E(x) = n\},$$

e E_n^* é o conjunto dos pontos que não estão em E e que entram em E pela primeira vez exatamente no momento n. É claro que estes conjuntos são mensuráveis e, portanto, ρ_E é função mensurável. Além disso, os conjuntos E_n, E_n^*, $n \geq 0$ são disjuntos dois-a-dois e a sua união é todo o espaço M. Portanto

$$\mu(M) = \sum_{n=0}^{\infty} \big(\mu(E_n) + \mu(E_n^*)\big) = \mu(E_0^*) + \sum_{n=1}^{\infty} \big(\mu(E_n) + \mu(E_n^*)\big). \tag{1.2.2}$$

Agora observe que

$$f^{-1}(E_n^*) = E_{n+1}^* \cup E_{n+1} \quad \text{para todo } n. \tag{1.2.3}$$

De fato, $f(y) \in E_n^*$ quer dizer que o primeiro iterado de $f(y)$ que está em E é $f^n(f(y)) = f^{n+1}(y)$ e isto ocorre se, e somente se, $y \in E_{n+1}^*$ ou $y \in E_{n+1}$. Isto prova a igualdade (1.2.3). Logo, pela invariância de μ,

$$\mu(E_n^*) = \mu(f^{-1}(E_n^*)) = \mu(E_{n+1}^*) + \mu(E_{n+1}) \quad \text{para todo } n.$$

Aplicando esta relação repetidas vezes, obtemos que

$$\mu(E_n^*) = \mu(E_m^*) + \sum_{i=n+1}^{m} \mu(E_i) \quad \text{para todo } m > n. \tag{1.2.4}$$

A relação (1.2.2) implica que $\mu(E_m^*) \to 0$ quando $m \to \infty$. Portanto, tomando o limite quando $m \to \infty$ na igualdade (1.2.4), obtemos:

$$\mu(E_n^*) = \sum_{i=n+1}^{\infty} \mu(E_i), \tag{1.2.5}$$

Para finalizar a demonstração, substituímos (1.2.5) na igualdade (1.2.2). Desta forma obtemos que

$$\mu(M) - \mu(E_0^*) = \sum_{n=1}^{\infty} \big(\sum_{i=n}^{\infty} \mu(E_i) \big) = \sum_{n=1}^{\infty} n\mu(E_n) = \int_E \rho_E \, d\mu,$$

como queríamos demonstrar. □

Em alguma situações, por exemplo quando o sistema (f, μ) é *ergódico* (esta propriedade será definida e estudada em detalhe mais tarde) o conjunto E_0^* tem medida zero. Então a conclusão do teorema de Kač diz que

$$\frac{1}{\mu(E)} \int_E \rho_E \, d\mu = \frac{\mu(M)}{\mu(E)} \tag{1.2.6}$$

para todo conjunto mensurável E. O lado esquerdo desta igualdade é o *tempo médio de retorno* a E. A igualdade (1.2.6) diz que *o tempo médio de retorno é inversamente proporcional à medida de E.*

Observação 1.2.3. Por definição, $E_n^* = f^{-n}(E) \setminus \cup_{k=0}^{n-1} f^{-k}(E)$. O fato de que a soma (1.2.2) é finita implica que a medida deste conjunto converge para zero quando $n \to \infty$. Isto será útil mais tarde.

1.2.3 Versão topológica

Agora suponhamos que M é um espaço topológico, munido da sua σ-álgebra de Borel $\mathcal{B}$. Dizemos que um ponto $x \in M$ é *recorrente* para uma transformação $f : M \to M$ se existe uma sequência $n_j \to \infty$ em $\mathbb{N}$ tal que $f^{n_j}(x) \to x$. Analogamente, dizemos que $x \in M$ é recorrente para um fluxo $(f^t)_t$ se existe uma sequência $t_j \to +\infty$ em $\mathbb{R}$ tal que $f^{t_j}(x) \to x$ quando $j \to \infty$.

No próximo teorema supomos que o espaço topológico M admite uma base enumerável de abertos, ou seja, existe uma família enumerável $\{U_k : k \in \mathbb{N}\}$ de abertos tal que todo aberto de M pode ser escrito como união de elementos U_k dessa família. Esta hipótese é satisfeita na maioria dos exemplos interessantes.

Teorema 1.2.4 (Recorrência de Poincaré)**.** *Suponhamos que M admite uma base enumerável de abertos. Seja $f : M \to M$ uma transformação mensurável e seja μ uma medida finita em M invariante por f. Então, μ-quase todo ponto $x \in M$ é recorrente para f.*

Demonstração. Para cada k representamos por $\tilde{U}_k$ o conjunto dos pontos $x \in U_k$ que nunca regressam a U_k. De acordo com o Teorema 1.2.1, todo $\tilde{U}_k$ tem medida nula. Consequentemente, a união enumerável

$$\tilde{U} = \bigcup_{k \in \mathbb{N}} \tilde{U}_k$$

tem medida nula. Portanto, para demonstrar o teorema será suficiente que mostremos que todo ponto x que não está em $\tilde{U}$ é recorrente. Isso é fácil, como vamos ver. Seja $x \in M \setminus \tilde{U}$ e seja U uma vizinhança qualquer de x. Por definição, existe algum elemento U_k da base de abertos tal que $x \in U_k$ e $U_k \subset U$. Como x não está em $\tilde{U}$, também temos que $x \notin \tilde{U}_k$. Em outras palavras, existe algum $n \geq 1$ tal que $f^n(x)$ está em U_k. Em particular, $f^n(x)$ também está em U. Como a vizinhança U é arbitrária, isto prova que x é um ponto recorrente. $\square$

Observe que as conclusões dos Teoremas 1.2.1 e 1.2.4 não são verdadeiras, em geral, se omitirmos a hipótese de que a medida μ é finita:

Exemplo 1.2.5. Seja $f : \mathbb{R} \to \mathbb{R}$ a translação de 1 unidade, isto é, a transformação definida por $f(x) = x + 1$ para todo $x \in \mathbb{R}$. É fácil verificar que f deixa invariante a medida de Lebesgue em $\mathbb{R}$ (que é infinita). Por outro lado, nenhum ponto é recorrente para f. Portanto, pelo teorema de recorrência, f não pode admitir nenhuma medida invariante finita.

No entanto, é possível estender estes enunciados para certos casos de medidas infinitas: veja o Exercício 1.6.7.

Para terminar, apresentamos uma versão puramente topológica do Teorema 1.2.4, chamada teorema de recorrência de Birkhoff, que não faz qualquer menção a medidas invariantes:

Teorema 1.2.6 (Recorrência de Birkhoff)**.** *Se $f : M \to M$ é uma transformação contínua num espaço métrico compacto M, então existe algum ponto $x \in M$ que é recorrente para f.*

Demonstração. Considere a família $\mathcal{I}$ de todos os conjuntos fechados não-vazios $X \subset M$ que são invariantes, no sentido de que $f(X) \subset X$. Esta família é não-vazia, uma vez que $M \in \mathcal{I}$. Afirmamos que um elemento $X \in \mathcal{I}$ é minimal para a relação de inclusão se, e somente se, a órbita de todo ponto $x \in X$ é densa em X. De fato, é claro que se X é fechado invariante então X contém o fecho da órbita de qualquer dos seus pontos. Logo, para ser minimal X precisa coincidir com qualquer desses fechos. Reciprocamente, pela mesma razão, se X coincide com o fecho da órbita de qualquer dos seus pontos então ele coincide com qualquer subconjunto fechado invariante, ou seja, X é minimal. Isto prova a nossa afirmação. Em particular, qualquer ponto x num conjunto minimal é recorrente. Logo, para provar o teorema basta mostrar que existe algum conjunto minimal.

Afirmamos que todo conjunto totalmente ordenado $\{X_\alpha\} \subset \mathcal{I}$ admite algum minorante. De fato, considere $X = \cap_\alpha X_\alpha$. Observe que X é não-vazio, uma vez que os X_α são compactos e constituem uma família totalmente ordenada. É claro que X é fechado e invariante por f e também que ele é um minorante para o conjunto $\{X_\alpha\}$. Isto prova a nossa afirmação. Agora podemos aplicar o Lema de Zorn para concluir que $\mathcal{I}$ realmente contém elementos minimais. $\square$

O Teorema 1.2.6 também segue imediatamente do Teorema 1.2.4 juntamente com o fato, que provaremos mais tarde, que toda transformação contínua num espaço métrico compacto admite alguma medida de probabilidade invariante.

1.3 Exemplos

Em seguida vamos descrever alguns exemplos simples de medidas invariantes por transformações ou por fluxos, que nos ajudam a interpretar o significado do teorema de recorrência de Poincaré, bem como obter conclusões interessantes.

1.3.1 Expansão decimal

O nosso primeiro exemplo é a transformação definida no intervalo $[0,1]$ do seguinte modo

$$f:[0,1]\to[0,1], \quad f(x)=10x-[10x]$$

onde $[10x]$ representa o maior inteiro menor ou igual a $10x$. Em outras palavras, f associa a cada $x\in[0,1]$ a parte fracionária de $10x$.

Afirmamos que a medida de Lebesgue μ no intervalo é invariante pela transformação f, isto é, ela satisfaz a condição

$$\mu(E)=\mu(f^{-1}(E)) \quad \text{para todo conjunto mensurável } E\subset M. \tag{1.3.1}$$

Esse fato pode ser verificado da seguinte forma. Comecemos por supor que E é um intervalo. Então, a pré-imagem $f^{-1}(E)$ consiste de dez intervalos, cada um deles dez vezes mais curto do que E. Logo, a medida de Lebesgue de $f^{-1}(E)$ é igual à medida de Lebesgue de E. Isto mostra que (1.3.1) é satisfeita no caso de intervalos. Como consequência, essa relação é satisfeita sempre que E é uma união finita de intervalos. Agora, a família das uniões finitas de intervalos é uma álgebra que gera a σ-álgebra de Borel de $[0,1]$. Portanto, para concluir a demonstração basta usar o seguinte fato geral:

Lema 1.3.1. *Seja $f:M\to M$ uma transformação mensurável e μ uma medida finita em M. Suponha que existe uma álgebra $\mathcal{A}$ de subconjuntos mensuráveis de M tal que $\mathcal{A}$ gera a σ-álgebra $\mathcal{B}$ de M e $\mu(E)=\mu(f^{-1}(E))$ para todo $E\in\mathcal{A}$. Então o mesmo vale para todo conjunto $E\in\mathcal{B}$, isto é, a medida μ é invariante por f.*

Demonstração. Comecemos por provar que $\mathcal{C}=\{E\in\mathcal{B}:\mu(E)=\mu(f^{-1}(E))\}$ é uma classe monótona. Para isso, seja $E_1\subset E_2\subset\dots$ uma sequência de elementos em $\mathcal{C}$ e seja $E=\cup_{i=1}^{\infty}E_i$. Pelo Teorema A.1.14 (veja o Exercício A.1.9), temos que

$$\mu(E)=\lim_i \mu(E_i) \quad \text{e} \quad \mu(f^{-1}(E))=\lim_i \mu(f^{-1}(E_i)).$$

Então, usando o fato de que $E_i \in \mathcal{C}$,

$$\mu(E) = \lim_i \mu(E_i) = \lim_i \mu(f^{-1}(E_i)) = \mu(f^{-1}(E)).$$

Logo $E \in \mathcal{C}$. De modo inteiramente análogo se mostra que a interseção de qualquer sequência decrescente de elementos de $\mathcal{C}$ está em $\mathcal{C}$. Isto prova que $\mathcal{C}$ é de fato uma classe monótona.

Agora é fácil obter a conclusão do lema. Note que $\mathcal{C}$ contém $\mathcal{A}$, por hipótese. Portanto, usando o teorema das classes monótonas (Teorema A.1.18), segue que $\mathcal{C}$ contém a σ-álgebra $\mathcal{B}$ gerada por $\mathcal{A}$. Isto é precisamente o que queríamos provar. □

Agora vamos explicar como, a partir do fato de que a medida de Lebesgue é invariante pela transformação f, podemos obter conclusões interessantes usando o teorema de recorrência de Poincaré. A função f tem uma relação direta com o algoritmo da expansão decimal: se x é dado por

$$x = 0, a_0 a_1 a_2 a_3 \cdots$$

com $a_i \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ e $a_i \neq 9$ para infinitos valores de i, então a sua imagem é dada por

$$f(x) = 0, a_1 a_2 a_3 \cdots .$$

Com isso, fica fácil escrever a expressão do iterado n-ésimo, para qualquer $n \geq 1$:

$$f^n(x) = 0, a_n a_{n+1} a_{n+2} \cdots \tag{1.3.2}$$

Agora, seja E o subconjunto dos $x \in [0, 1]$ cuja expansão decimal começa com o dígito 7, ou seja, tais que $a_0 = 7$. De acordo com o Teorema 1.2.1, quase todo elemento de E tem infinitos iterados que também estão em E. Levando em conta a expressão (1.3.2), isto quer dizer que existem infinitos valores de n tais que $a_n = 7$. Portanto, provamos que *quase todo número x cuja expansão decimal começa por 7 tem infinitos dígitos iguais a 7.*

Claro que no lugar de 7 podemos considerar qualquer outro dígito. Além disso, também podemos considerar blocos com vários dígitos (Exercício 1.6.14). Mais tarde provaremos um resultado muito mais forte: para quase todo número $x \in [0, 1]$, todo dígito aparece com frequência $1/10$ na expansão decimal de x.

1.3.2 Transformação de Gauss

O sistema que apresentamos nesta seção está relacionado com outro importante algoritmo em Teoria dos Números, a expansão de um número em fração contínua, cuja origem remonta ao problema de achar a melhor aproximação racional para um número real qualquer. Vamos descrever este algoritmo sucintamente.

Dado um número $x_0 \in (0, 1)$, sejam

$$a_1 = \left[\frac{1}{x_0}\right] \qquad \text{e} \qquad x_1 = \frac{1}{x_0} - a_1.$$

Note que a_1 é um número natural, $x_1 \in [0,1)$ e tem-se

$$x_0 = \frac{1}{a_1 + x_1}.$$

Supondo que x_1 seja diferente de zero, podemos repetir o processo, definindo

$$a_2 = \left[\frac{1}{x_1}\right] \qquad \text{e} \qquad x_2 = \frac{1}{x_1} - a_2.$$

Então

$$x_1 = \frac{1}{a_1 + x_2} \quad \text{e portanto} \quad x_0 = \cfrac{1}{a_1 + \cfrac{1}{a_2 + x_2}}.$$

Por recorrência, para cada $n \geq 1$ tal que $x_{n-1} \in (0,1)$ define-se

$$a_n = \left[\frac{1}{x_{n-1}}\right] \qquad \text{e} \qquad x_n = \frac{1}{x_{n-1}} - a_n = G(x_{n-1})$$

e tem-se

$$x_0 = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots + \cfrac{1}{a_n + x_n}}}}. \tag{1.3.3}$$

Pode mostrar-se que a sequência

$$z_n = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots + \cfrac{1}{a_n}}}} \tag{1.3.4}$$

converge para x_0 quando $n \to \infty$, e é usual traduzir este fato escrevendo

$$x_0 = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots + \cfrac{1}{a_n + \cfrac{1}{\cdots}}}}}, \tag{1.3.5}$$

que é chamada *expansão em fração contínua* de x_0.

Note que a sequência $(z_n)_n$ definida pela relação (1.3.4) consiste de números racionais. De fato, mostra-se que estes são os números racionais que melhor aproximam o número x_0, no sentido de que z_n está mais próximo de x_0 do que qualquer outro número racional com denominador menor ou igual que o denominador de z_n (escrito em forma irredutível). Observe também que para obter (1.3.5) supusemos que $x_n \in (0,1)$ para todo $n \in \mathbb{N}$. Se encontramos algum

$x_n = 0$, o processo para nesse momento e consideramos (1.3.3) a expansão em fração contínua de x_0. Claro que este último caso ocorre somente se x_0 é um número racional.

O algoritmo de expansão em fração contínua está intimamente conectado com o sistema dinâmico no intervalo $[0,1]$ que vamos descrever a seguir. A *transformação de Gauss* $G : [0,1] \to [0,1]$ é definida por

$$G(x) = \frac{1}{x} - \left[\frac{1}{x}\right] = \text{parte fracionária de } 1/x,$$

se $x \in (0,1]$ e $G(0) = 0$. O gráfico de G pode ser esboçado facilmente, a partir da seguinte observação: para todo x em cada intervalo $I_k = (1/(k+1), 1/k]$ a parte inteira de $1/x$ é igual a k e, portanto, $G(x) = 1/x - k$.

A expansão em fração contínua de qualquer número $x_0 \in (0,1)$ pode ser obtida a partir da transformação de Gauss, da seguinte forma: para cada $n \geq 1$ o número natural a_n é determinado por

$$G^{n-1}(x_0) \in I_{a_n}$$

e x_n é simplesmente o n-ésimo iterado $G^n(x_0)$ de x_0. Este processo termina se encontrarmos algum $x_n = 0$; como explicamos anteriormente, isto só pode acontecer se o número x_0 for racional (veja o Exercício 1.6.16). Em particular, existe um conjunto com medida de Lebesgue total tal que todos os iterados de G estão definidos para os pontos deste conjunto.

O que torna esta transformação interessante do ponto de vista da Teoria Ergódica é que G admite uma probabilidade invariante que é equivalente à medida de Lebesgue no intervalo. De fato, considere a medida definida por

$$\mu(E) = \int_E \frac{c}{1+x}\,dx \quad \text{para cada mensurável } E \subset [0,1], \tag{1.3.6}$$

onde c é uma constante positiva. Note que a integral está bem definida, já que a função integrando é contínua no intervalo $[0,1]$. Além disso, essa função toma valores no intervalo $[c/2, c]$ e, portanto,

$$\frac{c}{2}\, m(E) \leq \mu(E) \leq c\, m(E) \tag{1.3.7}$$

para todo conjunto mensurável $E \subset [0,1]$. Em particular, μ é de fato equivalente à medida de Lebesgue m, isto é, as duas medidas têm os mesmos conjuntos com medida nula.

Proposição 1.3.2. *A medida μ é invariante por G. Além disso, se escolhermos $c = 1/\log 2$ então μ é uma probabilidade.*

Demonstração. Vamos utilizar o seguinte lema:

Lema 1.3.3. *Seja $f : [0,1] \to [0,1]$ uma transformação tal que existem intervalos abertos $I_1, I_2, \dots$ disjuntos dois-a-dois tais que*

1. *a união* $\cup_k I_k$ *tem medida de Lebesgue total em* $[0,1]$ *e*

2. *a restrição* $f_k = f \mid I_k$ *a cada* I_k *é um difeomorfismo sobre* $(0,1)$.

Seja $\rho : [0,1] \to [0,\infty)$ *uma função integrável (para a medida de Lebesgue) com*

$$\rho(y) = \sum_{x \in f^{-1}(y)} \frac{\rho(x)}{|f'(x)|} \tag{1.3.8}$$

para quase todo $y \in [0,1]$. *Então a medida* $\mu = \rho dx$ *é invariante por* f.

Demonstração. Seja $\phi = \chi_E$ a função característica de um conjunto mensurável $E \subset [0,1]$ qualquer. Pela fórmula de mudança de variáveis,

$$\int_{I_k} \phi(f(x))\rho(x)\,dx = \int_0^1 \phi(y)\rho(f_k^{-1}(y))|(f_k^{-1})'(y)|\,dy.$$

Note que $(f_k^{-1})'(y) = 1/f'(f_k^{-1}(y))$. Portanto, a relação anterior implica que

$$\begin{aligned}\int_0^1 \phi(f(x))\rho(x)\,dx &= \sum_{k=1}^{\infty} \int_{I_k} \phi(f(x))\rho(x)\,dx \\ &= \sum_{k=1}^{\infty} \int_0^1 \phi(y) \frac{\rho(f_k^{-1}(y))}{|f'(f_k^{-1}(y))|}\,dy.\end{aligned} \tag{1.3.9}$$

Usando o teorema da convergência monótona (Teorema A.2.9) e a hipótese (1.3.8), vemos que a última expressão em (1.3.9) é igual a

$$\int_0^1 \phi(y) \sum_{k=1}^{\infty} \frac{\rho(f_k^{-1}(y))}{|f'(f_k^{-1}(y))|}\,dy = \int_0^1 \phi(y)\rho(y)\,dy.$$

Deste jeito mostramos que $\int_0^1 \phi(f(x))\rho(x)\,dx = \int_0^1 \phi(y)\rho(y)\,dy$. Como $\mu = \rho dx$ e $\phi = \mathcal{X}_E$, isto quer dizer que $\mu(f^{-1}(E)) = \mu(E)$ para todo conjunto mensurável $E \subset [0,1]$. Portanto, μ é invariante por f. $\square$

Para concluir a demonstração da Proposição 1.3.2 devemos mostrar que a condição (1.3.8) vale para $\rho(x) = c/(1+x)$ e $f = G$. Seja $I_k = (1/(k+1), 1/k)$ e seja G_k a restrição de G a I_k. Note que $G_k^{-1}(y) = 1/(y+k)$ para todo k. Note também que $G'(x) = (1/x)' = -1/x^2$ para todo $x \neq 0$. Portanto,

$$\sum_{k=1}^{\infty} \frac{\rho(G_k^{-1}(y))}{|G'(G_k^{-1}(y))|} = \sum_{k=1}^{\infty} \frac{c(y+k)}{y+k+1} \Big(\frac{1}{y+k}\Big)^2 = \sum_{k=1}^{\infty} \frac{c}{(y+k)(y+k+1)}. \tag{1.3.10}$$

Observando que

$$\frac{1}{(y+k)(y+k+1)} = \frac{1}{y+k} - \frac{1}{y+k+1}\,,$$

vemos que a última soma em (1.3.10) pode ser escrita na forma telescópica: todos os termos, exceto o primeiro, aparecem duas vezes, com sinais contrários, e portanto se cancelam. Logo a soma é igual ao primeiro termo:

$$\sum_{k=1}^{\infty} \frac{c}{(y+k)(y+k+1)} = \frac{c}{y+1} = \rho(y).$$

Isto mostra que a igualdade (1.3.8) é realmente satisfeita e, portanto, podemos usar o Lema 1.3.1 para concluir que μ é invariante.

Finalmente, usando a primitiva $c\log(1+x)$ da função $\rho(x)$ vemos que

$$\mu([0,1]) = \int_0^1 \frac{c}{1+x}\,dx = c\log 2.$$

Logo, escolhendo $c = 1/\log 2$ obtemos que μ é uma probabilidade. $\square$

Esta proposição permite utilizar ideias de Teoria Ergódica, aplicadas à transformação de Gauss, para obter conclusões interessantes em Teoria dos Números. Por exemplo (veja o Exercício 1.6.15), o número 7 aparece infinitas vezes na expansão em fração contínua de quase todo número $x_0 \in (1/8, 1/7)$, isto é, tem-se $a_n = 7$ para infinitos valores de $n \in \mathbb{N}$. Mais tarde provaremos um fato muito mais preciso, que implica o seguinte: para quase todo $x_0 \in (0,1)$ o número 7 aparece com frequência

$$\frac{1}{\log 2} \log \frac{64}{63}$$

na sua expansão em fração contínua. Tente intuir desde já de onde vem este número!

1.3.3 Rotações no círculo

Considere na reta $\mathbb{R}$ a relação de equivalência $\sim$ que identifica quaisquer números cuja diferença é um número inteiro, isto é:

$$x \sim y \quad \Leftrightarrow \quad x - y \in \mathbb{Z}.$$

Representamos por $[x] \in \mathbb{R}/\mathbb{Z}$ a classe de equivalência de qualquer $x \in \mathbb{R}$ e denotamos por $\mathbb{R}/\mathbb{Z}$ o espaço de todas as classes de equivalência. Este espaço será chamado de *círculo* e também será denotado por S^1. A razão de ser desta terminologia é que $\mathbb{R}/\mathbb{Z}$ pode ser identificado de maneira natural com o círculo unitário no plano complexo, por meio da aplicação

$$\phi : \mathbb{R}/\mathbb{Z} \to \{z \in \mathbb{C} : |z| = 1\}, \quad [x] \mapsto e^{2\pi x i}. \tag{1.3.11}$$

Note que ϕ está bem definida: a expressão $e^{2\pi x i}$ não depende da escolha do representante x na classe $[x]$, uma vez que a função $x \mapsto e^{2\pi x i}$ é periódica de período 1. Além disso, ϕ é uma bijeção.

O círculo herda da reta uma estrutura de grupo abeliano, dada pela operação

$$[x] + [y] = [x + y].$$

Observe que esta definição está bem formulada: a classe de equivalência do lado direito não depende da escolha dos representantes x e y das classes do lado esquerdo. Dado $\theta \in \mathbb{R}$, chamamos *rotação* de ângulo θ a transformação

$$R_\theta : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}, \quad [x] \mapsto [x + \theta] = [x] + [\theta].$$

Note que a aplicação que lhe corresponde em $\{z \in \mathbb{C} : |z| = 1\}$, via a identificação (1.3.11), é o que chamaríamos de rotação de ângulo $2\pi\theta$, ou seja, é a restrição ao círculo unitário da transformação $z \mapsto e^{2\pi\theta i} z$. É imediato da definição que R_0 é a identidade e $R_\theta \circ R_\tau = R_{\theta+\tau}$ para todo θ e τ. Em particular, toda rotação R_θ é invertível e a inversa é $R_{-\theta}$.

Também podemos munir S^1 com uma estrutura natural de espaço de probabilidade, da seguinte forma. Seja $\pi : \mathbb{R} \to S^1$ a projeção canônica que associa a cada $x \in \mathbb{R}$ a respectiva classe de equivalência $[x]$. Primeiramente, dizemos que um conjunto $E \subset S^1$ é mensurável se $\pi^{-1}(E)$ é um subconjunto mensurável da reta. Em seguida, seja m a medida de Lebesgue na reta. Definimos a *medida de Lebesgue* μ no círculo da seguinte forma:

$$\mu(E) = m\big(\pi^{-1}(E) \cap [k, k+1)\big) \quad \text{para qualquer } k \in \mathbb{Z}.$$

Note que o lado esquerdo desta igualdade não depende de k, uma vez, por definição, $\pi^{-1}(E) \cap [k, k+1) = \big(\pi^{-1}(E) \cap [0,1)\big) + k$ e a medida m é invariante por translações.

É claro da definição que μ é uma probabilidade. Além disso, μ é invariante por toda rotação R_θ (trata-se da única medida de probabilidade com esta propriedade, como veremos no Exercício 1.6.20). Isto pode ser mostrado da seguinte forma. Por definição, $\pi^{-1}(R_\theta^{-1}(E)) = \pi^{-1}(E) - \theta$ para todo conjunto mensurável $E \subset S^1$. Seja k a parte inteira de θ. Como m é invariante por translações,

$$\begin{aligned} m\big((\pi^{-1}(E) - \theta) \cap [0,1)\big) &= m\big(\pi^{-1}(E) \cap [\theta, \theta+1)\big) \\ &= m\big(\pi^{-1}(E) \cap [\theta, k+1)\big) + m\big(\pi^{-1}(E) \cap [k+1, \theta+1)\big). \end{aligned}$$

Note que $\pi^{-1}(E) \cap [k+1, \theta+1) = \big(\pi^{-1}(E) \cap [k, \theta)\big) + 1$. Portanto, a expressão no lado direito da igualdade anterior pode ser escrita como

$$m\big(\pi^{-1}(E) \cap [\theta, k+1)\big) + m\big(\pi^{-1}(E) \cap [k, \theta)\big) = m\big(\pi^{-1}(E) \cap [k, k+1)\big).$$

Combinando estas duas igualdades obtemos que

$$\mu\big(R_\theta^{-1}(E)\big) = m\big(\pi^{-1}(R_\theta^{-1}(E) \cap [0,1))\big) = m\big(\pi^{-1}(E) \cap [k, k+1)\big) = \mu(E)$$

para todo conjunto mensurável $E \subset S^1$.

A dinâmica da rotação $R_\theta : S^1 \to S^1$ apresenta dois comportamentos bem distintos, dependendo do valor de θ. Se θ é racional, digamos $\theta = p/q$ com $p \in \mathbb{Z}$ e $q \in \mathbb{N}$, então

$$R_\theta^q([x]) = [x + q\theta] = [x] \quad \text{para todo } [x].$$

Como consequência, todo ponto $x \in S^1$ é periódico de período q. No caso contrário temos:

Proposição 1.3.4. *Se θ é irracional então $\mathcal{O}([x]) = \{R_\theta^n([x]) : n \in \mathbb{N}\}$ é um subconjunto denso de $\mathbb{R}/\mathbb{Z}$ para todo $[x]$.*

Demonstração. Afirmamos que o conjunto $\mathcal{D} = \{m + n\theta : m \in \mathbb{Z}, n \in \mathbb{N}\}$ é denso em $\mathbb{R}$. De fato, considere um número qualquer $r \in \mathbb{R}$. Dado qualquer $\varepsilon > 0$, podemos escolher $p \in \mathbb{Z}$ e $q \in \mathbb{N}$ tais que $|q\theta - p| < \varepsilon$. Note que o número $a = q\theta - p$ é necessariamente diferente de zero, uma vez que θ é irracional. Suponhamos que a é positivo (o outro caso é análogo). Subdividindo a reta em intervalos de comprimento a, vemos que existe um número inteiro l tal que $0 \le r - la < a$. Isto implica que

$$|r - (lq\theta - lp)| = |r - la| < a < \varepsilon.$$

Como $m = lq$ e $n = -lq$ são inteiros e ε é arbitrário, isto mostra que r está no fecho do conjunto $\mathcal{D}$, para todo $r \in \mathbb{R}$.

Agora, dados $y \in \mathbb{R}$ e $\varepsilon > 0$, podemos tomar $r = y - x$ e, usando o parágrafo anterior, podemos encontrar $m, n \in \mathbb{Z}$ tais que $|m + n\theta - (y - x)| < \varepsilon$. Isto equivale a dizer que a distância de $[y]$ ao iterado $R_\theta^n([x])$ é menor que ε. Como x, y e ε são arbitrários, isto mostra que toda órbita $\mathcal{O}([x])$ é densa. $\square$

Em particular, segue que *todo* ponto do círculo é recorrente para R_θ (isto também é verdade quando θ é racional). A proposição anterior também terá várias implicações interessantes no estudo das medidas invariantes de R_θ. Entre outras coisas, veremos posteriormente que se θ é irracional então a medida de Lebesgue é a única medida de probabilidade que é preservada por R_θ. Relacionado com isso, veremos que as órbitas de R_θ se distribuem de modo uniforme em S^1.

1.3.4 Rotações em toros

As noções que acabamos de apresentar podem ser generalizadas para qualquer dimensão, como vamos explicar em seguida. Para cada $d \ge 1$, considere a relação de equivalência em $\mathbb{R}^d$ que identifica dois vetores se a sua diferença é um vetor com coordenadas inteiras:

$$(x_1, \ldots, x_d) \sim (y_1, \ldots, y_d) \quad \Leftrightarrow \quad (x_1 - y_1, \ldots, x_d - y_d) \in \mathbb{Z}^d.$$

Representamos por $[x]$ ou $[(x_1, \ldots, x_d)]$ a classe de equivalência de um vetor $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ qualquer. Chamamos *toro de dimensão* d ou, simplesmente, *d-toro* o espaço

$$\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d = (\mathbb{R}/\mathbb{Z})^d$$

das classes de equivalência. Seja m a medida de volume em $\mathbb{R}^d$. A operação

$$[(x_1, \ldots, x_d)] + [(y_1, \ldots, y_d)] = [(x_1 + y_1, \ldots, x_d + y_d)]$$

faz de $\mathbb{T}^d$ um grupo abeliano. A *rotação* associada a um vetor $\theta = (\theta_1, \ldots, \theta_d)$ é

$$R_\theta : \mathbb{T}^d \to \mathbb{T}^d, \quad R_\theta([x]) = [x] + [\theta].$$

A aplicação

$$\phi : [0,1]^d \to \mathbb{T}^d, \quad (x_1, \ldots, x_d) \mapsto [(x_1, \ldots, x_d)]$$

é sobrejetora e nos permite definir a medida de probabilidade de Lebesgue μ no d-toro, por meio da seguinte fórmula: $\mu(B) = m\big(\phi^{-1}(B)\big)$ para todo $B \subset \mathbb{T}^d$ tal que $\phi^{-1}(B)$ é mensurável. Esta medida é invariante por R_θ para todo θ.

Dizemos que um vetor $\theta = (\theta_1, \ldots, \theta_d)$ é *racionalmente independente* se para quaisquer números inteiros $n_0, n_1, \ldots, n_d$ temos que

$$n_0 + n_1\theta_1 + \cdots + n_d\theta_d = 0 \quad \Rightarrow \quad n_0 = n_1 = \cdots = n_d = 0.$$

Caso contrário dizemos que θ é racionalmente dependente. Pode mostrar-se que θ é racionalmente independente se, e somente se, a rotação R_θ é uma transformação minimal, ou seja, a órbita $\mathcal{O}([x]) = \{R_\theta^n([x]) : n \in \mathbb{N}\}$ é um subconjunto denso de $\mathbb{T}^d$ para todo $[x]$. A este respeito, veja os Exercícios 1.6.21-1.6.22 e também o Corolário 3.2.3.

1.3.5 Transformações conservativas

Seja M um aberto do espaço euclidiano $\mathbb{R}^d$ e seja $f : M \to M$ um difeomorfismo de classe C^1. Isto quer dizer que f é uma bijeção e tanto ele quanto a sua inversa são deriváveis com derivada contínua. Representaremos por vol a medida de Lebesgue, ou medida de volume, em M. A fórmula de mudança de variáveis afirma que, para qualquer conjunto mensurável $B \subset M$,

$$\operatorname{vol}(f(B)) = \int_B |\det Df| \, dx. \tag{1.3.12}$$

Daqui se deduz facilmente o seguinte fato:

Lema 1.3.5. *Um difeomorfismo $f : M \to M$ de classe C^1 deixa invariante a medida de volume se, e somente se, o valor absoluto $|\det Df|$ do seu jacobiano é constante igual a 1.*

Demonstração. Suponha primeiro que o valor absoluto do jacobiano é igual a 1 em todo ponto. Considere um conjunto mensurável E e seja $B = f^{-1}(E)$. A fórmula (1.3.12) dá que

$$\operatorname{vol}(E) = \int_B 1 \, dx = \operatorname{vol}(B) = \operatorname{vol}(f^{-1}(E)).$$

Isto significa que f deixa invariante o volume e, portanto, provamos a parte "se" do enunciado.

Para provar a parte “somente se”, suponha que $|\det Df|$ fosse maior que 1 em algum ponto x. Então, como o jacobiano é contínuo, existiria uma vizinhança U de x e algum número $\sigma > 1$ tais que

$$|\det Df(y)| \geq \sigma \quad \text{para todo } y \in U.$$

Então a fórmula (1.3.12) aplicada a $B = U$ daria

$$\operatorname{vol}(f(U)) \geq \int_U \sigma\, dx \geq \sigma \operatorname{vol}(U).$$

Denotando $E = f(U)$, isto implica que $\operatorname{vol}(E) > \operatorname{vol}(f^{-1}(E))$ e, portanto, f não deixa invariante o volume. Do mesmo modo se mostra que se o valor absoluto do jacobiano é menor que 1 em algum ponto então f não deixa invariante o volume. □

1.3.6 Fluxos conservativos

Agora vamos analisar a questão da invariância da medida de volume no caso de fluxos $f^t : M \to M$, $t \in \mathbb{R}$. Continuamos supondo que M é um aberto do espaço euclideano $\mathbb{R}^d$. Também suporemos que o fluxo é de classe C^1, no sentido de que a aplicação $(t, x) \mapsto f^t(x)$ é de classe C^1. Então cada transformação f^t é um difeomorfismo C^1: a inversa é f^{-t}. Como f^0 é a identidade e o jacobiano varia continuamente, obtemos que $\det Df^t(x) > 0$ em todo ponto.

Aplicando o Lema 1.3.5 neste contexto, obtemos que o fluxo deixa invariante a medida de volume se, e somente se,

$$\det Df^t(x) = 1 \quad \text{para todo } x \in U \text{ e todo } t \in \mathbb{R}. \tag{1.3.13}$$

No entanto esta conclusão não é muito útil na prática porque, em geral, não temos uma expressão explícita para f^t e, portanto, não é claro como verificar a condição (1.3.13). Felizmente, existe uma expressão razoavelmente explícita para o jacobiano, de que iremos falar em seguida, que pode ser usada em muitas situações interessantes.

Suponhamos que o fluxo $f^t : M \to M$ corresponde às trajetórias de um campo de vetores $F : M \to \mathbb{R}^d$ de classe C^1. Em outras palavras, $t \mapsto f^t(x)$ é a solução da equação diferencial

$$\frac{dy}{dt} = F(y) \tag{1.3.14}$$

que tem x como condição inicial (quando tratando de equações diferenciais sempre suporemos que as suas soluções estão definidas para todo tempo).

A *fórmula de Liouville* exprime o jacobiano de f^t em termos do *divergente* $\operatorname{div} F$ do campo de vetores:

$$\det Df^t(x) = \exp\Big(\int_0^t \operatorname{div} F(f^s(x))\, ds\Big) \quad \text{para todo } x \text{ e todo } t.$$

Lembre que o divergente de um campo de vetores F é o traço da sua matriz jacobiana, isto é

$$\operatorname{div} F = \frac{\partial F_1}{\partial x_1} + \cdots + \frac{\partial F_d}{\partial x_d}. \tag{1.3.15}$$

Combinando a fórmula de Liouville com (1.3.13) obtemos:

Lema 1.3.6 (Liouville)**.** *O fluxo $(f^t)_t$ associado a um campo de vetores F de classe C^1 deixa invariante a medida de volume se, e somente se, o divergente de F é identicamente nulo.*

Podemos generalizar esta discussão para o caso em que M é uma variedade riemanniana qualquer, de dimensão $d \geq 2$.

Por simplicidade, suponhamos que a variedade é orientável. Neste caso, a medida de Lebesgue é dada por uma d-forma diferenciável ω, chamada *forma de volume*, que se escreve em coordenadas locais como $\omega = \rho dx_1 \cdots dx_d$. Isto significa que o volume de qualquer conjunto mensurável B contido num domínio de coordenadas locais $(x_1, \ldots, x_d)$ é dado por

$$\operatorname{vol}(B) = \int_B \rho(x_1, \ldots, x_d)\, dx_1 \cdots dx_d.$$

Seja F um campo de vetores de classe C^1 em M. Escrevendo

$$F(x_1, \ldots, x_d) = (F_1(x_1, \ldots, x_d), \ldots, F_d(x_1, \ldots, x_d)),$$

podemos definir o divergente de F como sendo

$$\operatorname{div} F = \frac{\partial(\rho F)}{\partial x_1} + \cdots + \frac{\partial(\rho F)}{\partial x_d}$$

(a definição não depende da escolha das coordenadas locais). Então temos a seguinte versão do teorema de Liouville (a prova pode ser encontrada no livro de Sternberg [Ste58]):

Teorema 1.3.7 (Liouville)**.** *O fluxo $(f^t)_t$ associado a um campo de vetores F de classe C^1 preserva a medida de volume na variedade M se, e somente se,* $\operatorname{div} F = 0$ *em todo ponto.*

Então, segue do teorema de recorrência para fluxos que, se a variedade M tem volume finito (por exemplo, se M é compacta) e $\operatorname{div} F = 0$, então quase todo ponto é recorrente para o fluxo de F.

1.4 Teoremas de recorrência múltipla

Vamos considerar famílias finitas de transformações $f_i : M \to M$, $i = 1, \ldots, q$ que *comutam entre si*, isto é, tais que

$$f_i \circ f_j = f_j \circ f_i \quad \text{para todo } i, j \in \{1, \ldots, q\}.$$

O objetivo é mostrar que os resultados da Seção 1.2 se estendem para este contexto: obtemos pontos que são *simultaneamente* recorrentes por todas as transformações.

O primeiro resultado nesta linha generaliza o teorema de recorrência de Birkhoff (Teorema 1.2.6):

Teorema 1.4.1 (Recorrência múltipla de Birkhoff)**.** *Seja M um espaço métrico compacto e sejam $f_1, \dots, f_q : M \to M$ transformações contínuas que comutam entre si. Então existe $a \in M$ e existe uma sequência $(n_k)_k \to \infty$ tal que*

$$\lim_k f_i^{n_k}(a) = a \quad \textit{para todo } i = 1, \dots, q. \tag{1.4.1}$$

A demonstração deste teorema será dada na Seção 1.4.1. A seguir, discutimos a seguinte generalização do teorema de recorrência de Poincaré:

Teorema 1.4.2 (Recorrência múltipla de Poincaré)**.** *Seja $(M, \mathcal{B}, \mu)$ um espaço de probabilidade e sejam $f_i : M \to M$, $i = 1, \dots, q$ transformações mensuráveis que preservam μ e que comutam entre si. Então, para qualquer conjunto $E \subset M$ com medida positiva, existe $n \geq 1$ tal que*

$$\mu\big(E \cap f_1^{-n}(E) \cap \cdots \cap f_q^{-n}(E)\big) > 0.$$

Em outras palavras, existe algum tempo n tal que os iterados de um subconjunto com medida positiva de pontos de E retornam a E, *simultaneamente* para todas as transformações f_i, nesse momento n.

A demonstração do Teorema 1.4.2 não será apresentada aqui; veja o livro de Furstenberg [Fur77]. Vamos apenas mencionar algumas consequências diretas e, mais tarde, usaremos o teorema para provar o teorema de Szemerédi sobre existência de progressões aritméticas em subconjuntos 'densos' dos números inteiros.

Observe, primeiramente, que o conjunto dos retornos simultâneos é sempre infinito. De fato, seja n qualquer iterado como no enunciado. Aplicando o Teorema 1.4.2 ao conjunto $F = E \cap f_1^{-n}(E) \cap \cdots \cap f_q^{-n}(E)$, obtemos que existe $m \geq 1$ tal que

$$\begin{aligned} \mu\big(E \cap f_1^{-(m+n)}(E) \cap \cdots \cap f_q^{-(m+n)}(E)\big) \\ \geq \mu\big(F \cap f_1^{-m}(F) \cap \cdots \cap f_q^{-m}(F)\big) > 0. \end{aligned}$$

Em outras palavras, $m + n$ também é um retorno simultâneo a E, para algum subconjunto de E com medida positiva.

Segue que, para qualquer conjunto $E \subset M$ com $\mu(E) > 0$ e para μ-quase todo ponto $x \in E$ existem infinitos iterados n que são retornos simultâneos de x a E, ou seja, que satisfazem $f_i^n(x) \in E$ para todo $i = 1, \dots, q$. De fato, suponha que existisse um subconjunto $F \subset E$ com medida positiva tal que todo ponto de F tem um número finito de retornos simultâneos a E. Por um lado, a menos de substituir F por um subconjunto adequado, podemos supor que todos esses retornos simultâneos dos pontos de F ao conjunto E são menores que um dado

$k \geq 1$ fixado. Por outro lado, usando o parágrafo anterior, existe $n > k$ tal que $G = F \cap f_1^{-n}(F) \cap \cdots \cap f_q^{-n}(F)$ tem medida positiva. Ora, é imediato da definição que n é um retorno simultâneo a E para todo $x \in G$. Isto contradiz a escolha de F, provando a nossa afirmação.

Outro corolário simples é o teorema de recorrência múltipla de Birkhoff (Teorema 1.4.1). De fato, se $f_i : M \to M$, $i = 1, 2, \ldots, q$ são transformações contínuas num espaço métrico compacto que comutam entre si, então existe alguma probabilidade invariante μ comum a todas essas transformações. Considere qualquer base enumerável $\{U_k\}$ da topologia de M. De acordo com o parágrafo anterior, para cada k existe um conjunto $\tilde{U}_k \subset U_k$ com medida nula tal que todo ponto de $U_k \setminus \tilde{U}_k$ tem infinitos retornos simultâneos a U_k. Então $\tilde{U} = \cup_k \tilde{U}_k$ tem medida nula e todo ponto do seu complementar é simultaneamente recorrente, no sentido do Teorema 1.4.1.

1.4.1 Teorema de recorrência múltipla de Birkhoff

Vamos tratar o caso em que as transformações $f_1, \ldots, f_q$ são homeomorfismos de M, que é suficiente para os nossos objetivos no presente capítulo. O caso geral pode ser deduzido facilmente usando a ideia de extensão natural.

O teorema pode ser reformulado do seguinte modo útil. Considere a transformação $F : M^q \to M^q$ definida no espaço produto $M^q = M \times \cdots \times M$ por $F(x_1, \ldots, x_q) = (f_1(x_1), \ldots, f_q(x_q))$. Denote por Δ_q a *diagonal* de M^q, ou seja, o subconjunto dos pontos da forma $\tilde{x} = (x, \ldots, x)$. O Teorema 1.4.1 afirma, precisamente, que existe $\tilde{a} \in \Delta_q$ e existe $(n_k)_k \to \infty$ tal que

$$\lim_k F^{n_k}(\tilde{a}) = \tilde{a}. \tag{1.4.2}$$

A demonstração será por indução no número q de transformações. O caso $q = 1$ está contido no Teorema 1.2.6. A seguir, considere qualquer $q \geq 2$ e suponha que o enunciado é verdadeiro para qualquer família formada por $q - 1$ homeomorfismos que comutam entre si. Vamos provar que ele também é verdadeiro para a família $f_1, \ldots, f_q$.

Denote por $\mathcal{G}$ o grupo (abeliano) gerado pelos homeomorfismos $f_1, \ldots, f_q$. Dizemos que um conjunto $X \subset M$ é $\mathcal{G}$-invariante se $g(X) \subset X$ para todo $g \in \mathcal{G}$. Considerando também a inversa g^{-1}, vemos que isto implica $g(X) = X$ para todo $g \in \mathcal{G}$. Tal como fizemos no Teorema 1.2.6, podemos usar o lema de Zorn para concluir que existe algum conjunto $X \subset M$ não-vazio fechado $\mathcal{G}$-invariante minimal (Exercício 1.6.26). O enunciado do teorema não é afetado se substituirmos M por X. Portanto, não constitui restrição supor que o espaço ambiente M é minimal. Essa suposição será usada da seguinte forma:

Lema 1.4.3. *Se M é minimal então para todo aberto não-vazio $U \subset M$ existe um subconjunto finito $\mathcal{H} \subset \mathcal{G}$ tal que*

$$\bigcup_{h \in \mathcal{H}} h^{-1}(U) = M.$$

Demonstração. Dado qualquer $x \in M$, o fecho da órbita $\mathcal{G}(x) = \{g(x) : g \in \mathcal{G}\}$ é um subconjunto não-vazio de M, fechado e $\mathcal{G}$-invariante. Portanto, a hipótese de que M é minimal implica que a órbita $\mathcal{G}(x)$ é densa em M. Em particular, existe $g \in \mathcal{G}$ tal que $g(x) \in U$. Isto prova que $\{g^{-1}(U) : g \in \mathcal{G}\}$ é uma cobertura aberta de M. Por compacidade, segue que existe uma subcobertura finita. Essa é, precisamente, a afirmação no lema. $\square$

Consideraremos o produto M^q munido da distância dada por

$$d\big((x_1, \ldots, x_q), (y_1, \ldots, y_q)\big) = \max\{d(x_i, y_i) : 1 \le i \le q\}.$$

Note que a aplicação $M \to \Delta_q$, $x \mapsto \tilde{x} = (x, \ldots, x)$ é um homeomorfismo e, mesmo, uma isometria para esta escolha da distância. Todo aberto $U \subset M$ corresponde a um aberto $\widetilde{U} \subset \Delta_q$ via esse homeomorfismo. Dado qualquer $g \in \mathcal{G}$, representaremos por $\tilde{g} : M^q \to M^q$ o homeomorfismo definido por $\tilde{g}(x_1, \ldots, x_q) = (g(x_1), \ldots, g(x_q))$. O fato de que $\mathcal{G}$ é abeliano implica que $\tilde{g}$ comuta com F; note também que todo $\tilde{g}$ preserva a diagonal Δ_q. Então a conclusão do Lema 1.4.3 pode ser reescrita na seguinte forma:

$$\bigcup_{h \in \mathcal{H}} \tilde{h}^{-1}(\widetilde{U}) = \Delta_q. \tag{1.4.3}$$

Lema 1.4.4. *Dado $\varepsilon > 0$ existem $\tilde{x}, \tilde{y} \in \Delta_q$ e $n \ge 1$ tais que $d(F^n(\tilde{x}), \tilde{y}) < \varepsilon$.*

Demonstração. Defina $g_i = f_i \circ f_q^{-1}$ para cada $i = 1, \ldots, q-1$. A hipótese de que os f_i comutam entre si implica que o mesmo vale para os g_i. Então, pela hipótese de indução, existe $y \in M$ e $(n_k)_k \to \infty$ tal que

$$\lim_k g_i^{n_k}(y) = y \quad \text{para todo } i = 1, \ldots, q-1.$$

Denote $x_k = f_q^{-n_k}(y)$ e considere $\tilde{x}_k = (x_k, \ldots, x_k) \in \Delta_q$. Então,

$$\begin{aligned} F^{n_k}(\tilde{x}_k) &= (f_1^{n_k} f_q^{-n_k}(y), \ldots, f_{q-1}^{n_k} f_q^{-n_k}(y), f_q^{n_k} f_q^{-n_k}(y)) \\ &= (g_1^{n_k}(y), \ldots, g_{q-1}^{n_k}(y), y) \end{aligned}$$

converge para $(y, \ldots, y, y)$ quando $k \to \infty$. Isto prova o lema com $\tilde{x} = \tilde{x}_k$, $\tilde{y} = (y, \ldots, y, y)$ e $n = n_k$ para qualquer k suficientemente grande. $\square$

Em seguida, mostraremos que o ponto $\tilde{y}$ no Lema 1.4.4 é arbitrário:

Lema 1.4.5. *Dados $\varepsilon > 0$ e $\tilde{z} \in \Delta_q$ existem $\tilde{w} \in \Delta_q$ e $m \ge 1$ satisfazendo $d(F^m(\tilde{w}), \tilde{z}) < \varepsilon$.*

Demonstração. Dados $\varepsilon > 0$ e $\tilde{z} \in \Delta_q$, considere $\widetilde{U} =$ bola aberta de centro $\tilde{z}$ e raio $\varepsilon/2$. Pelo Lema 1.4.3 e pela observação (1.4.3), podemos encontrar um conjunto finito $\mathcal{H} \subset \mathcal{G}$ tal que os conjuntos $\tilde{h}^{-1}(\widetilde{U})$, $h \in \mathcal{H}$ cobrem Δ_q. Como os elementos de $\mathcal{G}$ são (uniformemente) contínuos, existe $\delta > 0$ tal que

$$d(\tilde{x}_1, \tilde{x}_2) < \delta \quad \Rightarrow \quad d(\tilde{h}(\tilde{x}_1), \tilde{h}(\tilde{x}_2)) < \varepsilon/2 \quad \text{para todo } h \in \mathcal{H}.$$

Pelo Lema 1.4.4 existem $\tilde{x}, \tilde{y} \in \Delta_q$ e $n \geq 1$ tais que $d(F^n(\tilde{x}), \tilde{y}) < \delta$. Fixe $h \in \mathcal{H}$ tal que $\tilde{y} \in \tilde{h}^{-1}(\widetilde{U})$. Então,

$$d\big(\tilde{h}(F^n(\tilde{x})), \tilde{z}\big) \leq d\big(\tilde{h}(F^n(\tilde{x})), \tilde{h}(\tilde{y})\big) + d\big(\tilde{h}(\tilde{y}), \tilde{z}\big) < \varepsilon/2 + \varepsilon/2.$$

Tome $\tilde{w} = \tilde{h}(\tilde{x})$. Como $\tilde{h}$ comuta com F^n, a desigualdade anterior dá que $d(F^n(\tilde{w}), \tilde{z}) < \varepsilon$, como queríamos provar. $\square$

Usando o Lema 1.4.5, mostraremos que é possível tomar $\tilde{x} = \tilde{y}$ no Lema 1.4.4:

Lema 1.4.6 (Bowen)**.** *Dado $\varepsilon > 0$ existem $\tilde{v} \in \Delta_q$ e $k \geq 1$ com $d(F^k(\tilde{v}), \tilde{v}) < \varepsilon$.*

Demonstração. Dados $\varepsilon > 0$ e $\tilde{z}_0 \in \Delta_q$, considere as sequências ε_j, m_j e $\tilde{z}_j$, $j \geq 1$, definidas por recorrência da seguinte forma. Inicialmente, tome $\varepsilon_1 = \varepsilon/2$.

- Pelo Lema 1.4.5 existem $\tilde{z}_1 \in \Delta_q$ e $m_1 \geq 1$ tais que $d(F^{m_1}(\tilde{z}_1), \tilde{z}_0) < \varepsilon_1$.
- Por continuidade da aplicação F^{m_1}, existe $\varepsilon_2 < \varepsilon_1$ tal que $d(\tilde{z}, \tilde{z}_1) < \varepsilon_2$ implica $d(F^{m_1}(\tilde{z}), \tilde{z}_0) < \varepsilon_1$.

Em seguida, dado qualquer $j \geq 2$:

- Pelo Lema 1.4.5 existem $\tilde{z}_j \in \Delta_q$ e $m_j \geq 1$ tais que $d(F^{m_j}(\tilde{z}_j), \tilde{z}_{j-1}) < \varepsilon_j$.
- Por continuidade de F^{m_j}, existe algum número $\varepsilon_{j+1} < \varepsilon_j$ tal que $d(\tilde{z}, \tilde{z}_j) < \varepsilon_{j+1}$ implica $d(F^{m_j}(\tilde{z}), \tilde{z}_{j-1}) < \varepsilon_j$.

Em particular, para quaisquer $i < j$,

$$d(F^{m_{i+1}+\cdots+m_j}(\tilde{z}_j), \tilde{z}_i) < \varepsilon_{i+1} \leq \frac{\varepsilon}{2}.$$

Como Δ_q é compacto, podemos encontrar i, j com $i < j$ tais que $d(\tilde{z}_i, \tilde{z}_j) < \varepsilon/2$. Tome $k = m_{i+1} + \cdots + m_j$. Então

$$d(F^k(\tilde{z}_j), \tilde{z}_j) \leq d(F^k(\tilde{z}_j), \tilde{z}_i) + d(\tilde{z}_i, \tilde{z}_j) < \varepsilon.$$

Isto completa a demonstração do lema. $\square$

Agora estamos prontos para concluir a demonstração do Teorema 1.4.1. Para tal, consideremos a função

$$\phi : \Delta_q \to [0, \infty), \quad \phi(\tilde{x}) = \inf\{d(F^n(\tilde{x}), \tilde{x}) : n \geq 1\}.$$

Observe que ϕ é semicontínua superiormente: dado qualquer $\varepsilon > 0$, todo ponto $\tilde{x}$ admite alguma vizinhança V tal que $\phi(\tilde{y}) < \phi(\tilde{x}) + \varepsilon$ para todo $y \in V$. Isso segue imediatamente do fato de que ϕ é dada pelo ínfimo de uma família de funções contínuas. Então (Exercício 1.6.28), ϕ admite algum ponto de continuidade $\tilde{a}$. Vamos mostrar que este ponto satisfaz a conclusão do Teorema 1.4.1.

Para isso, começamos por observar que $\phi(\tilde{a}) = 0$. De fato, suponha que $\phi(\tilde{a})$ é positivo. Então, por continuidade, existem $\beta > 0$ e uma vizinhança V de $\tilde{a}$ tais que $\phi(\tilde{y}) \geq \beta > 0$ para todo $\tilde{y} \in V$. Então,

$$d(F^n(\tilde{y}), \tilde{y}) \geq \beta \quad \text{para todo } y \in V \text{ e todo } n \geq 1. \tag{1.4.4}$$

Por outro lado, de acordo com (1.4.3), para todo $\tilde{x} \in \Delta_q$ existe $h \in \mathcal{H}$ tal que $\tilde{h}(\tilde{x}) \in V$. Como as transformações h são uniformemente contínuas, podemos fixar $\alpha > 0$ tal que

$$d(\tilde{z}, \tilde{w}) < \alpha \quad \Rightarrow \quad d\big(\tilde{h}(\tilde{z}), \tilde{h}(\tilde{w})\big) < \beta \quad \text{para todo } h \in \mathcal{H}. \tag{1.4.5}$$

Pelo Lema 1.4.6, existe $n \geq 1$ tal que $d(\tilde{x}, F^n(\tilde{x})) < \alpha$. Então, usando (1.4.5) e lembrando que F comuta com todo $\tilde{h}$,

$$d\big(\tilde{h}(\tilde{x}), F^n(\tilde{h}(\tilde{x}))\big) < \beta.$$

Isto contradiz (1.4.4). Esta contradição mostra que $\phi(\tilde{a}) = 0$, tal como afirmamos.

Em outras palavras, existe $(n_k)_k \to \infty$ tal que $d(F^{n_k}(\tilde{a}), \tilde{a}) \to 0$ quando $k \to \infty$. Isto significa que (1.4.2) é satisfeita e, portanto, a prova do teorema está completa.

1.5 Progressões aritméticas

Nesta seção vamos provar dois resultados fundamentais da Aritmética Combinatória, o teorema de van der Waerden e o teorema de Szemerédi, a partir dos teoremas de recorrência múltipla (Teorema 1.4.1 e Teorema 1.4.2) que foram apresentados na Seção 1.4.

Chamamos *partição* do conjunto dos números inteiros a qualquer família finita de conjuntos $S_1, \ldots, S_k \subset \mathbb{Z}$ disjuntos dois-a-dois e cuja união é todo o $\mathbb{Z}$. Lembre que uma *progressão aritmética* (finita) é uma sequência da forma

$$m + n, m + 2n, \ldots, m + qn, \quad \text{com } m \in \mathbb{Z} \text{ e } n, q \geq 1.$$

O número q é chamado *comprimento* da progressão.

O seguinte resultado foi obtido, originalmente, pelo matemático holandês Bartel van der Waerden [vdW27] nos anos 20 do século passado:

Teorema 1.5.1 (van der Waerden)**.** *Dada qualquer partição $\{S_1, \ldots, S_l\}$ de $\mathbb{Z}$ existe algum $j \in \{1, \ldots, l\}$ tal que S_j contém progressões aritméticas de todos os comprimentos. Em outras palavras, para todo $q \geq 1$ existem $m \in \mathbb{Z}$ e $n \geq 1$ tais que $m + in \in S_j$ para todo $1 \leq i \leq q$.*

Algum tempo depois, os matemáticos húngaros Pál Erdös e Pál Turán [ET36] formularam a seguinte conjectura, que é mais forte que o teorema de van der Waerden: *todo conjunto $S \subset \mathbb{Z}$ cuja densidade superior é positiva contém progressões aritméticas de comprimento arbitrário.* Esta conjectura foi demonstrada por outro matemático húngaro, Endre Szemerédi [Sze75], quase quatro décadas mais tarde. Para enunciarmos o teorema de Szemerédi precisamos introduzir a noção de densidade superior de um subconjunto de $\mathbb{Z}$.

Chamamos *intervalo* do conjunto $\mathbb{Z}$ dos números inteiros qualquer subconjunto I da forma $\{n \in \mathbb{Z} : a \leq n < b\}$, com $a \leq b$ em $\mathbb{Z}$. O *cardinal* do intervalo é o número $\#I = b - a$.

A *densidade superior* $D_s(S)$ de um subconjunto S de $\mathbb{Z}$ é o número

$$D_s(S) = \limsup_{\#I \to \infty} \frac{\#(S \cap I)}{\#I}$$

onde I representa qualquer intervalo em $\mathbb{Z}$. Do mesmo modo se define a *densidade inferior* $D_i(S)$, trocando limite superior por limite inferior. Em outras palavras, $D_s(S)$ é o maior número D tal que existe uma sequência de intervalos $I_j \subset \mathbb{Z}$ satisfazendo

$$\#I_j \to \infty \quad \text{e} \quad \frac{\#(S \cap I_j)}{\#I_j} \to D$$

e $D_i(S)$ é o menor número nessas condições.

No próximo lema colecionamos algumas propriedades simples destas noções. A demonstração do lema fica a cargo do leitor (Exercício 1.6.31).

Lema 1.5.2. *Tem-se* $0 \leq D_i(S) \leq D_s(S) \leq 1$ *e* $D_i(S) = 1 - D_s(\mathbb{Z} \setminus S)$ *para todo* $S \subset \mathbb{Z}$. *Além disso, se* $S_1, \ldots, S_l$ *é uma partição de* $\mathbb{Z}$ *então*

$$D_i(S_1) + \cdots + D_i(S_l) \leq 1 \leq D_s(S_1) + \cdots + D_s(S_l).$$

Exemplo 1.5.3. Seja S o conjunto dos números pares. Dado qualquer intervalo $I \subset \mathbb{Z}$, temos que $\#(S \cap I) = \#I/2$ se o cardinal de I é par e $\#(S \cap I) = (\#I \pm 1)/2$ se o cardinal de I é ímpar, onde o sinal $\pm$ é positivo se o menor elemento de I é um número par e é negativo caso contrário. Desta observação segue, imediatamente, que $D_s(S) = D_i(S) = 1/2$.

Exemplo 1.5.4. Seja S o seguinte subconjunto de $\mathbb{Z}$:

$$\{1, 3, 4, 7, 8, 9, 13, 14, 15, 16, 21, 22, 23, 24, 25, 31, 32, 33, 34, 35, 36, 43, \ldots\}.$$

Isto é, para cada $k \geq 1$ incluímos em S um bloco de k inteiros consecutivos e omitimos os k inteiros seguintes. Este conjunto contém intervalos com comprimento arbitrariamente grande. Portanto $D_s(S) = 1$. Por outro lado, o complementar de S também contém intervalos com comprimento arbitrariamente grande. Portanto, $D_i(S) = 1 - D_s(\mathbb{Z} \setminus S) = 0$.

Observe que em qualquer destes dois exemplos o conjunto S contém progressões aritméticas de qualquer comprimento. De fato, no Exemplo 1.5.3 o conjunto S até contém progressões aritméticas de comprimento infinito. Isso não é verdade no Exemplo 1.5.4, uma vez que nesse caso o complementar de S contém intervalos arbitrariamente longos.

Teorema 1.5.5 (Szemerédi)**.** *Se S é um subconjunto de $\mathbb{Z}$ com densidade superior positiva, então ele contém progressões aritméticas de comprimento arbitrário.*

Observe que o teorema de van der Waerden é realmente uma consequência fácil do teorema de Szemerédi. De fato, segue do Lema 1.5.2 que se $S_1, \ldots, S_l$ é uma partição de $\mathbb{Z}$ então existe j tal que $D_s(S_j) > 0$. Pelo Teorema 1.5.5, tal S_j contém progressões aritméticas de comprimento arbitrário.

As primeiras demonstrações destes resultados foram de natureza combinatória. Furstenberg (veja [Fur81]) observou que eles podem também ser deduzidos de ideias da Teoria Ergódica: mostraremos na Seção 1.5.1 como obter o teorema de van der Waerden a partir do teorema de recorrência múltipla de Birkhoff (Teorema 1.4.1); argumentos análogos dão o teorema de Szemerédi a partir do teorema de recorrência múltipla de Poincaré (Teorema 1.4.2), como veremos na Seção 1.5.2.

A teoria de Szemerédi continua sendo uma área de pesquisa muito ativa. Em particular, outras demonstrações do Teorema 1.5.5 vêm sendo dadas por diversos autores. Recentemente, a teoria culminou no seguinte resultado espetacular do matemático inglês Ben Green e do matemático australiano Terence Tao [GT08]: *existem progressões aritméticas arbitrariamente longas formadas por números primos.*

O conjunto dos números primos não tem densidade positiva e, portanto, o teorema de Green-Tao não é consequência do teorema de Szemerédi. No entanto, este último tem um papel importante na sua demonstração. Por outro lado, o teorema de Green-Tao é um caso particular de outra conjectura devida a Erdös: se $S \subset \mathbb{N}$ é tal que a soma dos inversos diverge, ou seja, tal que

$$\sum_{n \in S} \frac{1}{n} = \infty,$$

então S contém progressões aritméticas de qualquer comprimento. Esta afirmação mais geral permanece em aberto.

1.5.1 Teorema de van der Waerden

Nesta seção vamos demonstrar o Teorema 1.5.1. A ideia é reduzir a conclusão a uma afirmação sobre o deslocamento à esquerda

$$\sigma : \Sigma \to \Sigma, \quad (\alpha_n)_{n \in \mathbb{Z}} \mapsto (\alpha_{n+1})_{n \in \mathbb{Z}}$$

no espaço $\Sigma = \{1, 2, \ldots, l\}^{\mathbb{Z}}$ das sequências bilaterais com valores no conjunto $\{1, 2, \ldots, l\}$, a qual será provada por meio do teorema de recorrência múltipla de Birkhoff.

Observe que toda partição $\{S_1, \ldots, S_l\}$ de $\mathbb{Z}$ em l subconjuntos determina um elemento $\underline{\alpha} = (\alpha_n)_{n \in \mathbb{Z}}$ de Σ, definido por $\alpha_n = i \Leftrightarrow n \in S_i$. Reciprocamente, todo $\underline{\alpha} \in \Sigma$ define uma partição de $\mathbb{Z}$ em subconjuntos

$$S_i = \{n \in \mathbb{Z} : \alpha_n = i\}, \quad i = 1, \ldots, l.$$

Vamos mostrar que para todo $\underline{\alpha} \in \Sigma$ e todo $q \geq 1$, existem $m \in \mathbb{Z}$ e $n \geq 1$ tais que

$$\alpha_{m+n} = \cdots = \alpha_{m+qn}. \tag{1.5.1}$$

Em vista do que acabamos de observar, isto significa que para toda partição $\{S_1, \ldots, S_l\}$ e todo $q \geq 1$ existe $i \in \{1, \ldots, l\}$ tal que S_i contém alguma progressão aritmética de comprimento q. Como a família dos S_i é finita, isso implica que algum S_j contém progressões aritméticas de comprimento arbitrariamente grande. É claro que uma progressão aritmética de comprimento q contém progressões aritméticas de todos os comprimentos menores que q. Portanto, segue que S_j contém progressões aritméticas de todos os comprimentos, tal como é afirmado no teorema. Resta provar a afirmação em (1.5.1).

Para tal, consideremos em Σ a distância $d(\underline{\beta}, \underline{\gamma}) = 2^{-N(\underline{\beta}, \underline{\gamma})}$, onde

$$N(\underline{\beta}, \underline{\gamma}) = \max \big\{ N \geq 0 : \beta_n = \gamma_n \text{ para todo } n \in \mathbb{Z} \text{ com } |n| < N \big\}.$$

Note que

$$d(\underline{\beta}, \underline{\gamma}) < 1 \quad \text{se, e somente se, } \beta_0 = \gamma_0. \tag{1.5.2}$$

Como o espaço métrico (Σ, d) é compacto, o fecho $Z = \overline{\{\sigma^n(\underline{\alpha}) : n \in \mathbb{Z}\}}$ da trajetória de $\underline{\alpha}$ é também um compacto. Além disso, Z é invariante pelo deslocamento. Consideremos as transformações $f_1 = \sigma$, $f_2 = \sigma^2$, $\ldots$, $f_q = \sigma^q$ definidas de Z em Z. É claro que as f_i comutam entre si. Portanto, podemos aplicar o Teorema 1.4.1 para concluir que existe $\underline{\theta} \in Z$ e uma sequência $(n_k)_k \to \infty$ tal que

$$\lim_k f_i^{n_k}(\underline{\theta}) = \underline{\theta} \quad \text{para todo } i = 1, 2, \ldots, q.$$

Observe que $f_i^{n_j} = \sigma^{i\, n_j}$. Em particular, podemos fixar $n = n_j$ tal que os iterados $\sigma^n(\underline{\theta})$, $\sigma^{2n}(\underline{\theta})$, $\ldots$, $\sigma^{qn}(\underline{\theta})$ estão todos a distância menor que $1/2$ do ponto $\underline{\theta}$. Consequentemente,

$$d\big(\sigma^{in}(\underline{\theta}), \sigma^{jn}(\underline{\theta})\big) < 1 \quad \text{para todo } 1 \leq i, j \leq q.$$

Então, como $\underline{\theta}$ está no fecho Z da órbita de $\underline{\alpha}$, podemos encontrar $m \in \mathbb{Z}$ tal que $\sigma^m(\underline{\alpha})$ está tão próximo de $\underline{\theta}$ que

$$d\big(\sigma^{m+in}(\underline{\alpha}), \sigma^{m+jn}(\underline{\alpha})\big) < 1 \quad \text{para todo } 1 \leq i, j \leq q.$$

Levando em conta a observação (1.5.2) e a definição do deslocamento σ, isto quer dizer que $\alpha_{m+n} = \cdots = \alpha_{m+qn}$, como pretendíamos provar. Isto completa a demonstração do teorema de van der Waerden.

1.5.2 Teorema de Szemerédi

Agora vamos demonstrar o Teorema 1.5.5. Para isso, usaremos o mesmo tipo de dicionário entre partições de $\mathbb{Z}$ e sequências de inteiros que foi usado na seção anterior para provar o teorema de van der Waerden.

Considere S um conjunto com densidade superior positiva, isto é, tal que existe $c > 0$ e existem intervalos $I_j = [a_j, b_j)$ de $\mathbb{Z}$ tais que

$$\lim_j \#I_j = \infty \quad \text{e} \quad \lim_j \frac{\#(S \cap I_j)}{\#I_j} \geq c.$$

Associamos a S a sequência $\underline{\alpha} = (\alpha_j)_{j \in \mathbb{Z}} \in \Sigma = \{0,1\}^{\mathbb{Z}}$ definida por:

$$\alpha_j = 1 \Leftrightarrow j \in S.$$

Considere o deslocamento $\sigma : \Sigma \to \Sigma$ e o subconjunto $A = \{\underline{\alpha} \in \Sigma : \alpha_0 = 1\}$ de Σ. Note que A é um aberto e também um fechado, uma vez que tanto ele quanto o seu complementar são cilindros de Σ. Note também que, para qualquer $j \in \mathbb{Z}$,

$$\sigma^j(\underline{\alpha}) \in A \Leftrightarrow \alpha_j = 1 \Leftrightarrow j \in S.$$

Logo, para mostrar o teorema de Szemerédi basta provar que para todo $k \in \mathbb{N}$ existem $m \in \mathbb{Z}$ e $n \geq 1$ tais que

$$\sigma^{m+n}(\underline{\alpha}), \sigma^{m+2n}(\underline{\alpha}), \ldots, \sigma^{m+kn}(\underline{\alpha}) \in A. \tag{1.5.3}$$

Para tal, considere a sequência μ_j de probabilidades definidas em Σ por:

$$\mu_j = \frac{1}{\#I_j} \sum_{i \in I_j} \delta_{\sigma^i(\underline{\alpha})}. \tag{1.5.4}$$

Como o conjunto $\mathcal{M}_1(\Sigma)$ das probabilidades em Σ é compacto, a menos de substituir $(\mu_j)_j$ por uma subsequência, podemos supor que ela converge na topologia fraca* para alguma probabilidade μ de Σ.

Observe que μ é uma probabilidade σ-invariante pois, para toda função contínua $\varphi : \Sigma \to \mathbb{R}$, vale

$$\begin{aligned}\int (\varphi \circ \sigma)\, d\mu_j &= \frac{1}{\#I_j} \sum_{i \in I_j} \varphi(\sigma^i(\underline{\alpha})) + \frac{1}{\#I_j}\Big[\varphi(\sigma^{b_j}(\underline{\alpha})) - \varphi(\sigma^{a_j}(\underline{\alpha}))\Big] \\ &= \int \varphi\, d\mu_j + \frac{1}{\#I_j}\Big[\varphi(\sigma^{b_j}(\underline{\alpha})) - \varphi(\sigma^{a_j}(\underline{\alpha}))\Big]\end{aligned}$$

e, passando ao limite quando $j \to \infty$, isto dá que $\int (\varphi \circ \sigma)\, d\mu = \int \varphi\, d\mu$. Observe também que $\mu(A) > 0$. De fato, como A é fechado,

$$\mu(A) \geq \limsup_j \mu_j(A) = \limsup_j \frac{\#(S \cap I_j)}{\#I_j} \geq c.$$

Dado qualquer $k \geq 1$, considere as transformações $f_i = \sigma^i$ para $i = 1, \ldots, k$. É claro que estas transformações comutam entre si. Então, estamos em condições de aplicar o Teorema 1.4.2 e concluir que existe algum $n \geq 1$ tal que

$$\mu\big(A \cap \sigma^{-n}(A) \cap \cdots \cap \sigma^{-kn}(A)\big) > 0.$$

Como A é aberto, isto implica que

$$\mu_l\big(A \cap \sigma^{-n}(A) \cap \cdots \cap \sigma^{-kn}(A)\big) > 0$$

para qualquer l suficientemente grande. Pela definição de μ_l em (1.5.4), isto quer dizer que existe algum $m \in I_l$ tal que

$$\sigma^m(\underline{\alpha}) \in A \cap \sigma^{-n}(A) \cap \cdots \cap \sigma^{-kn}(A).$$

Em particular, $\sigma^{m+in}(\underline{\alpha}) \in A$ para todo $i = 1, \ldots, k$, como queríamos provar.

1.6 Exercícios

1.6.1. Seja $f : M \to M$ uma transformação mensurável. Mostre que uma medida de Dirac δ_p é invariante por f se, e somente se, p é ponto fixo de f. Mais geralmente, a probabilidade $\delta_{p,k} = k^{-1}\big(\delta_p + \delta_{f(p)} + \cdots + \delta_{f^{k-1}(p)}\big)$ é invariante por f se, e somente se, $f^k(p) = p$.

1.6.2. Prove a seguinte versão da Proposição 1.1.1. Sejam M um espaço métrico, $f : M \to M$ uma transformação mensurável e μ uma medida em M. Mostre que se

$$\int \phi \, d\mu = \int \phi \circ f \, d\mu$$

para toda função contínua limitada $\phi : M \to \mathbb{R}$ então f preserva a medida μ.

1.6.3. Prove que se $f : M \to M$ preserva uma medida μ então, dado qualquer $k \geq 2$, o iterado f^k preserva μ. Decida se a recíproca é verdadeira.

1.6.4. Suponha que $f : M \to M$ preserva uma probabilidade μ. Seja $B \subset M$ um conjunto mensurável que satisfaz qualquer uma das seguintes condições:

1. $\mu(B \setminus f^{-1}(B)) = 0$;
2. $\mu(f^{-1}(B) \setminus B) = 0$;
3. $\mu(B \Delta f^{-1}(B)) = 0$;
4. $f(B) \subset B$.

Mostre que existe $C \subset M$ tal que $f^{-1}(C) = C$ e $\mu(B \Delta C) = 0$.

1.6.5. Seja $f : U \to U$ um difeomorfismo C^1 de um aberto $U \subset \mathbb{R}^d$. Mostre que a medida de Lebesgue m é invariante por f se, e somente se, $|\det Df| \equiv 1$.

1.6.6. Mostre que o seguinte enunciado é equivalente ao Teorema 1.2.1, isto é, qualquer um deles pode ser obtido a partir do outro. Seja $f : M \to M$ uma transformação mensurável e seja μ uma medida invariante finita. Seja $E \subset M$ qualquer conjunto mensurável com $\mu(E) > 0$. Então existe $N \geq 1$ e um conjunto $D \subset E$ com medida positiva, tal que $f^N(x) \in E$ para todo ponto $x \in D$.

1.6.7. Seja $f : M \to M$ uma transformação invertível e suponha que μ é uma medida invariante não necessariamente finita. Seja $B \subset M$ um conjunto com medida finita. Mostre que, dado qualquer conjunto mensurável $E \subset M$ com medida positiva, quase todo ponto $x \in E$ regressa infinitas vezes a E ou tem apenas um número finito de iterados em B.

1.6.8. Seja $f : M \to M$ uma transformação invertível e suponha que μ é uma medida invariante σ-*finita*: existe uma sequência crescente de subconjuntos mensuráveis M_k com $\mu(M_k) < \infty$ para todo k e $\cup_k M_k = M$. Dizemos que um ponto x *vai para infinito* se, para qualquer k, existe apenas um número finito de iterados de x que estão em M_k. Mostre que, dado qualquer conjunto mensurável $E \subset M$ com medida positiva, quase todo ponto $x \in E$ regressa a E infinitas vezes ou vai para infinito.

1.6.9. Sejam $f : M \to M$ uma transformação, não necessariamente invertível, μ uma probabilidade invariante e $D \subset M$ um conjunto com medida positiva. Prove que quase todo ponto de D passa uma fração positiva do tempo em D:

$$\limsup_n \frac{1}{n} \#\{0 \le j \le n-1 : f^j(x) \in D\} > 0$$

para μ-quase todo ponto $x \in D$. [Observação: Dá para substituir lim sup por lim inf no enunciado, mas a prova desse fato terá que esperar até o Capítulo 2.]

1.6.10. Seja $f : M \to M$ uma transformação mensurável que preserva uma medida finita μ. Dado qualquer conjunto mensurável $A \subset M$ com $\mu(A) > 0$, seja $n_1 < n_2 < \cdots$ a sequência dos valores de n tais que $\mu(f^{-n}(A) \cap A) > 0$. O objetivo deste exercício é mostrar que o conjunto $V_A = \{n_1, n_2, \dots\}$ é *sindético*, ou seja, que existe $C > 0$ tal que $n_{i+1} - n_i \le C$ para todo i.

1. Mostre que para qualquer sequência crescente $k_1 < k_2 < \cdots$ existem $j > i \ge 1$ tal que $\mu(A \cap f^{-(k_j - k_i)}(A)) > 0$.
2. Dada qualquer sequência infinita $\ell = (l_j)_j$ de números naturais, denote por $S(\ell)$ o conjunto de todas as somas finitas de elementos contíguos de ℓ. Mostre que V_A intersecta $S(\ell)$ qualquer que seja ℓ.
3. Deduza que o conjunto V_A é sindético.

[Observação: Veremos outra demonstração deste fato no Exercício 2.3.2.]

1.6.11. Mostre que se $f : [0,1] \to [0,1]$ é uma transformação mensurável preservando a medida de Lebesgue m então m-quase todo ponto $x \in [0,1]$ satisfaz

$$\liminf_n n|f^n(x) - x| \le 1.$$

[Observação: Boshernitzan [Bos93] provou um resultado bastante mais geral, a saber que $\liminf_n n^{1/d} d(f^n(x), x) < \infty$ para μ-quase todo ponto e toda probabilidade μ invariante por $f : M \to M$, se M é um espaço métrico separável cuja medida de Hausdorff d-dimensional é σ-finita.]

1.6.12. Seja $\omega = (1+\sqrt{5})/2$ a *razão áurea* e seja $f : [0,1] \to [0,1]$ a transformação definida por $f(x) = (x+\omega) - [x+\omega]$. Dado x, verifique que $n|f^n(x) - x| = n^2|\omega - q_n|$ para todo n, onde $(q_n)_n \to \omega$ é a sequência de números racionais dada por $q_n = [x+n\omega]/n$. Usando que as raízes do polinômio $R(z) = z^2 - z - 1$ são ω e $\omega - \sqrt{5}$, mostre que $\liminf_n n^2|\omega - q_n| \geq 1/\sqrt{5}$. [Observação: Isto mostra que a constante 1 no Exercício 1.6.11 não pode ser substituída por nenhuma outra menor que $1/\sqrt{5}$. Não é conhecido se 1 é a menor constante que vale para *toda* transformação no intervalo.]

1.6.13. Utilizando o Lema 1.3.3, dê outra prova de que a transformação expansão decimal $f(x) = 10x - [10x]$ preserva a medida de Lebesgue no intervalo.

1.6.14. Prove que, para quase todo número $x \in [0,1]$ cuja expansão decimal contém o bloco 617 (por exemplo $x = 0{,}3375\mathbf{617}264\cdots$), esse bloco aparece infinitas vezes na expansão. Vá mais longe e mostre que, de fato, o bloco 617 aparece infinitas vezes na expansão decimal de quase todo $x \in [0,1]$.

1.6.15. Para (Lebesgue) quase todo número $x_0 \in (1/618, 1/617)$ o número 617 aparece infinitas vezes na sua expansão em fração contínua, isto é, $a_n = 617$ para infinitos valores de $n \in \mathbb{N}$.

1.6.16. Seja G a transformação de Gauss. Mostre que um número $x \in (0,1)$ é racional se, e somente se, existe $n \geq 1$ tal que $G^n(x) = 0$.

1.6.17. Considere a sequência $1, 2, 4, 8, \ldots, a_n = 2^n, \ldots$ das potências de 2. Mostre que dado qualquer dígito $i \in \{1, \ldots, 9\}$, existe uma quantidade infinita de valores n tais que a_n começa com este dígito.

1.6.18. Prove a seguinte extensão do Lema 1.3.3. Suponha que $f : M \to M$ é um difeomorfismo local de classe C^1 de uma variedade riemanniana compacta M. Seja vol a medida de volume em M e seja $\rho : M \to [0,\infty)$ uma função contínua. Mostre que f preserva a medida $\mu = \rho\,\mathrm{vol}$ se, e somente se,

$$\sum_{x \in f^{-1}(y)} \frac{\rho(x)}{|\det Df(x)|} = \rho(y) \quad \text{para todo } y \in M.$$

No caso em que f é invertível isto significa que f preserva a medida μ se, e somente se, $\rho(x) = \rho(f(x))|\det Df(x)|$ para todo $x \in M$.

1.6.19. Mostre que se A é uma matriz $n \times n$ com coeficientes inteiros e determinante diferente de zero, então a transformação $f_A : \mathbb{T}^d \to \mathbb{T}^d$ definida por $f_A([x]) = [A(x)]$ preserva a medida de Lebesgue de $\mathbb{T}^d$.

1.6.20. Mostre que a medida de Lebesgue em S^1 é a única probabilidade no círculo S^1 que é invariante por todas as rotações. De fato, ela é a única probabilidade invariante por todas as rotações *racionais* de S^1.

1.6.21. Suponha que $\theta = (\theta_1, \ldots, \theta_d)$ é vetor racionalmente dependente. Mostre que existe alguma função contínua $\varphi : \mathbb{T}^d \to \mathbb{C}$ não constante tal que $\varphi \circ R_\theta = \varphi$. Conclua que existem abertos U e V, não vazios, disjuntos e invariantes por R_θ, ou seja, tais que $R_\theta(U) = U$ e $R_\theta(V) = V$. Deduza que nenhuma órbita $\mathcal{O}([x])$ da rotação R_θ é densa em $\mathbb{T}^d$.

1.6.22. Suponha que $\theta = (\theta_1, \dots, \theta_d)$ é vetor racionalmente independente. Mostre que se $V \subset \mathbb{T}^d$ é aberto, não vazio, invariante por R_θ, então V é denso no toro. Conclua que $\cup_{n\in\mathbb{Z}} R_\theta^n(U)$ é denso no toro, qualquer que seja o aberto não vazio U. Conclua que existe $[x]$ cuja órbita $\mathcal{O}([x])$ pela rotação R_θ é densa em $\mathbb{T}^d$. Deduza que $\mathcal{O}([y])$ é densa em $\mathbb{T}^d$ para todo $[y]$.

1.6.23. Seja U um aberto de $\mathbb{R}^{2d}$ e $H : U \to \mathbb{R}$ uma função de classe C^2. Representamos as variáveis em $\mathbb{R}^{2d}$ por $(p_1, \dots, p_d, q_1, \dots, q_d)$. O campo de vetores hamiltoniano associado a H é definido por

$$F(p_1, \dots, p_d, q_1, \dots, q_d) = \left(\frac{\partial H}{\partial q_1}, \dots, \frac{\partial H}{\partial q_d}, -\frac{\partial H}{\partial p_1}, \dots, -\frac{\partial H}{\partial p_d}\right).$$

Verifique que o fluxo definido por F preserva o volume.

1.6.24. Seja $f : U \to U$ um difeomorfismo de classe C^1 preservando a medida de Lebesgue num aberto U de $\mathbb{R}^d$. Seja $H : U \to \mathbb{R}$ uma integral primeira de f, ou seja, uma função de classe C^1 tal que $H \circ f = H$. Seja c um valor regular de H e seja ds a medida de volume definida na hipersuperfície $H_c = H^{-1}(c)$ pela restrição da métrica riemanniana de $\mathbb{R}^d$. Mostre que a restrição de f à hipersuperfície H_c preserva a medida $ds/\|\operatorname{grad} H\|$.

1.6.25. Mostre, por meio de exemplos, que a conclusão do Teorema 1.4.1 é falsa, em geral, se as transformações f_i não comutam.

1.6.26. Seja $\mathcal{G}$ o grupo abeliano gerado por homeomorfismos $f_1, \dots, f_q : M \to M$ num espaço métrico compacto que comutam entre si. Mostre que existe $X \subset M$ minimal para a relação de inclusão na família dos fechados, $\mathcal{G}$-invariantes, não vazios.

1.6.27. Mostre que se $\varphi : M \to \mathbb{R}$ é uma função semicontínua superiormente num espaço métrico compacto então φ atinge o seu supremo, isto é, existe $p \in M$ tal que $\varphi(p) \geq \varphi(x)$ para todo $x \in M$.

1.6.28. Mostre que se $\varphi : M \to \mathbb{R}$ é uma função semicontínua (superiormente ou inferiormente) num espaço métrico compacto então o conjunto dos pontos de continuidade de φ contém uma interseção enumerável de subconjuntos abertos e densos de M. Em particular, o conjunto dos pontos de continuidade é denso em M.

1.6.29. Seja $f : M \to M$ uma transformação mensurável preservando uma medida finita μ. Dado $k \geq 1$ e $A \subset M$ com medida positiva, mostre que para quase todo $x \in A$ existe $n \geq 1$ tal que $f^{jn}(x) \in A$ para todo $1 \leq j \leq k$.

1.6.30. Sejam $f_1, \dots, f_q : M \to M$ homeomorfismos de um espaço métrico compacto que comutam entre si. Por definição, o *conjunto não errante* $\Omega(f_1, \dots, f_q)$ é o conjunto dos pontos $x \in M$ tais que para toda vizinhança U de x existem $n_1, \dots, n_q \geq 1$ tais que $f_1^{n_1} \cdots f_q^{n_q}(U)$ intersecta U. Prove que $\Omega(f_1, \dots, f_q)$ é um compacto, não-vazio.

1.6.31. Demonstre o Lema 1.5.2.

1.6.32. Mostre que a conclusão do Teorema 1.5.1 ainda vale para partições de subconjuntos finitos de $\mathbb{Z}$, desde que sejam suficientemente grandes. Mais precisamente: dados $q, l \geq 1$ existe $N \geq 1$ tal que, dada qualquer partição do conjunto $\{1, 2, \ldots, N\}$ em l subconjuntos, algum desses subconjuntos contém progressões aritméticas com comprimento q.

1.6.33. Um ponto $x \in M$ é dito *super não errante* se, dada qualquer vizinhança U de x e dado qualquer $k \geq 1$, existe $n \geq 1$ tal que $\cap_{j=0}^{k} f^{-jn}(U) \neq \emptyset$. Mostre que o teorema de van der Waerden é equivalente ao seguinte enunciado: toda aplicação invertível num espaço métrico compacto tem algum ponto super não errante.

1.6.34. Prove a seguinte generalização do teorema de van der Waerden para dimensão arbitrária (teorema de Grünwald): dada qualquer partição finita $\mathbb{N}^k = S_1 \cup \cdots \cup S_l$ e qualquer $q \geq 1$, existem $j \in \{1, \ldots, l\}$, $d \in \mathbb{N}$ e $b \in \mathbb{N}^k$ tais que

$$b + d(a_1, \ldots, a_k) \in S_j \quad \text{para quaisquer } 1 \leq a_i \leq q \text{ e } 1 \leq i \leq k.$$

1.7 Problemas Abertos

1.7.1 (Conjectura 2x, 3x de Furstemberg)**.** Considere as transformações $f(x) = 2x \mod 1$ e $g(x) = 3x \mod 1$ definidas no círculo S^1. É verdade que a única medida invariante para f e g que não é atômica é a medida de Lebesgue?

1.7.2. Determine se o número π possui infinitos dígitos 4.

1.7.3 (Conjectura de Erdos)**.** Todo conjunto $S \subset \mathbb{N}$ tal que $\sum\limits_{n \in S} 1/n < \infty$ contém progressões aritméticas de comprimento arbitrário.

Capítulo 2

Teoremas Ergódicos

Neste capítulo apresentaremos alguns dos resultados fundamentais da Teoria Ergódica. Para motivar o tipo de enunciado, consideremos um conjunto mensurável $E \subset M$ com medida positiva e um ponto $x \in M$ qualquer. Queremos analisar o conjunto dos iterados de x que visitam E, isto é,

$$\{j \geq 0 : f^j(x) \in E\}.$$

Por exemplo, o teorema de recorrência de Poincaré afirma que, para quase todo $x \in E$, este conjunto é infinito. Gostaríamos de ter informação mais precisa, de natureza quantitativa. Chamamos *tempo médio de visita* de x a E o valor de

$$\tau(E, x) = \lim_{n \to \infty} \frac{1}{n} \#\{0 \leq j < n : f^j(x) \in E\}. \tag{2.0.1}$$

No caso de fluxos temos uma noção análoga, definida por

$$\tau(E, x) = \lim_{T \to \infty} \frac{1}{T} m\big(\{0 \leq t \leq T : f^t(x) \in E\}\big) \tag{2.0.2}$$

(m é a medida de Lebesgue na reta). Seria interessante saber, por exemplo, em que condições este tempo médio de visita é positivo. Antes de abordar este problema, é necessário responder a uma questão ainda mais básica: os limites em (2.0.1)-(2.0.2) existem?

Estas perguntas remontam ao trabalho do físico austríaco Ludwig Boltzmann (1844-1906), fundador da teoria cinética dos gases. Boltzmann era partidário da teoria atômica, que na época ainda era muito controversa, segundo a qual a matéria gasosa está formada por um grande número de minúsculas partículas em movimento e que se chocam continuamente. Em princípio, seria possível descrever o comportamento de um gás aplicando as leis da Mecânica Newtoniana a cada uma das suas partículas (moléculas). Na prática isso não é realista, porque o número de moléculas é enorme.

O problema da teoria cinética dos gases era, então, explicar o comportamento dos gases no nível macroscópico, como resultado estatístico da combinação de todos esses movimentos das suas moléculas. Para formular matematicamente a sua teoria, Boltzmann precisou de fazer uma suposição, que ficou conhecida como *hipótese ergódica*. Em linguagem moderna, a hipótese ergódica afirma que, para os sistemas (fluxos hamiltonianos) que descrevem o movimento das partículas de um gás, *o tempo médio de visita a qualquer subconjunto mensurável E existe e é igual à medida de E, para quase todo ponto x.*

O esforço para validar (ou invalidar) esta hipótese conduziu a importantes avanços tanto em Matemática (Teoria Ergódica, Sistemas Dinâmicos) quanto em Física Teórica (Mecânica Estatística). O que nos interessa neste capítulo são os resultados matemáticos relativos à *existência* do tempo médio de visita. A questão de saber quando $\tau(E,x) = \mu(E)$ para quase todo x será tratada no Capítulo 3.

Representando por φ a função característica do conjunto E, podemos reescrever a expressão no lado direito de (2.0.1) como:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)). \tag{2.0.3}$$

Isto sugere uma generalização natural da nossa pergunta inicial: o limite acima existe para funções φ muito gerais, por exemplo, para todas as funções integráveis?

O teorema ergódico de von Neumann (Teorema 2.1.6) afirma que, de fato, o limite em (2.0.3) existe no espaço $L^2(\mu)$, para toda função $\varphi \in L^2(\mu)$. O teorema ergódico de Birkhoff (Teorema 2.2.3) vai mais longe e afirma que há convergência em μ-quase todo ponto, para toda função $\varphi \in L^1(\mu)$. Em particular, o limite em (2.0.1) está bem definido para μ-quase todo x (Teorema 2.2.1).

Daremos uma demonstração direta do teorema de von Neumann e também mostraremos como ele pode ser deduzido do teorema ergódico de Birkhoff. Quanto a este último, iremos obtê-lo como caso particular de um resultado ainda mais forte, o teorema ergódico subaditivo de Kingman (Teorema 2.3.3). Este teorema afirma que ψ_n/n converge em quase todo ponto, para qualquer sequência de funções ψ_n tal que $\psi_{m+n} \le \psi_m + \psi_n \circ f^m$.

2.1 Teorema ergódico de von Neumann

Nesta seção enunciamos e provamos o teorema ergódico de von Neumann.

2.1.1 Isometrias em espaços de Hilbert

Seja H um espaço de Hilbert e seja F um subespaço fechado de H. Então

$$H = F \oplus F^\perp, \tag{2.1.1}$$

onde $F^\perp = \{w \in H : v \cdot w = 0 \text{ para todo } v \in F\}$ é o complementar ortogonal de F. A projeção $P_F : H \to F$ associada à decomposição (2.1.1) é chamada *projeção ortogonal* sobre F. Ela está unicamente caracterizada por

$$\|x - P_F(x)\| = \min\{\|x - v\| : v \in F\}.$$

Observe que $P_F(v) = v$ para todo $v \in F$ e, por consequência, $P_F^2 = P_F$.

Exemplo 2.1.1. Considere o espaço de Hilbert $L^2(\mu)$, com o produto interno

$$\varphi \cdot \psi = \int \varphi \bar{\psi} \, d\mu.$$

Seja F o espaço das funções constantes. Dada qualquer $\varphi \in L^2(\mu)$, temos que $(P_F(\varphi) - \varphi) \cdot 1 = 0$, ou seja,

$$P_F(\varphi) \cdot 1 = \varphi \cdot 1.$$

Como $P_F(\varphi)$ é uma função constante, o lado esquerdo desta igualdade é igual a $P_F(\varphi)$. O lado direito é igual a $\int \varphi \, d\mu$. Portanto, a projeção ortogonal no subespaço F é dada por

$$P_F(\varphi) = \int \varphi \, d\mu.$$

O operador adjunto $U^* : H \to H$ de um operador linear contínuo $U : H \to H$ está definido pela relação

$$U^* u \cdot v = u \cdot Uv \quad \text{para todo } u, v \in H. \tag{2.1.2}$$

O operador U diz-se uma *isometria* se ele preserva o produto interno:

$$Uu \cdot Uv = u \cdot v \quad \text{para todo } u, v \in H. \tag{2.1.3}$$

Isso é equivalente a dizer que U preserva a norma de H. Outra condição equivalente é $U^*U = \mathrm{id}$. De fato,

$$Uu \cdot Uv = u \cdot v \text{ para todo } u, v \quad \Leftrightarrow \quad U^*Uu \cdot v = u \cdot v \text{ para todo } u, v.$$

A propriedade $U^*U = \mathrm{id}$ implica que U é injetivo; em geral, uma isometria pode não ser sobrejetiva.

Exemplo 2.1.2. Se $f : M \to M$ preserva uma medida μ então, o seu operador de Koopman $U_f : L^2(\mu) \to L^2(\mu)$ definido por $U_f g = g \circ f$ é uma isometria.

Chamamos conjunto dos *vetores invariantes* de um operador linear contínuo $U : H \to H$ ao subespaço

$$I(U) = \{v \in H : Uv = v\}.$$

Observe que $I(U)$ é um subespaço vetorial fechado, uma vez que U é contínuo. Quando U é uma isometria, temos que $I(U) = I(U^*)$:

Lema 2.1.3. *Se $U : H \to H$ é uma isometria então $Uv = v$ se, e somente se, $U^*v = v$.*

Demonstração. Como $U^*U = \mathrm{id}$, é claro que $Uv = v$ implica $U^*v = v$. Agora suponha que $U^*v = v$. Então $Uv \cdot v = v \cdot U^*v = v \cdot v = \|v\|^2$. Logo, usando que U preserva a norma de H,

$$\|Uv - v\|^2 = (Uv - v) \cdot (Uv - v) = \|Uv\|^2 - Uv \cdot v - v \cdot Uv + \|v\|^2 = 0.$$

Isto significa que $Uv = v$. $\square$

Para encerrar esta breve discussão, citamos um resultado clássico de Análise Funcional, devido a Marshall H. Stone, que permite reduzir o estudo dos operadores de Koopman de sistemas a tempo contínuo ao caso discreto.

Seja $U_t : H \to H$, $t \in \mathbb{R}$ um *grupo a* 1*-parâmetro* de operadores lineares num espaço de Banach: isto significa que $U_0 = \mathrm{id}$ e $U_{t+s} = U_t U_s$ para todo $t, s \in \mathbb{R}$. Dizemos que o grupo é *fortemente contínuo* se

$$\lim_{t \to t_0} U_t v = U_{t_0} v, \quad \text{para todo } t_0 \in \mathbb{R} \text{ e } v \in H.$$

Teorema 2.1.4 (Stone)**.** *Se $U_t : H \to H$, $t \in \mathbb{R}$ é um grupo a 1-parâmetro fortemente contínuo de operadores unitários num espaço de Hilbert complexo, existe um operador auto-adjunto A, definido num subespaço denso $D(A)$ de H, tal que $U_t \mid D(A) = e^{itA}$ para todo $t \in \mathbb{R}$.*

O leitor pode conferir a demonstração no livro de Yosida [Yos68, § IX.9] e damos uma aplicação simples no Exercício 2.3.5. O operador iA é chamado *gerador infinitesimal* do grupo. Ele pode ser calculado por meio da igualdade

$$iAv = \lim_{t \to 0} \frac{1}{t}\big(U_t v - v\big). \tag{2.1.4}$$

Veja em Yosida [Yos68, § IX.3] uma prova de que o limite no lado direito existe para todo v num subespaço denso de H.

Exemplo 2.1.5. Seja H o espaço de Banach das funções contínuas $\varphi : S^1 \to \mathbb{C}$, munido da norma da convergência uniforme. Defina $U_t(\varphi)(x) = \varphi(x + t)$ para cada função $\varphi \in H$. Observe que $(U_t)_t$ é um grupo a 1-parâmetro fortemente contínuo de isometrias de H. O gerador infinitesimal está dado por

$$iA\phi(x) = \lim_{t \to 0} \frac{1}{t}\big(U_t\phi(x) - \phi(x)\big) = \lim_{t \to 0} \frac{1}{t}\big(\phi(x + t) - \phi(x)\big) = \phi'(x).$$

Note que o seu domínio é o conjunto das funções de classe C^1, o qual é denso em H, como sabemos.

2.1.2 Enunciado e prova do teorema

O nosso primeiro teorema ergódico é:

Teorema 2.1.6 (von Neumann)**.** *Seja $U : H \to H$ uma isometria num espaço de Hilbert H, e seja P a projeção ortogonal sobre o subespaço $I(U)$ dos vetores invariantes por U. Então,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} U^j v = Pv \quad \textit{para todo } v \in H. \tag{2.1.5}$$

Demonstração. Seja $L(U)$ o conjunto dos vetores $v \in H$ da forma $v = Uu - u$ para algum $u \in H$ e seja $\bar{L}(U)$ o seu fecho. Afirmamos que

$$I(U) = \bar{L}(U)^{\perp}. \tag{2.1.6}$$

Isto pode ser verificado da seguinte forma. Considere quaisquer $v \in I(U)$ e $w \in \bar{L}(U)$. Pelo Lema 2.1.3, temos que $v \in I(U^*)$, ou seja $U^*v = v$. Além disso, por definição de $\bar{L}(U)$, existem $u_n \in H$, $n \geq 1$ tais que $(Uu_n - u_n)_n \to w$. Então

$$v \cdot (Uu_n - u_n) = v \cdot Uu_n - v \cdot u_n = U^*v \cdot u_n - v \cdot u_n = 0$$

para todo n e, como consequência, $v \cdot w = 0$. Isto prova que $I(U) \subset \bar{L}(U)^{\perp}$. Em seguida, considere qualquer $v \in \bar{L}(U)^{\perp}$. Então, em particular,

$$v \cdot (Uu - u) = 0 \quad \text{ou seja,} \quad U^*v \cdot u - v \cdot u = 0$$

para todo $u \in H$. Isto quer dizer que $U^*v = v$. Usando Lema 2.1.3 uma vez mais, deduzimos que $v \in I(U)$. Isto mostra que $\bar{L}(U)^{\perp} \subset I(U)$ e, portanto, a prova de (2.1.6) está completa. Como consequência, usando (2.1.1),

$$H = I(U) \oplus \bar{L}(U). \tag{2.1.7}$$

Agora vamos verificar a igualdade (2.1.5), sucessivamente, quando $v \in I(U)$, quando $v \in \bar{L}(U)$, e no caso geral. Suponha primeiro que $v \in I(U)$. Por um lado, $Pv = v$. Por outro lado,

$$\frac{1}{n} \sum_{j=0}^{n-1} U^j v = \frac{1}{n} \sum_{j=0}^{n-1} v = v$$

para todo n. Logo esta sequência converge para v quando $n \to \infty$. Isto prova (2.1.5) neste caso.

Em seguida suponha que $v \in L(U)$. Então, por definição, existe $u \in H$ tal que $v = Uu - u$. É imediato que

$$\frac{1}{n} \sum_{j=0}^{n-1} U^j v = \frac{1}{n} \sum_{j=0}^{n-1} \big(U^{j+1}u - U^j u\big) = \frac{1}{n}(U^n u - u).$$

A norma desta última expressão está majorada por $2\|u\|/n$ e, portanto, converge para zero quando $n \to \infty$. Isto mostra que

$$\lim_{n} \frac{1}{n} \sum_{j=0}^{n-1} U^j v = 0 \quad \text{para todo } v \in L(U). \tag{2.1.8}$$

Mais em geral, suponha que $v \in \bar{L}(U)$. Então, existem $v_k \in L(U)$ convergindo para v quando $k \to \infty$. Observe que

$$\Big\|\frac{1}{n}\sum_{j=0}^{n-1} U^j v - \frac{1}{n}\sum_{j=0}^{n-1} U^j v_k\Big\| \le \frac{1}{n}\sum_{j=0}^{n-1} \|U^j(v - v_k)\| \le \|v - v_k\|$$

para todo n e todo k. Juntamente com (2.1.8), isto implica que

$$\lim_n \frac{1}{n}\sum_{j=0}^{n-1} U^j v = 0 \quad \text{para todo } v \in \bar{L}(U). \tag{2.1.9}$$

Como a relação (2.1.6) implica que $Pv = 0$ para todo $v \in \bar{L}(U)$, isto mostra que (2.1.5) vale também quando $v \in \bar{L}(U)$.

O caso geral de (2.1.5) segue imediatamente, já que $H = I(U) \oplus \bar{L}(U)$. □

2.1.3 Convergência em $L^2(\mu)$

Dada uma transformação mensurável $f : M \to M$ que preserva uma probabilidade μ em M, dizemos que uma função mensurável $\psi : M \to \mathbb{R}$ é *invariante* se $\psi \circ f = \psi$ em μ-quase todo ponto. O seguinte resultado é um caso particular do Teorema 2.1.6:

Teorema 2.1.7. *Para qualquer $\varphi \in L^2(\mu)$, seja $\tilde{\varphi}$ a projeção ortogonal de φ no subespaço das funções invariantes. Então a sequência*

$$\frac{1}{n}\sum_{j=0}^{n-1} \varphi \circ f^j \tag{2.1.10}$$

converge para $\tilde{\varphi}$ no espaço $L^2(\mu)$. Se f é invertível, então a sequência

$$\frac{1}{n}\sum_{j=0}^{n-1} \varphi \circ f^{-j} \tag{2.1.11}$$

também converge para $\tilde{\varphi}$ em $L^2(\mu)$.

Demonstração. Seja $U = U_f : L^2(\mu) \to L^2(\mu)$ o operador de Koopman de uma transformação $f : M \to M$ que preserva uma medida finita μ. Note que uma função ψ está em $I(U)$ se, e somente se, $\psi \circ f = \psi$ em μ-quase todo ponto. Seja $\tilde{\varphi}$ a projeção ortogonal de φ em $I(U)$. Pelo Teorema 2.1.6, a sequência em (2.1.10) converge para $\tilde{\varphi}$ em $L^2(\mu)$. Isto prova a primeira afirmação.

A segunda afirmação é análoga, considerando $U = U_{f^{-1}}$, ou seja $U = U_f^{-1}$. Obtemos que a sequência em (2.1.11) converge para a projeção ortogonal de φ no espaço $I(U_f^{-1})$. Observando que $I(U_f^{-1}) = I(U_f)$, concluímos que o limite desta sequência é a mesma função $\tilde{\varphi}$ que obtivemos antes. □

2.2 Teorema ergódico de Birkhoff

O teorema que apresentamos nesta seção foi demonstrado por George David Birkhoff. É devida a ele a noção de distância projetiva, um dos maiores matemáticos americanos da sua geração e autor de muitas outras contribuições fundamentais em Dinâmica. O teorema de Birkhoff melhora bastante o teorema de von Neumann porque a sua conclusão é formulada em termos de convergência em μ-quase todo o ponto, que neste contexto é uma propriedade mais forte do que convergência em $L^2(\mu)$, conforme explicamos na Seção 2.2.3.

2.2.1 Tempo médio de visita

Começamos por enunciar a versão do teorema para tempos médios de visita:

Teorema 2.2.1 (Birkhoff)**.** *Seja $f: M \to M$ uma transformação mensurável e μ uma probabilidade invariante por f. Dado qualquer conjunto mensurável $E \subset M$, o tempo médio de visita*

$$\tau(E,x) = \lim_n \frac{1}{n} \#\{j = 0, 1, \ldots, n-1 : f^j(x) \in E\}$$

existe em μ-quase todo ponto $x \in M$. Além disso, $\int \tau(E,x)\, d\mu(x) = \mu(E)$.

Observe que se $\tau(E,x)$ existe para um certo ponto $x \in M$ então

$$\tau(E, f(x)) = \tau(E,x). \tag{2.2.1}$$

De fato, por definição,

$$\begin{aligned}
\tau(E, f(x)) &= \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \mathcal{X}_E(f^j(x)) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mathcal{X}_E(f^j(x)) - \frac{1}{n}\big[\mathcal{X}_E(x) - \mathcal{X}_E(f^n(x))\big] \\
&= \tau(E,x) - \lim_{n\to\infty} \frac{1}{n}\big[\mathcal{X}_E(x) - \mathcal{X}_E(f^n(x))\big].
\end{aligned}$$

Como a função característica é limitada, o último limite é igual a zero. Isto prova a igualdade (2.2.1).

O exemplo a seguir mostra que o tempo médio de visita não existe para *todo* ponto, em geral:

Exemplo 2.2.2. Considere o número $x \in (0,1)$ definido pela expansão decimal $x = 0, a_1 a_2 a_3 \ldots$, onde $a_i = 1$ se $2^k \le i < 2^{k+1}$ com k par e $a_i = 0$ se $2^k \le i < 2^{k+1}$ com k ímpar. Ou seja,

$$x = 0,100111100000000111111111111111110\ldots,$$

onde os blocos alternantes de 0s e de 1s têm comprimentos dados pelas sucessivas potências de dois. Seja $f : [0,1] \to [0,1]$ a transformação definida na Seção 1.3.1

e seja $E = [0, 1/10)$. Isto é, E é o conjunto dos pontos cuja expansão decimal começa com o dígito 0. É fácil ver que se $n = 2^k - 1$ com k par então

$$\frac{1}{n}\sum_{j=0}^{n-1} \mathcal{X}_E(f^j(x)) = \frac{2^1 + 2^3 + \cdots + 2^{k-1}}{2^k - 1} = \frac{2}{3}.$$

Por outro lado, se $n = 2^k - 1$ e k ímpar então

$$\frac{1}{n}\sum_{j=0}^{n-1} \mathcal{X}_E(f^j(x)) = \frac{2^1 + 2^3 + \cdots + 2^{k-2}}{2^k - 1} = \frac{2^k - 2}{3(2^k - 1)} \to \frac{1}{3}$$

quando $k \to \infty$. Assim, o tempo médio de visita de x ao conjunto E não existe.

2.2.2 Médias temporais

Conforme observamos anteriormente

$$\tau(E, x) = \lim_n \frac{1}{n}\sum_{j=0}^{n-1} \varphi(f^j(x)), \quad \text{onde } \varphi = \mathcal{X}_E.$$

O próximo enunciado generaliza o Teorema 2.2.1 para o caso em que φ é uma função integrável qualquer:

Teorema 2.2.3 (Birkhoff)**.** *Seja $f : M \to M$ uma transformação mensurável e seja μ uma probabilidade invariante por f. Dada qualquer função integrável $\varphi : M \to \mathbb{R}$, o limite*

$$\tilde{\varphi}(x) = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \varphi(f^j(x)) \tag{2.2.2}$$

existe em μ-quase todo ponto $x \in M$. Além disso, a função $\tilde{\varphi}$ definida desta forma é integrável e satisfaz

$$\int \tilde{\varphi}(x)\, d\mu(x) = \int \varphi(x)\, d\mu(x).$$

Um pouco mais adiante obteremos este teorema como caso particular de um resultado mais geral, o teorema ergódico subaditivo. O limite $\tilde{\varphi}$ é chamado *média temporal*, ou média orbital, de φ. A proposição a seguir mostra que as médias temporais são constantes ao longo de órbitas, em μ-quase todo ponto, generalizando a igualdade (2.2.1):

Proposição 2.2.4. *Seja $\varphi : M \to \mathbb{R}$ uma função integrável. Então,*

$$\tilde{\varphi}(f(x)) = \tilde{\varphi}(x) \quad \textit{para } \mu\textit{-quase todo ponto } x \in M. \tag{2.2.3}$$

Demonstração. Por definição,

$$\tilde{\varphi}(f(x)) = \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \varphi(f^j(x)) = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \varphi(f^j(x)) + \frac{1}{n}\big[\varphi(f^n(x)) - \varphi(x)\big]$$
$$= \tilde{\varphi}(x) + \lim_{n\to\infty} \frac{1}{n}\big[\varphi(f^n(x)) - \varphi(x)\big].$$

Vamos precisar do seguinte lema:

Lema 2.2.5. *Se ϕ é uma função integrável então $\lim_n (1/n)\phi(f^n(x)) = 0$ para μ-quase todo ponto $x \in M$.*

Demonstração. Fixe qualquer $\varepsilon > 0$. Como μ é invariante, temos que

$$\mu\big(\{x \in M : |\phi(f^n(x))| \geq n\varepsilon\}\big) = \mu\big(\{x \in M : |\phi(x)| \geq n\varepsilon\}\big)$$
$$= \sum_{k=n}^{\infty} \mu\big(\{x \in M : k \leq \frac{|\phi(x)|}{\varepsilon} < k+1\}\big).$$

Somando sobre todo $n \in \mathbb{N}$, obtemos que

$$\sum_{n=1}^{\infty} \mu\big(\{x \in M : |\phi(f^n(x))| \geq n\varepsilon\}\big) = \sum_{k=1}^{\infty} k\mu\big(\{x \in M : k \leq \frac{|\phi(x)|}{\varepsilon} < k+1\}\big)$$
$$\leq \int \frac{|\phi|}{\varepsilon}\, d\mu.$$

Como ϕ é integrável, por hipótese, todas estas expressões são finitas. Isso implica que o conjunto $B(\varepsilon)$ dos pontos x tais que $|\phi(f^n(x))| \geq n\varepsilon$ para infinitos valores de n tem medida nula (veja o Exercício A.1.6). Segue imediatamente da definição de $B(\varepsilon)$ que para todo $x \notin B(\varepsilon)$ existe algum $p \geq 1$ tal que $|\phi(f^n(x))| < n\varepsilon$ para todo $n \geq p$. Agora considere o conjunto $B = \cup_{i=1}^{\infty} B(1/i)$. Então B tem medida nula e para todo $x \notin B$ vale que $\lim_n (1/n)\phi(f^n(x)) = 0$. □

Aplicando o Lemma 2.2.5 à função $\phi = \varphi$ obtemos a igualdade (2.2.3). Isto completa a demonstração da Proposição 2.2.4. □

Em geral, o subconjunto com medida total onde vale a convergência (2.2.2) no Teorema 2.2.3 depende da função φ que estamos considerando. No entanto, em alguns casos é possível escolher esse conjunto independentemente da função. Um exemplo útil desta situação é o seguinte:

Teorema 2.2.6. *Suponha que M é um espaço métrico compacto e $f : M \to M$ é uma aplicação mensurável. Então existe um conjunto mensurável $G \subset M$ com $\mu(G) = 1$ tal que*

$$\frac{1}{n}\sum_{j=0}^{n-1} \varphi(f^j(x)) \to \tilde{\varphi}(x) \tag{2.2.4}$$

para todo $x \in G$ e toda função contínua $\varphi : M \to \mathbb{R}$.

Demonstração. Pelo teorema ergódico de Birkhoff, para cada função contínua φ existe um conjunto $G(\varphi) \subset M$ com $\mu(G(\varphi)) = 1$ tal que (2.2.4) é válido para todo $x \in G(\varphi)$. Sabemos que o espaço $C^0(M)$ das funções contínuas admite algum subconjunto $\{\varphi_k : k \in \mathbb{N}\}$ enumerável denso. Tomemos

$$G = \bigcap_{k=1}^{\infty} G(\varphi_k).$$

É claro que $\mu(G) = 1$. Portanto basta provar que (2.2.4) vale para toda função contínua φ sempre que $x \in G$. Isso pode ser feito da seguinte maneira. Dado $\varphi \in C^0(M)$ e qualquer $\varepsilon > 0$, tomemos $k \in \mathbb{N}$ tal que

$$\|\varphi - \varphi_k\| = \sup\big\{|\varphi(x) - \varphi_k(x)| : x \in M\big\} \le \varepsilon.$$

Então, dado qualquer ponto $x \in G$,

$$\limsup_n \frac{1}{n}\sum_{j=0}^{n-1} \varphi(f^j(x)) \le \lim_n \frac{1}{n}\sum_{j=0}^{n-1} \varphi_k(f^j(x)) + \varepsilon = \tilde{\varphi}_k(x) + \varepsilon$$
$$\liminf_n \frac{1}{n}\sum_{j=0}^{n-1} \varphi(f^j(x)) \ge \lim_n \frac{1}{n}\sum_{j=0}^{n-1} \varphi_k(f^j(x)) - \varepsilon = \tilde{\varphi}_k(x) - \varepsilon.$$

Isto implica que

$$\limsup_n \frac{1}{n}\sum_{j=0}^{n-1} \varphi(f^j(x)) - \liminf_n \frac{1}{n}\sum_{j=0}^{n-1} \varphi(f^j(x)) \le 2\varepsilon.$$

Como ε é arbitrário, segue que o limite $\tilde{\varphi}(x)$ existe, conforme afirmado. □

Em geral não é possível dizer nada sobre a velocidade da convergência no Teorema 2.2.3. Por exemplo, segue de um teorema de Kakutani e Petersen (confira as páginas 94 a 99 do livro de Petersen [Pet83]) que se a medida μ é ergódica [1] e não atômica então, dada qualquer sequência $(a_n)_n$ de números positivos com $\lim_n a_n = 0$, existe alguma função mensurável limitada φ tal que

$$\limsup_n \frac{1}{a_n}\Big|\frac{1}{n}\sum_{j=0}^{n-1} \varphi(f^j(x)) - \int \varphi\, d\mu\Big| = +\infty.$$

Outra observação interessante é que não existe um análogo do teorema ergódico de Birkhoff para medidas invariantes *infinitas*. De fato, suponha que μ é medida invariante σ-finita, mas infinita, de uma transformação $f : M \to M$. Dizemos que um conjunto mensurável $W \subset M$ é *errante* se as pré-imagens $f^{-i}(W)$, $i \ge 0$ são disjuntas duas-a-duas. Suponha que μ é ergódica e *conservativa*, ou seja, todo conjunto errante tem medida nula. Então, dada qualquer sequência $(a_n)_n$ de números positivos,

[1] Dizemos que uma medida invariante μ é *ergódica* se $f^{-1}(A) = A$ a menos de medida nula implica que $\mu(A) = 0$ ou $\mu(A^c) = 0$. O estudo das medidas ergódicas será o tema do próximo capítulo.

- ou, para toda $\varphi \in L^1(\mu)$,

$$\liminf_n \frac{1}{a_n} \sum_{j=0}^{n-1} \varphi \circ f^j = 0 \quad \text{em quase todo ponto;}$$

- ou existe $(n_k)_k \to \infty$ tal que, para toda $\varphi \in L^1(\mu)$,

$$\lim_k \frac{1}{a_{n_k}} \sum_{j=0}^{n_k-1} \varphi \circ f^j = \infty \quad \text{em quase todo ponto.}$$

Este resultado e outros fatos correlatos estão demonstrados na Seção 2.4 do livro de Aaronson [Aar97].

2.2.3 Teorema de von Neumann e consequências

O teorema de von Neumann (Teorema 2.1.7) também pode ser deduzido diretamente do teorema de Birkhoff, como vamos mostrar a seguir.

Considere qualquer função $\varphi \in L^2(\mu)$ e seja $\tilde{\varphi}$ a sua média temporal. Começamos por mostrar que $\tilde{\varphi} \in L^2(\mu)$ e a sua norma satisfaz $\|\tilde{\varphi}\|_2 \le \|\varphi\|_2$. Para isso, note que

$$|\tilde{\varphi}| \le \lim_n \frac{1}{n} \sum_{j=0}^{n-1} |\varphi \circ f^j| \quad \text{e, portanto,} \quad |\tilde{\varphi}|^2 \le \lim_n \Big(\frac{1}{n} \sum_{j=0}^{n-1} |\varphi \circ f^j| \Big)^2.$$

Então, pelo lema de Fatou (Teorema A.2.10),

$$\Big[\int |\tilde{\varphi}|^2 \, d\mu \Big]^{1/2} \le \liminf_n \Big[\int \Big(\frac{1}{n} \sum_{j=0}^{n-1} |\varphi \circ f^j| \Big)^2 d\mu \Big]^{1/2}. \tag{2.2.5}$$

Podemos usar a desigualdade de Minkowski para majorar a sequência do lado direito:

$$\Big[\int \Big(\frac{1}{n} \sum_{j=0}^{n-1} |\varphi \circ f^j| \Big)^2 d\mu \Big]^{1/2} \le \frac{1}{n} \sum_{j=0}^{n-1} \Big[\int |\varphi \circ f^j|^2 \, d\mu \Big]^{1/2}. \tag{2.2.6}$$

Como μ é invariante por f, a expressão do lado direito é igual a $\big[\int |\varphi|^2 \, d\mu \big]^{1/2}$. Portanto, (2.2.5) e (2.2.6) implicam que $\|\tilde{\varphi}\|_2 \le \|\varphi\|_2 < \infty$.

Agora vamos mostrar que $(1/n) \sum_{j=0}^{n-1} \varphi \circ f^j$ converge para $\tilde{\varphi}$ em $L^2(\mu)$. Inicialmente, suponha que a função φ é limitada, isto é, que existe $C > 0$ tal que $|\varphi| \le C$. Então

$$\Big| \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f^j \Big| \le C \quad \text{para todo } n \quad \text{e} \quad |\tilde{\varphi}| \le C.$$

Então podemos usar o teorema da convergência dominada (Teorema A.2.11) para concluir que

$$\lim_n \int \Big(\frac{1}{n}\sum_{j=0}^{n-1} \varphi \circ f^j - \tilde{\varphi}\Big)^2 d\mu = \int \Big(\lim_n \frac{1}{n}\sum_{j=0}^{n-1} \varphi \circ f^j - \tilde{\varphi}\Big)^2 d\mu = 0,$$

ou seja, que $(1/n)\sum_{j=0}^{n-1} \varphi \circ f^j$ converge para $\tilde{\varphi}$ em $L^2(\mu)$. Falta estender esta conclusão para uma função φ qualquer em $L^2(\mu)$. Para isso, consideremos uma sequência (φ_k) de funções limitadas tal que $(\varphi_k)_k$ converge para φ. Por exemplo

$$\varphi_k(x) = \begin{cases} \varphi(x) & \text{se } |\varphi(x)| \le k \\ 0 & \text{caso contrário.} \end{cases}$$

Denotemos por $\tilde{\varphi}_k$ as respectivas médias temporais. Dado qualquer $\varepsilon > 0$, fixemos k_0 tal que $\|\varphi - \varphi_k\|_2 < \varepsilon/3$ para todo $k \ge k_0$. Note que $\|(\varphi - \varphi_k) \circ f^j\|_2$ é igual a $\|\varphi - \varphi_k\|_2$ para todo $j \ge 0$, porque a medida μ é invariante. Logo,

$$\Big\|\frac{1}{n}\sum_{j=0}^{n-1}(\varphi - \varphi_k) \circ f^j\Big\|_2 \le \|\varphi - \varphi_k\|_2 < \varepsilon/3 \quad \text{para todo } n \ge 1 \text{ e } k \ge k_0. \tag{2.2.7}$$

Observe também que $\tilde{\varphi} - \tilde{\varphi}_k$ é a média temporal da função $\varphi - \varphi_k$. Portanto, o argumento do parágrafo anterior dá que

$$\|\tilde{\varphi} - \tilde{\varphi}_k\|_2 \le \|\varphi - \varphi_k\|_2 < \varepsilon/3 \quad \text{para todo } k \ge k_0. \tag{2.2.8}$$

Por hipótese, para cada $k \ge 1$ existe $n_0(k) \ge 1$ tal que

$$\Big\|\frac{1}{n}\sum_{j=0}^{n-1} \varphi_k \circ f^j - \tilde{\varphi}_k\Big\|_2 < \varepsilon/3 \quad \text{para todo } n \ge n_0(k). \tag{2.2.9}$$

Somando (2.2.7), (2.2.8), (2.2.9) obtemos

$$\Big\|\frac{1}{n}\sum_{j=0}^{n-1} \varphi \circ f^j - \tilde{\varphi}\Big\|_2 < \varepsilon \quad \text{para todo } n \ge n_0(k_0).$$

Isto completa a prova do teorema de von Neumann a partir do teorema de Birkhoff.

No Exercício 2.3.10 propomos uma generalização destas conclusões para um espaço $L^p(\mu)$ qualquer.

Corolário 2.2.7. *A média temporal $\tilde{\varphi}$ de qualquer função $\varphi \in L^2(\mu)$ coincide com a projeção ortogonal $P(\varphi)$ de φ no subespaço das funções invariantes.*

Demonstração. Por um lado, o Teorema 2.1.7 dá que $(1/n)\sum_{j=0}^{n-1} \varphi \circ f^j$ converge para $P(\varphi)$ em $L^2(\mu)$. Por outro lado, acabamos de mostrar que essa sequência converge para $\tilde{\varphi}$ em $L^2(\mu)$. Por unicidade do limite, $P(\varphi) = \tilde{\varphi}$. □

Corolário 2.2.8. *Se $f : M \to M$ é invertível então as médias temporais de qualquer função $\varphi \in L^2(\mu)$ para f e para f^{-1} coincidem em μ-quase todo ponto:*

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f^{-j} = \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f^{j} \quad \textit{em } \mu\textit{-quase todo ponto.} \tag{2.2.10}$$

Demonstração. O limite do lado esquerdo de (2.2.10) é a projeção ortogonal de φ no subespaço das funções invariantes por f^{-1}, enquanto que o limite do lado direito é a projeção ortogonal de φ no subespaço das funções invariantes por f. É claro que estes dois subespaços são exatamente o mesmo. Logo os dois limites coincidem em $L^2(\mu)$. □

2.3 Teorema ergódico subaditivo

Dizemos que uma sequência de funções $\varphi_n : M \to \mathbb{R}$ é *subaditiva* para uma transformação $f : M \to M$ se

$$\varphi_{m+n} \le \varphi_m + \varphi_n \circ f^m \quad \text{para todo } m, n \ge 1. \tag{2.3.1}$$

Exemplo 2.3.1. A sequência $\varphi_n : M \to \mathbb{R}$ diz-se *aditiva* se vale a igualdade em (2.3.1), ou seja, se $\varphi_{m+n} = \varphi_m + \varphi_n \circ f^m$ para todo $m, n \ge 1$. Por exemplo, toda soma temporal

$$\varphi_n(x) = \sum_{j=0}^{n-1} \varphi(f^j(x))$$

constitui uma sequência aditiva. É fácil verificar que toda sequência aditiva é desta forma, com $\varphi = \varphi_1$.

No próximo exemplo usamos a noção de *norma* de uma matriz quadrada, que é definida do seguinte modo. Seja A uma matriz quadrada de dimensão $d \ge 2$. Então

$$\|A\| = \sup \Big\{ \frac{\|Av\|}{\|v\|} : v \in \mathbb{R}^d \setminus \{0\} \Big\}. \tag{2.3.2}$$

Segue diretamente da definição que a norma do produto de duas matrizes é menor ou igual que o produto das normas dessas matrizes:

$$\|AB\| \le \|A\| \, \|B\| \,. \tag{2.3.3}$$

Exemplo 2.3.2. Seja $A : M \to \mathrm{GL}(d, \mathbb{R})$ uma função mensurável com valores no *grupo linear*, ou seja, o conjunto $\mathrm{GL}(d, \mathbb{R})$ das matrizes quadradas invertíveis de dimensão d. Defina

$$\phi^n(x) = A(f^{n-1}(x)) \cdots A(f(x))A(x)$$

para todo $n \ge 1$ e $x \in M$. Então a sequência $\varphi_n(x) = \log \|\phi^n(x)\|$ é subaditiva. De fato,

$$\phi^{m+n}(x) = \phi^n(f^m(x))\phi^m(x)$$

e, portanto, usando (2.3.3),

$$\begin{aligned}\varphi_{m+n}(x) &= \log\|\phi^n(f^m(x))\phi^m(x)\| \\ &\leq \log\|\phi^m(x)\| + \log\|\phi^n(f^m(x))\| = \varphi_m(x) + \varphi_n(f^m(x))\end{aligned}$$

para todo m, n e x.

Lembre que, dada uma função $\varphi : M \to \mathbb{R}$ representamos por $\varphi^+ : M \to \mathbb{R}$ a função definida por $\varphi^+(x) = \max\{\varphi(x), 0\}$.

Teorema 2.3.3 (Kingman)**.** *Seja μ uma probabilidade invariante para uma transformação $f : M \to M$ e seja $\varphi_n : M \to \mathbb{R}$, $n \geq 1$ uma sequência subaditiva de funções mensuráveis tal que $\varphi_1^+ \in L^1(\mu)$. Então a sequência $(\varphi_n/n)_n$ converge em μ-quase todo ponto para uma função f-invariante $\varphi : M \to [-\infty, +\infty)$. Além disso, $\varphi^+ \in L^1(\mu)$ e*

$$\int \varphi\, d\mu = \lim_n \frac{1}{n}\int \varphi_n\, d\mu = \inf_n \frac{1}{n}\int \varphi_n\, d\mu \in [-\infty, +\infty).$$

A prova do Teorema 2.3.3 que vamos apresentar é devida a Avila, Bochi [AB], os quais se inspiraram na demonstração do Teorema 2.2.3 dada por Katznelson, Weiss [KW82]. Um ponto importante é que o teorema ergódico de Birkhoff *não* é usado no argumento. Isso nos permitirá obter o teorema de Birkhoff como corolário do Teorema 2.3.3. Para a prova, veja [eKO14]. Vamos agora dar algumas aplicações do Teorema 2.3.3.

2.3.1 Expoentes de Lyapunov

Como observamos anteriormente, toda sequência de somas orbitais

$$\varphi_n = \sum_{j=0}^{n-1} \varphi \circ f^j, \quad n \geq 1$$

é aditiva e, em particular, subaditiva. Portanto, o teorema ergódico de Birkhoff (Teorema 2.2.3) é um caso particular do Teorema 2.3.3.

Outra corolário importante do teorema ergódico subaditivo é o teorema de Furstenberg-Kesten, que enunciaremos a seguir.

Seja $f : M \to M$ uma transformação mensurável e seja μ uma probabilidade invariante. Seja $\theta : M \to \mathrm{GL}(d)$ uma função mensurável com valores no conjunto $\mathrm{GL}(d)$ das matrizes quadradas invertíveis de dimensão d. O *cociclo* definido por θ sobre a transformação f é a sequência de funções dada por

$$\phi^n(x) = \theta(f^{n-1}(x)) \cdots \theta(f(x))\theta(x) \text{ para } n \geq 1 \quad \text{e} \quad \phi^0(x) = \mathrm{id}$$

para todo $x \in M$. Deixamos ao cuidado do leitor verificar que

$$\phi^{m+n}(x) = \phi^n(f^m(x)) \cdot \phi^m(x) \quad \text{para todo } m, n \in \mathbb{Z} \text{ e todo } x \in M. \tag{2.3.4}$$

Também é fácil verificar que, reciprocamente, qualquer sequência $(\phi^n)_n$ que satisfaz (2.3.4) é o cociclo definido por $\theta = \phi^1$ sobre a transformação f.

Teorema 2.3.4 (Furstenberg-Kesten). *Se* $\log^+ \|\theta\| \in L^1(\mu)$ *então*

$$\lambda_{\max}(x) = \lim_n \frac{1}{n} \log \|\phi^n(x)\|$$

existe em μ-quase todo ponto. Além disso, $\lambda^+_{\max} \in L^1(\mu)$ *e*

$$\int \lambda_{\max}\, d\mu = \lim_n \frac{1}{n} \int \log \|\phi^n\|\, d\mu = \inf_n \frac{1}{n} \int \log \|\phi^n\|\, d\mu.$$

Se $\log^+ \|\theta^{-1}\| \in L^1(\mu)$ *então*

$$\lambda_{\min}(x) = \lim_n -\frac{1}{n} \log \|\phi^n(x)^{-1}\|$$

existe em μ-quase todo ponto. Além disso, $\lambda_{\min} \in L^1(\mu)$ *e*

$$\int \lambda_{\min}\, d\mu = \lim_n -\frac{1}{n} \int \log \|(\phi^n)^{-1}\|\, d\mu = \sup_n -\frac{1}{n} \int \log \|(\phi^n)^{-1}\|\, d\mu.$$

Para deduzir este resultado do Teorema 2.3.3 basta observar que as sequências

$$\varphi_n^{\max}(x) = \log \|\phi^n(x)\| \quad \text{e} \quad \varphi_n^{\min}(x) = \log \|\phi^n(x)^{-1}\|$$

são subaditivas (lembre do Exemplo 2.3.2).

O teorema ergódico multiplicativo de Oseledets, que vamos enunciar a seguir, refina muito a conclusão do teorema de Furstenberg-Kesten. Ele afirma que, nas mesmas condições do Teorema 2.3.4, para μ-quase todo $x \in M$ existe um número inteiro positivo $k = k(x)$ e existem números reais $\lambda_1(x) > \cdots > \lambda_k(x)$ e uma filtração

$$\mathbb{R}^d = V_x^1 > \cdots > V_x^k > V_x^{k+1} = \{0\} \tag{2.3.5}$$

tal que, para todo $i \in \{1, \ldots, k\}$ e para μ-quase todo $x \in M$,

(a1) $k(f(x)) = k(x)$ e $\lambda_i(f(x)) = \lambda_i(x)$ e $\theta(x) \cdot V_x^i = V_{f(x)}^i$;

(b1) $\lim_n \frac{1}{n} \log \|\phi^n(x) v\| = \lambda_i(x)$ para todo $v \in V_x^i \setminus V_x^{i+1}$;

(c1) $\lim_n \frac{1}{n} \log |\det \phi^n(x)| = \sum_{i=1}^k d_i(x) \lambda_i(x)$, onde $d_i(x) = \dim V_x^i - \dim V_x^{i+1}$.

Além disso, os números $k(x)$ e $\lambda_1(x), \ldots, \lambda_k(x)$ e os subespaços $V_x^1, \ldots, V_x^k$ dependem mensuravelmente do ponto x.

Os números $\lambda_i(x)$ são chamados *expoentes de Lyapunov* de θ relativamente a f no ponto x. Eles satisfazem $\lambda_1 = \lambda_{\max}$ e $\lambda_k = \lambda_{\min}$. Por esta razão, também chamamos $\lambda_{\max}(x)$ e $\lambda_{\min}(x)$ de *expoentes de Lyapunov extremais* no ponto x. Cada $d_i(x)$ é chamado *multiplicidade* do expoente de Lyapunov $\lambda_i(x)$.

Quando f é invertível, podemos estender a sequência $(\phi^n)_n$ a todo o $\mathbb{Z}$, definindo

$$\phi^{-n}(x) = \phi^n(f^{-n}(x))^{-1} \quad \text{para todo } n \geq 1 \text{ e } x \in M.$$

Supondo também que $\log^+ \|\theta^{-1}\| \in L^1(\mu)$, é possível obter uma conclusão mais forte do que anteriormente: no lugar da filtração (2.3.5) obtemos uma decomposição

$$\mathbb{R}^d = E_x^1 \oplus \cdots \oplus E_x^k \tag{2.3.6}$$

tal que, para todo $i = 1, \ldots, k$,

(a2) $\theta(x) \cdot E_x^i = E_{f(x)}^i$ e $V_x^i = V_x^{i+1} \oplus E_x^i$; logo, $\dim E_x^i = \dim V_x^i - \dim V_x^{i+1}$;

(b2) $\displaystyle\lim_{n\to\pm\infty} \frac{1}{n} \log \|\phi^n(x)v\| = \lambda_i(x)$ para todo $v \in E_x^i$ diferente de zero;

(c2) $\displaystyle\lim_{n\to+\infty} \frac{1}{n} \log |\det \phi^n(x)| = \sum_{i=1}^{k} d_i(x)\lambda_i(x)$, onde $d_i(x) = \dim E_x^i$.

O leitor encontra uma discussão muita mais completa destes resultados, incluindo as demonstrações, no Capítulo 4 do livro [Via14].

2.3.2 Exercícios

2.3.1. Mostre que sob as hipóteses do teorema de von Neumann vale a seguinte conclusão mais forte:

$$\lim_{n-m\to\infty} \frac{1}{n-m} \sum_{j=m}^{n-1} \varphi \circ f^j \to P(\varphi).$$

2.3.2. Use o exercício anterior para mostrar que dado $A \subset M$ com $\mu(A) > 0$, o conjunto dos valores de $n \in \mathbb{N}$ tais que $\mu(A \cap f^{-n}(A)) > 0$ é sindético. [Observação: Já vimos outra prova deste fato no Exercício 1.6.10.]

2.3.3. Prove que o conjunto $F = \{\varphi \in L^1(\mu) : \varphi \text{ é } f\text{-invariante}\}$ é um subespaço fechado de $L^1(\mu)$.

2.3.4. Enuncie e prove uma versão do teorema de von Neumann para fluxos.

2.3.5. Seja $f^t : M \to M$, $t \in \mathbb{R}$ um fluxo contínuo num espaço métrico compacto M e seja μ uma probabilidade invariante. Verifique que o grupo a 1-parâmetro $U_t : L^2(\mu) \to L^2(\mu)$, $t \in \mathbb{R}$ dos operadores de Koopman $\varphi \mapsto U_t\varphi = \varphi \circ f^t$ é fortemente contínuo. Mostre que μ é ergódica se, e somente se, 0 é um autovalor simples do gerador infinitesimal desse grupo.

2.3.6. Seja $X = \{x_1, \ldots, x_r\}$ um conjunto finito e seja $\sigma : X \to X$ uma permutação. A permutação σ é chamada de *cíclica* se ela admite uma (única) órbita de cardinalidade r.

1. Dada uma permutação cíclica σ e uma função $\varphi : X \to \mathbb{R}$ prove que

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi(\sigma^i(x)) = \frac{\varphi(x_1) + \cdots + \varphi(x_r)}{r}.$$

2. Mais geralmente, prove que para toda permutação σ e função φ

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi(\sigma^i(x)) = \frac{\varphi(x) + \varphi(\sigma(x)) + \cdots + \varphi(\sigma^{p-1}(x))}{p},$$

onde $p \geq 1$ é a cardinalidade da órbita de x.

2.3.7. Verifique que o Lema 2.2.5 também pode ser deduzido do teorema ergódico de Birkhoff. Desta forma, também podemos enfraquecer a hipótese: basta supor que ϕ é mensurável e $\psi = \phi \circ f - \phi$ é integrável.

2.3.8. Uma função $\varphi : \mathbb{Z} \to \mathbb{R}$ é dita *uniformemente quase periódica* se para cada $\varepsilon > 0$ existe $L(\varepsilon) \in \mathbb{N}$ tal que todo intervalo $\{n+1, \ldots, n+L(\varepsilon)\}$ no conjunto $\mathbb{Z}$ tem algum elemento τ tal que $|\varphi(k+\tau) - \varphi(k)| < \varepsilon$ para todo $k \in \mathbb{Z}$. Chamaremos τ de um *ε-quase período* de f.

(a) Prove que se φ é uniformemente quase periódica então ela é limitada.

(b) Mostre que para todo $\varepsilon > 0$ existe $\rho \geq 1$ tal que

$$\Big|\frac{1}{\rho} \sum_{j=n\rho+1}^{(n+1)\rho} \varphi(j) - \frac{1}{\rho} \sum_{j=1}^{\rho} \varphi(j)\Big| < 2\varepsilon \quad \text{para todo } n \geq 1.$$

(c) Mostre que a sequência $(1/n)\sum_{j=1}^{n} \varphi(j)$ converge para algum número real quando $n \to \infty$.

(d) Mais geralmente, prove que $\lim_n (1/n)\sum_{k=1}^{n} \varphi(x+k)$ existe para todo $x \in \mathbb{Z}$ e é independente de x.

2.3.9. Prove que para Lebesgue quase todo ponto $x \in [0,1]$, a média geométrica dos números inteiros $a_1, \ldots, a_n, \ldots$ na expansão de x em fração contínua converge para algum valor, ou seja, existe $b \in \mathbb{R}$ tal que $\lim_n (a_1 a_2 \cdots a_n)^{1/n} = b$. [Observação: Compare com o Exercício 3.2.12.]

2.3.10. Seja $\varphi : M \to \mathbb{R}$ uma função integrável e seja $\tilde{\varphi}$ a sua média temporal, dada pelo Teorema 2.2.3. Mostre que se $\varphi \in L^p(\mu)$ para algum $p > 1$ então $\tilde{\varphi} \in L^p(\mu)$ e vale $\|\tilde{\varphi}\|_p \leq \|\varphi\|_p$. Além disso,

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

converge para $\tilde{\varphi}$ no espaço $L^p(\mu)$.

2.3.11. Prove o teorema de Birkhoff para fluxos: se μ é uma probabilidade invariante por um fluxo f e $\varphi \in L^1(\mu)$ então a função

$$\tilde{\varphi}(x) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \varphi(f^t(x))\, dt$$

está definida em μ-quase todo ponto e $\int \tilde{\varphi}\, d\mu = \int \varphi\, d\mu$.

2.3.12. Prove que se uma transformação contínua $f : M \to M$ de um espaço métrico compacto M só admite uma probabilidade invariante μ e ela é tal que $\mu(A) > 0$ para todo aberto não vazio $A \subset M$, então toda órbita de f é densa em M.

2.3.13. Dê uma demonstração direta do teorema ergódico de Birkhoff (Teorema 2.2.3), usando a abordagem da demonstração do Teorema 2.3.3.

2.3.14. Dada uma sequência subaditiva $(\varphi_n)_n$ com $\varphi_1^+ \in L^1(\mu)$, mostre que as funções

$$\varphi_- = \liminf_n \frac{\varphi_n}{n} \quad \text{e} \quad \varphi_+ = \limsup_n \frac{\varphi_n}{n}$$

são f-invariantes, isto é, $\varphi_-(x) = \varphi_- \circ f(x)$ e $\varphi_+(x) = \varphi_+ \circ f(x)$ para μ-quase todo $x \in M$.

2.3.15. Enuncie e prove o teorema ergódico subaditivo para fluxos.

2.3.16. Seja $f : M \to M$ um difeomorfismo de classe C^1 numa variedade compacta, preservando a medida de Lebesgue. Verifique que

$$\sum_{i=1}^{k(x)} d_i(x)\lambda_i(x) = 0 \quad \text{em } \mu\text{-quase todo ponto } x \in M$$

onde $\lambda_i(x)$, $i = 1, \dots, k(x)$ são os expoentes de Lyapunov de Df no ponto x e $d_i(x)$, $i = 1, \dots, k(x)$ são as respectivas multiplicidades.

2.3.17. Seja $(\varphi_n)_n$ uma sequência subaditiva de funções para uma transformação $f : M \to M$. Chamamos *constante temporal* de $(\varphi_n)_n$ ao limite

$$\lim_n \frac{1}{n} \int \varphi_n \, d\mu.$$

Supondo que o limite existe e é finito, mostre que podemos escrever $\varphi_n = \psi_n + \gamma_n$ para cada n, de tal forma que $(\psi_n)_n$ é uma sequência aditiva e $(\gamma_n)_n$ é uma sequência subaditiva com constante temporal igual a zero.

2.3.18. Nas condições do teorema de Furstenberg-Kesten, mostre que a sequência $\psi_n = (1/n) \log \|\phi^n\|$ é *uniformemente integrável*, no seguinte sentido: para todo $\varepsilon > 0$ existe $\delta > 0$ tal que

$$\mu(E) < \delta \quad \Rightarrow \quad \int_E \psi_n^+ \, d\mu < \varepsilon \text{ para todo } n.$$

2.3.19. Nas condições do teorema de Furstenberg-Kesten, para cada $k \geq 1$, seja Ψ_k a média temporal da função $\psi_k = (1/k) \log \|\phi^k\|$ relativamente à transformação f^k. Mostre que $\lambda_{\max}(x) \leq \Psi_k(x)$ para todo k e μ-quase todo x. Usando o Exercício 2.3.18, mostre que para todo $\rho > 0$ e μ-quase todo x existe k tal que $\Psi_k(x) \leq \lambda_{\max}(x) + \rho$.

Capítulo 3

Ergodicidade

Os teoremas apresentados no capítulo anterior dão plena justificativa à primeira parte da hipótese ergódica de Boltzmann: o tempo médio de visita $\tau(E, x)$ a um dado conjunto mensurável E está bem definido para quase todo ponto x. A segunda parte da hipótese ergódica, isto é, que o tempo médio de visita seja igual à medida de E para quase todo ponto x, é um enunciado de natureza diferente e será o tema do presente capítulo.

Ao longo do capítulo sempre suporemos que μ é uma medida de probabilidade invariante por uma transformação mensurável $f : M \to M$. Diremos que o sistema (f, μ) é *ergódico* se, dado qualquer conjunto mensurável E, temos $\tau(E, x) = \mu(E)$ para μ-quase todo ponto $x \in M$. Vamos ver que isto equivale a dizer que o sistema é dinamicamente indivisível, no sentido de que qualquer conjunto invariante tem medida nula ou medida total. Outras formulações equivalentes da propriedade de ergodicidade serão discutidas na Seção 3.1. Uma delas é que médias temporais coincidem com médias espaciais: para toda função integrável φ,

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) = \int \varphi \, d\mu \quad \text{em } \mu\text{-quase todo ponto.}$$

Na Seção 3.2 apresentaremos, por meio de exemplos, diversas técnicas para provar ou negar ergodicidade. A maioria será reutilizada posteriormente, em situações mais complexas. Em seguida adotaremos o seguinte ponto de vista: fixamos o sistema dinâmico e analisamos as propriedades das medidas ergódicas dentro do espaço de todas as medidas invariantes desse sistema dinâmico. As medidas ergódicas são precisamente os elementos extremais desse espaço.

3.1 Sistemas ergódicos

Conforme dissemos, a medida μ diz-se ergódica para f (ou f diz-se ergódica relativamente a μ) se o tempo médio de visita a qualquer conjunto mensurável coincide, em μ-quase todo ponto, com a medida desse conjunto. Nas duas subseções a seguir estudaremos diversas propriedades equivalentes a esta.

3.1.1 Conjuntos e funções invariantes

Dizemos que uma função mensurável $\varphi : M \to \mathbb{R}$ é *invariante* se $\varphi = \varphi \circ f$ em μ-quase todo ponto. Ou seja, a menos de um conjunto com medida nula, a função é constante em toda trajetória de f. Além disso, dizemos que um conjunto mensurável $B \subset M$ é *invariante* se a sua função característica $\mathcal{X}_B$ é uma função invariante. Em outras palavras, B é invariante se ele difere da sua pré-imagem $f^{-1}(B)$ por um conjunto de medida nula:

$$\mu(B \Delta f^{-1}(B)) = 0.$$

Veja no Exercício 1.6.4 formulações equivalentes desta propriedade. É fácil verificar que a família de todos os conjuntos invariantes é uma σ-álgebra, isto é, ela é fechada para o complementar e para uniões e interseções enumeráveis.

Exemplo 3.1.1. Seja f a expansão decimal, introduzida na Seção 1.3.1, e seja μ a medida de Lebesgue. Claramente, o conjunto $A = \mathbb{Q} \cap [0,1]$ dos números racionais é invariante. Outros exemplos interessantes são os conjunto de pontos $x = 0, a_1 a_2 \ldots$ em $[0,1]$ com uma dada proporção de dígitos a_i com cada valor $k \in \{0, \ldots, 9\}$. Mais precisamente, dado qualquer vetor $p = (p_0, \ldots, p_9)$ tal que $p_i \geq 0$ para todo i e $\sum_i p_i = 1$, defina

$$A_p = \{x : \lim_n \frac{1}{n} \#\{1 \leq i \leq n : a_i = k\} = p_k \text{ para } k = 0, \ldots, 9\}.$$

Para ver que A_p é invariante, observe que se $x = 0, a_1 a_2 \ldots$ então todo ponto $y \in f^{-1}(x)$ se escreve na forma $y = 0, b a_1 a_2 \ldots$ para algum $b \in \{0, \ldots, 9\}$. É claro que o dígito extra b não muda a frequência dos diversos valores 0, …, 9 na expansão decimal. Portanto $y \in A_p$ se, e somente se, $x \in A_p$.

Exemplo 3.1.2. Seja $\varphi : [0,1] \to \mathbb{R}$ uma função em $L^1(\mu)$. De acordo com o teorema ergódico de Birkhoff (Teorema 2.2.3), a sua média temporal $\tilde{\varphi}$ é uma função invariante. Então, todo conjunto de nível

$$B_c = \{x \in [0,1]; \tilde{\varphi}(x) = c\}$$

é invariante. Observe também que toda função invariante é desta forma: é fácil ver que se φ é invariante então ela coincide em μ-quase todo ponto com a sua média temporal $\tilde{\varphi}$.

A seguinte proposição coleta diversas maneiras equivalentes de definir ergodicidade. Dizemos que uma função φ é constante em μ-quase todo ponto se existe $c \in \mathbb{R}$ tal que $\varphi(x) = c$ para μ-quase todo $x \in M$.

Proposição 3.1.3. *Seja μ uma probabilidade invariante de uma transformação mensurável $f : M \to M$. As seguintes condições são equivalentes:*

(a) *Para todo conjunto mensurável $B \subset M$ tem-se $\tau(B,x) = \mu(B)$ para μ-quase todo ponto.*

(b) *Para todo conjunto mensurável $B \subset M$ a função $\tau(B,\cdot)$ é constante em μ-quase todo ponto.*

(c) *Para toda função integrável $\varphi : M \to \mathbb{R}$ tem-se $\tilde{\varphi}(x) = \int \varphi \, d\mu$ para μ-quase todo ponto.*

(d) *Para toda função integrável $\varphi : M \to \mathbb{R}$ a média temporal $\tilde{\varphi} : M \to \mathbb{R}$ é constante em μ-quase todo ponto.*

(e) *Para toda função integrável invariante $\psi : M \to \mathbb{R}$ tem-se $\psi(x) = \int \psi \, d\mu$ para μ-quase todo ponto.*

(f) *Toda função integrável invariante $\psi : M \to \mathbb{R}$ é constante em μ-quase todo ponto.*

(g) *Para todo subconjunto invariante A tem-se $\mu(A) = 0$ ou $\mu(A) = 1$.*

Demonstração. É imediato que (a) implica (b), que (c) implica (d) e que (e) implica (f). Também é claro que (e) implica (c) e (f) implica (d), porque a média temporal é uma função invariante (lembre da Proposição 2.2.4). Analogamente, (c) implica (a) e (d) implica (b), porque o tempo médio de visita é uma média temporal (da função característica de B). Agora basta provar as seguintes implicações:

(b) implica (g): Seja A um conjunto invariante. Então $\tau(A,x) = 1$ para μ-quase todo $x \in A$ e $\tau(A,x) = 0$ para μ-quase todo $x \in A^c$. Como $\tau(A,\cdot)$ é constante em μ-quase todo ponto, por hipótese, segue que $\mu(A) = 0$ ou $\mu(A) = 1$.

(g) implica (e): Seja ψ uma função integrável invariante. Então todo conjunto

$$B_c = \{x \in M : \psi(x) \leq c\}$$

é invariante. Logo, a hipótese implica que $\mu(B_c) \in \{0,1\}$ para todo $c \in \mathbb{R}$. Como $c \mapsto \mu(B_c)$ é não-decrescente, segue que existe $\bar{c} \in \mathbb{R}$ tal que $\mu(B_c) = 0$ para todo $c < \bar{c}$ e $\mu(B_c) = 1$ para todo $c \geq \bar{c}$. Então $\psi = \bar{c}$ em μ-quase todo ponto. Logo, $\int \psi \, d\mu = \bar{c}$ e, portanto, $\psi = \int \psi \, d\mu$ em μ-quase todo ponto. □

3.1.2 Caracterização espectral

A próxima proposição caracteriza a propriedade de ergodicidade por meio do operador de Koopman $U_f(\varphi) = \varphi \circ f$:

Proposição 3.1.4. *Seja μ uma probabilidade invariante de uma transformação mensurável $f : M \to M$. As seguintes condições são equivalentes:*

(a) *(f,μ) é ergódico.*

(b) Para qualquer par de conjuntos mensuráveis A e B vale

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} \mu(f^{-j}(A) \cap B) = \mu(A)\mu(B). \tag{3.1.1}$$

(c) Para quaisquer funções $\varphi \in L^p(\mu)$ e $\psi \in L^q(\mu)$, com $1/p + 1/q = 1$, vale

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} \int (U_f^j \varphi)\psi \, d\mu = \int \varphi \, d\mu \int \psi \, d\mu. \tag{3.1.2}$$

Demonstração. É claro que (c) implica (b): basta tomar $\varphi = \mathcal{X}_A$ e $\psi = \mathcal{X}_B$. Para mostrar que (b) implica (a), suponha que A é um conjunto invariante. Tomando $A = B$ na hipótese (b), obtemos que

$$\mu(A) = \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \mu(f^{-j}(A) \cap A) = \mu(A)^2.$$

Isto implica que $\mu(A) = 0$ ou $\mu(A) = 1$.

Agora resta provar que (a) implica (c). Considere $\varphi \in L^p(\mu)$ e $\psi \in L^q(\mu)$. Por ergodicidade e pelo teorema ergódico de Birkhoff (Teorema 2.2.3) temos que

$$\frac{1}{n} \sum_{j=0}^{n-1} U_f^j \varphi \to \int \varphi \, d\mu \tag{3.1.3}$$

em μ-quase todo ponto. Inicialmente, suponha que $|\varphi| \le k$ para algum $k \ge 1$. Então, para todo $n \in \mathbb{N}$,

$$\Big| \Big(\frac{1}{n} \sum_{j=0}^{n-1} U_f^j \varphi \Big) \psi \Big| \le k|\psi|.$$

Portanto, como $k|\psi| \in L^1(\mu)$, podemos usar o teorema da convergência dominada (Teorema A.2.11) para concluir que

$$\int \Big(\frac{1}{n} \sum_{j=0}^{n-1} U_f^j \varphi \Big) \psi \, d\mu \to \int \varphi \, d\mu \int \psi \, d\mu.$$

Isto prova a afirmação (3.1.2) quando φ é limitada. Falta remover esta última condição. Dado qualquer $\varphi \in L^p(\mu)$ e dado $k \ge 1$, defina

$$\varphi_k(x) = \begin{cases} k & \text{se } \varphi(x) > k \\ \varphi(x) & \text{se } \varphi(x) \in [-k, k] \\ -k & \text{se } \varphi(x) < -k. \end{cases}$$

Fixemos $\varepsilon > 0$. Pelo argumento anterior, para todo $k \geq 1$ vale que

$$\Big|\int (\frac{1}{n}\sum_{j=0}^{n-1} U_f^j \varphi_k)\psi \, d\mu - \int \varphi_k \, d\mu \int \psi \, d\mu\Big| < \varepsilon \tag{3.1.4}$$

se n é suficientemente grande (dependendo de k). Em seguida, observe que $\|\varphi_k - \varphi\|_p \to 0$ quando $k \to \infty$: isto é claro quando $p = \infty$, porque $\varphi_k = \varphi$ para todo $k > \|\varphi\|_\infty$; para $p < \infty$ use o teorema da convergência monótona (Teorema A.2.9). Logo, usando a desigualdade de Hölder, temos que

$$\Big|\int (\varphi_k - \varphi) \, d\mu \int \psi \, d\mu\Big| \leq \|\varphi_k - \varphi\|_p \, \Big|\int \psi \, d\mu\Big| < \varepsilon, \tag{3.1.5}$$

para todo k suficientemente grande. De modo semelhante,

$$\begin{aligned} \Big|\int \frac{1}{n}\sum_{j=0}^{n-1} U_f^j(\varphi_k - \varphi)\, \psi \, d\mu\Big| &\leq \frac{1}{n}\sum_{j=0}^{n-1} \Big|\int U_f^j(\varphi_k - \varphi)\, \psi \, d\mu\Big| \\ &\leq \frac{1}{n}\sum_{j=0}^{n-1} \|U_f^j(\varphi_k - \varphi)\|_p \, \|\psi\|_q \, d\mu \\ &= \|\varphi_k - \varphi\|_p \, \|\psi\|_q < \varepsilon, \end{aligned} \tag{3.1.6}$$

para todo n e todo k suficientemente grande, independente de n. Fixe k tal que (3.1.5) e (3.1.6) sejam válidas e, em seguida, tome n suficientemente grande para que (3.1.4) valha igualmente. Somando as três relações (3.1.4) a (3.1.6), obtemos que

$$\Big|\int (\frac{1}{n}\sum_{j=0}^{n-1} U_f^j \varphi)\psi \, d\mu - \int \varphi \, d\mu \int \psi \, d\mu\Big| < 3\varepsilon$$

para todo n suficientemente grande. Isto conclui a prova da condição (c). □

No caso $p = q = 2$, a condição (3.1.2) pode ser expressa em termos do produto interno $\cdot$ no espaço $L^2(\mu)$. Desta forma obtemos que (f, μ) é ergódico se, e somente se:

$$\lim_n \frac{1}{n}\sum_{j=0}^{n-1} \big[(U_f^n \varphi) - (\varphi \cdot 1)\big] \cdot \psi = 0 \quad \text{para todo } \varphi, \psi \in L^2(\mu). \tag{3.1.7}$$

Usaremos algumas vezes os seguintes fatos elementares: dados quaisquer conjuntos mensuráveis A e B,

$$\begin{aligned} |\mu(A) - \mu(B)| &= |\mu(A \setminus B) - \mu(B \setminus A)| \\ &\leq \mu(A \setminus B) + \mu(B \setminus A) = \mu(A \Delta B) \end{aligned} \tag{3.1.8}$$

e dados quaisquer conjuntos A_1, A_2, B_1, B_2,

$$(A_1 \cap A_2)\Delta(B_1 \cap B_2) \subset (A_1 \Delta B_1) \cup (A_2 \Delta B_2). \tag{3.1.9}$$

Corolário 3.1.5. *Suponha que a condição* (3.1.1) *na Proposição 3.1.4 é satisfeita para todo A e B em alguma álgebra $\mathcal{A}$ que gera a σ-álgebra dos conjuntos mensuráveis. Então (f,μ) é ergódico.*

Demonstração. Sejam A e B conjuntos mensuráveis quaisquer. Pelo teorema de aproximação (Teorema A.1.19), dado qualquer $\varepsilon > 0$ existem A_0 e B_0 em $\mathcal{A}$ tais que $\mu(A\Delta A_0) < \varepsilon$ e $\mu(B\Delta B_0) < \varepsilon$. Observe que

$$\begin{aligned}\left|\mu(f^{-j}(A)\cap B) - \mu(f^{-j}(A_0)\cap B_0)\right| &\le \mu(f^{-j}(A)\Delta f^{-j}(A_0)) + \mu(B\Delta B_0)\\ &= \mu(A\Delta A_0) + \mu(B\Delta B_0) < 2\varepsilon\end{aligned}$$

(a igualdade usa o fato de que μ é medida invariante) para todo j

$$|\mu(A)\mu(B) - \mu(A_0)\mu(B_0)| \le \mu(A\Delta A_0) + \mu(B\Delta B_0) < 2\varepsilon.$$

Então, a hipótese

$$\lim_n \frac{1}{n}\sum_{j=0}^{n-1}\mu(f^{-j}(A_0)\cap B_0) = \mu(A_0)\mu(B_0)$$

implica que

$$\begin{aligned}-4\varepsilon &\le \liminf_n \frac{1}{n}\sum_{j=0}^{n-1}\mu(f^{-j}(A)\cap B) - \mu(A)\mu(B)\\ &\le \limsup_n \frac{1}{n}\sum_{j=0}^{n-1}\mu(f^{-j}(A)\cap B) - \mu(A)\mu(B) \le 4\varepsilon.\end{aligned}$$

Como ε é arbitrário, isto prova que a condição (3.1.1) vale para todo par de conjuntos mensuráveis. De acordo com a Proposição 3.1.4, segue que o sistema é ergódico. □

De modo semelhante, basta verificar o item (c) da Proposição 3.1.4 em subconjuntos densos. A prova deste fato fica a cargo do leitor (veja o Exercício 3.1.3):

Corolário 3.1.6. *Suponha que a condição* (3.1.2) *na Proposição 3.1.4 é satisfeita para todo φ e ψ em subconjuntos densos de $L^p(\mu)$ e $L^q(\mu)$, respectivamente. Então (f,μ) é ergódico.*

3.1.3 Exercícios

3.1.1. Sejam $(M,\mathcal{A})$ um espaço mensurável e $f: M \to M$ uma transformação mensurável. Prove que se $p \in M$ é um ponto periódico de período k, então a medida $\mu_p = \frac{1}{k}(\delta_p + \delta_{f(p)} + \cdots + \delta_{f^{k-1}(p)})$ é ergódica.

3.1.2. Seja μ uma probabilidade invariante, não necessariamente ergódica, de uma transformação mensurável $f : M \to M$. Mostre que dados quaisquer conjuntos mensuráveis A e B existe o limite

$$\lim_n \frac{1}{n} \sum_{i=0}^{n-1} \mu(f^{-i}(A) \cap B).$$

3.1.3. Mostre que uma probabilidade invariante μ é ergódica para uma transformação f se, e somente se, ocorre qualquer uma das seguintes condições:

(a) $\mu(\bigcup_{n\geq 0} f^{-n}(A)) = 1$ para todo A mensurável com $\mu(A) > 0$;

(b) dados quaisquer conjuntos mensuráveis A, B com $\mu(A)\mu(B) > 0$, existe $n \geq 1$ tal que $\mu\big(f^{-n}(A) \cap B\big) > 0$;

(c) a convergência na condição (c) da Proposição 3.1.4 vale para alguma escolha de p, q e algum subconjunto denso de funções $\varphi \in L^p(\mu)$ e $\psi \in L^q(\mu)$;

(d) existe $p \in [1, \infty]$ tal que toda função invariante $\varphi \in L^p(\mu)$ é constante em μ-quase todo ponto;

(e) toda função integrável φ com $\varphi \circ f \geq \varphi$ em μ-quase todo ponto (ou $\varphi \circ f \leq \varphi$ em μ-quase todo ponto) é constante em μ-quase todo ponto.

3.1.4. Suponha que M é um espaço métrico. Prove que μ é ergódica para $f : M \to M$ se, e somente se, a média temporal de toda função uniformemente contínua limitada $\varphi : M \to \mathbb{R}$ é constante em μ-quase todo ponto.

3.1.5. Suponha que M é um espaço métrico. Chamamos *bacia* de uma probabilidade invariante μ ao conjunto $B(\mu)$ dos pontos $x \in M$ tais que

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) = \int \varphi \, d\mu$$

para toda função contínua limitada $\varphi : M \to \mathbb{R}$. Justifique que a bacia é um conjunto invariante. Além disso, se μ é ergódica então $B(\mu)$ tem μ-medida total.

3.1.6. Mostre que se μ e η são probabilidades ergódicas distintas de uma transformação $f : M \to M$, então η e μ são mutuamente singulares.

3.1.7. Seja μ uma probabilidade invariante de uma transformação $f : M \to M$. Mostre que a medida produto $\mu_2 = \mu \times \mu$ é invariante pela transformação $f_2 : M \times M \to M \times M$ definida por $f_2(x, y) = (f(x), f(y))$. Além disso, se (f_2, μ_2) é ergódico então (f, μ) é ergódico. A recíproca é verdadeira?

3.1.8. Seja $f : M \to M$ uma transformação preservando uma probabilidade μ. Suponha que (f^n, μ) é ergódico para todo $n \geq 1$. Mostre que se φ é uma autofunção não constante do operador de Koopman U_f então o autovalor *não* é raiz da unidade e qualquer conjunto onde φ é constante tem medida nula.

3.2 Exemplos

Nesta seção apresentamos, por meio de exemplos, diversos métodos para verificar se um dado sistema é ou não ergódico.

3.2.1 Rotações em toros

Consideremos inicialmente o caso de uma rotação $R_\theta : S^1 \to S^1$ no círculo $S^1 = \mathbb{R}/\mathbb{Z}$. Conforme observamos na Seção 1.3.3, a medida de Lebesgue m é invariante por R_θ. Queremos analisar o comportamento ergódico do sistema (R_θ, m) para os diferentes valores de θ.

Se θ é racional, digamos $\theta = p/q$ em forma irredutível, então $R_\theta^q(x) = x$ para todo $x \in S^1$. Então, dado qualquer segmento $I \subset S^1$ com comprimento menor que $1/q$, o conjunto

$$A = I \cup R_\theta(I) \cup \cdots \cup R_\theta^{q-1}(I)$$

é invariante e a sua medida de Lebesgue satisfaz $0 < m(A) < 1$. Assim, se θ é racional a medida de Lebesgue *não* é ergódica. A recíproca é muito mais interessante:

Proposição 3.2.1. *Se θ é irracional, então R_θ é ergódica para a medida de Lebesgue.*

Vamos mencionar duas demonstrações diferentes deste fato. A primeira, que detalharemos a seguir, usa fatos simples de Análise de Fourier. A segunda, que deixaremos como exercício (Exercício 3.2.6), é baseada num argumento de ponto de densidade semelhante ao que usaremos na Seção 3.2.2 para provar a ergodicidade da expansão decimal.

Como anteriormente, denotamos por $L^2(m)$ o espaço de Hilbert das funções mensuráveis ψ cujo quadrado é integrável, ou seja, tais que:

$$\int |\psi|^2 \, dm < \infty.$$

É conveniente considerarmos funções com valores em $\mathbb{C}$, e assim será feito ao longo da seção. Usaremos o fato bem conhecido de que a família de funções

$$\phi_k : S^1 \to \mathbb{C}, \ x \mapsto e^{2\pi i k x}, \quad k \in \mathbb{Z}$$

é uma base de Hilbert deste espaço: dado qualquer $\varphi \in L^2(m)$ existe uma única sequência $(a_k)_{k\in\mathbb{Z}}$ de números complexos tais que

$$\varphi(x) = \sum_{k\in\mathbb{Z}} a_k e^{2\pi i k x} \quad \text{para quase todo } x \in S^1. \tag{3.2.1}$$

Considere a expansão em série de Fourier (3.2.1) de uma função qualquer $\varphi \in L^2(m)$. Então

$$\varphi\big(R_\theta(x)\big) = \sum_{k\in\mathbb{Z}} a_k e^{2\pi i k\theta} e^{2\pi i k x}. \tag{3.2.2}$$

Suponha que φ é invariante. Então (3.2.1) e (3.2.2) coincidem. Pela unicidade dos coeficientes da expansão de Fourier, isto acontece se, e somente se,

$$a_k e^{2\pi i k\theta} = a_k \quad \text{para todo } k \in \mathbb{Z}.$$

A hipótese de que θ é irracional significa que $e^{2\pi i k\theta} \neq 1$ para todo $k \neq 0$. Então a relação que acabamos de obter implica que $a_k = 0$ para todo $k \neq 0$. Em outras palavras, $\varphi(z) = a_0$ para m-quase todo $z \in S^1$. Isto mostra que toda função invariante em $L^2(m)$ é constante em m-quase todo ponto. Em particular, a função característica $\varphi = \mathcal{X}_A$ de qualquer conjunto invariante $A \subset S^1$ é constante em m-quase todo ponto. Isto é o mesmo que dizer que A tem medida zero ou um. Logo, pela Proposição 3.1.3, temos que m é ergódica.

Estas observações estendem-se naturalmente às rotações no d-toro $\mathbb{T}^d$, para qualquer $d \geq 1$:

Proposição 3.2.2. *Se $\theta = (\theta_1, \ldots, \theta_d)$ é racionalmente independente então a rotação $R_\theta : \mathbb{T}^d \to \mathbb{T}^d$ é ergódica para a medida de Lebesgue.*

Isto pode ser provado por um argumento análogo ao do caso $d = 1$, usando o fato de que a família de funções

$$\phi_{k_1,\ldots,k_d} : \mathbb{T}^d \to \mathbb{C},\ (x_1, \ldots, x_d) \mapsto e^{2\pi i(k_1 x_1 + \cdots + k_d x_d)}, \quad (k_1, \ldots, k_d) \in \mathbb{Z}^d$$

é uma base de Hilbert do espaço $L^2(m)$ das funções $\varphi : \mathbb{T}^d \to \mathbb{C}$ com quadrado somável. Deixamos esta tarefa ao cuidado do leitor (Exercício 3.2.1).

Corolário 3.2.3. *Se $\theta = (\theta_1, \ldots, \theta_d)$ é racionalmente independente então a rotação $R_\theta : \mathbb{T}^d \to \mathbb{T}^d$ é minimal, ou seja, toda órbita $\mathcal{O}(x) = \{R_\theta^n(x) : n \in \mathbb{N}\}$ é densa em $\mathbb{T}^d$.*

Demonstração. Consideremos em $\mathbb{T}^d$ a distância plana, que é definida por

$$d([\xi], [\eta]) = \inf\{d(\xi', \eta') : \xi', \eta' \in \mathbb{R}^d, \xi' \sim \xi, \eta' \sim \eta\}.$$

Observe que esta distância é preservada por toda rotação. Seja $\{U_k : k \in \mathbb{N}\}$ uma base enumerável de abertos de $\mathbb{T}^d$ e seja m a medida de Lebesgue em $\mathbb{T}^d$. Por ergodicidade, existe $W \subset \mathbb{T}^d$, com medida de Lebesgue total, tal que $\tau(U_k, x) = m(U_k) > 0$ para todo k e todo $x \in W$. Em particular, a órbita de x é densa em $\mathbb{T}^d$ para todo $x \in W$. Agora considere um ponto arbitrário $x \in M$ e seja $y \in W$ qualquer. Então, para todo $\delta > 0$ existe $k \geq 1$ tal que $d(f^k(y), x) < \delta$. Segue então que $d(f^{n+k}(y), f^n(x)) < \delta$ para todo $n \geq 1$. Como a órbita de y é densa, isto implica que a órbita de x é δ-densa, ou seja, ela intersecta a δ-vizinhança de todo ponto. Como δ é arbitrário, isto implica que a órbita de x é densa no toro ambiente. $\square$

De fato as rotações irracionais no círculo ou, mais geralmente, num toro satisfazem uma condição muito mais forte do que ergodicidade: elas são *unicamente ergódicas*, o que quer dizer que elas têm uma única probabilidade invariante (que é a medida de Lebesgue, claro).

3.2.2 Expansão decimal

Considere a transformação $f : [0,1] \to [0,1]$, $f(x) = 10x - [10x]$ que gera a expansão decimal. Na Seção 1.3.1 verificamos que f preserva a medida de Lebesgue m. Afirmamos:

Proposição 3.2.4. *A transformação f é ergódica para a medida de Lebesgue m.*

Demonstração. De acordo com a Proposição 3.1.3, basta provar que todo conjunto invariante A tem medida total. O principal ingrediente é o teorema de derivação de Lebesgue (Teorema A.2.15), segundo o qual quase todo ponto de A é ponto de densidade de A. Mais precisamente (veja também o Exercício A.2.9), m-quase todo ponto $a \in A$ satisfaz

$$\lim_{\varepsilon \to 0} \inf \big\{ \frac{m(I \cap A)}{m(I)} : I \text{ intervalo tal que } a \in I \subset B(a,\varepsilon) \big\} = 1 \,. \tag{3.2.3}$$

Fixemos um ponto de densidade $a \in A$. Como o conjunto dos pontos da forma $m/10^k$, $k \in \mathbb{N}$, $0 \le m \le 10^k$ tem medida nula, podemos supor, sem qualquer restrição, que a não é desta forma. Consideremos a família de intervalos

$$I(k,m) = \big(\frac{m-1}{10^k}, \frac{m}{10^k} \big), \quad k \in \mathbb{N}, \quad m = 1, \dots, 10^k.$$

É claro que para cada $k \in \mathbb{N}$ existe um único $m = m_k$ tal que $I(k, m_k)$ contém o ponto a. Denotaremos $I_k = I(k, m_k)$. A propriedade (3.2.3) implica que

$$\frac{m(I_k \cap A)}{m(I_k)} \to 1 \quad \text{quando } k \to \infty.$$

Observe também que cada f^k é uma bijeção *afim* de I_k sobre o intervalo $(0,1)$. Isso tem a seguinte consequência, que é crucial para o nosso argumento:

Lema 3.2.5 (Distorção)**.** *Para todo $k \in \mathbb{N}$, vale*

$$\frac{m(f^k(E_1))}{m(f^k(E_2))} = \frac{m(E_1)}{m(E_2)} \tag{3.2.4}$$

para quaisquer subconjuntos mensuráveis E_1 e E_2 de I_k.

Aplicando este fato a $E_1 = I_k \cap A$ e $E_2 = I_k$ obtemos que

$$\frac{m\big(f^k(I_k \cap A)\big)}{m\big((0,1)\big)} = \frac{m(I_k \cap A)}{m(I_k)} \,.$$

Claro que $m\big((0,1)\big) = 1$. Além disso, como estamos supondo que A é invariante, $f^k(I_k \cap A)$ está contido em A. Deste modo obtemos que

$$m(A) \ge \frac{m(I_k \cap A)}{m(I_k)} \quad \text{para todo } k.$$

Como a sequência do lado direito converge para 1 quando $k \to \infty$, segue que $m(A) = 1$, como queríamos demonstrar. □

O Lema 3.2.5 depende do fato de que a transformação f é afim em cada intervalo $\big((m-1)/10, m/10\big)$ e isso pode dar a impressão de que o método de demonstração que acabamos de apresentar está restrito a uma classe muito particular de exemplos. De fato, não é assim, muito pelo contrário.

A razão é que existem muitas situações interessantes nas quais é possível obter uma versão apenas um pouco mais fraca do enunciado do Lema 3.2.5, mas que ainda é suficiente para concluir a demonstração da ergodicidade. Em poucas palavras, no lugar de afirmar que os dois lados de (3.2.4) são iguais, mostra-se, em muitos casos, que a razão entre os dois termos é limitada por alguma constante uniforme. Isso é chamado *propriedade de distorção limitada*. Como exemplo de aplicação destas ideias, na Seção 3.2.4 provaremos que a transformação de Gauss é ergódica.

Em seguida vamos dar uma aplicação da Proposição 3.2.4 no contexto da Teoria dos Números. Dizemos que um número $x \in \mathbb{R}$ é *normal na base* 10 se todo bloco de dígitos $(b_1, \ldots, b_l)$, $l \geq 1$ aparece com frequência 10^{-l} na expansão decimal de x. Números racionais nunca são normais, claro, e também é fácil dar exemplos irracionais, tais como $x = 0,101001000100001000001\cdots$. Além disso, não é difícil construir números normais como, por exemplo, a constante de Champernowne $x = 0,1234567891011121314151617181920 2122\cdots$, que é obtida por concatenação dos sucessivos números naturais.

No entanto, em geral é muito difícil decidir se um dado número irracional é normal ou não. Por exemplo, até hoje isso não é sabido para os números π, e e mesmo $\sqrt{2}$. Por outro lado, usando a proposição anterior é fácil mostrar que quase todo número é normal:

Proposição 3.2.6. *O conjunto dos números $x \in \mathbb{R}$ que são normais na base 10 tem medida de Lebesgue total.*

Demonstração. Como o fato de ser normal ou não é independente da parte inteira do número, só precisamos mostrar que quase todo $x \in [0,1]$ é normal. Considere $f : [0,1] \to [0,1]$ definida por $f(x) = 10x - [10x]$. Para cada bloco $(b_1, \ldots, b_l) \in \{0, \ldots, 9\}^l$ considere o intervalo

$$I_{b_1,\ldots,b_l} = \Big[\frac{\kappa}{10^l}, \frac{\kappa+1}{10^l}\Big) \quad \text{onde } \kappa = \sum_{i=1}^{l} b_i 10^{l-i}.$$

Recorde que se $x = 0, a_0 a_1 \cdots a_k a_{k+1} \cdots$ então $f^k(x) = 0, a_k a_{k+1} \cdots$ para cada $k \geq 1$. Portanto, $f^k(x) \in I_{b_1,\ldots,b_l}$ se, e somente se, o bloco de dígitos $(a_k, \ldots, a_{k+l-1})$ na expansão decimal de x coincide com $(b_1, \ldots, b_l)$. Logo, o tempo médio de visita $\tau(I_{b_1,\ldots,b_l}, x)$ é igual à frequência com que $(b_1, \ldots, b_l)$ ocorre na expansão decimal de x. Usando o teorema ergódico de Birkhoff e o fato de que a transformação f é ergódica relativamente à medida de Lebesgue m, concluímos que para cada $(b_1, \ldots, b_l)$ existe um subconjunto $B(b_1, \ldots, b_l)$ de $[0,1]$ com medida de Lebesgue total, tal que

$$\tau(I_{b_1,\ldots,b_l}, x) = m(I_{b_1,\ldots,b_l}) = \frac{1}{10^l} \quad \text{para todo } x \in B(b_1, \ldots, b_l).$$

Seja B a interseção dos $B(b_1, \ldots, b_l)$ sobre todos os valores de $b_1, \ldots, b_l$ em $\{0, \ldots, 9\}$ e todo $l \geq 1$. Então $m(B) = 1$ e todo $x \in B$ é normal na base 10. $\square$

Mais geralmente, para qualquer inteiro $d \geq 2$, dizemos que $x \in \mathbb{R}$ é *normal na base* d se todo bloco $(b_1, \ldots, b_l) \in \{0, \ldots, d-1\}^l$, $l \geq 1$ aparece com frequência d^{-l} na expansão de x na base d. Finalmente, dizemos que x é *normal* se ele é normal na base d para todo $d \geq 2$. Tudo que foi dito antes para $d = 10$ se estende imediatamente para qualquer d. Em particular, o conjunto dos números normais na base d tem medida de Lebesgue total para todo $d \geq 2$. Tomando a interseção sobre todos os valores de d concluímos que *Lebesgue quase todo número real é normal* (teorema normal de Borel).

3.2.3 Deslocamentos de Bernoulli

Seja $(X, \mathcal{C}, \nu)$ um espaço de probabilidade qualquer. Nesta seção consideramos o espaço produto $\Sigma = X^{\mathbb{N}}$, munido da σ-álgebra produto $\mathcal{B} = \mathcal{C}^{\mathbb{N}}$ e da medida produto $\mu = \nu^{\mathbb{N}}$ (veja a definição no Apêndice A.2.3). Isto quer dizer que Σ é o conjunto de todas as sequências $(x_n)_{n\in\mathbb{N}}$ com $x_n \in X$ para todo n. Por definição, $\mathcal{B}$ é a σ-álgebra gerada pelos cilindros

$$[m; A_m, \ldots, A_n] = \{(x_i)_{i\in\mathbb{N}} : x_i \in A_i \text{ para } m \leq i \leq n\}$$

onde $m \leq n$ e cada A_i é um elemento de $\mathcal{C}$. Além disso, μ é caracterizada por

$$\mu([m; A_m, \ldots, A_n]) = \prod_{i=m}^{n} \nu(A_i). \tag{3.2.5}$$

Podemos pensar nos elementos de Σ como representando os resultados de sequências de experimentos regidos por uma mesma distribuição de probabilidade ν: dado qualquer conjunto mensurável $A \subset X$, a probabilidade de obtermos $x_i \in A$ é igual a $\nu(A)$, qualquer que seja i. Além disso, os resultados dos sucessivos experimentos são independentes: de fato a relação (3.2.5) significa que a probabilidade de $x_i \in A_i$ para todo $m \leq i \leq n$ é o produto das probabilidades de cada um dos eventos $x_i \in A_i$ separadamente.

Nesta seção introduzimos uma dinâmica $\sigma : \Sigma \to \Sigma$ no espaço Σ, chamada deslocamento (ou "shift"), que preserva a medida μ. Chamaremos *deslocamento de Bernoulli* ao par (σ, μ). O principal resultado é que todo deslocamento de Bernoulli é ergódico.

Vale a pena observar que é possível substituir $\mathbb{N}$ por $\mathbb{Z}$ em toda a construção, ou seja, podemos considerar Σ como sendo o espaço das sequências bilaterais $(\ldots, x_{-n}, \ldots, x_0, \ldots, x_n, \ldots)$. A menos de pequenos ajustes, que deixamos a cargo do leitor, tudo o que vai ser dito em seguida permanece válido nesse caso. Além disso, no caso bilateral o deslocamento é uma aplicação invertível.

O *deslocamento* é a aplicação $\sigma : \Sigma \to \Sigma$ definida por

$$\sigma\big((x_n)_n\big) = (x_{n+1})_n.$$

Ou seja, σ envia a sequência $(x_0, x_1, \ldots, x_n, \ldots)$ na sequência $(x_1, \ldots, x_n, \ldots)$. Observe que a pré-imagem de qualquer cilindro ainda é um cilindro:

$$\sigma^{-1}([m; A_m, \ldots, A_n]) = [m+1; A_m, \ldots, A_n]. \tag{3.2.6}$$

Segue que σ é mensurável relativamente à σ-álgebra $\mathcal{B}$. Além disso,

$$\mu\big(\sigma^{-1}([m; A_m, \ldots, A_n])\big) = \nu(A_m) \cdots \nu(A_n) = \mu\big([m; A_m, \ldots, A_n]\big)$$

e (usando o Lema 1.3.1) isso assegura que a medida μ é invariante por σ.

Proposição 3.2.7. *Todo deslocamento de Bernoulli (σ, μ) é ergódico.*

Demonstração. Seja A um conjunto mensurável invariante qualquer. Queremos mostrar que $\mu(A) = 0$ ou $\mu(A) = 1$. Vamos usar o seguinte fato:

Lema 3.2.8. *Se B e C são uniões finitas de cilindros disjuntos dois-a-dois, então tem-se*

$$\mu(B \cap \sigma^{-j}(C)) = \mu(B)\mu(\sigma^{-j}(C)) = \mu(B)\mu(C),$$

para todo j suficientemente grande.

Demonstração. Para começar, suponhamos que B e C são ambos cilindros: $B = [k; B_k, \ldots, B_l]$ e $C = [m; C_m, \ldots, C_n]$. Então,

$$\sigma^{-j}(C) = [m+j; C_m, \ldots, C_n] \quad \text{para cada } j.$$

Considere qualquer j suficientemente grande para que $m + j > l$. Então,

$$\begin{aligned} B \cap \sigma^{-j}(C) &= \{(x_n)_n : x_k \in B_k, \ldots, x_l \in B_l, x_{m+j} \in C_m, \ldots, x_{n+j} \in C_n\} \\ &= [k; B_k, \ldots, B_l, X, \ldots, X, C_m, \ldots, C_n], \end{aligned}$$

onde X aparece exatamente $m + j - l - 1$ vezes. Pela definição (3.2.5), isto dá que

$$\mu(B \cap \sigma^{-j}(C)) = \prod_{i=k}^{l} \nu(B_i)\, 1^{m+j-l-1} \prod_{i=m}^{n} \nu(C_i) = \mu(B)\mu(C).$$

Isto prova a conclusão do lema quando os conjuntos envolvidos são cilindros. O caso geral segue imediatamente, pelo fato de μ ser finitamente aditiva. $\square$

Suponhamos, inicialmente, que o conjunto invariante A pertence à álgebra $\mathcal{B}_0$ das uniões finitas de cilindros disjuntos. Nesse caso podemos aplicar o lema anterior com $B = C = A$. Concluímos que $\mu(A \cap \sigma^{-j}(A)) = \mu(A)^2$ sempre que tomemos j suficientemente grande. Mas, como A é invariante, o lado esquerdo desta igualdade é $\mu(A)$. Desta forma obtemos que $\mu(A) = \mu(A)^2$, o que só pode acontecer se $\mu(A) = 0$ ou $\mu(A) = 1$.

Agora vamos fazer a prova quando A é um conjunto invariante mensurável qualquer. A ideia é aproximar o conjunto invariante por elementos da álgebra

$\mathcal{B}_0$, usando o teorema de aproximação (Teorema A.1.19): dado qualquer $\varepsilon > 0$ existe $B \in \mathcal{B}_0$ tal que $\mu(A\Delta B) < \varepsilon$. Fixemos j tal que

$$\mu(B \cap \sigma^{-j}(B)) = \mu(B)\mu(\sigma^{-j}(B)) = \mu(B)^2. \tag{3.2.7}$$

Usando (3.1.8) e (3.1.9) e o fato de que a medida μ é invariante por σ obtemos

$$\big|\mu(A \cap \sigma^{-j}(A)) - \mu(B \cap \sigma^{-j}(B))\big| \le 2\mu(A\Delta B) < 2\varepsilon \tag{3.2.8}$$

(um fato análogo foi deduzido durante a prova do Corolário 3.1.5). Além disso,

$$\big|\mu(A)^2 - \mu(B)^2\big| \le 2\big|\mu(A) - \mu(B)\big| < 2\varepsilon. \tag{3.2.9}$$

Juntando as relações (3.2.7), (3.2.8), (3.2.9), concluímos que $|\mu(A)-\mu(A)^2| < 4\varepsilon$. Como ε é arbitrário, deduzimos que $\mu(A) = \mu(A)^2$ e, portanto, ou $\mu(A) = 0$ ou $\mu(A) = 1$. □

Quando X é um espaço topológico, e $\mathcal{C}$ é a sua σ-álgebra de Borel, podemos munir Σ com a *topologia produto* que é, por definição, a topologia gerada pelos cilindros $[m; A_m, \ldots, A_n]$ onde os conjuntos $A_m, \ldots, A_n$ são abertos de X. A propriedade (3.2.6) implica que o deslocamento $\sigma : \Sigma \to \Sigma$ é contínuo para esta topologia. O teorema de Tychonoff (veja [Dug66]) afirma que Σ é compacto se X for compacto.

Um caso particular importante ocorre quando X é um conjunto finito munido da topologia discreta, na qual todo subconjunto é aberto. Dizemos que uma transformação $f : M \to M$ num espaço topológico M é *transitiva* se existe $x \in M$ cuja trajetória $f^n(x)$, $n \ge 1$ é densa em M. Deixamos a demonstração do próximo resultado a cargo do leitor (Exercício 3.2.2):

Proposição 3.2.9. *Seja X um conjunto finito e $\Sigma = X^{\mathbb{N}}$ ou $\Sigma = X^{\mathbb{Z}}$. Então o deslocamento $\sigma : \Sigma \to \Sigma$ é uma aplicação contínua e transitiva. Além disso, o conjunto dos pontos periódicos de σ é denso em Σ.*

A seguinte afirmação informal, uma variante do *paradoxo do macaco*, ilustra o significado da ergodicidade da medida μ: *Um macaco batendo ao acaso nas teclas de uma máquina de escrever durante tempo infinito acabará escrevendo o texto completo de "Os Lusíadas"* [1]*, quase certamente.*

Para "demonstrar" esta afirmação precisamos formulá-la de modo um pouco mais preciso. Os textos possivelmente digitados pelo macaco correspondem às sequências $(x_n)_{n\in\mathbb{N}}$ no conjunto (finito) X dos caracteres no teclado da máquina de escrever: letras, dígitos, espaço, sinais de pontuação, etc. Supomos que cada caracter $*$ no teclado tem uma probabilidade positiva p_* de ser digitado, a cada vez. Isto corresponde à medida de probabilidade

$$\nu = \sum_{*\in X} p_* \delta_*$$

[1] Poema épico monumental, em 10 cantos, de autoria do poeta português Luís de Camões, falecido em Lisboa em 1580.

no conjunto X dos caracteres. Também supomos que o caracter digitado a cada vez é independente dos caracteres anteriores. Então a distribuição das sequências $(x_n)_n$ está regida pela probabilidade de Bernoulli $\mu = \nu^{\mathbb{N}}$. Denotemos por $\sigma : \Sigma \to \Sigma$ a aplicação deslocamento no espaço $\Sigma = X^{\mathbb{N}}$.

O texto de "Os Lusíadas" corresponde a uma certa sequência finita (embora muito longa) de caracteres $(l_0, \ldots, l_N)$. Considere o cilindro $L = [0; l_0, \ldots, l_N]$. Então

$$\mu(L) = \prod_{j=1}^{N} p_{l_j}$$

é positivo (embora muito pequeno). Uma sequência $(x_n)_n$ contém o texto completo de "Os Lusíadas" precisamente se $\sigma^k\big((x_n)_n\big) \in L$ para algum $k \geq 0$. Pelo teorema ergódico de Birkhoff e pela ergodicidade de (σ, μ), o conjunto K dos valores de k para os quais isso acontece satisfaz

$$\lim_n \frac{1}{n} \#(K \cap [0, n-1]) = \mu(L) > 0. \tag{3.2.10}$$

com probabilidade total. Em particular, para quase toda sequência $(x_n)_n$ o conjunto K é infinito, o que significa que $(x_n)_n$ contém infinitas cópias de "Os Lusíadas". Na verdade, (3.2.10) conduz a uma conclusão ainda mais forte: sempre com probabilidade total, as cópias do nosso poema ocupam uma fração positiva (embora muito pequena) de todos os caracteres digitados. Em outras palavras, em média, o macaco digita uma nova cópia de "Os Lusíadas" a cada tantos (muitos) anos.

3.2.4 Transformação de Gauss

Como vimos na Seção 1.3.2, a transformação de Gauss $G(x) = 1/x - [1/x]$ admite uma probabilidade invariante μ que é equivalente à medida de Lebesgue, a saber:

$$\mu(E) = \frac{1}{\log 2} \int_E \frac{dx}{1+x}. \tag{3.2.11}$$

Proposição 3.2.10. *O sistema (G, μ) é ergódico.*

Este fato pode ser demonstrado por uma versão mais elaborada do método que usamos na Seção 3.2.2. Vamos esboçar o argumento da demonstração, focando na principal dificuldade adicional.

Seja A um conjunto invariante com medida positiva. Queremos mostrar que $\mu(A) = 1$. Em primeiro lugar, continua sendo verdade que para quase todo ponto $a \in [0, 1]$ existe uma sequência de intervalos I_k contendo a e tais que G^k envia I_k bijetivamente e diferenciavelmente sobre $(0, 1)$. Tais intervalos podem ser encontrados da seguinte forma. Primeiramente, considere

$$I(1, m) = \big(\frac{1}{m+1}, \frac{1}{m}\big),$$

para cada $m \geq 1$. Em seguida defina, por recorrência,

$$I(k, m_1, \ldots, m_k) = I(1, m_1) \cap G^{-k+1}\big(I(k-1, m_2, \ldots, m_k)\big)$$

para $m_1, \ldots, m_k \geq 1$. Então, basta tomar para I_k o intervalo $I(k, m_1, \ldots, m_k)$ que contém a. Isto está bem definido para todo $k \geq 1$ e todo ponto a no complementar de um conjunto enumerável, a saber, o conjunto $\cup_{k=0}^{\infty} G^{-k}(\{0,1\})$.

Por outro lado, embora a restrição de G^k a cada I_k seja uma bijeção diferenciável, ela não é afim. Por essa razão, não temos o análogo da relação (3.2.4) neste caso. Esta dificuldade é contornada por meio do seguinte resultado, que é um exemplo de controle da *distorção*: é importante notar que a constante K no enunciado é independente de I_k, E_1, E_2 e, sobretudo, k.

Proposição 3.2.11 (Distorção limitada)**.** *Existe uma constante $K > 1$ tal que para todo $k \geq 1$ e todo intervalo I_k tal que G^k restrita a I_k é uma bijeção diferenciável, tem-se*

$$\frac{\mu(G^k(E_1))}{\mu(G^k(E_2))} \leq K \frac{\mu(E_1)}{\mu(E_2)}$$

para quaisquer subconjuntos mensuráveis E_1 e E_2 de I_k.

Para a prova desta proposição precisamos de dois resultados auxiliares:

Lema 3.2.12. *Para todo $x \in (0,1]$ vale que*

$$|G'(x)| \geq 1 \quad e \quad |(G^2)'(x)| \geq 2 \quad e \quad |G''(x)/G'(x)^2| \leq 2.$$

Demonstração. Lembre que $G(x) = 1/x - m$ em cada intervalo $(1/(m+1), 1/m]$. Portanto

$$G'(x) = -\frac{1}{x^2} \quad \text{e} \quad G''(x) = \frac{2}{x^3}.$$

A primeira igualdade implica $|G'(x)| \geq 1$ para todo $x \in (0,1]$. Além disso $|G'(x)| \geq 2$ sempre que $x \leq 2/3$. Por outro lado, $x \geq 2/3$ implica que $G(x) = 1/x - 1 < 2/3$ e, por consequência, $G'(G(x)) \geq 2$. Combinando estas observações obtemos que $|(G^2)'(x)| = |G'(x)|\,|G'(G(x))| \geq 2$ para todo $x \in (0,1]$. Finalmente, $|G''(x)/G'(x)^2| = 2|x| \leq 2$ também para todo $x \in (0,1]$. □

Lema 3.2.13. *Existe uma constante $C > 1$ tal que para todo $k \geq 1$ e todo intervalo I_k tal que G^k restrita a I_k é uma bijeção diferenciável, tem-se*

$$\frac{|(G^k)'(x)|}{|(G^k)'(y)|} \leq C \quad \textit{para quaisquer } x \textit{ e } y \textit{ em } I_k.$$

Demonstração. Seja g uma inversa local de G, isto é, uma função diferenciável definida em algum intervalo e tal que $G(g(z)) = z$ para todo z no domínio de definição. Note que

$$\big[\log |G' \circ g(z)|\big]' = \frac{G''(g(z))\, g'(z)}{G'(g(z))} = \frac{G''(g(z))}{G'(g(z))^2}.$$

Portanto, a última estimativa no Lema 3.2.12 implica que

$$\big|\big[\log|G' \circ g(z)|\big]'\big| \le 2 \quad \text{para todo } g \text{ e todo } z. \tag{3.2.12}$$

Em outras palavras, toda função da forma $\log|G' \circ g|$ admite 2 como constante de Lipschitz. Observe também que se $x, y \in I_k$ então

$$\begin{aligned}\log\frac{|(G^k)'(x)|}{|(G^k)'(y)|} &= \sum_{j=0}^{k-1} \log|G'(G^j(x))| - \log|G'(G^j(y))| \\ &= \sum_{j=1}^{k} \log|G' \circ g_j(G^j(x))| - \log|G' \circ g_j(G^j(y))|\end{aligned}$$

onde g_j representa uma inversa local de G definida no intervalo $[G^j(x), G^j(y)]$. Usando a estimativa (3.2.12), obtemos que

$$\log\frac{|(G^k)'(x)|}{|(G^k)'(y)|} \le 2\sum_{j=1}^{k}|G^j(x) - G^j(y)| = 2\sum_{i=0}^{k-1}|G^{k-i}(x) - G^{k-i}(y)|. \tag{3.2.13}$$

Agora, as duas primeiras estimativas no Lema 3.2.12 implicam que

$$|G^k(x) - G^k(y)| \ge 2^{[i/2]}|G^{k-i}(x) - G^{k-i}(y)|$$

para todo $i = 0, \ldots, k$. Substituindo em (3.2.13), concluímos que

$$\log\frac{|(G^k)'(x)|}{|(G^k)'(y)|} \le 2\sum_{i=0}^{k-1} 2^{-[i/2]}|G^k(x) - G^k(y)| \le 8|G^k(x) - G^k(y)| \le 8.$$

Agora basta tomar $C = e^8$. □

Demonstração da Proposição 3.2.11. Seja m a medida de Lebesgue em $[0,1]$. O Lema 3.2.13 implica que

$$\frac{m(G^k(E_1))}{m(G^k(E_2))} = \frac{\int_{E_1}|(G^k)'|\,dm}{\int_{E_2}|(G^k)'|\,dm} \le C\frac{m(E_1)}{m(E_2)}.$$

Por outro lado, a definição (3.2.11) implica que

$$\frac{1}{2\log 2}m(E) \le \mu(E) \le \frac{1}{\log 2}m(E)$$

para todo conjunto mensurável $E \subset [0,1]$. Combinando estas duas relações, obtemos que

$$\frac{\mu(G^k(E_1))}{\mu(G^k(E_2))} \le 2\frac{m(G^k(E_1))}{m(G^k(E_2))} \le 2C\frac{m(E_1)}{m(E_2)} \le 4C\frac{\mu(E_1)}{\mu(E_2)}.$$

Assim, basta tomar $K = 4C$. □

Estamos prontos para concluir que (G, μ) é ergódica. Seja A um conjunto invariante por G com $\mu(A) > 0$. Então A também tem medida de Lebesgue positiva, uma vez que μ é absolutamente contínua com relação à medida de Lebesgue. Seja a um ponto de densidade de A cuja trajetória futura está contida no intervalo aberto $(0,1)$. Considere a sequência $(I_k)_k$ dos intervalos $I(k, m_1, \ldots, m_k)$ que contêm a. Segue do Lema 3.2.12 que

$$\operatorname{diam} I_k \le \sup \Big\{ \frac{1}{|(G^k)'(x)|} : x \in I_k \Big\} \le 2^{-[k/2]}$$

para todo $k \ge 1$. Em particular, o diâmetro de I_k converge para zero e, portanto,

$$\frac{\mu(I_k \cap A)}{\mu(I_k)} \to 1 \quad \text{quando } k \to \infty. \tag{3.2.14}$$

Tomemos $E_1 = I_k \cap A^c$ e $E_2 = I_k$. Pela Proposição 3.2.11,

$$\frac{\mu(G^k(I_k \cap A^c))}{\mu(G^k(I_k))} \le K \frac{\mu(I_k \cap A^c)}{\mu(I_k)}.$$

Observe que $G^k(I_k \cap A^c) = A^c$, a menos de um conjunto com medida nula, porque o conjunto A é invariante. Lembre também que $G^k(I_k) = (0,1)$, o qual tem medida total. Portanto, a desigualdade anterior pode ser escrita como

$$\mu(A^c) \le K \frac{\mu(I_k \cap A^c)}{\mu(I_k)} .$$

De acordo com (3.2.14), a expressão do lado direito converge para zero quando $k \to \infty$. Logo $\mu(A^c) = 0$, como queríamos demonstrar.

3.2.5 Endomorfismos lineares do toro

Lembre que chamamos toro de dimensão d ao quociente $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$, ou seja, o espaço das classes de equivalência da relação de equivalência definida em $\mathbb{R}^d$ por $x \sim e \Leftrightarrow x - y \in \mathbb{Z}^d$. Este quociente herda de $\mathbb{R}^d$ uma estrutura de variedade diferenciável de dimensão d. No que segue suporemos que $\mathbb{T}^d$ também está munido da métrica riemanniana plana, que o torna localmente isométrico ao espaço euclideano $\mathbb{R}^d$. Seja m a medida de Lebesgue associada a esta métrica riemanniana.

Seja A uma matriz d-por-d com coeficientes inteiros e determinante diferente de zero. Então $A(\mathbb{Z}^d) \subset \mathbb{Z}^d$ e, por consequência, A induz uma transformação

$$f_A : \mathbb{T}^d \to \mathbb{T}^d, \quad f_A([x]) = [A(x)]$$

onde $[x]$ denota a classe de equivalência que contém $x \in \mathbb{R}^d$. Chamamos tais transformações de *endomorfismos lineares* do toro. Note que f_A é diferenciável e a derivada $Df_A([x])$ em cada ponto está canonicamente identificada com A. Em particular, o jacobiano $\det Df_A([x])$ é constante igual a $\det A$. Isso também implica (Exercício 3.2.9) que o grau de f é igual a $|\det A|$. Portanto, f_A é invertível se, e somente se, $|\det A| = 1$. Neste caso, a sua inversa é a transformação $f_{A^{-1}}$ induzida pela matriz inversa A^{-1}; observe que A^{-1} também é uma matriz com coeficientes inteiros.

Em qualquer caso, f_A preserva a medida de Lebesgue em $\mathbb{T}^d$. Isto pode ser visto da seguinte forma. Como f_A é um difeomorfismo local, a pré-imagem de qualquer conjunto mensurável D com diâmetro suficientemente pequeno está formada por $|\det A| = \text{grau}(f_A)$ partes disjuntas D_i, cada uma das quais é enviada difeomorficamente sobre D. Pela fórmula de mudança de variável, $m(D) = |\det A|\, m(D_i)$ para todo i. Isto prova que $m(D) = m(f_A^{-1}(D))$ para todo conjunto mensurável D com diâmetro suficientemente pequeno. Logo, f_A preserva a medida m, tal como afirmamos. Agora vamos provar o seguinte fato:

Teorema 3.2.14. *O sistema (f_A, m) é ergódico se, e somente se, nenhum autovalor da matriz A é raiz da unidade.*

Demonstração. Suponha que nenhum autovalor de A é raiz da unidade. Considere qualquer função $\varphi \in L^2(m)$ e seja

$$\varphi([x]) = \sum_{k \in \mathbb{Z}^d} c_k e^{2\pi i (k \cdot x)}$$

a sua expansão em série de Fourier. Observe que $k \cdot x = k_1 x_1 + \cdots + k_d x_d$. Os coeficientes $c_k \in \mathbb{C}$ satisfazem

$$\sum_{k \in \mathbb{Z}^d} |c_k|^2 = \|\varphi\|_2^2 < \infty. \tag{3.2.15}$$

Então, a expansão em série de Fourier de $\varphi \circ f_A$ é:

$$\varphi(f_A([x])) = \sum_{k \in \mathbb{Z}^d} c_k e^{2\pi i (k \cdot A(x))} = \sum_{k \in \mathbb{Z}^d} c_k e^{2\pi i (A^*(k) \cdot x)},$$

onde A^* representa a adjunta de A. Suponha que φ é função invariante, isto é, $\varphi \circ f_A = \varphi$ em m-quase todo ponto. Então, por unicidade da expansão de Fourier, devemos ter

$$c_{A^*(k)} = c_k \quad \text{para todo } k \in \mathbb{Z}. \tag{3.2.16}$$

Afirmamos que a trajetória de todo $k \neq 0$ pela transformação A^* é infinita. De fato, se a trajetória de algum $k \neq 0$ fosse finita então deveriam existir $l, r \in \mathbb{Z}$ com $r > 0$ tais que $A^{(l+r)*}(k) = A^{l*}(k)$. Isto só poderia acontecer se A^* tivesse algum autovalor λ tal que $\lambda^r = 1$. Mas essa possibilidade está excluída, por hipótese, uma vez que A e A^* têm os mesmos autovalores. Logo, a trajetória de todo $k \neq 0$ é infinita, como afirmamos. Então a igualdade (3.2.16) juntamente com (3.2.15) implica que $c_k = 0$ para todo $k \neq 0$. Portanto, $\varphi = c_0$ em m-quase todo ponto. Isto prova a ergodicidade.

Para provar a recíproca, suponha que A admite algum autovalor que é uma raiz da unidade. Então o mesmo vale para A^* e, portanto, existe $r > 0$ tal que 1 é autovalor de A^{r*}. Como A^{r*} tem coeficientes inteiros, segue (veja o Exercício 3.2.8) que existe algum $k \in \mathbb{Z}^d \setminus \{0\}$ tal que $A^{r*}(k) = k$. Fixe k e considere a função $\varphi \in L^2(m)$ definida por

$$\varphi([x]) = \sum_{i=0}^{r-1} e^{2\pi i(A^{i*}(k)\cdot x)} = \sum_{i=0}^{r-1} e^{2\pi i(k\cdot A^i(x))}.$$

Então φ é uma função invariante por f_A mas não é constante em m-quase todo ponto. Logo, (f_A, m) não é ergódico. □

3.2.6 Argumento de Hopf

Nesta seção vamos apresentar outro método, mais geométrico, para demonstrar a ergodicidade de certos endomorfismos lineares do toro. Ele se aplica sempre que $|\det A| = 1$ e a matriz A é hiperbólica, ou seja, ela não tem autovalores de módulo 1. Mas a sua grande vantagem é que ele pode ser estendido a sistemas diferenciáveis muito mais gerais, não necessariamente lineares. A hipótese de que a matriz A é hiperbólica significa que o espaço $\mathbb{R}^d$ pode ser escrito como uma soma direta $\mathbb{R}^d = E^s \oplus E^u$ tal que:

1. $A(E^s) = E^s$ e todos os autovalores de $A \mid E^s$ têm módulo menor que 1;
2. $A(E^u) = E^u$ e todos os autovalores de $A \mid E^u$ têm módulo maior que 1.

Então existem constantes $C > 0$ e $\lambda < 1$ tais que

$$\begin{aligned} \|A^n(v^s)\| &\le C\lambda^n \|v^s\| \quad \text{para todo } v^s \in E^s \text{ e todo } n \ge 0, \\ \|A^{-n}(v^u)\| &\le C\lambda^n \|v^u\| \quad \text{para todo } v^u \in E^u \text{ e todo } n \ge 0. \end{aligned} \tag{3.2.17}$$

Exemplo 3.2.15. Considere $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. Os seus autovalores são

$$\lambda_u = \frac{3+\sqrt{5}}{2} > 1 > \lambda_s = \frac{3-\sqrt{5}}{2} > 0$$

e os respectivos autoespaços são:

$$E^u = \{(x,y) \in \mathbb{R}^2 : y = \frac{\sqrt{5}-1}{2}x\} \quad \text{e} \quad E^s = \{(x,y) \in \mathbb{R}^2 : y = -\frac{\sqrt{5}+1}{2}x\}.$$

A família de todos os subespaços afins de $\mathbb{R}^d$ da forma $v + E^s$, com $v \in \mathbb{R}^d$, define uma partição $\mathcal{F}^s$ de $\mathbb{R}^d$, que chamamos *folheação estável* e cujos elementos chamamos *folhas estáveis* de A. Ela é invariante por A, ou, seja, a imagem de qualquer folha estável é também uma folha estável. Além disso, pela propriedade (3.2.17), a transformação A contrai distâncias, uniformemente, dentro de cada folha. Analogamente, a família de todos os subespaços afins de $\mathbb{R}^d$ da forma $v + E^u$ com $v \in \mathbb{R}^d$ define uma partição $\mathcal{F}^u$ de $\mathbb{R}^d$, chamada *folheação instável.* Esta folheação também é invariante e a transformação A expande distâncias ao longo das suas folhas. Projetando $\mathcal{F}^s$ e $\mathcal{F}^u$ pela projeção canônica $\pi : \mathbb{R}^d \to \mathbb{T}^d$ obtemos folheações $\mathcal{W}^s$ e $\mathcal{W}^u$ do toro que chamamos *folheação estável* e *folheação instável* da transformação f_A. As observações anteriores mostram que estas folheações são invariantes por f_A. Além disso:

(a) $d(f_A^j(x), f_A^j(y)) \to 0$ quando $j \to +\infty$ para quaisquer pontos x e y em uma mesma folha estável;

(b) $d(f_A^j(y), f_A^j(z)) \to 0$ quando $j \to -\infty$ para quaisquer pontos y e z em uma mesma folha instável.

Vamos usar esta informação geométrica para provar que (f_A, m) é ergódica. Para isso, considere qualquer função contínua $\varphi : \mathbb{T}^d \to \mathbb{R}$ e considere as médias temporais

$$\varphi^+(x) = \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f_A^j(x)) \quad \text{e} \quad \varphi^-(x) = \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f_A^{-j}(x)),$$

definidas para m-quase todo $x \in \mathbb{T}^d$. Pelo Corolário 2.2.8, existe um conjunto $X \subset \mathbb{T}^d$ com medida total tal que

$$\varphi^+(x) = \varphi^-(x) \quad \text{para todo } x \in X. \tag{3.2.18}$$

Denotaremos por $\mathcal{W}^s(x)$ e $\mathcal{W}^u(x)$, respectivamente, a folha estável e a folha instável de f_A passando por cada ponto $x \in \mathbb{T}^d$.

Lema 3.2.16. *A função φ^+ é constante em toda folha de $\mathcal{W}^s$: se $\varphi^+(x)$ existe e $y \in \mathcal{W}^s(x)$ então $\varphi^+(y)$ existe e é igual a $\varphi^+(x)$. Analogamente, φ^- é constante em toda folha de $\mathcal{W}^u$.*

Demonstração. De acordo com a propriedade (a) acima, $d(f_A^j(x), f_A^j(y))$ converge para zero quando $j \to \infty$. Como φ é contínua (logo uniformemente contínua, uma vez que o domínio é compacto) isso implica que

$$\varphi(f_A^j(x)) - \varphi(f_A^j(y)) \to 0 \quad \text{quando } j \to \infty.$$

Por maioria de razão, o limite Cesàro

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f_A^j(x)) - \varphi(f_A^j(y))$$

também é zero. Isso implica $\varphi^+(y)$ existe e é igual a $\varphi^+(x)$. O argumento para φ^- é inteiramente análogo. □

Dado um subconjunto aberto R do toro e dado $x \in R$, denotamos por $\mathcal{W}^s(x, R)$ a componente conexa de $\mathcal{W}^s(x) \cap R$ que contém x e por $\mathcal{W}^u(x, R)$ a componente conexa de $\mathcal{W}^u(x) \cap R$ que contém x. Chamamos R de *retângulo* se $\mathcal{W}^s(x, R)$ intersecta $\mathcal{W}^u(y, R)$ num único ponto, para todo x e y em R.

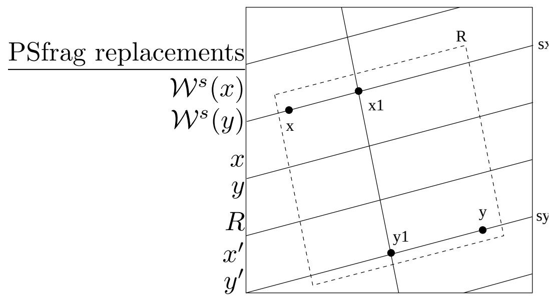

Figura 3.1: Retângulo em $\mathbb{T}^d$

Lema 3.2.17. *Dado qualquer retângulo $R \subset \mathbb{T}^d$, existe um conjunto mensurável $Y_R \subset X \cap R$ tal que $m(R \setminus Y_R) = 0$ e, dados quaisquer x e y em Y_R, existem pontos x' e y' em $X \cap R$ tais que $x' \in \mathcal{W}^s(x, R)$ e $y' \in \mathcal{W}^s(y, R)$ e $y' \in \mathcal{W}^u(x')$.*

Demonstração. Representemos por m_x^s a medida de Lebesgue na folha estável $\mathcal{W}^s(x)$ de cada ponto $x \in \mathbb{T}^d$. Note que $m(R \setminus X) = 0$, uma vez que X tem medida total em $\mathbb{T}^d$. Então, usando o teorema de Fubini,

$$m_x^s\big(\mathcal{W}^s(x, R) \setminus X\big) = 0 \quad \text{para } m\text{-quase todo } x \in R.$$

Defina $Y_R = \big\{x \in X \cap R : m_x^s\big(\mathcal{W}^s(x, R) \setminus X\big) = 0\big\}$. Então Y_R tem medida total em R. Dados $x, y \in R$ considere a aplicação

$$\pi : \mathcal{W}^s(x, R) \to \mathcal{W}^s(y, R), \quad \pi(x') = \text{interseção entre } \mathcal{W}^u(x', R) \text{ e } \mathcal{W}^s(y, R).$$

Esta aplicação é afim e, portanto, tem a seguinte propriedade, que chamamos *continuidade absoluta*:

$$m_x^s(E) = 0 \quad \Leftrightarrow \quad m_y^s(\pi(E)) = 0.$$

Em particular, a imagem de $\mathcal{W}^s(x, R) \cap X$ tem medida total em $\mathcal{W}^s(y, R)$ e, consequentemente, ela intersecta $\mathcal{W}^s(y, R) \cap X$. Em outras palavras, existe $x' \in \mathcal{W}^s(x, R) \cap X$ cuja imagem $y' = \pi(x')$ está em $\mathcal{W}^s(y, R) \cap X$. Observando que x' e y' estão na mesma folha instável, pela definição da π, vemos que estes pontos satisfazem as condições na conclusão do lema. $\square$

Considere um retângulo R qualquer. Dados quaisquer x, y em Y_R, considere os pontos x', y' em X dados pelo Lema 3.2.17. Usando também o Lema 3.2.16, obtemos:

$$\varphi^-(x) = \varphi^+(x) = \varphi^+(x') = \varphi^-(x') = \varphi^-(y') = \varphi^+(y') = \varphi^+(y) = \varphi^-(y).$$

Isto mostra que as funções φ^+ e φ^- coincidem uma com a outra e são constantes em Y_R. Agora seja $R_1, \ldots, R_N$ uma cobertura finita do toro por retângulos.

Considere o conjunto

$$Y = \bigcup_{j=1}^{N} Y_j, \qquad \text{onde } Y_j = Y_{R_j}.$$

Observe que $m(Y) = 1$, uma vez que $Y \cap R_j \supset Y_j$ tem medida total em R_j para todo j. Afirmamos que $\varphi^+ = \varphi^-$ é constante em todo o Y. De fato, dados quaisquer $k, l \in \{1, \dots, N\}$ podemos encontrar $j_0 = k, j_1, \dots, j_{n-1}, j_n = l$ tais que cada R_{j_i} intersecta $R_{j_{i-1}}$ (isto é uma simples consequência da conexidade por arcos do toro). Lembrando que R_j é aberto e Y_j é um subconjunto de medida total, obtemos que cada Y_{j_i} intersecta $Y_{j_{i-1}}$. Então, $\varphi^+ = \varphi^-$ é constante na união de todos os Y_{j_i}. Isto prova a nossa afirmação.

Desta forma, mostramos que as médias temporais $\varphi^{\pm}$ de qualquer função contínua φ são constantes em m-quase todo ponto. Consequentemente (veja o Exercício 3.1.4), o sistema (f_A, m) é ergódico.

3.2.7 Exercícios

3.2.1. Prove a Proposição 3.2.2.

3.2.2. Prove a Proposição 3.2.9.

3.2.3. Seja $I = [0, 1]$ e $f : I \to I$ a função definida por

$$f(x) = \begin{cases} 2x & \text{se } 0 \le x < 1/3 \\ 2x - 2/3 & \text{se } 1/3 \le x < 1/2 \\ 2x - 1/3 & \text{se } 1/2 \le x < 2/3 \\ 2x - 1 & \text{se } 2/3 \le x \le 1. \end{cases}$$

Mostre que f é ergódica relativamente à medida de Lebesgue m.

3.2.4. Seja X um conjunto finito e $\Sigma = X^{\mathbb{N}}$. Prove que todo subconjunto infinito compacto de Σ invariante pelo deslocamento $\sigma : \Sigma \to \Sigma$ contém algum ponto não-periódico.

3.2.5. Seja X um espaço topológico, munido da sua σ-álgebra de Borel $\mathcal{C}$, e seja $\Sigma = X^{\mathbb{N}}$. Mostre que se X tem base enumerável de abertos então a σ-álgebra de Borel de Σ (para a topologia produto) coincide com a σ-álgebra produto $\mathcal{B} = \mathcal{C}^{\mathbb{N}}$. O mesmo vale para $\Sigma = X^{\mathbb{Z}}$ e $\mathcal{B} = \mathcal{C}^{\mathbb{Z}}$.

3.2.6. Neste exercício propomos outra demonstração para a Proposição 3.2.1. Suponha que θ é irracional. Seja A um conjunto invariante com medida positiva. Lembrando que a órbita $\{R_\theta^n(a) : n \in \mathbb{Z}\}$ de todo $a \in S^1$ é densa em S^1, mostre que *nenhum* ponto de S^1 é ponto de densidade de A^c. Conclua que $\mu(A) = 1$.

3.2.7. Suponha que θ é irracional. Seja $\varphi : S^1 \to \mathbb{R}$ uma função contínua qualquer. Mostre que

$$\tilde{\varphi}(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(R_\theta^j(x)) \tag{3.2.19}$$

existe em *todo* ponto e, de fato, o limite é uniforme. Conclua que $\tilde{\varphi}$ é constante em todo ponto. Deduza que R_θ tem uma única probabilidade invariante.

3.2.8. Seja A uma matriz quadrada de dimensão d com coeficientes racionais e seja λ um autovalor racional. Mostre que existe algum autovetor com coeficientes inteiros, ou seja, algum $k \in \mathbb{Z}^d \setminus \{0\}$ tal que $Ak = \lambda k$.

3.2.9. Mostre que se $f : M \to M$ é um difeomorfismo local numa variedade riemanniana compacta então

$$\operatorname{grau}(f) = \int |\det Df| \, dm,$$

onde m representa a medida de volume induzida pela métrica riemanniana de M, normalizada de tal modo que $m(M) = 1$. Em particular, o grau do endomorfismo linear $f_A : \mathbb{T}^d \to \mathbb{T}^d$ associado a uma matriz A de dimensão d com coeficientes inteiros é igual a $|\det A|$.

3.2.10. Um número $x \in (0,1)$ tem expansão em fração contínua *de tipo limitado* se a sequência $(a_n)_n$ construída na Seção 1.3.2 é limitada. Prove que o conjunto $\mathcal{L} \subset (0,1)$ dos pontos com expansão em fração contínua de tipo limitado tem medida de Lebesgue zero.

3.2.11. Seja $f : M \to M$ uma transformação mensurável, μ uma medida invariante ergódica e $\varphi : M \to \mathbb{R}$ uma função tal que $\int \varphi \, d\mu = +\infty$. Prove que $\lim_n (1/n) \sum_{j=0}^{n-1} \varphi(f^j(x)) = +\infty$ para μ-quase todo $x \in M$.

3.2.12. Observe que o número b no Exercício 2.3.9 é independente de x num conjunto com medida de Lebesgue total. Prove que a média aritmética dos números $a_1, \ldots, a_n, \ldots$ vai para infinito: $\lim_n (1/n)(a_1 + \cdots + a_n) = +\infty$.

Capítulo 4

Entropia

A palavra *entropia* foi inventada em 1865 pelo físico e matemático alemão Rudolf Clausius, um dos pioneiros fundadores da Termodinâmica. Na teoria dos sistemas termodinâmicos em equilíbrio, a entropia é uma medida do grau de "desordem" do sistema. A segunda lei da Termodinâmica afirma que, quando um sistema isolado passa de um estado de equilíbrio a outro, a entropia do estado final é necessariamente maior do que a entropia do estado inicial. Por exemplo, quando juntamos dois recipientes contendo gases distintos, digamos oxigênio e nitrogênio, os dois gases se misturam até alcançar um novo equilíbrio macroscópico no qual ambos se encontram uniformemente distribuídos no conjunto dos dois recipientes. A entropia deste novo estado é superior à entropia do equilíbrio inicial, no qual os dois gases estavam separados.

Esta noção desempenha um papel de destaque em diversas outras áreas do conhecimento. Um exemplo importante, que iremos explorar na nossa apresentação, é a Teoria da Informação, desenvolvida a partir dos trabalhos do engenheiro americano Claude Shannon em meados do século 20. Mais ou menos ao mesmo tempo, os matemáticos russos Andrey Kolmogorov e Yakov Sinai estavam propondo uma definição de entropia de um sistema em Teoria Ergódica. O principal objetivo era fornecer um invariante de equivalência ergódica que, em particular, permitisse distinguir dois deslocamentos de Bernoulli. Esta noção é o tema do presente capítulo.

Na Seção 4.1 definiremos a entropia de uma transformação relativamente a uma probabilidade invariante, a partir de uma analogia com a Teoria da Informação. O teorema de Kolmogorov-Sinai, que discutiremos na Seção 4.2, constitui uma ferramenta fundamental para o cálculo da entropia de sistemas específicos. Na Seção 4.3 analisaremos a entropia de um ponto de vista mais local, que se relaciona diretamente com a formulação de Shannon. Em seguida, na Seção 4.4, ilustraremos alguns métodos de cálculo da entropia por meio de exemplos concretos.

4.1 Definição de entropia

Para motivar a definição de entropia de Kolmogorov-Sinai, vamos considerar a seguinte situação básica da Teoria da Informação. Consideremos um canal de comunicação que transmite, sucessivamente, certos símbolos. Esse canal pode ser um telégrafo transmitindo pontos e traços, segundo o antigo código Morse, uma fibra ótica, transmitindo zeros e uns, segundo o código binário ASCII, ou qualquer outro sistema de transmissão sequencial de informação. O objetivo é medir a *entropia* do canal, ou seja, a quantidade de informação transmitida, em média, a cada unidade de tempo.

4.1.1 Entropia em Teoria da Informação

Para formalizar esta ideia, suponhamos que os símbolos transmitidos pelo canal pertencem a um certo alfabeto $\mathcal{A}$ previamente definido. Nem todos os caracteres deste alfabeto têm a mesma frequência, ou seja, a mesma probabilidade de serem utilizados. Por exemplo, se o canal está transmitindo mensagens na língua portuguesa a letra A será utilizada com muito maior probabilidade que a letra Z. Portanto, nem todos os caracteres carregam a mesma quantidade de informação: quanto mais improvável é um caracter, menor é o número de palavras que o contêm e, portanto, mais informação está associada a esse caracter. Analogamente, quanto mais improvável for uma palavra, menor é o número de frases em que ela participa e, portanto, maior é a quantidade de informação associada a essa palavra.

Convém observar que quantidade de informação associada a cada caracter, ou a cada palavra, depende dos demais caracteres ou palavras. Por exemplo, se o canal está transmitindo em língua portuguesa e gera, sucessivamente, os caracteres I, N, V, A, R, I, A, N e T então o caracter seguinte deverá ser um E; neste caso, em vista dos caracteres transmitidos anteriormente, esta letra E não carrega informação adicional.[1]

Por outro lado, Se os caracteres transmitidos sucessivamente são independentes uns dos outros então a informação de cada um se soma à informação anterior. Por exemplo, se a transmissão reflete os resultados de lançamentos sucessivos de uma moeda justa, a informação correspondente ao resultado ($Cara$, $Coroa$, $Coroa$) deve ser igual à *soma* das informações correspondentes a cada um dos caracteres $Cara$, $Coroa$ e $Coroa$. Ora, por independência, a probabilidade do evento ($Cara$, $Coroa$, $Coroa$) é o *produto* das probabilidades dos eventos $Cara$, $Coroa$ e $Coroa$.

Isto sugere que a informação deve ser definida em termos do *logaritmo* da probabilidade. Em Teoria da Informação é usual considerar logaritmos na base 2, porque essencialmente todos os canais de informação que encontramos na

[1] Um dos autores deste livro participou uma vez numa "caçada ao tesouro"que consistia em buscar as diferentes letras do nome de um certo objeto matemático. Aconteceu que as três primeiras letras encontradas foram Z, Z e Z. Essa circunstância infeliz arruinou a continuação do jogo, pois as demais letras não acrescentariam qualquer informação: só existe um objeto objeto matemático cujo nome inclui três vezes a letra Z (o *puzzle de Yoccoz*).

prática são binários. No entanto, em Teoria Ergódica é mais comum considerar logaritmos naturais (base e), e nós faremos o mesmo.

Por definição, a *quantidade de informação* associada a um caracter $a \in \mathcal{A}$ está dada por

$$I(a) = -\log p_a \tag{4.1.1}$$

onde p_a é a probabilidade (frequência) do caracter a. A *informação média* associada ao alfabeto $\mathcal{A}$ é dada por

$$I(\mathcal{A}) = \sum_{a \in \mathcal{A}} p_a I(a) = \sum_{a \in \mathcal{A}} -p_a \log p_a. \tag{4.1.2}$$

Mais geralmente, a informação associada a uma palavra $a_1 \dots a_n$ é

$$I(a_1 \dots a_n) = -\log p_{a_1 \dots a_n} \tag{4.1.3}$$

onde $p_{a_1 \dots a_n}$ representa a probabilidade da palavra. No caso independente ela coincide com o produto $p_{a_1} \dots p_{a_n}$ das probabilidades de cada uma das letras, mas em geral os dois números são distintos. Denotando por $\mathcal{A}^n$ o conjunto de todas as palavras de comprimento n, definimos

$$I(\mathcal{A}^n) = \sum_{a_1, \dots, a_n} p_{a_1 \dots a_n} I(a_1, \dots, a_n) = \sum_{a_1, \dots, a_n} -p_{a_1 \dots a_n} \log p_{a_1 \dots a_n}. \tag{4.1.4}$$

Finalmente, a *entropia* do canal de comunicação é definida por:

$$I = \lim_n \frac{1}{n} I(\mathcal{A}^n). \tag{4.1.5}$$

Convidamos o leitor a verificar que a sequência $I(\mathcal{A}^n)$ é subaditiva e, portanto, o limite em (4.1.5) existe. Isso também está contido na teoria, muito mais geral, que desenvolveremos a seguir.

4.1.2 Entropia de uma partição

Queremos adaptar estas ideias ao nosso contexto em Teoria Ergódica. A principal diferença é que, enquanto em Teoria da Informação o alfabeto $\mathcal{A}$ é discreto (finito), em geral, esse não é necessariamente o caso para o espaço de estados da maioria dos sistemas dinâmicos interessantes. Esse ponto é resolvido fazendo uso de partições do espaço de estados.

Seja $(M, \mathcal{B}, \mu)$ um espaço de probabilidade. Neste capítulo, por *partição* sempre entenderemos uma família **enumerável** (finita ou infinita) $\mathcal{P}$ de subconjuntos mensuráveis de M disjuntos dois-a-dois e cuja união tem medida total. Denotamos por $\mathcal{P}(x)$ o elemento da partição que contém um dado ponto x. A *soma* $\mathcal{P} \vee \mathcal{Q}$ de duas partições $\mathcal{P}$ e $\mathcal{Q}$ é a partição cujos elementos são as interseções $P \cap Q$ com $P \in \mathcal{P}$ e $Q \in \mathcal{Q}$. Mais geralmente, dada qualquer família enumerável de partições $\mathcal{P}_n$, definimos

$$\bigvee_n \mathcal{P}_n = \big\{ \bigcap_n P_n : P_n \in \mathcal{P}_n \text{ para cada } n \big\}.$$

A cada partição $\mathcal{P}$ associamos a respectiva *função de informação*

$$I_{\mathcal{P}} : M \to \mathbb{R}, \quad I_{\mathcal{P}}(x) = -\log \mu(\mathcal{P}(x)). \tag{4.1.6}$$

É claro que a função $I_{\mathcal{P}}$ é mensurável. Então chamamos *entropia*, ou *informação média*, da partição $\mathcal{P}$ ao número

$$H_\mu(\mathcal{P}) = \int I_{\mathcal{P}}\, d\mu = \sum_{P \in \mathcal{P}} -\mu(P) \log \mu(P). \tag{4.1.7}$$

Como é usual na teoria da integral de Lebesgue, fazemos a convenção de que $0 \log 0 = \lim_{x\to 0} x \log x = 0$. Veja a Figura 4.1.

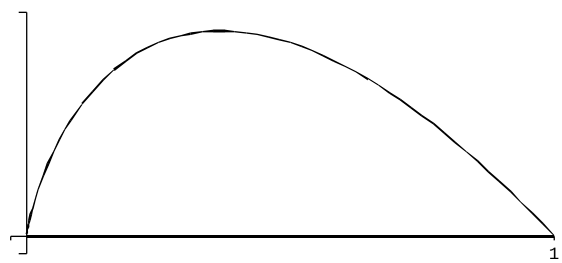

Figura 4.1: Gráfico da função $\phi(x) = -x \log x$

Relacionado com isto, a seguinte observação será útil em diversas ocasiões. Considere a função $\phi : (0, \infty) \to \mathbb{R}$ dada por $\phi(x) = -x \log x$. Derivando duas vezes vemos que $\phi'' < 0$. Portanto ϕ é côncava:

$$t_1\phi(x_1) + \cdots + t_k\phi(x_k) \le \phi(t_1x_1 + \cdots + t_kx_k) \tag{4.1.8}$$

para todo $x_1, \ldots, x_k > 0$ e $t_1, \ldots, t_k > 0$ com $t_1 + \cdots + t_k = 1$. Além disso, a concavidade é estrita: vale a igualdade em (4.1.8) se, e somente se, $x_1 = \cdots = x_k$.

Dizemos que duas partições $\mathcal{P}$ e $\mathcal{Q}$ são *independentes* se $\mu(P \cap Q) = \mu(P)\mu(Q)$ para todo $P \in \mathcal{P}$ e todo $Q \in \mathcal{Q}$. Nesse caso, $I_{\mathcal{P}\vee\mathcal{Q}} = I_{\mathcal{P}} + I_{\mathcal{Q}}$ e, portanto, $H_\mu(\mathcal{P} \vee \mathcal{Q}) = H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q})$. Em geral, vale a desigualdade $\le$ como veremos daqui a pouco.

Exemplo 4.1.1. Considere $M = [0, 1]$ munido da medida de Lebesgue. Para cada $n \ge 1$ considere a partição $\mathcal{P}^n$ nos subintervalos $\big((i-1)/10^n, i/10^n\big]$ com $1 \le i \le 10^n$. Então

$$H_\mu(\mathcal{P}^n) = \sum_{i=1}^{10^n} -10^{-n} \log 10^{-n} = n \log 10.$$

Exemplo 4.1.2. Seja $M = \{1, \ldots, d\}^{\mathbb{N}}$ munido de uma medida produto $\mu = \nu^{\mathbb{N}}$. Denotamos $p_i = \nu(\{i\})$ para cada $i \in \{1, \ldots, d\}$. Para cada $n \geq 1$, seja $\mathcal{P}^n$ a partição de M em cilindros $[0; a_1, \ldots, a_n]$ de comprimento n. A entropia de $\mathcal{P}^n$ é

$$\begin{aligned} H_\mu(\mathcal{P}^n) &= \sum_{a_1,\ldots,a_n} -p_{a_1} \ldots p_{a_n} \log(p_{a_1} \ldots p_{a_n}) \\ &= \sum_j \sum_{a_1,\ldots,a_n} -p_{a_1} \ldots p_{a_j} \ldots p_{a_n} \log p_{a_j} \\ &= \sum_j \sum_{a_j} -p_{a_j} \log p_{a_j} \sum_{a_i, i \neq j} p_{a_1} \ldots p_{a_{j-1}} p_{a_{j+1}} \ldots p_{a_n}. \end{aligned}$$

A última soma é igual a 1, uma vez que $\sum_i p_i = 1$. Portanto,

$$H_\mu(\mathcal{P}^n) = \sum_{j=1}^{n} \sum_{a_j=1}^{d} -p_{a_j} \log p_{a_j} = \sum_{j=1}^{n} \sum_{i=1}^{d} -p_i \log p_i = -n \sum_{i=1}^{d} p_i \log p_i.$$

Lema 4.1.3. *Toda partição finita tem entropia finita. De fato, $H_\mu(\mathcal{P}) \leq \log \#\mathcal{P}$ e vale a igualdade se, e somente se, $\mu(P) = 1/\#\mathcal{P}$ para todo $P \in \mathcal{P}$.*

Demonstração. Seja $\mathcal{P} = \{P_1, P_2, \ldots, P_n\}$ e considere os números $t_i = 1/n$ e $x_i = \mu(P_i)$. Pela propriedade de concavidade (4.1.8):

$$\frac{1}{n} H_\mu(\mathcal{P}) = \sum_{i=1}^{n} t_i \phi(x_i) \leq \phi\Big(\sum_{i=1}^{n} t_i x_i\Big) = \phi\Big(\frac{1}{n}\Big) = \frac{\log n}{n}.$$

Portanto, $H_\mu(\mathcal{P}) \leq \log n$. Além disso, a igualdade ocorre se, e somente se, $\mu(P_i) = 1/n$ para todo $i = 1, \ldots, n$. □

Exemplo 4.1.4. Considere $M = [0,1]$ munido da medida de Lebesgue μ. Observe que a série $\sum_{k=1}^{\infty} 1/(k(\log k)^2)$ é convergente. Seja c o valor da soma. Então podemos decompor $[0,1]$ em intervalos P_k com $\mu(P_k) = 1/(ck(\log k)^2)$ para todo k. Seja $\mathcal{P}$ a partição formada por estes subintervalos. Então,

$$H_\mu(\mathcal{P}) = \sum_{k=1}^{\infty} \frac{\log c + \log k + 2 \log \log k}{ck(\log k)^2}.$$

Pelo critério da razão, a série do lado direito tem o mesmo comportamento que a série $\sum_{k=1}^{\infty} 1/(k \log k)$ que, como sabemos, é divergente (use o critério da integral). Portanto, $H_\mu(\mathcal{P}) = \infty$.

Este exemplo mostra que partições infinitas podem ter entropia infinita. A partir daqui, ao longo de todo o capítulo, **sempre consideraremos partições** (enumeráveis) **com entropia finita**.

Chamamos *entropia condicional* de uma partição $\mathcal{P}$ com relação a uma partição $\mathcal{Q}$ ao número

$$H_\mu(\mathcal{P}/\mathcal{Q}) = \sum_{P \in \mathcal{P}} \sum_{Q \in \mathcal{Q}} -\mu(P \cap Q) \log \frac{\mu(P \cap Q)}{\mu(Q)}. \tag{4.1.9}$$

Intuitivamente, ele mede a informação adicional fornecida pela partição $\mathcal{P}$ uma vez conhecida a informação da partição $\mathcal{Q}$. É claro que $H_\mu(\mathcal{P}/\mathcal{M}) = H_\mu(\mathcal{P})$ para todo $\mathcal{P}$, onde $\mathcal{M}$ denota a partição trivial $\mathcal{M} = \{M\}$. Além disso, se $\mathcal{P}$ e $\mathcal{Q}$ são independentes então $H_\mu(\mathcal{P}/\mathcal{Q}) = H_\mu(\mathcal{P})$. Em geral, vale a desigualdade $\leq$ como veremos num instante.

Dadas duas partições $\mathcal{P}$ e $\mathcal{Q}$, dizemos que $\mathcal{P}$ é *menos fina* que $\mathcal{Q}$, e escrevemos $\mathcal{P} \prec \mathcal{Q}$, se todo elemento de $\mathcal{Q}$ está contido em algum elemento de $\mathcal{P}$, a menos de medida nula. A soma $\mathcal{P} \vee \mathcal{Q}$ é, precisamente, a menos fina de todas as partições $\mathcal{R}$ tais que $\mathcal{P} \prec \mathcal{R}$ e $\mathcal{Q} \prec \mathcal{R}$.

Lema 4.1.5. *Sejam $\mathcal{P}$, $\mathcal{Q}$ e $\mathcal{R}$ partições com entropia finita. Então,*

(a) $H_\mu(\mathcal{P} \vee \mathcal{Q}/\mathcal{R}) = H_\mu(\mathcal{P}/\mathcal{R}) + H_\mu(\mathcal{Q}/\mathcal{P} \vee \mathcal{R})$;

(b) se $\mathcal{P} \prec \mathcal{Q}$ então $H_\mu(\mathcal{P}/\mathcal{R}) \leq H_\mu(\mathcal{Q}/\mathcal{R})$ e $H_\mu(\mathcal{R}/\mathcal{P}) \geq H_\mu(\mathcal{R}/\mathcal{Q})$;

(c) $\mathcal{P} \prec \mathcal{Q}$ se, e somente se, $H_\mu(\mathcal{P}/\mathcal{Q}) = 0$.

Demonstração. Por definição,

$$\begin{aligned} H_\mu(\mathcal{P} \vee \mathcal{Q}/\mathcal{R}) &= \sum_{P,Q,R} -\mu(P \cap Q \cap R) \log \frac{\mu(P \cap Q \cap R)}{\mu(R)} \\ &= \sum_{P,Q,R} -\mu(P \cap Q \cap R) \log \frac{\mu(P \cap Q \cap R)}{\mu(P \cap R)} \\ &\qquad + \sum_{P,Q,R} -\mu(P \cap Q \cap R) \log \frac{\mu(P \cap R)}{\mu(R)}. \end{aligned}$$

A soma do lado direito pode ser reescrita como

$$\begin{aligned} \sum_{S \in \mathcal{P} \vee \mathcal{R}, Q \in \mathcal{Q}} -\mu(S \cap Q) \log \frac{\mu(S \cap Q)}{\mu(S)} &+ \sum_{P \in \mathcal{P}, R \in \mathcal{R}} -\mu(P \cap R) \log \frac{\mu(P \cap R)}{\mu(R)} \\ &= H_\mu(\mathcal{Q}/\mathcal{P} \vee \mathcal{R}) + H_\mu(\mathcal{P}/\mathcal{R}). \end{aligned}$$

Isto demonstra o item (a). Agora observe que se $\mathcal{P} \prec \mathcal{Q}$ então

$$\begin{aligned} H_\mu(\mathcal{P}/\mathcal{R}) &= \sum_P \sum_R \sum_{Q \subset P} -\mu(Q \cap R) \log \frac{\mu(P \cap R)}{\mu(R)} \\ &\leq \sum_P \sum_R \sum_{Q \subset P} -\mu(Q \cap R) \log \frac{\mu(Q \cap R)}{\mu(R)} = H_\mu(\mathcal{Q}/\mathcal{R}). \end{aligned}$$

Isto prova a primeira parte do item (b). Para provar a segunda parte, note que para quaisquer $P \in \mathcal{P}$ e $R \in \mathcal{R}$, tem-se

$$\frac{\mu(R \cap P)}{\mu(P)} = \sum_{Q \subset P} \frac{\mu(Q)}{\mu(P)} \frac{\mu(R \cap Q)}{\mu(Q)}.$$

Claro que $\sum_{Q\subset P} \mu(Q)/\mu(P) = 1$. Então, pela propriedade (4.1.8),

$$\phi\big(\frac{\mu(R\cap P)}{\mu(P)}\big) \geq \sum_{Q\subset P} \frac{\mu(Q)}{\mu(P)}\phi\big(\frac{\mu(R\cap Q)}{\mu(Q)}\big)$$

para todo $P \in \mathcal{P}$ e $R \in \mathcal{R}$. Consequentemente,

$$\begin{aligned} H_\mu(\mathcal{R}/\mathcal{P}) &= \sum_{P,R} \mu(P)\phi\big(\frac{\mu(R\cap P)}{\mu(P)}\big) \geq \sum_{P,R} \mu(P) \sum_{Q\subset P} \frac{\mu(Q)}{\mu(P)}\phi\big(\frac{\mu(R\cap Q)}{\mu(Q)}\big) \\ &= \sum_{Q,R} \mu(Q)\phi\big(\frac{\mu(R\cap Q)}{\mu(Q)}\big) = H_\mu(\mathcal{R}/\mathcal{Q}). \end{aligned}$$

Finalmente, segue da definição em (4.1.9) que $H_\mu(\mathcal{P}/\mathcal{Q}) = 0$ se, e somente se, para todo $P \in \mathcal{P}$ e todo $Q \in \mathcal{Q}$,

$$\mu(P\cap Q) = 0 \quad \text{ou então} \quad \frac{\mu(P\cap Q)}{\mu(Q)} = 1.$$

Em outras palavras, ou Q é disjunto de P (a menos de medida nula) ou Q está contido em P (a menos de medida nula). Isto quer dizer que $H_\mu(\mathcal{P}/\mathcal{Q}) = 0$ se, e somente se, $\mathcal{P} \prec \mathcal{Q}$. □

Em particular, tomando $\mathcal{P} = \mathcal{M}$ no item (b) do lema obtemos que

$$H_\mu(\mathcal{R}/\mathcal{Q}) \leq H_\mu(\mathcal{R}) \quad \text{para quaisquer partições } \mathcal{R} \text{ e } \mathcal{Q}. \tag{4.1.10}$$

Além disso, tomando $\mathcal{R} = \mathcal{M}$ no item (a), vem que

$$H_\mu(\mathcal{P}\vee\mathcal{Q}) = H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q}/\mathcal{P}) \leq H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q}). \tag{4.1.11}$$

Seja $f : M \to N$ uma transformação mensurável e seja μ uma probabilidade em M. Então $f_*\mu$ é uma probabilidade em N. Além disso, se $\mathcal{P}$ é uma partição de N então $f^{-1}(\mathcal{P}) = \{f^{-1}(P) : P \in \mathcal{P}\}$ é uma partição de M. Por definição,

$$\begin{aligned} H_\mu(f^{-1}(\mathcal{P})) &= \sum_{P\in\mathcal{P}} -\mu(f^{-1}(P)) \log \mu(f^{-1}(P)) \\ &= \sum_{P\in\mathcal{P}} -f_*\mu(P) \log f_*\mu(P) = H_{f^*\mu}(\mathcal{P}). \end{aligned} \tag{4.1.12}$$

Em particular, se $M = N$ e a medida μ é invariante por f então

$$H_\mu(f^{-1}(\mathcal{P})) = H_\mu(\mathcal{P}) \quad \text{para toda partição } \mathcal{P}. \tag{4.1.13}$$

Também precisaremos da seguinte propriedade de continuidade:

Lema 4.1.6. *Dado $k \geq 1$ e $\varepsilon > 0$ existe $\delta > 0$ tal que, para quaisquer partições finitas $\mathcal{P} = \{P_1, \dots, P_k\}$ e $\mathcal{Q} = \{Q_1, \dots, Q_k\}$,*

$$\mu(P_i \Delta Q_i) < \delta \text{ para todo } i = 1, \dots, k \quad \Rightarrow \quad H_\mu(\mathcal{Q}/\mathcal{P}) < \varepsilon.$$

Demonstração. Fixe $\varepsilon > 0$ e $k \geq 1$. Pela continuidade da função $\phi : [0,1] \to \mathbb{R}$, $\phi(x) = -x \log x$, existe $\rho > 0$ tal que $\phi(x) < \varepsilon/k^2$ para todo $x \in [0,\rho) \cup (1-\rho, 1]$. Tome $\delta = \rho/k$. Dadas partições $\mathcal{P}$ e $\mathcal{Q}$ como no enunciado, denote por $\mathcal{R}$ a partição cujos elementos são as interseções $P_i \cap Q_j$ com $i \neq j$ e também o conjunto $\cup_{i=1}^k P_i \cap Q_i$. Note que $\mu(P_i \cap Q_j) \leq \mu(P_i \Delta Q_i) < \delta$ para todo $i \neq j$ e

$$\mu\big(\bigcup_{i=1}^k P_i \cap Q_i\big) \geq \sum_{i=1}^k \big(\mu(P_i) - \mu(P_i \Delta Q_i)\big) > \sum_{i=1}^k \big(\mu(P_i) - \delta\big) = 1 - \rho.$$

Portanto,

$$H_\mu(\mathcal{R}) = \sum_{R \in \mathcal{R}} \phi(\mu(R)) < \#\mathcal{R} \frac{\varepsilon}{k^2} \leq \varepsilon.$$

É claro da definição que $\mathcal{P} \vee \mathcal{Q} = \mathcal{P} \vee \mathcal{R}$. Então, usando (4.1.11) e (4.1.10),

$$\begin{aligned} H_\mu(\mathcal{Q}/\mathcal{P}) &= H_\mu(\mathcal{P} \vee \mathcal{Q}) - H_\mu(\mathcal{P}) = H_\mu(\mathcal{P} \vee \mathcal{R}) - H_\mu(\mathcal{P}) \\ &= H_\mu(\mathcal{R}/\mathcal{P}) \leq H_\mu(\mathcal{R}) < \varepsilon. \end{aligned}$$

Isto prova o lema. □

4.1.3 Entropia de um sistema dinâmico

Seja $f : M \to M$ uma transformação mensurável preservando uma medida de probabilidade μ. A noção de entropia do sistema (f, μ), apresentada a seguir, é inspirada pela ideia de entropia de um canal de comunicação definida por (4.1.5).

Dada uma partição $\mathcal{P}$ de M com entropia finita, denotamos

$$\mathcal{P}^n = \bigvee_{i=0}^{n-1} f^{-i}(\mathcal{P}) \quad \text{para cada } n \geq 1.$$

Observe que o elemento $\mathcal{P}^n(x)$ que contém $x \in M$ está dado por:

$$\mathcal{P}^n(x) = \mathcal{P}(x) \cap f^{-1}(\mathcal{P}(f(x))) \cap \cdots \cap f^{-n+1}(\mathcal{P}(f^{n-1}(x))).$$

É claro que a sequência $\mathcal{P}^n$ é não-decrescente, ou seja, $\mathcal{P}^n \prec \mathcal{P}^{n+1}$ para todo n. Portanto, a sequência das entropias $H_\mu(\mathcal{P}^n)$ também é não-decrescente. Outro fato importante é que esta sequência é subaditiva:

Lema 4.1.7. *$H_\mu(\mathcal{P}^{m+n}) \leq H_\mu(\mathcal{P}^m) + H_\mu(\mathcal{P}^n)$ para todo $m, n \geq 1$.*

Demonstração. Por definição, $\mathcal{P}^{m+n} = \vee_{i=0}^{m+n-1} f^{-i}(\mathcal{P}) = \mathcal{P}^m \vee f^{-m}(\mathcal{P}^n)$. Portanto, usando (4.1.11),

$$H_\mu(\mathcal{P}^{m+n}) \leq H_\mu(\mathcal{P}^m) + H_\mu(f^{-m}(\mathcal{P}^n)). \tag{4.1.14}$$

Por outro lado, como a medida μ é invariante por f, a propriedade (4.1.13) implica que $H_\mu(f^{-m}(\mathcal{P}^n)) = H_\mu(\mathcal{P}^n)$ para todo m, n. Substituindo este fato em (4.1.14) obtemos a conclusão do lema. □

Como a sequência $H_\mu(\mathcal{P}^n)$ é subaditiva, temos que o limite

$$h_\mu(f,\mathcal{P}) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}^n) \tag{4.1.15}$$

existe e coincide com o ínfimo da sequência do lado direito. Chamamos $h_\mu(f,\mathcal{P})$ *entropia de f com relação à partição $\mathcal{P}$*. Observe que esta entropia é tanto maior quanto mais fina for a partição. De fato, se $\mathcal{P} \prec \mathcal{Q}$ então $\mathcal{P}^n \prec \mathcal{Q}^n$ para todo n. Usando o Lema 4.1.5, segue que $H_\mu(\mathcal{P}^n) \le H_\mu(\mathcal{Q}^n)$ para todo n. Consequentemente,

$$\mathcal{P} \prec \mathcal{Q} \quad \Rightarrow \quad h_\mu(f,\mathcal{P}) \le h_\mu(f,\mathcal{Q}). \tag{4.1.16}$$

Finalmente, a *entropia* do sistema (f,μ) é definida por

$$h_\mu(f) = \sup_{\mathcal{P}} h_\mu(f,\mathcal{P}), \tag{4.1.17}$$

onde o supremo é tomado sobre todas as partições com entropia finita. Uma observação útil é que a definição não é afetada se considerarmos o supremo apenas sobre as partições finitas (veja o Exercício 4.4.2).

Exemplo 4.1.8. Suponhamos que a medida invariante μ está suportada numa órbita periódica. Em outras palavras, existe x em M e $k \ge 1$ tal que $f^k(x) = x$ e a medida μ é dada por

$$\mu = \frac{1}{k}\big(\delta_x + \delta_{f(x)} + \cdots + \delta_{f^{k-1}(x)}\big).$$

Neste caso a medida só toma um número finito de valores. Consequentemente, a entropia $H_\mu(\mathcal{P})$ também só toma um número finito de valores quando consideramos todas as partições $\mathcal{P}$. Em particular, $\lim_n n^{-1} H_\mu(\mathcal{P}^n) = 0$ para toda partição $\mathcal{P}$. Isto prova que neste caso $h_\mu(f) = 0$.

Exemplo 4.1.9. Considere a transformação expansão decimal $f : [0,1] \to [0,1]$, dada por $f(x) = 10x - [10x]$. Como observamos anteriormente, f preserva a medida de Lebesgue no intervalo μ. Seja $\mathcal{P}$ a partição de $[0,1]$ nos intervalos da forma $\big((i-1)/10, i/10\big]$ com $i = 1,\dots,10$. Então $\mathcal{P}^n$ é a partição nos intervalos da forma $\big((i-1)/10^n, i/10^n\big]$ com $i = 1,\dots,10^n$. Usando o cálculo do Exemplo 4.1.1, obtemos que

$$h_\mu(f,\mathcal{P}) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}^n) = \log 10.$$

Usando a teoria que será desenvolvida na Seção 4.2 (teorema de Kolmogorov-Sinai e seus corolários), veremos que este é também o valor da entropia $h_\mu(f)$, ou seja, $\mathcal{P}$ realiza o supremo na definição (4.1.17).

Exemplo 4.1.10. Considere o deslocamento $\sigma : \Sigma \to \Sigma$ em $\Sigma = \{1,\dots,d\}^{\mathbb{N}}$ (ou $\Sigma = \{1,\dots,d\}^{\mathbb{Z}}$), munido de uma medida de Bernoulli $\mu = \nu^{\mathbb{N}}$ (respectivamente, $\mu = \nu^{\mathbb{Z}}$). Seja $\mathcal{P}$ a partição de Σ em cilindros $[0;a]$ com $a = 1,\dots,d$. Então $\mathcal{P}^n$

é a partição em cilindros $[0; a_1, \dots, a_n]$ de comprimento n. Usando o cálculo do Exemplo 4.1.2 concluímos que

$$h_\mu(\sigma, \mathcal{P}) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}^n) = \sum_{i=1}^d -p_i \log p_i. \tag{4.1.18}$$

A teoria que apresentaremos na Seção 4.2 permitirá concluir que este é também o valor da entropia $h_\mu(\sigma)$.

Segue da expressão (4.1.18) que para todo $x > 0$ existe algum deslocamento de Bernoulli (σ, μ) tal que $h_\mu(\sigma) = x$. Usaremos esta observação um certo número de vezes ao longo do texto.

Lema 4.1.11. *$h_\mu(f, \mathcal{Q}) \le h_\mu(f, \mathcal{P}) + H_\mu(\mathcal{Q}/\mathcal{P})$ para quaisquer partições $\mathcal{P}$ e $\mathcal{Q}$ com entropia finita.*

Demonstração. Pelo Lema 4.1.5, para todo $n \ge 1$ vale que

$$\begin{aligned} H_\mu\big(\mathcal{Q}^{n+1}/\mathcal{P}^{n+1}\big) &= H_\mu\big(\mathcal{Q}^n \vee f^{-n}(\mathcal{Q})/\mathcal{P}^n \vee f^{-n}(\mathcal{P})\big) \\ &\le H_\mu\big(\mathcal{Q}^n/\mathcal{P}^n\big) + H_\mu\big(f^{-n}(\mathcal{Q})/f^{-n}(\mathcal{P})\big). \end{aligned}$$

O último termo é igual a $H_\mu(\mathcal{Q}/\mathcal{P})$, porque a medida μ é invariante por f. Portanto, a relação anterior prova que

$$H_\mu\big(\mathcal{Q}^n/\mathcal{P}^n\big) \le n H_\mu\big(\mathcal{Q}/\mathcal{P}\big) \quad \text{para todo } n \ge 1. \tag{4.1.19}$$

Usando o Lema 4.1.5 uma vez mais, segue que

$$H_\mu(\mathcal{Q}^n) \le H_\mu(\mathcal{P}^n \vee \mathcal{Q}^n) = H_\mu(\mathcal{P}^n) + H_\mu(\mathcal{Q}^n/\mathcal{P}^n) \le H_\mu(\mathcal{P}^n) + n H_\mu(\mathcal{Q}/\mathcal{P}).$$

Dividindo por n e passando ao limite quando $n \to \infty$ obtemos a conclusão do lema. □

Lema 4.1.12. *$h_\mu(f, \mathcal{P}) = \lim_n H_\mu(\mathcal{P}/\bigvee_{j=1}^n f^{-j}(\mathcal{P}))$ para qualquer partição $\mathcal{P}$ com entropia finita.*

Demonstração. Usando o Lema 4.1.5(a) e o fato de que a medida μ é invariante:

$$\begin{aligned} H_\mu\big(\bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P})\big) &= H_\mu\big(\bigvee_{j=1}^{n-1} f^{-j}(\mathcal{P})\big) + H_\mu\big(\mathcal{P}/\bigvee_{j=1}^{n-1} f^{-j}(\mathcal{P})\big) \\ &= H_\mu\big(\bigvee_{j=0}^{n-2} f^{-j}(\mathcal{P})\big) + H_\mu\big(\mathcal{P}/\bigvee_{j=1}^{n-1} f^{-j}(\mathcal{P})\big) \end{aligned}$$

para todo n. Por recorrência, segue que

$$H_\mu\big(\bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P})\big) = H_\mu(\mathcal{P}) + \sum_{k=1}^{n-1} H_\mu\big(\mathcal{P}/\bigvee_{j=1}^{k} f^{-j}(\mathcal{P})\big).$$

Portanto, $h_\mu(f,\mathcal{P})$ é dada pelo limite Cesàro

$$h_\mu(f,\mathcal{P}) = \lim_n \frac{1}{n} H_\mu\big(\bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P})\big) = \lim_n \frac{1}{n} \sum_{k=1}^{n-1} H_\mu\big(\mathcal{P}/\bigvee_{j=1}^{k} f^{-j}(\mathcal{P})\big).$$

Por outro lado, o Lema 4.1.5(b) garante que a sequência $H_\mu(\mathcal{P}/\vee_{j=1}^n f^{-j}(\mathcal{P}))$ é decrescente. Em particular, $\lim_n H_\mu(\mathcal{P}/\vee_{j=1}^n f^{-j}(\mathcal{P}))$ existe e, consequentemente, coincide com o limite Cesàro na igualdade anterior. $\square$

Recorde que $\mathcal{P}^n = \vee_{j=0}^{n-1} f^{-j}(\mathcal{P})$. Quando $f : M \to M$ é invertível, também consideramos $\mathcal{P}^{\pm n} = \vee_{j=-n}^{n-1} f^{-j}(\mathcal{P})$.

Lema 4.1.13. *Se $\mathcal{P}$ é partição com entropia finita então $h_\mu(f,\mathcal{P}) = h_\mu(f,\mathcal{P}^k)$ para todo $k \geq 1$. Se f é invertível, também temos $h_\mu(f,\mathcal{P}) = h_\mu(f,\mathcal{P}^{\pm k})$ para todo $k \geq 1$.*

Demonstração. Observe que, dado qualquer $n \geq 1$,

$$\bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P}^k) = \bigvee_{j=0}^{n-1} f^{-j}\big(\bigvee_{i=0}^{k-1} f^{-i}(\mathcal{P})\big) = \bigvee_{l=0}^{n+k-2} f^{-l}(\mathcal{P}) = \mathcal{P}^{n+k-1}.$$

Portanto,

$$h_\mu\big(f,\mathcal{P}^k\big) = \lim_n \frac{1}{n} H_\mu\big(\mathcal{P}^{n+k-1}\big) = \lim_n \frac{1}{n} H_\mu\big(\mathcal{P}^n\big) = h_\mu\big(f,\mathcal{P}\big).$$

Isto prova a primeira parte do lema. Para provar a segunda parte, note que:

$$\bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P}^{\pm k}) = \bigvee_{j=0}^{n-1} f^{-j}\big(\bigvee_{i=-k}^{k-1} f^{-i}(\mathcal{P})\big) = \bigvee_{l=-k}^{n+k-2} f^{-l}(\mathcal{P}) = f^{-k}\big(\mathcal{P}^{n+2k-1}\big)$$

para todo n e todo k. Portanto,

$$h_\mu\big(f,\mathcal{P}^{\pm k}\big) = \lim_n \frac{1}{n} H_\mu\big(f^{-k}(\mathcal{P}^{n+2k-1})\big) = \lim_n \frac{1}{n} H_\mu\big(\mathcal{P}^{n+2k-1}\big) = h_\mu\big(f,\mathcal{P}\big)$$

(a segunda igualdade usa o fato de que μ é invariante por f). $\square$

Proposição 4.1.14. *Tem-se $h_\mu(f^k) = k h_\mu(f)$ para todo $k \in \mathbb{N}$. Se f é invertível então $h_\mu(f^k) = |k| h_\mu(f)$ para todo $k \in \mathbb{Z}$.*

Demonstração. É claro que a igualdade vale para $k = 0$, uma vez que $f^0 = \mathrm{id}$ e $h_\mu(\mathrm{id}) = 0$. No que segue, supomos que k é diferente de zero. Considere $g = f^k$ e seja $\mathcal{P}$ uma partição qualquer de M com entropia finita. Lembrando que $\mathcal{P}^k = \mathcal{P} \vee f^{-1}(\mathcal{P}) \vee \cdots \vee f^{-k+1}(\mathcal{P})$, vemos que

$$\mathcal{P}^{km} = \bigvee_{j=0}^{km-1} f^{-j}(\mathcal{P}) = \bigvee_{i=0}^{m-1} f^{-ki}\big(\bigvee_{j=0}^{k-1} f^{-j}(\mathcal{P})\big) = \bigvee_{i=0}^{m-1} g^{-i}(\mathcal{P}^k).$$

Portanto,

$$kh_\mu\big(f,\mathcal{P}\big) = \lim_m \frac{1}{m} H_\mu\big(\mathcal{P}^{km}\big) = h_\mu\big(g,\mathcal{P}^k\big). \tag{4.1.20}$$

Como $\mathcal{P} \prec \mathcal{P}^k$, esta igualdade implica que $h_\mu(g,\mathcal{P}) \leq kh_\mu(f,\mathcal{P}) \leq h_\mu(g)$ qualquer que seja $\mathcal{P}$. Tomando o supremo sobre as partições $\mathcal{P}$, segue que $h_\mu(g) \leq kh_\mu(f) \leq h_\mu(g)$. Isto prova que $kh_\mu(f) = h_\mu(g)$, conforme afirmado.

Agora suponha que f é invertível. Seja $\mathcal{P}$ uma partição qualquer de M com entropia finita. Para qualquer $n \geq 1$,

$$H_\mu\big(\vee_{j=0}^{n-1} f^{-j}(\mathcal{P})\big) = H_\mu\big(f^{-n+1}\big(\vee_{i=0}^{n-1} f^{i}(\mathcal{P})\big)\big) = H_\mu\big(\vee_{i=0}^{n-1} f^{i}(\mathcal{P})\big),$$

uma vez que a medida μ é invariante. Dividindo por n e passando ao limite quando $n \to \infty$, obtemos que

$$h_\mu(f,\mathcal{P}) = h_\mu(f^{-1},\mathcal{P}). \tag{4.1.21}$$

Tomando o supremo sobre estas partições $\mathcal{P}$, vem que $h_\mu(f) = h_\mu(f^{-1})$. Substituindo f for f^k e usando a primeira metade da proposição, concluímos que $h_\mu(f^{-k}) = h_\mu(f^k) = kh_\mu(f)$ para todo $k \in \mathbb{N}$. $\square$

4.2 Teorema de Kolmogorov-Sinai

Em geral, a principal dificuldade para calcular a entropia reside no cálculo do supremo na definição (4.1.17). Os métodos que vamos desenvolver nesta seção permitem simplificar a tarefa em muitos casos, identificando certas partições $\mathcal{P}$ que realizam o supremo, isto é, tais que $h_\mu(f,\mathcal{P}) = h_\mu(f)$. O resultado principal é o seguinte:

Teorema 4.2.1 (Kolmogorov-Sinai)**.** *Seja $\mathcal{P}_1 \prec \cdots \prec \mathcal{P}_n \prec \cdots$ uma sequência não-decrescente de partições com entropia finita tais que $\cup_{n=1}^\infty \mathcal{P}_n$ gera a σ-álgebra dos conjuntos mensuráveis, a menos de medida nula. Então*

$$h_\mu(f) = \lim_n h_\mu(f,\mathcal{P}_n).$$

Demonstração. O limite sempre existe, pois a propriedade (4.1.16) implica que a sequência $h_\mu(f,\mathcal{P}_n)$ é não decrescente. A desigualdade $\geq$ no enunciado é consequência direta da definição de entropia. Portanto, só precisamos mostrar que $h_\mu(f,\mathcal{Q}) \leq \lim_n h_\mu(f,\mathcal{P}_n)$ para toda partição $\mathcal{Q}$ com entropia finita. Vamos usar o seguinte fato, que é interessante em si mesmo:

Proposição 4.2.2. *Seja $\mathcal{A}$ uma álgebra que gera a σ-álgebra dos conjuntos mensuráveis, a menos de medida nula. Para toda partição $\mathcal{Q}$ com entropia finita e todo $\varepsilon > 0$ existe alguma partição finita $\mathcal{P} \subset \mathcal{A}$ tal que $H_\mu(\mathcal{Q}/\mathcal{P}) < \varepsilon$.*

Demonstração. O primeiro passo é reduzir o enunciado ao caso em que $\mathcal{Q}$ é finita. Denote por Q_j, $j = 1, 2, \ldots$ os elementos de $\mathcal{Q}$. Para cada $k \geq 1$, considere a partição finita

$$\mathcal{Q}_k = \{Q_1, \ldots, Q_k, M \setminus \cup_{j=1}^k Q_j\}.$$

Lema 4.2.3. $\lim_k H_\mu(\mathcal{Q}/\mathcal{Q}_k) = 0$ *para toda partição* $\mathcal{Q}$ *com entropia finita.*

Demonstração. Denote $Q_0 = M \setminus \cup_{j=1}^k Q_j$. Por definição,

$$H_\mu(\mathcal{Q}/\mathcal{Q}_k) = \sum_{i=0}^{k} \sum_{j \geq 1} -\mu(Q_i \cap Q_j) \log \frac{\mu(Q_i \cap Q_j)}{\mu(Q_i)}.$$

Todas as parcelas com $i \geq 1$ são nulas, uma vez que nesse caso $\mu(Q_i \cap Q_j)$ é igual a zero se $i \neq j$ e é igual a $\mu(Q_i)$ se $i = j$. Para $i = 0$ temos que $\mu(Q_i \cap Q_j)$ é igual a zero se $j \leq k$ e é igual a $\mu(Q_j)$ se $j > k$. Portanto,

$$H_\mu(\mathcal{Q}/\mathcal{Q}_k) = \sum_{j>k} -\mu(Q_j) \log \frac{\mu(Q_j)}{\mu(Q_0)} \leq \sum_{j>k} -\mu(Q_j) \log \mu(Q_j).$$

A hipótese de que $\mathcal{Q}$ tem entropia finita significa que a expressão do lado direito converge para zero quando $k \to \infty$. □

Dado $\varepsilon > 0$, fixe $k \geq 1$ tal que $H_\mu(\mathcal{Q}/\mathcal{Q}_k) < \varepsilon/2$. Considere qualquer $\delta > 0$. Pelo teorema de aproximação (Teorema A.1.19), para cada $i = 1, \ldots, k$ existe $A_i \in \mathcal{A}$ tal que

$$\mu(Q_i \Delta A_i) < \delta/(2k^2). \tag{4.2.1}$$

Defina $P_1 = A_1$ e $P_i = A_i \setminus \cup_{j=1}^{i-1} A_j$ para $i = 2, \ldots, k$ e também $P_0 = M \setminus \cup_{j=1}^{k-1} A_j$. É claro que $\mathcal{P} = \{P_1, \ldots, P_k, P_0\}$ é partição de M e que $P_i \in \mathcal{A}$ para todo i. Para $i = 1, \ldots, k$, temos que $P_i \Delta A_i = P_i \setminus A_i = A_i \cap \left(\cup_{j=1}^{i-1} A_j\right)$. Dado qualquer x neste conjunto, existe $j < i$ tal que $x \in A_i \cap A_j$. Como $Q_i \cap Q_j = \emptyset$, segue que $x \in (A_i \setminus Q_i) \cup (A_j \setminus Q_j)$. Isto prova que

$$P_i \Delta A_i \subset \cup_{j=1}^{i} (A_j \setminus Q_j) \subset \cup_{j=1}^{i} (A_j \Delta Q_j)$$

e, portanto, $\mu(P_i \Delta A_i) < i\delta/(2k^2) \leq \delta/(2k)$. Juntamente com (4.2.1), isto implica que

$$\mu(P_i \Delta Q_i) < \delta/(2k^2) + \delta/(2k) \leq \delta/k \quad \text{para } i = 1, \ldots, k. \tag{4.2.2}$$

Além disso, $P_0 \Delta Q_0 \subset \cup_{i=1}^k P_i \Delta Q_i$ uma vez que $\mathcal{P}$ e $\mathcal{Q}_k$ são partições de M. Logo, (4.2.2) implica

$$\mu(P_0 \Delta Q_0) < \delta. \tag{4.2.3}$$

De acordo com o Lema 4.1.6, supondo que $\delta > 0$ é suficientemente pequeno, as relações (4.2.2) e (4.2.3) implicam que $H_\mu(\mathcal{Q}_k/\mathcal{P}) < \varepsilon/2$. Então, pela desigualdade no Exercício 4.4.1,

$$H_\mu(\mathcal{Q}/\mathcal{P}) \leq H_\mu(\mathcal{Q}/\mathcal{Q}_k) + H_\mu(\mathcal{Q}_k/\mathcal{P}) < \varepsilon,$$

tal como afirmado. □

Corolário 4.2.4. *Se* $(\mathcal{P}_n)_n$ *é sequência de partições nas condições do Teorema 4.2.1 então* $\lim_n H_\mu(\mathcal{Q}/\mathcal{P}_n) = 0$ *para toda partição* $\mathcal{Q}$ *com entropia finita.*

Demonstração. Para cada n, seja $\mathcal{A}_n$ a álgebra gerada por $\mathcal{P}_n$. Note que os elementos de $\mathcal{A}_n$ são as uniões finitas de elementos de $\mathcal{P}_n$ juntamente com os complementos dessas uniões. Além disso, $\mathcal{A} = \cup_n \mathcal{A}_n$ é a álgebra gerada por $\cup_n \mathcal{P}_n$. Considere qualquer $\varepsilon > 0$. Pela Proposição 4.2.2, existe uma partição finita $\mathcal{P} \subset \mathcal{A}$ tal que $H_\mu(\mathcal{Q}/\mathcal{P}) < \varepsilon$. Logo, uma vez que $\mathcal{P}$ é finita, existe $m \geq 1$ tal que $\mathcal{P} \subset \mathcal{A}_m$ e, portanto, $\mathcal{P}$ é menos fina que $\mathcal{P}_m$. Então, usando o Lema 4.1.5,

$$H_\mu(\mathcal{Q}/\mathcal{P}_n) \leq H_\mu(\mathcal{Q}/\mathcal{P}_m) \leq H_\mu(\mathcal{Q}/\mathcal{P}) < \varepsilon \quad \text{para todo } n \geq m.$$

Isto completa a demonstração do corolário. □

Estamos prontos para concluir a demonstração do Teorema 4.2.1. Pelo Lema 4.1.11, temos que

$$h_\mu(f, \mathcal{Q}) \leq h_\mu(f, \mathcal{P}_n) + H_\mu(\mathcal{Q}/\mathcal{P}_n) \quad \text{para todo } n.$$

Passando ao limite quando $n \to \infty$, obtemos que $h_\mu(f, \mathcal{Q}) \leq \lim_n h_\mu(f, \mathcal{P}_n)$ para toda partição $\mathcal{Q}$ com entropia finita. □

4.2.1 Partições geradoras

Nesta seção, e nas seguintes, deduziremos diversas consequências úteis do Teorema 4.2.1.

Corolário 4.2.5. *Seja $\mathcal{P}$ uma partição com entropia finita tal que a união dos iterados $\mathcal{P}^n = \vee_{j=0}^{n-1} f^{-j}(\mathcal{P})$, $n \geq 1$ gera a σ-álgebra dos conjuntos mensuráveis, a menos de medida nula. Então $h_\mu(f) = h_\mu(f, \mathcal{P})$.*

Demonstração. Basta aplicar o Teorema 4.2.1 à sequência $\mathcal{P}^n$, lembrando que $h_\mu(f, \mathcal{P}^n) = h_\mu(f, \mathcal{P})$ para todo n, de acordo com o Lema 4.1.13. □

Corolário 4.2.6. *Suponha que o sistema (f, μ) é invertível. Seja $\mathcal{P}$ uma partição com entropia finita tal que a união dos iterados $\mathcal{P}^{\pm n} = \vee_{j=-n}^{n-1} f^{-j}(\mathcal{P})$, $n \geq 1$ gera a σ-álgebra dos conjuntos mensuráveis, a menos de medida nula. Então $h_\mu(f) = h_\mu(f, \mathcal{P})$.*

Demonstração. Basta aplicar o Teorema 4.2.1 à sequência $\mathcal{P}^{\pm n}$, lembrando que $h_\mu(f, \mathcal{P}^{\pm n}) = h_\mu(f, \mathcal{P})$ para todo n, de acordo com o Lema 4.1.13. □

Em particular, os Corolários 4.2.5 e 4.2.6 completam o cálculo da entropia da transformação expansão decimal e dos deslocamentos de Bernoulli, que iniciamos nos Exemplos 4.1.9 e 4.1.10, respectivamente.

Em qualquer dos casos nos Corolários 4.2.5 e 4.2.6 dizemos que $\mathcal{P}$ é uma *partição geradora*, ou um *gerador* do sistema. Note, no entanto, que isto contém um certo abuso de linguagem, já que as condições nos dois corolários não são equivalentes. Por exemplo, para o deslocamento em $M = \{1, \ldots, d\}^{\mathbb{Z}}$, a partição $\mathcal{P}$ em cilindros $\{[0; a] : a = 1, \ldots, d\}$ é tal que a união dos iterados bilaterais $\mathcal{P}^{\pm n}$ gera a σ-álgebra mas a união dos iterados unilaterais $\mathcal{P}^n$ não gera. Quando for necessário distinguir entre os dois conceitos, falaremos de gerador *unilateral* e gerador bilateral, respectivamente.

A este respeito também observamos que certos sistemas *invertíveis* admitem geradores unilaterais. Por exemplo, se $f : S^1 \to S^1$ é uma rotação irracional e $\mathcal{P} = \{I, S^1 \setminus I\}$ é uma partição do círculo em dois intervalos complementares, então $\mathcal{P}$ é gerador unilateral (e também bilateral, evidentemente). No entanto, este tipo de situação só é possível para sistemas com entropia nula:

Corolário 4.2.7. *Suponha que o sistema (f, μ) é invertível e existe alguma partição $\mathcal{P}$ com entropia finita tal que $\cup_{n=1}^{\infty} \mathcal{P}^n$ gera a σ-álgebra dos conjuntos mensuráveis, a menos de medida nula. Então $h_\mu(f) = 0$.*

Demonstração. Combinando o Lema 4.1.12 e o Corolário 4.2.5:

$$h_\mu(f) = h_\mu(f, \mathcal{P}) = \lim_n H_\mu(\mathcal{P}/f^{-1}(\mathcal{P}^n)).$$

Como $\cup_n \mathcal{P}^n$ gera a σ-álgebra $\mathcal{B}$ dos conjuntos mensuráveis, $\cup_n f^{-1}(\mathcal{P}^n)$ gera a σ-álgebra $f^{-1}(\mathcal{B})$. Mas $f^{-1}(\mathcal{B}) = \mathcal{B}$, uma vez que f é invertível. Logo, o Corolário 4.2.4 implica que $H_\mu(\mathcal{P}/f^{-1}(\mathcal{P}^n))$ converge para zero quando $n \to \infty$. Segue que $h_\mu(f) = h_\mu(f, \mathcal{P}) = 0$. □

Suponha que M é um espaço métrico, munido da sua σ-álgebra de Borel.

Corolário 4.2.8. *Seja $\mathcal{P}_1 \prec \cdots \prec \mathcal{P}_n \prec \cdots$ uma sequência não decrescente de partições com entropia finita tais que* $\operatorname{diam} \mathcal{P}_n(x) \to 0$ *para μ-quase todo $x \in M$. Então*

$$h_\mu(f) = \lim_n h_\mu(f, \mathcal{P}_n).$$

Demonstração. Seja U um aberto, não vazio, qualquer de M. A hipótese garante que para cada $x \in U$ existe $n(x)$ tal que o conjunto $P_x = \mathcal{P}_{n(x)}(x)$ está contido em U. É claro que P_x pertence à álgebra $\mathcal{A}$ gerada por $\cup_n \mathcal{P}_n$. Observe também que esta álgebra é enumerável, já que ela está formada pelas uniões finitas de elementos das partições $\mathcal{P}_n$ juntamente com os complementares de tais uniões. Em particular, o conjunto dos valores tomados por P_x é enumerável. Segue que $U = \cup_{x \in U} P_x$ está na σ-álgebra gerada $\mathcal{A}$. Isto prova que a σ-álgebra gerada por $\mathcal{A}$ contém todos os abertos e, portanto, contém todos os conjuntos borelianos. Agora, a conclusão segue de uma aplicação direta do Teorema 4.2.1. □

Exemplo 4.2.9. Seja $f : S^1 \to S^1$ um homeomorfismo e seja μ uma probabilidade invariante qualquer. Dada uma partição finita $\mathcal{P}$ de S^1 em subintervalos, denotemos por $x_1, \ldots, x_m$ os seus pontos extremos. Para qualquer $j \geq 1$, a partição $f^{-j}(\mathcal{P})$ está formada pelos subintervalos de S^1 determinados pelos pontos $f^{-j}(x_i)$. Isto implica que, para cada $n \geq 1$, os elementos de $\mathcal{P}^n$ têm os seus pontos extremos no conjunto

$$\{f^{-j}(x_i) : j = 0, \ldots, n-1 \text{ e } i = 1, \ldots, m\}.$$

Em particular, $\#\mathcal{P}^n \le mn$. Então, usando o Lema 4.1.3,

$$h_\mu(f,\mathcal{P}) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}^n) \le \lim_n \frac{1}{n} \log \#\mathcal{P}^n \le \lim_n \frac{1}{n} \log mn = 0.$$

Segue que $h_\mu(f) = 0$: para isso basta considerar qualquer sequência de partições finitas em intervalos com diâmetro indo para zero e aplicar o Corolário 4.2.8.

Corolário 4.2.10. *Seja $\mathcal{P}$ uma partição com entropia finita tal que, para μ-quase todo $x \in M$, tem-se* $\operatorname{diam} \mathcal{P}^n(x) \to 0$. *Então $h_\mu(f) = h_\mu(f,\mathcal{P})$.*

Demonstração. Basta aplicar o Corolário 4.2.8 à sequência $\mathcal{P}^n$, lembrando que $h_\mu(f,\mathcal{P}^n) = h_\mu(f,\mathcal{P})$ para todo n. □

Analogamente, se f é invertível e $\mathcal{P}$ é uma partição de M com entropia finita e tal que $\operatorname{diam} \mathcal{P}^{\pm n}(x) \to 0$ para μ-quase todo $x \in M$, então $h_\mu(f) = h_\mu(f,\mathcal{P})$.

Sabemos que geradores existem na maioria dos casos interessantes, embora possa ser difícil exibir um gerador explicitamente. De fato, suponha que o ambiente é um espaço de Lebesgue. Rokhlin [Rok67, §10] mostrou que se um sistema é aperiódico (isto é, o conjunto dos pontos periódicos tem medida nula) e quase todo ponto tem um conjunto enumerável (finito ou infinito) de pré-imagens, então existe algum gerador enumerável. Em particular, todo sistema invertível aperiódico admite algum gerador enumerável. Em geral, este gerador pode ter entropia infinita. Mas Rokhlin [Rok67, §10] também mostrou que todo sistema invertível aperiódico com entropia finita admite algum gerador bilateral com entropia finita. Além disso (Krieger [Kri70]), este gerador pode ser escolhido finito se o sistema for ergódico.

4.2.2 Semicontinuidade da entropia

A seguir vamos examinar a *função entropia*, que associa a cada medida invariante μ de uma transformação f dada o valor da respectiva entropia $h_\mu(f)$. Vamos ver que esta função não é contínua, em geral. No entanto, sob hipóteses bastante amplas ela é *semicontínua superiormente*: dado qualquer $\varepsilon > 0$, tem-se que $h_\nu(f) \le h_\mu(f) + \varepsilon$ para todo ν suficientemente próximo de μ. Isso vale, em particular, para a classe de transformações que chamamos de expansivas.

Começamos por mostrar, por meio de um exemplo, que a função entropia pode ser descontínua:

Exemplo 4.2.11. Seja $f : [0,1] \to [0,1]$ a transformação expansão decimal. Como vimos no Exemplo 4.1.9, a entropia de f relativamente à medida de Lebesgue m é $h_m(f) = \log 10$. Para cada $k \ge 1$, denote por F_k o conjunto dos pontos fixos do iterado f^k. Observe que F_k é um conjunto invariante com $\#F_k = 10^k$, e que estes conjuntos estão equidistribuídos no seguinte sentido: cada intervalo $[(i-1)/10^k, i/10^k]$ contém exatamente um ponto de F_k. Considere a sequência de medidas

$$\mu_k = \frac{1}{10^k} \sum_{x \in F_k} \delta_x.$$

As observações anteriores implicam que cada μ_k é uma probabilidade invariante e que a sequência $(\mu_k)_k$ converge para a medida de Lebesgue m na topologia fraca* (verifique). Como μ_k está suportada num conjunto finito, o mesmo argumento que usamos no Exemplo 4.1.8 prova que $h_{\mu_k}(f) = 0$ para todo k. Portanto, a entropia não varia continuamente com a medida invariante.

Por outro lado, em geral, considere qualquer partição finita $\mathcal{P}$ de M cujo *bordo*

$$\partial\mathcal{P} = \bigcup_{P\in\mathcal{P}} \partial P$$

satisfaz $\mu(\partial\mathcal{P}) = 0$. A função $\nu \mapsto \nu(P)$ é contínua no ponto μ, para todo $P \in \mathcal{P}$. Consequentemente, a função

$$\nu \mapsto H_\nu(\mathcal{P}) = \sum_{P\in\mathcal{P}} -\nu(P)\log\nu(P)$$

também é contínua em μ. A hipótese sobre $\mathcal{P}$ também implica que $\mu(\partial\mathcal{P}^n) = 0$ para todo $n \geq 1$, uma vez que

$$\partial\mathcal{P}^n \subset \partial\mathcal{P} \cup f^{-1}(\partial\mathcal{P}) \cup \cdots \cup f^{-n+1}(\partial\mathcal{P}).$$

Assim, a função $\nu \mapsto H_\nu(\mathcal{P}^n)$ é contínua para todo n.

Proposição 4.2.12. *Seja $\mathcal{P}$ uma partição finita tal que $\mu(\partial\mathcal{P}) = 0$. Então, a função $\nu \mapsto h_\nu(f,\mathcal{P})$ é semicontínua superiormente em μ.*

Demonstração. Basta lembrar que, por definição,

$$h_\nu(f,\mathcal{P}) = \inf_n \frac{1}{n} H_\nu(f,\mathcal{P})$$

e que o ínfimo de qualquer família de funções contínuas é uma função semicontínua superiormente. □

Corolário 4.2.13. *Suponha que existe uma partição finita $\mathcal{P}$ tal que $\mu(\partial\mathcal{P}) = 0$ e $\cup_n\mathcal{P}^n$ gera a σ-álgebra dos subconjuntos mensuráveis de M a menos de medida nula. Então a função $\eta \mapsto h_\eta(f)$ é semicontínua superiormente no ponto μ.*

Demonstração. Pela Proposição 4.2.12, dado $\varepsilon > 0$ existe uma vizinhança U de μ tal que $h_\nu(f,\mathcal{P}) \leq h_\mu(f,\mathcal{P}) + \varepsilon$ para todo $\nu \in V$. Temos $h_\mu(f,\mathcal{P}) \leq h_\mu(f)$, por definição. Pelo Corolário 4.2.5, a hipótese implica que $h_\nu(f,\mathcal{P}) = h_\nu(f)$ para todo ν. Portanto, $h_\nu(f) \leq h_\mu(f) + \varepsilon$ para todo $\nu \in V$. □

Agora suponhamos que M é um espaço métrico compacto e μ é uma probabilidade boreliana em M. Por definição, o diâmetro $\operatorname{diam}\mathcal{P}$ de uma partição $\mathcal{P}$ é o supremo dos diâmetros dos seus elementos. Então temos a seguinte versão mais especializada do corolário anterior:

Corolário 4.2.14. *Suponha que existe $\varepsilon_0 > 0$ tal que toda partição finita $\mathcal{P}$ com $\operatorname{diam}\mathcal{P} < \varepsilon_0$ satisfaz $\lim_n \operatorname{diam}\mathcal{P}^n = 0$. Então, a função $\mu \mapsto h_\mu(f)$ é semi-contínua superiormente. Consequentemente, essa função é limitada e o seu supremo é atingido para alguma medida μ.*

Demonstração. Como vimos no Corolário 4.2.10, a propriedade $\lim_n \operatorname{diam} \mathcal{P}^n = 0$ implica que $\cup_n \mathcal{P}^n$ gera a σ-álgebra dos conjuntos mensuráveis. Por outro lado, dada qualquer probabilidade invariante μ é fácil escolher[2]; seja $\mathcal{U}$ uma cobertura finita de M por tais bolas; tome para $\mathcal{P}$ a partição associada a $\mathcal{U}$, ou seja, a partição cujos elementos são os conjuntos maximais que, para cada $U \in \mathcal{U}$, estão contidos em U ou no complementar U^c; uma partição $\mathcal{P}$ com diâmetro menor que ε_0 e tal que $\mu(\partial\mathcal{P}) = 0$. Segue do corolário anterior que a função entropia é semicontínua superiormente em μ e, como μ é arbitrária, isso dá a primeira afirmação do enunciado.

As demais afirmações são consequências gerais da semicontinuidade, lembrando que o domínio da função entropia, ou seja, o espaço $\mathcal{M}_1(f)$ das probabilidades invariantes, é um espaço compacto. $\square$

Quando f é invertível podemos substituir $\mathcal{P}^n$ por $\mathcal{P}^{\pm n} = \vee_{j=-n}^{n-1} f^{-j}(\mathcal{P})$ no enunciado dos Corolários 4.2.13 e 4.2.14. A demonstração é análoga, usando o Corolário 4.2.6 e a versão do Corolário 4.2.10 para transformações invertíveis.

4.2.3 Transformações expansivas

Agora vamos discutir uma classe bastante ampla de transformações que satisfazem as condições do Corolário 4.2.14.

Uma transformação contínua $f : M \to M$ num espaço métrico é dita *expansiva* se existe $\varepsilon_0 > 0$ (chamada *constante de expansividade*) tal que, dados $x, y \in M$ com $x \neq y$ existe $n \in \mathbb{N}$ tal que $d(f^n(x), f^n(y)) \geq \varepsilon_0$. Ou seja, quaisquer duas órbitas distintas podem ser distinguidas, de forma macroscópica, em algum momento da iteração.

Quando a transformação f é invertível também temos uma versão bilateral da noção de expansividade, definida do seguinte modo: existe $\varepsilon_0 > 0$ tal que, dados $x, y \in M$ com $x \neq y$ existe $n \in \mathbb{Z}$ tal que $d(f^n(x), f^n(y)) \geq \varepsilon_0$. Esta propriedade sempre vale se a transformação for expansiva (no sentido anterior), já que $\mathbb{N} \subset \mathbb{Z}$.

Exemplo 4.2.15. Seja $\sigma : \Sigma \to \Sigma$ o deslocamento em $\Sigma = \{1, \ldots, d\}^{\mathbb{N}}$. Consideramos em Σ a distância $d((x_n)_n, (y_n)_n) = 2^{-N}$ onde N é o menor valor de n tal que $x_n \neq y_n$. Note que $d(\sigma^N(x), \sigma^N(y)) = 2^0 = 1$ se $x = (x_n)_n$ e $y = (y_n)_n$ são pontos distintos. Isto prova que σ é uma transformação expansiva, com $\varepsilon_0 = 1$ como constante de expansividade.

Analogamente, o deslocamento bilateral $\sigma : \Sigma \to \Sigma$ em $\Sigma = \{1, \ldots, d\}^{\mathbb{Z}}$ é expansivo no sentido bilateral (mas não no sentido unilateral).

Deixamos ao cuidado do leitor verificar (Exercício 4.4.6) que a expansão decimal $f(x) = 10x - [10x]$ também é expansiva. Por outro lado, rotações num toro nunca são expansivas.

[2]Por exemplo: para cada x escolha $r_x \in (0, \varepsilon_0)$ tal que o bordo da bola de centro x e raio r_x tenha medida nula

Proposição 4.2.16. *Seja $f : M \to M$ uma transformação expansiva num espaço métrico compacto e seja $\varepsilon_0 > 0$ uma constante de expansividade. Então tem-se* $\lim_n \operatorname{diam} \mathcal{P}^n = 0$ *para toda partição finita $\mathcal{P}$ com* $\operatorname{diam} \mathcal{P} < \varepsilon_0$.

Demonstração. É claro que a sequência $\operatorname{diam} \mathcal{P}^n$ é não crescente. Seja δ o seu ínfimo e suponhamos que $\delta > 0$. Então, para todo $n \geq 1$ existem pontos x_n e y_n tais que $d(x_n, y_n) > \delta/2$ mas x_n e y_n pertencem ao mesmo elemento de $\mathcal{P}^n$ e, portanto, satisfazem

$$d(f^j(x_n), f^j(y_n)) \leq \operatorname{diam} \mathcal{P} < \varepsilon_0 \quad \text{para todo } 0 \leq j < n.$$

Por compacidade, existe $(n_j)_j \to \infty$ tal que $(x_{n_j})_j$ e $(y_{n_j})_j$ convergem para pontos x e y, respectivamente. Então $x \neq y$ mas $d(f^j(x), f^j(y)) \leq \operatorname{diam} \mathcal{P} < \varepsilon_0$ para todo $j \geq 0$. Isto contradiz a hipótese de que ε_0 é constante de expansividade. □

Corolário 4.2.17. *Se $f : M \to M$ é uma transformação expansiva num espaço métrico compacto então a função entropia é semicontínua superiormente e existem probabilidades invariantes μ cuja entropia $h_\mu(f)$ é máxima entre todas as probabilidades invariantes de f.*

Demonstração. Combine a Proposição 4.2.16 com o Corolário 4.2.14. □

Se a transformação f é invertível e expansiva no sentido bilateral, podemos substituir $\mathcal{P}^n$ por $\mathcal{P}^{\pm n}$ na Proposição 4.2.16 e a conclusão do Corolário 4.2.17 também permanece válida tal como está enunciada.

4.3 Entropia local

O teorema de Shannon-McMillan-Breiman, que vamos discutir nesta seção, fornece uma visão complementar do conceito de entropia, mais detalhada e de natureza mais local. Também mencionaremos uma versão topológica dessa ideia, que é devida a Brin-Katok.

Teorema 4.3.1 (Shannon-McMillan-Breiman)**.** *Dada qualquer partição $\mathcal{P}$ com entropia finita, o limite*

$$h_\mu(f, \mathcal{P}, x) = \lim_n -\frac{1}{n} \log \mu(\mathcal{P}^n(x)) \quad \textit{existe em } \mu\textit{-quase todo ponto.} \tag{4.3.1}$$

A função $x \mapsto h_\mu(f, \mathcal{P}, x)$ é μ-integrável, e o limite também vale em $L^1(\mu)$. Além disso,

$$\int h_\mu(f, \mathcal{P}, x)\, d\mu(x) = h_\mu(f, \mathcal{P}).$$

Se (f, μ) é ergódico então $h_\mu(f, \mathcal{P}, x) = h_\mu(f, \mathcal{P})$ em μ-quase todo ponto.

Lembre que $\mathcal{P}^n(x) = \mathcal{P}(x) \cap f^{-1}(\mathcal{P}(f(x))) \cap \cdots \cap f^{-n+1}(\mathcal{P}(f^{n-1}(x)))$, ou seja, este conjunto está formado pelos pontos cujas trajetórias se mantêm "próximas" da trajetória de x durante n iterados, no sentido de que as duas visitam os mesmos elementos de $\mathcal{P}$. O Teorema 4.3.1 afirma que a medida deste conjunto tem uma taxa exponencial de decaimento bem definida: em μ-quase todo ponto,

$$\mu(\mathcal{P}^n(x)) \approx e^{-nh_\mu(f,\mathcal{P},x)} \quad \text{para todo } n \text{ grande.}$$

O teorema de Brin-Katok, que enunciamos a seguir, pertence à mesma família de resultados, mas usa uma noção distinta de proximidade.

Definição 4.3.2. Suponhamos que $f : M \to M$ é uma aplicação contínua num espaço métrico compacto. Dado $x \in M$, $n \geq 1$ e $\varepsilon > 0$, chamamos *bola dinâmica* de comprimento n e raio ε em torno de x ao conjunto:

$$B(x, n, \varepsilon) = \{y \in M : d(f^j(x), f^j(y)) < \varepsilon \text{ para todo } j = 0, 1, \ldots, n-1\}.$$

Em outras palavras, $B(x, n, \varepsilon) = \cap_{j=0}^{n-1} f^{-j}(B(f^j(x), \varepsilon))$. Defina:

$$h_\mu^+(f, \varepsilon, x) = \limsup_n -\frac{1}{n} \log \mu(B(x, n, \varepsilon)) \quad \text{e}$$
$$h_\mu^-(f, \varepsilon, x) = \liminf_n -\frac{1}{n} \log \mu(B(x, n, \varepsilon)).$$

Teorema 4.3.3 (Brin-Katok)**.** *Seja μ uma probabilidade invariante por f. Os limites*

$$\lim_{\varepsilon \to 0} h_\mu^+(f, \varepsilon, x) \quad e \quad \lim_{\varepsilon \to 0} h_\mu^-(f, \varepsilon, x)$$

existem e são iguais para μ-quase todo ponto. Denotando por $h_\mu(f, x)$ o seu valor comum, a função $h_\mu(f, \cdot)$ é integrável e tem-se

$$h_\mu(f) = \int h_\mu(f, x) d\mu(x).$$

A prova deste resultado pode ser encontrada no artigo original de Brin, Katok [BK83] e não será apresentada aqui.

Exemplo 4.3.4 (Translações em grupos compactos)**.** Seja G um grupo compacto metrizável e seja μ a respectiva medida de Haar. Toda translação de G, à esquerda ou à direita, tem entropia nula relativamente a μ. De fato, considere em G uma distância d invariante por translações . Então,

$$E_g^j(B(x, \varepsilon)) = B(E_g^j(x), \varepsilon)$$

para quaisquer $g \in G$ e $j \in \mathbb{Z}$. Consequentemente, $B(x, n, \varepsilon) = B(x, \varepsilon)$ para todo $n \geq 1$. Então,

$$h_\mu^\pm(E_g, \varepsilon, x) = \lim_n -\frac{1}{n} \log \mu(B(x, \varepsilon)) = 0$$

para todo $\varepsilon > 0$ e $x \in G$. Pelo teorema de Brin-Katok, segue que $h_\mu(E_g) = 0$. O mesmo argumento se aplica para translações D_g à direita.

4.4 Exemplos

Vamos agora ilustrar os resultados anteriores por meio de alguns exemplos.

4.4.1 Deslocamentos de Markov

Seja $\Sigma = \{1, \ldots, d\}^{\mathbb{N}}$ e seja $\sigma : \Sigma \to \Sigma$ a transformação deslocamento. Seja μ uma medida de Markov associada a uma matriz estocástica $P = (P_{i,j})_{i,j}$ e um vetor de probabilidade $p = (p_i)_i$. Vamos provar:

Proposição 4.4.1. $h_\mu(\sigma) = \sum_{a=1}^d p_a \sum_{b=1}^d -P_{a,b} \log P_{a,b}$.

Demonstração. Considere a partição $\mathcal{P}$ de Σ em cilindros $[0; a]$, $a = 1, \ldots, d$. Para cada n, o iterado $\mathcal{P}^n$ é a partição em cilindros $[0; a_1, \ldots, a_n]$ de comprimento n. Lembrando que $\mu([0; a_1, \ldots, a_n]) = p_{a_1} P_{a_1,a_2} \cdots P_{a_{n-1},a_n}$, vemos que

$$\begin{aligned} H_\mu(\mathcal{P}^n) &= \sum_{a_1,\ldots,a_n} -p_{a_1} P_{a_1,a_2} \cdots P_{a_{n-1},a_n} \log \left(p_{a_1} P_{a_1,a_2} \cdots P_{a_{n-1},a_n}\right) \\ &= \sum_{a_1} -p_{a_1} \log p_{a_1} \sum_{a_2,\ldots,a_n} P_{a_1,a_2} \cdots P_{a_{n-1},a_n} \\ &\qquad + \sum_{j=1}^{n-1} \sum_{a_j,a_{j+1}} - \log P_{a_j,a_{j+1}} \sum p_{a_1} P_{a_1,a_2} \cdots P_{a_{n-1},a_n}, \end{aligned} \tag{4.4.1}$$

onde a última soma é sobre todos os valores de $a_1, \ldots, a_{j-1}, a_{j+2}, \ldots, a_n$. Por um lado,

$$\sum_{a_2,\ldots,a_n} P_{a_1,a_2} \cdots P_{a_{n-1},a_n} = \sum_{a_n} P^n_{a_1,a_n} = 1$$

uma vez que P^n é uma matriz estocástica. Por outro lado, $P^* p = p$,

$$\begin{aligned} \sum p_{a_1} P_{a_1,a_2} \cdots P_{a_{n-1},a_n} &= \sum_{a_1,a_n} p_{a_1} P^j_{a_1,a_j} P_{a_j,a_{j+1}} P^{n-j-1}_{a_{j+1},a_n} \\ &= \sum_{a_1} p_{a_1} P^j_{a_1,a_j} P_{a_j,a_{j+1}} = p_{a_j} P_{a_j,a_{j+1}}, \end{aligned}$$

porque P^{n-j-1} é uma matriz estocástica e $pP^j = P^{j*} p = p$. Substituindo estas observações em (4.4.1), obtemos que

$$\begin{aligned} H_\mu(\mathcal{P}^n) &= \sum_{a_1} -p_{a_1} \log p_{a_1} + \sum_{j=1}^{n-1} \sum_{a_j,a_{j+1}} -p_{a_j} P_{a_j,a_{j+1}} \log P_{a_j,a_{j+1}} \\ &= \sum_{a} -p_a \log p_a + (n-1) \sum_{a,b} -p_a P_{a,b} \log P_{a,b}. \end{aligned}$$

Então $h_\mu(\sigma, \mathcal{P}) = \sum_{a,b} -p_a P_{a,b} \log P_{a,b}$. Como a família de todos os cilindros $[0; a_1, \ldots, a_n]$ gera a σ-álgebra de Σ, segue do Corolário 4.2.5 que $h_\mu(\sigma) = h_\mu(\sigma, \mathcal{P})$. Isto completa a prova do teorema. □

Esta conclusão permanece válida no caso de deslocamentos de Markov bilaterais, ou seja, em $\Sigma = \{1, \ldots, d\}^{\mathbb{Z}}$. O argumento é análogo, usando o Corolário 4.2.6.

4.4.2 Transformação de Gauss

Vamos calcular a entropia da transformação de Gauss $G(x) = (1/x) - [1/x]$ relativamente à probabilidade invariante

$$\mu(E) = \frac{1}{\log 2} \int_E \frac{dx}{1+x} \tag{4.4.2}$$

estudada nas Seções 1.3.2 e 3.2.4. O método que vamos apresentar se estende a uma classe bastante ampla de sistemas, incluindo as transformações expansoras do intervalo.

Seja $\mathcal{P}$ a partição nos intervalos $(1/(m+1), 1/m)$ para $m \geq 1$. Como antes, denotamos $\mathcal{P}^n = \vee_{j=0}^{n-1} G^{-j}(\mathcal{P})$. Os seguintes fatos serão usados a seguir:

(A) G^n envia cada $P_n \in \mathcal{P}^n$ difeomorficamente sobre $(0,1)$, para cada $n \geq 1$.

(B) $\operatorname{diam} \mathcal{P}^n \to 0$ quando $n \to \infty$.

(C) Existe $C > 1$ tal que $|(G^n)'(y)|/|(G^n)'(x)| \leq C$ para todo $n \geq 1$ e quaisquer x e y num mesmo elemento da partição $\mathcal{P}^n$.

(D) Existem $c_1, c_2 > 0$ tal que $c_1 m(P_n) \leq \mu(P_n) \leq c_2 m(P_n)$ para todo $n \geq 1$ e todo $P_n \in \mathcal{P}_n$, onde m representa a medida de Lebesgue.

É imediato da definição que cada $P \in \mathcal{P}$ é enviado por G difeomorficamente sobre $(0,1)$. A propriedade (A) é uma consequência, por indução em n. Usando (A) e o Lema 3.2.12, obtemos que

$$\operatorname{diam} P_n \leq \sup_{x \in P_n} \frac{1}{|(G^n)'(x)|} \leq 2^{-[n/2]}$$

para todo $n \geq 1$ e todo $P_n \in \mathcal{P}^n$. Isto implica (B). A propriedade (C) está dada pelo Lema 3.2.13. Finalmente, (D) segue diretamente de (4.4.2).

Proposição 4.4.2. $h_\mu(G) = \int \log |G'| \, d\mu$.

Demonstração. Consideremos a função $\psi_n(x) = -\log \mu(\mathcal{P}^n(x))$, para cada $n \geq 1$. Observe que:

$$H_\mu(\mathcal{P}^n) = \sum_{P_n \in \mathcal{P}^n} -\mu(P_n) \log \mu(P_n) = \int \psi_n(x) \, d\mu(x).$$

A propriedade (D) dá que

$$-\log c_1 \geq \psi_n(x) + \log m(\mathcal{P}^n(x)) \geq -\log c_2.$$

Pela propriedade (A), temos que $\log m(\mathcal{P}^n(x)) = -\log|(G^n)'(y)|$ para algum $y \in \mathcal{P}^n(x)$. Usando a propriedade (C), segue que

$$-\log c_1 + \log C \geq \psi_n(x) - \log|(G^n)'(x)| \geq -\log c_2 - \log C$$

para todo x e todo n. Por consequência,

$$-\log(c_1/C) \geq H_\mu(\mathcal{P}^n) - \int \log|(G^n)'|\, d\mu \geq \log(c_2 C) \tag{4.4.3}$$

para todo n. Uma vez que a medida μ é invariante por G,

$$\int \log|(G^n)'|\, d\mu = \sum_{j=0}^{n-1} \int \log|G'| \circ G^j\, d\mu = n \int |G'|\, d\mu.$$

Então dividindo (4.4.3) por n e passando ao limite quando $n \to \infty$,

$$h_\mu(f, \mathcal{P}) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}^n) = \int \log|G'|\, d\mu.$$

Agora a propriedade (B) garante que podemos aplicar o Corolário 4.2.10 para concluir que

$$h_\mu(G) = h_\mu(G, \mathcal{P}) = \int \log|G'|\, d\mu.$$

Isto completa a demonstração da proposição. □

A integral no enunciado da proposição pode ser calculada explicitamente: deixamos a cargo do leitor verificar que (use integração por partes e o fato de que $\sum_{j=1}^{\infty} 1/j^2 = \pi^2/6$)

$$h_\mu(G) = \int \log|G'|\, d\mu = \int_0^1 \frac{-2\log x\, dx}{(1+x)\log 2} = \frac{\pi^2}{6\log 2} \approx 5,46\ldots$$

Então, lembrando que (G, μ) é ergódico (Seção 3.2.4), segue do teorema de Shannon-McMillan-Breiman (Teorema 4.3.1) que

$$\lim_n -\frac{1}{n} \log \mu(\mathcal{P}^n(x)) = \frac{\pi^2}{6\log 2} \quad \text{para } \mu\text{-quase todo } x.$$

Como a medida μ é comparável com a medida de Lebesgue, a menos de um fator constante, isto quer dizer que

$$\operatorname{diam} \mathcal{P}^n(x) \approx e^{-\frac{\pi^2 n}{6\log 2}} \qquad (\text{a menos de um fator } e^{\pm\varepsilon n})$$

para μ-quase todo x e para n suficientemente grande. Observe que $\mathcal{P}^n(x)$ está formada pelos pontos y cuja expansão em fração contínua coincide com a expansão de x até a ordem n.

4.4.3 Endomorfismos lineares do toro

Dado um número real $x > 0$, denotamos $\log^+ x = \max\{\log x, 0\}$. Nesta seção provamos o seguinte resultado:

Proposição 4.4.3. *Seja $f_A : \mathbb{T}^d \to \mathbb{T}^d$ o endomorfismo induzido no toro $\mathbb{T}^d$ por alguma matriz invertível A com coeficientes inteiros. Seja μ a medida de Haar em $\mathbb{T}^d$. Então*

$$h_\mu(f_A) = \sum_{i=1}^{d} \log^+ |\lambda_i|$$

onde $\lambda_1, \ldots, \lambda_d$ são os autovalores de A, contados com multiplicidade.

Inicialmente, suponhamos que a matriz A é diagonalizável. Seja $v_1, \ldots, v_d$ uma base normada de $\mathbb{R}^d$ tal que $Av_i = \lambda_i v_i$ para cada i. Seja u o número de autovalores de A com valor absoluto estritamente maior que 1. Podemos supor que os autovalores estão numerados de tal forma que $|\lambda_i| > 1$ se, e somente se, $i \leq u$. Dado $x \in \mathbb{T}^d$, todo ponto y numa vizinhança de x pode ser escrito na forma

$$y = x + \sum_{i=1}^{d} t_i v_i$$

como $t_1, \ldots, t_d$ próximos de zero. Dado $\varepsilon > 0$, denotamos por $D(x, \varepsilon)$ o conjunto dos pontos y desta forma com $|t_i| < \varepsilon$ para todo $i = 1, \ldots, d$. Além disso, para cada $n \geq 1$, consideramos

$$D(x, n, \varepsilon) = \{y \in \mathbb{T}^d : f_A^j(y) \in D(f_A^j(x), \varepsilon) \text{ para todo } j = 0, \ldots, n-1\}.$$

Observe que $f_A^j(y) = f_A^j(x) + \sum_{i=1}^{d} t_i \lambda_i^j v_i$ para todo $n \geq 1$. Portanto,

$$D(x, n, \varepsilon) = \Big\{x + \sum_{i=1}^{d} t_i v_i : |\lambda_i^n t_i| < \varepsilon \text{ para } i \leq u \text{ e } |t_i| < \varepsilon \text{ para } i > u\Big\}.$$

Logo, existe uma constante $C_1 > 1$ que depende apenas de A, tal que

$$C_1^{-1} \varepsilon^d \prod_{i=1}^{u} |\lambda_i|^{-n} \leq \mu(D(x, n, \varepsilon)) \leq C_1 \varepsilon^d \prod_{i=1}^{u} |\lambda_i|^{-n}$$

para todo $x \in \mathbb{T}^d$, $n \geq 1$ e $\varepsilon > 0$. Também é claro que existe $C_2 > 1$ que depende apenas de A, tal que

$$B(x, C_2^{-1}\varepsilon) \subset D(x, \varepsilon) \subset B(x, C_2 \varepsilon)$$

para $x \in \mathbb{T}^d$ e $\varepsilon > 0$ pequeno. Então, $B(x, n, C_2^{-1}\varepsilon) \subset D(x, n, \varepsilon) \subset B(x, n, C_2\varepsilon)$ para todo $n \geq 1$. Combinando estas duas observações, e tomando $C = C_1 C_2^d$, obtemos que:

$$C^{-1} \varepsilon^d \prod_{i=1}^{u} |\lambda_i^{-n}| \leq \mu(B(x, n, \varepsilon)) \leq C \varepsilon^d \prod_{i=1}^{u} |\lambda_i^{-n}|$$

para todo $x \in \mathbb{T}^d$, $n \geq 1$ e $\varepsilon > 0$. Então,

$$h_\mu^+(f,\varepsilon,x) = h_\mu^-(f,\varepsilon,x) = \lim_n \frac{1}{n} \log \mu\big(B(x,n,\varepsilon)\big) = \sum_{i=1}^{u} \log|\lambda_i|$$

para todo $x \in \mathbb{T}^d$ e todo $\varepsilon > 0$ pequeno. Logo, usando o teorema de Brin-Katok (Teorema 4.3.3)

$$h_\mu(f) = h_\mu(f,x) = \sum_{i=1}^{u} \log|\lambda_i|$$

para μ-quase todo ponto x. Isto prova a Proposição 4.4.3 no caso diagonalizável.

O caso geral pode ser tratado de forma semelhante, escrevendo a matriz A na forma canônica de Jordan. Deixamos essa tarefa para o leitor (Exercício 4.4.18).

4.4.4 Exercícios

4.4.1. Prove que $H_\mu(\mathcal{P}/\mathcal{R}) \leq H_\mu(\mathcal{P}/\mathcal{Q}) + H_\mu(\mathcal{Q}/\mathcal{R})$ para quaisquer partições $\mathcal{P}$, $\mathcal{Q}$ e $\mathcal{R}$.

4.4.2. Mostre que o supremo de $h_\mu(f,\mathcal{P})$ sobre todas as partições finitas coincide com o supremo sobre todas as partições com entropia finita.

4.4.3. Verifique que $\lim_n H_\mu(\vee_{i=0}^{k-1} f^{-i}(\mathcal{P}) / \vee_{j=k}^{n} f^{-j}(\mathcal{P})) = kh(f,\mathcal{P})$ para toda partição $\mathcal{P}$ com entropia finita e todo $k \geq 1$.

4.4.4. Seja $f : M \to M$ uma transformação preservando uma medida de probabilidade μ.

(a) Suponha que existe um conjunto invariante $A \subset M$ com $\mu(A) \in (0,1)$. Sejam μ_A e μ_B as restrições normalizadas de μ aos conjuntos A e $B = A^c$, respectivamente. Mostre que $h_\mu(f) = \mu(A)h_{\mu_A}(f) + \mu(B)h_{\mu_B}(f)$.

(b) Suponha que μ é uma combinação convexa $\mu = \sum_{i=1}^n a_i\mu_i$ de medidas invariantes ergódicas $\mu_1, \ldots, \mu_n$. Mostre que $h_\mu(f) = \sum_{i=1}^n a_i h_{\mu_i}(f)$.

[Observação: Na Seção 4.5 obteremos resultados mais fortes.]

4.4.5. Sejam $(M,\mathcal{B},\mu)$ e $(N,\mathcal{C},\nu)$ espaços de probabilidade e $f : M \to M$ e $g : N \to N$ transformações mensuráveis preservando as medidas μ e ν, respectivamente. Dizemos que (g,ν) é um *fator* de (f,μ) se existe uma aplicação mensurável, não necessariamente invertível, $\phi : (M,\mathcal{B}) \to (N,\mathcal{C})$ tal que $\phi_*\mu = \nu$ e $\phi \circ f = g \circ \phi$ em quase todo ponto. Mostre que nesse caso $h_\nu(g) \leq h_\mu(f)$.

4.4.6. Mostre que a expansão decimal $f : [0,1] \to [0,1]$, $f(x) = 10x - [10x]$ é expansiva e exiba uma constante de expansividade.

4.4.7. Verifique que para todo $s > 0$ existe algum deslocamento de Bernoulli (σ,μ) cuja entropia é igual a s.

4.4.8. Seja $X = \{0\} \cup \{1/n : n \geq 1\}$ e considere o espaço $\Sigma = X^{\mathbb{N}}$ munido com a distância $d((x_n)_n, (y_n)_n) = 2^{-N}|x_N - y_N|$, onde $N = \min\{n \in \mathbb{N} : x_n \neq y_n\}$.

(a) Certifique-se de que o deslocamento $\sigma : \Sigma \to \Sigma$ não é expansivo.

(b) Para cada $k \geq 1$, seja ν_k a probabilidade em X que dá peso $1/2$ a cada um dos pontos $1/k$ e $1/(k+1)$. Use as medidas de Bernoulli $\mu_k = \nu_k^{\mathbb{N}}$ para concluir que a função entropia do deslocamento não é semicontínua superiormente.

(c) Seja μ a medida de Bernoulli associada a um vetor de probabilidade $(p_x)_{x \in X}$ tal que $\sum_{x \in X} -p_x \log p_x = \infty$. Mostre que $h_\mu(\sigma)$ é infinita.

4.4.9. Seja $f : S^1 \to S^1$ uma aplicação de recobrimento de grau $d \geq 2$ e seja μ uma probabilidade invariante por f. Mostre que $h_\mu(f) \leq \log d$.

4.4.10. Sejam $\mathcal{P}$ e $\mathcal{Q}$ duas partições com entropia finita. Mostre que se $\mathcal{P}$ é menos fina que $\vee_{j=0}^{\infty} f^{-j}(\mathcal{Q})$ então $h_\mu(f, \mathcal{P}) \leq h_\mu(f, \mathcal{Q})$.

4.4.11. Mostre que se $\mathcal{A}$ é uma álgebra geradora da σ-álgebra dos conjuntos mensuráveis, a menos de medida nula, então o supremo de $h_\mu(f, \mathcal{P})$ sobre as partições com entropia finita (ou mesmo sobre as partições finitas) $\mathcal{P} \subset \mathcal{A}$ coincide com $h_\mu(f)$.

4.4.12. Considere transformações $f : M \to M$ e $g : N \to N$ preservando medidas de probabilidade μ e ν, respectivamente. Considere $f \times g : M \times N \to M \times N$, dada por $(f \times g)(x, y) = (f(x), g(y))$. Mostre que $f \times g$ preserva a medida produto $\mu \times \nu$ e que $h_{\mu \times \nu}(f \times g) = h_\mu(f) + h_\nu(g)$.

4.4.13. Verifique que, para qualquer função mensurável $\varphi : M \to (0, \infty)$,

$$\int \varphi \, d\mu = \int_0^\infty \mu(\{x \in M : \varphi(x) > t\}) \, dt.$$

4.4.14. Use o Teorema 4.3.1 para calcular a entropia de um deslocamento de Bernoulli em $\Sigma = \{1, \ldots, d\}^{\mathbb{N}}$.

4.4.15. Mostre que a função $h_\mu(f, x)$ no Teorema 4.3.3 é f-invariante. Conclua que se (f, μ) é ergódico, então $h_\mu(f) = h_\mu(f, x)$ para μ-quase todo ponto x.

4.4.16. Suponha que (f, μ) é ergódico e seja $\mathcal{P}$ uma partição com entropia finita. Mostre que dado $\varepsilon > 0$ existe $k \geq 1$ tal que para todo $n \geq k$ existe $\mathcal{B}_n \subset \mathcal{P}^n$ tal que a união dos elementos de $\mathcal{B}_n$ tem medida maior que $1 - \varepsilon$ e

$$e^{-n(h_\mu(f,\mathcal{P})+\varepsilon)} < \mu(B) < e^{-n(h_\mu(f,\mathcal{P})-\varepsilon)} \quad \text{para todo } B \in \mathcal{B}_n.$$

4.4.17. Mostre que toda rotação $R_\theta : \mathbb{T}^d \to \mathbb{T}^d$ tem entropia nula relativamente à medida de Haar no toro $\mathbb{T}^d$. [Observação: Isto é um caso particular do Exemplo 4.3.4 mas para o enunciado presente não precisamos usar o teorema de Brin-Katok.]

4.4.18. Complete a demonstração da Proposição 4.4.3.

4.4.19. Seja $f : M \to M$ uma transformação mensurável e seja μ uma probabilidade invariante ergódica em M. Seja $B \subset M$ um conjunto mensurável com $\mu(B) > 0$, seja $g : B \to B$ a transformação de primeiro retorno a B e seja ν a restrição normalizada de μ ao conjunto B . Mostre que $h_\mu(f) = \nu(B)h_\nu(g)$.

4.4.20. Seja $f : M \to M$ uma transformação preservando medida num espaço de Lebesgue (M, μ). Seja $\hat{f} : \hat{M} \to \hat{M}$ a extensão natural de f e seja $\hat{\mu}$ o levantamento de μ. Mostre que $h_\mu(f) = h_{\hat{\mu}}(\hat{f})$.

4.4.21. Mostre que se f é o tempo-1 de um fluxo diferenciável numa superfície M então $h_\mu(f) = 0$ para toda medida invariante ergódica μ. [Observação: Usiando o Teorema 4.5.2 abaixo, segue que a entropia é zero para toda medida de probabilidade invariante.]

4.4.22. Verifique que a definição de jacobiano não depende da escolha da cobertura $\{U_k : k \geq 1\}$ por domínios de invertibilidade.

4.4.23. Seja $\sigma : \Sigma \to \Sigma$ a transformação deslocamento em $\Sigma = \{1, 2, \dots, d\}^{\mathbb{N}}$ e seja μ a medida de Markov associada a uma matriz aperiódica P. Encontre o jacobiano de f com relação a μ.

4.4.24. Seja $f : M \to M$ uma transformação localmente invertível e seja η uma probabilidade em M não singular com relação a f. Mostre que para toda função mensurável limitada $\psi : M \to \mathbb{R}$,

$$\int \psi \, d\eta = \int \sum_{z \in f^{-1}(x)} \frac{\psi}{J_\eta f}(z) d\eta(x).$$

4.4.25. Seja $f : M \to M$ uma transformação localmente invertível e seja η uma probabilidade em M não singular com relação a f. Mostre que η é invariante por f se, e somente se,

$$\sum_{z \in f^{-1}(x)} \frac{1}{J_\eta f(z)} = 1 \quad \text{para } \eta\text{-quase todo } x \in M.$$

Além disso se η é invariante então $J_\eta f \geq 1$ em μ-quase todo ponto.

4.4.26. Seja $f : M \to M$ uma transformação localmente invertível e seja η uma probabilidade boreliana em M não singular com relação a f. Mostre que, para todo $k \geq 1$, existe jacobiano de f^k com relação a η e ele é dado por

$$J_\eta f^j(x) = \prod_{j=0}^{k-1} J_\eta f(f^j(x)) \quad \text{para } \eta\text{-quase todo } x.$$

Supondo que f é invertível, o que pode ser dito a respeito do jacobiano de f^{-1} relativamente a η?

4.4.27. Sejam $f : M \to M$ e $g : N \to N$ transformações localmente invertíveis e sejam μ e ν medidas de probabilidade invariantes por f e g, respectivamente. Suponha que existe uma equivalência ergódica $\phi : M \to N$ entre os sistemas (f, μ) e (g, ν). Mostre que $J_\mu f = J_\nu g \circ \phi$ em μ-quase todo ponto.

4.4.28. Sejam $\sigma_k : \Sigma_k \to \Sigma_k$ e $\sigma_l : \Sigma_l \to \Sigma_l$ as transformações deslocamento em $\Sigma_k = \{1, \dots, k\}^{\mathbb{N}}$ e $\Sigma_l = \{1, \dots, l\}^{\mathbb{N}}$. Sejam μ_k e μ_l medidas de Bernoulli em Σ_k e Σ_l, respectivamente, associadas a vetores de probabilidade $p = (p_1, \dots, p_k)$ e $q = (q_1, \dots, q_l)$. Mostre que os sistemas (σ_k, μ_k) e (σ_l, μ_l) são ergodicamente equivalentes se, e somente se, $k = l$ e os vetores p e q se obtêm um do outro por permutação das componentes.

4.4.29. Seja μ uma probabilidade invariante por uma transformação localmente invertível $f : M \to M$. Mostre que existe $N \subset M$ com medida total tal que $N \subset f^{-1}(N)$ e μ restrita a N é não singular com relação à restrição $f : N \to N$. Conclua que f admite jacobiano com relação a μ.

4.5 Entropia e decomposição ergódica

Não é difícil mostrar que a entropia $h_\mu(f)$ é sempre uma função *afim* da medida invariante μ:

Proposição 4.5.1. *Sejam μ e ν probabilidades invariantes por uma transformação $f : M \to M$. Então $h_{t\mu+(1-t)\nu}(f) = th_\mu(f) + (1-t)h_\nu(f)$ para todo $0 < t < 1$.*

Demonstração. Defina $\phi(x) = -x \log x$, para $x > 0$. Por um lado, como a função ϕ é côncava,

$$\phi(t\mu(B) + (1-t)\nu(B)) \geq t\phi(\mu(B)) + (1-t)\phi(\nu(B))$$

para todo conjunto mensurável $B \subset M$. Por outro lado, dado qualquer conjunto mensurável $B \subset M$,

$$\begin{aligned}
&\phi\big(t\mu(B) + (1-t)\nu(B)\big) - t\phi\big(\mu(B)\big) - (1-t)\phi\big(\nu(B)\big) \\
&\quad = -t\mu(B)\log\frac{t\mu(B) + (1-t)\nu(B)}{\mu(B)} - (1-t)\nu(B)\log\frac{t\mu(B) + (1-t)\nu(B)}{\nu(B)} \\
&\quad \leq -t\mu(B)\log t - (1-t)\nu(B)\log(1-t)
\end{aligned}$$

porque a função $-\log$ é decrescente. Portanto, dada qualquer partição $\mathcal{P}$ com entropia finita,

$$\begin{aligned}
&H_{t\mu+(1-t)\nu}(\mathcal{P}) \geq tH_\mu(\mathcal{P}) + (1-t)H_\nu(\mathcal{P}) \quad \text{e} \\
&H_{t\mu+(1-t)\nu}(\mathcal{P}) \leq tH_\mu(\mathcal{P}) + (1-t)H_\nu(\mathcal{P}) - t\log t - (1-t)\log(1-t).
\end{aligned}$$

Consequentemente,

$$h_{t\mu+(1-t)\nu}(f, \mathcal{P}) = th_\mu(f, \mathcal{P}) + (1-t)h_\nu(f, \mathcal{P}). \tag{4.5.1}$$

Segue, imediatamente, que $h_{t\mu+(1-t)\nu}(f) \le t h_\mu(f) + (1-t) h_\nu(f)$. Além disso, (4.1.16) e (4.5.1) implicam que

$$h_{t\mu+(1-t)\nu}(f, \mathcal{P}_1 \vee \mathcal{P}_2) \ge t h_\mu(f, \mathcal{P}_1) + (1-t) h_\nu(f, \mathcal{P}_2)$$

para quaisquer partições $\mathcal{P}_1$ e $\mathcal{P}_2$. Tomando o supremo em $\mathcal{P}_1$ e $\mathcal{P}_2$ obtemos a desigualdade pretendida: $h_{t\mu+(1-t)\nu}(f) \ge t h_\mu(f) + (1-t) h_\nu(f)$. $\square$

Em particular, dado qualquer conjunto invariante $A \subset M$, temos

$$h_\mu(f) = \mu(A) h_{\mu_A}(f) + \mu(A^c) h_{\mu_{A^c}}(f), \tag{4.5.2}$$

onde μ_A e μ_{A^c} denotam as restrições normalizadas de μ ao conjunto A e ao seu complementar, respectivamente (este fato já foi obtido no Exercício 4.4.4). Outra consequência imediata é a seguinte versão da Proposição 4.5.1 para combinações convexas finitas:

$$\mu = \sum_{i=1}^n t_i \mu_i \quad \Rightarrow \quad h_\mu(f) = \sum_{i=1}^n t_i h_{\mu_i}(f), \tag{4.5.3}$$

quaisquer que sejam as probabilidades invariantes $\mu_1, \dots, \mu_n$ e os números positivos $t_1, \dots, t_n$ satisfazendo $\sum_{i=1}^n t_i = 1$.

Um fato muito mais profundo, devido a Konrad Jacobs [Jac60, Jac63], é que a propriedade de afinidade se estende para a decomposição ergódica dada pelo Teorema de Decomposição Ergódica:

Teorema 4.5.2 (Jacobs)**.** *Suponha que M é um espaço métrico completo separável. Dada qualquer probabilidade invariante μ, seja $\{\mu_P : P \in \mathcal{P}\}$ a sua decomposição ergódica. Então $h_\mu(f) = \int h_{\mu_P}(f)\, d\hat{\mu}(P)$ (quando um dos lados da igualdade é infinito o outro também é).*

Isso é consequência de um teorema geral sobre funcionais afins no espaço das probabilidades. Para mais detalhes, sugerimos o Capítulo de Entropia Métrica de [eKO14].

4.6 Jacobianos e fórmula de Rokhlin

Sejam U um aberto e m a medida de Lebesgue de $\mathbb{R}^d$ e seja $f : U \to U$ um difeomorfismo local. Pela fórmula de mudança de variáveis,

$$m(f(A)) = \int_A |\det Df(x)|\, dx \tag{4.6.1}$$

para todo A contido numa bola restrita à qual f é injetivo. A noção de jacobiano, que vamos apresentar a seguir, estende este tipo de relação para transformações e medidas muito mais gerais. Além de introduzirmos este conceito, mostraremos que jacobianos existem sob hipóteses bastante gerais. Mais ainda, é possível exprimir a entropia do sistema explicitamente em termos do jacobiano. Já encontramos uma manifestação interessante desse fato, na Proposição 4.4.2.

Seja $f : M \to M$ uma transformação mensurável. Diremos que f é *localmente invertível* se existe algum cobertura enumerável $\{U_k : U_k \geq 1\}$ de M por conjuntos mensuráveis tais que a restrição de f a cada U_k é uma bijeção sobre a sua imagem, a qual é um conjunto mensurável, e a inversa dessa bijeção também é mensurável. Os subconjuntos mensuráveis destes conjuntos U_k serão chamados *domínios de invertibilidade*. Note que se A é domínio de invertibilidade então $f(A)$ é um conjunto mensurável. Observe, igualmente, que se f é localmente invertível então a pré-imagem $f^{-1}(y)$ de qualquer $y \in M$ é enumerável: ela contém no máximo um ponto em cada U_k.

Seja η uma probabilidade em M, não necessariamente invariante por f. Uma função mensurável $\xi : M \to [0, \infty)$ é um *jacobiano* de f relativamente a η se a restrição de ξ a qualquer domínio de invertibilidade A é integrável com relação a η e satisfaz

$$\eta(f(A)) = \int_A \xi \, d\eta. \tag{4.6.2}$$

Note (Exercício 4.4.22) que a definição não depende da escolha de $\{U_k : k \geq 1\}$.

Exemplo 4.6.1. Seja $\sigma : \Sigma \to \Sigma$ o deslocamento em $\Sigma = \{1, 2, \ldots, d\}^{\mathbb{N}}$ e seja μ a medida de Bernoulli associada a um vetor de probabilidade $p = (p_1, \ldots, p_d)$. A restrição de σ a cada cilindro $[0; a]$ é uma transformação invertível. Além disso, dado qualquer cilindro $[0; a, a_1, \ldots, a_n] \subset [0; a]$,

$$\mu\big(\sigma([0; a, a_1, \ldots, a_n])\big) = p_{a_1} \cdots p_{a_n} = \frac{1}{p_a} \mu\big([0; a, a_1, \ldots, a_n]\big).$$

Deixamos ao cuidado do leitor deduzir que $\mu\big(\sigma(A)\big) = (1/p_a)\mu(A)$ para todo conjunto mensurável $A \subset [0; a]$. Portanto, a função $\xi((x_n)_n) = 1/p_{x_0}$ é um jacobiano de σ relativamente a μ.

Dizemos que uma medida η é *não singular* com relação à transformação f se a imagem de qualquer domínio de invertibilidade com medida nula também tem medida nula: se $\eta(A) = 0$ então $\eta(f(A)) = 0$. Por exemplo, se $f : U \to U$ é um difeomorfismo local num aberto de $\mathbb{R}^d$ e η é a medida de Lebesgue, então η é não singular. Para qualquer transformação localmente invertível, toda probabilidade invariante é não singular com relação à restrição de f a algum subconjunto invariante com medida total (Exercício 4.4.29).

Segue imediatamente da definição (4.6.2) que se f admite jacobiano com relação a uma medida η então essa medida é não singular. Vamos mostrar que a recíproca também é verdadeira:

Proposição 4.6.2. *Seja $f : M \to M$ uma transformação localmente invertível e seja η uma medida em M, não singular com relação a f. Então, existe algum jacobiano de f com relação a η e ele é essencialmente único: dois jacobianos quaisquer coincidem em η-quase todo ponto.*

Demonstração. Comecemos por provar a existência. Dada uma cobertura enumerável $\{U_k : k \geq 1\}$ de M por domínios de invertibilidade de f, defina $P_1 = U_1$ e $P_k = U_k \setminus (U_1 \cup \cdots \cup U_{k-1})$ para cada $k > 1$. Então, $\mathcal{P} = \{P_k : k \geq 1\}$ é uma partição de M formada por domínios de invertibilidade. Para cada $P_k \in \mathcal{P}$, represente por η_k a medida definida em P_k por $\eta_k(A) = \eta(f(A))$. Em outras palavras, η_k é a imagem por $(f \mid P_k)^{-1}$ da medida η restrita a $f(P_k)$. A hipótese de que η é não singular implica que cada η_k é absolutamente contínua com relação a η restrita a P_k:

$$\eta(A) = 0 \quad \Rightarrow \quad \eta_k(A) = \eta(f(A)) = 0$$

para todo conjunto mensurável $A \subset P_k$. Seja $\xi_k = d\eta_k/d(\eta \mid P_k)$ a derivada de Radon-Nikodym (Teorema A.2.18). Então ξ_k é uma função definida em P_k, integrável com relação a η e satisfazendo

$$\eta(f(A)) = \eta_k(A) = \int_A \xi_k \, d\eta \tag{4.6.3}$$

para todo conjunto mensurável $A \subset P_k$. Considere a função $\xi : M \to [0, \infty)$ cuja restrição a cada $P_k \in \mathcal{P}$ está dada por ξ_k. Todo subconjunto de U_k pode ser escrito como união disjunta de subconjuntos de $P_1, \ldots, P_k$. Aplicando (4.6.3) a cada um desses subconjuntos e somando as respectivas igualdades, obtemos que

$$\eta(f(A)) = \int_A \xi \, d\eta \quad \text{para todo conjunto mensurável } A \subset U_k \text{ e } k \geq 1.$$

Isto prova que ξ é um jacobiano de f relativamente a η.

Agora suponha que ξ e ζ são jacobianos de f relativamente a η e que existe $B \subset M$ com $\eta(B) > 0$ tal que $\xi(x) \neq \zeta(x)$ para todo $x \in B$. A menos de substituir B por um subconjunto adequado, e permutar os papéis de ξ e ζ se necessário, podemos supor que $\xi(x) < \zeta(x)$ para todo $x \in B$. De modo similar, podemos supor que B está contido em algum U_k. Então,

$$\eta(f(B)) = \int_B \xi \, d\eta < \int_B \zeta \, d\eta = \eta(f(B)).$$

Esta contradição prova que o jacobiano é essencialmente único. □

A partir desta proposição, usaremos a notação $J_\eta f$ para representar o (essencialmente único) jacobiano de f com relação a η, quando exista. Por definição, $J_\eta f$ é integrável em cada domínio de invertibilidade. Se f é tal que o número de pré-imagens de qualquer $y \in M$ é limitado então o jacobiano é (globalmente) integrável: representando por ℓ o número máximo de pré-imagens,

$$\int J_\eta f \, d\eta = \sum_k \int_{P_k} J_\eta f \, d\eta = \sum_k \eta(f(P_k)) \leq \ell,$$

já que cada ponto $y \in M$ pertence a não mais que ℓ imagens $f(P_k)$.

A seguinte observação será útil na sequência. Seja $Z \subset M$ o conjunto dos pontos onde o jacobiano $J_\eta f$ se anula. Cobrindo Z com uma família enumerável de domínios de invertibilidade e usando (4.6.2) vemos que $f(Z)$ é um conjunto mensurável e $\eta(f(Z)) = 0$. Em outras palavras, o conjunto dos pontos $y \in M$ tais que $J_\mu f(x) > 0$ para todo $x \in f^{-1}(y)$ tem medida total para η. Se a probabilidade η é invariante por f, também segue que $\eta(f^{-1}(f(Z))) = \eta(f(Z)) = 0$ e, consequentemente, $\eta(Z) = 0$.

O principal resultado desta seção é a seguinte fórmula para a entropia de uma medida invariante:

Teorema 4.6.3 (fórmula de Rokhlin)**.** *Seja $f : M \to M$ uma transformação localmente invertível e seja μ uma probabilidade invariante por f. Suponha que existe alguma partição $\mathcal{P}$ com entropia finita tal que $\cup_n \mathcal{P}^n$ gera a σ-álgebra de M e todo $P \in \mathcal{P}$ é domínio de invertibilidade de f. Então $h_\mu(f) = \int \log J_\mu f \, d\mu$.*

Demonstração. Consideremos a sequência de partições $\mathcal{Q}_n = \vee_{j=1}^n f^{-j}(\mathcal{P})$. Pelo Corolário 4.2.5 e pelo Lema 4.1.12,

$$h_\mu(f) = h_\mu(f, \mathcal{P}) = \lim_n H_\mu(\mathcal{P}/\mathcal{Q}_n). \tag{4.6.4}$$

Por definição (como anteriormente, $\phi(x) = -x \log x$)

$$\begin{aligned} H_\mu(\mathcal{P}/\mathcal{Q}_n) &= \sum_{P \in \mathcal{P}} \sum_{Q_n \in \mathcal{Q}_n} -\mu(P \cap Q_n) \log \frac{\mu(P \cap Q_n)}{\mu(Q_n)} \\ &= \sum_{P \in \mathcal{P}} \sum_{Q_n \in \mathcal{Q}_n} \mu(Q_n) \phi\Big(\frac{\mu(P \cap Q_n)}{\mu(Q_n)}\Big). \end{aligned} \tag{4.6.5}$$

Seja $e_n(\psi, x)$ a esperança condicional de uma função ψ relativamente à partição $\mathcal{Q}_n$ e seja $e(\psi, x)$ o seu limite quando n vai para infinito. É claro da definição que

$$\frac{\mu(P \cap Q_n)}{\mu(Q_n)} = e_n(\mathcal{X}_P, x) \quad \text{para todo } x \in Q_n \text{ e todo } Q_n \in \mathcal{Q}_n.$$

Portanto,

$$\sum_{P \in \mathcal{P}} \sum_{Q_n \in \mathcal{Q}_n} \mu(Q_n) \phi\Big(\frac{\mu(P \cap Q_n)}{\mu(Q_n)}\Big) = \sum_{P \in \mathcal{P}} \int \phi(e_n(\mathcal{X}_P, x)) \, d\mu(x). \tag{4.6.6}$$

O limite $e(\mathcal{X}_P, x) = \lim_n e_n(\mathcal{X}_P, x)$ existe para μ-quase todo x. Então, observando que a função ϕ é limitada, podemos usar o teorema da convergência dominada para deduzir das relações (4.6.4) – (4.6.6) que

$$h_\mu(f) = \sum_{P \in \mathcal{P}} \int \phi(e(\mathcal{X}_P, x)) \, d\mu(x). \tag{4.6.7}$$

Resta relacionar o integrando do lado direito com o jacobiano. Isso será feito por meio do Lema 4.6.5. Comecemos por provar as seguintes *fórmulas de mudança de variáveis*:

Lema 4.6.4. *Para qualquer probabilidade η em M não singular com relação a f e qualquer domínio de invertibilidade $A \subset M$ de f:*

(a) $\int_{f(A)} \varphi \, d\eta = \int_A (\varphi \circ f) J_\eta f \, d\eta$ *para toda função mensurável* $\varphi : f(A) \to \mathbb{R}$ *tal que as integrais estão definidas (podendo ser $\pm\infty$).*

(b) $\int_A \psi \, d\eta = \int_{f(A)} (\psi / J_\eta f) \circ (f \mid A)^{-1} \, d\eta$ *para qualquer função mensurável* $\psi : A \to \mathbb{R}$ *tal que as integrais estão definidas (podendo ser $\pm\infty$).*

Demonstração. A definição (4.6.2) dá que a fórmula no item (a) vale para a função característica $\varphi = \mathcal{X}_{f(A)}$, qualquer que seja A. Segue que ela vale para a função característica de qualquer subconjunto de $f(A)$, uma vez que ele pode ser escrito como $f(B)$ para algum domínio de invertibilidade $B \subset A$. Logo, por linearidade, a igualdade se estende para toda função simples definida em $f(A)$. Usando o teorema da convergência monótona, concluímos que ela vale para qualquer função mensurável não negativa. Usando linearidade uma vez mais, obtemos o caso geral do item (a).

Para deduzir o item (b), aplique o item (a) à função $\varphi = (\psi/J_\eta f) \circ (f \mid A)^{-1}$. Note que esta função está bem definida em η-quase todo ponto pois, conforme observamos anteriormente, $J_\eta f(x) > 0$ para todo x na pré-imagem de η-quase todo $y \in M$. $\square$

Lema 4.6.5. *Para toda função mensurável limitada $\psi : M \to \mathbb{R}$ e toda probabilidade η invariante por f,*

$$e(\psi, x) = \hat{\psi}(f(x)) \quad \textit{para } \eta\textit{-quase todo } x, \quad \textit{onde} \quad \hat{\psi}(y) = \sum_{z \in f^{-1}(y)} \frac{\psi}{J_\eta f}(z).$$

Demonstração. Lembre que $\mathcal{Q}_n = \vee_{j=1}^{n} f^{-j}(\mathcal{P})$. Também usaremos a sequência de partições $\mathcal{P}^n = \vee_{j=0}^{n-1} f^{-j}(\mathcal{P})$. Observe que $\mathcal{Q}_n(x) = f^{-1}(\mathcal{P}^{n-1}(f(x)))$ e $\mathcal{P}^n(x) = \mathcal{P}(x) \cap \mathcal{Q}_n(x)$ para todo n e todo x. Então,

$$\int_{\mathcal{P}^{n-1}(f(x))} \hat{\psi} \, d\eta = \sum_{P \in \mathcal{P}} \int_{f(P) \cap \mathcal{P}^{n-1}(f(x))} \frac{\psi}{J_\eta f} \circ (f \mid P)^{-1} \, d\eta.$$

Usando a fórmula de mudança de variáveis dada no Lema 4.6.4(b), a expressão do lado direito pode ser reescrita como

$$\sum_{P \in \mathcal{P}} \int_{P \cap \mathcal{Q}_n(x)} \psi(z) \, d\eta(z) = \int_{\mathcal{Q}_n(x)} \psi \, d\eta.$$

Portanto,

$$\int_{\mathcal{P}^{n-1}(f(x))} \hat{\psi} \, d\eta = \int_{\mathcal{Q}_n(x)} \psi \, d\eta. \tag{4.6.8}$$

Seja $e'_{n-1}(\hat{\psi}, x)$ a esperança condicional de $\hat{\psi}$ relativamente à partição $\mathcal{P}^{n-1}$ e seja $e'(\hat{\psi}, x)$ o seu limite quando n vai para infinito. A hipótese de que η é invariante dá que $\eta\big(\mathcal{P}^{n-1}(f(x))\big) = \eta\big(\mathcal{Q}_n(x)\big)$. Dividindo ambos os lados de (4.6.8) por este número, obtemos que

$$e'_{n-1}(\hat{\psi}, f(x)) = e_n(\psi, x) \quad \text{para todo } x \text{ e todo } n > 1. \tag{4.6.9}$$

Então, passando ao limite, $e'(\hat{\psi}, f(x)) = e(\psi, x)$ para η-quase todo x. Por outro lado, a hipótese implica que $e'(\hat{\psi}, y) = \hat{\psi}(y)$ para η-quase todo $y \in M$. □

Vamos aplicar este resultado a $\psi = \mathcal{X}_P$ e $\eta = \mu$. Como f é injetiva em todo elemento de $\mathcal{P}$, cada interseção $P \cap f^{-1}(y)$ ou é vazia ou contém exatamente um ponto. Portanto, segue do Lema 4.6.5 que $e(\mathcal{X}_P, x) = \hat{\mathcal{X}}_P(f(x))$, com

$$\hat{\mathcal{X}}_P(y) = \begin{cases} 1/J_\mu f\big((f \mid P)^{-1}(y)\big) & \text{se } y \in f(P) \\ 0 & \text{se } y \notin f(P). \end{cases}$$

Então, lembrando que a medida μ é invariante,

$$\begin{aligned} \int \phi(e(\mathcal{X}_P, x))\, d\mu(x) &= \int \phi(\hat{\mathcal{X}}_P(y))\, d\mu(y) \\ &= \int_{f(P)} \big(\frac{1}{J_\mu f} \log J_\mu f\big) \circ (f \mid P)^{-1}\, d\mu \quad = \int_P \log J_\mu f\, d\mu \end{aligned}$$

(a última igualdade usa a igualdade (b) no Lema 4.6.4). Substituindo esta expressão em (4.6.7), vem que

$$h_\mu(f) = \sum_{P \in \mathcal{P}} \int_P \log J_\mu f\, d\mu = \int \log J_\mu f\, d\mu,$$

tal como afirmado no teorema. □

Apêndice A

Elementos de Medida, Topologia e Análise

Vamos recordar algumas noções e resultados básicos da Teoria da Medida, da Topologia e da Análise Funcional que serão úteis ao longo de todo o livro. Nossa intenção é proporcionar ao leitor uma fonte de referência rápida sobre medida e integração para uma leitura autocontida do texto. O material nestes apêndices não é totalmente sequencial: pode acontecer que uma noção mencionada num apêndice esteja definida, ou discutida com mais profundidade, apenas num apêndice subsequente (confira o índice remissivo).

De modo geral, omitiremos as demonstrações dos resultados apresentados. No que se refere aos Apêndices A.1 e A.2, o leitor poderá consultar os livros de Castro [Cas04], Fernandez [Fer02], Halmos [Hal50], Royden [Roy63] ou Rudin [Rud87].

A.1 Espaços de medida

Os espaços de medida constituem o ambiente natural para a definição da integral de Lebesgue, que será o tema principal da próxima seção. Aqui apresentaremos, de modo sucinto, os fundamentos da teoria desses espaços.

Começaremos por introduzir e estudar as noções de *álgebra* e *σ-álgebra* de conjuntos, as quais conduzem ao conceito de espaço mensurável. A seguir, apresentaremos o conceito de medida e analisaremos algumas de suas propriedades. Em particular, mencionaremos alguns resultados sobre construção de medidas, incluindo as medidas de Lebesgue em espaços euclideanos. A última subseção é dedicada às aplicações mensuráveis, ou seja, as aplicações entre espaços mensuráveis preservando a estrutura desses espaços.

A.1.1 Espaços mensuráveis

Dado um subconjunto $A \subset X$ denotaremos por A^c o complementar $X \setminus A$ do conjunto A em relação a X.

Definição A.1.1. Uma *álgebra* de X é uma família $\mathcal{B}$ de subconjuntos de X que é fechada para as operações elementares de conjuntos e contém o conjunto vazio, isto é, tal que

- $\emptyset \in \mathcal{B}$;
- $A \in \mathcal{B}$ implica $A^c \in \mathcal{B}$;
- $A \in \mathcal{B}$ e $B \in \mathcal{B}$ implica $A \cup B \in \mathcal{B}$.

Observe que $A \cap B = (A^c \cup B^c)^c$ e $A \setminus B = A \cap B^c$ também estão em $\mathcal{B}$, quaisquer que sejam $A, B \in \mathcal{B}$. Além disso, por associatividade, a união e a interseção de qualquer número finito de elementos de $\mathcal{B}$ também estão em $\mathcal{B}$.

Definição A.1.2. Uma álgebra diz-se uma *σ-álgebra* de subconjuntos de X se também for fechada para as uniões enumeráveis:

- $A_j \in \mathcal{B}$ para $j = 1, \ldots, n, \ldots$ implica $\displaystyle\bigcup_{j=1}^{\infty} A_j \in \mathcal{B}$.

Observe que uma σ-álgebra $\mathcal{B}$ também é fechada para as interseções enumeráveis: se $A_j \in \mathcal{B}$ para $j = 1, \ldots, n, \ldots$ então $\cap_{j=1}^{\infty} A_j = \left(\cup_{j=1}^{\infty} A_j^c\right)^c$ também está em $\mathcal{B}$.

Definição A.1.3. Um *espaço mensurável* é uma dupla $(X, \mathcal{B})$ onde X é um conjunto e $\mathcal{B}$ é uma σ-álgebra de subconjuntos de X. Os elementos de $\mathcal{B}$ são chamados *conjuntos mensuráveis* do espaço.

Em seguida apresentaremos algumas construções de σ-álgebras.

Exemplo A.1.4. Seja X um conjunto qualquer.

1. Denotemos por 2^X a família de todos os subconjuntos de X. Então $\mathcal{B} = 2^X$ é claramente uma σ-álgebra.
2. $\mathcal{B} = \{\emptyset, X\}$ é também uma σ-álgebra.

Note que se $\mathcal{B}$ é uma álgebra de X então $\{\emptyset, X\} \subset \mathcal{B} \subset 2^X$. Portanto $\{\emptyset, X\}$ é a menor e 2^X é a maior de todas as álgebras de subconjuntos de X.

No enunciado a seguir, $\mathcal{I}$ é um conjunto qualquer; ele serve apenas para indexar os elementos da família.

Proposição A.1.5. *Considere uma família não vazia qualquer $\{\mathcal{B}_i : i \in \mathcal{I}\}$ de σ-álgebras. Então a interseção $\mathcal{B} = \cap_{i \in \mathcal{I}} \mathcal{B}_i$ é também uma σ-álgebra.*

Agora, dado um conjunto qualquer $\mathcal{E}$ de subconjuntos de X, podemos aplicar a Proposição A.1.5 à família de todas as σ-álgebras que contêm $\mathcal{E}$. Note que esta família é não vazia, uma vez que contém a σ-álgebra 2^X, pelo menos. De acordo com a observação anterior, a interseção de todas estas σ-álgebras é também uma σ-álgebra, e é claro que contém $\mathcal{E}$. Além disso, do modo como é construída, ela está contida em todas as σ-álgebras que contêm $\mathcal{E}$. Portanto é a menor σ-álgebra que contém $\mathcal{E}$. Isto conduz à seguinte definição:

Definição A.1.6. A *σ-álgebra gerada* por uma família $\mathcal{E}$ de subconjuntos de X é a menor σ-álgebra $\sigma(\mathcal{E})$ que contém a família $\mathcal{E}$, ou seja, é a interseção de todas as σ-álgebras que contêm $\mathcal{E}$.

Lembremos que um *espaço topológico* é uma dupla (X, τ) em que X é um conjunto e τ é uma família de subconjuntos de X, contendo $\{\emptyset, X\}$ e fechada por interseções finitas e por uniões quaisquer. Essa família τ é chamada *topologia* e os seus elementos são chamados *abertos* de X. Neste livro consideraremos apenas espaços topológicos *de Hausdorff*, ou seja, tais que, para qualquer par de pontos distintos existe algum par de abertos disjuntos, cada um deles contendo exatamente um dos pontos.

No contexto dos espaços topológicos é natural considerar a construção que acabamos de descrever tomando $\mathcal{E} = \tau$. Isto nos conduz à seguinte noção:

Definição A.1.7. A *σ-álgebra de Borel* (ou *σ-álgebra boreliana*) de um espaço topológico é a σ-álgebra $\sigma(\tau)$ gerada pela topologia τ, isto é, a menor σ-álgebra que contém todos os subconjuntos abertos. Neste caso, os conjuntos mensuráveis recebem o nome de *borelianos*.

Observe que os subconjuntos *fechados* de X, ou seja, os complementares dos subconjuntos abertos, também pertencem à σ-álgebra de Borel.

Analogamente à Proposição A.1.5, a interseção de qualquer família não-vazia $\{\tau_i : i \in \mathcal{I}\}$ de topologias de um mesmo conjunto X é também uma topologia de X. Logo, pelo mesmo argumento que foi usado anteriormente para σ-álgebras, dada qualquer família $\mathcal{E}$ de subconjuntos de X existe uma topologia $\tau(\mathcal{E})$ que é a menor topologia que contém $\mathcal{E}$. Ela é chamada *topologia gerada* por $\mathcal{E}$.

Exemplo A.1.8. Seja $(X, \mathcal{B})$ um espaço mensurável. O *limite superior* de uma sequência de conjuntos $E_n \in \mathcal{B}$ é o conjunto $\limsup_n E_n$ formado pelos pontos $x \in X$ tais que $x \in E_n$ para infinitos valores de n. Analogamente, o *limite inferior* da sequência é o conjunto $\liminf_n E_n$ dos pontos $x \in X$ tais que $x \in E_n$ para todo valor de n suficientemente grande. Em outras palavras,

$$\liminf_n E_n = \bigcup_{n \geq 1} \bigcap_{m \geq n} E_m \quad \text{e} \quad \limsup_n E_n = \bigcap_{n \geq 1} \bigcup_{m \geq n} E_m.$$

Observe que $\liminf_n E_n \subset \limsup_n E_n$ e que ambos estão em $\mathcal{B}$.

Exemplo A.1.9. A *reta estendida* $\bar{\mathbb{R}} = [-\infty, \infty]$ é a união de $\mathbb{R} = (-\infty, +\infty)$ com os pontos no infinito $-\infty$ e $+\infty$. Este espaço possui uma topologia natural, gerada pelos intervalos $[-\infty, b)$ e $(a, +\infty]$, com $a, b \in \mathbb{R}$. É fácil ver que, munida desta topologia, a reta estendida é homeomorfa a um intervalo compacto da reta: por exemplo, a função $\arctan : \mathbb{R} \to (-\pi/2, \pi/2)$ se estende imediatamente a um *homeomorfismo* (ou seja, uma bijeção contínua cuja inversa também é contínua) entre $\bar{\mathbb{R}}$ e $[-\pi/2, \pi/2]$. Sempre consideramos na reta estendida a σ-álgebra de Borel associada a essa topologia.

Claro que a reta $\mathbb{R}$ é um subespaço (tanto topológico quanto mensurável) da reta estendida. Os borelianos da reta formam uma grande gama de subconjuntos e poderia até pensar-se que todo subconjunto de $\mathbb{R}$ fosse boreliano. No entanto, isso não é verdade: um contraexemplo será construído no Exercício A.1.4.

A.1.2 Espaços de medida

Seja $(X, \mathcal{B})$ um espaço mensurável. As seguintes noções têm um papel central neste livro:

Definição A.1.10. Uma *medida* em $(X, \mathcal{B})$ é uma função $\mu : \mathcal{B} \to [0, +\infty]$ tal que $\mu(\emptyset) = 0$ e

$$\mu(\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \mu(A_j)$$

para qualquer família enumerável de conjuntos $A_j \in \mathcal{B}$ disjuntos dois-a-dois. Esta última propriedade é chamada *σ-aditividade*. A tripla $(X, \mathcal{B}, \mu)$ é chamada *espaço de medida*. Quando vale $\mu(X) < \infty$ dizemos que μ é uma medida *finita* e se $\mu(X) = 1$ dizemos que μ é uma *probabilidade*. Neste último caso, $(X, \mathcal{B}, \mu)$ é um *espaço de probabilidade*.

Exemplo A.1.11. Seja X um conjunto e consideremos a σ-álgebra $\mathcal{B} = 2^X$. Dado qualquer $p \in X$, consideremos a função $\delta_p : 2^X \to [0, +\infty]$ definida por:

$$\delta_p(A) = \begin{cases} 1 & \text{se } p \in A \\ 0 & \text{se } p \notin A. \end{cases}$$

É fácil ver que δ_p é uma medida. Ela é usualmente designada *medida de Dirac* no ponto p.

Definição A.1.12. Dizemos que uma medida μ é *σ-finita* se existe uma família enumerável de subconjuntos $A_1, \ldots, A_n, \ldots$ de X tal que $\mu(A_i) < \infty$ para todo $i \in \mathbb{N}$ e

$$X = \bigcup_{i=1}^{\infty} A_i.$$

A segunda propriedade na definição de medida (Definição A.1.10) é chamada *σ-aditividade*. Dizemos que uma função $\mu : \mathcal{B} \to [0, +\infty]$ é *finitamente aditiva* se

$$\mu(\bigcup_{j=1}^{N} A_j) = \sum_{j=1}^{N} \mu(A_j)$$

para qualquer família finita $A_1, \dots, A_N \in \mathcal{B}$ de subconjuntos disjuntos dois-a-dois. Note que se μ é σ-aditiva então ela também é finitamente aditiva e que se μ é finitamente aditiva e não é constante igual a $+\infty$ então $\mu(\emptyset) = 0$.

A principal ferramenta para construir medidas é o seguinte teorema:

Teorema A.1.13 (Extensão)**.** *Seja $\mathcal{A}$ uma álgebra de subconjuntos de X e seja $\mu_0 : \mathcal{A} \to [0, +\infty]$ uma função σ-aditiva com $\mu_0(X) < \infty$. Então existe uma única medida μ definida na σ-álgebra $\mathcal{B}$ gerada por $\mathcal{A}$ que é uma extensão de μ_0, ou seja, tal que $\mu(A) = \mu_0(A)$ para todo $A \in \mathcal{A}$.*

O Teorema A.1.13 ainda vale para medidas σ-finitas. Além disso, existe uma versão dele para funções finitamente aditivas: se μ_0 é finitamente aditiva então ela admite extensão finitamente aditiva à σ-álgebra $\mathcal{B}$. Porém, neste caso tal extensão pode não ser única.

Em geral, ao tentarmos mostrar que uma função definida numa σ-álgebra é uma medida, o mais difícil é verificar a σ-aditividade. O critério mais usado para esse efeito é o seguinte:

Teorema A.1.14 (Continuidade no vazio)**.** *Seja $\mathcal{A}$ uma álgebra de subconjuntos de X e seja $\mu : \mathcal{A} \to [0, +\infty)$ uma função finitamente aditiva com $\mu(X) < \infty$. Então μ é σ-aditiva se, e somente se,*

$$\lim_n \mu\big(A_n\big) = 0 \tag{A.1.1}$$

para toda sequência $A_1 \supset \cdots \supset A_j \supset \cdots$ de elementos de $\mathcal{A}$ com $\cap_{j=1}^{\infty} A_j = \emptyset$.

A prova deste teorema é proposta no Exercício A.1.7. No Exercício A.1.9 propomos algumas variações do enunciado.

Definição A.1.15. Dizemos que uma álgebra $\mathcal{A}$ é *compacta* se qualquer sequência decrescente $A_1 \supset \cdots \supset A_n \supset \cdots$ de elementos não vazios de $\mathcal{A}$ tem interseção não vazia.

Lembre que um espaço topológico K é *compacto* se toda cobertura aberta, ou seja, toda família de abertos cuja união é todo o K, admite alguma subcobertura finita, ou seja, alguma subfamília finita cuja união ainda é todo o K. Dizemos que um subconjunto K de um espaço topológico X é compacto se a topologia de X restrita a K torna este último um espaço topológico compacto. Todo subconjunto fechado de um espaço compacto é compacto. Na direção oposta, se o espaço X é de Hausdorff então todo subconjunto compacto é fechado. Outro fato importante é que a interseção $\cap_n K_n$ de qualquer sequência decrescente $K_1 \supset \cdots \supset K_n \supset \cdots$ de subconjuntos compactos é não vazia.

Exemplo A.1.16. Suponha que X é um espaço topológico de Hausdorff e todo elemento de $\mathcal{A}$ é compacto. Segue imediatamente do que acabamos de dizer que a álgebra $\mathcal{A}$ é compacta.

Segue do Teorema A.1.14 que se $\mathcal{A}$ é álgebra compacta então toda função finitamente aditiva $\mu : \mathcal{A} \to [0, +\infty)$ satisfazendo $\mu(X) < \infty$ é σ-aditiva.

Outro resultado relacionado, que será útil para o nosso estudo, é o teorema das classes monótonas, que enunciaremos a seguir.

Definição A.1.17. Dizemos que uma família não vazia $\mathcal{C}$ de subconjuntos de X é uma *classe monótona*, se $\mathcal{C}$ contém X e é fechada para as uniões e interseções enumeráveis monótonas, ou seja:

- dados subconjuntos $A_1 \subset A_2 \subset \cdots$ em $\mathcal{C}$, então $\cup_{n\geq 1} A_n \in \mathcal{C}$ e
- dados subconjuntos $A_1 \supset A_2 \supset \cdots$ em $\mathcal{C}$, então $\cap_{n\geq 1} A_n \in \mathcal{C}$.

Claramente, as famílias $\{\emptyset, X\}$ e 2^X são classes monótonas. Além disso, se $\{\mathcal{C}_i : i \in \mathcal{I}\}$ é uma família qualquer de classes monótonas, então $\cap_{i\in\mathcal{I}}\mathcal{C}_i$ é uma classe monótona. Portanto, dado um subconjunto $\mathcal{A}$ de 2^X, podemos sempre considerar a menor classe monótona que contém $\mathcal{A}$.

Teorema A.1.18 (Classes monótonas)**.** *A menor classe monótona que contém uma álgebra $\mathcal{A}$ coincide com a σ-álgebra $\sigma(\mathcal{A})$ gerada por $\mathcal{A}$.*

Outro resultado importante sobre σ-álgebras, que nos será útil mais tarde, afirma que todo elemento B da σ-álgebra gerada por uma álgebra é aproximado por algum elemento A da álgebra, no sentido de que a medida da diferença simétrica

$$A\Delta B = (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$$

pode ser tão pequena quanto se queira:

Teorema A.1.19 (Aproximação)**.** *Seja $(X, \mathcal{B}, \mu)$ um espaço de probabilidade e seja $\mathcal{A}$ uma álgebra que gera a σ-álgebra $\mathcal{B}$. Então para todo $\varepsilon > 0$ e todo $B \in \mathcal{B}$ existe $A \in \mathcal{A}$ tal que $\mu(A\Delta B) < \varepsilon$.*

Definição A.1.20. Um espaço de medida diz-se *completo* se todo subconjunto de um conjunto mensurável com medida nula também é mensurável.

É possível transformar qualquer espaço de medida $(X, \mathcal{B}, \mu)$ num espaço completo, do seguinte modo. A família $\bar{\mathcal{B}}$ de todos os conjuntos $A \subset X$ tais que existem $B_1, B_2 \in \mathcal{B}$ com $B_1 \subset A \subset B_2$ e $\mu(B_2 \setminus B_1) = 0$ é uma σ-álgebra que contém $\mathcal{B}$. Considere $\bar{\mu} : \bar{\mathcal{B}} \to [0, +\infty]$ dada por $\bar{\mu}(A) = \mu(B_1) = \mu(B_2)$, para quaisquer $B_1, B_2 \in \mathcal{B}$ nessas condições. Esta função está bem definida e é uma medida em $\bar{\mathcal{B}}$, cuja restrição a $\mathcal{B}$ coincide com μ. Por construção, $(X, \bar{\mathcal{B}}, \bar{\mu})$ é um espaço de medida completo.

Dados subconjuntos $\mathcal{U}_1$ e $\mathcal{U}_2$ da σ-álgebra $\mathcal{B}$, dizemos que $\mathcal{U}_1 \subset \mathcal{U}_2$ *a menos de medida nula* se para todo $B_1 \in \mathcal{U}_1$ existe $B_2 \in \mathcal{U}_2$ tal que $\mu(B_1 \Delta B_2) = 0$. Dizemos que $\mathcal{U}_1 = \mathcal{U}_2$ a menos de medida nula se $\mathcal{U}_1 \subset \mathcal{U}_2$ a menos de medida nula e $\mathcal{U}_2 \subset \mathcal{U}_1$ a menos de medida nula. Dizemos que um conjunto $\mathcal{U} \subset \mathcal{B}$ gera a σ-álgebra $\mathcal{B}$ a menos de medida nula se a σ-álgebra gerada por $\mathcal{U}$ é igual a $\mathcal{B}$ a menos de medida nula. Equivalentemente, $\mathcal{U}$ gera $\mathcal{B}$ a menos de medida nula se o completamento (no sentido do parágrafo anterior) da σ-álgebra gerada por $\mathcal{U}$ coincide com o completamento de $\mathcal{B}$.

Por definição, uma medida toma valores em $[0, \infty]$; sempre que for necessário enfatizar esse fato falaremos de *medida positiva*. Mas também é possível enfraquecer essa exigência e, de fato, tais generalizações são úteis no texto.

Chamaremos *medida com sinal* num espaço mensurável $(X, \mathcal{B})$ a toda função σ-aditiva $\mu : \mathcal{B} \to [-\infty, \infty]$ tal que $\mu(\emptyset) = 0$. Mais precisamente, μ pode tomar o valor $-\infty$ ou o valor $+\infty$, mas não ambos; esta última restrição é para evitar a "indeterminação" $\infty - \infty$ na condição de aditividade.

Teorema A.1.21 (Decomposição de Hahn)**.** *Se μ é medida com sinal então existem conjuntos mensuráveis $P, N \subset X$ tais que $P \cup N = X$ e $P \cap N = \emptyset$ e*

$$\mu(E) \geq 0 \text{ para todo } E \subset P \quad \text{e} \quad \mu(E) \leq 0 \text{ para todo } E \subset N.$$

Isto quer dizer que podemos escrever $\mu = \mu^+ - \mu^-$ onde μ^+ e μ^- são as medidas (positivas) definidas por

$$\mu^+(E) = \mu(E \cap P) \quad \text{e} \quad \mu^-(E) = -\mu(E \cap N).$$

Em particular, a soma $|\mu| = \mu^+ + \mu^-$ também é uma medida; ela é chamada de *variação total* da medida com sinal μ.

Se μ só toma valores em $(-\infty, \infty)$ então dizemos que ela é uma medida com sinal *finita*. Neste caso as medidas μ^+ e μ^- são finitas. O conjunto $\mathcal{M}(X)$ das medidas com sinal finitas é um espaço vetorial real e a função $\|\mu\| = |\mu|(X)$ é uma norma completa neste espaço (veja o Exercício A.1.10). Em outras palavras, $(\mathcal{M}(X), \|\cdot\|)$ é um espaço de Banach. Quando X é um espaço métrico compacto, este espaço de Banach é isomorfo ao dual do espaço $C^0(X)$ das funções reais contínuas em X (teorema de Riesz-Markov).

Mais geralmente, chamaremos *medida complexa* num espaço mensurável $(X, \mathcal{B})$ a qualquer função σ-aditiva $\mu : \mathcal{B} \to \mathbb{C}$. Observe que $\mu(\emptyset)$ é necessariamente zero. É claro que podemos escrever $\mu = \Re\mu + i\Im\mu$, onde a parte real $\Re\mu$ e a parte imaginária $\Im\mu$ são medidas com sinal finitas. A *variação total* de μ é a medida finita definida por

$$|\mu|(E) = \sup_{\mathcal{P}} \sum_{P \in \mathcal{P}} |\mu(P)|,$$

onde o supremo é tomado sobre todas as partições finitas ou enumeráveis do conjunto mensurável E em subconjuntos mensuráveis (esta definição coincide com aquela que demos acima no caso particular em que μ é real). A função $\|\mu\| = |\mu|(X)$ define uma norma no espaço vetorial das medidas complexas em X, que ainda denotaremos por $\mathcal{M}(X)$, e esta norma é completa. Quando X é um espaço métrico compacto, o espaço de Banach $(\mathcal{M}(X), \|\cdot\|)$ é isomorfo ao dual do espaço $C^0(X)$ das funções complexas contínuas em X (teorema de Riesz-Markov)

A.1.3 Medida de Lebesgue

A medida de Lebesgue formaliza a noção de volume de subconjuntos do espaço euclideano $\mathbb{R}^d$. Ela é definida do seguinte modo.

Consideremos $X = [0, 1]$ e seja $\mathcal{A}$ a família de todos os subconjuntos da forma $A = I_1 \cup \cdots \cup I_N$ onde $I_1, \ldots, I_N$ são intervalos disjuntos dois-a-dois. É fácil ver que $\mathcal{A}$ é uma álgebra de subconjuntos de X. Além disso, temos uma função $m_0 : \mathcal{A} \to [0,1]$ definida nesta álgebra por

$$m_0\big(I_1 \cup \cdots \cup I_N\big) = |I_1| + \cdots + |I_N|,$$

onde $|I_j|$ representa o comprimento de cada intervalo I_j. Note que $m_0(X) = 1$. No Exercício A.1.8 propomos ao leitor mostrar que m_0 é σ-aditiva.

Note que a σ-álgebra $\mathcal{B}$ gerada por $\mathcal{A}$ coincide com a σ-álgebra de Borel de X, já que todo conjunto aberto pode ser escrito como união enumerável de intervalos abertos disjuntos dois-a-dois. Então, pelo Teorema A.1.13, existe uma única probabilidade m definida em $\mathcal{B}$ que é uma extensão de m_0. Chamamos m de *medida de Lebesgue* em $[0,1]$.

Mais geralmente, definimos medida de Lebesgue m no *cubo* $X = [0,1]^d$ de qualquer dimensão $d \geq 1$ da seguinte maneira. Primeiramente, chamamos *retângulo* em X qualquer subconjunto da forma $R = I_1 \times \cdots \times I_d$ onde os I_j são intervalos, e definimos

$$m_0(R) = |I_1| \times \cdots \times |I_d|.$$

Em seguida, consideramos a álgebra $\mathcal{A}$ dos subconjuntos de $[0,1]^d$ da forma $A = R_1 \cup \cdots \cup R_N$, onde $R_1, \ldots, R_N$ são retângulos disjuntos dois-a-dois, e definimos

$$m_0(A) = m_0(R_1) + \cdots + m_0(R_N)$$

para todo A nessa álgebra. A σ-álgebra gerada por $\mathcal{A}$ coincide com a σ-álgebra de Borel de X. A *medida de Lebesgue* em $X = [0,1]^d$ é a extensão de m_0 a essa σ-álgebra.

Para definir a medida de Lebesgue em todo o espaço euclidiano $\mathbb{R}^d$, decompomos esse espaço em cubos de lado unitário

$$\mathbb{R}^d = \bigcup_{k_1 \in \mathbb{Z}} \cdots \bigcup_{k_d \in \mathbb{Z}} [k_1, k_1+1) \times \cdots \times [k_d, k_d+1).$$

Cada cubo $[k_1, k_1+1) \times \cdots \times [k_d, k_d+1)$ pode ser identificado com $[0,1)^d$ por meio da translação $T_{k_1,\ldots,k_d}(x) = x - (k_1, \ldots, k_d)$ que envia o ponto $(k_1, k_2, \ldots, k_d)$ na origem. Isso nos permite definir uma medida $m_{k_1,k_2,\ldots,k_d}$ em C, dada por

$$m_{k_1,k_2,\ldots,k_d}(B) = m_0\big(T_{k_1,\ldots,k_d}(B)\big)$$

para todo o conjunto mensurável $B \subset C$. Finalmente, dado qualquer conjunto mensurável $B \subset \mathbb{R}^d$, definimos

$$m(B) = \sum_{k_1 \in \mathbb{Z}} \cdots \sum_{k_d \in \mathbb{Z}} m_{k_1,\ldots,k_d}\big(B \cap [k_1, k_1+1) \times \cdots \times [k_d, k_d+1)\big).$$

Note que m não é uma medida finita, mas é uma medida σ-finita.

Exemplo A.1.22. Vale a pena lembrar uma construção alternativa clássica da medida de Lebesgue. Para mais detalhes, veja o Capítulo 2 do livro de

Royden [Roy63]. Chamamos *medida exterior de Lebesgue* de qualquer conjunto $E \subset \mathbb{R}^d$ ao número

$$m^*(E) = \inf \Big\{ \sum_k m_0(R_k) : \\ (R_k)_k \text{ é cobertura enumerável de } E \text{ por retângulos abertos} \Big\}.$$

Esta função está definida para todo $E \subset \mathbb{R}^d$, mas não é aditiva (embora seja enumeravelmente subaditiva). Dizemos que E é *conjunto mensurável de Lebesgue* se

$$m^*(A) = m^*(A \cap E) + m^*(A \cap E^c) \quad \text{para todo } A \subset \mathbb{R}^d.$$

Todo retângulo R é conjunto mensurável de Lebesgue e satisfaz $m^*(R) = m_0(R)$. A família $\mathcal{M}$ de todos os conjuntos mensuráveis de Lebesgue é uma σ-álgebra e a restrição de m^* a $\mathcal{M}$ é uma medida (σ-aditiva). Pela observação anterior, $\mathcal{M}$ contém todo conjunto de Borel de $\mathbb{R}^d$. A restrição de m^* à σ-álgebra de Borel $\mathcal{B}$ de $\mathbb{R}^d$ coincide com a medida de Lebesgue em $\mathbb{R}^d$.

De fato, $\mathcal{M}$ é o completamento da σ-álgebra de Borel relativamente à medida de Lebesgue. Esta e outras propriedades estão contidas no Exercício A.1.13.

Exemplo A.1.23. Seja $\phi : [0,1] \to \mathbb{R}$ uma função contínua e positiva. Dado qualquer intervalo I, com extremos $0 \le a < b \le 1$, defina

$$\mu_\phi(I) = \int_a^b \phi(x)\,dx \qquad \text{(integral de Riemann).}$$

Em seguida, estenda a definição de μ_ϕ para a álgebra $\mathcal{A}$ das uniões finitas $A = I_1 \cup \cdots \cup I_k$ de intervalos disjuntos dois-a-dois, por meio da relação

$$\mu_\phi(A) = \sum_{j=1}^k \mu_\phi(I_j).$$

As propriedades básicas da integral de Riemann nos dizem que μ_ϕ é finitamente aditiva. Deixamos para o leitor a tarefa de mostrar que a medida μ_ϕ é σ-aditiva na álgebra formada pelas uniões finitas de intervalos (veja o Exercício A.1.7). Além disso, $\mu_\phi(\emptyset) = 0$ e $\mu_\phi([0,1]) < \infty$ já que ϕ é contínua e, portanto limitada. Com o auxílio do Teorema A.1.13 podemos estender μ_ϕ para toda σ-álgebra dos borelianos de $[0,1]$.

Observe que a medida μ_ϕ que acabamos de construir tem a seguinte propriedade especial: se um conjunto $A \subset [0,1]$ tem medida de Lebesgue zero então $\mu_\phi(A) = 0$. Essa propriedade chama-se *continuidade absoluta* (com respeito à medida de Lebesgue) e é discutida com mais detalhe no Apêndice A.2.4.

Vamos agora exibir uma medida que, apesar de ser positiva em qualquer aberto, não é absolutamente contínua com respeito à medida de Lebesgue:

Exemplo A.1.24. Considere uma enumeração $\{r_1, r_2, \dots\}$ do conjunto $\mathbb{Q}$ dos números racionais. Defina μ por:

$$\mu(A) = \sum_{r_i \in A} \frac{1}{2^i}.$$

Observe que a medida de qualquer aberto da reta é positiva, pois necessariamente A contém algum r_i. Apesar disso, a medida de $\mathbb{Q}$ é

$$\mu(\mathbb{Q}) = \sum_{r_i \in \mathbb{Q}} \frac{1}{2^i} = 1.$$

Como $\mathbb{Q}$ tem medida de Lebesgue nula (por ser um conjunto enumerável), isto mostra que μ não é absolutamente contínua com respeito à medida de Lebesgue.

Este exemplo nos motiva a introduzir a definição de suporte de uma medida num espaço topológico (X, τ). Para isso, precisamos recordar alguns conceitos básicos de Topologia.

Um subconjunto τ' da topologia τ é chamado *base de abertos*, ou *base da topologia*, se para todo $x \in X$ e todo aberto U contendo x existe $U' \in \tau'$ tal que $x \in U' \subset U$. Dizemos que o espaço topológico admite *base enumerável de abertos* se tal subconjunto τ' pode ser escolhido enumerável. Um conjunto $V \subset X$ é uma *vizinhança* de um ponto $x \in X$ se existe algum aberto U tal que $x \in U \subset V$. Reciprocamente, um subconjunto de X é aberto se, e somente se, ele é vizinhança de cada um dos seus pontos. Uma família υ' de subconjuntos de X é uma *base de vizinhanças* de um ponto $x \in X$ se para toda vizinhança V existe algum $V' \in \upsilon'$ tal que $x \in V' \subset V$. Dizemos que x admite *base enumerável de vizinhanças* se υ' pode ser escolhida enumerável. Se o espaço topológico admite base enumerável de abertos então todo $x \in X$ admite alguma base enumerável de vizinhanças de x, a saber, a família dos elementos da base enumerável de abertos que contêm x.

Definição A.1.25. Seja (X, τ) um espaço topológico e seja μ uma medida na σ-álgebra de Borel de X. O *suporte* da medida μ é o conjunto $\operatorname{supp} \mu$ formado pelos pontos $x \in X$ tais que $\mu(V) > 0$ para qualquer vizinhança V de x.

Segue imediatamente da definição que o suporte de uma medida é um conjunto fechado. No Exemplo A.1.24 acima, o suporte da medida μ é a reta inteira, apesar de que $\mu(\mathbb{Q}) = 1$.

Proposição A.1.26. *Sejam X um espaço topológico com base enumerável de abertos e μ uma medida não nula em X. Então, o suporte $\operatorname{supp} \mu$ é não vazio.*

Demonstração. Se $\operatorname{supp} \mu$ é vazio, então para cada ponto $x \in X$ podemos encontrar uma vizinhança aberta V_x tal que $\mu(V_x) = 0$. Seja $\{A_j : j = 1, 2, \dots\}$ uma base enumerável da topologia de X. Então para cada x podemos escolher $i(x) \in \mathbb{N}$ tal que $x \in A_{i(x)} \subset V_x$. Logo,

$$X = \bigcup_{x \in X} V_x = \bigcup_{x \in X} A_{i(x)}$$

e portanto

$$\mu(X) = \mu\big(\bigcup_{x\in X} A_{i(x)}\big) \le \sum_{i=1}^{\infty} \mu(A_i) = 0.$$

Isto é uma contradição e, portanto, $\operatorname{supp}\mu$ não pode ser vazio. □

A.1.4 Aplicações mensuráveis

Aplicações mensuráveis têm um papel em Teoria da Medida análogo ao das aplicações contínuas em Topologia: mensurabilidade corresponde à ideia de que a aplicação preserva a família dos conjuntos mensuráveis, do mesmo modo que continuidade significa que a família dos subconjuntos abertos é preservada pela aplicação.

Definição A.1.27. Dados espaços mensuráveis $(X, \mathcal{B})$ e $(Y, \mathcal{C})$, dizemos que uma aplicação $f : X \to Y$ é *mensurável* se $f^{-1}(C) \in \mathcal{B}$ para todo $C \in \mathcal{C}$.

Em geral, o conjunto dos $C \in \mathcal{C}$ tais que $f^{-1}(C) \in \mathcal{B}$ é uma σ-álgebra. Logo, para provar que f é mensurável basta mostrar que $f^{-1}(C_0) \in \mathcal{B}$ para todo C_0 em algum família de conjuntos $\mathcal{C}_0 \subset \mathcal{C}$ que gera a σ-álgebra $\mathcal{C}$. Veja também o Exercício A.1.1.

Exemplo A.1.28. Uma função $f : X \to [-\infty, \infty]$ é mensurável se, e somente se, o conjunto $f^{-1}((c, +\infty])$ pertence a $\mathcal{B}$ para todo $c \in \mathbb{R}$. Isto segue da observação anterior, uma vez que os intervalos $(c, +\infty]$ geram a σ-álgebra de Borel da reta estendida (lembre do Exemplo A.1.9). Em particular, se a função f toma valores em $(-\infty, +\infty)$ então ela é mensurável se, e somente se, $f^{-1}((c, +\infty))$ pertence a $\mathcal{B}$ para todo $c \in \mathbb{R}$.

Exemplo A.1.29. Se X é um espaço topológico e $\mathcal{B}$ é a sua σ-álgebra de Borel, então toda função contínua $f : X \to \mathbb{R}$ é mensurável. De fato, continuidade significa que a pré-imagem de qualquer aberto de $\mathbb{R}$ é um aberto de X e, portanto, está em $\mathcal{B}$. Como os abertos geram a σ-álgebra de Borel de $\mathbb{R}$, segue que a pré-imagem de qualquer boreliano da reta também está em $\mathcal{B}$.

Exemplo A.1.30. Dado um conjunto $B \subset X$ definimos a *função característica* $\mathcal{X}_B : X \to \mathbb{R}$ de B por:

$$\mathcal{X}_B(x) = \begin{cases} 1, & \text{se } x \in B; \\ 0, & \text{caso contrário.} \end{cases}$$

Observe que a função $\mathcal{X}_B$ é mensurável se, e somente se, B é um subconjunto mensurável: de fato, $\mathcal{X}_B^{-1}(A) \in \{\emptyset, B, X \setminus B, X\}$ para qualquer $A \subset \mathbb{R}$.

Entre as propriedades básicas das funções mensuráveis temos:

Proposição A.1.31. *Sejam $f, g : X \to [-\infty, +\infty]$ funções mensuráveis e sejam $a, b \in \mathbb{R}$. Então também são mensuráveis as seguintes funções:*

$$(af + bg)(x) = af(x) + bg(x) \quad e \quad (f \cdot g)(x) = f(x) \cdot g(x).$$

Além disso, se $f_n : X \to [-\infty, +\infty]$ é uma sequência de funções mensuráveis, também são mensuráveis as seguintes funções:

$$s(x) = \sup\{f_n(x) : n \geq 1\} \quad e \quad i(x) = \inf\{f_n(x) : n \geq 1\},$$

$$f^*(x) = \limsup_n f_n(x) \quad e \quad f_*(x) = \liminf_n f_n(x).$$

Em particular, se $f(x) = \lim f_n(x)$ existe então f é mensurável.

As combinações lineares de funções características formam uma classe importante de funções mensuráveis:

Definição A.1.32. Dizemos que uma função $s : X \to \mathbb{R}$ é *simples* se existem constantes $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ e conjuntos mensuráveis $A_1, \dots, A_k \in \mathcal{B}$ disjuntos dois-a-dois tais que

$$s = \sum_{j=1}^{k} \alpha_j \mathcal{X}_{A_j}, \tag{A.1.2}$$

onde $\mathcal{X}_A$ é a função característica do conjunto A.

Note que toda função simples é mensurável. Na direção recíproca, o próximo resultado afirma que toda função mensurável é limite de alguma sequência de funções simples. Este fato será muito útil na próxima seção, quando definirmos a integral de Lebesgue.

Proposição A.1.33. *Seja $f : X \to [-\infty, +\infty]$ uma função mensurável. Então existe uma sequência $(s_n)_n$ de funções simples tal que $|s_n(x)| \leq |f(x)|$ para todo n e*

$$\lim_n s_n(x) = f(x) \text{ para todo } x \in X.$$

Se f é limitada, a sequência pode ser escolhida tal que a convergência seja uniforme. Se f é não negativa, podemos tomar $0 \leq s_1 \leq s_2 \leq \cdots \leq f$.

A demonstração desta proposição está proposta ao leitor no Exercício A.1.16.

A.1.5 Exercícios

A.1.1. Seja X um conjunto e $(Y, \mathcal{C})$ um espaço mensurável. Mostre que, dada qualquer transformação $f : X \to Y$, existe alguma σ-álgebra $\mathcal{B}$ em X tal que a transformação é mensurável relativamente às σ-álgebras $\mathcal{B}$ e $\mathcal{C}$.

A.1.2. Seja X um conjunto e considere a família de subconjuntos

$$\mathcal{B}_0 = \{A \subset X : A \text{ é finito ou } A^c \text{ é finito}\}.$$

Mostre que $\mathcal{B}_0$ é uma álgebra e que $\mathcal{B}_0$ é uma σ-álgebra se, e somente se, o conjunto X é finito. Mostre também que, em geral,

$$\mathcal{B}_1 = \{A \subset X : A \text{ é finito ou enumerável ou } A^c \text{ é finito ou enumerável}\}$$

é a σ-álgebra gerada pela álgebra $\mathcal{B}_0$.

A.1.3. Prove a Proposição A.1.5.

A.1.4. O objetivo deste exercício é exibir um subconjunto da reta que não é boreliano. Seja α um número irracional qualquer. Defina em $\mathbb{R}$ a seguinte relação: $x \sim y \Leftrightarrow$ existem $m, n \in \mathbb{Z}$ tais que $x - y = m + n\alpha$. Verifique que $\sim$ é uma relação de equivalência e toda classe de equivalência intersecta $[0,1)$. Seja E_0 um subconjunto de $[0,1)$ contendo exatamente um elemento de cada classe de equivalência (a existência de tal conjunto é consequência do axioma da escolha). Mostre que E_0 não é boreliano.

A.1.5. Seja $(X, \mathcal{B}, \mu)$ um espaço de medida. Mostre que se A_1,A_2, ... estão em $\mathcal{B}$ então

$$\mu(\bigcup_{j=1}^{\infty} A_j) \leq \sum_{j=1}^{\infty} \mu(A_j).$$

A.1.6 (Lema de Borel-Cantelli)**.** Seja $(E_n)_n$ uma família enumerável de conjuntos mensuráveis. Seja F o conjunto dos pontos que pertencem a E_n para infinitos valores de n, ou seja, $F = \limsup_n E_n = \cap_{k=1}^{\infty} \cup_{n=k}^{\infty} E_n$. Mostre que se $\sum_n \mu(E_n) < \infty$ então $\mu(F) = 0$.

A.1.7. Demonstre o Teorema A.1.14.

A.1.8. Seja $\mathcal{A}$ a coleção dos subconjuntos de $X = [0,1]$ que se escrevem como união finita de intervalos disjuntos. Verifique que $\mathcal{A}$ é uma álgebra de subconjuntos de X. Seja $m_0 : \mathcal{A} \to [0,1]$ a função definida nesta álgebra por

$$m_0\big(I_1 \cup \cdots \cup I_N\big) = |I_1| + \cdots + |I_N|,$$

onde $|I_j|$ representa o comprimento de I_j. Mostre que m_0 é σ-aditiva.

A.1.9. Seja $\mathcal{B}$ uma álgebra de subconjuntos de X e seja $\mu : \mathcal{B} \to [0, +\infty)$ uma função finitamente aditiva com $\mu(X) < \infty$. Mostre que μ é σ-aditiva se, e somente se, vale qualquer uma das condições abaixo:

1. $\lim_n \mu(A_n) = \mu(\cap_{j=1}^{\infty} A_j)$ para toda sequência $A_1 \supset \cdots \supset A_j \supset \cdots$ de elementos de $\mathcal{B}$;

2. $\lim_n \mu(A_n) = \mu(\cup_{j=1}^{\infty} A_j)$ para toda sequência $A_1 \subset \cdots \subset A_j \subset \cdots$ de elementos de $\mathcal{B}$.

A.1.10. Mostre que $\|\mu\| = |\mu|(X)$ define uma norma completa no espaço vetorial das medidas com sinal finitas num espaço mensurável $(X, \mathcal{B})$.

A.1.11. Seja $X = \{1, \ldots, d\}$ um conjunto finito, munido da topologia discreta, e seja $M = X^{\mathcal{I}}$ onde $\mathcal{I} = \mathbb{N}$ ou $\mathcal{I} = \mathbb{Z}$.

(a) Verifique que a igualdade (A.2.7) define uma distância em M e que a topologia que ela define coincide com a topologia produto em M. Descreva as bolas abertas e as bolas fechadas num ponto $x \in X^{\mathcal{I}}$.

(b) Mostre, sem usar o teorema de Tychonoff, que (M, d) é um espaço compacto.

(c) Seja $\mathcal{A}$ a álgebra gerada pelos cilindros elementares de M. Mostre que toda função aditiva $\mu : \mathcal{A} \to [0,1]$ com $\mu(M) = 1$ se estende a uma medida (σ-aditiva) de probabilidade na σ-álgebra boreliana de M.

A.1.12. Seja $K \subset [0,1]$ o *conjunto de Cantor*, isto é, $K = \cap_{n=0}^{\infty} K_n$ onde $K_0 = [0,1]$ e cada K_n é o conjunto obtido quando retiramos de cada componente conexa C de K_{n-1} o intervalo aberto cujo comprimento é um terço do comprimento de C e cujo centro coincide com o centro de C. Mostre que K tem medida de Lebesgue igual a zero.

A.1.13. Dado um conjunto $E \subset \mathbb{R}^d$, mostre que as seguintes condições são equivalentes:

(a) E é um subconjunto mensurável de Lebesgue.

(b) E está no completamento da σ-álgebra de Borel relativamente à medida de Lebesgue, ou seja, existem conjuntos borelianos $B_1, B_2 \subset \mathbb{R}^d$ tais que $B_1 \subset E \subset B_2$ e $m(B_2 \setminus B_1) = 0$.

(c) (Aproximação por cima por abertos) Para todo $\varepsilon > 0$ podemos encontrar um aberto A com $E \subset A$ e $m^*(A \setminus E) < \varepsilon$.

(d) (Aproximação por baixo por fechados) Para todo $\varepsilon > 0$ podemos encontrar um fechado F com $F \subset E$ e $m^*(E \setminus F) < \varepsilon$.

A.1.14. Prove a Proposição A.1.31.

A.1.15. Seja $g_n : M \to \mathbb{R}$, $n \geq 1$ uma sequência de funções mensuráveis tais que $f(x) = \sum_{n=1}^{\infty} g_n(x)$ converge em todo ponto. Mostre que a soma f é uma função mensurável.

A.1.16. Prove a Proposição A.1.33.

A.1.17. Seja $f : X \to X$ uma transformação mensurável e seja ν uma medida em X. Defina $(f_*\nu)(A) = \nu(f^{-1}(A))$. Mostre que $f_*\nu$ é uma medida e note que ela é finita se, e somente se, ν é finita.

A.1.18. Seja $\omega_5 : [0,1] \to [0,1]$ a função que associa a cada número $x \in [0,1]$ a *frequência superior de dígitos iguais 5* na expansão de x na base 10. Isto é, escrevendo $x = 0, a_0 a_1 a_2 \ldots$... com $a_i \neq 9$ para infinitos valores de i,

$$\omega_5(x) = \limsup_n \frac{1}{n} \#\{0 \leq j \leq n-1 : a_j = 5\}.$$

Prove que a função ω_5 é mensurável.

A.2 Integração em espaços de medida

Nesta seção definiremos a integral de Lebesgue de uma função em relação a uma medida. Ela generaliza a noção de integral de Riemann que é normalmente apresentada nos cursos de Cálculo ou em cursos introdutórios de Análise.

A motivação para fazermos esta generalização é que a integral de Riemann não está definida para muitas funções úteis, por exemplo, para funções características de conjuntos mensuráveis em geral (veja o Exemplo A.2.5 abaixo). Já a integral de Lebesgue faz sentido em toda a classe das funções mensuráveis, a qual, como vimos na Proposição A.1.31, é fechada para as principais operações da Análise.

Ainda nesta seção, enunciaremos alguns resultados importantes sobre o comportamento da integral relativamente à passagem ao limite de sequências. Também descreveremos a construção de medidas produto, tanto para famílias finitas quanto para famílias enumeráveis de medidas. Ao final, discutiremos os conceitos correlatos de continuidade absoluta e de derivação de Lebesgue.

A.2.1 Integral de Lebesgue

Ao longo desta seção $(X, \mathcal{B}, \mu)$ será sempre um espaço de medida. Vamos definir a noção de integral de Lebesgue por etapas. O primeiro passo trata da integral de uma função simples:

Definição A.2.1. Seja $s = \sum_{j=1}^{k} \alpha_j \mathcal{X}_{A_j}$ uma função simples. Então a *integral* de s em relação à medida μ é dada por:

$$\int s \, d\mu = \sum_{j=1}^{k} \alpha_j \mu(A_j).$$

É fácil verificar que esta definição é coerente: se duas combinações lineares de funções características definem uma mesma função então os valores das integrais obtidos a partir das duas combinações coincidem (Exercício A.2.1).

O próximo passo é definir integral de uma função mensurável não negativa. A ideia é definir a integral da função como sendo o limite das integrais de funções simples que a aproximam, utilizando a Proposição A.1.33:

Definição A.2.2. Seja $f : X \to [0, \infty]$ uma função mensurável não negativa. Então

$$\int f d\mu = \lim_n \int s_n d\mu,$$

onde $s_1 \leq s_2 \leq \ldots$ é uma sequência não decrescente de funções simples tal que $\lim_n s_n(x) = f(x)$ para todo $x \in X$.

Não é difícil verificar (Exercício A.2.2) que esta definição é coerente: o valor da integral não depende da escolha da sequência de funções simples crescendo para f.

Agora, para estender a definição de integral a qualquer função mensurável, observemos que dada uma função $f : X \to [-\infty, +\infty]$ sempre podemos escrever $f = f^+ - f^-$ com

$$f^+(x) = \max\{f(x), 0\} \quad \text{e} \quad f^-(x) = \max\{-f(x), 0\}.$$

É claro que as funções f^+ e f^- são não negativas. Além disso, pela Proposição A.1.31, elas são mensuráveis se f é mensurável.

Definição A.2.3. Seja $f : X \to [-\infty, +\infty]$ uma função mensurável. Então,

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu,$$

desde que pelo menos uma das integrais do lado direito seja finita (valem as convenções usuais $(+\infty) - a = +\infty$ e $a - (+\infty) = -\infty$ para todo $a \in \mathbb{R}$).

Definição A.2.4. Dizemos que uma função $f : X \to [-\infty, +\infty]$ é *integrável* se for mensurável e a sua integral for um número real. Denotamos o conjunto das funções integráveis por $\mathcal{L}^1(X, \mathcal{B}, \mu)$ ou, mais simplesmente, por $\mathcal{L}^1(\mu)$.

Dada uma função mensurável $f : X \to [-\infty, \infty]$ e um conjunto mensurável E definimos a *integral de f sobre E* por

$$\int_E f d\mu = \int f \mathcal{X}_E \, d\mu,$$

onde $\mathcal{X}_E$ é a função característica do conjunto E.

Exemplo A.2.5. Considere $X = [0, 1]$ munido da medida de Lebesgue m. Seja $f = \mathcal{X}_B$, onde B é o subconjunto dos números racionais. Então $m(B) = 0$ e portanto, usando a Definição A.2.2, a integral de Lebesgue de f é igual a zero. Por outro lado, um cálculo direto mostra que toda soma de Riemann inferior de f é igual a 0 mas toda soma de Riemann superior de f é igual a 1. Portanto, a integral de Riemann de f não está definida. Outra forma de chegar a esta mesma conclusão é utilizando o fato conhecido de que a integral de Riemann da função característica de um conjunto mensurável está definida se, e somente se, a sua fronteira tem medida nula. Note que no caso presente a fronteira de B é todo o $[0, 1]$ e, portanto, tem medida positiva.

Exemplo A.2.6. Sejam $x_1, \dots, x_m \in X$ e $p_1, \dots, p_m > 0$ com $p_1 + \cdots + p_m = 1$. Consideremos a medida de probabilidade μ definida em 2^X por

$$\mu = \sum_{i=1}^m p_i \delta_{x_i} \quad \text{onde } \delta_{x_i} \text{ é a medida delta de Dirac em } x_i.$$

Em outras palavras $\mu(A) = \sum_{x_i \in A} p_i$ para todo subconjunto A de X. Então, para qualquer função $f : X \to [-\infty, +\infty]$,

$$\int f \, d\mu = \sum_{i=1}^m p_i f(x_i).$$

Proposição A.2.7. *O conjunto $\mathcal{L}^1(\mu)$ das funções reais integráveis é um espaço vetorial real. Além disso, a aplicação $I : \mathcal{L}^1(\mu) \to \mathbb{R}$ dada por $I(f) = \int f \, d\mu$ é um funcional linear positivo, ou seja:*

(1) $\int af + bg \, d\mu = a \int f \, d\mu + b \int g \, d\mu$ e

(2) $\int f \, d\mu \geq \int g \, d\mu$ se $f(x) \geq g(x)$ para todo x.

Em particular, $|\int f \, d\mu| \leq \int |f| \, d\mu$ se $|f| \in \mathcal{L}^1(\mu)$. Além disso, $|f| \in \mathcal{L}^1(\mu)$ se, e somente se, $f \in \mathcal{L}^1(\mu)$.

A noção de integral pode ser estendida a uma classe ainda mais ampla de funções, de duas maneiras diferentes. Por um lado, podemos considerar funções mensuráveis complexas $f : X \to \mathbb{C}$. Nesse caso, dizemos que f é integrável se, e somente se, a parte real $\Re f$ e a parte imaginária $\Im f$ forem integráveis. Então, por definição,

$$\int f \, d\mu = \int \Re f \, d\mu + i \int \Im f \, d\mu.$$

Por outro lado, podemos considerar funções que não são necessariamente mensuráveis mas que coincidem com alguma função mensurável num subconjunto do domínio com medida total. Para explicar isto precisamos da seguinte noção, que será utilizada frequentemente ao longo do texto:

Definição A.2.8. Dizemos que uma propriedade é válida *em μ-quase todo ponto* se é válida em todo o X exceto, possivelmente, num conjunto de medida nula.

Por exemplo, dizemos que uma sequência de funções $(f_n)_n$ converge para uma função em μ-quase todo ponto se existe um conjunto mensurável N com $\mu(N) = 0$ tal que $f(x) = \lim_n f_n(x)$ para todo $x \in X \setminus N$. Analogamente, dizemos que duas funções f e g são iguais em μ-quase todo ponto se existe um conjunto mensurável N com $\mu(N) = 0$ tal que $f(x) = g(x)$ para todo $x \in X \setminus N$. Neste caso, supondo que as funções sejam integráveis, as suas integrais coincidem

$$\int f \, d\mu = \int g \, d\mu \quad \text{se } f = g \text{ em } \mu\text{-quase todo ponto.}$$

Esta observação permite definir integral para qualquer função f, possivelmente não mensurável, que é igual em μ-quase todo ponto a uma função mensurável g: basta tomar $\int f \, d\mu = \int g \, d\mu$.

Vamos encerrar esta seção observando que a noção de integral também pode ser estendida para medidas com sinal e para medidas complexas, do seguinte modo. Dada uma medida com sinal μ, considere a respectiva decomposição de Hahn $\mu = \mu^+ - \mu^-$. Dizemos que uma função ϕ é integrável para μ se ela é integrável para μ^+ e para μ^-. Então definimos:

$$\int \phi \, d\mu = \int \phi \, d\mu^+ - \int \phi \, d\mu^-.$$

De modo similar, dada uma medida complexa μ, dizemos que uma função ϕ é integrável para μ se ela é integrável para a parte real $\Re\mu$ e a parte imaginária $\Im\mu$ e, nesse caso, definimos

$$\int \phi \, d\mu = \int \phi \, d\Re\mu + i \int \phi \, d\Im\mu.$$

A.2.2 Teoremas de convergência

Nesta seção mencionamos três resultados muito importantes para o estudo da convergência de funções sob o sinal de integral. O primeiro deles lida com sequências monótonas de funções:

Teorema A.2.9 (Convergência monótona)**.** *Seja $f_n : X \to [-\infty, +\infty]$ uma sequência não-decrescente de funções mensuráveis não negativas e seja $f : X \to [-\infty, +\infty]$ a função definida por $f(x) = \lim_n f_n(x)$. Então*

$$\lim_n \int f_n \, d\mu = \int f(x) \, d\mu.$$

O próximo resultado vale para sequências mais gerais, não necessariamente monótonas:

Teorema A.2.10 (Lema de Fatou)**.** *Seja $f_n : X \to [0, +\infty]$ uma sequência de funções mensuráveis não negativas. Então, a função $f : X \to [-\infty, +\infty]$ definida por $f(x) = \liminf_n f_n(x)$ é integrável e vale*

$$\int \liminf_n f_n(x) \, d\mu \leq \liminf_n \int f_n \, d\mu.$$

O mais poderoso dos resultados nesta seção é o teorema da convergência dominada, que garante que podemos tomar o limite sob o sinal da integral sempre que a sequência de funções é majorada por alguma função integrável:

Teorema A.2.11 (Convergência dominada)**.** *Seja $f_n : X \to \mathbb{R}$ uma sequência de funções mensuráveis e suponha que existe uma função integrável g tal que $|f_n(x)| \leq |g(x)|$ para μ-quase todo x em X. Suponha também que a sequência $(f_n)_n$ converge em μ-quase todo ponto para uma função $f : X \to \mathbb{R}$. Então f é integrável e vale:*

$$\lim_n \int f_n \, d\mu = \int f \, d\mu.$$

A.2.3 Produto de medidas

Sejam $(X_j, \mathcal{A}_j, \mu_j)$, $j = 1, \ldots, n$ espaços de medida finita, isto é, tais que $\mu_j(X_j) < \infty$. É possível tornar o produto cartesiano $X_1 \times \cdots \times X_n$ um espaço de medida, da seguinte forma. Considere em $X_1 \times \cdots \times X_n$ a σ-álgebra gerada pela família de todos os conjuntos da forma $A_1 \times \cdots \times A_n$ com $A_j \in \mathcal{A}_j$. Ela é chamada *σ-álgebra produto* e é representada por $\mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_n$.

Teorema A.2.12. *Existe uma única medida μ em $(X_1 \times \cdots \times X_n, \mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_n)$ tal que $\mu(A_1 \times \cdots \times A_n) = \mu_1(A_1) \cdots \mu_n(A_n)$ para todo $A_1 \in \mathcal{A}_1, \ldots, A_n \in \mathcal{A}_n$. Em particular, μ é uma medida finita.*

A demonstração deste resultado (veja o Teorema 35.B do livro de Halmos [Hal50]) combina o teorema da extensão (Teorema A.1.13) com o teorema da convergência monótona (Teorema A.2.9). A medida μ no enunciado é representada por $\mu_1 \times \cdots \times \mu_n$ e é chamada *produto* das medidas $\mu_1, \ldots, \mu_n$. Desta forma, fica definido o *espaço de medida produto*

$$(X_1 \times \cdots \times X_n, \mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_n, \mu_1 \times \cdots \times \mu_n).$$

O Teorema A.2.12 permanece válido quando as medidas μ_j são apenas σ-finitas, exceto que neste caso a medida produto μ é apenas σ-finita.

Agora vamos descrever a construção do produto de uma família *enumerável* de espaços de medida. Na verdade, para isso nos restringiremos ao caso de probabilidades. Sejam $(X_j, \mathcal{B}_j, \mu_j)$, $j \in \mathcal{I}$ espaços de medida com $\mu_j(X_j) = 1$ para todo $j \in \mathcal{I}$. O conjunto de índices tanto pode ser $\mathcal{I} = \mathbb{N}$ como $\mathcal{I} = \mathbb{Z}$. Consideremos o produto cartesiano

$$\Sigma = \prod_{j \in \mathcal{I}} X_j = \{(x_j)_{j \in \mathcal{I}} : x_j \in X_j\}. \tag{A.2.1}$$

Chamamos *cilindros* de Σ os subconjuntos da forma

$$[m; A_m, \ldots, A_n] = \{(x_j)_{j \in \mathcal{I}} : x_j \in A_j \text{ para } m \leq j \leq n\} \tag{A.2.2}$$

onde $m \in \mathcal{I}$ e $n \geq m$ e $A_j \in \mathcal{B}_j$ para $m \leq j \leq n$. Note que o próprio X é um cilindro, por exemplo, $X = [1; X_1]$. Por definição, a *σ-álgebra produto* em Σ é a σ-álgebra $\mathcal{B}$ gerada pela família de todos os cilindros. A família $\mathcal{A}$ das uniões finitas de cilindros disjuntos dois-a-dois é uma álgebra e ela gera a σ-álgebra $\mathcal{B}$.

Teorema A.2.13. *Existe uma única medida μ em $(\Sigma, \mathcal{B})$ tal que*

$$\mu([m; A_m, \ldots, A_n]) = \mu_m(A_m) \cdots \mu_n(A_n) \tag{A.2.3}$$

para qualquer cilindro $[m; A_m, \ldots, A_n]$. Em particular, μ é uma probabilidade.

A prova deste teorema (veja o Teorema 38.B do livro de Halmos [Hal50]) usa o teorema de extensão (Teorema A.1.13) e o teorema da continuidade no vazio (Teorema A.1.14). A probabilidade μ é chamada *medida produto* e é representada por $\prod_{j \in \mathcal{I}} \mu_j$. O espaço de probabilidade $(\Sigma, \mathcal{B}, \mu)$ construído desta forma é denominado *espaço produto* dos espaços $(X_j, \mathcal{B}_j, \mu_j)$, $j \in \mathcal{I}$.

Um caso particular importante é quando os espaços $(X_i, \mathcal{B}_i, \mu_i)$ são todos iguais a um dado $(X, \mathcal{C}, \nu)$. Estes espaços podem ser usados para modelar sequências de experimentos aleatórios idênticos em que o resultado de cada experimento é independente dos demais. Supõe-se que cada experimento toma valores no conjunto X, com distribuição de probabilidade igual a ν. Neste contexto, a medida $\mu = \nu^{\mathcal{I}}$ costuma ser chamada *medida de Bernoulli* definida por ν. A propriedade (A.2.3) corresponde à igualdade

$$\mu([m; A_m, \ldots, A_n]) = \prod_{j=m}^{n} \nu(A_j), \tag{A.2.4}$$

a qual pode ser lida nos seguintes termos: a probabilidade de todo evento composto $\{x_m \in A_m, \ldots, x_n \in A_n\}$ é igual ao produto das probabilidades dos eventos individuais $x_i \in A_i$. Portanto, (A.2.4) traduz realmente a ideia de que os sucessivos experimentos são independentes.

Temos interesse especial no caso em que X é um conjunto finito, munido da σ-álgebra $\mathcal{C} = 2^X$, formada por todos os subconjuntos de X. Neste caso, é útil considerar os *cilindros elementares*

$$[m; a_m, \ldots, a_n] = \{(x_j)_{j\in\mathcal{I}} \in X : x_m = a_m, \ldots, x_n = a_n\}, \tag{A.2.5}$$

correspondentes a conjuntos A_j formados por um único ponto a_j. Observe que todo cilindro é uma união finita de cilindros elementares disjuntos dois-a-dois. Portanto, a σ-álgebra gerada pelos cilindros elementares coincide com a σ-álgebra gerada por todos os cilindros, e o mesmo vale para a álgebra gerada. Além disso, a relação (A.2.4) pode ser escrita

$$\mu([m; a_m, \ldots, a_n]) = p_{a_m} \cdots p_{a_n} \quad \text{onde } p_a = \nu(\{a\}) \text{ para } a \in X. \tag{A.2.6}$$

Considere o conjunto finito X munido da topologia discreta. A topologia produto em $\Sigma = X^{\mathcal{I}}$ coincide com a topologia gerada pelos cilindros elementares. Além disso (veja o Exercício A.1.11), ela coincide com a topologia associada à distância definida por

$$d\big((x_i)_{i\in\mathcal{I}}, (y_i)_{i\in\mathcal{I}}\big) = \theta^N, \tag{A.2.7}$$

onde $\theta \in (0,1)$ está fixado e $N = N((x_i)_{i\in\mathcal{I}}, (y_i)_{i\in\mathcal{I}}) \geq 0$ é o maior número inteiro tal que $x_i = y_i$ para todo $i \in \mathcal{I}$ com $|i| < N$.

A.2.4 Derivação de medidas

Seja m a medida de Lebesgue em $\mathbb{R}^d$. Dado um subconjunto mensurável A de $\mathbb{R}^d$, dizemos que um ponto $a \in \mathbb{R}^d$ é um *ponto de densidade* de A se este conjunto preenche a maior parte de qualquer pequena vizinhança de a, isto é,

$$\lim_{\delta\to 0} \frac{m(B(a,\delta) \cap A)}{m(B(a,\delta))} = 1. \tag{A.2.8}$$

Teorema A.2.14. *Seja A um subconjunto mensurável de $\mathbb{R}^d$ com medida de Lebesgue $m(A)$ maior que zero. Então m-quase todo ponto $a \in A$ é ponto de densidade de A.*

No Exercício A.2.11 sugerimos uma demonstração deste resultado. Ele pode também ser obtido como consequência direta do teorema que vamos enunciar a seguir. Dizemos que uma função $f : \mathbb{R}^d \to \mathbb{R}$ é *localmente integrável* se o produto $f\mathcal{X}_K$ é integrável para todo compacto $K \subset \mathbb{R}^d$.

Teorema A.2.15 (Derivação de Lebesgue)**.** *Seja $X = \mathbb{R}^d$, $\mathcal{B}$ a σ-álgebra de Borel e m a medida de Lebesgue em $\mathbb{R}^d$. Seja $f : X \to \mathbb{R}$ uma função localmente integrável. Então*

$$\lim_{r\to 0} \frac{1}{m(B(x,r))} \int_{B(x,r)} |f(y) - f(x)| dm = 0 \quad \textit{em } m\textit{-quase todo ponto.}$$

Em particular,

$$\lim_{r\to 0} \frac{1}{m(B(x,r))} \int_{B(x,r)} f(y) dm = f(x) \quad \textit{em } m\textit{-quase todo o ponto.}$$

O ingrediente fundamental na demonstração destes resultados é o seguinte fato geométrico:

Teorema A.2.16 (Lema de Vitali)**.** *Seja m a medida de Lebesgue em $\mathbb{R}^d$ e suponha que para cada $x \in \mathbb{R}^d$ é dada uma sequência $(B_n(x))_n$ de bolas centradas em x com raio convergindo para zero. Seja $A \subset \mathbb{R}^d$ um conjunto mensurável com $m(A) > 0$. Então, para cada $\varepsilon > 0$ existem sequências $(x_j)_j$ em $\mathbb{R}^d$ e $(n_j)_j$ em $\mathbb{N}$ tais que*

1. *as bolas $B_{n_j}(x_j)$ são disjuntas duas-a-duas;*
2. *$m\big(\cup_j B_{n_j}(x_j) \setminus A\big) < \varepsilon$ e $m\big(A \setminus \cup_j B_{n_j}(x_j)\big) = 0$.*

O teorema ainda é válido se, no lugar de bolas, tomarmos para $(B_n(x))_n$ qualquer sequência de conjuntos satisfazendo $\cap_n B_n(x) = \{x\}$ e

$$\sup_{x,n} \frac{\sup\{d(x,y) : y \in B_n(x)\}}{\inf\{d(x,z) : z \notin B_n(x)\}} < \infty.$$

O conjunto das medidas definidas num mesmo espaço mensurável possui a seguinte relação de ordem parcial:

Definição A.2.17. Sejam μ e ν duas medidas num mesmo espaço mensurável $(X, \mathcal{B})$.

Dizemos que ν é *absolutamente contínua* em relação a μ se todo conjunto mensurável E que satisfaz $\mu(E) = 0$ também satisfaz $\nu(E) = 0$. Nesse caso escrevemos $\nu \ll \mu$.

Dizemos que μ e ν são *equivalentes*, e escrevemos $\mu \sim \nu$, se cada uma delas for absolutamente contínua em relação à outra. Em outras palavras, duas medidas são equivalentes se elas têm os mesmos conjuntos com medida nula.

Outro resultado importante, conhecido por teorema de Radon-Nikodym, afirma que se $\nu \ll \mu$ então a medida ν pode ser vista como o produto de μ por uma certa função mensurável ρ:

Teorema A.2.18 (Radon-Nikodym)**.** *Se μ e ν são medidas finitas tais que $\nu \ll \mu$ então existe uma função mensurável $\rho : X \to [0, +\infty]$ tal que $\nu = \rho\mu$, ou seja, tal que*

$$\int \phi \, d\nu = \int \phi\rho \, d\mu \quad \textit{para toda função mensurável limitada } \phi : X \to \mathbb{R}. \tag{A.2.9}$$

Em particular, $\nu(E) = \int_E \rho \, d\mu$ para todo conjunto mensurável $E \subset X$. Além disso, ρ é essencialmente única: duas quaisquer funções que satisfazem (A.2.9) *são iguais em μ-quase todo ponto.*

Chamamos ρ de *densidade*, ou *derivada de Radon-Nikodym*, de ν relativamente a μ e escrevemos

$$\rho = \frac{d\nu}{d\mu}.$$

Definição A.2.19. Sejam μ e ν duas medidas num espaço mensurável $(X, \mathcal{B})$. Dizemos que μ e ν são *mutuamente singulares* se existem conjuntos mensuráveis disjuntos A e B tais que $A \cup B = X$ e $\mu(A) = 0$ e $\nu(B) = 0$. Nesse caso escrevemos $\mu \perp \nu$.

O teorema de decomposição de Lebesgue afirma que, dadas duas medidas finitas μ e ν, podemos escrever $\nu = \nu_a + \nu_s$ onde ν_a e ν_s são medidas finitas tais que $\nu_a \ll \mu$ e $\nu_s \perp \mu$. Combinando este resultado com o teorema de Radon-Nikodym, obtemos:

Teorema A.2.20 (Decomposição de Lebesgue)**.** *Dadas medidas finitas μ e ν, existe uma função mensurável $\rho : X \to [0, +\infty]$ e existe uma medida finita η satisfazendo $\nu = \rho\mu + \eta$ e $\eta \perp \mu$.*

A.2.5 Exercícios

A.2.1. Prove que a integral de uma função simples está bem definida: se duas combinações lineares de funções características definem uma mesma função, então os valores das integrais obtidos a partir das duas combinações coincidem.

A.2.2. Mostre que se $(r_n)_n$ e $(s_n)_n$ são sequências não decrescentes de funções simples não negativas convergindo em μ-quase todo ponto para uma mesma função $f : M \to [0, +\infty)$ então $\lim_n \int r_n \, d\mu = \lim_n \int s_n \, d\mu$.

A.2.3. Prove a Proposição A.2.7.

A.2.4. (Desigualdade de Tchebysheff-Markov) Seja $f : M \to \mathbb{R}$ uma função não-negativa integrável com respeito a uma medida finita μ. Então, dado qualquer número real $a > 0$,

$$\mu\big(\{x \in M : f(x) \geq a\}\big) \leq \frac{1}{a} \int_M f \, d\mu.$$

Em particular, se $\int |f| \, d\mu = 0$, então $\mu\big(\{x \in M : f(x) \neq 0\}\big) = 0$.

A.2.5. Seja f uma função integrável. Mostre que para todo $\varepsilon > 0$ existe $\delta > 0$ tal que $|\int_E f\,d\mu| < \varepsilon$ para todo conjunto mensurável E com $\mu(E) < \delta$.

A.2.6. Sejam $\psi_1, \dots, \psi_N : M \to \mathbb{R}$ funções mensuráveis limitadas num espaço de probabilidade $(M, \mathcal{B}, \mu)$. Mostre que para todo $\varepsilon > 0$ existem $x_1, \dots, x_s \in M$ e números positivos $\alpha_1, \dots, \alpha_s$ tais que $\sum_{j=1}^{s} \alpha_j = 1$ e

$$|\int \psi_i\,d\mu - \sum_{j=1}^{s} \alpha_j \psi_i(x_j)| < \varepsilon \quad \text{para todo } i = 1, \dots, N.$$

A.2.7. Prove o teorema da convergência dominada (Teorema A.2.11) a partir do lema de Fatou (Teorema A.2.10).

A.2.8. Um conjunto $\mathcal{F}$ de funções mensuráveis $f : M \to \mathbb{R}$ diz-se *uniformemente integrável* com respeito a uma probabilidade μ se para todo $\alpha > 0$ existe $C > 0$ tal que $\int_{\{|f|>C\}} |f|\,d\mu < \alpha$ para todo $f \in \mathcal{F}$. Mostre que

(a) $\mathcal{F}$ é uniformemente integrável com respeito a μ se, e somente se, existe $L > 0$ e para todo $\varepsilon > 0$ existe $\delta > 0$ tal que $\int |f|\,d\mu < L$ e $\int_A |f|\,d\mu < \varepsilon$ para todo $f \in \mathcal{F}$ e todo conjunto mensurável A com $\mu(A) < \delta$.

(b) Se existe alguma função g integrável com respeito a μ tal que $|f| \le |g|$ para todo $f \in \mathcal{F}$ (dizemos que $\mathcal{F}$ é dominado por g) então $\mathcal{F}$ é uniformemente integrável com respeito a μ.

(c) Se $\mathcal{F}$ é uniformemente integrável com respeito a μ então $\lim_n \int f_n\,d\mu = \int \lim f_n\,d\mu$ para qualquer sequência $(f_n)_n$ em $\mathcal{F}$ convergindo em μ-quase todo ponto.

A.2.9. Mostre que a é um ponto de densidade de um conjunto $A \subset \mathbb{R}^d$ se, e somente se,

$$\lim_{\delta \to 0} \inf \left\{ \frac{m(B \cap A)}{m(B)} : B \text{ bola contida em } B(a, \delta) \text{ e contendo } a \right\} = 1. \quad \text{(A.2.10)}$$

A.2.10. Seja $\mathcal{P}_n$, $n \ge 1$ uma sequência de partições enumeráveis de $\mathbb{R}^d$ em subconjuntos mensuráveis. Suponha que o diâmetro $\operatorname{diam} \mathcal{P}_n = \sup\{\operatorname{diam} P : P \in \mathcal{P}_n\}$ converge para zero quando $n \to \infty$. Mostre que, dado qualquer conjunto mensurável $A \subset \mathbb{R}^d$ com medida de Lebesgue positiva, é possível escolher conjuntos $P_n \in \mathcal{P}_n$, $n \ge 1$ de tal forma que $m(A \cap P_n)/m(P_n) \to 1$ quando $n \to \infty$.

A.2.11. Prove o Teorema A.2.14.

A.2.12. Seja x_1, $x_2 \in M$ e p_1, p_2, q_1, $q_2 > 0$ com $p_1 + p_2 = q_1 + q_2 = 1$. Considere as medidas de probabilidade μ e ν dadas por

$$\mu(A) = \sum_{x_i \in A} p_i, \qquad \nu(A) = \sum_{x_i \in A} q_i,$$

ou seja, $\mu = p_1\delta_{x_1} + p_2\delta_{x_2}$ e $\nu = q_1\delta_{x_1} + q_2\delta_{x_2}$. Mostre que $\nu \ll \mu$ e $\mu \ll \nu$ e calcule as respectivas derivadas de Radon-Nikodym.

A.2.13. Construa uma probabilidade μ absolutamente contínua com respeito à medida de Lebesgue m em $[0,1]$ tal que existe um conjunto mensurável $K \subset [0,1]$ com medida $\mu(K) = 0$ mas $m(K) = 1/2$. Em particular, m não é absolutamente contínua com respeito a μ. Poderíamos pedir que $m(K) = 1$?

A.2.14. Suponha que $f : X \to X$ é tal que existe uma cobertura enumerável de M por conjuntos mensuráveis B_n, $n \geq 1$, tais que a restrição de f a cada B_n é uma bijeção sobre a imagem, com inversa mensurável. Seja η uma probabilidade η em M tal que $A \subset B_n$ e $\eta(A) = 0$ implica $\eta(f(A)) = 0$. Mostre que existe uma função $J_\eta : X \to [0, +\infty]$ tal que

$$\int_{f(B_n)} \psi \, d\eta = \int_{B_n} (\psi \circ f) J_\eta \, d\eta$$

para toda função mensurável limitada $\psi : X \to \mathbb{R}$ e todo n. Além disso, J_η é essencialmente única.

A.2.15. Seja $\mu = \mu^+ - \mu^-$ a decomposição de Hahn de uma medida finita com sinal μ. Mostre que existem funções $\rho^\pm$ e $\tau^\pm$ tais que $\mu^+ = \rho^+|\mu| = \tau^+\mu$ e $\mu^- = \rho^-|\mu| = \tau^-\mu$. Que funções são estas?

A.2.16. Seja $(\mu_n)_n$ e $(\nu_n)_n$ duas sequências de medidas tais que $\mu = \sum_n \mu_n$ e $\nu = \sum_n \nu_n$ são medidas finitas. Escreva $\hat{\mu}_n = \sum_{i=1}^n \mu_i$ e $\hat{\nu}_n = \sum_{i=1}^n \nu_i$. Mostre que se $\hat{\mu}_n \ll \hat{\nu}_n$ para todo n então $\mu \ll \nu$ e

$$\frac{d\mu}{d\nu} = \lim_n \frac{d\hat{\mu}_n}{d\hat{\nu}_n} \quad \text{em } \nu\text{-quase todo ponto.}$$

Bibliografia

[Aar97] J. Aaronson. *An introduction to infinite ergodic theory*, volume 50 of *Mathematical Surveys and Monographs*. American Mathematical Society, 1997.

[AB] A. Avila and J. Bochi. Proof of the subadditive ergodic theorem. Preprint www.mat.puc-rio.br/∼jairo/.

[BK83] M. Brin and A. Katok. On local entropy. In *Geometric dynamics (Rio de Janeiro, 1981)*, volume 1007 of *Lecture Notes in Math.*, pages 30–38. Springer-Verlag, 1983.

[Bos93] M. Boshernitzan. Quantitative recurrence results. *Invent. Math.*, 113(3):617–631, 1993.

[Cas04] A. A. Castro. *Teoria da medida.* Projeto Euclides. IMPA, 2004.

[Dug66] J. Dugundji. *Topology.* Allyn and Bacon Inc., 1966.

[eKO14] M. Viana e K. Oliveira. *Fundamentos da Teoria Ergódica*, volume 1 of *Fronteiras da Matemática*. Sociedade Brasileira de Matemática, 2014.

[ET36] P. Erdös and P. Turán. On some sequences of integers. *J. London. Math. Soc.*, 11:261–264, 1936.

[Fer02] R. Fernandez. *Medida e integração.* Projeto Euclides. IMPA, 2002.

[Fur77] H. Furstenberg. Ergodic behavior and a theorem of Szemerédi on arithmetic progressions. *J. d'Analyse Math.*, 31:204–256, 1977.

[Fur81] H. Furstenberg. *Recurrence in ergodic theory and combinatorial number theory.* Princeton Univertsity Press, 1981.

[GT08] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math.*, 167:481–547, 2008.

[Hal50] P. Halmos. *Measure Theory.* D. Van Nostrand Company, 1950.

[Jac60] K. Jacobs. *Neuere Methoden und Ergebnisse der Ergodentheorie.* Ergebnisse der Mathematik und ihrer Grenzgebiete. N. F., Heft 29. Springer-Verlag, 1960.

[Jac63] K. Jacobs. *Lecture notes on ergodic theory, 1962/63. Parts I, II.* Matematisk Institut, Aarhus Universitet, Aarhus, 1963.

[Kri70] W. Krieger. On entropy and generators of measure-preserving transformations. *Trans. Amer. Math. Soc.*, 149:453–464, 1970.

[KW82] Y. Katznelson and B. Weiss. A simple proof of some ergodic theorems. *Israel J. Math.*, 42:291–296, 1982.

[Pet83] K. Petersen. *Ergodic theory.* Cambridge Univ. Press, 1983.

[Rok67] V. A. Rokhlin. Lectures on the entropy theory of measure-preserving transformations. *Russ. Math. Surveys*, 22 -5:1–52, 1967. Transl. from Uspekhi Mat. Nauk. 22 - 5 (1967), 3–56.

[Roy63] H. L. Royden. *Real analysis.* The Macmillan Co., 1963.

[Rud87] W. Rudin. *Real and complex analysis.* McGraw-Hill, 1987.

[Ste58] E. Sternberg. On the structure of local homeomorphisms of Euclidean n-space - II. *Amer. J. Math.*, 80:623–631, 1958.

[Sze75] S. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975.

[vdW27] B. van der Waerden. Beweis eibe Baudetschen Vermutung. *Nieuw Arch. Wisk.*, 15:212–216, 1927.

[Via14] M. Viana. *Lectures on Lyapunov Exponents.* Cambridge University Press, 2014.

[Yos68] K. Yosida. *Functional analysis.* Second edition. Die Grundlehren der mathematischen Wissenschaften, Band 123. Springer-Verlag, 1968.

Índice