

Fundamentos da Teoria Ergódica

Krerley Oliveira e Marcelo Viana

Prefácio

Em termos simples, a Teoria Ergódica é a disciplina matemática que estuda sistemas dinâmicos munidos de medidas invariantes. Começaremos por dar as definições precisas destas noções e por analisar as principais motivações para o seu estudo, após o que mencionaremos alguns momentos marcantes da história desta disciplina. Ao final do prefácio esboçaremos o conteúdo deste livro e a sua organização, bem como os requisitos desejáveis para o seu estudo.

Sistemas dinâmicos

Há várias definições, mais ou menos gerais, do que é um sistema dinâmico. Nós nos restringiremos a dois modelos principais. O primeiro deles, ao qual nos referiremos na maior parte do tempo, são as transformações $f : M \rightarrow M$ em algum espaço M . Heuristicamente, pensamos em f como associando a cada estado $x \in M$ do sistema o estado $f(x) \in M$ em que o sistema se encontrará uma unidade de tempo depois. Trata-se portanto de um modelo de dinâmica com tempo discreto.

Também consideraremos fluxos, que são modelos de sistemas dinâmicos com tempo contínuo. Lembre que um fluxo em M é uma família $f^t : M \rightarrow M$, $t \in \mathbb{R}$ de transformações satisfazendo

$$f^0 = \text{identidade} \quad \text{e} \quad f^t \circ f^s = f^{t+s} \quad \text{para todo } t, s \in \mathbb{R}. \quad (0.0.1)$$

Fluxos aparecem, por exemplo, associados a equações diferenciais: tome como f^t a transformação que associa a cada ponto x o valor no tempo t da solução da equação que passa por x no tempo zero.

Num caso e no outro, sempre suporemos que o sistema dinâmico é mensurável, ou seja, que o espaço M está munido de uma σ -álgebra de subconjuntos ditos *mensuráveis* e que essa σ -álgebra é preservada pela dinâmica: a pré-imagem de qualquer conjunto mensurável também é um conjunto mensurável. Na maior parte dos casos, M será um espaço topológico, ou até um espaço métrico, munido da menor σ -álgebra que contém todos os abertos (σ -álgebra de Borel). De fato, em muitas das situações que consideraremos ao longo do livro, suporemos mesmo que M é uma variedade e que a dinâmica é diferenciável.

Medidas invariantes

Sempre consideraremos medidas μ definida na σ -álgebra do espaço M . Dizemos que μ é uma probabilidade se $\mu(M) = 1$. Na maior parte dos casos trataremos com medidas finitas, isto é, tais que $\mu(M) < \infty$. Neste caso sempre podemos transformar μ numa probabilidade ν : para isso basta definir

$$\nu(E) = \frac{\mu(E)}{\mu(M)} \quad \text{para cada conjunto mensurável } E \subset M.$$

Em geral, uma medida μ diz-se invariante pela transformação f se

$$\mu(E) = \mu(f^{-1}(E)) \quad \text{para todo conjunto mensurável } E \subset M. \quad (0.0.2)$$

Heuristicamente, isto significa que a probabilidade de um ponto estar num dado conjunto é igual à probabilidade de que a sua imagem esteja nesse conjunto. Note que a definição (0.0.2) faz sentido, uma vez que, por hipótese, a pré-imagem de qualquer conjunto mensurável ainda é um conjunto mensurável.

No caso de fluxos, substituímos a relação (0.0.2) por

$$\mu(E) = \mu(f^{-t}(E)) \quad \text{para todo mensurável } E \subset M \text{ e todo } t \in \mathbb{R}. \quad (0.0.3)$$

Por que estudar medidas invariantes?

Como em todo ramo da Matemática, parte importante da motivação é intrínseca e estética: estas estruturas matemáticas têm propriedades profundas e surpreendentes que conduzem à demonstração de belíssimos teoremas. Igualmente fascinante, idéias e resultados da Teoria Ergódica se aplicam em outras áreas da Matemática que a priori nada têm de probabilístico, por exemplo a Combinatória e a Teoria dos Números.

Outra razão para este estudo é que muitos fenômenos importantes na Natureza e nas ciências experimentais são modelados por sistemas dinâmicos que deixam invariante alguma medida interessante. O exemplo mais importante, historicamente, veio da Física: sistemas hamiltonianos, que descrevem a evolução de sistemas conservativos na mecânica newtoniana, correspondem a fluxos que preservam uma medida natural, a medida de Liouville. Aliás veremos que sistemas dinâmicos muito gerais possuem medidas invariantes.

Ainda outra motivação fundamental para que nos interessemos por medidas invariantes é que o seu estudo pode conduzir a informação importante sobre o comportamento dinâmico do sistema, que dificilmente poderia ser obtida de outro modo. O Teorema de Recorrência de Poincaré, um dos primeiros que estudaremos neste livro, ilustra bem o que acabamos de dizer: ele afirma que a órbita de quase todo ponto, relativamente a qualquer medida invariante finita, regressa arbitrariamente perto do ponto inicial.

Breve apresentação histórica

A palavra *ergódico* é o resultado da concatenação de duas palavras gregas, *ergos* = trabalho e *odos* = caminho, e foi introduzida pelo físico L. Boltzmann, no

século 19, no seu trabalho sobre a teoria cinética dos gases. Os sistemas em que L. Boltzmann, J. C. Maxwell, J. C. Gibbs, os principais fundadores da teoria cinética, estavam interessados são descritos por um fluxo hamiltoniano, ou seja, uma equação diferencial da forma

$$\left(\frac{dq_1}{dt}, \dots, \frac{dq_n}{dt}, \frac{dp_1}{dt}, \dots, \frac{dp_n}{dt} \right) = \left(\frac{\partial H}{\partial p_1}, \dots, \frac{\partial H}{\partial p_n}, -\frac{\partial H}{\partial q_1}, \dots, -\frac{\partial H}{\partial q_n} \right).$$

Boltzmann acreditava que as órbitas típicas do fluxo preenchem toda a superfície de energia $H^{-1}(c)$ que as contém. A partir desta *hipótese ergódica*, ele deduziu que as médias temporais de grandezas observáveis (funções) ao longo de órbitas típicas coincidem com as respectivas médias espaciais na superfície de energia, um fato crucial para a sua formulação da teoria cinética.

De fato, esta hipótese é claramente falsa e, com o tempo, tornou-se usual chamar hipótese ergódica ao que seria uma consequência dela, a saber, que as médias temporais e espaciais são iguais. Sistemas para os quais vale esta igualdade foram chamados *ergódicos*. E pode dizer-se que boa parte da Teoria Ergódica, tal como ela se desenvolveu ao longo do século 20, foi motivada pelo problema de decidir se a maioria dos sistemas hamiltonianos, especialmente aqueles que aparecem na teoria cinética dos gases, são ergódicos ou não.

Um avanço fundamental ocorreu nos anos trinta, quando os matemáticos J. von Neumann e G. D. Birkhoff provaram que médias temporais existem para quase toda órbita. No entanto, em meados dos anos cinquenta, o grande matemático russo A. N. Kolmogorov observou que muitos sistemas hamiltonianos *não são* ergódicos. Este resultado espectacular foi muito expandido por V. Arnold e por J. Moser, no que veio a ser chamado teoria KAM em homenagem aos três.

Por outro lado, ainda nos anos trinta, E. Hopf tinha dado os primeiros exemplos importantes de sistemas hamiltonianos ergódicos, os fluxos geodésicos de superfícies com curvatura negativa. O seu resultado foi generalizado por D. Anosov, nos anos sessenta, para variedades de qualquer dimensão. De fato, Anosov tratou uma classe bem mais geral de sistemas, tanto com tempo contínuo como com tempo discreto, que são chamados sistemas de Anosov, ou sistemas globalmente hiperbólicos. Uma classe ainda mais ampla de sistemas, chamados uniformemente hiperbólicos, foi introduzida por S. Smale, e constituiu um importante foco da teoria dos Sistemas Dinâmicos ao longo das últimas décadas.

Nos anos setenta, Ya. Sinai desenvolveu a teoria das medidas de Gibbs dos sistemas de Anosov, conservativos ou dissipativos, que foi logo em seguida estendida por D. Ruelle e por R. Bowen para sistemas uniformemente hiperbólicos, constituindo uma das maiores realizações da teoria ergódica diferenciável. Não podemos deixar de mencionar, nesta breve lista de contribuições fundamentais, a introdução da noção de *entropia* por Kolmogorov e Sinai no final dos anos cinquenta, e a demonstração, por D. Ornstein cerca de dez anos depois, de que a entropia é um invariante completo para deslocamentos (“shifts”) de Bernoulli: dois deslocamentos de Bernoulli são equivalentes se, e somente se, eles têm a mesma entropia.

Histórico sucinto

Este livro foi desenvolvido a partir de notas de curso que escrevemos para os participantes de minicursos ministrados na Escola de Verão do Departamento de Matemática da Universidade Federal de Pernambuco (Recife), em janeiro de 2003, e do encontro Novos Talentos em Matemática da Fundação Calouste Gulbenkian (Lisboa), em setembro de 2004.

Nos dois casos, o público estava formado majoritariamente por alunos jovens que não tinham contato prévio com a Teoria Ergódica (em muitos casos nem mesmo com a Teoria da Medida) e tornava-se necessário fornecer material bastante acessível que permitisse a esses alunos acompanhar minimamente as ideias principais a serem expostas. Ainda neste estágio, o texto foi utilizado por colegas, tais como o professor Vanderlei Horita (UNESP), para ministrar minicursos a públicos com um perfil semelhante.

Ao longo do desenvolvimento do texto, buscamos preservar o caráter elementar dos capítulos iniciais, especialmente os Capítulos 1 e 2, de tal forma que eles possam ser utilizados de forma independente, com um mínimo de pré-requisitos.

Sobretudo a partir do minicurso ministrado no Colóquio Brasileiro de Matemática (IMPA, Rio de Janeiro) de 2005, este projeto foi adquirindo objetivos mais abrangentes. Gradualmente, fomos evoluindo para tentar apresentar num texto coerente, com formato de livro de texto, o material que consideramos formar o núcleo central da Teoria Ergódica. Para isso nos inspiramos fortemente na nossa própria experiência como pesquisadores da área, buscando reunir numa apresentação unificada as noções e resultados que se mostraram importantes para o extraordinário desenvolvimento que esta área tem vivido nas últimas décadas.

Uma preocupação importante foi tentar manter o texto o mais possível auto-contido. De fato, a Teoria Ergódica se apoia em diversas disciplinas da Matemática, com destaque para a Teoria da Medida, a Topologia e a Análise. Nos Apêndices coligimos as principais noções e resultados destas disciplinas que são úteis para o restante do texto. De um modo geral, as demonstrações são omitidas, já que existem diversos excelentes textos sobre estes temas. Uma exceção são os resultados sobre medidas em espaços métricos (Apêndice A.3), para as quais optamos por incluir provas dos fatos que mais nos interessam.

Por outro lado, pressupomos que o leitor conhece os conceitos e resultados fundamentais da Álgebra Linear, inclusive a forma canônica de Jordan.

Organização do texto

O corpo principal do livro está formado pelos Capítulos 1 a 12, que podem ser organizados do seguinte modo:

- Os Capítulos 1 a 4 formam uma espécie de ciclo básico, no qual apresentamos as noções e resultados fundamentais da Teoria Ergódica - invariância, recorrência e ergodicidade - bem como alguns exemplos principais. O Capítulo 3 introduz os resultados fundamentais (teoremas ergódicos) em torno dos quais está constituída toda a teoria.

- O Capítulo 4, onde introduzimos a noção de ergodicidade, é um dos pontos fulcrais deste texto. Os dois capítulos seguintes (Capítulos 5 e 6) desenvolvem alguns temas importantes relacionados com essa noção: decomposição de medidas invariantes em medidas ergódicas e sistemas admitindo uma única medida invariante.
- Os Capítulos 7 a 9 tratam temas bastante diversos - perda de memória, problema do isomorfismo e entropia - mas se estruturam de forma coerente em torno da ideia de estudar sistemas cada vez mais ‘caóticos’: sistemas misturadores, sistemas com espectro de Lebesgue, sistemas de Kolmogorov e sistemas de Bernoulli.
- O Capítulo 9 é outro ponto fulcral do texto. Além de apresentar a noção de entropia, buscamos dar ao leitor a oportunidade de observar este conceito riquíssimo sob diversos pontos de vista. Essa teoria se articula naturalmente com o conteúdo do Capítulo 10, onde desenvolvemos a vertente topológica da noção de entropia.
- Os Capítulos 11 e 12 são dedicados a uma classe paradigmática de sistemas, as transformações expansoras, que nos permitem exibir uma aplicação concreta (e espetacular!) de muitas das ideias gerais apresentadas ao longo do texto. Vemos o Teorema de Ruelle e suas aplicações como o culminar natural de todo o texto.

Exemplos e aplicações têm um papel fundamental em qualquer disciplina matemática e isso é particularmente verdade no caso da Teoria Ergódica. Por esta razão, dedicamos particular atenção à apresentação de situações concretas que ilustram e valorizam os resultados gerais. Tais exemplos e construções são introduzidos gradativamente, buscando para cada um o contexto que melhor realça a sua relevância. Tipicamente, eles reaparecem em capítulos subsequentes para ilustrar os conceitos fundamentais que vamos introduzindo.

Os exercícios incluídos em cada seção têm uma função tripla. Num nível mais rotineiro, eles permitem adquirir familiaridade com os conceitos e o uso dos resultados apresentados no texto. Também deixamos para os exercícios alguns argumentos e demonstrações que não são usados na sequência do texto ou que pertencem a áreas afins mais elementares (Topologia, Teoria da Medida etc). Finalmente, exercícios mais sofisticados testam a compreensão global da teoria apresentada. Para facilidade do leitor, numa seção ao final do livro apresentamos soluções mais ou menos detalhadas de todos os exercícios.

Como utilizar este livro

Os comentários a seguir se destinam, prioritariamente, ao leitor que vai utilizar este livro para ministrar um curso.

O conteúdo dos Capítulos 1 a 12 é adequado para um curso anual, ou uma sequência de dois cursos semestrais. Se o leitor dispõe desse tempo, poderá tentar cobrir a grande maioria do material, possivelmente reservando alguns tópicos para seminários apresentados pelos alunos. As seguintes seções

são especialmente adequadas para esse fim: Seção 1.5, Seção 2.5, Seção 3.4, Seção 4.4, Seção 6.4, Seção 7.3, Seção 7.4, Seção 8.3 Seção 8.4, Seção 8.5, Seção 9.5, Seção 9.7, Seção 10.4, Seção 10.5, Seção 11.1, Seção 11.3, Seção 12.3 e Seção 12.4. Neste formato, o teorema de Ruelle (Teorema 12.1) constitui a conclusão natural para o curso.

Os Apêndices fornecem referências para material que é pré-requisito para o curso. Em princípio, eles não serão objeto de apresentação em aula, exceto pontualmente, em caso de necessidade.

Caso o leitor disponha apenas de um semestre, será necessário selecionar o material mais fundamental para apresentação em aula. A sugestão dos autores é buscar cobrir o seguinte programa:

- Capítulo 1: Seções 1.1, 1.2 e 1.3.
- Capítulo 2 Seções 2.1 e 2.2.
- Capítulo 3: Seções 3.1, 3.2 e 3.3.
- Capítulo 4: Seções 4.1, 4.2 e 4.3.
- Capítulo 5: Seção 5.1 (mencionar o teorema de Rokhlin).
- Capítulo 6: Seções 6.1, 6.2 e 6.3.
- Capítulo 7: Seções 7.1 e 7.2.
- Capítulo 8: Seção 8.1 (mencionar o teorema de Ornstein).
- Capítulo 9: Seções 9.1, 9.2, 9.3 e 9.4.
- Capítulo 10: Seções 10.1 e 10.2.
- Capítulo 11: Seção 11.1.

Neste formato, o curso poderá ser encerrado com a demonstração do princípio variacional para a entropia (Teorema 10.1) ou com a construção de medidas invariantes absolutamente contínuas para transformações expansoras em variedades (Teorema 11.1.2).

Em qualquer dos casos, procuramos elaborar o texto de tal forma que o professor possa se concentrar na apresentação das ideias e resultados centrais, deixando a cargo do aluno estudar por si mesmo muitas das demonstrações e resultados complementares. A seção final, com as dicas e soluções dos exercícios, é parte desse esforço para facilitar o estudo autônomo do aluno. De fato, dedicamos bastante esforço a fazer que as demonstrações sejam amigáveis, detalhando cuidadosamente os argumentos e incluindo referências explícitas aos resultados anteriores que estão sendo utilizados, bem como aos pontos do texto onde as noções pertinentes foram introduzidas. Além disso, a par da presença regular de exemplos e dos exercícios ao final de cada seção, não hesitamos em apresentar a mesma noção de dois ou mais pontos de vista sempre que isso nos pareceu útil para a sua compreensão em profundidade.

Agradecimentos

A elaboração deste texto se estendeu por mais de uma década. Ao longo desse tempo recebemos sugestões, comentários e críticas construtivas de um grande número de colegas e alunos. A lista que mencionamos a seguir está certamente incompleta e desde já nos desculpamos a qualquer um que tenha sido omitido inadvertidamente.

A primeira versão do material nos Apêndices A.1-A.2 foi escrita por João Gouveia, Vítor Saraiva e Ricardo Andrade, os quais atuaram como monitores do minicurso no evento Novos Talentos em Matemática 2004 que mencionamos previamente.

Diversos colegas utilizaram versões variadas do texto para ministrar minicursos e nos brindaram com as conclusões de suas experiências. Além de Vanderlei Horita (UNESP), Nivaldo Muniz (UFMA) e Meysam Nassiri (Teerã), gostaríamos de ressaltar os copiosos comentários de Vítor Araújo (UFRJ e, agora, UFBA), que influenciaram significativamente o modo como o texto foi evoluindo. François Ledrappier (Paris) nos ajudou com algumas questões sobre sistemas de substituições.

Sucessivas gerações de alunos dos cursos de pós-graduação do IMPA e da UFAL nos facultaram testar o texto na sala de aula. O retorno dado por Aline Gomes Cerqueira, Ermerson Araujo, Ignacio Atal, Raphaël Cyna, Rafael Lucena e Xiao-Chuan Liu nos permitiu corrigir muitas das debilidades do texto. Edileno de Almeida Santos, Felipe Soares Guimarães, Fernando Nera Lenarduzzi, Ítalo Dowell Lira Melo, Marco Vinicius Bahi Aymone e Renan Henrique Finder escreveram boa parte das dicas para os exercícios dos Capítulos 1 a 8 e dos apêndices.

Krerley Oliveira ¹ e Marcelo Viana ²

¹Departamento de Matemática, Universidade Federal de Alagoas, Campus A. C. Simões s/n, 57072-090 Maceió, Brasil. krerley@mat.ufal.br.

²IMPA, Estrada D. Castorina 110, 22460-320 Rio de Janeiro, Brasil. viana@impa.br.

Conteúdo

1	Medidas Invariantes e Recorrência	1
1.1	Medidas Invariantes	2
1.1.1	Exercícios	3
1.2	Teorema de recorrência de Poincaré	4
1.2.1	Versão mensurável	4
1.2.2	Teorema de Kač	5
1.2.3	Versão topológica	7
1.2.4	Exercícios	8
1.3	Exemplos	9
1.3.1	Expansão decimal	10
1.3.2	Transformação de Gauss	11
1.3.3	Rotações no círculo	16
1.3.4	Rotações em toros	18
1.3.5	Transformações conservativas	18
1.3.6	Fluxos conservativos	19
1.3.7	Exercícios	21
1.4	Indução	22
1.4.1	Transformação de primeiro retorno	22
1.4.2	Transformações induzidas	24
1.4.3	Torres de Kakutani-Rokhlin	27
1.4.4	Exercícios	28
1.5	Teoremas de recorrência múltipla	29
1.5.1	Teorema de recorrência múltipla de Birkhoff	30
1.5.2	Exercícios	33
2	Existência de Medidas Invariantes	35
2.1	Topologia fraca*	36
2.1.1	Definição e propriedades da topologia fraca*	36
2.1.2	Teorema Portmanteau	37
2.1.3	A topologia fraca* é metrizável	39
2.1.4	A topologia fraca* é compacta	40
2.1.5	Teorema de Prohorov	42
2.1.6	Exercícios	43
2.2	Demonstração do teorema de existência	44

2.2.1	Exercícios	48
2.3	Comentários de Análise Funcional	49
2.3.1	Dualidade e topologias fracas	49
2.3.2	Operador de Koopman	50
2.3.3	Exercícios	52
2.4	Produtos semi-diretos e extensões naturais	53
2.4.1	Medidas em produtos semi-diretos	53
2.4.2	Extensões naturais	54
2.4.3	Exercícios	57
2.5	Progressões aritméticas	58
2.5.1	Teorema de van der Waerden	60
2.5.2	Teorema de Szemerédi	61
2.5.3	Exercícios	63
3	Teoremas Ergódicos	65
3.1	Teorema ergódico de Von Neumann	66
3.1.1	Isometrias em espaços de Hilbert	66
3.1.2	Enunciado e prova do teorema	68
3.1.3	Convergência em $L^2(\mu)$	69
3.1.4	Exercícios	70
3.2	Teorema ergódico de Birkhoff	71
3.2.1	Tempo médio de visita	71
3.2.2	Médias temporais	72
3.2.3	Teorema de von Neumann e consequências	75
3.2.4	Exercícios	77
3.3	Teorema ergódico subaditivo	78
3.3.1	Preparação da demonstração	79
3.3.2	Lema fundamental	81
3.3.3	Estimativa da função φ_-	82
3.3.4	Majoração da função φ_+	83
3.3.5	Expoentes de Lyapunov	85
3.3.6	Exercícios	86
3.4	Tempo discreto e tempo contínuo	87
3.4.1	Fluxos suspensão	87
3.4.2	Transformações de Poincaré	90
3.4.3	Exercícios	92
4	Ergodicidade	95
4.1	Sistemas ergódicos	96
4.1.1	Conjuntos e funções invariantes	96
4.1.2	Caracterização espectral	97
4.1.3	Exercícios	100
4.2	Exemplos	101
4.2.1	Rotações em toros	101
4.2.2	Expansão decimal	103
4.2.3	Deslocamentos de Bernoulli	105

4.2.4	Transformação de Gauss	109
4.2.5	Endomorfismos lineares do toro	112
4.2.6	Argumento de Hopf	113
4.2.7	Exercícios	117
4.3	Propriedades das medidas ergódicas	118
4.3.1	Exercícios	121
4.4	Comentários sobre sistemas conservativos	122
4.4.1	Sistemas hamiltonianos	122
4.4.2	Teoria de Kolmogorov-Arnold-Moser	125
4.4.3	Pontos periódicos elíticos	128
4.4.4	Fluxos geodésicos	132
4.4.5	Sistemas de Anosov	132
4.4.6	Bilhares	134
4.4.7	Exercícios	141
5	Decomposição Ergódica	143
5.1	Teorema da decomposição ergódica	143
5.1.1	Enunciado do teorema	144
5.1.2	Desintegração de uma medida	145
5.1.3	Partições mensuráveis	147
5.1.4	Prova do teorema da decomposição ergódica	148
5.1.5	Exercícios	150
5.2	Teorema da desintegração de Rokhlin	151
5.2.1	Esperanças condicionais	151
5.2.2	Critério de σ -aditividade	153
5.2.3	Construção das medidas condicionais	155
5.2.4	Exercícios	157
6	Unicidade Ergódica	159
6.1	Unicidade ergódica	159
6.1.1	Exercícios	161
6.2	Minimalidade	161
6.2.1	Exercícios	163
6.3	Medida de Haar	164
6.3.1	Rotações em toros	164
6.3.2	Grupos topológicos e grupos de Lie	165
6.3.3	Translações em grupos compactos metrizáveis	169
6.3.4	Odômetros	171
6.3.5	Exercícios	174
6.4	Teorema de Weyl	175
6.4.1	Ergodicidade	176
6.4.2	Unicidade ergódica	177
6.4.3	Demonstração do teorema de Weyl	180
6.4.4	Exercícios	181

7	Correlações	183
7.1	Sistemas misturadores	184
7.1.1	Propriedades	184
7.1.2	Mistura fraca	187
7.1.3	Caracterização espectral	190
7.1.4	Exercícios	191
7.2	Deslocamentos de Markov	192
7.2.1	Ergodicidade	196
7.2.2	Mistura	199
7.2.3	Exercícios	201
7.3	Intercâmbios de intervalos	202
7.3.1	Minimalidade e ergodicidade	205
7.3.2	Mistura	207
7.3.3	Exercícios	210
7.4	Decaimento de correlações	210
7.4.1	Exercícios	214
8	Sistemas Equivalentes	217
8.1	Equivalência ergódica	218
8.1.1	Exercícios	220
8.2	Equivalência espectral	220
8.2.1	Invariantes de equivalência espectral	221
8.2.2	Autovetores e mistura fraca	222
8.2.3	Exercícios	225
8.3	Espectro discreto	226
8.3.1	Exercícios	229
8.4	Espectro de Lebesgue	229
8.4.1	Exemplos e propriedades	230
8.4.2	O caso invertível	233
8.4.3	Exercícios	236
8.5	Espaços de Lebesgue e isomorfismo ergódico	237
8.5.1	Isomorfismo ergódico	237
8.5.2	Espaços de Lebesgue	239
8.5.3	Exercícios	244
9	Entropia	247
9.1	Definição de entropia	248
9.1.1	Entropia em Teoria da Informação	248
9.1.2	Entropia de uma partição	249
9.1.3	Entropia de um sistema dinâmico	254
9.1.4	Exercícios	258
9.2	Teorema de Kolmogorov-Sinai	259
9.2.1	Partições geradoras	260
9.2.2	Semicontinuidade da entropia	262
9.2.3	Transformações expansivas	264
9.2.4	Exercícios	265

9.3	Entropia local	266
9.3.1	Prova do teorema de Shannon-McMillan-Breiman	267
9.3.2	Exercícios	271
9.4	Exemplos	271
9.4.1	Deslocamentos de Markov	272
9.4.2	Transformação de Gauss	273
9.4.3	Endomorfismos lineares do toro	274
9.4.4	Exercícios	276
9.5	Entropia e equivalência	276
9.5.1	Automorfismos de Bernoulli	277
9.5.2	Sistemas com entropia nula	278
9.5.3	Sistemas de Kolmogorov	281
9.5.4	Sistemas exatos	286
9.5.5	Exercícios	287
9.6	Entropia e decomposição ergódica	287
9.6.1	Afinidade da decomposição ergódica	289
9.6.2	Demonstração do teorema de Jacobs	291
9.6.3	Exercícios	294
9.7	Jacobianos e fórmula de Rokhlin	295
9.7.1	Exercícios	299
10	Princípio variacional	301
10.1	Entropia topológica	302
10.1.1	Definição via coberturas abertas	302
10.1.2	Conjuntos geradores e conjuntos separados	305
10.1.3	Cálculo e propriedades	309
10.1.4	Exercícios	311
10.2	Exemplos	312
10.2.1	Transformações expansivas	312
10.2.2	Deslocamentos de tipo finito	314
10.2.3	Entropia topológica de fluxos	317
10.2.4	Transformações diferenciáveis	319
10.2.5	Endomorfismos lineares do toro	321
10.2.6	Exercícios	323
10.3	Pressão	324
10.3.1	Definição via coberturas abertas	324
10.3.2	Conjuntos geradores e conjuntos separados	326
10.3.3	Propriedades	328
10.3.4	Comentários sobre Mecânica Estatística	332
10.3.5	Exercícios	333
10.4	Princípio variacional	334
10.4.1	Prova da cota superior	336
10.4.2	Aproximando a pressão	338
10.4.3	Exercícios	341
10.5	Estados de equilíbrio	341
10.5.1	Exercícios	346

11	Transformações Expansoras	349
11.1	Transformações expansoras em variedades	350
11.1.1	Lema de distorção	351
11.1.2	Existência de medidas ergódicas	354
11.1.3	Unicidade e conclusão da prova	356
11.1.4	Exercícios	358
11.2	Dinâmica das transformações expansoras	359
11.2.1	Ramos inversos contrativos	362
11.2.2	Sombreamento e pontos periódicos	363
11.2.3	Decomposição dinâmica	367
11.2.4	Exercícios	370
11.3	Entropia e pontos periódicos	371
11.3.1	Taxa de crescimento dos pontos periódicos	372
11.3.2	Aproximação por medidas atômicas	373
11.3.3	Exercícios	375
12	Formalismo Termodinâmico	377
12.1	Teorema de Ruelle	378
12.1.1	Medida de referência	379
12.1.2	Distorção e propriedade de Gibbs	381
12.1.3	Densidade invariante	383
12.1.4	Construção do estado de equilíbrio	386
12.1.5	Pressão e autovalores	387
12.1.6	Unicidade do estado de equilíbrio	390
12.1.7	Exatidão	392
12.1.8	Medidas absolutamente contínuas	393
12.1.9	Exercícios	395
12.2	Teorema de Livsič	395
12.2.1	Exercícios	398
12.3	Decaimento de correlações	399
12.3.1	Distâncias projetivas	400
12.3.2	Cones de funções Hölder	406
12.3.3	Convergência exponencial	409
12.3.4	Exercícios	413
12.4	Dimensão de repulsores conformes	414
12.4.1	Dimensão de Hausdorff	414
12.4.2	Repulsores conformes	416
12.4.3	Distorção e conformalidade	418
12.4.4	Existência e unicidade de d_0	421
12.4.5	Cota superior	423
12.4.6	Cota inferior	424
12.4.7	Exercícios	425

A	Elementos de Medida, Topologia e Análise	427
A.1	Espaços de medida	427
A.1.1	Espaços mensuráveis	428
A.1.2	Espaços de medida	430
A.1.3	Medida de Lebesgue	433
A.1.4	Aplicações mensuráveis	437
A.1.5	Exercícios	438
A.2	Integração em espaços de medida	441
A.2.1	Integral de Lebesgue	441
A.2.2	Teoremas de convergência	443
A.2.3	Produto de medidas	444
A.2.4	Derivação de medidas	446
A.2.5	Exercícios	448
A.3	Medidas em espaços métricos	450
A.3.1	Medidas regulares	450
A.3.2	Espaços métricos separáveis completos	453
A.3.3	Espaço das funções contínuas	454
A.3.4	Exercícios	456
A.4	Variedades diferenciáveis	457
A.4.1	Variedades e aplicações diferenciáveis	457
A.4.2	Espaço tangente e aplicação derivada	459
A.4.3	Transversalidade	461
A.4.4	Variedades riemannianas	462
A.4.5	Exercícios	463
A.5	Espaços $L^p(\mu)$	464
A.5.1	Espaços $L^p(\mu)$ com $1 \leq p < \infty$	464
A.5.2	Produto interno em $L^2(\mu)$	465
A.5.3	Funções essencialmente limitadas	466
A.5.4	Convexidade	466
A.5.5	Exercícios	467
A.6	Espaços de Hilbert	468
A.6.1	Ortogonalidade	469
A.6.2	Dualidade	470
A.6.3	Exercícios	471
A.7	Teoremas espectrais	472
A.7.1	Medidas espectrais	472
A.7.2	Representação espectral	475
A.7.3	Exercícios	476
	Dicas e soluções dos exercícios	479

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 1.5 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 resultados tem importantes aplicações em Combinatória e Teoria dos Números, como veremos mais tarde.

A ideia de recorrência também está por trás das construções que apresentamos na Seção 1.4. 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}, μ) um espaço de medida e seja $f : M \rightarrow M$ uma transformação mensurável. Dizemos que a medida μ é *invariante* por f se

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

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 $f^t : M \rightarrow M$, onde $t \in \mathbb{R}$, satisfazendo as seguintes condições:

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

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 $X(\gamma(t)) = \gamma'$, 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

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

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

$$\int \phi d\mu = \int \phi \circ f d\mu. \quad (1.1.4)$$

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

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 $\phi : M \rightarrow \mathbb{R}$, considere uma sequência $(s_n)_n$ de funções simples convergindo para ϕ e tal que $|s_n| \leq |\phi|$ 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

1.1.1 Exercícios

1.1.1. Seja $f : M \rightarrow M$ 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 $\delta_{p,k} = k^{-1}(\delta_p + \delta_{f(p)} + \cdots + \delta_{f^{k-1}(p)})$ é invariante por f se, e somente se, $f^k(p) = p$.

1.1.2. Prove a seguinte versão da Proposição 1.1.1. Sejam M um espaço métrico, $f : M \rightarrow M$ 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 $\phi : M \rightarrow \mathbb{R}$ então f preserva a medida μ .

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

1.1.4. Suponha que $f : M \rightarrow M$ 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 $\mu(B \Delta C) = 0$.

1.1.5. Seja $f : U \rightarrow U$ 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, $|\det Df| \equiv 1$.

1.2 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 $f : M \rightarrow M$ 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 de 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 $m > n \geq 1$ tais que $f^{-m}(E_0)$ intersecta $f^{-n}(E_0)$. Seja x um ponto na intersecção 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

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

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:

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

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 $t_j \rightarrow +\infty$ 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 $f : M \rightarrow M$ 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* $\rho_E : E \rightarrow \mathbb{N} \cup \{\infty\}$ definida da seguinte forma:

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

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 exhibe o valor da sua integral. Para o enunciado precisamos da seguinte notação:

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

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 $f : M \rightarrow M$, 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

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

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

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

Agora observe que

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

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

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

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

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

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

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

como queríamos demonstrar. \square

Em algumas 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

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

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 $n \rightarrow \infty$. 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 $f : M \rightarrow M$ se existe uma sequência $n_j \rightarrow \infty$ em \mathbb{N} tal que $f^{n_j}(x) \rightarrow x$. Analogamente, dizemos que $x \in M$ é recorrente para um fluxo $(f^t)_t$ se existe uma sequência $t_j \rightarrow +\infty$ em \mathbb{R} tal que $f^{t_j}(x) \rightarrow x$ quando $j \rightarrow \infty$.

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 $f : M \rightarrow M$ 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 $f : \mathbb{R} \rightarrow \mathbb{R}$ 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 1.2.2.

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 $f : M \rightarrow M$ é uma transformação contínua num espaço métrico compacto M , então existe algum ponto $x \in X$ 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 $X = \bigcap_\alpha X_\alpha$. 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.

1.2.4 Exercícios

1.2.1. 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 $f : M \rightarrow M$ 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 $x \in D$.

1.2.2. Seja $f : M \rightarrow M$ 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 .

1.2.3. Seja $f : M \rightarrow M$ 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 $\bigcup_k M_k = M$. 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.

1.2.4. Sejam $f : M \rightarrow M$ 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 :

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

para μ -quase todo ponto $x \in D$. [Observação: Dá para substituir \limsup por \liminf no enunciado, mas a prova desse fato terá que esperar até o Capítulo 3.]

1.2.5. Seja $f : M \rightarrow M$ uma transformação mensurável que preserva uma medida finita μ . Dado qualquer conjunto mensurável $A \subset M$ com $\mu(A) > 0$, seja $n_1 < n_2 < \dots$ 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 $n_{i+1} - n_i \leq C$ para todo i .

1. Mostre que para qualquer sequência crescente $k_1 < k_2 < \dots$ existem $j > i \geq 1$ 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 3.1.2.]

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

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

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

1.2.7. Seja $\omega = (1 + \sqrt{5})/2$ a *razão áurea* e seja $f : [0, 1] \rightarrow [0, 1]$ 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 $(q_n)_n \rightarrow \omega$ 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 1.2.6 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.]

1.3 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

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

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 $10x$. O gráfico da transformação f está representado na Figura 1.1.

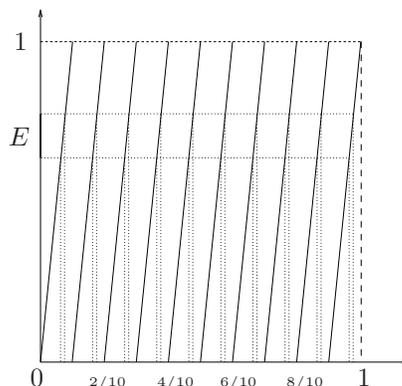


Figura 1.1: Transformação parte fracionária de $10x$

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

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

Esse fato pode ser verificado da seguinte forma. Começemos por supor que E é um intervalo. Então, conforme ilustrado na Figura 1.1, 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 $f : M \rightarrow M$ 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. Começemos 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. \square

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

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

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

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

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

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

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 1.3.2). 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

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

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

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

Então

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

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

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

e tem-se

$$x_0 = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\dots + \frac{1}{a_n + x_n}}}}. \quad (1.3.3)$$

Pode mostrar-se que a sequência

$$z_n = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\dots + \frac{1}{a_n}}}}. \quad (1.3.4)$$

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

$$x_0 = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\dots + \frac{1}{a_n + \dots}}}}, \quad (1.3.5)$$

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) suposemos que $x_n \in (0, 1)$ para todo $n \in \mathbb{N}$. Se encontramos algum $x_n = 0$, o processo pára 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* $G : [0, 1] \rightarrow [0, 1]$ 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, $G(x) = 1/x - k$. Veja a Figura 1.2.

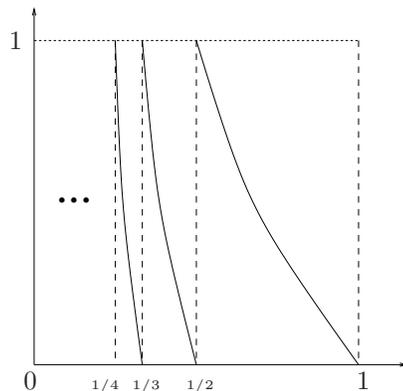


Figura 1.2: Transformação de Gauss

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 1.3.4). 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

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

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

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

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 $f : [0, 1] \rightarrow [0, 1]$ 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 $f_k = f | I_k$ a cada I_k é um difeomorfismo sobre $(0, 1)$.*

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

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

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

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

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,

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

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 1.3.3), 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:

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

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

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

Note que ϕ está bem definida: a expressão $e^{2\pi xi}$ não depende da escolha do representante x na classe $[x]$, uma vez que a função $x \mapsto e^{2\pi xi}$ 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 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

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

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 $\pi : \mathbb{R} \rightarrow S^1$ a projeção canônica que associa a cada $x \in \mathbb{R}$ a respectiva classe de equivalência $[x]$. Primeiramente, dizemos que um conjunto $E \subset \mathbb{R}$ 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:

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

Note que o lado esquerdo desta igualdade não depende de k , uma vez, por definição, $\pi^{-1}(E) \cap [k, k + 1) = (\pi^{-1}(E) \cap [0, 1)) + k$ 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 1.3.8). 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,

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

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

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

Combinando estas duas igualdades obtemos que

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

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

A dinâmica da rotação $R_\theta : S^1 \rightarrow S^1$ 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 $0 \leq r - la < a$. Isto implica que

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

Como $m = lq$ e $n = -lp$ 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 $d \geq 1$, considere a relação de equivalência em \mathbb{R}^d que identifica dois vetores se a sua diferença é um vetor com coordenadas inteiras:

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

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 desta relação. Seja m a medida de volume em \mathbb{R}^d . A operação

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

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

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

A aplicação

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

é sobrejetora e nos permite definir a medida de probabilidade de Lebesgue μ no d -toro, por meio da seguinte fórmula: $\mu(B) = m(\phi^{-1}(B))$ 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 $\theta = (\theta_1, \dots, \theta_d)$ é *racionalmente independente* se para quaisquer números inteiros n_0, n_1, \dots, n_d temos que

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

Caso contrário dizemos que θ é racionalmente dependente. Pode mostrar-se que θ é racionalmente independente se, e somente se, rotação 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 1.3.9- 1.3.10 e também o Corolário 4.2.3.

1.3.5 Transformações conservativas

Seja $M \subset \mathbb{R}^k$ um aberto do espaço euclidiano \mathbb{R}^d e seja $f : M \rightarrow M$ 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 $B \subset M$,

$$\text{vol}(f(B)) = \int_B |\det Df| dx. \quad (1.3.12)$$

Daqui se deduz facilmente o seguinte fato:

Lema 1.3.5. *Um difeomorfismo $f : M \rightarrow M$ 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 1 em todo ponto. Considere um conjunto mensurável E e seja $B = f^{-1}(E)$. A fórmula (1.3.12) dá que

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

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

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

Denotando $E = f(U)$, isto implica que $\text{vol}(E) > \text{vol}(f^{-1}(E))$ 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. \square

1.3.6 Fluxos conservativos

Agora vamos considerar o caso de fluxos $f^t : M \rightarrow M$, $t \in \mathbb{R}$. 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,

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

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 M é um aberto de \mathbb{R}^d e o fluxo $f^t : M \rightarrow M$ corresponde às trajetórias de um campo de vetores $F : M \rightarrow \mathbb{R}^d$ de classe C^1 . Em outras

palavras, $t \mapsto f^t(x)$ é a solução da equação diferencial

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

que tem x como condição inicial (quando tratando de equações diferenciáveis 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:

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

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

$$\operatorname{div} F = \frac{\partial F}{\partial x_1} + \cdots + \frac{\partial F}{\partial x_d}. \quad (1.3.15)$$

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, \dots, x_d) é dado por

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

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

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

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 .

1.3.7 Exercícios

1.3.1. 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.

1.3.2. Prove que, para quase todo número $x \in [0, 1]$ cuja expansão decimal contém o bloco 617 (por exemplo $x = 0,3375617264 \dots$), 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 $x \in [0, 1]$.

1.3.3. 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 $n \in \mathbb{N}$.

1.3.4. 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 $G^n(x) = 0$.

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

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

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

1.3.8. 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 .

1.3.9. Suponha que $\theta = (\theta_1, \dots, \theta_d)$ é vetor racionalmente dependente. Mostre que existe alguma função contínua $\varphi : \mathbb{T}^d \rightarrow \mathbb{C}$ 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 .

1.3.10. 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 $[y]$.

1.3.11. Seja U um aberto de \mathbb{R}^{2d} e $H : U \rightarrow \mathbb{R}$ 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.

1.3.12. Seja $f : U \rightarrow U$ um difeomorfismo de classe C^1 preservando a medida de Lebesgue num aberto U de \mathbb{R}^d . Seja $H : U \rightarrow \mathbb{R}$ 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 $ds/\|\text{grad } H\|$.

1.4 Indução

Nesta seção vamos descrever certas construções, baseadas no teorema de recorrência de Poincaré, que permitem associar a um dado sistema (f, μ) outros sistemas intimamente relacionados com ele, que chamamos *sistemas induzidos* por (f, μ) . O seu interesse resulta do seguinte. Por um lado, em muitos casos o sistema induzido pode ser construído de modo a ter melhores propriedades globais que o sistema original, o que torna a sua análise mais acessível. Por outro lado, a partir das propriedades do sistema induzido é possível obter conclusões interessantes a respeito do sistema original.

1.4.1 Transformação de primeiro retorno

Seja $f : M \rightarrow M$ uma transformação mensurável e μ uma probabilidade invariante. Seja $E \subset M$ um conjunto mensurável com $\mu(E) > 0$ e seja $\rho(x) = \rho_E(x)$ o tempo de primeiro retorno de x a E , tal como foi definido em (1.2.1). A *transformação de primeiro retorno* ao domínio E é definida por

$$g(x) = f^{\rho(x)}(x)$$

num subconjunto com medida total de E . Também denotamos por μ_E a restrição de μ aos subconjuntos mensuráveis de E .

Proposição 1.4.1. *A medida μ_E é invariante pela transformação $g : E \rightarrow E$.*

Demonstração. Para cada $k \geq 1$, denote por E_k o conjunto dos pontos $x \in E$ tais que $\rho(x) = k$. Por definição, $g(x) = f^k(x)$ para todo $x \in E_k$. Seja B um subconjunto mensurável qualquer de E . Então

$$\mu(g^{-1}(B)) = \sum_{k=1}^{\infty} \mu(f^{-k}(B) \cap E_k). \quad (1.4.1)$$

Por outro lado, como μ é f -invariante,

$$\mu(B) = \mu(f^{-1}(B)) = \mu(f^{-1}(B) \cap E_1) + \mu(f^{-1}(B) \setminus E). \quad (1.4.2)$$

Analogamente,

$$\begin{aligned} \mu(f^{-1}(B) \setminus E) &= \mu(f^{-2}(B) \setminus f^{-1}(E)) \\ &= \mu(f^{-2}(B) \cap E_2) + \mu(f^{-2}(B) \setminus (E \cup f^{-1}(E))). \end{aligned}$$

Substituindo em (1.4.2), obtemos

$$\mu(B) = \sum_{k=1}^2 \mu(f^{-k}(B) \cap E_k) + \mu(f^{-2}(B) \setminus \bigcup_{k=0}^1 f^{-k}(E)).$$

Repetindo este argumento sucessivamente, obtemos que

$$\mu(B) = \sum_{k=1}^n \mu(f^{-k}(B) \cap E_k) + \mu(f^{-n}(B) \setminus \bigcup_{k=0}^{n-1} f^{-k}(E)). \quad (1.4.3)$$

Vamos passar ao limite quando $n \rightarrow \infty$. É claro que a última parcela é majorada por $\mu(f^{-n}(B) \setminus \bigcup_{k=0}^{n-1} f^{-k}(E))$. Logo, usando a Observação 1.2.3, ela converge para zero quando $n \rightarrow \infty$. Deste modo, concluímos que

$$\mu(B) = \sum_{k=1}^{\infty} \mu(f^{-k}(B) \cap E_k).$$

Juntamente com (1.4.1), isto mostra que $\mu(g^{-1}(B)) = \mu(B)$ para todo subconjunto mensurável B de E . Isto é, a medida μ_E é invariante por g . \square

Exemplo 1.4.2. Considere a transformação $f : [0, \infty) \rightarrow [0, \infty)$ dada por

$$f(0) = 0 \quad \text{e} \quad f(x) = 1/x \text{ se } x \in (0, 1) \quad \text{e} \quad f(x) = x - 1 \text{ se } x \geq 1.$$

Considere $E = [0, 1]$. O tempo ρ de primeiro retorno a E é dado por

$$\rho(0) = 1 \quad \text{e} \quad \rho(x) = k + 1 \text{ se } x \in (1/(k + 1), 1/k] \text{ com } k \geq 1.$$

Então a transformação de primeiro retorno a E é dada por

$$g(0) = 0 \quad \text{e} \quad g(x) = 1/x - k \text{ se } x \in (1/(k + 1), 1/k] \text{ com } k \geq 1.$$

Em outras palavras, g é a transformação de Gauss. Vimos na Seção 1.3.2 que a transformação de Gauss admite uma probabilidade invariante equivalente à medida de Lebesgue em $[0, 1)$. Segue, usando as ideias que apresentaremos na próxima seção, que a transformação original f admite uma medida (infinita) invariante equivalente à medida de Lebesgue em $[0, \infty)$.

1.4.2 Transformações induzidas

Na direção contrária, dada uma medida invariante ν qualquer de $g : E \rightarrow E$, podemos construir uma certa medida invariante ν_ρ para $f : M \rightarrow M$. Aliás, para isto g não precisa ser a transformação de primeiro retorno: a construção que vamos apresentar a seguir é válida para qualquer *transformação induzida* de f , ou seja, qualquer transformação da forma

$$g : E \rightarrow E, \quad g(x) = f^{\rho(x)}(x), \quad (1.4.4)$$

onde $\rho : E \rightarrow \mathbb{N}$ é uma função mensurável (basta que esteja definida num subconjunto com medida total em E). Como antes, denotamos por E_k o conjunto dos $x \in E$ tais que $\rho(x) = k$. Então definimos:

$$\nu_\rho(B) = \sum_{n=0}^{\infty} \sum_{k>n} \nu(f^{-n}(B) \cap E_k), \quad (1.4.5)$$

para todo conjunto mensurável $B \subset M$.

Proposição 1.4.3. *A medida ν_ρ definida em (1.4.5) é invariante por f e satisfaz $\nu_\rho(M) = \int_E \rho d\nu$. Em particular, ν_ρ é finita se, e somente se, a função ρ é integrável com respeito a ν .*

Demonstração. Primeiro, provamos a invariância de ν_ρ . Pela definição (1.4.5),

$$\nu_\rho(f^{-1}(B)) = \sum_{n=0}^{\infty} \sum_{k>n} \nu(f^{-(n+1)}(B) \cap E_k) = \sum_{n=1}^{\infty} \sum_{k \geq n} \nu(f^{-n}(B) \cap E_k).$$

Podemos reescrever a expressão acima como:

$$\nu_\rho(f^{-1}(B)) = \sum_{n=1}^{\infty} \sum_{k>n} \nu(f^{-n}(B) \cap E_k) + \sum_{k=1}^{\infty} \nu(f^{-k}(B) \cap E_k). \quad (1.4.6)$$

A respeito da última parcela, observe que

$$\sum_{k=1}^{\infty} \nu(f^{-k}(B) \cap E_k) = \nu(g^{-1}(B)) = \nu(B) = \sum_{k=1}^{\infty} \nu(B \cap E_k),$$

uma vez que ν é invariante por g . Substituindo esta igualdade em (1.4.6), vemos que

$$\nu_\rho(f^{-1}(B)) = \sum_{n=1}^{\infty} \sum_{k>n} \nu(f^{-n}(B) \cap E_k) + \sum_{k=1}^{\infty} \nu(B \cap E_k) = \nu_\rho(B)$$

para todo conjunto mensurável $B \subset E$. A segunda afirmação na proposição é uma consequência direta das definições:

$$\nu_\rho(M) = \sum_{n=0}^{\infty} \sum_{k>n} \nu(f^{-n}(M) \cap E_k) = \sum_{n=0}^{\infty} \sum_{k>n} \nu(E_k) = \sum_{k=1}^{\infty} k\nu(E_k) = \int_E \rho d\nu.$$

Isto completa a demonstração. \square

É interessante analisar como esta construção se relaciona com a construção da seção anterior quando g é a transformação de primeiro retorno de f e a medida ν é a restrição $\mu|_E$ de alguma medida invariante μ de f :

Corolário 1.4.4. *Se g é a transformação de primeiro retorno de f a E , então*

1. $\nu_\rho(B) = \nu(B) = \mu(B)$ para todo conjunto mensurável $B \subset E$.
2. $\nu_\rho(B) \leq \mu(B)$ para todo conjunto mensurável $B \subset M$.

Demonstração. Por definição, $f^{-n}(E) \cap E_k = \emptyset$ para todo $0 < n < k$. Isto implica que, dado qualquer conjunto mensurável $B \subset E$, todas as parcelas com $n > 0$ na definição (1.4.5) são nulas. Logo, $\nu_\rho(B) = \sum_{k>0} \nu(B \cap E_k) = \nu(B)$ tal como afirmado na primeira parte do enunciado.

Considere qualquer conjunto mensurável $B \subset M$. Então,

$$\begin{aligned} \mu(B) &= \mu(B \cap E) + \mu(B \cap E^c) = \nu(B \cap E) + \mu(B \cap E^c) \\ &= \sum_{k=1}^{\infty} \nu(B \cap E_k) + \mu(B \cap E^c). \end{aligned} \quad (1.4.7)$$

Como μ é invariante, $\mu(B \cap E^c) = \mu(f^{-1}(B) \cap f^{-1}(E^c))$. Então, tal como na igualdade anterior

$$\begin{aligned} \mu(B \cap E^c) &= \mu(f^{-1}(B) \cap E \cap f^{-1}(E^c)) + \mu(f^{-1}(B) \cap E^c \cap f^{-1}(E^c)) \\ &= \sum_{k=2}^{\infty} \nu(f^{-1}(B) \cap E_k) + \mu(f^{-1}(B) \cap E^c \cap f^{-1}(E^c)). \end{aligned}$$

Substituindo em (1.4.7), obtemos

$$\mu(B) = \sum_{n=0}^1 \sum_{k>n} \nu(f^{-n}(B) \cap E_k) + \mu(f^{-1}(B) \cap \bigcap_{n=0}^1 f^{-n}(E^c)).$$

Repetindo este argumento sucessivamente, obtemos que

$$\begin{aligned} \mu(B) &= \sum_{n=0}^N \sum_{k>n} \nu(f^{-n}(B) \cap E_k) + \mu(f^{-N}(B) \cap \bigcap_{k=0}^N f^{-k}(E^c)) \\ &\geq \sum_{n=0}^N \sum_{k>n} \nu(f^{-n}(B) \cap E_k) \quad \text{para todo } N \geq 1. \end{aligned}$$

Passando ao limite quando $N \rightarrow \infty$, concluímos que $\mu(B) \geq \nu_\rho(B)$. \square

Além disso, o teorema de Kač nos diz que,

$$\nu_\rho(M) = \int_E \rho d\nu = \int_E \rho d\mu = \mu(M) - \mu(E_0^*).$$

Portanto, segue do Corolário 1.4.4 que $\nu_\rho = \mu$ se, e somente se, $\mu(E_0^*) = 0$.

Exemplo 1.4.5 (Manneville-Pomeau). Dado $d > 0$, seja a o único número em $(0, 1)$ que satisfaz $a(1 + a^d) = 1$. Então defina $f : [0, 1] \rightarrow [0, 1]$ da seguinte forma:

$$f(x) = x(1 + x^d) \text{ se } x \in [0, a] \quad \text{e} \quad f(x) = \frac{x - a}{1 - a} \text{ se } x \in (a, 1].$$

O gráfico de f está representado no lado esquerdo da Figura 1.3. Observe que $|f'(x)| \geq 1$ em todo ponto, sendo que a igualdade ocorre apenas para $x = 0$. Seja $(a_n)_n$ a sequência no intervalo $[0, a]$ definida por $a_1 = a$ e $f(a_{n+1}) = a_n$ para $n \geq 1$. Também escrevemos $a_0 = 1$. As propriedades desta sequência são estudadas no Exercício 1.4.2.

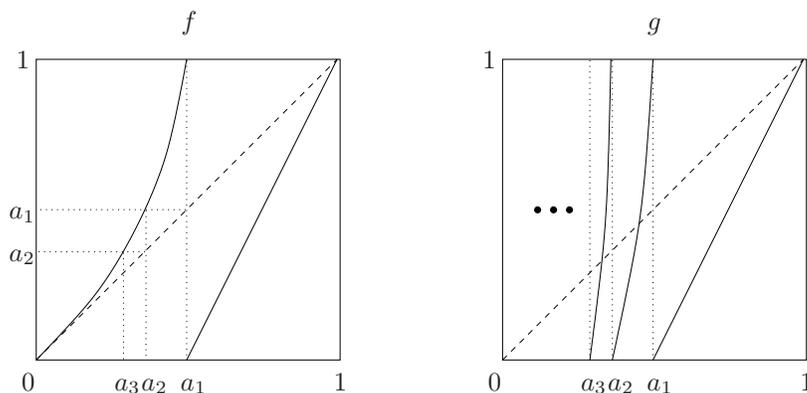


Figura 1.3: Construção de transformação induzida

Agora considere a aplicação $g(x) = f^{\rho(x)}(x)$, onde

$$\rho : [0, 1] \rightarrow \mathbb{N}, \quad \rho(x) = 1 + \min\{n \geq 0 : f^n(x) \in (a, 1]\}.$$

Em outras palavras, $\rho(x) = k$ e portanto $g(x) = f^k(x)$ para todo $x \in (a_k, a_{k-1}]$. O gráfico de g está representado no lado direito da Figura 1.3. Note que a restrição a cada intervalo $(a_k, a_{k-1}]$ é uma bijeção sobre $(0, 1]$. Um ponto crucial é que a transformação induzida g é *expansora*:

$$|g'(x)| \geq \frac{1}{1-a} > 1 \quad \text{para todo } x \in [0, 1].$$

Pode mostrar-se, usando as ideias que desenvolveremos no Capítulo 11, que g admite uma única probabilidade invariante ν equivalente à medida de Lebesgue em $(0, 1]$. Além disso, a densidade de ν relativamente à medida de Lebesgue está limitada de zero e infinito. Então a medida f -invariante ν_ρ em (1.4.5) é equivalente à medida de Lebesgue. Resulta (veja o Exercício 1.4.2) que esta medida é finita se, e somente se, $d \in (0, 1)$.

1.4.3 Torres de Kakutani-Rokhlin

É possível e útil generalizar a construção anterior ainda mais, abrindo mão da própria transformação $f : M \rightarrow M$. Mais precisamente, dada uma transformação $g : E \rightarrow E$, uma medida ν em E invariante por g , e uma função mensurável $\rho : E \rightarrow \mathbb{N}$, vamos mostrar como *construir* uma transformação $f : M \rightarrow M$ e uma medida ν_ρ invariante por f tais que E pode ser identificado com um subconjunto de M , g é a transformação de primeiro retorno de f a E , com tempo de retorno dado por ρ , e a medida ν_ρ restrita a E coincide com ν .

Esta transformação f é chamada *torre* de g com tempo ρ . A medida ν_ρ é finita se, e somente se, ρ é integrável com respeito a ν . Elas são construídas da seguinte forma. Começamos por definir:

$$\begin{aligned} M &= \{(x, n) : x \in E \text{ e } 0 \leq n < \rho(x)\} \\ &= \bigcup_{k=1}^{\infty} \bigcup_{n=0}^{k-1} E_k \times \{n\}. \end{aligned}$$

Ou seja, M consiste de k cópias de cada conjunto $E_k = \{x \in E : \rho(x) = k\}$, ‘empilhadas’ umas sobre as outras. Chamamos cada $\cup_{k>n} E_k \times \{n\}$ de *n-ésimo andar* de M . Veja a Figura 1.4.

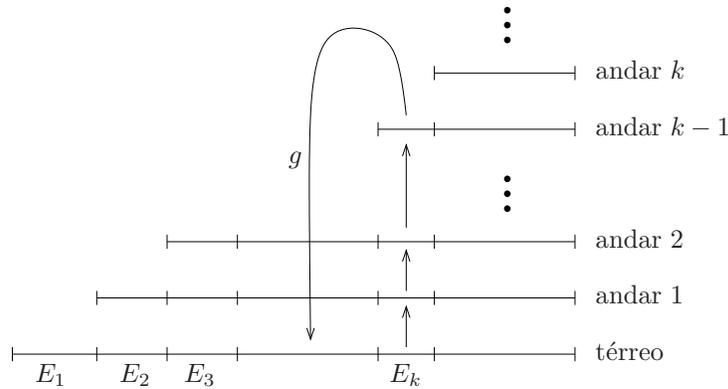


Figura 1.4: Torre de g com tempo ρ

Em seguida definimos $f : M \rightarrow M$ da seguinte forma:

$$f(x, n) = \begin{cases} (x, n + 1) & \text{se } n < \rho(x) - 1 \\ (g(x), 0) & \text{se } n = \rho(x) - 1 \end{cases} .$$

Em outras palavras, a dinâmica ‘eleva’ cada ponto (x, n) um andar de cada vez, até alcançar o andar $\rho(x) - 1$; a essa altura o ponto ‘cai’ diretamente para $(g(x), 0)$, no andar térreo. O andar térreo $E \times \{0\}$ se identifica naturalmente com o conjunto E . Além disso, a transformação de primeiro retorno a $E \times \{0\}$ corresponde precisamente a $g : E \rightarrow E$.

Finalmente, a medida ν_ρ fica definida por

$$\nu_\rho | (E_k \times \{n\}) = \nu | E_k$$

para todo $0 \leq n < k$. É claro que a restrição de ν_ρ ao andar térreo coincide com ν . Além disso, ν_ρ é invariante por f e

$$\nu_\rho(M) = \sum_{k=1}^{\infty} k\nu(E_k) = \int_E \rho d\nu.$$

Isto completa a construção da torre de Kakutani-Rokhlin.

1.4.4 Exercícios

1.4.1. Seja $f : S^1 \rightarrow S^1$ a transformação $f(x) = 2x \pmod{\mathbb{Z}}$. Mostre que a função $\tau(x) = \min\{k \geq 0 : f^k(x) \in (1/2, 1)\}$ é integrável relativamente à medida de Lebesgue. Enuncie e prove um resultado correspondente para qualquer transformação $g : S^1 \rightarrow S^1$ de classe C^1 que esteja C^1 próxima de f , ou seja, tal que $\sup_x \{\|g(x) - f(x)\|, \|g'(x) - f'(x)\|\}$ seja suficientemente pequeno.

1.4.2. Considere a medida ν_ρ e a sequência $(a_n)_n$ definidas no Exemplo 1.4.5. Verifique que ν_ρ é sempre σ -finita. Mostre que $(a_n)_n$ é decrescente e converge para zero. Além disso, existem $c_1, c_2, c_3, c_4 > 0$ tais que

$$c_1 \leq a_j j^{1/d} \leq c_2 \quad \text{e} \quad c_3 \leq (a_j - a_{j+1})j^{1+1/d} \leq c_4 \quad \text{para todo } j. \quad (1.4.8)$$

Deduz que a medida g -invariante ν_ρ é finita se, e somente se, $d \in (0, 1)$.

1.4.3. Seja $\sigma : \Sigma \rightarrow \Sigma$ a aplicação definida no espaço $\Sigma = \{1, \dots, d\}^{\mathbb{Z}}$ por $\sigma((x_n)_n) = (x_{n+1})_n$. Descreva a transformação g de primeiro retorno ao conjunto $\{(x_n)_n \in \Sigma : x_0 = 1\}$.

1.4.4. [Lema de Kakutani-Rokhlin] Seja $f : M \rightarrow M$ uma transformação invertível e seja μ uma medida de probabilidade invariante sem átomos tal que $\mu(\cup_{n \in \mathbb{N}} f^n(E)) = 1$ para todo $E \subset M$ com $\mu(E) > 0$. Mostre que para todo $n \geq 1$ e todo $\varepsilon > 0$ existe um conjunto mensurável $B \subset M$ tal que os iterados $B, f(B), \dots, f^{n-1}(B)$ são disjuntos dois-a-dois e o complementar da sua união tem medida menor que ε . Em particular, isto vale para todo sistema invertível *aperiódico*, ou seja, cujo conjunto dos pontos periódicos tem medida nula.

1.4.5. Seja $f : M \rightarrow M$ uma transformação e seja $(H_j)_{j \geq 1}$ uma coleção de subconjuntos de M tal que se $x \in H_n$ então $f^j(x) \in H_{n-j}$ para todo $0 \leq j < n$. Seja H o conjunto dos pontos que pertencem a H_j para infinitos de valor de j , ou seja, $H = \cap_{k=1}^{\infty} \cup_{j=k}^{\infty} H_j$. Para $y \in H$ defina $\tau(y) = \min\{j \geq 1 : y \in H_j\}$ e $T(y) = f^{\tau(y)}(y)$. Observe que T é uma aplicação de H em H . Além disso, mostre que

$$\limsup_n \frac{1}{n} \#\{1 \leq j \leq n : x \in H_j\} \geq \theta > 0 \quad \Rightarrow \quad \liminf_k \frac{1}{k} \sum_{i=0}^{k-1} \tau(T^i(x)) \leq \frac{1}{\theta}.$$

1.4.6. Seja $f : M \rightarrow M$ uma transformação preservando uma medida μ . Sejam $(H_j)_{j \geq 1}$ e $\tau : M \rightarrow \mathbb{N}$ como no Exercício 1.4.5. Considere a sequência de funções $(\tau_n)_n$ definida por $\tau_1(x) = \tau(x)$ e $\tau_n(x) = \tau(f^{\tau_{n-1}}(x)) + \tau_{n-1}(x)$ para $n > 1$. Suponha que

$$\limsup_n \frac{1}{n} \#\{1 \leq j \leq n : x \in H_j\} \geq \theta > 0 \quad \text{para } \mu \text{ quase todo } x \in M.$$

Mostre que $\tau_{n+1}(x)/\tau_n(x) \rightarrow 1$ para μ -quase todo ponto $x \in M$. [Observação: Dizemos que a sequência $(\tau_n)_n$ é *não lacunar*.]

1.5 Teoremas de recorrência múltipla

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

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

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

$$\lim_k f_i^{n_k}(a) = a \quad \text{para todo } i = 1, \dots, q. \quad (1.5.1)$$

A demonstração deste teorema será dada na Seção 1.5.1. A seguir, discutimos a seguinte generalização do teorema de recorrência de Poincaré:

Teorema 1.5.2 (Recorrência múltipla de Poincaré). *Seja (M, \mathcal{B}, μ) um espaço de probabilidade e sejam $f_i : M \rightarrow M$, $i = 1, \dots, q$ 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*

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

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 1.5.2 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 1.5.2 ao conjunto $F = E \cap f_1^{-n}(E) \cap \cdots \cap f_q^{-n}(E)$, obtemos que existe $m \geq 1$ tal que

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

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.

Outra consequência simples é o teorema de recorrência múltipla de Birkhoff (Teorema 1.5.1). De fato, se $f_i : M \rightarrow M$, $i = 1, 2, \dots, q$ são transformações contínuas num espaço métrico compacto que comutam entre si, então existe alguma probabilidade invariante μ comum a todas essas transformações (este fato será verificado no próximo capítulo, no Exercício 2.2.2). A partir daqui podemos argumentar exatamente como na demonstração do Teorema 1.2.4. Ou seja, 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 1.5.1.

1.5.1 Teorema de recorrência múltipla de Birkhoff

Vamos tratar o caso em que as transformações f_1, \dots, f_q são homeomorfismos de M , que é suficiente para os nossos objetivos no presente capítulo. O caso geral pode ser deduzido facilmente (veja o Exercício 2.4.7) usando a ideia de extensão natural, que apresentaremos no próximo capítulo.

O teorema pode ser reformulado do seguinte modo útil. Considere a transformação $F : M^q \rightarrow M^q$ definida no espaço produto $M^q = M \times \cdots \times M$ por $F(x_1, \dots, x_q) = (f_1(x_1), \dots, f_q(x_q))$. Denote por Δ_q a *diagonal* de M^q , ou seja, o subconjunto dos pontos da forma $\tilde{x} = (x, \dots, x)$. O Teorema 1.5.1 afirma, precisamente, que existe $\tilde{a} \in \Delta_q$ e existe $(n_k)_k \rightarrow \infty$ tal que

$$\lim_k F^{n_k}(\tilde{a}) = \tilde{a}. \quad (1.5.2)$$

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, \dots, f_q .

Denote por \mathcal{G} o grupo (abeliano) gerado pelos homeomorfismos f_1, \dots, 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 1.5.2). 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 1.5.3. *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

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

Note que a aplicação $M \rightarrow \Delta_q$, $x \mapsto \tilde{x} = (x, \dots, x)$ é um homeomorfismo e, mesmo, uma isometria para esta escolha da distância. Todo aberto $U \subset M$ corresponde a um aberto $\tilde{U} \subset \Delta_q$ via esse homeomorfismo. Dado qualquer $g \in \mathcal{G}$, representaremos por $\tilde{g} : M^q \rightarrow M^q$ o homeomorfismo definido por $\tilde{g}(x_1, \dots, x_q) = (g(x_1), \dots, g(x_q))$. 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 1.5.3 pode ser reescrita na seguinte forma:

$$\bigcup_{h \in \mathcal{H}} \tilde{h}^{-1}(\tilde{U}) = \Delta_q. \quad (1.5.3)$$

Lema 1.5.4. *Dado $\varepsilon > 0$ existem $\tilde{x}, \tilde{y} \in \Delta_q$ e $n \geq 1$ tais que $d(F^n(\tilde{x}), \tilde{y}) < \varepsilon$.*

Demonstração. Defina $g_i = f_i \circ f_q^{-1}$ para cada $i = 1, \dots, q - 1$. 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 $(n_k)_k \rightarrow \infty$ tal que

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

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

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

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

Em seguida, mostraremos que o ponto \tilde{y} no Lema 1.5.4 é arbitrário:

Lema 1.5.5. *Dados $\varepsilon > 0$ e $\tilde{z} \in \Delta_q$ existem $\tilde{w} \in \Delta_q$ e $m \geq 1$ satisfazendo $d(F^m(\tilde{w}), \tilde{z}) < \varepsilon$.*

Demonstração. Dados $\varepsilon > 0$ e $\tilde{z} \in \Delta_q$, considere $\tilde{U} =$ bola aberta de centro \tilde{z} e raio $\varepsilon/2$. Pelo Lema 1.5.3 e pela observação (1.5.3), podemos encontrar um conjunto finito $\mathcal{H} \subset \mathcal{G}$ tal que os conjuntos $\tilde{h}^{-1}(\tilde{U})$, $h \in \mathcal{H}$ 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 1.5.4 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 $\tilde{y} \in \tilde{h}^{-1}(\tilde{U})$. Então,

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

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 1.5.5, mostraremos que é possível tomar $\tilde{x} = \tilde{y}$ no Lema 1.5.4:

Lema 1.5.6 (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 1.5.5 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 de 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 1.5.5 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 $\varepsilon_{j+1} < \varepsilon_j$ 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 $i < j$,

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

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

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

Isto completa a demonstração do lema. \square

Agora estamos prontos para concluir a demonstração do Teorema 1.5.1. Para tal, consideremos a função

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

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 $\tilde{y} \in V$. Isso segue imediatamente do fato de que ϕ é dada pelo ínfimo de uma família de funções contínuas. Então (Exercício 1.5.4), ϕ admite algum ponto de continuidade \tilde{a} . Vamos mostrar que este ponto satisfaz a conclusão do Teorema 1.5.1.

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,

$$d(F^n(\tilde{y}), \tilde{y}) \geq \beta \quad \text{para todo } \tilde{y} \in V \text{ e todo } n \geq 1. \quad (1.5.4)$$

Por outro lado, de acordo com (1.5.3), 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

$$d(\tilde{z}, \tilde{w}) < \alpha \quad \Rightarrow \quad d(\tilde{h}(\tilde{z}), \tilde{h}(\tilde{w})) < \beta \quad \text{para todo } h \in \mathcal{H}. \quad (1.5.5)$$

Pelo Lema 1.5.6, existe $n \geq 1$ tal que $d(\tilde{x}, F^n(\tilde{x})) < \alpha$. Então, usando (1.5.5) e lembrando que F comuta com todo \tilde{h} ,

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

Isto contradiz (1.5.4). Esta contradição mostra que $\phi(\tilde{a}) = 0$, tal como afirmamos.

Em outras palavras, existe $(n_k)_k \rightarrow \infty$ tal que $d(F^{n_k}(\tilde{a}), \tilde{a}) \rightarrow 0$ quando $k \rightarrow \infty$. Isto significa que (1.5.2) é satisfeita e, portanto, a prova do teorema está completa.

1.5.2 Exercícios

1.5.1. Mostre, por meio de exemplos, que a conclusão do Teorema 1.5.1 é falsa, em geral, se as transformações f_i não comutam.

1.5.2. Seja \mathcal{G} o grupo abeliano gerado por homeomorfismos $f_1, \dots, f_q : M \rightarrow M$ 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.

1.5.3. Mostre que se $\varphi : M \rightarrow \mathbb{R}$ 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 $x \in M$.

1.5.4. Mostre que se $\varphi : M \rightarrow \mathbb{R}$ 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 .

1.5.5. Seja $f : M \rightarrow M$ 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 $1 \leq j \leq k$.

1.5.6. Sejam $f_1, \dots, f_q : M \rightarrow M$ 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 $f_1^{n_1} \dots f_q^{n_q}(U)$ intersecta U . Prove que $\Omega(f_1, \dots, f_q)$ é um compacto, não-vazio.

Capítulo 2

Existência de Medidas Invariantes

Neste capítulo provaremos o seguinte resultado, que garante a existência de medidas invariantes para uma classe muito ampla de transformações:

Teorema 2.1 (Existência de medidas invariantes). *Seja $f : M \rightarrow M$ uma transformação contínua num espaço métrico compacto. Então existe pelo menos uma medida de probabilidade em M que é invariante por f .*

O ponto principal na demonstração é considerar uma certa topologia no conjunto $\mathcal{M}_1(M)$ das medidas de probabilidade em M , que chamamos de topologia fraca*. A ideia é que duas medidas são consideradas próximas se as integrais que elas dão a (muitas) funções contínuas limitadas estão próximas. A definição precisa e as propriedades da topologia fraca* serão apresentadas na Seção 2.1. A propriedade crucial, que torna esta topologia tão útil para provar o teorema de existência, é que ela faz de $\mathcal{M}_1(M)$ um espaço compacto (Teorema 2.1.5).

A demonstração do Teorema 2.1 será dada na Seção 2.2. Também veremos, por meio de exemplos, que as hipóteses de continuidade e compacidade não podem ser omitidas.

Na Seção 2.3 vamos inserir a construção da topologia fraca* numa perspectiva mais ampla de Análise Funcional e também aproveitaremos a oportunidade para introduzir a noção de operador de Koopman de uma transformação, que será muito útil a seguir. Em particular, ele nos permite usar ferramentas de Análise Funcional para dar uma demonstração alternativa do Teorema 2.1, como veremos.

Na Seção 2.4 descreveremos construções explícitas de medidas invariantes para duas classes importantes de sistemas: produtos semi-diretos e extensões naturais (ou limites inversos) de transformações não invertíveis.

Finalmente, na Seção 2.5 discutiremos algumas aplicações importantes da ideia de recorrência múltipla (Seção 1.5) no âmbito da Aritmética Combinatória. O Teorema 2.1.5 tem um papel importante nos argumentos e essa é a razão de termos adiado esta discussão para o presente capítulo.

2.1 Topologia fraca*

Nesta seção M sempre será um espaço métrico. O nosso objetivo é definir a topologia fraca* no conjunto $\mathcal{M}_1(M)$ das medidas borelianas de probabilidade em M e discutir as suas propriedades principais.

Representamos por $d(\cdot, \cdot)$ a função distância em M e por $B(x, \delta)$ a bola de centro $x \in M$ e raio $\delta > 0$. Dado $B \subset M$, denotamos $d(x, B) = \inf\{d(x, y) : y \in B\}$ e chamamos δ -vizinhança de B ao conjunto B^δ dos pontos $x \in M$ tais que $d(x, B) < \delta$.

2.1.1 Definição e propriedades da topologia fraca*

Dada uma medida $\mu \in \mathcal{M}_1(M)$, um conjunto finito $\Phi = \{\phi_1, \dots, \phi_N\}$ de funções contínuas limitadas $\phi_i : M \rightarrow \mathbb{R}$ e um número $\varepsilon > 0$, definimos

$$V(\mu, \Phi, \varepsilon) = \{\nu \in \mathcal{M}_1(M) : |\int \phi_i d\nu - \int \phi_i d\mu| < \varepsilon \text{ para todo } i\}. \quad (2.1.1)$$

Note que a interseção de dois quaisquer conjuntos desta forma contém algum conjunto desta forma. Isto assegura que a família $\{V(\mu, \Phi, \varepsilon) : \Phi, \varepsilon\}$ pode ser tomada como base de vizinhanças de cada $\mu \in \mathcal{M}_1(M)$.

A topologia fraca* é a topologia definida por estas bases de vizinhanças. Em outras palavras, os abertos da topologia fraca* são os conjuntos $\mathcal{A} \subset \mathcal{M}_1(M)$ tais que para todo elemento $\mu \in \mathcal{A}$ existe algum $V(\mu, \Phi, \varepsilon)$ contido em \mathcal{A} . Observe que esta topologia depende apenas da topologia de M e não da sua distância. Observe também que ela é Hausdorff: a Proposição A.3.3 implica que se μ e ν são probabilidades distintas então existe $\varepsilon > 0$ e alguma função contínua limitada $\phi : M \rightarrow \mathbb{R}$ tal que $V(\mu, \{\phi\}, \varepsilon) \cap V(\nu, \{\phi\}, \varepsilon) = \emptyset$.

Lema 2.1.1. *Uma sequência $(\mu_n)_{n \in \mathbb{N}}$ converge para uma medida $\mu \in \mathcal{M}_1(M)$ na topologia fraca* se, e somente se,*

$$\int \phi d\mu_n \rightarrow \int \phi d\mu \quad \text{para toda função contínua limitada } \phi : M \rightarrow \mathbb{R}.$$

Demonstração. Para provar a parte “somente se”, considere qualquer função contínua limitada ϕ e forme o conjunto $\Phi = \{\phi\}$. Como $(\mu_n)_n \rightarrow \mu$, temos que dado qualquer $\varepsilon > 0$ existe uma ordem \bar{n} tal que $\mu_n \in V(\mu, \Phi, \varepsilon)$ para todo $n \geq \bar{n}$. Mas isto significa, precisamente, que

$$|\int \phi d\mu_n - \int \phi d\mu| < \varepsilon \quad \text{para todo } n \geq \bar{n}.$$

Em outras palavras, a sequência $(\int \phi d\mu_n)_n$ converge para $\int \phi d\mu$.

A recíproca afirma que se $(\int \phi d\mu_n)_n$ converge para $\int \phi d\mu$, para toda função contínua limitada ϕ então, dados quaisquer Φ e ε existe uma ordem a partir da qual $\mu_n \in V(\mu, \Phi, \varepsilon)$. Para verificar esse fato, escrevemos $\Phi = \{\phi_1, \dots, \phi_N\}$. A hipótese garante que para cada i existe \bar{n}_i tal que

$$|\int \phi_i d\mu_n - \int \phi_i d\mu| < \varepsilon \quad \text{para todo } n \geq \bar{n}_i.$$

Tomando $\bar{n} = \max\{\bar{n}_1, \dots, \bar{n}_N\}$, temos $\mu_n \in V(\mu, \Phi, \varepsilon)$ para todo $n \geq \bar{n}$. \square

2.1.2 Teorema Portmanteau

Vamos agora discutir outras maneiras úteis de definir a topologia fraca*. De fato, nas relações (2.1.3), (2.1.4), (2.1.5) e (2.1.2) apresentaremos outras escolhas naturais de vizinhanças de uma probabilidade $\mu \in \mathcal{M}_1$. Em seguida, no Teorema 2.1.2, provaremos que todas estas escolhas dão origem à mesma topologia em $\mathcal{M}_1(M)$, que é a topologia fraca*.

Uma variação direta da definição da topologia fraca* é obtida tomando como base de vizinhanças a família de conjuntos

$$V(\mu, \Psi, \varepsilon) = \{\eta \in \mathcal{M}_1(M) : |\int \psi_i d\eta - \int \psi_i d\mu| < \varepsilon \text{ para todo } i\}. \quad (2.1.2)$$

onde $\varepsilon > 0$ e $\Psi = \{\psi_1, \dots, \psi_N\}$ é uma família finita de funções Lipschitz. A próxima definição é formulada em termos dos fechados de M . Dada qualquer família finita $\mathcal{F} = \{F_1, \dots, F_N\}$ de fechados de M e dado qualquer $\varepsilon > 0$, considere

$$V_f(\mu, \mathcal{F}, \varepsilon) = \{\nu \in \mathcal{M}_1 : \nu(F_i) < \mu(F_i) + \varepsilon \text{ para todo } i\}. \quad (2.1.3)$$

A construção seguinte é análoga, apenas substituindo fechados por abertos. Dada qualquer família finita $\mathcal{A} = \{A_1, \dots, A_N\}$ de abertos de M e dado qualquer $\varepsilon > 0$, considere

$$V_a(\mu, \mathcal{A}, \varepsilon) = \{\nu \in \mathcal{M}_1 : \nu(A_i) > \mu(A_i) - \varepsilon \text{ para todo } i\}. \quad (2.1.4)$$

Chamamos *conjunto de continuidade* de μ qualquer conjunto boreliano B cujo bordo ∂B tem medida nula para μ . Dada uma família finita $\mathcal{B} = \{B_1, \dots, B_N\}$ de conjuntos de continuidade de μ e dado qualquer $\varepsilon > 0$, considere

$$V_c(\mu, \mathcal{B}, \varepsilon) = \{\nu \in \mathcal{M}_1 : |\mu(B_i) - \nu(B_i)| < \varepsilon \text{ para todo } i\}. \quad (2.1.5)$$

Dadas duas topologias \mathcal{T}_1 e \mathcal{T}_2 num mesmo conjunto, dizemos que \mathcal{T}_1 é *mais fraca* que \mathcal{T}_2 (ou \mathcal{T}_2 é *mais forte* que \mathcal{T}_1) se todo subconjunto que é aberto para \mathcal{T}_1 também é aberto para \mathcal{T}_2 . Dizemos que as duas topologias são equivalentes se elas contêm exatamente os mesmos abertos.

Teorema 2.1.2. *As topologias definidas pelas bases de vizinhanças (2.1.1), (2.1.3), (2.1.4), (2.1.5) e (2.1.2) são todas equivalentes.*

Demonstração. É claro que a topologia (2.1.2) é mais fraca que a topologia (2.1.1), já que toda função Lipschitz é contínua.

Para mostrar que a topologia (2.1.3) é mais fraca que a topologia (2.1.2), considere qualquer família finita $\mathcal{F} = \{F_1, \dots, F_N\}$ de subconjuntos fechados de M . De acordo com o Lema A.3.4, para cada $\delta > 0$ e cada i , existe uma função Lipschitz $\psi_i : M \rightarrow [0, 1]$ tal que $\mathcal{X}_{F_i} \leq \psi_i \leq \mathcal{X}_{F_i^\delta}$. Observe que $\cap_\delta F_i^\delta = F_i$,

porque F_i é fechado, e portanto $\mu(F_i^\delta) \rightarrow \mu(F_i)$ quando $\delta \rightarrow 0$. Fixe $\delta > 0$ suficientemente pequeno para que $\mu(F_i^\delta) < \varepsilon/2$ para todo i . Seja Ψ o conjunto das funções ψ_1, \dots, ψ_N obtidas deste modo. Observe que

$$\left| \int \psi_i d\nu - \int \psi_i d\mu \right| < \varepsilon/2 \Rightarrow \nu(F_i) - \mu(F_i^\delta) < \varepsilon/2 \Rightarrow \nu(F_i) \leq \mu(F_i) + \varepsilon$$

para todo i . Em outras palavras, $V(\mu, \Psi, \varepsilon/2)$ está contido em $V_f(\mu, \mathcal{F}, \varepsilon)$.

É fácil ver que as topologias (2.1.3) e (2.1.4) são equivalentes. De fato, seja $\mathcal{F} = \{F_1, \dots, F_n\}$ uma família finita de fechados e seja $\mathcal{A} = \{A_1, \dots, A_N\}$, onde cada A_i é o complementar de F_i . É claro que

$$\begin{aligned} V_f(\mu, \mathcal{F}, \varepsilon) &= \{\nu \in \mathcal{M}_1 : \nu(F_i) < \mu(F_i) + \varepsilon \text{ para todo } i\} \\ &= \{\nu \in \mathcal{M}_1 : \nu(A_i) > \mu(A_i) - \varepsilon \text{ para todo } i\} = V_a(\mu, \mathcal{A}, \varepsilon). \end{aligned}$$

Em seguida, vamos mostrar que a topologia (2.1.5) é mais fraca que as topologias equivalentes (2.1.3) e (2.1.4). Dada qualquer família finita $\mathcal{B} = \{B_1, \dots, B_N\}$ de conjuntos de continuidade de μ seja, para cada i , F_i o fecho e A_i o interior de B_i . Denote $\mathcal{F} = \{F_1, \dots, F_N\}$ e $\mathcal{A} = \{A_1, \dots, A_N\}$. Como $\mu(F_i) = \mu(B_i) = \mu(A_i)$,

$$\begin{aligned} \nu(F_i) < \mu(F_i) + \varepsilon &\Rightarrow \nu(B_i) < \mu(B_i) + \varepsilon \\ \nu(A_i) > \mu(A_i) - \varepsilon &\Rightarrow \nu(B_i) > \mu(B_i) - \varepsilon \end{aligned}$$

para todo i . Isto significa que $V_f(\mu, \mathcal{F}, \varepsilon) \cap V_a(\mu, \mathcal{A}, \varepsilon)$ está contido em $V_c(\mu, \mathcal{B}, \varepsilon)$.

Finalmente, provemos que a topologia (2.1.1) é mais fraca que a topologia (2.1.5). Seja $\Phi = \{\phi_1, \dots, \phi_N\}$ uma família finita de funções contínuas limitadas. Fixemos um inteiro ℓ tal que $\sup |\phi_i(x)| < \ell$ para todo i . Para cada i , as pré-imagens $\phi_i^{-1}(s)$, $s \in [-\ell, \ell]$ são disjuntas duas-a-duas. Portanto, $\mu(\phi_i^{-1}(s)) = 0$ exceto para um conjunto enumerável de valores de s . Em particular, podemos escolher $k \in \mathbb{N}$ e pontos $-\ell = t_0 < t_1 < \dots < t_{k-1} < t_k = \ell$ tais que $t_j - t_{j-1} < \varepsilon/2$ e $\mu(\{\phi_i^{-1}(t_j)\}) = 0$ para todo j . Então cada

$$B_{i,j} = \phi_i^{-1}((t_{j-1}, t_j])$$

é conjunto de continuidade para μ . Além disso,

$$\sum_{j=1}^k t_j \mu(B_{i,j}) \geq \int \phi_i d\mu \geq \sum_{j=1}^k t_{j-1} \mu(B_{i,j}) > \sum_{j=1}^k t_j \mu(B_{i,j}) - \varepsilon/2$$

e valem desigualdades análogas para as integrais relativamente a ν . Segue que

$$\left| \int \phi_i d\mu - \int \phi_i d\nu \right| \leq \sum_{j=1}^k \ell |\mu(B_{i,j}) - \nu(B_{i,j})| + \varepsilon/2 \quad (2.1.6)$$

para todo i . Denote $\mathcal{B} = \{B_{i,j} : i = 1, \dots, N \text{ e } j = 1, \dots, k\}$. Então a relação (2.1.6) implica que $V_c(\mu, \mathcal{B}, \varepsilon/(2k\ell))$ está contido em $V(\mu, \Phi, \varepsilon)$. \square

Veremos no Exercício 2.1.3 que a topologia fraca* em $\mathcal{M}_1(M)$ é separável sempre que o espaço métrico M é separável. Na próxima seção vamos mostrar que, nas mesmas condições, ela também é metrizável.

2.1.3 A topologia fraca* é metrizável

Agora suponha que o espaço métrico M é separável. Então $\mathcal{M}_1(M)$ munido da topologia fraca* é um espaço separável (Exercício 2.1.3). Nesta seção vamos mostrar que esse espaço também é metrizável: é possível exibir distâncias em $\mathcal{M}_1(M)$ que induzem a topologia fraca*.

Dados $\mu, \nu \in \mathcal{M}_1(M)$, defina $D(\mu, \nu)$ como sendo o ínfimo de todos os números $\delta > 0$ tais que

$$\mu(B) < \nu(B^\delta) + \delta \quad \text{e} \quad \nu(B) < \mu(B^\delta) + \delta \quad \text{para todo boreliano } B. \quad (2.1.7)$$

Lema 2.1.3. *A função D é uma distância em $\mathcal{M}_1(M)$.*

Demonstração. Começemos por mostrar que se $D(\mu, \nu) = 0$ então $\mu = \nu$. De fato, a hipótese implica que

$$\mu(B) \leq \nu(\bar{B}) \quad \text{e} \quad \nu(B) \leq \mu(\bar{B})$$

para todo subconjunto boreliano $B \subset M$, onde \bar{B} representa o fecho. Quando B é um fechado estas desigualdades significam que $\mu(B) = \nu(B)$. Como vimos anteriormente, duas medidas que coincidem nos subconjuntos fechados são necessariamente iguais.

Deixamos a verificação das demais condições na definição de distância ao cuidado do leitor (Exercício 2.1.5). \square

Esta distância é denominada *métrica de Levy-Prohorov*. No que segue representaremos por $B_D(\mu, r)$ a bola relativamente a D com centro em $\mu \in \mathcal{M}_1(M)$ e raio $r > 0$.

Proposição 2.1.4. *Se M é espaço métrico separável então a topologia induzida pela distância D coincide com a topologia fraca* em $\mathcal{M}_1(M)$.*

Demonstração. Seja $\varepsilon > 0$ e seja $\mathcal{F} = \{F_1, \dots, F_N\}$ uma família finita de fechados de M . Fixe $\delta \in (0, \varepsilon/2)$ tal que $\mu(F_i^\delta) < \mu(F_i) + \varepsilon/2$ para todo i . Se $\nu \in B_D(\mu, \delta)$ então

$$\nu(F_i) < \mu(F_i^\delta) + \delta < \mu(F_i) + \varepsilon \quad \text{para todo } i,$$

o que significa que $\nu \in V_f(\mu, \mathcal{F}, \varepsilon)$. Isto mostra que a topologia induzida pela distância D é mais forte que a topologia (2.1.3) a qual, como vimos anteriormente, é equivalente à topologia fraca*.

Resta provar que se M é separável então a topologia fraca* é mais forte que a topologia induzida por D . Dado $\varepsilon > 0$, fixemos $\delta \in (0, \varepsilon/3)$. Para isso, seja $\{p_1, p_2, \dots\}$ um subconjunto enumerável denso de M . Para cada j , as esferas $\partial B(p_j, r) \subset \{x : d(x, p_j) = r\}$, $r > 0$ são disjuntas duas-a-duas. Portanto, é possível encontrar $r > 0$ arbitrariamente pequeno tal que $\mu(\partial B(p_j, r)) = 0$ para todo j . Fixemos um tal r , com $r \in (0, \delta/3)$. A família $\{B(p_j, r) : j = 1, 2, \dots\}$ é uma cobertura enumerável de M por conjuntos de continuidade de μ . Fixemos $k \geq 1$ tal que o conjunto $U = \cup_{j=1}^k B(p_j, r)$ satisfaz

$$\mu(U) > 1 - \delta. \quad (2.1.8)$$

Em seguida, consideremos a partição (finita) \mathcal{P} de U definida pela família $\{B(p_j, r) : j = 1, \dots, k\}$: os elementos de \mathcal{P} são os conjuntos maximais $P \subset U$ tais que, para cada j , ou P está contido em $B(p_j, r)$ ou P é disjunto de $B(p_j, r)$. Veja a Figura 2.1. Em seguida, seja \mathcal{E} a família de todos as uniões finitas de elementos de \mathcal{P} . Note que o bordo de todo elemento de \mathcal{E} está contido na união dos bordos dos $B(p_j, r)$, $1 \leq j \leq k$ e, conseqüentemente, tem medida nula. Ou seja, todo elemento de \mathcal{E} é conjunto de continuidade de μ .

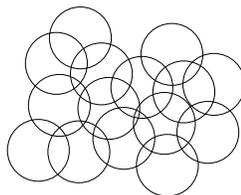


Figura 2.1: Partição definida por uma cobertura

Se $\nu \in V_c(\mu, \mathcal{E}, \delta)$ então

$$|\mu(E) - \nu(E)| < \delta \quad \text{para todo } E \in \mathcal{E}. \quad (2.1.9)$$

Em particular, (2.1.8) juntamente com (2.1.9) implicam que

$$\nu(U) > 1 - 2\delta. \quad (2.1.10)$$

Agora, dado um subconjunto boreliano B qualquer, denote por E_B a união dos elementos de \mathcal{P} que intersectam B . Então $E_B \in \mathcal{B}$ e portanto a relação (2.1.9) dá que

$$|\mu(E_B) - \nu(E_B)| < \delta.$$

Observe que B está contido em $E_B \cup U^c$. Além disso, $E_B \subset B^\delta$ porque todo elemento de \mathcal{P} tem diâmetro menor que $2r < \delta$. Estes fatos, juntamente com (2.1.8) e (2.1.10) implicam que

$$\begin{aligned} \mu(B) &\leq \mu(E_B) + \delta < \nu(E_B) + 2\delta \leq \nu(B^\delta) + 2\delta \\ \nu(B) &\leq \nu(E_B) + 2\delta < \mu(E_B) + 3\delta \leq \mu(B^\delta) + 3\delta. \end{aligned}$$

Como $3\delta < \varepsilon$, estas relações implicam que $\nu \in B_D(\mu, \varepsilon)$. □

Pode mostrar-se que se M é espaço métrico completo separável então a métrica de Levy-Prohorov em $\mathcal{M}_1(M)$ é completa (e separável, de acordo com o Exercício 2.1.3). Veja, por exemplo, o Teorema 6.8 do livro de Billingsley [Bil68].

2.1.4 A topologia fraca* é compacta

Nesta seção suporemos que o espaço métrico M é compacto. Vamos provar o seguinte fato:

Teorema 2.1.5. *O espaço $\mathcal{M}_1(M)$ munido da topologia fraca* é compacto.*

Como já sabemos que $\mathcal{M}_1(M)$ é metrizável, basta provar o seguinte:

Proposição 2.1.6. *Toda sequência $(\mu_k)_{k \in \mathbb{N}}$ em $\mathcal{M}_1(M)$ admite alguma subsequência que é convergente na topologia fraca*.*

Demonstração. Seja $\{\phi_n : n \in \mathbb{N}\}$ um subconjunto enumerável denso na bola unitária de $C^0(M)$ (lembre do Teorema A.3.13). Para cada $n \in \mathbb{N}$, a sequência de números reais $\int \phi_n d\mu_k$, $k \in \mathbb{N}$ é limitada por 1. Portanto, para cada $n \in \mathbb{N}$ existe uma sequência $(k_j^n)_{j \in \mathbb{N}}$ tal que

$$\int \phi_n d\mu_{k_j^n} \text{ converge para algum número } \Phi_n \in \mathbb{R} \text{ quando } j \rightarrow \infty.$$

Além disso, cada sequência $(k_j^{n+1})_{j \in \mathbb{N}}$ pode ser escolhida como subsequência da anterior $(k_j^n)_{j \in \mathbb{N}}$. Definamos $\ell_j = k_j^j$ para cada $j \in \mathbb{N}$. Por construção, $(\ell_j)_{j \in \mathbb{N}}$ é uma subsequência de $(k_j^n)_{j \in \mathbb{N}}$ a menos de um número finito de termos. Logo

$$\left(\int \phi_n d\mu_{\ell_j} \right)_j \rightarrow \Phi_n \text{ para todo } n \in \mathbb{N}.$$

Daqui se deduz facilmente que

$$\Phi(\varphi) = \lim_j \int \varphi d\mu_{\ell_j} \text{ existe} \quad (2.1.11)$$

para toda função $\varphi \in C^0(M)$. De fato, suponha primeiro que φ está na bola unitária de $C^0(M)$. Dado qualquer $\varepsilon > 0$ podemos encontrar $n \in \mathbb{N}$ tal que $\|\varphi - \phi_n\| \leq \varepsilon$. Então

$$\left| \int \varphi d\mu_{\ell_j} - \int \phi_n d\mu_{\ell_j} \right| \leq \varepsilon$$

para todo j . Como $\int \phi_n d\mu_{\ell_j}$ converge (para Φ_n), segue que

$$\limsup_j \int \varphi d\mu_{\ell_j} - \liminf_j \int \varphi d\mu_{\ell_j} \leq 2\varepsilon.$$

Como ε é arbitrário, concluímos que $\lim_j \int \varphi d\mu_{\ell_j}$ existe. Isto prova (2.1.11) quando a função está na bola unitária. O caso geral reduz-se imediatamente a esse, substituindo φ por $\varphi/\|\varphi\|$. Assim, completamos a prova de (2.1.11).

Finalmente, é claro que o operador $\Phi : C^0(M) \rightarrow \mathbb{R}$ definido por (2.1.11) é linear e positivo: $\Phi(\varphi) \geq \min \varphi \geq 0$ sempre que $\varphi \geq 0$ em todo ponto. Além disso, $\Phi(1) = 1$. Logo, pelo Teorema A.3.11, existe alguma probabilidade boreliana μ em M tal que $\Phi(\varphi) = \int \varphi d\mu$ para toda função contínua φ . Agora a igualdade em (2.1.11) pode ser reescrita

$$\int \varphi d\mu = \lim_j \int \varphi d\mu_{\ell_j} \text{ para toda } \varphi \in C^0(M).$$

De acordo com o Lema 2.1.1, isto quer dizer que a subsequência $(\mu_{\ell_j})_{j \in \mathbb{N}}$ converge para μ na topologia fraca*. \square

Conforme observamos anteriormente, o Teorema 2.1.5 é uma consequência imediata da proposição que acabamos de demonstrar.

2.1.5 Teorema de Prohorov

O teorema que vamos enunciar nesta seção dá um critério muito geral para compacidade de uma família de medidas de probabilidade. De fato, a classe de espaços métricos M a que ele se aplica inclui praticamente todos os exemplos interessantes.

Definição 2.1.7. Um conjunto \mathcal{M} de medidas num espaço topológico é *justo* se para todo $\varepsilon > 0$ existe um subconjunto compacto K tal que $\mu(K^c) < \varepsilon$ para toda medida $\mu \in \mathcal{M}$.

Note que quando \mathcal{M} se reduz a uma única medida esta definição corresponde exatamente à Definição A.3.6. Claramente, esta propriedade é hereditária: se um conjunto é justo então todo subconjunto dele também é justo. Note também que se M é um espaço métrico compacto então o espaço $\mathcal{M}_1(M)$ de todas as medidas de probabilidade constitui um conjunto justo. Portanto o resultado a seguir generaliza o Teorema 2.1.5:

Teorema 2.1.8 (Prohorov). *Seja M um espaço métrico separável completo. Um conjunto $\mathcal{K} \subset \mathcal{M}_1(M)$ é justo se, e somente se, toda sequência em \mathcal{K} possui alguma subsequência convergente em $\mathcal{M}_1(M)$.*

Demonstração. Vamos provar apenas a condição necessária, que é a parte mais útil do enunciado. No Exercício 2.1.8 convidamos o leitor a provar a recíproca.

Suponha que \mathcal{K} é justo. Considere uma sequência crescente $(K_l)_l$ de subconjuntos compactos de M tais que $\eta(K_l^c) \leq 1/l$ para todo l e todo $\eta \in \mathcal{K}$. Fixe uma sequência qualquer $(\mu_n)_n$ em \mathcal{K} . Afirmamos, inicialmente, que para cada l existe uma subsequência $(n_j)_j$ e existe uma medida ν_l em M tal que $\nu_l(K_l^c) = 0$ e $(\mu_{n_j} | K_l)_j$ converge para ν_l , no sentido de que

$$\int_{K_l} \psi d\mu_{n_j} \rightarrow \int_{K_l} \psi d\nu_l \quad \text{para toda função contínua } \psi : K_l \rightarrow \mathbb{R}. \quad (2.1.12)$$

Este fato é uma consequência simples do Teorema 2.1.5: a menos de restringir a uma subsequência, podemos supor que o limite $b_l = \lim_n \mu_n(K_l)$ existe (note que $1 \geq b_l \geq 1 - 1/l$); segue do teorema que a sequência de restrições normalizadas $((\mu_n | K_l) / \mu_n(K_l))_n$ admite subsequência convergente para alguma probabilidade $\eta_l \in \mathcal{M}_1(K_l)$; para concluir a prova da afirmação basta tomar considerar η_l uma probabilidade em M , com $\eta_l(K_l^c) = 0$, e tomar $\nu_l = b_l \eta_l$.

Em seguida, usando um argumento diagonal análogo ao da Proposição 2.1.6, podemos escolher a subsequência $(n_j)_j$ de tal forma que (2.1.12) vale, simultaneamente, para todo $l \geq 1$. Observe que a sequência $(\eta_l)_l$ é monótona não decrescente: dados $k > l$ e qualquer função contínua $\phi : M \rightarrow [0, 1]$,

$$\int \phi d\nu_l = \lim_j \int_{K_l} \phi d\mu_{n_j} \leq \lim_j \int_{K_k} \phi d\mu_{n_j} = \int \phi d\nu_k.$$

Analogamente, para quaisquer $k > l$ e qualquer função contínua $\phi : M \rightarrow [0, 1]$,

$$\int \phi d\nu_k - \int \phi d\nu_l = \lim_j \int_{K_k \setminus K_l} \phi d\mu_{n_j} \leq \limsup_j \mu_{n_j}(K_l^c) \leq 1/l.$$

Usando o Exercício A.3.5, podemos traduzir estes fatos em termos de medidas de conjuntos: para todo $k > l$ e todo boreliano $E \subset M$,

$$\nu_l(E) \leq \nu_k(E) \leq \nu_l(E) + 1/l. \quad (2.1.13)$$

Defina $\nu(E) = \lim_l \nu_l(E)$ para cada boreliano E . Afirmamos que ν é uma probabilidade em M . É imediato da definição que $\nu(\emptyset) = 0$ e que ν é aditiva. Além disso, $\nu(M) = \lim_l \nu(K_l) = \lim_l b_l = 1$. Para mostrar que ν é σ -aditiva, usaremos o critério da continuidade no vazio. Considere qualquer seqüência decrescente $(B_n)_n$ de borelianos de M com $\cap_n B_n = \emptyset$. Dado $\varepsilon > 0$, escolha l tal que $1/l < \varepsilon$. Como ν_l é σ -aditiva, o Teorema A.1.14 dá que $\nu_l(B_n) < \varepsilon$ para todo n suficientemente grande. Logo, $\nu(B_n) \leq \nu_l(B_n) + 1/l < 2\varepsilon$ para todo n suficientemente grande. Isto prova que $(\nu(B_n))_n$ converge para zero e, pelo Teorema A.1.14, segue que ν é realmente σ -aditiva.

A definição de ν implica (veja o Exercício 2.1.1 ou o Exercício 2.1.4) que $(\nu_l)_l$ converge para ν na topologia fraca*. Portanto, dado $\varepsilon > 0$ e dada uma função contínua limitada $\varphi : M \rightarrow \mathbb{R}$ qualquer, $|\int \varphi d\nu_l - \int \varphi d\nu| < \varepsilon$ para todo l suficientemente grande. Fixe l satisfazendo também $\sup |\varphi|/l < \varepsilon$. Então,

$$\left| \int \varphi d\mu_{n_j} - \int \varphi d\nu_l \right| \leq \left| \int_{K_l^c} \varphi d\mu_{n_j} \right| + \left| \int_{K_l} \varphi d\mu_{n_j} - \int_{K_l} \varphi d\nu_l \right| \leq 2\varepsilon$$

para todo j suficientemente grande. Isto mostra que $|\int \varphi d\mu_{n_j} - \int \varphi d\nu| < 3\varepsilon$ sempre que j é suficientemente grande e, portanto, $(\mu_{n_j})_j$ converge para ν na topologia fraca*. \square

2.1.6 Exercícios

2.1.1. Seja M um espaço métrico e seja $(\mu_n)_n$ uma seqüência em $\mathcal{M}_1(M)$. Mostre que as seguintes condições são todas equivalentes:

1. $(\mu_n)_n$ converge para uma probabilidade μ na topologia fraca*.
2. $\limsup_n \mu_n(F) \leq \mu(F)$ para todo fechado $F \subset M$.
3. $\liminf_n \mu_n(A) \geq \mu(A)$ para todo aberto $A \subset M$.
4. $\lim_n \mu_n(B) = \mu(B)$ para todo conjunto de continuidade B de μ .
5. $\lim_n \int \psi d\mu_n = \int \psi d\mu$ para toda função Lipschitz $\psi : M \rightarrow \mathbb{R}$.

2.1.2. Fixe qualquer subconjunto denso \mathcal{F} da bola unitária de $C^0(M)$. Mostre que uma seqüência $(\mu_n)_{n \in \mathbb{N}}$ de probabilidades em M converge na topologia fraca* para alguma $\mu \in \mathcal{M}_1(M)$ se, e somente se,

$$\int \phi d\mu_n \text{ converge para } \int \phi d\mu, \quad \text{para todo } \phi \in \mathcal{F}.$$

2.1.3. Mostre que o conjunto das medidas com suporte finito é denso em $\mathcal{M}_1(M)$, relativamente à topologia fraca*. Supondo que o espaço métrico M é separável, conclua que $\mathcal{M}_1(M)$ também é separável.

2.1.4. A topologia uniforme em $\mathcal{M}_1(M)$ é definida pelo sistema de vizinhanças

$$V_u(\mu, \varepsilon) = \{\nu \in \mathcal{M}_1(M) : |\mu(B) - \nu(B)| < \varepsilon \text{ para todo } B \in \mathcal{B}\}$$

e a topologia pontual é definida pelo sistema de vizinhanças

$$V_p(\mu, \mathcal{B}, \varepsilon) = \{\nu \in \mathcal{M}_1(M) : |\mu(B_i) - \nu(B_i)| < \varepsilon \text{ para } i\}$$

onde $\varepsilon > 0$, $n \geq 1$ e $\mathcal{B} = \{B_1, \dots, B_n\}$ é uma família finita de conjuntos mensuráveis. Verifique que a topologia uniforme é mais forte que a topologia pontual a qual, por sua vez, é mais forte que a topologia fraca*. Mostre, por meio de exemplos, que essas relações podem ser estritas.

2.1.5. Complete a demonstração do Lema 2.1.3.

2.1.6. Sejam V_k , $k = 1, 2, \dots$ variáveis aleatórias reais, ou seja, funções mensuráveis $V_k : (X, \mathcal{B}, \mu) \rightarrow \mathbb{R}$ em algum espaço de probabilidade (X, \mathcal{B}, μ) . A função de distribuição de V_k é a função monótona $F_k : \mathbb{R} \rightarrow [0, 1]$ definida por $F_k(a) = \mu(\{x \in X : V_k(x) \leq a\})$. Diz-se que $(V_k)_k$ converge em distribuição para uma variável aleatória V se $\lim_k F_k(a) = F(a)$ para todo ponto de continuidade a da função de distribuição F da variável aleatória V . O que isto tem que ver com a topologia fraca*?

2.1.7. Seja $(\mu_n)_{n \in \mathbb{N}}$ uma sequência de probabilidades convergindo para uma probabilidade μ na topologia fraca*. Seja B um conjunto de continuidade para μ com $\mu(B) > 0$. Mostre que as restrições normalizadas $(\mu_n | B)/\mu_n(B)$ convergem para a restrição normalizada $(\mu | B)/\mu(B)$ quando $n \rightarrow \infty$. O que pode ser dito se substituirmos conjunto de continuidade por subconjunto fechado, ou por subconjunto aberto?

2.1.8. (Recíproca do teorema de Prohorov) Mostre que se $\mathcal{K} \subset \mathcal{M}_1(M)$ é tal que toda sequência em \mathcal{K} admite alguma subsequência convergente em $\mathcal{M}_1(M)$, então \mathcal{K} é justo.

2.2 Demonstração do teorema de existência

Começamos por introduzir uma notação útil. Dado $f : M \rightarrow M$ e qualquer medida η em M denota-se por $f_*\eta$ e chama-se iterado (ou imagem) de η por f a medida definida por $f_*\eta(B) = \eta(f^{-1}(B))$ para cada conjunto mensurável $B \subset M$. Note que η é invariante por f se, e somente se, $f_*\eta = \eta$.

Lema 2.2.1. Sejam η uma medida e ϕ uma função mensurável limitada. Então

$$\int \phi df_*\eta = \int \phi \circ f d\eta. \quad (2.2.1)$$

Demonstração. Se ϕ é a função característica de um conjunto mensurável B então a relação (2.2.1) significa que $f_*\eta(B) = \eta(f^{-1}(B))$, o que é verdade por definição. Pela linearidade da integral, segue que (2.2.1) vale sempre que ϕ é uma função simples. Finalmente, como toda função mensurável limitada pode ser aproximada uniformemente por funções simples (veja a Proposição A.1.33), segue que a conclusão do lema é verdadeira em geral. \square

Proposição 2.2.2. *A aplicação $f_* : \mathcal{M}_1(M) \rightarrow \mathcal{M}_1(M)$ é contínua relativamente à topologia fraca*.*

Demonstração. Seja $\varepsilon > 0$ e $\Phi = \{\phi_1, \dots, \phi_n\}$ uma família qualquer de funções contínuas limitadas. Como f é contínua, a família $\Psi = \{\phi_1 \circ f, \dots, \phi_n \circ f\}$ também consiste de funções contínuas limitadas. Pelo lema anterior,

$$\left| \int \phi_i d(f_*\mu) - \int \phi_i d(f_*\nu) \right| = \left| \int (\phi_i \circ f) d\mu - \int (\phi_i \circ f) d\nu \right|$$

e portanto o lado esquerdo é menor que ε se o lado direito for menor que ε . Isto quer dizer que

$$f_*(V(\mu, \Psi, \varepsilon)) \subset V(f_*\mu, \Phi, \varepsilon) \quad \text{para todo } \mu, \Phi \text{ e } \varepsilon$$

e este último fato mostra que f_* é contínua. \square

A esta altura, o Teorema 2.1 pode ser deduzido de um resultado clássico sobre operadores contínuos em espaços vetoriais topológicos, conhecido como teorema do ponto fixo de Schauder-Tychonoff.

Um *espaço vetorial topológico* é um espaço vetorial V munido de uma topologia relativamente à qual as duas operações (adição de vetores e produto de um vetor por um escalar) são contínuas. Um conjunto $K \subset V$ diz-se *convexo* se $(1-t)x + ty \in K$ para todo $x, y \in K$ e todo $t \in [0, 1]$.

Teorema 2.2.3 (Schauder-Tychonoff). *Seja $F : V \rightarrow V$ uma transformação contínua num espaço vetorial topológico V . Suponha que existe um conjunto compacto convexo $K \subset V$ tal que $F(K) \subset K$. Então $F(v) = v$ para algum $v \in K$.*

O Teorema 2.1 corresponde ao caso em que $V = \mathcal{M}(M)$ é o espaço das medidas complexas, $K = \mathcal{M}_1(M)$ é o espaço das probabilidades em M , e $F = f_*$ é a ação de f em $\mathcal{M}(M)$.

Não entanto, a situação do Teorema 2.1 é bem mais simples do que o caso geral do teorema de Schauder-Tychonoff, porque o operador f_* além de ser contínuo também é *linear*. Isso permite dar uma demonstração direta e elementar do Teorema 2.1, que também tem a vantagem de fornecer alguma informação adicional sobre a medida invariante.

Para fazer isso, seja ν uma probabilidade qualquer em M : por exemplo, a medida de Dirac em um ponto qualquer. Forme a sequência de probabilidades

$$\mu_n = \frac{1}{n} \sum_{j=0}^{n-1} f_*^j \nu \quad (2.2.2)$$

onde $f_*^j \nu$ é a imagem de ν pelo iterado f^j . Pelo Teorema 2.1.5, esta sequência tem algum ponto de acumulação, ou seja, existe alguma subsequência $(n_k)_{k \in \mathbb{N}}$ e alguma probabilidade $\mu \in \mathcal{M}_1(M)$ tais que

$$\frac{1}{n_k} \sum_{j=0}^{n_k-1} f_*^j \nu \rightarrow \mu \quad (2.2.3)$$

na topologia fraca*. Agora é suficiente provar o seguinte:

Lema 2.2.4. *Todo ponto de acumulação de uma sequência $(\mu_n)_{n \in \mathbb{N}}$ do tipo (2.2.2) é uma probabilidade invariante por f .*

Demonstração. A relação (2.2.3) afirma que dada uma família $\Phi = \{\phi_1, \dots, \phi_n\}$ de funções contínuas limitadas e para todo $\varepsilon > 0$ tem-se

$$\left| \frac{1}{n_k} \sum_{j=0}^{n_k-1} \int (\phi_i \circ f^j) d\nu - \int \phi_i d\mu \right| < \varepsilon/2 \quad (2.2.4)$$

para todo i e todo k suficientemente grande. Pelo Lema 2.2.2, temos que

$$f_* \mu = f_* \left(\lim_k \frac{1}{n_k} \sum_{j=0}^{n_k-1} f_*^j \nu \right) = \lim_k \frac{1}{n_k} \sum_{j=1}^{n_k} f_*^j \nu. \quad (2.2.5)$$

Agora observe que

$$\begin{aligned} & \left| \frac{1}{n_k} \sum_{j=0}^{n_k-1} \int (\phi_i \circ f^j) d\nu - \frac{1}{n_k} \sum_{j=1}^{n_k} \int (\phi_i \circ f^j) d\nu \right| \\ &= \frac{1}{n_k} \left| \int \phi_i d\nu - \int (\phi_i \circ f^{n_k}) d\nu \right| \leq \frac{2}{n_k} \sup |\phi_i| \end{aligned}$$

e esta última expressão é menor que $\varepsilon/2$ para todo i e todo k suficientemente grande. Juntando este fato com (2.2.4), concluímos que

$$\left| \frac{1}{n_k} \sum_{j=1}^{n_k} \int (\phi_i \circ f^j) d\nu - \int \phi_i d\mu \right| < \varepsilon \quad (2.2.6)$$

para todo i e todo k suficientemente grande. Isto significa que

$$\frac{1}{n_k} \sum_{j=1}^{n_k} f_*^j \nu \rightarrow \mu$$

quando $k \rightarrow \infty$. Mas (2.2.5) significa que esta mesma sequência converge para $f_* \mu$. Por unicidade do limite, segue que $f_* \mu = \mu$. \square

Agora a demonstração do Teorema 2.1 está completa. Os exemplos simples a seguir mostram que nenhuma das duas hipóteses do teorema, continuidade e compacidade, pode ser omitida.

Exemplo 2.2.5. Considere $f : (0, 1] \rightarrow (0, 1]$ dada por $f(x) = x/2$. Suponha que f admite alguma probabilidade invariante; o objetivo é mostrar que isso não acontece. Pelo teorema de recorrência (Teorema 1.2.4), relativamente a essa probabilidade quase todo ponto de $(0, 1]$ é recorrente. Mas é imediato que não existe nenhum ponto recorrente: a órbita de qualquer $x \in (0, 1]$ converge para zero e, em particular, não acumula no ponto inicial x . Portanto, f é um exemplo de transformação contínua num espaço não compacto que não admite nenhuma probabilidade invariante.

Exemplo 2.2.6. Modificando um pouco o exemplo, podemos mostrar que o mesmo fenômeno pode ocorrer em espaços compactos, se a transformação não é contínua. Considere $f : [0, 1] \rightarrow [0, 1]$ dada por $f(x) = x/2$ se $x \neq 0$ e $f(0) = 1$. Pela mesma razão que antes, nenhum ponto $x \in (0, 1]$ é recorrente. Portanto, se existe alguma probabilidade invariante μ ela tem que dar peso total ao único ponto recorrente, que é $x = 0$. Em outras palavras, μ precisa ser a medida de Dirac δ_0 suportada em zero, que é definida por

$$\delta_0(E) = 1 \text{ se } 0 \in E \quad \text{e} \quad \delta_0(E) = 0 \text{ se } 0 \notin E.$$

Mas a medida δ_0 não é invariante por f : tomando $E = \{0\}$ temos que E tem medida 1 mas a sua pré-imagem $f^{-1}(E)$ é o conjunto vazio, que tem medida nula. Portanto, esta transformação também não tem nenhuma probabilidade invariante.

O nosso terceiro exemplo é de natureza um pouco diferente:

Exemplo 2.2.7. Consideremos $f : [0, 1] \rightarrow [0, 1]$ dada por $f(x) = x/2$. Trata-se de uma transformação contínua num espaço compacto. Logo, pelo Teorema 2.1, ela admite alguma probabilidade invariante. Pelos mesmos argumentos que usamos no caso anterior, se conclui que de fato há uma única probabilidade invariante, que é a medida de Dirac δ_0 suportada no ponto zero. Note que neste caso a medida δ_0 é de fato invariante.

Mencionamos este último caso para enfatizar as limitações do Teorema 2.1 (que são inerentes à sua grande generalidade): as medidas que ele garante existem podem ser completamente triviais; por exemplo, neste caso quando falamos de “quase todo ponto” estamos nos referindo apenas ao ponto $x = 0$. Por isso, um objetivo importante em Teoria Ergódica é encontrar medidas invariantes mais sofisticadas, com propriedades adicionais (por exemplo, serem equivalentes à medida de Lebesgue) que as tornem mais interessantes.

Como uma aplicação imediata do Teorema 2.1, temos a seguinte demonstração alternativa do teorema de recorrência de Birkhoff (Teorema 1.2.6). Suponha que $f : M \rightarrow M$ é uma transformação contínua num espaço métrico compacto. Pelo Teorema 2.1, existe alguma probabilidade f -invariante μ . Por outro lado, todo espaço métrico compacto admite uma base enumerável de abertos. Portanto, podemos aplicar o Teorema 1.2.4, para concluir que μ -quase todo ponto é recorrente. Em particular, o conjunto dos pontos recorrentes é não vazio, tal como afirma o Teorema 1.2.6.

2.2.1 Exercícios

2.2.1. Prove a seguinte generalização do Lema 2.2.4. Seja $f : M \rightarrow M$ uma transformação contínua num espaço métrico compacto, ν uma probabilidade em M e $(I_n)_n$ uma sequência de intervalos de números naturais tais que $\#I_n$ converge para infinito quando n vai para infinito. Então qualquer ponto de acumulação da sequência

$$\mu_n = \frac{1}{\#I_n} \sum_{j \in I_n} f_*^j \nu$$

é uma probabilidade f -invariante.

2.2.2. Seja $f_1, \dots, f_q : M \rightarrow M$ uma família finita qualquer de transformações contínuas num espaço métrico compacto que comutam entre si. Prove que existe alguma medida de probabilidade μ que é invariante por f_i para todo $i \in \{1, \dots, q\}$. Além disso, a conclusão permanece válida para qualquer família infinita enumerável $\{f_j : j \in \mathbb{N}\}$ de transformações que comutam entre si.

2.2.3. Seja $f : [0, 1] \rightarrow [0, 1]$ a transformação expansão decimal. Mostre que, para cada $k \geq 1$, existe alguma probabilidade invariante cujo suporte é formado por exatamente k pontos (em particular, f admite infinitas probabilidades invariantes). Determine se existem probabilidades invariantes μ tais que

- (a) o suporte de μ é infinito enumerável;
- (b) o suporte de μ é não enumerável mas tem interior vazio;
- (c) o suporte de μ tem interior não vazio mas μ é singular com relação à medida de Lebesgue m .

2.2.4. Prove o teorema de existência de medidas invariantes para fluxos: todo fluxo contínuo $(f^t)_{t \in \mathbb{R}}$ num espaço métrico compacto admite alguma medida de probabilidade invariante.

2.2.5. Mostre que a transformação $f : [-1, 1] \rightarrow [-1, 1]$, $f(x) = 1 - 2x^2$ admite alguma medida de probabilidade invariante equivalente à medida de Lebesgue no intervalo.

2.2.6. Seja $f : M \rightarrow M$ uma transformação mensurável invertível e seja m uma probabilidade em M tal que $m(A) = 0$ se, e somente se, $m(f(A)) = 0$. Dizemos que o par (f, m) é *totalmente dissipativo* se existe um conjunto mensurável $W \subset M$ cujos iterados $f^j(W)$, $j \in \mathbb{Z}$ são disjuntos dois-a-dois e a sua união tem medida total. Mostre que se (f, m) é totalmente dissipativo então f admite alguma medida invariante, σ -finita, equivalente à medida m . Essa medida é necessariamente infinita.

2.2.7. Seja $f : M \rightarrow M$ uma transformação mensurável invertível e seja m uma probabilidade em M tal que $m(A) = 0$ se, e somente se, $m(f(A)) = 0$. Dizemos que o par (f, m) é *conservativo* se não existe conjunto mensurável $W \subset M$ com medida positiva cujos iterados $f^j(W)$, $j \in \mathbb{Z}$ são disjuntos dois-a-dois. Mostre que se (f, m) é conservativo então, para todo conjunto mensurável $X \subset M$, m -quase todo ponto de X regressa a X infinitas vezes.

2.2.8. Suponha que (f, m) é conservativo. Mostre que f admite uma medida invariante σ -finita μ equivalente a m se, e somente se, existem conjuntos $X_1 \subset \dots \subset X_n \subset \dots$ com $M = \cup_n X_n$ e $m(X_n) < \infty$ para todo n , tais que a transformação de primeiro retorno f_n a cada X_n admite uma medida invariante μ_n finita e absolutamente contínua com respeito à restrição de m a X_n .

2.2.9. Dê exemplos de pares conservativos (f, m) tais que f não admite medidas invariantes finitas equivalentes a m . [Observação: Ornstein [Orn60] deu exemplos em que f não admite sequer medidas invariantes σ -finitas equivalentes a m .]

2.3 Comentários de Análise Funcional

A definição da topologia fraca* no espaço das probabilidades é um caso especial de uma construção geral em Análise Funcional que vamos lembrar a seguir. Ela nos conduzirá a introduzir uma certa isometria linear U_f do espaço $L^1(\mu)$, denominada *operador de Koopman* do sistema (f, μ) . Estes operadores são de grande utilidade em Teoria Ergódica pois permitem fazer uso de ferramentas de Análise no estudo das propriedades de medidas invariantes. Para ilustrar este fato, daremos uma demonstração alternativa do Teorema 2.1 a partir de propriedades espectrais do operador de Koopman.

2.3.1 Dualidade e topologias fracas

Seja E um espaço de Banach, ou seja, um espaço vetorial munido de uma norma completa. Seja E^* o seu dual, isto é, o espaço dos funcionais lineares contínuos. O dual é também um espaço de Banach, com a norma

$$\|g\| = \sup \left\{ \frac{|g(v)|}{\|v\|} : v \in E \setminus \{0\} \right\}. \quad (2.3.1)$$

A *topologia fraca* no espaço E é a topologia definida pela base de vizinhanças

$$V(v, \{g_1, \dots, g_N\}, \varepsilon) = \{w \in E : |g_i(v) - g_i(w)| < \varepsilon \text{ para todo } i\}, \quad (2.3.2)$$

onde $g_1, \dots, g_N \in E^*$. Em termos de sequências, ela satisfaz

$$(v_n)_n \rightarrow v \quad \Rightarrow \quad (g(v_n))_n \rightarrow g(v) \quad \text{para todo } g \in E^*.$$

A *topologia fraca** no dual E^* é a topologia definida pela base de vizinhanças

$$V^*(g, \{v_1, \dots, v_N\}, \varepsilon) = \{h \in E^* : |g(v_i) - h(v_i)| < \varepsilon \text{ para todo } i\}, \quad (2.3.3)$$

onde $v_1, \dots, v_N \in E$. Ela satisfaz

$$(g_n)_n \rightarrow g \quad \Rightarrow \quad (g_n(v))_n \rightarrow g(v) \quad \text{para todo } v \in E.$$

Esta última topologia é notável devido ao seguinte fato:

Teorema 2.3.1 (Banach-Alaoglu). *A bola unitária fechada de E^* é compacta para a topologia fraca*.*

A construção que realizamos anteriormente corresponde ao caso em que E é o espaço $C^0(M)$ das funções contínuas e E^* é o espaço $\mathcal{M}(M)$ das medidas complexas num espaço métrico compacto M : de acordo com o teorema de Riesz-Markov (Teorema A.3.12) $\mathcal{M}(M)$ corresponde ao dual de $C^0(M)$ quando identificamos cada medida $\mu \in \mathcal{M}(M)$ com o funcional linear $I_\mu(\phi) = \int \phi d\mu$. Note que a definição da norma (2.3.1) dá

$$\|\mu\| = \sup \left\{ \frac{|\int \phi d\mu|}{\sup |\phi|} : \phi \in C^0(M) \setminus \{0\} \right\}.$$

Em particular, o conjunto $\mathcal{M}_1(M)$ das medidas de probabilidade está contido na bola unitária de $\mathcal{M}(M)$. Como este subconjunto é fechado na topologia fraca*, concluímos que o Teorema 2.1.5 também segue diretamente do teorema de Banach-Alaoglu.

Agora consideremos uma transformação contínua $f : M \rightarrow M$ qualquer em M e consideremos a sua ação $f_* : \mathcal{M}(M) \rightarrow \mathcal{M}(M)$, $\mu \mapsto f_*\mu$ no espaço das medidas complexas. Então f_* é um operador linear em $\mathcal{M}(M)$ e é contínuo relativamente à topologia fraca*. Existe outro operador linear contínuo naturalmente associado a f , a saber $U_f : C^0(M) \rightarrow C^0(M)$, $\phi \mapsto \phi \circ f$. Agora observe que estes operadores são duais (lembre do Lema 2.2.1):

$$\int U_f(\phi) d\mu = \int (\phi \circ f) d\mu = \int \phi d(f_*\mu). \quad (2.3.4)$$

Estas considerações motivam a noção que apresentaremos na próxima seção.

2.3.2 Operador de Koopman

Sejam (M, \mathcal{B}) um espaço mensurável, $f : M \rightarrow M$ uma transformação mensurável e μ uma medida invariante por f . O *operador de Koopman* é o operador linear

$$U_f : L^1(\mu) \rightarrow L^1(\mu), \quad U_f(\phi) = \phi \circ f.$$

Note que U_f está bem definido e é uma *isometria*, isto é, ele preserva a norma do espaço $L^1(\mu)$:

$$\|U_f(\phi)\|_1 = \int |U_f(\phi)| d\mu = \int |\phi| \circ f d\mu = \int |\phi| d\mu = \|\phi\|_1 \quad (2.3.5)$$

uma vez que μ é invariante. Além disso, U_f é um operador linear *positivo*: $U_f(\phi) \geq 0$ em μ -quase todo ponto sempre que $\phi \geq 0$ em μ -quase todo ponto. Resumimos estes fatos na seguinte proposição:

Proposição 2.3.2. *O operador $U_f : L^1(M) \rightarrow L^1(M)$ induzido por f é linear, positivo e uma isometria.*

A propriedade (2.3.5) implica que o operador U_f é injetivo. Em geral U_f não é sobrejetivo (veja o Exercício 2.3.5). Claro que se f é invertível então U_f é um isomorfismo: o inverso é o operador de Koopman $U_{f^{-1}}$ da inversa de f .

Podemos também considerar versões do operador de Koopman definidas nos espaços $L^p(\mu)$,

$$U_f : L^p(\mu) \rightarrow L^p(\mu), \quad U_f(\phi) = \phi \circ f$$

para qualquer $p \in [1, \infty]$. A Proposição 2.3.2 permanece válida em todos estes casos: todos estes operadores são isometrias lineares positivas.

Quando M é um espaço métrico compacto e f é contínua, é particularmente interessante observar a ação de U_f restrita ao espaço $C^0(M)$ das funções contínuas:

$$U_f : C^0(M) \rightarrow C^0(M).$$

É claro que este operador é contínuo relativamente à norma da convergência uniforme. Conforme vimos anteriormente, o dual de $C^0(M)$ está identificado de maneira natural com o espaço $\mathcal{M}(M)$ das medidas complexas em M . Além disso, a relação (2.3.4) mostra que o operador dual

$$U_f^* : C^0(M)^* \rightarrow C^0(M)^*$$

corresponde precisamente à ação $f_* : \mathcal{M}(M) \rightarrow \mathcal{M}(M)$ da transformação f quando fazemos essa identificação. Esse fato vai nos permitir usar resultados de Teoria Espectral para dar outra demonstração do Teorema 2.1.

Para isso precisamos lembrar algumas noções da teoria de operadores positivos. O leitor interessado poderá encontrar mais detalhes e demonstrações no livro de Deimling [Dei85].

Seja E um espaço de Banach. Um subconjunto fechado e convexo C é chamado de *cone* de E , se ele satisfaz:

$$\lambda C \subset C \text{ para todo } \lambda \geq 0 \text{ e } C \cap (-C) = \{0\}. \quad (2.3.6)$$

Dizemos que o cone C é *normal* quando

$$\inf\{\|x + y\| : x, y \in C \text{ tais que } \|x\| = \|y\| = 1\} > 0.$$

Fixemos um cone C de E . Dado um operador linear contínuo $T : E \rightarrow E$, diremos que T é um *operador positivo sobre C* se $T(C) \subset C$. Dado um funcional linear contínuo $\phi : E \rightarrow \mathbb{R}$, diremos que ϕ é um *funcional positivo sobre C* se $\phi(v) \geq 0$ para todo $v \in C$. Por definição, o *cone dual* C^* é o cone em E^* formado por todos os funcionais positivos sobre C .

Exemplo 2.3.3. $C_+^0(M) = \{\varphi \in C^0(M) : \varphi \geq 0\}$ é um cone normal de $C^0(M)$ (Exercício 2.3.3). Pelo teorema de Riesz-Markov (Teorema A.3.11), o cone dual se identifica naturalmente com o espaço das medidas (positivas) finitas em M .

Denotaremos por $r(T)$ o *raio espectral* do operador linear contínuo T :

$$r(T) = \lim_n \sqrt[n]{\|T^n\|}.$$

Representando por $T^* : E^* \rightarrow E^*$ o operador dual de T , vale que $r(T) = r(T^*)$. O resultado a seguir é uma consequência do teorema de Banach-Mazur; veja a Proposição 7.2 no livro de Deimling [Dei85]:

Teorema 2.3.4. *Seja C um cone normal num espaço de Banach E e seja $T : E \rightarrow E$ um operador linear positivo sobre C . Então, $r(T^*)$ é autovalor do operador dual $T^* : E^* \rightarrow E^*$ e admite algum autovetor $v^* \in C^*$.*

Como aplicação deste resultado, vamos dar uma prova alternativa da existência de probabilidades invariantes. Considere o cone $C = C_+^0(M)$ em $E = C^0(M)$. Conforme já observamos, o cone dual C^* é o espaço das medidas positivas finitas em M . É claro da definição que o operador $T = U_f$ é positivo sobre C . Além disso, o seu raio espectral é igual a 1, uma vez $\sup |T(\varphi)| \leq \sup |\varphi|$ para todo $\varphi \in C^0(M)$ e $T(1) = 1$. Logo, pelo Teorema 2.3.4, existe alguma medida finita μ em M que é autovetor do operador dual $T^* = f_*$ associado ao autovalor 1. Em outras palavras, a medida μ é invariante. Multiplicando por uma constante adequada, podemos supor que μ é uma probabilidade.

2.3.3 Exercícios

2.3.1. Seja ℓ^1 o espaço das seqüências somáveis de números complexos, munido da norma $\|(a_n)_n\|_1 = \sum_{n=0}^{\infty} |a_n|$. Seja ℓ^∞ o espaço das seqüências limitadas e seja c_0 o espaço das seqüências convergentes para zero, ambos munidos da norma $\|(a_n)_n\|_\infty = \sup_{n \geq 0} |a_n|$.

(a) Verifique que ℓ^∞ , ℓ^1 e c_0 são espaços de Banach.

(b) Mostre que a aplicação $(a_n)_n \mapsto [(b_n)_n \mapsto \sum_n a_n b_n]$ define isomorfismos isométricos entre ℓ^∞ e o dual $(\ell^1)^*$ e entre ℓ^1 e o dual $(c_0)^*$.

2.3.2. Mostre que uma seqüência $(x^k)_k$ em ℓ^1 (escreva $x^k = (x_n^k)_n$ para cada k) é convergente na topologia da norma se, e somente se, ela é convergente na topologia fraca, ou seja, se $(\sum_n a_n x_n^k)_k$ converge qualquer que seja $(a_n)_n \in \ell^\infty$. [Observação: Isso não implica que as duas topologias sejam iguais. Porquê?] Mostre que a afirmação torna-se falsa se substituirmos a topologia fraca pela topologia fraca*.

2.3.3. Verifique que $C_+^0(M)$ é um cone normal.

2.3.4. Seja $R_\theta : S^1 \rightarrow S^1$ uma rotação irracional e seja m a medida de Lebesgue no círculo. Calcule os autovalores e os autovetores do operador de Koopman $U_\theta : L^2(m) \rightarrow L^2(m)$. Mostre que o espectro de U_θ coincide com o círculo unitário $\{z \in \mathbb{C} : |z| = 1\}$.

2.3.5. Mostre, por meio de exemplos, que o operador de Koopman U_f pode não ser sobrejetivo.

2.3.6. Seja $U : H \rightarrow H$ uma isometria de um espaço de Hilbert. Pelo Exercício A.6.8, a imagem de U é um subespaço fechado de H . Conclua que existem

subespaços fechados V e W tais que $U(V) = V$, os iterados de W são ortogonais a V e entre si, e

$$H = V \oplus \bigoplus_{n=0}^{\infty} U^n(W).$$

Além disso, U é um isomorfismo se, e somente se, $W = \{0\}$.

2.3.7. Seja $\phi : E \rightarrow \mathbb{R}$ um funcional convexo contínuo em um espaço de Banach separável E , derivável em todas as direções num ponto $u \in E$. Prove que existe, no máximo, um funcional linear limitado $T : E \rightarrow \mathbb{R}$ tangente a ϕ em u , ou seja, tal que $T(v) \leq \phi(u+v) - \phi(u)$ para todo $v \in E$. Se ϕ é derivável em u , a derivada $D\phi(u)$ é um funcional linear tangente a ϕ em u . [Observação: O teorema de suavidade de Mazur (Teorema 1.20 no livro de Phelps [Phe93]) afirma que o conjunto dos pontos onde ϕ é derivável e, conseqüentemente, existe um único funcional tangente a ϕ é residual em E .]

2.4 Produtos semi-diretos e extensões naturais

Nesta seção vamos descrever duas construções úteis em Teoria Ergódica. A primeira construção modela a situação em que temos dois sistemas dinâmicos acoplados da seguinte forma: um sistema é autônomo mas o outro não, porque a evolução do segundo depende da evolução do primeiro. A segunda construção associa um sistema invertível a qualquer sistema dinâmico dado; além disso, as medidas invariantes dos dois sistemas estão em correspondência biunívoca. Isto permite reduzir ao caso invertível muitos enunciados sobre sistemas gerais, não necessariamente invertíveis.

2.4.1 Medidas em produtos semi-diretos

Sejam (X, \mathcal{A}) e (Y, \mathcal{B}) espaços mensuráveis. Chamamos *produto semi-direto* a qualquer transformação mensurável $F : X \times Y \rightarrow X \times Y$ da forma $F(x, y) = (f(x), g(x, y))$. Representamos por $\pi : X \times Y \rightarrow X$ a projeção canônica. Por definição,

$$\pi \circ F = f \circ \pi. \quad (2.4.1)$$

Seja m uma probabilidade em $X \times Y$ invariante por F e seja $\mu = \pi_* m$ a sua projeção para X . Então, usando a invariância de m ,

$$f_* \mu = f_* \pi_* m = \pi_* F_* m = \pi_* m = \mu,$$

ou seja, μ é invariante por f . A próxima proposição dá uma recíproca parcial para esta conclusão: sob hipóteses apropriadas, toda a medida invariante por f é projeção de alguma medida invariante por F .

Proposição 2.4.1. *Suponha que X é um espaço métrico completo separável, Y é um espaço métrico compacto e F é contínua. Então, para toda medida de probabilidade μ em X invariante por f existe alguma medida m em $X \times Y$ invariante por F , tal que $\pi_* m = \mu$.*

Demonstração. Dada qualquer medida de probabilidade invariante μ em X , seja $\mathcal{K} \subset \mathcal{M}_1(X \times Y)$ o conjunto das medidas η em $X \times Y$ tais que $\pi_*\eta = \mu$. Considere qualquer $\eta \in \mathcal{K}$. Então, $\pi_*F_*\eta = f_*\pi_*\eta = f_*\mu = \mu$. Isto mostra que \mathcal{K} é invariante por F_* . Em seguida, note que a projeção $\pi : X \times Y \rightarrow X$ é contínua e, consequentemente, o operador π_* é contínuo relativamente à topologia fraca*. Portanto, \mathcal{K} é fechado em $\mathcal{M}_1(X \times Y)$. Pela Proposição A.3.7, dado qualquer $\varepsilon > 0$ existe um compacto $K \subset X$ tal que $\mu(K^c) < \varepsilon$. Então $K \times Y$ é compacto e $\eta((K \times Y)^c) = \mu(K^c) < \varepsilon$ para todo $\eta \in \mathcal{K}$. Isto prova que o conjunto \mathcal{K} é justo. Considere qualquer $\eta \in \mathcal{K}$. Pelo teorema de Prohorov (Teorema 2.1.8), a sequência

$$\frac{1}{n} \sum_{j=0}^{n-1} F_*^j \eta$$

admite algum ponto de acumulação $m \in \mathcal{K}$. Argumentando como na prova do Lema 2.2.4 concluímos que m é invariante por F . \square

2.4.2 Extensões naturais

Dada uma transformação sobrejetiva $f : M \rightarrow M$ é sempre possível encontrar uma extensão $\hat{f} : \hat{M} \rightarrow \hat{M}$ que é invertível. Por *extensão* queremos dizer que existe uma aplicação sobrejetiva $\pi : \hat{M} \rightarrow M$ tal que $\pi \circ \hat{f} = f \circ \pi$. Este fato é muito útil porque permite reduzir a prova de muitos enunciados do caso geral ao caso de transformações invertíveis. Conforme comentaremos no Exemplo 2.4.2, a hipótese de que f é sobrejetiva pode ser removida em muitos casos interessantes.

Para começar, tomamos para \hat{M} o conjunto de todas as *pré-órbitas* de f , ou seja, o conjunto de todas as sequências $(x_n)_{n \leq 0}$ indexadas pelos números inteiros não-positivos e satisfazendo $f(x_n) = x_{n+1}$ para todo $n < 0$. Consideramos a aplicação $\pi : \hat{M} \rightarrow M$ que associa a cada sequência $(x_n)_{n \leq 0}$ o seu termo x_0 de ordem zero. Observe que $\pi(\hat{M}) = M$. Finalmente, definimos $\hat{f} : \hat{M} \rightarrow \hat{M}$ como sendo o deslocamento à esquerda:

$$\hat{f}(\dots, x_n, \dots, x_0) = (\dots, x_n, \dots, x_0, f(x_0)). \quad (2.4.2)$$

É claro que \hat{f} está bem definida e satisfaz $\pi \circ \hat{f} = f \circ \pi$. Além disso, ela é invertível: a inversa é o deslocamento à direita

$$(\dots, y_n, \dots, y_{-1}, y_0) \mapsto (\dots, y_n, \dots, y_{-2}, y_{-1}).$$

Se M é um espaço mensurável podemos tornar \hat{M} um espaço mensurável, munindo-o da σ -álgebra gerada pelos *cilindros mensuráveis*

$$[A_k, \dots, A_0] = \{(x_n)_{n \leq 0} \in \hat{M} : x_i \in A_i \text{ para } i = k, \dots, 0\}, \quad (2.4.3)$$

onde $k \leq 0$ e A_k, \dots, A_0 são subconjuntos mensuráveis de M . Então π é aplicação mensurável, uma vez que

$$\pi^{-1}(A) = [A]. \quad (2.4.4)$$

Além disso, \hat{f} é mensurável se f for:

$$\hat{f}^{-1}([A_k, \dots, A_0]) = [A_k, \dots, A_{-2}, A_{-1} \cap f^{-1}(A_0)]. \quad (2.4.5)$$

A inversa de \hat{f} também é mensurável, já que

$$\hat{f}([A_k, \dots, A_0]) = [A_k, \dots, A_0, M]. \quad (2.4.6)$$

Analogamente, se M é um espaço topológico podemos tornar \hat{M} um espaço topológico, munindo-o da topologia gerada pelos *cilindros abertos* $[A_k, \dots, A_0]$, onde $k \leq 0$ e A_k, \dots, A_0 são subconjuntos abertos de M . As relações (2.4.4) e (2.4.6) mostram que π e \hat{f}^{-1} são contínuas, enquanto que (2.4.5) mostra que \hat{f} é contínua se f for. Observe que se M admite alguma base enumerável de abertos \mathcal{U} , então os cilindros $[A_k, \dots, A_0]$ com $k \geq 0$ e $A_0, \dots, A_k \in \mathcal{U}$ constituem uma base enumerável de abertos para \hat{M} .

Se M é um espaço métrico, com distância d , podemos tornar \hat{M} um espaço métrico munindo-o da distância

$$\hat{d}(\hat{x}, \hat{y}) = \sum_{n=-\infty}^0 2^n \min\{d(x_n, y_n), 1\} \quad (2.4.7)$$

onde $\hat{x} = (x_n)_{n \leq 0}$ e $\hat{y} = (y_n)_{n \leq 0}$. A seguinte observação é uma consequência imediata da definição: se \hat{x} e \hat{y} estão numa mesma pré-imagem $\pi^{-1}(x)$ então

$$\hat{d}(\hat{f}^j(\hat{x}), \hat{f}^j(\hat{y})) \leq 2^{-j} \hat{d}(\hat{x}, \hat{y}) \quad \text{para todo } j \geq 0.$$

Portanto, toda pré-imagem $\pi^{-1}(x)$ é um conjunto estável, ou seja, um conjunto onde a transformação \hat{f} é uniformemente contrativa.

Exemplo 2.4.2. Dada uma transformação $g : M \rightarrow M$ qualquer, considere $M_g = \bigcap_{n=1}^{\infty} g^n(M)$. É claro $g(M_g) \subset M_g$. Suponha que

- (a) M é compacto e g é contínua ou (b) $\#g^{-1}(y) < \infty$ para todo y .

Então (Exercício 2.4.3), a restrição $f = (g | M_g) : M_g \rightarrow M_g$ é sobrejetiva. Esta restrição contém a dinâmica interessante de g . Por exemplo, supondo que o conjunto $f^n(M)$ é mensurável para todo n , toda probabilidade invariante por g também é probabilidade invariante por f . De modo semelhante, todo ponto recorrente por g também é ponto recorrente por f , pelo menos no caso (a). Assim, também nos referiremos à extensão natural de $f = (g | M_g)$ como extensão natural de g .

Um conjunto $\Lambda \subset M$ tal que $f^{-1}(\Lambda) = \Lambda$ é chamado *conjunto invariante* de f . Vale uma noção análoga para a transformação \hat{f} . A próxima proposição mostra que todo conjunto invariante fechado de f admite um único levantamento a um conjunto invariante fechado de \hat{f} :

Proposição 2.4.3. *Suponha que M é um espaço topológico. Se $\Lambda \subset M$ é um conjunto fechado invariante de f então $\hat{\Lambda} = \pi^{-1}(\Lambda)$ é o único conjunto fechado invariante de \hat{f} satisfazendo $\pi(\hat{\Lambda}) = \Lambda$.*

Demonstração. Como π é contínua, se Λ é fechado então $\hat{\Lambda} = \pi^{-1}(\Lambda)$ também é fechado. Além disso, se Λ é invariante por f então $\hat{\Lambda}$ é invariante por \hat{f} :

$$\hat{f}^{-1}(\hat{\Lambda}) = (\pi \circ \hat{f})^{-1}(\Lambda) = (f \circ \pi)^{-1}(\Lambda) = \pi^{-1}(\Lambda) = \hat{\Lambda},$$

Na direção recíproca, seja $\hat{\Lambda}$ um conjunto invariante fechado de \hat{f} tal que $\pi(\hat{\Lambda}) = \Lambda$. É claro que $\hat{\Lambda} \subset \pi^{-1}(\Lambda)$. Para provar a outra inclusão, devemos mostrar que, dado qualquer $x_0 \in \Lambda$, se $\hat{x} \in \pi^{-1}(x_0)$ então $\hat{x} \in \hat{\Lambda}$. Escrevamos $\hat{x} = (x_n)_{n \leq 0}$. Considere $n \leq 0$ e qualquer vizinhança de \hat{x} da forma

$$V = [A_n, \dots, A_0], \quad A_n, \dots, A_0 \text{ abertos de } M.$$

Pela definição da extensão natural, $x_0 = f^{-n}(x_n)$ e, portanto, $x_n \in f^n(\Lambda) = \Lambda$. Então, a hipótese de que $\pi(\hat{\Lambda}) = \Lambda$ implica que existe algum ponto $\hat{y}_n \in \hat{\Lambda}$ tal que $\pi(\hat{y}_n) = x_n$. Como $\hat{\Lambda}$ é invariante por \hat{f} , temos que $\hat{f}^{-n}(\hat{y}_n) \in \hat{\Lambda}$. Além disso, a propriedade $\pi(\hat{y}_n) = x_n$ implica que

$$f^{-n}(\hat{y}_n) = (\dots, y_{n,k}, \dots, y_{n,-1}, y_{n,0} = x_n, x_{n-1}, \dots, x_{-1}, x_0).$$

Portanto $f^{-n}(\hat{y}_n) \in V$, já que V contém \hat{x} e a sua definição só envolve as coordenadas com índices $j \in \{n, \dots, 0\}$. Isto mostra que \hat{x} é acumulado por elementos de $\hat{\Lambda}$. Como $\hat{\Lambda}$ é fechado, segue que $\hat{x} \in \hat{\Lambda}$. \square

Agora seja $\hat{\mu}$ uma medida invariante por \hat{f} e seja $\mu = \pi_*\hat{\mu}$. A propriedade $\pi \circ \hat{f} = f \circ \pi$ implica que μ é invariante por f :

$$f_*\mu = f_*\pi_*\hat{\mu} = \pi_*\hat{f}_*\hat{\mu} = \pi_*\hat{\mu} = \mu.$$

Dizemos que $\hat{\mu}$ é um *levantamento* de μ . O próximo resultado, uma espécie de versão da Proposição 2.4.3 para medidas, é devido a Rokhlin [Rok61]:

Proposição 2.4.4. *Suponha que M é um espaço métrico completo separável e que $f : M \rightarrow M$ é contínua. Então, toda probabilidade μ invariante por f admite um único levantamento, ou seja, única medida $\hat{\mu}$ em \hat{M} invariante por \hat{f} e tal que $\pi_*\hat{\mu} = \mu$.*

A unicidade é imediata e não depende das hipóteses sobre o espaço M e a transformação f . De fato, se $\hat{\mu}$ é um levantamento de μ então (2.4.4) e (2.4.5) implicam que a medida de cada cilindro está unicamente determinada:

$$\hat{\mu}([A_k, \dots, A_0]) = \hat{\mu}([A_k \cap \dots \cap f^{-k}(A_0)]) = \mu(A_k \cap \dots \cap f^{-k}(A_0)). \quad (2.4.8)$$

A prova da existência usa ideias que serão desenvolvidas no Capítulo 5 e será proposta ao leitor no Exercício 5.2.4. Esses argumentos se estendem para sistemas em espaços de Lebesgue, conforme observaremos no Exercício 8.5.7. Mas o enunciado de existência não é válido para espaços de probabilidade arbitrários (veja o exemplo no Exercício 1.15 no livro de Przytycki, Urbański [PU10]).

2.4.3 Exercícios

2.4.1. Seja M um espaço métrico compacto e seja X um conjunto de aplicações contínuas $f : M \rightarrow M$, munido de uma probabilidade ν . Considere o produto semidireto $F : X^{\mathbb{N}} \times M \rightarrow X^{\mathbb{N}} \times M$ definido por $F((f_n)_n, x) = ((f_{n+1})_n, f_0(x))$. Mostre que F admite alguma probabilidade invariante m da forma $m = \nu^{\mathbb{N}} \times \mu$. Além disso, uma medida m dessa forma é invariante por F se, e somente se, a medida μ é *estacionária* para ν , isto é, $\mu(E) = \int f_*\mu(E) d\nu(f)$ para todo conjunto mensurável $E \subset M$.

2.4.2. Sejam $f : M \rightarrow M$ uma transformação sobrejetiva, $\hat{f} : \hat{M} \rightarrow \hat{M}$ a sua extensão natural e $\pi : \hat{M} \rightarrow M$ a projeção canônica. Mostre que se $g : N \rightarrow N$ é uma transformação invertível tal que $f \circ p = p \circ g$ para alguma aplicação $p : N \rightarrow M$ então existe uma única transformação $\hat{p} : N \rightarrow \hat{M}$ tal que $\pi \circ \hat{p} = p$ e $\hat{p} \circ g = \hat{f} \circ \hat{p}$. Suponha que M e N são espaços topológicos compactos e que as aplicações p e g são contínuas. Mostre que se p é sobrejetiva então \hat{p} é sobrejetiva (portanto, $g : N \rightarrow N$ é uma extensão de $\hat{f} : \hat{M} \rightarrow \hat{M}$).

2.4.3. Verifique as afirmações no Exemplo 2.4.2.

2.4.4. Mostre que se (M, d) é um espaço métrico separável completo então o mesmo vale para o espaço (\hat{M}, \hat{d}) das pré-órbitas de qualquer transformação contínua sobrejetiva $f : M \rightarrow M$.

2.4.5. O objetivo deste exercício e do seguinte é generalizar a noção de extensão natural para transformações que comutam entre si. Seja M um espaço compacto e sejam $f_1, \dots, f_q : M \rightarrow M$ transformações contínuas sobrejetivas que comutam entre si. Seja \hat{M} o conjunto das sequências $(x_{n_1, \dots, n_q})_{n_1, \dots, n_q \leq 0}$, indexadas pelas q -uplas de inteiros não positivos, tais que

$$f_i(x_{n_1, \dots, n_i, \dots, n_q}) = x_{n_1, \dots, n_i+1, \dots, n_q} \quad \text{para todo } i \text{ e todo } (n_1, \dots, n_q).$$

Seja $\pi : \hat{M} \rightarrow M$ a aplicação que envia $(x_{n_1, \dots, n_q})_{n_1, \dots, n_q \leq 0}$ no ponto $x_{0, \dots, 0}$. Para cada i , seja $\hat{f}_i : \hat{M} \rightarrow \hat{M}$ a transformação que envia $(x_{n_1, \dots, n_i, \dots, n_q})_{n_1, \dots, n_q \leq 0}$ em $(x_{n_1, \dots, n_i+1, \dots, n_q})_{n_1, \dots, n_q \leq 0}$.

- Mostre que \hat{M} é um espaço compacto. Além disso, \hat{M} é metrizable se M é metrizable.
- Mostre que cada $\hat{f}_i : \hat{M} \rightarrow \hat{M}$ é um homeomorfismo com $\pi \circ \hat{f}_i = f_i \circ \pi$. Além disso, estes homeomorfismos comutam entre si.
- Mostre que π é contínua e sobrejetiva. Em particular, \hat{M} é não vazio.

2.4.6. Seja M um espaço compacto e sejam $g_1, \dots, g_q : M \rightarrow M$ transformações contínuas que comutam entre si. Defina $M_g = \bigcap_{n=1}^{\infty} g_1^n \cdots g_q^n(M)$.

- Verifique que $M_g = \bigcap_{n_1, \dots, n_q} g_1^{n_1} \cdots g_q^{n_q}(M)$, onde a interseção é sobre todas as q -uplas (n_1, \dots, n_q) com $n_i \geq 1$ para todo i .

- (b) Mostre que $g_i(M_g) \subset M_g$ e que a restrição $f_i = g_i | M_g$ é sobrejetiva, para todo i .

[Observação: É claro que estas restrições f_i comutam entre si.]

2.4.7. Use a construção nos Exercícios 2.4.5 e 2.4.6 para estender a demonstração do Teorema 1.5.1 para o caso em que as transformações f_i não são necessariamente invertíveis.

2.5 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 1.5.1 e Teorema 1.5.2) que foram apresentados na Seção 1.5.

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

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

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 2.5.1 (van der Waerden). *Dada qualquer partição finita $\{S_1, \dots, S_l\}$ de \mathbb{Z} existe algum $j \in \{1, \dots, l\}$ 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 Erdős e Pál Turan [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 sequências 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

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

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

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

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 2.5.1).

Lema 2.5.2. *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, \dots, S_l é uma partição de \mathbb{Z} então*

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

Exemplo 2.5.3. 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 2.5.4. Seja S o seguinte subconjunto de \mathbb{Z} :

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

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 2.5.3 o conjunto S até contém progressões aritméticas de comprimento infinito. Isso não é verdade no Exemplo 2.5.4, uma vez que nesse caso o complementar de S contém intervalos arbitrariamente longos.

Teorema 2.5.5 (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 2.5.2 que se S_1, \dots, S_l é uma partição de \mathbb{Z} então existe j tal que $D_s(S_j) > 0$. Pelo Teorema 2.5.5, 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 2.5.1 como obter o teorema de van der Waerden a partir do teorema de recorrência múltipla de

Birkhoff (Teorema 1.5.1); argumentos análogos dão o teorema de Szemerédi a partir do teorema de recorrência múltipla de Poincaré (Teorema 1.5.2), como veremos na Seção 2.5.2.

A teoria de Szemerédi continua sendo uma área de pesquisa muito ativa. Em particular, outras demonstrações do Teorema 2.5.5 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 Erdős: 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.

2.5.1 Teorema de van der Waerden

Nesta seção vamos demonstrar o Teorema 2.5.1. A ideia é reduzir a conclusão a uma afirmação sobre o deslocamento à esquerda

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

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

Observe que toda partição $\{S_1, \dots, S_l\}$ 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

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

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

$$\alpha_{m+n} = \dots = \alpha_{m+qn}. \quad (2.5.1)$$

Em vista do que acabamos de observar, isto significa que para toda partição $\{S_1, \dots, S_l\}$ e todo $q \geq 1$ existe $i \in \{1, \dots, l\}$ 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 (2.5.1).

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

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

Note que

$$d(\underline{\beta}, \underline{\gamma}) < 1 \quad \text{se, e somente se, } \beta_0 = \gamma_0. \quad (2.5.2)$$

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 $f_1 = \sigma$, $f_2 = \sigma^2$, \dots , $f_q = \sigma^q$ definidas de Z em Z . É claro que as f_i comutam entre si. Portanto, podemos aplicar o Teorema 1.5.1 para concluir que existe $\underline{\theta} \in Z$ e uma seqüência $(n_k)_k \rightarrow \infty$ tal que

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

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

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

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

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

Levando em conta a observação (2.5.2) e a definição do deslocamento σ , isto quer dizer que $\alpha_{m+n} = \dots = \alpha_{m+qn}$, como pretendíamos provar. Isto completa a demonstração do teorema de van der Waerden.

2.5.2 Teorema de Szemerédi

Agora vamos demonstrar o Teorema 2.5.5. Para isso, usaremos o mesmo tipo de dicionário entre partições de \mathbb{Z} e seqüências 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 seqüência $\underline{\alpha} = (\alpha_j)_{j \in \mathbb{Z}} \in \Sigma = \{0, 1\}^{\mathbb{Z}}$ definida por:

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

Considere o deslocamento $\sigma : \Sigma \rightarrow \Sigma$ 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

$$\sigma^{m+n}(\underline{\alpha}), \sigma^{m+2n}(\underline{\alpha}), \dots, \sigma^{m+kn}(\underline{\alpha}) \in A. \quad (2.5.3)$$

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

$$\mu_j = \frac{1}{\#I_j} \sum_{i \in I_j} \delta_{\sigma^i(\underline{\alpha})} \quad (2.5.4)$$

Como o conjunto $\mathcal{M}_1(\Sigma)$ das probabilidades em Σ é compacto (Teorema 2.1.5), 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 $\varphi : \Sigma \rightarrow \mathbb{R}$, vale

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

e, passando ao limite quando $j \rightarrow \infty$, 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, o Teorema 2.1.2 garante que

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

Dado qualquer $k \geq 1$, considere as transformações $f_i = \sigma^i$ para $i = 1, \dots, k$. É claro que estas transformações comutam entre si. Então, estamos em condições de aplicar o Teorema 1.5.2 e concluir que existe algum $n \geq 1$ tal que

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

Como A é aberto, isto implica (Teorema 2.1.2) que

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

para qualquer l suficientemente grande. Pela definição de μ_l em (2.5.4), isto quer dizer que existe algum $m \in I_l$ tal que

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

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

2.5.3 Exercícios

2.5.1. Demonstre o Lema 2.5.2.

2.5.2. Mostre que a conclusão do Teorema 2.5.1 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 $\{1, 2, \dots, N\}$ em l subconjuntos, algum desses subconjuntos contém progressões aritméticas com comprimento q .

2.5.3. 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 $\bigcap_{j=0}^k f^{-jn}(U) \neq \emptyset$. Mostre que o Teorema de van der Warden é equivalente ao seguinte enunciado: toda aplicação invertível num espaço métrico compacto tem algum ponto super não errante.

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

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

Capítulo 3

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

$$\tau(E, x) = \lim_{n \rightarrow \infty} \frac{1}{n} \#\{0 \leq j < n : f^j(x) \in E\}. \quad (3.0.1)$$

No caso de fluxos temos uma noção análoga, definida por

$$\tau(E, x) = \lim_{T \rightarrow \infty} \frac{1}{T} m(\{0 \leq t \leq T : f^t(x) \in E\}) \quad (3.0.2)$$

(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 (3.0.1)-(3.0.2) 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 4.

Representando por φ a função característica do conjunto E , podemos reescrever a expressão no lado direito de (3.0.1) como:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)). \quad (3.0.3)$$

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 3.1.5) afirma que, de fato, o limite em (3.0.3) existe no espaço $L^2(\mu)$, para toda função $\varphi \in L^2(\mu)$. O teorema ergódico de Birkhoff (Teorema 3.2.3) 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 (3.0.1) está bem definido para μ -quase todo x (Teorema 3.2.1).

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 3.3.3). Este teorema afirma que ψ_n/n converge em quase todo ponto, para qualquer sequência de funções ψ_n tal que $\psi_{m+n} \leq \psi_m + \psi_n \circ f^m$.

Todos estes resultados permanecem válidos para fluxos, conforme comentaremos na Seção 3.4.

3.1 Teorema ergódico de Von Neumann

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

3.1.1 Isometrias em espaços de Hilbert

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

$$H = F \oplus F^\perp, \quad (3.1.1)$$

onde $F^\perp = \{w \in H : v \cdot w = 0 \text{ para todo } v \in F\}$ é o complementar ortogonal de F . A projeção $P_F : H \rightarrow F$ associada à decomposição (3.1.1) 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 3.1.1. Considere o espaço de Hilbert $L^2(\mu)$, com o produto interno

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

Se φ_0 é a função constante igual a 1 e F é o subespaço gerado por φ_0 em $L^2(\mu)$, ou seja, o espaço das funções constantes, então a projeção ortogonal $P_F(\varphi)$ é dada por

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

De fato, como F é gerado por φ_0 , temos que $P_F(\varphi) = c\varphi_0$ para algum $c \in \mathbb{R}$. Para calcular a constante c , note que

$$(P_F(\varphi) - \varphi) \cdot \varphi_0 = 0 \quad \Leftrightarrow \quad c = \frac{\varphi \cdot \varphi_0}{\varphi_0 \cdot \varphi_0} = \int \varphi d\mu.$$

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

$$U^*u \cdot v = u \cdot Uv \quad \text{para todo } u, v \in H. \quad (3.1.2)$$

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

$$Uu \cdot Uv = u \cdot v \quad \text{para todo } u, v \in H. \quad (3.1.3)$$

Isso é equivalente a dizer que U preserva a norma de H (veja o Exercício A.6.9). Outra condição equivalente é $U^*U = \text{id}$. 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 $U^*U = \text{id}$ implica que U é injetivo; em geral, uma isometria pode não ser sobrejetiva. Veja os Exercícios 2.3.5 e 2.3.6.

Exemplo 3.1.2. Se $f : M \rightarrow M$ preserva uma medida μ então, como vimos na Seção 2.3.2, o seu operador de Koopman $U_f : L^2(\mu) \rightarrow L^2(\mu)$ é uma isometria.

Chamamos conjunto dos *vetores invariantes* de um operador linear contínuo $U : H \rightarrow H$ 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 3.1.3. *Se $U : H \rightarrow H$ é uma isometria então $Uv = v$ se, e somente se, $U^*v = v$.*

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

$$\|Uv - v\|^2 = \langle Uv - v, Uv - v \rangle = \|Uv\|^2 - 2\langle Uv, v \rangle + \|v\|^2 = 0.$$

Isto significa que $Uv = v$. □

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:

Teorema 3.1.4. *Seja $U_t : H \rightarrow H$, $t \in \mathbb{R}$ um grupo a 1-parâmetro de operadores unitários num espaço de Hilbert. Suponha que o grupo é fortemente contínuo, isto é,*

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

Então existe um único operador auto-adjunto $A : H \rightarrow H$ tal que $U_t = e^{2\pi i t A}$ para todo $t \in \mathbb{R}$.

O leitor pode encontrar a demonstração no livro de Yosida [Yos68] e uma aplicação simples está dada no Exercício 3.1.5.

3.1.2 Enunciado e prova do teorema

Teorema 3.1.5 (von Neumann). *Seja $U : H \rightarrow H$ uma isometria num espaço de Hilbert H , e seja P a projeção ortogonal sobre o subespaço $I(U)$ dos vetores invariantes por U . Então,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} U^j v = Pv \quad \text{para todo } v \in H. \quad (3.1.4)$$

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

$$I(U) = \bar{L}(U)^\perp. \quad (3.1.5)$$

Isto pode ser verificado da seguinte forma. Considere quaisquer $v \in I(U)$ e $w \in \bar{L}(U)$. Pelo Lema 3.1.3, 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 $Uu_n - u_n \rightarrow w$. Então

$$\langle v, Uu_n - u_n \rangle = \langle v, Uu_n \rangle - \langle v, u_n \rangle = \langle U^*v, u_n \rangle - \langle v, u_n \rangle = 0$$

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

$$\langle v, Uu - u \rangle = 0 \quad \text{ou seja,} \quad \langle U^*v, u \rangle - \langle v, u \rangle = 0$$

para todo $u \in H$. Isto quer dizer que $U^*v = v$. Usando Lema 3.1.3 uma vez mais, deduzimos que $v \in I(U)$. Isto mostra que $\bar{L}(U)^\perp \subset I(U)$ e, portanto, a prova de (3.1.5) está completa. Como consequência, usando (3.1.1),

$$H = I(U) \oplus \bar{L}(U) \quad (3.1.6)$$

Agora vamos verificar a igualdade (3.1.4), 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 $n \rightarrow \infty$. Isto prova (3.1.4) 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

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

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

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} U^j v = 0 \quad \text{para todo } v \in L(U). \quad (3.1.7)$$

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

$$\left\| \frac{1}{n} \sum_{j=0}^{n-1} U^j v - \frac{1}{n} \sum_{j=0}^{n-1} U^j v_k \right\| \leq \frac{1}{n} \sum_{j=0}^{n-1} \|U^j (v - v_k)\| \leq \|v - v_k\|$$

para todo n e todo k . Juntamente com (3.1.7), isto implica que

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} U^j v = 0 \quad \text{para todo } v \in \bar{L}(U). \quad (3.1.8)$$

Como a relação (3.1.5) implica que $Pv = 0$ para todo $v \in \bar{L}(U)$, isto mostra que (3.1.4) vale também quando $v \in \bar{L}(U)$.

O caso geral de (3.1.4) segue imediatamente, já que $H = I(U) \oplus \bar{L}(U)$. \square

3.1.3 Convergência em $L^2(\mu)$

Dada uma transformação mensurável $f : M \rightarrow M$ que preserve uma probabilidade μ em M , dizemos que uma função mensurável $\psi : M \rightarrow \mathbb{R}$ é *invariante* se $\psi \circ f = \psi$ em μ -quase todo ponto. O seguinte resultado é um caso particular do Teorema 3.1.5:

Teorema 3.1.6. Para qualquer $\varphi \in L^2(\mu)$ a sequência

$$\frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f^j \quad (3.1.9)$$

converge em $L^2(\mu)$ para a projeção ortogonal $\tilde{\varphi}$ da função φ no subespaço das funções invariantes. Se f é invertível, então a sequência

$$\frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f^{-j} \quad (3.1.10)$$

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

Demonstração. Seja $U = U_f : L^2(\mu) \rightarrow L^2(\mu)$ o operador de Koopman de uma transformação $f : M \rightarrow M$ 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 3.1.5, a sequência em (3.1.9) 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 (3.1.10) 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. \square

3.1.4 Exercícios

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

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

3.1.2. 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 1.2.5].

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

3.1.4. Enuncie e prove uma versão do Teorema de Von Neumann para fluxos.

3.1.5. Seja μ uma probabilidade invariante por um fluxo contínuo $f_t : M \rightarrow M$ num espaço métrico compacto M . Sejam $U_t : L^2(\mu) \rightarrow L^2(\mu)$, $t \in \mathbb{R}$ o grupo a 1-parâmetro de operadores unitários dados por $U_t \varphi = \varphi \circ f_t$ e $A : L^2(\mu) \rightarrow L^2(\mu)$ o operador auto-adjunto associado (dado pelo Teorema 3.1.4). Mostre que 0 é um auto-valor simples de A se, e somente se, toda função $\varphi \in L^2(\mu)$ que satisfaz $U_t \varphi = \varphi$ para todo $t \in \mathbb{R}$ é constante em μ -quase todo ponto.

3.2 Teorema ergódico de Birkhoff

O teorema que apresentamos nesta seção foi demonstrado por George David Birkhoff¹, 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 é uma propriedade mais forte do que convergência em $L^2(\mu)$.

3.2.1 Tempo médio de visita

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

Teorema 3.2.1 (Birkhoff). *Seja $f : M \rightarrow M$ uma transformação mensurável e μ uma probabilidade invariante por f . Dado qualquer conjunto mensurável $E \subset M$, o tempo médio visita*

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

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

$$\tau(E, f(x)) = \tau(E, x). \quad (3.2.1)$$

De fato, por definição,

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

Como a função característica é limitada, o último limite é igual a zero. Isto prova a igualdade (3.2.1).

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

Exemplo 3.2.2. Considere o número $x \in (0, 1)$ definido pela expansão decimal $x = 0, a_1 a_2 a_3 \dots$, onde $a_i = 1$ se $2^k \leq i < 2^{k+1}$ com k par e $a_i = 0$ se $2^k \leq i < 2^{k+1}$ com k ímpar. Ou seja,

$$x = 0, 100111100000000111111111111110\dots,$$

¹Seu filho, Garret Birkhoff, também foi um matemático, bem conhecido por seus trabalhos em Álgebra. É devida a ele a noção de distância projetiva, que usaremos na Seção 12.3.

onde os blocos alternantes de 0s e de 1s têm comprimentos dados pelas sucessivas potências de dois. Seja $f : [0, 1] \rightarrow [0, 1]$ 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 $k = 2q$ então

$$\frac{1}{n} \sum_{j=0}^{n-1} \mathcal{X}_E(f^j(x)) = \frac{1 + 2^2 + 2^4 + \dots + 2^{k-2}}{2^k - 1} = \frac{1}{3}.$$

Por outro lado, se $n = 2^k - 1$ e $k = 2q + 1$ então, quando $q \rightarrow \infty$,

$$\frac{1}{n} \sum_{j=0}^{n-1} \mathcal{X}_E(f^j(x)) = \frac{1 + 2^2 + 2^4 + \dots + 2^{k-1}}{2^k - 1} = \frac{1}{3} \frac{2^{k+1} - 1}{2^k - 1} \rightarrow \frac{2}{3}.$$

Assim, o tempo médio de visita de x ao conjunto E não existe.

3.2.2 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 3.2.1 para o caso em que φ é uma função integrável qualquer:

Teorema 3.2.3 (Birkhoff). *Seja $f : M \rightarrow M$ uma transformação mensurável e seja μ uma probabilidade invariante por f . Dada qualquer função integrável $\varphi : M \rightarrow \mathbb{R}$, o limite*

$$\tilde{\varphi}(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) \tag{3.2.2}$$

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 (3.2.1):

Proposição 3.2.4. *Seja $\varphi : M \rightarrow \mathbb{R}$ uma função integrável. Então,*

$$\tilde{\varphi}(f(x)) = \tilde{\varphi}(x) \quad \text{para } \mu\text{-quase todo ponto } x \in M. \tag{3.2.3}$$

Demonstração. Por definição,

$$\begin{aligned}\tilde{\varphi}(f(x)) &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \varphi(f^j(x)) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) + \frac{1}{n} [\varphi(f^n(x)) - \varphi(x)] \\ &= \tilde{\varphi}(x) + \lim_{n \rightarrow \infty} \frac{1}{n} [\varphi(f^n(x)) - \varphi(x)]\end{aligned}$$

Vamos precisar do seguinte lema:

Lema 3.2.5. *Se ϕ é uma função integrável então $\lim_n n^{-1} \phi(f^n(x)) = 0$ para μ -quase todo ponto $x \in M$.*

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

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

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

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

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 $\lim_n (1/n)\phi(f^n(x)) = 0$. \square

Aplicando o Lemma 3.2.5 à função $\phi = \varphi$ obtemos a igualdade (3.2.3). Isto completa a demonstração da Proposição 3.2.4. \square

Em geral, o subconjunto com medida total onde vale a convergência (3.2.2) no Teorema 3.2.3 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 3.2.6. *Suponha que M é um espaço métrico compacto e $f : M \rightarrow M$ é uma aplicação contínua. Então existe um conjunto mensurável $G \subset M$ com $\mu(G) = 1$ tal que*

$$\frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) \rightarrow \tilde{\varphi}(x) \quad (3.2.4)$$

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

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 (3.2.4) é válido para todo $x \in G(\varphi)$. Pelo Teorema A.3.13, 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 (3.2.4) 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

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

Então, dado qualquer ponto $x \in G$,

$$\begin{aligned} \limsup_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) &\leq \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi_k(f^j(x)) + \varepsilon = \tilde{\varphi}_k(x) + \varepsilon \\ \liminf_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) &\geq \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi_k(f^j(x)) - \varepsilon = \tilde{\varphi}_k(x) - \varepsilon. \end{aligned}$$

Isto implica que

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

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

Em geral não é possível dizer nada sobre a velocidade da convergência no Teorema 3.2.3. 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 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

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

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.

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 $f : M \rightarrow M$. Dizemos que um conjunto mensurável $W \subset M$ é *errante* se as pré-imagens $f^{-i}(W)$, $i \geq 0$ 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,

- ou $\liminf_n \frac{1}{a_n} \sum_{j=0}^{n-1} \varphi \circ f^j = 0$ em quase todo ponto, para toda $\varphi \in L^1(\mu)$;
- ou existe $(n_k)_k \rightarrow \infty$ tal que $\lim_k \frac{1}{a_{n_k}} \sum_{j=0}^{n_k-1} \varphi \circ f^j = \infty$ em quase todo ponto, para toda $\varphi \in L^1(\mu)$.

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

3.2.3 Teorema de von Neumann e conseqüências

O teorema de von Neumann (Teorema 3.1.6) 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 $\|\tilde{\varphi}\|_2 \leq \|\varphi\|_2$. Para isso, note que

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

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

$$\left[\int |\tilde{\varphi}|^2 d\mu \right]^{1/2} \leq \liminf_n \left[\int \left(\frac{1}{n} \sum_{j=0}^{n-1} |\varphi \circ f^j| \right)^2 d\mu \right]^{1/2}. \quad (3.2.5)$$

Podemos usar a desigualdade de Minkowski (Teorema A.5.3) para majorar a seqüência do lado direito:

$$\left[\int \left(\frac{1}{n} \sum_{j=0}^{n-1} |\varphi \circ f^j| \right)^2 d\mu \right]^{1/2} \leq \frac{1}{n} \sum_{j=0}^{n-1} \left[\int |\varphi \circ f^j|^2 d\mu \right]^{1/2}. \quad (3.2.6)$$

Como μ é invariante por f , a expressão do lado direito é igual a $[\int |\varphi|^2 d\mu]^{1/2}$. Portanto, (3.2.5) e (3.2.6) implicam que $\|\tilde{\varphi}\|_2 \leq \|\varphi\|_2 < \infty$.

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 $|\varphi| \leq C$. Então

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

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

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

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

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

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 $k \geq k_0$. Note que $\|(\varphi - \varphi_k) \circ f^j\|_2$ é igual a $\|\varphi - \varphi_k\|_2$ para todo $j \geq 0$, porque a medida μ é invariante. Logo,

$$\left\| \frac{1}{n} \sum_{j=0}^{n-1} (\varphi - \varphi_k) \circ f^j \right\|_2 \leq \|\varphi - \varphi_k\|_2 < \varepsilon/3 \quad \text{para todo } n \geq 1 \text{ e } k \geq k_0. \quad (3.2.7)$$

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

$$\|\tilde{\varphi} - \tilde{\varphi}_k\|_2 \leq \|\varphi - \varphi_k\|_2 < \varepsilon/3 \quad \text{para todo } k \geq k_0. \quad (3.2.8)$$

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

$$\left\| \frac{1}{n} \sum_{j=0}^{n-1} \varphi_k \circ f^j - \tilde{\varphi}_k \right\|_2 < \varepsilon/3 \quad \text{para todo } n \geq n_0(k). \quad (3.2.9)$$

Somando (3.2.7), (3.2.8), (3.2.9) obtemos

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

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

No Exercício 3.2.5 propomos uma generalização destas conclusões para um espaço $L^p(\mu)$ qualquer.

Corolário 3.2.7. *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 3.1.6 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, $P(\varphi) = \tilde{\varphi}$. \square

Corolário 3.2.8. *Se $f : M \rightarrow M$ 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:*

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f^{-j} = \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f^j \quad \text{em } \mu\text{-quase todo ponto.} \quad (3.2.10)$$

Demonstração. O limite do lado esquerdo de (3.2.10) 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 $L^2(\mu)$. \square

3.2.4 Exercícios

3.2.1. Seja $X = \{x_1, \dots, x_r\}$ um conjunto finito e seja $\sigma : X \rightarrow X$ 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 $\varphi : X \rightarrow \mathbb{R}$ prove que

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

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

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi(\sigma^i(x)) = \frac{\varphi(x) + \varphi(\sigma(x)) + \dots + \varphi(\sigma^{p-1}(x))}{p}.$$

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

3.2.2. Verifique que o Lema 3.2.5 também pode ser deduzido do teorema ergódico de Birkhoff.

3.2.3. Uma função $\varphi : \mathbb{Z} \rightarrow \mathbb{R}$ é dita *uniformemente quase periódica* se para cada $\varepsilon > 0$ existe $L(\varepsilon) \in \mathbb{N}$ tal que todo intervalo $\{n+1, \dots, n+L(\varepsilon)\}$ 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

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

- (c) Mostre que a sequência $(1/n) \sum_{j=1}^n \varphi(j)$ converge para algum número real quando $n \rightarrow \infty$.
 (d) Mais geralmente, prove que $\lim_n (1/n) \sum_{l=1}^n \varphi(x+l)$ existe para todo $x \in \mathbb{Z}$ e é independente de x .

3.2.4. Prove que para Lebesgue quase todo ponto $x \in [0, 1]$, a média geométrica dos números inteiros a_1, \dots, a_n, \dots na expansão de x em fração contínua converge para algum valor, ou seja, existe $b \in \mathbb{R}$ tal que $\lim_n (a_1 a_2 \dots a_n)^{1/n} = b$. [Observação: Compare com o Exercício 4.2.11.]

3.2.5. Seja $\varphi : M \rightarrow \mathbb{R}$ uma função integrável e seja $\tilde{\varphi}$ a sua média temporal, dada pelo Teorema 3.2.3. 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 $L^p(\mu)$.

3.2.6. 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

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

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

3.2.7. Prove que se um homeomorfismo $f : M \rightarrow M$ de um espaço métrico M só admite uma medida 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 .

3.3 Teorema ergódico subaditivo

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

$$\varphi_{m+n} \leq \varphi_m + \varphi_n \circ f^m \quad \text{para todo } m, n \geq 1. \quad (3.3.1)$$

Exemplo 3.3.1. A sequência $\varphi_n : M \rightarrow \mathbb{R}$ diz-se *aditiva* se vale a igualdade em (3.3.1), ou seja, se $\varphi_{m+n} = \varphi_m + \varphi_n \circ f^m$ para todo $m, n \geq 1$. 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 $d \geq 2$. Então

$$\|A\| = \sup \left\{ \frac{\|Av\|}{\|v\|} : v \in \mathbb{R}^d \setminus \{0\} \right\} \quad (3.3.2)$$

Compare com a Equação 2.3.1. Segue diretamente da definição que a norma do produto de duas matrizes é menor ou igual que o produto das normas dessas matrizes:

$$\|AB\| \leq \|A\| \|B\|. \quad (3.3.3)$$

Exemplo 3.3.2. Seja $A : M \rightarrow \text{GL}(d)$ uma função mensurável com valores no conjunto $\text{GL}(d)$ das matrizes quadradas invertíveis de dimensão d . Defina $\phi^n(x) = A(f^{n-1}(x)) \cdots A(f(x))A(x)$ para todo $n \geq 1$ 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 (3.3.3),

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

para todo m, n e x .

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

Teorema 3.3.3 (Kingman). *Seja μ uma probabilidade invariante para uma transformação $f : M \rightarrow M$ e seja $\varphi_n : M \rightarrow \mathbb{R}$, $n \geq 1$ 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 $\varphi : M \rightarrow [-\infty, +\infty)$. 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 3.3.3 que vamos apresentar é devida a Avila, Bochi [AB], os quais se inspiraram na demonstração do Teorema 3.2.3 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 3.3.3.

3.3.1 Preparação da demonstração

Uma sequência $(a_n)_n$ em $[-\infty, +\infty)$ é dita *subaditiva* se vale $a_{m+n} \leq a_m + a_n$ para todo $m, n \geq 1$.

Lema 3.3.4. *Se $(a_n)_n$ é uma sequência subaditiva então*

$$\lim_n \frac{a_n}{n} = \inf_n \frac{a_n}{n} \in [-\infty, \infty). \quad (3.3.4)$$

Demonstração. Se $a_m = -\infty$ para algum m então, pela subaditividade, temos que $a_n = -\infty$ para todo $n > m$. Então os dois lados de (3.3.4) são iguais a $-\infty$, e portanto o lema é válido neste caso. A partir daqui suporemos que $a_n \in \mathbb{R}$ para todo n .

Seja $L = \inf_n (a_n/n) \in [-\infty, +\infty)$ e seja B qualquer número real maior do que L . Então podemos encontrar $k \geq 1$ tal que

$$\frac{a_k}{k} < B.$$

Para $n > k$, podemos escrever $n = kp + q$, onde p e q são números inteiros tais que $p \geq 1$ e $1 \leq q \leq k$. Então, por subaditividade,

$$a_n \leq a_{kp} + a_q \leq pa_k + a_q \leq pa_k + \alpha,$$

onde $\alpha = \max\{a_i : 1 \leq i \leq k\}$. Logo,

$$\frac{a_n}{n} \leq \frac{pk}{n} \frac{a_k}{k} + \frac{\alpha}{n}.$$

Observe que pk/n converge para 1 e α/n converge para zero quando $n \rightarrow \infty$. Portanto, uma vez que $a_k/k < B$, temos

$$L \leq \frac{a_n}{n} < B$$

para todo n suficientemente grande. Fazendo $B \rightarrow L$, concluímos que

$$\lim_n \frac{a_n}{n} = L = \inf_n \frac{a_n}{n}.$$

Isto completa o argumento. \square

Agora seja $(\varphi_n)_n$ como nas hipóteses do Teorema 3.3.3. Por subaditividade,

$$\varphi_n \leq \varphi_1 + \varphi_1 \circ f + \cdots + \varphi_1 \circ f^{n-1}.$$

Esta relação permanece válida quando colocamos φ_n^+ e φ_1^+ no lugar de φ_n e φ_1 . Logo, a hipótese de que $\varphi_1^+ \in L^1(\mu)$ implica que $\varphi_n^+ \in L^1(\mu)$ para todo n . Além disso, a hipótese de que $(\varphi_n)_n$ é subaditiva implica que

$$a_n = \int \varphi_n d\mu, \quad n \geq 1,$$

é uma sequência subaditiva em $[-\infty, +\infty)$. Logo, pelo Lema 3.3.4, o limite

$$L = \lim_n \frac{a_n}{n} = \inf_n \frac{a_n}{n} \in [-\infty, \infty).$$

existe. Defina $\varphi_- : M \rightarrow [-\infty, \infty]$ e $\varphi_+ : M \rightarrow [-\infty, \infty]$ por

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

É claro que $\varphi_-(x) \leq \varphi_+(x)$ para todo $x \in M$. Vamos provar que

$$\int \varphi_- d\mu \geq L \geq \int \varphi_+ d\mu, \tag{3.3.5}$$

desde que toda função φ_n seja limitada por baixo. Consequentemente, as duas funções φ_- e φ_+ coincidem em μ -quase todo ponto e a sua integral é igual a L . Desta forma o teorema ficará demonstrado neste caso, com $\varphi = \varphi_- = \varphi_+$ (o fato de que φ é f -invariante é provado no Exercício 3.3.1). Ao final, removeremos a condição de limitação usando um truque de truncagem.

3.3.2 Lema fundamental

Nesta seção suporemos que $\varphi_- > -\infty$ em todo ponto. Fixado $\varepsilon > 0$, defina, para cada $k \in \mathbb{N}$,

$$E_k = \{x \in M : \varphi_j(x) \leq j(\varphi_-(x) + \varepsilon) \text{ para algum } j \in \{1, \dots, k\}\}.$$

É claro que $E_k \subset E_{k+1}$ para todo k . Além disso, a definição de $\varphi_-(x)$ implica que $M = \cup_k E_k$. Logo $\mu(E_k) \rightarrow 1$ quando $k \rightarrow \infty$. Defina também

$$\psi_k(x) = \begin{cases} \varphi_-(x) + \varepsilon & \text{se } x \in E_k \\ \varphi_1(x) & \text{se } x \in E_k^c. \end{cases}$$

Segue da definição que $\psi_k(x) \geq \varphi_-(x) + \varepsilon$ para todo $x \in M$. O passo crucial na prova do teorema é a seguinte estimativa:

Lema 3.3.5. *Para todo $n > k \geq 1$ e μ -quase todo $x \in M$,*

$$\varphi_n(x) \leq \sum_{i=0}^{n-k-1} \psi_k(f^i(x)) + \sum_{i=n-k}^{n-1} \max\{\psi_k, \varphi_1\}(f^i(x)).$$

Demonstração. Tome $x \in M$ tal que $\varphi_-(x) = \varphi_-(f^j(x))$ para todo $j \geq 1$ (isso vale em μ -quase todo ponto; veja o Exercício 3.3.1). Considere a sequência, possivelmente finita, de número inteiros

$$m_0 \leq n_1 < m_1 \leq n_2 < m_2 \leq \dots \tag{3.3.6}$$

definida indutivamente da seguinte forma (veja também a Figura 3.1).

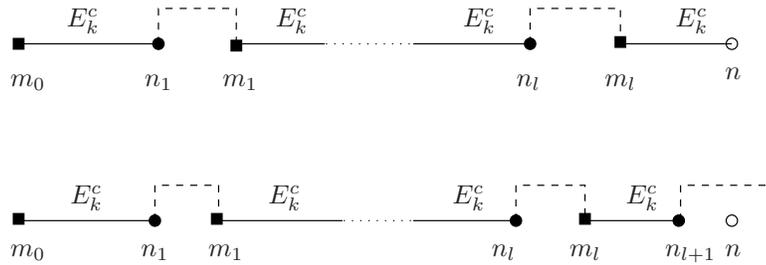


Figura 3.1: Decomposição da trajetória de um ponto

Defina $m_0 = 0$. Seja n_j o menor inteiro maior ou igual a m_{j-1} tal que $f^{n_j}(x) \in E_k$ (caso exista). Então, pela definição de E_k , existe m_j tal que $1 \leq m_j - n_j \leq k$ e

$$\varphi_{m_j - n_j}(f^{n_j}(x)) \leq (m_j - n_j)(\varphi_-(f^{n_j}(x)) + \varepsilon). \tag{3.3.7}$$

Isto completa a definição da sequência (3.3.6). Agora, dado $n \geq k$, seja $l \geq 0$ o maior número inteiro tal que $m_l \leq n$. Pela subaditividade,

$$\varphi_{n_j - m_{j-1}}(f^{m_{j-1}}(x)) \leq \sum_{i=m_{j-1}}^{n_j-1} \varphi_1(f^i(x))$$

para todo $j = 1, \dots, l$ tal que $m_{j-1} \neq n_j$, e analogamente para $\varphi_{n-m_l}(f^{m_l}(x))$. Assim,

$$\varphi_n(x) \leq \sum_{i \in I} \varphi_1(f^i(x)) + \sum_{j=1}^l \varphi_{m_j - n_j}(f^{n_j}(x)) \quad (3.3.8)$$

onde $I = \cup_{j=1}^l [m_{j-1}, n_j) \cup [m_l, n)$. Observe que

$$\varphi_1(f^i(x)) = \psi_k(f^i(x)) \quad \text{para todo } i \in \cup_{j=1}^l [m_{j-1}, n_j) \cup [m_l, \min\{n_{l+1}, n\}),$$

já que $f^i(x) \in E_k^c$ em todos esses casos. Além disso, como φ_- é constante em órbitas (veja o Exercício 3.3.1) e $\psi_k \geq \varphi_- + \varepsilon$, a relação (3.3.7) nos dá que

$$\varphi_{m_j - n_j}(f^{n_j}(x)) \leq \sum_{i=n_j}^{m_j-1} (\varphi_-(f^i(x)) + \varepsilon) \leq \sum_{i=n_j}^{m_j-1} \psi_k(f^i(x))$$

para todo $j = 1, \dots, l$. Deste modo, usando (3.3.8), concluímos que

$$\varphi_n(x) \leq \sum_{i=0}^{\min\{n_{l+1}, n\}-1} \psi_k(f^i(x)) + \sum_{i=n_{l+1}}^{n-1} \varphi_1(f^i(x)).$$

Como $n_{l+1} > n - k$, o lema está provado. \square

3.3.3 Estimativa da função φ_-

Na direção de provar (3.3.5), nesta seção vamos provar o seguinte lema:

Lema 3.3.6. $\int \varphi_- d\mu = L$

Demonstração. Suponha, por um instante, que φ_n/n está uniformemente limitado por baixo, ou seja, que existe $\kappa > 0$ tal que $\varphi_n/n \geq -\kappa$ para todo n . Aplicando o lema de Fatou (Teorema A.2.10) à sequência de funções não-negativas $\varphi_n/n + \kappa$, obtemos que φ_- é integrável e

$$\int \varphi_- d\mu \leq \lim_n \int \frac{\varphi_n}{n} d\mu = L.$$

Para provar a outra desigualdade, observe que o Lema 3.3.5 implica que

$$\frac{1}{n} \int \varphi_n d\mu \leq \frac{n-k}{n} \int \psi_k d\mu + \frac{k}{n} \int \max\{\psi_k, \varphi_1\} d\mu. \quad (3.3.9)$$

Note que $\max\{\psi_k, \varphi_1\} \leq \max\{\varphi_- + \varepsilon, \varphi_1^+\}$ e que esta última função é integrável. Portanto, o \limsup da última parcela em (3.3.9) quando $n \rightarrow \infty$ é menor ou igual que zero. Então, fazendo $n \rightarrow \infty$ obtemos que $L \leq \int \psi_k d\mu$ para todo k . Logo, fazendo $k \rightarrow \infty$, concluímos que

$$L \leq \int \varphi_- d\mu + \varepsilon.$$

Finalmente, fazendo $\varepsilon \rightarrow 0$ obtemos que $L \leq \int \varphi_- d\mu$. Isto prova o lema quando φ_n/n está uniformemente limitado por baixo.

Agora resta remover essa hipótese. Defina, para cada $\kappa > 0$,

$$\varphi_n^\kappa = \max\{\varphi_n, -\kappa n\} \quad \text{e} \quad \varphi_-^\kappa = \max\{\varphi_-, -\kappa\}.$$

A sequência $(\varphi_n^\kappa)_n$ satisfaz as hipóteses do Teorema 3.3.3, ou seja, ela é subaditiva e a parte positiva de φ_1^κ é integrável. Além disso, $\varphi_-^\kappa = \liminf_n (\varphi_n^\kappa/n)$. Então, o argumento do parágrafo anterior mostra que

$$\int \varphi_-^\kappa d\mu = \inf_n \frac{1}{n} \int \varphi_n^\kappa d\mu. \quad (3.3.10)$$

Pelo teorema da convergência monótona (Teorema A.2.9), também temos que

$$\int \varphi_n d\mu = \inf_\kappa \int \varphi_n^\kappa d\mu \quad \text{e} \quad \int \varphi_- d\mu = \inf_\kappa \int \varphi_-^\kappa d\mu. \quad (3.3.11)$$

Combinando as relações (3.3.10) e (3.3.11), obtemos que

$$\int \varphi_- d\mu = \inf_\kappa \int \varphi_-^\kappa = \inf_\kappa \inf_n \frac{1}{n} \int \varphi_n^\kappa d\mu = \inf_n \frac{1}{n} \int \varphi_n d\mu = L.$$

Isto completa a demonstração do lema. \square

3.3.4 Majoração da função φ_+

Para completar a prova de (3.3.5), vamos mostrar que $\int \varphi_+ d\mu \leq L$ desde que $\inf_n \varphi_n$ seja finito para todo n . Começamos por provar o seguinte resultado auxiliar:

Lema 3.3.7. *Para todo k fixado,*

$$\limsup_n \frac{\varphi_{kn}}{n} = k \limsup_n \frac{\varphi_n}{n}.$$

Demonstração. A desigualdade \leq é clara, uma vez que φ_{kn}/kn é subsequência de φ_n/n . Para mostrar a desigualdade contrária, escrevemos $n = kq_n + r_n$ com $r_n \in \{1, \dots, k\}$. Pela subaditividade,

$$\varphi_n \leq \varphi_{kq_n} + \varphi_{r_n} \circ f^{kq_n} \leq \varphi_{kq_n} + \psi \circ f^{kq_n}$$

onde $\psi = \max\{\varphi_1^+, \dots, \varphi_k^+\}$. Observe que $n/q_n \rightarrow k$ quando $n \rightarrow \infty$. Além disso, como $\psi \in L^1(\mu)$, podemos usar o Lema 3.2.5 para ver que $\psi \circ f^n/n$ converge para zero em μ -quase todo ponto. Assim, dividindo a relação anterior por n e tomando o \limsup quando $n \rightarrow \infty$ nós obtemos que

$$\limsup_n \frac{1}{n} \varphi_n \leq \limsup_n \frac{1}{n} \varphi_{kq_n} + \limsup_n \frac{1}{n} \psi \circ f^{kq_n} = \frac{1}{k} \limsup_q \frac{1}{q} \varphi_{kq},$$

como afirmado no lema. \square

Lema 3.3.8. *Suponha que $\inf_n \varphi_n > -\infty$ para todo n . Então $\int \varphi_+ d\mu \leq L$.*

Demonstração. Para cada k fixado e $n \geq 1$, considere $\theta_n = -\sum_{j=0}^{n-1} \varphi_k \circ f^{jk}$. Observe que

$$\int \theta_n d\mu = -n \int \varphi_k d\mu \quad \text{para todo } n, \quad (3.3.12)$$

uma vez que f^k preserva a medida μ . Como a sequência $(\varphi_n)_n$ é subaditiva, $\theta_n \leq -\varphi_{kn}$ para todo n . Logo, usando o Lema 3.3.7,

$$\theta_- = \liminf_n \frac{\theta_n}{n} \leq -\limsup_n \frac{\varphi_{kn}}{n} = -k \limsup_n \frac{\varphi_n}{n} = -k\varphi_+$$

e, portanto,

$$\int \theta_- d\mu \leq -k \int \varphi_+ d\mu. \quad (3.3.13)$$

Observe também que a sequência $(\theta_n)_n$ é aditiva: $\theta_{m+n} = \theta_m + \theta_n \circ f^{km}$ para todo $m, n \geq 1$. Como $\theta_1 = -\varphi_k$ é majorada por $-\inf \varphi_k$, também temos que a função θ_1^+ é limitada e, por consequência, integrável. Assim, podemos aplicar o Lema 3.3.6, juntamente com a igualdade (3.3.12), para concluir que

$$\int \theta_- d\mu = \liminf_n \int \frac{\theta_n}{n} d\mu = - \int \varphi_k d\mu. \quad (3.3.14)$$

Juntando as relações (3.3.13) e (3.3.14) obtemos que

$$\int \varphi_+ d\mu \leq \frac{1}{k} \int \varphi_k d\mu.$$

Finalmente, tomando o ínfimo sobre k obtemos que $\int \varphi_+ d\mu \leq L$. \square

Os Lemas 3.3.6 e 3.3.8 provam a relação (3.3.5) e, portanto, o Teorema 3.3.3 quando $\inf \varphi_k > -\infty$ para todo k . No caso geral, defina

$$\varphi_n^\kappa = \max\{\varphi_n, -\kappa n\} \quad \text{e} \quad \varphi_-^\kappa = \max\{\varphi_-, -\kappa\} \quad \text{e} \quad \varphi_+^\kappa = \max\{\varphi_+, -\kappa\}$$

para cada constante $\kappa > 0$. Os argumentos anteriores podem ser aplicados à sequência $(\varphi_n^\kappa)_n$ para todo $\kappa > 0$ fixado. Portanto, $\varphi_+^\kappa = \varphi_-^\kappa$ em μ -quase todo ponto para todo $\kappa > 0$. Como $\varphi_-^\kappa \rightarrow \varphi_-$ e $\varphi_+^\kappa \rightarrow \varphi_+$ quando $\kappa \rightarrow \infty$, segue que $\varphi_- = \varphi_+$ em μ -quase todo ponto. A prova do Teorema 3.3.3 está completa.

3.3.5 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 3.2.3) é um caso particular do Teorema 3.3.3.

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

Seja $f : M \rightarrow M$ uma transformação mensurável e seja μ uma probabilidade invariante. Seja $\theta : M \rightarrow \text{GL}(d)$ uma função mensurável com valores no conjunto $\text{GL}(d)$ das matrizes quadradas invertíveis de dimensão d . Seja $\theta^{-1} : M \rightarrow \text{GL}(d)$ a função definida por $\theta^{-1}(x) =$ matriz inversa de $\theta(x)$. O cociclo definido por θ é a sequência de funções

$$\phi^n(x) = \theta(f^{n-1}(x)) \cdots \theta(f(x))\theta(x) \quad \text{e} \quad \phi^{-n}(x) = \text{inversa de } \phi^n(f^{-n}(x))$$

para todo $n \geq 1$ e $x \in M$. Deixamos ao cuidado do leitor verificar que

$$\phi^{m+n}(x) = \phi^n(f^m(x)) \cdot \phi^m(x) \quad \text{para todo } m, n \in \mathbb{Z} \text{ e todo } x \in M. \quad (3.3.15)$$

Aliás, é por esta razão que a sequência $(\phi^n)_n$ é chamada de cociclo.

Teorema 3.3.9 (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

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

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

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

Para deduzir este resultado do Teorema 3.3.3 basta observar que as sequências

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

são subaditivas (lembre do Exemplo 3.3.2).

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 3.3.9, para μ -quase todo $x \in M$ existe um número inteiro positivo $k = k(x)$ e existem números reais $\lambda_1(x) > \dots > \lambda_k(x)$ e uma filtração

$$\mathbb{R}^d = V_x^1 > \dots > V_x^k > V_x^{k+1} = \{0\} \quad (3.3.16)$$

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

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

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

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

Além disso, os números $k(x)$ e $\lambda_1(x), \dots, \lambda_k(x)$ e os subespaços V_x^1, \dots, 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, supondo também que $\log^+ \|\theta^{-1}\| \in L^1(\mu)$, é possível obter uma conclusão mais forte: no lugar da filtração (3.3.16) obtemos uma decomposição

$$\mathbb{R}^d = E_x^1 \oplus \dots \oplus E_x^k \quad (3.3.17)$$

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

$$(a2) \quad \theta(x) \cdot E_x^i = E_{f(x)}^i \text{ e } V_x^i = V_x^{i+1} \oplus E_x^i;$$

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

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

As expressões (3.3.16) e (3.3.17) estão relacionadas por $V_x^i = E_x^i \oplus V_x^{i+1}$. Em particular, $\dim E_x^i = \dim V_x^i - \dim V_x^{i+1}$.

3.3.6 Exercícios

3.3.1. 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 $x \in M$.

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

3.3.3. Seja $f : M \rightarrow M$ 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.

3.3.4. Chamamos *constante temporal* de uma sequência subaditiva de funções $(\varphi_n)_n$ 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.

3.3.5. 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

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

3.3.6. Nas condições do teorema de Furstenberg-Kesten, para cada $k \geq 1$, seja Ψ_k a média temporal de ψ_k relativamente a f^k . Mostre que $\lambda_{\max}(x) \leq \Psi_k(x)$ para todo k e μ -quase todo x . Usando o Exercício 3.3.5, mostre que para todo $\rho > 0$ e μ -quase todo x existe k tal que $\Psi_k(x) \leq \lambda_{\max}(x) + \rho$.

3.4 Tempo discreto e tempo contínuo

A maior parte do tempo focamos a nossa apresentação no contexto dos sistemas dinâmicos com tempo discreto. No entanto, quase tudo que foi dito até aqui se estende, de forma mais ou menos direta, para sistemas com tempo contínuo. A razão das duas teorias serem tão semelhantes é que é possível relacionar sistemas de um tipo com sistemas do outro tipo, por meio de certas construções que vamos apresentar a seguir. Por simplicidade, nos ateremos ao caso de sistemas invertíveis.

3.4.1 Fluxos suspensão

A nossa primeira construção associa a cada transformação invertível $f : M \rightarrow M$ e cada função mensurável $\tau : M \rightarrow (0, \infty)$ um fluxo $g^t : N \rightarrow N$, $t \in \mathbb{R}$ chamado *suspensão de f com tempo de retorno τ* , cujas propriedades de recorrência estão

diretamente ligadas às propriedades da transformação f . Em particular, a cada medida μ invariante por f vamos associar uma medida ν invariante pelo fluxo. Na construção suporemos que a função τ é tal que, para todo $x \in M$,

$$\sum_{j=1}^{\infty} \tau(f^j(x)) = \sum_{j=1}^{\infty} \tau(f^{-j}(x)) = +\infty. \quad (3.4.1)$$

Esse é o caso, por exemplo, se τ estiver afastada de zero.

O primeiro passo é construir o domínio N do fluxo suspensão. Consideremos a transformação $F : M \times \mathbb{R} \rightarrow M \times \mathbb{R}$ dada por $F(x, s) = (f(x), s - \tau(x))$. Observe que F é invertível. Seja \sim a relação de equivalência definida em $M \times \mathbb{R}$ por

$$(x, s) \sim (\tilde{x}, \tilde{s}) \Leftrightarrow \text{existe } n \in \mathbb{Z} \text{ tal que } F^n(x, s) = (\tilde{x}, \tilde{s}).$$

Denotamos por N o conjunto das classes de equivalência desta relação e por $\pi : M \times \mathbb{R} \rightarrow N$ a projeção canônica que associa a cada $(x, s) \in M \times \mathbb{R}$ a sua classe de equivalência.

Agora considere o fluxo $G^t : M \times \mathbb{R} \rightarrow M \times \mathbb{R}$ dado por $G^t(x, s) = (x, s + t)$. É imediato que $G^t \circ F = F \circ G^t$ para todo $t \in \mathbb{R}$. Isto garante que G^t , $t \in \mathbb{R}$ induz um fluxo g^t , $t \in \mathbb{R}$ no espaço quociente N , dado por

$$g^t(\pi(x, s)) = \pi(G^t(x, s)) \quad \text{para todo } x \in M \text{ e } s, t \in \mathbb{R}. \quad (3.4.2)$$

De fato, se $\pi(x, s) = \pi(\tilde{x}, \tilde{s})$ então existe $n \in \mathbb{Z}$ tal que $F^n(x, s) = (\tilde{x}, \tilde{s})$. Logo,

$$G^t(\tilde{x}, \tilde{s}) = G^t \circ F^n(x, s) = F^n \circ G^t(x, s)$$

e, portanto, $\pi(G^t(x, s)) = \pi(G^t(\tilde{x}, \tilde{s}))$. Isto mostra que o fluxo g^t , $t \in \mathbb{R}$ está realmente bem definido.

Para compreender melhor como este fluxo se relaciona com a transformação f , precisamos apresentar a construção de um ponto de vista mais concreto. Consideremos o domínio $D = \{(x, s) \in M \times \mathbb{R} : 0 \leq s < \tau(x)\}$. Afirmamos que D é um *domínio fundamental* para a relação de equivalência \sim , ou seja, ele contém exatamente um representante de cada classe de equivalência. A unicidade do representante é imediata: basta observar que se $(x, s) \in D$ então $F^n(x, s) = (x_n, s_n)$ com $s_n < 0$ para todo $n > 0$ e $s_n > \tau(f^n(x))$ para todo $n > 0$. Para provar a existência, precisamos da condição (3.4.1): ela garante que os iterados $(x_n, s_n) = F^n(x, s)$ de qualquer (x, s) satisfazem

$$\lim_{n \rightarrow +\infty} s_n = -\infty \quad \text{e} \quad \lim_{n \rightarrow -\infty} s_n = +\infty;$$

tomando n máximo tal que $s_n \geq 0$, temos que $(x_n, s_n) \in D$. Desta forma, fica provada a nossa afirmação, a qual significa que a restrição da projeção π ao domínio D é uma bijeção sobre N . Portanto, podemos identificar N com D e, em particular, podemos considerar g^t , $t \in \mathbb{R}$ como um fluxo em D .

Da mesma forma, podemos identificar M com o subconjunto $\Sigma = \pi(M \times \{0\})$ de N . Observando que

$$g^{\tau(x)}(\pi(x, 0)) = \pi(x, \tau(x)) = \pi(f(x), 0) \quad (3.4.3)$$

vemos que, com esta identificação, a transformação $f : M \rightarrow M$ corresponde à transformação de primeiro retorno (ou *transformação de Poincaré*) do fluxo suspensão a Σ . Veja a Figura 3.2.

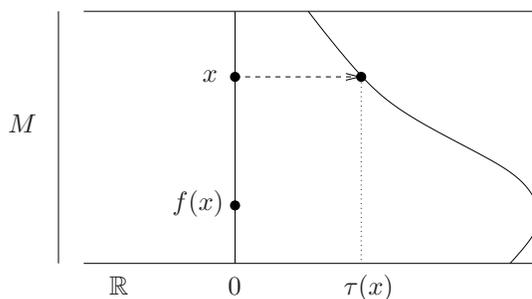


Figura 3.2: Fluxo suspensão

Agora seja μ uma medida em M invariante por f . Denotemos por ds a medida de Lebesgue na reta \mathbb{R} . É claro que a medida (infinita) $\mu \times ds$ é invariante pelo fluxo G^t , $t \in \mathbb{R}$. Além disso, ela é invariante pela transformação F , uma vez que μ é invariante por f . Chamamos *suspensão de μ com tempo de retorno τ* a medida ν definida em N por

$$\nu = \pi_*(\mu \times ds \mid D). \quad (3.4.4)$$

Em outras palavras, ν é a medida dada por

$$\int \psi d\nu = \int d\mu(x) \int_0^{\tau(x)} \psi(\pi(x, s)) ds$$

para cada função mensurável limitada $\psi : N \rightarrow (0, \infty)$. Em particular,

$$\nu(N) = \int 1 d\nu = \int \tau(x) d\mu(x) \quad (3.4.5)$$

é finito se, e somente se, a função τ é integrável para μ .

Proposição 3.4.1. *O fluxo g^t , $t \in \mathbb{R}$ preserva a medida ν .*

Demonstração. Fixemos $t \in \mathbb{R}$. Dado qualquer conjunto mensurável $B \subset N$, seja $\hat{B} = \pi^{-1}(B) \cap D$. Pela definição de ν , temos que $\nu(B) = (\mu \times ds)(\hat{B})$. Para cada $n \in \mathbb{Z}$ seja \hat{B}_n o conjunto dos $(x, s) \in \hat{B}$ tais que $G^{-t}(x, s) \in F^n(D)$ e seja $B_n = \pi(\hat{B}_n)$. Como D é um domínio fundamental, $\{\hat{B}_n : n \in \mathbb{Z}\}$ é uma partição de \hat{B} e $\{B_n : n \in \mathbb{Z}\}$ é uma partição de B . Além disso, $\hat{B}_n = \pi^{-1}(B_n) \cap D$ e, portanto, $\nu(B_n) = (\mu \times ds)(\hat{B}_n)$ para todo n . A definição do fluxo suspensão dá que

$$\pi^{-1}(g^{-t}(B_n)) = G^{-t}(\pi^{-1}(B_n)) = G^{-t}\left(\bigcup_{k \in \mathbb{Z}} F^k(\hat{B}_n)\right) = \bigcup_{k \in \mathbb{Z}} F^k(G^{-t}(\hat{B}_n)).$$

Observando que $F^{-n}(G^{-t}(\hat{B}_n)) \subset D$, concluímos que

$$\nu(g^{-t}(B_n)) = (\mu \times ds)(\pi^{-1}(g^{-t}(B_n)) \cap D) = (\mu \times ds)(F^{-n}(G^{-t}(\hat{B}_n))).$$

Como a medida $\mu \times ds$ é invariante por F e por G^t , a última expressão é igual a $(\mu \times ds)(\hat{B}_n)$. Portanto,

$$\nu(g^{-t}(B)) = \sum_{n \in \mathbb{Z}} \nu(g^{-t}(B_n)) = \sum_{n \in \mathbb{Z}} (\mu \times ds)(\hat{B}_n) = (\mu \times ds)(\hat{B}) = \nu(B).$$

Isto prova que ν é invariante pelo fluxo g^t , $t \in \mathbb{R}$. \square

No Exercício 3.4.2 convidamos o leitor a relacionar as propriedades de recorrência dos sistemas (f, μ) e (g^t, ν) .

3.4.2 Transformações de Poincaré

A seguir, apresentamos uma espécie de inversa da construção descrita na seção anterior. Seja $g^t : N \rightarrow N$, $t \in \mathbb{R}$ um fluxo mensurável e seja ν uma medida invariante. Seja $\Sigma \subset N$ uma *seção transversal* do fluxo, ou seja, um subconjunto de N tal que para todo $x \in \Sigma$ existe $\tau(x) \in (0, \infty]$ tal que $g^t(x) \notin \Sigma$ para todo $t \in (0, \tau(x))$ e $g^{\tau(x)}(x) \in \Sigma$ sempre que $\tau(x)$ for finito. Chamamos $\tau(x)$ de *tempo de primeiro retorno* de x a Σ . O nosso objetivo é construir, a partir de ν , uma medida μ invariante para a *transformação de primeiro retorno* (ou *transformação de Poincaré*)

$$f : \{x \in \Sigma : \tau(x) < \infty\} \rightarrow \Sigma, \quad f(x) = g^{\tau(x)}(x).$$

Observe que esta transformação é injetiva.

Para cada $\rho > 0$, denotamos $\Sigma_\rho = \{x \in \Sigma : \tau(x) \geq \rho\}$. Dados $A \subset \Sigma_\rho$ e $\delta \in (0, \rho]$, denotamos $A_\delta = \{g^t(x) : x \in A \text{ e } 0 \leq t < \delta\}$. Observe que a aplicação $(x, t) \mapsto g^t(x)$ é uma bijeção de $A \times [0, \delta)$ em A_δ . Suporemos que Σ está munida de uma σ -álgebra de subconjuntos mensuráveis para a qual:

1. a função τ e as transformações f e f^{-1} são mensuráveis;
2. se $A \subset \Sigma_\rho$ é mensurável então $A_\delta \subset N$ é mensurável, para todo $\delta \in (0, \rho]$.

Lema 3.4.2. *Seja A um subconjunto mensurável de Σ_ρ para algum $\rho > 0$. Então, a função $\delta \mapsto \nu(A_\delta)/\delta$ é constante no intervalo $(0, \rho]$.*

Demonstração. Considere qualquer $\delta \in (0, \rho]$ e qualquer $l \geq 1$. É claro que

$$A_\delta = \bigcup_{i=0}^{l-1} g^{i\delta/l}(A_{\delta/l})$$

e esta união é disjunta. Usando que ν é invariante pelo fluxo g^t , $t \in \mathbb{R}$, concluímos que $\nu(A_\delta) = l\nu(A_{\delta/l})$ para todo $\delta \in (0, \rho]$ e todo $l \geq 1$. Então, $\nu(A_{r\delta}) = r\nu(A_\delta)$ para todo $\delta \in (0, \rho]$ e todo número racional $r \in (0, 1)$. Usando que os dois lados desta relação variam monotonamente com r , concluímos que a igualdade permanece válida para todo número real $r \in (0, 1)$. Isso implica a conclusão do lema. \square

Dado qualquer subconjunto mensurável A de Σ_ρ , $\rho > 0$, definimos

$$\mu(A) = \frac{\nu(A_\delta)}{\delta} \quad \text{para qualquer } \delta \in (0, \rho]. \quad (3.4.6)$$

Em seguida, dado qualquer subconjunto mensurável A de Σ , definimos

$$\mu(A) = \sup_\rho \mu(A \cap \Sigma_\rho). \quad (3.4.7)$$

Veja a Figura 3.3. Deixamos ao cuidado do leitor verificar que μ é uma medida em Σ . Nós a chamaremos de *transporte* de ν pelo fluxo através de Σ .

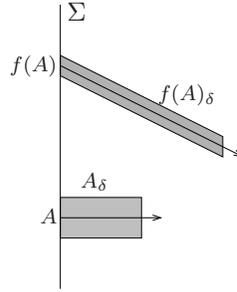


Figura 3.3: Medida transporte através de uma seção transversal

Proposição 3.4.3. *Suponhamos que a medida ν é finita. Então a medida μ em Σ é invariante pela transformação de Poincaré f .*

Demonstração. Começamos por observar que a transformação f é essencialmente sobrejetiva: o complementar da imagem $f(\Sigma)$ tem medida nula. De fato, suponha que existe um conjunto E com $\mu(E) > 0$ contido em $\Sigma \setminus f(\Sigma)$. Não é restrição supor que $E \subset \Sigma_\rho$ para algum $\rho > 0$. Então, $\nu(E_\rho) > 0$. Como ν é finita, por hipótese, podemos aplicar o teorema de recorrência de Poincaré ao fluxo g^{-t} , $t \in \mathbb{R}$. Obtemos que existe $z \in E_\rho$ tal que $g^{-s}(z) \in E_\rho$ para valores de $s > 0$ arbitrariamente grandes. Por definição, $z = g^t(y)$ para algum $y \in E$ e algum $t \in (0, \rho]$. Por construção, a trajetória passada de y intersecta Σ e, portanto, existe $x \in \Sigma$ tal que $f(x) = y$. Isto contradiz a escolha de E . Logo a nossa afirmação está provada.

Dado um conjunto mensurável $B \subset \Sigma$, denotemos $A = f^{-1}(B)$. Além disso, dado $\varepsilon > 0$, consideremos uma partição enumerável de B em subconjuntos mensuráveis B^i satisfazendo as seguintes condições: para cada i existe $\rho_i > 0$ tal que

1. B^i e $A^i = f^{-1}(B^i)$ estão contidos em Σ_{ρ_i} ;
2. $\sup(\tau | A^i) - \inf(\tau | A^i) < \varepsilon \rho_i$.

Em seguida, escolha $t_i < \inf(\tau | A^i) \leq \sup(\tau | A^i) < s_i$ tais que $s_i - t_i < \varepsilon \rho_i$. Fixe $\delta_i = \rho_i/2$. Então, usando o fato de que f é essencialmente sobrejetiva,

$$g^{t_i}(A_{\delta_i}^i) \supset B_{\delta_i - (s_i - t_i)}^i \quad \text{e} \quad g^{s_i}(A_{\delta_i}^i) \subset B_{\delta_i + (s_i - t_i)}^i.$$

Logo, usando a hipótese de que ν é invariante,

$$\begin{aligned} \nu(A_{\delta_i}^i) &= \nu(g^{t_i}(A_{\delta_i}^i)) \geq \nu(B_{\delta_i - (s_i - t_i)}^i) \\ \nu(A_{\delta_i}^i) &= \nu(g^{s_i}(A_{\delta_i}^i)) \leq \nu(B_{\delta_i + (s_i - t_i)}^i). \end{aligned}$$

Dividindo por δ_i obtemos que

$$\begin{aligned} \mu(A^i) &\geq 1 - \frac{(s_i - t_i)}{\delta} \mu(B^i) > (1 - 2\varepsilon)\mu(B^i) \\ \mu(A^i) &\leq 1 + \frac{(s_i - t_i)}{\delta} \mu(B^i) < (1 + 2\varepsilon)\mu(B^i). \end{aligned}$$

Finalmente, somando sobre todos os valores de i , concluímos que

$$(1 - 2\varepsilon)\mu(A) \leq \mu(B) \leq (1 + 2\varepsilon)\mu(A).$$

Como ε é arbitrário, isto prova que a medida ν é invariante por f . \square

3.4.3 Exercícios

3.4.1. Verifique que a função μ definida em (3.4.6)-(3.4.7) é de fato uma medida.

3.4.2. No contexto da Seção 3.4.1, suponha que M é um espaço topológico e que $f : M \rightarrow M$ e $\tau : M \rightarrow (0, \infty)$ são contínuas. Seja $g^t : N \rightarrow N$ o fluxo de suspensão e seja ν a suspensão de uma medida boreliana μ invariante por f .

- Mostre que se $x \in M$ é recorrente para a transformação f então $\pi(x, s) \in N$ é recorrente para o fluxo g^t , qualquer que seja $s \in \mathbb{R}$.
- Mostre que $\pi(x, s) \in N$ é recorrente para o fluxo g^t , para algum $s \in \mathbb{R}$, então $x \in M$ é recorrente para f .
- Conclua que o conjunto dos pontos recorrentes de f tem medida total para μ se, e somente se, o conjunto dos pontos recorrentes de g^t , $t \in \mathbb{R}$ tem medida total para ν . Em particular, isto acontece se pelo menos uma das medidas μ ou ν é finita.

3.4.3. Seja $g^t : N \rightarrow N$, $t \in \mathbb{R}$ o fluxo definido por um campo de vetores X de classe C^1 numa variedade riemanniana compacta N . Suponha que este fluxo preserva a medida de volume ν associada à métrica riemanniana. Seja Σ uma hipersuperfície de N transversal a X e seja ν_Σ a medida de volume em Σ associada à restrição da métrica riemanniana. Defina $\phi : \Sigma \rightarrow (0, \infty)$ por $\phi(y) = |X(y) \cdot n(y)|$, onde $n(\cdot)$ é um campo de vetores unitário ortogonal a Σ . Mostre que a medida $\eta = \phi \nu_\Sigma$ é invariante pela transformação de Poincaré $f : \Sigma \rightarrow \Sigma$ do fluxo. De fato, η coincide com o transporte de ν pelo fluxo através de Σ .

3.4.4. A seguinte construção tem um papel relevante na teoria de intercâmbios de intervalos. Seja $\hat{N} \subset \mathbb{R}_+^4$ o conjunto das 4-uplas $(\lambda_1, \lambda_2, h_1, h_2)$ de números reais positivos, munido da medida de volume usual $\hat{\nu} = d\lambda_1 d\lambda_2 dh_1 dh_2$. Defina

$$F : \hat{N} \rightarrow \hat{N}, \quad F(\lambda_1, \lambda_2, h_1, h_2) = \begin{cases} (\lambda_1 - \lambda_2, \lambda_2, h_1, h_1 + h_2) & \text{se } \lambda_1 > \lambda_2 \\ (\lambda_1, \lambda_2 - \lambda_1, h_1 + h_2, h_2) & \text{se } \lambda_1 < \lambda_2 \end{cases}$$

(F não está definida quando $\lambda_1 = \lambda_2$.) Seja N o quociente de \hat{N} pela relação de equivalência $z \sim \tilde{z} \Leftrightarrow F^n(z) = \tilde{z}$ para algum $n \in \mathbb{Z}$ e seja $\pi : \hat{N} \rightarrow N$ a projeção canônica. Defina

$$G^t : \hat{N} \rightarrow \hat{N}, t \in \mathbb{R}, \quad G^t(\lambda_1, \lambda_2, h_1, h_2) = (e^t \lambda_1, e^t \lambda_2, e^{-t} h_1, e^{-t} h_2).$$

Seja $\hat{a} : \hat{N} \rightarrow (0, \infty)$ o funcional dado por $\hat{a}(\lambda_1, \lambda_2, h_1, h_2) = \lambda_1 h_1 + \lambda_2 h_2$. Para cada $c > 0$, seja \hat{N}_c o subconjunto dos $x \in \hat{N}$ tais que $\hat{a}(x) = c$, seja $\hat{\nu}_c$ a medida de volume definida em \hat{N}_c pela restrição da métrica Riemanniana de \mathbb{R}_+^4 e seja $\hat{\eta}_c = \hat{\nu}_c / \|\text{grad } \hat{a}\|$.

- (a) Mostre que F preserva o funcional \hat{a} e, portanto, existe um funcional $a : N \rightarrow (0, \infty)$ tal que $a \circ \pi = \hat{a}$. Mostre que G^t comuta com F e preserva \hat{a} . Logo, $(G^t)_t$ induz um fluxo $(g^t)_t$ no espaço quociente N , o qual preserva o funcional a . Justifique que F e $(G^t)_t$ preservam $\hat{\nu}$ e $\hat{\eta}_c$ para todo c .
- (b) Verifique que $D = \{(\lambda_1, \lambda_2, h_1, h_2) : \lambda_1 + \lambda_2 \geq 1 > \max\{\lambda_1, \lambda_2\}\}$ é um domínio fundamental para \sim . Considere a medida $\nu = \pi_*(\hat{\nu} \mid D)$ em N . Justifique que a definição não depende da escolha do domínio fundamental e mostre que ν é invariante pelo fluxo $(g^t)_t$. A medida ν é finita?
- (c) Verifique que $\Sigma = \pi(\{(\lambda_1, \lambda_2, h_1, h_2) : \lambda_1 + \lambda_2 = 1\})$ é uma seção transversal para $(g^t)_t$. Calcule a transformação de Poincaré $f : \Sigma \rightarrow \Sigma$ e a respectiva função tempo de primeiro retorno τ . Calcule a medida transporte μ de ν pelo fluxo $(g^t)_t$ através de Σ . A medida μ é finita?
- (d) Para cada $c > 0$, seja $N_c = \pi(\hat{N}_c)$ e $\eta_c = \pi_*(\hat{\eta}_c \cap D)$. Mostre que N_c e η_c são invariantes por $(g^t)_t$, para todo $c > 0$. Verifique que $\eta_c(N_c) < \infty$ para todo c . Conclua que ν -quase todo ponto é recorrente pelo fluxo $(g^t)_t$.

Capítulo 4

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 $f : M \rightarrow M$. 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 4.1. Uma delas é que médias temporais coincidem com médias espaciais: para toda função integrável φ ,

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

Nas Seção 4.2 apresentaremos, por meio de exemplos, diversas técnicas para provar ou desprovar 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. Como veremos na Seção 4.3, as medidas ergódicas são precisamente os elementos extremos desse espaço.

Na Seção 4.4 daremos um breve esboço do desenvolvimento histórico da teoria, no âmbito dos sistemas conservativos. Os principais marcos são a *teoria KAM*, assim denominada em homenagem a Andrey Kolmogorov, Vladimir Arnold e Jürgen Moser, e a *dinâmica hiperbólica*, iniciada por Steven Smale, Dmitry Anosov, Yakov Sinai e seus colaboradores. As duas teorias lidam com tipos

distintos de comportamento dinâmico, elítico e hiperbólico, e obtêm conclusões opostas: em linhas gerais, sistemas hiperbólicos são ergódicos mas sistemas elíticos não são.

4.1 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.

4.1.1 Conjuntos e funções invariantes

Dizemos que uma função mensurável $\varphi : M \rightarrow \mathbb{R}$ 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 χ_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 1.1.4 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 4.1.1. 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. Outro exemplo interessante é o conjunto dos pontos $x = 0, a_0 a_1 \dots$ em $[0, 1]$ tais que a proporção de dígitos a_i com cada valor $k \in \{0, \dots, 9\}$ é prescrita. Ou seja, dado qualquer vetor $p = (p_0, \dots, p_9)$ tal que $p_i \geq 0$ para todo i e $\sum_i p_i = 1$, defina

$$A_p = \left\{ x : \lim_{n \rightarrow \infty} \frac{1}{n} \#\{0 \leq i \leq n-1 : a_i = k\} = p_k \text{ para } k = 0, \dots, 9 \right\}.$$

Para ver que A_p é invariante, observe que se $x = 0, a_0 a_1 \dots$ então todo ponto $y \in f^{-1}(x)$ se escreve na forma $y = 0, b a_0 a_1 \dots$ para algum $b \in \{0, \dots, 9\}$. É claro que o dígito extra b não muda a frequência dos diversos valores $0, \dots, 9$ na expansão decimal. Portanto $y \in A_p$ se, e somente se, $x \in A_p$.

Exemplo 4.1.2. Seja $\varphi : [0, 1] \rightarrow \mathbb{R}$ uma função em $L^1(\mu)$. De acordo com o teorema ergódico de Birkhoff (Teorema 3.2.3), 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 4.1.3. *Seja μ uma probabilidade invariante de uma transformação mensurável $f : M \rightarrow M$. 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 $\varphi : M \rightarrow \mathbb{R}$ tem-se $\tilde{\varphi}(x) = \int \varphi d\mu$ para μ -quase todo ponto.*
- (d) *Para toda função integrável $\varphi : M \rightarrow \mathbb{R}$ a média temporal $\tilde{\varphi} : M \rightarrow \mathbb{R}$ é constante em μ -quase todo ponto.*
- (e) *Para toda função integrável invariante $\psi : M \rightarrow \mathbb{R}$ tem-se $\psi(x) = \int \psi d\mu$ para μ -quase todo ponto.*
- (f) *Toda função integrável invariante $\psi : M \rightarrow \mathbb{R}$ 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 3.2.4). 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. \square

4.1.2 Caracterização espectral

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

Proposição 4.1.4. *Seja μ uma probabilidade invariante de uma transformação mensurável $f : M \rightarrow M$. As seguintes condições são equivalentes:*

(a) (f, μ) é ergódico.

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

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} \mu(f^{-j}(A) \cap B) = \mu(A)\mu(B). \quad (4.1.1)$$

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

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} \int (U_f^j \varphi) \psi d\mu = \int \varphi d\mu \int \psi d\mu. \quad (4.1.2)$$

Demonstração. É claro que (c) implica (b): basta tomar $\varphi = \chi_A$ e $\psi = \chi_B$. 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 3.2.3) temos que

$$\frac{1}{n} \sum_{j=0}^{n-1} U_f^j \varphi \rightarrow \int \varphi d\mu \quad (4.1.3)$$

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

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

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

$$\int \left(\frac{1}{n} \sum_{j=0}^{n-1} U_f^j \varphi \right) \psi d\mu \rightarrow \int \varphi d\mu \int \psi d\mu.$$

Isto prova a afirmação (4.1.2) quando φ é limitada. Falta remover esta última condição. Dado qualquer $\varphi \in L^p(\mu)$ e dado $k \geq 1$, 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

$$\left| \int \left(\frac{1}{n} \sum_{j=0}^{n-1} U_f^j \varphi_k \right) \psi \, d\mu - \int \varphi_k \, d\mu \int \psi \, d\mu \right| < \varepsilon \quad (4.1.4)$$

se n é suficientemente grande (dependendo de k). Em seguida, observe que $\|\varphi_k - \varphi\|_p \rightarrow 0$ quando $k \rightarrow \infty$. Logo, usando a desigualdade de Hölder (Teorema A.5.5), temos que

$$\left| \int (\varphi_k - \varphi) \, d\mu \int \psi \, d\mu \right| \leq \|\varphi_k - \varphi\|_p \left| \int \psi \, d\mu \right| < \varepsilon, \quad (4.1.5)$$

para todo k suficientemente grande. De modo semelhante,

$$\begin{aligned} \left| \int \frac{1}{n} \sum_{j=0}^{n-1} U_f^j (\varphi_k - \varphi) \psi \, d\mu \right| &\leq \frac{1}{n} \sum_{j=0}^{n-1} \left| \int U_f^j (\varphi_k - \varphi) \psi \, d\mu \right| \\ &\leq \frac{1}{n} \sum_{j=0}^{n-1} \|U_f^j (\varphi_k - \varphi)\|_p \|\psi\|_q \, d\mu \\ &= \|\varphi_k - \varphi\|_p \|\psi\|_q < \varepsilon, \end{aligned} \quad (4.1.6)$$

para todo n e todo k suficientemente grande, independente de n . Fixe k tal que (4.1.5) e (4.1.6) sejam válidas e, em seguida, tome n suficientemente grande para que (4.1.4) valha igualmente. Somando as três relações (4.1.4) a (4.1.6), obtemos que

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

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

No caso $p = q = 2$, a condição (4.1.2) 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:

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} [(U_f^j \varphi) - (\varphi \cdot 1)] \cdot \psi = 0 \quad \text{para todo } \varphi, \psi \in L^2(\mu). \quad (4.1.7)$$

Usaremos algumas vezes o seguinte fato elementar: dados quaisquer conjuntos mensuráveis A e B ,

$$\begin{aligned} |\mu(A) - \mu(B)| &= |\mu(A \setminus B) - \mu(B \setminus A)| \\ &\leq |\mu(A \setminus B)| + |\mu(B \setminus A)| = \mu(A \Delta B). \end{aligned} \quad (4.1.8)$$

Corolário 4.1.5. *Suponha que a condição (4.1.1) na Proposição 4.1.4 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. Só precisamos mostrar que a condição (4.1.1) se estende a todo par de borelianos. 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

$$|\mu(f^{-j}(A) \cap B) - \mu(f^{-j}(A_0) \cap B_0)| \leq \mu(A\Delta A_0) + \mu(B\Delta B_0) < 2\varepsilon$$

para todo j e $|\mu(A)\mu(B) - \mu(A_0)\mu(B_0)| \leq \mu(A\Delta A_0) + \mu(B\Delta B_0) < 2\varepsilon$. 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

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

Como ε é arbitrário, isto prova a nossa afirmação. \square

De modo semelhante, basta verificar o item (c) da Proposição 4.1.4 em subconjuntos densos. A prova deste fato fica a cargo do leitor (veja o Exercício 4.1.3):

Corolário 4.1.6. *Suponha que a condição (4.1.2) na Proposição 4.1.4 é satisfeita para todo φ e ψ em subconjuntos densos de $L^p(\mu)$ e $L^q(\mu)$, respectivamente. Então (f, μ) é ergódico.*

4.1.3 Exercícios

4.1.1. Sejam (M, \mathcal{A}) um espaço mensurável e $f : M \rightarrow M$ 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.

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

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

4.1.3. 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 $\mu(f^{-n}(A) \cap B) > 0$;
- (c) a convergência na condição (c) da Proposição 4.1.4 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.

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

4.1.5. Dada uma probabilidade invariante μ , chamamos *bacia* de μ o conjunto $B(\mu)$ dos pontos $x \in M$ tais que

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

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

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

4.1.7. Seja μ uma probabilidade invariante de uma transformação $f : M \rightarrow M$. Mostre que a medida produto $\mu_2 = \mu \times \mu$ é invariante pela transformação $f_2 : M \times M \rightarrow M \times M$ 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?

4.1.8. Seja $f : M \rightarrow M$ 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.

4.2 Exemplos

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

4.2.1 Rotações em toros

Consideremos inicialmente o caso de uma rotação $R_\theta : S^1 \rightarrow S^1$ 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

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

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 4.2.1. *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 4.2.6), é baseada num argumento de ponto de densidade semelhante ao que usaremos na Seção 4.2.2 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

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

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

$$\varphi(x) = \sum_{k \in \mathbb{Z}} a_k e^{2\pi i k x} \quad \text{para quase todo } x \in S^1. \quad (4.2.1)$$

Considere a expansão em série de Fourier (4.2.1) de uma função qualquer $\varphi \in L^2(m)$. Então

$$\varphi(R_\theta(x)) = \sum_{k \in \mathbb{Z}} a_k e^{2\pi i k \theta} e^{2\pi i k x}. \quad (4.2.2)$$

Suponha que φ é invariante. Então (4.2.1) e (4.2.2) 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 $z \in S^1$. Em particular, a função

característica $\varphi = \chi_A$ 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 4.1.3, temos que m é ergódica.

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

Proposição 4.2.2. *Se $\theta = (\theta_1, \dots, \theta_d)$ é racionalmente independente então a rotação $R_\theta : \mathbb{T}^d \rightarrow \mathbb{T}^d$ 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

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

é uma base de Hilbert do espaço $L^2(m)$ das funções $\varphi : \mathbb{T}^d \rightarrow \mathbb{C}$ com quadrado somável. Deixamos esta tarefa ao cuidado do leitor (Exercício 4.2.1).

Corolário 4.2.3. *Se $\theta = (\theta_1, \dots, \theta_d)$ é racionalmente independente então a rotação $R_\theta : \mathbb{T}^d \rightarrow \mathbb{T}^d$ 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, nos toros satisfazem uma propriedade 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). Sistemas unicamente ergódicos serão estudados no Capítulo 6.

4.2.2 Expansão decimal

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

Proposição 4.2.4. *A transformação f é ergódica para a medida de Lebesgue m .*

Demonstração. De acordo com a Proposição 4.1.3, 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

$$\liminf_{\varepsilon \rightarrow 0} \left\{ \frac{m(I \cap A)}{m(I)} : I \text{ intervalo tal que } a \in I \subset B(a, \varepsilon) \right\} = 1. \quad (4.2.3)$$

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

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

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 (4.2.3) implica que

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

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

$$\frac{m(f^k(E_1))}{m(f^k(E_2))} = \frac{m(E_1)}{m(E_2)} \quad (4.2.4)$$

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

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

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

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

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

O Lema 4.2.5 depende do fato de que a transformação f é afim em cada intervalo $((m-1)/10, m/10)$ 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, mas que ainda é suficiente para concluir a demonstração da ergodicidade. Em poucas palavras, no lugar de afirmar que os dois lados de (4.2.4) são iguais, mostra-se, em muitos casos, que a razão entre os dois termos é limitada por alguma constante uniforme. Isso é chamado de *propriedade de distorção limitada*. Como exemplo de aplicação destas ideias, na Seção 4.2.4 provaremos que a transformação de Gauss é ergódica.

Em seguida vamos dar uma aplicação da Proposição 4.2.4 no contexto da Teoria dos Números. Dizemos que um número $x \in \mathbb{R}$ é *balanceado* (ou *normal*) se todo dígito aparece com a mesma frequência, $1/10$, na sua expansão decimal. É fácil dar exemplos de números balanceados ou não-balanceados mas, em geral, é muito difícil decidir se um dado número irracional é balanceado ou não. Por exemplo, não é sabido até hoje se o número π é balanceado.

Por outro lado, a proposição anterior nos permite mostrar, facilmente, que quase todo número é balanceado. Este resultado é conhecido como teorema normal de Borel:

Proposição 4.2.6. *O conjunto dos números $x \in \mathbb{R}$ não balanceados tem medida de Lebesgue nula.*

Demonstração. Como o fato de ser balanceado é independente da parte inteira do número, só precisamos mostrar que quase todo $x \in [0, 1]$ é balanceado. Considere $f : [0, 1] \rightarrow [0, 1]$ definida por $f(x) = 10x - [10x]$. Para cada dígito $j \in \{0, \dots, 9\}$ considere o intervalo $I_j = [j/10, (j+1)/10)$. Recorde que se $x = 0, a_0 a_1 \dots a_k a_{k+1} \dots$ então $f^k(x) = 0, a_k a_{k+1} \dots$ para cada $k \geq 1$. Portanto, $f^k(x) \in I_j$ se, e somente se, o j -ésimo dígito da expansão decimal de x é igual a j . Consequentemente, o tempo médio de visita $\tau(I_j, x)$ é exatamente a frequência do dígito j na expansão decimal de x . Usando o teorema ergódico de Birkhoff, e o fato de que a transformação f é ergódica para a medida de Lebesgue m , concluímos que para cada $j \in \{0, \dots, 9\}$ existe um subconjunto B_j de $[0, 1]$ com $m(B_j) = 1$ tal que

$$\tau(I_j, x) = m(I_j) = \frac{1}{10} \quad \text{para todo } x \in B_j.$$

Então $B = B_0 \cap B_1 \cap \dots \cap B_9$ também tem $m(B) = 1$, e todo número $x \in B$ é balanceado. \square

4.2.3 Deslocamentos de Bernoulli

Seja (X, \mathcal{C}, ν) 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 na Seção 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

$$[m; A_m, \dots, A_n] = \{(x_i)_{i \in \mathbb{N}} : x_i \in A_i \text{ para } m \leq i \leq n\}$$

onde $m \leq n$ e cada A_i é um elemento de \mathcal{C} . Além disso, μ é caracterizada por

$$\mu([m; A_m, \dots, A_n]) = \prod_{i=m}^n \nu(A_i). \quad (4.2.5)$$

Podemos pensar nos elementos de Σ como representando os resultados de seqüências 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 (4.2.5) 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 $\sigma : \Sigma \rightarrow \Sigma$ no espaço Σ , chamada deslocamento (ou “shift”) de Bernoulli, que preserva a medida μ . O principal resultado é que o sistema obtido deste modo 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 seqüências bilaterais $(\dots, x_{-n}, \dots, x_0, \dots, x_n, \dots)$. 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 de Bernoulli é uma aplicação invertível.

O deslocamento de Bernoulli é a dupla (σ, μ) onde $\sigma : \Sigma \rightarrow \Sigma$ é a aplicação definida por

$$\sigma((x_n)_n) = (x_{n+1})_n.$$

Ou seja, σ envia a seqüência $(x_0, x_1, \dots, x_n, \dots)$ na seqüência (x_1, \dots, x_n, \dots) . Observe que a pré-imagem de qualquer cilindro ainda é um cilindro:

$$\sigma^{-1}([m; A_m, \dots, A_n]) = [m+1; A_m, \dots, A_n]. \quad (4.2.6)$$

Segue que σ é mensurável relativamente à σ -álgebra \mathcal{B} . Além disso,

$$\mu(\sigma^{-1}([m; A_m, \dots, A_n])) = \nu(A_m) \cdots \nu(A_n) = \mu([m; A_m, \dots, A_n])$$

e (usando o Lema 1.3.1) isso assegura que a medida μ é invariante por σ .

Proposição 4.2.7. *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 4.2.8. *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: $B = [k; B_k, \dots, B_l]$ e $C = [m; C_m, \dots, C_n]$. Então,

$$\sigma^{-j}(C) = [m + j; C_m, \dots, C_n] \quad \text{para cada } j.$$

Consider qualquer j suficientemente grande para que $m + j > l$. Então,

$$\begin{aligned} B \cap \sigma^{-j}(C) &= \{(x_n)_n : x_k \in B_k, \dots, x_l \in B_l, x_{m+j} \in C_m, \dots, x_{n+j} \in C_n\} \\ &= [k; B_k, \dots, B_l, X, \dots, X, C_m, \dots, C_n], \end{aligned}$$

onde X aparece exatamente $m + j - l - 1$ vezes. Pela definição (4.2.5), 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 $B = C = A$. Concluimos 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

$$\mu(B \cap \sigma^{-j}(B)) = \mu(B)\mu(\sigma^{-m}(B)) = \mu(B)^2. \quad (4.2.7)$$

Observe que a diferença simétrica $(A \cap \sigma^{-j}(A)) \Delta (B \cap \sigma^{-j}(B))$ está contida em

$$(A \Delta B) \cup (\sigma^{-j}(A) \Delta \sigma^{-j}(B)) = (A \Delta B) \cup \sigma^{-j}(A \Delta B).$$

Isto, juntamente com o fato de que μ é invariante por f , implica que

$$|\mu(A \cap \sigma^{-j}(A)) - \mu(B \cap \sigma^{-j}(B))| \leq 2\mu(A \Delta B) < 2\varepsilon \quad (4.2.8)$$

(lembre da relação (4.1.8)). Além disso,

$$|\mu(A)^2 - \mu(B)^2| \leq 2|\mu(A) - \mu(B)| < 2\varepsilon. \quad (4.2.9)$$

Juntando as relações (4.2.7), (4.2.8), (4.2.9), concluimos 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 $\mu(A) = 1$. \square

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 $[m; A_m, \dots, A_n]$ onde os conjuntos A_m, \dots, A_n são abertos de X . A propriedade (4.2.6) implica que o deslocamento $\sigma : \Sigma \rightarrow \Sigma$ 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 $f : M \rightarrow M$ é *transitiva* se existe $x \in M$ cuja trajetória $f^n(x)$, $n \geq 1$ é densa em Σ . Deixamos a demonstração do próximo resultado a cargo do leitor (Exercício 4.2.2):

Proposição 4.2.9. *Seja X um conjunto finito e $\Sigma = X^{\mathbb{N}}$ ou $\Sigma = X^{\mathbb{Z}}$. Então o deslocamento $\sigma : \Sigma \rightarrow \Sigma$ é uma aplicação contínua e transitiva. Além disso, o conjunto dos pontos periódicos de σ é denso em Σ .*

A seguinte afirmação, que é uma das variantes do *paradoxo do macaco*, ilustra o significado da ergodicidade da medida μ : *Se colocarmos um macaco para digitar texto durante um tempo infinito então, com probabilidade total, ele acabará digitando “Os Lusíadas”¹ e, de fato, o fará infinitas vezes.*

Para “demonstrar” esta afirmação precisamos formular a situação de modo um pouco mais preciso. Consideramos que o macaco se encontra perante o teclado, apertando uma tecla após outra, ao acaso. O texto digitado é, portanto, uma sequência $(x_n)_{n \in \mathbb{N}}$ onde x_n pertence ao conjunto (finito) X dos caracteres no teclado: letras, espaço, hífen, sinais de pontuação, etc. Supomos que cada caracter i no teclado tem uma probabilidade positiva p_i de ser digitado, a cada vez. Isto corresponde a uma medida de probabilidade

$$\nu = \sum_{i \in X} p_i \delta_i$$

no conjunto X dos caracteres. Também supomos que a escolha de cada tecla é independente das teclas digitadas anteriormente (o macaco não sabe o que está escrevendo, ele apenas aperta teclas ao acaso). Isto quer dizer que as possíveis sequências $(x_n)_n$ são regidas pela probabilidade de Bernoulli $\mu = \nu^{\mathbb{N}}$. Denotamos por $\sigma : \Sigma \rightarrow \Sigma$ a aplicação deslocamento no espaço $\Sigma = X^{\mathbb{N}}$.

O texto de “Os Lusíadas” corresponde a uma certa sequência finita (mas longa) de caracteres (l_0, \dots, l_N) . Consideremos o cilindro $L = [0; l_0, \dots, l_N]$. Então

$$\mu(L) = \prod_{j=1}^N p_{l_j} > 0.$$

Uma sequência $(x_n)_n$ contém o texto de “Os Lusíadas”, começando no k -ésimo caracter, precisamente se $\sigma^k((x_n)_n) \in L$. Pelo teorema ergódico de Birkhoff

¹Poema épico monumental, em 10 cantos, de autoria do poeta português Luis de Camões, falecido em Lisboa em 1580.

e pela ergodicidade de (σ, μ) , com probabilidade total isso acontece para um conjunto K de valores de k satisfazendo

$$\lim_n \frac{1}{n} \#(K \cap [0, n-1]) = \mu(L) > 0. \quad (4.2.10)$$

Em particular, o conjunto K é infinito. Ou seja, a sequência digitada contém infinitas cópias de “Os Lusíadas”, conforme afirmamos. Na verdade, (4.2.10) prova ainda mais: sempre com probabilidade total, as cópias do nosso poema ocupam uma fração positiva de todos os caracteres digitados. Em outras palavras, em média, o macaco digita uma nova cópia de “Os Lusíadas” a cada tantos anos.

4.2.4 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:

$$\mu(E) = \frac{1}{\log 2} \int_E \frac{dx}{1+x}. \quad (4.2.11)$$

Proposição 4.2.10. *O sistema (G, μ) é ergódico.*

Este fato pode ser demonstrado por uma versão mais elaborada do método que usamos na Seção 4.2.2. 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

$$I(1, m) = \left(\frac{1}{m+1}, \frac{1}{m} \right),$$

para cada $m \geq 1$. Em seguida defina, por recorrência,

$$I(k, m_1, \dots, m_k) = I(1, m_1) \cap G^{-k+1}(I(k-1, m_2, \dots, m_k))$$

para $m_1, \dots, m_k \geq 1$. Então, basta tomar para I_k o intervalo $I(k, m_1, \dots, m_k)$ 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 (4.2.4) 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 4.2.11 (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 4.2.12. *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

$$G'(x) = -\frac{1}{x^2} \quad e \quad G''(x) = \frac{2}{x^3}.$$

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 $x \in (0, 1]$. \square

Lema 4.2.13. *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*

$$\frac{|(G^k)'(x)|}{|(G^k)'(y)|} \leq C \quad \text{para quaisquer } x \text{ e } y \text{ em } I_k.$$

Consequentemente, $\text{diam } I_k \leq C$.

Demonstração. Seja g um 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

$$[\log |G' \circ g(z)|]' = \frac{G''(g(z))g'(z)}{G'(g(z))} = \frac{G''(g(z))}{G'(g(z))^2}.$$

Portanto, a última estimativa no Lema 4.2.12 implica que

$$|[\log |G' \circ g(z)|]'| \leq 2 \quad \text{para todo } g \text{ e todo } z. \quad (4.2.12)$$

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 (4.2.12), obtemos que

$$\log \frac{|(G^k)'(x)|}{|(G^k)'(y)|} \leq 2 \sum_{j=1}^k |G^j(x) - G^j(y)| = 2 \sum_{i=0}^{k-1} |G^{k-i}(x) - G^{k-i}(y)|. \quad (4.2.13)$$

Agora, as duas primeiras estimativas no Lema 4.2.12 implicam que

$$|G^k(x) - G^k(y)| \geq 2^{[i/2]} |G^{k-i}(x) - G^{k-i}(y)|$$

para todo $i = 0, \dots, k$. Substituindo em (4.2.13), concluimos que

$$\log \frac{|(G^k)'(x)|}{|(G^k)'(y)|} \leq 2 \sum_{i=0}^{k-1} 2^{-[i/2]} |G^k(x) - G^k(y)| \leq 8 |G^k(x) - G^k(y)| \leq 8.$$

Agora basta tomar $C = e^8$. \square

Demonstração da Proposição 4.2.11. Seja m a medida de Lebesgue em $[0, 1]$. O Lema 4.2.13 implica que

$$\frac{m(G^k(E_1))}{m(G^k(E_2))} = \frac{\int_{E_1} |(G^k)'| dm}{\int_{E_2} |(G^k)'| dm} \leq C \frac{m(E_1)}{m(E_2)}.$$

Por outro lado, a definição (4.2.11) implica que

$$\frac{1}{2 \log 2} m(E) \leq \mu(E) \leq \frac{1}{\log 2} m(E),$$

para todo conjunto mensurável $E \subset [0, 1]$. Combinando estas duas relações, obtemos que

$$\frac{\mu(G^k(E_1))}{\mu(G^k(E_2))} \leq 2 \frac{m(G^k(E_1))}{m(G^k(E_2))} \leq 2C \frac{m(E_1)}{m(E_2)} \leq 4C \frac{\mu(E_1)}{\mu(E_2)}.$$

Assim, basta tomar $K = 4C$. \square

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 $I(k, m_1, \dots, m_k)$ que contêm a . Segue do Lema 4.2.12 que

$$\text{diam } I_k \leq \sup \left\{ \frac{1}{|(G^k)'(x)|} : x \in I_k \right\} \leq 2^{-[k/2]}$$

para todo $k \geq 1$. Em particular, o diâmetro de I_k converge para zero e, portanto,

$$\frac{\mu(I_k \cap A)}{\mu(I_k)} \rightarrow 1 \quad \text{quando } k \rightarrow \infty. \quad (4.2.14)$$

Tomemos $E_1 = I_k \cap A^c$ e $E_2 = I_k$. Pela Proposição 4.2.11,

$$\frac{\mu(G^k(I_k \cap A^c))}{\mu(G^k(I_k))} \leq K \frac{\mu(I_k \cap A^c)}{\mu(I_k)}.$$

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

$$\mu(A^c) \leq K \frac{\mu(I_k \cap A^c)}{\mu(I_k)}.$$

De acordo com (4.2.14), a expressão do lado direito converge para zero quando $k \rightarrow \infty$. Logo $\mu(A^c) = 0$, como queríamos demonstrar.

4.2.5 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 $x \sim y \Leftrightarrow x - y \in \mathbb{Z}^d$. 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 euclidiano \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

$$f_A : \mathbb{T}^d \rightarrow \mathbb{T}^d, \quad f_A([x]) = [A(x)]$$

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 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 $|\det A|$ (= grau de 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 $m(D) = m(f^{-1}(D))$ para todo domínio D suficientemente pequeno. Logo f preserva a medida m . Agora vamos provar o seguinte fato:

Teorema 4.2.14. *O sistema (f_A, m) é ergódico se, e somente se, nenhum autovalor da matriz A é raiz da unidade.*

Demonstração. 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

$$\sum_{k \in \mathbb{Z}^d} |c_k|^2 = \|\varphi\|_2^2 < \infty. \quad (4.2.15)$$

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 μ -quase todo ponto. Então, por unicidade da expansão de Fourier, devemos ter

$$c_{A^*(k)} = c_k \quad \text{para todo } k \in \mathbb{Z}. \quad (4.2.16)$$

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 $l, m \in \mathbb{Z}$ com $m > 0$ tais que $A^{(l+m)^*}(k) = A^{l^*}(k)$. Isto só poderia acontecer se A^* tivesse algum autovalor λ tal que $\lambda^m = 1$. 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 (4.2.16) juntamente com (4.2.15) 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 $m \geq 1$ tal que 1 é autovalor de A^{m^*} . Como A^{m^*} tem coeficientes inteiros, segue (veja o Exercício 4.2.8) que existe algum $k \in \mathbb{Z}^d \setminus \{0\}$ tal que $A^{m^*}(k) = k$. Fixe k e considere a função $\varphi \in L^2(m)$ definida por

$$\varphi([x]) = \sum_{i=0}^{m-1} e^{2\pi i(x \cdot A^{i^*}(k))} = \sum_{i=0}^{m-1} e^{2\pi i(A^i(x) \cdot k)}$$

Então φ é uma função invariante por f_A mas não é constante. Logo, f_A não é ergódica. \square

4.2.6 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. Algumas dessas extensões serão mencionadas na Seção 4.4.

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 $A|E^s$ têm módulo menor que 1;
2. $A(E^u) = E^u$ e todos os autovalores de $A|E^u$ têm módulo maior que 1.

Então existem constantes $C > 0$ e $\lambda < 1$ tais que

$$\begin{aligned} \|A^n(v^s)\| &\leq C\lambda^n\|v^s\| && \text{para todo } v^s \in E^s \text{ e todo } n \geq 0, \\ \|A^{-n}(v^u)\| &\leq C\lambda^n\|v^u\| && \text{para todo } v^u \in E^u \text{ e todo } n \geq 0. \end{aligned} \quad (4.2.17)$$

Exemplo 4.2.15. 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:

$$E^u = \{(x, y) \in \mathbb{R}^2 : y = \frac{\sqrt{5}-1}{2}x\} \quad e \quad E^s = \{(x, y) \in \mathbb{R}^2 : y = -\frac{\sqrt{5}+1}{2}x\}.$$

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 (4.2.17), 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.

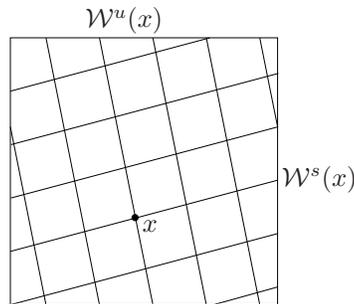


Figura 4.1: Folheação estável e folheação instável no toro

Projetando \mathcal{F}^s e \mathcal{F}^u pela projeção canônica $\pi : \mathbb{R}^d \rightarrow \mathbb{T}^d$ obtemos folheações \mathcal{W}^s e \mathcal{W}^u do toro que chamamos *folheação estável* e *folheação instável* da transformação f_A . Veja a Figura 4.1. As observações anteriores mostram que estas folheações são invariantes por f_A . Além disso:

- (a) dados dois pontos quaisquer x e y na mesma folha estável, tem-se que $d(f_A^j(x), f_A^j(y)) \rightarrow 0$ quando $j \rightarrow +\infty$;
- (b) dados dois pontos quaisquer y e z na mesma folha instável, tem-se que $d(f_A^j(y), f_A^j(z)) \rightarrow 0$ quando $j \rightarrow -\infty$.

Vamos usar esta informação geométrica para provar que (f_A, m) é ergódica. Para isso, considere qualquer função contínua $\varphi : \mathbb{T}^d \rightarrow \mathbb{R}$ 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 3.2.8, existe um conjunto $X \subset \mathbb{T}^d$ com medida total tal que

$$\varphi^+(x) = \varphi^-(x) \quad \text{para todo } x \in X. \quad (4.2.18)$$

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 4.2.16. *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 $j \rightarrow \infty$. Como φ é contínua (logo uniformemente contínua, uma vez que o domínio é compacto) isso implica que

$$\varphi(f_A^j(x)) - \varphi(f_A^j(y)) \rightarrow 0 \quad \text{quando } j \rightarrow \infty.$$

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. \square

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 . Veja a Figura 4.2.

Lema 4.2.17. *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 $y' \in \mathcal{W}^u(x')$.*

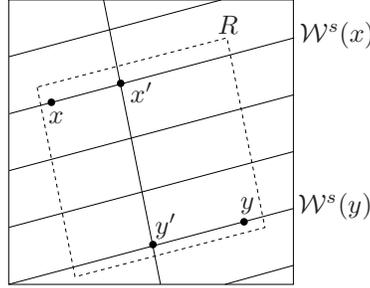


Figura 4.2: Retângulo em \mathbb{T}^d

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,

$$m_x^s(\mathcal{W}^s(x, R) \setminus X) = 0 \quad \text{para } m\text{-quase todo } x \in R.$$

Defina $Y_R = \{x \in X \cap R : m_x^s(\mathcal{W}^s(x, R) \setminus X) = 0\}$. Então Y_R tem medida total em R . Dados $x, y \in R$ considere a aplicação

$$\pi : \mathcal{W}^s(x, R) \rightarrow \mathcal{W}^s(y, R), \quad \pi(x') = \text{interseção entre } \mathcal{W}^u(x', R) \text{ e } \mathcal{W}^s(y, R).$$

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, conseqüentemente, 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 4.2.17. Usando também o Lema 4.2.16, 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, \dots, R_N uma cobertura finita do toro por retângulos. Considere o conjunto

$$Y = \bigcup_{j=1}^N Y_j, \quad \text{onde } Y_j = Y_{R_j}.$$

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 X_j é um subconjunto de medida total, obtemos que cada X_{j_i} intersecta $X_{j_{i-1}}$. Então, $\varphi^+ = \varphi^-$ é constante na união de todos os X_{j_i} . Isto prova a nossa afirmação.

Desta forma, mostramos que as médias temporais φ^\pm de qualquer função contínua φ são constantes em m -quase todo ponto. Consequentemente (veja o Exercício 4.1.4), o sistema (f_A, m) é ergódico.

4.2.7 Exercícios

4.2.1. Prove a Proposição 4.2.2.

4.2.2. Prove a Proposição 4.2.9.

4.2.3. Seja $I = [0, 1]$ e $f : I \rightarrow I$ a função definida por

$$f(x) = \begin{cases} 2x & \text{se } 0 \leq x < 1/3 \\ 2x - 2/3 & \text{se } 1/3 \leq x < 1/2 \\ 2x - 1/3 & \text{se } 1/2 \leq x < 2/3 \\ 2x - 1 & \text{se } 2/3 \leq x \leq 1. \end{cases}$$

Mostre que f é ergódica relativamente à medida de Lebesgue m .

4.2.4. Seja $\Sigma = \{1, 2, \dots, k\}^{\mathbb{N}}$. Prove que todo subconjunto X infinito, compacto, invariante pelo deslocamento $\sigma : \Sigma \rightarrow \Sigma$ contém algum ponto não-periódico.

4.2.5. 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 $\mathcal{B} = \mathcal{C}^{\mathbb{Z}}$.

4.2.6. Neste exercício propomos outra demonstração para a Proposição 4.2.1. 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 $\mu(A) = 1$.

4.2.7. Suponha que θ é irracional. Seja $\varphi : S^1 \rightarrow \mathbb{R}$ uma função contínua qualquer. Mostre que

$$\tilde{\varphi}(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(R_\theta^j(x)) \quad (4.2.19)$$

existe em *todo* ponto e, de fato, o limite é uniforme. Justifique o fato que $\tilde{\varphi}$ é constante em todo ponto. Deduza que R_θ tem uma única probabilidade invariante.

4.2.8. 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 $Ak = \lambda k$.

4.2.9. 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.

4.2.10. Seja $f : M \rightarrow M$ uma transformação mensurável, seja μ uma medida invariante ergódica e seja $\varphi : M \rightarrow \mathbb{R}$ 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 $x \in M$.

4.2.11. Observe que o número b no Exercício 3.2.4 é independente de x num conjunto com medida de Lebesgue total. Prove que a média aritmética dos números a_1, \dots, a_n, \dots vai para infinito: $\lim_n (1/n)(a_1 + \dots + a_n) = +\infty$.

4.3 Propriedades das medidas ergódicas

Nesta seção consideramos que a transformação $f : M \rightarrow M$ está fixada e analisamos o espaço $\mathcal{M}_1(f)$ das probabilidades invariantes por f , especialmente o subconjunto $\mathcal{M}_e(f)$ das probabilidades ergódicas.

Lembre que uma medida ν diz-se *absolutamente contínua* com relação a outra medida μ se $\mu(E) = 0$ implica $\nu(E) = 0$. Nesse caso escrevemos $\nu \ll \mu$. Esta relação é transitiva: se $\nu \ll \mu$ e $\mu \ll \lambda$ então $\nu \ll \lambda$. O primeiro resultado afirma que probabilidades ergódicas são minimais para esta relação de ordem:

Lema 4.3.1. *Se μ e ν são probabilidades invariantes tais que μ é ergódica e ν é absolutamente contínua com relação a μ , então $\mu = \nu$.*

Demonstração. Seja $\varphi : M \rightarrow \mathbb{R}$ uma função mensurável limitada qualquer. Como μ é invariante e ergódica, a média temporal

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

é constante: $\tilde{\varphi}(x) = \int \varphi d\mu$ em μ -quase todo ponto. Segue que esta igualdade também vale em ν -quase todo ponto, já que $\nu \ll \mu$. Em particular,

$$\int \varphi d\nu = \int \tilde{\varphi} d\nu = \int \varphi d\mu$$

(a primeira igualdade é parte do teorema ergódico de Birkhoff). Portanto, as integrais de φ com relação a μ e em relação a ν coincidem, qualquer que seja a função mensurável limitada φ . Em particular, considerando funções características, concluímos que $\mu = \nu$. \square

É claro que se μ_1 e μ_2 são probabilidades invariantes com respeito à transformação f então $(1-t)\mu_1 + t\mu_2$ também é probabilidade invariante, qualquer que seja $t \in (0, 1)$. Isto significa que o espaço $\mathcal{M}_1(f)$ das probabilidades invariantes é *convexo*. A proposição que apresentamos a seguir afirma que as medidas ergódicas são os elementos extremos deste convexo:

Proposição 4.3.2. *Uma probabilidade invariante μ é ergódica se, e somente se, não é possível escrevê-la na forma $\mu = (1-t)\mu_1 + t\mu_2$ com $t \in (0, 1)$ e μ_1, μ_2 probabilidades invariantes distintas.*

Demonstração. Para provar a parte “se”, suponha que μ não é ergódica. Então existe algum conjunto invariante A com $0 < \mu(A) < 1$. Defina μ_1 e μ_2 como sendo as restrições normalizadas de μ a A e ao seu complementar, respectivamente:

$$\mu_1(E) = \frac{\mu(E \cap A)}{\mu(A)} \quad \text{e} \quad \mu_2(E) = \frac{\mu(E \cap A^c)}{\mu(A^c)}.$$

Como A e A^c são conjuntos invariantes e μ é medida invariante, μ_1 e μ_2 são também probabilidades invariantes. Além disso,

$$\mu = \mu(A)\mu_1 + \mu(A^c)\mu_2$$

e portanto μ não é extremal. Para provar a recíproca, suponha que μ é ergódica e temos $\mu = (1-t)\mu_1 + t\mu_2$ com $t \in (0, 1)$. É claro que $\mu(E) = 0$ implica $\mu_1(E) = \mu_2(E) = 0$, ou seja, μ_1 e μ_2 são absolutamente contínuas com relação a μ . Logo, pelo Lema 4.3.1, $\mu_1 = \mu = \mu_2$. Isto prova que μ é extremal. \square

Também observamos que medidas ergódicas distintas “vivem” em subconjuntos disjuntos do espaço M (veja também o Exercício 4.3.6):

Lema 4.3.3. *Suponha que a σ -álgebra de M admite um gerador enumerável. Seja $\{\mu_i : i \in \mathcal{I}\}$ uma família qualquer de probabilidades invariantes e ergódicas, todas distintas. Então as medidas μ_i são mutuamente singulares: existem subconjuntos mensuráveis invariantes $\{P_i : i \in \mathcal{I}\}$ disjuntos dois-a-dois e tais que $\mu_i(P_i) = 1$ para todo $i \in \mathcal{I}$.*

Demonstração. Seja Γ um gerador enumerável da σ -álgebra de M e seja \mathcal{A} a álgebra gerada por Γ . Note que \mathcal{A} é enumerável, já que ela coincide com a união das álgebras geradas pelos subconjuntos finitos de Γ . Para cada $i \in \mathcal{I}$, defina

$$P_i = \{x \in M : \tau(A, x) = \mu_i(A) \text{ para todo } A \in \mathcal{A}\}.$$

Por ergodicidade, $\mu_i(P_i) = 1$ para todo $i \in \mathcal{I}$. Além disso, se existe $x \in P_i \cap P_j$ então $\mu_i(A) = \tau(A, x) = \mu_j(A)$ para todo $A \in \mathcal{A}$; em outras palavras, $\mu_i = \mu_j$. Isto prova que os P_i são disjuntos dois-a-dois. \square

Agora suponha que $f : M \rightarrow M$ é uma aplicação contínua num espaço topológico. Dizemos que a aplicação f é *transitiva* se existe algum $x \in M$ tal que $\{f^n(x) : n \in \mathbb{N}\}$ é denso em M .

O lema a seguir dá uma caracterização útil da transitividade. Lembre que um espaço topológico M é chamado de *espaço de Baire* se a interseção de qualquer família enumerável de abertos densos é densa em M . Todo espaço métrico completo é um espaço de Baire e o mesmo vale para todo espaço topológico localmente compacto (veja [Dug66]).

Lema 4.3.4. *Suponha que M é um espaço de Baire com base enumerável de abertos. Então $f : M \rightarrow M$ é transitiva se, e somente se, para todo par de abertos U e V existe $k \geq 1$ tal que $f^{-k}(U)$ intersecta V .*

Demonstração. Suponha que f é transitiva e seja $x \in M$ um ponto cuja órbita $\{f^n(x) : n \in \mathbb{N}\}$ é densa em M . Então existe $m \geq 1$ tal que $f^m(x) \in V$ e (usando que $\{f^n(x) : n > m\}$ também é denso) existe $n > m$ tal que $f^n(x) \in U$. Tome $k = n - m$. Então $f^m(x) \in f^{-k}(U) \cap V$. Isto prova a parte ‘somente se’ do enunciado.

Para provar a recíproca, seja $\{U_j : j \in \mathbb{N}\}$ uma base enumerável de abertos de M . A hipótese garante que o aberto $\cup_{k=1}^{\infty} f^{-k}(U_j)$ é denso em M para todo $j \in \mathbb{N}$. Então a interseção

$$X = \bigcap_{j=1}^{\infty} \bigcup_{k=1}^{\infty} f^{-k}(U_j)$$

é um subconjunto denso de M . Em particular, ele é não vazio. Por outro lado, por definição, se $x \in X$ então para todo $j \in \mathbb{N}$ existe algum $k \geq 1$ tal que $f^k(x) \in U_j$. Como os U_j constituem uma base de vizinhanças de M , isto significa que $\{f^k(x) : k \in \mathbb{N}\}$ é denso em M . \square

Proposição 4.3.5. *Suponha que M é um espaço de Baire com base enumerável de abertos. Então a restrição de f ao suporte de μ é transitiva.*

Demonstração. Comece por notar que $\text{supp } \mu$ tem base enumerável de abertos, por ser um subespaço de M , e é um espaço de Baire, uma vez que é fechado em M . Sejam U e V abertos de $\text{supp } \mu$. Pela definição do suporte, $\mu(U) > 0$ e $\mu(V) > 0$. Defina $B = \cup_{k=1}^{\infty} f^{-k}(U)$. Então $\mu(B) > 0$, porque $B \supset U$, e $f^{-1}(B) \subset B$. Por ergodicidade (veja o Exercício 1.1.4) segue que $\mu(B) = 1$. Então B deve intersectar V . Isto prova que existe $k \geq 1$ tal que $f^{-k}(U)$ intersecta V . Pelo Lema 4.3.4, segue que a transformação $f : \text{supp } \mu \rightarrow \text{supp } \mu$ é transitiva. \square

4.3.1 Exercícios

4.3.1. Seja $f : M \rightarrow M$ uma aplicação mensurável num espaço topológico M com base enumerável de abertos e seja μ uma medida de probabilidade ergódica para f . Mostre que a órbita $\{f^n(x) : n \geq 0\}$ de μ -quase todo ponto $x \in M$ é densa no suporte de μ .

4.3.2. Seja $f : M \rightarrow M$ uma transformação contínua em um espaço métrico compacto. Fixada uma função $\varphi : M \rightarrow \mathbb{R}$, prove que existe uma probabilidade invariante μ_φ tal que

$$\int \varphi d\mu_\varphi = \sup_{\eta \in \mathcal{M}_1(f)} \int \varphi d\eta.$$

4.3.3. Seja $g : E \rightarrow E$ uma transformação induzida por $f : M \rightarrow M$, ou seja uma transformação da forma $g(x) = f^{\rho(x)}(x)$ com $\rho : E \rightarrow \mathbb{N}$ (veja a Seção 1.4.2). Seja ν uma probabilidade invariante de g e seja ν_g a medida invariante de f definida por (1.4.5). Suponha que $\nu_\rho(M) < \infty$ e denote $\mu = \nu_\rho/\nu_\rho(M)$. Mostre que (f, μ) é ergódico se, e somente se, (g, ν) é ergódico.

4.3.4. Seja $f : M \rightarrow M$ uma transformação contínua num espaço métrico completo separável. Dada uma probabilidade invariante μ , seja $\hat{\mu}$ o seu levantamento para a extensão natural $\hat{f} : \hat{M} \rightarrow \hat{M}$ (veja a Seção 2.4.2). Mostre que $(\hat{f}, \hat{\mu})$ é ergódico se, e somente se, (f, μ) é ergódico.

4.3.5. Seja $f : M \rightarrow M$ uma transformação e seja μ uma medida invariante. Seja $g^t : N \rightarrow N$, $t \in \mathbb{R}$ um fluxo suspensão de f e seja ν a suspensão correspondente da medida μ (veja a Seção 3.4.1). Suponha que $\nu(N) < \infty$ e denote $\hat{\nu} = \nu/\nu(N)$. Mostre que $\hat{\nu}$ é ergódica para o fluxo $(g^t)_t$ se, e somente se, μ é ergódica para f .

4.3.6. Mostre que no caso de famílias finitas ou enumeráveis de medidas ergódicas a conclusão do Lema 4.3.3 vale mesmo que a σ -álgebra não seja enumeravelmente gerada.

4.3.7. Dê exemplo de um espaço métrico M e uma transformação $f : M \rightarrow M$, tais que existe uma sequência de medidas borelianas μ_n ergódicas para f que convergem, na topologia fraca*, para uma medida μ que não é ergódica.

4.3.8. Seja M um espaço métrico, $f : M \rightarrow M$ uma transformação contínua e μ uma probabilidade invariante e ergódica. Mostre que, dada qualquer probabilidade ν em M absolutamente contínua com respeito a μ , mas não necessariamente invariante,

$$\frac{1}{n} \sum_{j=0}^{n-1} f_*^j \nu \text{ converge para } \mu \text{ na topologia fraca*}.$$

4.3.9. Considere $X = \{1, \dots, d\}$ e seja $\sigma : \Sigma \rightarrow \Sigma$ o deslocamento em $\Sigma = X^{\mathbb{N}}$ ou $\Sigma = X^{\mathbb{Z}}$.

- (1) Mostre que para todo $\delta > 0$ existe $k \geq 1$ tal que, dados $x^1, \dots, x^s \in \Sigma$ e $m_1, \dots, m_s \geq 1$, existe um ponto periódico $y \in \Sigma$ com período n_s e satisfazendo $d(f^{j+n_i}(y), f^j(x^i)) < \delta$ para todo $0 \leq j < m_i$, onde $n_1 = 0$ e $n_i = (m_1 + k) + \dots + (m_{i-1} + k)$ para $1 < i \leq s$.
- (2) Sejam $\varphi : \Sigma \rightarrow \mathbb{R}$ uma função contínua e $\tilde{\varphi}$ a sua média de Birkhoff. Mostre, que dados $\varepsilon > 0$, $x^1, \dots, x^s \in \Sigma$ tais que a média de Birkhoff está definida e $\alpha^1, \dots, \alpha^s > 0$ tais que $\sum_i \alpha^i = 1$, existe um ponto periódico $y \in \Sigma$ satisfazendo $|\tilde{\varphi}(y) - \sum_i \alpha^i \tilde{\varphi}(x^i)| < \varepsilon$.
- (3) Conclua que o conjunto $\mathcal{M}_e(\sigma)$ das medidas invariantes ergódicas é denso no espaço de todas as probabilidades invariantes $\mathcal{M}_1(\sigma)$.

4.4 Comentários sobre sistemas conservativos

O teorema ergódico de Birkhoff, provado nos anos trinta do século 20, deu sólida fundamentação matemática para o enunciado da hipótese ergódica de Boltzmann, mas deixou totalmente em aberto a questão da sua *veracidade*. Nesta seção vamos dar um breve panorama dos principais resultados obtidos desde então nesta direção, no contexto de sistemas *conservativos*, isto é, sistemas dinâmicos que preservam uma medida de volume numa variedade.

Começamos por observar que, num certo sentido topológico abstrato, a maioria dos sistemas conservativos são ergódicos. Esse é o sentido do teorema que vamos enunciar a seguir, provado no início dos anos 1940 por John Oxtoby and Stanislaw Ulam [OU41]. Lembre que um subconjunto de um espaço de Baire é chamado *residual* se ele pode ser escrito como uma interseção enumerável de subconjuntos abertos e densos. Por definição de espaço de Baire, todo subconjunto residual é denso.

Teorema 4.4.1 (Oxtoby, Ulam). *Para toda variedade riemanniana compacta M , existe um subconjunto residual \mathcal{R} do espaço $\text{Homeo}_{\text{vol}}(M)$ dos homeomorfismos conservativos de M tal que todo elemento de \mathcal{R} é ergódico.*

Os resultados que apresentaremos a seguir implicam que a conclusão deste teorema torna-se falsa quando substituímos $\text{Homeo}(M)$ pelo espaço $\text{Difeo}_{\text{vol}}^k(M)$ dos difeomorfismos conservativos de classe C^k , pelo menos para $k > 3$. Praticamente nada é sabido a este respeito nos casos $k = 2$ e $k = 3$. Por outro lado, Artur Avila, Sylvain Crovisier e Amie Wilkinson anunciaram recentemente uma versão C^1 do teorema anterior: *para toda variedade riemanniana compacta M , existe um subconjunto residual \mathcal{R} do espaço $\text{Difeo}_{\text{vol}}^1(M)$ dos difeomorfismos conservativos de classe C^1 tal que todo $f \in \mathcal{R}$ com entropia $h_{\text{vol}}(f)$ positiva é ergódico.* A noção de entropia será estudada no Capítulo 9.

4.4.1 Sistemas hamiltonianos

Os sistemas em que Boltzmann estava interessado, relativos ao movimento das moléculas de gases podem, em princípio, ser descritos pelas leis da mecânica

clássica newtoniana. No chamado formalismo hamiltoniano da mecânica clássica, os estados do sistema são representados por meio das “coordenadas generalizadas” q_1, \dots, q_d e dos “momentos generalizados” p_1, \dots, p_d e a sua evolução é descrita pelas soluções das equações de Hamilton-Jacobi:

$$\frac{dq_j}{dt} = \frac{\partial H}{\partial p_j} \quad \text{e} \quad \frac{dp_j}{dt} = -\frac{\partial H}{\partial q_j}, \quad j = 1, \dots, d, \quad (4.4.1)$$

onde H (a energia total do sistema) é uma função de classe C^2 das variáveis $q = (q_1, \dots, q_d)$ e $p = (p_1, \dots, p_d)$; o inteiro $d \geq 1$ é o número de graus de liberdade do sistema.

Exemplo 4.4.2 (Pêndulo). Tome $d = 1$ e $H(q, p) = p^2/2 - g \cos q$, onde g é uma constante positiva e $(q, p) \in \mathbb{R}^2$. As equações de Hamilton-Jacobi

$$\frac{dq}{dt} = p \quad \text{e} \quad \frac{dp}{dt} = -g \sin q$$

descrevem o movimento de um pêndulo sujeito a um campo gravitacional constante: a coordenada q mede o ângulo relativo à posição de equilíbrio (estável) e p mede o momento angular. O *hamiltoniano* H corresponde à energia total do sistema: $H =$ energia cinética + energia potencial.

Note que H sempre é uma *integral primeira* do sistema, ou seja, ela é constante ao longo das trajetórias do fluxo:

$$\frac{dH}{dt} = \sum_{j=1}^d \frac{\partial}{\partial q_j} \frac{dq_j}{dt} + \frac{\partial}{\partial p_j} \frac{dp_j}{dt} \equiv 0.$$

Então, podemos considerar a restrição do fluxo a cada *hipersuperfície de energia* $H_c = \{(q, p) : H(q, p) = c\}$. A medida de volume $dq_1 \cdots dq_d dp_1 \cdots dp_d$ é chamada *medida de Liouville*. Observando que o campo de vetores

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

tem divergente nulo (lembre a Seção 1.3.6), concluímos que o fluxo preserva a medida de Liouville. Em consequência (veja o Exercício 1.3.12), a restrição do fluxo a cada hipersuperfície de energia H_c também tem uma medida invariante μ_c , que é dada por

$$\mu_c(E) = \int_E \frac{ds}{\|\text{grad } H\|} \quad \text{para todo conjunto mensurável } E \subset H_c$$

onde ds representa o elemento de volume na hipersuperfície. Então é natural perguntar se, em geral, sistemas hamiltonianos são ergódicos relativamente à medida invariante em (quase) toda hipersuperfície de energia.

O primeiro resultado importante nesta direção foi anunciado por Andrey Kolmogorov em 1954 e foi, logo em seguida, substanciado pelos trabalhos de

Vladimir Arnold e Jürgen Moser. Isto conduziu a uma teoria muito profunda, que é conhecida como Teoria KAM em homenagem aos seus fundadores, e para qual contribuíram de maneira decisiva diversos outros matemáticos, tais como Helmut Rüssmann, Michael Herman, Eduard Zehnder, Jean-Christophe Yoccoz, Jürgen Pöschel, entre outros. Inicialmente, a teoria lida com sistemas “quase integráveis”.

Um sistema hamiltoniano diz-se *integrável* (no sentido de Liouville) se ele admite d integrais primeiras I_1, \dots, I_d :

- *independentes*: ou seja, tais que os gradientes

$$\text{grad } I_j = \left(\frac{\partial I_j}{\partial q_1}, \frac{\partial I_j}{\partial p_1}, \dots, \frac{\partial I_j}{\partial q_d}, \frac{\partial I_j}{\partial p_d} \right), \quad 1 \leq j \leq d,$$

são linearmente independentes em todo ponto num subconjunto aberto e denso do domínio;

- *em involução*: ou seja, tais que os colchetes de Poisson

$$\{I_j, I_k\} = \sum_{i=1}^d \left[\frac{\partial I_j}{\partial q_i} \frac{\partial I_k}{\partial p_i} - \frac{\partial I_j}{\partial p_i} \frac{\partial I_k}{\partial q_i} \right].$$

são todos identicamente nulos.

Segue dos comentários anteriores que todo sistema com $d = 1$ grau de liberdade é integrável: o próprio hamiltoniano H é uma integral primeira. Outro exemplo importante:

Exemplo 4.4.3. Para qualquer número $d \geq 1$ de graus de liberdade, suponha que o hamiltoniano H depende apenas das variáveis $p = (p_1, \dots, p_d)$. Então as equações de Hamilton-Jacobi (4.4.1) se reduzem a

$$\frac{dq_j}{dt} = \frac{\partial H}{\partial p_j}(p) \quad \text{e} \quad \frac{dp_j}{dt} = -\frac{\partial H}{\partial q_j}(p) = 0.$$

A segunda equação significa, precisamente, que cada p_j é uma integral primeira; é fácil ver que estas integrais primeiras são independentes e estão em involução. Observe também que na primeira equação o termo da direita é independente do tempo. Logo, a solução da equação é

$$q_j(t) = q_j(0) + \frac{\partial H}{\partial p_j}(p(0)) t.$$

Como vamos comentar a seguir, este exemplo é totalmente típico de sistemas integráveis.

Um teorema clássico de Liouville afirma que se o sistema é integrável então as equações de Hamilton-Jacobi podem ser resolvidas completamente por quadraturas. Na demonstração (veja o livro de Arnold [Arn78]) são construídas

funções $\varphi = (\varphi_1, \dots, \varphi_d) \in \mathbb{T}^d$ as quais, juntamente com as integrais primeiras $I = (I_1, \dots, I_d) \in \mathbb{R}^d$, constituem *coordenadas canônicas* do sistema (chamadas na literatura de *coordenadas ação-ângulo*). Com isto, queremos dizer que

$$\Psi : (q, p) \mapsto (\varphi, I)$$

é uma mudança de coordenadas (ou seja, um difeomorfismo) que preserva a forma das equações de Hamilton-Jacobi: (4.4.1) traduz-se para

$$\frac{d\varphi_j}{dt} = \frac{\partial H'}{\partial I_j} \quad \text{e} \quad \frac{dI_j}{dt} = -\frac{\partial H'}{\partial \varphi_j}, \quad (4.4.2)$$

onde $H' = H \circ \Psi^{-1}$ é a expressão do hamiltoniano nas novas coordenadas. Como as I_j são integrais primeiras, a segunda equação dá que

$$0 = \frac{dI_j}{dt} = -\frac{\partial H'}{\partial \varphi_j}.$$

Isto quer dizer que H não depende das variáveis φ e, portanto, estamos numa situação similar ao Exemplo 4.4.3. Cada trajetória do fluxo hamiltoniano está restrita a um toro $\{I = \text{const}\}$ e, de acordo com a primeira equação em (4.4.2), é linear na coordenada φ :

$$\varphi_j(t) = \varphi_j(0) + t\omega_j(I), \quad \text{onde } \omega_j(I) = \frac{\partial H'}{\partial I_j}(I).$$

Em termos das coordenadas originais (q, p) , concluímos que as trajetórias do fluxo hamiltoniano são dadas por

$$t \mapsto \Psi^{-1}(\varphi(0) + \omega(I)t, I) = \Phi_{\varphi(0), I}(\omega(I)t) \quad (4.4.3)$$

onde $\Phi_{\varphi(0), I} : \mathbb{R}^d \rightarrow M$ é uma função \mathbb{Z}^d -periódica e $\omega(I) = (\omega_1(I), \dots, \omega_d(I))$ é chamado *vetor de frequências*. Dizemos que a trajetória é *quase periódica*.

4.4.2 Teoria de Kolmogorov-Arnold-Moser

É claro que sistemas integráveis nunca são ergódicos. Mas isso, por si só, não seria impedimento para que a maioria dos sistemas hamiltonianos fossem ergódicos, já que integrabilidade é uma propriedade muito rara. No entanto, o resultado fundamental que vamos enunciar em seguida afirma que sistemas integráveis genéricos são *robustamente* não ergódicos: todo fluxo hamiltoniano próximo também é não ergódico.

Seja H_0 um hamiltoniano integrável, escrito em coordenadas de ação-ângulo (φ, I) . Mais precisamente, seja B^d uma bola em \mathbb{R}^d e suponha que $H_0(\varphi, I)$ está definido para todo $(\varphi, I) \in \mathbb{T}^d \times B^d$ mas depende apenas da coordenada I . Dizemos que H_0 é *não-degenerado* se a sua matriz hesseana é invertível:

$$\det \left(\frac{\partial^2 H_0}{\partial I_i \partial I_j} \right)_{i,j} \neq 0 \quad \text{em todo ponto.} \quad (4.4.4)$$

Observe que a matriz hessiana de H_0 coincide com a matriz jacobiana de $\omega(I)$. Portanto, a *condição de torção* (4.4.4) significa que a aplicação que associa a cada valor de I o respectivo vetor de frequências $\omega(I)$ é um difeomorfismo local.

Teorema 4.4.4. *Seja $H_0(I)$ um hamiltoniano integrável, não degenerado, de classe C^∞ . Então existe uma vizinhança \mathcal{V} de H_0 no espaço $C^\infty(\mathbb{T}^d \times B^d, \mathbb{R})$ tal que para todo $H \in \mathcal{V}$ existe um conjunto compacto $K \subset \mathbb{T}^d \times B^d$ satisfazendo:*

- (a) *K é uma união de toros diferenciáveis da forma $\{(\varphi, u(\varphi)) : \varphi \in \mathbb{T}^d\}$ cada um dos quais é invariante pelo fluxo hamiltoniano de H ;*
- (b) *a restrição do fluxo hamiltoniano a cada um destes toros é conjugada a um fluxo linear em \mathbb{T}^d ;*
- (c) *o conjunto K tem volume positivo e, de fato, o volume do complementar converge para zero quando $H \rightarrow H_0$.*

Claramente, a existência um conjunto K como no enunciado implica que o fluxo hamiltoniano de H não é ergódico. Podemos resumir o conteúdo do teorema dizendo que a maioria dos toros invariantes persiste quando perturbamos um sistema integrável não degenerado. A prova mostra a persistência ou não de um dado toro está intimamente ligada às propriedades aritméticas do respectivo vetor de frequências, da seguinte forma.

Dados $c > 0$ e $\tau > 0$, dizemos que um vetor $\omega \in \mathbb{R}^d$ é (c, τ) -*diofantino* se

$$|k \cdot \omega| \geq \frac{c}{\|k\|^\tau} \quad \text{para todo } k \in \mathbb{Z}^d, \quad (4.4.5)$$

onde $\|k\| = |k_1| + \dots + |k_d|$. Vetores diofantinos são racionalmente independentes; de fato a condição (4.4.5) significa que, num certo sentido, ω é mal aproximado por vetores racionalmente dependentes. Dizemos que ω é τ -diofantino se ele é (c, τ) -diofantino para algum $c > 0$. O conjunto dos vetores τ -diofantinos é não vazio se, e somente se, $\tau \geq d - 1$; além disso, ele tem medida total em \mathbb{R}^d sempre que τ é estritamente maior que $d - 1$ (veja o Exercício 4.4.1).

Na prova do Teorema 4.4.4 mostra-se que, dados $c > 0$ e $\tau \geq d - 1$ e dado qualquer compacto $\Omega \subset \omega(B^d)$, podemos encontrar uma vizinhança \mathcal{V} de H_0 tal que, para todo $H \in \mathcal{V}$ e todo vetor (c, τ) -diofantino ω , o fluxo de H admite um toro invariante diferenciável, restrito ao qual o fluxo hamiltoniano de H é conjugado ao fluxo linear $t \mapsto \varphi(t) = \varphi(0) + t\omega$.

A seguir discutimos uma versão do Teorema 4.4.4 para sistemas com tempo discreto, mais precisamente, para transformações simpléticas. Lembre que uma *forma simplética* numa variedade M é uma 2-forma diferencial θ não degenerada: para todo $x \in M$ e todo $u \neq 0$, existe v tal que $\theta_x(u, v) \neq 0$. Existência de uma forma simplética implica que a dimensão de M é par: $\dim M = 2d$; além disso, a d -ésima potência $\theta^d = \theta \wedge \dots \wedge \theta$ é uma forma de volume em M . Uma transformação diferenciável $f : M \rightarrow M$ diz-se *simplética* se ela preserva a forma simplética: $\theta_x(u, v) = \theta_{f(x)}(Df(x)u, Df(x)v)$ para todo $x \in M$ e quaisquer $u, v \in T_x M$. Então, em particular, f preserva a forma de volume θ^d .

Exemplo 4.4.5. Seja $M = \mathbb{R}^{2d}$, com coordenadas $(q_1, \dots, q_d, p_1, \dots, p_d)$, e seja θ a forma diferencial definida por

$$\theta_x = dq_1 \wedge dp_1 + \dots + dq_d \wedge dp_d \quad (4.4.6)$$

para todo x . Então θ é uma forma simplética em M . De fato, um teorema clássico de Darboux afirma que para toda forma simplética numa variedade existe uma atlas da variedade tal que a expressão da forma simplética em qualquer carta local desse atlas coincide com (4.4.6). Qualquer transformação da forma

$$f(q_1, \dots, q_d, p_1, \dots, p_d) = (q_1 + \omega_1(p), \dots, q_d + \omega_d(p), p_1, \dots, p_d)$$

é simplética relativamente a θ . Usando

$$Df \cdot \frac{\partial}{\partial q_j} = \frac{\partial}{\partial q_j} \quad \text{e} \quad Df \cdot \frac{\partial}{\partial p_j} = \frac{\partial}{\partial p_j} + \frac{\partial \omega_j}{\partial q_j} \frac{\partial}{\partial q_j},$$

vemos que f é simplética relativamente à forma θ .

Por analogia com o caso de fluxos, dizemos que uma transformação f_0 é *integrável* se existem coordenadas $(q, p) \in \mathbb{T}^d \times B^d$ tais que $f_0(q, p) = (q + \omega(p), p)$ para todo (q, p) . Além disso, dizemos que f_0 é *não degenerada* se a aplicação $p \mapsto \omega(p)$ é um difeomorfismo local.

Teorema 4.4.6. *Seja f_0 uma transformação integrável não degenerada de classe C^∞ . Então existe uma vizinhança \mathcal{V} de f_0 no espaço $C^\infty(\mathbb{T}^d \times B^d, \mathbb{R}^d)$ tal que para toda transformação simplética² $f \in \mathcal{V}$ existe um conjunto compacto $K \subset \mathbb{T}^d \times B^d$ satisfazendo:*

- (a) *K é uma união de toros diferenciáveis da forma $\{(q, u(q)) : q \in \mathbb{T}^d\}$, cada um dos quais é invariante for f ;*
- (b) *a restrição de f a cada um destes toros é conjugada a uma translação em \mathbb{T}^d ;*
- (c) *o conjunto K tem volume positivo e, de fato, o volume do complementar converge para zero quando $f \rightarrow f_0$.*

Analogamente ao caso anterior, a presença de tal conjunto K implica que f não é ergódica. Outra analogia é que o conjunto K está formado por toros restritos aos quais a dinâmica é conjugada a uma rotação diofantina.

Os Teoremas 4.4.4 e 4.4.6 se estendem para sistemas de classe C^r com r finito mas suficientemente grande (dependendo da dimensão). Por exemplo, a versão do Teorema 4.4.6 para $d = 1$ é verdadeira para $r > 3$ e falsa para $r < 3$; no caso de transformações de classe C^3 , as conclusões (a) e (b) do teorema permanecem verdadeiras, mas não a conclusão (c).

²Relativamente à forma simplética canônica (4.4.6).

4.4.3 Pontos periódicos elícticos

Os resultados enunciados acima podem ser aplicados, em particular, para descrever o comportamento de sistemas conservativos na vizinhança de pontos periódicos elícticos. Vamos explicar isto sucintamente, começando pelo caso simplético em dimensão 2.

Quando M é uma superfície, as noções de forma simplética e de forma de área coincidem. Portanto, uma transformação $f : M \rightarrow M$ é simplética se, e somente se, ela preserva área. Seja $\zeta \in M$ um ponto fixo *elíctico*, ou seja, tal que os autovalores de $Df(\zeta)$ estão no círculo unitário. Sejam λ e $\bar{\lambda}$ os autovalores. Dizemos que o ponto fixo ζ é *não degenerado* se $\lambda^k \neq 1$ para todo $1 \leq k \leq 4$. Então, pelo teorema da forma normal de Birkhoff (veja o Apêndice 7 no livro de Arnold [Arn78]), existem coordenadas canônicas $(x, y) \in \mathbb{R}^2$ na vizinhança do ponto fixo, com $\zeta = (0, 0)$ e tais que, relativamente às coordenadas “polares” $(\theta, \rho) \in S^1 \times \mathbb{R}$ definidas por

$$x = \sqrt{\rho} \cos \theta \quad \text{e} \quad y = \sqrt{\rho} \sin \theta$$

a transformação f se escreve na seguinte forma:

$$f(\theta, \rho) = (\theta + \omega_0 + \omega_1 \rho, \rho) + R(\theta, \rho) \quad \text{com} \quad |R(\theta, \rho)| \leq C|\rho|^2. \quad (4.4.7)$$

Observe que a *forma normal* $(\theta, \rho) \mapsto (\theta + \omega_0 + \omega_1 \rho, \rho)$ é integrável. Além disso, ela satisfaz a condição de torção (4.4.4), desde que $\omega_1 \neq 0$ (esta última condição não depende da escolha das coordenadas canônicas, apenas da transformação f). Então podemos aplicar os métodos do Teorema 4.4.6 para concluir que existe um conjunto K com área positiva formado por círculos invariantes com *números de rotação diofantinos*, isto é, tais que a restrição de f a cada um destes círculos é conjugada a uma rotação diofantina. Mais ainda, o ponto fixo ζ é ponto de densidade deste conjunto:

$$\lim_{r \rightarrow 0} \frac{m(B(\zeta, r) \setminus K)}{m(B(\zeta, r))} = 0,$$

onde $B(\zeta, r)$ representa a bola de raio $r > 0$ em torno de ζ .

Vamos nos referir a pontos ζ nas condições do parágrafo anterior como *pontos fixos elícticos genéricos*. Uma consequência importante do que acabamos de dizer é que *pontos fixos elícticos genéricos de transformações que preservam área são estáveis*: a trajetória de qualquer ponto próximo de ζ permanece próxima de ζ para sempre, uma vez que está “presa” dentro de algum círculo invariante pequeno. Este fato não se estende para dimensão maior, como comentaremos mais tarde.

Ainda em dimensão dois, queremos mencionar outros fenômenos dinâmicos importantes que ocorrem na vizinhança de pontos fixos elícticos genéricos. Começamos por apresentar uma ferramenta muito útil, o chamado *último teorema de Poincaré* ou *teorema do ponto fixo de Poincaré-Birkhoff*. O enunciado e certos casos particulares da demonstração foram propostos por Poincaré alguns

meses antes da sua morte; o caso geral foi demonstrado por Birkhoff [Bir13] no ano seguinte.

Seja $A = S^1 \times [a, b]$, com $0 < a < b$, e seja $f : A \rightarrow A$ um homeomorfismo que preserva cada uma das componentes conexas do bordo do anel A . Dizemos que f é um *homeomorfismo de torção* se ele gira as componentes conexas do bordo em sentidos opostos, ou seja, se existe algum levantamento $F : \mathbb{R} \times [a, b] \rightarrow \mathbb{R} \times [a, b]$, $F(\theta, \rho) = (\Theta(\theta, \rho), R(\theta, \rho))$ da transformação f ao recobrimento universal do anel, tal que

$$[\Theta(\theta, a) - \theta][\Theta(\theta, b) - \theta] < 0 \quad \text{para todo } \theta \in \mathbb{R}. \quad (4.4.8)$$

Teorema 4.4.7 (Ponto fixo de Poincaré-Birkhoff). *Se $f : A \rightarrow A$ é um homeomorfismo de torção que preserva área então f admite, pelo menos, dois pontos fixos no interior de A .*

Conforme comentado anteriormente, todo ponto fixo elítico genérico ζ é acumulado por círculos invariantes com números de rotação irracionais diofantinos. Dois quaisquer desses círculos delimitam um anel em torno de ζ . Aplicando o Teorema 4.4.7 (ou, mais precisamente, o corolário no Exercício 4.4.6) conclui-se que cada um desses anéis contém, pelo menos, um par de órbitas periódicas com o mesmo período.

Em certo sentido, estes pares de órbitas periódicas são o que resta dos círculos invariantes da forma normal com números de rotação *racionais*, destruídos pela adição do termo R em (4.4.7). Os seus períodos convergem para infinito quando nos aproximamos de ζ . Genericamente, uma das órbitas periódicas é hiperbólica (pontos de sela) e a outra elítica. Um exemplo está esboçado na Figura 4.3: o ponto fixo elítico ζ está rodeado por uma órbita periódica hiperbólica e uma órbita periódica elítica, assinaladas com as letras p e q , respectivamente, ambas com período 4. Na figura também estão representadas dois círculos invariantes em torno de ζ .

O matemático suíço Eduard Zehnder mostrou que, genericamente, as órbitas periódicas hiperbólicas apresentam pontos homoclínicos transversais, ou seja, as respectivas variedades estáveis e instáveis se intersectam transversalmente, tal como está representado na Figura 4.3. Isto implica que a geometria das variedades estáveis e instáveis é extremamente complexa. As órbitas periódicas elíticas podem ser obtidas satisfazendo as condições de genericidade referidas anteriormente. Isto quer dizer que toda a complexidade dinâmica que estamos descrevendo na vizinhança de ζ se reproduz na vizinhança de cada uma dessas órbitas elíticas “satélites”.

Além disso, a teoria desenvolvida pelo físico francês Serge Aubry e pelo matemático americano John Mather mostra que ζ também é acumulado por certos conjuntos invariantes totalmente desconexos, restrita a cada um dos quais a transformação f é minimal (todas as órbitas são densas). Em certo sentido, estes *conjuntos de Aubry-Mather* correspondem aos círculos invariantes da forma normal com números de rotação irracionais *não diofantinos*, igualmente destruídos pela adição do termo R em (4.4.7).

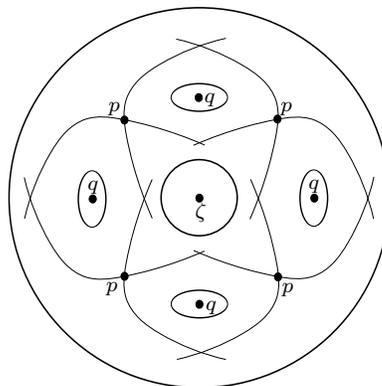


Figura 4.3: Círculos invariantes, órbitas periódicas e interseções homoclínicas na vizinhança de um ponto fixo elítico genérico

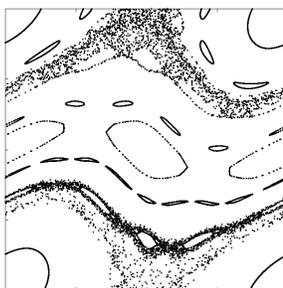


Figura 4.4: Evidência computacional da presença de círculos invariantes, ilhas elíticas e interseções homoclínicas transversais

A Figura 4.4 resume e ilustra boa parte do que acabamos de dizer. As trajetórias calculadas por computador representadas na figura permitem intuir a presença de círculos invariantes, de satélites elíticos, com seus próprios círculos invariantes, e até mesmo de órbitas hiperbólicas com as respectivas interseções homoclínicas transversais. Também observamos na figura a presença de trajetórias com comportamento “caótico”, aparentemente relacionadas com essas interseções transversais.

Mais geralmente, seja $f : M \rightarrow M$ um difeomorfismo simplético numa variedade simplética M de dimensão $2d \geq 2$ qualquer. Dizemos que um ponto fixo $\zeta \in M$ de f é *elítico* se todos os autovalores da derivada $Df(\zeta)$ estão no círculo unitário. Sejam $\lambda_1, \bar{\lambda}_1, \dots, \lambda_d, \bar{\lambda}_d$ esses autovalores. Dizemos que ζ é *não degenerado* se $\lambda_1^{k_1} \dots \lambda_d^{k_d} \neq 1$ para todo $(k_1, \dots, k_d) \in \mathbb{Z}^d$ com $|k_1| + \dots + |k_d| \leq 4$ (em particular, os autovalores são todos distintos). Então, pelo teorema da forma normal de Birkhoff (Apêndice 7 no livro de Arnold [Arn78]), existem co-

ordenadas canônicas $(x_1, \dots, x_d, y_1, \dots, y_d) \in \mathbb{R}^{2d}$ na vizinhança de ζ tais que, relativamente às coordenadas “polares” $(\theta, \rho) \in \mathbb{T}^d \times \mathbb{R}^d$ definidas por

$$x_j = \sqrt{\rho_j} \cos \theta_j \quad \text{e} \quad y_j = \sqrt{\rho_j} \sin \theta_j, \quad j = 1, \dots, d,$$

a transformação f se escreve:

$$f(\theta, \rho) = (\theta + \omega_0 + \omega_1(\rho), \rho) + R(\theta, \rho) \quad \text{onde} \quad \|R(\theta, \rho)\| \leq \text{const} \|\rho\|^2,$$

$\omega_0 \in \mathbb{R}^d$ e ω_1 é uma aplicação linear em \mathbb{R}^d . Supondo que ω_1 seja um isomorfismo (mais uma condição genérica sobre a transformação f), temos que a forma normal

$$(\theta, \rho) = (\theta + \omega_0 + \omega_1(\rho), \rho)$$

é integrável e satisfaz a condição de torção 4.4.4. Aplicando as ideias do Teorema 4.4.6, concluímos que ζ é ponto de densidade de um conjunto K formado por toros invariantes de dimensão d , restritos aos quais a transformação f é conjugada a uma rotação diofantina.

Em particular, transformações simpléticas com pontos fixos (ou periódicos) elíticos genéricos nunca são ergódicas. Observe, por outro lado que quando $d > 1$ um toro de dimensão d não separa o espaço ambiente M em duas componentes conexas. Portanto, o argumento que usamos atrás para mostrar que pontos fixos elíticos genéricos em superfícies são estáveis não se estende para dimensão superior. De fato, é sabido que quando $d > 1$ certas trajetórias começando perto do ponto fixo podem escapar para longe dele. Isto está relacionado com o fenômeno chamado de *difusão de Arnold*, que constitui um tema muito ativo de pesquisa nesta área.

Finalmente, vamos observar que a teoria descrita anteriormente também se aplica a sistemas conservativos com tempo contínuo. Dizemos que um ponto estacionário ζ de um fluxo hamiltoniano é *elítico* se todos os autovalores da derivada do campo de vetores no ponto ζ são números imaginários puros. Argumentos semelhantes ao do caso de tempo discreto mostram que, sob hipóteses genéricas, o ponto estacionário elítico é ponto de densidade de um conjunto formado por toros invariantes de dimensão d restritos a cada um dos quais o fluxo hamiltoniano é conjugado a um fluxo linear.

Além disso, valem resultados análogos para trajetórias periódicas de fluxos hamiltonianos. Um modo de obter tais resultados é considerando uma seção transversal ao fluxo cortando a trajetória periódica, e aplicando as ideias anteriores à transformação de Poincaré correspondente. Concluímos que, sob certas condições genéricas, trajetórias periódicas elíticas de fluxos hamiltonianos são acumuladas por conjuntos com medida positiva formados por toros invariantes do fluxo.

A teoria de Kolmogorov, Arnold, Moser tem muitas outras aplicações, numa grande diversidade de situações em Matemática, que fogem ao escopo deste texto. O leitor interessado poderá obter informação mais completa nas seguintes referências: Arnold [Arn78], Bost [Bos86], Yoccoz [Yoc92], de la Llave [dlL93] e Arnold, Kozlov, Neishtadt [AKN06], entre outras.

4.4.4 Fluxos geodésicos

Seja M uma variedade riemanniana compacta. O fibrado tangente TM é o espaço das duplas (x, v) em que $x \in M$ e v é um vetor tangente à variedade no ponto x . Denotamos por T^1M o fibrado tangente unitário, formado pelas duplas $(x, v) \in TM$ com $\|v\| = 1$. Segue da teoria das equações diferenciais que para cada $(x, v) \in TM$ existe uma única geodésica $\gamma_{x,v} : \mathbb{R} \rightarrow M$ na variedade tal que $\gamma_{x,v}(0) = x$ e $\dot{\gamma}_{x,v}(0) = v$. Além disso, a família de transformações definida por

$$f^t : (x, v) \mapsto (\gamma_{x,v}(t), \dot{\gamma}_{x,v}(t))$$

é um fluxo no fibrado tangente TM , que é chamado *fluxo geodésico* de M .

Equivalentemente, o fluxo geodésico pode ser definido como o fluxo hamiltoniano em TM correspondente à função hamiltoniana $H(x, v) = \|v\|^2$. Portanto, $(f^t)_t$ preserva a medida de Liouville do fibrado tangente, a qual pode ser descrita da seguinte forma. A métrica riemanniana induz uma medida de volume dx na variedade M e também induz um volume dv em cada espaço tangente T_xM . A medida de Liouville de TM é dada, localmente, pelo produto $dx dv$. Além disso, a restrição m da medida de Liouville ao fibrado tangente unitário é dada, localmente, pelo produto $dx d\alpha$, onde $d\alpha$ é a medida de ângulo na esfera unitária de T_xM . O fato de que H é uma integral primeira significa que a norma $\|v\|$ é constante ao longo de trajetórias do fluxo. Em particular, $(f^t)_t$ deixa invariante o fibrado tangente unitário. Além disso, ele preserva a medida de Liouville m restrita a T^1M .

No entanto, o comportamento dos fluxos geodésicos é, geralmente, muito distinto da dinâmica dos sistemas quase integráveis que descrevemos na Seção 4.4.2. Por exemplo, o matemático austríaco Eberhard Hopf [Hop39] mostrou em 1939 que se M é uma superfície compacta com *curvatura gaussiana negativa* em todo ponto então o fluxo geodésico é ergódico. Este teorema foi estendido para variedades de qualquer dimensão quase três décadas depois, por meio do seguinte resultado notável do matemático soviético Dmitry Anosov [Ano67]:

Teorema 4.4.8 (Anosov). *Seja M uma variedade compacta com curvatura seccional negativa. Então o fluxo geodésico no fibrado tangente unitário é ergódico para a medida de Liouville em T^1M .*

Desta forma, os fluxos geodésicos em variedades de curvatura negativa constituíram a primeira classe importante de sistemas hamiltonianos para os quais a hipótese ergódica pode ser comprovada rigorosamente.

4.4.5 Sistemas de Anosov

O primeiro ingrediente fundamental na demonstração do Teorema 4.4.8 é mostrar que todo fluxo geodésico de uma variedade com curvatura negativa é *uniformemente hiperbólico*. Isto quer dizer que toda trajetória γ do fluxo está contida em subvariedades invariantes $W^s(\gamma)$ e $W^u(\gamma)$ que se intersectam transversalmente ao longo de γ e tais que

- toda trajetória em $W^s(\gamma)$ é exponencialmente assintótica a γ no futuro
- toda trajetória em $W^u(\gamma)$ é exponencialmente assintótica a γ no passado

(veja a Figura 4.5), com taxas exponenciais de convergência uniformes para toda trajetória γ . Além disso, o fluxo geodésico é transitivo. A segunda parte da demonstração do Teorema 4.4.8 consiste então em mostrar que todo fluxo uniformemente hiperbólico (*fluxo de Anosov*), transitivo, de classe C^2 que preserva volume é ergódico. Comentaremos essa questão num instante.

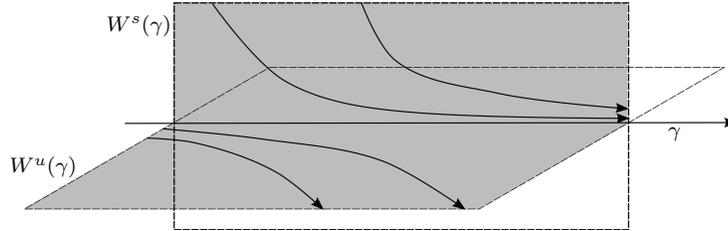


Figura 4.5: Comportamento hiperbólico

Existe uma noção correspondente para sistemas com tempo discreto: dizemos que um difeomorfismo $f : N \rightarrow N$ numa variedade riemanniana compacta é *uniformemente hiperbólico* (*difeomorfismo de Anosov*) se o espaço tangente à variedade em todo ponto $z \in N$ admite uma decomposição em soma direta $T_z N = E_z^s \oplus E_z^u$ tal que a decomposição é invariante pela derivada do difeomorfismo:

$$Df(z)E_z^s = E_{f(z)}^s \quad \text{e} \quad Df(z)E_z^u = E_{f(z)}^u \quad \text{para todo } z \in N, \quad (4.4.9)$$

e a derivada contrai E_z^s e expande E_z^u , uniformemente:

$$\sup_{z \in N} \|Df(z) | E_z^s\| < 1 \quad \text{e} \quad \sup_{z \in N} \|Df(z)^{-1} | E_z^u\| < 1 \quad (4.4.10)$$

(para alguma escolha da norma compatível com a métrica riemanniana de M).

Prova-se que para cada $z \in N$ o conjunto $W^s(z)$ dos pontos cuja trajetória futura é assintótica à trajetória de z forma uma subvariedade diferenciável de N tangente a E_z^s no ponto z ; analogamente, o conjunto $W^u(z)$ dos pontos cuja trajetória passada é assintótica à trajetória de z é uma subvariedade diferenciável tangente a E_z^u no ponto z . Estas subvariedades formam laminações (ou seja, decomposições de N em subvariedades diferenciáveis) que são invariantes pelo difeomorfismo:

$$f(W^s(z)) = W^s(f(z)) \quad \text{e} \quad f(W^u(z)) = W^u(f(z)) \quad \text{para todo } z \in N.$$

Anosov também provou que todo difeomorfismo uniformemente hiperbólico, transitivo, de classe C^2 que preserva o volume é ergódico. A ferramenta técnica crucial é o seguinte teorema provado por Anosov, Sinai [AS67]:

Teorema 4.4.9 (Continuidade absoluta). *As laminações estável e instável de qualquer difeomorfismo (ou fluxo) de Anosov de classe C^2 são absolutamente contínuas:*

1. *se $X \subset N$ tem volume zero então $X \cap W^s(x)$ tem volume zero dentro de $W^s(x)$ para quase todo $x \in N$;*
2. *se $Y \subset \Sigma$ é um subconjunto com volume zero de uma subvariedade Σ transversal à laminação estável, então a união das variedades estáveis dos pontos de Y tem volume zero em N ;*

e analogamente para a laminação instável.

A ergodicidade pode então ser deduzida por meio do argumento de Hopf, que já ilustramos na Seção 4.2.6 num caso particular. Dada qualquer função contínua $\varphi : N \rightarrow \mathbb{R}$, seja E_φ o conjunto dos pontos $z \in N$ onde as médias orbitais no futuro e no passado, $\varphi^+(z)$ e $\varphi^-(z)$ estão definidas e coincidem. E_φ tem volume total, como vimos no Corolário 3.2.8. Observe também que φ^+ é constante em cada variedade estável e φ^- é constante em cada variedade instável. Logo pela primeira parte do Teorema 4.4.9, a interseção $Y_z = W^u(z) \cap E_\varphi$ tem volume total em $W^u(z)$ para quase todo $z \in N$. Além disso, $\varphi^- = \varphi^+$ é constante em cada Y_z . Fixe qualquer z nestas condições. A hipótese de transitividade implica que a união das variedades estáveis dos pontos de $W^u(z)$ é toda a variedade N . Logo, usando a segunda parte do Teorema 4.4.9, a união das variedades estáveis dos pontos de Y_z tem volume total em N . Além disso, φ^+ é constante nessa união. Isto mostra que a média orbital de φ é constante em quase todo ponto, para toda função contínua. Portanto, f é ergódico.

Terminamos esta seção observando que todos os exemplos conhecidos de difeomorfismos de Anosov são transitivos mas isso não é verdade no caso de fluxos de Anosov (veja Verjovsky [Ver99]). Outro problema em aberto é se a ergodicidade ainda vale quando o difeomorfismo ou fluxo de Anosov é apenas de classe C^1 . É sabido que neste caso o teorema de continuidade absoluta (Teorema 4.4.9) é falso, em geral (veja [Bow75b, RY80]).

4.4.6 Bilhares

Como vimos nas Seções 4.4.2 e 4.4.3, sistemas não ergódicos são bastante comuns no mundo dos fluxos hamiltonianos e das transformações simpléticas. No entanto, por si só isto não basta para invalidar a hipótese ergódica de Boltzmann no contexto em que ela foi formulada. De fato, gases ideais são sistemas de um tipo especial: é concebível que ergodicidade possa ser típico nesse âmbito mais restrito, mesmo não o sendo para sistemas hamiltonianos gerais.

Na década de 1960, o matemático e físico teórico soviético Yakov Sinai [Sin63] conjecturou que os sistemas hamiltonianos formados por esferas que chocam elasticamente entre si são ergódicos. Esse modelo para o comportamento de gases ideais havia sido proposto originalmente pelo cientista americano Josiah Willard Gibbs que, juntamente com Boltzmann e com o matemático e físico

teórico escocês James Clark Maxwell, criou a área da Mecânica Estatística. Ele será apresentado de forma precisa no Exemplo 4.4.11. A *hipótese ergódica de Boltzmann-Sinai*, como é conhecida a conjectura de Sinai, é o tema principal da presente seção.

De fato, discutiremos a questão da ergodicidade para sistemas um pouco mais gerais, chamados bilhares, cuja definição formal foi dada pela primeira vez por Birkhoff, nos anos 1930.

Na sua forma mais simples, um bilhar é dado por um domínio conexo limitado $\Omega \subset \mathbb{R}^2$, a *mesa do bilhar*, cujo bordo $\partial\Omega$ está formado por um número finito de curvas diferenciáveis. Chamamos *cantos* aos pontos onde o bordo não é diferenciável; por hipótese eles formam um conjunto finito $\mathcal{C} \subset \partial\Omega$. Consideramos uma partícula pontual em movimento retilíneo uniforme dentro de Ω , com choques elásticos com o bordo. Isto é, a cada encontro com $\partial\Omega \setminus \mathcal{C}$ a partícula é refletida, de tal forma que o ângulo de incidência é igual ao ângulo de reflexão. Quando a partícula acerta um dos cantos ela é absorvida: a trajetória não está definida a partir daí.

Denotamos por \mathbf{n} o campo de vetores unitário normal ao bordo $\partial\Omega$ e apontando “para dentro” de Ω . Ele define uma orientação em $\partial\Omega \setminus \mathcal{C}$: um vetor t tangente ao bordo é positivo se a base $\{t, \mathbf{n}\}$ de \mathbb{R}^2 é positiva. É claro que o movimento da partícula fica totalmente caracterizado pela sequência de choques com o bordo. Além disso, cada choque pode ser descrito pela posição $s \in \partial\Omega$ e pelo ângulo de reflexão $\theta \in (-\pi/2, \pi/2)$. Portanto, a evolução do bilhar é regida pela transformação

$$f : (\partial\Omega \setminus \mathcal{C}) \times (-\pi/2, \pi/2) \rightarrow \partial\Omega \times (-\pi/2, \pi/2), \quad (4.4.11)$$

que a cada choque (s, θ) associa o choque subsequente (s', θ') . Veja a Figura 4.6.

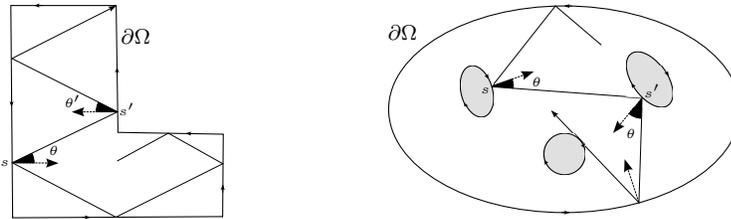


Figura 4.6: Dinâmica de bilhares

No exemplo na esquerda da Figura 4.6 a mesa é poligonal, ou seja, o seu bordo é conexo e está formado por um número finito de segmentos de reta. A trajetória que está representada na figura acerta um dos cantos do bilhar. Trajetórias próximas dela que colidem com o bordo em lados distintos do canto apresentam ângulos de incidência muito diferentes. Em particular, a transformação do bilhar (4.4.11) não é contínua. No exemplo na direita da Figura 4.6 o bordo tem quatro componentes conexas, cada umas das quais é uma curva diferenciável. Logo, não existem cantos. Considere a trajetória representada

na figura, tangenciando uma das componentes conexas do bordo. Trajetórias próximas dela colidem com componentes conexas distintas do bordo. Portanto, a transformação do bilhar apresenta descontinuidades também neste caso.

Exemplo 4.4.10. (Mesa circular) No lado esquerdo da Figura 4.7 representamos o bilhar na bola unitária $\Omega \subset \mathbb{R}^2$. A respectiva transformação está dada por

$$f : (s, \theta) \mapsto (s - (\pi - 2\theta), \theta);$$

no lado direito da Figura 4.7 descrevemos geometricamente o comportamento desta transformação. Observe que f preserva a medida de área $ds d\theta$. Note que f é integrável (no sentido da Seção 4.4.2) e, em particular, a medida de área não é ergódica. Note também que f satisfaz a condição de torção (4.4.4). Veremos daqui a pouco, no Teorema 4.4.12, que todo bilhar plano preserva uma medida natural, equivalente à medida de área em $\partial\Omega \times (-\pi/2, \pi/2)$. A partir das observações anteriores, a teoria KAM permite mostrar que muitos bilhares com mesas quase circulares não são ergódicos relativamente a essa medida invariante.

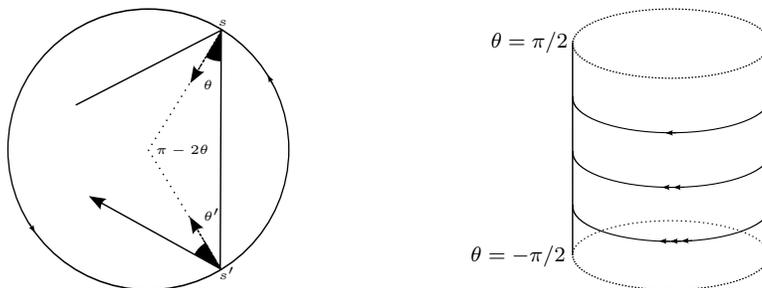


Figura 4.7: Bilhar em mesa circular

A definição de bilhar se estende imediatamente para domínios conexos limitados Ω em qualquer espaço euclidiano \mathbb{R}^d , $d \geq 1$, cujo bordo está formado por um número finito de hipersuperfícies diferenciáveis que se intersectam ao longo de subvariedades de codimensão maior que 1. Denotamos por \mathcal{C} a união destas subvariedades. Tal como antes, munimos $\partial\Omega$ com a orientação induzida pelo vetor unitário normal \mathbf{n} que aponta “para dentro”. Reflexões elásticas no bordo são definidas pela condição de que a trajetória incidente, a trajetória refletida e o vetor normal \mathbf{n} são coplanares e o ângulo de incidência é igual ao ângulo de reflexão. A transformação do bilhar é definida como em (4.4.11), tendo como domínio

$$\{(s, v) \in (\partial\Omega \setminus \mathcal{C}) \times S^{d-1} : v \cdot \mathbf{n}(s) > 0\}.$$

Mais geralmente ainda, podemos tomar como mesa de bilhar qualquer domínio conexo limitado numa variedade riemanniana, cujo bordo está formado por um número finito de hipersuperfícies diferenciáveis. As definições são análogas, exceto que as trajetórias entre choques consecutivos com o bordo são dadas por geodésicas da variedade.

Exemplo 4.4.11 (Gases ideais e bilhares). Idealmente, um gás está formado por um grande número N de moléculas ($N \approx 10^{27}$) que se deslocam em movimento retilíneo e uniforme e se chocam de forma elástica. Veja o lado direito da Figura 4.8. Por simplicidade, vamos supor que as moléculas são esferas idênticas, contidas no toro³ de dimensão $d \geq 2$. Suponhamos, igualmente, que todas as moléculas se deslocam com velocidade escalar igual a 1. Este sistema pode ser descrito por meio de um bilhar, da seguinte forma.

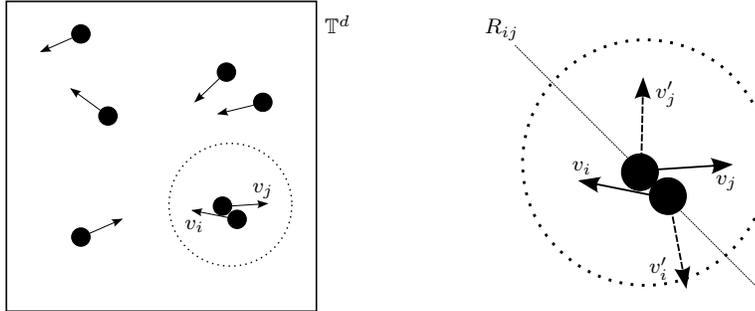


Figura 4.8: Modelo de um gás ideal

Para $1 \leq i \leq N$, represente por $p_i \in \mathbb{T}^d$ a posição do centro da i -ésima molécula M_i . Seja $\rho > 0$ o raio de cada molécula. Então, cada estado do sistema fica inteiramente descrito por um valor de $p = (p_1, \dots, p_N)$ no conjunto

$$\Omega = \{p = (p_1, \dots, p_N) \in \mathbb{T}^{Nd} : \|p_i - p_j\| \geq 2\rho \text{ para todo } i \neq j\}$$

(este conjunto é conexo, desde que ρ seja suficientemente pequeno). Na ausência de choques, o ponto p se desloca ao longo de uma reta, com velocidade com norma constante igual a 1. Quando duas moléculas M_i e M_j se chocam temos que $\|p_i - p_j\| = 2\rho$ e os vetores de velocidade das duas moléculas mudam da seguinte forma. Sejam v_i e v_j os vetores de velocidade imediatamente antes do choque e seja R_{ij} a reta que passa por p_i e p_j . A hipótese de que o choque é elástico quer dizer que os vetores de velocidade v'_i e v'_j logo após o choque são tais que: (veja o lado direito da Figura 4.8):

- (i) as componentes dos vetores v_i e v'_i na direção da reta R_{ij} são simétricas e o mesmo vale para v_j e v'_j ;
- (ii) as componentes dos vetores v_i e v'_i na direção ortogonal à reta R_{ij} são iguais e o mesmo vale para v_j e v'_j .

Isto significa (Exercício 4.4.4) que o ponto p tem uma reflexão elástica na hipersuperfície $\{p \in \partial\Omega : \|p_i - p_j\| = 2\rho\}$ do bordo de Ω . Portanto, o movimento do ponto p corresponde exatamente à evolução do bilhar na mesa Ω .

³Podemos substituir \mathbb{T}^d por recipientes fisicamente mais plausíveis, como o cubo $[0, 1]^d$ de dimensão d , por exemplo. Mas a análise que segue fica um pouco mais complicada nesse caso, porque precisamos levar em conta também os choques das bolas com as paredes do recipiente.

O fato que vamos enunciar em seguida coloca os bilhares dentro do domínio de interesse da Teoria Ergódica. Seja ds a medida de volume induzida em $\partial\Omega$ pela métrica riemanniana ambiente; no caso bidimensional, ds é, simplesmente, o comprimento de arco. Denotemos por $d\theta$ a medida de ângulo em cada hemisfério $\{v \in S^{d-1} : v \cdot \mathbf{n}(s) > 0\}$.

Teorema 4.4.12. *A transformação f preserva a medida $\nu = \cos\theta ds d\theta$ no domínio $\{(s, v) \in \partial\Omega \times S^{d-1} : v \cdot \mathbf{n}(s) > 0\}$.*

Vamos esboçar a demonstração no caso de bilhares planos (ou seja, quando $\Omega \subset \mathbb{R}^2$), deixando a cargo do leitor verificar que os argumentos se estendem de maneira natural para dimensão arbitrária.

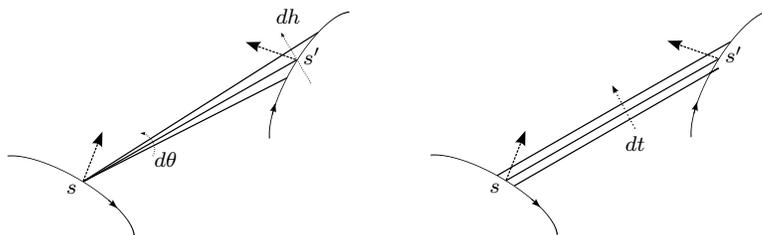


Figura 4.9: Cálculo da derivada da transformação do bilhar

Considere uma família de trajetórias saindo de um único ponto do bordo (ou seja, com s fixado), parametrizada pelo ângulo de reflexão θ , conforme representado no lado esquerdo da Figura 4.9. Designe por $\ell(s, s')$ o comprimento do segmento ligando s a s' . Então $\ell(s, s')d\theta = dh = \cos\theta' ds'$ e, portanto,

$$\frac{\partial s'}{\partial \theta} = \frac{\ell(s, s')}{\cos \theta'}.$$

Para calcular a derivada de θ' em relação a θ , observe que a variação de θ' é a soma de duas componentes: a primeira corresponde à variação de θ enquanto que a segunda provém da variação do vetor normal $\mathbf{n}(s')$ à medida que o ponto de incidência varia. Pela definição de curvatura, esta segunda componente é igual a $\kappa(s')ds'$. Portanto, $d\theta' = d\theta + \kappa(s')ds'$ e, conseqüentemente,

$$\frac{\partial \theta'}{\partial \theta} = 1 + \kappa(s') \frac{\partial s'}{\partial \theta} = 1 + \kappa(s') \frac{\ell(s, s')}{\cos \theta'}.$$

Resumindo, temos que

$$Df(s, \theta) \frac{\partial}{\partial \theta} = \frac{\ell(s, s')}{\cos \theta'} \frac{\partial}{\partial s'} + (1 + \kappa(s') \frac{\ell(s, s')}{\cos \theta'}) \frac{\partial}{\partial \theta'} \quad (4.4.12)$$

Em seguida, considere uma família de trajetórias paralelas, parametrizada pelo comprimento de arco t na direção ortogonal, conforme representado no lado direito da Figura 4.9. As variações de s e s' ao longo desta família são dadas

por $-\cos\theta ds = dt = \cos\theta' ds'$. Como as trajetórias têm direção constante, as variações de θ e θ' provêm apenas das variações dos vetores normais $\mathbf{n}(s)$ e $\mathbf{n}(s')$ quando s e s' variam. Ou seja, $d\theta = \kappa(s)ds$ e $d\theta' = \kappa(s')ds'$. Portanto,

$$Df(s, \theta) \left(-\frac{1}{\cos\theta} \frac{\partial}{\partial s} - \frac{\kappa(s)}{\cos\theta} \frac{\partial}{\partial\theta} \right) = \frac{1}{\cos\theta'} \frac{\partial}{\partial s'} + \frac{\kappa(s')}{\cos\theta'} \frac{\partial}{\partial\theta'}. \quad (4.4.13)$$

Seja $J(s, \theta)$ a matriz da derivada $Df(s, \theta)$ relativamente às bases $\{\partial/\partial s, \partial/\partial\theta\}$ e $\{\partial/\partial s', \partial/\partial\theta'\}$. As relações (4.4.13) e (4.4.13) dão que

$$\det J(s, \theta) = \frac{\begin{vmatrix} \frac{\ell(s, s')}{\cos\theta'} & \frac{1}{\cos\theta'} \\ 1 + \kappa(s') \frac{\ell(s, s')}{\cos\theta'} & \frac{\kappa(s')}{\cos\theta'} \end{vmatrix}}{\begin{vmatrix} 0 & -\frac{1}{\cos\theta} \\ 1 & -\frac{\kappa(s)}{\cos\theta} \end{vmatrix}} = \frac{\cos\theta}{\cos\theta'}. \quad (4.4.14)$$

Então, pela fórmula de mudança de variáveis na integral,

$$\begin{aligned} \int \varphi d\nu &= \int \varphi(s', \theta') \cos\theta' ds' d\theta' = \int \varphi(f(s, \theta)) \cos\theta' \frac{\cos\theta}{\cos\theta'} ds d\theta \\ &= \int \varphi(f(s, \theta)) \cos\theta ds d\theta = \int (\varphi \circ f) d\nu. \end{aligned}$$

Isto mostra que f preserva a medida $\nu = \cos\theta ds d\theta$, conforme afirmamos.

Dizemos que um bilhar é *dispersivo* se o seu bordo é estritamente convexo em todo ponto, quando observado do interior. No caso plano, com as convenções sobre orientação que fizemos anteriormente, isto significa que a curvatura κ é negativa em todo ponto. A Figura 4.10 apresenta dois exemplos. No primeiro $\Omega \subset \mathbb{R}^2$ e o bordo é uma curva conexa formada por cinco segmentos de curva diferenciáveis. No segundo exemplo $\Omega \subset \mathbb{T}^2$ e o bordo tem três componentes conexas, todas diferenciáveis e convexas.

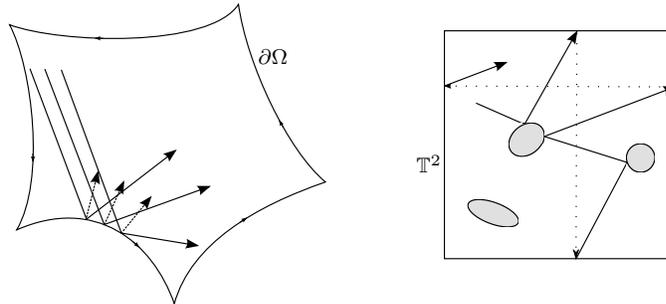


Figura 4.10: Bilhares dispersivos

A classe dos bilhares dispersivos foi introduzida por Sinai em artigo [Sin70] publicado em 1970. A razão de ser da denominação é que para tais bilhares feixes

de trajetórias paralelas se tornam divergentes após reflexão no bordo do bilhar, conforme ilustrado na Figura 4.10. Sinai observou que bilhares dispersivos são sistemas hiperbólicos, num sentido não uniforme: fibrados invariantes E_z^s e E_z^u como em (4.4.9) existem em *quase todo* ponto; no lugar de (4.4.10) temos que a derivada é *assintoticamente* contrativa ao longo de E_z^s e expansiva ao longo de E_z^u , ou seja, para iterados suficientemente grandes (dependendo do ponto z).

Os bilhares associados a gases ideais com $N = 2$ moléculas (Exemplo 4.4.11) são dispersivos: é fácil ver que $\{(p_1, p_2) \in \mathbb{R}^{2d} : \|p_1 - p_2\| = 2\rho\}$ é uma hipersuperfície convexa. Consequentemente, esses bilhares são hiperbólicos, no sentido do parágrafo anterior. Usando uma versão sutil do argumento de Hopf, Sinai mostrou em [Sin70] que tais bilhares são ergódicos, pelo menos no caso $d = 2$. Esta conclusão foi estendida para qualquer $d \geq 2$ por Sinai e seu estudante Nikolai Chernov [SC87], ainda no caso $N = 2$. Esta foi a primeira classe de bilhares para os quais ergodicidade foi provada rigorosamente.

O caso $N \geq 3$ da hipótese ergódica de Boltzmann-Sinai é bem mais difícil, porque os bilhares correspondentes não são dispersivos: a hipersuperfície

$$\{(p_1, p_2, \dots, p_N) \in \mathbb{R}^{Nd} : \|p_1 - p_2\| = 2\rho\}$$

tem geometria cilíndrica, com curvatura nula na direção das variáveis p_i , $i > 2$. Tais bilhares são chamados *semi-dispersivos*. A maior parte dos resultados nesse caso é devida aos matemáticos húngaros András Krámli, Nándor Simányi y Domoko Szász. Nos artigos [KSS91, KSS92] eles provaram hiperbolicidade e ergodicidade para $N = 3$ e também para $N = 4$ supondo $d \geq 3$. Posteriormente, Simányi [Sim02] provou hiperbolicidade para o caso geral: qualquer número de esferas em qualquer dimensão. A prova da ergodicidade continua em aberto, em geral, embora existam muitos outros resultados parciais.

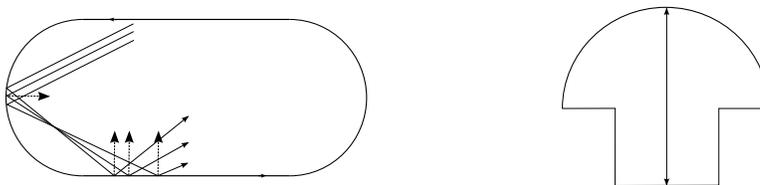


Figura 4.11: Estádio e cogumelo de Bunimovich

Atualmente são conhecidos diversos bilhares ergódicos que não são dispersivos, inclusive tais que a curvatura é maior ou igual que zero em todo ponto. O exemplo mais conhecido é o *estádio de Bunimovich*, cujo bordo está formado por dois semicírculos e dois segmentos de reta. Este bilhar é hiperbólico, mas esta propriedade resulta de um mecanismo diferente, chamado *desfocalização*: um feixe de trajetórias paralelas que se reflete num segmento côncavo da parede da mesa começa por se concentrar (focar), mas depois se dispersa. Veja a Figura 4.11. Outro exemplo interessante é o *cogumelo de Bunimovich*, no qual os comportamentos hiperbólico e elítico coexistem, em conjuntos invariantes disjuntos com medida positiva.

4.4.7 Exercícios

4.4.1. Dizemos que $\omega \in \mathbb{R}^d$ é τ -diofantino se ele é (c, τ) -diofantino, ou seja, se ele satisfaz (4.4.5), para algum $c > 0$. Prove que o conjunto dos vetores τ -diofantinos é não vazio se, e somente se, $\tau \geq d - 1$. Além disso, esse conjunto tem medida total em \mathbb{R}^d sempre que τ é estritamente maior que $d - 1$.

4.4.2. Considere um bilhar em uma mesa retangular. Verifique que toda trajetória (que nunca acerta um dos vértices) ou é periódica ou é densa na mesa do bilhar.

4.4.3. Mostre que todo bilhar em uma mesa triangular acutângula tem alguma trajetória periódica. [Observação: O mesmo vale para triângulos retângulos. Mas, este problema está em aberto no caso de triângulos obtusângulos.]

4.4.4. Considere o modelo de gás ideal descrito no Exemplo 4.4.11. Verifique que os choques elásticos entre duas moléculas quaisquer correspondem a reflexões elásticas do bilhar no bordo de Ω .

4.4.5. Demonstre o Teorema 4.4.7, sob a hipótese adicional de que a função $\rho \mapsto \Theta(\theta, \rho)$ é monótona (crescente ou decrescente) para todo $\theta \in \mathbb{R}$.

4.4.6. Considere o contexto do Teorema 4.4.7 mas, no lugar de (4.4.8), suponha que f gira as componentes conexas do bordo de A com velocidades distintas: existe um levantamento $F : \mathbb{R} \times [a, b] \rightarrow \mathbb{R} \times [a, b]$ e existem $p, q \in \mathbb{Z}$ com $q \geq 1$, tais que, escrevendo $F^q = (\Theta^q, R^q)$,

$$[\Theta^q(\theta, a) - p - \theta] [\Theta^q(\theta, b) - p - \theta] < 0 \quad \text{para todo } \theta \in \mathbb{R}. \quad (4.4.15)$$

Mostre que f tem pelo menos duas órbitas de período q no interior de A .

4.4.7. Seja Ω um domínio convexo no plano cujo bordo $\partial\Omega$ é uma curva diferenciável. Mostre que o bilhar em Ω tem infinitas órbitas periódicas.

Capítulo 5

Decomposição Ergódica

No caso de subconjuntos convexos de espaços vetoriais com dimensão finita tem-se que todo elemento do convexo pode ser escrito como combinação convexa dos elementos extremais. Por exemplo, todo ponto num triângulo no plano pode ser escrito como combinação convexa dos vértices do triângulo. Na sequência dos resultados da Seção 4.3, é natural perguntar se a mesma propriedade vale no espaço das probabilidades invariantes, ou seja, se toda medida invariante é uma combinação linear de medidas ergódicas.

O teorema da decomposição ergódica que vamos provar neste capítulo (Teorema 5.1.3) mostra que a resposta é afirmativa, exceto que o número de “parcelas” nesta combinação não é necessariamente finito, nem mesmo enumerável. Este teorema tem diversas importantes aplicações; em particular, ele permite reduzir a demonstração de muitos resultados ao caso em que o sistema é ergódico.

Provaremos o teorema da decomposição ergódica a partir de outro resultado importante da Teoria da Medida, chamado teorema da desintegração de Rokhlin. A instância mais simples deste teorema ocorre quando temos uma partição finita de um espaço de medida (M, μ) num número finito de subconjuntos mensuráveis P_1, \dots, P_N com medida positiva. Então, evidentemente, podemos escrever μ como combinação linear

$$\mu = \mu(P_1)\mu_1 + \dots + \mu(P_N)\mu_N$$

das suas restrições normalizadas $\mu_i(E) = \mu(E \cap P_i)/\mu(P_i)$ a cada um dos elementos da partição. O teorema da desintegração de Rokhlin (Teorema 5.1.11) afirma que este tipo de desintegração da probabilidade é possível para qualquer partição \mathcal{P} (possivelmente não enumerável!) que possa ser obtida como limite de uma sequência crescente de partições finitas.

5.1 Teorema da decomposição ergódica

Antes de enunciarmos o teorema da decomposição ergódica vamos analisar um par de exemplos que ajudam a motivar e delimitar o seu enunciado:

Exemplo 5.1.1. Considere $f : [0, 1] \rightarrow [0, 1]$ dada por $f(x) = x^2$. As medidas de Dirac δ_0 e δ_1 são invariantes e ergódicas para f . Também é claro que $x = 0$ e $x = 1$ são os únicos pontos recorrentes por f e portanto toda probabilidade invariante μ satisfaz $\mu(\{0, 1\}) = 1$. Em particular, $\mu = \mu(\{0\})\delta_0 + \mu(\{1\})\delta_1$ é uma combinação convexa (finita) de medidas ergódicas.

Exemplo 5.1.2. Considere $f : \mathbb{T}^2 \rightarrow \mathbb{T}^2$ dada por $f(x, y) = (x + y, y)$. A medida de Lebesgue m no toro é preservada por f . Observe que todo círculo horizontal $H_y = S^1 \times \{y\}$ é invariante por f e a restrição $f : H_y \rightarrow H_y$ é a rotação R_y . Seja m_y a medida de Lebesgue em H_y . Observe que m_y também é invariante por f . Além disso, m_y é ergódica sempre que y é irracional. Por outro lado, pelo teorema de Fubini,

$$m(E) = \int m_y(E) dy \quad \text{para todo conjunto mensurável } E. \quad (5.1.1)$$

A igualdade não é afetada se considerarmos a integral restrita ao subconjunto dos valores irracionais de y . Então (5.1.1) apresenta μ como uma combinação convexa (não-enumerável) de medidas ergódicas.

5.1.1 Enunciado do teorema

Antes de enunciar o teorema da decomposição ergódica, precisamos introduzir alguma terminologia. No que segue (M, \mathcal{B}, μ) será um espaço de probabilidade e \mathcal{P} será uma partição de M em conjuntos mensuráveis. Denotaremos por $\pi : M \rightarrow \mathcal{P}$ a projeção natural que associa a cada ponto $x \in M$ o elemento $\mathcal{P}(x)$ da partição que o contém. Esta projeção permite munir \mathcal{P} de uma estrutura de espaço de probabilidade, da seguinte forma. Primeiramente, dizemos que um subconjunto \mathcal{Q} de \mathcal{P} é mensurável se, e somente se, a pré-imagem

$$\pi^{-1}(\mathcal{Q}) = \text{união dos elementos } P \text{ de } \mathcal{P} \text{ que pertencem a } \mathcal{Q}$$

é um subconjunto mensurável de M . É fácil ver que esta definição está correta: a família $\hat{\mathcal{B}}$ dos subconjuntos mensuráveis é uma σ -álgebra em \mathcal{P} . Em seguida, definimos a *medida quociente* $\hat{\mu}$ por

$$\hat{\mu}(\mathcal{Q}) = \mu(\pi^{-1}(\mathcal{Q})) \quad \text{para cada } \mathcal{Q} \in \hat{\mathcal{B}}.$$

Teorema 5.1.3 (Decomposição ergódica). *Seja M um espaço métrico completo separável, $f : M \rightarrow M$ uma transformação mensurável e μ uma probabilidade invariante. Então existe um conjunto mensurável $M_0 \subset M$ com $\mu(M_0) = 1$, uma partição \mathcal{P} de M_0 em subconjuntos mensuráveis e uma família de probabilidades $\{\mu_P : P \in \mathcal{P}\}$ em M , satisfazendo*

- (a) $\mu_P(P) = 1$ para $\hat{\mu}$ -quase todo $P \in \mathcal{P}$;
- (b) $P \mapsto \mu_P(E)$ é mensurável, para todo conjunto mensurável $E \subset M$;
- (c) μ_P é invariante e ergódica para $\hat{\mu}$ -quase todo $P \in \mathcal{P}$;

(d) $\mu(E) = \int \mu_P(E) d\hat{\mu}(P)$, para todo conjunto mensurável $E \subset M$.

A relação (d) significa que μ é uma combinação convexa das várias probabilidades ergódicas μ_P , em que cada μ_P entra com “peso” igual a $\hat{\mu}(P)$. O item (b) assegura que a integral em (d) está bem definida. Observe também que a afirmação em (b) implica que a aplicação $\mathcal{P} \rightarrow \mathcal{M}_1(M)$ dada por $P \mapsto \mu_P$ é mensurável (veja o Exercício 5.1.3).

5.1.2 Desintegração de uma medida

Vamos deduzir o Teorema 5.1.3 de um resultado de Teoria da Probabilidade, o teorema da desintegração de Rokhlin, que tem muitas outras aplicações. Para enunciar este teorema precisamos da seguinte noção.

Definição 5.1.4. Uma *desintegração* de μ relativamente a uma partição \mathcal{P} é uma família $\{\mu_P : P \in \mathcal{P}\}$ de probabilidades em M tal que, para todo conjunto mensurável $E \subset M$:

- (a) $\mu_P(P) = 1$ para $\hat{\mu}$ -quase todo $P \in \mathcal{P}$;
- (b) a aplicação $\mathcal{P} \rightarrow \mathbb{R}$, definida por $P \mapsto \mu_P(E)$ é mensurável;
- (c) $\mu(E) = \int \mu_P(E) d\hat{\mu}(P)$.

Lembre que a partição \mathcal{P} tem uma estrutura natural de espaço de probabilidade, com uma σ -álgebra $\hat{\mathcal{B}}$ e uma probabilidade $\hat{\mu}$. As μ_P são chamadas *probabilidades condicionais* de μ relativamente a \mathcal{P} .

Exemplo 5.1.5. Seja $\mathcal{P} = \{P_1, \dots, P_n\}$ uma partição finita de M em subconjuntos mensuráveis com $\mu(P_i) > 0$ para todo i . A medida quociente $\hat{\mu}$ é dada por $\hat{\mu}(\{P_i\}) = \mu(P_i)$. Considere a restrição normalizada μ_i de μ a cada P_i :

$$\mu_i(E) = \frac{\mu(E \cap P_i)}{\mu(P_i)} \quad \text{para cada conjunto mensurável } E \subset M.$$

Então $\{\mu_1, \dots, \mu_n\}$ é uma desintegração da medida μ relativamente a \mathcal{P} , já que $\mu(E) = \sum_{i=1}^n \hat{\mu}(\{P_i\})\mu_i(E)$ para todo conjunto mensurável $E \subset M$.

Esta construção se estende imediatamente ao caso de partições enumeráveis. No próximo exemplo tratamos um caso não enumerável:

Exemplo 5.1.6. Seja $M = \mathbb{T}^2$ e seja \mathcal{P} a partição de M em círculos horizontais $S^1 \times \{y\}$, $y \in S^1$. Seja m a medida de Lebesgue em \mathbb{T}^2 , seja \hat{m} a medida de Lebesgue em S^1 , e denote por m_y a medida de Lebesgue (medida de comprimento) em cada círculo horizontal $S^1 \times \{y\}$. Pelo teorema de Fubini,

$$m(E) = \int m_y(E) d\hat{m}(y) \quad \text{para todo conjunto mensurável } E \subset \mathbb{T}^2.$$

Logo $\{m_y : y \in S^1\}$ é uma desintegração de m relativamente a \mathcal{P} .

A próxima proposição mostra que desintegrações são essencialmente únicas, quando existem. A hipótese da proposição é muito geral: ela é satisfeita, por exemplo, sempre que M é um espaço topológico com base enumerável de abertos e \mathcal{B} é a sua σ -álgebra de Borel:

Proposição 5.1.7. *Suponha que a σ -álgebra \mathcal{B} admite algum gerador enumerável. Se $\{\mu_P : P \in \mathcal{P}\}$ e $\{\mu'_P : P \in \mathcal{P}\}$ são desintegrações de μ com respeito a \mathcal{P} , então $\mu_P = \mu'_P$ para $\hat{\mu}$ -quase todo $P \in \mathcal{P}$.*

Demonstração. Seja Γ um gerador enumerável de \mathcal{B} e seja \mathcal{A} a álgebra gerada por Γ . Note que \mathcal{A} é enumerável, uma vez que ela coincide com a união das álgebras (finitas) geradas pelos subconjuntos finitos de Γ . Para cada $A \in \mathcal{A}$ considere os conjuntos

$$\mathcal{Q}_A = \{P \in \mathcal{P} : \mu_P(A) > \mu'_P(A)\} \quad \text{e} \quad \mathcal{R}_A = \{P \in \mathcal{P} : \mu_P(A) < \mu'_P(A)\}.$$

Se $P \in \mathcal{Q}_A$ então P está contido em $\pi^{-1}(\mathcal{Q}_A)$ e, usando a propriedade (a) na definição de desintegração, $\mu_P(A \cap \pi^{-1}(\mathcal{Q}_A)) = \mu_P(A)$. Caso contrário, P é disjunto de $\pi^{-1}(\mathcal{Q}_A)$ e, portanto, $\mu_P(A \cap \pi^{-1}(\mathcal{Q}_A)) = 0$. Além disso, valem enunciados análogos com μ'_P no lugar de μ_P . Logo, usando a propriedade (c) na definição de desintegração,

$$\mu(A \cap \pi^{-1}(\mathcal{Q}_A)) = \begin{cases} \int_{\mathcal{P}} \mu_P(A \cap \pi^{-1}(\mathcal{Q}_A)) d\hat{\mu}(P) = \int_{\mathcal{Q}_A} \mu_P(A) d\hat{\mu}(P) \\ \int_{\mathcal{P}} \mu'_P(A \cap \pi^{-1}(\mathcal{Q}_A)) d\hat{\mu}(P) = \int_{\mathcal{Q}_A} \mu'_P(A) d\hat{\mu}(P). \end{cases}$$

Como $\mu_P(A) > \mu'_P(A)$ para todo $P \in \mathcal{Q}_A$, isto implica que $\hat{\mu}(\mathcal{Q}_A) = 0$ para todo $A \in \mathcal{A}$. Um argumento análogo mostra que $\hat{\mu}(\mathcal{R}_A) = 0$ para todo $A \in \mathcal{A}$. Então

$$\bigcup_{A \in \mathcal{A}} \mathcal{Q}_A \cup \mathcal{R}_A$$

também é um subconjunto de \mathcal{P} com medida nula. Para todo P no complementar deste subconjunto, as medidas μ_P e μ'_P coincidem na álgebra geradora \mathcal{A} e, portanto, coincidem em toda a σ -álgebra \mathcal{B} . \square

Por outro lado, desintegrações podem não existir:

Exemplo 5.1.8. Seja $f : S^1 \rightarrow S^1$ uma rotação irracional e seja \mathcal{P} a partição de S^1 cujos elementos são as órbitas $\{f^n(x) : n \in \mathbb{Z}\}$ de f . Suponha que existe uma desintegração $\{\mu_P : P \in \mathcal{P}\}$ da medida de Lebesgue μ com relação a \mathcal{P} . Considere os iterados $\{f_*\mu_P : P \in \mathcal{P}\}$ das probabilidades condicionais. Como os elementos da partição são conjuntos invariantes, $f_*\mu_P(P) = \mu_P(P) = 1$ para $\hat{\mu}$ -quase todo P . É claro que, dado qualquer conjunto mensurável $E \subset M$,

$$P \mapsto f_*\mu_P(E) = \mu_P(f^{-1}(E))$$

é uma função mensurável. Além disso, como μ é uma medida invariante,

$$\mu(E) = \mu(f^{-1}(E)) = \int \mu_P(f^{-1}(E)) d\hat{\mu}(P) = \int f_*\mu_P(E) d\hat{\mu}(P).$$

Estas observações mostram que $\{f_*\mu_P : P \in \mathcal{P}\}$ é uma desintegração de μ relativamente a \mathcal{P} . Por unicidade (Proposição 5.1.7), segue que $f_*\mu_P = \mu_P$ para $\hat{\mu}$ -quase todo P . Ou seja, quase toda probabilidade condicional μ_P é invariante. Isto é uma contradição, uma vez que $P = \{f^n(x) : n \in \mathbb{Z}\}$ é um conjunto infinito enumerável e, portanto, não pode existir nenhuma probabilidade invariante que dê peso positivo a P .

O teorema de Rokhlin afirma que desintegrações sempre existem se a partição \mathcal{P} for o limite de uma sequência crescente de partições enumeráveis e se o espaço M for razoável. O enunciado preciso será dado na próxima seção.

5.1.3 Partições mensuráveis

Dizemos que \mathcal{P} é uma *partição mensurável* se, restrita a algum subconjunto de M com medida total, ela é o limite de uma sequência crescente de partições enumeráveis. Mais precisamente, a partição é mensurável se existe algum conjunto mensurável $M_0 \subset M$ com medida total tal que, restrito a M_0 ,

$$\mathcal{P} = \bigvee_{n=1}^{\infty} \mathcal{P}_n$$

para alguma sequência crescente $\mathcal{P}_1 \prec \mathcal{P}_2 \prec \dots \prec \mathcal{P}_n \prec \dots$ de partições enumeráveis. Lembre que $\mathcal{P}_i \prec \mathcal{P}_{i+1}$ significa que todo elemento de \mathcal{P}_{i+1} está contido em algum elemento de \mathcal{P}_i . Então dizemos que \mathcal{P}_i é *menos fina* do que \mathcal{P}_{i+1} . Além disso, $\bigvee_{n=1}^{\infty} \mathcal{P}_n$ é a partição menos fina tal que

$$\mathcal{P}_n \prec \bigvee_{n=1}^{\infty} \mathcal{P}_n \quad \text{para todo } n.$$

Os seus elementos são as interseções não vazias da forma $\bigcap_{n=1}^{\infty} P_n$ com $P_n \in \mathcal{P}_n$ para todo n . Veja também o Exercício 5.1.1.

É imediato da definição que toda partição finita ou enumerável é mensurável. É fácil dar exemplos de partições mensuráveis não enumeráveis:

Exemplo 5.1.9. Seja $M = \mathbb{T}^2$, munido da medida de Lebesgue m , e seja \mathcal{P} a partição de M nos círculos horizontais $S^1 \times \{y\}$. Então \mathcal{P} é uma partição mensurável. Para ver isso, considere

$$\mathcal{P}_n = \{S^1 \times I(i, n) : i = 1, \dots, 2^n\},$$

onde $I(i, n)$, $1 \leq i \leq 2^n$ é o segmento de $S^1 = \mathbb{R}/\mathbb{Z}$ correspondente ao intervalo $[(i-1)/2^n, i/2^n) \subset \mathbb{R}$. A sequência $(\mathcal{P}_n)_n$ é crescente e $\mathcal{P} = \bigvee_{n=1}^{\infty} \mathcal{P}_n$.

Por outro lado, nem todas as partições são mensuráveis:

Exemplo 5.1.10. Seja $f : M \rightarrow M$ uma transformação mensurável tal que (f, μ) é ergódica. Seja \mathcal{P} a partição de M cujos elementos são as órbitas de f . Então \mathcal{P} não é mensurável, a menos que f possua alguma órbita com medida

total. De fato, suponha que existe uma sequência $\mathcal{P}_1 \prec \mathcal{P}_2 \prec \cdots \prec \mathcal{P}_n \prec \cdots$ de partições enumeráveis tal que $\mathcal{P} = \bigvee_{n=1}^{\infty} \mathcal{P}_n$ restrito a algum conjunto com medida total. Esta última condição implica que quase toda órbita de f está contida em algum elemento P_n da partição \mathcal{P}_n . Em outras palavras, a menos de um conjunto com medida nula, todo elemento de \mathcal{P}_n é invariante por f . Por ergodicidade, segue que para cada n existe exatamente um $P_n \in \mathcal{P}_n$ tal que $\mu(P_n) = 1$. Denote $P = \bigcap_{n=1}^{\infty} P_n$. Então P é um elemento da partição $\bigvee_{n=1}^{\infty} \mathcal{P}_n = \mathcal{P}$, ou seja, é uma órbita de f e tem $\mu(P) = 1$.

Teorema 5.1.11 (Desintegração de Rokhlin). *Suponha que o espaço métrico M é completo separável e que \mathcal{P} é partição mensurável. Então a probabilidade μ admite alguma desintegração relativamente a \mathcal{P} .*

O Teorema 5.1.11 será provado na Seção 5.2. Pode mostrar-se que a hipótese de que \mathcal{P} é mensurável é também necessária para a conclusão do teorema (veja o Exercício 5.2.2).

5.1.4 Prova do teorema da decomposição ergódica

Neste momento vamos usar o Teorema 5.1.11 para provar o teorema da decomposição ergódica. Seja \mathcal{U} uma base enumerável de abertos de M e seja \mathcal{A} a álgebra gerada por \mathcal{U} . Note que \mathcal{A} é enumerável e que ela gera a σ -álgebra de Borel de M . Pelo teorema ergódico de Birkhoff, para cada $A \in \mathcal{A}$ existe um conjunto $M_A \subset M$ com $\mu(M_A) = 1$ tal o tempo médio de visita $\tau(A, x)$ está bem definido para todo $x \in M_A$. Considere $M_0 = \bigcap_{A \in \mathcal{A}} M_A$. Note que $\mu(M_0) = 1$, uma vez que a interseção é enumerável.

Agora considere a partição \mathcal{P} de M_0 definida da seguinte forma: dois pontos $x, y \in M_0$ estão no mesmo elemento de \mathcal{P} se, e somente se, $\tau(A, x) = \tau(A, y)$ para todo $A \in \mathcal{A}$. Afirmamos que esta partição é mensurável. Para mostrar esse fato, considere uma enumeração qualquer $\{A_k : k \in \mathbb{N}\}$ dos elementos da álgebra \mathcal{A} e seja $\{q_k : k \in \mathbb{N}\}$ uma enumeração dos números racionais. Para cada $n \in \mathbb{N}$, considere a partição \mathcal{P}_n de M_0 definida da seguinte forma: dois pontos $x, y \in M_0$ estão no mesmo elemento de \mathcal{P}_n se, e somente se, para todo $i, j \in \{1, \dots, n\}$,

$$\text{ou } \tau(A_i, x) \leq q_j \text{ e } \tau(A_i, y) \leq q_j$$

$$\text{ou } \tau(A_i, x) > q_j \text{ e } \tau(A_i, y) > q_j.$$

É claro que cada \mathcal{P}_n é uma partição finita (com não mais de 2^{n^2} elementos). Segue imediatamente da definição que x e y estão no mesmo elemento de $\bigvee_{n=1}^{\infty} \mathcal{P}_n$ se, e somente se, $\tau(A_i, x) = \tau(A_i, y)$ para todo i . Em outras palavras,

$$\mathcal{P} = \bigvee_{n=1}^{\infty} \mathcal{P}_n.$$

Observe também que os elementos de \mathcal{P} são conjuntos invariantes por f , já que os tempos médios de visita são constantes ao longo de órbitas.

Então, pelo Teorema 5.1.11, existe alguma desintegração $\{\mu_P : P \in \mathcal{P}\}$ de μ relativamente a \mathcal{P} . Os itens (a), (b) e (d) do Teorema 5.1.3 estão contidos na definição de desintegração. Para provar o item (c) basta mostrar que μ_P é invariante e ergódica para $\hat{\mu}$ -quase todo P , que é o que faremos a seguir.

Considere a família de probabilidades $\{f_*\mu_P : P \in \mathcal{P}\}$. Note que

$$f_*\mu_P(P) = \mu_P(f^{-1}(P)) = \mu_P(P) = 1$$

uma vez que todo $P \in \mathcal{P}$ é um conjunto invariante. Além disso, dado qualquer conjunto mensurável $E \subset M$, a função

$$P \mapsto f_*\mu_P(E) = \mu_P(f^{-1}(E))$$

é mensurável e, usando o fato de que μ é invariante por f ,

$$\mu(E) = \mu(f^{-1}(E)) = \int \mu_P(f^{-1}(E)) d\hat{\mu}(P) = \int f_*\mu_P(E) d\hat{\mu}(P).$$

Isto mostra que $\{f_*\mu_P : P \in \mathcal{P}\}$ é uma desintegração de μ relativamente a \mathcal{P} . Por unicidade (Proposição 5.1.7), segue que $f_*\mu_P = \mu_P$ para quase todo P .

Resta provar que μ_P é ergódica para quase todo P . Como $\mu(M_0) = 1$, temos que $\mu_P(M_0 \cap P) = 1$ para quase todo P . Logo, basta provar que, dado qualquer $P \in \mathcal{P}$ e qualquer conjunto mensurável $E \subset M$, o tempo médio de visita $\tau(E, x)$ está bem definido para todo $x \in M_0 \cap P$ e é constante nesse conjunto. Fixado P , denotemos por \mathcal{C} a classe dos conjuntos mensuráveis E para os quais valem estas propriedades. Por construção, \mathcal{C} contém a álgebra geradora \mathcal{A} . Observe que se $E_1, E_2 \in \mathcal{C}$ com $E_1 \supset E_2$ então $E_1 \setminus E_2 \in \mathcal{C}$:

$$\tau(E_1 \setminus E_2, x) = \tau(E_1, x) - \tau(E_2, x)$$

está definido e é constante em $M_0 \cap P$. Em particular, \mathcal{C} é fechada por passagem ao complementar. Analogamente, \mathcal{C} é fechada por uniões enumeráveis disjuntas: se $E_j \in \mathcal{C}$ são disjuntos dois-a-dois então

$$\tau\left(\bigcup_j E_j, x\right) = \sum_j \tau(E_j, x)$$

está definido e é constante em $M_0 \cap P$. É fácil deduzir que \mathcal{C} é uma classe monótona: sejam $A_n, B_n \in \mathcal{C}$ com $A_n \subset A_{n+1}$ e $B_n \supset B_{n+1}$ para todo n ; pelas duas observações anteriores,

$$\bigcup_{n=1}^{\infty} A_n = A_1 \cup \bigcup_{n=1}^{\infty} (A_{n+1} \setminus A_n) \in \mathcal{C} \quad \text{e} \quad \bigcap_{n=1}^{\infty} B_n = \left(\bigcup_{n=1}^{\infty} B_n^c\right)^c \in \mathcal{C}.$$

Pelo Teorema A.1.18, segue que \mathcal{C} contém a σ -álgebra de Borel de M .

Isto conclui a demonstração do Teorema 5.1.3 a partir do Teorema 5.1.11.

5.1.5 Exercícios

5.1.1. Mostre que uma partição \mathcal{P} é mensurável se, e somente se, existem subconjuntos mensuráveis $M_0, E_1, E_2, \dots, E_n, \dots$ tais que $\mu(M_0) = 1$ e, restrito a M_0 ,

$$\mathcal{P} = \bigvee_{n=1}^{\infty} \{E_n, M \setminus E_n\}.$$

5.1.2. Seja μ uma probabilidade ergódica para uma transformação f e seja $k \geq 2$. Como é a decomposição ergódica de μ para o iterado f^k ?

5.1.3. Seja M um espaço métrico e seja X um espaço mensurável. Mostre que as seguintes condições são equivalentes:

- (a) a aplicação $\nu : X \rightarrow \mathcal{M}_1(M)$, $x \mapsto \nu_x$ é mensurável;
- (b) a aplicação $X \rightarrow \mathbb{R}$, $x \mapsto \int \varphi d\nu_x$ é mensurável, para toda função contínua limitada $\varphi : M \rightarrow \mathbb{R}$;
- (c) a aplicação $X \rightarrow \mathbb{R}$, $x \mapsto \int \psi d\nu_x$ é mensurável, para toda função mensurável limitada $\psi : M \rightarrow \mathbb{R}$;
- (d) a aplicação $X \rightarrow \mathbb{R}$, $x \mapsto \nu_x(E)$ é mensurável, para todo conjunto mensurável $E \subset M$.

5.1.4. Mostre que se $\{\mu_P : P \in \mathcal{P}\}$ é uma desintegração de μ então

$$\int \psi d\mu = \int \left(\int \psi d\mu_P \right) d\hat{\mu}(P)$$

para toda função mensurável limitada $\psi : M \rightarrow \mathbb{R}$.

5.1.5. Seja μ uma probabilidade invariante para uma transformação mensurável $f : M \rightarrow M$. Seja $\hat{f} : \hat{M} \rightarrow \hat{M}$ a extensão natural de f e seja $\hat{\mu}$ o levantamento de μ (Seção 2.4.2). Qual é a relação entre a decomposição ergódica de μ e a decomposição ergódica de $\hat{\mu}$?

5.1.6. Quando M é um espaço métrico compacto, podemos obter a decomposição ergódica tomando para M_0 o conjunto dos pontos $x \in M$ tais que

$$\mu_x = \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \delta_{f^j(x)}$$

existe na topologia fraca* e tomando para \mathcal{P} a partição de M_0 definida por $\mathcal{P}(x) = \mathcal{P}(y) \Leftrightarrow \mu_x = \mu_y$. Verifique os detalhes desta demonstração alternativa do Teorema 5.1.3 para espaços métricos compactos.

5.1.7. Seja $\sigma : \Sigma \rightarrow \Sigma$ o deslocamento em $\Sigma = \{1, \dots, d\}^{\mathbb{Z}}$. Considere a partição \mathcal{W}^s de Σ em “conjuntos estáveis”

$$\mathcal{W}^s((a_n)_n) = \{(x_n)_n : x_n = a_n \text{ para todo } n \geq 0\}.$$

Dada qualquer probabilidade μ invariante por σ , seja $\{\mu_P : P \in \mathcal{P}\}$ uma decomposição ergódica de μ . Mostre que $\mathcal{P} \prec \mathcal{W}^s$ a menos de medida nula.

5.2 Teorema da desintegração de Rokhlin

Agora vamos demonstrar o Teorema 5.1.11. Fixe uma seqüência crescente qualquer $\mathcal{P}_1 \prec \mathcal{P}_2 \prec \dots \prec \mathcal{P}_n \prec \dots$ de partições enumeráveis tal que $\mathcal{P} = \bigvee_{n=1}^{\infty} \mathcal{P}_n$ restrito a algum conjunto $M_0 \subset M$ com medida total. Usaremos $\mathcal{P}_n(x)$ para denotar o elemento de \mathcal{P}_n que contém um dado ponto $x \in M$.

5.2.1 Esperanças condicionais

Seja $\psi : M \rightarrow \mathbb{R}$ uma função mensurável limitada qualquer. Para cada $n \geq 1$, defina $e_n(\psi) : M \rightarrow \mathbb{R}$ da seguinte forma:

$$e_n(\psi, x) = \begin{cases} \frac{1}{\mu(\mathcal{P}_n(x))} \int_{\mathcal{P}_n(x)} \psi d\mu & \text{se } \mu(\mathcal{P}_n(x)) > 0 \\ 0 & \text{caso contrário.} \end{cases} \quad (5.2.1)$$

Como as partições \mathcal{P}_n são enumeráveis, o segundo caso da definição se aplica somente num conjunto de pontos com medida μ igual a zero. Observe também que $e_n(\psi)$ é constante em cada $P_n \in \mathcal{P}_n$; denotamos por $E_n(\psi, P_n)$ o valor desta constante. Então,

$$\int \psi d\mu = \sum_{P_n} \int_{P_n} \psi d\mu = \sum_{P_n} \mu(P_n) E_n(\psi, P_n) = \int e_n(\psi) d\mu \quad (5.2.2)$$

para todo $n \in \mathbb{N}$ (as somas envolvem somente os $P_n \in \mathcal{P}_n$ com medida positiva).

Lema 5.2.1. *Dada qualquer função mensurável limitada $\psi : M \rightarrow \mathbb{R}$, existe um subconjunto M_ψ de M com $\mu(M_\psi) = 1$ tal que*

- (a) $e(\psi, x) = \lim_n e_n(\psi, x)$ existe para todo $x \in M_\psi$.
- (b) $e(\psi) : M_\psi \rightarrow \mathbb{R}$ é mensurável e é constante em cada $P \in \mathcal{P}$.
- (c) $\int \psi d\mu = \int e(\psi) d\mu$.

Demonstração. Inicialmente, suponha que $\psi \geq 0$. Para cada $\alpha < \beta$, seja $S(\alpha, \beta)$ o conjunto dos pontos $x \in M$ tais que

$$\liminf_n e_n(\psi, x) < \alpha < \beta < \limsup_n e_n(\psi, x).$$

É claro que a seqüência $e_n(\psi, x)$ diverge se, e somente se, $x \in S(\alpha, \beta)$ para algum par de números racionais $\alpha < \beta$. Em outras palavras, o limite $e(\psi, x)$ existe se, e somente se, x pertence à interseção M_ψ de todos os $S(\alpha, \beta)^c$ com $\alpha < \beta$ racionais. Como se trata de uma interseção enumerável, para provar que $\mu(M_\psi) = 1$ basta mostrar que $\mu(S(\alpha, \beta)) = 0$ para todo $\alpha < \beta$. Isso é o que faremos a seguir.

Como α e β estão fixados, podemos escrever $S = S(\alpha, \beta)$. Dado $x \in S$, fixe uma seqüência de inteiros $1 \leq a_1^x < b_1^x < \dots < a_i^x < b_i^x < \dots$ tais que

$$e_{a_i^x}(\psi, x) < \alpha \quad \text{e} \quad e_{b_i^x}(\psi, x) > \beta \quad \text{para todo } i \geq 1.$$

Defina A_i como sendo a união dos elementos $A_i(x) = \mathcal{P}_{a_i^x}(x)$ e B_i como sendo a união dos elementos $B_i(x) = \mathcal{P}_{b_i^x}(x)$ obtidos deste modo, para todos os pontos $x \in S$. Por construção, $S \subset A_{i+1} \subset B_i \subset A_i$ para todo $i \geq 1$. Em particular, S está contido no conjunto

$$\tilde{S} = \bigcap_{i=1}^{\infty} B_i = \bigcap_{i=1}^{\infty} A_i.$$

Como a sequência \mathcal{P}_n , $n \geq 1$, é crescente, dados dois quaisquer dos conjuntos $A_i(x) = \mathcal{P}_{a_i^x}(x)$ que formam A_i , ou eles são disjuntos ou um deles está contido no outro. Então os conjuntos $A_i(x)$ maximais são disjuntos dois-a-dois e, portanto, constituem uma partição de A_i . Logo, somando apenas sobre estes conjuntos maximais com medida positiva,

$$\int_{A_i} \psi d\mu = \sum_{A_i(x)} \int_{A_i(x)} \psi d\mu \leq \sum_{A_i(x)} \alpha\mu(A_i(x)) = \alpha\mu(A_i),$$

para qualquer $i \geq 1$. Analogamente,

$$\int_{B_i} \psi d\mu = \sum_{B_i(x)} \int_{B_i(x)} \psi d\mu \geq \sum_{B_i(x)} \beta\mu(B_i(x)) = \beta\mu(B_i).$$

Como $A_i \supset B_i$ e nós estamos supondo que $\psi \geq 0$, segue que

$$\alpha\mu(A_i) \geq \int_{A_i} \psi d\mu \geq \int_{B_i} \psi d\mu \geq \beta\mu(B_i),$$

para todo $i \geq 1$. Tomando o limite quando $i \rightarrow \infty$, obtemos que $\alpha\mu(\tilde{S}) \geq \beta\mu(\tilde{S})$. Isto implica que $\mu(\tilde{S}) = 0$ e, portanto, $\mu(S) = 0$. Isto prova a afirmação quando ψ é não-negativa. O caso geral segue imediatamente, uma vez que sempre podemos escrever $\psi = \psi^+ - \psi^-$, onde ψ^\pm são mensuráveis, não-negativas e limitadas. Note que $e_n(\psi) = e_n(\psi^+) - e_n(\psi^-)$ para todo $n \geq 1$ e, portanto, a conclusão do lema é verdadeira para ψ se ela vale para ψ^+ e ψ^- . Isto conclui a prova da afirmação (a).

As demais afirmações são consequências simples da definição. A mensurabilidade de $e(\psi)$ segue diretamente da Proposição A.1.31. Dado que \mathcal{P}_n é menos fina que \mathcal{P} , é claro que $e_n(\psi)$ é constante em cada $P \in \mathcal{P}$, restrito a um subconjunto de M com medida total. Logo o mesmo vale para $e(\psi)$. Isto prova (b). Observe também que $|e_n(\psi)| \leq \sup|\psi|$ para todo $n \geq 1$. Logo, podemos usar o teorema da convergência dominada para passar ao limite em (5.2.2). Desta forma obtemos a afirmação (c). \square

Aqui estamos especialmente interessados no caso em que ψ é uma função característica: $\psi = \chi_A$ para algum conjunto mensurável $A \subset M$. Neste caso a definição significa que

$$e(\psi, x) = \lim_n \frac{\mu(\mathcal{P}_n(x) \cap A)}{\mu(\mathcal{P}_n(x))}. \quad (5.2.3)$$

Denotamos por \mathcal{P}_A o conjunto dos elementos P da partição \mathcal{P} que intersectam M_ψ . Observe que $\hat{\mu}(\mathcal{P}_A) = 1$. Além disso, definimos $E(A) : \mathcal{P}_A \rightarrow \mathbb{R}$ colocando $E(A, P) = e(\psi, x)$ para qualquer $x \in M_\psi \cap P$. Observe que $e(\psi) = E(A) \circ \pi$. Logo, a função $E(A)$ é mensurável e satisfaz:

$$\int \psi d\mu = \int e(\psi) d\mu = \int E(A) d\hat{\mu}. \quad (5.2.4)$$

5.2.2 Critério de σ -aditividade

A hipótese de que o ambiente M é um espaço métrico completo separável intervém na prova por meio do importante critério de σ -aditividade que vamos enunciar e provar a seguir:

Proposição 5.2.2. *Seja M um espaço métrico completo separável e seja \mathcal{A} a álgebra gerada por uma base enumerável $\mathcal{U} = \{U_k : k \in \mathbb{N}\}$ de abertos de M . Seja $\mu : \mathcal{A} \rightarrow [0, 1]$ uma função aditiva com $\mu(\emptyset) = 0$. Então μ se estende a uma medida de probabilidade na σ -álgebra de Borel de M .*

A ideia da prova é a seguinte. Consideraremos o espaço produto $\Sigma = \{0, 1\}^{\mathbb{N}}$, munido da topologia gerada pelos cilindros

$$[a_1, \dots, a_s] = \{(i_k)_{k \in \mathbb{N}} : i_1 = a_1, \dots, i_s = a_s\}, \quad s \geq 1.$$

Note que Σ é um compacto (Exercício A.1.11). Usando o fato de que M é um espaço métrico completo, mostraremos que a aplicação

$$\gamma : M \rightarrow \Sigma, \quad \gamma(x) = (\mathcal{X}_{U_k}(x))_{k \in \mathbb{N}}$$

é um mergulho mensurável de M em Σ . Além disso, a função μ dá origem a uma função aditiva ν definida na álgebra \mathcal{A}_Σ gerada pelos cilindros de Σ . Esta álgebra é compacta (Definição A.1.15), uma vez que todo elemento é compacto. Logo, ν se estende a uma medida de probabilidade na σ -álgebra de Borel de Σ , que ainda representaremos por ν . Mostraremos que a imagem $\gamma(M)$ tem medida total para ν . Então a imagem $\gamma_*^{-1}\nu$ é uma probabilidade na σ -álgebra de Borel de M . Finalmente, verificaremos que esta probabilidade é uma extensão da função μ .

Passemos a detalhar estes argumentos. No que segue, dado qualquer conjunto $A \subset M$, denotaremos $A^1 = A$ e $A^0 = A^c$.

Lema 5.2.3. *A imagem $\gamma(M)$ é um subconjunto boreliano de Σ .*

Demonstração. Seja $x \in M$ e $(i_k)_k = \gamma(x)$. É claro que

$$(A) \quad \bigcap_{j=1}^k U_j^{i_j} \neq \emptyset \text{ para todo } k \in \mathbb{N},$$

pois x pertence à interseção. Além disso, como \mathcal{U} é uma base de abertos de M ,

$$(B) \quad \text{existe algum } k \text{ tal que } i_k = 1 \text{ e } \text{diam } U_k \leq 1 \text{ e}$$

(C) para todo k tal que $i_k = 1$ existe $l > k$ tal que $i_l = 1$ e $\bar{U}_l \subset U_k$ e $\text{diam } U_l \leq \text{diam } U_k/2$.

Reciprocamente, suponha que $(i_k)_k \in \Sigma$ satisfaz as condições (A), (B) e (C). Vamos mostrar que existe $x \in M$ tal que $\gamma(x) = (i_k)_k$. Para isso, defina

$$F_n = \bigcap_{k=1}^n V_k,$$

onde $V_k = U_k^c$ se $i_k = 0$ e $V_k = \bar{U}_{l(k)}$ se $i_k = 1$. Então $(F_n)_n$ é uma sequência decrescente de conjuntos fechados. A condição (A) garante que $F_n \neq \emptyset$ para todo $n \geq 1$. As condições (B) e (C) implicam que o diâmetro de F_n converge para zero. Então, como M é um espaço completo, a interseção $\bigcap_n F_n$ contém algum ponto x . Por construção, F_n está contido em $\bigcap_{k=1}^n U_k^{i_k}$ para todo n . Segue que

$$x \in \bigcap_{k=1}^{\infty} U_k^{i_k} \quad \text{ou seja} \quad \gamma(x) = (i_k)_k.$$

Desta forma, mostramos que a imagem de γ é caracterizada perfeitamente pelas condições (A), (B) e (C).

Para concluir a demonstração basta mostrar que o subconjunto descrito por cada uma destas condições pode ser construído a partir dos cilindros por meio de uniões e interseções enumeráveis. Dado $k \in \mathbb{N}$, seja $N(k)$ o conjunto das k -uplas (a_1, \dots, a_k) em $\{0, 1\}$ tais que $U_1^{a_1} \cap \dots \cap U_k^{a_k} \neq \emptyset$. A condição (A) corresponde ao subconjunto

$$\bigcap_{k=1}^{\infty} \bigcup_{(a_1, \dots, a_k) \in N(k)} [a_1, \dots, a_k].$$

Seja $D = \{k \in \mathbb{N} : \text{diam } U_k \leq 1\}$. Então a condição (B) corresponde a

$$\bigcup_{k \in D} \bigcup_{(a_1, \dots, a_{k-1})} [a_1, \dots, a_{k-1}, 1].$$

Finalmente, dado $k \in \mathbb{N}$, seja $L(k)$ o conjunto dos $l > k$ tais que $\bar{U}_l \subset U_k$ e $\text{diam } U_l \leq \text{diam } U_k/2$. A condição (C) corresponde ao subconjunto

$$\bigcap_{k=1}^{\infty} \bigcup_{a_1, \dots, a_{k-1}} \left([a_1, \dots, a_{k-1}, 0] \cup \bigcup_{l \in L(k)} \bigcup_{a_{k+1}, \dots, a_{l-1}} [a_1, \dots, a_{k-1}, 1, a_{k+1}, \dots, a_{l-1}, 1] \right).$$

Isto completa a demonstração do lema. □

Corolário 5.2.4. *A aplicação $\gamma : M \rightarrow \Sigma$ é uma bijeção mensurável com inversa mensurável.*

Demonstração. Dados quaisquer pontos $x \neq y$ em M , existe algum $k \in \mathbb{N}$ tal que U_k contém um dos pontos mas não o outro. Isto mostra que γ é injetiva. Para todo $s \geq 1$ e $a_1, \dots, a_s \in \{0, 1\}$,

$$\gamma^{-1}([a_1, \dots, a_s]) = U_1^{a_1} \cap \dots \cap U_s^{a_s}. \quad (5.2.5)$$

Isto garante que a aplicação γ é mensurável, já que os cilindros geram a σ -álgebra de Borel de Σ . Em seguida, observe que

$$\gamma(U_1^{a_1} \cap \dots \cap U_s^{a_s}) = [a_1, \dots, a_s] \cap \gamma(M) \quad \text{para todo } s, a_1, \dots, a_s. \quad (5.2.6)$$

Usando o Lema 5.2.3, segue que $\gamma(U_1^{a_1} \cap \dots \cap U_s^{a_s})$ é um boreliano de Σ para todo s, a_1, \dots, a_s . Isto mostra que a transformação γ^{-1} é mensurável. \square

Usando este fato, vamos agora provar que μ se estende a uma medida de probabilidade na σ -álgebra de Borel de M . Para isso, consideremos a álgebra \mathcal{A}_Σ gerada pelos cilindros de Σ . Note que os seus elementos são as uniões finitas de cilindros. Em particular, todos os elementos de \mathcal{A}_Σ são compactos e, conseqüentemente, \mathcal{A}_Σ é uma álgebra compacta (Definição A.1.15). Defina

$$\nu([a_1, \dots, a_s]) = \mu(U_1^{a_1} \cap \dots \cap U_s^{a_s}), \quad (5.2.7)$$

para cada $s \geq 1$ e $a_1, \dots, a_s \in \{0, 1\}$. Então ν é uma função aditiva no conjunto dos cilindros, tomando valores em $[0, 1]$. Ela se estende de modo natural a uma função aditiva na álgebra \mathcal{A}_Σ , que ainda denotaremos por ν .

É claro que $\nu(\Sigma) = 1$. Além disso, como a álgebra \mathcal{A}_Σ é compacta, podemos usar o Teorema A.1.14 para concluir que a função $\nu : \mathcal{A}_\Sigma \rightarrow [0, 1]$ é σ -aditiva. Logo, pelo Teorema A.1.13, ν se estende a uma medida de probabilidade definida na σ -álgebra de Borel de Σ . Dada qualquer cobertura \mathcal{C} de $\gamma(M)$ por cilindros, segue da definição (5.2.7) que

$$\nu\left(\bigcup_{C \in \mathcal{C}} C\right) = \mu\left(\bigcup_{C \in \mathcal{C}} \gamma^{-1}(C)\right) = \mu(M) = 1.$$

Tomando o ínfimo sobre todas as coberturas, concluímos que $\nu(\gamma(M)) = 1$.

Então, pelo Corolário 5.2.4, a imagem $\gamma_*^{-1}\nu$ é uma probabilidade boreliana em M . Por definição e pela relação (5.2.6)

$$\begin{aligned} \gamma_*^{-1}\nu(U_1^{a_1} \cap \dots \cap U_s^{a_s}) &= \nu(\gamma(U_1^{a_1} \cap \dots \cap U_s^{a_s})) = \nu([a_1, \dots, a_s] \cap \gamma(M)) \\ &= \nu([a_1, \dots, a_s]) = \mu(U_1^{a_1} \cap \dots \cap U_s^{a_s}) \end{aligned}$$

para quaisquer s, a_1, \dots, a_s . Isto implica que $\gamma_*^{-1}\nu$ é uma extensão da função $\mu : \mathcal{A} \rightarrow [0, 1]$. Portanto, a prova da Proposição 5.2.2 está completa.

5.2.3 Construção das medidas condicionais

Seja $\mathcal{U} = \{U_k : k \in \mathbb{N}\}$ uma base de abertos de M e seja \mathcal{A} a álgebra gerada por \mathcal{U} . É claro que \mathcal{A} gera a σ -álgebra de Borel de M . Observe também que

\mathcal{A} é enumerável: ela coincide com a união das álgebras (finitas) geradas pelos subconjuntos $\{U_k : 1 \leq k \leq n\}$, para cada $n \geq 1$. Defina:

$$\mathcal{P}_* = \bigcap_{A \in \mathcal{A}} \mathcal{P}_A$$

Então $\hat{\mu}(\mathcal{P}_*) = 1$, já que a interseção é enumerável. Para cada $P \in \mathcal{P}_*$, defina:

$$\mu_P : \mathcal{A} \rightarrow [0, 1], \quad \mu_P(A) = E(A, P). \quad (5.2.8)$$

Em particular, $\mu_P(M) = E(M, P) = 1$. É claro que μ_P é uma função aditiva: a definição (5.2.3) dá que

$$A \cap B = \emptyset \quad \Rightarrow \quad E(A \cup B, P) = E(A, P) + E(B, P) \quad \text{para todo } P \in \mathcal{P}_*.$$

Pela Proposição 5.2.2, segue que esta função se estende a uma medida de probabilidade (que ainda denotaremos por μ_P) definida na σ -álgebra de Borel de M . Resta verificar que esta família de medidas $\{\mu_P : P \in \mathcal{P}_*\}$ satisfaz as condições na definição de desintegração (Definição 5.1.4).

Comecemos pela condição (a). Seja $P \in \mathcal{P}_*$ e, para cada $n \in \mathbb{N}$, seja P_n o elemento da partição \mathcal{P}_n que contém P . Observe que se $A \in \mathcal{A}$ é tal que $A \cap P_n = \emptyset$ para algum n , então,

$$\mu_P(A) = E(A, P) = \lim_m \frac{\mu(A \cap P_m)}{\mu(P_m)} = 0,$$

já que $P_m \subset P_n$ para todo $m \geq n$. Fixe n . Para cada $s \geq 1$, seja P_n^s a união de todos os conjuntos da forma $U_1^{a_1} \cap \dots \cap U_s^{a_s}$ que intersectam P_n . Pela observação que acabamos de fazer, os cilindros de comprimento s que não estão em P_n^s têm medida nula para μ_P . Portanto, $\mu_P(P_n^s) = 1$ para todo $s \geq 1$. Passando ao limite quando $s \rightarrow \infty$, concluímos que $\mu_P(U) = 1$ para todo conjunto aberto U que contém P_n . Como a medida μ_P é regular (Proposição A.3.2), segue que $\mu_P(P_n) = 1$. Passando ao limite quando $n \rightarrow \infty$, obtemos que $\mu_P(P) = 1$ para todo $P \in \mathcal{P}_*$.

Agora tratamos das duas condições (b) e (c). Por construção (lembre o Lema 5.2.1), dado qualquer $A \in \mathcal{A}$, a função $P \mapsto \mu_P(A) = E(A, P)$ é mensurável e satisfaz

$$\mu(A) = \int E(A, P) d\hat{\mu}(P) = \int \mu_P(A) d\hat{\mu}(P).$$

Afirmamos que a família dos subconjuntos de M para os quais valem estas duas propriedades é uma classe monótona. De fato, suponha que B é a união de uma sequência crescente $(B_j)_j$ de conjuntos para os quais estas são propriedades são válidas. Então, pela Proposição A.1.31

$$P \mapsto \mu_P(B) = \sup_j \mu_P(B_j) \quad \text{é uma função mensurável}$$

e, usando o teorema da convergência monótona,

$$\mu(B) = \lim_n \mu(B_n) = \lim_n \int \mu_P(B_n) d\hat{\mu} = \int \lim_n \mu_P(B_n) d\hat{\mu} = \int \mu_P(B) d\hat{\mu}.$$

Analogamente, se B é a intersecção de uma seqüência decrescente $(B_j)_j$ de conjuntos para os quais as propriedades (b) e (c) são válidas então $P \mapsto \mu_P(B)$ é mensurável e $\mu(B) = \int \mu_P(B) d\hat{\mu}(P)$. Isto implica que as duas propriedades permanecem válidas em toda a classe monótona gerada por \mathcal{A} , ou seja (Teorema A.1.18), em toda a σ -álgebra de Borel de M .

A prova do Teorema 5.1.11 está completa.

5.2.4 Exercícios

5.2.1. Sejam \mathcal{P} e \mathcal{Q} partições mensuráveis de (M, \mathcal{B}, μ) tais que $\mathcal{P} \prec \mathcal{Q}$ a menos de medida nula. Seja $\{\mu_P : P \in \mathcal{P}\}$ uma desintegração de μ relativamente a \mathcal{P} e, para cada $P \in \mathcal{P}$, seja $\{\mu_{P,Q} : Q \in \mathcal{Q}, Q \subset P\}$ uma desintegração de μ_P relativamente a \mathcal{Q} . Seja $\pi : \mathcal{Q} \rightarrow \mathcal{P}$ a projeção canônica, tal que $Q \subset \pi(Q)$ para quase todo $Q \in \mathcal{Q}$. Mostre que $\{\mu_{\pi(Q),Q} : Q \in \mathcal{Q}\}$ é uma desintegração de μ relativamente a \mathcal{Q} .

5.2.2. (Recíproca do teorema de Rokhlin) Seja M um espaço métrico completo separável. Mostre que se \mathcal{P} satisfaz a conclusão do Teorema 5.1.11, isto é, se μ admite uma desintegração relativamente a \mathcal{P} , então a partição \mathcal{P} é mensurável.

5.2.3. Seja $\mathcal{P}_1 \prec \dots \prec \mathcal{P}_n \prec \dots$ uma seqüência crescente de partições enumeráveis tais que a união $\cup_n \mathcal{P}_n$ gera a σ -álgebra \mathcal{B} dos conjuntos mensuráveis a menos de medida nula. Mostre que a esperança condicional $e(\psi) = \lim_n e_n(\psi)$ coincide com ψ em quase todo ponto, para toda função mensurável limitada.

5.2.4. Prove a Proposição 2.4.4, usando a Proposição 5.2.2.

Capítulo 6

Unicidade Ergódica

Este capítulo é dedicado a uma classe especial de sistemas dinâmicos, caracterizados pela propriedade de possuírem exatamente uma probabilidade invariante. Inicialmente, na Seção 6.1, daremos algumas formulações equivalentes desta propriedade e analisaremos as propriedades da única medida invariante.

A relação entre unicidade ergódica e minimalidade é outro tema importante deste capítulo. Um sistema dinâmico diz-se *minimal* se toda órbita é densa no espaço ambiente. Veremos na Seção 6.2 que todo sistema unicamente ergódico é minimal restrito ao suporte da medida invariante, mas a recíproca não é verdadeira em geral.

A principal construção de transformações unicamente ergódicos é de natureza algébrica. Na Seção 6.3 introduziremos a noção de medida de Haar de um grupo topológico. Veremos que toda translação transitiva num grupo topológico compacto metrizável é minimal e mesmo unicamente ergódica: a medida de Haar é a única probabilidade invariante.

Na Seção 6.4 apresentaremos uma aplicação notável da ideia de unicidade ergódica no domínio da Aritmética: o teorema de Hermann Weyl sobre equidistribuição dos valores de funções polinomiais definidas nos números inteiros.

Ao longo do capítulo, a menos de menção em contrário, suporemos que M é um espaço métrico compacto e $f : M \rightarrow M$ é uma transformação contínua.

6.1 Unicidade ergódica

Dizemos que uma transformação $f : M \rightarrow M$ é *unicamente ergódica* se admite exatamente uma medida de probabilidade invariante. Vale uma noção inteiramente análoga para fluxos. A razão de ser da denominação é que a probabilidade invariante μ é necessariamente ergódica. De fato, suponha que existisse $A \subset M$ invariante com $0 < \mu(A) < 1$. Então a restrição normalizada de μ a A , definida por

$$\mu_A(E) = \frac{\mu(E \cap A)}{\mu(A)} \quad \text{para cada conjunto mensurável } E \subset M$$

seria uma probabilidade invariante, distinta de μ , o que estaria em contradição com a unicidade de μ .

Proposição 6.1.1. *As seguintes condições são equivalentes:*

- (a) f admite uma única probabilidade invariante;
- (b) f admite uma única probabilidade ergódica;
- (c) para toda função contínua $\varphi : M \rightarrow \mathbb{R}$, a sequência das médias orbitais $n^{-1} \sum_{j=0}^{n-1} \varphi(f^j(x))$ converge em todo ponto para uma constante.
- (d) para toda função contínua $\varphi : M \rightarrow \mathbb{R}$, a sequência das médias orbitais $n^{-1} \sum_{j=0}^{n-1} \varphi \circ f^j$ converge uniformemente para uma constante;

Demonstração. É fácil ver que (b) implica (a). De fato, como toda medida invariante é uma combinação convexa de medidas ergódicas (Teorema 5.1.3), se existe uma única probabilidade ergódica então a probabilidade invariante é, igualmente, única. É claro que (d) implica (c), uma vez que convergência uniforme implica convergência pontual. Para ver que (c) implica (b), suponha que μ e ν são probabilidades invariantes ergódicas de f . Então, dada qualquer função contínua $\varphi : M \rightarrow \mathbb{R}$,

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

Como, por hipótese, o limite não depende do ponto x , segue que

$$\int \varphi d\mu = \int \varphi d\nu$$

para toda função contínua φ . Usando a Proposição A.3.3, obtemos que $\mu = \nu$.

Resta mostrar que (a) implica (d). Comece por lembrar que f admite alguma probabilidade invariante μ (pelo Teorema 2.1). A ideia é mostrar que se (d) não vale então existe outra probabilidade ν diferente de μ e, portanto, (a) também não vale. Suponha então que (d) não vale, isto é, que existe alguma função contínua $\varphi : M \rightarrow \mathbb{R}$ tal que $n^{-1} \sum_{j=0}^{n-1} \varphi \circ f^j$ não converge uniformemente para nenhuma constante; em particular, não converge uniformemente para $\int \varphi d\mu$. Por definição, isto significa que existe $\varepsilon > 0$ tal que para todo $k \geq 1$ existe $n_k \geq k$ e existe $x_k \in M$ tal que

$$\left| \frac{1}{n_k} \sum_{j=0}^{n_k-1} \varphi(f^j(x_k)) - \int \varphi d\mu \right| \geq \varepsilon. \quad (6.1.1)$$

Consideremos a sequência de probabilidades

$$\nu_k = \frac{1}{n_k} \sum_{j=0}^{n_k-1} \delta_{f^j(x_k)}.$$

Como o espaço $\mathcal{M}_1(M)$ das probabilidades em M é compacto para a topologia fraca* (Teorema 2.1.5), a menos de substituir esta sequência por uma subsequência, podemos supor que ela converge para alguma probabilidade ν em M . Pelo Lema 2.2.4 aplicado à medida de Dirac δ_x , temos que a probabilidade ν é invariante por f . Por outro lado, o fato de que $(\nu_k)_k$ converge para ν na topologia fraca* implica que

$$\int \varphi d\nu = \lim_k \int \varphi d\nu_k = \lim_k \frac{1}{n_k} \sum_{j=0}^{n_k-1} \varphi(f^j(x_k)).$$

Então, lembrando (6.1.1), temos que

$$\left| \int \varphi d\nu - \int \varphi d\mu \right| \geq \varepsilon.$$

Em particular, $\nu \neq \mu$. Isto conclui o argumento. \square

6.1.1 Exercícios

6.1.1. Dê exemplo de uma transformação $f : M \rightarrow M$ num espaço métrico compacto tal que $(1/n) \sum_{j=0}^{n-1} \varphi \circ f^j$ converge uniformemente, qualquer que seja $\varphi : M \rightarrow \mathbb{R}$ contínua, mas f não é unicamente ergódica.

6.1.2. Seja $f : M \rightarrow M$ uma transformação contínua transitiva em um espaço métrico compacto. Mostre que se $(1/n) \sum_{j=0}^{n-1} \varphi \circ f^j$ converge uniformemente, qualquer que seja $\varphi : M \rightarrow \mathbb{R}$ contínua, então f é unicamente ergódica.

6.1.3. Seja $f : M \rightarrow M$ um homeomorfismo isométrico num espaço métrico compacto M . Mostre que se μ é uma medida invariante ergódica para f então, para cada $n \in \mathbb{N}$, a função $\varphi(x) = d(x, f^n(x))$ é constante no suporte de μ .

6.2 Minimalidade

Seja $\Lambda \subset M$ um conjunto invariante fechado de $f : M \rightarrow M$. Dizemos que Λ é *minimal* se ele coincide com o fecho da órbita $\{f^n(x) : n \geq 0\}$ de todo ponto $x \in \Lambda$. Dizemos que a transformação f é minimal se o ambiente M for um conjunto minimal.

Lembre que o suporte de uma medida μ é o conjunto dos pontos $x \in M$ tais que $\mu(V) > 0$ para toda vizinhança V de x . Segue imediatamente da definição que o complementar do suporte é um conjunto aberto: se $x \notin \text{supp } \mu$ então existe uma vizinhança aberta V tal que $\mu(V) = 0$; então V está toda contida no complementar do suporte. Portanto $\text{supp } \mu$ é um conjunto fechado.

Também é fácil ver que o suporte de qualquer medida invariante é um conjunto invariante, no seguinte sentido: $f(\text{supp } \mu) \subset \text{supp } \mu$. De fato, seja $x \in \text{supp } \mu$ e seja V uma vizinhança qualquer de $y = f(x)$. Como f é contínua, $f^{-1}(V)$ é uma vizinhança de x . Então $\mu(f^{-1}(V)) > 0$ uma vez que $x \in \text{supp } \mu$. Logo, usando que μ é invariante, $\mu(V) > 0$. Isto prova que $y \in \text{supp } \mu$.

Proposição 6.2.1. *Se $f : M \rightarrow M$ é unicamente ergódica então o suporte da única probabilidade invariante μ é um conjunto minimal.*

Demonstração. Suponha que existe $x \in \text{supp } \mu$ cuja órbita $\{f^j(x) : j \geq 0\}$ não é densa no suporte de μ . Isto significa que existe algum subconjunto aberto U de M tal que $U \cap \text{supp } \mu$ é não vazio e

$$f^j(x) \notin U \cap \text{supp } \mu \quad \text{para todo } j \geq 0. \quad (6.2.1)$$

Seja ν um ponto de acumulação qualquer da sequência de probabilidades

$$\nu_n = n^{-1} \sum_{j=0}^{n-1} \delta_{f^j(x)}, \quad n \geq 1$$

relativamente à topologia fraca*. Tais pontos de acumulação existem, pelo Teorema 2.1.5, e ν é uma probabilidade invariante, pelo Lema 2.2.4. A condição (6.2.1) significa que $\nu_n(U) = 0$ para todo $n \geq 1$. Logo, usando o Teorema 2.1.2 (veja também o item 3 do Exercício 2.1.1) temos que $\nu(U) = 0$. Isto implica que nenhum ponto de U está no suporte de μ , contradizendo o fato de que $U \cap \text{supp } \mu$ é não vazio. \square

A recíproca da Proposição 6.2.1 é falsa em geral:

Teorema 6.2.2 (Furstenberg). *Existe um difeomorfismo analítico $f : \mathbb{T}^2 \rightarrow \mathbb{T}^2$ que é minimal, preserva a medida de Lebesgue m no toro, mas não é ergódico para m . Em particular, f não é unicamente ergódico.*

No que resta desta seção daremos um esboço sucinto da demonstração deste resultado. A demonstração detalhada pode ser encontrada no artigo original de Furstenberg [Fur61] ou no livro de Mañé [Mañ87]. Na Seção 7.3.1 mencionaremos outros exemplos de transformações minimais que não são unicamente ergódicas.

Para provar o Teorema 6.2.2, buscaremos uma transformação f da forma $f(x, y) = (x + \alpha, y + \phi(x))$ onde α é um número irracional e $\phi : S^1 \rightarrow \mathbb{R}$ é uma função analítica com $\int \phi(x) dx = 0$. Observe que f preserva a medida de Lebesgue de \mathbb{T}^2 . Consideremos também a transformação $f_0 : \mathbb{T}^2 \rightarrow \mathbb{T}^2$ dada por $f_0(x, y) = (x + \alpha, y)$. Note que f_0 não possui órbitas densas em \mathbb{T}^2 e que o sistema (f_0, m) não é ergódico.

Consideremos a equação cohomológica

$$u(x + \alpha) - u(x) \equiv \phi(x). \quad (6.2.2)$$

Se ϕ e α são tais que (6.2.2) admite alguma solução mensurável $u : S^1 \rightarrow \mathbb{R}$ então (f_0, m) e (f, m) são ergodicamente equivalentes (veja o Exercício 6.2.1) e, conseqüentemente, (f, m) não é ergódico. Por outro lado, pode mostrar-se que se (6.2.2) não admite solução contínua então f é minimal (a recíproca deste fato é o Exercício 6.2.2). Portanto, basta que encontremos ϕ e α tais que a equação cohomológica admita solução mensurável mas não solução contínua.

É conveniente expressar estas condições em termos da expansão de Fourier $\phi(x) = \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n x}$. Para garantir que ϕ é analítica é suficiente exigir que:

$$\text{existe } \rho < 1 \text{ tal que } |a_n| \leq \rho^n \text{ para } n \text{ suficientemente grande.} \quad (6.2.3)$$

De fato, nesse caso a série $\sum_{n \in \mathbb{Z}} a_n z^n$ converge uniformemente em toda coroa circular $\{z \in \mathbb{C} : r \leq |z| \leq r^{-1}\}$ com $r > \rho$. Em particular, a sua soma no círculo unitário (a qual coincide com ϕ) é uma função analítica. Como queremos que ϕ tenha média nula e tome valores na reta real, também devemos exigir:

$$a_0 = 0 \quad \text{e} \quad a_{-n} = \bar{a}_n \text{ para todo } n \geq 1. \quad (6.2.4)$$

De acordo com o Exercício 6.2.3, a equação cohomológica admite uma solução no espaço $L^2(m)$ se, e somente se,

$$\sum_{n=1}^{\infty} \left| \frac{a_n}{e^{2\pi i n \alpha} - 1} \right|^2 < \infty. \quad (6.2.5)$$

Além disso, essa solução está unicamente determinada: $u = \sum_{n \in \mathbb{Z}} b_n e^{2\pi i n x}$ com

$$b_n = \frac{a_n}{e^{2\pi i n \alpha} - 1} \quad \text{para todo } n \in \mathbb{Z}. \quad (6.2.6)$$

O teorema de Fejér's (veja [Zyg68]) afirma que se uma função u é contínua então a seqüência das somas parciais da sua expansão de Fourier converge Cesàro uniformemente para u :

$$\frac{1}{n} \sum_{k=1}^n \left(\sum_{j=-k}^k b_j e^{2\pi i j x} \right) \text{ converge uniformemente para } u(x). \quad (6.2.7)$$

Logo, para garantir que u não é contínua é suficiente exigir:

$$\left(\sum_{j=-k}^k b_j \right)_k \text{ não é Cesàro convergente.} \quad (6.2.8)$$

Desta forma, o problema fica reduzido a encontrar α e $(a_n)_n$ satisfazendo (6.2.3), (6.2.4), (6.2.5) e (6.2.8). O Exercício 6.2.4 dá uma ideia das questões envolvidas na escolha destes objetos.

6.2.1 Exercícios

6.2.1. Mostre que se u é solução mensurável da equação cohomológica (6.2.2) então $h : \mathbb{T}^2 \rightarrow \mathbb{T}^2$, $h(x, y) = (x, y + u(x))$ é uma equivalência ergódica entre (f_0, m) e (f, m) . Em particular, (f, m) não pode ser ergódico.

6.2.2. Mostre que se u é solução contínua da equação cohomológica (6.2.2) então $h : \mathbb{T}^2 \rightarrow \mathbb{T}^2$, $h(x, y) = (x, y + u(x))$ é uma conjugação topológica entre f_0 e f . Em particular, f não pode ser transitivo.

6.2.3. Verifique que se $u(x) = \sum_{n \in \mathbb{Z}} b_n e^{2\pi i n x}$ é solução de (6.2.2) então

$$b_n = \frac{a_n}{e^{2\pi i n \alpha} - 1} \quad \text{para todo } n \in \mathbb{Z}. \quad (6.2.9)$$

Além disso, $u \in L^2(m)$ se, e somente se, $\sum_{n=1}^{\infty} |b_n|^2 < \infty$.

6.2.4. Dizemos que um número irracional α é *diofantino* se existem $c > 0$ e $\tau > 0$ tais que $|q\alpha - p| \geq c|q|^{-\tau}$ para quaisquer $p, q \in \mathbb{Z}$ com $q \neq 0$. Mostre que a condição (6.2.5) é satisfeita sempre que α é diofantino e ϕ satisfaz (6.2.3).

6.2.5. (Teorema de Gottschalk) Seja $f : M \rightarrow M$ uma aplicação contínua em um espaço métrico compacto M . Mostre que o fecho da órbita de um ponto $x \in M$ é um conjunto minimal se, e somente se, $R_\varepsilon = \{n \in \mathbb{Z} : d(x, f^n(x)) < \varepsilon\}$ é um conjunto sindético para todo $\varepsilon > 0$.

6.2.6. Seja $f : M \rightarrow M$ uma aplicação contínua em um espaço métrico compacto M . Dizemos que $x, y \in M$ estão *próximos* se $\inf_n d(f^n(x), f^n(y)) = 0$. Mostre que para todo $x \in M$ tal que o fecho da sua órbita é um conjunto minimal, toda vizinhança U de x e todo ponto y próximo de x , existe uma sequência crescente $(n_i)_i$ tal que $f^{n_{i_1} + \dots + n_{i_k}}(x)$ e $f^{n_{i_1} + \dots + n_{i_k}}(y)$ estão em U para quaisquer $i_1 < \dots < i_k$ e $k \geq 1$.

6.2.7. Um teorema de Auslander e Ellis (veja [Fur81, Teorema 8.7]) afirma que nas condições do Exercício 6.2.6 o fecho da órbita de todo $y \in M$ contém algum ponto x que está próximo de y e cujo fecho da órbita é um conjunto minimal. Deduza o seguinte refinamento do teorema de van der Waerden (Teorema de Hindman): dada qualquer decomposição $\mathbb{N} = S_1 \cup \dots \cup S_q$ do conjunto dos números naturais em conjuntos disjuntos dois-a-dois, existe j tal que S_j contém uma sequência $n_1 < \dots < n_i < \dots$ tal que $n_{i_1} + \dots + n_{i_k} \in S_j$ para todo $k \geq 1$ e quaisquer $i_1 < \dots < i_k$.

6.3 Medida de Haar

Nesta seção veremos que grupos topológicos ou diferenciáveis compactos possuem uma probabilidade notável, chamada *medida de Haar*, que é invariante pelas translações e pelos endomorfismos sobrejetivos do grupo. Supondo que o grupo é metrizável, toda translação transitiva é unicamente ergódica, com a medida de Haar como única probabilidade invariante.

6.3.1 Rotações em toros

Fixe $d \geq 1$ e um vetor racionalmente independente $\theta = (\theta_1, \dots, \theta_d)$. Conforme vimos na Seção 4.2.1, a rotação $R_\theta : \mathbb{T}^d \rightarrow \mathbb{T}^d$ é ergódica para a medida de Lebesgue m no toro. Agora o nosso objetivo é mostrar que, na verdade, R_θ é unicamente ergódica.

De acordo com a Proposição 6.1.1, basta mostrar que para toda função contínua $\varphi : \mathbb{T}^d \rightarrow \mathbb{R}$ existe $c_\varphi \in \mathbb{R}$ tal que

$$\varphi_n = \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ R_\theta^j \text{ converge para } c_\varphi \text{ em todo ponto.} \quad (6.3.1)$$

Tomemos $c_\varphi = \int \varphi dm$. Por ergodicidade, a sequência $(\varphi_n)_n$ das médias orbitais converge para c_φ em m -quase todo ponto. Em particular, $\varphi_n(x) \rightarrow c_\varphi$ para um conjunto denso de valores de $x \in \mathbb{T}^d$.

Seja d a distância induzida no toro $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$ pela norma usual em \mathbb{R}^d : a distância entre dois pontos quaisquer do toro é o mínimo das distâncias entre os seus representantes em \mathbb{R}^d . É claro que a rotação R_θ preserva esta distância:

$$d(R_\theta(x), R_\theta(y)) = d(x, y) \quad \text{para todo } x, y \in \mathbb{T}^d.$$

Então, usando que φ é contínua, dado qualquer $\varepsilon > 0$ podemos encontrar $\delta > 0$ tal que

$$d(x, y) < \delta \quad \Rightarrow \quad d(R_\theta^j(x), R_\theta^j(y)) < \delta \quad \Rightarrow \quad |\varphi(R_\theta^j(x)) - \varphi(R_\theta^j(y))| < \varepsilon$$

para todo $j \geq 0$. Mas então,

$$d(x, y) < \delta \quad \Rightarrow \quad |\varphi_n(x) - \varphi_n(y)| < \varepsilon \quad \text{para todo } n \geq 1.$$

Como ε não depende de n , isto mostra que a sequência $(\varphi_n)_n$ é equicontínua.

Isto nos permite usar o Teorema de Ascoli para provar a afirmação (6.3.1), do seguinte modo. Suponha que existe $\bar{x} \in \mathbb{T}^d$ tal que $(\varphi_n(\bar{x}))_n$ não converge para c_φ . Então existe $c \neq c_\varphi$ e alguma subsequência $(n_k)_k$ tal que $\varphi_{n_k}(\bar{x})$ converge para c quando $k \rightarrow \infty$. Pelo teorema de Ascoli, a menos de tomarmos uma subsequência, podemos supor que $(\varphi_{n_k})_k$ é uniformemente convergente. Seja ψ o seu limite. Então ψ é uma função contínua tal que $\psi(x) = c_\varphi$ para um conjunto denso de valores de $x \in \mathbb{T}^d$ mas $\psi(\bar{x}) = c$ é diferente de c_φ . É claro que tal função não existe. Esta contradição prova a nossa afirmação de que R_θ é unicamente ergódica.

6.3.2 Grupos topológicos e grupos de Lie

Lembre que um *grupo topológico* é um grupo (G, \cdot) munido de uma topologia relativamente à qual as operações

$$G \times G \rightarrow G, (g, h) \mapsto gh \quad \text{e} \quad G \rightarrow G, g \mapsto g^{-1} \quad (6.3.2)$$

são contínuas. Em tudo o que segue suporemos que a topologia é tal que todo conjunto pontual é um conjunto fechado. Quando G é uma variedade e as operações em (6.3.2) são diferenciáveis, dizemos que (G, \cdot) é um *grupo de Lie*. Veja o Exercício 6.3.1.

O espaço euclidiano \mathbb{R}^d é um grupo topológico, e até um grupo de Lie, para a adição e o mesmo vale para o toro \mathbb{T}^d . Lembre que \mathbb{T}^d é o quociente de \mathbb{R}^d pelo seu subgrupo \mathbb{Z}^d . Esta construção pode ser generalizada da seguinte forma:

Exemplo 6.3.1. Dado qualquer subgrupo normal fechado H de um grupo topológico G , seja G/H o conjunto das classes de equivalência para a relação de equivalência definida em G por $x \sim y \Leftrightarrow x^{-1}y \in H$. Represente por xH a classe de equivalência que contém cada $x \in G$. Considere a seguinte operação de grupo em G/H :

$$xH \cdot yH = (x \cdot y)H.$$

A hipótese de que H é subgrupo normal assegura que esta operação está bem definida. Seja $\pi : G \mapsto G/H$ a projeção canônica, dada por $\pi(x) = xH$. Considere em G/H a topologia quociente, definida da seguinte forma: uma função $\psi : G/H \rightarrow X$ é contínua se, e somente se, $\psi \circ \pi : G \rightarrow X$ é contínua. A hipótese de que H é fechada assegura que todo subconjunto pontual é um subconjunto fechado de G/H . Segue facilmente das definições que G/H é um grupo topológico. Lembre também que se G é abeliano então todos os seus subgrupos são normais.

Exemplo 6.3.2. O conjunto $G = \text{GL}(d, \mathbb{R})$ das matrizes reais invertíveis de dimensão d é um grupo de Lie para a operação de multiplicação de matrizes, chamado *grupo linear real* de dimensão d . De fato, G pode ser identificado com um aberto do espaço euclidiano $\mathbb{R}^{(d^2)}$ e portanto tem uma estrutura natural de variedade. Além disso, segue diretamente das definições que a multiplicação de matrizes e a aplicação $A \mapsto A^{-1}$ são diferenciáveis para esta estrutura diferenciável. G contém diversos subgrupos de Lie importantes, tais como o *grupo especial linear* $\text{SL}(d, \mathbb{R})$ das matrizes com determinante igual a 1, e o *grupo ortogonal* $\text{O}(d, \mathbb{R})$ formado pelas matrizes ortogonais.

Dado qualquer $g \in G$, chamamos *g-translação à esquerda* e *g-translação à direita*, respectivamente, as aplicações

$$E_g : G \rightarrow G, E_g(h) = gh \quad \text{e} \quad D_g : G \rightarrow G, D_g(h) = hg.$$

Um *endomorfismo* de G é uma aplicação contínua $\phi : G \rightarrow G$ que também preserva a operação de grupo, isto é, tal que $\phi(gh) = \phi(g)\phi(h)$. Quando ϕ é invertível, ou seja, uma bijeção cuja inversa também é um endomorfismo, dizemos que se trata de um *automorfismo*.

Exemplo 6.3.3. Seja $A \in \text{GL}(d, \mathbb{Z})$, isto é, uma matriz invertível de dimensão d com coeficientes inteiros. Então, como vimos na Seção 4.2.5, A induz um endomorfismo $f_A : \mathbb{T}^d \rightarrow \mathbb{T}^d$. Pode mostrar-se que todo endomorfismo do toro é desta forma.

Um grupo topológico é *localmente compacto* se todo $g \in G$ possui alguma vizinhança compacta. Por exemplo, todo grupo de Lie é localmente compacto. Por outro lado, o grupo aditivo dos racionais, considerado com a topologia induzida pela reta, não é localmente compacto.

O seguinte teorema é o ponto de partida da teoria ergódica dos grupos topológicos localmente compactos:

Teorema 6.3.4 (Haar). *Seja G um grupo localmente compacto.*

- (a) Existe uma medida boreliana μ_G em G que é invariante pelas translações à esquerda, finita em conjuntos compactos e positiva em conjuntos abertos;
- (b) Se η é uma medida invariante pelas translações à esquerda e finita em conjuntos compactos então $\eta = c\mu_G$ para algum $c > 0$.
- (c) $\mu_G(G) < \infty$ se, e somente se, G é compacto.

Vamos esboçar a demonstração dos itens (a) e (b) no caso particular em que G é um grupo de Lie. Neste caso, como veremos μ_G é uma medida de volume em G . A demonstração do item (c), no caso geral, é proposta no Exercício 6.3.4.

Começando pelo item (a), seja e o elemento neutro e seja $d \geq 1$ a dimensão do grupo de Lie. Considere um produto interno qualquer \cdot no espaço tangente T_eG . Para cada $g \in G$, representemos por $\mathcal{E}_g : T_eG \rightarrow T_gG$ a derivada da translação à esquerda E_g no ponto e . Em seguida, consideremos o produto interno definido em T_gG do seguinte modo:

$$u \cdot v = \mathcal{E}_g^{-1}(u) \cdot \mathcal{E}_g^{-1}(v) \quad \text{para todo } u, v \in T_gG.$$

É claro que este produto interno depende diferenciavelmente de g . Portanto, ele define uma métrica riemanniana em G . Também é claro da construção que esta métrica é invariante por translações à esquerda: notando que $\mathcal{E}_{hg} = DE_h(g)\mathcal{E}_g$, vemos que

$$\begin{aligned} DE_h(g)(u) \cdot DE_h(g)(v) &= \mathcal{E}_{hg}^{-1}DE_h(g)(u) \cdot \mathcal{E}_{hg}^{-1}DE_h(g)(v) \\ &= \mathcal{E}_g^{-1}(u) \cdot \mathcal{E}_g^{-1}(v) = u \cdot v \end{aligned}$$

para quaisquer $g, h \in G$ e $u, v \in T_gG$. Seja μ_G a medida de volume associada a esta métrica riemanniana. Esta medida pode ser caracterizada do seguinte modo. Dadas coordenadas locais $x = (x_1, \dots, x_d)$ em G , considere

$$\rho(x) = \det \begin{pmatrix} g_{1,1}(x) & \cdots & g_{1,d}(x) \\ \cdots & \cdots & \cdots \\ g_{d,1}(x) & \cdots & g_{d,d}(x) \end{pmatrix} \quad \text{onde } g_{i,j} = \frac{\partial}{\partial x_i} \cdot \frac{\partial}{\partial x_j}.$$

Então $\mu_G(B) = \int_B |\rho(x)| dx_1 \cdots dx_d$, para qualquer conjunto mensurável B contido no domínio das coordenadas locais. Notando que a função ρ é contínua e não nula, para toda carta local, segue que μ_G é positiva em abertos e finita em compactos. Além disso, como a métrica riemanniana é invariante por translações à esquerda, a medida μ_G também é invariante por translações à esquerda.

Agora passamos ao item (b) do Teorema 6.3.4. Seja ν qualquer medida nas condições do enunciado. Representamos por $B(g, r)$ a bola aberta de centro g e raio r , para a distância associada à métrica riemanniana. Em outras palavras, $B(g, r)$ é o conjunto dos pontos em G que podem ser ligados a g por alguma curva de comprimento menor que r . Fixe $\rho > 0$ tal que $\nu(B(e, \rho))$ é finita (tal ρ existe porque G é localmente compacto e ν é finita em compactos). Afirmamos que

$$\limsup_{r \rightarrow 0} \frac{\nu(B(g, r))}{\mu_G(B(g, r))} \leq \frac{\nu(B(e, \rho))}{\mu_G(B(e, \rho))} \quad (6.3.3)$$

para todo $g \in G$. Isto pode ser visto da seguinte forma. Observe que o lado esquerdo da desigualdade não depende de g , uma vez que tanto as duas medidas quanto a distância são invariantes por translações à esquerda. Portanto, basta considerar o caso $g = e$. Seja $(r_n)_n$ qualquer sequência convergindo para zero e tal que:

$$\lim_n \frac{\nu(B(e, r_n))}{\mu_G(B(e, r_n))} = \limsup_{r \rightarrow 0} \frac{\nu(B(e, r))}{\mu_G(B(e, r))}. \quad (6.3.4)$$

Pelo lema de Vitali (Teorema A.2.16), podemos encontrar $(g_j)_j$ em $B(e, \rho)$ e $(n_j)_j$ em \mathbb{N} tais que

1. as bolas $B(g_j, r_{n_j})$ estão contidas em $B(e, \rho)$ e são disjuntas dois-a-dois;
2. a união dessas bolas tem μ_G -medida total em $B(e, \rho)$.

Além disso, dado qualquer $a \in \mathbb{R}$ menor que o limite em (6.3.4), podemos supor que os n_j são suficientemente grandes para que $\nu(B(g_j, r_{n_j})) \geq a\mu_G(B(g_j, r_{n_j}))$ para todo j . Segue que

$$\nu(B(e, \rho)) \geq \sum_j \nu(B(g_j, r_{n_j})) \geq \sum_j a\mu_G(B(g_j, r_{n_j})) = a\mu_G(B(e, \rho)).$$

Como a pode ser tomado arbitrariamente próximo de (6.3.4), isto prova a afirmação (6.3.3).

Em seguida, afirmamos que ν é absolutamente contínua com relação a μ_G . De fato, seja b qualquer número maior que o quociente no lado direito de (6.3.3). Dado qualquer conjunto mensurável $B \subset G$ com $\mu_G(B) = 0$, e dado qualquer $\varepsilon > 0$, seja $\{B(g_j, r_j) : j\}$ uma cobertura de B por bolas de raio pequeno, tal que $\nu(B(g_j, r_j)) \leq b\mu_G(B(g_j, r_j))$ e $\sum_j \mu_G(B(g_j, r_j)) \leq \varepsilon$. Então,

$$\nu(B) \leq \sum_j \nu(B(g_j, r_j)) \leq b \sum_j \mu_G(B(g_j, r_j)) \leq b\varepsilon.$$

Como $\varepsilon > 0$ é arbitrário, segue que $\nu(B) = 0$. Portanto, $\nu \ll \mu_G$, conforme foi afirmado. Agora, pelo teorema de derivação de Lebesgue (Teorema A.2.15)

$$\frac{d\nu}{d\mu_G}(g) = \lim_{r \rightarrow 0} \frac{1}{\mu_G(B(g, r))} \int_{B(g, r)} \frac{d\nu}{d\mu_G} d\mu_G = \lim_{r \rightarrow 0} \frac{\nu(B(g, r))}{\mu_G(B(g, r))}$$

para μ -quase todo $g \in G$. O limite no lado esquerdo não depende de g e, pela afirmação (6.3.3), ele é finito. Seja $c \in \mathbb{R}$ esse limite. Então $\nu = c\mu_G$, conforme afirmado no item (b) do Teorema 6.3.4.

No caso em que o grupo G é compacto, segue do Teorema 6.3.4 que existe uma única probabilidade que é invariante pelas translações à esquerda, positiva em abertos e finita em compactos. Esta probabilidade μ_G é chamada *medida de Haar*. Ela disfruta de algumas propriedades adicionais: Por exemplo, a medida de Lebesgue normalizada é a medida de Haar no toro \mathbb{T}^d . Veja também os Exercícios 6.3.5 e 6.3.6.

Corolário 6.3.5. *Suponha que G é compacto. Então a medida de Haar μ_G é invariante pelas translações à direita e por todo endomorfismo sobrejetivo de G .*

Demonstração. Dado qualquer $g \in G$, considere a probabilidade $(D_g)_*\mu_G$. Observe que $E_h \circ D_g = D_g \circ E_h$ para todo $h \in G$. Logo,

$$(E_h)_*(D_g)_*\mu_G = (D_g)_*(E_h)_*\mu_G = (D_g)_*\mu_G.$$

Em outras palavras, $(D_g)_*\mu_G$ é invariante por toda a translação à esquerda. Por unicidade, segue que $(D_g)_*\mu_G = \mu_G$ para todo $g \in G$, tal como afirmado.

Dado qualquer homomorfismo sobrejetivo $\phi : G \rightarrow G$, considere a probabilidade $\phi_*\mu_G$. Dado qualquer $h \in G$, escolha algum $g \in \phi^{-1}(h)$. Observe que $E_h \circ \phi = \phi \circ E_g$. Logo,

$$(E_h)_*\phi_*\mu_G = \phi_*(E_g)_*\mu_G = \phi_*\mu_G.$$

Em outras palavras, $\phi_*\mu_G$ é invariante por toda a translação à esquerda. Por unicidade, segue que $\phi_*\mu_G = \mu_G$, tal como afirmado. \square

Mais geralmente, se não suposermos, que G é compacto, o argumento do Corolário 6.3.5 mostra que para cada $g \in G$ existe $\lambda(g) > 0$ tal que

$$(E_g)_*\mu_G = \lambda(g)\mu_G.$$

A aplicação $G \rightarrow (0, \infty)$, $g \mapsto \lambda(g)$ é um homomorfismo de grupo.

6.3.3 Translações em grupos compactos metrizáveis

Nesta seção sempre suporemos que o grupo topológico é compacto e metrizável. Começamos por observar que é sempre possível escolher a distância de tal forma que ela seja invariante por toda translação:

Lema 6.3.6. *Se G é um grupo topológico compacto metrizável então existe alguma distância compatível com a topologia de G relativamente à qual todas as translações, tanto à esquerda quanto à direita, são isometrias.*

Demonstração. Seja $(U_n)_n$ uma base de vizinhanças do elemento neutro 1 de G . Pelo Lema A.3.4, para cada n existe uma função contínua $\varphi_n : G \rightarrow [0, 1]$ tal que $\varphi_n(1) = 0$ e $\varphi_n(z) = 1$ para todo $z \in G \setminus U_n$. Defina

$$\varphi : G \rightarrow [0, 1], \quad \varphi(z) = \sum_{n=1}^{\infty} 2^{-n} \varphi_n(z).$$

Então, φ é contínua e $\varphi(1) = 0 < \varphi(z)$ para todo $z \neq 1$. Agora defina

$$d(x, y) = \sup\{|\varphi(gxh) - \varphi(gyh)| : (g, h) \in G^2\} \quad (6.3.5)$$

para cada $x, y \in G$. O supremo é finito, por compacidade de G . É fácil ver que d é uma distância em G . De fato, note que se $d(x, y) = 0$ então $\varphi(gxh) = \varphi(gyh)$

para todo $g, h \in G$. Em particular, tomando $g = 1$ e $h = y^{-1}$, vem que $\varphi(xy^{-1}) = \varphi(1)$. Pela construção de φ , isto implica que $x = y$. Os demais axiomas da noção de distância seguem diretamente da definição de d . Também é claro da definição que d é invariante por translações à esquerda e à direita.

Resta mostrar que a distância d é compatível com a topologia do grupo G . É fácil verificar que, dada qualquer vizinhança V de um ponto $x \in G$, existe $\delta > 0$ tal que $B(x, \delta) \subset V$. De fato, como $U = x^{-1}V$ é uma vizinhança de $1 \in G$, as propriedades de φ garantem que existe $\delta > 0$ tal que $\varphi(z) \leq 1 - \delta$ para todo $z \notin U$. Então, $y \notin V$ implica que $\varphi(x^{-1}y) \leq 1 - \delta$ ou, em outras palavras, que $|\varphi(1) - \varphi(x^{-1}y)| \geq \delta$. Tomando $g = x^{-1}$ e $h = 1$ na definição (6.3.5), vemos que esta última desigualdade implica que $d(x, y) \geq \delta$, ou seja, $y \notin B(x, \delta)$. Agora, verifiquemos a recíproca: dados $x \in G$ e $\delta > 0$, existe alguma vizinhança V de x contida em $B(x, \delta)$. Por continuidade, para cada par $(g, h) \in G^2$ existe uma vizinhança aberta $U \times V \times W$ de (g, x, h) em G^3 tal que

$$|\varphi(gxh) - \varphi(g'x'h')| \leq \delta/2 \quad \text{para todo } (g', x', h') \in U \times V \times W. \quad (6.3.6)$$

Os conjuntos $U \times W$ assim obtidos, com x fixado e g, h variáveis, formam uma cobertura aberta de G^2 . Seja $U_i \times W_i$, $i = 1, \dots, k$ uma subcobertura finita e sejam V_i , $i = 1, \dots, k$ as vizinhanças de x correspondentes. Tome $V = \bigcap_{i=1}^k V_i$. Considere $y \in V$. Dado qualquer $(g, h) \in G^2$, a condição (6.3.6) implica que $|\varphi(gxh) - \varphi(gyh)| \leq \delta/2$. Logo $d(x, y) \leq \delta/2$ e, portanto, $y \in B(x, \delta)$. \square

Exemplo 6.3.7. Dada uma matriz $A \in \text{GL}(d, \mathbb{R})$, represente por $\|A\|$ a sua norma como operador, ou seja, $\|A\| = \sup\{\|Av\| : \|v\| = 1\}$. Observe que $\|OA\| = \|A\| = \|AO\|$ para todo O no grupo ortogonal $O(d, \mathbb{R})$. Defina

$$d(A, B) = \log(1 + \|A^{-1}B - \text{id}\| + \|B^{-1}A - \text{id}\|).$$

Então d é uma distância em $\text{GL}(d, \mathbb{R})$, invariante por translações à esquerda:

$$d(CA, CB) = \log(1 + \|A^{-1}C^{-1}CB - \text{id}\| + \|B^{-1}C^{-1}CA - \text{id}\|) = d(A, B)$$

para todo $C \in \text{GL}(d, \mathbb{R})$. Esta distância não é invariante à direita em $\text{GL}(d, \mathbb{R})$ (Exercício 6.3.3). No entanto, ela é invariante à direita (e à esquerda) se restringirmos ao grupo ortogonal $O(d, \mathbb{R})$: para todo $O \in O(d, \mathbb{R})$,

$$\begin{aligned} d(AO, CO) &= \log(1 + \|O^{-1}A^{-1}BO - \text{id}\| + \|O^{-1}B^{-1}AO - \text{id}\|) \\ &= \log(1 + \|O^{-1}(A^{-1}B - \text{id})O\| + \|O^{-1}(B^{-1}A - \text{id})O\|) \\ &= d(A, B). \end{aligned}$$

Teorema 6.3.8. *Seja G um grupo topológico compacto metrizável e seja $g \in G$. Então são equivalentes:*

- (a) E_g é unicamente ergódica;
- (b) E_g é ergódica com respeito a μ_G ;
- (c) o subgrupo $\{g^n : n \in \mathbb{Z}\}$ gerado por g é denso em G ;

Demonstração. É claro que (a) implica (b). Para provar que (b) implica (c), considere a distância invariante d dada pelo Lema 6.3.6. Seja H o fecho do $\{g^n : n \in \mathbb{Z}\}$ e considere a função contínua

$$\varphi(x) = \min\{d(x, y); y \in H\}.$$

Observe que esta função é invariante para E_g : usando que $gH = H$, obtemos:

$$\begin{aligned}\varphi(x) &= \min\{d(x, y) : y \in H\} = \min\{d(gx, gy) : y \in H\} \\ &= \min\{d(gx, z) : z \in H\} = \varphi(gx) \quad \text{para todo } x \in G.\end{aligned}$$

Como H é fechado, $\varphi(x) = 0$ se, e somente se, $x \in H$. Se $H \neq G$ então $\mu_G(H \neq G) > 0$, já que a medida de Haar é positiva em abertos. Nesse caso, a função φ não é constante em μ_G -quase todo ponto e, portanto, E_g não pode ser ergódica para μ_G .

Finalmente, para mostrar que (c) implica (a), vamos mostrar que se μ é uma probabilidade invariante por E_g , então $\mu = \mu_G$. Para ver isso, basta verificar que μ é invariante para toda translação à esquerda em G . Fixe $h \in G$. Pela invariância de μ , temos

$$\int \varphi(x) d\mu(x) = \int \varphi(g^n x) d\mu(x)$$

para todo $n \in \mathbb{N}$ e toda função contínua $\varphi : G \rightarrow \mathbb{R}$. Por outro lado, a hipótese garante que existe uma sequência de números naturais $n_j \rightarrow \infty$ tal que $g^{n_j} \rightarrow h$. Dada qualquer função contínua (logo, uniformemente contínua) $\varphi : G \rightarrow \mathbb{R}$ e dado $\varepsilon > 0$, fixe $\delta > 0$ tal que $|\varphi(x) - \varphi(y)| < \varepsilon$ sempre que $d(x, y) < \delta$. Se j é suficientemente grande,

$$d(g^{n_j} x, hx) = d(g^{n_j}, h) < \delta \quad \text{para todo } x \in G.$$

Logo, $|\varphi(g^{n_j} x) - \varphi(hx)| < \varepsilon$ para todo $x \in G$, portanto,

$$\left| \int (\varphi(x) - \varphi(hx)) d\mu \right| = \left| \int (\varphi(g^{n_j} x) - \varphi(hx)) d\mu \right| < \varepsilon.$$

Como ε é arbitrário, segue que $\int \varphi d\mu = \int \varphi \circ E_h d\mu$ para toda função contínua φ e todo $h \in G$. Isto implica que μ é invariante por E_h para todo $h \in G$, conforme afirmado. \square

6.3.4 Odômetros

Os odômetros, ou máquinas de somar, modelam sistemas tais como o contador de quilômetros percorridos de um automóvel, ou o registro de consumo de eletricidade de um prédio: a sua dinâmica consiste em fazer avançar o contador de uma unidade. A principal diferença com relação à realidade é que estes contadores idealizados comportam um número ilimitado de dígitos.

Fixe $d \geq 2$, que representa a base de numeração (por exemplo, $d = 10$). Seja $X = \{0, 1, \dots, d-1\}$, munido com a topologia discreta. Considere o conjunto

$M = X^{\mathbb{N}}$ de todas as sequências $\alpha = (\alpha_n)_n$ com valores em X , munido com a topologia produto. Esta topologia é metrizável: ela é compatível, por exemplo, com distância definida em M por

$$d(\alpha, \alpha') = 2^{-N(\alpha, \alpha')} \quad \text{onde} \quad N(\alpha, \alpha') = \min\{j \geq 0 : \alpha_j \neq \alpha'_j\} \quad (6.3.7)$$

Observe também que M é compacto, pois é um produto de compactos (teorema de Tychonoff).

Introduzimos em M a seguinte operação de “soma com transporte”: dados $\alpha = (\alpha_n)_n$ e $\beta = (\beta_n)_n$ em M , definimos $\alpha + \beta = (\gamma_n)_n$ da seguinte forma. Primeiramente,

- se $\alpha_0 + \beta_0 < d$ então $\gamma_0 = \alpha_0 + \beta_0$ e $\delta_1 = 0$;
- se $\alpha_0 + \beta_0 \geq d$ então $\gamma_0 = \alpha_0 + \beta_0 - d$ e $\delta_1 = 1$.

Em seguida, para cada $n \geq 1$,

- se $\alpha_n + \beta_n + \delta_n < d$ então $\gamma_n = \alpha_n + \beta_n + \delta_n$ e $\delta_{n+1} = 0$;
- se $\alpha_n + \beta_n + \delta_n \geq d$ então $\gamma_n = \alpha_n + \beta_n + \delta_n - d$ e $\delta_{n+1} = 1$.

A sequência auxiliar $(\delta_n)_n$ corresponde, precisamente, aos transportes. A aplicação $+ : M \times M \rightarrow M$ definida deste modo torna M um grupo topológico abeliano e a distância (6.3.7) é invariante por translações (Exercício 6.3.8).

Agora considere a “translação de 1” $f : M \rightarrow M$ definida por

$$f((\alpha_n)_n) = (\alpha_n)_n + (1, 0, \dots, 0, \dots) = (0, \dots, 0, \alpha_k + 1, \alpha_{k+1}, \dots, \alpha_n, \dots)$$

onde $k \geq 0$ é o menor valor de n tal que $\alpha_n < d - 1$; se não existe tal k , ou seja, se $(\alpha_n)_n$ é a sequência constante igual a $d - 1$, então a imagem $f((\alpha_n)_n)$ é a sequência constante igual a 0. Deixamos ao cuidado do leitor verificar que esta transformação $f : M \rightarrow M$ é unicamente ergódica (Exercício 6.3.9).

É possível generalizar esta construção um pouco mais, da seguinte forma. Tome $M = \prod_{n=0}^{\infty} \{0, 1, \dots, d_n - 1\}$, onde $(d_n)_n$ é uma sequência qualquer de números inteiros maiores que 1. Tal como no caso particular acima, este conjunto tem uma estrutura de grupo compacto abeliano metrizável e a “translação de 1” é unicamente ergódica.

Exemplo 6.3.9. Uma *pilha (simples)* num intervalo ¹ I é uma família ordenada \mathcal{S} de subintervalos I_0, \dots, I_{k-1} , com o mesmo comprimento, disjuntos dois-a-dois e cuja união é I . Escrevemos $I_k = I_0$. Associamos a \mathcal{S} a transformação $f : I \rightarrow I$ cuja restrição a cada I_j é a translação que envia I_j em I_{j+1} . Graficamente, representamos os subintervalos “empilhados” uns sobre os outros, com a *base* I_0 em baixo e o *topo* I_{k-1} em cima. Então f não é mais que a translação “para cima”, exceto no topo da pilha. Veja a Figura 6.1.

Consideremos uma sequência $(\mathcal{S}_n)_n$ de pilhas num mesmo intervalo I , construída da seguinte forma. Fixe um inteiro $d \geq 2$. Tome $\mathcal{S}_0 = \{I\}$. Para cada

¹Aqui consideramos todos os intervalos fechados à esquerda e abertos à direita.

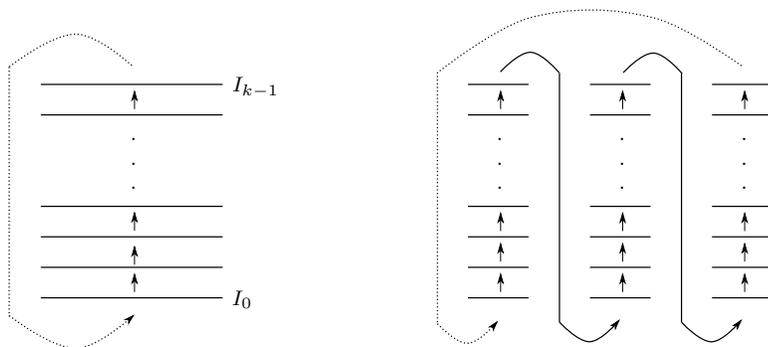


Figura 6.1: Exemplo do método de empilhamento

$n \geq 1$, tome para \mathcal{S}_n a pilha obtida dividindo \mathcal{S}_{n-1} em d colunas, todas com a mesma largura, e empilhando essas colunas umas sobre as outras. Este procedimento está descrito na Figura 6.1 para $d = 3$. Seja $f_n : I \rightarrow I$ a transformação associada a cada \mathcal{S}_n . Propomos ao leitor mostrar (Exercício 6.3.10) que a sequência $(f_n)_n$ converge em todo ponto para uma transformação $f : I \rightarrow I$, a qual preserva a medida de Lebesgue. Além disso, f é unicamente ergódica.

Esta é apenas uma aplicação simples do chamado *método de empilhamento*, que constitui uma ferramenta muito eficaz para produzir exemplos com propriedades interessantes. O leitor pode encontrar uma discussão detalhada deste método na Seção 6 do livro de Friedman [Fri69]. Outra aplicação, um pouco mais elaborada, será dada no Exemplo 8.2.3.

Exemplo 6.3.10 (Substituições). Vamos mencionar brevemente uma construção de natureza combinatória que generaliza a definição do odômetro, proporcionando muitos outros exemplos interessantes de sistemas minimais e unicamente ergódicos. Para maior informação, inclusive sobre as relações entres estes sistemas e o odômetro, recomendamos ao leitor o livro de Queffélec [Que87] e o artigo de Ferenczi, Fisher, Talet [FFT09].

Chamaremos *substituição* em um alfabeto finito \mathcal{A} a qualquer aplicação que associa a cada letra α do alfabeto uma palavra $s(\alpha)$ formada por um número finito de letras de \mathcal{A} . Alguns exemplos, no caso $\mathcal{A} = \{0, 1\}$: *substituição de Thue-Morse* $s(0) = 01$ e $s(1) = 10$; *substituição de Fibonacci* $s(0) = 01$ e $s(1) = 0$; *substituição de Feigenbaum* $s(0) = 11$ e $s(1) = 10$; *substituição de Cantor* $s(0) = 010$ e $s(1) = 111$; e *substituição de Chacon* $s(0) = 0010$ e $s(1) = 1$. Podemos iterar uma substituição, definindo $s^1(\alpha) = s(\alpha)$ e

$$s^{k+1}(\alpha) = s(\alpha_1) \cdots s(\alpha_n) \quad \text{se} \quad s^k(\alpha) = \alpha_1 \cdots \alpha_n.$$

Diremos que a substituição s é *primitiva* (ou *aperiódica*) se existe $k \geq 1$ tal que para quaisquer $\alpha, \beta \in \mathcal{A}$ a palavra $s^k(\alpha)$ contém a letra β .

Considere \mathcal{A} munido da topologia discreta e seja $\Sigma = \mathcal{A}^{\mathbb{N}}$ o espaço das seqüências em \mathcal{A} , munido da topologia produto. Representemos por $S : \Sigma \rightarrow \Sigma$

a aplicação induzida nesse espaço por uma dada substituição s : a imagem de cada $(a_0, \dots, a_n, \dots) \in \Sigma$ é a sequência das letras da palavra obtida quando concatenamos as palavras finitas $s(a_n)$. Suponha que existe alguma letra $\alpha_0 \in \mathcal{A}$ tal que a palavra $s(\alpha_0)$ tem comprimento maior que um e começa com a letra α_0 . Esse é o caso de todos os exemplos listados acima. Então (Exercício 6.3.11), S admite um único ponto fixo $x = (x_n)_n$ com $x_0 = \alpha_0$.

Considere a restrição $\sigma : \mathcal{X} \rightarrow \mathcal{X}$ do deslocamento $\sigma : \Sigma \rightarrow \Sigma$ ao fecho $\mathcal{X} \subset \Sigma$ da órbita $\{\sigma^n(x) : n \geq 0\}$ do ponto x . Se a substituição s é primitiva então $\sigma : \mathcal{X} \rightarrow \mathcal{X}$ é minimal e unicamente ergódico (veja a Seção 5 em [Que87]). Isso vale, por exemplo, no caso das sequências de Thue-Morse, Fibonacci e Feigenbaum.

6.3.5 Exercícios

6.3.1. Seja G uma variedade e seja \cdot uma operação de grupo em G tal que $(g, h) \mapsto g \cdot h$ é de classe C^1 . Mostre que $g \mapsto g^{-1}$ também é de classe C^1 .

6.3.2. Seja G um espaço topológico compacto e seja \cdot uma operação de grupo em G tal que a aplicação $(g, h) \mapsto g \cdot h$ é contínua. Mostre que $g \mapsto g^{-1}$ também é contínua.

6.3.3. Mostre que a distância d no Exemplo 6.3.7 não é invariante à direita.

6.3.4. Prove o item (c) do Teorema 6.3.4: um grupo localmente compacto G é compacto se, e somente se, a sua medida de Haar é finita.

6.3.5. Identifique $\text{GL}(1, \mathbb{R})$ com o grupo multiplicativo $\mathbb{R} \setminus \{0\}$. Verifique que a medida μ definida por

$$\int_{\text{GL}(1, \mathbb{R})} \varphi d\mu = \int_{\mathbb{R} \setminus \{0\}} \frac{\varphi(x)}{|x|} dx$$

é invariante pelas translações de $\text{GL}(1, \mathbb{R})$. Encontre uma medida invariante pelas translações de $\text{GL}(1, \mathbb{C})$.

6.3.6. Identifique $\text{GL}(2, \mathbb{R})$ com $\{(a_{11}, a_{12}, a_{21}, a_{22}) \in \mathbb{R}^4 : a_{11}a_{22} - a_{12}a_{21} \neq 0\}$, de tal modo que $\det(a_{11}, a_{12}, a_{21}, a_{22}) = a_{11}a_{22} - a_{12}a_{21}$. Mostre que a medida μ definida por

$$\int_{\text{GL}(2, \mathbb{R})} \varphi d\mu = \int \frac{\varphi(x_{11}, x_{12}, x_{21}, x_{22})}{|\det(x_{11}, x_{12}, x_{21}, x_{22})|^2} dx_{11} dx_{12} dx_{21} dx_{22}$$

é invariante pelas translações à esquerda e à direita de $\text{GL}(2, \mathbb{R})$. Encontre uma medida invariante pelas translações de $\text{GL}(2, \mathbb{C})$.

6.3.7. Seja G um grupo compacto metrizável e seja $g \in G$. Verifique que são equivalentes:

- (1) E_g é unicamente ergódica;

(2) E_g é transitiva: existe $x \in G$ tal que $\{g^n x : n \in \mathbb{Z}\}$ é densa em G ;

(3) E_g é minimal: $\{g^n y : n \in \mathbb{Z}\}$ é densa em G para todo $y \in G$.

6.3.8. Mostre que a operação $+ : M \times M \rightarrow M$ definida na Seção 6.3.4 é contínua e mune M com uma estrutura de grupo abeliano. Além disso, toda translação neste grupo preserva a distância definida em (6.3.7).

6.3.9. Seja $f : M \rightarrow M$ um odômetro, tal como foi definido na Seção 6.3.4. Dados b_0, \dots, b_{k-1} em $\{0, \dots, 9\}$, represente por $[b_0, \dots, b_{k-1}]$ o conjunto das seqüências $\beta \in M$ com $\beta_0 = b_0, \dots, \beta_{k-1} = b_{k-1}$. Mostre que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \#\{0 \leq j < n : f^j(x) \in [b_0, \dots, b_{k-1}]\} = \frac{1}{10^k}$$

para todo $x \in M$. Além disso, este limite é uniforme. Conclua que f admite uma única probabilidade invariante e calcule essa probabilidade explicitamente.

6.3.10. Verifique as afirmações no Exemplo 6.3.9.

6.3.11. Prove que se s é uma substituição num alfabeto finito \mathcal{A} e $\alpha \in \mathcal{A}$ é tal que $s(\alpha)$ tem comprimento maior que um e começa com a letra α , então a transformação $S : \Sigma \rightarrow \Sigma$ definida no Exemplo 6.3.10 admite um único ponto fixo que começa com a letra $\alpha \in \mathcal{A}$.

6.4 Teorema de Weyl

Nesta seção vamos utilizar ideias discutidas anteriormente para provar um belo teorema de Hermann Weyl [Wey16] sobre a distribuição dos valores de funções polinomiais restritas aos números inteiros.

Considere qualquer função polinomial $P : \mathbb{R} \rightarrow \mathbb{R}$ com coeficientes reais e grau $d \geq 1$:

$$P(x) = a_0 + a_1x + a_2x^2 + \dots + a_dx^d.$$

Compondo P com a projeção canônica $\mathbb{R} \rightarrow S^1$, obtemos uma função polinomial $P_* : \mathbb{R} \rightarrow S^1$ com valores no círculo $S^1 = \mathbb{R}/\mathbb{Z}$. Defina:

$$z_n = P_*(n), \quad \text{para cada } n \geq 1.$$

Podemos pensar em z_n como sendo a parte fracionária do número real $P(n)$. Estamos interessados em entender como a seqüência $(z_n)_n$ se distribui no círculo.

Definição 6.4.1. Dizemos que uma seqüência $(x_n)_n$ em S^1 é *equidistribuída* se para qualquer função contínua $\varphi : S^1 \rightarrow \mathbb{R}$ tem-se

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \varphi(x_j) = \int \varphi(x) dx.$$

De acordo com o Exercício 6.4.1, isto equivale a dizer que, para todo segmento $I \subset S^1$, a fração dos termos da sequência que estão em I é igual ao comprimento $m(I)$ do segmento.

Teorema 6.4.2 (Weyl). *Se algum dos coeficientes a_1, a_2, \dots, a_d é irracional então a sequência $z_n = P_*(n)$, $n \in \mathbb{N}$ é equidistribuída.*

Para desenvolvermos a nossa intuição sobre este teorema, comecemos por considerar o caso especial $d = 1$. Neste caso a função polinomial resume-se a $P(x) = a_0 + a_1x$. Consideremos a transformação

$$f : S^1 \rightarrow S^1, \quad f(\theta) = \theta + a_1.$$

Por hipótese, o coeficiente a_1 é irracional. Portanto, como vimos na Seção 6.3.1, esta transformação admite uma única probabilidade invariante, que é a medida de Lebesgue m . Consequentemente, dada qualquer função contínua $\varphi : S^1 \rightarrow \mathbb{R}$ e dado qualquer ponto $\theta \in S^1$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \varphi(f^j(\theta)) = \int \varphi dm.$$

Considere $\theta = a_0$. Então, $f^j(\theta) = a_0 + a_1j = z_j$. Logo, a relação anterior dá

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \varphi(z_j) = \int \varphi dm.$$

Isto é precisamente o que significa dizer que z_j é equidistribuída.

6.4.1 Ergodicidade

Agora vamos estender os argumentos acima para qualquer grau $d \geq 1$. Para isso introduzimos a transformação $f : \mathbb{T}^d \rightarrow \mathbb{T}^d$ definida no toro d -dimensional \mathbb{T}^d pela seguinte expressão:

$$f(\theta_1, \theta_2, \dots, \theta_d) = (\theta_1 + \alpha, \theta_2 + \theta_1, \dots, \theta_d + \theta_{d-1}), \quad (6.4.1)$$

onde α é um número irracional que será escolhido mais tarde. Note que f é invertível: a inversa está dada por

$$f^{-1}(\theta_1, \theta_2, \dots, \theta_d) = (\theta_1 - \alpha, \theta_2 - \theta_1 + \alpha, \dots, \theta_d - \theta_{d-1} + \dots + (-1)^{d-1}\theta_1 + (-1)^d\alpha).$$

Note também que a derivada de f em cada ponto é dada pela matriz

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & 1 \end{pmatrix}$$

cujo determinante é 1. Isto garante que f preserva a medida de Lebesgue no toro (lembre do Lema 1.3.5).

Proposição 6.4.3. *A medida de Lebesgue em \mathbb{T}^d é ergódica para f .*

Demonstração. Vamos usar um argumento de expansão em série de Fourier, análogo ao da Proposição 4.2.1. Seja $\varphi : \mathbb{T}^d \rightarrow \mathbb{R}$ uma função em $L^2(m)$. Escrevemos

$$\varphi(\theta) = \sum_{n \in \mathbb{Z}^d} a_n e^{2\pi i n \cdot \theta}$$

onde $\theta = (\theta_1, \dots, \theta_d)$ e $n = (n_1, \dots, n_d)$ e $n \cdot \theta = n_1\theta_1 + \dots + n_d\theta_d$ e

$$\sum_{n \in \mathbb{Z}^d} |a_n|^2 = \int |\varphi(\theta)|^2 d\theta_1 \cdots d\theta_d < \infty. \quad (6.4.2)$$

Observe que

$$\begin{aligned} \varphi(f(\theta)) &= \sum_{n \in \mathbb{Z}^d} a_n e^{2\pi i (n_1(\theta_1 + \alpha) + n_2(\theta_2 + \theta_1) + \dots + n_d(\theta_d + \theta_{d-1}))} \\ &= \sum_{n \in \mathbb{Z}^d} a_n e^{2\pi i n_1 \alpha} e^{2\pi i L(n) \cdot \theta} \end{aligned}$$

onde $L(n) = (n_1 + n_2, n_2 + n_3, \dots, n_{d-1} + n_d, n_d)$. Suponhamos que a função φ é invariante, isto é, $\varphi \circ f = \varphi$ em quase todo ponto. Então,

$$a_n e^{2\pi i n_1 \alpha} = a_{L(n)} \quad \text{para todo } n \in \mathbb{Z}^d. \quad (6.4.3)$$

Isto implica que a_n e $a_{L(n)}$ têm o mesmo valor absoluto. Por outro lado, a relação de integrabilidade (6.4.2) implica que existe no máximo um número finito de termos com um dado valor absoluto não-nulo. Concluímos que $a_n = 0$ para todo $n \in \mathbb{Z}^d$ cuja órbita $L^j(n)$, $j \in \mathbb{Z}$ seja infinita. Observando a expressão de L deduzimos que $a_n = 0$ exceto, possivelmente, se $n_2 = \dots = n_d = 0$. Além disso, para os valores de n restantes, ou seja, para $n = (n_1, 0, \dots, 0)$, tem-se que $L(n) = n$ e portanto a relação (6.4.3) torna-se

$$a_n = a_n e^{2\pi i n_1 \alpha}.$$

Como α é irracional, o último fator é diferente de 1 sempre que n_1 é não-nulo. Portanto esta relação dá que $a_n = 0$ também para $n = (n_1, 0, \dots, 0)$ com $n_1 \neq 0$. Deste modo, mostramos que se φ é uma função invariante então todos os termos da sua expansão em série de Fourier se anulam exceto, possivelmente, o termo constante. Isto mostra que φ é constante em quase todo ponto, e isso prova que a medida de Lebesgue é ergódica para f . \square

6.4.2 Unicidade ergódica

O próximo passo da demonstração do Teorema 6.4.2 é o seguinte resultado:

Proposição 6.4.4. *A transformação f é unicamente ergódica: a medida de Lebesgue no toro é a sua única probabilidade invariante.*

Demonstração. A demonstração será por indução no grau d do polinômio P . O caso de grau 1 foi tratado anteriormente. Portanto, só precisamos explicar como o caso de grau d pode ser deduzido do caso de grau $d-1$. Para isso, escrevemos $\mathbb{T}^d = \mathbb{T}^{d-1} \times S^1$ e

$$f : \mathbb{T}^{d-1} \times S^1 \rightarrow \mathbb{T}^{d-1} \times S^1, \quad f(\theta_0, \eta) = (f_0(\theta_0), \eta + \theta_{d-1}), \quad (6.4.4)$$

onde $\theta_0 = (\theta_1, \dots, \theta_{d-1})$ e $f_0(\theta_0) = (\theta_1 + \alpha, \theta_2 + \theta_1, \dots, \theta_{d-1} + \theta_{d-2})$. Por indução, a transformação

$$f_0 : \mathbb{T}^{d-1} \rightarrow \mathbb{T}^{d-1}$$

é unicamente ergódica. Representamos por $\pi : \mathbb{T}^d \rightarrow \mathbb{T}^{d-1}$ a projeção $\pi(\theta) = \theta_0$.

Lema 6.4.5. *Se μ é uma probabilidade invariante por f então a projeção $\pi_*\mu$ coincide com a medida de Lebesgue m_0 em \mathbb{T}^{d-1} .*

Demonstração. Dado qualquer conjunto mensurável $E \subset \mathbb{T}^{d-1}$,

$$(\pi_*\mu)(f_0^{-1}(E)) = \mu(\pi^{-1}f_0^{-1}(E)).$$

Usando $\pi \circ f = f_0 \circ \pi$ e o fato de que μ é f -invariante, se verifica que a expressão do lado direito é igual a

$$\mu(f^{-1}\pi^{-1}(E)) = \mu(\pi^{-1}(E)) = (\pi_*\mu)(E).$$

Portanto $(\pi_*\mu)(f_0^{-1}(E)) = (\pi_*\mu)(E)$ para todo subconjunto mensurável E , ou seja, $\pi_*\mu$ é probabilidade f_0 -invariante. Como f_0 é unicamente ergódico, segue que $\pi_*\mu$ coincide com a medida de Lebesgue m_0 em \mathbb{T}^{d-1} . \square

Agora suponhamos que μ , além de invariante, também é ergódica para f . Pelo Teorema 3.2.6, e por ergodicidade, o conjunto $G(\mu) \subset M$ dos pontos $\theta \in \mathbb{T}^d$ tais que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(\theta)) = \int \varphi d\mu \quad \text{para toda função contínua } \varphi : \mathbb{T}^d \rightarrow \mathbb{R} \quad (6.4.5)$$

tem medida total. Seja $G_0(\mu)$ o conjunto dos $\theta_0 \in \mathbb{T}^{d-1}$ tais que $G(\mu)$ intersecta $\{\theta_0\} \times S^1$. Em outras palavras, $G_0(\mu) = \pi(G(\mu))$. É claro que $\pi^{-1}(G_0(\mu))$ contém $G(\mu)$ e, portanto, tem medida total. Logo, usando o Lema 6.4.5,

$$m_0(G_0(\mu)) = \mu(\pi^{-1}(G_0(\mu))) = 1. \quad (6.4.6)$$

Pelas mesmas razões, esta relação também vale para a medida de Lebesgue:

$$m_0(G_0(m)) = m(\pi^{-1}(G_0(m))) = 1. \quad (6.4.7)$$

Uma consequência direta das igualdades (6.4.6) e (6.4.7) é que a intersecção de $G_0(\mu)$ e $G_0(m)$ tem medida m_0 total. Logo, em particular, estes conjuntos não podem ser disjuntos. Seja θ_0 um ponto qualquer na intersecção. Por definição, $G(\mu)$ intersecta $\{\theta_0\} \times S^1$. Mas o próximo resultado afirma que $G(m)$ contém $\{\theta_0\} \times S^1$:

Lema 6.4.6. Se $\theta_0 \in G_0(m)$ então $\{\theta_0\} \times S^1$ está contido em $G(m)$.

Demonstração. A observação crucial é que a medida m é invariante por toda a transformação da forma

$$R_\beta : \mathbb{T}^{d-1} \times S^1 \rightarrow \mathbb{T}^{d-1} \times S^1, \quad (\zeta, \eta) \mapsto (\zeta, \eta + \beta).$$

A hipótese $\theta_0 \in G_0(m)$ significa que existe algum $\eta \in S^1$ tal que $(\theta_0, \eta) \in G(m)$, ou seja,

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(\theta_0, \eta)) = \int \varphi dm$$

para toda função contínua $\varphi : \mathbb{T}^d \rightarrow \mathbb{R}$. Qualquer outro ponto de $\{\theta_0\} \times S^1$ pode ser escrito como $(\theta_0, \eta + \beta) = R_\beta(\theta_0, \eta)$ para algum $\beta \in S^1$. Recordando (6.4.1), vemos que

$$f(R_\beta(\tau_0, \zeta)) = (\tau_1 + \alpha, \tau_2 + \tau_1, \dots, \tau_{d-1} + \tau_{d-2}, \zeta + \beta + \tau_{d-1}) = R_\beta(f(\tau_0, \zeta))$$

para todo $(\tau_0, \zeta) \in \mathbb{T}^{d-1} \times S^1$. Logo, por indução,

$$f^j(\theta_0, \eta + \beta) = f^j(R_\beta(\theta_0, \eta)) = R_\beta(f^j(\theta_0, \eta))$$

para todo $j \geq 1$. Portanto, dada qualquer função contínua $\varphi : \mathbb{T}^d \rightarrow \mathbb{R}$,

$$\begin{aligned} \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(\theta_0, \eta + \beta)) &= \lim_n \frac{1}{n} \sum_{j=0}^{n-1} (\varphi \circ R_\beta)(f^j(\theta_0, \eta)) \\ &= \int (\varphi \circ R_\beta) dm = \int \varphi dm. \end{aligned}$$

Isto prova que $(\theta_0, \eta + \beta)$ está no conjunto $G(m)$ para todo $\beta \in S^1$, conforme foi afirmado. \square

Segue do que dissemos até agora que $G(\mu)$ e $G(m)$ se intersectam em algum ponto de $\{\theta_0\} \times S^1$. Tendo em vista a definição (6.4.5), isto implica que as duas medidas têm a mesma integral para cada função contínua. De acordo com a Proposição A.3.3, isto implica que $\mu = m$, como queríamos demonstrar. \square

Corolário 6.4.7. A órbita de todo ponto $\theta \in \mathbb{T}^d$ é equidistribuída no toro \mathbb{T}^d , ou seja, para toda função contínua $\psi : \mathbb{T}^d \rightarrow \mathbb{R}$ tem-se

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} \psi(f^j(\theta)) = \int \psi dm.$$

Demonstração. Isto segue imediatamente das Proposições 6.1.1 e 6.4.4. \square

6.4.3 Demonstração do teorema de Weyl

Para completarmos a demonstração do Teorema 6.4.2, introduzimos as funções polinomiais p_1, \dots, p_d definidas por

$$\begin{aligned} p_d(x) &= P(x) \quad \text{e} \\ p_{j-1}(x) &= p_j(x+1) - p_j(x) \quad \text{para } j = 2, \dots, d. \end{aligned} \quad (6.4.8)$$

Lema 6.4.8. *O polinômio $p_j(x)$ tem grau j , para todo $1 \leq j \leq d$. Além disso, $p_1(x) = \alpha x + \beta$ com $\alpha = d!a_d$.*

Demonstração. Por definição, $p_d(x) = P(x)$ tem grau d . Logo, para mostrar a primeira afirmação basta mostrar que se $p_j(x)$ tem grau j então $p_{j-1}(x)$ tem grau $j-1$. Para isso, escreva

$$p_j(x) = b_j x^j + b_{j-1} x^{j-1} + \dots + b_0,$$

onde $b_j \neq 0$. Então

$$\begin{aligned} p_j(x+1) &= b_j (x+1)^j + b_{j-1} (x+1)^{j-1} + \dots + b_0 \\ &= b_j x^j + (j b_j + b_{j-1}) x^{j-1} + \dots + b_0. \end{aligned}$$

Subtraindo uma expressão da outra, obtemos que

$$p_{j-1}(x) = (j b_j) x^{j-1} + b'_{j-2} x^{j-2} + \dots + b'_0$$

tem grau $j-1$. Isto prova a primeira afirmação no lema. Este cálculo também mostra que o coeficiente guia de cada $p_{j-1}(x)$ se obtém multiplicando por j o coeficiente guia de $p_j(x)$. Consequentemente, o coeficiente guia de p_1 deve ser igual a $d!a_d$. Isto prova a segunda afirmação. \square

Lema 6.4.9. *Para todo $n \geq 0$,*

$$f^n(p_1(0), p_2(0), \dots, p_d(0)) = (p_1(n), p_2(n), \dots, p_d(n)).$$

Demonstração. A demonstração será por indução em n . Como o caso $n = 0$ é óbvio, só precisamos tratar do passo indutivo. Lembre que f foi definida em (6.4.1). Se

$$f^{n-1}(p_1(0), p_2(0), \dots, p_d(0)) = (p_1(n-1), p_2(n-1), \dots, p_d(n-1))$$

então $f^n(p_1(0), p_2(0), \dots, p_d(0))$ é igual a

$$(p_1(n-1) + \alpha, p_2(n-1) + p_1(n-1), \dots, p_d(n-1) + p_{d-1}(n-1)).$$

Usando a definição (6.4.8) e o Lema 6.4.8, obtemos que esta expressão é igual a

$$(p_1(n), p_2(n), \dots, p_d(n)),$$

e isto prova o lema. \square

Finalmente, estamos prontos para provar que a sequência $z_n = P_*(n)$, $n \in \mathbb{N}$ é equidistribuída. Vamos tratar dois casos separadamente.

Em primeiro lugar, suponha que o coeficiente guia a_d de $P(x)$ é irracional. Então o número α no Lema 6.4.8 é irracional e, portanto, os resultados da Seção 6.4.2 são válidos para a transformação $f : \mathbb{T}^d \rightarrow \mathbb{T}^d$. Seja $\varphi : S^1 \rightarrow \mathbb{R}$ uma função contínua qualquer. Considere $\psi : \mathbb{T}^d \rightarrow \mathbb{R}$ definida por

$$\psi(\theta_1, \theta_2, \dots, \theta_d) = \varphi(\theta_d).$$

Fixemos $\theta = (p_1(0), p_2(0), \dots, p_d(0))$. Usando o Lema 6.4.9 e o Corolário 6.4.7, obtemos que

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi(z_n) = \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \psi(f^j(\theta)) = \int \psi \, dm = \int \varphi \, dx.$$

Isto termina a demonstração do Teorema 6.4.2 no caso em que a_d é irracional.

Agora suponha que a_d é racional, digamos $a_d = p/q$ com $p \in \mathbb{Z}$ e $q \in \mathbb{N}$. É claro que podemos escrever z_n como uma soma

$$z_n = x_n + y_n, \quad x_n = a_d n^d \quad \text{e} \quad y_n = Q_*(n)$$

onde $Q(x) = a_0 + a_1 x + \dots + a_{d-1} x^{d-1}$ e $Q_* : \mathbb{R} \rightarrow S^1$ é dada por $Q_* = \pi \circ Q$. Observe, em primeiro lugar, que

$$x_{n+q} - x_n = \frac{p}{q}(n+q)^d - \frac{p}{q}n^d$$

é um número inteiro, para todo $n \in \mathbb{N}$. Isto significa que a sequência x_n , $n \in \mathbb{N}$ é periódica de período q no círculo \mathbb{R}/\mathbb{Z} . Em particular, ela toma no máximo q valores distintos. Observe também que, como a_d é racional, a hipótese do teorema implica que algum dos coeficientes a_1, \dots, a_{d-1} de Q é irracional. Logo, por indução no grau, temos que y_n , $n \in \mathbb{N}$ é equidistribuída. Mais do que isso, as subsequências

$$y_{qn+r} = Q_*(qn+r), \quad n \in \mathbb{Z}$$

são equidistribuídas para todo $r \in \{0, 1, \dots, q-1\}$. Na verdade, estas sequências podem ser escritas como $y_{qn+r} = Q_*^{(r)}(n)$ para algum polinômio $Q^{(r)}$ que também tem grau $d-1$ (verifique) e, portanto, a hipótese de indução se aplica a elas também. Destas duas observações segue que toda subsequência z_{qn+r} , $n \in \mathbb{Z}$ é equidistribuída. Consequentemente, z_n , $n \in \mathbb{N}$ também é equidistribuída. Isto completa a prova do Teorema 6.4.2.

6.4.4 Exercícios

6.4.1. Mostre que uma sequência $(z_j)_j$ é equidistribuída no círculo se, e somente se, para todo segmento $I \subset S^1$ tem-se

$$\lim_n \frac{1}{n} \#\{1 \leq j \leq n : z_j \in I\} = m(I)$$

onde $m(I)$ representa o comprimento de I .

6.4.2. Mostre que a sequência $(\sqrt{n} \bmod \mathbb{Z})_n$ é equidistribuída no círculo. O mesmo vale para a sequência $(\log n \bmod \mathbb{Z})_n$?

6.4.3. Koksma [Kok35] mostrou que a sequência $(a^n \bmod \mathbb{Z})_n$ é equidistribuída no círculo, para Lebesgue quase todo $a > 1$. Isso não é verdade para *todo* $a > 1$. De fato, considere a *razão áurea* $a = (1 + \sqrt{5})/2$. Verifique que a sequência $(a^n \bmod \mathbb{Z})_n$ converge para $0 \in S^1$ quando $n \rightarrow \infty$; em particular, ela não é equidistribuída no círculo.

Capítulo 7

Correlações

Os modelos de sistemas dinâmicos em que estamos interessados mais diretamente, transformações e fluxos, são determinísticos: o estado do sistema em qualquer momento determina toda a trajetória futura; quando o sistema é invertível, a trajetória passada fica igualmente determinada. No entanto, estes sistemas podem apresentar também comportamento de tipo estocástico (ou seja, “parcialmente aleatório”): num nível mais grosseiro que o das trajetórias individuais, informação sobre o passado vai sendo esquecida à medida que o sistema é iterado. Este é o tema do presente capítulo.

Em Teoria da Probabilidade, chamamos *correlação* de duas variáveis aleatórias, X e Y , ao número

$$C(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Note que a expressão $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])$ é positiva se X e Y estão do mesmo lado (maior ou menor) das respectivas médias, $\mathbb{E}[X]$ e $\mathbb{E}[Y]$, e é negativa no caso contrário. Portanto, o sinal de $C(X, Y)$ indica se as duas variáveis apresentam, predominantemente, o mesmo comportamento ou comportamentos opostos, relativamente às suas médias; correlação próxima de zero sinaliza que os dois comportamentos estão pouco ou nada relacionados um com o outro.

Dada uma probabilidade invariante μ de um sistema dinâmico $f : M \rightarrow M$ e dadas funções mensuráveis $\varphi, \psi : M \rightarrow \mathbb{R}$, queremos analisar a evolução das correlações

$$C_n(\varphi, \psi) = C(\varphi \circ f^n, \psi)$$

quando o tempo n vai para infinito. Podemos pensar em φ e ψ como grandezas que medimos no sistema, tais como a temperatura, o pH, a energia cinética etc. Então $C_n(\varphi, \psi)$ mede como o valor de φ em tempo n se correlaciona com o valor de ψ em tempo zero, até que ponto um valor “influencia” o outro. Por exemplo, se $\varphi = \mathcal{X}_A$ e $\psi = \mathcal{X}_B$ são funções características, então $\psi(x)$ dá informação sobre a posição do ponto inicial x , enquanto que $\varphi(f^n(x))$ informa sobre a posição do seu n -ésimo iterado $f^n(x)$. Se a correlação $C_n(\varphi, \psi)$ for pequena então a primeira informação será de pouca utilidade para fazer previsões quanto

ao segundo evento. Como veremos, esse tipo de comportamento, em que as correlações se aproximam de zero à medida que n aumenta, é bastante comum em modelos importantes.

Começaremos por introduzir as noções de sistema (fortemente) misturador e fracamente misturador e por estudar as suas propriedades básicas (Seção 7.1). Nas Seções 7.2 e 7.3 discutiremos estas noções no contexto dos deslocamentos de Markov, que generalizam os deslocamentos de Bernoulli, e dos intercâmbios de intervalos, uma extensão da classe das rotações no círculo. Na Seção 7.4 analisaremos, em termos quantitativos, a velocidade de decaimento das correlações para certas classes de funções.

7.1 Sistemas misturadores

Seja $f : M \rightarrow M$ uma transformação mensurável e seja μ uma probabilidade invariante. A *sequência de correlações* de duas funções mensuráveis $\varphi, \psi : M \rightarrow \mathbb{R}$ é definida por

$$C_n(\varphi, \psi) = \int (\varphi \circ f^n)\psi d\mu - \int \varphi d\mu \int \psi d\mu, \quad n \in \mathbb{N}. \quad (7.1.1)$$

Dizemos que o sistema (f, μ) é *misturador* se

$$\lim_n C_n(\mathcal{X}_A, \mathcal{X}_B) = \lim_n \mu(f^{-n}(A) \cap B) - \mu(A)\mu(B) = 0, \quad (7.1.2)$$

para quaisquer conjuntos mensuráveis $A, B \subset M$. Em outras palavras, quando n cresce, a probabilidade do evento $\{x \in B \text{ e } f^n(x) \in A\}$ converge para o produto das probabilidades dos eventos $\{x \in B\}$ e $\{f^n(x) \in A\}$.

Analogamente, dado um fluxo $f^t : M \rightarrow M$, $t \in \mathbb{R}$ e uma probabilidade invariante μ , definimos

$$C_t(\varphi, \psi) = \int (\varphi \circ f^t)\psi d\mu - \int \varphi d\mu \int \psi d\mu, \quad t \in \mathbb{R} \quad (7.1.3)$$

e dizemos que o sistema (f^t, μ) é *misturador* se

$$\lim_{t \rightarrow +\infty} C_t(\mathcal{X}_A, \mathcal{X}_B) = \lim_{t \rightarrow +\infty} \mu(f^{-t}(A) \cap B) - \mu(A)\mu(B) = 0, \quad (7.1.4)$$

para quaisquer conjuntos mensuráveis $A, B \subset M$.

7.1.1 Propriedades

Um sistema misturador é necessariamente ergódico. De fato, suponha que existe algum conjunto invariante $A \subset M$ com $0 < \mu(A) < 1$. Tomando $B = A^c$ vem que $f^{-n}(A) \cap B = \emptyset$ para todo n . Logo, $\mu(f^{-n}(A) \cap B) = 0$ para todo n , enquanto que $\mu(A)\mu(B) \neq 0$. Em particular, (f, μ) não é misturador. O exemplo a seguir mostra que ergodicidade é uma propriedade estritamente mais fraca:

Exemplo 7.1.1. Considere $\theta \in \mathbb{R}$ irracional. Como vimos anteriormente, a rotação R_θ no círculo S^1 é ergódica para a medida de Lebesgue m . No entanto, (R_θ, m) não é misturador. De fato, se $A, B \subset S^1$ são dois intervalos pequenos (com comprimento menor que $1/10$, digamos) então $R_\theta^{-n}(A) \cap B$ é vazio e, portanto, $m(R_\theta^{-n}(A) \cap B) = 0$ para infinitos valores de n . Como $m(A)m(B) \neq 0$, segue que o limite em (7.1.2) não se verifica.

É claro da definição (7.1.2) que se (f, μ) é misturador então (f^k, μ) é misturador, para todo $k \in \mathbb{N}$. A afirmação correspondente para ergodicidade é falsa: a aplicação $f(x) = 1 - x$ no conjunto $\{0, 1\}$ é ergódica para a medida $(\delta_0 + \delta_1)/2$ mas o segundo iterado f^2 não é.

Lema 7.1.2. *Suponha que $\lim_n \mu(f^{-n}(A) \cap B) = \mu(A)\mu(B)$ para todo par de conjuntos A e B em alguma álgebra \mathcal{A} geradora da σ -álgebra dos conjuntos mensuráveis. Então (f, μ) é misturador.*

Demonstração. Seja \mathcal{C} a família de todos os conjuntos mensuráveis A tais que $\mu(f^{-n}(A) \cap B) \rightarrow \mu(A)\mu(B)$ para todo $B \in \mathcal{A}$. Por hipótese, \mathcal{C} contém \mathcal{A} . Afirmamos que \mathcal{C} é uma classe monótona. De fato, sejam $A_1 \subset \dots \subset A_k \subset \dots$ elementos de \mathcal{C} e seja $A = \cup_k A_k$. Dado $\varepsilon > 0$, existe $k_0 \geq 1$ tal que

$$\mu(A) - \mu(A_k) = \mu(A \setminus A_k) < \varepsilon$$

para todo $k \geq k_0$. Além disso, para todo $n \geq 1$,

$$\begin{aligned} \mu(f^{-n}(A) \cap B) - \mu(f^{-n}(A_k) \cap B) &= \mu(f^{-n}(A \setminus A_k) \cap B) \\ &\leq \mu(f^{-n}(A \setminus A_k)) = \mu(A \setminus A_k) < \varepsilon. \end{aligned}$$

Para $k \geq k_0$ fixado, o fato de que $A_k \in \mathcal{C}$ garante que existe $n(k) \geq$ tal que

$$|\mu(f^{-n}(A_k) \cap B) - \mu(A_k)\mu(B)| < \varepsilon \quad \text{para todo } n \geq n(k).$$

Somando estas três desigualdades concluímos que

$$|\mu(f^{-n}(A) \cap B) - \mu(A)\mu(B)| < 3\varepsilon \quad \text{para todo } n \geq n(k_0).$$

Como $\varepsilon > 0$ é arbitrário, isto mostra que $A \in \mathcal{C}$. Da mesma forma se mostra que a interseção de qualquer sequência decrescente de elementos de \mathcal{C} ainda é um elemento de \mathcal{C} . Portanto \mathcal{C} é uma classe monótona, tal como afirmamos. Pelo teorema das classes monótonas (Teorema A.1.18), segue que \mathcal{C} contém todo o conjunto mensurável: para todo conjunto mensurável A tem-se

$$\lim_n \mu(f^{-n}(A) \cap B) = \mu(A)\mu(B) \quad \text{para todo } B \in \mathcal{A}.$$

Resta deduzir que esta propriedade vale para todo conjunto mensurável B . Isto é inteiramente análogo aos argumentos que acabamos de detalhar, pelo que deixaremos a verificação a cargo do leitor. \square

Exemplo 7.1.3. Todo deslocamento de Bernoulli (lembre da Seção 4.2.3) é misturador. De fato, dados dois cilindros $A = [p; A_p, \dots, A_q]$ e $B = [r; B_r, \dots, B_s]$ quaisquer, tem-se

$$\begin{aligned}\mu(f^{-n}(A) \cap B) &= \mu([r; B_r, \dots, B_s, X, \dots, X, A_p, \dots, A_q]) \\ &= \mu([r; B_r, \dots, B_s])\mu([p; A_p, \dots, A_q]) = \mu(A)\mu(B)\end{aligned}$$

para todo $n > s - p$. Seja \mathcal{A} a álgebra gerada pelos cilindros: os seus elementos são as uniões finitas disjuntas de cilindros. Segue do que acabamos de dizer que $\mu(f^{-n}(A) \cap B) = \mu(A)\mu(B)$ para todo par de conjuntos $A, B \in \mathcal{A}$ e todo n suficientemente grande. Como \mathcal{A} gera a σ -álgebra dos conjuntos mensuráveis, podemos usar o Lema 7.1.2 para concluir que o sistema é misturador, tal como foi afirmado.

Exemplo 7.1.4. Seja $g : S^1 \rightarrow S^1$ a transformação definida por $g(x) = kx$, onde $k \geq 2$ é um número inteiro, e seja m a medida de Lebesgue m no círculo. O sistema (g, m) é equivalente a um deslocamento de Bernoulli, no seguinte sentido. Considere $X = \{0, 1, \dots, k-1\}$ e seja $f : M \rightarrow M$ a aplicação deslocamento em $M = X^{\mathbb{N}}$. Considere a medida produto $\mu = \nu^{\mathbb{N}}$ em M , onde ν é a probabilidade definida por $\nu(A) = \#A/k$ para todo $A \subset X$. A aplicação

$$h : M \rightarrow S^1, \quad h((a_n)_n) = \sum_{n=1}^{\infty} \frac{a_{n-1}}{k^n}$$

é uma bijeção, restrita a um subconjunto com medida total, e tanto ela quanto a sua inversa são mensuráveis. Além disso, $h_*\mu = m$ e $h \circ f = g \circ h$ em quase todo ponto. Dizemos que h é uma *equivalência ergódica* entre (g, m) e (f, μ) . Por meio dela, propriedades podem ser traduzidas de um sistema para o outro. Em particular, lembrando do Exemplo 7.1.3, obtemos que (g, m) é misturador: dados quaisquer conjuntos mensuráveis $A, B \subset S^1$,

$$\begin{aligned}m(g^{-n}(A) \cap B) &= \mu(h^{-1}(g^{-n}(A) \cap B)) = \mu(f^{-n}(h^{-1}(A)) \cap h^{-1}(B)) \\ &\rightarrow \mu(h^{-1}(A))\mu(h^{-1}(B)) = m(A)m(B) \quad \text{quando } n \rightarrow \infty.\end{aligned}$$

Exemplo 7.1.5. Para endomorfismos sobrejetivos do toro (Seção 4.2.5) as propriedades de mistura e ergodicidade são equivalentes: o sistema (f_A, m) é misturador se, e somente se, nenhum autovalor da matriz A é raiz da unidade (compare com o Teorema 4.2.14). Este fato está proposto no Exercício 7.1.4 e um resultado ainda mais forte será proposto no Exercício 8.4.2. Mais geralmente, um endomorfismo sobrejetivo de um grupo compacto é misturador para a medida de Haar se, e somente se, ele é ergódico. De fato valem resultados bem mais fortes, como comentaremos na Seção 9.5.3.

Vamos também discutir uma versão topológica da noção de sistema misturador. Para isso, suponha que o ambiente M é um espaço topológico. Uma transformação $f : M \rightarrow M$ é dita *topologicamente misturadora* se dados quaisquer abertos não-vazios $U, V \subset M$, existe $n_0 \in \mathbb{N}$ tal que $f^{-n}(U) \cap V$ é não

vazio para todo $n \geq n_0$. Isto é parecido, mas estritamente mais forte do que a hipótese do Lema 4.3.4: lá pedimos que $f^{-n}(U)$ intersecte V para *algum* n , enquanto que agora queremos que isso aconteça para todo n suficientemente grande.

Proposição 7.1.6. *Se (f, μ) é misturadora então a restrição de f ao suporte de μ é topologicamente misturadora.*

Demonstração. Denote por $X = \text{supp}(\mu)$. Sejam $A, B \subset X$ conjuntos abertos. Temos que $\mu(A), \mu(B) > 0$, por definição de $\text{supp}(\mu)$. Assim, como μ é misturadora, temos que existe n_0 tal que $n \geq n_0$ vale que $\mu(f^{-n}(A) \cap B) > \mu(A)\mu(B)/2 > 0$. Em particular, $\mu(f^{-n}(A) \cap B) \neq \emptyset$, como queríamos demonstrar. \square

Segue diretamente desta proposição que se f possui uma medida μ misturadora positiva em abertos, então f é topologicamente misturadora. Por exemplo, dado qualquer conjunto finito $X = \{1, \dots, d\}$, o deslocamento

$$f : X^{\mathbb{Z}} \rightarrow X^{\mathbb{Z}} \quad (\text{ou } f : X^{\mathbb{N}} \rightarrow X^{\mathbb{N}})$$

é topologicamente misturador. De fato, para qualquer probabilidade ν suportada em todo o X , a medida de Bernoulli $\mu = \nu^{\mathbb{N}}$ (ou $\mu = \nu^{\mathbb{Z}}$) é positiva em abertos e misturadora, como vimos no Exemplo 7.1.3. Analogamente, pelo Exemplo 7.1.4, toda transformação $f : S^1 \rightarrow S^1$ da forma $f(x) = kx$ com $k \geq 2$ é topologicamente misturadora.

Exemplo 7.1.7. Translações num grupo metrizável G nunca são topologicamente misturadoras. De fato considere qualquer translação à esquerda E_g (o caso de translações à direita é análogo). Podemos supor que $g \neq e$, pois caso contrário é evidente que E_g não é topologicamente misturadora. Fixemos uma distância d invariante pelas translações do grupo G (lembre do Lema 6.3.6) e seja $\alpha = d(e, g^{-1})$. Considere $U = V =$ bola de centro e e raio $\alpha/4$. Todo $E_g^{-n}(U)$ é uma bola de raio $\alpha/4$. Suponha que $E_g^{-n}(U)$ intersecta V . Então $E_g^{-n}(U)$ está contida na bola de raio $3\alpha/4$ e, portanto, $E_g^{-n-1}(U)$ está contida na bola de raio $3\alpha/4$ em torno de g^{-1} . Consequentemente, $E_g^{-n-1}(U)$ não intersecta V . Como n é arbitrário, isto mostra que E_g não é topologicamente misturadora.

7.1.2 Mistura fraca

Dizemos que o sistema (f, μ) é *fracamente misturador*, se dados quaisquer conjuntos mensuráveis $A, B \subset M$ então:

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} |C_j(\mathcal{X}_A, \mathcal{X}_B)| = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} |\mu(f^{-j}(A) \cap B) - \mu(A)\mu(B)| = 0. \quad (7.1.5)$$

É claro da definição que todo sistema misturador é também fracamente misturador. Por outro lado, todo sistema fracamente misturador é ergódico. De fato,

se $A \subset M$ é conjunto invariante então

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} |C_j(\mathcal{X}_A, \mathcal{X}_{A^c})| = \mu(A)\mu(A^c)$$

e, portanto, a hipótese implica que $\mu(A) = 0$ ou $\mu(A^c) = 0$.

Exemplo 7.1.8. Translações em grupos compactos metrizáveis nunca são fracamente misturadoras, relativamente à medida de Haar μ (ou qualquer outra medida invariante positiva em abertos). De fato, conforme observado no Exemplo 7.1.7, é sempre possível escolher abertos U e V tais que $f^{-n}(U) \cap V$ é vazio para um em cada dois valores consecutivos de n . Então,

$$\liminf_n \frac{1}{n} \sum_{j=0}^{n-1} |\mu(f^{-j}(U) \cap V) - \mu(U)\mu(V)| \geq \frac{1}{2}\mu(U)\mu(V) > 0.$$

Desta forma obtemos diversos exemplos de sistemas ergódicos, e até unicamente ergódicos, que não são fracamente misturadores.

Conforme veremos na Seção 7.3.2, a família dos intercâmbios de intervalos contem diversos sistemas fracamente misturadores (e unicamente ergódicos) que não são misturadores.

A demonstração do resultado a seguir é análoga à do Lemma 7.1.2 e fica a cargo do leitor:

Lema 7.1.9. *Suponha que $\lim_n n^{-1} \sum_{j=0}^{n-1} |\mu(f^{-j}(A) \cap B) - \mu(A)\mu(B)| = 0$ para todo par de conjuntos A e B em alguma álgebra \mathcal{A} geradora da σ -álgebra dos conjuntos mensuráveis. Então (f, μ) é fracamente misturador.*

Exemplo 7.1.10. Dado um sistema (f, μ) , consideremos a transformação produto $f_2 : M \times M \rightarrow M \times M$ dado por $f_2(x, y) = (f(x), f(y))$. É fácil ver que f_2 preserva a medida produto $\mu_2 = \mu \times \mu$. Se (f_2, μ_2) é ergódico então (f, μ) é ergódico: basta notar que se $A \subset M$ é conjunto invariante para f com $\mu(A) \in (0, 1)$ então $A \times A$ é conjunto invariante para f_2 com $\mu_2(A \times A) \in (0, 1)$.

A recíproca não é verdadeira em geral, ou seja, (f_2, μ_2) pode não ser ergódico mesmo que (f, μ) seja ergódico. Por exemplo, se f é uma rotação irracional em S^1 e d é uma distância invariante por rotações, então qualquer vizinhança $\{(x, y) : d(x, y) < r\}$ da diagonal é um conjunto invariante para $f \times f$.

O próximo resultado mostra que este tipo de fenômeno não ocorre na categoria dos sistemas fracamente misturadores:

Proposição 7.1.11. *As seguintes condições são equivalentes:*

- (a) (f, μ) é fracamente misturador;
- (b) (f_2, μ_2) é fracamente misturador;
- (c) (f_2, μ_2) é ergódico.

Demonstração. Para provar que (a) implica (b), considere quaisquer conjuntos mensuráveis $A, B, C, D \subset M$. Então:

$$\begin{aligned} & |\mu_2(f_2^{-j}(A \times B) \cap (C \times D)) - \mu_2(A \times B)\mu_2(C \times D)| \\ &= |\mu(f^{-j}(A) \cap C)\mu(f^{-j}(B) \cap D) - \mu(A)\mu(B)\mu(C)\mu(D)| \\ &\leq |\mu(f^{-j}(A) \cap C) - \mu(A)\mu(C)| + |\mu(f^{-j}(B) \cap D) - \mu(B)\mu(D)|. \end{aligned}$$

Portanto, a hipótese (a) implica que

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} |\mu_2(f_2^{-j}(A \times B) \cap (C \times D)) - \mu_2(A \times B)\mu_2(C \times D)| = 0.$$

Segue que

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} |\mu_2(f_2^{-j}(X) \cap Y) - \mu_2(X)\mu_2(Y)| = 0$$

para quaisquer X, Y na álgebra gerada pelos produtos $E \times F$ de subconjuntos mensuráveis de M , ou seja, a álgebra das uniões finitas disjuntas de tais produtos. Como esta álgebra gera a σ -álgebra dos subconjuntos mensuráveis de $M \times M$, podemos usar o Lema 7.1.9 para concluir que (f_2, μ_2) é fracamente misturador.

É imediato que (b) implica (c). Para provar que (c) implica (a), observe que

$$\begin{aligned} & \sum_{j=0}^{n-1} [\mu(f^{-j}(A) \cap B) - \mu(A)\mu(B)]^2 \\ &= \sum_{j=0}^{n-1} [\mu(f^{-j}(A) \cap B)^2 - 2\mu(A)\mu(B)\mu(f^{-j}(A) \cap B) + \mu(A)\mu(B)^2] \end{aligned}$$

Pode ser reescrito como

$$\begin{aligned} & \sum_{j=0}^{n-1} [\mu_2(f_2^{-j}(A \times A) \cap (B \times B)) - \mu_2(A \times A)\mu_2(B \times B)] \\ & \quad - 2\mu(A)\mu(B) \sum_{j=0}^{n-1} [\mu(f^{-j}(A) \cap B) - \mu(A)\mu(B)]. \end{aligned}$$

Como que (f_2, μ_2) é ergódico e, conseqüentemente, (f, μ) também é, concluímos que

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

para quaisquer conjuntos mensuráveis $A, B \subset M$. Usando o Exercício 7.1.2, obtemos que (f, μ) é fracamente misturador. \square

7.1.3 Caracterização espectral

Nesta seção discutiremos formulações equivalentes das noções de sistema misturador e fracamente misturador, em termos do operador de Koopman.

Proposição 7.1.12. *As seguintes condições são equivalentes:*

- (a) (f, μ) é misturador.
- (b) Existem $p, q \in [1, \infty]$ com $1/p + 1/q = 1$ tais que $C_n(\varphi, \psi) \rightarrow 0$ para quaisquer $\varphi \in L^p(\mu)$ e $\psi \in L^q(\mu)$.
- (c) A condição do item (b) vale para φ em algum subconjunto denso de $L^p(\mu)$ e ψ em algum subconjunto denso de $L^q(\mu)$.

Demonstração. A condição (a) é o caso particular de (b) para funções características. Como as correlações $(\varphi, \psi) \mapsto C_n(\varphi, \psi)$ são funções bilineares, a condição (a) implica que $C_n(\varphi, \psi) \rightarrow 0$ para quaisquer funções simples φ e ψ . Isto implica (c), uma vez que as funções simples formam um subconjunto denso de $L^r(\mu)$, para qualquer $r \geq 1$.

Para mostrar que (c) implica (b), comecemos por observar que as correlações $C_n(\varphi, \psi)$ são funções equicontínuas de φ e ψ . De fato, dadas $\varphi_1, \varphi_2 \in L^p(\mu)$ e $\psi_1, \psi_2 \in L^q(\mu)$, as desigualdade de Hölder (Teorema A.5.5) dá que

$$\left| \int (\varphi_1 \circ f^n) \psi_1 d\mu - \int (\varphi_2 \circ f^n) \psi_2 d\mu \right| \leq \|\varphi_1 - \varphi_2\|_p \|\psi_1\|_q + \|\varphi_2\|_p \|\psi_1 - \psi_2\|_q.$$

Além disso,

$$\left| \int \varphi_1 d\mu \int \psi_1 d\mu - \int \varphi_2 d\mu \int \psi_2 d\mu \right| \leq \|\varphi_1 - \varphi_2\|_1 \|\psi_1\|_1 + \|\varphi_2\|_1 \|\psi_1 - \psi_2\|_1.$$

Somando estas desigualdades, e notando que $\|\cdot\|_1 \leq \|\cdot\|_r$ para todo $r \geq 1$, obtemos que:

$$\left| C_n(\varphi_1, \psi_1) - C_n(\varphi_2, \psi_2) \right| \leq 2\|\varphi_1 - \varphi_2\|_p \|\psi_1\|_q + 2\|\varphi_2\|_p \|\psi_1 - \psi_2\|_q \quad (7.1.6)$$

para todo $n \geq 1$. Então, dado $\varepsilon > 0$ e dados quaisquer $\varphi \in L^p(\mu)$ e $\psi \in L^q(\mu)$, podemos tomar φ' e ψ' nos subconjuntos densos da mencionados na hipótese tais que

$$\|\varphi - \varphi'\|_p < \varepsilon \quad \text{e} \quad \|\psi - \psi'\|_q < \varepsilon.$$

Em particular, $\|\varphi'\|_p < \|\varphi\|_p + \varepsilon$ e $\|\psi'\|_q < \|\psi\|_q + \varepsilon$. Então, (7.1.6) dá que

$$|C_n(\varphi, \psi)| \leq |C_n(\varphi', \psi')| + 2\varepsilon(\|\varphi\|_p + \|\psi\|_q + 2\varepsilon) \quad \text{para todo } n.$$

Além disso, por hipótese, temos $|C_n(\varphi', \psi')| < \varepsilon$ para todo n suficientemente grande. Como ε é arbitrário, estas duas desigualdades implicam que $C_n(\varphi, \psi)$ converge para zero quando $n \rightarrow \infty$. Isto prova a propriedade (b). \square

O mesmo argumento prova a seguinte versão da Proposição 7.1.12 para a propriedade de mistura fraca:

Proposição 7.1.13. *As seguintes condições são equivalentes:*

- (a) (f, μ) é fracamente misturador.
- (b) Existem $p, q \in [1, \infty]$ com $1/p + 1/q = 1$ tais que $(1/n) \sum_{j=1}^n |C_j(\varphi, \psi)|$ converge para 0 para quaisquer $\varphi \in L^p(\mu)$ e $\psi \in L^q(\mu)$.
- (c) A condição do item (b) vale em algum subconjunto denso de $L^p(\mu)$ e algum subconjunto denso de $L^q(\mu)$.

No caso $p = q = 2$, podemos expressar as correlações em termos do produto interno \cdot no espaço de Hilbert $L^2(\mu)$:

$$C_n(\varphi, \psi) = [U_f^n \varphi - (\varphi \cdot 1)] \cdot \psi \quad \text{para todo } \varphi, \psi \in L^2(\mu).$$

Portanto, a Proposição 7.1.12 dá que (f, μ) é misturador se, e somente se,

$$\lim_n [U_f^n \varphi - (\varphi \cdot 1)] \cdot \psi = 0 \quad \text{para todo } \varphi, \psi \in L^2(\mu) \quad (7.1.7)$$

e a Proposição 7.1.13 dá que (f, μ) é fracamente misturador se, e somente se,

$$\lim_n \frac{1}{n} \sum_{j=1}^n |[U_f^j \varphi - (\varphi \cdot 1)] \cdot \psi| = 0 \quad \text{para todo } \varphi, \psi \in L^2(\mu). \quad (7.1.8)$$

A condição (7.1.7) significa que $U_f^n \varphi$ converge *fracamente* para $\varphi \cdot 1 = \int \varphi d\mu$, enquanto que (7.1.8) é uma versão Cesaro dessa afirmação. Compare as duas condições com a caracterização da ergodicidade em (4.1.7).

Corolário 7.1.14. *Seja $f : M \rightarrow M$ uma transformação misturadora para uma probabilidade invariante μ . Seja ν uma probabilidade qualquer em M , absolutamente contínua em relação a μ . Então $f_*^n \nu$ converge pontualmente para μ , ou seja, $\nu(f^{-n}(B)) \rightarrow \mu(B)$ para todo conjunto mensurável $B \subset M$.*

Demonstração. Considere $\varphi = \chi_B$ e $\psi = d\nu/d\mu$. Note que $\varphi \in L^\infty(\mu)$ e $\psi \in L^1(\mu)$. Logo, pela Proposição 7.1.12,

$$\int (\chi_B \circ f^n) \frac{d\nu}{d\mu} d\mu = \int (U_f^n \varphi) \psi d\mu \rightarrow \int \varphi d\mu \int \psi d\mu = \int \chi_B d\mu \int \frac{d\nu}{d\mu} d\mu.$$

A sequência do lado esquerdo coincide com $\int (\chi_B \circ f^n) d\nu = \nu(f^{-n}(B))$. O lado direito é igual a $\mu(B) \int 1 d\nu = \mu(B)$. \square

7.1.4 Exercícios

7.1.1. Mostre que (f, μ) é misturador se, e somente se, $\mu(f^{-n}(A) \cap A) \rightarrow \mu(A)^2$ para todo conjunto mensurável A .

7.1.2. Seja $(a_n)_n$ uma sequência limitada de números reais. Demonstre que $\lim_n (1/n) \sum_{j=1}^n |a_j| = 0$ se, e somente se, existe $E \subset \mathbb{N}$ com densidade zero no infinito (ou seja, satisfazendo $\lim_n (1/n) \#(E \cap \{0, \dots, n-1\}) = 0$) tal que a restrição de $(a_n)_n$ ao complementar de E converge para zero quando $n \rightarrow \infty$. Deduza que

$$\lim_n \frac{1}{n} \sum_{j=1}^n |a_j| = 0 \quad \Leftrightarrow \quad \lim_n \frac{1}{n} \sum_{j=1}^n (a_j)^2 = 0.$$

7.1.3. Prove que se μ é fracamente misturadora para f então μ é fracamente misturadora para todo iterado f^k , $k \geq 1$.

7.1.4. Mostre que se nenhum autovalor de $A \in \text{SL}(d, \mathbb{R})$ é uma raiz da unidade então o endomorfismo linear $f_A : \mathbb{T}^d \rightarrow \mathbb{T}^d$ induzido por A é misturador, com respeito à medida de Haar.

7.1.5. Seja $f : M \rightarrow M$ uma transformação mensurável num espaço métrico. Verifique que uma probabilidade invariante μ é misturadora se, e somente se, $(f_*^n \eta)_n$ converge para μ na topologia fraca* para toda probabilidade η absolutamente contínua com respeito a μ .

7.1.6 (Teorema de Von Neuman múltiplo). Mostre que se (f, μ) é fracamente misturador então

$$\frac{1}{N} \sum_{n=0}^{N-1} (\varphi_1 \circ f^n) \cdots (\varphi_k \circ f^{kn}) \rightarrow \int \varphi_1 d\mu \cdots \int \varphi_k d\mu \quad \text{em } L^2(\mu),$$

para quaisquer funções mensuráveis limitadas $\varphi_1, \dots, \varphi_k$.

7.2 Deslocamentos de Markov

Nesta seção introduzimos uma importante classe de sistemas, que generaliza a classe de deslocamentos de Bernoulli. Como vimos anteriormente, deslocamentos de Bernoulli modelam sequências de experimentos idênticos em que o resultado de cada experimento é *independente* dos demais. Na definição dos deslocamentos de Markov abandonamos essa condição de independência, mas supomos que cada resultado depende apenas do resultado imediatamente anterior. Mais geralmente, deslocamentos de Markov podem ser usados para modelar processos, ditos com *memória finita*, isto é, tais que existe $k \geq 1$ tal que cada resultado depende apenas dos k resultados imediatamente anteriores. A este respeito, veja o Exercício 7.2.4.

Para definir um deslocamento de Markov, consideremos um espaço mensurável (X, \mathcal{A}) e seja $\Sigma = X^{\mathbb{N}}$ (ou $\Sigma = X^{\mathbb{Z}}$) o espaço das sequências em X , munido da σ -álgebra produto. Consideraremos o deslocamento

$$\sigma : \Sigma \rightarrow \Sigma, \quad \sigma((x_n)_n) = (x_{n+1})_n.$$

Suponha que é dada uma família $\{P(x, \cdot) : x \in X\}$ de probabilidades em X , chamadas *probabilidades de transição*, dependendo mensuravelmente do ponto

x . Heuristicamente, dado um conjunto mensurável $E \subset X$, o número $P(x, E)$ representa a probabilidade de $x_{n+1} \in E$ sabendo que $x_n = x$. Uma probabilidade p em X é chamada *medida estacionária*, relativamente à família de probabilidades de transição, se ela satisfaz

$$\int P(x, E) dp(x) = p(E), \quad \text{para todo conjunto mensurável } E \subset X. \quad (7.2.1)$$

Heuristicamente, isto significa que, relativamente a p , a probabilidade do evento $x_{n+1} \in E$ é igual à probabilidade do evento $x_n \in E$.

Fixe uma medida estacionária p qualquer (supondo que exista) e então defina

$$\begin{aligned} \mu([m; A_m, \dots, A_n]) &= \\ &= \int_{A_m} dp(x_m) \int_{A_{m+1}} dP(x_m, x_{m+1}) \cdots \int_{A_n} dP(x_{n-1}, x_n) \end{aligned} \quad (7.2.2)$$

para todo cilindro $[m; A_m, \dots, A_n]$ de Σ . Pode mostrar-se que esta função se estende a uma probabilidade na σ -álgebra gerado pelos cilindros (confira o Exercício 7.2.1). Esta probabilidade é invariante pelo deslocamento σ , uma vez que o lado direito de (7.2.2) não depende de m . Toda probabilidade μ obtida desta forma é chamada *medida de Markov*; além disso, o sistema (σ, μ) é chamado *deslocamento de Markov*.

Exemplo 7.2.1. (Medida de Bernoulli) Suponha que $P(x, \cdot)$ não depende de x , ou seja, existe uma probabilidade ν em X tal que $P(x, \cdot) = \nu$ para todo $x \in X$. Então

$$\int P(x, E) dp(x) = \int \nu(E) dp(x) = \nu(E)$$

para toda probabilidade p e todo conjunto mensurável $E \subset X$. Portanto, existe exatamente uma medida estacionária, a saber $p = \nu$. A definição (7.2.2) dá

$$\begin{aligned} \mu([m; A_m, \dots, A_n]) &= \int_{A_m} d\nu(x_m) \int_{A_{m+1}} d\nu(x_{m+1}) \cdots \int_{A_n} d\nu(x_n) \\ &= \nu(A_m)\nu(A_{m+1}) \cdots \nu(A_n). \end{aligned}$$

Exemplo 7.2.2. Suponha que o conjunto X é finito, digamos $X = \{1, \dots, d\}$ para algum $d \geq 2$. Qualquer família de probabilidades de transição $P(x, \cdot)$ em X fica completamente caracterizada pelos valores

$$P_{i,j} = P(i, \{j\}), \quad 1 \leq i, j \leq d. \quad (7.2.3)$$

Além disso, uma medida p em X fica completamente caracterizada pelos valores $p_i = p(\{i\})$, $1 \leq i \leq d$. Com esta notação, a definição (7.2.1) traduz-se por

$$\sum_{i=1}^d p_i P_{i,j} = p_j, \quad \text{para todo } 1 \leq j \leq d. \quad (7.2.4)$$

Além disso, a medida de Markov μ fica dada por

$$\mu([m; a_m, \dots, a_n]) = p_{a_m} P_{a_m, a_{m+1}} \cdots P_{a_{n-1}, a_n}. \quad (7.2.5)$$

Na sequência iremos nos restringir a deslocamentos de Markov *finitos*, ou seja, ao contexto do Exemplo 7.2.2. Consideramos o conjunto X munido da topologia discreta e da respectiva σ -álgebra de Borel. Observe que a matriz

$$P = (P_{i,j})_{1 \leq i,j \leq d}$$

definida por (7.2.3) satisfaz as seguintes condições:

- (i) $P_{i,j} \geq 0$ para todo $1 \leq i \leq j$;
- (ii) $\sum_{j=1}^d P_{i,j} = 1$ para todo $1 \leq i \leq d$.

Dizemos que P é uma *matriz estocástica*. Reciprocamente, qualquer matriz satisfazendo (i) e (ii) define uma família de probabilidades de transição no conjunto X . Observe também que, denotando $p = (p_1, \dots, p_d)$, a relação (7.2.4) corresponde a

$$P^*p = p, \tag{7.2.6}$$

onde P^* representa a transposta da matriz P . Em outras palavras, as medidas estacionárias correspondem precisamente aos autovetores da matriz transposta para o autovalor 1. O seguinte resultado clássico permite mostrar que tais autovalores sempre existem:

Teorema 7.2.3 (Perron-Fröbenius). *Seja A uma matriz $d \times d$ com entradas não-negativas. Então existe $\lambda \geq 0$ e existe algum vetor $v \neq 0$ com entradas não-negativas tal que $Av = \lambda v$ e $\lambda \geq |\gamma|$ para todo autovalor γ de A .*

Se A admite alguma potência cujas entradas são positivas então $\lambda > 0$ e existe algum autovetor v com entradas positivas. De fato, $\lambda > |\gamma|$ para qualquer outro autovalor γ de A . Além disso, o autovalor λ tem multiplicidade 1 e é o único autovalor de A que admite algum autovetor com entradas não-negativas.

Uma demonstração do teorema de Perron-Fröbenius pode ser encontrada no livro de Meyers [Mey00], por exemplo. Aplicando este teorema à matriz $A = P^*$, concluímos que existem $\lambda \geq 0$ e $p \neq 0$ com $p_i \geq 0$ para todo i , tais que

$$\sum_{i=1}^d p_i P_{i,j} = \lambda p_j, \quad \text{para todo } 1 \leq j \leq d.$$

Somando sobre $i = 1, \dots, d$ obtemos que

$$\sum_{j=1}^d \sum_{i=1}^d p_i P_{i,j} = \lambda \sum_{j=1}^d p_j.$$

Usando a propriedade (ii) da matriz estocástica, o lado esquerdo desta igualdade pode ser escrito como

$$\sum_{i=1}^d p_i \sum_{j=1}^d P_{i,j} = \sum_{i=1}^d p_i.$$

Comparando as duas últimas igualdades, e lembrando que a soma das entradas de p é um número positivo, concluímos que $\lambda = 1$. Isto prova a nossa afirmação de que sempre existem vetores $p \neq 0$ satisfazendo (7.2.6).

Quando P^n tem entradas positivas para algum $n \geq 1$, segue do Teorema 7.2.3 que o autovetor é único, a menos de produto por um escalar, e pode ser escolhido com entradas positivas.

Exemplo 7.2.4. Em geral, p não é único e também pode não existir nenhum autovalor com entradas positivas. Por exemplo, considere:

$$P = \begin{pmatrix} 1-a & a & 0 & 0 & 0 \\ b & 1-b & 0 & 0 & 0 \\ 0 & 0 & 1-c & c & 0 \\ 0 & 0 & d & 1-d & 0 \\ e & 0 & 0 & 0 & 1-e \end{pmatrix}$$

onde $a, b, c, d, e \in (0, 1)$. Um vetor $p = (p_1, p_2, p_3, p_4, p_5)$ satisfaz $P^*p = p$ se, e somente se, $ap_1 = bp_2$ e $cp_3 = dp_4$ e $p_5 = 0$. Portanto, o autoespaço tem dimensão 2 e nenhum autovetor tem entradas positivas.

Por outro lado, suponha que p é tal que $p_i = 0$ para algum i e seja μ a respectiva medida de Markov. Seja $\Sigma_i = (X \setminus \{i\})^{\mathbb{N}}$ (ou $\Sigma_i = (X \setminus \{i\})^{\mathbb{Z}}$). Então $\mu(\Sigma_i) = 1$, uma vez que $\mu([n; i]) = p_i = 0$ para todo n . Isto significa que podemos eliminar o símbolo i , obtendo um sistema equivalente ao original. Portanto, a menos de remover um certo número de símbolos supérfluos do conjunto X , sempre podemos considerar que o autovetor p tem entradas positivas.

Denotemos por Σ_P o conjunto de todas as sequências $(x_n)_n \in \Sigma$ que satisfazem

$$P_{x_n, x_{n+1}} > 0 \quad \text{para todo } n, \quad (7.2.7)$$

ou seja, tais que todas as transições são “permitidas” por P . É claro da definição que Σ_P é invariante pelo deslocamento σ . As transformações $\sigma : \Sigma_P \rightarrow \Sigma_P$ construídas desta forma são chamadas *deslocamentos de tipo finito* e serão estudadas em mais detalhe na Seção 10.2.2.

Lema 7.2.5. *O conjunto Σ_P é fechado em Σ e, dada qualquer solução p de $P^*p = p$ com entradas positivas, o suporte da respectiva medida de Markov μ coincide com Σ_P .*

Demonstração. Seja $x^k = (x_n^k)_n$, $k \in \mathbb{N}$ uma sequência qualquer em Σ_P e suponha que ela converge em Σ para algum $x = (x_n)_n$. pela definição da topologia em Σ , isto quer dizer que para todo n existe $k_n \geq 1$ tal que $x_n^k = x_n$ para todo $k \geq k_n$. Então, dado qualquer n , tomando $k \geq \max\{k_n, k_{n+1}\}$ concluímos que $P_{x_n, x_{n+1}} = P_{x_n^k, x_{n+1}^k} > 0$. Isto mostra que $x \in \Sigma_P$.

Para provar a segunda parte do lema, lembre que os cilindros $[m; x_m, \dots, x_n]$ formam uma base de vizinhanças de qualquer $x = (x_n)_n$ em Σ . Se $x \in \Sigma_P$ então

$$\mu([m; x_m, \dots, x_n]) = p_{x_m} P_{x_m, x_{m+1}} \cdots P_{x_{n-1}, x_n} > 0$$

para todo cilindro e , portanto, $x \in \text{supp } \mu$. Se $x \notin \Sigma_P$ então existe n tal que $P_{x_n, x_{n+1}} = 0$. Nesse caso, $\mu([n; x_n, x_{n+1}]) = 0$ e, portanto, $x \notin \text{supp } \mu$. \square

Exemplo 7.2.6. Na situação do Exemplo 7.2.4 há três possibilidades para o suporte de uma medida de Markov. Se $p = (p_1, p_2, 0, 0, 0)$ com $p_1, p_2 > 0$ então podemos eliminar os símbolos 3, 4, 5. Todas as sequências nos símbolos 1, 2 são admissíveis. Logo $\text{supp } \mu = \{1, 2\}^{\mathbb{N}}$. Analogamente, se $p = (0, 0, p_3, p_4, 0)$ com $p_3, p_4 > 0$ então $\text{supp } \mu = \{3, 4\}^{\mathbb{N}}$. Nos demais casos, $p = (p_1, p_2, p_3, p_4, 0)$ com $p_1, p_2, p_3, p_4 > 0$. Eliminando o símbolo 5, temos que o conjunto das sequências admissíveis é

$$\Sigma_P = \{1, 2\}^{\mathbb{N}} \cup \{3, 4\}^{\mathbb{N}}.$$

Ambos os conjuntos nesta união têm medida positiva. Portanto, neste caso o deslocamento de Markov (σ, μ) não é ergódico. Mas segue da teoria que apresentaremos na próxima seção que nos dois primeiros casos o sistema (σ, μ) é ergódico.

No próximo lema colecionamos algumas propriedades simples de matrizes estocásticas que serão úteis a seguir:

Lema 7.2.7. *Seja P uma matriz estocástica e seja $p = (p_1, \dots, p_d)$ uma solução de $P^*p = p$. Para cada $n \geq 0$, denote por $P_{i,j}^n$, $1 \leq i, j \leq d$ as entradas da matriz P^n . Então:*

- (a) $\sum_{j=1}^d P_{i,j}^n = 1$ para todo $1 \leq i \leq d$ e todo $n \geq 1$;
- (b) $\sum_{i=1}^d p_i P_{i,j}^n = p_j$ para todo $1 \leq j \leq d$ e todo $n \geq 1$;
- (c) o hiperplano $H = \{(h_1, \dots, h_d) : h_1 + \dots + h_d = 0\}$ é invariante por P^* .

Demonstração. A condição (ii) na definição de matriz estocástica pode ser escrita como $Pu = u$, onde $u = (1, \dots, 1)$. Então $P^n u = u$ para todo $n \geq 1$. Isto é apenas outra maneira de escrever o item (a). Analogamente, $P^*p = p$ implica que $(P^*)^n p = p$ para todo $n \geq 1$, o que é outra maneira de escrever o item (b). Observe que H é o complemento ortogonal do vetor u . Como u é invariante por P , segue que H é invariante pela matriz transposta P^* . \square

7.2.1 Ergodicidade

Nesta seção sempre suporemos que $p = (p_1, \dots, p_d)$ é uma solução de $P^*p = p$ com $p_i > 0$ para todo i , normalizada de tal forma que $\sum_i p_i = 1$. Seja μ a respectiva medida de Markov. Queremos entender que condições a matriz estocástica P deve satisfazer para que (σ, μ) seja ergódico.

Dizemos que a matriz estocástica P é *irreduzível* se para todo $1 \leq i, j \leq d$ existe $n \geq 0$ tal que $P_{i,j}^n > 0$. Em outras palavras, P é irreduzível se for possível passar de qualquer resultado i a qualquer resultado j num certo número n de passos (que depende de i e j).

Teorema 7.2.8. *O deslocamento de Markov (σ, μ) é ergódico se, e somente se, a matriz P é irreduzível.*

O restante da presente seção é dedicado à prova deste teorema. Começamos por provar a seguinte estimativa útil:

Lema 7.2.9. *Sejam $A = [m; a_m, \dots, a_q]$ e $B = [r; b_r, \dots, b_s]$ cilindros de Σ com $r > q$. Então:*

$$\mu(A \cap B) = \mu(A)\mu(B) \frac{P_{a_q, b_r}^{r-q}}{p_{b_r}}.$$

Demonstração. Podemos escrever $A \cap B$ como uma união disjunta

$$A \cap B = \bigcup_x [m; a_m, \dots, a_q, x_{n+1}, \dots, x_{r-1}, b_r, \dots, b_s],$$

sobre todos os $x = (x_{n+1}, \dots, x_{r-1}) \in X^{r-n-1}$. Então,

$$\begin{aligned} \mu(A \cap B) &= \sum_x p_{a_m} P_{a_m, a_{m+1}} \cdots P_{a_{q-1}, a_q} P_{a_q, x_{n+1}} \cdots P_{x_{r-1}, b_r} P_{b_r, b_{r+1}} \cdots P_{b_{s-1}, b_s} \\ &= \mu(A) \sum_x P_{a_q, x_{n+1}} \cdots P_{x_{r-1}, b_r} \frac{1}{p_{b_r}} \mu(B). \end{aligned}$$

A soma nesta última expressão é igual a P_{a_q, b_r}^{r-q} . Portanto,

$$\mu(A \cap B) = \mu(A)\mu(B) P_{a_q, b_r}^{r-q} / p_{b_r},$$

tal como afirmado. \square

Lema 7.2.10. *Uma matriz estocástica P é irredutível se, e somente se,*

$$\lim_n \frac{1}{n} \sum_{l=0}^{n-1} P_{i,j}^l = p_j \quad \text{para todo } 1 \leq i, j \leq d. \quad (7.2.8)$$

Demonstração. Suponha que vale (7.2.8). Lembre que $p_j > 0$ para todo j . Então, dados quaisquer $1 \leq i, j \leq d$, tem-se $P_{i,j}^l > 0$ para infinitos valores de l . Em particular, P é irredutível.

Para provar a recíproca, considere $A = [0; i]$ e $B = [1; j]$. Pelo Lema 7.2.9:

$$\frac{1}{n} \sum_{l=0}^{n-1} \mu(A \cap \sigma^{-l}(B)) = \frac{1}{p_j} \mu(A)\mu(B) \frac{1}{n} \sum_{l=0}^{n-1} P_{i,j}^l.$$

De acordo com o Exercício 4.1.2, o lado esquerdo desta igualdade converge quando $n \rightarrow \infty$. Portanto,

$$Q_{i,j} = \lim_n \frac{1}{n} \sum_{l=0}^{n-1} P_{i,j}^l$$

existe para todo $1 \leq i, j \leq d$. Considere a matriz $Q = (Q_{i,j})_{i,j}$, ou seja,

$$Q = \lim_n \frac{1}{n} \sum_{l=0}^{n-1} P^l. \quad (7.2.9)$$

Usando o Lema 7.2.7(b) e passando ao limite quando $n \rightarrow \infty$, obtemos que

$$\sum_{i=1}^d p_i Q_{i,j} = p_j \quad \text{para todo } 1 \leq j \leq d. \quad (7.2.10)$$

Observe também que, dado qualquer $k \geq 1$,

$$P^k Q = \lim_n \frac{1}{n} \sum_{l=0}^{n-1} P^{k+l} = \lim_n \frac{1}{n} \sum_{l=0}^{n-1} P^l = Q. \quad (7.2.11)$$

Segue que $Q_{i,j}$ não depende de i . De fato, suponha que existem r e s tais que $Q_{r,j} < Q_{s,j}$. Claro que podemos escolher s de modo que o lado direito desta desigualdade tenha o valor máximo possível. Como P é irredutível, existe k tal que $P_{s,r}^k > 0$. Logo, usando (7.2.11) seguido do Lema 7.2.7(a),

$$Q_{s,j} = \sum_{i=1}^d P_{s,i}^k Q_{i,j} < \left(\sum_{i=1}^d P_{s,i}^k \right) Q_{s,j} = Q_{s,j},$$

o que é uma contradição. Portanto $Q_{i,j}$ não depende de i . Seja $Q_j = Q_{i,j}$ para qualquer i . A propriedade (7.2.10) dá que

$$p_j = \sum_{i=1}^d Q_{i,j} p_i = Q_j \left(\sum_{i=1}^d p_i \right) = Q_j,$$

para todo j . Isto termina a prova do lema. \square

Demonstração do Teorema 7.2.8. Suponhamos que μ é ergódica. Seja $A = [0; i]$ e $B = [1; j]$. Pela Proposição 4.1.4,

$$\lim_n \frac{1}{n} \sum_{l=0}^{n-1} \mu(A \cap \sigma^{-l}(B)) = \mu(A)\mu(B) = p_i p_j. \quad (7.2.12)$$

Por outro lado, pelo Lema 7.2.9, temos que $\mu(A \cap \sigma^{-l}(B)) = p_i P_{i,j}^l$. Substituindo em (7.2.12) e cancelando p_i em ambos os lados,

$$\lim_n \frac{1}{n} \sum_{l=0}^{n-1} P_{i,j}^l = p_j.$$

Note que j é arbitrário. Portanto, pelo Lema 7.2.10, isto prova que P é irredutível.

Agora suponhamos que a matriz P é irredutível. Queremos concluir que μ é ergódica. De acordo com o Corolário 4.1.5, basta provar que

$$\lim_n \frac{1}{n} \sum_{l=0}^{n-1} \mu(A \cap \sigma^{-l}(B)) = \mu(A)\mu(B) \quad (7.2.13)$$

para quaisquer A e B na álgebra gerada pelos cilindros. Como os elementos desta álgebra são as uniões finitas disjuntas de cilindros, é suficiente considerar o caso em que A e B são cilindros, digamos $A = [m; a_m, \dots, a_q]$ e $B = [r; b_r, \dots, b_s]$. Observe também que a validade de (7.2.13) não muda se substituirmos B por alguma pré-imagem $\sigma^{-j}(B)$. Portanto, não é restrição supormos que $r > q$. Então, pelo Lema 7.2.9,

$$\frac{1}{n} \sum_{l=0}^{n-1} \mu(A \cap \sigma^{-l}(B)) = \mu(A)\mu(B) \frac{1}{p_{b_r}} \frac{1}{n} \sum_{l=0}^{n-1} P_{a_q, b_r}^{r-q+l}$$

para todo n . Pelo Lema 7.2.10,

$$\lim_n \frac{1}{n} \sum_{l=0}^{n-1} P_{a_q, b_r}^{r-q+l} = \lim_n \frac{1}{n} \sum_{l=0}^{n-1} P_{a_q, b_r}^l = p_{b_r}.$$

Isto prova a propriedade (7.2.13) para os cilindros A e B . \square

7.2.2 Mistura

Nesta seção caracterizamos os deslocamentos de Markov misturadores, em termos da respectiva matriz estocástica P . Continuamos supondo que p é uma solução normalizada de $P^*p = p$ com entradas positivas e que μ é a respectiva medida de Markov.

Dizemos que a matriz estocástica P é *aperiódica* se existe $n \geq 1$ tal que $P_{i,j}^n > 0$ para todo $1 \leq i, j \leq d$. Em outras palavras, P é aperiódica se tem alguma potência P^n com entradas positivas. A relação entre as noções de aperiodicidade e irreducibilidade é analisada no Exercício 7.2.6.

Teorema 7.2.11. *O deslocamento de Markov (σ, μ) é misturador se, e somente se, a matriz P é aperiódica.*

Para a demonstração do Teorema 7.2.11 precisamos do seguinte fato:

Lema 7.2.12. *Uma matriz estocástica P é aperiódica se, e somente se,*

$$\lim_l P_{i,j}^l = p_j \quad \text{para todo } 1 \leq i, j \leq d. \quad (7.2.14)$$

Demonstração. Como $p_j > 0$ para todo j , por hipótese, é claro que (7.2.14) implica que $P_{i,j}^l > 0$ para todo i, j e todo l suficientemente grande.

Agora suponha que P é aperiódica. Então podemos aplicar o teorema de Perron-Frobenius (Teorema 7.2.3) à matriz $A = P^*$. Como p é autovetor de A com entradas positivas, obtemos que $\lambda = 1$ e todos os demais autovalores de A são menores que 1 em valor absoluto. Pelo Lema 7.2.7(c), o hiperplano H dos vetores (h_1, \dots, h_d) tais que $h_1 + \dots + h_d = 0$ é invariante por A . É claro que H é transversal à direção de p . Então a decomposição

$$\mathbb{R}^d = \mathbb{R}p \oplus H \quad (7.2.15)$$

é invariante por A e a restrição de A ao hiperplano H é uma contração, no sentido de que o seu raio espectral é menor que 1. Segue deste fato que a sequência $(A^l)_l$ converge para a projeção na primeira coordenada de (7.2.15), ou seja, para a matriz B caracterizada por $Bp = p$ e $Bh = 0$ para todo $h \in H$. Em outras palavras, $(P^l)_l$ converge para B^* . Observe que

$$B_{i,j} = p_i \quad \text{para todo } 1 \leq i, j \leq d.$$

Portanto, $\lim_n P_{i,j}^l = B_{j,i} = p_j$ para todo i, j . \square

Demonstração do Teorema 7.2.11. Suponhamos que a medida μ é misturadora. Seja $A = [0; i]$ e $B = [1; j]$. Pelo Lema 7.2.9, temos que $\mu(A \cap \sigma^{-l}(B)) = p_i P_{i,j}^l$ para todo l . Portanto,

$$p_i \lim_l P_{i,j}^l = \lim_l \mu(A \cap \sigma^{-l}(B)) = \mu(A)\mu(B) = p_i p_j.$$

Cancelando p_i em ambos os lados, obtemos que $\lim_l P_{i,j}^l = p_j$. De acordo com o Lema 7.2.12, isto prova que P é aperiódica.

Agora suponhamos que a matriz P é aperiódica. Queremos concluir que μ é misturadora. De acordo com o Lema 7.1.2, basta provar que

$$\lim_l \mu(A \cap \sigma^{-l}(B)) = \mu(A)\mu(B) \quad (7.2.16)$$

para quaisquer A e B na álgebra gerada pelos cilindros. Uma vez que os elementos desta álgebra são as uniões finitas disjuntas de cilindros, basta tratar o caso em que A e B são cilindros, digamos $A = [m; a_m, \dots, a_q]$ e $B = [r; b_r, \dots, b_s]$. Pelo Lema 7.2.9,

$$\mu(A \cap \sigma^{-l}(B)) = \mu(A)\mu(B) \frac{1}{p_{b_r}} P_{a_q, b_r}^{r-q+l}$$

para todo $l > q - r$. Então, usando o Lema 7.2.12,

$$\begin{aligned} \lim_l \mu(A \cap \sigma^{-l}(B)) &= \mu(A)\mu(B) \frac{1}{p_{b_r}} \lim_l P_{a_q, b_r}^{r-q+l} \\ &= \mu(A)\mu(B) \frac{1}{p_{b_r}} \lim_l P_{a_q, b_r}^l = \mu(A)\mu(B) \end{aligned}$$

Isto prova a propriedade (7.2.16) para os cilindros A e B . \square

Exemplo 7.2.13. No Exemplo 7.2.4 obtivemos diversos tipos de medidas de Markov, dependendo da escolha do autovetor p . No primeiro, $p = (p_1, p_2, 0, 0, 0)$ e a medida μ está suportada em $\{1, 2\}^{\mathbb{N}}$. Uma vez removidos os símbolos supérfluos 3, 4, 5, a matriz estocástica se reduz a

$$P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}.$$

Como esta matriz é irredutível, a medida de Markov μ é misturadora. O segundo tipo é inteiramente análogo. No terceiro caso $p = (p_1, p_2, p_3, p_4, 0)$ e, removido o símbolo supérfluo 5, a matriz estocástica se reduz a

$$P = \begin{pmatrix} 1-a & a & 0 & 0 \\ b & 1-b & 0 & 0 \\ 0 & 0 & 1-c & c \\ 0 & 0 & d & 1-d \end{pmatrix}.$$

Esta matriz não é irredutível e, portanto, as medidas de Markov que se obtêm neste caso não são ergódicas (lembre também do Exemplo 7.2.6).

Exemplo 7.2.14. Não é difícil dar exemplos de matrizes irredutíveis que não são aperiódicas:

$$P = \begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix}.$$

De fato, vemos que $P_{i,j}^n > 0$ se, e somente se, n tem a mesma paridade que $i - j$. Note que

$$P^2 = \begin{pmatrix} 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{pmatrix}.$$

O Exercício 7.2.6 mostra que toda matriz irredutível tem uma forma deste tipo.

7.2.3 Exercícios

7.2.1. Seja $X = \{1, \dots, d\}$, seja $P = (P_{i,j})_{i,j}$ uma matriz estocástica e seja $p = (p_i)_i$ um vetor tal que $P^*p = p$. Mostre que a definição

$$\mu([m; a_m, \dots, a_n]) = p_{a_m} P_{a_m, a_{m+1}} \cdots P_{a_{n-1}, a_n}$$

se estende a uma medida na σ -álgebra boreliana de $\Sigma = X^{\mathbb{N}}$ (ou $\Sigma = X^{\mathbb{Z}}$), invariante pelo deslocamento $\sigma : \Sigma \rightarrow \Sigma$.

7.2.2. Prove que um deslocamento de Markov fracamente misturador é, de fato, misturador.

7.2.3. Seja μ uma medida de Markov para um deslocamento $\sigma : X^{\mathbb{Z}} \rightarrow X^{\mathbb{Z}}$, com $X = \{1, \dots, d\}$. Isso garante que μ também é uma medida de Markov para a inversa $\sigma^{-1} : \Sigma \rightarrow \Sigma$?

7.2.4. Seja X um conjunto finito e $\Sigma = X^{\mathbb{Z}}$ (ou $\Sigma = X^{\mathbb{N}}$). Seja μ uma probabilidade em Σ , invariante pelo deslocamento $\sigma : \Sigma \rightarrow \Sigma$. Dado $k \geq 0$, dizemos que μ tem *memória* k se

$$\frac{\mu([m-l; a_{m-l}, \dots, a_{m-1}, a_m])}{\mu([m-l; a_{m-l}, \dots, a_{m-1}])} = \frac{\mu([m-k; a_{m-k}, \dots, a_{m-1}, a_m])}{\mu([m-k; a_{m-k}, \dots, a_{m-1}])}$$

para todo $l \geq k$, todo m e todo $(a_n)_n \in \Sigma$ (por convenção, a igualdade vale sempre que um dos denominadores for zero). Verifique que as medidas com memória zero são as medidas de Bernoulli e as medidas com memória um são as medidas de Markov. Mostre que toda medida com memória $k \geq 2$ é equivalente a uma medida de Markov no espaço $\tilde{\Sigma} = \tilde{X}^{\mathbb{Z}}$ (ou $\tilde{\Sigma} = \tilde{X}^{\mathbb{N}}$), onde $\tilde{X} = X^k$.

7.2.5. O objetivo é mostrar que o conjunto das medidas com memória finita é denso no espaço $\mathcal{M}_1(\sigma)$ das medidas invariantes pelo deslocamento $\sigma : \Sigma \rightarrow \Sigma$. Dada qualquer medida invariante μ e dado $k \geq 1$, considere a função μ_k definida no conjunto de todos os cilindros por

- $\mu_k = \mu$ para cilindros com comprimento menor ou igual a k ;
- para todo $l \geq k$, todo m e todo $(a_n)_n \in \Sigma$,

$$\frac{\mu_k([m-l; a_{m-l}, \dots, a_{m-1}, a_m])}{\mu_k([m-l; a_{m-l}, \dots, a_{m-1}])} = \frac{\mu([m-k; a_{m-k}, \dots, a_{m-1}, a_m])}{\mu([m-k; a_{m-k}, \dots, a_{m-1}])}$$

Mostre que μ_k se estende a uma medida de probabilidade na σ -álgebra de Borel de Σ , invariante pelo deslocamento e com memória k . Mostre também que $\lim_k \mu_k = \mu$ na topologia fraca*.

7.2.6. Seja P uma matriz estocástica irredutível. O objetivo é mostrar que existem $\kappa \geq 1$ e uma partição de X em κ subconjuntos, tais que a restrição de P^κ a cada um destes subconjuntos é aperiódica. Para isso:

- (1) Para cada $i \in X$, defina $R(i) = \{n \geq 1 : P_{i,i}^n > 0\}$. Mostre que $R(i)$ é fechado por adição: se $n_1, n_2 \in R(i)$ então $n_1 + n_2 \in R(i)$.
- (2) Seja $R \subset \mathbb{N}$ um conjunto fechado por adição e seja $\kappa \geq 1$ o máximo divisor comum dos seus elementos. Mostre que existe $m \geq 1$ tal que $R \cap [m, \infty) = \kappa\mathbb{N} \cap [m, \infty)$.
- (3) Mostre que o máximo divisor comum κ dos elementos de $R(i)$ não depende de $i \in X$ e que P é aperiódica se, e somente se, $\kappa = 1$.
- (4) Suponha que $\kappa \geq 2$. Encontre uma partição $\{X_r : 0 \leq r < \kappa\}$ de X tal que a restrição de P^κ a cada X_r é aperiódica.

7.3 Intercâmbios de intervalos

Por definição, um *intercâmbio de intervalos* é uma bijeção do intervalo $[0, 1)$ com um número finito de descontinuidades e cuja restrição a cada subintervalo de continuidade é uma translação. Veja na Figura 7.1 um exemplo com 4 subintervalos de continuidade. Para fixar ideias, sempre suporemos que a transformação é contínua à direita, ou seja, que os subintervalos de continuidade são fechados à esquerda e abertos à direita.

Como consequência direta da definição, todo intercâmbio de intervalo preserva a medida de Lebesgue em $[0, 1)$. Estas transformações exibem uma dinâmica

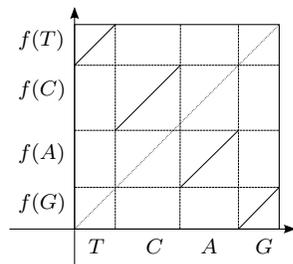


Figura 7.1: Intercâmbio de intervalos

muito rica e têm conexões importantes com diversos outros sistemas, tais como os bilhares em mesas poligonais, os fluxos conservativos em superfícies e os fluxos de Teichmüller. Por exemplo, a construção que esboçamos a seguir mostra que os intercâmbios de intervalos aparecem naturalmente como transformações de Poincaré de campos de vetores conservativos em superfícies.

Exemplo 7.3.1. Seja S uma superfície orientável e seja ω uma forma de área em S , ou seja, uma 2-forma diferencial que nunca se anula. Podemos associar a todo campo de vetores X uma 1-forma diferencial β , definida por

$$\beta_x(v) = \omega_x(X(x), v) \quad \text{para cada vetor } v \in T_x S.$$

Observe que X e β têm os mesmos zeros e que nos demais pontos o núcleo de β coincide a direção do campo de vetores. A 1-forma β permite definir um “comprimento transversal” de curvas $c : [a, b] \rightarrow S$, da seguinte forma:

$$\ell(c) = \int_c \beta = \int_a^b \beta_{c(t)}(\dot{c}(t)) dt.$$

Note que as trajetórias do fluxo têm comprimento transversal nulo mas para curvas transversais ao fluxo, ℓ é equivalente à medida de Lebesgue na curva. Pode mostrar-se que a 1-forma β é fechada se, e somente se, X preserva área (veja o Exercício 7.3.1). Então, usando o teorema de Green, as transformações de Poincaré do fluxo preservam o comprimento transversal. Com uma hipótese adicional sobre os zeros de X , a transformação $f : \Sigma \rightarrow \Sigma$ de primeiro retorno a uma seção transversal qualquer está definida e é contínua exceto num número finito de pontos da seção Σ . Então, parametrizando Σ pelo comprimento transversal de arco, f é um intercâmbio de intervalos.

Cada intercâmbio de intervalos fica determinado por dois ingredientes. O primeiro deles, de natureza combinatória, diz respeito ao número de subintervalos de continuidade e à ordem destes subintervalos e de suas imagens dentro do intervalo $[0, 1)$. Isso pode ser informado da seguinte forma: atribuímos a cada subintervalo um rótulo (uma letra, por exemplo), atribuímos à sua imagem o

mesmo rótulo e listamos esses rótulos em duas linhas horizontais, na ordem respectiva. Por exemplo, no caso da Figura 7.1, obtemos

$$\pi = \begin{pmatrix} T & C & A & G \\ G & A & C & T \end{pmatrix}.$$

Note que a escolha dos rótulos é arbitrária. Denotamos por \mathcal{A} , e chamamos de *alfabeto*, o conjunto dos rótulos escolhidos.

O segundo ingrediente, de natureza métrica, diz respeito aos comprimentos dos subintervalos. Ele pode ser informado por meio de um vetor de coordenadas positivas, indexadas pelos rótulos escolhidos e que determinam o comprimento do intervalo respectivo (e também da sua imagem). No caso da Figura 7.1 este *vetor de comprimentos* tem a forma

$$\lambda = (\lambda_T, \lambda_C, \lambda_A, \lambda_G).$$

A soma das coordenadas do vetor de comprimentos é sempre igual a 1.

Então, o intercâmbio de intervalos $f : [0, 1) \rightarrow [0, 1)$ associada a cada par (π, λ) fica definida do seguinte modo. Para cada rótulo $\alpha \in \mathcal{A}$, represente por I_α o respectivo subintervalo e defina $w_\alpha = v_1 - v_0$, onde v_0 é a soma dos comprimentos λ_β correspondentes aos rótulos β à esquerda de α na linha superior de π e v_1 é a soma dos comprimentos λ_γ correspondentes aos rótulos γ à esquerda de α na linha inferior de π . Então

$$f(x) = x + w_\alpha \quad \text{para todo } x \in I_\alpha.$$

O vetor $w = (w_\alpha)_{\alpha \in \mathcal{A}}$ é chamado *vetor de translações*. Observe que, para cada π fixado, o vetor de translações é uma função linear do vetor de comprimentos $\lambda = (\lambda_\alpha)_{\alpha \in \mathcal{A}}$.

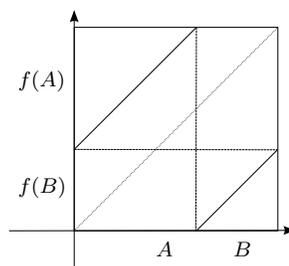


Figura 7.2: Rotação vista como intercâmbio de dois intervalos

Exemplo 7.3.2. O caso mais simples de intercâmbio de intervalos ocorre quando temos apenas dois subintervalos de continuidade. Veja a Figura 7.2. Escolhendo o alfabeto $\mathcal{A} = \{A, B\}$, temos

$$\pi = \begin{pmatrix} A & B \\ B & A \end{pmatrix} \quad \text{e} \quad f(x) = \begin{cases} x + \lambda_B & \text{para } x \in I_A \\ x - \lambda_A = x + \lambda_B - 1 & \text{para } x \in I_B. \end{cases}$$

Esta transformação corresponde precisamente à rotação R_{λ_A} , se identificarmos $[0, 1)$ com o círculo S^1 do modo natural. Neste sentido, os intercâmbios de intervalos são uma generalização das rotações do círculo.

7.3.1 Minimalidade e ergodicidade

Como vimos anteriormente, uma rotação R_θ do círculo é minimal se, e somente se, θ é irracional. Além disso, nesse caso R_θ também é unicamente ergódica. Dado que quase todo número é irracional, isto significa que minimalidade e unicidade ergódica são típicas na família das rotações. Nesta seção vamos discutir como estes fatos se estendem ao contexto mais amplo dos intercâmbios de intervalos.

Começamos por uma observação que não tem análogo no caso das rotações. Dizemos que uma combinatória π é *reduzível* se existe alguma posição tal que os rótulos à esquerda dessa posição nas duas linhas de π são exatamente os mesmos. Por exemplo,

$$\pi = \begin{pmatrix} B & X & O & L & F & D \\ X & O & B & F & D & L \end{pmatrix}$$

é reduzível, pois os rótulos à esquerda da quarta posição são os mesmos nas duas linhas: B , O e X . Como consequência, qualquer que seja o vetor de comprimentos λ , o intercâmbio de intervalos f definido por (π, λ) deixa o subintervalo $I_B \cup I_O \cup I_X$ invariante. Em particular, f não pode ser minimal e nem sequer transitivo. No que segue sempre suporemos que a combinatória π é irredutível.

Uma conjectura natural é que o intercâmbio de intervalos seja minimal sempre que o vetor de comprimentos $\lambda = (\lambda_\alpha)_{\alpha \in \mathcal{A}}$ é racionalmente independente, ou seja, se

$$\sum_{\alpha \in \mathcal{A}} n_\alpha \lambda_\alpha \neq 0$$

para todo vetor $(n_\alpha)_{\alpha \in \mathcal{A}}$ não nulo com coordenadas inteiras. Isto é verdade mas, de fato, a hipótese de independência racional é demasiado forte: vamos apresentar uma condição um pouco mais geral que ainda implica minimalidade.

Denotemos por ∂I_α o extremo inferior de cada subintervalo I_α . Dizemos que um par (π, λ) satisfaz a *condição de Keane* se as trajetórias destes pontos são disjuntas:

$$f^m(\partial I_\alpha) \neq \partial I_\beta \quad \text{para todo } m \geq 1 \text{ e quaisquer } \alpha, \beta \in \mathcal{A} \text{ com } \partial I_\beta \neq 0 \quad (7.3.1)$$

(note que sempre existem $\bar{\alpha}$ e $\bar{\beta}$ tais que $f(\partial I_{\bar{\alpha}}) = 0 = \partial I_{\bar{\beta}}$). Deixamos a demonstração do seguinte lema ao cuidado do leitor (Exercício 7.3.2).

Lema 7.3.3. (1) Se (π, λ) satisfaz a condição de Keane então π é irredutível.

(2) Se π é irredutível e λ é racionalmente independente então (π, λ) satisfaz a condição de Keane.

Como o subconjunto dos vetores racionalmente independentes tem medida de Lebesgue total, obtemos que a condição de Keane é satisfeita para quase todo vetor de comprimentos λ , sempre que π é irredutível.

Exemplo 7.3.4. No caso de dois subintervalos (lembre o Exemplo 7.3.2), o intercâmbio de intervalos tem a forma $f^m(x) = x + m\lambda_B$ módulo números inteiros. Então, a condição de Keane significa que

$$m\lambda_B \neq \lambda_A + n \quad \text{e} \quad \lambda_A + m\lambda_B \neq \lambda_A + n$$

para todo $m \in \mathbb{N}$ e todo $n \in \mathbb{Z}$. É claro que isto acontece se, e somente se, o vetor (λ_A, λ_B) é racionalmente independente.

Exemplo 7.3.5. Para intercâmbios de 3 ou mais intervalos a condição de Keane é estritamente mais fraca do que a independência racional do vetor de comprimentos. Considere, por exemplo,

$$\pi = \begin{pmatrix} A & B & C \\ C & A & B \end{pmatrix}.$$

Então $f^m(x) = x + m\lambda_C$ módulo números inteiros e, portanto, a condição de Keane significa que os números

$$\{m\lambda_C, \lambda_A + m\lambda_C, \lambda_A + \lambda_B + m\lambda_C\} \quad \text{é disjunto de} \quad \{\lambda_A + n, \lambda_A + \lambda_B + n\}$$

para todo $m \in \mathbb{N}$ e $n \in \mathbb{Z}$. De modo equivalente,

$$p\lambda_C \notin \{q, \lambda_A + q\} \quad \text{para todo } p \in \mathbb{Z} \text{ e } q \in \mathbb{Z}.$$

Isto pode valer mesmo quando $(\lambda_A, \lambda_B, \lambda_C)$ é racionalmente dependente.

O seguinte resultado foi obtido por Michael Keane [Kea75]:

Teorema 7.3.6 (Keane). *Se (π, λ) satisfaz a condição de Keane então o intercâmbio de intervalos f é minimal.*

Exemplo 7.3.7. A condição de Keane não é necessária para a minimalidade. Por exemplo, considere o intercâmbio de intervalos definido por (π, λ) , onde

$$\pi = \begin{pmatrix} A & B & C & D \\ D & C & B & A \end{pmatrix}$$

$\lambda_A = \lambda_C$, $\lambda_B = \lambda_D$ e $\lambda_A/\lambda_B = \lambda_C/\lambda_D$ é irracional. Então (π, λ) não satisfaz a condição de Keane e, no entanto, f é minimal.

Conforme observamos anteriormente, toda rotação minimal é também unicamente ergódica. Isto ainda é válido para intercâmbios de 3 intervalos, mas não no caso geral. De fato, Keane deu um exemplo de um intercâmbio de 4 intervalos onde π é irredutível, λ é racionalmente independente e, no entanto, f tem duas probabilidades ergódicas.

Keane conjecturou que, assim mesmo, deveria ser verdade que quase todo intercâmbio de intervalos é unicamente ergódico. Um resultado fundamental, obtido independentemente por Howard Masur [Mas82] e William Veech [Vee82], prova que esta conjectura é verdadeira:

Teorema 7.3.8 (Masur, Veech). *Suponha que π é irredutível. Então, para Lebesgue quase todo vetor de comprimentos λ , o intercâmbio de intervalos definido por (π, λ) é unicamente ergódico.*

Um pouco antes, Michael Keane and Gérard Rauzy [KR80] haviam provado que unicidade ergódica vale para um conjunto residual (segunda categoria de Baire) de vetores de comprimento sempre que a combinatória é irredutível.

7.3.2 Mistura

Os intercâmbios de intervalos fornecem muitos exemplos de sistemas unicamente ergódicos e fracamente misturadores que não são (fortemente) misturadores.

De fato, como acabamos de comentar, pelo teorema de Masur-Veech (Teorema 7.3.8) quase todo intercâmbio de intervalos é unicamente ergódico. Outro teorema profundo, devido a Artur Avila e Giovanni Forni [AF07], afirma que, se excetuarmos as rotações (mais precisamente: os intercâmbios de intervalos com uma única descontinuidade), quase todo o intercâmbio de intervalos é fracamente misturador. A versão topológica deste fato havia sido provada por Arnaldo Nogueira e Donald Rudolph [NR97].

Por outro lado, um resultado de Anatole Katok [Kat80] que discutiremos a seguir afirma que intercâmbios de intervalo nunca são misturadores:

Teorema 7.3.9. *Seja $f : [0, 1) \rightarrow [0, 1)$ um intercâmbio de intervalos e seja μ uma probabilidade invariante. Então (f, μ) não é misturador.*

Demonstração. É claro que podemos supor que μ é ergódica, pois caso contrário a conclusão é óbvia. Se μ tem algum átomo então ela está suportada numa órbita periódica e, portanto, não é misturadora. Logo, podemos supor que μ é não atômica. Denote por m a medida de Lebesgue no intervalo e considere a aplicação

$$h : [0, 1) \rightarrow [0, 1), \quad h(x) = \mu([0, x]).$$

Então h é um homeomorfismo tal que $h_*\mu = m$. Consequentemente, a aplicação $g = h \circ f \circ h^{-1} : [0, 1) \rightarrow [0, 1)$ tem um número finito de pontos de descontinuidade e preserva a medida de Lebesgue. Em particular, a restrição de g a cada intervalo de continuidade é uma translação. Portanto, g também é um intercâmbio de intervalos. É claro que (f, μ) é misturador se, e somente se, (g, m) é misturador. Portanto, para provar o Teorema 7.3.9 não é restrição supor que μ é a medida de Lebesgue m . Faremos isso de ora em diante.

Nosso objetivo é achar um conjunto mensurável X tal que $m(X \cap f^{-n}(X))$ não converge para $m(X)^2$ quando $n \rightarrow \infty$. Seja $d = \#\mathcal{A}$.

Lema 7.3.10. *Todo intervalo $J = [a, b)$ contido em algum I_β admite uma partição $\{J_1, \dots, J_s\}$ em não mais que $d+2$ subintervalos da forma $J_i = [a_i, b_i)$ e admite números naturais $t_1, \dots, t_s \geq 1$ tais que*

(a) $f^n(J_i) \cap J = \emptyset$ para todo $0 < n < t_i$ e $1 \leq i \leq s$;

(b) $f^{t_i} \upharpoonright J_i$ é uma translação para todo $1 \leq i \leq s$;

- (c) $\{f^{t_1}(J_1), \dots, f^{t_s}(J_s)\}$ é uma partição de J ;
- (d) os intervalos $f^n(J_i)$, $1 \leq i \leq s$, $0 \leq n < t_i$ são disjuntos dois-a-dois;
- (e) $\bigcup_{n=0}^{\infty} f^n(J) = \bigcup_{i=1}^s \bigcup_{n=0}^{t_i-1} f^n(J_i)$.

Demonstração. Seja B o conjunto formado pelos pontos extremos a, b de J juntamente com os pontos ∂I_α , $\alpha \in \mathcal{A}$ excetuada a origem. Então $\#B \leq d+1$. Seja $B_J \subset J$ o conjunto dos pontos $x \in J$ para os quais existe $m \geq 1$ tal que $f^m(x) \in B$ e $f^n(x) \notin J$ para todo $0 < n < m$. O fato de que f é injetiva, juntamente com a definição de m , implica que a aplicação

$$B_J \rightarrow B, \quad x \mapsto f^m(x)$$

é injetiva. Em particular, $\#B_J \leq \#B$. Considere a partição de J em subintervalos $J_i = [a_i, b_i]$ com extremos a_i, b_i no conjunto $B_J \cup \{a, b\}$. Esta partição tem no máximo $d+2$ elementos. Pelo teorema de recorrência de Poincaré, para cada i existe $t_i \geq 1$ tal que $f^{t_i}(J_i)$ intersecta J . Tome t_i mínimo com esta propriedade. O item (a) do lema é uma consequência imediata. Pela definição de B_J , a restrição de f^{t_i} ao intervalo J_i é uma translação, tal como afirmado no item (b), e a sua imagem está contida em J . Além disso, as imagens $f^{t_i}(J_i)$, $1 \leq i \leq s$ são disjuntas duas-a-duas, uma vez que f é injetiva e os t_i são os tempos de primeiro retorno a J . Em particular,

$$\sum_{i=1}^s m(f^{t_i}(J_i)) = \sum_{i=1}^s m(J_i) = m(J)$$

e portanto $\bigcup_{i=1}^s f^{t_i}(J_i) = J$. Isto prova o item (c). O item (d) também segue diretamente do fato de que f é injetiva e os t_i são os tempos de primeiro retorno a J . Finalmente, o item (e) é uma consequência direta do item (c). \square

Considere um intervalo J qualquer contido em algum I_β . Por ergodicidade, o conjunto invariante $\bigcup_{n=0}^{\infty} f^n(J)$ tem medida total. Pelo item (e) do Lema 7.3.10, este conjunto é uma união finita de intervalos fechados à esquerda e abertos à direita. Portanto,

$$\bigcup_{n=0}^{\infty} f^n(J) = \bigcup_{i=1}^s \bigcup_{n=0}^{t_i-1} f^n(J_i) = I$$

Logo, pelo item (d) do lema, a família $\mathcal{P}_J = \{f^n(J_i) : 1 \leq i \leq s \text{ e } 0 \leq n < t_i\}$ é uma partição de I .

Lema 7.3.11. *Dados $\delta > 0$ e $N \geq 1$ podemos escolher o intervalo J de tal forma que $\text{diam } \mathcal{P}_J < \delta$ e $t_i \geq N$ para todo i .*

Demonstração. É claro que $\text{diam } f^n(J_i) = \text{diam } J_i \leq \text{diam } J$ para todo i e todo n . Logo, $\text{diam } \mathcal{P}_J < \delta$ desde que escolhamos J com diâmetro menor que δ . Para obter a segunda propriedade no enunciado, tome qualquer ponto $x \in I$ tal que $f^n(x) \neq \partial I_\alpha$ para todo $0 \leq n < N$ e todo $\alpha \in \mathcal{A}$. Afirmamos que $f^n(x) \neq x$

para todo $0 < n < N$. Caso contrário, como f^n é uma translação na vizinhança de x , teríamos $f^n(y) = y$ para todo ponto y nessa vizinhança, contradizendo a hipótese de que (f, m) é ergódica. Isto prova a nossa afirmação. Agora basta tomar $J = [x, x + \varepsilon)$ com $\varepsilon < \min_{0 < n < N} d(x, f^n(x))$ para garantir que $t_i \geq N$ para todo i . \square

Lema 7.3.12. *Para cada $1 \leq i \leq s$ existe $s_i \leq d + 2$ e existem números naturais $\{t_{i,1}, \dots, t_{i,s_i}\}$ tais que $t_{i,j} \geq t_i$ e, dado qualquer conjunto A na álgebra \mathcal{A}_J gerada por \mathcal{P}_J , existe $t_{i,j}$ tal que*

$$m(A \cap f^{-t_{i,j}}(A)) \geq \frac{1}{(d+2)^2} m(A). \quad (7.3.2)$$

Demonstração. Aplicando o Lema 7.3.10 a cada um dos intervalos J_i , $1 \leq i \leq s$ obtemos $s_i \leq d + 2$, uma partição $\{J_{i,j} : 1 \leq j \leq s_i\}$ do intervalo J_i e números naturais $t_{i,j}$ tais que cada $t_{i,j}$ é o tempo de primeiro retorno dos pontos de $J_{i,j}$ a J_i . É claro que $t_{i,j} \geq t_i$, já que t_i é o tempo de primeiro retorno de qualquer ponto de J_i ao intervalo J . O fato de que $J_{i,j} \subset f^{-t_{i,j}}(J_i)$ implica que

$$f^n(J_i) = \bigcup_{j=1}^{s_i} f^n(J_{i,j}) \subset \bigcup_{j=1}^{s_i} f^{-t_{i,j}}(f^n(J_i)) \quad \text{para todo } n \geq 0.$$

Como a álgebra \mathcal{A}_J está formada pelas uniões disjuntas finitas dos intervalos $f^n(J_i)$, $0 \leq n < t_i$, segue que

$$A \subset \bigcup_{i=1}^s \bigcup_{j=1}^{s_i} f^{-t_{i,j}}(A) \quad \text{para todo } A \in \mathcal{A}_J.$$

Em particular, $m(A) \leq \sum_{i=1}^s \sum_{j=1}^{s_i} m(A \cap f^{-t_{i,j}}(A))$. Lembrando que $s \leq d + 2$ e $s_i \leq d + 2$ para todo i , isto implica (7.3.2). \square

Estamos prontos para concluir a prova do Teorema 7.3.9. Para isso, fixemos um conjunto mensurável $X \subset [0, 1)$ com

$$0 < m(X) < \frac{1}{4(d+2)^2}.$$

Pelo Lema 7.3.11, dado qualquer $N \geq 1$ podemos encontrar um intervalo $J \subset [0, 1)$ tal que todos os tempos de retorno $t_i \geq N$ e existe algum $A \in \mathcal{A}_J$ tal que

$$m(X \Delta A) < \frac{1}{4} m(X)^2. \quad (7.3.3)$$

Aplicando o Lema 7.3.12, obtemos que existe $t_{i,j} \geq t_i \geq N$ tal que:

$$\begin{aligned} m(X \cap f^{-t_{i,j}}(X)) &\geq m(A \cap f^{-t_{i,j}}(A)) - 2m(X \Delta A) \\ &\geq \frac{1}{(d+2)^2} m(A) - \frac{1}{2} m(X)^2. \end{aligned}$$

A relação (7.3.3) implica que $m(A) \geq (3/4)m(X)$. Portanto,

$$\begin{aligned} m(X \cap f^{-t_{ij}}(X)) &\geq \frac{3}{4} \frac{1}{(d+2)^2} m(X) - \frac{1}{2} m(X)^2 \\ &\geq 3m(X)^2 - \frac{1}{2} m(X)^2 \geq 2m(X)^2 \end{aligned}$$

Isto prova que $\limsup_n m(X \cap f^{-n}(X)) \geq 2m(X)^2$ e, portanto, o sistema (f, m) não é misturador. \square

7.3.3 Exercícios

7.3.1. Seja ω uma forma de área numa superfície. Seja X um campo de vetores diferenciável em S e seja β a 1-forma diferencial definida em S por $\beta_x = \omega_x(X(x), \cdot)$. Mostre que β é fechada se, e somente se, X preserva área.

7.3.2. Prove o Lema 7.3.3.

7.3.3. Mostre que se (π, λ) satisfaz a condição de Keane então f não tem pontos periódicos. [Observação: Este é um passo na demonstração do Teorema 7.3.6.]

7.3.4. Seja $f : [0, 1] \rightarrow [0, 1]$ uma transformação de intercâmbio de intervalos irredutível e seja $a \in (0, 1)$ o maior dos pontos de descontinuidade de f ou f^{-1} . A *renormalização de Rauzy-Veech* $R(f) : [0, 1] \rightarrow [0, 1]$ é definida por $R(f)(x) = g(ax)/a$, onde g é a transformação de primeiro retorno de f ao intervalo $[0, a)$. Verifique que $R(f)$ é transformação de intercâmbio de intervalos, com o mesmo número de subintervalos de descontinuidade que f , ou menor. Se f é descrita pelos dados (π, λ) , como podemos descrever $R(f)$?

7.3.5. Dado $d \geq 2$ e dada uma bijeção $\sigma : \mathbb{N} \rightarrow \mathbb{N}$ sem pontos periódicos, considere a transformação $f : [0, 1] \rightarrow [0, 1]$ tal que cada $f(x)$ é obtido permutando os dígitos da expansão de x na base d , conforme determinado por σ . Mais precisamente, se $x = \sum_{n=1}^{\infty} a_n d^{-n}$ com $a_n \in \{0, \dots, d-1\}$ então $f(x) = \sum_{n=1}^{\infty} a_{\sigma(n)} d^{-n}$. Mostre que f preserva a medida de Lebesgue m no intervalo e que (f, m) é misturador.

7.4 Decaimento de correlações

Nesta seção discutimos a velocidade de decaimento das sequências de correlações $C_n(\varphi, \psi)$, num sistema misturador. Relacionado com o fato de que estamos lidando com sistemas determinísticos, não seria realista esperar estimativas interessantes que valham para funções arbitrárias. Mas, como veremos, tais estimativas existem em muitos casos importantes, se restringirmos φ, ψ a subconjuntos adequados de funções. Dado que as correlações $(\varphi, \psi) \mapsto C_n(\varphi, \psi)$ são funções bilineares, é natural considerar subconjuntos que são subespaços vetoriais.

Dizemos que (f, μ) tem *decaimento exponencial de correlações* num dado espaço vetorial \mathcal{V} se existe $\lambda < 1$ e para todo $\varphi, \psi \in \mathcal{V}$ existe $A(\varphi, \psi) > 0$ tal que

$$|C_n(\varphi, \psi)| \leq A(\varphi, \psi) \lambda^n \quad \text{para todo } n \geq 1. \quad (7.4.1)$$

Temos noções similares em que a exponencial λ^n é substituída por qualquer outra sequência convergindo para zero.

Para ilustrar a teoria vamos analisar a questão do decaimento de correlações no contexto dos deslocamentos de Markov unilaterais. Isso nos permitirá introduzir diversas ideias que serão úteis posteriormente em situações mais gerais.

Seja $f : M \rightarrow M$ o deslocamento em $M = X^{\mathbb{N}}$, onde $X = \{1, \dots, d\}$ é um conjunto finito. Seja $P = (P_{i,j})_{i,j}$ uma matriz estocástica aperiódica e seja $p = (p_i)_i$ o autovetor positivo de P^* , normalizado por $p_1 + \dots + p_d = 1$. Seja μ a medida de Markov definida em M por (7.2.2).

Considere $L = G^{-1}P^*G$, onde G é a matriz diagonal cujas entradas são p_1, \dots, p_d . As entradas de L são dadas por:

$$L_{i,j} = \frac{p_j}{p_i} P_{j,i} \quad \text{para cada } 1 \leq i, j \leq d.$$

Lembre que denotamos $u = (1, \dots, 1)$ e $H = \{(h_1, \dots, h_d) : h_1 + \dots + h_d = 0\}$. Seja

$$V = \{(v_1, \dots, v_d) : p_1 v_1 + \dots + p_d v_d = 0\}.$$

Então $G(u) = p$ e $G(V) = H$. Lembrando (7.2.15), segue que a decomposição

$$\mathbb{R}^d = \mathbb{R}u \oplus V \tag{7.4.2}$$

é invariante por L e todos os autovalores da restrição de L a V são menores que 1 em valor absoluto. Dizemos que o operador L tem *lacuna espectral*: o maior autovalor é simples e o restante do espectro está contido num disco fechado com raio estritamente menor.

Chamamos *operador de transferência* à aplicação linear \mathcal{L} que a cada função $\psi : M \rightarrow \mathbb{R}$ associa a função $\mathcal{L}\psi : M \rightarrow \mathbb{R}$ definida por

$$\mathcal{L}\psi(x_1, \dots, x_n, \dots) = \sum_{x_0=1}^d L_{x_1, x_0} \psi(x_0, x_1, \dots, x_n, \dots). \tag{7.4.3}$$

Este operador é dual do operador de Koopman U_f , no seguinte sentido:

$$\int \varphi(\mathcal{L}\psi) d\mu = \int (U_f \varphi) \psi d\mu \tag{7.4.4}$$

para quaisquer funções mensuráveis limitadas φ, ψ . Para provar este fato basta considerar o caso em que φ e ψ são funções *localmente constantes*, isto é, funções que dependem apenas de um número finito de coordenadas. Isto é porque toda função mensurável limitada é limite uniforme de alguma sequência de funções localmente constantes. Então, considere funções φ e ψ que dependem apenas das primeiras n coordenadas. Pela definição da medida de Markov,

$$\int \varphi(\mathcal{L}\psi) d\mu = \sum_{a_1, \dots, a_n} p_{a_1} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n} \varphi(a_1, \dots, a_n) \mathcal{L}\psi(a_1, \dots, a_n).$$

Usando a definição do operador de transferência, o lado direito desta expressão é igual a

$$\sum_{a_0, a_1, \dots, a_n} p_{a_0} P_{a_0, a_1} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n} \varphi(a_1, \dots, a_n) \psi(a_0, a_1, \dots, a_n).$$

Observe que $\varphi(a_1, \dots, a_n) = U_f \varphi(a_0, a_1, \dots, a_n)$. Portanto, usando mais uma vez a definição da medida de Markov, esta última expressão é igual a $\int (U_f \varphi) \psi d\mu$. Isto prova a propriedade de dualidade (7.4.4).

Como consequência, podemos escrever a sequência de correlações em termos dos iterados do operador de transferência:

$$C_n(\varphi, \psi) = \int (U_f^n \varphi) \psi - \int \varphi d\mu \int \psi d\mu = \int \varphi (\mathcal{L}^n \psi - \int \psi d\mu) d\mu. \quad (7.4.5)$$

A propriedade $Lu = u$ significa que $\sum_j L_{i,j} = 1$ para todo j . Isto tem a seguinte consequência útil:

$$\sup |\mathcal{L}\psi| \leq \sup |\psi| \quad \text{para todo } \psi. \quad (7.4.6)$$

Tomando $\varphi \equiv 1$ em (7.4.4) obtemos o seguinte caso particular, que também será útil a seguir:

$$\int \mathcal{L}\psi d\mu = \int \psi d\mu \quad \text{para todo } \psi. \quad (7.4.7)$$

Agora denotemos por \mathcal{E}_0 o conjunto das funções ψ que dependem apenas da primeira coordenada. A aplicação $\psi \mapsto (\psi(1), \dots, \psi(d))$ é um isomorfismo entre \mathcal{E}_0 e o espaço euclidiano \mathbb{R}^d . Além disso, a definição

$$\mathcal{L}\psi(x_1) = \sum_{x_0=1}^d L_{x_1, x_0} \psi(x_0)$$

mostra que a restrição do operador de transferência a \mathcal{E}_0 corresponde precisamente ao operador $L : \mathbb{R}^d \rightarrow \mathbb{R}^d$. Note ainda que o hiperplano $V \subset \mathbb{R}^d$ corresponde ao subconjunto dos $\psi \in \mathcal{E}_0$ tais que $\int \psi d\mu = 0$. Considere em \mathcal{E}_0 a norma definida por $\|\psi\|_0 = \sup |\psi|$.

Fixe qualquer número λ entre 1 e o raio espectral de L restrito a V . Toda função $\psi \in \mathcal{E}_0$ pode ser escrita:

$$\psi = c + v \quad \text{com } c = \int \psi d\mu \in \mathbb{R}u \quad \text{e } v = \psi - \int \psi d\mu \in V.$$

Então a propriedade de lacuna espectral implica que existe $B > 1$ tal que

$$\sup |\mathcal{L}^n \psi - \int \psi d\mu| \leq B \|\psi\|_0 \lambda^n \quad \text{para todo } n \geq 1. \quad (7.4.8)$$

Usando (7.4.5), segue que

$$|C_n(\varphi, \psi)| \leq B \|\varphi\|_0 \|\psi\|_0 \lambda^n \quad \text{para todo } n \geq 1.$$

Desta forma, provamos que *o deslocamento de Markov tem decaimento exponencial de correlações em \mathcal{E}_0* .

Com um pouco mais de esforço, é possível melhorar este resultado, estendendo a conclusão a um espaço muito maior de funções. Considere em M a distância definida por

$$d((x_n)_n, (y_n)_n) = 2^{-N(x,y)} \quad \text{onde } N(x,y) = \min\{n \geq 0 : x_n \neq y_n\}.$$

Fixe $\theta > 0$ qualquer e denote por \mathcal{E} o conjunto das funções φ que são θ -Hölder, isto é, tais que

$$K_\theta(\varphi) = \sup \left\{ \frac{|\varphi(x) - \varphi(y)|}{d(x,y)^\theta} : x \neq y \right\} \quad \text{é finito.}$$

É claro que \mathcal{E} contém todas as funções localmente constantes. Afirmamos:

Teorema 7.4.1. *O deslocamento de Markov (f, μ) tem decaimento exponencial de correlações no espaço \mathcal{E} das funções θ -Hölder, qualquer que seja $\theta > 0$.*

Observe que $\mathcal{L}(\mathcal{E}) \subset \mathcal{E}$. A função $\|\psi\| = \sup |\psi| + K_\theta(\psi)$ é uma norma completa em \mathcal{E} (Exercício 7.4.1) e o operador linear $\mathcal{L} : \mathcal{E} \rightarrow \mathcal{E}$ é contínuo relativamente a esta norma. Um modo de provar o teorema é mostrando que este operador tem lacuna espectral, com decomposição invariante

$$\mathcal{E} = \mathbb{R}u \oplus \left\{ \psi \in \mathcal{E} : \int \psi d\mu = 0 \right\}.$$

Feito isso, exatamente o mesmo argumento que utilizamos anteriormente para \mathcal{E}_0 prova o decaimento exponencial de correlações em \mathcal{E} . Não apresentaremos os detalhes aqui (mas voltaremos a este tema, num contexto muito mais geral, ao final da Seção 12.3). Em vez disso, vamos dar uma prova direta de que (7.4.8) pode ser estendida ao espaço \mathcal{E} .

Dado $\psi \in \mathcal{E}$ e dado $x = (x_1, \dots, x_n, \dots) \in M$, temos

$$\mathcal{L}^k \psi(x) = \sum_{a_1, \dots, a_k=1}^d L_{x_1, a_k} \cdots L_{a_2, a_1} \psi(a_1, \dots, a_k, x_1, \dots, x_n, \dots)$$

para todo $k \geq 1$. Então, dado $y = (y_1, \dots, y_n, \dots)$ com $x_1 = y_1 = j$,

$$|\mathcal{L}^k \psi(x) - \mathcal{L}^k \psi(y)| \leq \sum_{a_1, \dots, a_k=1}^d L_{j, a_k} \cdots L_{a_2, a_1} K_\theta(\psi) 2^{-k\theta} d(x, y)^\theta.$$

Usando a propriedade $\sum_{i=1}^d L_{j,i} = 1$, concluímos que

$$|\mathcal{L}^k \psi(x) - \mathcal{L}^k \psi(y)| \leq K_\theta(\psi) 2^{-k\theta} d(x, y)^\theta \leq K_\theta(\psi) 2^{-k\theta}. \quad (7.4.9)$$

Dada qualquer função φ , denotamos por $\pi\varphi$ a função que depende apenas da primeira coordenada e coincide com a média de φ em cada cilindro $[0; i]$:

$$\pi\varphi(i) = \frac{1}{p_i} \int_{[0; i]} \varphi d\mu.$$

É claro que $\sup |\pi\varphi| \leq \sup |\varphi|$ e $\int \pi\varphi d\mu = \int \varphi d\mu$. A desigualdade (7.4.9) implica que

$$\sup |\mathcal{L}^k \psi - \pi(\mathcal{L}^k \psi)| \leq K_\theta(\psi) 2^{-k\theta} \quad \text{para todo } k \geq 1.$$

Então, usando a propriedade (7.4.6),

$$\sup |\mathcal{L}^{k+l} \psi - \mathcal{L}^l \pi(\mathcal{L}^k \psi)| \leq K_\theta(\psi) 2^{-k\theta} \quad \text{para todo } k, l \geq 1. \quad (7.4.10)$$

Além disso, as propriedades (7.4.6) e (7.4.7) implicam que

$$\sup |\pi(\mathcal{L}^k \psi)| \leq \sup |\psi| \quad \text{e} \quad \int \pi(\mathcal{L}^k \psi) d\mu = \int \psi d\mu.$$

Portanto a propriedade (7.4.8) dá que

$$\sup |\mathcal{L}^l \pi(\mathcal{L}^k \psi) - \int \psi d\mu| \leq B \sup |\psi| \lambda^l \quad \text{para todo } l \geq 1. \quad (7.4.11)$$

Somando (7.4.10) e (7.4.11), obtemos que

$$\sup |\mathcal{L}^{k+l} \psi - \int \psi d\mu| \leq K_\theta(\psi) 2^{-k\theta} + B \sup |\psi| \lambda^l \quad \text{para todo } k, l \geq 1.$$

Fixe $\sigma < 1$ tal que $\sigma^2 \geq \max\{2^{-\theta}, \lambda\}$. Então a desigualdade anterior dá

$$\sup |\mathcal{L}^n \psi - \int \psi d\mu| \leq B \|\psi\| \sigma^{n-1} \quad \text{para todo } n. \quad (7.4.12)$$

Agora o Teorema 7.4.1 segue do mesmo argumento que usamos anteriormente para \mathcal{E}_0 , com (7.4.12) no lugar de (7.4.8).

7.4.1 Exercícios

7.4.1. Mostre que $\|\varphi\| = \sup |\varphi| + K_\theta(\varphi)$ define uma norma completa no espaço \mathcal{E} das funções θ -Hölder e o operador de transferência \mathcal{L} é contínuo relativamente a esta norma.

7.4.2. Seja $f : M \rightarrow M$ um difeomorfismo local numa variedade compacta M e seja $d \geq 2$ o seu grau. Suponha que existe $\sigma > 1$ tal que $\|Df(x)v\| \geq \sigma\|v\|$ para todo $x \in M$ e todo vector v tangente a M no ponto x . Fixe $\theta > 0$ e seja \mathcal{E} o espaço das funções θ -Hölder $\varphi : M \rightarrow \mathbb{R}$. Para cada $\varphi \in \mathcal{E}$, defina

$$\mathcal{L}\varphi : M \rightarrow \mathbb{R}, \quad \mathcal{L}\varphi(y) = \frac{1}{d} \sum_{x \in f^{-1}(y)} \varphi(x).$$

- (1) Mostre que $\sup \mathcal{L}\varphi \leq \sup \varphi$ e $\inf \mathcal{L}\varphi \geq \inf \varphi$ e $K_\theta(\mathcal{L}\varphi) \leq \sigma^{-\theta} K_\theta(\varphi)$ para toda $\varphi \in \mathcal{E}$.
- (2) Conclua que $\mathcal{L} : E \rightarrow E$ é um operador linear contínuo (relativamente à norma definida no Exercício 7.4.1) com $\|\mathcal{L}\| = 1$.

- (3) Mostre que, para toda $\varphi \in E$, a sequência $(\mathcal{L}^n \varphi)_n$ converge para uma constante $\nu_\varphi \in \mathbb{R}$ quando $n \rightarrow \infty$. Mais ainda, existe $C > 0$ tal que

$$\|\mathcal{L}^n \varphi - \nu_\varphi\| \leq C \sigma^{-n\theta} \|\varphi\| \quad \text{para todo } n \text{ e todo } \varphi \in \mathcal{E}.$$

- (4) Conclua que o operador $\mathcal{L} : \mathcal{E} \rightarrow \mathcal{E}$ tem a propriedade da lacuna espectral.
- (5) Mostre que a aplicação $\varphi \mapsto \nu_\varphi$ se estende a uma medida de probabilidade boreliana em M (lembre do Teorema A.3.12).

Capítulo 8

Sistemas Equivalentes

Neste capítulo discutimos o problema da isomorfismo em Teoria Ergódica: quando é que dois sistemas (f, μ) e (g, ν) devem ser considerados o “mesmo” e como decidir, para sistemas dados, se eles estão nessas condições?

A noção fundamental é chamada *equivalência ergódica*: dois sistemas são ergodicamente equivalentes se, restritas a subconjuntos com medida total, as respectivas transformações são conjugadas por uma aplicação invertível que preserva as medidas invariantes. Tal aplicação permite traduzir as propriedades de qualquer dos sistemas para o outro sistema.

Embora esta seja uma noção natural de isomorfismo no contexto da Teoria Ergódica, ela não é fácil de manejar. Em geral, dados dois sistemas equivalentes, a única forma de provar esse fato é exibindo a equivalência de forma mais ou menos explícita. Por outro lado, o modo mais usual de mostrar que dois sistemas não são equivalentes é encontrando alguma propriedade que está presente num deles mas não no outro.

Neste contexto, é muito útil considerar uma noção mais fraca, chamada de *equivalência espectral*: dois sistemas são espectralmente equivalentes se os seus operadores de Koopman são conjugados por algum operador unitário. Dois sistemas ergodicamente equivalentes são, necessariamente, espectralmente equivalentes, mas a recíproca não é verdadeira.

A ideia de equivalência espectral conduz a uma família rica de invariantes, ligados ao espectro do operador de Koopman, que precisam ser iguais nos dois sistemas para que eles possam ser equivalentes e que, portanto, podem ser usados para excluir essa possibilidade. Outros invariantes, de natureza não espectral, têm igualmente um papel fundamental. O mais importante de todos, a entropia, será tratado no Capítulo 9.

As noções de equivalência ergódica e equivalência espectral, e as relações entre elas, serão estudadas nas Seções 8.1 e 8.2, respectivamente. Nas Seções 8.3 e 8.4 estudaremos duas classes de sistemas com características dinâmicas opostas: as transformações com espectro discreto, que incluem as translações ergódicas em grupos abelianos compactos, e as transformações com espectro de Lebesgue, que têm os deslocamentos de Bernoulli como principal exemplo.

Estas duas classes de sistemas, bem como outras que introduzimos anteriormente (sistemas ergódicos, fortemente misturadores, fracamente misturadores) são invariantes de equivalência espectral e, portanto, também de equivalência ergódica. Finalmente, na Seção 8.5 discutiremos uma terceira noção de equivalência, que chamamos *isomorfismo ergódico*, especialmente no contexto dos espaços de Lebesgue.

8.1 Equivalência ergódica

Sejam μ e ν probabilidades invariantes por transformações $f : M \rightarrow M$ e $g : N \rightarrow N$, respectivamente. Dizemos que os sistemas (f, μ) e (g, ν) são *ergodicamente equivalentes* se podemos escolher conjuntos mensuráveis $X \subset M$ e $Y \subset N$ com $\mu(M \setminus X) = 0$ e $\nu(N \setminus Y) = 0$, e uma bijeção mensurável $\phi : X \rightarrow Y$ com inversa mensurável, de tal forma que

$$\phi_*\mu = \nu \quad \text{e} \quad \phi \circ f = g \circ \phi.$$

Deixamos a cargo do leitor verificar que esta é, de fato, uma relação de equivalência, ou seja, uma relação reflexiva, simétrica e transitiva.

Observe também que os conjuntos X e Y na definição podem ser escolhidos invariantes por f e g , respectivamente. De fato, considere $X_0 = \bigcap_{n=0}^{+\infty} f^{-n}(X)$. É claro da definição que $X_0 \subset X$ e $f(X_0) \subset X_0$. Como $\mu(X) = 1$ e a interseção é enumerável, temos que $\mu(X_0) = 1$. Analogamente, $Y_0 = \bigcap_{n=0}^{+\infty} g^{-n}(Y)$ é subconjunto mensurável de Y tal que $\nu(Y_0) = 1$ e $g(Y_0) \subset Y_0$. Além disso, por construção, $Y_0 = \phi(X_0)$. Portanto, a restrição de ϕ a X_0 ainda é uma bijeção sobre Y_0 .

Exemplo 8.1.1. Considere $f : [0, 1] \rightarrow [0, 1]$ definida por $f(x) = 10x - [10x]$. Como vimos anteriormente (Seção 1.3.1) esta transformação preserva a medida de Lebesgue m em $[0, 1]$. Escrevendo um número $x \in [0, 1]$ em sua expansão decimal $x = 0, a_0a_1a_2 \dots$, a transformação f corresponde simplesmente a deslocar os dígitos de x uma unidade para a esquerda. Isso nos motiva a considerar:

$$\phi : \{0, 1, \dots, 9\}^{\mathbb{N}} \rightarrow [0, 1], \quad \phi((a_n)_n) = \sum_{n=0}^{\infty} \frac{a_n}{10^{n+1}} = 0, a_0a_1a_2 \dots$$

É claro que ϕ é sobrejetiva. Por outro lado, ela não é injetiva, uma vez que certos números reais possuem mais que uma expansão decimal: por exemplo, $0, 1000000 \dots = 0, 099999 \dots$. De fato, isso acontece somente se o número admite uma expansão decimal finita, ou seja, tal que todos os dígitos a partir de certa ordem são nulos. Esses números formam um conjunto enumerável e, portanto, são irrelevantes do ponto de vista da medida de Lebesgue. Mais precisamente, consideremos o conjunto $X \subset \{0, 1, \dots, 9\}^{\mathbb{N}}$ das sequências com um número infinito de símbolos diferentes de zero e o conjunto $Y \subset [0, 1]$ dos números cuja expansão decimal é infinita (logo, única). Então a restrição de ϕ a X é uma bijeção sobre Y .

É fácil verificar que tanto ϕ quanto a sua inversa são mensuráveis: use o fato de que a imagem da interseção de X com cada cilindro $[0; a_0, \dots, a_{m-1}]$ é a interseção de Y com um intervalo de comprimento 10^{-m} . Esta observação também mostra que $\phi_*m = \nu$, onde ν representa a medida de Bernoulli em $\{0, 1, \dots, 9\}^{\mathbb{N}}$ que dá igual peso a todos os dígitos. Além disso, se denotarmos por σ o deslocamento em $\{0, 1, \dots, 9\}^{\mathbb{N}}$, temos que

$$\phi \circ \sigma((a_n)_n) = 0, a_1 a_2 \dots a_n \dots = f \circ \phi((a_n)_n)$$

para todo $(a_n)_n \in X$. Isto prova que (f, m) é ergodicamente equivalente ao deslocamento de Bernoulli (σ, ν) .

Suponha que (f, μ) e (g, ν) são ergodicamente equivalentes. Um conjunto mensurável $A \subset M$ é invariante para $f : M \rightarrow M$ se, e somente se, $\phi(A)$ é invariante para $g : N \rightarrow N$. Além disso, $\nu(\phi(A)) = \mu(A)$. Portanto, (f, μ) é ergódico se, e somente se, (g, ν) é ergódico. As conclusões correspondentes para as propriedades de mistura e mistura fraca são igualmente fáceis de obter. De fato, praticamente todas as propriedades que estudamos neste curso são *invariantes de equivalência ergódica*, isto é, se elas valem para um sistema então também valem para qualquer outro que seja ergodicamente equivalente a esse. Uma exceção é a unicidade ergódica, que tem uma natureza diferente, já que é uma propriedade apenas da transformação.

Isto também quer dizer que essas propriedades podem ser utilizadas para tentar distinguir sistemas que não sejam ergodicamente equivalentes. Ainda assim, essa é usualmente uma tarefa difícil. Por exemplo, nada do que vimos até agora permitir responder à seguinte questão: os deslocamentos

$$\sigma : \{1, 2\}^{\mathbb{Z}} \rightarrow \{1, 2\}^{\mathbb{Z}} \quad \text{e} \quad \zeta : \{1, 2, 3\}^{\mathbb{Z}} \rightarrow \{1, 2, 3\}^{\mathbb{Z}}, \quad (8.1.1)$$

munidos das medidas de Bernoulli que dão pesos iguais a todos os símbolos, são ergodicamente equivalentes? É fácil ver que σ e ζ não são *topologicamente* conjugadas (por exemplo: ζ tem três pontos fixos, enquanto que σ tem apenas dois) mas a existência de uma equivalência ergódica é uma questão muito mais delicada. De fato, este tipo de pergunta motiva a maior parte do conteúdo do presente capítulo e acabará conduzindo à noção de entropia, que estudaremos no Capítulo 9.

Exemplo 8.1.2. Seja $\sigma : M \rightarrow M$ o deslocamento em $M = X^{\mathbb{N}}$ e seja $\mu = \nu^{\mathbb{N}}$ uma medida de Bernoulli. Seja $\hat{\sigma} : \hat{M} \rightarrow \hat{M}$ a extensão natural de σ e seja $\hat{\mu}$ o levantamento de $\hat{\mu}$ (Seção 2.4.2). Além disso, seja $\tilde{\sigma} : \tilde{M} \rightarrow \tilde{M}$ o deslocamento em $\tilde{M} = X^{\mathbb{Z}}$ e seja $\tilde{\mu} = \nu^{\mathbb{Z}}$ a medida de Bernoulli correspondente. Então, $(\hat{\sigma}, \hat{\mu})$ é ergodicamente equivalente a $(\tilde{\sigma}, \tilde{\mu})$. A equivalência pode ser construída da seguinte forma.

Por definição, \hat{M} é o espaço das pré-órbitas de σ , ou seja, das sequências $\hat{x} = (\dots, x_{-n}, \dots, x_0)$ em M tais que $\sigma(x_{-j}) = x_{-j+1}$ para todo $j \geq 1$. Além disso, cada x_{-j} é uma sequência $(x_{-j,i})_{i \in \mathbb{N}}$ em X . Então, a relação anterior quer dizer que

$$x_{-j,i+1} = x_{-j+1,i} \quad \text{para todo } i, j \in \mathbb{N}. \quad (8.1.2)$$

Considere a aplicação $\phi : \hat{M} \rightarrow \tilde{M}$, $\hat{x} \mapsto \tilde{x}$ dada por

$$\tilde{x}_n = x_{0,n} = x_{-1,n+1} = \cdots \quad \text{e} \quad \tilde{x}_{-n} = x_{-n,0} = x_{-n-1,1} = \cdots$$

Deixamos ao cuidado do leitor verificar que ϕ é realmente uma equivalência ergódica entre a extensão natural $(\hat{\sigma}, \hat{\mu})$ e o deslocamento bilateral $(\tilde{\sigma}, \tilde{\mu})$.

8.1.1 Exercícios

8.1.1. Seja $f : [0, 1] \rightarrow [0, 1]$ a transformação definida por $f(x) = 2x - [2x]$ e seja m a medida de Lebesgue em $[0, 1]$. Exiba uma transformação $g : [0, 1] \rightarrow [0, 1]$ e uma probabilidade ν invariante por g , tais que (g, ν) é ergodicamente equivalente a (f, μ) e o suporte de ν tem interior vazio.

8.1.2. Considere os deslocamentos unilaterais $f : \{1, \dots, k\}^{\mathbb{N}} \rightarrow \{1, \dots, k\}^{\mathbb{N}}$ e $g : \{1, \dots, l\}^{\mathbb{N}} \rightarrow \{1, \dots, l\}^{\mathbb{N}}$ munidos de medidas de Bernoulli μ e ν , respectivamente. Mostre que, para todo conjunto $X \subset \{1, \dots, k\}^{\mathbb{N}}$ com $f^{-1}(X) = X$ e $\mu(X) = 1$, existe $x \in X$ tal que $\#(X \cap f^{-1}(x)) = k$. Conclua que se $k \neq l$ então (f, μ) e (g, ν) não podem ser ergodicamente equivalentes.

8.1.3. Seja $X = \{1, \dots, d\}$ e considere o deslocamento $\sigma : X^{\mathbb{N}} \rightarrow X^{\mathbb{N}}$ munido de uma medida de Markov μ . Dado qualquer cilindro $C = [0; c_0, \dots, c_l]$ em $X^{\mathbb{N}}$, seja μ_C a restrição normalizada de μ a C . Mostre que existe transformação induzida $\sigma_C : C \rightarrow C$ (veja a Seção 1.4.2) que preserva μ_C e tal que (σ_C, μ_C) é ergodicamente equivalente a um deslocamento de Bernoulli $(\sigma_{\mathbb{N}}, \nu)$ em $\mathbb{N}^{\mathbb{N}}$.

8.2 Equivalência espectral

Sejam $f : M \rightarrow M$ e $g : N \rightarrow N$ transformações preservando probabilidades μ e ν , respectivamente. Sejam $U_f : L^2(\mu) \rightarrow L^2(\mu)$ e $U_g : L^2(\nu) \rightarrow L^2(\nu)$ os respectivos operadores de Koopman. Dizemos que (f, μ) e (g, ν) são *espectralmente equivalentes* se existe algum operador unitário $L : L^2(\mu) \rightarrow L^2(\nu)$ tal que

$$U_g \circ L = L \circ U_f. \tag{8.2.1}$$

Deixamos para o leitor verificar que a relação definida desta forma é, de fato, uma relação de equivalência.

É fácil ver que se dois sistemas são ergodicamente equivalentes então eles são espectralmente equivalentes. De fato, suponha que existe uma aplicação invertível $h : M \rightarrow N$ tal que $\phi_*\mu = \nu$ e $\phi \circ f = g \circ \phi$. Então, o operador de Koopman

$$U_\phi : L^2(\nu) \rightarrow L^2(\mu), \quad U_\phi(\psi) = \psi \circ \phi$$

é uma isometria e é invertível: a inversa é o operador de Koopman associado a ϕ^{-1} . Em outras palavras, U_ϕ é um operador unitário. Além disso,

$$U_f \circ U_\phi = U_{\phi \circ f} = U_{g \circ \phi} = U_\phi \circ U_g.$$

Portanto, $L = U_\phi$ é uma equivalência espectral entre os dois sistemas.

A recíproca é falsa, como veremos. Por exemplo, todos os deslocamentos de Bernoulli invertíveis enumeravelmente gerados são espectralmente equivalentes (Corolário 8.4.12); no entanto, nem todos têm a mesma entropia (Exemplo 9.1.10) e, portanto, nem todos são ergodicamente equivalentes.

8.2.1 Invariantes de equivalência espectral

Lembre que o espectro $\text{esp}(A)$ de um operador linear $A : E \rightarrow E$ num espaço de Banach complexo E é constituído pelos números complexos λ tais que $A - \lambda \text{id}$ não é invertível. Dizemos que $\lambda \in \text{esp}(A)$ é um autovalor se $A - \lambda \text{id}$ não é injetiva, isto é, se existe $v \neq 0$ tal que $Av = \lambda v$. Nesse caso a dimensão do núcleo de $A - \lambda \text{id}$ é chamada *multiplicidade* do autovalor.

Por definição, o *espectro* de um sistema (f, μ) é o espectro do operador de Koopman associado $U_f : L^2(\mu) \rightarrow L^2(\mu)$. Se (f, μ) é espectralmente equivalente a (g, ν) então os dois sistemas têm o mesmo espectro: a relação (8.2.1) implica que

$$(U_g - \lambda \text{id}) = L \circ (U_f - \lambda \text{id}) \circ L^{-1} \quad (8.2.2)$$

e, portanto, $U_g - \lambda \text{id}$ é invertível se, e somente se, $U_f - \lambda \text{id}$ é invertível. De fato, o espectro em si mesmo é um invariante muito grosseiro: em particular, todos os sistemas ergódicos invertíveis sem átomos têm o mesmo espectro (Exercício 8.2.1). No entanto, a respectiva medida espectral fornece invariantes muito úteis, como veremos a seguir. O mais simples deles é o conjunto dos átomos da medida espectral, ou seja, o conjunto dos autovalores do operador de Koopman. Note que (8.2.2) também mostra que λ é autovalor de U_f se, e somente se, ele é autovalor de U_g ; além disso, nesse caso as duas multiplicidades são iguais.

Observe que 1 é sempre um autovalor do operador de Koopman, uma vez que $U_f \varphi = \varphi$ para toda função constante φ . Pela Proposição 4.1.3(e), o sistema (f, μ) é ergódico se, e somente se, os únicos autovetores são as funções constantes, ou seja, se este autovalor tem multiplicidade 1. Então, por (8.2.2) o autovalor 1 tem multiplicidade 1 também para U_g , qualquer que seja o sistema (g, ν) espectralmente equivalente a (f, μ) . Isto mostra que a ergodicidade é um *invariante de equivalência espectral*.

Analogamente, suponha que o sistema (f, μ) é misturador. Então, pela Proposição 7.1.12,

$$\lim_n U_f^n \varphi \cdot \psi = \int \varphi d\mu \int \psi d\nu$$

para todo $\varphi, \psi \in L^2(\mu)$. Agora suponha que (g, ν) é espectralmente equivalente a (f, μ) . O operador L em (8.2.1) envia autovetores do autovalor 1 de U_f em autovetores do autovalor 1 de U_g . Como os dois sistemas são ergódicos, isto quer dizer que L envia funções constantes em funções constantes. Como L é um operador unitário,

$$U_g^n \varphi \cdot \psi = L(U_f^n \varphi) \cdot L\psi = U_f^n(L\varphi) \cdot L\psi$$

e, portanto, $\lim_n U_g^n \varphi \cdot \psi = \int L\varphi d\mu \int L\psi d\mu$. Além disso,

$$\int L\varphi d\mu = L\varphi \cdot 1 = L\varphi \cdot L1 = \varphi \cdot 1 = \int \varphi d\nu$$

e, analogamente, $\int L\psi d\mu = \int \psi d\mu$. Deste modo, provamos que

$$\lim_n U_g^n \varphi \cdot \psi = \int \varphi d\mu \int \psi d\mu,$$

ou seja, (g, ν) também é misturador. Isto mostra que a propriedade de mistura também é um invariante de equivalência espectral.

O mesmo argumento pode ser usado para a propriedade de mistura fraca. Mas o teorema que provaremos na Seção 8.2.2 dá uma demonstração mais interessante do fato de que esta propriedade é um invariante de equivalência espectral.

8.2.2 Autovetores e mistura fraca

Como vimos anteriormente, o operador de Koopman $U_f : L^2(\mu) \rightarrow L^2(\mu)$ de um sistema (f, μ) é uma isometria, ou seja, ele satisfaz $U_f^* U_f = \text{id}$. Se f é invertível então o operador de Koopman é unitário, ou seja, temos $U_f^* U_f = U_f U_f^* = \text{id}$. Em particular, nesse caso U_f é normal. Então, a propriedade de mistura fraca admite a seguinte caracterização interessante:

Teorema 8.2.1. *Um sistema invertível (f, μ) é fracamente misturador se, e somente se, as funções constantes são os únicos autovetores do operador de Koopman.*

Em particular, o sistema (f, μ) é fracamente misturador se, e somente se, ele é ergódico e 1 é o único autovalor de U_f .

Demonstração. Suponha que (f, μ) é fracamente misturador. Seja $\varphi \in L^2(\mu)$ qualquer autofunção (não nula) de U_f e seja λ o respectivo autovalor. Então,

$$\int \varphi d\mu = \int U_f \varphi d\mu = \lambda \int \varphi d\mu$$

e isto implica que $\int \varphi d\mu = 0$ ou $\lambda = 1$. No primeiro caso,

$$C_j(\varphi, \bar{\varphi}) = \left| \int (U_f^j \varphi) \bar{\varphi} d\mu \right| = |\lambda^j \int \varphi \bar{\varphi} d\mu| = \int |\varphi|^2 d\mu$$

para todo $j \geq 1$ (lembrando que $|\lambda| = 1$). Mas então,

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} C_j(\varphi, \bar{\varphi}) = \int |\varphi|^2 d\mu > 0,$$

contradizendo a hipótese de que o sistema é fracamente misturador. No segundo caso, usando que o sistema é ergódico, obtemos que φ é constante em μ -quase

todo ponto. Isto mostra que se o sistema é fracamente misturador então os únicos autovetores são as funções constantes.

Agora suponha que os únicos autovetores de U_f são as funções constantes. Para concluir que (f, μ) é fracamente misturador, devemos mostrar que

$$\frac{1}{n} \sum_{j=0}^{n-1} C_j(\varphi, \psi)^2 \rightarrow 0 \quad \text{para quaisquer } \varphi, \psi \in L^2(\mu)$$

(lembre do Exercício 7.1.2). É imediato da definição que

$$C_j(\varphi, \psi) = C_j(\varphi', \psi) \quad \text{onde } \varphi' = \varphi - \int \varphi d\mu$$

e a integral de φ' é nula. Logo, não constitui restrição supor que $\int \varphi d\mu = 0$. Então, usando a relação (A.7.6) para o operador unitário $L = U_f$, obtemos:

$$C_j(\varphi, \psi)^2 = \left| \int (U_f^j \varphi) \psi d\mu \right|^2 = \left| \int_{\mathbb{C}} z^j d\theta(z) \right|^2,$$

onde $\theta = E\varphi \cdot \psi$. A expressão à direita pode ser reescrita na seguinte forma:

$$\int_{\mathbb{C}} z^j d\theta(z) \int_{\mathbb{C}} \bar{z}^j d\bar{\theta}(z) = \int_{\mathbb{C}} \int_{\mathbb{C}} z^j \bar{w}^j d\theta(z) d\bar{\theta}(w).$$

Portanto, dado qualquer $n \geq 1$,

$$\frac{1}{n} \sum_{j=0}^{n-1} C_j(\varphi, \psi)^2 = \int_{\mathbb{C}} \int_{\mathbb{C}} \frac{1}{n} \sum_{j=0}^{n-1} (z\bar{w})^j d\theta(z) d\bar{\theta}(w). \quad (8.2.3)$$

Afirmamos que a medida $\theta = E\varphi \cdot \psi$ é não atômica. De fato, suponha que existe $\lambda \in \mathbb{C}$ tal que $\theta(\{\lambda\}) \neq 0$. Então, $E(\{\lambda\}) \neq 0$ e então podemos usar a Proposição A.7.8 para concluir que a função $E(\{\lambda\})\varphi$ é um autovetor de U_f . Pela hipótese sobre o operador U_f , isto implica que $E(\{\lambda\})\varphi$ é constante em μ -quase todo ponto. Logo,

$$E(\{\lambda\})\varphi \cdot \varphi = E(\{\lambda\})\varphi \int \bar{\varphi} d\mu = 0.$$

O Lema A.7.3 também dá que

$$E(\{\lambda\})\varphi \cdot \varphi = E(\{\lambda\})^2 \varphi \cdot \varphi = E(\{\lambda\})\varphi \cdot E(\{\lambda\})\varphi.$$

Juntando estas duas igualdades, concluímos que $E(\{\lambda\})\varphi = 0$, o qual contradiz a hipótese. Desta forma, fica provada a nossa afirmação.

A sequência $\frac{1}{n} \sum_{j=0}^{n-1} (z\bar{w})^j$ em (8.2.3) é limitada e (veja o Exercício 8.2.6) converge para zero no complementar da diagonal $\Delta = \{(z, w) : z = w\}$. Além disso, a diagonal tem medida nula:

$$(\theta \times \bar{\theta})(\Delta) = \int \theta(\{y\}) d\bar{\theta}(y) = 0.$$

porque θ é não atômica. Então, podemos usar o teorema da convergência dominada para concluir que (8.2.3) converge para zero quando $n \rightarrow \infty$. Isto prova que (f, μ) é fracamente misturador se U_f não tem autovetores não constantes. \square

Suponha que M é um espaço topológico. Dizemos que uma aplicação contínua $f : M \rightarrow M$ é *topologicamente fracamente misturadora* se o operador de Koopman U_f não possui autofunções contínuas não constantes. O seguinte fato é uma consequência direta do Teorema 8.2.1:

Corolário 8.2.2. *Se (f, μ) é fracamente misturadora então a restrição de f ao suporte de μ é topologicamente fracamente misturadora.*

Demonstração. Seja φ uma autofunção contínua de U_f . Pelo Teorema 8.2.1, a função φ é constante em μ -quase todo ponto. Logo, por continuidade, φ é constante (em todo ponto) no suporte de μ . \square

Vimos na Seção 7.3 que quase todo intercâmbio de intervalos é fracamente misturador mas não é misturador. A seguir descreveremos uma construção explícita, baseada numa extensão das ideias que vislumbramos no Exemplo 6.3.9. O leitor interessado pode encontrar esta e outras variações dessas ideias na Seção 7.4 do livro de Kalikow e McCutcheon [KM10]

Exemplo 8.2.3 (R. V. Chacon). Considere a sequência $(\mathcal{S}_n)_n$ de pilhas definida da seguinte forma. Primeiramente, $\mathcal{S}_1 = \{[0, 2/3]\}$. Em seguida, para cada $n \geq 1$, seja \mathcal{S}_n a pilha obtida dividindo \mathcal{S}_{n-1} em 3 colunas, com a mesma largura, e empilhando essas colunas umas sobre as outras, apenas com um intervalo adicional inserido entre a segunda e a terceira colunas, tal como está representado na Figura 8.1.

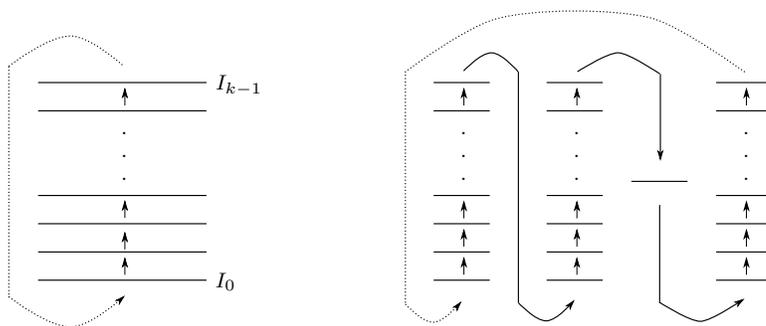


Figura 8.1: Distinguindo mistura fraca e mistura

Por exemplo, $\mathcal{S}_2 = \{[0, 2/9), [2/9, 4/9), [6/9, 8/9), [4/9, 6/9)\}$ e

$$\begin{aligned} \mathcal{S}_3 = \{ & [0, 2/27), [6/27, 8/27), [18/27, 20/27), [12/27, 14/27), [2/27, 4/27), \\ & [8/27, 10/27), [20/27, 22/27), [14/27, 16/27), [24/27, 26/27), \\ & [4/27, 6/27), [10/27, 12/27), [22/27, 24/27), [16/27, 18/27)\}. \end{aligned}$$

Note que cada \mathcal{S}_n é uma pilha no intervalo $J_n = [0, 1 - 3^{-n}]$. A sequência $(f_n)_n$ de transformações associadas a tais pilhas converge em todo ponto para uma transformação $f : [0, 1) \rightarrow [0, 1)$, a qual preserva a medida de Lebesgue m . Este sistema (f, m) é fracamente misturador mas não é misturador (Exercício 8.2.7).

8.2.3 Exercícios

8.2.1. Seja (f, μ) um sistema ergódico invertível sem átomos. Mostre que todo λ no círculo unitário $\{z \in \mathbb{C} : |z| = 1\}$ é *autovalor aproximado* do operador de Koopman $U_f : L^2(\mu) \rightarrow L^2(\mu)$: existe alguma sequência $(\varphi_n)_n$ tal que $\|\varphi_n\| \rightarrow 1$ e $\|U_f \varphi_n - \lambda \varphi_n\| \rightarrow 0$. Em particular, o espectro de U_f coincide com o círculo unitário.

8.2.2. Seja m a medida de Lebesgue no círculo e seja $U_\alpha : L^2(m) \rightarrow L^2(m)$ o operador de Koopman da rotação irracional $R_\alpha : S^1 \rightarrow S^1$. Calcule os autovalores de U_α e deduza que (R_α, m) e (R_β, m) são espectralmente equivalentes se, e somente se, $\alpha = \pm\beta$. [Observação: O Corolário 8.3.6 fornece um enunciado mais completo.]

8.2.3. Seja m a medida de Lebesgue no círculo e, para cada inteiro $k \geq 2$, seja $U_k : L^2(m) \rightarrow L^2(m)$ o operador de Koopman da transformação $f_k : S^1 \rightarrow S^1$ dada por $f_k(x) = kx \pmod{\mathbb{Z}}$. Verifique que se $p \neq q$ então (f_p, m) e (f_q, m) não são ergodicamente equivalentes. Mostre que, para qualquer $k \geq 2$,

$$L^2(m) = \{\text{constantes}\} \oplus \bigoplus_{j=0}^{\infty} U_k^j(H_k)$$

onde $H_k = \{\sum_{n \in \mathbb{Z}} a_n e^{2\pi i n x} : a_n = 0 \text{ se } k \nmid n\}$ e os termos da soma direta são ortogonais dois-a-dois. Conclua que (f_p, m) e (f_q, m) são espectralmente equivalentes para quaisquer p e q .

8.2.4. Seja $f : S^1 \rightarrow S^1$ a transformação dada por $f(x) = kx \pmod{\mathbb{Z}}$ e seja μ a medida de Lebesgue. Mostre que (f, μ) é fracamente misturador se, e somente se, $|k| \geq 2$.

8.2.5. Prove que dada f invertível, se μ é ergódico para todo iterado f^n e existe $C > 0$ tal que

$$\limsup_n \mu(f^{-n}(A) \cap B) \leq C \mu(A) \mu(B),$$

para quaisquer conjuntos mensuráveis A e B , então μ é fracamente misturadora. [Observação: Este enunciado é devido a Ornstein [Orn72]. De fato, ele mostrou que nas hipóteses do exercício, o sistema é (fortemente) misturador.]

8.2.6. Sejam z e w dois números complexos com módulo 1. Verifique que

$$1. \lim_n \frac{1}{n} \sum_{j=0}^{n-1} |z^j - 1| = 0 \text{ se, e somente se, } z = 1;$$

$$2. \lim_n \frac{1}{n} \sum_{j=0}^{n-1} (z\bar{w})^j = 0 \text{ se } z \neq w.$$

8.2.7. Considere o sistema (f, m) no Exemplo 8.2.3. Mostre que

- (a) o sistema (f, m) é ergódico;
- (b) os únicos autovetores do operador de Koopman $U_f : L^1(m) \rightarrow L^1(m)$ são as funções constantes; logo (f, m) é fracamente misturador;
- (b) $\limsup_n m(f^n(A) \cap A) \geq 2/27$ se tomarmos $A = [0, 2/9]$; portanto (f, m) não é misturador.

8.3 Espectro discreto

Nesta seção e na próxima estudaremos os dois casos extremos, no que se refere ao tipo da medida espectral do operador de Koopman: sistemas com *espectro discreto*, cuja medida espectral é puramente atômica, e sistemas com *espectro de Lebesgue*, cuja medida espectral é equivalente à medida de Lebesgue no círculo unitário.

Começamos por descrever algumas propriedades dos autovalores e autovetores do operador de Koopman. É claro que todo autovalor pertence ao círculo unitário, uma vez que U_f é uma isometria.

Proposição 8.3.1. *Se $\varphi_1, \varphi_2 \in L^2(\mu)$ satisfazem $U_f \varphi_1 = \lambda_1 \varphi_1$ e $U_f \varphi_2 = \lambda_2 \varphi_2$ com $\lambda_1 \neq \lambda_2$, então $\varphi_1 \cdot \varphi_2 = 0$. Além disso, os autovalores de U_f constituem um subgrupo do círculo unitário.*

Se o sistema (f, μ) é ergódico então todo o autovalor de U_f é simples e o módulo de toda autofunção é constante em μ -quase todo ponto.

Demonstração. A primeira afirmação segue da igualdade

$$\varphi_1 \cdot \varphi_2 = U_f \varphi_1 \cdot U_f \varphi_2 = \lambda_1 \varphi_1 \cdot \lambda_2 \varphi_2 = \lambda_1 \bar{\lambda}_2 (\varphi_1 \cdot \varphi_2) = \lambda_1 \lambda_2^{-1} (\varphi_1 \cdot \varphi_2),$$

uma vez que $\lambda_1 \lambda_2^{-1} \neq 1$. Essa igualdade também mostra que o conjunto dos autovalores é fechado para a operação $(\lambda_1, \lambda_2) \mapsto \lambda_1 \lambda_2^{-1}$. Lembrando que 1 sempre é um autovalor, segue que esse conjunto é um grupo.

Agora suponha que (f, μ) é ergódico. Suponha que $U_f \varphi = \lambda \varphi$. Então, $U_f(|\varphi|) = |U_f \varphi| = |\lambda \varphi| = |\varphi|$ em μ -quase todo ponto. Por ergodicidade, isto implica que $|\varphi|$ é constante em μ -quase todo ponto. Em seguida, suponha que $U_f \varphi_1 = \lambda \varphi_1$ e $U_f \varphi_2 = \lambda \varphi_2$ e as funções φ_1 e φ_2 não são identicamente nulas. Como $|\varphi_2|$ é constante em μ -quase todo ponto, temos $\varphi_2(x) \neq 0$ para μ -quase todo x . Então φ_1/φ_2 está bem definida. Além disso,

$$U_f\left(\frac{\varphi_1}{\varphi_2}\right) = \frac{U_f(\varphi_1)}{U_f(\varphi_2)} = \frac{\lambda \varphi_1}{\lambda \varphi_2} = \frac{\varphi_1}{\varphi_2}.$$

Por ergodicidade, segue que o quociente é constante em μ -quase todo ponto. Ou seja, $\varphi_1 = c\varphi_2$ para algum $c \in \mathbb{C}$. □

Dizemos que um sistema (f, μ) tem *espectro discreto* se os autovetores do operador de Koopman $U_f : L^2(\mu) \rightarrow L^2(\mu)$ geram o espaço $L^2(\mu)$. Observe que isto implica que U_f é invertível e, portanto, é um operador unitário. A terminologia é justificada pela observação a seguir (lembre do Teorema A.7.9):

Proposição 8.3.2. *Um sistema (f, μ) tem espectro discreto se, e somente se, o seu operador de Koopman U_f admite representação espectral*

$$T : \bigoplus_j L^2(\sigma_j)^{\chi_j} \rightarrow \bigoplus_j L^2(\sigma_j)^{\chi_j}, \quad (\varphi_{j,l})_{j,l} \mapsto (z \mapsto z\varphi_{j,l}(z))_{j,l} \quad (8.3.1)$$

onde cada σ_j é uma medida de Dirac num ponto do círculo unitário.

Demonstração. Suponhamos que U_f admite uma representação espectral (8.3.1) onde cada $\sigma_j = \delta_{\lambda_j}$ para algum λ_j no círculo unitário. Cada $L^2(\sigma_j)^{\chi_j}$ pode ser identificado, de maneira canônica, com um subespaço da soma direta. A restrição de T a esse subespaço coincide com $\lambda_j \text{id}$, uma vez que

$$z\varphi_{j,l}(z) = \lambda_j\varphi_{j,l}(z) \quad \text{em } \sigma_j\text{-quase todo ponto.} \quad (8.3.2)$$

Seja $(v_{j,l})_l$ uma base de Hilbert de $L^2(\sigma_j)^{\chi_j}$. Então, a família $(v_{j,l})_{j,l}$ é uma base de Hilbert da soma direta, constituída por autovetores de T . Como T é unitariamente conjugado a U_f , segue que $L^2(\mu)$ admite uma base de Hilbert formada por autovetores do operador de Koopman.

Agora suponha que (f, μ) tem espectro discreto. Sejam $(\lambda_j)_j$ os autovalores de U_f e, para cada j , seja $\sigma_j = \delta_{\lambda_j}$ e seja χ_j a dimensão de Hilbert do autoespaço $\ker(U_f - \lambda_j \text{id})$. Note que o espaço $L^2(\sigma_j)$ é uni-dimensional, pois toda função é constante em σ_j -quase todo ponto. Portanto, a dimensão de $L^2(\delta_{\lambda_j})^{\chi_j}$ também é igual a χ_j . Logo, existe isomorfismo unitário

$$L_j : \ker(U_f - \lambda_j \text{id}) \rightarrow L^2(\delta_{\lambda_j})^{\chi_j}.$$

É claro que $L_j \circ U_f \circ L_j^{-1} = \lambda_j \text{id}$. Ou seja, recordando a observação (8.3.2),

$$L_j \circ U_f \circ L_j^{-1} : (\varphi_{j,l})_l \mapsto (z \mapsto \lambda_j\varphi_{j,l}(z))_l = (z \mapsto z\varphi_{j,l}(z))_l. \quad (8.3.3)$$

Os autoespaços $\ker(U_f - \lambda_j \text{id})$ geram $L^2(\mu)$, por hipótese, e são ortogonais dois-a-dois, pela Proposição 8.3.1. Logo, podemos combinar os operadores L_j para obter um isomorfismo unitário $L : L^2(\mu) \rightarrow \bigoplus_j L^2(\sigma_j)^{\chi_j}$. A relação (8.3.3) dá que

$$L \circ U_f \circ L^{-1} : (\varphi_{j,l})_{j,l} \mapsto (z \mapsto z\varphi_{j,l}(z))_l.$$

é representação espectral de U_f da forma pretendida. \square

Exemplo 8.3.3. Seja m a medida de Lebesgue no toro \mathbb{T}^d . Como sabemos, o espaço $L^2(m)$ é gerado pela família de Fourier $\{\phi_k(x) = e^{2\pi i k \cdot x} : k \in \mathbb{Z}^d\}$. Seja $f = R_\theta : \mathbb{T}^d \rightarrow \mathbb{T}^d$ a rotação correspondente a um dado vetor $\theta = (\theta_1, \dots, \theta_d)$. Então,

$$U_f \phi_k(x) = \phi_k(x + \theta) = e^{2\pi i k \cdot \theta} \phi_k(x) \quad \text{para todo } x \in \theta.$$

Isto mostra que todo ϕ_k é autovetor de U_f e, portanto, (f, m) tem espectro discreto. Note que o grupo dos autovalores é

$$G_\theta = \{e^{2\pi i k \cdot \theta} : k \in \mathbb{Z}^d\}, \quad (8.3.4)$$

ou seja, o subgrupo do círculo unitário gerado por $\{e^{2\pi i \theta_j} : j = 1, \dots, d\}$.

Mais geralmente, toda translação ergódica num grupo abeliano compacto tem espectro discreto. Reciprocamente, todo sistema ergódico com espectro discreto é ergodicamente isomorfo a uma translação num grupo abeliano compacto. Outro resultado interessante é que todo subgrupo do círculo unitário é o grupo de autovalores de algum sistema ergódico com espectro discreto. Estes fatos são provados na Seção 3.3 do livro de Peter Walters [Wal82].

Proposição 8.3.4. *Suponha que (f, μ) e (g, ν) são ergódicos e têm espectro discreto. Então (f, μ) e (g, ν) são espectralmente equivalentes se, e somente se, os seus operadores de Koopman $U_f : L^2(\mu) \rightarrow L^2(\mu)$ e $U_g : L^2(\nu) \rightarrow L^2(\nu)$ têm os mesmos autovalores.*

Demonstração. É claro que se os operadores de Koopman são conjugados então eles têm os mesmos autovalores. Para provar a recíproca, sejam $(\lambda_j)_j$ os autovalores dos dois operadores. Pela Proposição 8.3.2, os autovalores são simples. Para cada j , sejam u_j e v_j vetores unitários nos núcleos $\ker(U_f - \lambda_j \text{id})$ e $\ker(U_g - \lambda_j \text{id})$, respectivamente. Então $(u_j)_j$ e $(v_j)_j$ são bases de Hilbert de $L^2(\mu)$ e de $L^2(\nu)$, respectivamente. Considere o isomorfismo $L : L^2(\mu) \rightarrow L^2(\nu)$ definido por $L(u_j) = v_j$. Este operador é unitário, uma vez que envia uma base de Hilbert numa base de Hilbert, e satisfaz

$$L \circ U_f(u_j) = L(\lambda_j u_j) = \lambda_j v_j = U_g(v_j) = U_g \circ L(u_j)$$

para todo j . Por linearidade, segue que $L \circ U_f = U_g \circ L$. Portanto, (f, μ) e (g, ν) são espectralmente equivalentes. \square

Corolário 8.3.5. *Se (f, μ) é ergódico, invertível e tem espectro discreto então (f, μ) é espectralmente equivalente a (f^{-1}, μ) .*

Demonstração. É claro que λ é autovalor de U_f se, e somente se, λ^{-1} é autovalor de $U_{f^{-1}}$; além disso, os autovetores são os mesmos. Como os conjuntos de autovalores são grupos, segue que os dois operadores têm os mesmos autovalores e os mesmos autovetores. Aplique a Proposição 8.3.4. \square

Seja m a medida de Lebesgue no toro \mathbb{T}^d . A Proposição 8.3.4 também nos permite classificar as rotações irracionais do toro a menos de equivalência ergódica ou espectral:

Corolário 8.3.6. *Sejam $\theta = (\theta_1, \dots, \theta_d)$ and $\tau = (\tau_1, \dots, \tau_d)$ vetores racionalmente independentes e sejam R_θ e R_τ as respectivas as rotações no toro \mathbb{T}^d . As seguintes condições são equivalentes:*

(a) (R_θ, m) e (R_τ, m) são ergodicamente equivalentes;

(b) (R_θ, m) e (R_τ, m) são espectralmente equivalentes;

(c) existe $L \in \text{SL}(d, \mathbb{Z})$ tal que $\theta = L\tau$ módulo \mathbb{Z}^d .

A demonstração fica a cargo do leitor (Exercício 8.3.2). No caso particular do círculo, obtemos que duas rotações irracionais R_θ e R_τ são equivalentes se, e somente se, $R_\theta = R_\tau$ ou então $R_\theta = R_\tau^{-1}$. Veja também o Exercício 8.3.3.

8.3.1 Exercícios

8.3.1. Suponha que (f, μ) tem espectro discreto e que o espaço de Hilbert $L^2(\mu)$ é separável (isto vale, por exemplo, se a σ -álgebra dos conjuntos mensuráveis admite um gerador enumerável). Mostre que existe uma sequência $(n_k)_k$ convergindo para infinito tal que $\|U_f^{n_k} \varphi - \varphi\|_2$ converge para zero quando $k \rightarrow \infty$, para toda $\varphi \in L^2(\mu)$.

8.3.2. Prove o Corolário 8.3.6.

8.3.3. Seja m a medida de Lebesgue em S^1 e sejam $\theta = p/q$ e $\tau = r/s$ dois números racionais, onde $\text{mdc}(p, q) = 1 = \text{mdc}(r, s)$. Mostre que as rotações (R_θ, m) e (R_τ, m) são ergodicamente equivalentes se, e somente se, os denominadores q e s são iguais.

8.4 Espectro de Lebesgue

Definição 8.4.1. Seja $U : H \rightarrow H$ uma isometria num espaço de Hilbert. Dizemos que U tem *espectro de Lebesgue* se existe algum subespaço fechado $E \subset H$ tal que

(a) $U(E) \subset E$;

(b) $\bigcap_{n \in \mathbb{N}} U^n(E) = \{0\}$;

(c) $\sum_{n \in \mathbb{N}} U^{-n}(E) = H$.

A razão de ser desta denominação será explicada na Proposição 8.4.10.

Dado uma probabilidade μ , representemos por $L_0^2(\mu) = L_0^2(M, \mathcal{B}, \mu)$ o complemento ortogonal, no espaço $L^2(\mu) = L^2(M, \mathcal{B}, \mu)$, do subespaço das funções constantes. Em outras palavras:

$$L_0^2(\mu) = \left\{ \varphi \in L^2(\mu) : \int \varphi d\mu = 0 \right\}.$$

Note que $L_0^2(\mu)$ é invariante pelo operador de Koopman: $\varphi \in L_0^2(\mu)$ se, e somente se, $U_f \varphi \in L_0^2(\mu)$. Dizemos que o sistema (f, μ) tem *espectro de Lebesgue* se a restrição do operador de Koopman a $L_0^2(\mu)$ tem espectro de Lebesgue.

8.4.1 Exemplos e propriedades

Começamos por observar que todo deslocamento de Bernoulli tem espectro de Lebesgue. Tratamos separadamente o caso unilateral e o caso bilateral.

Exemplo 8.4.2. Considere um deslocamento unilateral $\sigma : X^{\mathbb{N}} \rightarrow X^{\mathbb{N}}$, munido de uma medida de Bernoulli $\mu = \nu^{\mathbb{N}}$. Tome $E = L_0^2(\mu)$. As condições (a) e (c) na definição são evidentes. Para provar a condição (b), considere qualquer função $\varphi \in L_0^2(\mu)$ na interseção, ou seja, tal que para todo $n \in \mathbb{N}$ existe uma função $\psi_n \in L_0^2(\mu)$ satisfazendo $\varphi = \psi_n \circ \sigma^n$. Queremos mostrar que φ é constante em μ -quase todo ponto. Para cada $c \in \mathbb{R}$, considere

$$A_c = \{x \in M : \varphi(x) > c\}.$$

Para cada $n \in \mathbb{N}$, podemos escrever $A_c = \sigma^{-n}(\{x \in M : \psi_n(x) > c\})$. Então A_c pertence à σ -álgebra gerada pelos cilindros da forma $[n; C_n, \dots, C_m]$ com $m \geq n$. Consequentemente, $\mu(A_c \cap C) = \mu(A_c)\mu(C)$ para todo cilindro C da forma $C = [0; C_0, \dots, C_{n-1}]$. Como n é arbitrário e os cilindros constituem uma família geradora, segue que $\mu(A_c \cap B) = \mu(A_c)\mu(B)$ para todo conjunto mensurável $B \subset M$. Tomando $B = A_c$ concluímos que $\mu(A_c) = \mu(A_c)^2$, ou seja, $\mu(A_c) \in \{0, 1\}$ para todo $c \in \mathbb{R}$. Isto prova que φ é constante em μ -quase todo ponto, tal como afirmamos.

Exemplo 8.4.3. Agora considere um deslocamento bilateral $\sigma : X^{\mathbb{Z}} \rightarrow X^{\mathbb{Z}}$, munido de uma medida de Bernoulli $\mu = \nu^{\mathbb{Z}}$. Seja \mathcal{A} a σ -álgebra gerada pelos cilindros da forma $[0; C_0, \dots, C_m]$ com $m \geq 0$. Denotamos por $L_0^2(M, \mathcal{A}, \mu)$ o espaço das funções $\varphi \in L_0^2(\mu)$ que são mensuráveis relativamente à σ -álgebra \mathcal{A} (em outras palavras, $\varphi(x)$ depende apenas das coordenadas $x_n, n \geq 0$ do ponto). Tome $E = L_0^2(M, \mathcal{A}, \mu)$. A condição (a) da definição é imediata. A condição (b) segue dos mesmos argumentos que usamos no Exemplo 8.4.2. Para provar a condição (c), note que $\cup_n U_{\sigma^{-n}}(E)$ contém as funções características de todos os cilindros. Portanto, ela contém todas as combinações lineares de funções características dos conjuntos na álgebra gerada pelos cilindros. Isto implica que a união é densa em $L_0^2(\mu)$, conforme queríamos provar.

Lema 8.4.4. Se (f, μ) tem espectro de Lebesgue então $\lim_n U_f^n \varphi \cdot \psi = 0$ para todo $\varphi \in L_0^2(\mu)$ e todo $\psi \in L^2(\mu)$.

Demonstração. Observe que a sequência $U_f^n \varphi \cdot \psi$ é limitada. De fato, pela desigualdade de Cauchy-Schwartz (Teorema A.5.4):

$$|U_f^n \varphi \cdot \psi| \leq \|U_f^n \varphi\|_2 \|\psi\|_2 = \|\varphi\|_2 \|\psi\|_2 \quad \text{para todo } n.$$

Então, basta provar que toda subsequência convergente $U_f^{n_j} \varphi \cdot \psi$ converge para zero. Analogamente, o conjunto $\{U_f^n \varphi : n \in \mathbb{N}\}$ é limitado em $L^2(\mu)$, porque U_f é uma isometria. Pelo teorema de Banach-Alaoglu (Teoremas A.6.1 e 2.3.1), toda sequência nesse conjunto admite subsequência que converge fracamente. Logo, não é restrição supor que $U_f^{n_j} \varphi$ converge fracamente para algum $\hat{\varphi} \in L^2(\mu)$.

Seja E um subespaço satisfazendo as condições na Definição 8.4.1. Inicialmente, suponha que $\varphi \in U_f^{-k}(E)$ para algum k . Então $U_f^{n_j} \varphi \in U_f^{n_j-k}(E)$. Logo, dado qualquer $l \in \mathbb{N}$, temos que $U_f^{n_j} \varphi \in U_f^l(E)$ para todo j suficientemente grande. Segue (veja o Exercício A.6.8) que $\hat{\varphi} \in U_f^l(E)$, para todo $l \in \mathbb{N}$. Pela condição (b) na definição, isto implica que $\hat{\varphi} = 0$ em μ -quase todo ponto. Em particular, $\lim_j U_f^{n_j} \varphi \cdot \psi = \hat{\varphi} \cdot \psi = 0$.

Agora considere $\varphi \in L_0^2(\mu)$ qualquer. Pela condição (c) na definição, para todo $\varepsilon > 0$ existem $k \in \mathbb{N}$ e $\varphi_k \in U_f^{-k}(E)$ tais que $\|\varphi - \varphi_k\|_2 \leq \varepsilon$. Usando a desigualdade de Cauchy-Schwarz uma vez mais:

$$|U_f^n \varphi \cdot \psi - U_f^n \varphi_k \cdot \psi| \leq \|\varphi - \varphi_k\|_2 \|\psi\|_2 \leq \varepsilon \|\psi\|_2$$

para todo n . Lembrando que $\lim_n U_f^n \varphi_k \cdot \psi = 0$ (pelo parágrafo anterior), obtemos que

$$-\varepsilon \|\psi\|_2 \leq \liminf_n U_f^n \varphi \cdot \psi \leq \limsup_n U_f^n \varphi \cdot \psi \leq \varepsilon \|\psi\|_2.$$

Fazendo $\varepsilon \rightarrow 0$, segue que $\lim_n U_f^n \varphi \cdot \psi = 0$, como queríamos demonstrar. \square

Corolário 8.4.5. *Se (f, μ) tem espectro de Lebesgue então ele é misturador.*

Demonstração. Basta observar que

$$C_n(\varphi, \psi) = |U_f^n \varphi \cdot \psi - \left(\int \varphi d\mu \right) \cdot \psi| = |U_f^n (\varphi - \int \varphi d\mu) \cdot \psi|$$

e a função $\varphi' = \varphi - \int \varphi d\mu$ está em $L_0^2(\mu)$. \square

No Exemplo 8.4.13 descreveremos exemplos de sistemas misturadores que não têm espectro de Lebesgue. Portanto, a recíproca do Corolário 8.4.5 é falsa, em geral.

A classe dos sistemas com espectro de Lebesgue é invariante por equivalência espectral. De fato, suponha que (f, μ) tem espectro de Lebesgue e que (g, ν) é um sistema espectralmente equivalente. Seja $L : L^2(\mu) \rightarrow L^2(\nu)$ um operador unitário conjugando os operadores de Koopman U_f e U_g . Segue da hipótese e do Corolário 8.4.5 que (f, μ) é fracamente misturador. Logo, pelo Teorema 8.2.1, as constantes são os únicos autovetores de U_f . Então o mesmo vale para U_g e a conjugação L envia constantes em constantes. Então, como L é unitário, a sua restrição ao complemento ortogonal $L_0^2(\mu)$ é um operador unitário sobre $L_0^2(\nu)$. Agora, dado qualquer subespaço $E \subset L_0^2(\mu)$ satisfazendo as condições (a), (b), (c) da definição para U_f , é claro que o subespaço $L(E) \subset L_0^2(\nu)$ satisfaz as condições correspondentes para U_g . Logo (g, ν) tem espectro de Lebesgue.

Dados subespaços fechados $V \subset W$ de um espaço de Hilbert H , denotamos por $W \ominus V$ o complemento ortogonal de V em W , ou seja,

$$W \ominus V = W \cap V^\perp = \{w \in W : v \cdot w = 0 \text{ para todo } v \in V\}.$$

A demonstração do seguinte fato será discutida na próxima seção:

Proposição 8.4.6. *Se $U : H \rightarrow H$ é uma isometria e E_1 e E_2 são subespaços de H satisfazendo as condições da Definição 8.4.1, então $E_1 \ominus U(E_1)$ e $E_2 \ominus U(E_2)$ têm a mesma dimensão de Hilbert (que é chamada posto do operador U).*

Então, chamamos *posto* de um sistema (f, μ) com espectro de Lebesgue ao posto do respectivo operador de Koopman restrito a $L_0^2(\mu)$. É claro que o posto é menor ou igual que a dimensão do espaço $L^2(\mu)$. Em particular, se $L^2(\mu)$ é separável então o posto é finito ou enumerável. A maior parte dos exemplos interessantes cai nesta categoria:

Exemplo 8.4.7. Suponha que o espaço de probabilidade (M, \mathcal{B}, μ) é *enumeravelmente gerado*, ou seja, que existe uma família enumerável G de subconjuntos mensuráveis tais que todo elemento de \mathcal{B} coincide, a menos de medida nula, com um elemento da σ -álgebra gerada por G . Então $L^2(\mu)$ é separável: a álgebra \mathcal{A} gerada por G é enumerável e as combinações lineares com coeficientes racionais das funções características dos elementos de \mathcal{A} formam um subconjunto enumerável denso de $L^2(\mu)$.

É interessante observar que não são conhecidos sistemas com espectro de Lebesgue com posto finito. No caso dos deslocamentos de Bernoulli, o posto coincide com a dimensão do respectivo $L^2(\mu)$:

Exemplo 8.4.8. Seja (σ, μ) um deslocamento de Bernoulli unilateral (no caso bilateral valem considerações análogas). Como vimos no Exemplo 8.4.2, podemos tomar $E = L_0^2(\mu)$. Então, denotando $x = (x_1, \dots, x_n, \dots)$ e lembrando que $\mu = \nu^{\mathbb{N}}$,

$$\begin{aligned} \varphi \in E \ominus U_\sigma(E) &\Leftrightarrow \int \varphi(x_0, x) \psi(x) d\mu(x_0, x) = 0 & \forall \psi \in L_0^2(\mu) \\ &\Leftrightarrow \int \left(\int \varphi(x_0, x) d\nu(x_0) \right) \psi(x) d\mu(x) = 0 & \forall \psi \in L_0^2(\mu). \end{aligned}$$

Logo, $E \ominus U_\sigma(E) = \{ \varphi \in L^2(\mu) : \int \varphi(x_0, x) d\nu(x_0) = 0 \text{ para } \mu\text{-quase todo } x \}$. Afirmamos que $\dim(E \ominus U_\sigma(E)) = \dim L^2(\mu)$. A desigualdade \leq é imediata. Para provar a outra desigualdade, fixe qualquer função mensurável $\phi : X \rightarrow \mathbb{R}$ com $\int \phi d\nu = 0$ e $\int \phi^2 d\nu = 1$. Considere a aplicação linear $I : L^2(\mu) \rightarrow L^2(\mu)$ que a cada $\psi \in L^2(\mu)$ associa a função $I\psi(x_0, x) = \phi(x_0)\psi(x)$. As hipóteses sobre ϕ implicam que:

$$I\psi \in E \ominus U_\sigma(E) \quad \text{e} \quad \|I\psi\|_2 = \|\psi\|_2 \quad \text{para todo } \psi \in L^2(\mu).$$

Isto mostra que $E \ominus U_\sigma(E)$ contém um subespaço isométrico a $L^2(\mu)$ e, portanto, $\dim E \ominus U_\sigma(E) \geq \dim L^2(\mu)$. Isto conclui o argumento.

Dizemos que o deslocamento é *de tipo enumerável* se o espaço de probabilidade X é enumeravelmente gerado. Isto é automático, por exemplo, se X for finito ou enumerável. Nesse caso, o espaço $\Sigma = X^{\mathbb{N}}$ (ou $\Sigma = X^{\mathbb{Z}}$) também é enumeravelmente gerado: se G é um gerador enumerável de X então os cilindros

$[m; C_m, \dots, C_n]$ com $C_j \in G$ constituem um gerador enumerável de M . Então, como comentamos no Exemplo 8.4.7, o espaço $L^2(\mu)$ é separável. Portanto, segue do Exemplo 8.4.8 que *todo deslocamento de Bernoulli de tipo enumerável tem espectro de Lebesgue com posto enumerável*.

8.4.2 O caso invertível

Nesta seção suporemos que o sistema (f, μ) é invertível. Neste contexto, a noção de espectro de Lebesgue pode ser formulada de maneira mais transparente.

Proposição 8.4.9. *Seja $U : H \rightarrow H$ um operador unitário num espaço de Hilbert H . Então U tem espectro de Lebesgue se, e somente se, existe um subespaço fechado $F \subset H$ tal que os iterados $U^k(F)$, $k \in \mathbb{Z}$ são ortogonais dois-a-dois e satisfazem*

$$H = \bigoplus_{k \in \mathbb{Z}} U^k(F).$$

Demonstração. Suponha que existe algum subespaço F como no enunciado. Tome $E = \bigoplus_{k=0}^{\infty} U^k(F)$. A condição (a) na Definição 8.4.1 é imediata:

$$U(E) = \bigoplus_{k=1}^{\infty} U^k(F) \subset E.$$

Quanto à condição (b), note que $\varphi \in \bigcap_{n=0}^{\infty} U^n(E)$ significa que $\varphi \in \bigoplus_{k=n}^{\infty} U^k(F)$ para todo $n \geq 0$. Isto implica que φ é ortogonal a $U^k(F)$ para todo $k \in \mathbb{Z}$. Logo $\varphi = 0$. Finalmente, por hipótese, podemos escrever qualquer $\varphi \in H$ como soma ortogonal $\varphi = \sum_{k \in \mathbb{Z}} \varphi_k$ com $\varphi_k \in U^k(F)$ para todo k . Então

$$\sum_{k=-n}^{\infty} \varphi_k \in \bigoplus_{k=-n}^{\infty} U^k(F) = U^{-n}(E)$$

para todo n e a sequência do lado esquerdo converge para φ quando $n \rightarrow \infty$. Isto dá a condição (c) na definição.

Agora provaremos a recíproca. Dado E satisfazendo as condições (a), (b) e (c) na definição, tome $F = E \ominus U(E)$. É fácil ver que os iterados de F são ortogonais dois-a-dois. Afiramos que

$$\bigoplus_{k=0}^{\infty} U^k(F) = E. \quad (8.4.1)$$

De fato, considere qualquer $v \in E$. É imediato da definição de F que, existem sequências $v_n \in U^n(F)$ e $w_n \in U^n(E)$ tais que $v = v_0 + \dots + v_{n-1} + w_n$ para cada $n \geq 1$. Queremos mostrar que $(w_n)_n$ converge para zero, para concluirmos que $v = \sum_{j=0}^{\infty} v_j$. Para isso, observe que

$$\|v\|^2 = \sum_{j=0}^{n-1} \|v_j\|^2 + \|w_n\|^2 \quad \text{para todo } n$$

e, portanto, a série $\sum_{j=0}^{\infty} \|v_j\|^2$ é somável. Dado $\varepsilon > 0$, fixe $m \geq 1$ tal que a soma dos termos com $j \geq m$ é menor que ε . Para todo $n \geq m$,

$$\|w_m - w_n\|^2 = \|v_m + \cdots + v_{n-1}\|^2 = \|v_m\|^2 + \cdots + \|v_{n-1}\|^2 < \varepsilon.$$

Isto prova que $(w_n)_n$ é sequência de Cauchy em H . Seja w o seu limite. Como $w_n \in U^n(E) \subset U^m(E)$ para todo $m \leq n$, passando ao limite obtemos que $w \in U^m(E)$ para todo m . Pela condição (b) na hipótese, isto implica que $w = 0$. Portanto, a prova da afirmação (8.4.1) está completa. Para concluir a demonstração da proposição basta observar que

$$\bigoplus_{k \in \mathbb{Z}} U^k(F) = \sum_{n=0}^{\infty} \bigoplus_{k=-n}^{\infty} U^k(F) = \sum_{n=0}^{\infty} U^{-n}(E).$$

A condição (c) na hipótese implica que este subespaço coincide com H . \square

Em particular, um sistema invertível (f, μ) tem espectro de Lebesgue se, e somente se, existe um subespaço fechado $F \subset L_0^2(\mu)$ tal que

$$L_0^2(\mu) = \bigoplus_{k \in \mathbb{Z}} U_f^k(F). \quad (8.4.2)$$

O próximo resultado é a razão pela qual sistemas com espectro de Lebesgue são denominados desse modo e também conduz naturalmente à noção de posto:

Proposição 8.4.10. *Seja $U : H \rightarrow H$ um operador unitário num espaço de Hilbert. Seja λ a medida de Lebesgue no círculo unitário. Então U tem espectro de Lebesgue se, e somente se, ele admite a representação espectral*

$$T : L^2(\lambda)^\chi \rightarrow L^2(\lambda)^\chi \quad (\varphi_\alpha)_\alpha \mapsto (z \mapsto z\varphi_\alpha(z))_\alpha$$

para algum cardinal χ . Além disso, χ está unicamente determinado por U .

Demonstração. Começemos provar a afirmação ‘se’. Como sabemos, a família $\{z^n : n \in \mathbb{Z}\}$ é uma base de Hilbert (base de Fourier) do espaço $L^2(\lambda)$. Seja V_n o subespaço unidimensional gerado por $\varphi(z) = z^n$. Então, $L^2(\lambda) = \bigoplus_{n \in \mathbb{Z}} V_n$ e, portanto,

$$L^2(\lambda)^\chi = \left(\bigoplus_{n \in \mathbb{Z}} V_n \right)^\chi = \bigoplus_{n \in \mathbb{Z}} V_n^\chi \quad (8.4.3)$$

(W^χ representa a soma direta ortogonal de χ cópias do espaço W). Além disso, a restrição de T a cada V_n^χ é um operador unitário sobre V_{n+1}^χ . Tome $F' = V_0^\chi$. A relação (8.4.3) significa que os iterados $T^n(F') = V_n^\chi$ são ortogonais dois-a-dois e a sua soma direta ortogonal é o espaço $L^2(\lambda)^\chi$. Usando a conjugação de T ao operador de Koopman em $L_0^2(\mu)$, concluímos que existe um subespaço F nas condições da Proposição 8.4.9.

Reciprocamente, suponha que existe F nas condições da Proposição 8.4.9. Seja $\{v_q : q \in Q\}$ uma base de Hilbert de F . Então $\{U^n(v_q) : n \in \mathbb{Z}, q \in Q\}$ é uma base de Hilbert de H . Dado $q \in Q$, represente por δ_q o elemento do espaço

$L^2(\lambda)^Q$ que é igual a 1 na coordenada q e igual a zero nas demais coordenadas. Defina

$$L : H \rightarrow L^2(\lambda)^Q, \quad L(U^n(v_q)) = z^n \delta_q \quad \text{para cada } n \in \mathbb{Z} \text{ e } q \in Q.$$

Observe que L é um operador unitário, uma vez que $\{z^n \delta_q\}$ é uma base de Hilbert de $L^2(\lambda)^Q$. Observe também que $LU = TL$. Isto fornece a representação espectral no enunciado da proposição, com χ igual ao cardinal do conjunto Q , ou seja, igual à dimensão de Hilbert do subespaço F . \square

Seja $E \subset H$ qualquer subespaço satisfazendo as condições na Definição 8.4.1. Então, a diferença ortogonal $F = E \ominus U(E)$ satisfaz a conclusão da Proposição 8.4.9, conforme vimos na demonstração dessa proposição. Além disso, de acordo com a demonstração da Proposição 8.4.10, podemos tomar o cardinal χ igual à dimensão de Hilbert de F . Como χ está unicamente determinado, o mesmo vale para a dimensão de Hilbert de $E \ominus U(E)$. Isto prova a Proposição 8.4.6 no caso invertível. A demonstração do caso geral está proposta no Exercício 8.4.3.

Acabamos de mostrar que o posto de um sistema com espectro de Lebesgue está bem definido. Vamos ver em seguida que, para sistemas invertíveis, o posto é um invariante completo de equivalência espectral:

Corolário 8.4.11. *Dois sistemas invertíveis com espectro de Lebesgue são espectralmente equivalentes se, e somente se, eles têm o mesmo posto.*

Demonstração. Claro que dois sistemas invertíveis são espectralmente equivalentes se, e somente se, eles admitem a mesma representação espectral. Pela Proposição 8.4.10 isto acontece se, e somente se, o valor do cardinal χ é o mesmo para os dois sistemas, ou seja, se o posto é o mesmo. \square

Corolário 8.4.12. *Todos os deslocamentos de Bernoulli bilaterais de tipo enumerável são espectralmente equivalentes.*

Demonstração. Todos os deslocamentos de Bernoulli de tipo enumerável têm posto enumerável, como vimos na seção anterior. \square

Demonstrações dos fatos citados a seguir podem ser encontradas no livro de Ricardo Mañé [Mañ87, Seção II.10]:

Exemplo 8.4.13 (Deslocamentos gaussianos). Dizemos que uma matriz infinita $A = (a_{i,j})_{i,j \in \mathbb{Z}}$ é *positiva definida* se a restrição $A_{m,n} = (a_{i,j})_{m \leq i,j < n}$ é positiva definida para quaisquer $m < n$. Dizemos que A é *simétrica* se $a_{i,j} = a_{j,i}$ para quaisquer $i, j \in \mathbb{Z}$. Seja μ uma probabilidade boreliana em $\Sigma = \mathbb{R}^{\mathbb{Z}}$ (valem considerações análogas para $\Sigma = \mathbb{R}^{\mathbb{N}}$). Dizemos que μ é uma *medida gaussiana* se existe alguma matriz simétrica positiva definida A tal que $\mu([m; B_m, \dots, B_{n-1}])$ é igual a

$$\frac{1}{(\det A_{m,n})^{1/2}} \frac{1}{(2\pi)^{(n-m)/2}} \int_{B_m \times \dots \times B_{n-1}} \exp\left(-\frac{1}{2}(A_{m,n}^{-1}z \cdot z)\right) dz$$

para quaisquer $m < n$ e quaisquer conjuntos mensuráveis $B_m, \dots, B_{n-1} \subset \mathbb{R}$. A razão de ser do fator do lado esquerdo da integral é explicada no Exercício 8.4.4. A é chamada *matriz de covariâncias* de μ . Ela está unicamente determinada, via

$$a_{i,j} = \int x_i x_j d\mu(x) \quad \text{para cada } i, j \in \mathbb{Z}.$$

Para cada matriz simétrica positiva definida A existe uma única probabilidade gaussiana μ que tem A como matriz de covariâncias. Além disso, μ é invariante pelo deslocamento $\sigma : \Sigma \rightarrow \Sigma$ se, e somente se, $a_{i,j} = a_{i+1,j+1}$ para quaisquer $i, j \in \mathbb{Z}$. Nesse caso, as propriedades do sistema (σ, μ) estão diretamente ligadas ao comportamento da *sequência de covariâncias*

$$\alpha_n = a_{n,0} = U_\sigma^n x_0 \cdot x_0 \quad \text{para cada } n \geq 0.$$

Em particular, (f, μ) é misturador se, e somente se, a sequência de covariâncias converge para zero.

Agora, o Exercício 8.4.5 mostra que se (f, μ) tem espectro de Lebesgue então a sequência de covariâncias é gerada por alguma medida absolutamente contínua ν no círculo unitário, no seguinte sentido:

$$\alpha_n = \int z^n d\nu(z) \quad \text{para cada } n \geq 0.$$

(O lema de Riemman-Lebesgue afirma que se ν é uma probabilidade absolutamente contínua relativamente à medida de Lebesgue λ no círculo unitário então a sequência $\int z^n d\nu(z)$ converge para zero quando $n \rightarrow \infty$.) Mas o Exercício 8.4.6 mostra que nem toda sequência que converge para zero é desta forma. Portanto, existem deslocamentos gaussianos (σ, μ) que são misturadores mas não têm espectro de Lebesgue.

8.4.3 Exercícios

8.4.1. Mostre que todo deslocamento de Markov misturador tem espectro de Lebesgue com posto enumerável. [Observação: Na Seção 9.5.3 mencionaremos resultados mais fortes.]

8.4.2. Seja μ a medida de Haar em \mathbb{T}^d e seja $f_A : \mathbb{T}^d \rightarrow \mathbb{T}^d$ um endomorfismo sobrejetivo. Suponha que nenhum autovalor da matriz A é raiz da unidade. Verifique que toda órbita de A^t no conjunto $\mathbb{Z}^d \setminus \{0\}$ é infinita e use esse fato para concluir que (f_A, μ) tem espectro de Lebesgue. Reciprocamente, se (f_A, μ) tem espectro de Lebesgue então nenhum autovalor de A é raiz da unidade.

8.4.3. Complete a demonstração da Proposição 8.4.6, usando o Exercício 2.3.6 para reduzir o caso geral ao caso invertível.

8.4.4. Verifique que $\int_{\mathbb{R}} e^{-x^2/2} dx = \sqrt{2\pi}$. Use esse fato para mostrar que se A é matriz simétrica positiva definida de dimensão $d \geq 1$ então

$$\int_{\mathbb{R}^d} \exp\left(-\frac{(A^{-1}z \cdot z)}{2}\right) dz = (\det A)^{1/2} (2\pi)^{d/2}$$

8.4.5. Seja (f, μ) um sistema invertível com espectro de Lebesgue. Mostre que para toda $\varphi \in L_0^2(\mu)$ existe uma probabilidade ν absolutamente contínua relativamente à medida de Lebesgue λ no círculo unitário tal que $U_f^n \varphi \cdot \varphi = \int z^n d\nu(z)$ para todo $n \in \mathbb{Z}$.

8.4.6. Seja λ a medida de Lebesgue no círculo unitário. Considere o operador linear $F : L^1(\lambda) \rightarrow c_0$ definido por

$$F(\varphi) = \left(\int z^n \varphi(z) d\lambda(z) \right)_n .$$

Mostre que F é contínuo e injetivo mas não é sobrejetivo. Portanto, nem toda sequência de números complexos $(\alpha_n)_n$ convergindo para zero pode ser escrita como $\alpha_n = \int z^n d\nu(z)$ para $n \geq 0$, para alguma probabilidade ν absolutamente contínua relativamente a λ .

8.5 Espaços de Lebesgue e isomorfismo ergódico

O principal tema desta seção são os *espaços de Lebesgue* (também chamados *espaços de probabilidade padrão*), uma classe de espaços de probabilidade introduzida pelo matemático soviético Vladimir A. Rokhlin [Rok62]. Estes espaços têm um papel de destaque na Teoria da Medida por duas razões: por um lado, eles apresentam propriedades muito melhores do que os espaços de probabilidade gerais; por outro lado, eles incluem a grande maioria dos exemplos interessantes. Em particular, todo espaço métrico completo separável, munido de uma probabilidade boreliana, é um espaço de Lebesgue.

Inicialmente, discutiremos uma terceira noção de equivalência, situada entre a equivalência ergódica e a equivalência espectral, que chamamos *isomorfismo ergódico*. Um dos que para transformações em espaços de Lebesgue as noções de equivalência ergódica e de isomorfismo ergódico coincidem.

8.5.1 Isomorfismo ergódico

Seja (M, \mathcal{B}, μ) um espaço de probabilidade. Denotamos por $\tilde{\mathcal{B}}$ o quociente da σ -álgebra pela relação de equivalência $A \sim B \Leftrightarrow \mu(A \Delta B) = 0$. Observe que se $A_k \sim B_k$ para todo $k \in \mathbb{N}$ então $\cup_k A_k \sim \cup_k B_k$ e $\cap_k A_k \sim \cap_k B_k$ e $A_k^c \sim B_k^c$ para todo $k \in \mathbb{N}$. Portanto, as operações usuais da teoria de conjuntos estão bem definidas no quociente $\tilde{\mathcal{B}}$. Além disso, a medida μ induz uma medida $\tilde{\mu}$ em $\tilde{\mathcal{B}}$. A dupla $(\tilde{\mathcal{B}}, \tilde{\mu})$ é chamada de *álgebra de medida* do espaço de probabilidade.

Agora sejam (M, \mathcal{B}, μ) e (N, \mathcal{C}, ν) dois espaços de probabilidade e sejam $(\tilde{\mathcal{B}}, \tilde{\mu})$ e $(\tilde{\mathcal{C}}, \tilde{\nu})$ as respectivas álgebras de medida. Um *homomorfismo* de álgebras de medida é uma aplicação $H : \tilde{\mathcal{B}} \rightarrow \tilde{\mathcal{C}}$ que preserva as operações de união, interseção e complementar e também preserva as medidas: $\mu(B) = \nu(H(B))$ para todo $B \in \tilde{\mathcal{B}}$. Se H é uma bijeção, dizemos que se trata de um *isomorfismo* de álgebras de medida. Nesse caso a inversa H^{-1} também é um isomorfismo de álgebras de medida.

Toda aplicação mensurável $h : M \rightarrow N$ satisfazendo $h_*\mu = \nu$ define um homomorfismo $\tilde{h} : \tilde{\mathcal{C}} \rightarrow \tilde{\mathcal{B}}$, via $B \mapsto h^{-1}(B)$. Além disso, se h é invertível então \tilde{h} é um isomorfismo. Do mesmo modo, transformações $f : M \rightarrow M$ e $g : N \rightarrow N$ preservando as medidas dos respectivos espaços de probabilidade definem homomorfismos $\tilde{f} : \tilde{\mathcal{B}} \rightarrow \tilde{\mathcal{B}}$ e $\tilde{g} : \tilde{\mathcal{C}} \rightarrow \tilde{\mathcal{C}}$, respectivamente. Dizemos que os sistemas (f, μ) e (g, ν) são *ergodicamente isomorfos* se estes homomorfismos são conjugados, ou seja, se $\tilde{f} \circ H = H \circ \tilde{g}$ para algum isomorfismo $H : \tilde{\mathcal{C}} \rightarrow \tilde{\mathcal{B}}$.

Sistemas ergodicamente equivalentes são sempre ergodicamente isomorfos: dada qualquer equivalência ergódica h , basta tomar $H = \tilde{h}$. Também temos a seguinte relação entre isomorfismo ergódico e equivalência espectral:

Proposição 8.5.1. *Se dois sistemas (f, μ) e (g, ν) são ergodicamente isomorfos então eles são espectralmente equivalentes.*

Demonstração. Seja $H : \tilde{\mathcal{C}} \rightarrow \tilde{\mathcal{B}}$ um isomorfismo tal que $\tilde{f} \circ H = H \circ \tilde{g}$. Considere o operador linear $L : L^2(\nu) \rightarrow L^2(\mu)$ construído da seguinte forma. Inicialmente, $L(\mathcal{X}_C) = \mathcal{X}_{H(C)}$ para todo $C \in \tilde{\mathcal{C}}$. Note que $\|L(\mathcal{X}_C)\| = \|\mathcal{X}_C\|$. Estenda a definição ao conjunto das funções simples, preservando a linearidade:

$$L\left(\sum_{j=1}^k c_j \mathcal{X}_{C_j}\right) = \sum_{j=1}^k c_j \mathcal{X}_{H(C_j)} \quad \text{para quaisquer } k \geq 1, c_j \in \mathbb{R} \text{ e } C_j \in \tilde{\mathcal{C}}.$$

A definição não depende da representação da função simples como combinação linear de funções características (Exercício 8.5.1). Além disso, $\|L(\varphi)\| = \|\varphi\|$ para toda função simples. Lembre que o conjunto das funções simples é denso em $L^2(\nu)$. Então, por continuidade, L se estende de modo único a uma isometria linear definida em todo o $L^2(\nu)$. Observe que esta isometria é invertível: a inversa é construída do mesmo modo, a partir da inversa de H . Finalmente,

$$U_f \circ L(\mathcal{X}_C) = U_f(\mathcal{X}_{H(C)}) = \mathcal{X}_{\tilde{f}(H(C))} = \mathcal{X}_{H(\tilde{g}(C))} = L(\mathcal{X}_{\tilde{g}(C)}) = L \circ U_g(\mathcal{X}_C)$$

para todo $C \in \tilde{\mathcal{C}}$. Por linearidade, segue que $U_f \circ L(\varphi) = L \circ U_g(\varphi)$ para toda função simples; então, por continuidade, o mesmo vale para todo $\varphi \in L^2(\nu)$. \square

Em resumo, temos a seguinte relação entre as três relações de equivalência:

$$\text{equivalência ergódica} \Rightarrow \text{isomorfismo ergódico} \Rightarrow \text{equivalência espectral}.$$

No que segue vamos discutir algumas recíprocas parciais, começando pela relação entre isomorfismo ergódico e equivalência espectral.

O seguinte resultado de Paul Halmos e John von Neumann [HvN42] amplia a Proposição 8.3.4 e mostra que para sistemas com espectro discreto as noções de isomorfismo ergódico e de equivalência espectral coincidem. O leitor pode encontrar a demonstração na Seção 3.2 do livro de Peter Walters [Wal75].

Teorema 8.5.2 (Espectro discreto). *Se (f, μ) e (g, ν) são sistemas ergódicos com espectro discreto então as seguintes condições são equivalentes:*

1. (f, μ) e (g, ν) são espectralmente equivalentes;
2. os operadores de Koopman de (f, μ) e (g, ν) têm os mesmos autovalores;
3. (f, μ) e (g, ν) são ergódicamente isomorfos.

Em particular, todo sistema ergódico invertível com espectro discreto é ergodicamente isomorfo ao seu inverso.

8.5.2 Espaços de Lebesgue

Seja (M, \mathcal{B}, μ) um espaço de probabilidade qualquer. Inicialmente, suponha que a medida μ é não atômica, ou seja que $\mu(\{x\}) = 0$ para todo $x \in M$. Suponha que é dada uma sequência crescente $\mathcal{P}_1 \prec \cdots \prec \mathcal{P}_n \prec \cdots$ de partições finitas de M em conjuntos mensuráveis. Dizemos que a sequência é *separadora* se dados pontos distintos $x, y \in M$ existe $n \geq 1$ tal que $\mathcal{P}_n(x) \neq \mathcal{P}_n(y)$. Em outras palavras, os elementos não vazios da partição $\bigvee_{n=1}^{\infty} \mathcal{P}_n$ contêm um único ponto.

Seja $M_{\mathcal{P}}$ o subconjunto obtido quando removemos de M todos os $P \in \bigcup_n \mathcal{P}_n$ com medida nula. Observe que $M_{\mathcal{P}}$ tem medida total. Representamos por $\mathcal{B}_{\mathcal{P}}$ e $\mu_{\mathcal{P}}$ as restrições de \mathcal{B} e μ , respectivamente, a $M_{\mathcal{P}}$. Seja m a medida de Lebesgue em \mathbb{R} . A próxima proposição significa que a sequência separadora nos permite representar o espaço de probabilidade $(M_{\mathcal{P}}, \mathcal{B}_{\mathcal{P}}, \mu_{\mathcal{P}})$ como uma espécie de subespaço da reta. Dizemos “espécie” porque, em geral, a imagem $\iota(M_{\mathcal{P}})$ não é um subconjunto mensurável de \mathbb{R} .

Proposição 8.5.3. *Dada qualquer sequência separadora $(\mathcal{P}_n)_n$, existe um compacto totalmente desconexo $K \subset \mathbb{R}$ e existe uma aplicação injetiva mensurável $\iota : M_{\mathcal{P}} \rightarrow K$ tal que para todo $P \in \bigcup_n \mathcal{P}_n$ o fecho da imagem $\iota(P)$ é um aberto fechado de K com $m(\overline{\iota(P)}) = \mu(P)$. Em particular, $\iota_*\mu$ coincide com a restrição da medida de Lebesgue m ao conjunto K .*

Demonstração. Escreva $\alpha_n = 1 + 1/n$ para $n \geq 1$. Vamos construir uma sequência de aplicações bijetivas $\psi_n : \mathcal{P}_n \rightarrow \mathcal{I}_n$, $n \geq 1$ satisfazendo:

- (a) cada \mathcal{I}_n é uma família finita de intervalos compactos, disjuntos dois-a-dois;
- (b) cada elemento de \mathcal{I}_n , $n > 1$ está contido em algum elemento de \mathcal{I}_{n-1} ;
- (c) $m(\psi_n(P)) = \alpha_n \mu(P)$ para todo $P \in \mathcal{P}_n$ e todo $n \geq 1$.

Para isso, comecemos por escrever $\mathcal{P}_1 = \{P_1, \dots, P_N\}$. Consideremos qualquer família $\mathcal{I}_1 = \{I_1, \dots, I_N\}$ de intervalos compactos, disjuntos dois-a-dois tais que $m(I_j) = \alpha_1 \mu(P_j)$ para todo j . Seja $\psi_1 : \mathcal{P}_1 \rightarrow \mathcal{I}_1$ a aplicação que a cada P_j associa o respectivo I_j . Agora suponhamos que, para um dado $n \geq 1$, foram construídas aplicações ψ_1, \dots, ψ_n satisfazendo (a), (b), (c). Para cada $P \in \mathcal{P}_n$, seja $I = \psi_n(P)$ e sejam P_1, \dots, P_N os elementos de \mathcal{P}_{n+1} contidos em P . Tomemos intervalos $I_1, \dots, I_N \subset I$ compactos, disjuntos dois-a-dois e satisfazendo $m(I_j) = \alpha_{n+1} \mu(P_j)$ para cada $j = 1, \dots, N$. Isto é possível porque, por hipótese de indução,

$$m(I) = \alpha_n \mu(P) = \alpha_n \sum_{j=1}^N \mu(P_j) > \alpha_{n+1} \sum_{j=1}^N \mu(P_j).$$

Então, defina $\psi_{n+1}(P_j) = I_j$ para cada $j = 1, \dots, N$. Repetindo este procedimento para cada $P \in \mathcal{P}_n$ completamos a definição de ψ_{n+1} e de \mathcal{I}_{n+1} . É claro que as condições (a), (b), (c) são preservadas. Isto encerra a construção.

Agora, seja $K = \bigcap_n \bigcup_{I \in \mathcal{I}_n} I$. É claro que K é compacto e que a sua interseção com qualquer $I \in \mathcal{I}_n$ é um aberto fechado de K . Além disso,

$$\max\{m(I) : I \in \mathcal{I}_n\} = \alpha_n \max\{\mu(P) : P \in \mathcal{P}_n\} \rightarrow 0 \quad \text{quando } n \rightarrow \infty \quad (8.5.1)$$

porque a sequência $(\mathcal{P}_n)_n$ é separadora e a medida μ é não atômica. Logo K é totalmente desconexo. Para cada $x \in M_{\mathcal{P}}$, os intervalos $\psi_n(\mathcal{P}_n(x))$ formam uma sequência decrescente de compactos cujos comprimentos decrescem para zero. Definimos $\iota(x)$ como sendo o único ponto em $\bigcap_n \psi_n(\mathcal{P}_n(x))$. A hipótese de que a sequência é separadora garante que ι é injetiva: se $x \neq y$ então existe $n \geq 1$ tal que $\mathcal{P}_n(x) \cap \mathcal{P}_n(y) = \emptyset$ e, portanto, $\iota(x) \neq \iota(y)$. Por construção, a pré-imagem de $K \cap I$ está em $\bigcup_n \mathcal{P}_n$ para todo $I \in \bigcup_n \mathcal{I}_n$. Considere a álgebra \mathcal{A} das uniões finitas disjuntas de conjuntos $K \cap I$ desta forma. Esta álgebra é geradora e acabamos de verificar que $\iota^{-1}(A)$ é um conjunto mensurável para todo $A \in \mathcal{A}$. Portanto, a transformação ι é mensurável.

Para verificar as propriedades no enunciado da proposição, comecemos por notar que, para todo $n \geq 1$ e $P \in \mathcal{P}_n$,

$$\overline{\iota(P)} = \bigcap_{k=n}^{\infty} \bigcup_Q \psi_k(Q), \quad (8.5.2)$$

onde a união é sobre todos os $Q \in \mathcal{P}_k$ que estão contidos em P . Para obter a inclusão \subset basta notar que $\iota(P) = \bigcup_Q \iota(Q)$ e que $\iota(Q) \subset \psi(Q)$ para todo $Q \in \mathcal{P}_k$ e todo k . A recíproca segue do fato de que $\iota(P)$ intersecta todo $\psi_k(Q)$ (a interseção contém $\iota(Q)$) e de que o comprimento dos $\psi_k(Q)$ converge para zero quando $k \rightarrow \infty$. Deste modo, fica provada a igualdade (8.5.2). Segue que

$$m(\overline{\iota(P)}) = \lim_k m\left(\bigcup_Q \psi_k(Q)\right) = \lim_k \sum_Q \alpha_k \mu(Q) = \lim_k \alpha_k \mu(P) = \mu(P).$$

Além disso, (8.5.2) significa que $\overline{\iota(P)} = \bigcap_{k=n}^{\infty} \bigcup_I I$, onde a união é sobre todos os $I \in \mathcal{I}_k$ que estão contidos em $\psi_n(P)$. O lado direito desta igualdade coincide com

$K \cap \psi_n(P)$ e, portanto, é um aberto fechado de K . Também segue da construção que $\iota^{-1}(\iota(P)) = P$. Consequentemente, $\iota_*\mu(\iota(P)) = \mu(P) = m(\iota(P))$ para todo $P \in \cup_n \mathcal{P}_n$. Como a álgebra das uniões finitas disjuntas de conjuntos $\iota(P)$ gera a estrutura mensurável de K , concluímos que $\iota_*\mu = m \upharpoonright K$. \square

Dizemos que um espaço de probabilidade sem átomos (M, \mathcal{B}, μ) é um *espaço de Lebesgue* se, para alguma sequência separadora, $\iota(M_{\mathcal{P}})$ é um conjunto mensurável de Lebesgue. De fato, esta propriedade não depende da escolha da sequência geradora (nem das famílias \mathcal{I}_n na demonstração da Proposição 8.5.3), mas não demonstraremos esse fato; o leitor interessado poderá encontrar uma demonstração em [Rok62, §2.2]. O Exercício 8.5.6 mostra que é possível definir espaço de Lebesgue de forma um pouco mais direta, sem usar a Proposição 8.5.3.

Note que se $\iota(M_{\mathcal{P}})$ é mensurável então $\iota(P) = \iota(M_{\mathcal{P}}) \cap \psi_n(P)$ é mensurável para todo $P \in \mathcal{P}_n$ e todo n . Logo, a inversa ι^{-1} também é uma transformação mensurável. Além disso, $m(\iota(M_{\mathcal{P}})) = \mu(M_{\mathcal{P}}) = 1 = m(K)$. Portanto, todo espaço de Lebesgue (M, \mathcal{B}, μ) é equivalente, como espaço de medida, a um subconjunto mensurável de um compacto totalmente desconexo da reta.

Observe que se M tem cardinalidade estritamente maior que a cardinalidade do contínuo então (M, \mathcal{B}, μ) não admite nenhuma sequência separadora e, portanto, não pode ser um espaço de Lebesgue. No Exercício 8.5.8 propomos outra construção de espaços de probabilidade que não são espaços de Lebesgue. Apesar de exemplos como estes, praticamente todos os espaços de probabilidade com que lidamos são espaços de Lebesgue:

Teorema 8.5.4. *Se M é um espaço métrico completo separável e μ é uma probabilidade boreliana sem átomos então (M, \mathcal{B}, μ) é um espaço de Lebesgue.*

Demonstração. Seja $X \subset M$ subconjunto enumerável denso e seja $\{B_n : n \in \mathbb{N}\}$ uma enumeração do conjunto das bolas $B(x, 1/k)$ com $x \in X$ e $k \geq 1$. Vamos construir uma sequência crescente $(\mathcal{P}_n)_n$ de partições finitas tal que:

- (i) \mathcal{P}_n é mais fina do que $\{B_1, B_1^c\} \vee \dots \vee \{B_n, B_n^c\}$ e
- (ii) $E_n = \{x \in M : \mathcal{P}_n(x) \text{ não é compacto}\}$ satisfaz $\mu(E_n) \leq 2^{-n}$.

Para isso, comece por considerar $\mathcal{Q}_1 = \{B_1, B_1^c\}$. Pela Proposição A.3.7, existem conjuntos compactos $K_1 \subset B_1$ e $K_2 \subset B_1^c$ tais que $\mu(B_1 \setminus K_1) \leq 2^{-1}\mu(B_1)$ e $\mu(B_1^c \setminus K_2) \leq 2^{-1}\mu(B_1^c)$. Então tome $\mathcal{P}_1 = \{K_1, B_1 \setminus K_1, K_2, B_1^c \setminus K_2\}$. Agora, para cada $n \geq 1$, suponha que já construímos partições $\mathcal{P}_1 \prec \dots \prec \mathcal{P}_n$ satisfazendo (i) e (ii). Considere a partição $\mathcal{Q}_{n+1} = \mathcal{P}_n \vee \{B_{n+1}, B_{n+1}^c\}$ e sejam Q_1, \dots, Q_m os seus elementos. Pela Proposição A.3.7, existem compactos $K_j \subset Q_j$ tais que $\mu(Q_j \setminus K_j) \leq 2^{-(n+1)}\mu(Q_j)$ para todo $j = 1, \dots, m$. Tome

$$\mathcal{P}_{n+1} = \{K_1, Q_1 \setminus K_1, \dots, K_m, Q_m \setminus K_m\}.$$

É claro que \mathcal{P}_{n+1} satisfaz (i) e (ii). Portanto, a nossa construção está completa.

Resta mostrar que a existência de tal sequência $(\mathcal{P}_n)_n$ implica a conclusão do teorema. A propriedade (i) garante que a sequência é separadora. Seja

$\iota : M_{\mathcal{P}} \rightarrow K$ uma aplicação como na Proposição 8.5.3. Fixe $N \geq 1$ qualquer. Considere qualquer ponto $y \in K \setminus \iota(M_{\mathcal{P}})$. Para cada $n > N$, seja I_n o intervalo da família \mathcal{I}_n que contém y e seja P_n o elemento de \mathcal{P}_n tal que $\psi_n(P_n) = I_n$. Note que os P_n formam uma sequência decrescente. Se eles fossem todos compactos, existiria $x \in \bigcap_{n>N} P_n$ e, por definição, $\iota(x)$ seria igual a y . Como estamos supondo que y não está na imagem de ι , isto prova que existe algum $l > N$ tal que P_l não é compacto. Tome $l > N$ mínimo nestas condições e seja $I_l = \psi_l(P_l)$. Lembre que $m(I_l) = \alpha_l \mu(P_l) \leq 2\mu(P_l)$. Sejam \tilde{I}_N e \tilde{P}_N as uniões de todos estes I_l e P_l , respectivamente, quando fazemos variar y em $K \setminus \iota(M_{\mathcal{P}})$. Por um lado, \tilde{I}_N contém $K \setminus \iota(M_{\mathcal{P}})$; por outro lado, \tilde{P}_N está contido em $\bigcup_{l>N} E_l$. Além disso, os I_l são disjuntos dois-a-dois (porque tomamos l mínimo) e o mesmo vale para os P_l . Logo,

$$m(\tilde{I}_N) \leq 2\mu(\tilde{P}_N) \leq 2\mu\left(\bigcup_{l>N} E_n\right) \leq 2^{-N+1}.$$

Então, a interseção $\bigcap_N \tilde{I}_N$ tem medida de Lebesgue nula e contém $K \setminus \iota(M_{\mathcal{P}})$. Como K é um boreliano, isto mostra que $\iota(M_{\mathcal{P}})$ é um conjunto mensurável de Lebesgue. \square

O próximo resultado mostra que todos os espaços de Lebesgue sem átomos são isomorfos.

Proposição 8.5.5. *Se (M, \mathcal{B}, μ) é um espaço de Lebesgue sem átomos, existe uma aplicação mensurável invertível $h : M \rightarrow [0, 1]$ (definida entre subconjuntos de medida total) tal que $h_*\mu$ coincide com a medida de Lebesgue em $[0, 1]$.*

Demonstração. Seja $\iota : M_{\mathcal{P}} \rightarrow K$ uma aplicação como na Proposição 8.5.3. Considere a aplicação $g : K \rightarrow [0, 1]$ definida por $g(x) = m([a, x] \cap K)$, onde $a = \min K$. É imediato da definição que g é não decrescente e Lischitz:

$$g(x_2) - g(x_1) = m([x_1, x_2] \cap K) \leq x_2 - x_1,$$

para quaisquer $x_1 < x_2$ em K . Em particular, g é mensurável. Por monotonia, a pré-imagem de qualquer intervalo $[y_1, y_2] \subset [0, 1]$ é um conjunto $[x_1, x_2] \cap K$ com $x_1, x_2 \in K$ e $g(x_1) = y_1$ e $g(x_2) = y_2$. Em particular,

$$m([x_1, x_2] \cap K) = g(x_2) - g(x_1) = y_2 - y_1 = m([y_1, y_2]).$$

Isto mostra que $g_*(m \upharpoonright K) = m \upharpoonright [0, 1]$. Seja Y o conjunto dos $y \in [0, 1]$ tais que $g^{-1}(\{y\}) = [x_1, x_2] \cap K$ com $x_1, x_2 \in K$ e $x_1 < x_2$. Seja $X = g^{-1}(Y)$. Então, $m(X) = m(Y) = 0$ porque Y é enumerável. Além disso, a restrição $g : K \setminus X \rightarrow [0, 1] \setminus Y$ é bijetiva. A sua inversa é não decrescente e, portanto, mensurável. Agora, tome $h = g \circ \iota$. Segue das observações anteriores que

$$h : M_{\mathcal{P}} \setminus \iota^{-1}(X) \rightarrow g(\iota(M_{\mathcal{P}})) \setminus Y$$

é uma bijeção mensurável com inversa mensurável tal que $h_*\mu = m \upharpoonright [0, 1]$. \square

Agora considere um espaço de probabilidade (M, \mathcal{B}, μ) qualquer, possivelmente com átomos. Seja $A \subset M$ o conjunto de todos os átomos; note que A é finito ou, no máximo, enumerável. Se o espaço é puramente atômico, ou seja, se $\mu(A) = 1$ então, por definição, ele é um espaço de Lebesgue. Mais geralmente, seja $M' = M \setminus A$, seja \mathcal{B}' a restrição de \mathcal{B} a M' e seja μ' a restrição normalizada de μ a \mathcal{B}' . Por definição, (M, \mathcal{B}, μ) é um espaço de Lebesgue se (M', \mathcal{B}', μ') é um espaço de Lebesgue.

O Teorema 8.5.4 continua válido no caso geral: todo espaço métrico completo separável munido de uma probabilidade boreliana, possivelmente com átomos, é um espaço de Lebesgue. Além disso, a Proposição 8.5.5 admite a seguinte generalização para o caso atômico: se (M, \mathcal{B}, μ) é um espaço de Lebesgue e $A \subset M$ representa o conjunto dos átomos da medida μ , então existe uma aplicação mensurável invertível $h : M \rightarrow [0, 1 - \mu(A)] \cup A$ tal que $h_*\mu$ coincide com m no intervalo $[0, 1 - \mu(A)]$ e coincide com μ em A .

Proposição 8.5.6. *Sejam (M, \mathcal{B}, μ) e (N, \mathcal{C}, ν) dois espaços de Lebesgue e seja $H : \tilde{\mathcal{C}} \rightarrow \tilde{\mathcal{B}}$ um isomorfismo entre as respectivas álgebras de medida. Então existe uma aplicação mensurável invertível $h : M \rightarrow N$ tal que $h_*\mu = \nu$ e $H = \tilde{h}$ para todo $C \in \tilde{\mathcal{C}}$. Além disso, h é essencialmente única: duas quaisquer aplicações satisfazendo estas propriedades coincidem em μ -quase todo ponto*

Vamos esboçar a demonstração desta proposição no caso não atômico. Os argumentos se baseiam nas ideias e notações da demonstração da Proposição 8.5.3.

Começemos pela unicidade. Sejam $h_1, h_2 : M \rightarrow N$ duas aplicações tais que $(h_1)_*\mu = (h_2)_*\mu = \nu$. Suponha que $h_1(x) \neq h_2(x)$ para todo x num conjunto $E \subset M$ com $\mu(E) > 0$. Seja $(Q_n)_n$ uma sequência separadora em (N, \mathcal{C}, ν) . Então, $Q_n(h_1(x)) \neq Q_n(h_2(x))$ para todo $x \in E$ e todo n suficientemente grande. Logo, podemos fixar n (grande) e $E' \subset E$ com $\mu(E') > 0$ tal que $Q_n(h_1(x)) \neq Q_n(h_2(x))$ para todo $x \in E'$. Consequentemente, existe $Q \in Q_n$ e $E'' \subset E'$ com $\mu(E'') > 0$ tal que $Q_n(h_1(x)) = Q$ e $Q_n(h_2(x)) \neq Q$ para todo $x \in E''$. Portanto, $E'' \subset h_1^{-1}(Q) \setminus h_2^{-1}(Q)$. Isto implica que $h_1(Q) \neq h_2(Q)$ e, logo, $\tilde{h}_1 \neq \tilde{h}_2$.

Passemos agora a comentar a existência. Sejam $(\mathcal{P}'_n)_n$ e $(\mathcal{Q}'_n)_n$ sequências separadoras em (M, \mathcal{B}, μ) e (N, \mathcal{C}, ν) , respectivamente. Defina $\mathcal{P}_n = \mathcal{P}'_n \vee H(\mathcal{Q}'_n)$ e $\mathcal{Q}_n = \mathcal{Q}'_n \vee H^{-1}(\mathcal{P}'_n)$. Então $(\mathcal{P}_n)_n$ e $(\mathcal{Q}_n)_n$ também são sequências separadoras e $\mathcal{P}_n = H(\mathcal{Q}_n)$ para cada n . Seja $\iota : M_{\mathcal{P}} \rightarrow K$ uma aplicação como na Proposição 8.5.3 e seja $\psi_n : \mathcal{P}_n \rightarrow \mathcal{I}_n$, $n \geq 1$ a família de bijeções utilizada na sua construção. Sejam $j : N_{\mathcal{Q}} \rightarrow L$ e $\varphi_n : \mathcal{Q}_n \rightarrow \mathcal{J}_n$ objetos correspondentes para (N, \mathcal{C}, ν) . Como estamos supondo que (M, \mathcal{B}, μ) e (N, \mathcal{C}, ν) são espaços de Lebesgue, ι e j são aplicações invertíveis sobre subconjuntos de medida total. Recorde também que $m(\psi_n(P)) = \alpha_n \mu(P)$ para cada $P \in \mathcal{P}_n$ e, analogamente, $m(\varphi_n(Q)) = \alpha_n \nu(Q)$ para cada $Q \in \mathcal{Q}_n$. Logo, $m(\psi_n(P)) = m(\varphi_n(Q))$ se $P = H(Q)$. Então, para cada n ,

$$\psi_n \circ H \circ \varphi_n^{-1} : \mathcal{J}_n \rightarrow \mathcal{I}_n \quad (8.5.3)$$

é uma bijeção que preserva o comprimento. Dados $z \in K$ e $n \geq 1$, seja I_n o elemento de \mathcal{I}_n que contém z e seja J_n o elemento de \mathcal{J}_n que lhe corresponde,

via (8.5.3). Por construção, $(J_n)_n$ é uma seqüência encaixada de intervalos compactos cujo comprimento converge para zero. Seja $\phi(z)$ o único ponto na sua interseção. Desta forma fica definida uma aplicação mensurável $\phi : K \rightarrow L$ que preserva a medida de Lebesgue. É claro da construção que ϕ é invertível e a inversa também é mensurável. Agora basta tomar $h = j^{-1} \circ \phi \circ \iota$.

Resta verificar que h é invertível. Aplicando a construção do parágrafo anterior à inversa H^{-1} encontramos $h' : N \rightarrow M$ tal que $h'_* \nu = \mu$ e $H^{-1} = \tilde{h}'$. Então, $\widetilde{h' \circ h} = \tilde{h} \circ \tilde{h}' = \text{id}$ e $\widetilde{h \circ h'} = \tilde{h}' \circ \tilde{h} = \text{id}$. Por unicidade, segue que $h' \circ h = \text{id}$ e $h \circ h' = \text{id}$ em quase todo ponto.

Corolário 8.5.7. *Sejam (M, \mathcal{B}, μ) e (N, \mathcal{C}, ν) dois espaços de Lebesgue e sejam $f : M \rightarrow M$ e $g : N \rightarrow N$ transformações mensuráveis preservando as medidas nos respectivos espaços. Então (f, μ) e (g, ν) são ergodicamente equivalentes se, e somente se, são ergodicamente isomorfos.*

Demonstração. Só precisamos mostrar que se os sistemas são ergodicamente ergódicos então eles são ergodicamente equivalentes. Seja $H : \tilde{\mathcal{C}} \rightarrow \tilde{\mathcal{B}}$ um isomorfismo ergódico. Pela Proposição 8.5.6, existe uma aplicação mensurável invertível $h : M \rightarrow N$ tal que $h_* \mu = \nu$ e $H = \tilde{h}$. Então

$$\widetilde{h \circ f} = \tilde{f} \circ \tilde{h} = \tilde{f} \circ H = H \circ \tilde{g} = \tilde{h} \circ \tilde{g} = \widetilde{g \circ h}.$$

Pela unicidade na Proposição 8.5.6, segue que $h \circ f = g \circ h$ em μ -quase todo ponto. Isto mostra que h é uma equivalência ergódica. \square

8.5.3 Exercícios

8.5.1. Seja $H : \tilde{\mathcal{C}} \rightarrow \tilde{\mathcal{B}}$ um homomorfismo de álgebras de medida. Mostre que

$$\sum_{i=1}^l b_i \mathcal{X}_{B_i} = \sum_{j=1}^k c_j \mathcal{X}_{C_j} \quad \Rightarrow \quad \sum_{i=1}^l b_i \mathcal{X}_{H(B_i)} = \sum_{j=1}^k c_j \mathcal{X}_{H(C_j)}.$$

8.5.2. Verifique que o homomorfismo de álgebra de medida $\tilde{g} : \tilde{\mathcal{C}} \rightarrow \tilde{\mathcal{B}}$ induzido por uma aplicação mensurável $g : M \rightarrow N$ que preserva medida é injetivo. Suponha que N é espaço de Lebesgue. Mostre que, dada outra aplicação mensurável $h : M \rightarrow N$ que preserva medida, os respectivos homomorfismos \tilde{g} e \tilde{h} coincidem se, e somente se, $g = h$ em quase todo ponto.

8.5.3. Seja $f : M \rightarrow M$ transformação mensurável num espaço de Lebesgue (M, \mathcal{B}, μ) , preservando a medida μ . Mostre que (f, μ) é invertível em quase todo ponto (ou seja, existe um subconjunto invariante de medida total restrito ao qual f é uma bijeção mensurável com inversa mensurável) se, e somente se, o respectivo homomorfismo de álgebras de medida $\tilde{f} : \tilde{\mathcal{B}} \rightarrow \tilde{\mathcal{B}}$ é sobrejetivo.

8.5.4. Mostre que o operador de Koopman de um sistema (f, μ) é sobrejetivo se, e somente se, o respectivo homomorfismo de álgebras de medida $\tilde{f} : \tilde{\mathcal{B}} \rightarrow \tilde{\mathcal{B}}$ é sobrejetivo. Em espaços de Lebesgue isso acontece se, e somente se, o sistema é invertível em quase todo ponto.

8.5.5. Mostre que todo sistema (f, μ) com espectro discreto num espaço de Lebesgue é invertível em quase todo ponto.

8.5.6. Dada uma sequência separadora $\mathcal{P}_1 \prec \dots \prec \mathcal{P}_n \prec \dots$, chamamos *cadeia* a qualquer sequência $(P_n)_n$ com $P_n \in \mathcal{P}_n$ e $P_{n+1} \subset P_n$ para todo n . Dizemos que a cadeia é vazia se $\bigcap_n P_n = \emptyset$. Considere a aplicação $\iota : M_{\mathcal{P}} \rightarrow K$ construída na Proposição 8.5.3. Mostre que $\iota(M_{\mathcal{P}})$ é conjunto mensurável de Lebesgue e $m(K \setminus \iota(M_{\mathcal{P}})) = 0$ se, e somente se, as cadeias vazias têm medida nula, no seguinte sentido: para todo $\delta > 0$ existe $B \subset M$ tal que B é uma união de elementos de $\bigcup_n \mathcal{P}_n$ com $\mu(B) < \delta$ e se $(P_n)_n$ é uma cadeia vazia então $P_n \subset B$ para todo n suficientemente grande.

8.5.7. Prove a seguinte extensão da Proposição 2.4.4: Se $f : M \rightarrow M$ preserva uma probabilidade μ e (M, μ) é um espaço de Lebesgue, então μ admite um (único) levantamento $\hat{\mu}$ para a extensão natural $\hat{f} : \hat{M} \rightarrow \hat{M}$.

8.5.8. Seja M um subconjunto de $[0, 1]$ com medida exterior $m^*(M) = 1$ mas que não é conjunto mensurável de Lebesgue. Considere a σ -álgebra \mathcal{M} dos conjuntos da forma $M \cap B$, onde B é subconjunto mensurável de Lebesgue \mathbb{R} . Verifique que $\mu(M \cap B) = m(B)$ define uma probabilidade em (M, \mathcal{M}) tal que (M, \mathcal{M}, μ) não é um espaço de Lebesgue.

Capítulo 9

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 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 soviéticos 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 9.1 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 9.2, constitui uma ferramenta fundamental para o cálculo da entropia de sistemas específicos. Na Seção 9.3 analisaremos a entropia de um ponto de vista mais local, que se relaciona diretamente com a formulação de Shannon. Em seguida, na Seção 9.4, ilustraremos alguns métodos de cálculo da entropia por meio de exemplos concretos.

Na Seção 9.5 discutiremos o papel da entropia como invariante de equivalência ergódica. O grande destaque é o teorema de Ornstein (Teorema 9.5.2) segundo o qual dois deslocamentos de Bernoulli bilaterais são ergodicamente

equivalentes se, e somente se, eles têm a mesma entropia. Nessa seção também introduzimos a classe dos sistemas de Kolmogorov, a qual inclui (estritamente) os deslocamentos de Bernoulli e está contida (também estritamente) na classe dos sistemas com espectro de Lebesgue.

Nas duas últimas seções apresentaremos dois tópicos complementares que serão úteis posteriormente. O primeiro (Seção 9.6) é o teorema de Jacobs, segundo o qual a entropia se comporta de modo afim relativamente à decomposição ergódica. O outro (Seção 9.7) diz respeito à noção de jacobiano e suas relações com a entropia.

9.1 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.

9.1.1 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

¹Um dos autores deste livro participou uma vez num jogo que consistia em seguir pistas para adivinhar, sucessivamente, certas letras que formariam o nome de um objeto matemático. Aconteceu que as três primeiras letras obtidas foram Z, Z e Z. Essa circunstância arruinou a continuação do jogo, pois as demais letras não acrescentariam qualquer informação: o único objeto matemático cuja nome inclui três vezes a letra Z é o *Puzzle de Yoccoz*.

Por outro lado, quando 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 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

$$I(a) = -\log p_a \quad (9.1.1)$$

onde p_a é a probabilidade (frequência) do caracter a . A *informação média* associada ao alfabeto \mathcal{A} é dada por

$$I(\mathcal{A}) = \sum_a -p_a I(a) = \sum_a -p_a \log p_a. \quad (9.1.2)$$

Mais geralmente, a informação associada a uma palavra $a_1 \dots a_n$ é

$$I(a_1 \dots a_n) = -\log p_{a_1 \dots a_n} \quad (9.1.3)$$

onde a probabilidade $p_{a_1 \dots a_n}$ da palavra é, usualmente, maior que o produto $p_{a_1} \dots p_{a_n}$ das probabilidades das suas letras (vale a igualdade no caso independente). Denotando por \mathcal{A}^n o conjunto de todas as palavras de comprimento n , definimos

$$I(\mathcal{A}^n) = \sum_{a_1, \dots, a_n} -p_{a_1 \dots a_n} I(a_1, \dots, a_n) = \sum_{a_1, \dots, a_n} -p_{a_1 \dots a_n} \log p_{a_1 \dots a_n}. \quad (9.1.4)$$

Finalmente, a *entropia* do canal de comunicação é definida por:

$$I = \lim_n \frac{1}{n} I(\mathcal{A}^n). \quad (9.1.5)$$

9.1.2 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, finitas ou enumeráveis, do espaço de estados.

Seja (M, \mathcal{B}, μ) um espaço de probabilidade. Neste capítulo, por *partição* sempre entenderemos uma família finita ou enumerável \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 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

$$\bigvee_n \mathcal{P}_n = \left\{ \bigcap_n P_n : P_n \in \mathcal{P}_n \text{ para cada } n \right\}$$

A cada partição \mathcal{P} associamos a respectiva *função de informação*

$$I_{\mathcal{P}} : M \rightarrow \mathbb{R}, \quad I_{\mathcal{P}}(x) = -\log \mu(\mathcal{P}(x)). \quad (9.1.6)$$

É claro que a função $I_{\mathcal{P}}$ é mensurável. Então chamamos *entropia*, ou *informação média*, da partição \mathcal{P} ao número

$$H_{\mu}(\mathcal{P}) = \int I_{\mathcal{P}} d\mu = \sum_{P \in \mathcal{P}} -\mu(P) \log \mu(P). \quad (9.1.7)$$

Como é usual na teoria da integral de Lebesgue, fazemos a convenção de que $0 \log 0 = \lim_{x \rightarrow 0} x \log x = 0$. Veja a Figura 9.1.

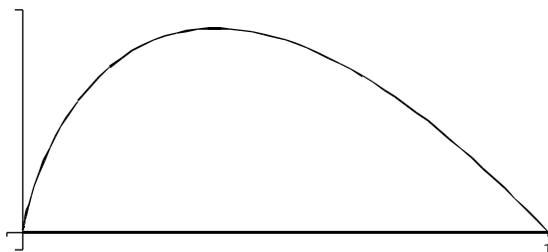


Figura 9.1: Gráfico da função $\phi(x) = -x \log x$

Relacionado com isto, a seguinte observação será útil em diversas ocasiões. Considere a função $\phi : (0, \infty) \rightarrow \mathbb{R}$ dada por $\phi(x) = -x \log x$. Derivando duas vezes vemos que $\phi'' < 0$. Portanto ϕ é côncava:

$$t_1 \phi(x_1) + \cdots + t_k \phi(x_k) \leq \phi(t_1 x_1 + \cdots + t_k x_k) \quad (9.1.8)$$

para todo $x_1, \dots, x_k > 0$ e $t_1, \dots, t_k > 0$ com $t_1 + \cdots + t_k = 1$. Além disso, a concavidade é estrita: vale a igualdade em (9.1.8) 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 \leq como veremos daqui a pouco.

Exemplo 9.1.1. Considere $M = [0, 1]$ munido da medida de Lebesgue. Para cada $n \geq 1$ considere a partição \mathcal{P}^n nos subintervalos $((i-1)/10^n, i/10^n]$ com $1 \leq i \leq 10^n$. Então

$$H_\mu(\mathcal{P}^n) = \sum_{i=1}^{10^n} -10^{-n} \log 10^{-n} = n \log 10.$$

Exemplo 9.1.2. Seja $M = \{1, \dots, d\}^{\mathbb{N}}$ munido de uma medida produto $\mu = \nu^{\mathbb{N}}$. Denotamos $p_i = \nu(\{i\})$ para cada $i \in \{1, \dots, d\}$. Para cada $n \geq 1$, seja \mathcal{P}^n a partição de M em cilindros $[0; a_1, \dots, a_n]$ de comprimento n . A entropia de \mathcal{P}^n é

$$\begin{aligned} H_\mu(\mathcal{P}^n) &= \sum_{a_1, \dots, a_n} -p_{a_1} \dots p_{a_n} \log(p_{a_1} \dots p_{a_n}) \\ &= \sum_j \sum_{a_1, \dots, a_n} -p_{a_1} \dots p_{a_j} \dots p_{a_n} \log p_{a_j} \\ &= \sum_j \sum_{a_j} -p_{a_j} \log p_{a_j} \sum_{a_i, i \neq j} p_{a_1} \dots p_{a_{j-1}} p_{a_{j+1}} \dots p_{a_n}. \end{aligned}$$

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 9.1.3. *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 $\mathcal{P} = \{P_1, P_2, \dots, P_n\}$ e considere os números $t_i = 1/n$ e $x_i = \mu(P_i)$. Pela desigualdade de Jensen (Teorema A.5.8):

$$\frac{1}{n} H_\mu(\mathcal{P}) = \sum_{i=1}^n t_i \phi(x_i) \leq \phi\left(\sum_{i=1}^n t_i x_i\right) = \phi\left(\frac{1}{n}\right) = \frac{\log n}{n}.$$

Portanto, $H_\mu(\mathcal{P}) \leq \log n$. Além disso, a igualdade ocorre se, e somente se, $\mu(P_i) = 1/n$ para todo $i = 1, \dots, n$. \square

O exemplo a seguir mostra que partições enumeráveis podem ter entropia infinita. A partir daqui, ao longo de todo o capítulo, sempre consideraremos partições (finitas ou enumeráveis) com entropia finita. Alguns enunciados serão restritos ao caso de partições finitas.

Exemplo 9.1.4. 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 intervalos. 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, $H_{\mu}(\mathcal{P}) = \infty$.

Chamamos *entropia condicional* de uma partição \mathcal{P} com relação a uma partição \mathcal{Q} ao número

$$H_{\mu}(\mathcal{P}/\mathcal{Q}) = \sum_{P \in \mathcal{P}} \sum_{Q \in \mathcal{Q}} -\mu(P \cap Q) \log \frac{\mu(P \cap Q)}{\mu(Q)}. \quad (9.1.9)$$

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 9.1.5. *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,

$$\begin{aligned} H_{\mu}(\mathcal{P} \vee \mathcal{Q}/\mathcal{R}) &= \sum_{P, Q, R} -\mu(P \cap Q \cap R) \log \frac{\mu(P \cap Q \cap R)}{\mu(R)} \\ &= \sum_{P, Q, R} -\mu(P \cap Q \cap R) \log \frac{\mu(P \cap Q \cap R)}{\mu(P \cap R)} \\ &\quad + \sum_{P, Q, R} -\mu(P \cap Q \cap R) \log \frac{\mu(P \cap R)}{\mu(R)}. \end{aligned}$$

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 desigualdade de Jensen (Teorema A.5.8),

$$\phi\left(\frac{\mu(R \cap P)}{\mu(P)}\right) \geq \sum_{Q \subset P} \frac{\mu(Q)}{\mu(P)} \phi\left(\frac{\mu(R \cap Q)}{\mu(Q)}\right)$$

para todo $P \in \mathcal{P}$ e $R \in \mathcal{R}$. Consequentemente,

$$\begin{aligned} H_\mu(\mathcal{R}/\mathcal{P}) &= \sum_{P,R} \mu(P) \phi\left(\frac{\mu(R \cap P)}{\mu(P)}\right) \geq \sum_{P,R} \mu(P) \sum_{Q \subset P} \frac{\mu(Q)}{\mu(P)} \phi\left(\frac{\mu(R \cap Q)}{\mu(Q)}\right) \\ &= \sum_{Q,R} \mu(Q) \phi\left(\frac{\mu(R \cap Q)}{\mu(Q)}\right) = H_\mu(\mathcal{R}/\mathcal{Q}). \end{aligned}$$

Finalmente, segue da definição (9.1.9) 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, $\mathcal{P} \prec \mathcal{Q}$. \square

Em particular, tomando $\mathcal{Q} = \mathcal{M}$ no item (b) do lema obtemos que

$$H_\mu(\mathcal{R}/\mathcal{P}) \leq H_\mu(\mathcal{R}) \quad \text{para quaisquer partições } \mathcal{R} \text{ e } \mathcal{P}. \quad (9.1.10)$$

Além disso, tomando $\mathcal{R} = \mathcal{M}$ no item (a), vem que

$$H_\mu(\mathcal{P} \vee \mathcal{Q}) = H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q}/\mathcal{P}) \leq H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q}). \quad (9.1.11)$$

Seja $f : M \rightarrow N$ 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:

$$\begin{aligned} H_\mu(f^{-1}(\mathcal{P})) &= \sum_{P \in \mathcal{P}} -\mu(f^{-1}(P)) \log \mu(f^{-1}(P)) \\ &= \sum_{P \in \mathcal{P}} -f_*\mu(P) \log f_*\mu(P) = H_{f_*\mu}(\mathcal{P}). \end{aligned} \quad (9.1.12)$$

Em particular, se $M = N$ e a medida μ é invariante por f então

$$H_\mu(f^{-1}(\mathcal{P})) = H_\mu(\mathcal{P}) \quad \text{para toda partição } \mathcal{P}. \quad (9.1.13)$$

Também precisaremos da seguinte propriedade de continuidade:

Lema 9.1.6. *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 $\phi : [0, 1] \rightarrow \mathbb{R}$, $\phi(x) = -x \log x$, 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

$$\mu\left(\bigcup_{i=1}^k P_i \cap Q_i\right) \geq \sum_{i=1}^k (\mu(P_i) - \mu(P_i \Delta Q_i)) > \sum_{i=1}^k (\mu(P_i) - \delta) = 1 - \rho$$

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 (9.1.11) e (9.1.10),

$$\begin{aligned} H_\mu(\mathcal{Q}/\mathcal{P}) &= H_\mu(\mathcal{P} \vee \mathcal{Q}) - H_\mu(\mathcal{P}) = H_\mu(\mathcal{P} \vee \mathcal{R}) - H_\mu(\mathcal{P}) \\ &= H_\mu(\mathcal{R}/\mathcal{P}) \leq H_\mu(\mathcal{R}) < \varepsilon. \end{aligned}$$

Isto prova o lema. □

9.1.3 Entropia de um sistema dinâmico

Seja $f : M \rightarrow M$ 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 (9.1.5).

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:

$$\mathcal{P}^n(x) = \mathcal{P}(x) \cap f^{-1}(\mathcal{P}(f(x))) \cap \dots \cap f^{-n+1}(\mathcal{P}(f^{n-1}(x))).$$

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 9.1.7. $H_\mu(\mathcal{P}^{m+n}) \leq H_\mu(\mathcal{P}^m) + H_\mu(\mathcal{P}^n)$ para todo $m, n \geq 1$.

Demonstração. Por definição, $\mathcal{P}^{m+n} = \bigvee_{i=0}^{m+n-1} f^{-i}(\mathcal{P}) = \mathcal{P}^m \vee f^{-m}(\mathcal{P}^n)$. Portanto, usando (9.1.11),

$$H_\mu(\mathcal{P}^{m+n}) \leq H_\mu(\mathcal{P}^m) + H_\mu(f^{-m}(\mathcal{P}^n)). \quad (9.1.14)$$

Por outro lado, como μ é invariante por f , a propriedade (9.1.13) implica que $H_\mu(f^{-m}(\mathcal{P}^n)) = H_\mu(\mathcal{P}^n)$ para todo m, n . Substituindo este fato em (9.1.14) obtemos a conclusão do lema. \square

Chamamos *entropia de f com respeito à medida μ e à partição \mathcal{P}* o limite

$$h_\mu(f, \mathcal{P}) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}^n) = \inf_n \frac{1}{n} H_\mu(\mathcal{P}^n). \quad (9.1.15)$$

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 o n . Usando o Lema 9.1.5, segue que $H_\mu(\mathcal{P}^n) \leq H_\mu(\mathcal{Q}^n)$ para todo n . Consequentemente,

$$\mathcal{P} \prec \mathcal{Q} \quad \Rightarrow \quad h_\mu(f, \mathcal{P}) \leq h_\mu(f, \mathcal{Q}) \quad (9.1.16)$$

Finalmente, a *entropia* do sistema (f, μ) é definida por

$$h_\mu(f) = \sup_{\mathcal{P}} h_\mu(f, \mathcal{P}), \quad (9.1.17)$$

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 9.1.1).

Exemplo 9.1.8. Suponhamos que a medida invariante μ está suportada numa órbita periódica. Em outras palavras, existe x em M e $k \geq 1$ tal que $f^k(x) = x$ e a medida μ é dada por

$$\mu = \frac{1}{k} (\delta_x + \delta_{f(x)} + \cdots + \delta_{f^{k-1}(x)}).$$

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 enumeráveis \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 9.1.9. Considere a transformação expansão decimal $f : [0, 1] \rightarrow [0, 1]$, 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 $((i-1)/10, i/10]$ com $i = 1, \dots, 10$. Então \mathcal{P}^n é a partição nos intervalos da forma $((i-1)/10^n, i/10^n]$ com $i = 1, \dots, 10^n$. Usando o cálculo do Exemplo 9.1.1, 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 9.2, veremos que este é também o valor da entropia $h_\mu(f)$, ou seja, \mathcal{P} realiza o supremo na definição (9.1.17).

Exemplo 9.1.10. Considere o deslocamento $\sigma : \Sigma \rightarrow \Sigma$ 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 9.1.2 concluímos que

$$h_\mu(\sigma, \mathcal{P}) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}^n) = \sum_{i=1}^d -p_i \log p_i. \quad (9.1.18)$$

A teoria que apresentaremos na Seção 9.2 permitirá concluir que este é também o valor da entropia $h_\mu(\sigma)$.

Segue da expressão (9.1.18) 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 9.1.11. $h_\mu(f, \mathcal{Q}) \leq h_\mu(f, \mathcal{P}) + H_\mu(\mathcal{Q}/\mathcal{P})$ para quaisquer partições \mathcal{P} e \mathcal{Q} com entropia finita.

Demonstração. Pelo Lema 9.1.5, para todo $n \geq 1$ vale que

$$\begin{aligned} H_\mu(\mathcal{Q}^{n+1}/\mathcal{P}^{n+1}) &= H_\mu(\mathcal{Q}^n \vee f^{-n}(\mathcal{Q})/\mathcal{P}^n \vee f^{-n}(\mathcal{P})) \\ &\leq H_\mu(\mathcal{Q}^n/\mathcal{P}^n) + H_\mu(f^{-n}(\mathcal{Q})/f^{-n}(\mathcal{P})) \end{aligned}$$

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

$$H_\mu(\mathcal{Q}^n/\mathcal{P}^n) \leq nH_\mu(\mathcal{Q}/\mathcal{P}) \quad \text{para todo } n \geq 1. \quad (9.1.19)$$

Usando o Lema 9.1.5 uma vez mais, segue que

$$H_\mu(\mathcal{Q}^n) \leq H_\mu(\mathcal{P}^n \vee \mathcal{Q}^n) = H_\mu(\mathcal{P}^n) + H_\mu(c\mathcal{Q}^n/\mathcal{P}^n) \leq H_\mu(\mathcal{P}^n) + nH_\mu(\mathcal{Q}/\mathcal{P}).$$

Dividindo por n e passando ao limite quando $n \rightarrow \infty$ obtemos a conclusão do lema. \square

Lema 9.1.12. $h_\mu(f, \mathcal{P}) = \lim_n H_\mu(\mathcal{P}/\bigvee_{j=1}^n f^{-j}(\mathcal{P}))$ para qualquer partição \mathcal{P} com entropia finita.

Demonstração. Usando o Lema 9.1.5(a) e o fato de que a medida μ é invariante:

$$\begin{aligned} H_\mu\left(\bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P})\right) &= H_\mu\left(\bigvee_{j=1}^{n-1} f^{-j}(\mathcal{P})\right) + H_\mu\left(\mathcal{P}/\bigvee_{j=1}^{n-1} f^{-j}(\mathcal{P})\right) \\ &= H_\mu\left(\bigvee_{j=0}^{n-2} f^{-j}(\mathcal{P})\right) + H_\mu\left(\mathcal{P}/\bigvee_{j=1}^{n-1} f^{-j}(\mathcal{P})\right) \end{aligned}$$

para todo n . Por recorrência, segue que

$$H_\mu\left(\bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P})\right) = H_\mu(\mathcal{P}) + \sum_{k=1}^{n-1} H_\mu\left(\mathcal{P} / \bigvee_{j=1}^k f^{-j}(\mathcal{P})\right).$$

Portanto, $h_\mu(f, \mathcal{P})$ é dada pelo limite Cesaro

$$h_\mu(f, \mathcal{P}) = \lim_n \frac{1}{n} H_\mu\left(\bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P})\right) = \lim_n \frac{1}{n} \sum_{k=1}^{n-1} H_\mu\left(\mathcal{P} / \bigvee_{j=1}^k f^{-j}(\mathcal{P})\right).$$

Por outro lado, o Lema 9.1.5(b) garante que a sequência $H_\mu(\mathcal{P} / \bigvee_{j=1}^n f^{-j}(\mathcal{P}))$ é decrescente. Em particular, $\lim_n \bigvee_{j=1}^n f^{-j}(\mathcal{P})$ existe e, conseqüentemente, coincide com o limite Cesaro na igualdade anterior. \square

Recorde que $\mathcal{P}^n = \bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P})$. Quando $f : M \rightarrow M$ é invertível, também consideramos $\mathcal{P}^{\pm n} = \bigvee_{j=-n}^{n-1} f^{-j}(\mathcal{P})$.

Lema 9.1.13. *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 $n \geq 1$,

$$\bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P}^k) = \bigvee_{j=0}^{n-1} f^{-j}\left(\bigvee_{i=0}^{k-1} f^{-i}(\mathcal{P})\right) = \bigvee_{l=0}^{n+k-2} f^{-l}(\mathcal{P}) = \mathcal{P}^{n+k-1}.$$

Portanto,

$$h_\mu(f, \mathcal{P}^k) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}^{n+k-1}) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}^n) = h_\mu(f, \mathcal{P}).$$

Isto prova a primeira parte do lema. Para provar a segunda parte, note que:

$$\bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P}^{\pm k}) = \bigvee_{j=0}^{n-1} f^{-j}\left(\bigvee_{i=-k}^{k-1} f^{-i}(\mathcal{P})\right) = \bigvee_{l=-k}^{n+k-2} f^{-l}(\mathcal{P}) = f^{-k}(\mathcal{P}^{n+2k-1})$$

para todo n e todo k . Portanto,

$$h_\mu(f, \mathcal{P}^{\pm k}) = \lim_n \frac{1}{n} H_\mu(f^{-k}(\mathcal{P}^{n+2k-1})) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}^{n+2k-1}) = h_\mu(f, \mathcal{P}).$$

(a segunda igualdade usa o fato de que μ é invariante por f). \square

Proposição 9.1.14. *Tem-se $h_\mu(f^k) = kh_\mu(f)$ 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. Considere $g = f^k$ e seja \mathcal{P} uma partição qualquer de M com entropia finita. Lembrando que $\mathcal{P}^k = \mathcal{P} \vee f^{-1}(\mathcal{P}) \vee \dots \vee f^{-k+1}(\mathcal{P})$, vemos que

$$\mathcal{P}^{km} = \bigvee_{i=0}^{km-1} f^{-j}(\mathcal{P}) = \bigvee_{i=0}^{m-1} f^{-ki} \left(\bigvee_{j=0}^{k-1} f^{-j}(\mathcal{P}) \right) = \bigvee_{i=0}^{m-1} g^{-i}(\mathcal{P}^k).$$

Portanto,

$$kh_\mu(f, \mathcal{P}) = \lim_m \frac{1}{m} H_\mu(\mathcal{P}^{km}) = h_\mu(g, \mathcal{P}^k). \quad (9.1.20)$$

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 $h_\mu(g) \leq kh_\mu(f) \leq h_\mu(g)$. Ist 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 $n \geq 1$,

$$H_\mu(\bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P})) = H_\mu(f^{-n+1}(\bigvee_{i=0}^{n-1} f^i(\mathcal{P}))) = H_\mu(\bigvee_{i=0}^{n-1} f^i(\mathcal{P})),$$

uma vez que a medida μ é invariante. Dividindo por n e passando ao limite quando $n \rightarrow \infty$, obtemos que

$$h_\mu(f, \mathcal{P}) = h_\mu(f^{-1}, \mathcal{P}). \quad (9.1.21)$$

Tomando o supremo sobre estas partições \mathcal{P} , vem que $h_\mu(f) = h_\mu(f^{-1})$. Substituindo f for f^k e usando o item (a), segue que $h_\mu(f^{-k}) = h_\mu(f^k) = kh_\mu(f)$ para todo $k \in \mathbb{N}$. \square

9.1.4 Exercícios

9.1.1. 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.

9.1.2. Verifique que $\lim_n H_\mu(\bigvee_{i=0}^{k-1} f^{-i}(\mathcal{P}) / \bigvee_{j=k}^n f^{-j}(\mathcal{P})) = kh(f, \mathcal{P})$ para toda partição \mathcal{P} com entropia finita e todo $k \geq 1$.

9.1.3. Seja $f : M \rightarrow M$ 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 μ_1, \dots, μ_n . Mostre que $h_\mu(f) = \sum_{i=1}^n a_i h_{\mu_i}(f)$.

[Observação: Na Seção 9.6 obteremos resultados mais fortes.]

9.1.4. Considere transformações $f : M \rightarrow M$ e $g : N \rightarrow N$ preservando medidas de probabilidade μ e ν , respectivamente. Considere $f \times g : M \times M \rightarrow N \times N$, 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 $h_{\mu \times \nu}(f \times g) = h_\mu(f) + h_\nu(g)$.

9.1.5. Sejam (M, \mathcal{B}, μ) e (N, \mathcal{C}, ν) espaços de probabilidade e $f : M \rightarrow M$ e $g : N \rightarrow N$ 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, $\phi : (M, \mathcal{B}) \rightarrow (N, \mathcal{C})$ tal que $\phi_*\mu = \nu$ e $\phi \circ f = g \circ \phi$ em quase todo ponto. Mostre que nesse caso $h_\nu(g) \leq h_\mu(f)$.

9.2 Teorema de Kolmogorov-Sinai

Em geral, a principal dificuldade no cálculo da entropia reside no cálculo do supremo na definição (9.1.17). Os métodos que vamos desenvolver nesta seção permitem simplificar a tarefa em muitos casos de interesse, 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 9.2.1 (Kolmogorov-Sinai). *Seja $\mathcal{P}_1 \prec \dots \prec \mathcal{P}_n \prec \dots$ uma seqüência 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 (9.1.16) implica que a seqüência $h_\mu(f, \mathcal{P}_n)$ é não decrescente. Vamos usar o seguinte fato:

Lema 9.2.2. $\lim_n H_\mu(\mathcal{Q}/\mathcal{P}_n) = 0$ para qualquer partição finita \mathcal{Q} .

Demonstração. Escreva $\mathcal{Q} = \{Q_1, \dots, Q_k\}$. Dado qualquer $\varepsilon > 0$, fixe $\delta > 0$ como no Lema 9.1.6. Seja \mathcal{A} a álgebra formada pelas uniões finitas de elementos de $\cup_n \mathcal{P}_n$. Por hipótese, \mathcal{A} gera a σ -álgebra de todos os conjuntos mensuráveis. Logo, pelo teorema de aproximação (Teorema A.1.19), para cada $i = 1, \dots, k$ existe $A_i \in \mathcal{A}$ tal que

$$\mu(Q_i \Delta A_i) < \delta/(4k). \quad (9.2.1)$$

O fato de que os Q_i são uma cobertura de M garante que os A_i estão perto de o serem também:

$$\mu(A_i \cap (\cup_{j \neq i} A_j)) \leq \mu(\cup_{j=1}^n (A_j \setminus Q_j)) < \delta/4 \quad \text{para todo } i \quad (9.2.2)$$

$$\text{e } \mu(M \setminus \cup_{i=1}^k A_i) \leq \mu(\cup_{i=1}^k (Q_i \setminus A_i)) < \delta/4. \quad (9.2.3)$$

A seguir, defina

$$Q'_i = \begin{cases} A_1 & \text{para } i = 1 \\ A_i \setminus \cup_{j=1}^{i-1} A_j & \text{para } 1 < i < k \\ M \setminus \cup_{j=1}^{k-1} A_j & \text{para } i = k \end{cases}$$

Então $\mathcal{Q}' = \{Q'_1, \dots, Q'_k\}$ é uma partição de M . Afirmamos que

$$\mu(A_i \Delta Q'_i) < \delta/2 \quad \text{para todo } i = 1, \dots, k. \quad (9.2.4)$$

Isto é trivial para $i = 1$. Para $i > 1$ temos que $A_i \setminus Q'_i$ está contido em $A_i \cap (\cup_{j < i} A_j)$. Logo, usando (9.2.2), obtemos que $\mu(A_i \setminus Q'_i) < \delta/4$. Isto prova a afirmação para todo $1 < i < k$, uma vez que nesse caso $Q'_i \setminus A_i = \emptyset$. Finalmente, para $i = k$, temos que $Q'_k \setminus A_k$ está contido no complementar de $\cup_{i=1}^k A_i$. Logo, usando (9.2.3), vemos que $\mu(Q'_k \setminus A_k) < \delta/4$. Somando esta estimativa com a anterior, vem que $\mu(A_k \Delta Q'_k) < \delta/2$. Isto completa a prova da afirmação (9.2.4).

Combinando as desigualdades (9.2.1) e (9.2.4), obtemos que $\mu(Q_i \Delta Q'_i) < \delta$ para todo $i = 1, \dots, k$. Agora, é claro que $Q'_i \in \mathcal{A}$ para todo i . Então, como se trata de uma família finita, podemos encontrar $m \geq 1$ tal que todo Q'_i é uma união de elementos de \mathcal{P}_m . Em outras palavras, a partição $\mathcal{Q}' = \{Q'_1, \dots, Q'_k\}$ é menos fina do que \mathcal{P}_m . Então, pelos Lemas 9.1.5 e 9.1.6,

$$H_\mu(\mathcal{Q}/\mathcal{P}_n) \leq H_\mu(\mathcal{Q}/\mathcal{P}_m) \leq H_\mu(\mathcal{Q}/\mathcal{Q}') < \varepsilon \quad \text{para todo } n \geq m.$$

Isto completa a demonstração do lema. \square

Pelo Lema 9.1.11, também 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 $n \rightarrow \infty$ e, então, tomando o supremo sobre todas as partições finitas \mathcal{Q} , obtemos a conclusão do teorema. \square

9.2.1 Partições geradoras

Nesta seção, e nas seguintes, vamos deduzir diversas consequências úteis do Teorema 9.2.1.

Corolário 9.2.3. *Seja \mathcal{P} uma partição com entropia finita tal que a união dos seus iterados $\mathcal{P}^n = \bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P})$, $n \geq 1$ gera a σ -álgebra dos conjuntos mensuráveis. Então $h_\mu(f) = h_\mu(f, \mathcal{P})$.*

Demonstração. Basta aplicar o Teorema 9.2.1 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 9.1.13. \square

Corolário 9.2.4. *Suponha que o sistema (f, μ) é invertível. Seja \mathcal{P} uma partição com entropia finita tal que a união dos iterados $\mathcal{P}^{\pm n} = \bigvee_{j=-n}^{n-1} f^{-j}(\mathcal{P})$, $n \geq 1$ gera a σ -álgebra dos conjuntos mensuráveis. Então $h_\mu(f) = h_\mu(f, \mathcal{P})$.*

Demonstração. Basta aplicar o Teorema 9.2.1 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 9.1.13. \square

Em particular, os Corolários 9.2.3 e 9.2.4 completam o cálculo da entropia da transformação expansão decimal e dos deslocamentos de Bernoulli, que iniciamos nos Exemplos 9.1.9 e 9.1.10, respectivamente.

Em qualquer dos casos nos Corolários 9.2.3 e 9.2.4 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, se tomarmos $M = \{1, \dots, d\}^{\mathbb{Z}}$ então a partição \mathcal{P} em cilindros $\{[0; a] : a = 1, \dots, d\}$ 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 as dois conceitos, falaremos de gerador *unilateral* e gerador bilateral, respectivamente.

A este respeito também observamos que certos sistemas *invertíveis* admitem geradores unilaterais, ou seja, tais que a união dos iterados \mathcal{P}^n , $n \geq 1$ gera a σ -álgebra dos conjuntos mensuráveis. Por exemplo, se $f : S^1 \rightarrow S^1$ 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 comportamento só é possível para sistemas com entropia nula:

Corolário 9.2.5. *Suponha que $f : M \rightarrow M$ 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 de M a menos de medida nula. Então $h_{\mu}(f) = 0$.*

Demonstração. Combinando o Lema 9.1.12 e o Corolário 9.2.3:

$$h_{\mu}(f) = h_{\mu}(f, \mathcal{P}) = \lim_n H_{\mu}(\mathcal{P}/f^{-1}(\mathcal{P}^n)).$$

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 Teorema 9.2.1 implica que $H_{\mu}(\mathcal{P}/f^{-1}(\mathcal{P}^n))$ converge para zero quando $n \rightarrow \infty$. Segue que $h_{\mu}(f) = 0$. \square

Suponha que M é um espaço métrico, munido da sua σ -álgebra de Borel.

Corolário 9.2.6. *Seja $\mathcal{P}_1 \prec \dots \prec \mathcal{P}_n \prec \dots$ uma sequência não decrescente de partições com entropia finita tais que $\text{diam } \mathcal{P}_n(x) \rightarrow 0$ para μ -quase todo $x \in M$. Então*

$$h_{\mu}(f) = \lim_n h_{\mu}(f, \mathcal{P}_n).$$

Demonstração. Seja U um aberto qualquer de M . A hipótese garante que para cada x 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 . Em particular, o conjunto dos valores tomados por P_x é enumerável. Segue que $U = \cup_{x \in U} P_x$ também está na álgebra \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 9.2.1. \square

Exemplo 9.2.7. Seja $f : S^1 \rightarrow S^1$ um homeomorfismo e seja μ uma probabilidade invariante qualquer. Dada uma partição finita \mathcal{P} de S^1 em subintervalos, denotemos por x_1, \dots, x_m os seus pontos extremos. Para qualquer $j \geq 1$, a

partição $f^{-j}(\mathcal{P})$ está formada pelo 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

$$\{f^{-j}(x_i) : j = 0, \dots, n-1 \text{ e } i = 1, \dots, m\}.$$

Em particular, $\#\mathcal{P}^n \leq mn$. Então, usando o Lema 9.1.3,

$$h_\mu(f, \mathcal{P}) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}^n) \leq \lim_n \frac{1}{n} \#\mathcal{P}^n = \lim_n \frac{1}{n} \log kn = 0.$$

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 9.2.6.

Corolário 9.2.8. *Seja \mathcal{P} uma partição com entropia finita tal que, para μ -quase todo $x \in M$, tem-se $\text{diam } \mathcal{P}^n(x) \rightarrow 0$. Então $h_\mu(f) = h_\mu(f, \mathcal{P})$.*

Demonstração. Basta aplicar o Corolário 9.2.6 à sequência \mathcal{P}^n , lembrando que $h_\mu(f, \mathcal{P}^n) = h_\mu(f, \mathcal{P})$ para todo n . \square

Analogamente, se f é invertível e \mathcal{P} é uma partição com entropia finita tal que $\text{diam } \mathcal{P}^{\pm n}(x) \rightarrow 0$ 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 [Rok67a, §10] mostrou que se um sistema é aperiódico (isto é, o conjunto dos pontos periódicos tem medida nula) e quase todo ponto tem um número finito ou enumerável de pré-imagens, então existe algum gerador finito ou enumerável. Em particular, todo sistema invertível aperiódico admite algum gerador finito ou enumerável. Em geral, este gerador pode ter entropia infinita. Mas Rokhlin [Rok67a, §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.

9.2.2 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 $h_\nu(f) \leq h_\mu(f) + \varepsilon$ para todo ν suficientemente próximo de μ . Isso vale, em particular, para a classe de transformações que chamamos de expansivas. Estes fatos têm consequências importantes, que serão exploradas na Seção 9.2.3 e, mais tarde, nas Seções 9.6 e 10.5.

Começamos por mostrar, por meio de um exemplo, que a função entropia pode ser descontínua:

Exemplo 9.2.9. Seja $f : [0, 1] \rightarrow [0, 1]$ a transformação expansão decimal. Como vimos no Exemplo 9.1.9, a entropia de f relativamente à medida de

Lebesgue m é $h_m(f) = \log 10$. Para cada $k \geq 1$, 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 seqüência 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 seqüência $(\mu_k)_k$ 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 9.1.8 prova que $h_{\mu_k}(f) = 0$ para todo k . Portanto, a entropia não varia continuamente com a medida invariante.

Por outro lado, considere qualquer partição finita \mathcal{P} de M cujo *bordo*

$$\partial\mathcal{P} = \bigcup_{P \in \mathcal{P}} \partial P$$

satisfaz $\mu(\partial\mathcal{P}) = 0$. Pelo Teorema 2.1.2 ou, mais precisamente, pelo fato de que a topologia (2.1.5) é equivalente à topologia fraca*, 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

$$\partial\mathcal{P}^n \subset \partial\mathcal{P} \cup f^{-1}(\partial\mathcal{P}) \cup \dots \cup f^{-n+1}(\partial\mathcal{P}).$$

Assim, a função $\nu \mapsto H_\nu(\mathcal{P}^n)$ é contínua para todo n .

Proposição 9.2.10. *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. \square

Corolário 9.2.11. *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 $\mu \mapsto h_\mu(f)$ é semicontínua superiormente no ponto μ .*

Demonstração. Pela Proposição 9.2.10, 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 $\nu \in U$. Temos $h_\mu(f, \mathcal{P}) \leq h_\mu(f)$, por definição. Pelo Corolário 9.2.3, 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 $\nu \in U$. \square

Agora suponhamos que M é um espaço métrico compacto e μ é uma probabilidade boreliana em M . Nesse caso temos a seguinte versão mais especializada do corolário anterior:

Corolário 9.2.12. *Suponha que existe $\varepsilon_0 > 0$ tal que toda partição finita \mathcal{P} com $\text{diam } \mathcal{P} < \varepsilon_0$ satisfaz $\lim_n \text{diam } \mathcal{P}^n = 0$. 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 9.2.8, a propriedade $\lim_n \text{diam } \mathcal{P}^n = 0$ 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 uma partição \mathcal{P} com diâmetro menor que ε_0 e tal que $\mu(\partial \mathcal{P}) = 0$. 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; 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 . 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 das probabilidades invariantes, é um espaço compacto. \square

Quando f é invertível podemos substituir \mathcal{P}^n por $\mathcal{P}^{\pm n} = \bigvee_{j=-n}^{n-1} f^{-j}(\mathcal{P})$ no enunciado dos Corolários 9.2.11 e 9.2.12. A demonstração é análoga, usando as versões dos Corolários 9.2.3 e 9.2.8 para transformações invertíveis.

9.2.3 Transformações expansivas

Agora vamos discutir uma classe bastante ampla de transformações que satisfazem as condições do Corolário 9.2.12.

Uma transformação contínua $f : M \rightarrow M$ num espaço métrico é dita *expansiva* se existe ε_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 ε_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 9.2.13. Seja $\sigma : \Sigma \rightarrow \Sigma$ o deslocamento em $\Sigma = \{1, \dots, d\}^{\mathbb{N}}$. Consideramos em Σ a distância $d((x_n)_n, (y_n)_n) = 2^{-N}$ onde N é o menor valor de n tal que $x_n \neq y_n$. $x = (x_n)_n$ e $y = (y_n)_n$ são pontos distintos então $d(f^N(x), f^N(y)) = 2^0 = 1$. Isto prova que f é uma transformação expansiva, com $\varepsilon_0 = 1$ como constante de expansividade.

Analogamente, o deslocamento bilateral $\sigma : \Sigma \rightarrow \Sigma$ em $\Sigma = \{1, \dots, d\}^{\mathbb{Z}}$ é expansivo no sentido bilateral (mas não no sentido unilateral).

Deixamos ao cuidado do leitor verificar (Exercício 9.2.1) que a expansão decimal $f(x) = 10x - [10x]$ também é expansiva. Por outro lado, isometrias de qualquer espaço métrico nunca são expansivas.

Proposição 9.2.14. *Seja $f : M \rightarrow M$ uma transformação expansiva num espaço métrico compacto e seja $\varepsilon_0 > 0$ uma constante de expansividade. Então tem-se $\lim_n \text{diam } \mathcal{P}^n = 0$ para toda partição finita \mathcal{P} com $\text{diam } \mathcal{P} < \varepsilon_0$.*

Demonstração. É claro que a sequência $\text{diam } \mathcal{P}^n$ 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

$$d(f^j(x_n), f^j(y_n)) \leq \text{diam } \mathcal{P} < \varepsilon_0 \quad \text{para todo } 0 \leq j < n.$$

Por compacidade, existe $(n_j)_j \rightarrow \infty$ tal que $(x_{n_j})_j$ e $(y_{n_j})_j$ convergem para pontos x e y , respectivamente. Então $x \neq y$ mas $d(f^j(x), f^j(y)) \leq \text{diam } \mathcal{P} < \varepsilon_0$ para todo $j \geq 0$. Isto contradiz a hipótese de que ε_0 é constante de expansividade. Deste modo, fica provada a primeira parte da proposição. A prova da segunda parte é análoga. \square

Corolário 9.2.15. *Se $f : M \rightarrow M$ 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 .*

Se a transformação f é invertível e expansiva no sentido bilateral, podemos substituir \mathcal{P}^n por $\mathcal{P}^{\pm n}$ na Proposição 9.2.14 e a conclusão do Corolário 9.2.15 também permanece válida tal como está enunciada.

9.2.4 Exercícios

9.2.1. Mostre que a expansão decimal $f : [0, 1] \rightarrow [0, 1]$, $f(x) = 10x - [10x]$ é expansiva e exiba uma constante de expansividade.

9.2.2. Verifique que para todo $s > 0$ existe algum deslocamento de Bernoulli (σ, μ) cuja entropia é igual a s .

9.2.3. 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 $N = \min\{n \in \mathbb{N} : x_n \neq y_n\}$.

- Certifique-se de que o deslocamento $\sigma : \Sigma \rightarrow \Sigma$ não é expansivo.
- 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.
- 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.

9.2.4. Seja $f : S^1 \rightarrow S^1$ uma aplicação de recobrimento de grau $d \geq 2$ e seja μ uma probabilidade invariante por f . Mostre que $h_\mu(f) \leq \log d$.

9.2.5. Sejam \mathcal{P} e \mathcal{Q} partições finitas. Mostre que se $\mathcal{P} \prec \bigvee_{j=0}^{\infty} f^{-j}(\mathcal{Q})$ então $h_\mu(f, \mathcal{P}) \leq h_\mu(f, \mathcal{Q})$.

9.3 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 9.3.1 (Shannon-McMillan-Breiman). *Dada qualquer partição \mathcal{P} com entropia finita, o limite*

$$h_\mu(f, \mathcal{P}, x) = \lim_n -\frac{1}{n} \log \mu(\mathcal{P}^n(x)) \quad \text{existe em } \mu\text{-quase todo ponto.} \quad (9.3.1)$$

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 $\mathcal{P}^n(x) = \mathcal{P}(x) \cap f^{-1}(\mathcal{P}(f(x))) \cap \dots \cap f^{-n+1}(\mathcal{P}(f^{n-1}(x)))$, ou seja, este conjunto está formado pelos pontos cuja trajetória se mantém “próxima” da trajetória de x durante n iterados, no sentido de que as duas visitam os mesmos elementos de \mathcal{P} . O Teorema 9.3.1 afirma que a medida deste conjunto tem uma taxa exponencial de decaimento bem definida: em μ -quase todo ponto,

$$\mu(\mathcal{P}^n(x)) \approx e^{-nh(f, \mathcal{P}, x)} \quad \text{para todo } n \text{ grande.}$$

A demonstração do teorema será apresentada na Seção 9.3.1.

O teorema de Brin-Katok, que enunciamos a seguir, pertence à mesma família de resultados, mas usa uma noção distinta de proximidade.

Definição 9.3.2. Suponhamos que $f : M \rightarrow M$ 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:

$$B(x, n, \varepsilon) = \{y \in M : d(f^j(x), f^j(y)) < \varepsilon \text{ para todo } j = 0, 1, \dots, n-1\}.$$

Em outras palavras, $B(x, n, \varepsilon) = \bigcap_{j=0}^{n-1} f^{-j}(B(f^j(x), \varepsilon))$. Defina:

$$h_\mu^+(f, \varepsilon, x) = \limsup_n -\frac{1}{n} \log \mu(B(x, n, \varepsilon))$$

$$h_\mu^-(f, \varepsilon, x) = \liminf_n -\frac{1}{n} \log \mu(B(x, n, \varepsilon)).$$

Teorema 9.3.3 (Brin-Katok). *Seja μ uma medida invariante por f . Os limites*

$$\lim_{\varepsilon \rightarrow 0} h_{\mu}^{+}(f, \varepsilon, x) \quad e \quad \lim_{\varepsilon \rightarrow 0} h_{\mu}^{-}(f, \varepsilon, x)$$

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 9.3.4 (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 (lembre do Lema 6.3.6). Então,*

$$E_g^j(B(x, \varepsilon)) = B(E_g(x), \varepsilon)$$

para todo $g \in G$, $x \in G$ e $\varepsilon > 0$. 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.

9.3.1 Prova do teorema de Shannon-McMillan-Breiman

Considere a sequência de funções $\varphi_n : M \rightarrow \mathbb{R}$ definida por

$$\varphi_n(x) = -\log \frac{\mu(\mathcal{P}^n(x))}{\mu(\mathcal{P}^{n-1}(f(x)))}.$$

Pela propriedade de cancelamento telescópico,

$$-\frac{1}{n} \log \mu(\mathcal{P}^n(x)) = -\frac{1}{n} \log \mu(\mathcal{P}(f^{n-1}(x))) + \frac{1}{n} \sum_{j=0}^{n-2} \varphi_{n-j}(f^j(x)) \quad (9.3.2)$$

para todo n e todo x .

Lema 9.3.5. *A sequência $(-1/n) \log \mu(\mathcal{P}(f^{n-1}(x)))$ converge para zero em μ -quase todo ponto e em $L^1(\mu)$.*

Demonstração. Comece por observar que a função $x \mapsto -\log \mu(\mathcal{P}(x))$ é integrável:

$$\int |\log \mu(\mathcal{P}(x))| d\mu(x) = \int -\log \mu(\mathcal{P}(x)) d\mu(x) = H_{\mu}(\mathcal{P}) < \infty.$$

Usando o Lema 3.2.5, segue que $-1/(n-1) \log \mu(\mathcal{P}(f^{n-1}(x)))$ converge para zero em μ -quase todo ponto. Além disso, é claro que esta conclusão não é afetada quando substituímos $n-1$ por n no denominador. Isto prova a convergência em μ -quase todo ponto no enunciado do lema. Em seguida, usando o fato de que a medida μ é invariante,

$$\| -\frac{1}{n} \log \mu(\mathcal{P}(f^{n-1}(x))) \|_1 = \frac{1}{n} \int -\log \mu(\mathcal{P}(f^{n-1}(x))) d\mu(x) = \frac{1}{n} H_\mu(\mathcal{P}).$$

converge para zero quando $n \rightarrow \infty$, uma vez que $H_\mu(\mathcal{P}) < \infty$. Isto prova a convergência em $L^1(\mu)$. \square

A seguir, vamos mostrar que o último termo de (9.3.2) também converge em μ -quase todo ponto e em $L^1(\mu)$.

Lema 9.3.6. *O limite $\varphi(x) = \lim_n \varphi_n(x)$ existe em μ -quase todo ponto.*

Demonstração. Para cada $n > 1$, denote por \mathcal{Q}_n a partição de M definida por

$$\mathcal{Q}_n(x) = f^{-1}(\mathcal{P}^{n-1}(f(x))) = f^{-1}(\mathcal{P}(f(x))) \cap \dots \cap f^{-n+1}(\mathcal{P}(f^{n-1}(x))).$$

Note que $\mu(\mathcal{P}^{n-1}(f(x))) = \mu(\mathcal{Q}_n(x))$ e que $\mathcal{P}^n(x) = \mathcal{P}(x) \cap \mathcal{Q}_n(x)$. Portanto,

$$\frac{\mu(\mathcal{P}^n(x))}{\mu(\mathcal{P}^{n-1}(f(x)))} = \frac{\mu(\mathcal{P}(x) \cap \mathcal{Q}_n(x))}{\mu(\mathcal{Q}_n(x))}. \quad (9.3.3)$$

Para cada $P \in \mathcal{P}$ e cada $n > 1$, considere a esperança condicional (lembre da Seção 5.2.1)

$$e_n(\mathcal{X}_P, x) = \frac{1}{\mu(\mathcal{Q}_n(x))} \int_{\mathcal{Q}_n(x)} \mathcal{X}_P d\mu = \frac{\mu(P \cap \mathcal{Q}_n(x))}{\mu(\mathcal{Q}_n(x))}.$$

Comparando com (9.3.3) vemos que

$$e_n(\mathcal{X}_P, x) = \frac{\mu(\mathcal{P}^n(x))}{\mu(\mathcal{P}^{n-1}(f(x)))} \quad \text{para todo } x \in P.$$

Pelo Lema 5.2.1, o limite $e(\mathcal{X}_P, x) = \lim_n e_n(\mathcal{X}_P, x)$ existe para μ -quase todo $x \in M$ e, em particular, para μ -quase todo $x \in P$. Como $P \in \mathcal{P}$ é arbitrário, isto prova que

$$\lim_n \frac{\mu(\mathcal{P}^n(x))}{\mu(\mathcal{P}^{n-1}(f(x)))}$$

existe para μ -quase todo ponto. Tomando logaritmos, obtemos que $\lim_n \varphi_n(x)$ existe para μ -quase todo ponto, conforme afirmado. \square

Lema 9.3.7. *A função $\Phi = \sup_n \varphi_n$ é integrável.*

Demonstração. Tal como no lema anterior, vamos usar as partições \mathcal{Q}_n definidas por $\mathcal{Q}_n(x) = f^{-1}(\mathcal{P}^{n-1}(f(x)))$. Fixe qualquer $P \in \mathcal{P}$. Dados $x \in P$ e $t > 0$, é claro que $\Phi(x) > t$ se, e somente se, $\varphi_n(x) > t$ para algum n . Além disso,

$$\varphi_n(x) > t \Leftrightarrow \mu(P \cap \mathcal{Q}_n(x)) < e^{-t} \mu(\mathcal{Q}_n(x))$$

e, nesse caso, $\varphi_n(y) > t$ para todo $y \in P \cap \mathcal{Q}_n(x)$. Portanto, podemos escrever o conjunto $\{x \in P : \Phi(x) > t\}$ como uma união disjunta $\cup_j (P \cap Q_j)$, onde cada Q_j pertence a alguma partição $\mathcal{Q}_{n(j)}$ e

$$\mu(P \cap Q_j) < e^{-t} \mu(Q_j) \quad \text{para todo } j.$$

Consequentemente, para todo $t > 0$ e todo $P \in \mathcal{P}$,

$$\mu(\{x \in P : \Phi(x) > t\}) = \sum_j \mu(P \cap Q_j) < e^{-t} \sum_j \mu(Q_j) \leq e^{-t}. \quad (9.3.4)$$

Então (veja o Exercício 9.3.1):

$$\begin{aligned} \int \Phi d\mu &= \sum_{P \in \mathcal{P}} \int_P \Phi d\mu = \sum_{P \in \mathcal{P}} \int_0^\infty \mu(\{x \in P : \Phi(x) > t\}) dt \\ &\leq \sum_{P \in \mathcal{P}} \int_0^\infty \min\{e^{-t}, \mu(P)\} dt. \end{aligned}$$

A última integral pode ser reescrita do seguinte modo:

$$\int_0^{-\log \mu(P)} \mu(P) dt + \int_{-\log \mu(P)}^\infty e^{-t} dt = -\mu(P) \log \mu(P) + \mu(P).$$

Combinando estas duas relações:

$$\int \Phi d\mu \leq \sum_{P \in \mathcal{P}} -\mu(P) \log \mu(P) + \mu(P) = H_\mu(\mathcal{P}) + 1 < \infty.$$

Isto prova o lema, uma vez que Φ é não negativa. \square

Lema 9.3.8. *A função φ é integrável e $(\varphi_n)_n$ converge para φ em $L^1(\mu)$.*

Demonstração. Vimos no Lema 9.3.6 que $(\varphi_n)_n$ converge para φ em μ -quase todo ponto. Como $0 \leq \varphi_n \leq \Phi$ para todo n , também temos $0 \leq \varphi \leq \Phi$. Em particular, φ é integrável. Além disso, $|\varphi - \varphi_n| \leq \Phi$ para todo n e, portanto, podemos usar o teorema da convergência dominada (Teorema A.2.11) para concluir que

$$\lim_n \int |\varphi - \varphi_n| d\mu = \int \lim_n |\varphi - \varphi_n| d\mu = 0.$$

Isto prova a convergência em $L^1(\mu)$. \square

Lema 9.3.9. Em μ -quase todo ponto e em $L^1(\mu)$,

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-2} \varphi_{n-j}(f^j(x)) = \lim_n \frac{1}{n} \sum_{j=0}^{n-2} \varphi(f^j(x)).$$

Demonstração. Pelo teorema ergódico de Birkhoff (Teorema 3.2.3), o limite do lado direito existe em μ -quase todo ponto e em $L^1(\mu)$, de fato, ele é igual à média temporal da função φ . Portanto, basta mostrar que a diferença

$$\frac{1}{n} \sum_{j=0}^{n-2} (\varphi_{n-j} - \varphi) \circ f^j \quad (9.3.5)$$

converge para zero em μ -quase todo ponto e em $L^1(\mu)$. Como a medida μ é invariante, $\|(\varphi_{n-j} - \varphi) \circ f^j\|_1 = \|\varphi_{n-j} - \varphi\|_1$ para todo j . Logo,

$$\left\| \frac{1}{n} \sum_{j=0}^{n-2} (\varphi_{n-j} - \varphi) \circ f^j \right\|_1 \leq \frac{1}{n} \sum_{j=0}^{n-2} \|\varphi_{n-j} - \varphi\|_1.$$

Pelo Lema 9.3.8 a sequência do lado direito converge para zero. Isto implica que (9.3.5) converge para zero em $L^1(\mu)$. Resta provar a convergência em μ -quase todo ponto.

Para cada $k \geq 2$ fixado, considere $\Phi_k = \sup_{i>k} |\varphi_i - \varphi|$. Note que $\Phi_k \leq \Phi$ e, portanto, $\Phi_k \in L^1(\mu)$. Além disso:

$$\begin{aligned} \frac{1}{n} \sum_{j=0}^{n-2} |\varphi_{n-j} - \varphi| \circ f^j &= \frac{1}{n} \sum_{j=0}^{n-k-1} |\varphi_{n-j} - \varphi| \circ f^j + \frac{1}{n} \sum_{j=n-k}^{n-2} |\varphi_{n-j} - \varphi| \circ f^j \\ &\leq \frac{1}{n} \sum_{j=0}^{n-k-1} \Phi_k \circ f^j + \frac{1}{n} \sum_{j=n-k}^{n-2} \Phi \circ f^j. \end{aligned}$$

Pelo teorema ergódico de Birkhoff, o primeiro termo do lado direito converge para a média temporal $\tilde{\Phi}_k$ em μ -quase todo ponto. Pelo Lema 3.2.5, o último termo converge para zero em μ -quase todo ponto: o lema implica que $n^{-1}\Phi \circ f^{n-i}$ converge para zero para qualquer i fixado. Logo,

$$\limsup_n \frac{1}{n} \sum_{j=0}^{n-2} |\varphi_{n-j} - \varphi|(f^j(x)) \leq \tilde{\Phi}_k(x) \quad \text{em } \mu\text{-quase todo ponto.} \quad (9.3.6)$$

Afirmamos que $\lim_k \tilde{\Phi}_k(x) = 0$ em μ -quase todo ponto. De fato, a sequência $(\Phi_k)_k$ é não-crescente e, pelo Lema 9.3.6, ela converge para zero em μ -quase todo ponto. Pelo teorema da convergência monótona (Teorema A.2.9), segue que $\int \Phi_k d\mu \rightarrow 0$ quando $k \rightarrow \infty$. Outra consequência é que $(\tilde{\Phi}_k)_k$ é não-crescente. Logo, usando o teorema da convergência monótona juntamente com o teorema ergódico de Birkhoff:

$$\int \lim_k \tilde{\Phi}_k d\mu = \lim_k \int \tilde{\Phi}_k d\mu = \lim_k \int \Phi_k d\mu = 0.$$

Como $\tilde{\Phi}_k$ é não negativa, segue que $\lim_k \tilde{\Phi}_k = 0$ em μ -quase todo ponto, tal como foi afirmado. Portanto, (9.3.6) implica que

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-2} |\varphi_{n-j} - \varphi| \circ f^j = 0$$

em μ -quase todo ponto. Isto completa a demonstração do lema. \square

Segue da igualdade (9.3.2) e dos Lemas 9.3.5 e 9.3.9 que

$$h_\mu(f, \mathcal{P}, x) = \lim_n -\frac{1}{n} \log \mu(\mathcal{P}^n(x))$$

existe em μ -quase todo ponto e em $L^1(\mu)$; de fato, ele coincide com a média temporal $\tilde{\varphi}(x)$ da função φ . Então, em particular,

$$\begin{aligned} \int h_\mu(f, \mathcal{P}, x) d\mu(x) &= \lim_n \frac{1}{n} \int -\log \mu(\mathcal{P}^n(x)) d\mu(x) \\ &= \lim_n \frac{1}{n} H_\mu(\mathcal{P}^n) = h_\mu(f, \mathcal{P}). \end{aligned}$$

Além disso, se (f, μ) é ergódico então $h(f, \mathcal{P}, x) = \tilde{\varphi}(x)$ é constante em μ -quase todo ponto. Ou seja, nesse caso $h_\mu(f, \mathcal{P}, x) = h_\mu(f, \mathcal{P})$ para μ -quase todo ponto. Isto encerra a demonstração do Teorema 9.3.1.

9.3.2 Exercícios

9.3.1. Verifique que, para qualquer função integrável $\varphi : M \rightarrow (0, \infty)$,

$$\int \varphi d\mu = \int_0^\infty \mu(\{x \in M : \varphi(x) > t\}) dt.$$

9.3.2. Use o Teorema 9.3.1 para calcular a entropia de um deslocamento de Bernoulli em $\Sigma = \{1, \dots, d\}^{\mathbb{N}}$.

9.3.3. Mostre que a função $h_\mu(f, x)$ no Teorema 9.3.3 é f -invariante. Conclua que se (f, μ) é ergódico, então $h_\mu(f) = h_\mu(f, x)$ para μ -quase todo ponto x .

9.3.4. 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

$$e^{-n(h_\mu(f, \mathcal{P}) + \varepsilon)} < \mu(B) < e^{-n(h_\mu(f, \mathcal{P}) - \varepsilon)} \quad \text{para todo } B \in \mathcal{B}_n.$$

9.4 Exemplos

Vamos agora ilustrar os resultados anteriores por meio de alguns exemplos.

9.4.1 Deslocamentos de Markov

Seja $\Sigma = \{1, \dots, d\}^{\mathbb{N}}$ e seja $\sigma : \Sigma \rightarrow \Sigma$ 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 9.4.1. $h_\mu(\sigma) = \sum_{a=1}^d p_a \sum_{b=1}^d -P_{a,b} \log P_{a,b}$.

Demonstração. Considere a partição \mathcal{P} de Σ em cilindros $[0; a]$, $a = 1, \dots, d$. Para cada n , o iterado \mathcal{P}^n é a partição em cilindros $[0; a_1, \dots, a_n]$ de comprimento n . Lembrando que $\mu([0; a_1, \dots, a_n]) = p_{a_1} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n}$, vemos que

$$\begin{aligned} H_\mu(\mathcal{P}^n) &= \sum_{a_1, \dots, a_n} -p_{a_1} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n} \log (p_{a_1} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n}) \\ &= \sum_{a_1} -p_{a_1} \log p_{a_1} \sum_{a_2, \dots, a_n} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n} \\ &\quad + \sum_{j=1}^n \sum_{a_j, a_{j+1}} -\log P_{a_j, a_{j+1}} \sum p_{a_1} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n} \end{aligned} \quad (9.4.1)$$

onde a última soma é sobre todos os valores de $a_1, \dots, a_{j-1}, a_{j+2}, \dots, a_n$. Por um lado,

$$\sum_{a_2, \dots, a_n} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n} = \sum_{a_n} P_{a_1, a_n}^n = 1$$

uma vez que P^n é uma matriz estocástica. De modo semelhante, lembrando também que $P^*p = p$,

$$\begin{aligned} \sum p_{a_1} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n} &= \sum_{a_1, a_n} p_{a_1} P_{a_1, a_j}^j P_{a_j, a_{j+1}} P_{a_{j+1}, a_n}^{n-j-1} \\ &= \sum_{a_1} p_{a_1} P_{a_1, a_j}^j P_{a_j, a_{j+1}} = p_{a_j} P_{a_j, a_{j+1}} \end{aligned}$$

Lembrando que $P^*p = p = p$, vemos que a última expressão é igual a p_{a_j} . Substituindo estas observações em (9.4.1), obtemos que

$$\begin{aligned} H_\mu(\mathcal{P}^n) &= \sum_{a_1} -p_{a_1} \log p_{a_1} + \sum_{j=1}^n \sum_{a_j, a_{j+1}} -p_{a_j} P_{a_j, a_{j+1}} \log P_{a_j, a_{j+1}} \\ &= -1 + n \sum_{a,b} -p_a P_{a,b} \log P_{a,b} \end{aligned}$$

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 $[0; a_1, \dots, a_n]$ gera a σ -álgebra de Σ , segue do Corolário 9.2.3 que $h_\mu(\sigma) = h_\mu(\sigma, \mathcal{P})$. Isto completa a prova do teorema. \square

Esta conclusão permanece válida no caso de deslocamentos de Markov bilaterais, ou seja, em $\Sigma = \{1, \dots, d\}^{\mathbb{Z}}$. O argumento é análogo, usando o Corolário 9.2.4.

9.4.2 Transformação de Gauss

Vamos calcular a entropia da transformação de Gauss $G(x) = (1/x) - [1/x]$ relativamente à probabilidade invariante

$$\mu(E) = \frac{1}{\log 2} \int_E \frac{dx}{1+x} \quad (9.4.2)$$

estudada nas Seções 1.3.2 e 4.2.4. O método que vamos apresentar se estende a uma classe bastante ampla de sistemas, incluindo as transformações expansoras do intervalo, que serão definidas e discutidas no Exemplo 11.1.16.

Seja \mathcal{P} a partição nos intervalos $(1/(m+1), 1/m)$ para $m \geq 1$. Como antes, denotamos $\mathcal{P}^n = \bigvee_{j=0}^{n-1} G^{-j}(\mathcal{P})$. 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) $\text{diam } P_n \rightarrow 0$ quando $n \rightarrow \infty$.
- (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 4.2.12, obtemos que

$$\text{diam } P_n \leq \sup_{x \in P_n} \frac{1}{|(G^n)'(x)|} \leq 2^{-[n/2]}$$

para todo $n \geq 1$ e todo $P_n \in \mathcal{P}^n$. Isto implica (B). A propriedade (C) está dada pelo Lema 4.2.13. Finalmente, (D) segue diretamente de (9.4.2).

Proposição 9.4.2. $h_\mu(G) = \int \log |G'| d\mu$.

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,

$$-\log(Cc_1) \geq H_\mu(\mathcal{P}^n) - \int \log |(G^n)'| d\mu \geq \log(C/c_2) \quad (9.4.3)$$

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 (9.4.3) por n e passando ao limite quando $n \rightarrow \infty$,

$$h_\mu(f, \mathcal{P}) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}^n) = \int \log |G'| d\mu.$$

Agora a propriedade (C) garante que podemos aplicar o Corolário 9.2.8 para concluir que

$$h_\mu(G) = h_\mu(G, \mathcal{P}) = \int \log |G'| d\mu.$$

Isto completa a demonstração da proposição. \square

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 $\sum_{j=1}^{\infty} 1/j^2 = \pi^2/6$)

$$h_\mu(G) = \int \log |G'| d\mu = \int_0^1 \frac{-2 \log x dx}{(1+x) \log 2} = \frac{\pi^2}{6 \log 2} \approx 5,46 \dots$$

Então, lembrando que (G, μ) é ergódico (Seção 4.2.4), segue do teorema de Shannon-McMillan-Breiman (Teorema 9.3.1) 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

$$\text{diam } \mathcal{P}^n(x) \approx \exp -\frac{\pi^2 n}{6 \log 2}$$

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 .

9.4.3 Endomorfismos lineares do toro

Dado um número real $x > 0$, denotamos $\log^+ x = \max\{\log x, 0\}$. Nesta seção provamos o seguinte resultado:

Proposição 9.4.3. *Seja $f_A : \mathbb{T}^d \rightarrow \mathbb{T}^d$ 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 $\lambda_1, \dots, \lambda_d$ são os autovalores de A , contados com multiplicidade.

Inicialmente, suponhamos que a matriz A é diagonalizável. Seja v_1, \dots, 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, \dots, 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 $i = 1, \dots, d$. Além disso, para cada $n \geq 1$, consideramos

$$D(x, n, \varepsilon) = \{y \in \mathbb{T}^d : f_A^j(y) \in D(f_A^j(x), \varepsilon) \text{ para todo } j = 0, \dots, n-1\}.$$

Observe que $f_A^j(y) = f_A^j(x) + \sum_{i=1}^d t_i \lambda_i^n v_i$ para todo $n \geq 1$. Portanto,

$$D(x, n, \varepsilon) = \left\{x + \sum_{i=1}^d t_i v_i : |\lambda_i^n t_i| < \varepsilon \text{ para } i \leq u \text{ e } |t_i| < \varepsilon \text{ para } i > u\right\}.$$

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 $C_2 > 1$ que depende apenas de A , tal que

$$B(x, C_2^{-1} \varepsilon) \subset D(x, \varepsilon) \subset B(x, C_2 \varepsilon)$$

para todo $x \in \mathbb{T}^d$ e todo $\varepsilon > 0$ pequeno. Então, $B(x, n, \varepsilon/C_2) \subset D(x, n, \varepsilon) \subset B(x, n, C_2 \varepsilon)$ 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,

$$h_\mu^+(f, \varepsilon, x) = h_\mu^-(f, \varepsilon, x) = \lim_n \frac{1}{n} \log \mu(B(x, n, \varepsilon)) = \sum_{i=1}^u \log |\lambda_i|$$

para todo $x \in \mathbb{T}$ e todo $\varepsilon > 0$ pequeno. Logo, usando o teorema de Brin-Katok (Teorema 9.3.3)

$$h_\mu(f) = h_\mu(f, x) = \sum_{i=1}^u \log |\lambda_i|$$

para μ -quase todo ponto x . Isto prova a Proposição 9.4.3 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 9.4.2).

9.4.4 Exercícios

9.4.1. Mostre que toda rotação $R_\theta : \mathbb{T}^d \rightarrow \mathbb{T}^d$ tem entropia nula relativamente à medida de Haar no toro \mathbb{T}^d . [Observação: Isto é um caso particular do Exemplo 9.3.4 mas para o enunciado presente não precisamos usar o teorema de Brin-Katok.]

9.4.2. Complete a demonstração da Proposição 9.4.3.

9.4.3. Seja $f : M \rightarrow M$ uma transformação mensurável e seja μ uma probabilidade invariante ergódica em M . Seja $B \subset M$ um conjunto mensurável com $\mu(B) > 0$, seja $g : B \rightarrow B$ a transformação de primeiro retorno a B e seja ν a restrição normalizada de μ ao conjunto B (lembre das Seção 1.4.1). Mostre que $h_\mu(f) = \nu(B)h_\nu(g)$.

9.4.4. Seja $f : M \rightarrow M$ uma transformação preservando medida num espaço de Lebesgue (M, μ) . Seja $\hat{f} : \hat{M} \rightarrow \hat{M}$ a extensão natural de f e seja $\hat{\mu}$ o levantamento de μ (Exercício 8.5.7). Mostre que $h_\mu(f) = h_{\hat{\mu}}(\hat{f})$.

9.5 Entropia e equivalência

A entropia foi introduzida em Teoria Ergódica com o objetivo principal de distinguir sistemas que não são ergodicamente equivalentes, especialmente no caso de sistemas que são espectralmente equivalentes e que, portanto, não podem ser distinguidos por meio de invariantes espectrais. É fácil que a entropia é, de fato, um invariante de equivalência ergódica:

Proposição 9.5.1. *Sejam $f : M \rightarrow M$ e $g : N \rightarrow N$ transformações preservando probabilidades μ em M e ν em N . Se (f, μ) é ergodicamente equivalente a (g, ν) , então $h_\mu(f) = h_\nu(g)$.*

Demonstração. Seja $\phi : M \rightarrow N$ uma equivalência ergódica entre os dois sistemas. Isto significa que $\phi_*\mu = \nu$ e existem conjuntos $X \subset M$ e $Y \subset N$ com medida total nos respectivos espaços, tais que ϕ é uma bijeção mensurável de X em Y , com inversa mensurável. Além disso, como observamos na Seção 8.1, os conjuntos X e Y podem ser escolhidos invariantes. Seja \mathcal{P} uma partição de M com entropia finita para μ . A sua restrição a X_1 é uma partição de (X, μ) .

A respectiva imagem $\mathcal{Q} = \phi(\mathcal{P})$ é uma partição de (Y, ν) que, naturalmente, também podemos ver como uma partição de (N, ν) . Note que

$$H_\nu(\mathcal{Q}) = \sum_{Q \in \mathcal{Q}} -\nu(Q) \log \nu(Q) = \sum_{P \in \mathcal{P}} -\mu(P) \log \mu(P) = H_\mu(\mathcal{P}).$$

Como $\mathcal{Q}^n = \vee_{j=0}^{n-1} g^{-j}(\mathcal{Q}) = \phi(\vee_{j=0}^{n-1} f^{-j}(\mathcal{P})) = \phi(\mathcal{P}^n)$ para todo n , também segue que

$$h_\nu(g, \mathcal{Q}) = \lim_n \frac{1}{n} H_\nu(\mathcal{Q}^n) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}^n) = h_\mu(f, \mathcal{P}).$$

Tomando o supremo sobre todos os \mathcal{P} , concluímos que $h_\nu(g) \geq h_\mu(\mathcal{P}^n)$. A desigualdade recíproca é inteiramente análoga. \square

Usando esta observação, Kolmogorov e Sinai concluíram que nem todos os deslocamentos de Bernoulli bilaterais são ergodicamente equivalentes apesar de que, como vimos no Corolário 8.4.12, todos são espectralmente equivalentes. Isto também mostra que a equivalência espectral é estritamente mais fraca que a equivalência ergódica. De fato, conforme observamos no Exemplo 9.1.10, para todo $x > 0$ existe algum deslocamento de Bernoulli bilateral (σ, μ) tal que $h_\mu(\sigma) = x$. Portanto, uma única classe de equivalência espectral contém todo um contínuo de classes de equivalência ergódica.

9.5.1 Automorfismos de Bernoulli

A recíproca da Proposição 9.5.1 é falsa, em geral. Por exemplo, vimos no Exemplo 9.2.7 (e no Corolário 9.2.5) que todas as rotações do círculo têm entropia nula. Mas é claro que uma rotação irracional nunca é ergodicamente equivalente a uma rotação racional, uma vez que a primeira é ergódica e a segunda não é. Aliás, o Corolário 8.3.6 mostra que as rotações irracionais também não são ergodicamente equivalentes entre si, em geral. O caso das rotações racionais será tratado no Exercício 8.3.3.

No entanto, um resultado notável devido a Donald Ornstein [Orn70] afirma que a entropia é um invariante completo para os deslocamentos de Bernoulli bilaterais:

Teorema 9.5.2 (Ornstein). *Dois deslocamentos de Bernoulli bilaterais em espaços de Lebesgue são ergodicamente equivalentes se, e somente se, as suas entropias são iguais.*

Chamaremos *automorfismo de Bernoulli* todo sistema ergodicamente equivalente a um deslocamento de Bernoulli bilateral. Na sequência encontraremos diversos exemplos de sistemas nesta classe. O teorema de Ornstein pode ser reformulado na seguinte forma: dois automorfismos de Bernoulli em espaços de Lebesgue são ergodicamente equivalentes se, e somente se, as suas entropias são iguais.

Vale a pena observar que o teorema de Ornstein não se estende para deslocamentos de Bernoulli unilaterais. De fato, o Exercício 8.1.2 mostra que no caso não invertível existem outros invariantes de equivalência, tais como o grau da transformação (o número de pré-imagens).

William Parry e Peter Walters [PW72b, PW72a, Wal73] mostraram, entre outros resultados, que dois deslocamentos de Bernoulli unilaterais correspondentes a vetores de probabilidade $p = (p_1, \dots, p_k)$ e $q = (q_1, \dots, q_l)$ são ergodicamente equivalentes se, e somente se, $k = l$ e o vetor p é uma permutação do vetor q . A demonstração deste fato será proposta ao leitor no Exercício 9.7.8, depois que tenhamos introduzido a noção de jacobiano.

9.5.2 Sistemas com entropia nula

Nesta seção vamos estudar algumas propriedades de sistemas com entropia igual a zero. O principal resultado é que tais sistemas são invertíveis em quase todo ponto, se o ambiente é um espaço de Lebesgue. Vale a pena comparar este enunciado com o Corolário 9.2.5. Ao final da seção discutiremos brevemente os tipos espectrais dos sistemas com entropia nula.

No que segue (M, \mathcal{B}, μ) é um espaço de probabilidade e $f : M \rightarrow M$ é uma transformação mensurável preservando a medida μ . Ao final da seção suporemos que (M, \mathcal{B}, μ) é um espaço de Lebesgue.

Lema 9.5.3. *Para todo $\varepsilon > 0$ existe $\delta > 0$ tal que se \mathcal{P} e \mathcal{Q} são partições com entropia finita e $H_\mu(\mathcal{P}/\mathcal{Q}) < \delta$, então para todo $P \in \mathcal{P}$ existe uma união P' de elementos de \mathcal{Q} satisfazendo $\mu(P \Delta P') < \varepsilon$.*

Demonstração. Seja $s = 1 - \varepsilon/2$ e $\delta = -\varepsilon \log s/2$. Para cada $P \in \mathcal{P}$ considere

$$\mathcal{S} = \{Q \in \mathcal{Q} : \mu(P \cap Q) \geq s\mu(Q)\}.$$

Seja P' a união dos elementos de \mathcal{S} . Por um lado,

$$\mu(P' \setminus P) = \sum_{Q \in \mathcal{S}} \mu(Q \setminus P) \leq \sum_{Q \in \mathcal{S}} (1 - s)\mu(Q) \leq \frac{\varepsilon}{2}. \quad (9.5.1)$$

Por outro lado,

$$\begin{aligned} H_\mu(\mathcal{P}/\mathcal{Q}) &= \sum_{R \in \mathcal{P}} \sum_{Q \in \mathcal{Q}} -\mu(R \cap Q) \log \frac{\mu(R \cap Q)}{\mu(Q)} \\ &\geq \sum_{Q \notin \mathcal{S}} -\mu(P \cap Q) \log s \geq -\mu(P \setminus P') \log s. \end{aligned}$$

Isto implica que

$$\mu(P \setminus P') \leq \frac{H_\mu(\mathcal{P}/\mathcal{Q})}{-\log s} < \frac{\delta}{-\log s} = \frac{\varepsilon}{2}. \quad (9.5.2)$$

Juntando (9.5.1) e (9.5.2) obtemos a conclusão do lema. \square

O próximo lema significa que a taxa média $h_\mu(f, \mathcal{P})$ de informação (relativamente à partição \mathcal{P}) gerada pelo sistema a cada iteração é nula se, e somente se, *o futuro determina o presente*, no sentido de que a informação relativa ao iterado zero pode ser deduzida do conjunto das informações relativas aos iterados futuros.

Lema 9.5.4. *Seja \mathcal{P} uma partição finita. Então $h_\mu(f, \mathcal{P}) = 0$ se, e somente se, $\mathcal{P} \prec \bigvee_{j=1}^{\infty} f^{-j}(\mathcal{P})$.*

Demonstração. Suponha que $h_\mu(f, \mathcal{P}) = 0$. Usando o Lema 9.1.12, vemos que $H_\mu(\mathcal{P}/\bigvee_{j=1}^n f^{-j}(\mathcal{P}))$ converge para zero quando $n \rightarrow \infty$. Então, pelo Lema 9.5.3, para cada $k \geq 1$ existe $n_k \geq 1$ e para todo $P \in \mathcal{P}$ existe uma união P_k de elementos de $\bigvee_{j=1}^{n_k} f^{-j}(\mathcal{P})$ tal que $\mu(P \Delta P_k) < 2^{-k}$. É claro que cada P_k é uma união de elementos de $\bigvee_{j=1}^{\infty} f^{-j}(\mathcal{P})$ e, portanto, o mesmo vale para cada $\bigcup_{k=n}^{\infty} P_k$ e também para $P_* = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} P_k$. Além disso,

$$\mu(P \setminus \bigcup_{k=n}^{\infty} P_k) = 0 \quad \text{e} \quad \mu(\bigcup_{k=n}^{\infty} P_k \setminus P) \leq 2^{-n}$$

para todo n e, conseqüentemente, $\mu(P \Delta P_*) = 0$. Isto mostra que todo elemento de \mathcal{P} coincide, a menos de medida nula, com uma união de elementos de $\bigvee_{j=1}^{\infty} f^{-j}(\mathcal{P})$, tal como afirmado na parte ‘somente se’ do enunciado.

Agora suponha que $\mathcal{P} \prec \bigvee_{j=1}^{\infty} f^{-j}(\mathcal{P})$. Sejam P_1, \dots, P_ℓ os elementos de \mathcal{P} (podemos supor que $\mu(P_i) > 0$ para todo i , claro). Dado qualquer $\varepsilon > 0$, tomemos $\delta > 0$ como no Lema 9.1.6. Para cada $n \geq 1$ e cada $i = 1, \dots, \ell$, seja $Q_{i,n}$ a união dos elementos de $\bigvee_{j=1}^n f^{-j}(\mathcal{P})$ que intersectam P_i . Cada $(Q_{i,n})_n$ é uma seqüência não crescente; por hipótese, a sua interseção coincide com P_i a menos de medida nula. Então,

$$\sum_{i=1}^d \mu(Q_{i,n} \setminus P_i) < \frac{\delta}{2} \quad \text{para todo } n \text{ suficientemente grande.} \quad (9.5.3)$$

Defina $R_{1,n} = Q_{1,n}$ e $R_{i,n} = Q_{i,n} \setminus (Q_{1,n} \cup \dots \cup Q_{i-1,n})$ para $i = 2, \dots, \ell$. É claro da construção que $\mathcal{R}_n = \{R_{1,n}, \dots, R_{\ell,n}\}$ é uma partição de M menos fina do que $\bigvee_{j=1}^n f^{-j}(\mathcal{P})$. Afirmamos que $\mu(P_i \Delta R_{i,n}) < \delta$ para todo i e todo n suficientemente grande. De fato, é claro que $\mu(R_{i,n} \setminus P_i) \leq \mu(Q_{i,n} \setminus P_i) < \delta/2$ para todo i . Por outro lado,

$$P_i \setminus R_{i,n} \subset P_i \bigcap \bigcup_{j < i} Q_{j,n} \subset \bigcup_{j < i} Q_{j,n} \setminus P_j.$$

Pela relação (9.5.3), segue que $\mu(P_i \setminus R_{i,n}) < \delta/2$. Isto prova a nossa afirmação a qual, juntamente com os Lemas 9.1.5 e 9.1.6, implica que

$$H_\mu(\mathcal{P}/\bigvee_{j=1}^n f^{-j}(\mathcal{P})) \leq H_\mu(\mathcal{P}/\mathcal{R}_n) < \varepsilon$$

se n é suficientemente grande. Agora, como $\varepsilon > 0$ é arbitrário, o Lema 9.1.12 dá que $h_\mu(f, \mathcal{P}) = 0$. \square

Como consequência, obtemos que todo sistema com entropia nula é invertível em quase todo ponto:

Proposição 9.5.5. *Seja (M, \mathcal{B}, μ) um espaço de Lebesgue e seja $f : M \rightarrow M$ uma transformação mensurável preservando a medida μ . Se $h_\mu(f) = 0$ então (f, μ) é invertível: existe uma transformação mensurável $g : M \rightarrow M$ que preserva a medida μ e satisfaz $f \circ g = g \circ f = \text{id}$ em μ -quase todo ponto.*

Demonstração. Considere o homomorfismo $\tilde{f} : \tilde{\mathcal{B}} \rightarrow \tilde{\mathcal{B}}$ induzido por f na álgebra de medida de \mathcal{B} (estas noções foram introduzidas na Seção 8.5). Lembre que \tilde{f} é sempre injetivo (Exercício 8.5.2). Dado qualquer $B \in \mathcal{B}$, considere a partição $\mathcal{P} = \{B, B^c\}$. A hipótese $h_\mu(f) = 0$ implica que $h_\mu(f, \mathcal{P}) = 0$ e então, pelo Lema 9.5.4, $\mathcal{P} \prec \bigvee_{j=1}^{\infty} f^{-j}(\mathcal{P})$. Isto implica que $\mathcal{P} \subset f^{-1}(\mathcal{B})$, já que $f^{-j}(\mathcal{P}) \subset f^{-1}(\mathcal{B})$ para todo $j \geq 1$. Fazendo variar B , concluímos que $\mathcal{B} \subset f^{-1}(\mathcal{B})$. Em outras palavras, o homomorfismo \tilde{f} é sobrejetivo. Então \tilde{f} é um isomorfismo de álgebras de medida. Então, pela Proposição 8.5.6, existe alguma aplicação mensurável $g : M \rightarrow M$, preservando a medida μ , tal que o respectivo homomorfismo de álgebra de medida $\tilde{g} : \tilde{\mathcal{B}} \rightarrow \tilde{\mathcal{B}}$ é o inverso de \tilde{f} . Em outras palavras, $\tilde{f} \circ \tilde{g} = \tilde{g} \circ \tilde{f} = \text{id}$. Então (Exercício 8.5.2) $f \circ g = g \circ f = \text{id}$, tal como afirmamos. \square

Estes argumentos também provam o seguinte fato que será útil a seguir:

Corolário 9.5.6. *Nas condições da Proposição 9.5.5, toda σ -álgebra $\mathcal{A} \subset \mathcal{B}$ que satisfaz $f^{-1}(\mathcal{A}) \subset \mathcal{A}$ a menos de medida nula também satisfaz $f^{-1}(\mathcal{A}) = \mathcal{A}$ a menos de medida nula.*

O Exercício 9.1.5 implica que se (f, μ) tem medida nula então o mesmo vale para qualquer fator. Portanto, o seguinte fato também é uma consequência imediata da proposição:

Corolário 9.5.7. *Nas condições da Proposição 9.5.5, todo fator de (f, μ) é invertível.*

Ainda não está totalmente compreendido como a entropia se relaciona com o tipo de espectro num sistema, mas existem diversos resultados parciais, especialmente no que se refere a sistemas com entropia nula.

Rokhlin [Rok67a, § 14] mostrou que todo sistema ergódico com espectro discreto definido num espaço de Lebesgue tem entropia nula. Isto também pode ser deduzido do fato de que, conforme mencionamos na Seção 8.3, todo sistema ergódico com espectro discreto é ergodicamente isomorfo a uma translação num grupo abeliano compacto. Como vimos no Corolário 8.5.7, em espaços de Lebesgue isomorfismo ergódico implica equivalência ergódica. Lembre também que sistemas com espectro discreto em espaços de Lebesgue são sempre invertíveis (Exercício 8.5.5).

No mesmo trabalho de Rokhlin também é provado que sistemas invertíveis com espectro singular definidos em espaços de Lebesgue têm entropia nula e o mesmo vale para sistemas com espectro de Lebesgue com posto finito (caso

existam). O caso de posto infinito será o foco da próxima seção. Mencionaremos que existem sistemas com espectro de Lebesgue com posto infinito e entropia nula. Por outro lado introduziremos a subclasse importante dos sistemas ditos de Kolmogorov, para os quais a entropia é necessariamente positiva, num sentido forte.

9.5.3 Sistemas de Kolmogorov

Seja (M, \mathcal{B}, μ) um espaço de probabilidade não trivial, ou seja, tal que nem todo conjunto mensurável tem medida 0 ou 1. Representamos por $\vee_{\alpha} \mathcal{U}_{\alpha}$ a σ -álgebra gerada por uma família qualquer de subconjuntos U_{α} de \mathcal{B} . Seja $f : M \rightarrow M$ uma transformação preservando a medida μ .

Definição 9.5.8. Dizemos que (f, μ) é um *sistema de Kolmogorov* se existe alguma σ -álgebra $\mathcal{A} \subset \mathcal{B}$ tal que

- (a) $f^{-1}(\mathcal{A}) \subset \mathcal{A}$ a menos de medida nula;
- (b) $\bigcap_{n=0}^{\infty} f^{-n}(\mathcal{A}) = \{\emptyset, M\}$ a menos de medida nula;
- (c) $\bigvee_{n=0}^{\infty} \{B \in \mathcal{B} : f^{-n}(B) \in \mathcal{A}\} = \mathcal{B}$ a menos de medida nula.

Deixamos ao cuidado do leitor verificar que esta propriedade é um invariante de equivalência ergódica (o mesmo *não* vale para equivalência espectral, como observaremos daqui a pouco).

Observe que se (f, μ) é sistema de Kolmogorov então (f^k, μ) é sistema de Kolmogorov, para todo $k \geq 1$. De fato, se \mathcal{A} satisfaz a condição (a) então $f^{-k}(\mathcal{A}) \subset \mathcal{A}$. Além disso, a condição (a) juntamente com (b) e (c) implica que

$$\begin{aligned} \bigcap_{n=0}^{\infty} f^{-kn}(\mathcal{A}) &= \bigcap_{n=0}^{\infty} f^{-n}(\mathcal{A}) = \{\emptyset, M\} \quad \text{e} \\ \bigvee_{n=0}^{\infty} \{B \in \mathcal{B} : f^{-kn}(B) \in \mathcal{A}\} &= \bigvee_{n=0}^{\infty} \{B \in \mathcal{B} : f^{-n}(B) \in \mathcal{A}\} = \mathcal{B} \end{aligned}$$

a menos de medida nula. Dizemos que (f, μ) é um *automorfismo de Kolmogorov* se é um sistema de Kolmogorov invertível. Veremos daqui a pouco que nesse caso o inverso (f^{-1}, μ) também é um sistema de Kolmogorov.

Proposição 9.5.9. *Todo sistema de Kolmogorov tem espectro de Lebesgue com posto infinito. Em particular, se a σ -álgebra \mathcal{B} é enumeravelmente gerada então o posto é enumerável.*

Demonstração. Seja $\mathcal{A} \subset \mathcal{B}$ uma σ -álgebra satisfazendo as condições na Definição 9.5.8. Seja $E = L_0^2(M, \mathcal{A}, \mu)$ o subespaço das funções $\varphi \in L_0^2(M, \mathcal{B}, \mu)$ que são \mathcal{A} -mensuráveis, ou seja, tais que a pré-imagem $\varphi^{-1}(B)$ de qualquer boreliano $B \subset \mathbb{R}$ está em \mathcal{A} a menos de medida nula.

Começemos por observar que $U_f(L_0^2(M, \mathcal{A}, \mu)) = L_0^2(M, f^{-1}(\mathcal{A}), \mu)$. De fato, é claro que se φ é \mathcal{A} -mensurável então $U_f \varphi = \varphi \circ f$ é $f^{-1}(\mathcal{A})$ -mensurável. A

inclusão \subset segue imediatamente. Reciprocamente, dado qualquer $B \in f^{-1}(\mathcal{A})$, tome $A \in \mathcal{A}$ tal que $B = f^{-1}(A)$ e seja $c = \mu(A) = \mu(B)$. Então $\mathcal{X}_B - c = U_f(\mathcal{X}_A - c)$ está em $U_f(L_0^2(M, \mathcal{A}, \mu))$. Isto implica a outra inclusão. Então, a hipótese de que $f^{-1}(\mathcal{A}) \subset \mathcal{A}$ a menos de medida nula garante que $U_f(E) \subset E$.

Segue do parágrafo anterior que $U_f^n(L_0^2(M, \mathcal{A}, \mu)) = L_0^2(M, f^{-n}(\mathcal{A}), \mu)$ para todo $n \geq 0$. Logo,

$$\bigcap_{n=0}^{\infty} U_f^n(L_0^2(M, \mathcal{A}, \mu)) = L_0^2(M, \bigcap_{n=0}^{\infty} f^{-n}(\mathcal{A}), \mu).$$

Então, a hipótese de que $\bigcap_{n=0}^{\infty} f^{-n}(\mathcal{A}) = \{0, M\}$ a menos de medida nula implica que $\bigcap_{n=0}^{\infty} U_f^n(E) = \{0\}$.

Agora considere $\mathcal{A}_n = \{B \in \mathcal{B} : f^{-n}(B) \in \mathcal{A}\}$. A sequência $(\mathcal{A}_n)_n$ é não decrescente, porque $f^{-1}(\mathcal{A}) \subset \mathcal{A}$. Além disso, cada φ é \mathcal{A}_n -mensurável se, e somente se, $U_f^n \varphi = \varphi \circ f^n$ é \mathcal{A} -mensurável. Isto mostra que $U_f^{-n}(L_0^2(M, \mathcal{A}, \mu)) = L_0^2(M, \mathcal{A}_n, \mu)$ para cada n . Observe igualmente que

$$\sum_{n=0}^{\infty} L_0^2(M, \mathcal{A}_n, \mu) = L_0^2(M, \bigvee_{n=0}^{\infty} \mathcal{A}_n, \mu). \quad (9.5.4)$$

De fato, é claro que $L_0^2(M, \mathcal{A}_k, \mu) \subset L_0^2(M, \bigvee_{n=0}^{\infty} \mathcal{A}_n, \mu)$ para todo k , uma vez que \mathcal{A}_k está contida em $\bigvee_{n=0}^{\infty} \mathcal{A}_n$. A inclusão \subset é uma consequência imediata desta observação, uma vez que $L_0^2(M, \bigvee_{n=0}^{\infty} \mathcal{A}_n, \mu)$ é um espaço de Banach. Agora considere qualquer $A \in \bigvee_{n=0}^{\infty} \mathcal{A}_n$. O teorema de aproximação (Teorema A.1.19) dá que para cada $\varepsilon > 0$ existe n e existe $A_n \in \mathcal{A}_n$ tal que $\mu(A \Delta A_n) < \varepsilon$. Então $(\mathcal{X}_{A_n})_n$ converge para \mathcal{X}_A na norma L^2 , e isso implica que $\mathcal{X}_A \in \sum_{n=0}^{\infty} L_0^2(M, \mathcal{A}_n, \mu)$. A inclusão \supset é uma consequência desta observação. Em vista da igualdade (9.5.4), a hipótese de que $\bigvee_{n=0}^{\infty} \mathcal{A}_n = \mathcal{B}$ a menos de medida nula dá que $\sum_{n=0}^{\infty} U_f^{-n}(E) = L_0^2(M, \mathcal{B}, \mu)$.

Isto conclui a demonstração de que E satisfaz as condições na Definição 8.4.1. Resta mostrar que $F = E \ominus U_f(E)$ tem dimensão infinita. Quando \mathcal{B} é enumeravelmente gerado $L_0^2(M, \mathcal{B}, \mu)$ é separável (Exemplo 8.4.7) e, por consequência, a dimensão de Hilbert de F será necessariamente enumerável.

Inicialmente, observe que $F \neq \{0\}$. Caso contrário, teríamos $U_f(E) = E$ e, portanto, $U_f^n(E) = E$ para todo $n \geq 1$. Pela condição (b), isso daria que $E = \bigcap_n U_f^n(E) = \{0\}$. Então, pela condição (c), teríamos $L_0^2(M, \mathcal{B}, \mu) = \{0\}$ e isso estaria em contradição com a hipótese de que o espaço de probabilidade é não trivial.

Seja φ um elemento não nulo de F , fixado de uma vez por todas. Seja N o conjunto dos $x \in M$ tais que $\varphi(x) \neq 0$. Então $N \in \mathcal{A}$ e $\mu(N) > 0$. Será conveniente considerar o espaço $E' = L^2(M, \mathcal{A}, \mu) = E \oplus \{\text{constantes}\}$. Observe que F coincide com $E' \ominus U_f(E')$, pois o operador de Koopman preserva a reta das funções constantes. Seja E'_N o subespaço das funções $\psi \in E'$ que se anulam fora de N , ou seja, tais que $\psi(x) = 0$ para todo $x \in N^c$. Afirmamos que E'_N tem dimensão infinita. Para a prova precisamos do seguinte fato:

Lema 9.5.10. *Seja \mathcal{A} qualquer σ -álgebra satisfazendo as condições na Definição 9.5.8. Então para todo $A \in \mathcal{A}$ com $\mu(A) > 0$ existe $B \subset A$ tal que $0 < \mu(B) < \mu(A)$.*

Demonstração. Suponha que \mathcal{A} admite algum elemento A com medida positiva e que não satisfaz a conclusão do lema. Afirmamos que $A \cap f^{-k}(A)$ tem medida nula para todo $k \geq 1$. Então

$$\mu(f^{-i}(A) \cap f^{-j}(A)) = \mu(A \cap f^{-j+i}(A)) = 0 \quad \text{para todo } 0 \leq i < j.$$

Como $\mu(f^{-j}(A)) = \mu(A)$ para todo $j \geq 0$, isto implica que a medida μ é infinita, o que é uma contradição. Portanto, para terminar a demonstração do lema basta provar a nossa afirmação.

Para isso, note que a condição (a) implica que $f^{-k}(A) \in f^{-k}(\mathcal{A}) \subset \mathcal{A}$. Então, $f^{-k}(A)$ precisa ter medida nula ou medida total em A :

$$\mu(A \cap f^{-k}(A)) = 0 \quad \text{ou} \quad \mu(A \setminus f^{-k}(A)) = 0.$$

Logo, para provar a afirmação basta excluir a segunda possibilidade. Suponha que $\mu(A \setminus f^{-k}(A)) = 0$. Então (Exercício 1.1.4), existe $B \in \mathcal{A}$ tal que $\mu(A \Delta B) = 0$ e $f^{-k}(B) = B$. Então $B = f^{-nk}(B)$ para todo $n \geq 1$ e, portanto,

$$B \in \bigcap_{n \in \mathbb{N}} f^{-nk}(\mathcal{A}) = \bigcap_{n \in \mathbb{N}} f^{-n}(\mathcal{A}).$$

Pela condição (b), isto quer dizer que B tem medida 0 ou 1. Como $\mu(B) = \mu(A)$ é positivo, segue que $\mu(A) = \mu(B) = 1$. Então, a hipótese que fizemos a respeito de A implica que a σ -álgebra \mathcal{A} só contém conjuntos com medida 0 ou 1. Pela condição (c), segue que \mathcal{B} só contém conjuntos com medida 0 ou 1, o que contradiz a suposição de que o espaço de probabilidade não é trivial. \square

Usando o Lema 9.5.10 podemos encontrar conjuntos $A_j \in \mathcal{A}$, $j \geq 1$ com medida positiva, contidos em N e disjuntos dois-a-dois. Então \mathcal{X}_{A_j} está em E'_N para todo j . Além disso, $A_i \cap A_j = \emptyset$ acarreta $\mathcal{X}_{A_i} \cdot \mathcal{X}_{A_j} = 0$ para todo $i \neq j$. Segue que E'_N tem dimensão infinita, tal como afirmamos.

Agora represente por $U_f(E')_N$ o subespaço das funções $\psi \in U_f(E')$ que se anulam fora de N . Seja $F_N = E'_N \ominus U_f(E')_N$. O fato de que $\dim E'_N = \infty$ assegura que $\dim F_N = \infty$ ou $\dim U_f(E')_N = \infty$ (ou ambos). Vamos mostrar que qualquer destas alternativas implica que $\dim F = \infty$.

Para tratar a primeira alternativa, basta provar que $F_N \subset F$. Mais ainda, como é claro que $F_N \subset E'$, basta verificar que F_N é ortogonal a $U_f(E')$. Sejam $\xi \in F_N$ e $\eta \in E'$ quaisquer. A função $(U_f \eta) \mathcal{X}_N = U_f(\eta \mathcal{X}_{f^{-1}(N)})$ está em $U_f(E')$ e se anula fora de N ; em outras palavras, ela está em $U_f(E')_N$. Como ξ se anula fora de N e ξ é ortogonal a $U_f(E')_N$, concluímos que $\xi \cdot U_f \eta = \xi \cdot (U_f \eta) \mathcal{X}_N = 0$. Isto termina o argumento neste caso.

Resta tratar a segunda alternativa. Se $\dim U_f(E')_N = \infty$, podemos tomar $\{\xi_k : k \geq 1\} \subset E'$ tal que $\{U_f \xi_k : k \geq 1\}$ é um subconjunto linearmente independente de $U_f(E')_N$. Não constitui restrição supor que as funções ξ_k são limitadas: se necessário, substitua ξ_k por $\xi_k \mathcal{X}_k$, onde \mathcal{X}_k é a função característica de

$\xi_k^{-1}(\{z : |z| \leq R_k\})$; a independência linear é preservada desde que a sequência $(R_k)_k$ convirja rapidamente para infinito. Então as funções $U_f \xi_k$ também são limitadas e $\{\varphi(U_f \xi_k) : k \geq 1\}$ é um subconjunto linearmente independente de E' . Além disso, dado qualquer $\eta \in E'$,

$$\varphi(U_f \xi_k) \cdot (U_f \eta) = \int \varphi(\xi_k \circ f)(\bar{\eta} \circ f) d\mu = \int \varphi(\xi_k \bar{\eta}) \circ f d\mu = \varphi \cdot U_f(\bar{\xi}_k \eta).$$

A última expressão é igual a zero, pois $\bar{\xi}_k \eta \in E'$ e a função $\varphi \in F$ é ortogonal a $U_f(E')$. Fazendo variar $\eta \in E'$, concluímos que $\varphi(U_f \xi_k)$ é ortogonal a $U_f(E')$ para todo k . Isto mostra que $\{\varphi(U_f \xi_k) : k \geq 1\}$ está contido em F e, portanto, $\dim F = \infty$ também neste caso. \square

Dizemos que uma partição finita ou enumerável de (M, \mathcal{B}, μ) é trivial se todos os seus elementos têm medida 0 ou 1.

Proposição 9.5.11. *Um sistema (f, μ) num espaço de Lebesgue é sistema de Kolmogorov se, e somente se, $h_\mu(f, \mathcal{P}) > 0$ para toda partição não trivial com entropia finita. Em particular, todo sistema de Kolmogorov tem entropia positiva.*

Este resultado é devido a Pinsker [Pin60] e a Rokhlin, Sinai [RS61]. A demonstração também pode ser encontrada no artigo de Rokhlin [Rok67a, § 13]. Mas assinalamos que a última parte do enunciado é uma consequência imediata das ideias na Seção 9.5.2. De fato, suponha que (f, μ) é um sistema de Kolmogorov com entropia nula. Pelo Corolário 9.5.6, qualquer σ -álgebra \mathcal{A} satisfazendo a condição (a) da Definição 9.5.8 também satisfaz $f^{-1}(\mathcal{A}) = \mathcal{A}$ a menos de medida nula. Então a condição (b) implica que \mathcal{A} é trivial e, pela condição (c), a própria σ -álgebra \mathcal{B} é trivial (contradizendo a hipótese que fizemos no início desta seção).

Segue da Proposição 9.5.11 e da relação (9.1.21) que o inverso de um automorfismo de Kolmogorov também é um automorfismo de Kolmogorov. Ao contrário do que acontece para automorfismos de Bernoulli (Exercício 9.5.1), no caso Kolmogorov os dois sistemas (f, μ) e (f^{-1}, μ) não são ergodicamente equivalentes, em geral.

Exemplo 9.5.12. O primeiro exemplo de sistema invertível com espectro de Lebesgue enumerável mas que não é um sistema de Kolmogorov foi construído por Girsanov em 1959, mas nunca foi publicado. Outro exemplo foi exibido por Newton, Parry [NP66] alguns anos depois. Trata-se de um fator de um certo deslocamento gaussiano (lembre do Exemplo 8.4.13) com espectro de Lebesgue enumerável mas cuja entropia é nula. Uma construção diferente foi dada por Gurevič [Gur61]: ele provou que o fluxo horocíclico em superfícies de curvatura negativa constante tem entropia nula; alguns anos antes, Parasyuk [Par53] havia mostrado que tais fluxos têm espectro de Lebesgue enumerável.

Como vimos no Teorema 8.4.11, todos os sistemas com espectro de Lebesgue enumerável são espectralmente equivalentes. Portanto, uma consequência interessante da existência de sistemas como no Exemplo 9.5.12 é que a propriedade de ser um sistema de Kolmogorov *não* é um invariante de equivalência espectral.

Exemplo 9.5.13. Verificamos nos Exemplos 8.4.2 e 8.4.3 que os deslocamentos de Bernoulli têm espectro de Lebesgue. Nos dois casos, unilateral e bilateral, exibimos subespaços de $L_0^2(M, \mathcal{B}, \mu)$ da forma $E = L_0^2(M, \mathcal{A}, \mu)$ para alguma σ -álgebra $\mathcal{A} \subset \mathcal{B}$. Portanto, o mesmo argumento prova que todo deslocamento de Bernoulli é um sistema de Kolmogorov. Em particular, todo automorfismo de Bernoulli é um automorfismo de Kolmogorov.

Existem sistemas de Kolmogorov invertíveis que não são automorfismos de Bernoulli. O primeiro exemplo, descoberto por Ornstein, é muito elaborado. A seguinte construção simples é devida a Kalikow [Kal82]:

Exemplo 9.5.14. Seja $\sigma : \Sigma \rightarrow \Sigma$ o deslocamento em $\Sigma = \{1, 2\}^{\mathbb{Z}}$ e seja μ a medida de Bernoulli associada ao vetor de probabilidade $p = (1/2, 1/2)$. Considere a aplicação $f : \Sigma \times \Sigma \rightarrow \Sigma \times \Sigma$ definida da seguinte forma:

$$f((x_n)_n, (y_n)_n) = (\sigma((x_n)_n), \sigma^{\pm 1}((y_n)_n))$$

onde o sinal é $-$ se $x_0 = 1$ e $+$ se $x_0 = 2$. Esta transformação preserva a medida $\mu \times \mu$. O sistema (f, μ) é um automorfismo de Kolmogorov mas não é um automorfismo de Bernoulli.

Considere qualquer automorfismo de Kolmogorov que não é um automorfismo de Bernoulli e seja $s > 0$ a sua entropia. Considere qualquer automorfismo de Bernoulli cuja entropia é igual a s (veja Exercício 9.2.2). Os dois sistemas têm a mesma entropia mas eles não são ergodicamente equivalentes, uma vez que ser automorfismo de Bernoulli é um invariante de equivalência ergódica. Portanto, a entropia não é um invariante completo de equivalência ergódica para automorfismos de Kolmogorov. Mais ainda, existe uma família não enumerável de automorfismos de Kolmogorov que não são ergodicamente conjugados e, no entanto, têm a mesma entropia (veja Ornstein, Shields [OS73]).

As propriedades dos automorfismos de Bernoulli descritas no Exercício 9.5.1 também não se estendem ao caso Kolmogorov: existe automorfismo de Kolmogorov que não é equivalente ao seu inverso (veja Ornstein, Shields [OS73]) e também existe automorfismo de Kolmogorov que não admite raiz k -ésima para nenhum valor de $k \geq 1$ (Clark [Cla72]).

Para encerrar esta seção, vamos discutir a propriedade de Kolmogorov para duas classes específicas de sistemas: deslocamentos de Markov e automorfismos de grupos compactos.

Relativamente à primeira delas, Friedman e Ornstein [FO70] provaram que todo deslocamento de Markov bilateral misturador é um automorfismo de Bernoulli. Lembre (Teorema 7.2.11) que um deslocamento de Markov é misturador se, e somente se, a respectiva matriz estocástica é aperiódica. Segue do teorema de Friedman, Ornstein que a entropia ainda é um invariante completo de equivalência ergódica no contexto mais amplo dos deslocamentos de Markov bilaterais misturadores. Outra consequência interessante é que todo deslocamento de Markov bilateral misturador é um automorfismo de Kolmogorov. Observamos, no entanto, que esta consequência admite uma prova direta relativamente fácil (veja o Exercício 9.5.4).

Quanto à segunda classe de sistemas, temos que todo automorfismo ergódico de um grupo compacto é um automorfismo de Kolmogorov. Este fato foi provado por Rokhlin [Rok67b] para grupos abelianos e por Yuzvinskii [Yuz68] no caso geral. De fato, automorfismos ergódicos de grupos compactos metrizáveis são automorfismos de Bernoulli (Lind [Lin77] e Miles, Thomas [MT78]). Em particular, todo automorfismo linear ergódico do toro \mathbb{T}^d é um automorfismo de Bernoulli; isto havia sido provado por Katznelson [Kat71]. Lembre (Teorema 4.2.14) que f_A é ergódico se, e somente se, nenhum autovalor da matriz A é raiz da unidade.

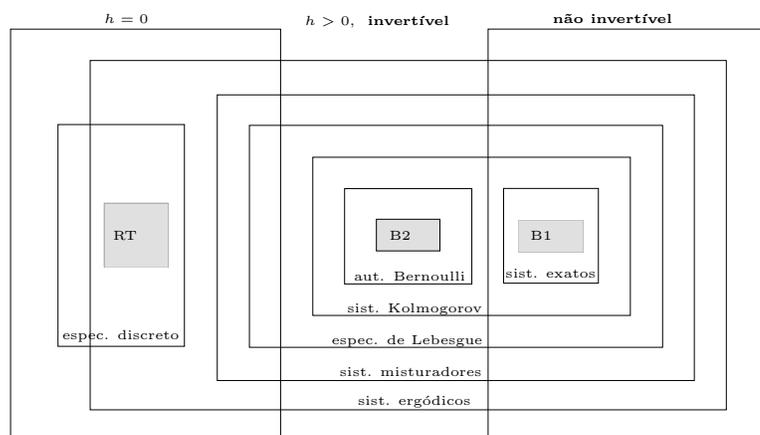


Figura 9.2: Relações entre diversas classes de sistemas (B1 = deslocamentos unilaterais, B2 = deslocamentos bilaterais, RT = rotações em \mathbb{T}^d)

9.5.4 Sistemas exatos

Dizemos que um sistema de Kolmogorov é *exato* se podemos tomar para \mathcal{A} a σ -álgebra \mathcal{B} de todos os conjuntos mensuráveis. Note que neste caso as condições (a) e (c) da Definição 9.5.8 são automaticamente satisfeitas. Portanto, um sistema (f, μ) é exato se, e somente se, a σ -álgebra \mathcal{B} é tal que $\bigcap_{n=0}^{\infty} f^{-n}(\mathcal{B})$ é trivial, ou seja, contém apenas conjuntos com medida 0 ou 1. Equivalentemente, (f, μ) é exato se, e somente se

$$\bigcap_{n=0}^{\infty} U_f^n(L_0^2(M, \mathcal{B}, \mu)) = \{0\}.$$

Esta observação também mostra que, ao contrário do que acontece com a propriedade de Kolmogorov, a exatidão é um invariante de equivalência espectral.

Verificamos no Exemplo 8.4.2 que os deslocamentos de Bernoulli unilaterais têm espectro de Lebesgue. Lembre que, para isso, consideramos o subespaço $E = L_0^2(M, \mathcal{B}, \mu)$. Portanto, o mesmo argumento prova que todo deslocamento

de Bernoulli unilateral é um sistema exato. Uma classe mais geral de exemplos, transformações expansoras munidas de seus estados de equilíbrio, será estudada no Capítulo 12.

É imediato que sistemas invertíveis nunca são exatos. De fato, no caso invertível $f^{-n}(\mathcal{B}) = \mathcal{B}$ a menos de medida nula, para todo n ; portanto, a condição de exatidão corresponde a dizer que a σ -álgebra \mathcal{B} é trivial (o que está excluído, por hipótese).

9.5.5 Exercícios

9.5.1. Mostre que se (f, μ) é um automorfismo de Bernoulli então ele é ergodicamente equivalente ao seu inverso (f^{-1}, μ) . Além disso, para todo $k \geq 1$ existe um automorfismo de Bernoulli (g, ν) tal que (g^k, ν) é ergodicamente equivalente a (f, μ) (em outras palavras, g é raiz k -ésima de f). [Observação: Ornstein provou que, reciprocamente, toda raiz k -ésima de um automorfismo de Bernoulli é um automorfismo de Bernoulli; veja [Orn74]]

9.5.2. Use a ideia de ponto de densidade para mostrar que a transformação expansão decimal $f(x) = 10x - [10x]$, munida da medida de Lebesgue, é exata.

9.5.3. Mostre que a transformação de Gauss é exata, relativamente à sua medida invariante absolutamente contínua μ .

9.5.4. Mostre que o deslocamento de Markov bilateral associado a qualquer matriz estocástica aperiódica P é um automorfismo de Kolmogorov.

9.5.5. Mostre que o deslocamento de Markov unilateral associado a qualquer matriz estocástica aperiódica P é um sistema exato.

9.5.6. Prove que se (f, μ) é exato então $h_\mu(f, \mathcal{P}) > 0$ para toda partição finita \mathcal{P} não trivial.

9.6 Entropia e decomposição ergódica

Não é difícil mostrar que a entropia $h_\mu(f)$ é sempre uma função *afim* da medida invariante μ :

Proposição 9.6.1. *Sejam μ e ν probabilidades invariantes por uma transformação $f : M \rightarrow M$. 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 $B \subset M$,

$$\begin{aligned} & \phi(t\mu(B) + (1-t)\nu(B)) - t\phi(\mu(B)) - (1-t)\phi(\nu(B)) \\ &= -t\mu(B) \log \frac{t\mu(B) + (1-t)\nu(B)}{\mu(B)} - (1-t)\nu(B) \log \frac{t\mu(B) + (1-t)\nu(B)}{\nu(B)} \\ &\leq -t\mu(B) \log t - (1-t)\nu(B) \log(1-t). \end{aligned}$$

porque a função $-\log$ é decrescente. Portanto, dada qualquer partição finita ou enumerável \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,

$$h_{t\mu+(1-t)\nu}(f, \mathcal{P}) = th_\mu(f, \mathcal{P}) + (1-t)h_\nu(f, \mathcal{P}). \quad (9.6.1)$$

Segue, imediatamente, que $h_{t\mu+(1-t)\nu}(f) \leq th_\mu(f) + (1-t)h_\nu(f)$. Além disso, (9.1.16) e (9.6.1) implicam que

$$h_{t\mu+(1-t)\nu}(f, \mathcal{P}_1 \vee \mathcal{P}_2) \geq th_\mu(f, \mathcal{P}_1) + (1-t)h_\nu(f, \mathcal{P}_2)$$

para quaisquer partições \mathcal{P}_1 e \mathcal{P}_2 . Tomando o supremo em \mathcal{P}_1 e \mathcal{P}_2 obtemos que $h_{t\mu+(1-t)\nu}(f) \geq th_\mu(f) + (1-t)h_\nu(f)$. \square

Em particular, dado qualquer conjunto invariante $A \subset M$, temos

$$h_\mu(f) = \mu(A)h_{\mu_A}(f) + \mu(A^c)h_{\mu_{A^c}}(f), \quad (9.6.2)$$

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 9.1.3). Outra consequência imediata é a seguinte versão da Proposição 9.6.1 para combinações convexas finitas:

$$\mu = \sum_{i=1}^n t_i \mu_i \quad \Rightarrow \quad h_\mu(f) = \sum_{i=1}^n t_i h_{\mu_i}(f), \quad (9.6.3)$$

quaisquer que sejam as probabilidades invariantes μ_1, \dots, μ_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 5.1.3:

Teorema 9.6.2 (Jacobs). *Suponha que M é um espaço métrico 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).*

Vamos deduzir este resultado de um teorema geral sobre funcionais afins no espaço das probabilidades, que enunciaremos na Seção 9.6.1 e cuja demonstração será dada na Seção 9.6.2.

9.6.1 Afinidade da decomposição ergódica

Seja M um espaço métrico separável completo. Vimos no Lema 2.1.3 que a topologia fraca* no espaço das probabilidades $\mathcal{M}_1(M)$ é metrizável. Além disso (Exercício 2.1.3), o espaço métrico $\mathcal{M}_1(M)$ é separável.

Seja W uma probabilidade na σ -álgebra de Borel de $\mathcal{M}_1(M)$. Chamamos *baricentro* de W à probabilidade $\text{bar}(W) \in \mathcal{M}_1(M)$ dada por

$$\int \psi d\text{bar}(W) = \int \left(\int \psi d\eta \right) dW(\eta) \quad (9.6.4)$$

para toda função mensurável limitada $\varphi : M \rightarrow \mathbb{R}$. Fica a cuidado do leitor (Exercício 9.6.1) verificar que esta relação determina $\text{bar}(W)$ univocamente e que o baricentro é uma função afim da medida (Exercício 9.6.2).

Exemplo 9.6.3. Se W é uma medida de Dirac, ou seja, se $W = \delta_\nu$ para algum $\nu \in \mathcal{M}_1(M)$, então $\text{bar}(W) = \nu$. Usando o Exercício 9.6.2 obtemos a seguinte generalização: se $W = \sum_{i=1}^{\infty} t_i \delta_{\nu_i}$ com $t_i \geq 0$ e $\sum_{i=1}^{\infty} t_i = 1$ e $\nu_i \in \mathcal{M}_1(M)$ para todo i , então $\text{bar}(W) = \sum_{i=1}^{\infty} t_i \nu_i$.

Exemplo 9.6.4. Seja $\{\mu_P : P \in \mathcal{P}\}$ a decomposição ergódica de uma medida μ invariante por uma transformação mensurável $f : M \rightarrow M$, com medida quociente $\hat{\mu}$. Seja W a imagem da probabilidade $\hat{\mu}$ pela aplicação $\mathcal{P} \rightarrow \mathcal{M}$ que associa a cada $P \in \mathcal{P}$ a probabilidade condicional μ_P . Então (Exercício 5.1.4),

$$\int \psi d\mu = \int \left(\int \psi d\mu_P \right) d\hat{\mu}(P) = \int \left(\int \psi d\eta \right) dW(\eta)$$

para toda função $\psi : M \rightarrow \mathbb{R}$ mensurável limitada. Isto significa que μ é o baricentro de W .

Diremos que $\mathcal{M} \subset \mathcal{M}_1(M)$ é *fortemente convexo* se $\sum_{i=1}^{\infty} t_i \nu_i \in \mathcal{M}$ quaisquer que sejam $\nu_i \in \mathcal{M}$ e $t_i \geq 0$ com $\sum_{i=1}^{\infty} t_i = 1$.

Teorema 9.6.5. *Seja \mathcal{M} um subconjunto fortemente convexo de $\mathcal{M}_1(M)$ e seja $H : \mathcal{M} \rightarrow \mathbb{R}$ um funcional afim, não negativo e semicontínuo superiormente. Então,*

$$H(\text{bar}(W)) = \int H(\eta) dW(\eta).$$

para toda probabilidade W em $\mathcal{M}_1(M)$ tal que $W(\mathcal{M}) = 1$ e $\text{bar}(W) \in \mathcal{M}$.

Antes de provarmos este resultado, vamos explicar como o Teorema 9.6.2 pode ser obtido a partir dele. O passo essencial é o seguinte lema:

Lema 9.6.6. $h_\mu(f, \mathcal{Q}) = \int h_{\mu_P}(f, \mathcal{Q}) d\hat{\mu}(P)$ para toda partição finita \mathcal{Q} de M .

Demonstração. Consideremos $\mathcal{M} = \mathcal{M}_1(f)$, o subespaço das probabilidades invariantes, e $H : \mathcal{M} \rightarrow \mathbb{R}$ definido por $H(\eta) = h_\eta(f, \mathcal{Q})$. Seja W a imagem da medida quociente $\hat{\mu}$ pela aplicação $\mathcal{P} \rightarrow \mathcal{M}$ que associa a cada $P \in \mathcal{P}$ a probabilidade condicional μ_P . É claro que \mathcal{M} é fortemente convexo, $W(\mathcal{M}) = 1$

e (lembre do Exemplo 9.6.4) $\text{bar}(W) = \mu$ está em \mathcal{M} . A Proposição 9.6.1 dá que H é afim e é claro que ele é não negativo. Para podermos aplicar o Teorema 9.6.5, resta verificar que H também é semicontínuo superiormente.

Inicialmente, suponha que f é o deslocamento num espaço $\Sigma = X^{\mathbb{N}}$, onde X é um conjunto finito, e que \mathcal{Q} é a partição em cilindros $[0; a]$, $a \in X$. A propriedade desta partição que nos interessa é que os seus elementos são abertos e fechados de Σ . Em outras palavras, $\partial Q = \emptyset$ para todo $Q \in \mathcal{Q}$. Pela Proposição 9.2.10, segue que a aplicação $\eta \mapsto H(\eta) = h_\eta(f, \mathcal{Q})$ é semicontínua superiormente em todo ponto de \mathcal{M} . Então, estamos em condições de aplicar o Teorema 9.6.5 ao funcional H . Desta forma obtemos que

$$\begin{aligned} h_\mu(f, \mathcal{Q}) &= H(\mu) = H(\text{bar}(W)) = \int H(\eta) dW(\eta) \\ &= \int H(\mu_P) d\hat{\mu}(P) = \int h_{\mu_P}(f, \mathcal{Q}) d\hat{\mu}(P). \end{aligned}$$

Agora tratamos o caso geral, por redução ao caso do parágrafo anterior. Dada a partição finita \mathcal{Q} , considere $\Sigma = \mathcal{Q}^{\mathbb{N}}$ e

$$h : M \rightarrow \Sigma, \quad h(x) = (\mathcal{Q}(f^n(x)))_{n \in \mathbb{N}}.$$

Observe que $h \circ f = \sigma \circ h$, onde $\sigma : \Sigma \rightarrow \Sigma$ representa o deslocamento. A cada medida η em M podemos associar a medida $\eta' = h_*\eta$ em Σ . A relação anterior garante que se η é invariante por f então η' é invariante por σ . Além disso, se η é ergódica então η' é ergódica. De fato, se $B' \subset \Sigma$ é um conjunto invariante por σ então $B = h^{-1}(B')$ é um conjunto invariante por f . Supondo que η é ergódica, segue que $\eta'(B') = \eta(B)$ é igual a 0 ou 1; logo, η' é ergódica.

Por construção, $\mathcal{Q} = h^{-1}(\mathcal{Q}')$, onde \mathcal{Q}' representa a partição de Σ nos cilindros $[0; Q]$, $Q \in \mathcal{Q}$. Mais geralmente, $\bigvee_{j=0}^{n-1} f^{-j}(\mathcal{Q}) = h^{-1}(\bigvee_{j=0}^{n-1} \sigma^{-j}(\mathcal{Q}'))$ e, portanto,

$$H_\eta\left(\bigvee_{j=0}^{n-1} f^{-j}(\mathcal{Q})\right) = H_{\eta'}\left(\bigvee_{j=0}^{n-1} \sigma^{-j}(\mathcal{Q}')\right)$$

para todo $n \in \mathbb{N}$. Dividindo por n e passando ao limite,

$$h_\eta(f, \mathcal{Q}) = h_{\eta'}(\sigma, \mathcal{Q}') \quad \text{para todo } \eta \in \mathcal{M}. \quad (9.6.5)$$

Denotamos $\mu' = h_*\mu$ e $\mu'_P = h_*(\mu_P)$ para cada P . Para toda função mensurável limitada $\psi : \Sigma \rightarrow \mathbb{R}$,

$$\begin{aligned} \int \psi d\mu' &= \int (\psi \circ h) d\mu = \int \left(\int (\psi \circ h) d\mu_P \right) d\hat{\mu}(P) \\ &= \int \left(\int \psi d\mu'_P \right) d\hat{\mu}(P). \end{aligned} \quad (9.6.6)$$

Como as medidas μ'_P são ergódicas, a relação (9.6.6) significa que $\{\mu'_P : P \in \mathcal{P}\}$ é a decomposição ergódica de μ' . Então, de acordo com o parágrafo anterior,

$h_{\mu'}(\sigma, \mathcal{Q}') = \int h_{\mu'_P}(\sigma, \mathcal{Q}') d\hat{\mu}(P)$. Pela relação (9.6.5) aplicada a $\eta = \mu$ e a $\eta = \mu_P$, isto pode ser reescrito na seguinte forma:

$$h_{\mu}(\sigma, \mathcal{Q}) = \int h_{\mu_P}(\sigma, \mathcal{Q}) d\hat{\mu}(P).$$

Esta última igualdade é, precisamente, o que pretendíamos provar. \square

Prosseguindo com a demonstração do Teorema 9.6.2, considere qualquer sequência crescente $\mathcal{Q}_1 \prec \dots \prec \mathcal{Q}_n \prec \dots$ de partições finitas de M tais que $\text{diam } \mathcal{Q}_n(x)$ converge para zero em todo $x \in M$ (tal sequência pode ser construída a partir de uma família de bolas centradas nos pontos de um subconjunto enumerável denso, com raios convergindo para zero). Pelo Lema 9.6.6,

$$h_{\mu}(f, \mathcal{Q}_n) = \int h_{\mu_P}(f, \mathcal{Q}_n) d\hat{\mu}(P) \tag{9.6.7}$$

para todo n . De acordo com (9.1.16), a sequência $h_{\eta}(f, \mathcal{Q}_n)$ é não decrescente, qualquer que seja a medida invariante η . Além disso, pelo Corolário 9.2.3, o seu limite é igual a $h_{\eta}(f)$. Logo, podemos passar ao limite em (9.6.7) com a ajuda do teorema da convergência monótona. Deste modo obtemos que

$$h_{\mu}(f) = \int h_{\mu_P}(f) d\hat{\mu}(P),$$

como queríamos demonstrar. Note que o argumento permanece válido mesmo que algum dos lados da igualdade seja infinito (então o outro também é).

Desta forma, conseguimos reduzir a demonstração do Teorema 9.6.2 a provar o Teorema 9.6.5.

9.6.2 Demonstração do teorema de Jacobs

Agora vamos demonstrar o Teorema 9.6.5. Começemos por provar que a função baricentro é contínua, no sentido de que se W está concentrada em uma vizinhança V de uma dada medida ν então o baricentro de W também está próximo de ν . Mais precisamente:

Lema 9.6.7. *Seja W uma probabilidade em $\mathcal{M}_1(M)$ e seja $\nu \in \mathcal{M}_1(M)$. Dado qualquer conjunto finito $\Phi = \{\phi_1, \dots, \phi_N\}$ de funções contínuas limitadas e dado qualquer $\varepsilon > 0$, considere a vizinhança $V = V(\nu, \Phi, \varepsilon)$ definida em (2.1.1). Se $W(V) = 1$, então $\text{bar}(W) \in V$.*

Demonstração. Considere qualquer $i = 1, \dots, N$. Pela definição de baricentro, e a hipótese de que o complementar de V tem medida nula,

$$\begin{aligned} \left| \int \phi_i d\text{bar}(W) - \int \phi_i d\nu \right| &= \left| \int \left(\int \phi_i d\eta \right) dW(\eta) - \int \left(\int \phi_i d\nu \right) dW(\eta) \right| \\ &\leq \int_V \left| \int \phi_i d\eta - \int \phi_i d\nu \right| dW(\eta). \end{aligned}$$

Pela definição de V , a última expressão é menor que ε . Portanto,

$$\left| \int \phi_i d\text{bar}(W) - \int \phi_i d\nu \right| < \varepsilon$$

para todo $i = 1, \dots, N$. Em outras palavras, $\text{bar}(W) \in V$. \square

Também usaremos a seguinte propriedade simples de funcionais afins não negativos:

Lema 9.6.8. *Considere probabilidades $\nu_i \in \mathcal{M}$, $i \geq 1$ e números não negativos t_i , $i \geq 1$ com $\sum_{i=1}^{\infty} t_i = 1$. Então, para qualquer funcional afim não negativo $H : \mathcal{M} \rightarrow \mathbb{R}$,*

$$H\left(\sum_{i=1}^{\infty} t_i \nu_i\right) \geq \sum_{i=1}^{\infty} t_i H(\nu_i).$$

Demonstração. Para $n \geq 1$, seja $s_n = \sum_{i=1}^n t_i$. Tome $R_n = (1 - s_n)^{-1} \sum_{i>n} t_i \nu_i$ se $s_n < 1$; caso contrário, escolha R_n arbitrariamente. Então,

$$\sum_{i=1}^{\infty} t_i \nu_i = \sum_{i=1}^n t_i \nu_i + (1 - s_n) R_n.$$

Como H é afim, e a expressão do lado direito é uma combinação convexa (finita), segue que

$$H\left(\sum_{i=1}^{\infty} t_i \nu_i\right) = \sum_{i=1}^n t_i H(\nu_i) + (1 - s_n) H(R_n) \geq \sum_{i=1}^n t_i H(\nu_i)$$

para todo n . Agora basta fazer n tender para infinito. \square

Corolário 9.6.9. *Se $H : \mathcal{M} \rightarrow \mathbb{R}$ é um funcional afim não negativo então H é limitado.*

Demonstração. Suponha que H não é limitado. Então existem $\nu_i \in \mathcal{M}$ tais que $H(\nu_i) \geq 2^i$ para todo $i \geq 1$. Considere $\nu = \sum_{i=1}^{\infty} 2^{-i} \nu_i$. Pelo Lema 9.6.8,

$$H(\nu) \geq \sum_{i=1}^{\infty} 2^{-i} H(\nu_i) = \infty.$$

Isto contradiz o fato de que $H(\nu)$ é finito. \square

Agora estamos prontos para demonstrar o Teorema 9.6.5. Começemos pela desigualdade \geq . Escreva $\mu = \text{bar}(W)$ e considere qualquer $\varepsilon > 0$. Pela hipótese de semicontinuidade, existe $\delta > 0$ e uma família finita $\Phi = \{\phi_1, \dots, \phi_N\}$ de funções contínuas limitadas tais que

$$H(\eta) < H(\mu) + \varepsilon \quad \text{para todo } \eta \in \mathcal{M} \cap V(\mu, \Phi, \delta). \quad (9.6.8)$$

Como $\mathcal{M}_1(M)$ é espaço métrico separável, o mesmo vale para o subespaço \mathcal{M} . Então podemos considerar uma base enumerável de abertos $\{V_1, \dots, V_n, \dots\}$ de \mathcal{M} , com as seguintes propriedades:

- (i) cada V_n está contido em $\mathcal{M} \cap V(\nu_n, \Phi, \delta)$ para algum $\nu_n \in \mathcal{M}$;
- (ii) e $H(\eta) < H(\nu_n) + \varepsilon$ para todo $\eta \in V_n$.

Considere a partição enumerável $\{P_1, \dots, P_n, \dots\}$ definida por $P_1 = \mathcal{M} \cap V_1$ e $P_n = \mathcal{M} \cap V_n \setminus (V_1 \cup \dots \cup V_{n-1})$ para todo $n > 1$. É claro que as propriedades (i) e (ii) permanecem válidas se substituirmos V_n por P_n . Afirmamos que

$$\sum_n W(P_n)\nu_n \in V(\mu, \Phi, \delta) \tag{9.6.9}$$

De fato, observe que

$$\left| \int \phi_i d\mu - \sum_n W(P_n) \int \phi_i d\nu_n \right| = \left| \sum_n \int_{P_n} \left(\int \phi_i d\eta - \int \phi_i d\nu_n \right) dW(\eta) \right|$$

para todo i . Portanto, a propriedade (i) garante que

$$\left| \int \phi_i d\mu - \sum_n W(P_n) \int \phi_i d\nu_n \right| < \sum_n \delta W(P_n) = \delta \quad \text{para todo } i,$$

que é o que significa (9.6.9). Então, combinando (9.6.8), (9.6.9) e o Lema 9.6.8,

$$\sum_n W(P_n)H(\nu_n) \leq H\left(\sum_n W(P_n)\nu_n\right) < H(\mu) + \varepsilon.$$

Por outro lado, a propriedade (ii) implica que

$$\begin{aligned} \int H(\eta) dW(\eta) - \sum_n W(P_n)H(\nu_n) &= \sum_n \int_{P_n} (H(\eta) - H(\nu_n)) dW\mu(\eta) \\ &< \sum_n \varepsilon W(P_n) = \varepsilon. \end{aligned}$$

Somando as duas últimas desigualdades, $\int H(\eta) dW(\eta) < H(\mu) + 2\varepsilon$. Como $\varepsilon > 0$ é arbitrário, isto implica que $H(\mu) \geq \int H(\eta) dW(\eta)$.

Falta provar a desigualdade \leq no Teorema 9.6.2. Considere qualquer sequência $(\mathcal{P}_n)_n$ de partições finitas de \mathcal{M} tal que o diâmetro de $\mathcal{P}_n(\nu)$ converge para zero quando n vai para infinito, qualquer que seja $\nu \in \mathcal{M}$. Por exemplo, $\mathcal{P}_n = \bigvee_{i=1}^n \{V_i, V_i^c\}$, onde $\{V_n : n \geq 1\}$ é qualquer base enumerável de abertos de \mathcal{M} . Para cada n fixado, considere a restrição normalizada W_P da medida W a cada conjunto $P \in \mathcal{P}_n$ (consideraremos apenas conjuntos com medida positiva; a união de todos os elementos de $\cup_n \mathcal{P}_n$ com $W(P) = 0$ pode ser negligenciada, pois tem medida nula):

$$W_P(A) = \frac{W(A \cap P)}{W(P)} \quad \text{para cada conjunto mensurável } A \subset M.$$

É claro que $W = \sum_{P \in \mathcal{P}_n} W(P)W_P$. Como o baricentro é uma função afim (Exercício 9.6.2), segue que

$$\text{bar}(W) = \sum_{P \in \mathcal{P}_n} W(P) \text{bar}(W_P)$$

e, portanto,

$$H(\text{bar}(W)) = \sum_{P \in \mathcal{P}_n} W(P)H(\text{bar}(W_P)).$$

Defina $H_n(\eta) = H(\text{bar}(W_{\mathcal{P}_n(\eta)}))$, para cada $\eta \in \mathcal{M}$. Então a última igualdade acima pode ser reescrita na seguinte forma:

$$H(\text{bar}(W)) = \int H_n(\eta) dW(\eta) \quad \text{para todo } n. \quad (9.6.10)$$

Segue imediatamente da definição que $0 \leq H_n(\eta) \leq \sup H$ para todo n e todo η . Lembre que $\sup H < \infty$ (Corolário 9.6.9). Afirmamos que

$$\limsup_n H_n(\eta) \leq H(\eta) \quad \text{para todo } \eta \in \mathcal{M}. \quad (9.6.11)$$

Isto pode ser visto da seguinte forma. Dada qualquer vizinhança $V = V(\eta, \Phi, \varepsilon)$ de η , temos que $\mathcal{P}_n(\eta) \subset V$ para todo n suficientemente grande, uma vez que o diâmetro de $\mathcal{P}_n(\eta)$ converge para zero. Então (sempre supondo que $W(\mathcal{P}_n(\eta))$ é positiva), $W_{\mathcal{P}_n(\eta)}(V) \geq W_{\mathcal{P}_n(\eta)}(\mathcal{P}_n(\eta)) = 1$. Pelo Lema 9.6.7, segue que $\text{bar}(W_{\mathcal{P}_n(\eta)}) \in V$ para todo n suficientemente grande. Agora basta usar a hipótese de que H é semicontínuo superiormente para obter (9.6.11). Consequentemente (aplique o lema de Fatou à sequência $-H_n + \sup H$),

$$\limsup_n \int H_n(\eta) dW(\eta) \leq \int \limsup_n H_n(\eta) dW(\eta) \leq \int H(\eta) dW(\eta). \quad (9.6.12)$$

Combinando as relações (9.6.10) e (9.6.12), obtemos que

$$H(\text{bar}(W)) \leq \int H(\eta) dW(\eta),$$

tal como queríamos demonstrar.

A demonstração dos Teoremas 9.6.2 e 9.6.5 está completa.

9.6.3 Exercícios

9.6.1. Verifique que, dada qualquer probabilidade W no espaço $\mathcal{M}_1(M)$, existe uma única probabilidade $\text{bar}(W) \in \mathcal{M}_1(M)$ em M satisfazendo (9.6.4).

9.6.2. Mostre que a função baricentro é afim, no seguinte forte: se W_i , $i \geq 1$ são probabilidades em $\mathcal{M}_1(M)$ e t_i , $i \geq 1$ são números não negativos com $\sum_{i=1}^{\infty} t_i = 1$, então

$$\text{bar}\left(\sum_{i=1}^{\infty} t_i W_i\right) = \sum_{i=1}^{\infty} t_i \text{bar}(W_i).$$

9.6.3. Mostre que se $\mathcal{M} \subset \mathcal{M}_1(M)$ é conjunto convexo fechado então \mathcal{M} é fortemente convexo. Além disso, nesse caso $W(\mathcal{M}) = 1$ implica $\text{bar}(W) \in \mathcal{M}$.

9.6.4. A desigualdade \geq no Teorema 9.6.2 pode ser obtida por um argumento mais direto:

1. Lembrando que a função $\phi(x) = -x \log x$ é côncava, mostre que vale $H_\mu(\mathcal{Q}) \geq \int H_{\mu_P}(\mathcal{Q}) d\hat{\mu}(P)$ para toda partição finita \mathcal{Q} .
2. Deduza que $h_\mu(f, \mathcal{Q}) \geq \int h_{\mu_P}(f, \mathcal{Q}) d\hat{\mu}(P)$ para toda partição finita \mathcal{Q} .
3. Conclua que $h_\mu(f) \geq \int h_{\mu_P}(f) d\hat{\mu}(P)$.

9.6.5. A desigualdade \leq no Teorema 9.6.2 está baseada na propriedade de que $h_\mu(f, \mathcal{Q}) \leq \int h_{\mu_P}(f, \mathcal{Q}) d\hat{\mu}(P)$ para toda partição finita \mathcal{Q} , a qual é parte do Lema 9.6.6. Aponte o que está errado na "demonstração alternativa" a seguir.

Seja \mathcal{Q} uma partição finita. O teorema de Shannon-McMillan-Breimen garante que $h_\mu(f, \mathcal{Q}) = \int h_\mu(f, \mathcal{Q}, x) d\mu(x)$, onde

$$h_\mu(f, \mathcal{Q}, x) = \lim_n -\frac{1}{n} \log \mu(\mathcal{Q}^n(x)) = \lim_n -\frac{1}{n} \log \int \mu_P(\mathcal{Q}^n(x)) d\hat{\mu}(P).$$

Pela desigualdade de Jensen aplicada à função convexa $\psi(x) = -\log x$,

$$\lim_n -\frac{1}{n} \log \int \mu_P(\mathcal{Q}^n(x)) d\hat{\mu}(P) \leq \lim_n \int -\frac{1}{n} \log \mu_P(\mathcal{Q}^n(x)) d\hat{\mu}(P).$$

Usando o fato de que $h_{\mu_P}(f, \mathcal{Q}) = h_{\mu_P}(f, \mathcal{Q}, x)$ em quase todo ponto (porque toda μ_P é ergódica),

$$\begin{aligned} \lim_n \int -\frac{1}{n} \log \mu_P(\mathcal{Q}^n(x)) d\hat{\mu}(P) &= \int \lim_n -\frac{1}{n} \log \mu_P(\mathcal{Q}^n(x)) d\hat{\mu}(P) \\ &= \int h_{\mu_P}(f, \mathcal{Q}) d\hat{\mu}(P). \end{aligned}$$

Isto mostra que $h_\mu(f, \mathcal{Q}, x) \leq \int h_{\mu_P}(f, \mathcal{Q}) d\hat{\mu}(P)$ para toda partição finita \mathcal{Q} e para μ -quase todo x . Consequentemente, $h_\mu(f, \mathcal{Q}) \leq \int h_{\mu_P}(f, \mathcal{Q}) d\hat{\mu}(P)$ para toda partição finita \mathcal{Q} .

9.7 Jacobianos e fórmula de Rokhlin

Seja U um aberto do \mathbb{R}^d , seja m a medida de Lebesgue e seja $f : U \rightarrow U$ um difeomorfismo local. Pela fórmula de mudança de variáveis,

$$m(f(A)) = \int_A |\det Df(x)| dx \quad (9.7.1)$$

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 9.4.2.

Seja $f : M \rightarrow M$ 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 injetividade*. 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 $\xi : M \rightarrow [0, \infty)$ 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

$$\eta(f(A)) = \int_A \xi d\eta \quad (9.7.2)$$

Note (Exercício 9.7.1) que a definição não depende da escolha de $\{U_k : k \geq 1\}$.

Exemplo 9.7.1. Seja $\sigma : \Sigma \rightarrow \Sigma$ o deslocamento em $\Sigma = \{1, 2, \dots, d\}^{\mathbb{N}}$ e seja μ a medida de Bernoulli associada a um vetor de probabilidade $p = (p_1, \dots, p_d)$. Consideramos Σ munido da distância (A.2.7) para algum valor (arbitrário) de θ . A restrição de σ a cada cilindro $[0; a]$ é uma transformação invertível. Além disso, dado qualquer cilindro $[0; a, a_1, \dots, a_n] \subset [0; a]$,

$$\mu(\sigma([0; a, a_1, \dots, a_n])) = p_{a_1} \cdots p_{a_n} = \frac{1}{p_a} \mu([0; a, a_1, \dots, a_n]).$$

Deixamos ao cuidado do leitor deduzir que $\mu(\sigma(A)) = (1/p_a)\mu(A)$ 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 $f : U \rightarrow U$ é um difeomorfismo local num aberto de \mathbb{R}^d e η é a medida de Lebesgue, então η é não singular. Também é fácil ver que toda probabilidade invariante é não singular.

Segue imediatamente da definição (9.7.2) 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 9.7.2. *Seja $f : M \rightarrow M$ uma transformação localmente invertível e seja η uma medida boreliana 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. Começemos 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 em definida em P_k por $\eta_k(A) = \eta(f(A))$. Em outras palavras, η_k é a imagem por $(f|_{P_k})^{-1}$ 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 $\xi_k = d\eta_k/d(\eta|_{P_k})$ a derivada de Radón-Nykodim (Teorema A.2.18). Então ξ_k é uma função definida em P_k , integrável com relação a η e satisfazendo

$$\eta(f(A)) = \eta_k(A) = \int_A \xi_k d\eta \quad (9.7.3)$$

para todo conjunto mensurável $A \subset P_k$. Considere a função $\xi : M \rightarrow [0, \infty)$ 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, \dots, P_k . Aplicando (9.7.3) 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 e 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. \square

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 $f(P_k)$.

O principal resultado desta seção é a seguinte fórmula para a entropia de uma medida invariante:

Teorema 9.7.3 (fórmula de Rokhlin). *Seja $f : M \rightarrow M$ uma transformação localmente invertível e seja μ uma probabilidade invariante por f . Suponha que existe alguma partição finita ou enumerável \mathcal{P} 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 9.2.3 e pelo Lema 9.1.12,

$$h_\mu(f) = h_\mu(f, \mathcal{P}) = \lim_n H_\mu(\mathcal{P}/\mathcal{Q}_n). \quad (9.7.4)$$

Por definição (como anteriormente, $\phi(x) = -x \log x$)

$$\begin{aligned} H_\mu(\mathcal{P}/\mathcal{Q}_n) &= \sum_{P \in \mathcal{P}} \sum_{Q_n \in \mathcal{Q}_n} -\mu(P \cap Q_n) \log \frac{\mu(P \cap Q_n)}{\mu(Q_n)} \\ &= \sum_{P \in \mathcal{P}} \sum_{Q_n \in \mathcal{Q}_n} \mu(Q_n) \phi\left(\frac{\mu(P \cap Q_n)}{\mu(Q_n)}\right) \end{aligned} \quad (9.7.5)$$

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 (estas noções foram introduzidas na Seção 5.2.1: veja (5.2.1) e o Lema 5.2.1). 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,

$$\sum_{P \in \mathcal{P}} \sum_{Q_n \in \mathcal{Q}_n} \mu(Q_n) \phi\left(\frac{\mu(P \cap Q_n)}{\mu(Q_n)}\right) = \sum_{P \in \mathcal{P}} \int \phi(e_n(\mathcal{X}_P, x)) d\mu(x). \quad (9.7.6)$$

Pelo Lema 5.2.1, 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 (9.7.4) – (9.7.6) que

$$h_\mu(f) = \sum_{P \in \mathcal{P}} \int \phi(e(\mathcal{X}_P, x)) d\mu(x). \quad (9.7.7)$$

Resta relacionar o integrando do lado direito com o jacobiano. Isso será feito por meio do seguinte lema:

Lema 9.7.4. *Para toda função mensurável limitada $\psi : M \rightarrow \mathbb{R}$ e toda probabilidade boreliana η invariante por f ,*

$$e(\psi, x) = \hat{\psi}(f(x)) \quad \text{para } \eta\text{-quase todo } x, \quad \text{onde } \hat{\psi}(y) = \sum_{z \in f^{-1}(y)} \frac{\psi}{J_\eta f}(z).$$

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,

$$\int_{\mathcal{P}^{n-1}(f(x))} \hat{\psi} d\eta = \sum_{P \in \mathcal{P}} \int_{f(P) \cap \mathcal{P}^{n-1}(f(x))} \frac{\psi}{J_\eta f} \circ (f|_P)^{-1} d\eta.$$

Usando a fórmula de mudança de variáveis dada no Exercício 9.7.3(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,

$$\int_{\mathcal{P}^{n-1}(f(x))} \hat{\psi} d\eta = \int_{\mathcal{Q}_n(x)} \psi d\eta. \quad (9.7.8)$$

A hipótese de que η é invariante dá que $\eta(\mathcal{P}^{n-1}(f(x))) = \eta(\mathcal{Q}_n(x))$. Dividindo ambos os lados de (9.7.8) por este número, obtemos que

$$e_n(\psi, x) = e'_{n-1}(\hat{\psi}, f(x)) \quad \text{para todo } x \text{ e todo } n > 1. \quad (9.7.9)$$

Então, passando ao limite, $e(\psi, x) = e'(\hat{\psi}, f(x))$ para η -quase x . Por outro lado, de acordo com o Exercício 5.2.3, a hipótese implica que $e'(\hat{\psi}, y) = \hat{\psi}(y)$ para η -quase todo $y \in M$. \square

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 9.7.4 que $e(\mathcal{X}_P, x) = \hat{\mathcal{X}}_P(f(x))$, com

$$\hat{\mathcal{X}}_P(y) = \begin{cases} 1/J_\mu f((f|P)^{-1}(y)) & \text{se } y \in f(P) \\ 0 & \text{se } y \notin f(P). \end{cases}$$

Então, lembrando que a medida μ é invariante,

$$\begin{aligned} \int \phi(e(\mathcal{X}_P, x)) d\mu(x) &= \int \phi(\hat{\mathcal{X}}_P(y)) d\mu(y) \\ &= \int_{f(P)} \left(\frac{1}{J_\mu f} \log J_\mu f \right) \circ (f|P)^{-1} d\mu = \int_P \log J_\mu f d\mu \end{aligned}$$

(a última igualdade usa a igualdade (b) no Exercício 9.7.3). Substituindo esta expressão em (9.7.7), 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. \square

9.7.1 Exercícios

9.7.1. Verifique que a definição de jacobiano não depende da escolha da cobertura $\{U_k : k \geq 1\}$ por domínios de invertibilidade.

9.7.2. Seja $\sigma : \Sigma \rightarrow \Sigma$ 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 .

9.7.3. Seja $f : M \rightarrow M$ uma transformação localmente invertível e η uma probabilidade boreliana em M não singular com relação a f . Mostre que valem as seguintes fórmulas de mudança de variáveis:

- (a) $\int_{f(A)} \varphi d\eta = \int_A (\varphi \circ f) J_\eta f d\eta$ para todo domínio de invertibilidade $A \subset M$ e toda função mensurável $\varphi : f(A) \rightarrow \mathbb{R}$ tal que as integrais estão definidas (podendo ser $\pm\infty$).
- (b) $\int_A \psi d\eta = \int_{f(A)} (\psi / J_\eta f) \circ (f|_A)^{-1} d\eta$ para qualquer função mensurável $\psi : A \rightarrow \mathbb{R}$ tal que as integrais estão definidas (podendo ser $\pm\infty$).

9.7.4. Seja $f : M \rightarrow M$ uma transformação localmente invertível e seja η uma probabilidade boreliana em M não singular com relação a f . Mostre que para toda função mensurável limitada $\psi : M \rightarrow \mathbb{R}$,

$$\int \psi d\eta = \int \sum_{z \in f^{-1}(x)} \frac{\psi}{J_\eta f}(z) d\eta(x).$$

9.7.5. Seja $f : M \rightarrow M$ uma transformação localmente invertível e seja η uma probabilidade boreliana 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.

9.7.6. Seja $f : M \rightarrow M$ 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

$$J_\eta f^k(x) = \prod_{j=0}^{k-1} J_\eta f(f^j(x)) \quad \text{para } \eta\text{-quase todo } x.$$

Supondo que f é invertível, o que pode ser dito a respeito do jacobiano de f^{-1} relativamente a η ?

9.7.7. Sejam $f : M \rightarrow M$ e $g : N \rightarrow N$ transformações localmente invertíveis em espaços métricos e sejam μ e ν medidas invariantes por f e g , respectivamente. Suponha que existe uma equivalência ergódica $\phi : M \rightarrow N$ entre os sistemas (f, μ) e (g, ν) . Mostre que $J_\mu f = J_\nu g \circ \phi$ em μ -quase todo ponto.

9.7.8. Sejam $\sigma_k : \Sigma_k \rightarrow \Sigma_k$ e $\sigma_l : \Sigma_l \rightarrow \Sigma_l$ 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 se 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.

Capítulo 10

Princípio variacional

Em 1965, os pesquisadores R. Adler, A. Konheim e M. McAndrew da IBM propuseram [AKM65] uma noção de *entropia topológica*, inspirada na entropia de Kolmogorov-Sinai que estudamos no capítulo anterior, mas cuja definição não envolve qualquer medida invariante. Esta noção se aplica a qualquer transformação contínua num espaço topológico compacto.

Posteriormente, Efim Dinaburg [Din70] e Rufus Bowen [Bow71, Bow75a] deram uma definição diferente, mas equivalente, para transformações contínuas em espaços métricos compactos. Apesar de ser um pouco mais restrita, ela tem a vantagem de tornar mais transparente o significado deste conceito: a entropia topológica é a taxa de crescimento exponencial do número de órbitas que são distinguíveis dentro de um certo grau de precisão, arbitrariamente pequeno. Além disso, Bowen estendeu a definição para espaços não compactos, o que também é muito útil em aplicações.

Estas definições da entropia topológica e suas propriedades serão estudadas na Seção 10.1 onde, em particular, também veremos que a entropia topológica é um invariante de equivalência (conjugação) topológica. Na Seção 10.2 analisaremos diversos exemplos concretos.

A principal dessas propriedades é a seguinte relação notável entre a entropia topológica e as entropias da transformação relativamente às suas medidas invariantes:

Teorema 10.1 (Princípio Variacional). *Se $f : M \rightarrow M$ uma transformação contínua num espaço métrico compacto então a sua entropia topológica $h(f)$ coincide com o supremo das entropias $h_\mu(f)$ da transformação f relativamente a todas as probabilidades invariantes.*

Este fato foi demonstrado originalmente por Dinaburg [Din70, Din71], Goodman [Goo71a] e Goodwin [Goo71b]. Adiamos a apresentação da demonstração para o próximo capítulo, no qual iremos obter um resultado mais abrangente, o princípio variacional para a pressão, que é devido a Walters [Wal75].

A *pressão* $P(f, \phi)$ é uma versão ponderada da entropia topológica $h(f)$, onde os “pesos” são determinados por uma função contínua $\phi : M \rightarrow \mathbb{R}$, que é cha-

mada de *potencial*. Estudaremos estas noções e suas propriedades na Seção 10.3. A entropia topológica, que foi introduzida no capítulo anterior, corresponde ao caso particular em que o potencial é identicamente nulo. A ideia de pressão foi trazida da Mecânica Estatística para a Teoria Ergódica pelo matemático e físico teórico David Ruelle, um dos criadores da teoria ergódica diferenciável, e foi depois estendida pelo matemático britânico Peter Walters.

O princípio variacional (Teorema 10.1) se generaliza para o contexto da pressão, como veremos na Seção 10.4: para toda função contínua ϕ ,

$$P(f, \phi) = \sup \left\{ h_\mu(f) + \int \phi d\mu : \mu \text{ é invariante por } f \right\}. \quad (10.0.1)$$

Uma probabilidade invariante μ é chamada *estado de equilíbrio* para o potencial ϕ se ela realiza o supremo em (10.0.1), ou seja, se $h_\mu(f) + \int \phi d\mu = P(f, \phi)$. As propriedades do conjunto dos estados de equilíbrio serão estudadas na Seção 10.5.

10.1 Entropia topológica

Inicialmente, apresentaremos as definições de Adler-Konheim-McAndrew e de Bowen-Dinaburg e provaremos que elas são equivalentes quando o ambiente é um espaço métrico compacto.

10.1.1 Definição via coberturas abertas

A definição original da entropia topológica é muito semelhante à definição da entropia de Kolmogorov-Sinai, com coberturas abertas no lugar de partições em conjunto mensuráveis.

Seja M um espaço topológico compacto. Chamamos *cobertura aberta* de M qualquer família α de abertos cuja união é todo o M . Por compacidade, toda cobertura aberta admite uma *subcobertura* (isto é, uma subfamília que ainda é uma cobertura) com um número finito de elementos. Chamamos *entropia* da cobertura α ao número

$$H(\alpha) = \log N(\alpha), \quad (10.1.1)$$

onde $N(\alpha)$ é o menor número tal que α admite alguma subcobertura finita com esse número de elementos.

Dadas duas coberturas abertas α e β , dizemos que α é *menos fina* que β , e escrevemos $\alpha \prec \beta$, se todo elemento de β está contido em algum elemento de α . Por exemplo, se β é subcobertura de α então $\alpha \prec \beta$. Pelo Exercício 10.1.1,

$$\alpha \prec \beta \quad \Rightarrow \quad H(\alpha) \leq H(\beta). \quad (10.1.2)$$

Dadas coberturas $\alpha_1, \dots, \alpha_n$, denotamos por $\alpha_1 \vee \dots \vee \alpha_n$ a sua *soma*, isto é, a cobertura cujos elementos são as interseções $A_1 \cap \dots \cap A_n$ com $A_j \in \alpha_j$ para cada j . Note que $\alpha_j \prec \alpha_1 \vee \dots \vee \alpha_n$ para todo j .

Seja $f : M \rightarrow M$ uma transformação contínua. Se α é uma cobertura aberta de M então $f^{-j}(\alpha) = \{f^{-j}(A) : A \in \alpha\}$ também é uma cobertura aberta. Para cada $n \geq 1$, denotamos

$$\alpha^n = \alpha \vee f^{-1}(\alpha) \vee \dots \vee f^{-n+1}(\alpha).$$

Usando o Exercício 10.1.2, vemos que

$$H(\alpha^{m+n}) = H(\alpha^m \vee f^{-m}(\alpha^n)) \leq H(\alpha^m) + H(f^{-m}(\alpha^n)) \leq H(\alpha^m) + H(\alpha^n)$$

para todo $m, n \geq 1$. Em outras palavras, a sequência $H(\alpha^n)$ é subaditiva. Consequentemente (Lema 3.3.4),

$$h(f, \alpha) = \lim_n \frac{1}{n} H(\alpha^n) = \inf_n \frac{1}{n} H(\alpha^n) \quad (10.1.3)$$

sempre existe. Ele é chamado entropia de f com respeito à cobertura α . A relação (10.1.2) implica que

$$\alpha \prec \beta \quad \Rightarrow \quad h(f, \alpha) \leq h(f, \beta). \quad (10.1.4)$$

Finalmente, definimos a *entropia topológica* de f como sendo

$$h(f) = \sup\{h(f, \alpha) : \alpha \text{ é cobertura aberta de } M\}. \quad (10.1.5)$$

Em particular, se β é subcobertura de α então $h(f, \alpha) \leq h(f, \beta)$. Portanto, a definição (10.1.5) não muda se restringirmos o supremo às coberturas abertas finitas.

Observe que a entropia $h(f)$ é um número não negativo, podendo ser infinito (veja o Exercício 10.1.5).

Exemplo 10.1.1. Seja $f : S^1 \rightarrow S^1$ um homeomorfismo qualquer (por exemplo, uma rotação R_θ) e seja α uma cobertura do círculo formada por um número finito de intervalos abertos. Seja $\partial\alpha$ o conjunto formado pelos pontos extremos desses intervalos. Para cada $n \geq 1$, a cobertura α^n está formada por intervalos, cujos pontos extremos estão em

$$\partial\alpha^n = \partial\alpha \cup f^{-1}(\partial\alpha) \cup \dots \cup f^{-n+1}(\partial\alpha).$$

Note que $\#\alpha^n \leq \#\partial\alpha^n \leq n\#\partial\alpha$. Portanto,

$$h(f, \alpha) = \lim_n \frac{1}{n} H(\alpha^n) \leq \liminf_n \frac{1}{n} \log \#\alpha^n \leq \liminf_n \frac{1}{n} \log n = 0.$$

Veremos na Proposição 10.1.9 que $h(f) = \lim_k h(f, \alpha_k)$ para qualquer sequência de coberturas abertas α_k com $\text{diam } \alpha_k \rightarrow 0$. Então, considerando coberturas abertas por intervalos de comprimento menor que $1/k$, concluímos do cálculo anterior que $h(f) = 0$ para todo homeomorfismo do círculo.

Exemplo 10.1.2. Seja $\Sigma = \{1, \dots, d\}^{\mathbb{N}}$ e seja α a cobertura de Σ pelos cilindros $[0; a]$, $a = 1, \dots, d$. Considere o deslocamento $\sigma : \Sigma \rightarrow \Sigma$. Para cada n , a cobertura α^n está formada pelos cilindros de comprimento n :

$$\alpha^n = \{[0; a_0, \dots, a_{n-1}] : a_j = 1, \dots, d\}.$$

Portanto, $H(\alpha^n) = \log \#\alpha^n = \log d^n$ e, conseqüentemente, $h(f, \alpha) = \log d$. Observe também que $\text{diam } \alpha^n$ converge para zero quando $n \rightarrow \infty$ (relativamente à distância definida em (A.2.7)). Segue do Corolário 10.1.10, que provaremos daqui a pouco, que $h(f) = h(f, \alpha) = \log d$. O mesmo vale para o deslocamento bilateral $\sigma : \Sigma \rightarrow \Sigma$ em $\Sigma = \{1, \dots, d\}^{\mathbb{Z}}$.

Agora vamos mostrar que a entropia topológica é um invariante de equivalência topológica. Sejam $f : M \rightarrow M$ e $g : N \rightarrow N$ transformações contínuas em espaço topológicos compactos M e N . Dizemos que g é um *fator topológico* de f se existe uma aplicação contínua sobrejetiva $h : M \rightarrow N$ satisfazendo $h \circ f = g \circ h$. Se h pode ser escolhida invertível (homeomorfismo), dizemos que as duas transformações são *topologicamente equivalentes*, ou *topologicamente conjugadas*, e chamamos h de *conjugação topológica* entre f e g .

Proposição 10.1.3. *Se g é um fator topológico de f então $h(g) \leq h(f)$. Em particular, se f e g são topologicamente equivalentes então $h(f) = h(g)$.*

Demonstração. Seja $\theta : M \rightarrow N$ uma aplicação contínua sobrejetiva tal que $\theta \circ f = g \circ \theta$. Dada uma cobertura α de N , a família

$$\theta^{-1}(\alpha) = \{\theta^{-1}(A) : A \in \alpha\}$$

é cobertura aberta de M . Dados quaisquer conjuntos $A_0, A_1, \dots, A_{n-1} \in \alpha$, temos que:

$$\theta^{-1}\left(\bigcap_{j=0}^{n-1} g^{-j}(A_j)\right) = \bigcap_{j=0}^{n-1} \theta^{-1}(g^{-j}(A_j)) = \bigcap_{j=0}^{n-1} f^{-j}(\theta^{-1}(A_j)).$$

Por definição, $\theta^{-1}(\alpha^n)$ está formada pelos conjuntos no lado esquerdo desta igualdade, enquanto que os conjuntos no lado direito constituem $\theta^{-1}(\alpha)^n$. Portanto, $\theta^{-1}(\alpha^n) = \theta^{-1}(\alpha)^n$. Como θ é sobrejetivo, uma família $\gamma \subset \alpha^n$ cobre N se, e somente se, $\theta^{-1}(\gamma)$ cobre M . Portanto, $\theta(\theta^{-1}(\alpha)^n) = \theta(\theta^{-1}(\alpha^n)) = \theta(\alpha^n)$. Como n é arbitrário, segue $\theta(f, \theta^{-1}(\alpha)) = \theta(g, \alpha)$. Então, tomando o supremo sobre todas as coberturas α de N :

$$h(g) = \sup_{\alpha} h(g, \alpha) = \sup_{\alpha} h(f, \theta^{-1}(\alpha)) \leq h(f).$$

Isto prova a primeira parte da proposição. A segunda parte é uma consequência imediata, uma vez que nesse caso f também é um fator de g . \square

A recíproca da Proposição 10.1.3 é falsa, em geral. Por exemplo, todos os homeomorfismos do círculo têm entropia topológica igual a zero (lembre do Exemplo 10.1.1) mas eles não são necessariamente topologicamente equivalentes (por exemplo, a identidade e a rotação R_θ com θ diferente de zero).

10.1.2 Conjuntos geradores e conjuntos separados

A seguir apresentamos a definição de entropia topológica de Bowen-Dinaburg. Seja $f : M \rightarrow M$ uma transformação contínua num espaço métrico M , não necessariamente compacto, e seja $K \subset M$ um subconjunto compacto qualquer. Quando M é compacto basta considerar $K = M$, como veremos em (10.1.12).

Dados $\varepsilon > 0$ e $n \in \mathbb{N}$, dizemos que um conjunto $E \subset M$ é um (n, ε) -gerador de K , se para todo $x \in K$ existe $a \in E$ tal que $d(f^i(x), f^i(y)) < \varepsilon$ para todo $i \in \{0, \dots, n-1\}$. Em outras palavras,

$$K \subset \bigcup_{a \in E} B(a, n, \varepsilon),$$

onde $B(a, n, \varepsilon) = \{x \in M : d(f^i(x), f^i(y)) < \varepsilon \text{ para } i = 0, \dots, n-1\}$ é a *bola dinâmica* de centro a , comprimento n e raio ε . Note que $\{B(x, n, \varepsilon) : x \in K\}$ é uma cobertura aberta de K . Logo, por compacidade, sempre existem conjuntos (n, ε) -geradores finitos. Denotamos por $g_n(f, \varepsilon, K)$ a menor cardinalidade de um conjunto (n, ε) -gerador de K . Definimos

$$g(f, \varepsilon, K) = \limsup_n \frac{1}{n} \log g_n(f, \varepsilon, K). \quad (10.1.6)$$

Observe que a função $\varepsilon \mapsto g(f, \varepsilon, K)$ é monótona não crescente. De fato, é claro da definição que se $\varepsilon_1 < \varepsilon_2$ então todo conjunto (n, ε_1) -gerador também é (n, ε_2) -gerador. Portanto, $g_n(f, \varepsilon_1, K) \geq g_n(f, \varepsilon_2, K)$ para todo $n \geq 1$ e, passando ao limite, $g(f, \varepsilon_1, K) \geq g(f, \varepsilon_2, K)$. Isto garante, em particular, que

$$g(f, K) = \lim_{\varepsilon \rightarrow 0} g(f, \varepsilon, K) \quad (10.1.7)$$

existe. Finalmente, definimos

$$g(f) = \sup\{g(f, K) : K \subset M \text{ compacto}\}. \quad (10.1.8)$$

Também introduzimos a seguinte noção dual. Dados $\varepsilon > 0$ e $n \in \mathbb{N}$, dizemos que um conjunto $E \subset K$ é (n, ε) -separado se dados $x, y \in E$, existe $j \in \{0, \dots, n-1\}$ tal que $d(f^j(x), f^j(y)) \geq \varepsilon$. Em outras palavras, se $x \in E$ então $B(x, n, \varepsilon)$ não contém nenhum outro ponto de E . Denotamos por $s_n(f, \varepsilon, K)$ a máxima cardinalidade de um conjunto (n, ε) -separado. Definimos

$$s(f, \varepsilon, K) = \limsup_n \frac{1}{n} \log s_n(f, \varepsilon, K). \quad (10.1.9)$$

É claro que se $0 < \varepsilon_1 < \varepsilon_2$, então todo conjunto (n, ε_2) -separado também é (n, ε_1) -separado. Portanto, $s_n(f, \varepsilon_1, K) \geq s_n(f, \varepsilon_2, K)$ para todo $n \geq 1$ e, passando ao limite, $s(f, \varepsilon_1, K) \geq s(f, \varepsilon_2, K)$. Em particular,

$$s(f, K) = \lim_{\varepsilon \rightarrow 0} s(f, \varepsilon, K) \quad (10.1.10)$$

sempre existe. Finalmente, definimos

$$s(f) = \sup\{s(f, K) : K \subset M \text{ compacto}\}. \quad (10.1.11)$$

É claro que $g(f, K_1) \leq g(f, K_2)$ e $s(f, K_1) \leq s(f, K_2)$ se $K_1 \subset K_2$. Em particular:

$$g(f) = g(f, M) \quad \text{e} \quad s(f) = s(f, M) \quad \text{se } M \text{ é compacto.} \quad (10.1.12)$$

Outra observação interessante é que (Exercício 10.1.7) as definições (10.1.8) e (10.1.11) não são afetadas quando restringimos os supremos a conjuntos compactos com diâmetro pequeno.

Proposição 10.1.4. *Tem-se $g(f, K) = s(f, K)$ para todo compacto $K \subset M$. Consequentemente, $g(f) = s(f)$.*

Demonstração. Precisamos do seguinte lema:

Lema 10.1.5. *$g_n(f, \varepsilon, K) \leq s_n(f, \varepsilon, K) \leq g_n(f, \varepsilon/2, K)$ para todo $n \geq 1$, todo $\varepsilon > 0$ e todo compacto $K \subset M$.*

Demonstração. Inicialmente, seja $E \subset K$ um conjunto (n, ε) -separado com cardinalidade máxima. Então, dado qualquer $y \in K \setminus E$, temos que $E \cup \{y\}$ não é (n, ε) -separado. Portanto, existe $x \in E$ e existe $i \in \{0, \dots, n-1\}$ tal que $d(f^i(x), f^i(y)) < \varepsilon$. Isto prova que E é um conjunto (n, ε) -gerador de K e, por sua vez, isso implica que $g_n(f, \varepsilon, K) \leq \#E = s_n(f, \varepsilon, K)$.

Para provar a outra desigualdade, seja $E \subset K$ um conjunto (n, ε) -separado e seja $F \subset M$ um conjunto $(n, \varepsilon/2)$ -gerador de K . A hipótese garante que, dado qualquer $x \in E$ existe algum ponto $y \in F$ tal que $d(f^i(x), f^i(y)) < \varepsilon/2$ para todo $i \in \{0, \dots, n-1\}$. Defina uma aplicação $\phi : E \rightarrow F$ considerando $\phi(x)$ como sendo qualquer ponto y nessas condições. Afirmamos que a aplicação ϕ é injetiva. De fato, suponha que $x, z \in E$ são tais que $\phi(x) = y = \phi(z)$. Então,

$$d(f^i(x), f^i(z)) \leq d(f^i(x), f^i(y)) + d(f^i(y), f^i(z)) < \varepsilon/2 + \varepsilon/2$$

para todo $i \in \{0, \dots, n-1\}$. Como E é (n, ε) -separado, isso implica que $x = z$. Portanto, ϕ é injetiva, tal como afirmamos. Segue que $\#E \leq \#F$. Como E e F são arbitrários, isto prova que $s_n(f, \varepsilon, K) \leq g_n(f, \varepsilon/2, K)$. \square

Então, dado qualquer $\varepsilon > 0$ e qualquer compact $K \subset M$,

$$\begin{aligned} g(f, \varepsilon, K) &= \limsup_n \frac{1}{n} \log g_n(f, \varepsilon, K) \leq \limsup_n \frac{1}{n} \log s_n(f, \varepsilon, K) \\ &\leq \limsup_n \frac{1}{n} \log g_n(f, \varepsilon/2, K) = g(f, \varepsilon/2, K). \end{aligned}$$

Tomando o limite quando $\varepsilon \rightarrow 0$, obtemos

$$g(f, K) = \lim_{\varepsilon \rightarrow 0} g(f, \varepsilon, K) \leq \lim_{\varepsilon \rightarrow 0} s(f, \varepsilon, K) = s(f) \leq \lim_{\varepsilon \rightarrow 0} g(f, \varepsilon/2, K) = g(f, K).$$

Isto prova a primeira parte da proposição. A segunda parte é uma consequência imediata. \square

Proposição 10.1.6. *Se M é espaço métrico compacto, $h(f) = g(f) = s(f)$.*

Demonstração. Pela Proposição 10.1.4, basta mostrar que $s(f) \leq h(f) \leq g(f)$.

Comece por fixar $\varepsilon > 0$ e $n \geq 1$. Seja $E \subset M$ um subconjunto (n, ε) -separado e seja α qualquer cobertura aberta de M com diâmetro menor que ε . Se x e y estão no mesmo elemento de α^n então

$$d(f^i(x), f^i(y)) \leq \text{diam } \alpha < \varepsilon \quad \text{para todo } i = 0, \dots, n-1.$$

Em particular, cada elemento de α^n contém no máximo um elemento de E e, portanto, $\#E \leq N(\alpha^n)$. Tomando E com cardinalidade maximal, concluímos que $s_n(f, \varepsilon, M) \leq N(\alpha^n)$ para todo $n \geq 1$. Consequentemente,

$$s(f, \varepsilon, M) = \limsup_n \frac{1}{n} \log s_n(f, \varepsilon, M) \leq \lim_n \frac{1}{n} \log N(\alpha^n) = h(f, \alpha) \leq h(f).$$

Fazendo $\varepsilon \rightarrow 0$ obtemos que $s(f) = s(f, M) \leq h(f)$.

Em seguida, dada qualquer cobertura aberta α de M , seja $\varepsilon > 0$ um número de Lebesgue para α , ou seja, um número positivo tal que toda bola de raio ε está contida em algum elemento de α . Seja $E \subset M$ um conjunto (n, ε) -gerador de M com cardinalidade minimal. Para cada $x \in E$ e $i = 0, \dots, n-1$, existe $A_{x,i} \in \alpha$ tal que $B(f^i(x), \varepsilon)$ está contida em $A_{x,i}$. Então

$$B(x, n, \varepsilon) \subset \bigcap_{i=0}^{n-1} f^{-i}(A_{x,i}).$$

Então, a hipótese de que E é gerador implica que $\gamma = \{\bigcap_{i=0}^{n-1} f^{-i}(A_{x,i}) : x \in E\}$ é uma cobertura de M . Como $\gamma \subset \alpha^n$, segue que $N(\alpha^n) \leq \#E = g_n(f, \varepsilon, M)$ para todo n . Portanto,

$$\begin{aligned} h(f, \alpha) &= \lim_n \frac{1}{n} \log N(\alpha^n) \leq \liminf_n \frac{1}{n} \log g_n(f, \varepsilon, M) \\ &\leq \limsup_n \frac{1}{n} \log g_n(f, \varepsilon, M) = g(f, \varepsilon, M). \end{aligned} \tag{10.1.13}$$

Fazendo $\varepsilon \rightarrow 0$, vem que $h(f, \alpha) \leq g(f, M) = g(f)$. Como a cobertura α é arbitrária, segue que $h(f) \leq g(f)$. \square

Definimos a *entropia topológica* de uma transformação contínua $f : M \rightarrow M$ num espaço métrico M como sendo $g(f) = s(f)$. A Proposição 10.1.6 mostra que esta definição é compatível com aquela que demos na Seção 10.1.1 para transformações em espaços topológicos compactos. Uma diferença relevante é que, enquanto no caso compacto a entropia topológica depende apenas da topologia (porque $h(f)$ é definida apenas em termos dos abertos), no caso não compacto a entropia topológica pode depender também da função distância em M . A este respeito, veja os Exercícios 10.1.3 e 10.1.4. Eles também mostram que, no caso não compacto, a entropia topológica é um invariante de conjugação *uniformemente* contínua mas não necessariamente de conjugação topológica.

Exemplo 10.1.7. Suponha que $f : M \rightarrow M$ não expande distâncias, ou seja, que $d(f(x), f(y)) \leq d(x, y)$ para todo $x, y \in M$. Então a entropia topológica de f é igual a zero. De fato, a hipótese implica que $B(x, n, \varepsilon) = B(x, \varepsilon)$ para todo $n \geq 1$. Logo, um conjunto E é (n, ε) -gerador se, e somente se, ele é $(1, \varepsilon)$ -gerador. Em particular, a sequência $g_n(f, \varepsilon, K)$ não depende de n e, portanto, $g(f, \varepsilon, K) = 0$ para todo $\varepsilon > 0$ e todo compacto K . Fazendo $\varepsilon \rightarrow 0$ e tomando o supremo sobre K , obtemos que $g(f) = 0$. Analogamente, $s(f) = 0$.

Destacam-se dois casos particulares importantes: *contrações*, tais que existe $\lambda < 1$ satisfazendo $d(f(x), f(y)) \leq \lambda d(x, y)$ para todo $x, y \in M$; e *isometrias*, tais que $d(f(x), f(y)) = d(x, y)$ para todo $x, y \in M$. Vimos no Lema 6.3.6 que todo grupo compacto metrizável admite uma distância relativamente à qual toda translação é uma isometria. Portanto, também segue das observações anteriores que toda translação num grupo topológico metrizável tem entropia topológica igual a zero.

Lembrando que $g(f) = g(f, M)$ e $s(f) = s(f, M)$, vemos que a conclusão da Proposição 10.1.6 pode ser reescrita do seguinte modo:

$$h(f) = \lim_{\varepsilon \rightarrow 0} \limsup_n \frac{1}{n} \log g_n(f, \varepsilon, M) = \lim_{\varepsilon \rightarrow 0} \limsup_n \frac{1}{n} \log s_n(f, \varepsilon, M).$$

Mas a partir da demonstração da proposição também podemos obter a seguinte igualdade relacionada:

Corolário 10.1.8. *Se $f : M \rightarrow M$ é uma transformação contínua num espaço métrico compacto então*

$$h(f) = \lim_{\varepsilon \rightarrow 0} \liminf_n \frac{1}{n} \log g_n(f, \varepsilon, M) = \lim_{\varepsilon \rightarrow 0} \liminf_n \frac{1}{n} \log s_n(f, \varepsilon, M).$$

Demonstração. A relação (10.1.13) dá que

$$h(f, \alpha) \leq \liminf_n \frac{1}{n} \log g_n(f, \varepsilon, M)$$

sempre que $\varepsilon > 0$ é um número de Lebesgue para a cobertura α . Fazendo $\varepsilon \rightarrow 0$, concluímos que

$$h(f) \leq \lim_{\varepsilon \rightarrow 0} \liminf_n \frac{1}{n} \log g_n(f, \varepsilon, M). \quad (10.1.14)$$

A primeira desigualdade no Lema 10.1.5 implica que

$$\lim_{\varepsilon \rightarrow 0} \liminf_n \frac{1}{n} \log g_n(f, \varepsilon, M) \leq \lim_{\varepsilon \rightarrow 0} \liminf_n \frac{1}{n} \log s_n(f, \varepsilon, M). \quad (10.1.15)$$

É claro que

$$\lim_{\varepsilon \rightarrow 0} \liminf_n \frac{1}{n} \log s_n(f, \varepsilon, M) \leq \lim_{\varepsilon \rightarrow 0} \limsup_n \frac{1}{n} \log s_n(f, \varepsilon, M). \quad (10.1.16)$$

Como acabamos de observar, a expressão do lado direito é igual a $h(f)$. Portanto, as desigualdades (10.1.14)-(10.1.16) implicam a conclusão. \square

10.1.3 Cálculo e propriedades

A próxima proposição e o seu corolário simplificam substancialmente o cálculo da entropia topológica em exemplos concretos. Quando M é um espaço métrico, chamamos *diâmetro* de uma cobertura aberta ao supremo dos diâmetros dos seus elementos.

Proposição 10.1.9. *Suponha que M é um espaço métrico compacto. Seja $(\beta_k)_k$ qualquer sequência de coberturas abertas de M tal que $\text{diam } \beta_k$ converge para zero. Então,*

$$h(f) = \sup_k h(f, \beta_k) = \lim_k h(f, \beta_k).$$

Demonstração. Dada qualquer cobertura aberta α , seja $\varepsilon > 0$ um número de Lebesgue de α . Tome $n \geq 1$ tal que $\text{diam } \beta_k < \varepsilon$ para todo $k \geq n$. Pela definição de número de Lebesgue, segue que todo elemento de β_k está contido em algum elemento de α . Em outras palavras, $\alpha \prec \beta_k$ e, portanto, $h(f, \beta_k) \geq h(f, \alpha)$. Lembrando a definição (10.1.5), isto prova que

$$\liminf_k h(f, \beta_k) \geq h(f).$$

Também é claro das definições que $h(f) \geq \sup_k h(f, \beta_k) \geq \limsup_k h(f, \beta_k)$. Combinando estas duas observações obtemos a conclusão do lema. \square

Corolário 10.1.10. *Suponha M é um espaço métrico compacto. Se β é cobertura aberta tal que*

(1) *o diâmetro de $\beta^k = \bigvee_{j=0}^{k-1} f^{-j}(\beta)$ converge para zero quando $k \rightarrow \infty$, ou*

(2) *$f : M \rightarrow M$ é um homeomorfismo e o diâmetro de $\beta^{\pm k} = \bigvee_{j=-k}^{k-1} f^{-j}(\beta)$ converge para zero quando $k \rightarrow \infty$,*

então $h(f) = h(f, \beta)$.

Demonstração. No caso (1), a Proposição 10.1.9 e o Exercício 10.1.6 dão:

$$h(f) = \lim_k h(f, \beta^k) = h(f, \beta).$$

A prova no caso (2) é análoga. \square

A seguir vamos verificar que a entropia topológica se comporta como seria de se esperar relativamente a iterados positivos, pelo menos quando a transformação é uniformemente contínua:

Proposição 10.1.11. *Se $f : M \rightarrow M$ é uma transformação uniformemente contínua num espaço métrico, então $h(f^k) = kh(f)$ para todo $k \in \mathbb{N}$.*

Demonstração. Fixe $k \geq 1$ e seja $K \subset M$ um conjunto compacto qualquer. Considere quaisquer $n \geq 1$ e $\varepsilon > 0$. É claro que se $E \subset M$ é (nk, ε) -gerador

de K para a transformação f então ele também é (n, ε) -gerador de K para o iterado f^k . Portanto, $g_n(f^k, \varepsilon, K) \leq g_{nk}(f, \varepsilon, K)$. Logo,

$$g(f^k, \varepsilon, K) = \lim_n \frac{1}{n} g_n(f^k, \varepsilon, K) \leq \lim_n \frac{1}{n} g_{nk}(f, \varepsilon, K) = kg(f, \varepsilon, K).$$

Fazendo $\varepsilon \rightarrow 0$ e tomando o supremo sobre K , vemos que $h(f^k) \leq kh(f)$.

A prova da outra desigualdade usa a hipótese de que f é uniformemente contínua. Tome $\delta > 0$ tal que $d(x, y) < \delta$ implica $d(f^j(x), f^j(y)) < \varepsilon$ para todo $j \in \{0, \dots, k-1\}$. Se $E \subset M$ é (n, δ) -gerador de K para f^k então E é (nk, ε) -gerador de K para f . Portanto, $g_{nk}(f, \varepsilon, K) \leq g_n(f^k, \delta, K)$. Isto implica que $kg(f, \varepsilon, K) \leq g(f, \delta, K)$. Fazendo ε e δ ir para zero, obtemos que $kg(f, K) \leq g(f^k, K)$ para todo compacto K . Logo $kh(f) \leq h(f^k)$. \square

Em particular, a Proposição 10.1.11 vale para toda transformação contínua num espaço métrico compacto. Por outro lado, no caso de homeomorfismos em espaços compactos a conclusão se estende aos iterados negativos:

Proposição 10.1.12. *Se M é um espaço métrico compacto e $f : M \rightarrow M$ é um homeomorfismo então $h(f^{-1}) = h(f)$. Consequentemente, $h(f^n) = |n|h(f)$ para todo $n \in \mathbb{Z}$.*

Demonstração. Seja α uma cobertura aberta de M . Para todo $n \geq 1$, denotemos

$$\alpha_+^n = \alpha \vee f^{-1}(\alpha) \vee \dots \vee f^{-n+1}(\alpha) \quad \text{e} \quad \alpha_-^n = \alpha \vee f(\alpha) \vee \dots \vee f^{n-1}(\alpha)$$

Observe que $\alpha_-^n = f^{n-1}(\alpha_+^n)$. Mais ainda, γ é uma subcobertura finita de α_+^n se, e somente se, $f^{n-1}(\gamma)$ é uma subcobertura finita de α_-^n . Como as duas subcoberturas têm a mesma cardinalidade, segue que $H(\alpha_+^n) = H(\alpha_-^n)$. Portanto,

$$h(f, \alpha) = \lim_n \frac{1}{n} H(\alpha_+^n) = \lim_n \frac{1}{n} H(\alpha_-^n) = h(f^{-1}, \alpha).$$

Como α é arbitrária, isto prova que $h(f) = h(f^{-1})$. A segunda parte do enunciado segue desse fato e da Proposição 10.1.11. \square

A afirmação da Proposição 10.1.12 é falsa, em geral, no caso não compacto:

Exemplo 10.1.13. Considere $M = \mathbb{R}$ munido da distância usual e $f : \mathbb{R} \rightarrow \mathbb{R}$ dada por $f(x) = 2x$. Vamos verificar que $h(f) \neq h(f^{-1})$. Para isso, seja $K = [0, 1]$ e, dados $n \geq 1$ e $\varepsilon > 0$, suponha que $E \subset \mathbb{R}$ é conjunto (n, ε) -gerador para K . Em particular, todo ponto de $f^{n-1}(K) = [0, 2^{n-1}]$ está a distância menor que ε de algum ponto de $f^{n-1}(E)$. Logo,

$$2\varepsilon \# E = 2\varepsilon \# f^{n-1}(E) \geq 2^{n-1}.$$

Isto prova que $g_n(f, \varepsilon, K) \geq 2^{n-2}/\varepsilon$ para todo n e, portanto, $g(f, \varepsilon, K) \geq \log 2$. Segue que $h(f) \geq g(f, K) \geq \log 2$. Por outro lado, f^{-1} é uma contração e, portanto, segue do Exemplo 10.1.7 a sua entropia $h(f^{-1})$ é nula.

10.1.4 Exercícios

10.1.1. Seja M um espaço topológico compacto. Se α e β são coberturas abertas de M tais que $\alpha \prec \beta$, então $H(\alpha) \leq H(\beta)$.

10.1.2. Seja M um espaço topológico compacto. Se α e β são coberturas abertas de M e $f : M \rightarrow M$ é transformação contínua então $H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$ e $H(f^{-1}(\beta)) \leq H(\beta)$. Se f é sobrejetiva, vale a igualdade $H(f^{-1}(\beta)) = H(\beta)$.

10.1.3. Considere $M = (0, \infty)$ e $f : M \rightarrow M$ dada por $f(x) = 2x$. Calcule a entropia de f quando consideramos em M :

1. a distância usual $d(x, y) = |x - y|$;
2. a distância $d(x, y) = |\log x - \log y|$.

[Observação: Logo, em espaços não compactos a entropia topológica pode depender da função distância, e não apenas da topologia.]

10.1.4. Considere em M duas distâncias d_1 e d_2 que são uniformemente equivalentes: para todo $\varepsilon > 0$ existe $\delta > 0$ tal que

$$d_1(x, y) < \delta \Rightarrow d_2(x, y) < \varepsilon \quad \text{e} \quad d_2(x, y) < \delta \Rightarrow d_1(x, y) < \varepsilon.$$

Mostre que se $f : M \rightarrow M$ é contínua relativamente a qualquer das duas distâncias então o valor da entropia é o mesmo para as duas distâncias.

10.1.5. Sejam $f : M \rightarrow M$ e $g : N \rightarrow N$ transformações contínuas. Mostre que se existe uma aplicação injetiva contínua $\psi : N \rightarrow M$ tal que $\psi \circ g = f \circ \psi$ então $h(f) \geq h(g)$. Use esse fato para mostrar que a entropia topológica do deslocamento $\sigma : [0, 1]^{\mathbb{Z}} \rightarrow [0, 1]^{\mathbb{Z}}$, munido com a distância entre $x = (x_n)$ e $y = (y_n)$ definida por

$$d(x, y) = \sum_{n \in \mathbb{Z}} 2^{-|n|} |x_n - y_n|,$$

é infinita (apesar de que σ é um homeomorfismo num compacto).

10.1.6. Seja M um espaço métrico compacto e $f : M \rightarrow M$ uma função contínua. Mostre que $h(f, \alpha) = h(f, \alpha^k)$ para toda cobertura aberta α e para todo $k \geq 1$. Além disso, se $f : M \rightarrow M$ é um homeomorfismo então $h(f, \alpha) = h(f, \alpha^{\pm k})$ para todo $k \geq 1$, onde $\alpha^{\pm k} = \bigvee_{j=-k}^{k-1} f^{-j}(\alpha)$.

10.1.7. Mostre que se K, K_1, \dots, K_l são conjuntos compactos tais que K está contido em $K_1 \cup \dots \cup K_l$ então $g(f, K) \leq \max_j g(f, K_j)$. Conclua que, dado qualquer $\delta > 0$,

$$g(f) = \sup\{g(f, K) : K \text{ compacto com } \text{diam } K < \delta\}$$

e analogamente para $s(f)$.

10.1.8. Prove que a transformação *logística* $f : [0, 1] \rightarrow [0, 1]$, $f(x) = 4x(1 - x)$ é topologicamente conjugada à transformação $g : [0, 1] \rightarrow [0, 1]$ definida por $g(x) = 1 - |2x - 1|$. Use esse fato para calcular $h(f)$.

10.1.9. Seja \mathcal{A} um alfabeto finito e seja $\sigma : \Sigma \rightarrow \Sigma$ o deslocamento em $\Sigma = \mathcal{A}^{\mathbb{N}}$. A *complexidade* de um elemento $x \in \Sigma$ é definida por $c(x) = \lim_n n^{-1} \log c_n(x)$, onde $c_n(x)$ é o número de palavras distintas com comprimento n que aparecem em x . Mostre que este limite existe e que ele coincide com a entropia topológica da restrição $\sigma : \mathcal{X} \rightarrow \mathcal{X}$ do deslocamento ao fecho \mathcal{X} da órbita de x . [Observação: Uma aplicação importante é no contexto do Exemplo 6.3.10, onde x é o ponto fixo de uma substituição.]

10.1.10. Verifique que se θ é o ponto fixo da substituição de Fibonacci em $\mathcal{A} = \{0, 1\}$ então $c_n(\theta) = n + 1$ para todo n e, portanto, a complexidade $c(\theta)$ é igual a zero. Logo, a entropia do deslocamento $\sigma : \mathcal{X} \rightarrow \mathcal{X}$ associado à substituição de Fibonacci é nula.

10.2 Exemplos

Vamos ilustrar as ideias discutidas anteriormente por meio de alguns exemplos.

10.2.1 Transformações expansivas

Lembre (Seção 9.2.3) que uma transformação contínua $f : M \rightarrow M$ num espaço métrico compacto é dita *expansiva* se existe $\varepsilon_0 > 0$ tal que $d(f^j(x), f^j(y)) < \varepsilon_0$ para todo $j \in \mathbb{N}$ implica que $x = y$. Quando $f : M \rightarrow M$ é invertível, dizemos que ela é expansiva no sentido bilateral, se existe $\varepsilon_0 > 0$ tal que $d(f^j(x), f^j(y)) < \varepsilon_0$ para todo $j \in \mathbb{Z}$ implica que $x = y$. Nos dois casos, ε_0 é chamado *constante de expansividade* de f .

Proposição 10.2.1. *Se $\varepsilon_0 > 0$ é uma constante de expansividade para f então*

(a) $h(f) = h(f, \alpha)$ para toda cobertura aberta α com diâmetro menor que ε_0 ;

(b) $h(f) = g(f, \varepsilon, M) = s(f, \varepsilon, M)$ para todo $\varepsilon < \varepsilon_0/2$.

Em particular, $h(f) < \infty$.

Demonstração. Seja α qualquer cobertura aberta de M com $\text{diam } \alpha < \varepsilon_0$. Afir-mamos que $\lim_k \text{diam } \alpha^k = 0$. De fato, suponha que isso não é verdade. É claro que a sequência dos diâmetros é não crescente. Então existe $\delta > 0$ e para cada $k \geq 1$ existem pontos x_k e y_k num mesmo elemento de α^k tais que $d(x_k, y_k) \geq \delta$. Por compacidade, podemos escolher uma subsequência $(k_j)_j$ tal que existem $x = \lim_j x_{k_j}$ e $y = \lim_j y_{k_j}$. Observe que $x \neq y$, de fato $d(x, y) \geq \delta$. Por outro lado, a escolha de x_k e y_k num mesmo elemento de α^k implica que

$$d(f^i(x_k), f^i(y_k)) \leq \text{diam } \alpha \quad \text{para todo } 0 \leq i < k.$$

Passando ao limite, vem que $d(f^i(x), f^i(y)) \leq \text{diam } \alpha < \varepsilon_0$ para todo $i \geq 0$. Isto contradiz a hipótese de que ε_0 é uma constante de expansividade para f . Esta contradição prova a nossa afirmação. Usando a Proposição 10.1.9, segue que $h(f) = h(f, \alpha)$, tal como afirma o item (a).

Para provar o item (b), seja α a cobertura de M formada pelas bolas de raio ε . Note que α^n contém toda bola dinâmica $B(x, n, \varepsilon)$:

$$B(x, n, \varepsilon) = \bigcap_{j=0}^{n-1} f^{-j}(B(f^j(x), \varepsilon)) \quad \text{e cada } B(f^j(x), \varepsilon) \in \alpha.$$

Se E é um conjunto (n, ε) -gerador de M então $\{B(a, n, \varepsilon) : a \in E\}$ é cobertura aberta de M ; pelo que acabamos de dizer, trata-se de uma subcobertura de α . Portanto (lembre também do Lema 10.1.5),

$$N(\alpha^n) \leq g_n(f, \varepsilon, M) \leq s_n(f, \varepsilon, M) \quad \text{para todo } n.$$

Passando ao limite, obtemos que $h(f, \alpha) \leq g(f, \varepsilon, M) \leq s(f, \varepsilon, M)$. Lembre que $s(f, \varepsilon, M) \leq s(f, M) = h(f)$. Como $\text{diam } \alpha < \varepsilon_0$, a primeira parte da proposição dá que $h(f) = h(f, \alpha)$. Estas relações implicam o item (b).

A última afirmação na proposição é uma consequência imediata, uma vez que $g(f, \varepsilon, K)$, $s(f, \varepsilon, K)$ e $h(f, \alpha)$ são sempre finitos. \square

Veja no Exercício 10.2.7 uma extensão deste resultado para transformações *h-expansivas*, devida a Rufus Bowen [Bow72].

Em seguida, vamos provar que para transformações expansivas a entropia topológica é uma cota superior para a taxa de crescimento do número de pontos periódicos. Denotaremos por $\text{Fix}(f^n)$ o conjunto dos pontos $x \in M$ tais que $f^n(x) = x$.

Proposição 10.2.2. *Se M é um espaço métrico compacto e $f : M \rightarrow M$ é expansiva então*

$$\limsup_n \frac{1}{n} \log \# \text{Fix}(f^n) \leq h(f).$$

Demonstração. Seja α uma cobertura de M com $\text{diam } \alpha < \varepsilon_0$, onde ε_0 é uma constante de expansividade de f . Afirmamos que cada elemento de α^n contém no máximo um ponto de $\text{Fix}(f^n)$. De fato, se $x, y \in \text{Fix}(f^n)$ estão no mesmo elemento de α^n , então $d(f^i(x), f^i(y)) < \text{diam } \alpha < \varepsilon_0$ para todo $i = 0, \dots, n-1$. Como $f^n(x) = x$ e $f^n(y) = y$, segue que $d(f^i(x), f^i(y)) < \varepsilon_0$ para todo $i \geq 0$. Por expansividade, isso implica que $x = y$, o que prova a nossa afirmação. Segue que

$$\limsup_n \frac{1}{n} \log \# \text{Fix}(f^n) \leq \limsup_n \frac{1}{n} \log \# N(\alpha^n) = h(f, \alpha).$$

Tomando o limite quando o diâmetro de α vai para zero, obtemos a conclusão da proposição. \square

Em algumas situações interessantes, pode mostrar-se que a entropia é igual à taxa de crescimento do número de pontos periódicos:

$$\lim_n \frac{1}{n} \log \# \text{Fix}(f^n) = h(f). \quad (10.2.1)$$

Esse é o caso, por exemplo, dos deslocamentos de tipo finito que estudamos na Seção 10.2.2. Lembre da Proposição 10.2.5. Mais geralmente, a igualdade (10.2.1) vale sempre que a $f : M \rightarrow M$ é uma transformação expansora num espaço compacto, conforme veremos na Seção 11.3.

10.2.2 Deslocamentos de tipo finito

Seja $X = \{1, \dots, d\}$ um conjunto finito e seja $A = (A_{i,j})_{i,j}$ uma matriz quadrada de dimensão d cujos coeficientes tomam apenas os valores 0 ou 1 e tal que nenhuma linha é identicamente nula: para todo i existe j tal que $A_{i,j} = 1$. A este tipo de matriz, chamamos *matriz de transição*. Considere o subconjunto Σ_A de $\Sigma = X^{\mathbb{N}}$ das sequências $(x_n)_n$ que são *A-admissíveis*, ou seja, tais que

$$A_{x_n, x_{n+1}} = 1 \quad \text{para todo } n \in \mathbb{N}. \quad (10.2.2)$$

É claro que Σ_A é invariante pelo deslocamento $\sigma : \Sigma \rightarrow \Sigma$, no sentido de que $\sigma(\Sigma_A) \subset \Sigma_A$. Note também que Σ_A é fechado em Σ e, portanto, é um espaço métrico compacto (um fato análogo foi provado no Lema 7.2.5).

Denotaremos por $\sigma_A : \Sigma_A \rightarrow \Sigma_A$ a restrição do deslocamento σ a este compacto invariante. Ela é chamada de *deslocamento unilateral de tipo finito* associado a A . O deslocamento bilateral de tipo finito associado a uma matriz de transição A é definido de maneira análoga, considerando $\Sigma = X^{\mathbb{Z}}$ e exigindo (10.2.2) para todo $n \in \mathbb{Z}$. Neste caso também exigimos, como parte da definição de matriz de transição, que as colunas (não apenas as linhas) de A sejam não nulas.

A restrição do deslocamento $\sigma : \Sigma \rightarrow \Sigma$ ao suporte de qualquer medida de Markov é um deslocamento de tipo finito:

Exemplo 10.2.3. Dada uma matriz estocástica $P = (P_{i,j})_{i,j}$ defina $A = (A_{i,j})_{i,j}$ por

$$A_{i,j} = \begin{cases} 1 & \text{se } P_{i,j} > 0 \\ 0 & \text{se } P_{i,j} = 0. \end{cases}$$

Note que A é uma matriz de transição: a definição de matriz estocástica implica que nenhuma linha de P é identicamente nula (no caso bilateral precisamos supor que nenhuma coluna de P é identicamente nula; isso é automático, por exemplo, se a matriz P é aperiódica). Comparando (7.2.7) e (10.2.2) vemos que uma sequência é A -admissível se, e somente se, ela é P -admissível. Seja μ a medida de Markov determinada por um vetor de probabilidade $p = (p_j)_j$ com entradas positivas e tal que $P^*p = p$ (Exemplo 7.2.2). Pelo Lema 7.2.5, o suporte de μ coincide com o conjunto $\Sigma_A = \Sigma_P$ das sequências admissíveis.

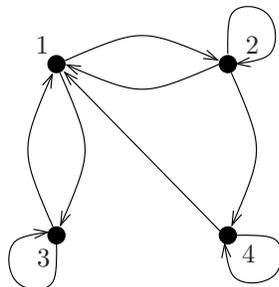


Figura 10.1: Grafo associado a uma matriz de transição

É útil associar a uma matriz de transição A o grafo orientado

$$G_A = \{(a, b) \in X \times X : A_{a,b} = 1\}.$$

Em outras palavras, G_A é o grafo cujos vértices são os pontos de $X = \{1, \dots, d\}$ e tal que existe uma aresta do vértice a para o vértice b se, e somente se, $A_{a,b} = 1$. Por exemplo, a Figura 10.1 descreve o grafo associado à matriz

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Chamamos *caminho de comprimento* $l \geq 1$ no grafo G_A a qualquer seqüência a_0, \dots, a_l em X tal que $A_{a_{i-1}, a_i} = 1$ para todo i , isto é, tal que sempre existe uma aresta ligando a_{i-1} a a_i . Dados $a, b \in X$ e $l \geq 1$, qual é o número $A_{a,b}^l$ de caminhos de comprimento l começando em a e terminando em b (isto é, com $a_0 = a$ e $a_l = b$)? Para responder a esta pergunta, observe que

1. $A_{a,b}^1 = 1$ se existe aresta ligando a a b e $A_{a,b}^1 = 0$ caso contrário. Em outras palavras, $A_{a,b}^1 = A_{a,b}$ para todo a, b .
2. Os caminhos de comprimento $l + m$ começando em a e terminando em b são as concatenações dos caminhos de comprimento l começando em a e terminando em algum ponto $z \in X$ com os caminhos de comprimento m começando nesse ponto z e terminando em b . Portanto,

$$A_{a,b}^{l+m} = \sum_{z=1}^d A_{a,z}^l A_{z,b}^m \quad \text{para todo } a, b \in X \text{ e todo } l, m \geq 1.$$

Destas duas observações segue, imediatamente, que $A_{a,b}^l$ é, precisamente, o coeficiente na linha a e coluna b da matriz A^l .

As propriedades topológicas básicas dos deslocamentos de tipo finito serão analisadas no Exercício 10.2.2. Na próxima proposição calcularemos a entropia topológica destas transformações. Para o enunciado precisamos de alguns comentários prévios sobre matrizes de transição.

Lembre que o *raio espectral* $\rho(B)$ de uma aplicação linear $B : \mathbb{R}^d \rightarrow \mathbb{R}^d$ (isto é, o máximo dos valores absolutos dos autovalores de B) é dado por

$$\rho(B) = \lim_n \|B^n\|^{1/n} = \lim_n |\operatorname{trc} B^n|^{1/n}, \quad (10.2.3)$$

onde $\|\cdot\|$ representa uma norma qualquer no espaço vetorial das aplicações lineares e trc designa o traço da matriz. Todas as normas são equivalentes, já que estamos em dimensão finita. Em geral, consideramos a *norma de operador* $\|B\| = \sup\{\|Bv\|/\|v\| : v \neq 0\}$ mas também será útil considerar a norma definida por

$$\|B\|_s = \sum_{i,j=1}^d |B_{i,j}|.$$

Agora suponha que A é uma matriz de transição. Como os coeficientes de A são não negativos, podemos usar o teorema de Perron-Frobenius (Teorema 7.2.3), para concluir que A admite um autovalor não negativo λ_A que é igual ao raio espectral. Pela definição de matriz de transição, também temos que as linhas da matriz A são não nulas. Então o mesmo vale para A^n , qualquer que seja $n \geq 1$ (Exercício 10.2.4). Isto implica que todos os coeficientes do vetor $A^n(1, \dots, 1)$ são (inteiros) positivos e, portanto,

$$\|A^n\| \geq \frac{\|A^n(1, \dots, 1)\|}{\|(1, \dots, 1)\|} \geq 1 \quad \text{para todo } n \geq 1.$$

Usando (10.2.3), obtemos que $\lambda_A = \rho(A) \geq 1$ para toda matriz de transição A .

Proposição 10.2.4. *A entropia topológica de um deslocamento de tipo finito $\sigma_A : \Sigma_A \rightarrow \Sigma_A$ é dada por $h(f_A) = \log \lambda_A$ onde λ_A é o maior autovalor da matriz de transição A .*

Demonstração. Tratamos o caso de deslocamentos unilaterais. O caso bilateral é análogo e fica a cargo do leitor. Consideremos a cobertura aberta α de Σ_A formada pelas restrições

$$[0; a]_A = \{(x_j)_j \in \Sigma_A : x_0 = a\}$$

dos cilindros $[0; a]$ de M . Para cada $n \geq 1$, a cobertura aberta α^n está formada pelas restrições

$$[0; a_0, \dots, a_{n-1}]_A = \{(x_j)_j \in \Sigma_A : x_j = a_j \text{ para } j = 0, \dots, n-1\}.$$

dos cilindros de comprimento n . Observe que este conjunto é não vazio se, e somente se, a_0, \dots, a_{n-1} é um caminho (de comprimento $n-1$) no grafo G_A : é evidente que esta condição é necessária; para ver que é suficiente, lembre

que por definição para todo i existe j tal que $A_{i,j} = 1$. Como os cilindros são disjuntos dois-a-dois, esta observação mostra que $N(\alpha^n)$ é igual ao número total de caminhos de comprimento $n - 1$ no grafo G_A , ou seja,

$$N(\alpha^n) = \sum_{i,j=1}^d A_{i,j}^{n-1} = \|A^{n-1}\|_s.$$

Pela fórmula do raio espectral (10.2.3), segue que

$$h(f, \alpha) = \lim_n \frac{1}{n} \log N(\alpha^n) = \lim_n \frac{1}{n} \log \|A^{n-1}\|_s = \log \rho(A) = \log \lambda_A.$$

Finalmente, como $\text{diam } \alpha^n \rightarrow 0$, o Corolário 10.1.10 dá que $h(\sigma_A) = h(\sigma_A, \alpha)$. \square

Proposição 10.2.5. *Se $\sigma_A : \Sigma_A \rightarrow \Sigma_A$ é um deslocamento de tipo finito então*

$$h(\sigma_A) = \lim_n \frac{1}{n} \log \# \text{Fix}(\sigma_A^n).$$

Demonstração. Tratamos o caso de deslocamentos unilaterais, deixando o caso bilateral a cargo do leitor. Note que $(x_k)_k \in \Sigma_A$ é ponto fixo de σ_A^n se, e somente se, $x_k = x_{k-n}$ para todo $k \geq n$. Em particular, cada cilindro $[0; a_0, \dots, a_{n-1}]_A$ contém no máximo um elemento de $\text{Fix}(\sigma_A^n)$. Além disso, existe um ponto fixo no cilindro se, e somente se, a_0, \dots, a_{n-1}, a_0 é um caminho (de comprimento n) no grafo G_A . Isto prova que

$$\# \text{Fix}(\sigma_A^n) = \sum_{i=1}^d A_{i,i}^n = \text{trc } A^n$$

para todo n . Consequentemente,

$$\lim_n \frac{1}{n} \log \# \text{Fix}(\sigma_A^n) = \lim_n \frac{1}{n} \log \text{trc } A^n = \log \rho(A).$$

Agora, a conclusão segue diretamente da proposição anterior. \square

10.2.3 Entropia topológica de fluxos

A definição da entropia topológica se estende facilmente para o contexto de fluxos contínuos $\phi = \{\phi^t : M \rightarrow M : t \in \mathbb{R}\}$ num espaço métrico M . Dados $x \in M$, $T > 0$ e $\varepsilon > 0$, chamamos *bola dinâmica* de centro x , comprimento T e raio $\varepsilon > 0$ ao conjunto

$$B(x, T, \varepsilon) = \{y \in M : d(\phi^t(x), \phi^t(y)) < \varepsilon \text{ para todo } 0 \leq t \leq T\}.$$

Seja K um subconjunto compacto de M . Dizemos que $E \subset M$ é conjunto (T, ε) -gerador para K se

$$K \subset \bigcup_{x \in E} B(x, T, \varepsilon)$$

e dizemos que $E \subset K$ é (T, ε) -separado se a bola dinâmica $B(x, T, \varepsilon)$ de cada $x \in E$ não contém nenhum outro elemento de E .

Denotamos por $g_T(\phi, \varepsilon, K)$ a menor cardinalidade de um conjunto (T, ε) -gerador para K e por $s_T(\phi, \varepsilon, K)$ a maior cardinalidade de um conjunto (T, ε) -separado $E \subset K$. Então tomamos

$$g(\phi, K) = \lim_{\varepsilon \rightarrow 0} \limsup_{T \rightarrow \infty} \frac{1}{T} \log g_T(\phi, \varepsilon, K) \quad \text{e}$$

$$s(\phi, K) = \lim_{\varepsilon \rightarrow 0} \limsup_{T \rightarrow \infty} \frac{1}{T} \log s_T(\phi, \varepsilon, K)$$

e definimos $g(\phi) = \sup_K g(\phi, K)$ e $s(\phi) = \sup_K s(\phi, K)$, onde os supremos são tomados sobre todos compactos $K \subset M$. Tal como no caso de tempo discreto (Proposição 10.1.4), estes dois últimos números coincidem.

Por definição, a *entropia topológica* do fluxo ϕ é $h(\phi) = g(\phi) = s(\phi)$. No enunciado a seguir supomos que o fluxo é *uniformemente contínuo*, ou seja, que para todo $T > 0$ e todo $\varepsilon > 0$ existe $\delta > 0$ tal que

$$d(x, y) < \delta \quad \Rightarrow \quad d(\phi^t(x), \phi^t(y)) < \varepsilon \quad \text{para todo } t \in [-T, T].$$

Observe que isto é automático para fluxos contínuos quando o espaço M é compacto.

Proposição 10.2.6. *Se o fluxo ϕ é uniformemente contínuo então a entropia topológica $h(\phi)$ do fluxo coincide com a entropia topológica $h(\phi^1)$ da sua aplicação tempo 1.*

Demonstração. Vamos mostrar que $g(\phi, K) = g(\phi^1, K)$ para todo compacto $K \subset M$. Isso implica a conclusão da proposição.

Se $E \subset M$ é (T, ε) -gerador para K relativamente ao fluxo ϕ então E é (n, ε) -gerador para K relativamente ao tempo 1, qualquer que seja $n \leq T + 1$. Em particular, $g_n(\phi^1, \varepsilon, K) \leq g_n(\phi, \varepsilon, K)$. Segue que

$$\limsup_n \frac{1}{n} \log g_n(\phi^1, \varepsilon, K) \leq \limsup_{T \rightarrow \infty} \frac{1}{T} \log g_T(\phi, \varepsilon, K)$$

e portanto $g(\phi^1, K) \leq g(\phi, K)$.

Na prova da desigualdade contrária usamos a hipótese de continuidade uniforme. Dado $\varepsilon > 0$, fixe $\delta \in (0, \varepsilon)$ tal que se $d(x, y) < \delta$ então $d(\phi^t(x), \phi^t(y)) < \varepsilon$ para todo $t \in [0, 1]$. Se $E \subset M$ é conjunto (n, δ) -gerador para K relativamente a ϕ^1 então E é conjunto (T, ε) -gerador para K relativamente ao fluxo ϕ , qualquer que seja $T \leq n$. Em particular, $g_T(\phi, \varepsilon, K) \leq g_n(\phi^1, \delta, K)$. Segue que

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \log g_T(\phi, \varepsilon, K) \leq \limsup_n \frac{1}{n} \log g_n(\phi^1, \delta, K)$$

(dada uma sequência $(T_j)_j$ que realiza o supremo do lado esquerdo, considere $n_j = [T_j] + 1$). Fazendo $\varepsilon, \delta \rightarrow 0$, obtemos que $g(\phi, K) \leq g(\phi^1, K)$. \square

10.2.4 Transformações diferenciáveis

Nesta seção supomos que M é uma variedade Riemanniana, ou seja, uma variedade diferenciável de dimensão finita munida de um produto interno no espaço tangente $T_x M$ em cada ponto x , tal que o produto interno depende diferenciavelmente de $x \in M$.

Se $f : M \rightarrow M$ é uma aplicação diferenciável, a sua derivada em cada ponto x é uma aplicação linear $Df(x) : T_x M \rightarrow T_{f(x)} M$ entre dois espaços de Hilbert (de dimensão finita). Denotamos por $\|Df(x)\|$ a sua norma de operador:

$$\|Df(x)\| = \sup \left\{ \frac{\|Df(x)v\|}{\|v\|} : v \in T_x M \text{ e } v \neq 0 \right\}.$$

O nosso objetivo é mostrar que esta norma determina uma cota superior para a entropia topológica. Dado um número real $x > 0$, denotamos $\log^+ x = \max\{\log x, 0\}$.

Proposição 10.2.7. *Seja $f : M \rightarrow M$ uma aplicação diferenciável numa variedade Riemanniana de dimensão d tal que $\|Df\|$ é limitado. Então*

$$h(f) \leq d \log^+ \sup \|Df\| < \infty.$$

Demonstração. Seja $L = \sup\{\|Df(x)\| : x \in M\}$. Pelo teorema do valor médio,

$$d(f(x), f(y)) \leq Ld(x, y) \quad \text{para todo } x, y \in M.$$

Se $L \leq 1$ então, como vimos no Exemplo 10.1.7, a entropia de f é nula, tal como é afirmado na proposição.

A partir de agora, suponhamos que $L > 1$. Seja \mathcal{A} um atlas da variedade M formado por cartas diferenciáveis $\varphi : (-2, 2)^d \rightarrow M$. Dado qualquer conjunto compacto $K \subset M$, podemos encontrar uma família finita $\mathcal{A}_K \subset \mathcal{A}$ tal que

$$\{\varphi((-1, 1)^d) : \varphi \in \mathcal{A}_K\}$$

cobre K . Fixe $B > 0$ tal que $d(\varphi(u), \varphi(v)) \leq Bd(u, v)$ para todo $u, v \in [-1, 1]^d$ e todo $\varphi \in \mathcal{A}_K$. Dados $n \geq 1$ e $\varepsilon > 0$, fixe $\delta = (\varepsilon/B\sqrt{d})L^{-n}$. Representamos por $\delta\mathbb{Z}^d$ o conjunto dos pontos da forma $(\delta k_1, \dots, \delta k_d)$ com $k_j \in \mathbb{Z}$ para todo $j = 1, \dots, d$. Seja $E \subset M$ a união das imagens $\varphi(\delta\mathbb{Z}^d \cap (-1, 1)^d)$, com $\varphi \in \mathcal{A}_K$.

Note que todo ponto de $(-1, 1)^d$ está a distância menor que $\delta\sqrt{d}$ de algum ponto de $\delta\mathbb{Z}^d \cap (-1, 1)^d$. Portanto, dado qualquer $\varphi \in \mathcal{A}_K$, todo $x \in \varphi((-1, 1)^d)$ está a distância menor que $B\delta\sqrt{d}$ de algum $a \in \varphi(\delta\mathbb{Z}^d \cap (-1, 1)^d)$. Então, pela escolha de δ ,

$$d(f^j(x), f^j(a)) \leq L^j B\delta\sqrt{d} < L^n B\delta\sqrt{d} = \varepsilon$$

para todo $j = 0, \dots, n-1$. Isto prova que E é (n, ε) -gerador para K . Por outro lado, por construção:

$$\#E \leq \#\mathcal{A}_K \#(\delta\mathbb{Z}^d \cap (-1, 1)^d) \leq \#\mathcal{A}_K (2/\delta)^d \leq \#\mathcal{A}_K (2B\sqrt{d}L^n/\varepsilon)^d.$$

Logo a expressão do lado direito é uma cota superior para $g_n(f, \varepsilon, K)$ e, por consequência,

$$g(f, \varepsilon, K) \leq \limsup_n \frac{1}{n} \log(2B\sqrt{d}L^n/\varepsilon)^d = d \log L.$$

Fazendo $\varepsilon \rightarrow 0$ e tomando o supremo sobre K , vem que $h(f) \leq d \log L$. \square

Um enunciado central na teoria da entropia topológica é a seguinte conjectura proposta por Michael Shub [Shu74]:

Conjectura 10.2.8 (Conjectura da entropia). Se $f : M \rightarrow M$ é um difeomorfismo de classe C^1 numa variedade Riemanniana de dimensão d , então

$$h(f) \geq \max_{1 \leq k \leq d} \log \rho(f_k) \quad (10.2.4)$$

onde $\rho(f_k)$ denota o raio espectral da ação $f_k : H_k(M) \rightarrow H_k(M)$ induzida por f na homologia real de dimensão k .

Embora sejam conhecidos diversas respostas parciais e resultados relacionados, positivos e negativos, o enunciado completo desta conjectura permanece em aberto. Vamos resumir alguns dos fatos conhecidos.

É sabido que a desigualdade (10.2.4) é verdadeira para um subconjunto aberto e denso do espaço dos homeomorfismos em qualquer variedade com dimensão $d \neq 4$. Além disso ela é verdadeira para todo homeomorfismo em certas classes de variedades, tais como as esferas e as infranilvariedades [MP77b, MP77a, MP08]. Por outro lado, Shub [Shu74] apresentou um homeomorfismo Lipschitz, com entropia nula, que não satisfaz a condição (10.2.4). Veja o Exercício 10.2.6.

Um modo útil de analisar (10.2.4) é comparando a entropia com cada um dos raios espectrais $\rho(f_k)$. O caso $k = d$ é relativamente fácil. De fato, para qualquer aplicação contínua f numa variedade de dimensão d , o raio espectral $\rho(f_d)$ é igual ao valor absoluto $|\deg f|$ do grau da aplicação. Em particular, a desigualdade $h(f) \geq \log \rho(f_d)$ é trivial no caso de homeomorfismos. Para aplicações contínuas não invertíveis, a entropia topológica pode ser menor que o logaritmo do valor absoluto do grau. Mas foi mostrado em [MP77b] que para aplicações diferenciáveis sempre vale $h(f) \geq \log |\deg f|$.

Anthony Manning [Man75] provou que a desigualdade $h(f) \geq \log \rho(f_1)$ é verdadeira para todo homeomorfismo numa variedade de qualquer dimensão d . Segue que $h(f) \geq \log \rho(f_{d-1})$, uma vez que o teorema de dualidade de Poincaré implica

$$\rho(f_k) = \rho(f_{d-k}) \quad \text{para todo } 0 < k < d.$$

Em particular, o teorema de Manning juntamente com as observações no parágrafo anterior prova que a conjectura da entropia é válida para homeomorfismos em qualquer variedade de dimensão $d \leq 3$.

Rufus Bowen [Bow78] provou que para qualquer homeomorfismo numa variedade $h(f)$ é maior ou igual que o logaritmo da taxa de crescimento do grupo

fundamental. Mostra-se que esta taxa de crescimento é maior ou igual que o raio espectral $\rho(f_1)$. Portanto, este resultado de Bowen implica o teorema de Manning que acabamos de mencionar.

O principal resultado relativo à conjectura da estabilidade é o teorema de Yosef Yomdin [Yom87], segundo o qual a conjectura é verdadeira para todo difeomorfismo de classe C^∞ . O ingrediente crucial da prova é uma relação entre a entropia topológica $h(f)$ e a *taxa de crescimento de volume* do difeomorfismo. Esta última noção é definida do seguinte modo. Para cada $1 \leq k < d$, seja B^k a bola unitária em \mathbb{R}^k . Represente por $v(\sigma)$ o volume (k -dimensional) da imagem de um mergulho diferenciável qualquer $\sigma : B^k \rightarrow M$. Então defina

$$v_k(f) = \sup_{\sigma} \limsup_n \frac{1}{n} \log v(f^n \circ \sigma),$$

onde o supremo é tomado sobre todos os mergulhos $\sigma : B^k \rightarrow M$ de classe C^∞ . Defina também $v(f) = \max\{v_k(f) : 1 \leq k < d\}$. Não é difícil verificar que

$$\log \rho(f_k) \leq v_k(f) \quad \text{para todo } 1 \leq k < d. \quad (10.2.5)$$

Sheldon Newhouse [New88] provou que $h(f) \leq v(f)$ para todo difeomorfismo C^r com $r > 1$. Por outro lado, Yomdin [Yom87] provou a desigualdade recíproca

$$v(f) \leq h(f), \quad (10.2.6)$$

para todo difeomorfismo de classe C^∞ (esta desigualdade é falsa, em geral, no caso C^r com $r < \infty$). Combinando (10.2.5) com (10.2.6) obtemos a conjectura da entropia (10.2.4) para todo difeomorfismo de classe C^∞ .

No que diz respeito a sistemas de classe C^1 , também é sabido que a desigualdade (10.2.4) é verdadeira para todo difeomorfismo Axioma A sem ciclos [SW75], para certos difeomorfismos parcialmente hiperbólicos [SX10] e, mais geralmente, para qualquer difeomorfismo longe de tangências homoclínicas [LVY].

10.2.5 Endomorfismos lineares do toro

Nesta seção provamos o seguinte resultado:

Proposição 10.2.9. *Seja $f_A : \mathbb{T}^d \rightarrow \mathbb{T}^d$ o endomorfismo induzido no toro \mathbb{T}^d por alguma matriz invertível A com coeficientes inteiros. Então*

$$h(f_A) = \sum_{j=1}^d \log^+ |\lambda_j|. \quad (10.2.7)$$

onde $\lambda_1, \dots, \lambda_d$ são os autovalores de A , contados com multiplicidade.

Vimos na Proposição 9.4.3 que a entropia de f_A relativamente à medida de Haar μ é igual à expressão no lado direito de (10.2.7). Pelo princípio variacional, que provaremos na Seção 10.4, a entropia topológica é maior ou igual do que

a entropia da transformação relativamente a qualquer probabilidade invariante. Isto implica que

$$h(f_A) \geq h_\mu(f) = \sum_{j=1}^d \log^+ |\lambda_j|.$$

Na presente seção nos concentramos na demonstração da desigualdade oposta:

$$h(f_A) \leq \sum_{j=1}^d \log^+ |\lambda_j|. \quad (10.2.8)$$

Inicialmente, suponha que A é diagonalizável, ou seja, que existe uma base v_1, \dots, v_d de \mathbb{R}^d com $Av_i = \lambda_i v_i$ para cada i . Não constitui restrição supor que $\|v_i\| = 1$ para todo i , e faremos isso no que segue. Além disso, renumerando os autovalores se necessário, podemos supor que $|\lambda_i| > 1$ de $1 \leq i \leq u$ e $|\lambda_i| \leq 1$ para todo $i > u$. Seja e_1, \dots, e_d a base canônica de \mathbb{R}^d e seja $P : \mathbb{R}^d \rightarrow \mathbb{R}^d$ o isomorfismo linear definido por $P(e_i) = v_i$ para cada i . Então $P^{-1}AP$ é uma matriz diagonal. Fixe $L > 0$ suficientemente grande para que $P((0, L)^d)$ contenha algum cubo unitário $\prod_{i=1}^d [b_i, b_i + 1]^d$. Veja a Figura 10.2. Seja $\pi : \mathbb{R}^d \rightarrow \mathbb{T}^d$ a projeção canônica. Então, $\pi P((0, L)^d)$ contém todo o toro \mathbb{T}^d .

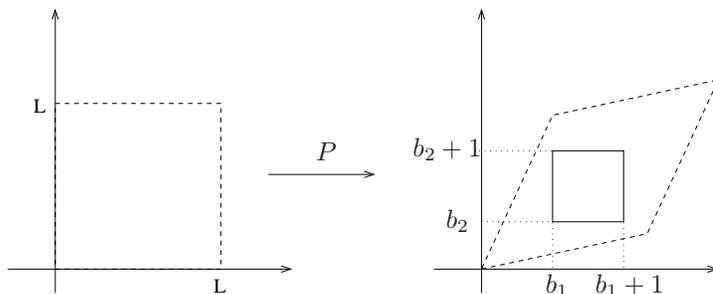


Figura 10.2: Construindo um conjunto (n, ε) -gerador em \mathbb{T}^d

Dados $n \geq 1$ e $\varepsilon > 0$, fixe $\delta > 0$ tal que $\|P\|\delta\sqrt{d} < \varepsilon$. Além disso, para cada $i = 1, \dots, d$, tome

$$\delta_i = \begin{cases} \delta |\lambda_i|^{-n} & \text{se } i \leq u \\ \delta & \text{se } i > u \end{cases}$$

Considere o conjunto

$$E = \pi P(\{(k_1 \delta_1, \dots, k_d \delta_d) \in (0, L)^d : k_1, \dots, k_d \in \mathbb{Z}\})$$

Observe também que, dado qualquer $j \geq 1$,

$$f^j(E) \subset \pi P(\{(k_1 \lambda_1^j \delta_1, \dots, k_d \lambda_d^j \delta_d) : k_1, \dots, k_d \in \mathbb{Z}\}).$$

Considere $0 \leq j < n$. Por construção, $|k_i \lambda_i^j \delta_i| \leq \delta$ para todo $i = 1, \dots, d$. Portanto, todo ponto de \mathbb{R}^d está a distância menor ou igual que $\delta\sqrt{d}$ de algum

ponto da forma $(k_1\lambda_1^j\delta_1, \dots, k_d\lambda_d^j\delta_d)$. Então (veja a Figura 10.2), para cada $x \in \mathbb{T}^d$ podemos encontrar $a \in E$ tal que $d(f^j(x), f^j(a)) \leq \|P\|\delta\sqrt{d}$ para todo $0 \leq j < n$. Isto mostra que E é um conjunto (n, ε) -gerador de \mathbb{T}^d . Por outro lado,

$$\#E \leq \prod_{i=1}^d \frac{L}{\delta_i} = \left(\frac{L}{\delta}\right)^d \prod_{i=1}^u |\lambda_i|^n.$$

Estas observações mostram que $g_n(f_A, \varepsilon, \mathbb{T}^d) \leq (L/\delta)^d \prod_{i=1}^u |\lambda_i|^n$ para todo $n \geq 1$ e todo $\varepsilon > 0$. Logo,

$$h(f) = \lim_{\varepsilon \rightarrow 0} \limsup_n \frac{1}{n} g_n(f_A, \varepsilon, \mathbb{T}^d) \leq \sum_{i=1}^u \log |\lambda_i| = \sum_{i=1}^d \log^+ |\lambda_i|.$$

Isto prova a Proposição 10.2.9 no caso em que A é diagonalizável.

O caso geral pode ser tratado de forma semelhante, escrevendo a matriz A na forma canônica de Jordan.

10.2.6 Exercícios

10.2.1. Sejam (M_i, d_i) , $i = 1, 2$ espaços métricos e sejam $f_i : M_i \rightarrow M_i$ transformações contínuas. Seja $M = M_1 \times M_2$ e seja d a distância definida em M por

$$d((x_1, x_2), (y_1, y_2)) = \max\{d_1(x_1, y_1), d_2(x_2, y_2)\}$$

e considere $f : M \rightarrow M$ definida por $f(x_1, x_2) = (f_1(x_1), f_2(x_2))$. Mostre que $h(f) \leq h(f_1) + h(f_2)$ e que vale a igualdade se algum dos espaços for compacto.

10.2.2. Seja $\sigma_A : \Sigma_A \rightarrow \Sigma_A$ um deslocamento de tipo finito, unilateral ou bilateral. Dizemos que a matriz de transição A é *irredutível* se para todo $i, j \in X$ existe $n \geq 1$ tal que $A_{i,j}^n > 0$ e dizemos que A é *aperiódica* se existe $n \geq 1$ tal que $A_{i,j}^n > 0$ para todo $i, j \in X$. Mostre que:

- (a) Se A é irredutível, o conjunto dos pontos periódicos de σ_A é denso em Σ_A .
- (b) σ_A é transitivo se, e somente se, A é irredutível.
- (c) σ_A é topologicamente misturador se, e somente se, A é aperiódico.

[Observação: A condição (b) significa que o grafo orientado G_A é conexo: dados $a, b \in X$ existe algum caminho em G_A começando em a e terminando em b .]

10.2.3. Seja M um espaço métrico compacto. Mostre que, dado qualquer $\varepsilon > 0$, a restrição da função entropia topológica $f \mapsto h(f)$ ao conjunto das transformações contínuas $f : M \rightarrow M$ que são ε -expansivas é semicontínua superiormente (relativamente à topologia da convergência uniforme).

10.2.4. Mostre que se A é matriz de transição então, para todo $k \geq 1$, nenhuma linha de A^k é identicamente nula. O mesmo vale para as colunas de A^k , $k \geq 1$, se supusermos que A é matriz de transição no sentido bilateral.

10.2.5. (a) Seja $f : M \rightarrow M$ um homeomorfismo local sobrejetivo num espaço métrico compacto e seja $d = \inf_y \#f^{-1}(y)$. Prove que $h(f) \geq \log d$.

(b) Seja $f : S^1 \rightarrow S^1$ uma aplicação contínua no círculo. Mostre que $h(f)$ é maior ou igual que o logaritmo do grau topológico de f : $h(f) \geq \log |\deg f|$.

[Observação: Misiurewicz, Przytycki [MP77b] provaram que a desigualdade $h(f) \geq \log |\deg f|$ vale para toda aplicação $f : M \rightarrow M$ de classe C^1 numa variedade compacta.]

10.2.6. Considere a aplicação $f : \overline{\mathbb{C}} \rightarrow \overline{\mathbb{C}}$ definida por $f(z) = z^d/(2|z|^{d-1})$, com $d \geq 2$. Prove que a entropia topológica de f é zero, mas o grau de f é d . Porquê isso não contradiz o Exercício 10.2.5?

10.2.7. Seja $f : M \rightarrow M$ uma aplicação contínua de um espaço métrico compacto M . Dado $\varepsilon > 0$, defina

$$g_*(f, \varepsilon) = \sup\{g(f, B(x, \infty, \varepsilon)) : x \in M\}$$

onde $B(x, \infty, \varepsilon)$ representa o conjunto dos $y \in M$ tais que $d(f^i(x), f^i(y)) \leq \varepsilon$ para todo $n \geq 0$. Bowen [Bow72] mostrou que, dados $b > 0$ e $\delta > 0$, existe $c > 0$ tal que $\log g_n(f, \delta, B(x, n, \varepsilon)) < c + (g_*(f, \varepsilon) + b)n$ para todo $x \in M$ e todo $n \geq 1$. Usando este fato, prove que $h(f) \leq g(f, \varepsilon, M) + g_*(f, \varepsilon)$. Dizemos que f é *h-expansiva* se $g_*(f, \varepsilon) = 0$ para algum $\varepsilon > 0$. Conclua que então $h(f) = g(f, \varepsilon, M)$. [Observação: Isto generaliza a Proposição 10.2.1, pois toda transformação expansiva é *h-expansiva*.]

10.3 Pressão

Nesta seção introduzimos uma generalização importante do conceito de entropia topológica, que é denominada *pressão* (ou *pressão topológica*), e estudamos as suas propriedades principais.

10.3.1 Definição via coberturas abertas

Seja $f : M \rightarrow M$ uma transformação contínua num espaço métrico compacto. Chamamos *potencial* em M a qualquer função contínua $\phi : M \rightarrow \mathbb{R}$. Para cada $n \in \mathbb{N}$, definimos $\phi_n : M \rightarrow \mathbb{R}$ por $\phi_n = \sum_{i=0}^{n-1} \phi \circ f^i$. Além disso, dado qualquer conjunto não vazio $C \subset M$, denotamos

$$\phi_n(C) = \sup\{\phi_n(x) : x \in C\}. \quad (10.3.1)$$

Dada uma cobertura aberta α de M definimos

$$P_n(f, \phi, \alpha) = \inf \left\{ \sum_{U \in \gamma} e^{\phi_n(U)} : \gamma \text{ é subcobertura finita de } \alpha^n \right\}. \quad (10.3.2)$$

Esta sequência é subaditiva (Exercício 10.3.1) e, portanto, o limite

$$P(f, \phi, \alpha) = \lim_n \frac{1}{n} \log P_n(f, \phi, \alpha) \quad (10.3.3)$$

existe. Finalmente, chamamos *pressão* do potencial ϕ relativamente a f ao limite $P(f, \phi)$ de $P(f, \phi, \alpha)$ quando o diâmetro de α vai para zero. A existência deste limite é garantida pelo seguinte lema:

Lema 10.3.1. *Existe $\lim_{\text{diam } \alpha \rightarrow 0} P(f, \phi, \alpha)$, ou seja, existe $P(f, \phi) \in [0, \infty]$ tal que*

$$\lim_k P(f, \phi, \alpha_k) = P(f, \phi)$$

para toda sequência $(\alpha_k)_k$ de coberturas abertas com $\text{diam } \alpha_k \rightarrow 0$.

Demonstração. Sejam $(\alpha_k)_k$ e $(\beta_l)_l$ seqüências quaisquer de coberturas abertas com diâmetros convergindo para zero. Dado qualquer $\varepsilon > 0$ fixe $\delta > 0$ tal que $|\phi(x) - \phi(y)| \leq \varepsilon$ sempre que $d(x, y) \leq \delta$. Por hipótese, $\text{diam } \alpha_k < \delta$ para todo k suficientemente grande. Para k fixado, seja $\rho > 0$ um número de Lebesgue para α_k . Por hipótese, $\text{diam } \beta_l < \rho$ para todo l suficientemente grande. Pela definição de número de Lebesgue, todo $B \in \beta_l$ está contido em algum $A \in \alpha_k$. Observe que $\phi_n(A) \leq n\varepsilon + \phi_n(B)$ para todo $n \geq 1$, uma vez que $\text{diam } \alpha_k < \delta$. Isto implica que

$$P_n(f, \phi, \alpha_k) \leq e^{n\varepsilon} P_n(f, \phi, \beta_l) \quad \text{para todo } n \geq 1$$

e, portanto, $P(f, \phi, \alpha_k) \leq \varepsilon + P(f, \phi, \beta_l)$. Fazendo $l \rightarrow \infty$ e depois $k \rightarrow \infty$, obtemos que

$$\limsup_k P(f, \phi, \alpha_k) \leq \varepsilon + \liminf_l P(f, \phi, \beta_l).$$

Como $\varepsilon > 0$ é arbitrário, segue que $\limsup_k P(f, \phi, \alpha_k) \leq \liminf_l P(f, \phi, \beta_l)$. Permutando os papéis das duas seqüências de coberturas, concluímos que os limites $\lim_k P(f, \phi, \alpha_k)$ e $\lim_l P(f, \phi, \beta_l)$ existem e são iguais. \square

Antes de prosseguirmos, vamos mencionar algumas consequências simples das definições. A primeira delas é que a pressão do potencial nulo coincide com a entropia topológica. De fato, é imediato de (10.3.2) que $P_n(f, 0, \alpha) = N(\alpha^n)$ para todo $n \geq 1$ e, portanto, $P(f, 0, \alpha) = h(f, \alpha)$ para toda cobertura aberta α . Portanto,

$$P(f, 0) = h(f). \quad (10.3.4)$$

Dada qualquer constante $c \in \mathbb{R}$, temos $P_n(f, \phi + c, \alpha) = e^{cn} P_n(f, \phi, \alpha)$ para todo $n \geq 1$ e, portanto, $P(f, \phi + c, \alpha) = P(f, \phi, \alpha) + c$ para toda cobertura aberta α . Logo,

$$P(f, \phi + c) = P(f, \phi) + c. \quad (10.3.5)$$

Analogamente, se $\phi \leq \psi$ então $P_n(f, \phi, \alpha) \leq P_n(f, \psi, \alpha)$ para todo $n \geq 1$ e, portanto, $P(f, \phi, \alpha) = P(f, \psi, \alpha)$ para toda cobertura aberta α . Ou seja,

$$\phi \leq \psi \quad \Rightarrow \quad P(f, \phi) \leq P(f, \psi). \quad (10.3.6)$$

Em particular, como $\inf \phi \leq \phi \leq \sup \phi$, temos

$$h(f) + \inf \phi \leq P(f, \phi) \leq h(f) + \sup \phi \quad (10.3.7)$$

para todo potencial ϕ . Um corolário interessante é que se $h(f)$ é finita então $P(f, \phi) < \infty$ para todo potencial ϕ e caso contrário $P(f, \phi) = \infty$ para todo potencial ϕ . Um exemplo desta último caso está dado no Exercício 10.1.5.

Outra consequência direta da definição é que a pressão é um invariante de equivalência topológica:

Proposição 10.3.2. *Sejam $f : M \rightarrow M$ e $g : N \rightarrow N$ transformações contínuas em espaços métricos compactos. Se existe um homeomorfismo $h : M \rightarrow N$ tal que $h \circ f = g \circ h$ então $P(g, \phi) = P(f, \phi \circ h)$ para todo potencial ϕ em N .*

Observação 10.3.3. É possível substituir o supremo pelo ínfimo em (10.3.1), embora a definição da pressão fique um pouco mais complicada. Para ver isso, considere:

$$\underline{P}_n(f, \phi, \alpha) = \inf \left\{ \sum_{U \in \gamma} e^{\phi_n(U)} : \gamma \text{ é subcobertura finita de } \alpha^n \right\}$$

onde $\underline{\phi}_n(C) = \inf\{\phi_n(x) : x \in C\}$. A sequência $\underline{P}_n(f, \phi, \alpha)$ pode não ser subaditiva. Por outro lado, como ϕ é (uniformemente) contínua, dado qualquer $\varepsilon > 0$ existe $\delta > 0$ tal que $\underline{\phi}_n(C) \leq \phi_n(C) \leq n\varepsilon + \underline{\phi}_n(C)$ sempre que $\text{diam } C \leq \delta$. Então,

$$\underline{P}_n(f, \phi, \alpha) \leq P_n(f, \phi, \alpha) \leq e^{n\varepsilon} \underline{P}_n(f, \phi, \alpha)$$

para toda cobertura aberta α com $\text{diam } \alpha \leq \delta$. Consequentemente,

$$\begin{aligned} P(f, \phi) &= \lim_{\text{diam } \alpha \rightarrow 0} \limsup_n \frac{1}{n} \log \underline{P}_n(f, \phi, \alpha) \\ &= \lim_{\text{diam } \alpha \rightarrow 0} \liminf_n \frac{1}{n} \log \underline{P}_n(f, \phi, \alpha). \end{aligned} \quad (10.3.8)$$

10.3.2 Conjuntos geradores e conjuntos separados

Agora vamos apresentar duas definições alternativas da pressão, em termos de conjuntos geradores e conjuntos separados. Tal como antes, $f : M \rightarrow M$ é uma transformação contínua num espaço métrico compacto e $\phi : M \rightarrow \mathbb{R}$ é uma função contínua.

Dados $n \geq 1$ e $\varepsilon > 0$, defina

$$\begin{aligned} G_n(f, \phi, \varepsilon) &= \inf \left\{ \sum_{x \in E} e^{\phi_n(x)} : E \text{ é conjunto } (n, \varepsilon)\text{-gerador de } M \right\} \text{ e} \\ S_n(f, \phi, \varepsilon) &= \sup \left\{ \sum_{x \in E} e^{\phi_n(x)} : E \text{ é conjunto } (n, \varepsilon)\text{-separado em } M \right\}. \end{aligned} \quad (10.3.9)$$

Em seguida, defina

$$\begin{aligned} G(f, \phi, \varepsilon) &= \limsup_n \frac{1}{n} \log G_n(f, \phi, \varepsilon) \text{ e} \\ S(f, \phi, \varepsilon) &= \limsup_n \frac{1}{n} \log S_n(f, \phi, \varepsilon) \end{aligned} \quad (10.3.10)$$

e também

$$G(f, \phi) = \lim_{\varepsilon \rightarrow 0} G(f, \phi, \varepsilon) \quad \text{e} \quad S(f, \phi) = \lim_{\varepsilon \rightarrow 0} S(f, \phi, \varepsilon) \quad (10.3.11)$$

(os limites existem, porque as funções são monótonas em ε).

Note que $G_n(f, 0, \varepsilon) = g_n(f, \varepsilon)$ e $S_n(f, 0, \varepsilon) = s_n(f, \varepsilon)$ para todo $n \geq 1$ e todo $\varepsilon > 0$. Portanto (Proposição 10.1.6), $G(f, 0) = g(f)$ e $S(f, 0) = s(f)$ são iguais à entropia topológica $h(f)$. Em geral, temos:

Proposição 10.3.4. $P(f, \phi) = G(f, \phi) = S(f, \phi)$ para todo potencial ϕ em M .

Demonstração. Considere $n \geq 1$ e $\varepsilon > 0$. É claro das definições que todo conjunto (n, ε) -separado maximal é (n, ε) -gerador. Então,

$$\begin{aligned} S_n(f, \phi, \varepsilon) &= \sup \left\{ \sum_{x \in E} e^{\phi_n(x)} : E \text{ é conjunto } (n, \varepsilon)\text{-separado} \right\} \\ &= \sup \left\{ \sum_{x \in E} e^{\phi_n(x)} : E \text{ é } (n, \varepsilon)\text{-separado maximal} \right\} \\ &\geq \inf \left\{ \sum_{x \in E} e^{\phi_n(x)} : E \text{ é } (n, \varepsilon)\text{-gerador} \right\} = G_n(f, \phi, \varepsilon). \end{aligned} \quad (10.3.12)$$

Isto implica que $G(f, \phi, \varepsilon) \leq S(f, \phi, \varepsilon)$. Passando ao limite quando $\varepsilon \rightarrow 0$, obtemos $G(f, \phi) \leq S(f, \phi)$.

Em seguida provamos que $S(f, \phi) \leq P(f, \phi)$. Sejam ε e δ números positivos tais que $d(x, y) \leq \delta$ implica $|\phi(x) - \phi(y)| \leq \varepsilon$. Seja α qualquer cobertura aberta de M com $\text{diam } \alpha < \delta$. Seja $E \subset M$ qualquer conjunto (n, δ) -separado. Dada qualquer subcobertura γ de α^n , é claro que todo ponto de E está contido em algum elemento de γ . Por outro lado, a hipótese de que E é (n, δ) -separado implica que cada elemento de γ contém no máximo um elemento de E . Portanto,

$$\sum_{x \in E} e^{\phi_n(x)} \leq \sum_{U \in \gamma} e^{\phi_n(U)}.$$

Tomando o supremo em E e o ínfimo em γ , obtemos que

$$S_n(f, \phi, \delta) \leq P_n(f, \phi, \alpha). \quad (10.3.13)$$

Segue que $S(f, \phi, \delta) \leq P(f, \phi, \alpha)$. Fazendo $\delta \rightarrow 0$ (logo $\text{diam } \alpha \rightarrow 0$), concluímos que $S(f, \phi) \leq P(f, \phi)$, conforme afirmado.

Finalmente, provamos que $P(f, \phi) \leq G(f, \phi)$. Sejam ε e δ números positivos tais que $d(x, y) \leq \delta$ implica $|\phi(x) - \phi(y)| \leq \varepsilon$. Seja α qualquer cobertura aberta de M com $\text{diam } \alpha < \delta$. Seja $\rho > 0$ um número de Lebesgue de α e seja $E \subset M$ um conjunto (n, ρ) -gerador qualquer de M . Para cada $x \in E$ e $i = 0, \dots, n-1$, existe $A_{x,i} \in \alpha$ tal que $B(f^i(x), \rho)$ está contida em $A_{x,i}$. Denotamos,

$$\gamma(x) = \bigcap_{i=0}^{n-1} f^{-i}(A_{x,i}).$$

Note que $\gamma(x) \in \alpha^n$ e que $B(x, n, \rho) \subset \gamma(x)$. Logo, a hipótese de que E é (n, ρ) -gerador implica que $\gamma = \{\gamma(x) : x \in E\}$ é uma subcobertura de α . Note também que

$$\phi_n(\gamma(x)) \leq n\varepsilon + \phi_n(x) \quad \text{para todo } x \in E,$$

uma vez que $\text{diam } A_{x,i} < \delta$ para todo i . Segue que

$$\sum_{U \in \gamma} e^{\phi_n(U)} \leq e^{n\varepsilon} \sum_{x \in E} e^{\phi_n(x)}.$$

Isto prova que $P_n(f, \phi, \alpha) \leq e^{n\varepsilon} G_n(f, \phi, \rho)$ para todo $n \geq 1$ e, por consequência,

$$P(f, \phi, \alpha) \leq \varepsilon + \liminf_n \frac{1}{n} G_n(f, \phi, \rho) \leq \varepsilon + G(f, \phi, \rho). \quad (10.3.14)$$

Fazendo $\rho \rightarrow 0$ vem que $P(f, \phi, \alpha) \leq \varepsilon + G(f, \phi)$. Então, fazendo ε, δ e $\text{diam } \alpha$ ir para zero, $P(f, \phi) \leq G(f, \phi)$. \square

A conclusão da Proposição 10.3.4 pode ser reescrita da seguinte forma:

$$P(f, \phi) = \lim_{s \rightarrow 0} \limsup_n \frac{1}{n} G_n(f, \phi, s) = \lim_{s \rightarrow 0} \limsup_n \frac{1}{n} S_n(f, \phi, s). \quad (10.3.15)$$

Mas as relações (10.3.14) e (10.3.12) na demonstração também dão que

$$P(f, \phi) \leq \lim_{s \rightarrow 0} \liminf_n \frac{1}{n} G_n(f, \phi, s) \leq \lim_{s \rightarrow 0} \liminf_n \frac{1}{n} S_n(f, \phi, s).$$

Comparando com (10.3.15), obtemos:

$$P(f, \phi) = \lim_{s \rightarrow 0} \liminf_n \frac{1}{n} G_n(f, \phi, s) = \lim_{s \rightarrow 0} \liminf_n \frac{1}{n} S_n(f, \phi, s). \quad (10.3.16)$$

10.3.3 Propriedades

Começamos por provar uma versão para a pressão do Corolário 10.1.10:

Proposição 10.3.5. *Seja $f : M \rightarrow M$ uma aplicação contínua num espaço métrico compacto. Seja β uma cobertura aberta de M tal que*

- (1) $\text{diam } \beta^k$ converge para zero quando $k \rightarrow \infty$, ou
- (2) $f : M \rightarrow M$ é um homeomorfismo e $\text{diam } \beta^{\pm k} \rightarrow 0$ quando $k \rightarrow \infty$.

Então $P(f, \phi) = P(f, \phi, \beta)$ para todo potencial ϕ em M .

Demonstração. Inicialmente, provamos o seguinte lemma:

Lema 10.3.6. $P(f, \phi, \alpha^k) = P(f, \phi, \alpha)$ para toda cobertura aberta α e todo $k \geq 1$.

Demonstração. Por definição, para todo $n \geq 1$:

$$P_n(f, \phi, \alpha^k) = \inf \left\{ \sum_{U \in \gamma} e^{\phi_n(U)} : \gamma \text{ subcobertura finita de } (\alpha^k)^n \right\} \text{ e}$$

$$P_{n+k}(f, \phi, \alpha) = \inf \left\{ \sum_{U \in \gamma} e^{\phi_{n+k}(U)} : \gamma \text{ subcobertura finita de } \alpha^{n+k} \right\}.$$

É claro que $(\alpha^k)^n = \alpha^{n+k}$. Então, denotando $L = \sup |\phi|$,

$$e^{-kL} P_n(f, \phi, \alpha^k) \leq P_{n+k}(f, \phi, \alpha) \leq e^{kL} P_n(f, \phi, \alpha^k)$$

para todo $n \geq 1$. Isto implica que $P(f, \phi, \alpha^k) = P(f, \phi, \alpha)$. \square

Os Lemas 10.3.1 e 10.3.6 dão que $P(f, \phi) = \lim_k P(f, \phi, \beta^k) = P(f, \phi, \beta)$. Isto prova o item (1) da Proposição 10.3.5. Para o item (2) também precisamos do seguinte fato:

Lema 10.3.7. *Se f é um homeomorfismo então $P(f, \phi, f^{-1}(\alpha)) = P(f, \phi, \alpha)$ para toda cobertura aberta α .*

Demonstração. Por definição, dado qualquer $n \geq 1$,

$$P_n(f, \phi, \alpha) = \inf \left\{ \sum_{U \in \gamma} e^{\phi_n(U)} : \gamma \text{ subcobertura finita de } \alpha^n \right\} \text{ e}$$

$$P_n(f, \phi, f^{-1}(\alpha)) = \inf \left\{ \sum_{V \in \delta} e^{\phi_n(V)} : \delta \text{ subcobertura finita de } (f^{-1}(\alpha))^n \right\}.$$

Seja $L = \sup |\phi|$. Note que $(f^{-1}(\alpha))^n = f^{-1}(\alpha^n)$ e toda subcobertura finita de $f^{-1}(\alpha^n)$ tem a forma $f^{-1}(\gamma)$ para alguma cobertura finita γ de α^n . Além disso,

$$\phi_n(U) - 2L \leq \phi_n(f^{-1}(U)) \leq \phi_n(U) + 2L$$

para qualquer $U \subset M$. Segue que,

$$e^{-2L} P_n(f, \phi, \alpha) \leq P_n(f, \phi, f^{-1}(\alpha)) \leq e^{2L} P_n(f, \phi, \alpha)$$

para todo $n \geq 1$. Logo, $P(f, \phi, f^{-1}(\alpha)) = P(f, \phi, \alpha)$. \square

Corolário 10.3.8. *Se f é um homeomorfismo então $P(f, \phi, \alpha^{\pm k}) = P(f, \phi, \alpha)$ para toda cobertura aberta α e todo $k \geq 1$.*

Demonstração. É claro da definição que $\alpha^{\pm k} = f^{-k}(\alpha^{2k})$. Portanto, a conclusão do corolário é uma consequência direta dos Lemas 10.3.6 e 10.3.7. \square

O item (2) da Proposição 10.3.5 é uma consequência direta do Lema 10.3.1 e do Corolário 10.3.8:

$$P(f, \phi) = \lim_k P(f, \phi, \beta^{\pm k}) = P(f, \phi, \beta).$$

Isto conclui a prova da proposição. \square

Proposição 10.3.9. *Seja $f : M \rightarrow M$ uma transformação contínua num espaço métrico compacto e seja ϕ um potencial em M . Então:*

$$(1) P(f^k, \phi_k) = kP(f, \phi) \text{ para todo } k \geq 1.$$

$$(2) \text{ Se } f \text{ é um homeomorfismo então } P(f^{-1}, \phi) = P(f, \phi).$$

Demonstração. Seja α uma cobertura aberta qualquer de M e seja $\beta = \alpha^k$. Dado um potencial ϕ em M , denote $\psi = \phi_k$. Observe que $\beta^n = \alpha^{kn}$ para cada n e que $\psi_n = \phi_{kn}$, onde $\psi_n = \sum_{j=0}^{n-1} \psi \circ f^j$. Então,

$$\begin{aligned} P_n(f^k, \psi, \beta) &= \inf \left\{ \sum_{U \in \gamma} e^{\psi_n(U)} : \gamma \subset \beta^n \right\} \\ &= \inf \left\{ \sum_{U \in \gamma} e^{\phi_{kn}(U)} : \gamma \subset \alpha^{kn} \right\} = P_{kn}(f, \phi, \alpha). \end{aligned}$$

Consequentemente, $P(f^k, \psi, \beta) = kP(f, \phi, \alpha)$ qualquer que seja α . Fazendo $\text{diam } \alpha \rightarrow 0$ (note que $\text{diam } \beta \rightarrow 0$), deduzimos que $P(f^k, \psi) = kP(f, \phi)$. Isto prova o item (1).

Para provar o item (2), seja α uma cobertura aberta de M . Para todo $n \geq 1$, denote

$$\alpha_-^n = \alpha \vee f(\alpha) \vee \dots \vee f^{n-1}(\alpha) \quad \text{e} \quad \phi_n^- = \sum_{j=0}^{n-1} \phi \circ f^{-j}.$$

Observe que $\alpha_-^n = f^{n-1}(\alpha_-^n)$ e que γ é uma subcobertura finita de α_-^n se, e somente se, $\delta = f^{n-1}(\gamma)$ é uma subcobertura finita de α_-^n . Além disso,

$$\phi_n(U) = \phi_n^-(f^{n-1}(U)),$$

qualquer que seja $U \subset M$. Combinando estes fatos, obtemos que

$$\begin{aligned} P_n(f, \phi, \alpha) &= \inf \left\{ \sum_{U \in \gamma} e^{\phi_n(U)} : \gamma \text{ subcobertura finita de } \alpha_-^n \right\} \\ &= \inf \left\{ \sum_{V \in \delta} e^{\phi_n^-(V)} : \delta \text{ subcobertura finita de } \alpha_-^n \right\} = P_n(f^{-1}, \phi, \alpha) \end{aligned}$$

para todo $n \geq 1$. Então $P(f, \phi, \alpha) = P(f^{-1}, \phi, \alpha)$ e, fazendo $\text{diam } \alpha \rightarrow 0$, vem que $P(f, \phi) = P(f^{-1}, \phi)$. \square

A seguir fixamos a transformação $f : M \rightarrow M$ e consideramos $P(f, \cdot)$ como função no espaço $C^0(M, \mathbb{R})$ das funções contínuas, munido da norma do supremo

$$\|\varphi\| = \sup\{|\varphi(x)| : x \in M\}.$$

Vimos em (10.3.7) que se a entropia topológica $h(f)$ é infinita então a pressão é identicamente igual a ∞ . No que segue suporemos que $h(f)$ é finita. Então, $P(f, \phi)$ é finita para todo potencial ϕ .

Proposição 10.3.10. *A função pressão é Lipschitz, com constante de Lipschitz igual a 1: $|P(f, \phi) - P(f, \psi)| \leq \|\phi - \psi\|$ para quaisquer potenciais ϕ e ψ .*

Demonstração. Claro que $\phi \leq \psi + \|\phi - \psi\|$. Logo, por (10.3.5) e (10.3.6), temos $P(f, \phi) \leq P(f, \psi) + \|\phi - \psi\|$. Trocando os papéis de ϕ e ψ obtemos a outra desigualdade. \square

Proposição 10.3.11. *A função pressão é convexa:*

$$P(f, (1-t)\phi + t\psi) \leq (1-t)P(f, \phi) + tP(f, \psi)$$

para quaisquer potenciais ϕ e ψ em M e para todo $0 \leq t \leq 1$.

Demonstração. Escreva $\xi = (1-t)\phi + t\psi$. Então $\xi_n = (1-t)\phi_n + t\psi_n$ para todo $n \geq 1$ e, portanto, $\xi_n(U) \leq (1-t)\phi_n(U) + t\psi_n(U)$ para todo $U \subset M$. Logo, pela desigualdade de Hölder (Teorema A.5.5),

$$\sum_{U \in \gamma} e^{\xi_n(U)} \leq \left(\sum_{U \in \gamma} e^{\phi_n(U)} \right)^{1-t} \left(\sum_{U \in \gamma} e^{\psi_n(U)} \right)^t$$

para qualquer família finita γ de subconjuntos de M . Isto implica que, dada qualquer cobertura aberta α ,

$$P_n(f, \xi, \alpha) \leq P_n(f, \phi, \alpha)^{1-t} P_n(f, \psi, \alpha)^t$$

para todo $n \geq 1$ e, portanto, $P(f, \xi, \alpha) \leq (1-t)P(f, \phi, \alpha) + tP(f, \psi, \alpha)$. Passando ao limite quando $\text{diam } \alpha \rightarrow 0$, obtemos a conclusão da proposição. \square

Dizemos que dois potenciais $\phi, \psi : M \rightarrow \mathbb{R}$ são *cohomólogos* se existe alguma função contínua $u : M \rightarrow \mathbb{R}$ tal que $\phi = \psi + u \circ f - u$. Note que esta é uma relação de equivalência no espaço dos potenciais (Exercício 10.3.2).

Proposição 10.3.12. *Seja $f : M \rightarrow M$ uma transformação contínua num espaço topológico compacto. Se $\phi, \psi : M \rightarrow \mathbb{R}$ são potenciais cohomólogos então $P(\phi, f) = P(\psi, f)$.*

Demonstração. Se $\psi = \phi + u \circ f - u$ então $\psi_n(x) = \phi_n(x) + u(f^n(x)) - u(x)$ para todo $n \in \mathbb{N}$. Seja $K = \sup |u|$. Então $|\psi_n(C) - \phi_n(C)| \leq 2K$ para todo conjunto $C \subset M$. Logo, para qualquer cobertura aberta γ ,

$$e^{-2K} \sum_{U \in \gamma} e^{\phi_n(U)} \leq \sum_{U \in \gamma} e^{\psi_n(U)} \leq e^{2K} \sum_{U \in \gamma} e^{\phi_n(U)}.$$

Isto implica que, dada qualquer cobertura aberta α de M ,

$$e^{-2K} P_n(f, \phi, \alpha) \leq P_n(f, \psi, \alpha) \leq e^{2K} P_n(f, \phi, \alpha)$$

para todo n . Logo, $P(f, \phi, \alpha) = P(f, \psi, \alpha)$ para todo α e, conseqüentemente, $P(f, \phi) = P(f, \psi)$. \square

10.3.4 Comentários sobre Mecânica Estatística

Façamos uma pausa para explicar a conexão entre o conceito matemático de pressão e as questões da Física que estão na origem de boa parte da teoria apresentada neste capítulo e no capítulo anterior. O leitor interessado em aprofundar este tema deve consultar o trabalho clássico de David Ruelle [Rue04].

A Mecânica Estatística (no equilíbrio) tem como objetivo descrever as propriedades de sistemas físicos formados por um grande número de unidades (ou partículas, ou sítios) que interagem entre si. A *constante de Avogadro* $6,022 \times 10^{23}$ dá uma ideia do que se entende por ‘grande’ em exemplos concretos.

O maior desafio é tentar compreender os fenômenos físicos de *transição de fase* (ou seja, de mudança de estado físico) como, por exemplo, a passagem da água do estado líquido ao estado sólido. Os métodos matemáticos desenvolvidos para este fim têm se revelado muito frutíferos em outros contextos, tais como a teoria quântica de campos e, mais próximo do âmbito deste livro, a teoria ergódica do sistemas dinâmicos uniformemente hiperbólicos (Bowen [Bow75a]).

Para efeitos de modelagem matemática, é conveniente supor que o conjunto M das unidades que formam o sistema é infinito. Os exemplos mais estudados são os reticulados $M = \mathbb{Z}^d$; tais modelos costumam ser chamados *cristais reticulados*. Além disso, é usual supor que cada unidade $x \in M$ admite um conjunto finito Ω_x de valores possíveis. Por exemplo, $\Omega_x = \{-1, +1\}$ no caso de *sistemas de spin* (-1 significa que a partícula x tem spin para baixo, $+1$ significa que o spin aponta para cima) e $\Omega = \{0, 1\}$ no caso de *gases reticulados* (1 significa que o sítio x está ocupado por uma molécula do gás, 0 significa que o sítio está vazio).

Então o *espaço de configurações* do sistema é um subconjunto Ω do produto $\prod_{x \in M} \Omega_x$ e um *estado* do sistema é uma medida de probabilidade em Ω . Um *estado de equilíbrio* descreve uma configuração macroscópica do sistema que pode ser fisicamente observada, isto é, que de fato ocorre na natureza. E uma transição de fases corresponde à coexistência de mais que um estado de equilíbrio.

Segundo o *princípio variacional* da Mecânica Estatística, que remonta à *lei do menor esforço* de Maupertuis, estados de equilíbrio são caracterizados por minimizarem uma certa grandeza fundamental, tal como, por exemplo, a energia livre de Gibbs ou a pressão. No caso de cristais reticulados, é usual supor que o espaço de configurações Ω é invariante pelos deslocamentos

$$(x_n^1, \dots, x_n^{i-1}, x_n^i, x_n^{i+1}, \dots, x_n^d) \mapsto (x_n^1, \dots, x_n^{i-1}, x_{n+1}^i, x_n^{i+1}, \dots, x_n^d).$$

Isto permite inserir tais sistemas no âmbito da Teoria Ergódica. Em particular, prova-se que sob hipóteses adequadas os estados de equilíbrio são medidas de um certo tipo, chamadas *estados de Gibbs*, que são invariantes pelos deslocamentos.

Vamos descrever o conceito de estado de Gibbs e detalhar um pouco mais as ideias anteriores no contexto dos cristais reticulados. Antes disso, a título de motivação adicional, consideremos um modelo particularmente simples, formado por uma única unidade com conjunto Ω finito. A *entropia* de uma probabilidade

μ em Ω é o número

$$S(\mu) = \sum_{\xi \in \Omega} -\mu(\xi) \log \mu(\xi).$$

A cada estado $\xi \in \Omega$ corresponde um valor $E(\xi)$ para a energia. Denotamos por $E(\mu)$ a energia média:

$$E(\mu) = \sum_{\xi \in \Omega} \mu(\xi) E(\xi).$$

Supomos que o sistema está a uma temperatura absoluta T , que é mantida constante ao longo do tempo. A *energia livre de Gibbs* está definida por:

$$G(\mu) = E(\mu) - \kappa T S(\mu)$$

onde $\kappa = 1.380 \times 10^{-23} m^2 kgs^{-2} K^{-1}$ é a chamada *constante de Boltzmann*. Em outras palavras, denotando $\beta = 1/(\kappa T)$, temos que

$$-\beta G(\mu) = \sum_{\xi \in \Omega} \mu(\xi) [-\beta E(\xi) - \log \mu(\xi)]. \quad (10.3.17)$$

É fácil mostrar que a expressão (10.3.17) é máxima (e, portanto, a energia livre de Gibbs $G(\mu)$ é mínima) se, e somente se:

$$\mu(\xi) = \frac{e^{-\beta E(\xi)}}{\sum_{\eta \in \Omega} e^{-\beta E(\eta)}} \quad \text{para todo } \xi \in \Omega \quad (10.3.18)$$

(veja o Lema 10.4.4 abaixo). Portanto, a *distribuição de Gibbs* μ dada por (10.3.18) é o único estado de equilíbrio do sistema.

10.3.5 Exercícios

10.3.1. Verifique que a sequência $\log P_n(f, \phi, \alpha)$ é subaditiva.

10.3.2. Verifique que a relação de cohomologia

$$\phi \sim \psi \Leftrightarrow \psi = \phi + u \circ f - u \quad \text{para alguma função contínua } u : M \rightarrow \mathbb{R}$$

é uma relação de equivalência.

10.3.3. Sejam $f_i : M_i \rightarrow M_i$, $i = 1, 2$ transformações contínuas em espaços métricos compactos e, para cada i , seja ϕ_i um potencial em M_i . Defina

$$\begin{aligned} f_1 \times f_2 : M_1 \times M_2 &\rightarrow M_1 \times M_2, & f_1 \times f_2(x_1, x_2) &= (f_1(x_1), f_2(x_2)) \\ \phi_1 \times \phi_2 : M_1 \times M_2 &\rightarrow \mathbb{R}, & \phi_1 \times \phi_2(x_1, x_2) &= \phi_1(x_1) + \phi_2(x_2). \end{aligned}$$

Mostre que $P(f_1 \times f_2, \phi_1 \times \phi_2) = P(f_1, \phi_1) + P(f_2, \phi_2)$.

10.3.4. Enuncie e prove um resultado análogo ao obtido na Proposição 10.2.1, para a pressão de uma função ϕ com respeito a uma transformação ε_0 -expansiva.

10.3.5. Seja M um espaço métrico compacto. Mostre que, dada qualquer função contínua $\phi : M \rightarrow \mathbb{R}$ e dado qualquer $\varepsilon_0 > 0$, a restrição da função pressão $f \mapsto P(f, \phi)$ ao conjunto das transformações contínuas $f : M \rightarrow M$ que são ε_0 -expansivas é semicontínua superiormente (relativamente à topologia da convergência uniforme).

10.3.6. Considere a transformação $f : S^1 \rightarrow S^1$ definida por $f(x) = 2x \pmod{\mathbb{Z}}$. Prove que se $\phi : S^1 \rightarrow \mathbb{R}$ é uma função Hölder, então

$$P(f, \phi) = \lim_n \frac{1}{n} \log \sum_{p \in \text{Fix}(f^n)} e^{\phi_n(p)}.$$

[Observação: Encontraremos um resultado mais geral no Exercício 11.3.4.]

10.4 Princípio variacional

O princípio variacional para a pressão, que enunciamos a seguir, foi provado originalmente por Ruelle [Rue73] num contexto mais restrito, sendo estendido por Walters [Wal75] para o contexto que consideramos aqui:

Teorema 10.4.1 (Princípio Variacional). *Seja $f : M \rightarrow M$ uma transformação contínua num espaço métrico compacto e seja $\mathcal{M}(f)$ o conjunto das medidas de probabilidade invariantes por f . Então, para toda função contínua $\phi : M \rightarrow \mathbb{R}$,*

$$P(\phi, f) = \sup\{h_\nu(f) + \int \phi d\nu : \nu \in \mathcal{M}(f)\}.$$

O Teorema 10.1 corresponde ao caso particular $\phi \equiv 0$. Em particular, segue que f tem entropia topológica nula se, e somente se, $h_\nu(f) = 0$ para toda probabilidade invariante ν . Esse é o caso, por exemplo, dos homeomorfismos do círculo (Exemplo 10.1.1) e das translações em grupos compactos metrizáveis (Exemplo 10.1.7). A hipótese de compacidade é fundamental, uma vez que existem transformações sem medidas invariantes e com entropia topológica positiva. Veja o Exercício 10.4.4.

Nas Seções 10.4.1 e 10.4.2 apresentaremos uma demonstração do Teorema 10.4.1 devida a Misiurewicz [Mis76]. Antes disso, vamos mencionar um par de consequências.

Corolário 10.4.2. *Seja $f : M \rightarrow M$ uma transformação contínua num espaço métrico compacto e seja $\mathcal{M}_e(f)$ o conjunto das medidas de probabilidade invariantes e ergódicas. Então,*

$$P(\phi, f) = \sup\{h_\nu(f) + \int \phi d\nu : \nu \in \mathcal{M}_e(f)\}.$$

Demonstração. Dada qualquer $\nu \in \mathcal{M}(f)$, seja $\{\nu_P : P \in \mathcal{P}\}$ a sua decomposição ergódica. Pelos Teoremas 5.1.3 e 9.6.2,

$$h_\nu(f) + \int \phi d\nu = \int \left(h_{\nu_P}(f) + \int \phi d\nu_P \right) d\hat{\mu}(P).$$

Isto implica que

$$\sup\{h_\nu(f) + \int \phi d\nu : \nu \in \mathcal{M}(f)\} \leq \sup\{h_\nu(f) + \int \phi d\nu : \nu \in \mathcal{M}_e(f)\}.$$

A desigualdade recíproca é trivial, uma vez que $\mathcal{M}_e(f) \subset \mathcal{M}(f)$. Agora basta aplicar o Teorema 10.4.1. \square

Outra consequência interessante é que para transformações com entropia topológica finita a função pressão determina o conjunto as medidas de probabilidade invariantes:

Corolário 10.4.3 (Walters). *Seja $f : M \rightarrow M$ uma transformação contínua num espaço métrico compacto, com entropia topológica $h(f) < \infty$. Seja η uma medida com sinal finita em M . Então η é uma probabilidade invariante por f se, e somente se, $\int \phi d\eta \leq P(f, \phi)$ para toda função contínua $\phi : M \rightarrow \mathbb{R}$.*

Demonstração. A afirmação ‘somente se’ é uma consequência imediata do Teorema 10.4.1: se η é uma probabilidade invariante então

$$P(f, \phi) \geq h_\eta(f) + \int \phi d\eta \geq \int \phi d\eta$$

para toda função contínua ϕ . Em seguida, provamos a recíproca.

Seja η uma medida com sinal, finita tal que $\int \phi d\eta \leq P(f, \phi)$ para toda ϕ . Considere qualquer $\phi \geq 0$. Para quaisquer $c > 0$ e $\varepsilon > 0$,

$$c \int (\phi + \varepsilon) d\eta = - \int -c(\phi + \varepsilon) d\eta \geq -P(f, -c(\phi + \varepsilon)).$$

Pela relação (10.3.7), temos que

$$P(f, -c(\phi + \varepsilon)) \leq h(f) + \sup(-c(\phi + \varepsilon)) = h(f) - c \inf(\phi + \varepsilon).$$

Portanto, $c \int (\phi + \varepsilon) d\eta \geq -h(f) + \inf(\phi + \varepsilon)$. Quando $c > 0$ é suficientemente grande o lado direito desta desigualdade é positivo. Logo $\int (\phi + \varepsilon) d\eta > 0$. Como $\varepsilon > 0$ é arbitrário, isto implica que $\int \phi d\eta \geq 0$ para todo potencial $\phi \geq 0$. Logo, η é uma medida positiva.

O próximo passo é mostrar que η é uma probabilidade. Por hipótese,

$$\int c d\eta \leq P(f, c) = h(f) + c$$

para todo $c \in \mathbb{R}$. Para $c > 0$ isto implica $\eta(M) \leq 1 + h(f)/c$. Passando ao limite quando $c \rightarrow +\infty$ obtemos que $\eta(M) \leq 1$. Analogamente, considerando $c < 0$ e passando ao limite quando $c \rightarrow -\infty$, vem que $\eta(M) \geq 1$. Portanto, η é uma probabilidade, conforme afirmado.

Resta provar que η é invariante por f . Por hipótese, dado $c \in \mathbb{R}$ e qualquer potencial ϕ ,

$$c \int (\phi \circ f - \phi) d\eta \leq P(f, c(\phi \circ f - \phi)).$$

Pela Proposição 10.3.12, a expressão no lado direito é igual a $P(f, 0) = h(f)$. Para $c > 0$, isto implica

$$\int (\phi \circ f - \phi) d\eta \leq \frac{h(f)}{c}$$

e, passando ao limite quando $c \rightarrow +\infty$, segue que $\int (\phi \circ f - \phi) d\eta \leq 0$. O mesmo argumento aplicado a $-\phi$ dá que $\int (\phi \circ f - \phi) d\eta \geq 0$. Logo $\int \phi \circ f d\eta = \int \phi d\eta$ para todo potencial ϕ . Isto implica (Proposição A.3.3) que $f_*\eta = \eta$. \square

10.4.1 Prova da cota superior

Nesta seção provamos que, dada qualquer probabilidade invariante ν ,

$$h_\nu(f) + \int \phi d\nu \leq P(f, \phi). \quad (10.4.1)$$

Para isso, seja $\mathcal{P} = \{P_1, \dots, P_s\}$ uma partição finita qualquer. Vamos mostrar que se α é uma cobertura aberta de M com diâmetro suficientemente pequeno, dependendo apenas de \mathcal{P} , então

$$h_\nu(f, \mathcal{P}) + \int \phi d\nu \leq \log 4 + P(f, \phi, \alpha). \quad (10.4.2)$$

Fazendo $\text{diam } \alpha \rightarrow 0$, segue que $h_\nu(f, \mathcal{P}) + \int \phi d\nu \leq \log 4 + P(f, \phi)$ para toda partição finita \mathcal{P} . Logo, $h_\nu(f) + \int \phi d\nu \leq \log 4 + P(f, \phi)$. Agora substitua f por f^k e o potencial ϕ por ϕ_k . Note que $\int \phi_k d\nu = k \int \phi d\nu$, uma vez que ν é invariante por f . Usando também as Proposições 9.1.14 e 10.3.9, vem que

$$kh_\nu(f, \mathcal{P}) + k \int \phi d\nu \leq \log 4 + kP(f, \phi)$$

para todo $k \geq 1$. Dividindo por k e passando ao limite quando $k \rightarrow \infty$ obtemos a desigualdade (10.4.1).

Agora resta provar (10.4.2). Usaremos o seguinte fato elementar:

Lema 10.4.4. *Sejam a_1, \dots, a_k números reais e sejam p_1, \dots, p_k números não negativos tais que $p_1 + \dots + p_k = 1$. Seja $A = \sum_{i=1}^k e^{a_i}$. Então,*

$$\sum_{i=1}^k p_i(a_i - \log p_i) \leq \log A.$$

Além disso, vale a igualdade se, e somente se, $p_j = e^{a_j}/A$ para todo j .

Demonstração. Escreva $t_i = e^{a_i}/A$ e $x_i = p_i/e^{a_i}$. Note que $\sum_{i=1}^k t_i = 1$. Pela propriedade de concavidade (9.1.8),

$$\sum_{i=1}^k t_i \phi(x_i) \leq \phi\left(\sum_{i=1}^k t_i x_i\right).$$

Note que $t_i \phi(x_i) = (p_i/A)(a_i - \log p_i)$ e que $\sum_{i=1}^k t_i x_i = 1/A$. Portanto, a desigualdade anterior pode ser reescrita na seguinte forma:

$$\sum_{i=1}^k \frac{p_i}{A} (a_i - \log p_i) \leq \frac{1}{A} \log A.$$

Multiplicando por A obtemos a desigualdade no enunciado do lema. Além disso, vale a igualdade se, e somente se, os x_i são todos iguais, ou seja se existe c tal que $p_i = ce^{a_i}$ para todo i . Somando sobre $i = 1, \dots, k$ vemos que nesse caso $c = 1/A$, conforme afirmado no enunciado. \square

Como a medida ν é regular (Proposição A.3.2), dado $\varepsilon > 0$ podemos encontrar compactos $Q_i \subset P_i$ tais que $\nu(P_i \setminus Q_i) < \varepsilon$ para todo $i = 1, \dots, s$. Seja Q_0 o complementar de $\cup_{i=1}^s Q_i$ e seja $P_0 = \emptyset$. Então $\mathcal{Q} = \{Q_0, Q_1, \dots, Q_s\}$ é uma partição finita de M satisfazendo $\nu(P_i \Delta Q_i) < s\varepsilon$ para todo $i = 0, 1, \dots, s$. Logo, pelo Lema 9.1.6,

$$H_\nu(\mathcal{P}/\mathcal{Q}) \leq \log 2$$

desde que $\varepsilon > 0$ seja suficientemente pequeno (dependendo apenas de s). Fixe ε e \mathcal{Q} a partir daqui e suponha que a cobertura α satisfaz

$$\text{diam } \alpha < \min\{d(Q_i, Q_j) : 1 \leq i < j \leq s\}. \quad (10.4.3)$$

Pelo Lema 9.1.11, temos que $h_\nu(f, \mathcal{P}) \leq h_\nu(f, \mathcal{Q}) + H_\nu(\mathcal{P}/\mathcal{Q}) \leq h_\nu(f, \mathcal{Q}) + \log 2$. Logo, para provar (10.4.2) basta mostrar que

$$h_\nu(f, \mathcal{Q}) + \int \phi d\nu \leq \log 2 + P(f, \phi, \alpha). \quad (10.4.4)$$

Para esse fim, observe que

$$H_\nu(\mathcal{Q}^n) + \int \phi_n d\nu \leq \sum_{Q \in \mathcal{Q}^n} \nu(Q) (-\log \mu(Q) + \phi_n(Q))$$

para todo $n \geq 1$. Então, pelo Lema 10.4.4,

$$H_\nu(\mathcal{Q}^n) + \int \phi_n d\nu \leq \log \left(\sum_{Q \in \mathcal{Q}^n} e^{\phi_n(Q)} \right). \quad (10.4.5)$$

Seja γ uma subcobertura finita qualquer de α^n . Para cada $Q \in \mathcal{Q}^n$ considere qualquer ponto x_Q no fecho de Q tal que $\phi_n(x_Q) = \phi_n(Q)$ (lembre que $\phi_n(Q)$ denota o supremo de ϕ_n no conjunto Q). Considere $U_Q \in \gamma$ tal que $x_Q \in U_Q$. Então, $\phi_n(Q) \leq \phi_n(U_Q)$ para todo $Q \in \mathcal{Q}^n$. A condição (10.4.3) implica que cada elemento de α intersecta o fecho de não mais que dois elementos de \mathcal{Q} . Portanto, cada elemento de α^n intersecta o fecho de, no máximo, 2^n elementos de \mathcal{Q}^n . Em particular, para cada $U \in \gamma$ existem não mais que 2^n elementos Q de \mathcal{Q}^n tais que $U_Q = U$. Portanto:

$$\sum_{Q \in \mathcal{Q}^n} e^{\phi_n(Q)} \leq 2^n \sum_{U \in \gamma} e^{\phi_n(U)}, \quad (10.4.6)$$

para qualquer subcobertura finita γ de α^n . Combinando (10.4.5) e (10.4.6):

$$H_\nu(\mathcal{Q}^n) + \int \phi_n d\nu \leq n \log 2 + \log P_n(f, \phi, \alpha).$$

Dividindo por n e passando ao limite quando $n \rightarrow \infty$, obtemos (10.4.4). Isto completa a demonstração da cota superior (10.4.1).

10.4.2 Aproximando a pressão

Para terminar a prova do Teorema 10.4.1, mostraremos agora que para todo $\varepsilon > 0$ existe uma probabilidade μ invariante por f e tal que

$$h_\mu(f) + \int \phi d\mu \geq S(f, \phi, \varepsilon) \quad (10.4.7)$$

Claramente, isto implica que o supremo dos valores de $h_\nu(f) + \int \phi d\nu$ quando ν varia em $\mathcal{M}(f)$ é maior ou igual que $S(f, \phi) = P(f, \phi)$.

Para cada $n \geq 1$, seja E um conjunto (n, ε) -separado tal que

$$\sum_{y \in E} e^{\phi_n(y)} \geq \frac{1}{2} S_n(f, \phi, \varepsilon). \quad (10.4.8)$$

Denotaremos por A a expressão no lado esquerdo desta desigualdade. Considere as medidas de probabilidade ν_n e μ_n definidas em M por:

$$\nu_n = \frac{1}{A} \sum_{x \in E} e^{\phi_n(x)} \delta_x \quad \text{e} \quad \mu_n = \sum_{j=0}^{n-1} f_*^j \nu_n.$$

Como o espaço das probabilidades é compacto (Teorema 2.1.5), e lembrando a definição (10.3.10), podemos escolher uma subseqüência $(n_j)_j \rightarrow \infty$ tal que

1. $\frac{1}{n_j} \log S_{n_j}(f, \phi, \varepsilon)$ converge para $S(f, \phi, \varepsilon)$ e
2. μ_{n_j} converge para alguma probabilidade μ na topologia fraca*.

Vamos mostrar que tal probabilidade μ é invariante por f e satisfaz (10.4.7). para facilitar a leitura, dividimos o argumento em quatro passos.

Passo 1: Provamos que μ é invariante. Seja $\varphi : M \rightarrow \mathbb{R}$ uma função contínua qualquer. Para cada $n \geq 1$,

$$\int \varphi d(f_* \mu_n) = \frac{1}{n} \sum_{j=1}^n \int \varphi \circ f^j d\nu_n = \int \varphi d\mu_n + \frac{1}{n} \left(\int \varphi \circ f^n d\nu_n - \int \varphi d\nu_n \right)$$

e, por consequência,

$$\left| \int \varphi d(f_* \mu_n) - \int \varphi d\mu_n \right| \leq \frac{2}{n} \sup |\varphi|.$$

Restringindo a $n = n_j$ e passando ao limite quando $j \rightarrow \infty$, vemos que $\int \varphi df_* \mu = \int \varphi d\mu$ para toda função contínua $\varphi : M \rightarrow \mathbb{R}$. Portanto (lembre a Proposição A.3.3), $f_* \mu = \mu$ tal como foi afirmado.

Passo 2: Estimamos a entropia relativamente a ν_n . Seja \mathcal{P} qualquer partição finita de M tal que $\text{diam } \mathcal{P} < \varepsilon$ e $\mu(\partial\mathcal{P}) = 0$. A primeira condição implica que cada elemento de \mathcal{P}^n contém no máximo um elemento de E . Por outro lado, todo elemento de E está contido em algum elemento de \mathcal{P}^n , evidentemente. Logo,

$$\begin{aligned} H_{\nu_n}(\mathcal{P}^n) &= \sum_{x \in E} -\nu_n(\{x\}) \log \nu_n(\{x\}) = \sum_{x \in E} -\frac{1}{A} e^{\phi_n(x)} \log \left(\frac{1}{A} e^{\phi_n(x)} \right) \\ &= \log A - \frac{1}{A} \sum_{x \in E} e^{\phi_n(x)} \phi_n(x) = \log A - \int \phi_n d\nu_n \end{aligned} \quad (10.4.9)$$

(a última igualdade decorre diretamente da definição de ν_n).

Passo 3: Calculamos a entropia relativamente a μ_n . Considere $1 \leq k < n$. Para cada $r \in \{0, \dots, k-1\}$, seja $q_r \geq 0$ o maior número inteiro tal que $r + kq_r \leq n$. Em outras palavras, $q_r = \lfloor (n-r)/k \rfloor$. Então,

$$\mathcal{P}^n = \mathcal{P}^r \vee \left[\bigvee_{j=0}^{q_r-1} f^{-(kj+r)}(\mathcal{P}^k) \right] \vee f^{-(kq_r+r)}(\mathcal{P}^{n-(kq_r+r)})$$

(o primeiro termo não existe se $r = 0$ e o terceiro não existe se $n = kq_r + r$). Portanto,

$$H_{\nu_n}(\mathcal{P}^n) \leq \sum_{j=0}^{q_r-1} H_{\nu_n}(f^{-(kj+r)}(\mathcal{P}^k)) + H_{\nu_n}(\mathcal{P}^r) + H_{\nu_n}(f^{-(kq_r+r)}(\mathcal{P}^{n-(kq_r+r)})).$$

Claro que $\#\mathcal{P}^r \leq (\#\mathcal{P})^k$. Usando o Lema 9.1.3, segue $H_{\nu_n}(\mathcal{P}^r) \leq k \log \#\mathcal{P}$. Pela mesma razão, o último termo na desigualdade anterior também é limitado por $k \log \#\mathcal{P}$. Então, usando também a propriedade (9.1.12),

$$H_{\nu_n}(\mathcal{P}^n) \leq \sum_{j=0}^{q_r-1} H_{f_*^{(kj+r)} \nu_n}(\mathcal{P}^k) + 2k \log \#\mathcal{P} \quad (10.4.10)$$

para todo $r \in \{0, \dots, k-1\}$. Agora, é claro que todo número $i \in \{0, \dots, n-1\}$ se escreve de maneira única na forma $i = kj + r$ com $0 \leq j \leq q_r - 1$. Então, somando (10.4.10) sobre todos os valores de r ,

$$kH_{\nu_n}(\mathcal{P}^n) \leq \sum_{i=0}^{n-1} H_{f_*^i \nu_n}(\mathcal{P}^k) + 2k^2 \log \#\mathcal{P}. \quad (10.4.11)$$

A propriedade (9.1.8) de concavidade da função $\phi(x) = -x \log x$ implica que

$$\frac{1}{n} \sum_{i=0}^{n-1} H_{f_*^i \nu_n}(\mathcal{P}^k) \leq H_{\mu_n}(\mathcal{P}^k).$$

Combinando esta desigualdade com (10.4.11), vemos que

$$\frac{1}{n}H_{\nu_n}(\mathcal{P}_n) \leq \frac{1}{k}H_{\mu_n}(\mathcal{P}^k) + \frac{2k}{n} \log \#\mathcal{P}.$$

Por outro lado, pela definição de μ_n ,

$$\frac{1}{n} \int \phi_n d\nu_n = \frac{1}{n} \sum_{j=0}^{n-1} \int \phi \circ f^j d\nu_n = \int \phi d\mu_n.$$

Somando com a relação anterior, obtemos

$$\frac{1}{n}H_{\nu_n}(\mathcal{P}_n) + \frac{1}{n} \int \phi_n d\nu_n \leq \frac{1}{k}H_{\mu_n}(\mathcal{P}^k) + \int \phi d\mu_n + \frac{2k}{n} \log \#\mathcal{P}. \quad (10.4.12)$$

Passo 4: Traduzimos as estimativas anteriores para a medida limite μ . Combinando (10.4.9) e (10.4.12), vem que

$$\frac{1}{k}H_{\mu_n}(\mathcal{P}^k) + \int \phi d\mu_n \geq \frac{1}{n} \log A - \frac{2k}{n} \log \#\mathcal{P}.$$

Pela escolha de E em (10.4.8), segue que

$$\frac{1}{k}H_{\mu_n}(\mathcal{P}^k) + \int \phi d\mu_n \geq \frac{1}{n} \log S_n(f, \phi, \varepsilon) - \frac{1}{n} \log 2 - \frac{2k}{n} \log \#\mathcal{P}. \quad (10.4.13)$$

A escolha da partição \mathcal{P} implica que $\mu(\partial\mathcal{P}^k) = 0$ para todo $k \geq 1$, já que

$$\partial\mathcal{P}^k \subset \mathcal{P} \cup f^{-1}(\mathcal{P}) \cup \dots \cup f^{-k+1}(\mathcal{P}).$$

Então, $\mu(P) = \lim_j \mu_{n_j}(P)$ para todo $P \in \mathcal{P}^k$ e, portanto,

$$H_\mu(\mathcal{P}^k) = \lim_j H_{\mu_{n_j}}(\mathcal{P}^k).$$

Como a função ϕ é contínua, também temos $\int \phi d\mu = \lim_j \int \phi d\mu_{n_j}$. Portanto, restringindo (10.4.13) à subsequência $(n_j)_j$ e passando ao limite quando $j \rightarrow \infty$,

$$\frac{1}{k}H_\mu(\mathcal{P}^k) + \int \phi d\mu \geq S(f, \phi, \varepsilon).$$

Passando ao limite quando $k \rightarrow \infty$, vem que

$$h_\mu(f, \mathcal{P}) + \int \phi d\mu \geq S(f, \phi, \varepsilon).$$

Agora, fazendo $\varepsilon \rightarrow 0$ (e, conseqüentemente, $\text{diam } \mathcal{P} \rightarrow 0$), obtemos (10.4.7).

Isto completa a demonstração do Princípio Variacional (Teorema 10.4.1).

10.4.3 Exercícios

10.4.1. Seja $f : M \rightarrow M$ uma aplicação contínua em um espaço métrico compacto M . Verifique que $P(f, \varphi) \leq h(f) + \sup \varphi$ para toda função contínua $\varphi : M \rightarrow \mathbb{R}$.

10.4.2. Se $f : M \rightarrow M$ é uma transformação contínua em um espaço métrico compacto e $X \subset M$ é um conjunto positivamente invariante, ou seja, tal que $f(X) \subset X$, então $P(f | X, \varphi | X) \leq P(f, \varphi)$.

10.4.3. Dê outra demonstração da Proposição 10.3.12, usando o princípio variacional.

10.4.4. Exiba um exemplo de transformação contínua $f : M \rightarrow M$ em um espaço métrico não compacto M , tal que f não possui medida invariante mas a entropia $h(f)$ é positiva. Conclua que o princípio variacional não vale quando M não é compacto.

10.4.5. Dados números $\alpha, \beta > 0$ tais que $\alpha + \beta < 1$, defina

$$g : [0, \alpha] \cup [1 - \beta, 1] \rightarrow [0, 1] \quad f(x) = \begin{cases} x/\alpha & \text{se } x \in [0, \alpha] \\ (x - 1)/\beta + 1 & \text{se } x \in [1 - \beta, 1]. \end{cases}$$

Seja $K \subset [0, 1]$ o conjunto de Cantor formado pelos pontos x tais que $g^n(x)$ está definido para todo $n \geq 0$ e seja $f : K \rightarrow K$ a restrição de g . Calcule a função $\psi : \mathbb{R} \rightarrow \mathbb{R}$ dada por $\psi(t) = P(f, -t \log g')$. Verifique que ψ é convexa e decrescente e admite um (único) zero em $(0, 1)$. Mostre que $h_\mu(f) < \int \log g' d\mu$ para toda probabilidade μ invariante por f .

10.4.6. Seja $f : M \rightarrow M$ uma transformação contínua em um espaço métrico compacto, tal que o conjunto das medidas invariantes ergódicas é finito. Mostre que todo potencial contínuo $\varphi : M \rightarrow \mathbb{R}$ admite algum estado de equilíbrio.

10.5 Estados de equilíbrio

Seja $f : M \rightarrow M$ uma transformação contínua num espaço métrico compacto. Uma medida invariante de probabilidade μ é dita um *estado de equilíbrio* para um potencial $\phi : M \rightarrow \mathbb{R}$ se ela realiza o supremo no princípio variacional, ou seja, se

$$h_\mu(f) + \int \phi d\mu = P(\phi, f) = \sup\{h_\nu(f) + \int \phi d\nu : \nu \in \mathcal{M}(f)\}.$$

No caso particular $\phi \equiv 0$ também dizemos que μ é uma *medida de máxima entropia*. Nesta seção estudamos as propriedades fundamentais do conjunto $\mathcal{E}(f, \phi)$ dos estados de equilíbrio. Começemos por alguns exemplos simples.

Exemplo 10.5.1. Se $f : M \rightarrow M$ tem entropia topológica nula, toda probabilidade invariante μ é medida de máxima entropia, já que $h_\mu(f) = 0 = h(f)$. Para um potencial qualquer $\phi : M \rightarrow \mathbb{R}$,

$$P(f, \phi) = \sup\left\{\int \phi d\nu : \nu \in \mathcal{M}(f)\right\}.$$

Logo, ν é estado de equilíbrio se, e somente se, ν maximiza a integral de ϕ . Como a função $\nu \mapsto \int \phi d\nu$ é contínua e $\mathcal{M}(f)$ é compacto, relativamente à topologia fraca*, máximos existem para todo potencial ϕ .

Exemplo 10.5.2. Seja $f_A : M \rightarrow M$ o endomorfismo linear induzido no toro \mathbb{T}^d por alguma matriz invertível A com coeficientes inteiros. Seja μ a medida de Haar em \mathbb{T}^d . Pelas Proposições 9.4.3 e 10.2.9,

$$h_\mu(f_A) = \sum_{i=1}^d \log^+ |\lambda_i| = h(f)$$

onde $\lambda_1, \dots, \lambda_d$ são os autovalores de A . Em particular, a medida de Haar é medida de máxima entropia para f .

Exemplo 10.5.3. Seja $\sigma : \Sigma \rightarrow \Sigma$ o deslocamento em $\Sigma = \{1, \dots, d\}^{\mathbb{N}}$ e seja μ a medida de Bernoulli dada por um vetor de probabilidade $p = (p_1, \dots, p_d)$. Como vimos no Exemplo 9.1.10,

$$h_\mu(\sigma, \mathcal{P}) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}^n) = \sum_{i=1}^d -p_i \log p_i.$$

Deixamos a cargo do leitor (Exercício 10.5.1) verificar que esta função atinge o seu máximo exatamente quando os p_i são todos iguais a $1/d$. Além disso, nesse caso $h_\mu(\sigma) = \log d$. Lembre também que (Exemplo 10.1.2) $h(\sigma) = \log d$. Portanto, a medida de Bernoulli dada pelo vetor $p = (1/d, \dots, 1/d)$ é a única medida de máxima entropia entre todas as medidas de Bernoulli. De fato, segue da teoria que desenvolveremos no Capítulo 12, que ela é a única medida de máxima entropia entre *todas* as medidas invariantes.

Começamos com a seguinte extensão do Princípio Variacional:

Proposição 10.5.4. Para todo potencial $\phi : M \rightarrow \mathbb{R}$,

$$P(f, \phi) = \sup\{h_\mu(f) + \int \phi d\mu : \mu \text{ invariante e ergódica para } f\}.$$

Demonstração. Considere o funcional definido $\Psi : \mathcal{M}_1(f) \rightarrow \mathbb{R}$ definido por $\Psi(\mu) = h_\mu(f) + \int \phi d\mu$. Para cada probabilidade invariante μ , seja $\{\mu_P : P \in \mathcal{P}\}$ a respectiva decomposição ergódica. Segue do Teorema 9.6.2 que

$$\Psi(\mu) = \int \Psi(\mu_P) d\hat{\mu}(P). \quad (10.5.1)$$

Esta relação implica que o supremo de Ψ sobre todas as probabilidades invariantes é menor ou igual que o supremo de Ψ sobre as probabilidades invariantes e ergódicas. Como a desigualdade oposta é trivial, segue que os dois supremos são iguais. Pelo princípio variacional (Teorema 10.4.1), o supremo de Ψ sobre todas as probabilidades invariantes é igual a $P(f, \phi)$. A conclusão da proposição segue imediatamente destas observações. \square

Proposição 10.5.5. *Suponha que $h(f) < \infty$. Então o conjunto dos estados de equilíbrio para qualquer potencial $\phi : M \rightarrow \mathbb{R}$ é um subconjunto convexo de $\mathcal{M}(f)$: mais precisamente, dado $t \in (0, 1)$ e dadas $\mu_1, \mu_2 \in \mathcal{M}(f)$,*

$$(1-t)\mu_1 + t\mu_2 \in \mathcal{E}(f, \phi) \quad \Leftrightarrow \quad \{\mu_1, \mu_2\} \subset \mathcal{E}(f, \phi).$$

Além disso, uma probabilidade invariante μ está em $\mathcal{E}(f, \phi)$ se, e somente se, quase toda componente ergódica de μ está em $\mathcal{E}(f, \phi)$.

Demonstração. A hipótese de que a entropia topológica é finita assegura que $P(f, \phi) < \infty$ para todo potencial ϕ , como vimos em (10.3.7). Consideremos o funcional $\Psi(\mu) = h_\mu(f) + \int \phi d\mu$ introduzido na prova da proposição anterior. Pela Proposição 9.6.1, este funcional é linear convexo:

$$\Psi((1-t)\mu_1 + t\mu_2) = (1-t)\Psi(\mu_1) + t\Psi(\mu_2)$$

para todo $t \in (0, 1)$ e quaisquer $\mu_1, \mu_2 \in \mathcal{M}(f)$. Então $\Psi((1-t)\mu_1 + t\mu_2)$ é igual ao supremo de Ψ se, e somente se, $\Psi(\mu_1)$ e $\Psi(\mu_2)$ são iguais a esse supremo. Isto prova a primeira parte da proposição. A prova da segunda parte é análoga: a relação (10.5.1) dá que $\Psi(\mu) = \sup \Psi$ se, e somente se, $\Psi(\mu_P) = \sup \Psi$ para $\hat{\mu}$ -quase todo P . \square

Corolário 10.5.6. *Se $\mathcal{E}(f, \phi)$ é não vazio então ele contém probabilidades invariantes ergódicas. Além disso, os elementos extremais do convexo $\mathcal{E}(f, \phi)$ são precisamente as medidas ergódicas contidas nele.*

Demonstração. Para provar a primeira afirmação basta considerar as componentes ergódicas de qualquer elemento de $\mathcal{E}(f, \phi)$. Passemos a provar a segunda afirmação. Se $\mu \in \mathcal{E}(f, \phi)$ é ergódica então (Proposição 4.3.2) μ é um elemento extremal de $\mathcal{M}(f)$. Com maior razão, μ é um elemento extremal de $\mathcal{E}(f, \phi)$. Reciprocamente, se $\mu \in \mathcal{E}(f, \phi)$ não é ergódica então podemos escrever

$$\mu = (1-t)\mu_1 + t\mu_2, \quad \text{com } 0 < t < 1 \text{ e } \mu_1, \mu_2 \in \mathcal{M}(f).$$

Pela Proposição 10.5.5 temos que $\mu_1, \mu_2 \in \mathcal{E}(f, \phi)$ e, portanto, μ não é elemento extremal do conjunto $\mathcal{E}(f, \phi)$. \square

Em geral, o conjunto dos estados de equilíbrio *pode* ser vazio. O primeiro exemplo foi dado por Gurevič. A construção que apresentamos a seguir está no livro de Walters [Wal82]:

Exemplo 10.5.7. Seja $f_n : M_n \rightarrow M_n$ uma sequência de homeomorfismos em espaços métricos compactos tal que a sequência $(h(f_n))_n$ é crescente e limitada. A ideia é construir um espaço métrico M e um homeomorfismo $f : M \rightarrow M$ com as seguintes características:

- M é a união de todos os M_n com um ponto adicional, que representamos por ∞ , com uma função distância tal que $(M_n)_n$ converge para ∞ .
- f fixa o ponto ∞ e a sua restrição a cada M_n coincide com f_n .

Veremos então que $f : M \rightarrow M$ não tem nenhuma medida de máxima entropia. Pela Proposição 10.5.5, basta mostrar que nenhuma probabilidade invariante e ergódica realiza $h(f)$.

Passemos aos detalhes. Seja d_n a distância em cada espaço métrico M_n . Não é restrição supor que $d_n \leq 1$ para todo n . Defina $M = \cup_n M_n \cup \{\infty\}$ e considere a distância d definida em M por:

$$d(x, y) = \begin{cases} n^{-2}d_n(x, y) & \text{se } x \in X_n \text{ e } y \in X_n \text{ com } n \geq 1 \\ \sum_{i=n}^m i^{-2} & \text{se } x \in X_n \text{ e } y \in X_m \text{ com } n \leq m \\ \sum_{i=n}^{\infty} i^{-2} & \text{se } x \in X_n \text{ e } y = \infty. \end{cases}$$

Deixamos ao cuidado do leitor verificar que d é realmente uma distância em M e que (M, d) é um espaço compacto. Seja $\beta = \sup_n h(f_n)$. Como os conjuntos $\{\infty\}$ e M_n , $n \geq 1$ são invariantes e cobrem todo o M , qualquer probabilidade ergódica μ de f satisfaz $\mu(\{\infty\}) = 1$, ou então $\mu(M_n) = 1$ para algum $n \geq 1$. No primeiro caso, $h_\mu(f) = 0$. No segundo, μ pode ser vista como uma probabilidade invariante de f_n e, portanto, $h_\mu(f) \leq h(f_n)$. Em particular, $h_\mu(f) < \beta$ para toda probabilidade μ invariante e ergódica para f . A observação anterior também mostra que

$$\begin{aligned} & \sup\{h_\mu(f) : \mu \text{ invariante e ergódica para } f\} \\ &= \sup_n \sup\{h_\mu(f) : \mu \text{ invariante e ergódica para } f_n\}. \end{aligned}$$

De acordo com a Proposição 10.5.4, isto significa que $h(f) = \sup_n h(f_n) = \beta$. Desta forma, fica provado que nenhuma medida invariante e ergódica de f realiza a entropia topológica.

Mas existe uma classe ampla de transformações para as quais a existência de estados de equilíbrio está garantida para todo potencial:

Lema 10.5.8. *Se a função entropia de f é semicontínua superiormente então $\mathcal{E}(f, \phi)$ é compacto (na topologia fraca*) e não vazio, para qualquer potencial ϕ .*

Demonstração. Seja $(\mu_n)_n$ uma sequência em $\mathcal{M}(f)$ tal que

$$h_{\mu_n}(f) + \int \phi d\mu_n \text{ converge para } P(f, \phi).$$

Por compacidade de $\mathcal{M}_1(f)$ (Teorema 2.1.5), a sequência admite algum ponto de acumulação μ . A semicontinuidade da entropia, juntamente com a continuidade da integral, implica que

$$h_\mu(f) + \int \phi d\mu \geq \liminf_n h_{\mu_n}(f) + \int \phi d\mu_n = P(f, \phi).$$

Portanto, μ é um estado de equilíbrio, tal como afirmado. Analogamente, tomando uma sequência qualquer $(\nu_n)_n$ em $\mathcal{E}(f, \phi)$ vemos que qualquer ponto de acumulação ν é um estado de equilíbrio. Isto mostra que $\mathcal{E}(f, \phi)$ é fechado e, conseqüentemente, compacto. \square

Corolário 10.5.9. *Suponha que $f : M \rightarrow M$ é uma transformação contínua expansiva num espaço métrico compacto M . Então todo potencial $\phi : M \rightarrow \mathbb{R}$ admite algum estado de equilíbrio.*

Demonstração. Basta combinar o Corolário 9.2.15 com o Lema 10.5.8. \square

A unicidade é um problema mais complexo. É muito fácil exibir transformações com infinitos estados de equilíbrio ergódicos. Por exemplo, seja $f : S^1 \rightarrow S^1$ um homeomorfismo do círculo com infinitos pontos fixos. As medidas de Dirac suportadas nesses pontos são probabilidades invariantes ergódicas. Como a entropia topológica $h(f)$ é igual a zero, essas medidas são estados de equilíbrio para todo potencial que atinja o seu máximo nesses pontos (Exemplo 10.5.1).

No entanto este tipo de exemplo é trivial, porque a transformação não é transitiva. É natural perguntar se uma propriedade de indivisibilidade, tal como transitividade ou mistura topológica, garante unicidade do estado de equilíbrio. Resulta que isso não é verdade. O primeiro contra-exemplo (chamado deslocamento de Dyck) foi exibido por Krieger [Kri75]. A seguir apresentaremos uma construção particularmente transparente e flexível, devida a Haydn [Hay]. Outros exemplos interessantes foram estudados por Hofbauer [Hof77].

Exemplo 10.5.10 (Deslocamento verde e amarelo). Tome $X = \{0, 1, 2, 3, 4\}$ e considere os subconjuntos $V = \{1, 2\}$ (símbolos *verdes*) e $A = \{3, 4\}$ (símbolos *amarelos*). Vamos exibir um compacto $\mathcal{H} \subset X^{\mathbb{Z}}$ invariante pelo deslocamento em $X^{\mathbb{Z}}$ e tal que a restrição $\sigma : \mathcal{H} \rightarrow \mathcal{H}$ é topologicamente misturadora e admite duas medidas invariantes mutuamente singulares, μ_v e μ_a , tais que

$$h_{\mu_v}(\sigma) = h_{\mu_a}(\sigma) = \log 2 = h(\sigma).$$

Passemos a descrever o exemplo. Por definição, $\mathcal{H} = V^{\mathbb{Z}} \cup A^{\mathbb{Z}} \cup \mathcal{H}_0$, onde \mathcal{H}_0 está formado pelas sequências $x \in X^{\mathbb{Z}}$ que satisfazem a seguinte regra: sempre que um bloco com m símbolos de uma dada cor, verde ou amarela, é seguido por um bloco com n símbolos da outra cor, amarela ou verde, os dois estão separados por um bloco formado por, pelo menos, $m + n$ zeros. Em outras

palavras, são admitidas as seguintes configurações nas sequências $x \in \mathcal{H}_0$:

$$x = (\dots, 0, v_1, \dots, v_m, \underbrace{0, \dots, 0}_k, a_1, \dots, a_n, 0 \dots) \quad \text{ou}$$

$$x = (\dots, 0, a_1, \dots, a_m, \underbrace{0, \dots, 0}_k, v_1, \dots, v_n, 0 \dots)$$

com $v_i \in V$, $a_j \in A$ e $k \geq m + n$. Observe que uma sequência $x \in \mathcal{H}_0$ pode começar e/ou terminar com um bloco infinito de zeros mas não pode começar nem terminar com um bloco infinito de uma dada cor (verde ou amarela). É claro que \mathcal{H} é invariante pelo deslocamento. Haydn [Hay] provou que (veja o Exercício 10.5.6):

- (a) o deslocamento $\sigma : \mathcal{H} \rightarrow \mathcal{H}$ é topologicamente misturador;
- (b) $h(\sigma) = \log 2$.

Sabemos que $V^{\mathbb{Z}}$ e $A^{\mathbb{Z}}$ suportam medidas de Bernoulli μ_v e μ_a com entropia igual a $\log 2$. Então μ_v e μ_a são medidas de máxima entropia para $\sigma : \mathcal{H} \rightarrow \mathcal{H}$ e são mutuamente singulares.

Claramente, esta construção pode ser modificada para obter sistemas com qualquer número de medidas ergódicas de máxima entropia. Haydn [Hay] também mostrou como adaptá-la para construir exemplos de multiplicidade de estados de equilíbrio para outros potenciais.

Nos Capítulos 11 e 12 estudaremos uma classe de transformações, chamadas *expansoras*, para as quais tem-se unicidade do estado de equilíbrio para todo potencial Hölder. Em particular, essas transformações são *intrinsecamente ergódicas*, isto é, elas admitem uma única medida de máxima entropia.

10.5.1 Exercícios

10.5.1. Mostre que, entre as medidas de Bernoulli do deslocamento $\sigma : \Sigma \rightarrow \Sigma$ no espaço $\Sigma = \{1, \dots, d\}^{\mathbb{Z}}$, a de maior entropia é dada pelo vetor de probabilidade $(1/d, \dots, 1/d)$.

10.5.2. Seja $\sigma : \Sigma \rightarrow \Sigma$ o deslocamento em $\Sigma = \{1, \dots, d\}^{\mathbb{Z}}$ e seja ϕ um potencial localmente constante em Σ (ou seja, ϕ é constante em cada cilindro $[0; i]$). Calcule $P(f, \phi)$ e mostre que existe algum estado de equilíbrio que é uma medida de Bernoulli.

10.5.3. Seja $\sigma : \Sigma \rightarrow \Sigma$ o deslocamento em $\Sigma = \{1, \dots, d\}^{\mathbb{N}}$. Dizemos que uma probabilidade invariante μ é um *estado de Gibbs* para um potencial φ relativamente a σ se existem $P \in \mathbb{R}$ e $K > 0$ tais que

$$K^{-1} \leq \frac{\mu(C)}{\exp(\varphi_n(x) - nP)} \leq K \quad (10.5.2)$$

para todo cilindro $C = [0; i_0, \dots, i_{n-1}]$ e qualquer $x \in C$. Prove que se μ é um estado de Gibbs então $h_\mu(\sigma) + \int \varphi d\mu$ é igual à constante P em (10.5.2).

Portanto, μ é um estado de equilíbrio se, e somente se, $P = P(\sigma, \varphi)$. Prove que para cada escolha da constante P existe no máximo um estado de Gibbs ergódico.

10.5.4. Sejam $f : M \rightarrow M$ uma transformação contínua em um espaço métrico compacto e $\phi : M \rightarrow \mathbb{R}$ uma função contínua. Se μ é um estado de equilíbrio para ϕ , então o funcional $F_\mu : C^0(M) \rightarrow \mathbb{R}$ definido por $F_\mu(\psi) = \int \psi d\mu$ é tal que $F_\mu(\psi) \leq P(f, \phi + \psi) - P(f, \phi)$ para todo $\psi \in C^0(M)$. Conclua que se a função pressão $P(f, \cdot) : C^0(M) \rightarrow \mathbb{R}$ é derivável em todas as direções num ponto ϕ , então ϕ admite no máximo um estado de equilíbrio.

10.5.5. Sejam $f : M \rightarrow M$ uma transformação contínua em um espaço métrico compacto. Mostre que o subconjunto das funções $\phi \in C^0(M)$ para as quais existe um único estado de equilíbrio é residual em $C^0(M)$.

10.5.6. Verifique as afirmações (a) e (b) do Exemplo 10.5.10.

Capítulo 11

Transformações Expansoras

As transformações $f : M \rightarrow M$ que estudaremos nos dois capítulos finais deste livro têm como propriedade característica expandirem distâncias entre pontos próximos: elas são tais que existe uma constante $\sigma > 1$ tal que

$$d(f(x), f(y)) \geq \sigma d(x, y)$$

sempre que a distância entre os pontos x e y é pequena (daremos a definição precisa daqui a pouco). Esta classe de transformações tem um papel fundamental na teoria, por diversas razões.

Por um lado, como veremos, as transformações expansoras exibem um comportamento dinâmico muito rico, tanto do ponto de vista métrico e topológico quanto do ponto de vista ergódico. Desta forma, elas nos fornecem um contexto natural e interessante para implementar muitas das ideias e métodos que foram introduzidos ao longo do texto. Por outro lado, as transformações expansoras permitem estabelecer paradigmas que são úteis para compreender muitos outros sistemas, tecnicamente mais complexos.

Um caso particular importante são as transformações diferenciáveis em variedades, caracterizadas pelo fato de que

$$\|Df(x)v\| \geq \sigma \|v\|$$

para todo $x \in M$ e todo vetor v tangente a M no ponto x . Focaremos este caso na Seção 11.1. O principal resultado (Teorema 11.1.2) é que, sob a hipótese de que o jacobiano $\det Df$ é Hölder, a transformação f admite uma única probabilidade invariante absolutamente contínua relativamente à medida de Lebesgue. Além disso, essa probabilidade é ergódica e positiva nos abertos de M .

Na Seção 11.2 estenderemos a noção de transformação expansora para espaços métricos e daremos uma descrição global da dinâmica topológica dessas transformações baseada no estudo dos seus pontos periódicos. O principal objetivo é mostrar que a dinâmica global sempre pode ser reduzida ao caso topologicamente exato (Teoremas 11.2.15). Na Seção 11.3 complementaremos essa análise mostrando que a entropia topológica destas transformações está dada pela taxa de crescimento do número de pontos periódicos.

No Capítulo 12 prosseguiremos este estudo do ponto de vista ergódico, desenvolvendo o chamado *formalismo termodinâmico* das transformações expansoras.

11.1 Transformações expansoras em variedades

Seja M uma variedade compacta e seja $f : M \rightarrow M$ uma transformação de classe C^1 . Dizemos que f é *expansora* se existe $\sigma > 1$ e alguma métrica Riemanniana em M tal que

$$\|Df(x)v\| \geq \sigma\|v\| \quad \text{para todo } x \in M \text{ e todo } v \in T_x M. \quad (11.1.1)$$

Em particular, f é um difeomorfismo local: a condição (11.1.1) implica que $Df(x)$ é um isomorfismo para todo $x \in M$. No que segue chamaremos *medida de Lebesgue* em M à medida de volume m induzida por tal métrica Riemanniana. A escolha precisa da métrica não será muito importante, uma vez que as medidas de volume associadas a distintas métricas Riemannianas são todas equivalentes.

Exemplo 11.1.1. Seja $f_A : \mathbb{T}^d \rightarrow \mathbb{T}^d$ o endomorfismo linear do toro induzido por uma matriz invertível A com coeficientes inteiros. Suponha que todos os autovalores $\lambda_1, \dots, \lambda_d$ de F têm módulo maior que 1. Então, dado qualquer $1 < \sigma < \inf_i |\lambda_i|$, existe um produto interno em \mathbb{R}^d relativamente ao qual $\|Av\| \geq \sigma\|v\|$. (Suponha que os autovalores são reais. Considere uma base de \mathbb{R}^d que coloca A em forma canônica de Jordan: $A = D + \varepsilon N$ com D diagonal e N nilpotente. O produto interno relativamente ao qual essa base é ortonormal tem a propriedade requerida, se $\varepsilon > 0$ é suficientemente pequeno. O leitor não terá dificuldade para estender este argumento ao caso em que existem autovalores complexos.) Isto mostra que a transformação f_A é expansora.

É claro da definição que toda transformação que esteja suficientemente próxima, na topologia C^1 , de uma transformação expansora ainda é expansora. Portanto, a observação no Exemplo 11.1.1 fornece todo um aberto de exemplos de transformações expansoras. Um resultado clássico de Michael Shub [Shu69] afirma que vale uma espécie de recíproca: toda transformação expansora no toro \mathbb{T}^d é topologicamente conjugada a um endomorfismo linear expansor f_A .

Dada uma probabilidade invariante μ , chamamos *bacia* de μ o conjunto $B(\mu)$ dos pontos $x \in M$ tais que

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

para toda função contínua $\varphi : M \rightarrow \mathbb{R}$. Note que a bacia sempre é um conjunto invariante. Se μ é ergódica, então $B(\mu)$ tem μ -medida total (Exercício 4.1.5).

Teorema 11.1.2. *Seja $f : M \rightarrow M$ uma transformação diferenciável expansora numa variedade compacta conexa M . Suponha que o jacobiano $x \mapsto \det Df(x)$ é Hölder. Então f admite uma única probabilidade invariante μ absolutamente contínua com relação à medida de Lebesgue m . Além disso, μ é ergódica, o*

seu suporte coincide com M e a sua bacia tem medida de Lebesgue total na variedade.

A estratégia da demonstração do Teorema 11.1.2 é a seguinte. É fácil ver (Exercício 11.1.1) que a pré-imagem por f de um conjunto com medida de Lebesgue m nula também tem medida de Lebesgue nula. Isto significa que a imagem $f_*\nu$ por f de qualquer medida ν absolutamente contínua com relação a m também é absolutamente contínua com relação a m . Em particular, a n -ésima imagem $f_*^n m$ é sempre absolutamente contínua com relação à medida de Lebesgue m .

Provaremos na Proposição 11.1.7 que a *densidade* (ou seja, a derivada de Radon-Nikodym) de cada $f_*^n m$ com relação a m é limitada por uma constante independente de $n \geq 1$. Deduziremos desse fato que todo ponto de acumulação, na topologia fraca*, da sequência

$$\frac{1}{n} \sum_{j=0}^{n-1} f_*^j m$$

é uma probabilidade invariante absolutamente contínua com relação à medida de Lebesgue, com densidade limitada pela mesma constante.

Um argumento adicional provará que tal ponto de acumulação é único e que ele satisfaz todas as propriedades no enunciado do teorema.

11.1.1 Lema de distorção

Dando início à demonstração do Teorema 11.1.2, provaremos o seguinte fato elementar (note que a condição (11.1.1) significa que no caso de uma transformação expansora podemos tomar $\sigma > 1$):

Lema 11.1.3. *Seja $f : M \rightarrow M$ um difeomorfismo local numa variedade Riemanniana compacta e seja $\sigma > 0$ um minorante para a conorma $\|Df^{-1}\|^{-1}$ da derivada de f . Então existe $k \geq 1$ tal que todo ponto $y \in M$ tem exatamente k pré-imagens por f . Além disso, existe $\rho > 0$ tal que, para qualquer pré-imagem x de um ponto $y \in M$, existe uma aplicação $h : B(y, \rho) \rightarrow M$ de classe C^1 tal que $f \circ h = \text{id}$, $h(y) = x$ e*

$$d(h(y_1), h(y_2)) \leq \sigma^{-1} d(y_1, y_2) \quad \text{para todo } y_1, y_2 \in B(y, \rho). \quad (11.1.2)$$

Demonstração. Como a derivada Df é um isomorfismo em todo ponto, dado qualquer $x \in M$ existe $\rho > 0$ tal que f envia alguma vizinhança $V(x)$ de x difeomorficamente sobre a bola de raio ρ em torno de $y = f(x)$. Por compacidade, podemos tomar ρ independente de x (Exercício 11.2.3). Então o número de pré-imagens de qualquer $y \in M$ deve ser limitado. Logo, existe $n \geq 1$ tal que

$$B_n = \{y \in M : \#f^{-1}(y) = n\}$$

é não vazio. Como este conjunto é aberto e fechado e M é conexo, segue que $B_n = M$. Em outras palavras, todos os pontos têm exatamente n pré-imagens.

Finalmente, seja $h = (f \mid V(x))^{-1}$. Então $\|Dh(z)\| = \|Df(h(z))^{-1}\| \leq \sigma^{-1}$ para todo z no domínio de h . Pelo teorema da média, isto implica que h verifica a propriedade (11.1.2). \square

Transformações h como neste enunciado são chamadas *ramos inversos* de f . O número $k = \#f^{-1}(y)$ de pré-imagens de um ponto qualquer de M é chamado *grau* da aplicação.

Agora suponha que f é uma transformação expansora. Conforme já observamos, neste caso podemos tomar $\sigma > 1$. Então, a propriedade (11.1.2) significa que os ramos inversos são contrações, com taxa uniforme de contração. Em particular, podemos definir ramos inversos h^n de qualquer iterado f^n , $n \geq 1$, da seguinte forma. Dado $y \in M$ e $x \in f^{-n}(y)$, sejam h_1, \dots, h_n ramos inversos de f com

$$h_j(f^{n-j+1}(x)) = f^{n-j}(x)$$

para todo $1 \leq j \leq n$. Como cada h_j é uma contração, a sua imagem está contida numa bola de raio menor que ρ em torno de $f^{n-j}(x)$. Então $h^n = h_n \circ \dots \circ h_1$ está bem definida na bola de raio ρ em torno de y . É claro que $f^n \circ h^n = \text{id}$ e $h^n(y) = x$. Além disso, cada h^n é uma contração:

$$d(h^n(y_1), h^n(y_2)) \leq \sigma^{-n} d(y_1, y_2) \quad \text{para todo } y_1, y_2 \in B(y, \rho).$$

Lema 11.1.4. *Se $f : M \rightarrow M$ é uma transformação expansora C^1 numa variedade compacta então f é expansiva.*

Demonstração. Pelo Lema 11.1.3, existe $\rho > 0$ tal que, para qualquer pré-imagem x de um ponto $y \in M$, existe uma aplicação $h : B(y, \rho) \rightarrow M$ de classe C^1 tal que $f \circ h = \text{id}$ e $h(y) = x$ e

$$d(h(y_1), h(y_2)) \leq \sigma^{-1} d(y_1, y_2) \quad \text{para todo } y_1, y_2 \in B(y, \rho).$$

Assim, se $d(f^n(x), f^n(y)) \leq \rho$ para todo $n \geq 0$ então

$$d(x, y) \leq \sigma^{-n} d(f^n(x), f^n(y)) \leq \sigma^{-n} \rho,$$

o que implica imediatamente que $x = y$. \square

O próximo resultado fornece um bom controle da distorção de iterados de f e seus ramos inversos, que é crucial para a demonstração do Teorema 11.1.2. Este é o único lugar da demonstração onde se usa a hipótese de que o jacobiano $x \mapsto \det Df(x)$ é Hölder. Note que, uma vez que f é um difeomorfismo local e M é compacta, o jacobiano é limitado de zero e de infinito. Portanto, o logaritmo $\log |\det Df|$ também é Hölder: existem $C_0 > 0$ e $\nu > 0$ tais que

$$\left| \log |\det Df(x)| - \log |\det Df(y)| \right| \leq C_0 d(x, y)^\nu \quad \text{para quaisquer } x, y \in M.$$

Proposição 11.1.5 (Lema de distorção). *Existe $C_1 > 0$ tal que, dado qualquer $n \geq 1$, qualquer $y \in M$, e qualquer ramo inverso $h^n : B(y, \rho) \rightarrow M$ de f^n , tem-se*

$$\log \frac{|\det Dh^n(y_1)|}{|\det Dh^n(y_2)|} \leq C_1 d(y_1, y_2)^\nu \leq C_1 (2\rho)^\nu$$

para todo $y_1, y_2 \in B(y, \rho)$.

Demonstração. Escrevemos h^n como composição $h^n = h_n \circ \cdots \circ h_1$ de ramos inversos de f . Também escrevemos $h^i = h_i \circ \cdots \circ h_1$ para $1 \leq i < n$, bem como $h^0 = \text{id}$. Então

$$\log \frac{|\det Dh^n(y_1)|}{|\det Dh^n(y_2)|} = \sum_{i=1}^n \log |\det Dh_i(h^{i-1}(y_1))| - \log |\det Dh_i(h^{i-1}(y_2))|.$$

Note que $\log |\det Dh_i| = -\log |\det Df| \circ h_i$ e que, pelo Lema 11.1.3, cada h_j é uma σ^{-1} -contração. Logo,

$$\log \frac{|\det Dh^n(y_1)|}{|\det Dh^n(y_2)|} \leq \sum_{i=1}^n C_0 d(h^i(y_1), h^i(y_2))^\nu \leq \sum_{i=1}^n C_0 \sigma^{-i\nu} d(y_1, y_2)^\nu.$$

Portanto, para provar o lema basta tomar $C_1 = C_0 \sum_{i=1}^{\infty} \sigma^{-i\nu}$. \square

O significado geométrico desta proposição fica ainda mais transparente no seguinte corolário:

Corolário 11.1.6. *Existe $C_2 > 0$ tal que, para todo $y \in M$ e quaisquer conjuntos mensuráveis $B_1, B_2 \subset B(y, \rho)$,*

$$\frac{1}{C_2} \frac{m(B_1)}{m(B_2)} \leq \frac{m(h^n(B_1))}{m(h^n(B_2))} \leq C_2 \frac{m(B_1)}{m(B_2)}.$$

Demonstração. Tome $C_2 = \exp(2C_1(2\rho)^\nu)$. Segue da Proposição 11.1.5 que

$$\begin{aligned} m(h^n(B_1)) &= \int_{B_1} |\det Dh^n| dm \leq \exp(C_1(2\rho)^\nu) |\det Dh^n(y)| m(B_1) \quad \text{e} \\ m(h^n(B_2)) &= \int_{B_2} |\det Dh^n| dm \geq \exp(-C_1(2\rho)^\nu) |\det Dh^n(y)| m(B_2) \end{aligned}$$

Dividindo as duas desigualdades, vem que

$$\frac{m(h^n(B_1))}{m(h^n(B_2))} \leq C_2 \frac{m(B_1)}{m(B_2)}.$$

Invertendo os papéis de B_1 e B_2 obtemos a outra desigualdade. \square

O próximo resultado, que é uma consequência do lema de distorção, mostra que as imagens $f_*^n m$ têm densidades uniformemente limitadas:

Proposição 11.1.7. *Existe $C_2 > 0$ tal que $(f_*^n m)(B) \leq C_2 m(B)$ para todo conjunto mensurável $B \subset M$ e todo $n \geq 1$.*

Demonstração. Podemos, sem restrição, supor que B está contido em alguma bola $B_0 = B(z, \rho)$ de raio ρ centrada em algum ponto $z \in M$. Usando o Corolário 11.1.6 vemos que

$$\frac{m(h^n(B))}{m(h^n(B_0))} \leq C_2 \frac{m(B)}{m(B_0)},$$

para todo ramo inverso h^n de f^n no ponto z . Além disso, também temos que $(f_*^n m)(B) = m(f^{-n}(B))$ é a soma de $m(h^n(B))$ sobre todos os ramos inversos, e analogamente para B_0 . Deste modo, obtemos que

$$\frac{(f_*^n m)(B)}{(f_*^n m)(B_0)} \leq C_2 \frac{m(B)}{m(B_0)}.$$

Claro que $(f_*^n m)(B_0) \leq (f_*^n m)(M) = 1$. Além disso, a medida de Lebesgue das bolas com um raio fixado ρ está limitada de zero por alguma constante $\alpha_0 > 0$ que só depende de ρ . Então, para obter a conclusão da proposição basta tomar $C_2 = \exp(C_1(2\rho)^\nu)/\alpha_0$. \square

Também precisamos do resultado auxiliar enunciado a seguir. Lembre que, dada uma função φ e uma medida ν , representamos por $\varphi\nu$ a medida definida por $(\varphi\nu)(B) = \int_B \varphi d\nu$.

Lema 11.1.8. *Seja ν uma probabilidade num espaço métrico compacto X , e seja $\varphi : X \rightarrow [0, +\infty)$ uma função integrável com respeito a ν . Seja $\mu_i, i \geq 1$, uma sequência de probabilidades em X convergindo para uma probabilidade μ na topologia fraca*. Se $\mu_i \leq \varphi\nu$ para todo $i \geq 1$ então $\mu \leq \varphi\nu$.*

Demonstração. Seja B um conjunto mensurável qualquer. Para cada $\varepsilon > 0$, seja K_ε um subconjunto compacto de B tal que $\mu(B \setminus K_\varepsilon)$ e $(\varphi\nu)(B \setminus K_\varepsilon)$ são ambos menores que ε . Então seja A_ε uma vizinhança aberta de K_ε da forma $A_\varepsilon = \{z : d(z, K_\varepsilon) < r\}$, com $r > 0$ suficientemente pequeno para que a medida de $A_\varepsilon \setminus K_\varepsilon$ seja menor que ε , tanto para μ como para $\varphi\nu$. O conjunto dos valores de r para os quais o bordo de A_ε tem μ -medida positiva é finito ou enumerável (Exercício A.3.2). Logo, a menos de mudar r ligeiramente, se necessário, podemos supor que o bordo de A_ε tem medida zero. Então, $\mu = \lim \mu_i$ implica $\mu(A_\varepsilon) = \lim \mu_i(A_\varepsilon) \leq (\varphi\nu)(A_\varepsilon)$. Fazendo $\varepsilon \rightarrow 0$ obtemos que $\mu(B) \leq (\varphi\nu)(B)$. \square

Aplicando este lema na nossa situação, obtemos

Corolário 11.1.9. *Todo ponto de acumulação μ da sequência $n^{-1} \sum_{j=0}^{n-1} f_*^j m$ é uma probabilidade invariante para f absolutamente contínua com relação à medida de Lebesgue.*

Demonstração. Tomemos φ constante igual a C_2 e $\nu = m$. Tomemos também $\mu_i = n_i^{-1} \sum_{j=0}^{n_i-1} f_*^j m$, para qualquer subsequência $(n_i)_i$ tal que $(\mu_i)_i$ converge para uma medida μ . A Proposição 11.1.7 garante que $\mu_i \leq \varphi\nu$. Então também temos $\mu \leq \varphi\nu = C_2 m$, pelo Lema 11.1.8. Isto implica que $\mu \ll m$, com densidade limitada por C_2 . \square

11.1.2 Existência de medidas ergódicas

Agora vamos mostrar que a medida μ que acabamos de construir é a única probabilidade invariante absolutamente contínua com relação à medida de Lebesgue e, além disso, é ergódica para f .

Começamos por fixar uma partição $\mathcal{P}_0 = \{U_1, \dots, U_s\}$ de M em regiões com interior não vazio e diâmetro menor que ρ . Então, para cada $n \geq 1$, definimos \mathcal{P}_n como sendo a partição de M que consiste das imagens de cada um dos U_i , $1 \leq i \leq s$, pelos respectivos ramos inversos de f^n . O diâmetro da partição \mathcal{P}_n , ou seja, o supremo dos diâmetros dos seus elementos, é menor que $\rho\sigma^{-n}$.

Lema 11.1.10. *Seja \mathcal{P}_n , $n \geq 1$, uma sequência de partições num espaço métrico compacto, com diâmetros convergindo para zero quando $n \rightarrow \infty$. Seja ν uma probabilidade nesse espaço, e seja B qualquer conjunto mensurável com $\nu(B) > 0$. Então existem $V_n \in \mathcal{P}_n$, para $n \geq 1$, tais que*

$$\nu(V_n) > 0 \quad e \quad \frac{\nu(B \cap V_n)}{\nu(V_n)} \rightarrow 1 \quad \text{quando } n \rightarrow \infty.$$

Demonstração. Dado qualquer $0 < \varepsilon < \nu(B)$, seja K_ε um subconjunto compacto de B com $\nu(B \setminus K_\varepsilon) < \varepsilon$. Como o diâmetro das partições converge para zero, a medida da união $K_{\varepsilon,n}$ de todos os elementos de \mathcal{P}_n que intersectam K_ε satisfaz $\nu(K_{\varepsilon,n} \setminus K_\varepsilon) < \varepsilon$ para todo n suficientemente grande. Suponha que

$$\nu(K_\varepsilon \cap V_n) \leq \frac{\nu(B) - \varepsilon}{\nu(B) + \varepsilon} \nu(V_n)$$

para todo $V_n \in \mathcal{P}_n$ que intersecta K_ε . Seguiria que

$$\begin{aligned} \nu(K_\varepsilon) &\leq \sum_{V_n} \nu(K_\varepsilon \cap V_n) \leq \sum_{V_n} \frac{\nu(B) - \varepsilon}{\nu(B) + \varepsilon} \nu(V_n) = \frac{\nu(B) - \varepsilon}{\nu(B) + \varepsilon} \nu(K_{\varepsilon,n}) \\ &\leq \frac{\nu(B) - \varepsilon}{\nu(B) + \varepsilon} (\nu(K_\varepsilon) + \varepsilon) \leq \nu(B) - \varepsilon < \nu(K_\varepsilon). \end{aligned}$$

Esta contradição mostra que deve existir algum $V_n \in \mathcal{P}_n$ tal que

$$\nu(V_n) \geq \nu(B \cap V_n) \geq \nu(K_\varepsilon \cap V_n) > \frac{\nu(B) - \varepsilon}{\nu(B) + \varepsilon} \nu(V_n)$$

e isto implica $\nu(V_n) > 0$. Fazendo $\varepsilon \rightarrow 0$ obtemos a conclusão do lema. \square

Nos enunciados a seguir, diremos que um conjunto mensurável $A \subset M$ é invariante por $f : M \rightarrow M$ se $f^{-1}(A) = A$ a menos de conjunto com medida de Lebesgue nula. Usaremos o fato (Exercício 11.1.1) de que se A é conjunto invariante então $f(A) = A$ a menos de conjunto com medida de Lebesgue nula.

Lema 11.1.11. *Se $A \subset M$ é um conjunto invariante para f com medida de Lebesgue positiva, então A tem medida de Lebesgue total em algum $U_i \in \mathcal{P}_0$, ou seja, existe $1 \leq i \leq s$ tal que $m(U_i \setminus A) = 0$.*

Demonstração. Pelo Lema 11.1.10, podemos escolher $V_n \in \mathcal{P}_n$ de tal forma que $m(V_n \setminus A)/m(V_n)$ converge para zero quando $n \rightarrow \infty$. Seja $U_{i(n)} = f^n(V_n)$.

Pela Proposição 11.1.5 aplicada ao ramo inverso de f^n que envia $U_{i(n)}$ em V_n , concluímos que

$$\frac{m(U_{i(n)} \setminus A)}{m(U_{i(n)})} \leq \frac{m(f^n(V_n \setminus A))}{m(f^n(V_n))} \leq \exp(C_1(2\rho)^\nu) \frac{m(V_n \setminus A)}{m(V_n)}$$

também converge para zero. Como \mathcal{P}_0 é finito, deve existir $1 \leq i \leq s$ tal que $i(n) = i$ para infinitos valores de n . Então, $m(U_i \setminus A) = 0$. \square

Corolário 11.1.12. *A transformação $f : M \rightarrow M$ admite alguma probabilidade invariante ergódica e absolutamente contínua com relação à medida de Lebesgue.*

Demonstração. Segue do lema anterior que existem no máximo $s = \#\mathcal{P}_0$ conjuntos invariantes com medida de Lebesgue positiva disjuntos dois-a-dois. Portanto, M pode ser particionado num número finito de conjuntos invariantes A_1, \dots, A_r , $r \leq s$ com medida de Lebesgue positiva e que são minimais, no sentido de que não existem subconjuntos invariantes $B_i \subset A_i$ satisfazendo $0 < m(B_i) < m(A_i)$. Dada qualquer medida invariante absolutamente contínua μ , existe algum i tal que $\mu(A_i) > 0$. Então a restrição normalizada

$$\mu_i(B) = \frac{\mu(B \cap A_i)}{\mu(A_i)}$$

de μ a cada A_i é invariante e absolutamente contínua. Além disso, μ_i é ergódica, porque A_i é minimal. \square

11.1.3 Unicidade e conclusão da prova

O argumento anterior também mostra que existe apenas um número finito de probabilidades ergódicas e absolutamente contínuas. O último passo da demonstração é mostrar que, de fato, tal probabilidade é única. Para isso usamos o fato de que f é *topologicamente exata*:

Lema 11.1.13. *Dado qualquer aberto não vazio $U \subset M$, existe $N \geq 1$ tal que $f^N(U) = M$.*

Demonstração. Seja $x \in U$ e $r > 0$ tais que a bola de raio r centrada em x está contida em U . Dado qualquer $n \geq 1$, suponha que $f^n(U)$ não cobre toda a variedade. Então existe alguma curva γ ligando $f^n(x)$ a um ponto $y \in M \setminus f^n(U)$, e essa curva pode ser tomada com comprimento menor que $\text{diam } M + 1$. Levantando γ pelo difeomorfismo local f^n , obtemos uma curva γ_n ligando x a um ponto $y_n \in M \setminus U$. Então $r \leq \text{comp}(\gamma_n) \leq \sigma^{-n}(\text{diam } M + 1)$. Isto dá uma cota superior para o valor possível de n . Logo, $f^n(U) = M$ para todo n suficientemente grande, como afirmamos. \square

Corolário 11.1.14. *Se $A \subset M$ é um conjunto invariante com medida de Lebesgue positiva, então A tem medida de Lebesgue total na variedade M .*

Demonstração. Seja U o interior de um conjunto U_i como no Lema 11.1.11, e seja $N \geq 1$ tal que $f^N(U) = M$. Então $m(U \setminus A) = 0$, e portanto $M \setminus A = f^N(U) \setminus f^N(A) \subset f^N(U \setminus A)$ também tem medida de Lebesgue zero, pois f é difeomorfismo local. \square

O próximo enunciado completa a demonstração do Teorema 11.1.2:

Corolário 11.1.15. *Seja μ qualquer probabilidade invariante absolutamente contínua. Então μ é ergódica e a sua bacia $B(\mu)$ tem medida de Lebesgue total em M . Consequentemente, μ é única. Além disso, o seu suporte é toda a variedade M .*

Demonstração. Se A é um subconjunto invariante qualquer então, pelo Corolário 11.1.14, A tem medida de Lebesgue zero ou A^c tem medida de Lebesgue zero. Uma vez que μ é absolutamente contínua, segue que $\mu(A) = 0$ ou $\mu(A^c) = 0$. Isto prova que μ é ergódica, logo $\mu(B(\mu)) = 1$. Então $B(\mu)$ é um conjunto invariante com medida de Lebesgue positiva e, consequentemente, deve ter medida de Lebesgue total. Analogamente, como o suporte de μ é um conjunto compacto invariante, ele tem que coincidir com M .

Finalmente, sejam μ e ν duas probabilidades invariantes absolutamente contínuas quaisquer. Segue do Corolário 11.1.15 que elas são ergódicas e que as suas bacias se intersectam. Dado qualquer ponto x na interseção $B(\mu) \cap B(\nu)$,

$$\mu = \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \delta_{f^j(x)} = \nu$$

(o limite é na topologia fraca*). Por unicidade do limite, segue que $\mu = \nu$. \square

Em geral, dizemos que uma probabilidade invariante μ de um difeomorfismo local f é uma *medida física* se a sua bacia tem medida de Lebesgue positiva. Segue do Corolário 11.1.15 que no presente contexto existe uma única medida física, que é a medida invariante absolutamente contínua μ , e a sua bacia tem mesmo medida de Lebesgue total. Este último fato pode ser reescrito na seguinte forma:

$$\frac{1}{n} \sum_{j=0}^{n-1} \delta_{f^j(x)} \rightarrow \mu \quad \text{para Lebesgue quase todo } x.$$

No Capítulo 12 reobteremos esta probabilidade invariante absolutamente contínua μ por outra via (Proposição 12.1.20), a qual também mostra que a densidade $h = d\mu/dm$ é Hölder e está limitada de zero. Em particular, μ é equivalente à medida de Lebesgue m , não apenas absolutamente contínua. Veremos na Seção 12.1.7 que o sistema (f, μ) é exato, não apenas ergódico. Além disso, (Lema 12.1.12) o seu jacobiano é dado por $J_\mu f = |\det Df|(h \circ f)/h$. Então, a fórmula de Rokhlin (Teorema 9.7.3) dá que

$$h_\mu(f) = \int \log J_\mu f \, d\mu = \int \log |\det Df| \, d\mu + \int \log(h \circ f) \, d\mu - \int \log h \, d\mu,$$

ou seja (porque μ é invariante), $h_\mu(f) = \int \log |\det Df| d\mu$.

Na verdade, os fatos que acabamos de anunciar já podem ser provados com os métodos de que dispomos neste momento. A título de ilustração, convidamos o leitor a fazer isso mesmo (Exercícios 11.1.3 a 11.1.6), no contexto das transformações expansoras do intervalo, as quais, tecnicamente, são um pouco mais simples do que transformações expansoras numa variedade qualquer.

Exemplo 11.1.16. Dizemos que uma transformação $f : [0, 1] \rightarrow [0, 1]$ é *transformação expansora do intervalo* se existe uma família (finita ou enumerável) \mathcal{P} de subintervalos abertos disjuntos dois-a-dois, cuja união tem medida de Lebesgue total em $[0, 1]$, satisfazendo:

- (a) A restrição de f a cada $P \in \mathcal{P}$ é um difeomorfismo sobre $(0, 1)$; denotaremos por $f_P^{-1} : (0, 1) \rightarrow P$ a sua inversa.
- (b) Existem $C > 0$ e $\theta > 0$ tais que, para todo x, y e todo $P \in \mathcal{P}$

$$|\log |D(f_P^{-1})(x)| - \log |D(f_P^{-1})(y)|| \leq C|x - y|^\theta.$$

- (c) Existem $c > 0$ e $\sigma > 1$ tais que, para todo n e todo x :

$$|Df^n(x)| \geq c\sigma^n \quad (\text{se a derivada está definida.})$$

Esta classe inclui a expansão decimal e a transformação de Gauss, entre outras. As suas propriedades são analisadas nos Exercícios 11.1.3 a 11.1.5.

O Exercício 11.1.6 trata de uma classe um pouco mais geral de transformações, na qual substituímos a condição (a) por

- (a') Existe $\delta > 0$ tal que a restrição de f a cada $P \in \mathcal{P}$ é um difeomorfismo sobre um intervalo $f(P)$ com comprimento maior que δ e que contém todos os elementos de \mathcal{P} que ele intersecta.

11.1.4 Exercícios

11.1.1. Seja $f : M \rightarrow M$ um difeomorfismo local numa variedade compacta e seja m a medida de Lebesgue em M . Verifique as seguintes propriedades:

1. Se $m(B) = 0$ então $m(f^{-1}(B)) = 0$.
2. Se B é mensurável então $f(B)$ é mensurável.
3. Se $m(B) = 0$ então $m(f(B)) = 0$.
4. Se $A = B$ a menos de medida de Lebesgue nula então $f(A) = f(B)$ e $f^{-1}(A) = f^{-1}(B)$, a menos de conjuntos com medida de Lebesgue nula.
5. Se A é conjunto invariante então $f(A) = A$ a menos de conjunto com medida de Lebesgue nula.

11.1.2. Seja $f : M \rightarrow M$ uma transformação de classe C^1 tal que existem $\sigma > 1$ e $k \geq 1$ tais que $\|Df^k(x)v\| \geq \sigma\|v\|$ para todo $x \in M$ e todo $v \in T_pM$. Mostre que existe $\theta > 1$ e existe uma norma Riemanniana $\langle \cdot, \cdot \rangle$ equivalente a $\|\cdot\|$, tal que $\langle Df(x)v, v \rangle \geq \theta\langle v, v \rangle$ para todo $x \in M$ e todo $v \in T_pM$.

11.1.3. Mostre que se $f : [0, 1] \rightarrow [0, 1]$ é uma transformação expansora do intervalo e m representa a medida de Lebesgue em $[0, 1]$, então existe uma função $\rho : (0, 1) \rightarrow (0, \infty)$ tal que $\log \rho$ é limitada e Hölder e $\mu = \rho m$ é uma probabilidade invariante por f .

11.1.4. Mostre que a medida μ no Exercício 11.1.3 é exata e é a única probabilidade invariante de f absolutamente contínua com relação à medida m .

11.1.5. Mostre que transformações expansoras do intervalo satisfazem a fórmula de Rokhlin: se μ é a única medida absolutamente contínua invariante com respeito a Lebesgue e $\log |f'| \in L^1(\mu)$, então $h_\mu(f) = \int \log |f'| d\mu$.

11.1.6. Prove a seguinte generalização dos Exercícios 11.1.3 e 11.1.4: se f satisfaz as condições (a'), (b) e (c) no Exemplo 11.1.16 então existe um número finito, não nulo, de probabilidades absolutamente contínuas ergódicas e toda probabilidade invariante absolutamente contínua é combinação convexa dessas probabilidades ergódicas.

11.2 Dinâmica das transformações expansoras

Nesta seção estenderemos a noção de transformação expansora para espaços métricos compactos e mencionaremos alguns exemplos interessantes. Neste contexto geral, uma transformação expansora pode não ser nem sequer transitiva (compare com o Lema 11.1.13). No entanto, veremos nos Teoremas 11.2.14 e 11.2.15, que a sua dinâmica sempre pode ser reduzida ao caso topologicamente exato. Isto é relevante porque nos principais resultados das próximas seções precisaremos supor que a transformação é topologicamente exata (ou, equivalentemente, topologicamente misturadora; veja o Exercício 11.2.2).

Uma transformação contínua $f : M \rightarrow M$ num espaço métrico compacto M é dita *expansora* se existem constantes $\sigma > 1$ e $\rho > 0$ tais que para todo $p \in M$, a imagem da bola $B(p, \rho)$ contém uma vizinhança do fecho de $B(f(p), \rho)$ e

$$d(f(x), f(y)) \geq \sigma d(x, y) \quad \text{para todo } x, y \in B(p, \rho). \quad (11.2.1)$$

Toda transformação diferenciável expansora, no sentido da Seção 11.1, também é expansora neste sentido:

Exemplo 11.2.1. Seja $f : M \rightarrow M$ uma aplicação de classe C^1 numa variedade riemanniana compacta tal que $\|Df(x)v\| \geq \sigma\|v\|$ para todo $x \in M$ e todo $v \in T_xM$, onde σ é uma constante maior que 1. Denote $K = \sup \|Df\|$ (note que $K > 1$). Considere qualquer $\rho > 0$ suficientemente pequeno para que a restrição de f a toda bola $B(p, 2K\rho)$ seja um difeomorfismo sobre a imagem. Considere $y \in B(f(p), \sigma\rho)$ qualquer e seja $\gamma : [0, 1] \rightarrow B(f(p), \rho)$ uma geodésica

minimizante (isto é, que realiza a distância entre os seus pontos) com $\gamma(0) = f(p)$ e $\gamma(1) = y$. Pela escolha de ρ , existe uma curva diferenciável $\beta : [0, \delta] \rightarrow B(p, \rho)$ tal que $\beta(0) = p$ e $f(\beta(t)) = \gamma(t)$ para todo t . Observe que (usamos $\ell(\cdot)$ para representar o comprimento de uma curva),

$$d(p, \beta(t)) \leq \ell(\beta | [0, t]) \leq \sigma^{-1} \ell(\gamma | [0, t]) = \sigma^{-1} t d(f(p), y) < t\rho$$

para todo t . Isto mostra que podemos tomar $\delta = 1$. Então $\beta(1) \in B(p, \rho)$ e $f(\beta(1)) = \gamma(1) = y$. Desta forma, fica provado que $f(B(p, \rho))$ contém $B(f(p), \sigma\rho)$ a qual, evidentemente, é uma vizinhança de $\bar{B}(f(p), \rho)$. Agora considere $x, y \in B(p, \rho)$ quaisquer. Note que $d(f(x), f(y)) < 2K\rho$. Seja $\gamma : [0, 1] \rightarrow B(x, 2K\rho)$ uma geodésica minimizante ligando $f(x)$ a $f(y)$. Argumentando como no parágrafo anterior, encontramos uma curva diferenciável $\beta : [0, 1] \rightarrow B(x, 2K\rho)$ ligando x a y e tal que $f(\beta(t)) = \gamma(t)$ para todo t . Então,

$$d(f(x), f(y)) = \ell(\gamma) \geq \sigma \ell(\beta) \geq \sigma d(x, y).$$

Isto completa a demonstração de que f é uma transformação expansora.

O seguinte fato é útil para a construção de exemplos:

Lema 11.2.2. *Suponha que $f : M \rightarrow M$ é uma transformação expansora e que $\Lambda \subset M$ é um compacto tal que $f^{-1}(\Lambda) = \Lambda$. Então a restrição $f : \Lambda \rightarrow \Lambda$ também é uma transformação expansora.*

Demonstração. É claro que a condição (11.2.1) permanece válida na restrição. Resta verificar que $f(\Lambda \cap B(p, \rho))$ contém uma vizinhança de $\Lambda \cap \bar{B}(f(p), \rho)$ em Λ . Por hipótese, $f(B(p, \rho))$ contém uma vizinhança V de $\bar{B}(f(p), \rho)$. Então $\Lambda \cap V$ é uma vizinhança de $\Lambda \cap \bar{B}(f(p), \rho)$. Além disso, dado qualquer $y \in \Lambda \cap V$ existe $x \in B(p, \rho)$ tal que $f(x) = y$. Como $f^{-1}(\Lambda) = \Lambda$, este ponto está necessariamente em Λ . Isto prova que $\Lambda \cap V$ está contida em $f(\Lambda \cap B(p, \rho))$. Logo a restrição de f a Λ é uma transformação expansora, tal como afirmado \square

Não é possível substituir a hipótese do Lema 11.2.2 por $f(\Lambda) = \Lambda$. Veja o Exercício 11.2.4.

Exemplo 11.2.3. Seja $J \subset [0, 1]$ uma união finita de (dois ou mais) intervalos compactos disjuntos. Considere uma aplicação $f : J \rightarrow [0, 1]$ tal que a restrição de f a cada componente conexa de J é um difeomorfismo sobre $[0, 1]$. Veja a Figura 11.1. Suponha que existe $\sigma > 1$ tal que

$$|f'(x)| \geq \sigma \quad \text{para todo } x \in J. \quad (11.2.2)$$

Denote $\Lambda = \bigcap_{n=0}^{\infty} f^{-n}(J)$. Isto é, Λ o conjunto dos pontos x cujos iterados $f^n(x)$ estão definidos para todo $n \geq 0$. Segue imediatamente da definição que Λ é compacto (pode mostrar-se que é um conjunto de Cantor) e que $f^{-1}(\Lambda) = \Lambda$. A restrição $f : \Lambda \rightarrow \Lambda$ é uma transformação expansora. De fato, fixe $\rho > 0$ menor que a distância mínima entre duas quaisquer componentes conexas de J . Então qualquer bola de raio ρ em Λ está contida numa única componente conexa de J ; logo, por (11.2.2), ela é dilatada a taxa maior ou igual que σ .

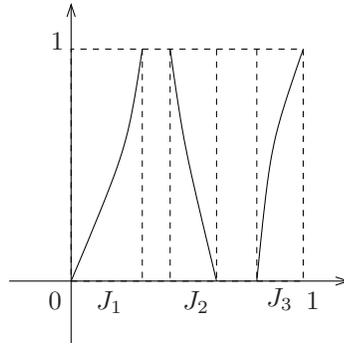


Figura 11.1: Tranformação expansora num conjunto de Cantor

Exemplo 11.2.4. Seja $\sigma : \Sigma_A \rightarrow \Sigma_A$ o deslocamento unilateral de tipo finito associado a uma matriz de transição A (estas noções foram introduzidas na Seção 10.2.2). Consideramos em Σ_A a distância definida por

$$d((x_n)_n, (y_n)_n) = 2^{-N}, \quad N = \inf\{n \in \mathbb{N} : x_n \neq y_n\}. \quad (11.2.3)$$

Então f_A é uma transformação expansora. De fato, fixe $\rho \in (1/2, 1)$ e $\sigma = 2$. A bola de raio ρ em torno de qualquer ponto $(p_n)_n \in \Sigma_A$ é o cilindro $[0; p_0]_A$ que contém esse ponto. A definição (11.2.3) dá que

$$d((x_{n+1})_n, (y_{n+1})_n) = 2d((x_n)_n, (y_n)_n)$$

para quaisquer $(x_n)_n$ e $(y_n)_n$ no cilindro $[0; p_0]_A$. Além disso, $f_A([0; p_0]_A)$ é a união de todos os cilindros $[0; q]$ tais que $A_{p_0, q} = 1$. Em particular, ela contém o cilindro $[0; p_1]_A$. Como os cilindros são abertos e fechados de Σ_A , isto mostra que a imagem da bola de raio ρ em torno de $(p_n)_n$ contém uma vizinhança do fecho da bola de raio ρ em torno de $(p_{n+1})_n$. Isto completa a verificação de que todo deslocamento de tipo finito é uma transformação expansora.

Exemplo 11.2.5. Seja $f : S^1 \rightarrow S^1$ um difeomorfismo local de classe C^2 com grau maior que 1 (o grau é o número de pré-imagens de todo ponto). Suponha que os pontos periódicos de f são hiperbólicos, ou seja, $|(f^n)'(x)| \neq 1$ para todo $x \in \text{Fix}(f^n)$ e todo $n \geq 1$. Seja Λ o complementar da união das bacias de atração dos pontos periódicos atratores de f . Então a restrição $f : \Lambda \rightarrow \Lambda$ é uma transformação expansora. Este fato é consequência de um teorema profundo devido a Ricardo Mañé [Mañ85].

Para transformações expansoras em espaços métricos o número de pré-imagens de um ponto $y \in M$ pode variar com y (a menos que o espaço métrico seja conexo; veja o Exercício 11.2.1 e compare também com o Lema 11.1.3). Por exemplo, para um deslocamento de tipo finito $\sigma : \Sigma_A \rightarrow \Sigma_A$ (Exemplo 11.2.4) o número de pré-imagens de um ponto $y = (y_n)_n \in \Sigma_A$ é igual ao número de símbolos i tais que $A_{i, y_0} = 1$; em geral, este número varia com y_0 .

Por outro lado, é fácil ver que o número de pré-imagens é sempre finito e até limitado: basta considerar uma cobertura finita do compacto M por bolas de raio ρ e observar que qualquer ponto tem no máximo uma pré-imagem em cada uma dessas bolas. Por um pequeno abuso de linguagem, chamaremos *grau* de uma transformação expansora $f : M \rightarrow M$ ao número máximo de pré-imagens de qualquer ponto, isto é

$$\text{grau}(f) = \max\{\#f^{-1}(y) : y \in M\}. \quad (11.2.4)$$

11.2.1 Ramos inversos contrativos

Por definição, toda transformação expansora f é injetiva em cada bola $B(p, \rho)$. Logo, a restrição de f a qualquer compacto $K \subset B(p, \rho)$ é um homeomorfismo sobre a imagem. Tome para K a pré-imagem do fecho $\bar{B}(f(p), \rho)$ da bola centrada em $f(p)$. Chamamos *ramo inverso de f em p* à inversa $h_p : \bar{B}(f(p), \rho) \rightarrow B(p, \rho)$ da restrição de f a K . Claro que $h_p(f(p)) = p$ e que $f \circ h_p = \text{id}$. A condição (11.2.1) implica

$$d(h_p(z), h_p(w)) \leq \sigma^{-1}d(z, w) \quad \text{para todo } z, w \in B(f(p), \rho). \quad (11.2.5)$$

Lema 11.2.6. *Se $f : M \rightarrow M$ é expansora então, para todo $y \in M$,*

$$f^{-1}(B(y, \rho)) = \bigcup_{x \in f^{-1}(y)} h_x(B(y, \rho)).$$

Demonstração. A relação $f \circ h_x = \text{id}$ implica que $h_x(B(y, \rho))$ está contido na pré-imagem de $B(y, \rho)$ para todo $x \in f^{-1}(y)$. Para provar a outra inclusão, seja z qualquer ponto tal que $f(z) \in B(y, \rho)$. Pela definição de transformação expansora, $f(B(z, \rho))$ contém $B(f(z), \rho)$ e, portanto, contém y . Seja $h_z : B(f(z), \rho) \rightarrow M$ o ramo inverso de f que envia $f(z)$ em z e seja $x = h_z(y)$. Tanto z quanto $h_x(f(z))$ estão em $B(x, \rho) \cap f^{-1}(z)$. Como f é injetiva em cada bola de raio ρ , segue que $z = h_x(f(z))$. Isto completa a demonstração. \square

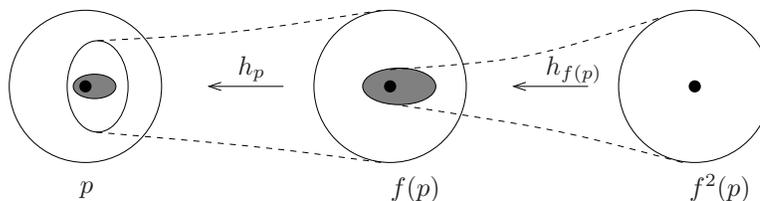


Figura 11.2: Ramos inversos de f^n

Mais geralmente, dado qualquer $n \geq 1$, chamamos *ramo inverso de f^{-n} em p* à composição

$$h_p^n = h_{f^{n-1}(p)} \circ \cdots \circ h_p : \bar{B}(f^n(p), \rho) \rightarrow B(p, \rho)$$

dos ramos inversos de f nos iterados de p . Observe que $h_p^n(f^n(p)) = p$ e que $f^n \circ h_p^n = \text{id}$. Além disso, $f^j \circ h_p^n = h_{f^j(p)}^{n-j}$ para cada $0 \leq j < n$. Logo,

$$d(f^j \circ h_p^n(z), f^j \circ h_p^n(w)) \leq \sigma^{j-n} d(z, w) \quad (11.2.6)$$

para todo $z, w \in B(f^n(p), \rho)$ e todo $0 \leq j \leq n$.

Lema 11.2.7. *Se $f : M \rightarrow M$ é expansora, $f^n(B(p, n+1, \varepsilon)) = B(f^n(p), \varepsilon)$ para todo $p \in M$, $n \geq 0$ e $\varepsilon \in (0, \rho]$.*

Demonstração. A inclusão $f^n(B(p, n+1, \varepsilon)) \subset B(f^n(p), \varepsilon)$ é uma consequência trivial da definição da bola dinâmica. Para provar a recíproca, considere o ramo inverso $h_p^n : \bar{B}(f^n(p), \rho) \rightarrow B(p, \rho)$. Dado qualquer $y \in B(f^n(p), \varepsilon)$ considere $x = h_p^n(y)$. Então, $f^n(x) = y$ e, por (11.2.6),

$$d(f^j(x), f^j(p)) \leq \sigma^{j-n} d(f^n(x), f^n(p)) \leq d(y, f^n(p)) < \varepsilon$$

para todo $0 \leq j \leq n$. Isto mostra que $x \in B(p, n+1, \varepsilon)$. \square

Corolário 11.2.8. *Toda transformação expansora é expansiva.*

Demonstração. Suponha que $d(f^n(z), f^n(w)) < \rho$ para todo $n \geq 0$. Isso implica que $z = h_w^n(f^n(z))$ para todo $n \geq 0$. Então, a propriedade (11.2.6) dá que

$$d(z, w) \leq \sigma^{-n} d(f^n(z), f^n(w)) < \rho \sigma^{-n}.$$

Fazendo $n \rightarrow \infty$, concluímos que $z = w$. Portanto, ρ é uma constante de expansividade para f . \square

11.2.2 Sombreamento e pontos periódicos

Dado $\delta > 0$, chamamos δ -pseudo-órbita da transformação $f : M \rightarrow M$ qualquer sequência $(x_n)_{n \geq 0}$ tal que

$$d(f(x_n), x_{n+1}) < \delta \quad \text{para todo } n \geq 0.$$

Dizemos que a δ -pseudo-órbita é *periódica* se existe $\kappa \geq 1$ tal que $x_n = x_{n+\kappa}$ para todo $n \geq 0$. Claro que toda órbita é uma δ -pseudo-órbita, para todo $\delta > 0$. Para transformações expansoras temos uma espécie de recíproca: toda pseudo-órbita permanece próxima de (dizemos que é *sombreada* por) alguma órbita da transformação:

Proposição 11.2.9 (Lema do sombreamento). *Suponha que $f : M \rightarrow M$ é uma transformação expansora. Para todo $\varepsilon > 0$ existe $\delta > 0$ tal que para toda δ -pseudo-órbita $(x_n)_n$ existe $x \in M$ tal que $d(f^n(x), x_n) < \varepsilon$ para todo $n \geq 0$.*

Se ε é suficientemente pequeno para que 2ε seja uma constante de expansividade de f então o ponto x é único. Se, além disso, a pseudo-órbita é periódica então x é ponto periódico.

Demonstração. Não é restrição considerar ε menor que ρ . Fixe $\delta > 0$ de tal modo que $\sigma^{-1}\varepsilon + \delta < \varepsilon$. Para cada $n \geq 0$, seja $h_n : \bar{B}(f(x_n), \rho) \rightarrow B(x_n, \rho)$ o ramo contrativo de f^{-1} em x_n . A propriedade (11.2.5) garante que

$$h_n(\bar{B}(f(x_n), \varepsilon)) \subset B(x_n, \sigma^{-1}\varepsilon) \quad \text{para todo } n \geq 1. \quad (11.2.7)$$

Como $d(x_n, f(x_{n-1})) < \delta$, segue que

$$h_n(\bar{B}(f(x_n), \varepsilon)) \subset B(f(x_{n-1}), \varepsilon) \quad \text{para todo } n \geq 1. \quad (11.2.8)$$

Então, podemos considerar a composição $h^n = h_0 \circ \dots \circ h_{n-1}$. Além disso, (11.2.8) implica que a sequência de compactos $K_n = h^n(\bar{B}(f(x_n), \varepsilon))$ é encaixada. Tome x na interseção. Para todo $n \geq 0$, temos que $x \in K_{n+1}$ e, portanto, $f^n(x)$ pertence a

$$f^n \circ h^{n+1}(\bar{B}(f(x_n), \varepsilon)) = h_n(\bar{B}(f(x_n), \varepsilon)).$$

Por (11.2.7), isto implica que $d(f^n(x), x_n) < \varepsilon$ para todo $n \geq 0$.

As restantes afirmações são consequências simples. Se x' é outro ponto satisfazendo a conclusão da proposição então

$$d(f^n(x), f^n(x')) \leq d(f^n(x), x_n) + d(f^n(x'), x_n) < 2\varepsilon \quad \text{para todo } n \geq 0.$$

Por expansividade, segue que $x = x'$. Finalmente, se a pseudo-órbita é periódica, com período $\kappa \geq 1$, também temos que

$$d(f^n(f^\kappa(x)), x_n) = d(f^{n+\kappa}(x), x_{n+\kappa}) < \varepsilon \quad \text{para todo } n \geq 0.$$

Por unicidade, obtemos que $f^\kappa(x) = x$. □

Vale a pena observar que δ depende linearmente de ε : a prova da Proposição 11.2.9 dá que podemos tomar $\delta = c\varepsilon$, onde $c > 0$ só depende de σ .

Chamamos pré-órbita de um ponto $x \in M$ qualquer sequência $(x_{-n})_{n \geq 0}$ tal que $x_0 = x$ e $f(x_{-n}) = x_{-n+1}$ para todo $n \geq 1$.

Lema 11.2.10. *Se $d(x, y) < \rho$ então, dada qualquer pré-órbita $(x_{-n})_n$ de x , existe uma pré-órbita $(y_{-n})_n$ de y assintótica a $(x_{-n})_n$, isto é, tal que $d(x_{-n}, y_{-n})$ converge para 0 quando $n \rightarrow \infty$.*

Demonstração. Para cada $n \geq 1$, seja $h_n : B(x, \rho) \rightarrow M$ o ramo contrativo de f^{-n} satisfazendo $h_n(x) = x_{-n}$. Defina $y_{-n} = h_n(y)$. É imediato que $d(x_{-n}, y_{-n}) \leq \sigma^{-n}d(x, y)$. □

Teorema 11.2.11. *Seja $f : M \rightarrow M$ uma transformação expansora num espaço métrico compacto e seja $\Lambda \subset M$ o fecho do conjunto dos pontos periódicos de f . Então, $f(\Lambda) = \Lambda$ e a restrição $f : \Lambda \rightarrow \Lambda$ é uma transformação expansora.*

Demonstração. É claro que $f(\Lambda)$ está contido em Λ : se um ponto x é acumulado por pontos periódicos p_n então $f(x)$ é acumulado pelas imagens $f(p_n)$, que também são pontos periódicos. Além disso, como $f(\Lambda)$ é um compacto que contém todos os pontos periódicos, ele contém Λ . Isto mostra que $f(\Lambda) = \Lambda$.

A seguir vamos provar que a restrição $f : \Lambda \rightarrow \Lambda$ é expansora. É claro que a propriedade (11.2.1) permanece válida na restrição. Para concluir devemos mostrar que existe $r \leq \rho$ (tomaremos $r = \sigma^{-1}\rho$) tal que, para todo $x \in \Lambda$, a imagem $f(\Lambda \cap B(x, r))$ contém uma vizinhança de $\Lambda \cap \bar{B}(x, r)$. Começamos pelo seguinte lema:

Lema 11.2.12. *Seja p um ponto periódico e seja $h_p : \bar{B}(f(p), \rho) \rightarrow B(p, \rho)$ o ramo contrativo de f^{-1} em p . Se $y \in B(f(p), \rho)$ é um ponto periódico então $h_p(y) \in \Lambda$.*

Demonstração. Escreveremos $x = h_p(y)$ e $q = f(p)$. Considere qualquer $\varepsilon > 0$ tal que 2ε é constante de expansividade de f . Tome δ dado pelo lema de sombreamento (Proposição 11.2.9). Pelo Lema 11.2.10, existe uma pré-órbita $(x_{-n})_n$ de x assintótica à pré-órbita periódica $(\bar{p}_n)_n$ de p . Em particular,

$$d(x_{-k+1}, q) = d(x_{-k+1}, p_{-k+1}) < \delta \tag{11.2.9}$$

para qualquer múltiplo suficientemente grande k do período de p . Analogamente, existe uma pré-órbita $(q_{-n})_n$ de x assintótica à pré-órbita periódica $(\bar{p}_n)_n$ de p . Fixe qualquer múltiplo l do período de y tal que

$$d(q_{-l}, f(x)) = d(q_{-l}, y) < \delta. \tag{11.2.10}$$

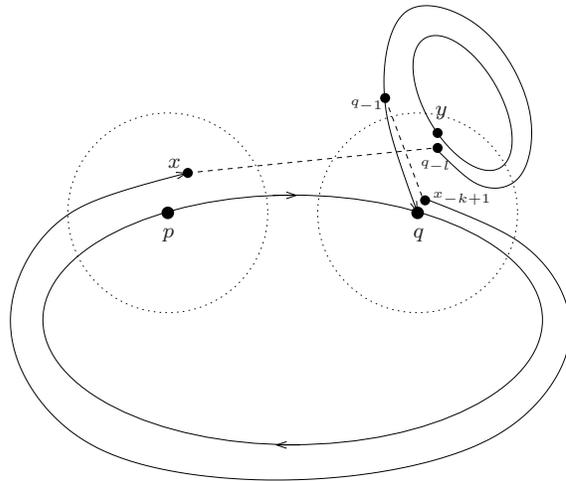


Figura 11.3: Construindo órbitas periódicas

Agora considere a sequência periódica $(z_n)_n$, de período $k + l$, dada por

$$z_0 = x, z_1 = q_{-l}, \dots, z_l = q_{-1}, z_{l+1} = x_{-k+1}, \dots, z_{l+k-1} = x_{-1}, z_{k+l} = x.$$

Veja a Figura 11.3. Afirmamos que $(z_n)_n$ é uma δ -pseudo-órbita. De fato, se n é um múltiplo de $k + l$ então, por (11.2.10),

$$d(f(z_n), z_{n+1}) = d(f(x), q_{-1}) = d(y, q_{-1}) < \delta.$$

Se n é congruente com l módulo $k + l$ então, por (11.2.9),

$$d(f(z_n), z_{n+1}) = d(f(q_{-1}), x_{-k+1}) = d(q, x_{-k+1}) < \delta.$$

Em todos os outros casos, $f(z_n) = z_{n+1}$. Isto prova a nossa afirmação. Logo, pela Proposição 11.2.9, existe um ponto periódico z tal que $d(f^n(z), z_n) < \varepsilon$ para todo $n \geq 0$. Em particular, $d(z, x) < \varepsilon$. Como $\varepsilon > 0$ é arbitrário, isto mostra que x está no fecho do conjunto dos pontos periódicos, conforme afirmado. \square

Corolário 11.2.13. *Seja $z \in \Lambda$ e seja $h_z : \bar{B}(f(z), \rho) \rightarrow B(z, \rho)$ o ramo contrativo de f^{-1} em z . Se $w \in \Lambda \cap B(f(z), \rho)$ então $h_z(w) \in \Lambda$.*

Demonstração. Como $z \in \Lambda$, podemos encontrar algum ponto periódico p suficientemente próximo para que $w \in B(f(p), \rho)$ e $h_p(w) = h_z(w)$. Como $w \in \Lambda$ podemos encontrar pontos periódicos $y_n \in B(f(p), \rho)$ convergindo para w . Pelo Lema 11.2.12, temos $h_p(y_n) \in \Lambda$ para todo n . Passando ao limite, concluímos que $h_p(w) \in \Lambda$. \square

Estamos prontos para terminar a demonstração do Teorema 11.2.11. Tome $r = \sigma^{-1}\rho$. A propriedade (11.2.6) implica que $h_z(B(f(z), \rho))$ está contida em $B(z, r)$, para todo $z \in \Lambda$. Então, o Corolário 11.2.13 implica que $f(\Lambda \cap B(z, r))$ contém $\Lambda \cap B(f(z), \rho)$. Este último conjunto é uma vizinhança de $\bar{B}(f(z), r)$. \square

Teorema 11.2.14. *Seja $f : M \rightarrow M$ uma transformação expansora num espaço métrico compacto e seja $\Lambda \subset M$ o fecho do conjunto dos pontos periódicos de f . Então,*

$$M = \bigcup_{k=0}^{\infty} f^{-k}(\Lambda).$$

Demonstração. Dado qualquer $x \in M$, seja $\omega(x)$ o seu conjunto ω -limite, ou seja, o conjunto dos pontos de acumulação dos iterados $f^n(x)$ quando $n \rightarrow \infty$. Inicialmente, mostraremos que $\omega(x) \subset \Lambda$. Então deduziremos que $f^k(x) \in \Lambda$ para algum $k \geq 0$.

Seja $\varepsilon > 0$ tal que 2ε é constante de expansividade para f . Tome $\delta > 0$ dado pelo lema de sombreamento (Proposição 11.2.9) e tome $\alpha \in (0, \delta)$ tal que $d(f(z), f(w)) < \delta$ sempre que $d(z, w) < \alpha$. Seja y um ponto qualquer em $\omega(x)$. Existem $r \geq 0$ e $s \geq 1$ tais que

$$d(f^r(x), y) < \alpha \quad \text{e} \quad d(f^{r+s}(x), y) < \alpha.$$

Considere a sequência periódica (z_n) , de período s , dada por

$$z_0 = y, z_1 = f^{r+1}(x), \dots, z_{s-1} = f^{r+s-1}(x), z_s = y.$$

Observe que $d(f(z_0), z_1) = d(f(y), f^{r+1}(x)) < \delta$ (porque $d(y, f^r(x)) < \alpha$) e $d(f(z_{s-1}), z_s) = d(f^{r+s}(x), y) < \alpha < \delta$ e $f(z_n) = z_{n+1}$ nos demais casos. Portanto, $(z_n)_n$ é uma δ -pseudo-órbita. Então, pela Proposição 11.2.9, existe algum ponto periódico z tal que $d(y, z) < \varepsilon$. Fazendo $\varepsilon \rightarrow 0$, concluímos que y é acumulado por pontos periódicos, ou seja, $y \in \Lambda$.

Sejam $\varepsilon > 0$ e $\delta > 0$ como antes. Não é restrição supor que $\delta < \varepsilon$. Tome $\beta \in (0, \delta/2)$ tal que $d(f(z), f(w)) < \delta/2$ sempre que $d(z, w) < \beta$. Como $\omega(x)$ está contido em Λ , existem $k \geq 1$ e pontos $w_n \in \Lambda$ tais que $d(f^{n+k}(x), w_n) < \beta$ para todo $n \geq 0$. Observe que

$$d(f(w_n), w_{n+1}) \leq d(f(w_n), f^{n+k+1}(x)) + d(f^{n+k+1}(x), w_{n+1}) < \delta/2 + \beta < \delta$$

para todo $n \geq 0$. Portanto, $(w_n)_n$ é uma δ -pseudo-órbita em Λ . Como a restrição de f a Λ é expansora (Teorema 11.2.11), segue da Proposição 11.2.9 aplicada à restrição que existe $w \in \Lambda$ tal que $d(f^n(w), w_n) < \varepsilon$ para todo $n \geq 0$. Então,

$$d(f^n(f^k(x)), f^n(w)) \leq d(f^{n+k}(x), w_n) + d(w_n, f^n(w)) < \beta + \varepsilon < 2\varepsilon$$

para todo $n \geq 0$. Então, por expansividade, $f^k(x) = w$. \square

11.2.3 Decomposição dinâmica

O Teorema 11.2.14 mostra que a dinâmica interessante de qualquer transformação expansora está localizada no fecho Λ dos pontos periódicos. Em particular, $\text{supp } \mu \subset \Lambda$ para toda probabilidade invariante f . Além disso (Teorema 11.2.11), a restrição de f a Λ ainda é uma transformação expansora. Portanto, a menos de substituir M por Λ , não constitui restrição supor que os pontos periódicos de f são densos em M .

Teorema 11.2.15 (Decomposição dinâmica). *Seja $f : M \rightarrow M$ uma transformação expansora cujo conjunto dos pontos periódicos é denso em M . Então existe uma partição de M em conjuntos compactos não-vazios $M_{i,j}$, com $1 \leq i \leq k$ e $1 \leq j \leq m(i)$, tais que*

- (a) $M_i = \bigcup_{j=1}^{m(i)} M_{i,j}$ é invariante for f , para todo i ;
- (b) $f(M_{i,j}) = M_{i,j+1}$ se $j < m(i)$ e $f(M_{i,m(i)}) = M_{i,1}$, para todo i, j ;
- (c) cada restrição $f : M_i \rightarrow M_i$ é uma transformação expansora transitiva;
- (d) cada $f^{m(i)} : M_{i,j} \rightarrow M_{i,j}$ é uma transformação expansora topologicamente exata.

Além disso, o número k , os números $m(i)$ e os conjuntos $M_{i,j}$ são únicos a menos de renumeração.

Demonstração. Consideremos a relação \sim definida da seguinte forma no conjunto dos pontos periódicos de f . Dados dois pontos periódicos p e q , sejam $(\bar{p}_n)_n$ e $(\bar{q}_n)_n$, respectivamente, as suas pré-órbitas periódicas. Por definição, $p \sim q$ se, e somente se, existem pré-órbitas $(p_n)_n$ de p e $(q_n)_n$ de q tais que

$$d(p_n, \bar{q}_n) \rightarrow 0 \quad \text{e} \quad d(\bar{p}_n, q_n) \rightarrow 0. \quad (11.2.11)$$

Afirmamos que \sim é uma relação de equivalência. É claro da definição que a relação \sim é reflexiva e simétrica. Agora suponha que $p \sim q$ e $q \sim r$. Então existem pré-órbitas $(q_n)_n$ de q e $(r_n)_n$ de r assintóticas às pré-órbitas periódicas $(\bar{p}_n)_n$ de p e $(\bar{q}_n)_n$ de q . Fixe $k \geq 1$ múltiplo dos períodos de p e q tal que $d(r_k, q_k) < \rho$. Note que $q_k = q$, uma vez que k é múltiplo do período de q . Então, pelo Lema 11.2.10, existe uma pré-órbita $(r'_n)_n$ do ponto $r' = r_k$ que $d(r'_n, q_n) \rightarrow 0$. Então $d(r'_n, \bar{p}_n) \rightarrow 0$. Considere a pré-órbita $(r''_n)_n$ de r definida por

$$r''_n = \begin{cases} r_n & \text{se } n < k \\ r'_{n-k} & \text{se } n \geq k. \end{cases}$$

Como k é múltiplo do período de p , temos $d(r''_n, \bar{p}_n) = d(r_{n-k}, \bar{p}_{n-k})$ para todo $n \geq k$. Portanto, $(r''_n)_n$ é assintótica a $(\bar{p}_n)_n$. Analogamente se constrói uma pré-órbita $(p''_n)_n$ de p assintótica a $(\bar{r}_n)_r$. Portanto, $p \sim r$, o que mostra que a relação \sim é transitiva. Deste modo fica provada a nossa afirmação.

Em seguida, afirmamos que $p \sim q$ se, e somente se, $f(p) \sim f(q)$. Comece por supor que $p \sim q$ e sejam $(p_n)_n$ e $(q_n)_n$ pré-órbitas de p e q satisfazendo (11.2.11). As pré-órbitas periódicas de $p' = f(p)$ e $q' = f(q)$ são, respectivamente,

$$\bar{p}'_n = \begin{cases} f(p) & \text{se } n = 0 \\ \bar{p}_{n-1} & \text{se } n \geq 1 \end{cases} \quad \text{e} \quad \bar{q}'_n = \begin{cases} f(q) & \text{se } n = 0 \\ \bar{q}_{n-1} & \text{se } n \geq 1. \end{cases}$$

Considere as pré-órbitas de p e q , respectivamente, dadas por

$$p'_n = \begin{cases} f(p) & \text{se } n = 0 \\ p_{n-1} & \text{se } n \geq 1 \end{cases} \quad \text{e} \quad q'_n = \begin{cases} f(q) & \text{se } n = 0 \\ q_{n-1} & \text{se } n \geq 1. \end{cases}$$

É claro que $(p'_n)_n$ é assintótico a $(\bar{q}'_n)_n$ e $(q'_n)_n$ é assintótico a $(\bar{p}'_n)_n$. Logo $f(p) \sim f(q)$. Agora suponha que $f(p) \sim f(q)$. Pelo que acabamos de provar, $f^k(p) \sim f^k(q)$ para todo $k \geq 1$. Tomando k um múltiplo comum dos períodos de p e q obtemos que $p \sim q$. Isto prova a nossa afirmação. Note que ela significa que a imagem e a pré-imagem de qualquer classe de equivalência são classes de equivalência.

Observe também que se $d(p, q) < \rho$ então $p \sim q$. De fato, pelo Lema 11.2.10 podemos encontrar uma pré-órbita de q assintótica à pré-órbita periódica de p e, do mesmo modo, uma pré-órbita de p assintótica à pré-órbita periódica de q . Como M é compacto, segue que as classes de equivalência são em número finito. Além disso, se A e B são duas classes de equivalência, \bar{A} e \bar{B} são disjuntos e sua distância é pelo menos ρ . Como $p \sim q$ se, e somente se, $f(p) \sim f(q)$, a função f permuta os fechos das classes de equivalência.

Deste modo, podemos enumerar os fechos das classes de equivalência como $M_{i,j}$ com $1 \leq i \leq k$ e $1 \leq j \leq m(i)$, de tal modo que

$$f(M_{i,j}) = M_{i,j+1} \text{ para } j < m(i) \quad \text{e} \quad f(M_{i,m(i)}) = M_{i,1}. \quad (11.2.12)$$

As propriedades (a) e (b) no enunciado do teorema são consequências imediatas.

Passamos a provar a propriedade (c). Como os M_i são disjuntos dois-a-dois, segue de (11.2.12) que $f^{-1}(M_i) = M_i$ para todo i . Logo, o Lema 11.2.2 dá que $f : M_i \rightarrow M_i$ é transformação expansora. Pelo Lema 4.3.4, para mostrar que esta transformação é transitiva basta mostrar que dados quaisquer abertos U e V de M_i existe $n \geq 1$ tal que $f^n(U)$ intersecta V . Não é restrição supor que $U \subset M_{i,j}$ para algum j . Além disso, a menos de substituir V por alguma pré-imagem $f^{-k}(V)$, podemos supor que V está contido no mesmo $M_{i,j}$. Consideremos pontos periódicos $p \in U$ e $q \in V$. Pela definição das classes de equivalência, existe alguma pré-órbita $(q_{-n})_n$ de q assintótica à pré-órbita periódica $(\bar{p}_{-n})_n$ de p . Em particular, podemos encontrar n arbitrariamente grande tal que $q_{-n} \in U$. Então, $q \in f^n(U) \cap V$. Portanto, $f : M_i \rightarrow M_i$ é transitiva.

Agora provamos a propriedade (d). Como os $M_{i,j}$ são disjuntos dois-a-dois, segue de (11.2.12) que $f^{-m(i)}(M_{i,j}) = M_{i,j}$ para todo i . Logo (Lema 11.2.2), $g = f^{m(i)} : M_{i,j} \rightarrow M_{i,j}$ é uma transformação expansora. Resta provar que g é topologicamente exata. Seja U um aberto não vazio de $M_{i,j}$ e seja p um ponto periódico de f em U . Por (11.2.12), o período κ é um múltiplo de $m(i)$, digamos $\kappa = sm(i)$. Seja q um ponto periódico qualquer de f em $M_{i,j}$. Pela definição da relação de equivalência, existe alguma pré-órbita $(q_{-n})_n$ de q assintótica à pré-órbita periódica $(\bar{p}_{-n})_n$ de p . Em particular, $d(q_{-\kappa n}, p) \rightarrow 0$ quando $n \rightarrow \infty$. Então, $h_q^{\kappa n}(\bar{B}(q, \rho))$ está contido em U para todo n suficientemente grande. Isto implica que $g^{sn}(U) = f^{\kappa n}(U)$ contém $B(q, \rho)$ para todo n suficientemente grande. Como $M_{i,j}$ é compacto, podemos encontrar uma cobertura finita por bolas de raio ρ centradas em pontos periódicos. Aplicando o argumento anterior a cada uma desses pontos periódicos, deduzimos que $g^{sn}(U)$ contém $M_{i,j}$ para todo n suficientemente grande. Portanto, g é topologicamente exata.

Resta provar a unicidade. Seja $N_{r,s}$, com $1 \leq r \leq l$ e $1 \leq s \leq n(r)$, outra partição nas condições do enunciado. Inicialmente, consideremos as partições $\mathcal{M} = \{M_i : 1 \leq i \leq k\}$ e $\mathcal{N} = \{N_r : 1 \leq r \leq l\}$, onde $N_r = \bigcup_{s=1}^{n(r)} N_{r,s}$. Dados quaisquer i e r , os conjuntos M_i e N_r são abertos, fechados, invariantes e transitivos. Afirmamos que ou $M_i \cap N_r = \emptyset$ ou $M_i = N_r$. De fato, como a interseção é aberta, se ela é não vazia então ela intersecta qualquer órbita densa em M_i (ou N_r). Como a interseção também é fechada e invariante, segue que ela contém M_i (e N_r). Em outras palavras, $M_i = N_r$. Isto prova a nossa afirmação. Segue que as partições \mathcal{M} e \mathcal{N} coincidem, isto é, $k = l$ e $M_i = N_i$ a menos de renumeração. Agora fixemos i . A transformação f permuta os $M_{i,j}$ e os $N_{i,s}$ ciclicamente, com períodos $m(i)$ e $n(i)$. Como cada uma destas famílias de conjuntos constitui uma partição de M_i , isso só é possível se $m(i) = n(i)$. Como $f^{m(i)}$ é transitivo em cada $M_{i,j}$ e cada $N_{i,s}$, o mesmo argumento da primeira parte deste parágrafo mostra que, dados quaisquer j e s , ou $M_{i,j} \cap N_{i,s} = \emptyset$ ou $M_{i,j} = N_{i,s}$. Segue que as famílias $M_{i,j}$ e $N_{i,s}$ coincidem, a menos de renumeração cíclica. \square

A seguinte consequência do teorema contém o Lema 11.1.13:

Corolário 11.2.16. *Se M é conexo e $f : M \rightarrow M$ é transformação expansora então o conjunto dos pontos periódicos é denso em M e f é topologicamente exata.*

Demonstração. Afiramos que Λ é um subconjunto aberto de $f^{-1}(\Lambda)$. Para provar isso, considere $\delta \in (0, \rho)$ tal que $d(x, y) < \delta$ implica $d(f(x), f(y)) < \rho$. Suponha que $x \in f^{-1}(\Lambda)$ é tal que $d(x, \Lambda) < \delta$. Então existe $z \in \Lambda$ tal $d(x, z) < \delta < \rho$ e portanto $d(f(x), f(z)) < \rho$. Aplicando o Corolário 11.2.13 com $w = f(x)$, vem que $x = h_z(w) \in \Lambda$. Portanto, Λ contém a sua δ -vizinhança dentro de $f^{-1}(\Lambda)$. Isso implica a nossa afirmação.

Então o conjunto $S = f^{-1}(\Lambda) \setminus \Lambda$ é fechado em $f^{-1}(\Lambda)$ e, conseqüentemente, é fechado em M . Então, $f^{-n}(S)$ é fechado em M para todo $k \geq 0$. Pelo Teorema 11.2.14 temos que M é a união enumerável disjunta dos fechados Λ e $f^{-n}(S)$, $k \geq 0$. Pelo teorema de Baire, algum destes abertos tem interior não vazio. Como f é uma aplicação aberta, segue que Λ tem interior não vazio. Agora considere a restrição $f : \Lambda \rightarrow \Lambda$, que é uma transformação expansora, e seja $\{\Lambda_{i,j} : 1 \leq i \leq k, 1 \leq j \leq m(i)\}$ a decomposição do domínio Λ dada pelo Teorema 11.2.15. Então algum $\Lambda_{i,j}$ contém algum aberto V de M . Como $f^{m(i)}$ é topologicamente exata, $f^{nm(i)}(V) = \Lambda_{i,j}$ para algum $n \geq 1$. Usando mais uma vez o fato de que a transformação f é aberta, segue que o compacto $M_{i,j}$ é um aberto de M . Por conexidade, segue que $M = \Lambda_{i,j}$. Isto implica que $\Lambda = M$ e $f : M \rightarrow M$ é topologicamente exata. \square

11.2.4 Exercícios

11.2.1. Mostre que se $f : M \rightarrow M$ é um homeomorfismo local num espaço métrico compacto e conexo, então o número de pré-imagens $\#f^{-1}(y)$ é o mesmo para todo $y \in M$.

11.2.2. Mostre que se uma transformação expansora é topologicamente misturadora então ela é topologicamente exata.

11.2.3. Seja $f : M \rightarrow M$ uma transformação topologicamente exata num espaço métrico compacto. Mostre que para todo $r > 0$ existe $N \geq 1$ tal que $f^N(B(x, r)) = M$ para todo $x \in M$.

11.2.4. Considere a transformação expansora $f : S^1 \rightarrow S^1$ dada por $f(x) = 2x \pmod{\mathbb{Z}}$. Dê exemplo de conjunto compacto $\Lambda \subset S^1$ tal que $f(\Lambda) = \Lambda$ mas a restrição $f : \Lambda \rightarrow \Lambda$ não é uma transformação expansora.

11.2.5. Seja $f : M \rightarrow M$ uma transformação expansora e seja Λ o fecho do conjunto dos pontos periódicos de f . Mostre que $h(f) = h(f|_{\Lambda})$.

11.2.6. Seja $f : M \rightarrow M$ uma transformação expansora tal que o conjunto dos pontos periódicos é denso em M e sejam $M_i, M_{i,j}$ os subconjuntos compactos dados pelo Teorema 11.2.15. Mostre que, $h(f) = \max_i h(f|_{M_i})$ e

$$h(f|_{M_i}) = \frac{1}{m(i)} h(f^{m(i)}|_{M_{i,j}}) \quad \text{para quaisquer } i, j.$$

11.2.7. Seja $f_A : \Sigma_A \rightarrow \Sigma_A$ um deslocamento de tipo finito. Interprete a decomposição dada pelo Teorema 11.2.15 em termos da matriz A .

11.3 Entropia e pontos periódicos

Nesta seção analisaremos de um ponto de vista quantitativo a distribuição dos pontos periódicos de uma transformação expansora $f : M \rightarrow M$.

Vamos mostrar (Seção 11.3.1) que a taxa de crescimento do número de pontos periódicos é igual à entropia topológica; compare este enunciado com a discussão na Seção 10.2.1. Outra conclusão interessante (Seção 11.3.2) é que toda medida invariante pode ser aproximada, na topologia fraca*, por medidas invariantes suportadas em órbitas periódicas. Estes resultados se baseiam na seguinte propriedade:

Proposição 11.3.1. *Seja $f : M \rightarrow M$ uma transformação expansora topologicamente exata. Então, dado qualquer $\varepsilon > 0$ existe $\kappa \geq 1$ tal que, dados quaisquer $x_1, \dots, x_s \in M$, quaisquer $n_1, \dots, n_s \geq 1$ e quaisquer $k_1, \dots, k_s \geq \kappa$, existe um ponto $p \in M$ tal que, escrevendo $m_j = \sum_{i=1}^j n_i + k_i$ para $j = 1, \dots, s$ e $m_0 = 0$,*

$$(i) \quad d(f^{m_{j-1}+i}(p), f^i(x_j)) < \varepsilon \text{ para } 0 \leq i < n_j \text{ e } 1 \leq j \leq s$$

$$(ii) \quad \text{e } f^{m_s}(p) = p.$$

Demonstração. Dado $\varepsilon > 0$, considere $\delta > 0$ como no lema de sombreamento (Proposição 11.2.9). Sem perda de generalidade, podemos supor que $\delta < \varepsilon$ e que 2ε é uma constante de expansividade de f (lembre do Lema 11.1.4). Como f é topologicamente exata, dado qualquer $z \in M$ existe $\kappa \geq 1$ tal que $f^k(B(z, \delta)) = M$ para todo $k \geq \kappa$. Além disso, pela compacidade de M (veja o Exercício 11.2.3), podemos escolher κ dependendo apenas de δ . Considere $x_j, n_j, k_j \geq \kappa$, $j = 1, \dots, s$ como no enunciado. Em particular, para cada $j = 1, \dots, s-1$ existe $y_j \in B(f^{n_j}(x_j), \delta)$ tal que $f^{k_j}(y_j) = x_{j+1}$. Analogamente, existe $y_s \in B(f^{n_s}(x_s), \delta)$ tal que $f^{k_s}(y_s) = x_1$. Considere a δ -pseudo-órbita periódica $(z_n)_{n \geq 0}$ definida por

$$z_n = \begin{cases} f^{n-m_{j-1}}(x_j) & \text{para } 0 \leq n - m_{j-1} < n_j \text{ e } j = 1, \dots, s \\ f^{n-m_{j-1}-n_j}(y_j) & \text{para } 0 \leq n - m_{j-1} - n_j < k_j \text{ e } j = 1, \dots, s \\ z_{n-m_s} & \text{para } n \geq m_s. \end{cases}$$

Pela segunda parte do lema de sombreamento, existe algum ponto periódico $p \in M$, de período m_s , cuja trajetória ε -sombreia esta pseudo-órbita periódica $(z_n)_n$. Em particular, valem as condições (i) e (ii) no enunciado. \square

A propriedade na conclusão da Proposição 11.3.1 foi introduzida por Rufus Bowen [Bow71] e é denominada *especificação por pontos periódicos*. Quando vale somente a propriedade (i), sem exigir que o ponto p seja periódico, dizemos que f tem a propriedade de *especificação*.

11.3.1 Taxa de crescimento dos pontos periódicos

Seja $f : M \rightarrow M$ uma transformação expansora. Então f é expansiva (pelo Lema 11.1.4) e, portanto, segue da Proposição 10.2.2 que a taxa de crescimento do número de pontos periódicos é majorada pela entropia topológica:

$$\limsup_n \frac{1}{n} \log \# \text{Fix}(f^n) \leq h(f). \quad (11.3.1)$$

Nesta seção vamos provar que, de fato, vale a igualdade em (11.3.1). Começamos pelo caso topologicamente exato, o qual admite um enunciado ainda mais preciso:

Proposição 11.3.2. *Para toda transformação expansora topologicamente exata $f : M \rightarrow M$,*

$$\lim_n \frac{1}{n} \log \# \text{Fix}(f^n) = h(f).$$

Demonstração. Dado $\varepsilon > 0$, fixe $\kappa \geq 1$ satisfazendo a conclusão da Proposição 11.3.1 com $\varepsilon/2$ no lugar de ε . Para cada $n \geq 1$, seja E um conjunto (n, ε) -separado maximal. Pela proposição, para cada $x \in E$ existe $p(x) \in B(x, n, \varepsilon/2)$ com $f^{n+\kappa}(p(x)) = p(x)$. Afirmamos que a aplicação $x \mapsto p(x)$ é injetiva. De fato, considere qualquer $y \in E \setminus \{x\}$. Como o conjunto E foi tomado maximal, $B(x, n, \varepsilon/2) \cap B(y, n, \varepsilon/2) = \emptyset$. Isso implica que $p(x) \neq p(y)$, o que prova a nossa afirmação. Segue, em particular, que

$$\# \text{Fix}(f^{n+\kappa}) \geq \#E = s_n(f, \varepsilon, M) \quad \text{para todo } n \geq 1$$

(lembre da definição (10.1.9) na Seção 10.2.1). Então,

$$\liminf_n \frac{1}{n} \log \# \text{Fix}(f^{n+\kappa}) \geq \liminf_n \frac{1}{n} \log s_n(f, \varepsilon, M).$$

Fazendo $\varepsilon \rightarrow 0$ e usando o Corolário 10.1.8, obtemos que

$$\liminf_n \frac{1}{n} \log \# \text{Fix}(f^{n+\kappa}) \geq \lim_{\varepsilon \rightarrow 0} \liminf_n \frac{1}{n} \log s_n(f, \varepsilon, M) = h(f). \quad (11.3.2)$$

Juntamente com (11.3.1), isto implica a conclusão da proposição. \square

A conclusão da Proposição 11.3.2 não vale, em geral, se f não é topologicamente exata. Por exemplo, dada uma transformação expansora g qualquer, considere $f : M \times \{0, 1\} \rightarrow M \times \{0, 1\}$ definida por $f(x, i) = (g(x), 1 - i)$. Então f é transformação expansora, mas todos os seus pontos periódicos têm período par. Em particular, neste caso

$$\liminf_n \frac{1}{n} \log \# \text{Fix}(f^n) = 0.$$

No entanto, a próxima proposição, incluindo a sua demonstração, mostra que este tipo de exemplo é o pior que pode ocorrer.

Proposição 11.3.3. *Para toda transformação expansora,*

$$\limsup_n \frac{1}{n} \log \# \text{Fix}(f^n) = h(f).$$

Demonstração. Pelo Teorema 11.2.11, a restrição f ao conjunto dos pontos periódicos é uma transformação expansora. De acordo com o Exercício 11.2.5 essa restrição tem a mesma entropia que f . Evidentemente, as duas transformações têm os mesmos pontos periódicos. Portanto, a menos de substituir f pela restrição, podemos supor que o conjunto dos pontos periódicos é denso em M . Então, pelo teorema de decomposição dinâmica (Teorema 11.2.11) é possível escrever M como união disjunta de conjuntos compactos $M_{i,j}$, com $1 \leq i \leq k$ e $1 \leq j \leq m(i)$, tais que cada $f^{m(i)} : M_{i,j} \rightarrow M_{i,j}$ é uma transformação expansora topologicamente exata. De acordo com o Exercício 11.2.6, existe algum $1 \leq i \leq k$ tal que

$$h(f) = \frac{1}{m(i)} h(f^{m(i)} | M_{i,1}). \quad (11.3.3)$$

É claro que

$$\begin{aligned} \limsup_n \frac{1}{n} \log \# \text{Fix}(f^n) &\geq \limsup_n \frac{1}{nm(i)} \log \# \text{Fix}(f^{nm(i)}) \\ &\geq \frac{1}{m(i)} \limsup_n \frac{1}{n} \log \# \text{Fix}((f^{m(i)} | M_{i,1})^n). \end{aligned} \quad (11.3.4)$$

Além disso, a Proposição 11.3.2 aplicada a $f^{m(i)} : M_{i,1} \rightarrow M_{i,1}$ dá que

$$\lim_n \frac{1}{n} \log \# \text{Fix}((f^{m(i)} | M_{i,1})^n) = h(f^{m(i)} | M_{i,1}). \quad (11.3.5)$$

Combinando (11.3.3)–(11.3.5) obtemos que

$$\limsup_n \frac{1}{n} \log \# \text{Fix}(f^n) \geq h(f), \quad (11.3.6)$$

tal como queríamos provar. \square

11.3.2 Aproximação por medidas atômicas

Dado um ponto periódico p de período n , considere a medida μ_p definida por

$$\mu_p = \frac{1}{n} (\delta_p + \delta_{f(p)} + \cdots + \delta_{f^{n-1}(p)}).$$

Claramente, a medida μ_p é invariante por f e é ergódica. Vamos mostrar que se f é expansora então o conjunto das medidas desta forma é denso no espaço $\mathcal{M}_1(f)$ das probabilidades invariantes:

Teorema 11.3.4. *Seja $f : M \rightarrow M$ uma transformação expansora. Então toda probabilidade μ invariante por f é aproximada na topologia fraca* por probabilidades invariantes suportadas em órbitas periódicas.*

Demonstração. Seja $\varepsilon > 0$ e seja $\Phi = \{\phi_1, \dots, \phi_N\}$ uma família finita de funções contínuas em M . Queremos mostrar que a vizinhança $V(\mu, \Phi, \varepsilon)$ definida em (2.1.1) contém alguma medida μ_p suportada numa órbita periódica. Pelo teorema de Birkhoff, para μ -quase todo ponto $x \in M$,

$$\tilde{\phi}_i(x) = \lim_n \frac{1}{n} \sum_{t=0}^{n-1} \phi_i(f^t(x)) \quad \text{existe para todo } i. \quad (11.3.7)$$

Fixe $C > \sup |\phi_i| \geq \sup |\tilde{\phi}_i|$ e tome $\delta > 0$ tal que

$$d(x, y) < \delta \quad \Rightarrow \quad |\phi_i(x) - \phi_i(y)| < \frac{\varepsilon}{5} \quad \text{para todo } i. \quad (11.3.8)$$

Fixe $\kappa = \kappa(\delta) \geq 1$ dado pela propriedade de especificação (Proposição 11.3.1). Escolha pontos x_j , $1 \leq j \leq s$ satisfazendo (11.3.7) e números positivos α_j , $1 \leq j \leq s$ tais que $\sum_j \alpha_j = 1$ e

$$\left| \int \tilde{\phi}_i d\mu - \sum_{j=1}^s \alpha_j \tilde{\phi}_i(x_j) \right| < \frac{\varepsilon}{5} \quad \text{para todo } i \quad (11.3.9)$$

(lembre do Exercício A.2.6). Tome $k_j \equiv \kappa$ e escolha inteiros n_j muito maiores do que κ , de tal modo que

$$\left| \frac{n_j}{m_s} - \alpha_j \right| < \frac{\varepsilon}{5Cs} \quad (11.3.10)$$

(lembre que $m_s = \sum_j (n_j + k_j) = s\kappa + \sum_j n_j$) e, usando (11.3.8),

$$\left| \sum_{t=0}^{n_j-1} \phi_i(f^t(x_j)) - n_j \tilde{\phi}_i(x_j) \right| < \frac{\varepsilon}{5} n_j \quad \text{para } 1 \leq i \leq N. \quad (11.3.11)$$

Combinando (11.3.9) e (11.3.10) e o fato de que $\int \tilde{\phi}_i d\mu = \int \phi_i d\mu$.

$$\left| \int \phi_i d\mu - \sum_{j=1}^s \frac{n_j}{m_s} \tilde{\phi}_i(x_j) \right| < \frac{\varepsilon}{5} + \frac{\varepsilon}{5Cs} s \sup |\tilde{\phi}_i| < \frac{2\varepsilon}{5}. \quad (11.3.12)$$

Pela Proposição 11.3.1, existe algum ponto periódico $p \in M$ com período m_s tal que $d(f^{m_j-1+t}(p), f^t(x_j)) < \delta$ para $0 \leq t < n_j$ e $1 \leq j \leq s$. Então, a propriedade (11.3.8) implica que

$$\left| \sum_{t=0}^{n_j-1} \phi_i(f^{m_j-1+t}(p)) - \sum_{t=0}^{n_j-1} \phi_i(f^t(x_j)) \right| < \frac{\varepsilon}{5} n_j \quad \text{para } 1 \leq j \leq s.$$

Somando esta relação com (11.3.11) obtemos

$$\left| \sum_{t=0}^{n_j-1} \phi_i(f^{m_j-1+t}(p)) - n_j \tilde{\phi}_i(x_j) \right| < \frac{2\varepsilon}{5} n_j \quad \text{para } 1 \leq j \leq s. \quad (11.3.13)$$

Como $\sum_j \alpha_j = 1$, a condição (11.3.10) implica que

$$s\kappa = m_s - \sum_{j=1}^s n_j < \frac{\varepsilon}{5C} m_s.$$

Então, (11.3.13) implica que

$$\left| \sum_{i=0}^{m_s-1} \phi_i(f^i(p)) - \sum_{j=1}^s n_j \tilde{\phi}_i(x_j) \right| < \frac{2\varepsilon}{5} \sum_{j=1}^s n_j + s\kappa \sup |\tilde{\phi}_i| < \frac{3\varepsilon}{5} m_s. \quad (11.3.14)$$

Seja μ_p a probabilidade invariante suportada na órbita de p . O primeiro termo em (11.3.14) coincide com $m_s \int \phi_i d\mu_p$. Portanto, somando as desigualdades (11.3.12) e (11.3.14), concluímos que

$$\left| \int \phi_i d\mu_p - \int \phi_i d\mu \right| < \frac{2\varepsilon}{5} + \frac{3\varepsilon}{5} = \varepsilon \quad \text{para todo } 1 \leq i \leq N.$$

Isto significa que $\mu_p \in V(\mu, \Phi, \varepsilon)$, como queríamos demonstrar. \square

11.3.3 Exercícios

11.3.1. Seja $f : M \rightarrow M$ uma transformação contínua num espaço métrico compacto M . Verifique que se algum iterado f^l , $l \geq 1$ tem a propriedade de especificação, ou especificação por pontos periódicos, então f também tem.

11.3.2. Seja $f : M \rightarrow M$ uma transformação contínua num espaço métrico com a propriedade de especificação. Mostre que f é topologicamente misturadora.

11.3.3. Seja $f : M \rightarrow M$ uma transformação expansora topologicamente misturadora e seja $\varphi : M \rightarrow \mathbb{R}$ uma função contínua. Suponha que existem probabilidades μ_1, μ_2 invariantes por f tais que $\int \varphi d\mu_1 \neq \int \varphi d\mu_2$. Mostre que então existe $x \in M$ tal que a média temporal de φ na órbita de x não converge. [Observação: Pode mostrar-se que o conjunto M_φ dos pontos onde a média temporal de φ não converge é grande: ele tem entropia total e também dimensão de Hausdorff total; veja [BS00].]

11.3.4. Prove a seguinte generalização da Proposição 11.3.2: se $f : M \rightarrow M$ é uma transformação expansora topologicamente exata então

$$P(f, \phi) = \lim_k \frac{1}{k} \log \sum_{p \in \text{Fix}(f^k)} e^{\phi_k(p)} \quad \text{para toda função Hölder } \phi : M \rightarrow \mathbb{R}.$$

11.3.5. Seja $f : M \rightarrow M$ uma transformação expansora diferenciável em uma variedade compacta M . Mostre que f admite:

- Uma vizinhança \mathcal{U}_0 relativamente à topologia da convergência uniforme tal que f é fator topológico de toda transformação $g \in \mathcal{U}_0$. Em particular, $h(g) \geq h(f)$ para todo $g \in \mathcal{U}_0$.
- Uma vizinhança \mathcal{U}_1 relativamente à topologia C^1 tal que toda $g \in \mathcal{U}_1$ é topologicamente conjugada a f . Em particular, a entropia topológica é constante em \mathcal{U}_1 .

Capítulo 12

Formalismo Termodinâmico

Neste capítulo desenvolvemos a teoria ergódica das transformações expansoras em espaços métricos compactos. Esta teoria teve sua inspiração na Mecânica Estatística e, por isso, é frequentemente chamada Formalismo Termodinâmico. Alertamos, no entanto, que esta última expressão se aplica num contexto bastante mais amplo que também inclui, em particular, os difeomorfismos e fluxos chamados uniformemente hiperbólicos (a este respeito, veja a excelente monografia de Rufus Bowen [Bow75a]).

O principal resultado deste capítulo é o seguinte teorema de David Ruelle, que será provado na Seção 12.1 (a noção de estado de Gibbs também será definida na Seção 12.1):

Teorema 12.1 (Ruelle). *Seja $f : M \rightarrow M$ uma transformação expansora topologicamente exata num espaço métrico compacto e seja $\varphi : M \rightarrow \mathbb{R}$ uma função Hölder. Então existe um único estado de equilíbrio μ para φ . Além disso, a medida μ é exata, está suportada em todo o M e é um estado de Gibbs.*

Lembre que uma transformação expansora é topologicamente exata se ela for topologicamente misturadora (Exercício 11.2.2).

No caso particular em que M é uma variedade e f é diferenciável, o estado de equilíbrio μ_u do potencial $\varphi_u = -\log |\det Df|$ coincide com a medida invariante absolutamente contínua dada pelo Teorema 11.1.2. Em particular, ele é a única medida física da transformação. Estes fatos serão vistos na Seção 12.1.8.

O teorema de Livsič, que será provado na Seção 12.2, complementa o teorema de Ruelle de maneira muito elegante: dois potenciais φ e ψ têm o mesmo estado de equilíbrio se, e somente se, a sua diferença é cohomóloga a uma constante. Em outras palavras, isto acontece se, e somente se, $\varphi - \psi = c + u \circ f - u$ para algum $c \in \mathbb{R}$ e alguma função contínua u . Além disso, notavelmente, basta verificar essa condição nas órbitas periódicas de f .

Também veremos, na Seção 12.3, que o sistema (f, μ) no teorema apresenta decaimento exponencial de correlações para todas as funções Hölder.

Encerraremos o capítulo (Seção 12.4) com uma aplicação destas ideias a uma classe de objetos geométricos e dinâmicos chamados *repulsores conformes*.

Mostraremos que vale a *fórmula de Bowen-Manning*: a dimensão de Hausdorff do repulsor é dada pelo único zero da função $t \mapsto P(f, t\varphi_u)$.

12.1 Teorema de Ruelle

Seja $f : M \rightarrow M$ uma transformação expansora topologicamente exata e seja φ um potencial Hölder. No que segue, $\rho > 0$ e $\sigma > 1$ são as mesmas constantes da definição (11.2.1). Representaremos por $S_n\varphi$ as somas orbitais de φ :

$$S_n\varphi(x) = \sum_{j=0}^{n-1} \varphi(f^j(x)). \quad (12.1.1)$$

Antes de entrarmos nos detalhes da demonstração do Teorema 12.1 vamos esboçar as suas linhas gerais. Os argumentos giram em torno do *operador de transferência* (ou *operador de Ruelle-Perron-Frobenius*) que é o operador linear $\mathcal{L} : C^0(M) \rightarrow C^0(M)$ definido no espaço $C^0(M)$ das funções contínuas complexas por

$$\mathcal{L}g(y) = \sum_{x \in f^{-1}(y)} e^{\varphi(x)} g(x). \quad (12.1.2)$$

Observe que \mathcal{L} está bem definido: $\mathcal{L}g \in C^0(M)$ sempre que $g \in C^0(M)$. De fato, como vimos no Lema 11.2.6, para cada $y \in M$ existem ramos inversos $h_i : B(y, \rho) \rightarrow M$, $i = 1, \dots, k$ da transformação f tais que $\cup_{i=1}^k h_i(B(y, \rho))$ coincide com a pré-imagem da bola $B(y, \rho)$. Então

$$\mathcal{L}g = \sum_{i=1}^k (e^\varphi g) \circ h_i \quad (12.1.3)$$

restrito a $B(y, \rho)$ e, claramente, esta expressão define uma função contínua.

É claro da definição que \mathcal{L} é um operador positivo: se $g(x) \geq 0$ para todo $x \in M$ então $\mathcal{L}g(y) \geq 0$ para todo $y \in M$. Também é fácil verificar que \mathcal{L} é um operador contínuo: de fato,

$$\|\mathcal{L}g\| = \sup |\mathcal{L}g| \leq \text{grau}(f) e^{\sup \varphi} \sup |g| = \text{grau}(f) e^{\sup \varphi} \|g\| \quad (12.1.4)$$

para todo $g \in C^0(M)$ e isso significa que $\|\mathcal{L}\| \leq \text{grau}(f) e^{\sup \varphi}$.

De acordo com o teorema de Riesz-Markov (Teorema A.3.12), o dual do espaço de Banach $C^0(M)$ se identifica com o espaço vetorial $\mathcal{M}(M)$ das medidas borelianas complexas. Então, o *dual* do operador de transferência é o operador linear $\mathcal{L}^* : \mathcal{M}(M) \rightarrow \mathcal{M}(M)$ definido por

$$\int g d(\mathcal{L}^*\eta) = \int (\mathcal{L}g) d\eta \quad \text{para todo } g \in C^0(M) \text{ e } \eta \in \mathcal{M}(M). \quad (12.1.5)$$

Este operador linear é positivo, no sentido de que se η é uma medida positiva então $\mathcal{L}^*\eta$ também é uma medida positiva.

O primeiro passo da demonstração (Seção 12.1.1) será mostrar que \mathcal{L}^* admite uma automedida positiva ν associada a um autovalor positivo λ . Veremos que tal medida admite jacobiano positivo e Hölder e está suportada em todo o espaço M . Além disso (Seção 12.1.2), ela é um *estado de Gibbs*: existem constantes $K \geq 1$ e $P \in \mathbb{R}$ tais que

$$K^{-1} \leq \frac{\nu(B(x, n, \varepsilon))}{\exp(S_n \varphi(x) - nP)} \leq K \quad \text{para todo } x \in M \text{ e todo } n, \quad (12.1.6)$$

onde $B(x, n, \varepsilon)$ é a bola dinâmica definida em (9.3.2). De fato, $P = \log \lambda$.

Por trás da demonstração da propriedade de Gibbs estão certos resultados sobre controle da distorção que também serão cruciais para mostrar (Seção 12.1.3) que o próprio operador de transferência \mathcal{L} admite uma autofunção associada ao autovalor λ . Esta função é estritamente positiva e Hölder. A medida $\mu = h\nu$ será o estado de equilíbrio que buscamos (Seção 12.1.4). Segue facilmente das propriedades de h que esta medida μ é invariante, suportada em todo o M e um estado de Gibbs. Além disso, $h_\mu(f) + \int \varphi d\mu = P$.

Para concluir que μ é de fato um estado de equilíbrio precisamos verificar que P é igual à pressão $P(f, \varphi)$. Isso será feito (Seção 12.1.5) com a ajuda da fórmula de Rokhlin (Teorema 9.7.3), a qual também nos permitirá concluir que se η é estado de equilíbrio então η/h é automedida de \mathcal{L}^* associada ao autovalor $\lambda = \log P(f, \varphi)$. Este último resultado será o ingrediente crucial para demonstrar que o estado de equilíbrio é único (Seção 12.1.6).

Finalmente (Seção 12.1.7), verificaremos que o sistema (f, μ) é exato. O controle de distorção terá, mais uma vez, um papel crucial. Na Seção 12.1.8, comentaremos o caso especial $\varphi = -\log |\det f|$, quando f é transformação expansora numa variedade Riemanniana. Neste caso a medida de referência ν é a própria medida de Lebesgue na variedade e, portanto, o estado de equilíbrio é uma medida invariante equivalente à medida de Lebesgue. Logo, μ coincide com a medida invariante que construímos na Seção 11.1.

Antes de passarmos a detalhar estes argumentos é conveniente fazermos dois comentários breves. Primeiramente, note que a existência de estado de equilíbrio segue imediatamente do Corolário 10.5.9, já que vimos no Lema 11.1.4 que toda transformação expansora é expansiva. No entanto este fato não será usado na demonstração: em vez disso, na Seção 12.1.4 apresentaremos uma construção muito mais explícita do estado de equilíbrio.

O outro comentário diz respeito à fórmula de Rokhlin. Seja \mathcal{P} uma partição finita qualquer de M com $\text{diam } \mathcal{P} < \rho$. Para cada $n \geq 1$, todo elemento de $\mathcal{P}^n = \bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P})$ está contido na imagem $h^{n-1}(P)$ de algum $P \in \mathcal{P}$ por algum ramo inverso h^{n-1} do iterado f^{n-1} . Em particular, $\text{diam } \mathcal{P}^n < \sigma^{-n+1}\rho$ para todo n . Então \mathcal{P} satisfaz as hipóteses do Teorema 9.7.3 *em todo ponto*. Logo, a fórmula de Rokhlin vale para toda probabilidade invariante de f .

12.1.1 Medida de referência

Lembre que $C_+^0(M)$ representa o cone das funções contínuas positivas. Conforme já observamos, este cone é preservado pelo operador de transferência \mathcal{L} . O cone

dual (lembre do Exemplo 2.3.3) é definido por

$$C_+^0(M)^* = \{\eta \in C^0(M)^* : \eta(\psi) \geq 0 \text{ para todo } \psi \in C_+^0(M)\}.$$

e pode ser visto como o cone das medidas borelianas positivas finitas. Segue diretamente de (12.1.5) que $C_+^0(M)^*$ é preservado pelo operador dual \mathcal{L}^* .

Lema 12.1.1. *Considere o raio espectral $\lambda = r(\mathcal{L}^*) = r(\mathcal{L})$. Então existe alguma probabilidade ν em M tal que $\mathcal{L}^*\nu = \lambda\nu$.*

Demonstração. Como vimos no Exercício 2.3.3, o cone $C_+^0(M)$ é normal. Logo, estamos em condições de aplicar o Teorema 2.3.4 com $E = C^0(M)$, $C = C_+^0(M)$ e $T = \mathcal{L}$. A conclusão do teorema significa que \mathcal{L}^* admite algum autovetor $\nu \in C_+^0(M)^*$ correspondente ao autovalor λ . Conforme acabamos de explicar, ν se identifica com uma medida positiva finita. Normalizando ν , podemos supor que se trata de uma probabilidade. \square

No Exercício 12.1.2 propomos uma prova alternativa para o Lema 12.1.1, a partir do teorema de Tychonoff-Schauder (Teorema 2.2.3).

Exemplo 12.1.2. Seja $f : M \rightarrow M$ um difeomorfismo local numa variedade Riemanniana compacta M . Considere o operador de transferência \mathcal{L} associado ao potencial $\varphi = -\log |\det Df|$. A medida de Lebesgue m de M é uma automedida do operador dual, correspondente ao autovalor $\lambda = 1$:

$$\mathcal{L}^*m = m. \tag{12.1.7}$$

Para verificar esse fato, basta mostrar que $\mathcal{L}^*m(E) = m(E)$ para todo conjunto mensurável E contido na imagem de uma bola $B(y, \rho)$ por algum ramo inverso $h_j : B(y, \rho) \rightarrow M$ (pois, pela compacidade de M , todo conjunto mensurável pode ser escrito como união finita disjunta de subconjuntos E deste tipo). Ora, usando a expressão (12.1.3),

$$\mathcal{L}^*m(E) = \int \mathcal{X}_E d(\mathcal{L}^*m) = \int (\mathcal{L}\mathcal{X}_E) dm = \int \sum_{i=1}^k \frac{\mathcal{X}_E}{|\det Df|} \circ h_i dm.$$

Então, pela escolha de E e pela fórmula de mudança de variáveis,

$$\mathcal{L}^*m(E) = \int \frac{\mathcal{X}_E}{|\det Df|} \circ h_j dm = \int \mathcal{X}_E dm = m(E).$$

Isto prova que m é, de fato, um ponto fixo de \mathcal{L}^* .

O Exercício 12.1.3 fornece uma conclusão análoga para medidas de Markov.

A partir daqui sempre suporemos que ν é uma *medida de referência*, ou seja, uma probabilidade satisfazendo $\mathcal{L}^*\nu = \lambda\nu$ para *algum* $\lambda > 0$. Uma das nossas conclusões ao final da demonstração do teorema de Ruelle será que λ está unicamente determinado (em vista do Lema 12.1.1, isso quer dizer que λ é necessariamente igual ao raio espectral de \mathcal{L} e \mathcal{L}^*) e que a própria medida ν também é única.

Inicialmente, vamos mostrar que existe jacobiano de f relativamente a ν , o qual pode ser escrito explicitamente em termos do autovalor λ e do potencial φ :

Lema 12.1.3. *A transformação $f : M \rightarrow M$ admite jacobiano relativamente a ν , dado por $J_\nu f = \lambda e^{-\varphi}$.*

Demonstração. Seja A um domínio de invertibilidade qualquer de f . Seja $(g_n)_n$ uma sequência de funções contínuas convergindo em ν -quase todo ponto para a função característica de A e tal que $\sup |g_n| \leq 1$ para todo n (veja o Exercício A.3.5). Observe que

$$\mathcal{L}(e^{-\varphi} g_n)(y) = \sum_{x \in f^{-1}(y)} g_n(x).$$

A expressão do lado direito é limitada pelo grau de f , que foi definido em (11.2.4), e ela converge para $\chi_{f(A)}(y)$ em ν -quase todo ponto. Logo, usando o teorema da convergência dominada,

$$\int \lambda e^{-\varphi} g_n d\nu = \int e^{-\varphi} g_n d(\mathcal{L}^* \nu) = \int \mathcal{L}(e^{-\varphi} g_n) d\nu$$

converge para $\nu(f(A))$. Como a expressão do lado esquerdo converge para $\int_A \lambda e^{-\varphi} d\nu$, concluímos que

$$\nu(f(A)) = \int_A \lambda e^{-\varphi} d\nu,$$

o que prova a afirmação. \square

O próximo lema se aplica, em particular, à medida de referência ν :

Lema 12.1.4. *Seja $f : M \rightarrow M$ uma transformação expansora topologicamente exata e seja η qualquer probabilidade boreliana tal que existe jacobiano de f relativamente a η . Então η está suportada em todo o M .*

Demonstração. Suponha, por absurdo, que existe algum aberto $U \subset M$ tal que $\eta(U) = 0$. Note que f é uma aplicação aberta, uma vez que é um homeomorfismo local. Então, a imagem $f(U)$ também é um aberto. Além disso, podemos cobrir U com uma união finita de domínios de invertibilidade A . Para cada um deles,

$$\eta(f(A)) = \int_A J_\eta f d\eta = 0.$$

Portanto, $\eta(f(U)) = 0$. Por indução, segue que $\eta(f^n(U)) = 0$ para todo $n \geq 0$. Por outro lado, como supomos que f é topologicamente exata, existe $k \geq 1$ tal que $f^k(U) = M$. Como $\eta(M) = 1$, isto gera uma contradição. \square

12.1.2 Distorção e propriedade de Gibbs

Nesta seção vamos provar alguns resultados de controle da distorção limitada que têm um papel central na nossa construção. A hipótese de que φ é Hölder é crítica a esta altura: estes resultados são falsos, em geral, se o potencial é apenas

contínuo. Como uma primeira aplicação do controle de distorção mostraremos que ν é um estado de Gibbs.

Fixe constantes $K_0 > 0$ e $\alpha > 0$ tais que $|\varphi(z) - \varphi(w)| \leq K_0 d(z, w)^\alpha$ para quaisquer $z, w \in M$.

Lema 12.1.5. *Existe $K_1 > 0$ tal que para todo $n \geq 1$, todo $x \in M$ e todo $y \in B(x, n + 1, \rho)$,*

$$|S_n \varphi(x) - S_n \varphi(y)| \leq K_1 d(f^n(x), f^n(y))^\alpha.$$

Demonstração. Por hipótese, $d(f^i(x), f^i(y)) < \rho$ para todo $0 \leq i \leq n$. Então, para cada $j = 1, \dots, n$, o ramo contrativo $h_j : B(f^n(x), \rho) \rightarrow M$ de f^j que envia $f^n(x)$ em $f^{n-j}(x)$ também envia $f^n(y)$ em $f^{n-j}(y)$. Logo, lembrando (11.2.6), temos que $d(f^{n-j}(x), f^{n-j}(y)) \leq \sigma^{-j} d(f^n(x), f^n(y))$ para todo $j = 1, \dots, n$. Então,

$$\begin{aligned} |S_n \varphi(x) - S_n \varphi(y)| &\leq \sum_{j=1}^n |\varphi(f^{n-j}(x)) - \varphi(f^{n-j}(y))| \\ &\leq \sum_{j=1}^n K_0 \sigma^{-j\alpha} d(f^n(x), f^n(y))^\alpha. \end{aligned}$$

Portanto, basta tomar $K_1 \geq K_0 \sum_{j=0}^{\infty} \sigma^{-j\alpha}$. \square

Como consequência do Lema 12.1.5, obtemos a seguinte variação da Proposição 11.1.5 em que o jacobiano usual com respeito à medida de Lebesgue é substituído pelo jacobiano relativamente à medida de referência ν :

Corolário 12.1.6. *Existe $K_2 > 0$ tal que para todo $n \geq 1$, todo $x \in M$ e todo $y \in B(x, n + 1, \rho)$,*

$$K_2^{-1} \leq \frac{J_\nu f^n(x)}{J_\nu f^n(y)} \leq K_2.$$

Demonstração. Da expressão do jacobiano no Lema 12.1.3 segue que (lembre do Exercício 9.7.6)

$$J_\nu f^n(z) = \lambda^n e^{-S_n \varphi(z)} \quad \text{para todo } z \in M \text{ e todo } n \geq 1. \quad (12.1.8)$$

Então o Lema 12.1.5 dá que

$$\left| \log \frac{J_\nu f^n(x)}{J_\nu f^n(y)} \right| = |S_n \varphi(x) - S_n \varphi(y)| \leq K_1 d(f^n(x), f^n(y))^\alpha \leq K_1 \rho^\alpha.$$

Assim, basta escolher $K_2 = \exp(K_1 \rho^\alpha)$. \square

Agora podemos mostrar que ν é um estado de Gibbs:

Lema 12.1.7. *Para todo $\varepsilon > 0$ suficientemente pequeno existe $K_3 = K_3(\varepsilon) > 0$ tal que, escrevendo $P = \log \lambda$,*

$$K_3^{-1} \leq \frac{\nu(B(x, n, \varepsilon))}{\exp(S_n \varphi(x) - nP)} \leq K_3 \quad \text{para todo } x \in M \text{ e todo } n \geq 1.$$

Demonstração. Considere $\varepsilon < \rho$. Então $f \mid B(y, \varepsilon)$ é injetiva para todo $y \in M$ e, conseqüentemente, $f^n \mid B(x, n, \varepsilon)$ é injetiva para todo $x \in M$ e todo n . Então,

$$\nu(f^n(B(x, n, \varepsilon))) = \int_{B(x, n, \varepsilon)} J_\nu f^n(y) d\nu(y).$$

Pelo Corolário 12.1.6, o valor de $J_\nu f^n$ num ponto qualquer $y \in B(x, n, \varepsilon)$ difere de $J_\nu f^n(x)$ por um fator limitado pela constante K_2 . Segue que

$$K_2^{-1} \nu(f^n(B(x, n, \varepsilon))) \leq J_\nu f^n(x) \nu(B(x, n, \varepsilon)) \leq K_2 \nu(f^n(B(x, n, \varepsilon))). \quad (12.1.9)$$

Agora, $J_\nu f^n(x) = \lambda^n e^{-S_n \varphi(x)} = \exp(nP - S_n \varphi(x))$, como vimos em (12.1.8). Pelo Lema 11.2.7 também temos que $f^n(B(x, n, \varepsilon)) = f(B(f^{n-1}(x), \varepsilon))$ e, portanto,

$$\nu(f^n(B(x, n, \varepsilon))) = \int_{B(f^{n-1}(x), \varepsilon)} J_\nu f d\nu \quad (12.1.10)$$

para todo $x \in M$ e todo n . É claro que o lado esquerdo de (12.1.10) é majorado por 1. Além disso, $J_\nu f = \lambda e^{-\varphi}$ é limitado de zero e $\{\nu(B(y, \varepsilon)) : y \in M\}$ também é limitado de zero (pelo Exercício 12.1.1 e o Lema 12.1.4). Portanto o lado direito de (12.1.10) é minorado por algum número $a > 0$. Usando estas observações em (12.1.9), obtemos

$$K_2^{-1} a \leq \frac{\nu(B(x, n, \varepsilon))}{\exp(S_n \varphi(x) - nP)} \leq K_2.$$

Agora basta tomar $K_3 = \max\{K_2/a, K_2\}$. □

12.1.3 Densidade invariante

Em seguida vamos mostrar que o operador \mathcal{L} admite alguma auto-função positiva associada ao autovalor $\lambda > 0$. Ela será construída como um ponto de acumulação Cesaro da sequência de funções $\lambda^{-n} \mathcal{L}^n 1$. Para mostrar que existe algum ponto de acumulação, começamos por provar que esta sequência é uniformemente limitada e equicontínua.

Lema 12.1.8. *Existe $K_4 > 0$ tal que*

$$-K_4 d(y_1, y_2)^\alpha \leq \log \frac{\mathcal{L}^n 1(y_1)}{\mathcal{L}^n 1(y_2)} \leq K_4 d(y_1, y_2)^\alpha$$

para todo $n \geq 1$ e quaisquer $y_1, y_2 \in M$ com $d(y_1, y_2) < \rho$.

Demonstração. Segue de (12.1.3) que, dada qualquer função contínua g ,

$$\mathcal{L}^n g = \sum_i (e^{S_n \varphi} g) \circ h_i^n \quad \text{restrito a cada bola } B(y, \rho),$$

onde a soma é sobre os ramos inversos $h_i^n : B(y, \rho) \rightarrow M$ do iterado f^n . Em particular,

$$\frac{\mathcal{L}^n 1(y_1)}{\mathcal{L}^n 1(y_2)} = \frac{\sum_i e^{S_n \varphi(h_i^n(y_1))}}{\sum_i e^{S_n \varphi(h_i^n(y_2))}}.$$

Pelo Lema 12.1.5, para cada um desses ramos inversos h_i^n tem-se

$$|S_n \varphi(h_i^n(y_1)) - S_n \varphi(h_i^n(y_2))| \leq K_1 d(y_1, y_2)^\alpha.$$

Consequentemente,

$$e^{-K_1 d(y_1, y_2)^\alpha} \leq \frac{\mathcal{L}^n 1(y_1)}{\mathcal{L}^n 1(y_2)} \leq e^{K_1 d(x_1, x_2)^\alpha}.$$

Portanto, basta tomar $K_4 \geq K_1$. \square

Segue que a sequência $\lambda^{-n} \mathcal{L}^n 1$ é limitada de zero e de infinito:

Corolário 12.1.9. *Existe $K_5 > 0$ tal que $K_5^{-1} \leq \lambda^{-n} \mathcal{L}^n 1(x) \leq K_5$ para todo $n \geq 1$ e quaisquer $x, y \in M$.*

Demonstração. Comece por observar que, para todo $n \geq 1$,

$$\int \mathcal{L}^n 1 d\nu = \int 1 d(\mathcal{L}^{*n} \nu) = \int \lambda^n d\nu = \lambda^n.$$

Em particular, para todo $n \geq 1$,

$$\min_{y \in M} \lambda^{-n} \mathcal{L}^n 1(y) \leq 1 \leq \max_{y \in M} \lambda^{-n} \mathcal{L}^n 1(y). \quad (12.1.11)$$

Como f é topologicamente exata, existe $N \geq 1$ tal que $f^N(B(x, \rho)) = M$ para todo $x \in M$ (confira o Exercício 11.2.3). Agora, dados $x, y \in M$ quaisquer, podemos encontrar $x' \in B(x, \rho)$ tal que $f^N(x') = y$. Então, por um lado,

$$\mathcal{L}^{n+N} 1(y) = \sum_{z \in f^{-N}(y)} e^{S_N \varphi(z)} \mathcal{L}^n 1(z) \geq e^{S_N \varphi(x')} \mathcal{L}^n 1(x') \geq e^{-cN} \mathcal{L}^n 1(x').$$

Por outro lado, o Lema 12.1.8 dá que $\mathcal{L}^n 1(x') \geq \mathcal{L}^n 1(x) \exp(-K_4 \rho^\alpha)$. Tome $c = \sup |\varphi|$ e $K \geq \exp(K_4 \rho^\alpha) e^{cN} \lambda^N$. Combinando as desigualdades anteriores vem que

$$\mathcal{L}^{n+N} 1(y) \geq \exp(-K_4 \rho^\alpha) e^{-cN} \mathcal{L}^n 1(x) \geq K^{-1} \lambda^N \mathcal{L}^n 1(x)$$

para todo $x, y \in M$. Portanto, para todo $n \geq 1$,

$$\min \lambda^{-(n+N)} \mathcal{L}^{n+N} 1 \geq K^{-1} \max \lambda^{-n} \mathcal{L}^n 1. \quad (12.1.12)$$

Combinando (12.1.11) e (12.1.12) obtemos,

$$\begin{aligned} \max \lambda^{-n} \mathcal{L}^n 1 &\leq K \min \lambda^{-(n+N)} \mathcal{L}^{n+N} 1 \leq K \quad \text{para todo } n \geq 1 \\ \min \lambda^{-n} \mathcal{L}^n 1 &\geq K^{-1} \max \lambda^{-n+N} \mathcal{L}^{n-N} 1 \geq K^{-1} \quad \text{para todo } n > N. \end{aligned}$$

Para terminar a demonstração, só falta estender esta última estimativa para os valores $n = 1, \dots, N$. Para isso, observe que cada $\mathcal{L}^n 1$ é uma função contínua e positiva. Logo, pela compacidade de M , o mínimo de $\mathcal{L}^n 1$ é positivo para todo n . Então, podemos tomar $K_5 \geq K$ tal que $\min \lambda^{-n} \mathcal{L}^n 1 \geq K_5^{-1}$ para todo $n = 1, \dots, N$. \square

Segue imediatamente do Corolário 12.1.9 que o autovalor λ está unicamente determinado. Pelo Lema 12.1.1, isto implica que $\lambda = r(\mathcal{L}) = r(\mathcal{L}^*)$. Também veremos daqui a pouco que $\lambda = e^{P(f,\varphi)}$.

Lema 12.1.10. *Existe $K_6 > 0$ tal que*

$$|\lambda^{-n} \mathcal{L}^n 1(x) - \lambda^{-n} \mathcal{L}^n 1(y)| \leq K_6 d(x, y)^\alpha \quad \text{para quaisquer } n \geq 1 \text{ e } x, y \in M.$$

Em particular, a sequência $\lambda^{-n} \mathcal{L}^n 1$ é equicontínua.

Demonstração. Suponha inicialmente que $d(x, y) < \rho$. Pelo Lema 12.1.8,

$$\mathcal{L}^n 1(x) \leq \mathcal{L}^n 1(y) \exp(K_4 d(x, y)^\alpha)$$

e, portanto,

$$\lambda^{-n} \mathcal{L}^n 1(x) - \lambda^{-n} \mathcal{L}^n 1(y) \leq [\exp(K_4 d(x, y)^\alpha) - 1] \lambda^{-n} \mathcal{L}^n 1(y).$$

Tome $K > 0$ tal que $|\exp(K_4 t) - 1| \leq K|t|$ sempre que $|t| \leq \rho^\alpha$. Então, usando o Corolário 12.1.9,

$$\lambda^{-n} \mathcal{L}^n 1(x) - \lambda^{-n} \mathcal{L}^n 1(y) \leq K K_5 d(x, y)^\alpha.$$

Invertendo os papéis de x e y concluímos que

$$|\lambda^{-n} \mathcal{L}^n 1(x) - \lambda^{-n} \mathcal{L}^n 1(y)| \leq K K_5 d(x, y)^\alpha \quad \text{sempre que } d(x, y) < \rho.$$

Quando $d(x, y) \geq \rho$ o Corolário 12.1.9 dá que

$$|\lambda^{-n} \mathcal{L}^n 1(x) - \lambda^{-n} \mathcal{L}^n 1(y)| \leq 2K_5 \leq 2K_5 \rho^{-\alpha} d(x, y)^\alpha.$$

Logo, basta tomar $K_6 \geq \max\{K K_5, 2K_5 \rho^{-\alpha}\}$ para obter a primeira parte do enunciado. A segunda parte é consequência imediata. \square

Estamos prontos para mostrar que o operador \mathcal{L} admite alguma autofunção associada ao autovalor λ . O Corolário 12.1.9 e o Lema 12.1.10 implicam que a média

$$h_n = \frac{1}{n} \sum_{i=0}^{n-1} \lambda^{-i} \mathcal{L}^i 1,$$

define uma sequência limitada e equicontínua. Então, pelo teorema de Ascoli-Arzelá, existe alguma subsequência $(h_{n_i})_i$ convergindo uniformemente para uma função contínua h .

Lema 12.1.11. *A função h satisfaz $\mathcal{L}h = \lambda h$. Além disso, $\int h \, d\nu = 1$ e*

$$K_5^{-1} \leq h(x) \leq K_5 \quad \text{e} \quad |h(x) - h(y)| \leq K_6 d(x, y)^\alpha \quad \text{para todo } x, y \in M.$$

Demonstração. Considere qualquer subsequência $(h_{n_i})_i$ convergindo para h . Como o operador \mathcal{L} é contínuo,

$$\begin{aligned}\mathcal{L}h &= \lim_i \mathcal{L}h_{n_i} = \lim_i \frac{1}{n_i} \sum_{k=0}^{n_i-1} \lambda^{-k} \mathcal{L}^{k+1} 1 = \lim_i \frac{\lambda}{n_i} \sum_{k=1}^{n_i} \lambda^{-k} \mathcal{L}^k 1 \\ &= \lim_i \frac{\lambda}{n_i} \sum_{k=0}^{n_i-1} \lambda^{-k} \mathcal{L}^k 1 + \frac{\lambda}{n_i} (\lambda^{-n_i} \mathcal{L}^{n_i} 1 - 1).\end{aligned}$$

O primeiro termo do lado direito converge para λh e o segundo converge para zero, uma vez que a sequência $\lambda^{-n} \mathcal{L}^n 1$ é limitada. Portanto, $\mathcal{L}h = \lambda h$ tal como afirmamos.

Pela definição de ν , temos $\int \lambda^{-n} \mathcal{L}^n 1 d\nu = \int \lambda^{-n} d(\mathcal{L}^{*n} \nu) = \int 1 d\nu = 1$ para todo $n \in \mathbb{N}$. Segue que $\int h_n d\nu = 1$ para todo n e, usando o teorema da convergência dominada, $\int h d\nu = 1$. As demais afirmações no enunciado seguem, de modo inteiramente análogo, do Corolário 12.1.9 e do Lema 12.1.10. \square

12.1.4 Construção do estado de equilíbrio

Considere a medida definida por $\mu = h\nu$, ou seja

$$\mu(A) = \int_A h d\nu \quad \text{para cada conjunto mensurável } A \subset M.$$

Veremos a seguir que μ é estado de equilíbrio para o potencial φ e verifica todas as demais condições no Teorema 12.1.

Do Lema 12.1.11 vem que $\mu(M) = \int h d\nu = 1$ e, portanto, μ é uma medida de probabilidade. Além disso,

$$K_5^{-1} \nu(A) \leq \mu(A) \leq K_5 \nu(A) \quad (12.1.13)$$

para todo conjunto mensurável $A \subset M$. Em particular, μ é equivalente à medida de referência ν . Este fato, juntamente com o Lema 12.1.4 dá que $\text{supp } \mu = M$. Também segue da relação (12.1.13), juntamente com o Lema 12.1.7, que μ é um estado de Gibbs: tomando $L = K_5 K$ obtemos que

$$L^{-1} \leq \frac{\mu(B(x, n, \varepsilon))}{\exp(S_n \varphi(x) - nP)} \leq L, \quad (12.1.14)$$

para todo $x \in M$, $n \geq 1$ e $\varepsilon > 0$. Lembre que $P = \log \lambda$.

Lema 12.1.12. *A probabilidade μ é invariante por f . Além disso, f admite jacobiano relativamente a μ , dado por $J_\mu f = \lambda e^{-\varphi}(h \circ f)/h$.*

Demonstração. Comece por notar que $\mathcal{L}((g_1 \circ f)g_2) = g_1 \mathcal{L}g_2$, quaisquer que sejam as funções contínuas $g_1, g_2 : M \rightarrow \mathbb{R}$: para todo $y \in M$,

$$\begin{aligned}\mathcal{L}((g_1 \circ f)g_2)(y) &= \sum_{x \in f^{-1}(y)} e^{\varphi(x)} g_1(f(x)) g_2(x) \\ &= g_1(y) \sum_{x \in f^{-1}(y)} e^{\varphi(x)} g_2(x) = g_1(y) \mathcal{L}g_2(y).\end{aligned} \quad (12.1.15)$$

Então, para toda função contínua $g : M \rightarrow \mathbb{R}$,

$$\begin{aligned} \int (g \circ f) d\mu &= \lambda^{-1} \int (g \circ f) h d(\mathcal{L}^* \nu) = \lambda^{-1} \int \mathcal{L}((g \circ f)h) d\nu \\ &= \lambda^{-1} \int g \mathcal{L}h d\nu = \int gh d\nu = \int g d\mu. \end{aligned}$$

Isto prova que a probabilidade μ é invariante por f (lembre da Proposição A.3.3).

Para provar a segunda afirmação, considere qualquer domínio de invertibilidade A de f . Então, usando o Exercício 9.7.3(a),

$$\mu(f(A)) = \int_{f(A)} 1 d\mu = \int_{f(A)} h d\nu = \int_A J_\nu f (h \circ f) d\nu = \int_A J_\nu f \frac{h \circ f}{h} d\mu.$$

Pelo Lema 12.1.3, isto significa que

$$J_\mu f = J_\nu f \frac{h \circ f}{h} = \lambda e^{-\varphi} \frac{h \circ f}{h},$$

tal como foi afirmado. \square

Corolário 12.1.13. *A medida invariante $\mu = h\nu$ satisfaz $h_\mu(f) + \int \varphi d\mu = P$.*

Demonstração. Combinando a fórmula de Rokhlin (Teorema 9.7.3) com a segunda parte do Lema 12.1.12,

$$h_\mu(f) = \int \log J_\mu f d\mu = \log \lambda - \int \varphi d\mu + \int (\log h \circ f - \log h) d\mu.$$

Como μ é invariante e $\log h$ é limitada (Corolário 12.1.9), a última parcela é igual a zero. Portanto, $h_\mu(f) = P - \int \varphi d\mu$ conforme enunciado. \square

Para completar a prova de que $\mu = h\nu$ é um estado de equilíbrio resta verificar que $P = \log \lambda$ é igual à pressão $P(f, \varphi)$. Isso será feito abaixo, no Corolário 12.1.15.

12.1.5 Pressão e autovalores

Seja η qualquer probabilidade invariante de f satisfazendo

$$h_\eta(f) + \int \varphi d\eta \geq P \quad (12.1.16)$$

(por exemplo: a medida μ construída na seção anterior). Seja $g_\eta = 1/J_\eta f$ e considere também a função $g = \lambda^{-1} e^\varphi h / (h \circ f)$. Observe que

$$\sum_{x \in f^{-1}(y)} g(x) = \frac{1}{\lambda h(y)} \sum_{x \in f^{-1}(y)} e^{\varphi(x)} h(x) = \frac{\mathcal{L}h(y)}{\lambda h(y)} = 1 \quad (12.1.17)$$

para todo $y \in M$. Além disso, como η é invariante por f (Exercício 9.7.5),

$$\sum_{x \in f^{-1}(y)} g_\eta(x) = 1 \quad \text{para } \eta\text{-quase todo } y \in M. \quad (12.1.18)$$

Usando (12.1.16) e a fórmula de Rokhlin (Teorema 9.7.3),

$$0 \leq h_\eta(f) + \int \varphi d\eta - P = \int (-\log g_\eta + \varphi - \log \lambda) d\eta. \quad (12.1.19)$$

Pela definição de g e a hipótese de que η é invariante, a integral do lado direito de (12.1.19) é igual a

$$\int (-\log g_\eta + \log g + \log h \circ f - \log h) d\eta = \int \log \frac{g}{g_\eta} d\eta. \quad (12.1.20)$$

Lembrando a definição de g_η , o Exercício 9.7.4 dá que

$$\int \log \frac{g}{g_\eta} d\eta = \int \left(\sum_{x \in f^{-1}(y)} g_\eta(x) \log \frac{g}{g_\eta}(x) \right) d\eta(y) \quad (12.1.21)$$

A esta altura precisamos do seguinte fato elementar:

Lema 12.1.14. *Sejam $p_i, b_i, i = 1, \dots, k$ números reais positivos tais que $\sum_{i=1}^k p_i = 1$. Então $\sum_{i=1}^k p_i \log b_i \leq \log(\sum_{i=1}^k p_i b_i)$ e a igualdade acontece se, e somente se, os números b_j forem todos iguais a $\sum_{i=1}^k p_i b_i$.*

Demonstração. Tome $a_i = \log(p_i b_i)$ no Lema 10.4.4. Então a desigualdade no Lema 10.4.4 corresponde exatamente à desigualdade no presente lema. Além disso, vale a igualdade se, e somente se,

$$p_j = \frac{e^{a_j}}{\sum_i e^{a_i}} \Leftrightarrow p_j = \frac{p_j b_j}{\sum_i p_i b_i} \Leftrightarrow b_j = \sum_i p_i b_i$$

para todo $j = 1, \dots, n$. □

Para cada $y \in M$, tome $p_i = g_\eta(x_i)$ e $b_i = \log(g(x_i)/g_\eta(x_i))$, onde os x_i são as pré-imagens de y . A igualdade (12.1.18) significa que $\sum_i p_i = 1$ para η -quase todo y . Então podemos aplicar o Lema 12.1.14:

$$\begin{aligned} \sum_{x \in f^{-1}(y)} g_\eta(x) \log \frac{g}{g_\eta}(x) &\leq \log \sum_{x \in f^{-1}(y)} g_\eta(x) \frac{g}{g_\eta}(x) \\ &= \log \sum_{x \in f^{-1}(y)} g(x) = 0 \end{aligned} \quad (12.1.22)$$

para η -quase todo y ; na última igualdade usamos (12.1.17). Combinando as relações (12.1.19)–(12.1.22) obtemos:

$$h_\eta(f) + \int \varphi d\eta - P = \int \log \frac{g}{g_\eta} d\eta = 0. \quad (12.1.23)$$

Corolário 12.1.15. $P(f, \varphi) = P = \log r(\mathcal{L})$.

Demonstração. Por (12.1.23), temos $h_\eta(f) + \int \varphi d\eta = P$ para toda probabilidade invariante η tal que $h_\eta(f) + \int \varphi d\eta \geq P$. Pelo princípio variacional (Teorema 10.4.1), segue que $P(f, \varphi) = P$. A segunda igualdade já havia sido observada ao final do Corolário 12.1.9. \square

A esta altura completamos a demonstração de que a medida $\mu = h\nu$ construída na seção anterior é um estado de equilíbrio para φ . O enunciado a seguir resulta do mesmo tipo de ideias e será a base para provarmos que esse estado de equilíbrio é único:

Corolário 12.1.16. *Se η é estado de equilíbrio para φ então $\text{supp } \eta = M$ e*

$$J_\eta f = \lambda e^{-\varphi}(h \circ f)/h \quad \text{e} \quad \mathcal{L}^*(\eta/h) = \lambda(\eta/h).$$

Demonstração. A primeira afirmação no enunciado é uma consequência imediata da segunda e do Lema 12.1.4.

A igualdade (12.1.23) também implica que vale a igualdade em (12.1.22) para quase todo $y \in M$. De acordo com o Lema 12.1.14, isso acontece se, e somente se, os números $b_i = \log(g(x_i)/g_\eta(x_i))$ são todos iguais. Em outras palavras, para η -quase todo $y \in M$ existe um número $c(y)$ tal que

$$\frac{g(x)}{g_\eta(x)} = c(y) \quad \text{para todo } x \in f^{-1}(y).$$

Além disso, lembrando as igualdades (12.1.19) e (12.1.20),

$$c(y) = \sum_{x \in f^{-1}(y)} c(y)g_\eta(x) = \sum_{x \in f^{-1}(y)} g(x) = 1$$

para η -quase todo y . Segue que $g_\eta = g$ em η quase todo ponto, ou seja, a função $1/g = \lambda e^{-\varphi}(h \circ f)/h$ é um jacobiano de f relativamente a η . Isto prova a segunda afirmação.

Para provar a terceira afirmação, seja $\xi : M \rightarrow \mathbb{R}$ uma função contínua qualquer. Por um lado, usando a definição do operador de transferência

$$\int \xi d\mathcal{L}^*\left(\frac{\eta}{h}\right) = \int \frac{1}{h}(\mathcal{L}\xi) d\eta = \int \frac{1}{h(y)} \left(\sum_{x \in f^{-1}(y)} e^{\varphi(x)} \xi(x) \right) d\eta(y). \quad (12.1.24)$$

Pela definição da função g ,

$$\frac{e^{\varphi(x)}}{h(y)} = \frac{\lambda g(x)}{h(x)}.$$

Substituindo esta igualdade em (12.1.24), obtemos

$$\int \xi d\mathcal{L}^*\left(\frac{\eta}{h}\right) = \int \left(\sum_{x \in f^{-1}(y)} \frac{\lambda g \xi}{h}(x) \right) d\eta(y). \quad (12.1.25)$$

Então, lembrando que $g = g_\eta = 1/J_\eta f$, podemos usar o Exercício 9.7.4 para concluir que

$$\int \xi d\mathcal{L}^*\left(\frac{\eta}{h}\right) = \int \left(\sum_{x \in f^{-1}(y)} \frac{\lambda g \xi}{h}(x) \right) d\eta(y) = \int \frac{\lambda \xi}{h} d\eta.$$

Como a função contínua ξ é arbitrária, isto mostra que $\mathcal{L}^*(\eta/h) = \lambda(\eta/h)$, tal como afirmamos. \square

12.1.6 Unicidade do estado de equilíbrio

Começemos por provar o seguinte controle da distorção:

Corolário 12.1.17. *Existe $K_7 > 0$ que para todo estado de equilíbrio η , todo $n \geq 1$, todo $x \in M$ e todo $y \in B(x, n+1, \rho)$,*

$$K_7^{-1} \leq \frac{J_\eta f^n(x)}{J_\eta f^n(y)} \leq K_7.$$

Demonstração. Pelo Corolário 12.1.16,

$$J_\eta f^n = \lambda e^{-S_n \varphi} \frac{h \circ f^n}{h} = J_\nu f^n \frac{h \circ f^n}{h}$$

para cada $n \geq 1$. Então, usando o Corolário 12.1.6 e o Lema 12.1.11,

$$K_2^{-1} K_5^{-4} \leq \frac{J_\eta f^n(x)}{J_\eta f^n(y)} = \frac{J_\nu f^n(x) h(f^n(x)) h(y)}{J_\nu f^n(y) f(f^n(y)) h(x)} \leq K_2 K_5^4.$$

Portanto, basta tomar $K_7 = K_2 K_5^4$. \square

Lema 12.1.18. *Todos os estados de equilíbrio de φ são equivalentes.*

Demonstração. Considere uma partição finita \mathcal{P} de M tal que todo $P \in \mathcal{P}$ tem interior não vazio e diâmetro menor que ρ . Como $\text{supp } \eta_1 = \text{supp } \eta_2 = M$ (pelo Corolário 12.1.16) temos que $\{\eta_i(P) : i = 1, 2 \text{ e } P \in \mathcal{P}\}$ é limitado de zero. Consequentemente, existe $C_1 > 0$ tal que

$$\frac{1}{C_1} \leq \frac{\eta_1(P)}{\eta_2(P)} \leq C_1 \quad \text{para todo } P \in \mathcal{P}. \quad (12.1.26)$$

Vamos mostrar que esta relação ainda vale para todo subconjunto mensurável de M , a menos de substituirmos C_1 por uma constante conveniente $C_2 > C_1$.

Para cada $n \geq 1$, seja \mathcal{Q}_n a partição de M formada pelas imagens $h^n(P)$ dos elementos de P pelos ramos inversos h^n do iterado f^n . Pela definição de jacobiano, $\eta_i(P) = \int_{h^n(P)} J_{\eta_i} f^n d\eta_i$. Logo, usando o Corolário 12.1.17,

$$K_7^{-1} J_{\eta_i} f^n(x) \leq \frac{\eta_i(P)}{\eta_i(h^n(P))} \leq K_7 J_{\eta_i} f^n(x)$$

para qualquer $x \in h^n(P)$. Lembrando que $J_{\eta_1}f = J_{\eta_2}f$ (Corolário 12.1.16), segue que

$$K_7^{-2} \leq \frac{\eta_2(P)\eta_1(h^n(P))}{\eta_1(P)\eta_2(h^n(P))} \leq K_7^2. \quad (12.1.27)$$

Combinando (12.1.26) e (12.1.27) e tomando $C_2 = C_1K_7^2$ vem que

$$\frac{1}{C_2} \leq \frac{\eta_1(h^n(P))}{\eta_2(h^n(P))} \leq C_2 \quad (12.1.28)$$

para todo $P \in \mathcal{P}$, todo ramo inverso h^n de f^n e todo $n \geq 1$. Em outras palavras, (12.1.26) vale para todo elemento de \mathcal{Q}_n , com C_2 no lugar de C_1 .

Agora observe que $\text{diam } \mathcal{Q}_n < \sigma^{-n}\rho$ para todo n . Dado qualquer conjunto mensurável B e dado qualquer $\delta > 0$, podemos usar a Proposição A.3.2 para encontrar um compacto $F \subset B$ e um aberto $A \supset B$ tais que $\eta_i(A \setminus F) < \delta$ para $i = 1$ e para $i = 2$. Seja Q_n a união de todos os elementos da partição \mathcal{Q}_n que intersectam F . É claro que $Q_n \supset F$ e, supondo que n é suficientemente grande, $Q_n \subset A$. Então,

$$\eta_1(B) \leq \eta_1(A) < \eta_1(Q_n) + \delta \quad \text{e} \quad \eta_2(B) \geq \eta_2(F) > \eta_2(Q_n) - \delta.$$

A relação (12.1.28) dá que $\eta_1(Q_n) \leq C_2\eta_2(Q_n)$, uma vez que Q_n é uma união (disjunta) de elementos de \mathcal{Q}_n . Combinando estas três desigualdades, obtemos

$$\eta_1(B) < C_2(\eta_2(B) + \delta) + \delta.$$

Como δ é arbitrário, concluímos que $\eta_1(B) \leq C_2\eta_2(B)$ para todo conjunto mensurável $B \subset M$. Permutando os papeéis das duas medidas também obtemos que $\eta_2(B) \leq C_2\eta_1(B)$ para todo conjunto mensurável $B \subset M$.

Estas desigualdades mostram que quaisquer dois estados de equilíbrio são equivalentes, com derivadas de Radon-Nikodym afastada de zero e infinito. \square

Combinando os Lemas 4.3.3 e 12.1.18 obtemos que todos os estados de equilíbrio *ergódicos* são iguais. Por outro lado, como vimos na Proposição 10.5.5, as componentes ergódicas de um estado de equilíbrio são estados de equilíbrio (ergódicos, claro). Segue que existe um único estado de equilíbrio, tal como afirmamos.

Como consequência, a medida de referência ν também é única: se existissem duas medidas de referência distintas ν_1 e ν_2 então $\mu_1 = h\nu_1$ e $\mu_2 = h\nu_2$ seriam estados de equilíbrio distintos. Analogamente, a autofunção positiva h é única a menos de produto por constante positiva.

Podemos dar outra demonstração da unicidade que dispensa o uso da Proposição 10.5.5 e, portanto, do teorema de Jacobs. De fato, segue dos resultados na próxima seção que o estado de equilíbrio μ na Seção 12.1.4 é ergódico. Pelo Lema 12.1.18 isso implica que todos os estados de equilíbrio são ergódicos. Em vista do Lema 4.3.3, segue que eles são todos iguais.

12.1.7 Exatidão

Finalmente, vamos provar que o sistema (f, μ) é exato. Lembre que isto quer dizer que se $B \subset M$ é tal que existem conjuntos mensuráveis B_n satisfazendo $B = f^{-n}(B_n)$ para todo $n \geq 1$, então B tem medida 0 ou 1.

Seja B um conjunto nessas condições e suponha que $\mu(B) > 0$. Seja \mathcal{P} uma partição finita de M por conjuntos com interior não vazio e diâmetro menor que ρ . Para cada n , seja \mathcal{Q}_n a partição de M cujos elementos são as imagens $h^n(P)$ dos conjuntos $P \in \mathcal{P}$ pelos ramos inversos h^n do iterado f^n .

Lema 12.1.19. *Para todo $\varepsilon > 0$ e todo $n \geq 1$ suficientemente grande existe algum $h^n(P) \in \mathcal{Q}_n$ tal que*

$$\mu(B \cap h^n(P)) > (1 - \varepsilon)\mu(h^n(P)). \quad (12.1.29)$$

Demonstração. Fixe $\varepsilon > 0$. Como a medida μ é regular (Propriedade A.3.2), dado qualquer $\delta > 0$ existe algum compacto $F \subset B$ e algum aberto $A \supset B$ satisfazendo $\mu(A \setminus F) < \delta$. Como supomos que $\mu(B) > 0$, esta desigualdade implica que $\mu(F) > (1 - \varepsilon)\mu(A)$, desde que δ seja suficientemente pequeno. Fixe $\delta > 0$ nessas condições. Note que $\text{diam } \mathcal{Q}_n < \sigma^{-n}\rho$. Então, para todo n suficientemente grande, qualquer elemento $h^n(P)$ de \mathcal{Q}_n que intersecta F está contido em A . Suponha que (12.1.29) fosse falsa para todo $h^n(P)$. Então, somando sobre todos $h^n(P)$ que intersectam F ,

$$\begin{aligned} \mu(F) &\leq \sum_{P, h^n} \mu(F \cap h^n(P)) \leq \sum_{P, h^n} \mu(B \cap h^n(P)) \\ &\leq (1 - \varepsilon) \sum_{P, h^n} \mu(h^n(P)) \leq (1 - \varepsilon)\mu(A). \end{aligned}$$

Esta contradição prova que (12.1.29) é válida para algum $h^n(P) \in \mathcal{Q}_n$. \square

Considere qualquer $h^n(P) \in \mathcal{Q}_n$ satisfazendo (12.1.29). Como $B = f^{-n}(B_n)$ e $f^n \circ h^n = \text{id}$ no seu domínio, temos que $f^n(h^n(P) \setminus B) = P \setminus B_n$. Então, aplicando o Corolário 12.1.17 à medida $\eta = \mu$,

$$\begin{aligned} \mu(P \setminus B_n) &= \int_{h^n(P) \setminus B} J_\mu f^n d\mu \leq K_7 \mu(h^n(P) \setminus B) J_\mu f^n(x) \\ \text{e } \mu(P) &= \int_{h^n(P)} J_\mu f^n d\mu \geq K_7^{-1} \mu(h^n(P)) J_\mu f^n(x) \end{aligned} \quad (12.1.30)$$

para qualquer $x \in h^n(P)$. Combinando (12.1.29) e (12.1.30),

$$\frac{\mu(P \setminus B_n)}{\mu(P)} \leq K_7^2 \frac{\mu(h^n(P) \setminus B)}{\mu(h^n(P))} \leq K_7^2 \varepsilon.$$

Resumindo, mostramos que, dado qualquer $\varepsilon > 0$ e qualquer $n \geq 1$ suficientemente grande existe algum $P \in \mathcal{P}$ tal que $\mu(P \setminus B_n) \leq K_7^2 \varepsilon \mu(P)$.

Como a partição \mathcal{P} é finita, segue que existe algum $P \in \mathcal{P}$ e alguma sequência $(n_j)_j \rightarrow \infty$ tal que

$$\mu(P \setminus B_{n_j}) \rightarrow 0 \quad \text{quando } j \rightarrow \infty. \quad (12.1.31)$$

Fixemos P a partir daqui. Como P tem interior não vazio e f é topologicamente exata, por hipótese, existe $N \geq 1$ tal que $f^N(P) = M$. Seja $P = P_1 \cup \dots \cup P_s$ uma partição finita de P em domínios de invertibilidade de f^N . Os Corolários 12.1.9 e 12.1.16 dão que $J_\mu f^N = \lambda^N e^{-S_N \varphi}(h \circ f^N)/f$ é uma função limitada de zero e infinito. Note também que $f^N(P_i \setminus B_{n_j}) = f^N(P_i) \setminus B_{n_j+N}$, uma vez que $f^{-n}(B_n) = B$ para todo n . Combinando estas duas observações com (12.1.31), obtemos que, dado qualquer $i = 1, \dots, s$,

$$\mu(f^N(P_i) \setminus B_{n_j+N}) = \mu(f^N(P_i \setminus B_{n_j})) = \int_{P_i \setminus B_{n_j}} \log J_\mu f^N d\mu$$

converge para zero quando $j \rightarrow \infty$. Agora, $\{f^N(P_i) : i = 1, \dots, s\}$ é uma cobertura finita de M por conjuntos mensuráveis. Portanto, esta última conclusão implica que $\mu(M \setminus B_{n_j+N})$ converge para zero, ou seja, que $\mu(B) = \mu(B_{n_j+N})$ converge para 1 quando $j \rightarrow \infty$. Isto significa que $\mu(B) = 1$, claro.

A demonstração do Teorema 12.1 está completa.

12.1.8 Medidas absolutamente contínuas

Nesta última seção sobre o teorema de Ruelle vamos discutir brevemente o caso particular em que $f : M \rightarrow M$ é um difeomorfismo local numa variedade compacta e $\varphi = -\log |\det Df|$. Suporemos sempre que f é tal que o potencial φ é Hölder. O primeiro objetivo é comparar as conclusões do teorema de Ruelle neste caso com os resultados da Seção 11.1:

Proposição 12.1.20. *A probabilidade absolutamente contínua invariante de f coincide com o estado de equilíbrio μ do potencial $\varphi = -\log |\det Df|$. Consequentemente, ela é equivalente à medida de Lebesgue m , com densidade $d\mu/dm$ Hölder e limitada de zero e infinito, e ela é exata.*

Demonstração. Vimos no Exemplo 12.1.2 que a medida de Lebesgue m é autovetor do dual \mathcal{L}^* do operador de transferência correspondente ao potencial $\varphi = -\log |\det Df|$: mais precisamente,

$$\mathcal{L}^* m = m.$$

Aplicando a teoria anterior (do Lema 12.1.3 em diante) com $\lambda = 1$ e $\nu = m$, encontramos uma função Hölder $h : M \rightarrow \mathbb{R}$, limitada de zero e infinito, tal que $\mathcal{L}h = h$ e a medida $\mu = hm$ é o estado de equilíbrio do potencial φ . Lembrando do Corolário 11.1.15, segue que μ também é a única probabilidade invariante de f absolutamente contínua com relação a m . O fato de que h é positiva implica que μ e m são equivalentes. A exatidão foi provada na Seção 12.1.7. \square

Também segue da demonstração do Teorema 12.1 que

$$h_\mu(f) - \int \log |\det Df| d\mu = P(f, -\log |\det Df|) = \log \lambda = 0. \quad (12.1.32)$$

Seja $\tilde{\varphi}$ a média temporal da função φ , dada pelo teorema ergódico de Birkhoff. Então,

$$\int \log |\det Df| d\mu = \int -\varphi d\mu = \int -\tilde{\varphi} d\mu. \quad (12.1.33)$$

Além disso,

$$-\tilde{\varphi}(x) = \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \log |\det Df(f^j(x))| = \lim_n \frac{1}{n} \log |\det Df^n(x)| \quad (12.1.34)$$

em μ -quase todo ponto. No contexto dos nossos comentários sobre o teorema de Oseledets (veja a relação (c1) na Seção 3.3.5) mencionamos que

$$\lim_n \frac{1}{n} \log |\det Df^n(x)| = \sum_{i=1}^{k(x)} d_i(x) \lambda_i(x), \quad (12.1.35)$$

onde $\lambda_1(x), \dots, \lambda_{k(x)}(x)$ são os expoentes de Lyapunov da transformação f no ponto x e $d_1(x), \dots, d_{k(x)}(x)$ são as respectivas multiplicidades. Combinando as relações (12.1.32)-(12.1.35), obtemos que

$$h_\mu(f) = \int \left(\sum_{i=1}^{k(x)} d_i(x) \lambda_i(x) \right) d\mu(x). \quad (12.1.36)$$

Como estas funções são invariantes (veja a relação (a1) na Seção 3.3.5) e a medida μ é ergódica, temos que $k(x)$, $\lambda_i(x)$ e $d_i(x)$ são constantes em μ -quase todo ponto. Representaremos por k , λ_i e d_i os seus valores. Então, (12.1.36) se traduz no seguinte teorema:

Teorema 12.1.21. *Seja $f : M \rightarrow M$ uma transformação expansora numa variedade Riemanniana compacta, tal que a derivada Df é Hölder. Seja μ a única probabilidade invariante absolutamente contínua com relação à medida de Lebesgue em M . Então*

$$h_\mu(f) = \sum_{i=1}^k d_i \lambda_i, \quad (12.1.37)$$

onde λ_i , $i = 1, \dots, k$ são os expoentes de Lyapunov de f em μ -quase todo ponto e d_i , $i = 1, \dots, k$ são as respectivas multiplicidades.

Este resultado é um caso particular de um teorema bem mais geral, conhecido como *fórmula da entropia de Pesin* (veja Pesin [Pes77], Mañé [Mañ81]), segundo o qual

$$h_\mu(f) = \int \left(\sum_{i=1}^{k(x)} d_i(x) \lambda_i^+(x) \right) d\mu(x), \quad \lambda_i^+ = \max\{0, \lambda_i\}, \quad (12.1.38)$$

sempre que f é uma transformação (não necessariamente expansora) diferenciável com derivada Hölder e μ é uma probabilidade absolutamente contínua com relação à medida de Lebesgue. No caso de transformações expansoras todos os expoentes de Lyapunov são positivos e, portanto, $\lambda_i^+ = \lambda_i$ para todo i .

De fato, para que valha a fórmula da entropia de Pesin (12.1.38) é suficiente que a medida μ seja “absolutamente contínua ao longo das variedades instáveis”. Mais ainda, esta condição suficiente também é necessária. Veja Ledrappier, Young [Led84, LY85].

12.1.9 Exercícios

12.1.1. Mostre que se η é uma medida boreliana num espaço métrico compacto então para todo $\varepsilon > 0$ existe $b > 0$ tal que $\eta(B(y, \varepsilon)) > b$ para todo $y \in \text{supp } \eta$.

12.1.2. Seja $f : M \rightarrow M$ uma transformação expansora. Considere o operador não linear $G : \mathcal{M}_1(M) \rightarrow \mathcal{M}_1(M)$ definido no espaço $\mathcal{M}_1(M)$ das probabilidades borelianas em M por

$$G(\eta) = \frac{\mathcal{L}^*(\eta)}{\int \mathcal{L}1 d\eta},$$

Use o teorema de Tychonoff-Schauder (Teorema 2.2.3) para concluir que G admite algum ponto fixo. Deduza o Lema 12.1.1.

12.1.3. Seja $\sigma : \Sigma_A \rightarrow \Sigma_A$ o deslocamento unilateral de tipo finito associado a uma dada matriz de transições A (lembre da Seção 10.2.2). Seja P uma matriz estocástica tal que $P_{i,j} = 0$ sempre que $A_{i,j} = 0$. Considere o operador de transferência \mathcal{L} associado ao potencial localmente constante

$$\varphi(i_0, i_1, \dots, i_n, \dots) = -\log \frac{P_{i_1}}{P_{i_0, i_1}}.$$

Mostre que $\mathcal{L}^* \mu = \mu$, onde μ é a medida de Markov μ associada à matriz P .

12.1.4. Seja λ um número positivo qualquer e seja ν uma probabilidade boreliana satisfazendo $\mathcal{L}^* \nu = \lambda \nu$. Mostre que, dada qualquer $u \in L^1(\nu)$ e qualquer função contínua $v : M \rightarrow \mathbb{R}$,

$$\int (u \circ f)v d\nu = \int u(\lambda^{-1} \mathcal{L}v) d\nu.$$

12.2 Teorema de Livsič

Nesta seção vamos discutir a seguinte questão: quando é que os estados de equilíbrio μ_ϕ e μ_ψ de dois potenciais Hölder ϕ e ψ são iguais? Observe que, como se trata de medidas ergódicas, μ_ϕ e μ_ψ ou são iguais ou são mutuamente singulares (Lema 4.3.3).

Lembre que dois potenciais dizem-se cohomólogos relativamente a f se a sua diferença pode ser escrita na forma $u \circ f - u$ para alguma função contínua $u : M \rightarrow \mathbb{R}$.

Proposição 12.2.1. *Um potencial $\varphi : M \rightarrow \mathbb{R}$ é cohomólogo a zero se, e somente se, $S_n\varphi(x) = 0$ para todo $x \in \text{Fix}(f^n)$ e todo $n \geq 1$.*

Demonstração. É claro que se $\varphi = u \circ f - u$ para algum u então

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

para todo $x \in M$ tal que $f^n(x) = x$. A recíproca é muito mais interessante.

Suponha que $S_n\varphi(x) = 0$ para todo $x \in \text{Fix}(f^n)$ e todo $n \geq 1$. Considere qualquer ponto $z \in M$ cuja órbita é densa em M ; tal ponto existe porque f é topologicamente misturadora e, conseqüentemente, transitiva. Defina u na órbita de z por meio da relação

$$u(f^n(z)) = u(z) + S_n\varphi(z), \quad (12.2.1)$$

onde $u(z)$ é arbitrário. Observe que

$$u(f^{n+1}(z)) - u(f^n(z)) = S_{n+1}\varphi(z) - S_n\varphi(z) = \varphi(f^n(z)) \quad (12.2.2)$$

para todo $n \geq 0$. Em outras palavras, vale a condição de cohomologia

$$\phi - \psi = u \circ f - u \quad (12.2.3)$$

sobre a órbita de z . Para estender esta relação a M , usaremos o seguinte fato:

Lema 12.2.2. *A função u é uniformemente contínua na órbita de z .*

Demonstração. Dado qualquer $\varepsilon \in (0, \rho)$, tome $\delta > 0$ dado pelo lema de sombreamento (Proposição 11.2.9). Suponha que $k \geq 0$ e $l \geq 1$ são tais que $d(f^k(z), f^{k+l}(z)) < \delta$. Então a sequência periódica $(x_n)_n$ de período l dada por

$$x_0 = f^k(z), x_1 = f^{k+1}(z), \dots, x_{l-1} = f^{k+l-1}(z), x_l = f^k(z)$$

é uma δ -pseudo-órbita. Logo, pela Proposição 11.2.9, existe algum $x \in \text{Fix}(f^l)$ tal que $d(f^j(x), f^{k+j}(z)) < \varepsilon$ para todo $j \geq 0$. Como tomamos $\varepsilon < \rho$, isto também implica que $x = h_l(f^l(x))$, onde $h_l : B(f^{k+l}(z), \rho) \rightarrow M$ representa o ramo inverso de f^l que envia $f^{k+l}(z)$ em $f^k(z)$. Por (11.2.6), segue que

$$d(f^j(x), f^{k+j}(z)) \leq \sigma^{j-l} d(f^l(x), f^{k+l}(z)) \quad \text{para todo } 0 \leq j \leq l. \quad (12.2.4)$$

Pela definição (12.2.1),

$$u(f^{k+l}(z)) - u(f^k(z)) = S_{k+l}\varphi(z) - S_k\varphi(z) = S_l\varphi(f^k(z)). \quad (12.2.5)$$

Fixe constantes $C > 0$ e $\nu > 0$ tais que $|\varphi(x) - \varphi(y)| \leq Cd(x, y)^\nu$ para quaisquer $x, y \in M$. Então,

$$|S_l\varphi(f^k(z)) - S_l\varphi(x)| \leq \sum_{j=0}^{l-1} |\varphi(f^{k+j}(z)) - \varphi(f^j(x))| \leq \sum_{j=0}^{l-1} Cd(f^j(x), f^{k+j}(z))^\nu.$$

Usando (12.2.4), segue que

$$|S_l \varphi(f^k(z)) - S_l \varphi(x)| \leq \sum_{j=0}^k C \sigma^{\nu(j-l)} d(x, f^{k+l}(z))^\nu \leq C_1 \varepsilon^\nu \quad (12.2.6)$$

onde $C_1 = C \sum_{i=0}^{\infty} \sigma^{-i\nu}$. Lembre também que, por hipótese, $S_l \psi(x) = 0$. Logo, combinando (12.2.5) e (12.2.6), obtemos que $|u(f^{k+l}(z)) - u(f^k(z))| \leq C_1 \varepsilon^\nu$. Isto conclui a prova do lema. \square

Segue do Lema 12.2.2 que u admite uma única extensão ao fecho da órbita de z , ou seja, ao espaço ambiente M . Então, pela continuidade de φ e u , a relação de cohomologia (12.2.3) se estende a todo o M . Isto prova a Proposição 12.2.1. \square

Teorema 12.2.3 (Livsič). *Seja $f : M \rightarrow M$ uma transformação expansora topologicamente misturadora num espaço métrico compacto e sejam ϕ e ψ dois potenciais em M . As seguintes condições são equivalentes:*

- (a) $\mu_\phi = \mu_\psi$;
- (b) existe $c \in \mathbb{R}$ e uma função qualquer $u : M \rightarrow \mathbb{R}$ tal que $\phi - \psi = c + u \circ f - u$;
- (c) $\phi - \psi$ é cohomólogo a alguma constante $c \in \mathbb{R}$;
- (d) existe $c \in \mathbb{R}$ e uma função Hölder $u : M \rightarrow \mathbb{R}$ tal que $\phi - \psi = c + u \circ f - u$;
- (e) existe $c \in \mathbb{R}$ tal que $S_n \phi(x) - S_n \psi(x) = cn$ para todo $x \in \text{Fix}(f^n)$ e $n \geq 1$.

Além disso, as constantes c em (b), (c), (d) e (e) coincidem.

Demonstração. É claro que (d) implica (c) e (c) implica (b).

Se $\phi - \psi = c + u \circ f - u$ para alguma função u então, dado $x \in \text{Fix}(f^n)$ qualquer,

$$S_n \phi(x) - S_n \psi(x) = \sum_{j=0}^{n-1} (\phi - \psi)(f^j(x)) = \sum_{j=0}^{n-1} (c + u(f^{j+1}(x)) - u(f^j(x))).$$

Como $f^n(x) = x$, a soma dos dois últimos termos sobre todo $j = 0, \dots, n-1$ é nula. Portanto, $S_n \phi(x) - S_n \psi(x) = cn$. Isto prova que (b) implica (e).

Suponha que $S_n \phi(x) - S_n \psi(x) = cn$ para todo $x \in \text{Fix}(f^n)$ e todo $n \geq 0$. Isto significa que a função $\varphi = \phi - \psi - c$ satisfaz $S_n \varphi(x) = 0$ para todo $x \in \text{Fix}(f^n)$ e todo $n \geq 0$. Note também que φ é Hölder. Logo, pela Proposição 12.2.1, existe uma função contínua $u : M \rightarrow \mathbb{R}$ tal que $\varphi = u \circ f - u$. Em outras palavras, $\phi - \psi$ é cohomólogo a c . Isto mostra que (e) implica (c).

Segue de (10.3.5) e da Proposição 10.3.12 que se ϕ é cohomóloga a $\psi + c$ então

$$P(f, \phi) = P(f, \psi + c) = P(f, \psi) + c.$$

Por outro lado, dada qualquer probabilidade invariante ν ,

$$h_\nu(f) + \int \phi d\nu = h_\nu(f) + \int (\psi + c) d\nu = h_\nu(f) + \int \psi d\nu + c.$$

Portanto, ν é estado de equilíbrio para ϕ se, e somente se, ν é estado de equilíbrio para ψ . Isto mostra que (c) implica (a).

Se μ_ϕ e μ_ψ coincidem então, evidentemente, elas têm o mesmo jacobiano. Pelo Lema 12.1.12, isto quer dizer que

$$\lambda_\phi e^{-\phi} \frac{h_\phi \circ f}{h_\phi} = \lambda_\psi e^{-\psi} \frac{h_\psi \circ f}{h_\psi}. \quad (12.2.7)$$

Seja $c = \log \lambda_\phi - \log \lambda_\psi$ e seja $u = \log h_\phi - \log h_\psi$. Estes objetos estão bem definidos, uma vez que λ_ϕ , λ_ψ , h_ϕ e h_ψ são positivos. Além disso, como as funções h_ϕ e h_ψ são Hölder e limitadas de zero e de infinito (Corolário 12.1.9), a função u é Hölder. Finalmente, (12.2.7) pode ser reescrita na seguinte forma:

$$\phi - \psi = c + \log u \circ f - u.$$

Isto mostra que (a) implica (d). A prova do teorema está completa. \square

Temos a seguinte consequência no contexto diferenciável. Seja M uma variedade compacta e seja $f : M \rightarrow M$ uma transformação diferenciável expansora cujo jacobiano $\det Df$ é Hölder.

Corolário 12.2.4. *A medida invariante absolutamente contínua μ coincide com a medida de máxima entropia se, e somente se, existe $c \in \mathbb{R}$ tal que*

$$|\det Df^n(x)| = e^{cn} \quad \text{para todo } x \in \text{Fix}(f^n) \text{ e todo } n \geq 1.$$

Demonstração. Como vimos na Proposição 12.1.20, μ é o estado de equilíbrio do potencial $\varphi = -\log |\det Df|$. É claro que a medida de máxima entropia μ_0 é o estado de equilíbrio da função nula. Observe que

$$S_n \varphi(x) = \sum_{j=0}^{n-1} \log |\det Df(f^j(x))| = \log |\det Df^n(x)|.$$

Portanto, o Teorema 12.2.3 dá que $\mu = \mu_0$ se, e somente se, existe $c \in \mathbb{R}$ tal que $\log |\det Df^n(x)| = 0 + cn$ para todo $x \in \text{Fix}(f^n)$ e todo $n \geq 1$. \square

12.2.1 Exercícios

12.2.1. Considere o deslocamento bilateral $\sigma : \Sigma \rightarrow \Sigma$ em $\Sigma = \{1, \dots, d\}^{\mathbb{Z}}$. Mostre que para toda função Hölder $\varphi : \Sigma \rightarrow \mathbb{R}$, existe uma função Hölder $\varphi^+ : \Sigma \rightarrow \mathbb{R}$, cohomóloga a φ e tal que $\varphi^+(x) = \varphi^+(y)$ sempre que $x = (x_i)_{i \in \mathbb{Z}}$ e $y = (y_i)_{i \in \mathbb{Z}}$ são tais que $x_i = y_i$ para $i \geq 0$.

12.2.2. Prove que se $\varphi, \psi : M \rightarrow \mathbb{R}$ são funções tais que existem constantes C, L tais que $|S_n \varphi(x) - S_n \psi(x) - nC| \leq L$ para todo $x \in M$, então $P(f, \varphi) = P(f, \psi) + C$ e φ é cohomólogo a $\psi + C$.

12.2.3. Dado $k \geq 2$, seja $f : S^1 \rightarrow S^1$ a transformação dada por $f(x) = kx \pmod{\mathbb{Z}}$. Seja $g : S^1 \rightarrow S^1$ uma transformação expansora diferenciável de grau k . Mostre que f e g são topologicamente conjugadas.

12.2.4. Dado $k \geq 2$, seja $f : S^1 \rightarrow S^1$ a transformação dada por $f(x) = kx \pmod{\mathbb{Z}}$. Seja $g : S^1 \rightarrow S^1$ uma transformação expansora diferenciável de grau k , com derivada Hölder. Mostre que as seguintes condições são equivalentes

- (a) f e g são conjugadas por algum difeomorfismo;
- (b) f e g são conjugadas por algum homeomorfismo absolutamente contínuo com inverso absolutamente contínuo;
- (c) $(g^n)'(p) = k^n$ para todo $p \in \text{Fix}(f^n)$

12.3 Decaimento de correlações

Seja $f : M \rightarrow M$ uma transformação expansora topologicamente exata e seja $\varphi : M \rightarrow \mathbb{R}$ um potencial Hölder. Continuaremos representando por ν a medida de referência (Seção 12.1.1) e por μ o estado de equilíbrio (Seção 12.1.4) do potencial φ . Lembre que $\mu = h\nu$, onde a função h é limitada de zero e infinito (Corolário 12.1.9). Em particular, $L^1(\mu) = L^1(\nu)$.

Dados $b > 0$ e $\beta > 0$, diremos que uma função $g : M \rightarrow \mathbb{R}$ é (b, β) -Hölder se

$$|g(x) - g(y)| \leq bd(x, y)^\beta \quad \text{para quaisquer } x, y \in M. \quad (12.3.1)$$

Diremos que g é β -Hölder se ela é (b, β) -Hölder para algum $b > 0$. Nesse caso, representaremos por $H_\beta(g)$ a menor de tais constantes b . Além disso, fixando $\rho > 0$ como em (11.2.1), representaremos por $H_{\beta, \rho}(g)$ a menor constante b tal que a desigualdade em (12.3.1) vale para quaisquer $x, y \in M$ com $d(x, y) < \rho$.

A sequência de correlações, relativamente à medida invariante μ , de duas funções g_1 e g_2 foi definida em (7.1.1):

$$C_n(g_1, g_2) = \left| \int (g_1 \circ f^n) g_2 d\mu - \int g_1 d\mu \int g_2 d\mu \right|,$$

Também consideraremos uma noção semelhante para a medida de referência ν :

$$B_n(g_1, g_2) = \left| \int (g_1 \circ f^n) g_2 d\nu - \int g_1 d\mu \int g_2 d\nu \right|.$$

Nesta seção vamos provar que estas sequências decaem exponencialmente:

Teorema 12.3.1 (Convergência exponencial para equilíbrio). *Dado $\beta \in (0, \alpha]$ existe $\Lambda < 1$ e para toda função β -Hölder $g_2 : M \rightarrow \mathbb{C}$ existe $K_1(g_2) > 0$ tal que*

$$B_n(g_1, g_2) \leq K_1(g_2) \Lambda^n \int |g_1| d\nu \quad \text{para toda } g_1 \in L_1(\nu) \text{ e todo } n \geq 1.$$

A prova deste resultado será apresentada nas Seções 12.3.1 a 12.3.3. A demonstração fornecerá uma expressão explícita para o fator $K_1(g_2)$. Observe também que

$$B_n(g_1, g_2) = \left| \int g_1 d(f_*^n(g_2\nu)) - \int g_1 d\mu \int g_2 d\nu \right|.$$

Então, a conclusão do Teorema 12.3.1 pode ser interpretada do seguinte modo: os iterados de qualquer medida da forma $g_2\nu$ convergem exponencialmente para a medida invariante $\mu \int g_2 d\nu$.

Teorema 12.3.2 (Decaimento exponencial de correlações). *Para todo $\beta \in (0, \alpha]$ existe $\Lambda < 1$ e para toda função β -Hölder $g_2 : M \rightarrow \mathbb{C}$ existe $K_2(g_2) > 0$ tal que*

$$C_n(g_1, g_2) \leq K_2(g_2)\Lambda^n \int |g_1| d\mu \quad \text{para toda } g_1 \in L_1(\mu) \text{ e todo } n \geq 1.$$

Em particular, para todo par de funções β -Hölder g_1 e g_2 , existe $K(g_1, g_2) > 0$ tal que $C_n(g_1, g_2) \leq K(g_1, g_2)\Lambda^n$ para todo $n \geq 1$.

Demonstração. Lembre que $\mu = h\nu$ e que, de acordo com o Corolário 12.1.9, a função h é α -Hölder e satisfaz $K_1^{-1} \leq h \leq K_5$ para algum $K_5 > 0$. Logo, g_2 é β -Hölder se, e somente se, g_2h é β -Hölder (veja o Exercício 12.3.5). Além disso,

$$\begin{aligned} C_n(g_1, g_2) &= \int (g_1 \circ f^n)g_2 d\mu - \int g_1 d\mu \int g_2 d\mu \\ &= \int (g_1 \circ f^n)(g_2h) d\nu - \int g_1 d\mu \int (g_2h) d\nu = B_n(g_1, g_2h). \end{aligned}$$

Portanto, segue do Teorema 12.3.1 que

$$C_n(g_1, g_2) \leq K_1(g_2h)\Lambda^n \int |g_1| d\nu \leq K_1(g_2h)/K_5\Lambda^n \int |g_1| d\mu.$$

Isto prova a primeira parte do teorema, com $K_2(g_2) = K_1(g_2h)/K_5$. A segunda parte é uma consequência imediata: se g_1 é β -Hölder então $g_1 \in L^1(\mu)$ e então basta tomar $K(g_1, g_2) = K_2(g_2) \int |g_1| d\mu$. \square

Antes de passarmos à prova do Teorema 12.3.1, queremos fazer alguns comentários breves. A questão do decaimento de correlações já foi discutida na Seção 7.4, sob a ótica da propriedade de lacuna espectral. Aqui usaremos uma abordagem diferente. A demonstração do teorema estará baseada na noção de distância projetiva associada a um cone, a qual foi introduzida por Garret Birkhoff [Bir67]. Esta noção nos permitirá obter a convergência exponencial para o equilíbrio, (da qual segue o decaimento exponencial de correlações, como acabamos de demonstrar) sem que tenhamos que analisar o espectro do operador de transferência. Mas a propriedade da lacuna espectral vale também no presente contexto e, de fato, é possível deduzi-la a partir da abordagem que apresentaremos a seguir. Voltaremos brevemente a este tema ao final da Seção 12.3.

12.3.1 Distâncias projetivas

Seja E um espaço de Banach. Chamaremos *cone* a qualquer subconjunto convexo C de E satisfazendo:

$$tC \subset C \text{ para todo } t > 0 \quad \text{e} \quad \bar{C} \cap (-\bar{C}) = \{0\}, \quad (12.3.2)$$

onde \bar{C} representa o fecho de C (anteriormente consideramos apenas cones fechados mas agora é conveniente remover essa exigência). Dados $v_1, v_2 \in C$, defina

$$\alpha(v_1, v_2) = \sup\{t > 0 : v_2 - tv_1 \in C\} \quad \text{e} \quad \beta(v_1, v_2) = \inf\{s > 0 : sv_1 - v_2 \in C\}.$$

A Figura 12.1 ajuda a compreender o significado geométrico destes números. Por convenção, $\alpha(v_1, v_2) = 0$ se $v_2 - tv_1 \notin C$ para todo $t > 0$ e $\beta(v_1, v_2) = +\infty$ se $sv_1 - v_2 \notin C$ para todo $s > 0$.

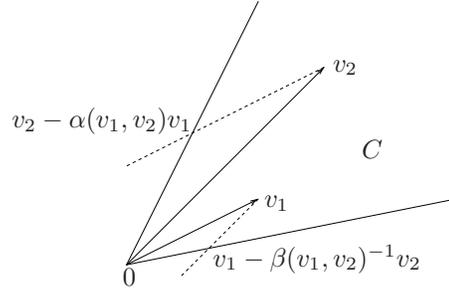


Figura 12.1: Definição da distância projetiva num cone C

Note que $\alpha(v_1, v_2)$ é sempre finito. De fato, $\alpha(v_1, v_2) = +\infty$ significaria que existe uma sequência $(t_n)_n \rightarrow +\infty$ com $v_2 - t_n v_1 \in C$ para todo n . Então, $s_n = 1/t_n$ seria uma sequência de números positivos convergindo para zero, tais que $s_n v_2 - v_1 \in C$ para todo n . Isso implicaria que $-v_1 \in \bar{C}$, contradizendo a segunda condição em (12.3.2). Um argumento análogo mostra que $\beta(v_1, v_2)$ é sempre positivo: $\beta(v_1, v_2) = 0$ implicaria $-v_2 \in \bar{C}$.

Dado qualquer cone $C \subset E$ e quaisquer $v_1, v_2 \in C \setminus \{0\}$, definimos

$$\theta(v_1, v_2) = \log \frac{\beta(v_1, v_2)}{\alpha(v_1, v_2)}, \quad (12.3.3)$$

com $\theta(v_1, v_2) = +\infty$ sempre que $\alpha(v_1, v_2) = 0$ ou $\beta(v_1, v_2) = +\infty$. As observações no parágrafo anterior asseguram que $\theta(v_1, v_2)$ está sempre bem definido. Chamamos θ de *distância projetiva* associada ao cone C . Esta terminologia é justificada pela próxima proposição, a qual mostra que θ define uma distância no quociente projetivo de $C \setminus \{0\}$, ou seja, no conjunto das classes de equivalência da relação \sim definida por $v_1 \sim v_2 \Leftrightarrow v_1 = tv_2$ para algum $t > 0$.

Proposição 12.3.3. *Se C é um cone então*

- (a) $\theta(v_1, v_2) = \theta(v_2, v_1)$ para quaisquer $v_1, v_2 \in C$;
- (b) $\theta(v_1, v_2) + \theta(v_2, v_3) \geq \theta(v_1, v_3)$ para quaisquer $v_1, v_2, v_3 \in C$;
- (c) $\theta(v_1, v_2) \geq 0$ para quaisquer $v_1, v_2 \in C$;

(d) $\theta(v_1, v_2) = 0$ se, e somente se, existe $t > 0$ tal que $v_1 = tv_2$;

(e) $\theta(t_1v_2, t_2v_2) = \theta(v_1, v_2)$ para quaisquer $v_1, v_2 \in C$ e $t_1, t_2 > 0$.

Demonstração. Se $\alpha(v_2, v_1) > 0$ então

$$\begin{aligned}\alpha(v_2, v_1) &= \sup\{t > 0 : v_1 - tv_2 \in C\} = \sup\{t > 0 : \frac{1}{t}v_1 - v_2 \in C\} \\ &= (\inf\{s > 0 : sv_1 - v_2 \in C\})^{-1} = \beta(v_1, v_2)^{-1}.\end{aligned}$$

Além disso,

$$\begin{aligned}\alpha(v_2, v_1) = 0 &\Leftrightarrow v_1 - tv_2 \notin C \text{ para todo } t > 0 \\ &\Leftrightarrow sv_1 - v_2 \notin C \text{ para todo } s > 0 \Leftrightarrow \beta(v_1, v_2) = +\infty.\end{aligned}$$

Portanto, $\alpha(v_2, v_1) = \beta(v_1, v_2)^{-1}$ em todos os casos. Permutando os papéis de v_1 e v_2 , também temos que $\beta(v_2, v_1) = \alpha(v_1, v_2)^{-1}$ para quaisquer $v_1, v_2 \in C$. A parte (a) da proposição é uma consequência imediata destas observações.

Em seguida, afirmamos que $\alpha(v_1, v_2)\alpha(v_2, v_3) \leq \alpha(v_1, v_3)$ para quaisquer $v_1, v_2, v_3 \in C$. Isto é óbvio se $\alpha(v_1, v_2) = 0$ ou $\alpha(v_2, v_3) = 0$; portanto, podemos supor que $\alpha(v_1, v_2) > 0$ e $\alpha(v_2, v_3) > 0$. Então, por definição, existem sequências crescentes de números positivos $(r_n)_n \rightarrow \alpha(v_1, v_2)$ e $(s_n)_n \rightarrow \alpha(v_2, v_3)$ tais que

$$v_2 - r_nv_1 \in C \quad \text{e} \quad v_3 - s_nv_2 \in C \quad \text{para todo } n \geq 1.$$

Como C é convexo, segue que $v_3 - s_nr_nv_1 \in C$ e, portanto, $s_nr_n \leq \alpha(v_1, v_3)$, para todo $n \geq 1$. Passando ao limite quando $n \rightarrow +\infty$, obtemos a afirmação. Um argumento análogo mostra que $\beta(v_1, v_2)\beta(v_2, v_3) \geq \beta(v_1, v_3)$ para quaisquer $v_1, v_2, v_3 \in C$. A parte (b) da proposição segue imediatamente destas desigualdades.

A parte (c) significa, simplesmente, que $\alpha(v_1, v_2) \leq \beta(v_1, v_2)$ para quaisquer $v_1, v_2 \in C$. Para provar isto, considere $t > 0$ e $s > 0$ tais que $v_2 - tv_1 \in C$ e $sv_1 - v_2 \in C$. Então, por convexidade, $(s-t)v_1 \in C$. Se $s-t$ fosse negativo, teríamos que $-v_1 \in C$, o que estaria em contradição com a última parte de (12.3.2). Portanto, $s \geq t$ para quaisquer t e s nas condições acima. Isto implica que $\alpha(v_1, v_2) \leq \beta(v_1, v_2)$.

Sejam $v_1, v_2 \in C$ tais que $\theta(v_1, v_2) = 0$. Então $\alpha(v_1, v_2) = \beta(v_1, v_2) = \gamma$ para algum $\gamma \in (0, +\infty)$. Logo, existe uma sequência crescente $(t_n)_n \rightarrow \gamma$ e uma sequência decrescente $(s_n)_n \rightarrow \gamma$ com

$$v_2 - t_nv_1 \in C \quad \text{e} \quad s_nv_1 - v_2 \in C \quad \text{para todo } n \geq 1.$$

Escrevendo $v_2 - t_nv_1 = (v_2 - \gamma v_1) + (\gamma - t_n)v_1$, concluímos que $v_2 - \gamma v_1$ está no fecho \bar{C} de C . Analogamente, $\gamma v_1 - v_2 \in \bar{C}$. Pela segunda parte de (12.3.2), segue que $v_2 - \gamma v_1 = 0$. Isto prova a parte (d) da proposição.

Finalmente, considere quaisquer $t_1, t_2 > 0$ e $v_1, v_2 \in C$. Por definição,

$$\alpha(t_1v_1, t_2v_2) = \frac{t_2}{t_1}\alpha(v_1, v_2) \quad \text{e} \quad \beta(t_1v_1, t_2v_2) = \frac{t_2}{t_1}\beta(v_1, v_2).$$

Logo, $\theta(t_1v_1, t_2v_2) = \theta(v_1, v_2)$, como afirmado na parte (e) da proposição. \square

Exemplo 12.3.4. Considere o cone $C = \{(x, y) \in E : y > |x|\}$ em $E = \mathbb{R}^2$. O quociente projetivo de C pode ser identificado com o intervalo $(-1, 1)$, por meio de $(x, 1) \mapsto x$. Dados $-1 < x_1 \leq x_2 < 1$, temos que

$$\begin{aligned}\alpha((x_1, 1), (x_2, 1)) &= \sup\{t > 0 : (x_2, 1) - t(x_1, 1) \in C\} \\ &= \sup\{t > 0 : 1 - t \geq |x_2 - tx_1|\} = \frac{1 - x_2}{1 - x_1} \\ \text{e } \beta((x_1, 1), (x_2, 1)) &= \frac{x_2 + 1}{x_1 + 1}.\end{aligned}$$

Portanto,

$$\theta((x_1, 1), (x_2, 1)) = \log R(-1, x_1, x_2, 1), \quad (12.3.4)$$

onde

$$R(a, b, c, d) = \frac{(c - a)(d - b)}{(b - a)(d - c)}$$

representa a *razão cruzada* de quatro números reais $a < b \leq c < d$.

No Exercício 12.3.2 propomos ao leitor verificar um fato semelhante onde o intervalo é substituído pelo disco unitário $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$.

Exemplo 12.3.5. Seja $E = C^0(M)$ o espaço das funções contínuas num espaço métrico compacto M . Considere o cone $C_+ = \{g \in E : g(x) > 0 \text{ para } x \in M\}$. Para quaisquer $g_1, g_2 \in C_+$,

$$\begin{aligned}\alpha(g_1, g_2) &= \sup\{t > 0 : (g_2 - tg_1)(x) > 0 \text{ para todo } x \in M\} \\ &= \inf\left\{\frac{g_2}{g_1}(x) : x \in M\right\} \\ \text{e } \beta(g_1, g_2) &= \sup\left\{\frac{g_2}{g_1}(x) : x \in M\right\}.\end{aligned}$$

Portanto,

$$\theta(g_1, g_2) = \log \frac{\sup(g_2/g_1)}{\inf(g_2/g_1)} = \log \sup\left\{\frac{g_2(x)g_1(y)}{g_1(x)g_2(y)} : x, y \in M\right\} \quad (12.3.5)$$

Esta distância projetiva é completa (Exercício 12.3.3) mas existem cones cujas distâncias projetivas não são completas (Exercício 12.3.4).

Agora, vamos observar que a distância projetiva varia de forma monótona com o cone. De fato, sejam C_1 e C_2 dois cones com $C_1 \subset C_2$ e sejam $\alpha_i(\cdot, \cdot)$, $\beta_i(\cdot, \cdot)$, $\theta_i(\cdot, \cdot)$, $i = 1, 2$ as respectivas funções, tal como definidas anteriormente. É claro das definições que, dados quaisquer $v_1, v_2 \in C_2$,

$$\alpha_1(v_1, v_2) \leq \alpha_2(v_1, v_2) \quad \text{e} \quad \beta_1(v_1, v_2) \geq \beta_2(v_1, v_2)$$

e, conseqüentemente, $\theta_1(v_1, v_2) \geq \theta_2(v_1, v_2)$.

Mais geralmente, sejam C_1 e C_2 cones em espaços de Banach E_1 e E_2 , respectivamente, e seja $L : E_1 \rightarrow E_2$ um operador linear tal que $L(C_1) \subset C_2$. Então

$$\begin{aligned}\alpha_1(v_1, v_2) &= \sup\{t > 0 : v_2 - tv_1 \in C_1\} \\ &\leq \sup\{t > 0 : L(v_2 - tv_1) \in C_2\} \\ &= \sup\{t > 0 : L(v_2) - tL(v_1) \in C_2\} = \alpha_2(L(v_1), L(v_2))\end{aligned}$$

e, analogamente, $\beta_1(v_1, v_2) \geq \beta_2(L(v_1), L(v_2))$. Consequentemente,

$$\theta_2(L(v_1), L(v_2)) \leq \theta_1(v_1, v_2) \quad \text{para quaisquer } v_1, v_2 \in C_1. \quad (12.3.6)$$

Em geral, a desigualdade (12.3.6) não é estrita. No entanto, de acordo com a próxima proposição, isso acontece sempre que $L(C_1)$ tem diâmetro finito em C_2 , relativamente à distância projetiva; de fato, nesse caso L é uma contração relativamente às distâncias projetivas θ_1 e θ_2 . Lembre que a função tangente hiperbólica \tanh é definida por

$$\tanh x = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad \text{para todo } x \in \mathbb{R}.$$

Note que a função \tanh toma valores no intervalo $(0, 1)$.

Proposição 12.3.6. *Sejam C_1 e C_2 cones em espaços de Banach E_1 e E_2 , respectivamente, e seja $L : E_1 \rightarrow E_2$ um operador linear tal que $L(C_1) \subset C_2$. Suponha que $D = \sup\{\theta_2(L(v_1), L(v_2)) : v_1, v_2 \in C_1\}$ é finito. Então,*

$$\theta_2(L(v_1), L(v_2)) \leq \tanh\left(\frac{D}{4}\right)\theta_1(v_1, v_2) \quad \text{para quaisquer } v_1, v_2 \in C_1.$$

Demonstração. Sejam $v_1, v_2 \in C_1$. Não é restrição supor que $\alpha_1(v_1, v_2) > 0$ e $\beta_1(v_1, v_2) < +\infty$ pois, caso contrário, $\theta_1(v_1, v_2) = +\infty$ e não há nada a provar. Então existem uma sequência crescente $(t_n)_n \rightarrow \alpha_1(v_1, v_2)$ e uma sequência decrescente $(s_n)_n \rightarrow \beta_1(v_1, v_2)$, tais que

$$v_2 - t_n v_1 \in C_1 \quad \text{e} \quad s_n v_1 - v_2 \in C_1.$$

Em particular, $\theta_2(L(v_2 - t_n v_1), L(s_n v_1 - v_2)) \leq D$ para todo $n \geq 1$. Fixe qualquer $D_0 > D$. Então podemos escolher números positivos T_n e S_n tais que

$$\begin{aligned}L(s_n v_1 - v_2) - T_n L(v_2 - t_n v_1) &\in C_2, \\ S_n L(v_2 - t_n v_1) - L(s_n v_1 - v_2) &\in C_2,\end{aligned} \quad (12.3.7)$$

e $\log(S_n/T_n) \leq D_0$ para todo $n \geq 1$. A primeira parte de (12.3.7) dá que

$$(s_n + t_n T_n)L(v_1) - (1 + T_n)L(v_2) \in C_2$$

e, por definição de $\beta_2(\cdot, \cdot)$, isto implica

$$\beta_2(L(v_1), L(v_2)) \leq \frac{s_n + t_n T_n}{1 + T_n}.$$

Analogamente, a segunda parte de (12.3.7) implica que

$$\alpha_2(L(v_1), L(v_2)) \geq \frac{s_n + t_n S_n}{1 + S_n}.$$

Portanto, $\theta_2(L(v_1), L(v_2))$ não pode exceder

$$\log \left(\frac{s_n + t_n T_n}{1 + T_n} \cdot \frac{1 + S_n}{s_n + t_n S_n} \right) = \log \left(\frac{s_n/t_n + T_n}{1 + T_n} \cdot \frac{1 + S_n}{s_n/t_n + S_n} \right).$$

O último termo pode ser reescrito como

$$\begin{aligned} \log \left(\frac{s_n}{t_n} + T_n \right) - \log(1 + T_n) - \log \left(\frac{s_n}{t_n} + S_n \right) + \log(1 + S_n) &= \\ &= \int_0^{\log(s_n/t_n)} \left(\frac{e^x dx}{e^x + T_n} - \frac{e^x dx}{e^x + S_n} \right), \end{aligned}$$

e esta expressão é menor ou igual que

$$\sup_{x>0} \frac{e^x(S_n - T_n)}{(e^x + T_n)(e^x + S_n)} \log \left(\frac{s_n}{t_n} \right).$$

Agora usaremos os seguintes fatos elementares:

$$\sup_{y>0} \frac{y(S_n - T_n)}{(y + T_n)(y + S_n)} = \frac{1 - \sqrt{T_n/S_n}}{1 + \sqrt{T_n/S_n}} \leq \frac{1 - e^{-D_0/2}}{1 + e^{-D_0/2}} = \tanh \frac{D_0}{4}.$$

De fato, o supremo é atingido quando $y = \sqrt{S_n T_n}$ e a desigualdade é consequência de que $\log(S_n/T_n) \leq D_0$. Isto prova que

$$\theta_2(L(v_1), L(v_2)) \leq \tanh \left(\frac{D_0}{4} \right) \log \left(\frac{s_n}{t_n} \right).$$

Note também que $\theta(v_1, v_2) = \lim_n \log(s_n/t_n)$, devido à nossa escolha de s_n e t_n . Logo, passando ao limite quando $n \rightarrow \infty$ e depois fazendo $D_0 \rightarrow D$, obtemos a conclusão da proposição. \square

Exemplo 12.3.7. Seja C_+ o cone das funções contínuas positivas em M . Para cada $L > 1$, seja $C(L) = \{g \in C_+ : \sup |g| \leq L \inf |g|\}$. Então $C(L)$ tem diâmetro finito em C_+ , para todo $L > 1$. De fato, vimos no Exemplo 12.3.5 que a distância projetiva θ associada a C_+ é dada por

$$\theta(g_1, g_2) = \log \sup \left\{ \frac{g_2(x)g_1(y)}{g_1(x)g_2(y)} : x, y \in M \right\}.$$

Em particular, $\theta(g_1, g_2) \leq 2 \log L$ para quaisquer $g_1, g_2 \in C(L)$.

12.3.2 Cones de funções Hölder

Seja $f : M \rightarrow M$ uma transformação expansora topologicamente exata e sejam $\rho > 0$ e $\sigma > 1$ as constantes na definição (11.2.1). Seja $\mathcal{L} : C^0(M) \rightarrow C^0(M)$ o operador de transferência associado a um potencial Hölder $\varphi : M \rightarrow M$. Fixe constantes $K_0 > 0$ e $\alpha > 0$ tais que

$$|\varphi(x) - \varphi(y)| \leq K_0 d(x, y)^\alpha \quad \text{para quaisquer } x, y \in M.$$

Dados $b > 0$ e $\beta > 0$, vamos representar por $C(b, \beta)$ o conjunto das funções positivas $g \in C^0(M)$ cujo logaritmo é (b, β) -Hölder em bolas de raio ρ , ou seja, tais que

$$|\log g(x) - \log g(y)| \leq bd(x, y)^\beta \quad \text{sempre que } d(x, y) < \rho. \quad (12.3.8)$$

Lema 12.3.8. *Para quaisquer $b > 0$ e $\beta > 0$, o conjunto $C(b, \beta)$ é um cone em $E = C^0(M)$ e a respectiva distância projetiva é dada por*

$$\theta(g_1, g_2) = \log \frac{\beta(g_1, g_2)}{\alpha(g_1, g_2)}$$

onde $\alpha(g_1, g_2)$ é o ínfimo e $\beta(g_1, g_2)$ é o supremo do conjunto

$$\left\{ \frac{g_2}{g_1}(x), \frac{\exp(bd(x, y)^\beta)g_2(x) - g_2(y)}{\exp(bd(x, y)^\beta)g_1(x) - g_1(y)} : x \neq y \text{ e } d(x, y) < \rho \right\}.$$

Demonstração. É claro que $g \in C$ implica $tg \in C$ para todo $t > 0$. Além disso, o fecho de C está contido no conjunto das funções não negativas e, portanto, $-\bar{C} \cap \bar{C}$ só contém a função nula. Para concluir que C é um cone resta verificar que ele é convexo. Considere quaisquer $g_1, g_2 \in C(b, \beta)$. A definição (12.3.8) quer dizer que

$$\exp(-bd(x, y)^\beta) \leq \frac{g_i(x)}{g_i(y)} \leq \exp(bd(x, y)^\beta)$$

para $i = 1, 2$ e quaisquer $x, y \in M$ com $d(x, y) < \rho$. Então, dados $t_1, t_2 > 0$,

$$\exp(-bd(x, y)^\beta) \leq \frac{t_1 g_1(x) + t_2 g_2(x)}{t_1 g_1(y) + t_2 g_2(y)} \leq \exp(bd(x, y)^\beta)$$

para quaisquer $x, y \in M$ com $d(x, y) < \rho$. Logo, $t_1 g_1 + t_2 g_2$ is in $C(b, \beta)$,

Passamos agora a calcular a distância projetiva. Por definição, $\alpha(g_1, g_2)$ é o supremo dos números $t > 0$ satisfazendo as seguintes três condições:

$$\begin{aligned} (g_2 - tg_1)(x) > 0 &\Leftrightarrow t < \frac{g_2(x)}{g_1(x)} \\ \frac{(g_2 - tg_1)(x)}{(g_2 - tg_1)(y)} &\leq \exp(bd(x, y)^\beta) \Leftrightarrow t \leq \frac{\exp(bd(x, y)^\beta)g_2(y) - g_2(x)}{\exp(bd(x, y)^\beta)g_1(y) - g_1(x)} \\ \frac{(g_2 - tg_1)(x)}{(g_2 - tg_1)(y)} &\geq \exp(-bd(x, y)^\beta) \Leftrightarrow t \leq \frac{\exp(bd(x, y)^\beta)g_2(x) - g_2(y)}{\exp(bd(x, y)^\beta)g_1(x) - g_1(y)} \end{aligned}$$

para quaisquer $x, y \in M$ com $x \neq y$ e $d(x, y) < \rho$. Logo, $\alpha(g_1, g_2)$ é igual a

$$\inf \left\{ \frac{g_2(x)}{g_1(x)}, \frac{\exp(bd(x, y)^\beta)g_2(x) - g_2(y)}{\exp(bd(x, y)^\beta)g_1(x) - g_1(y)} : x \neq y \text{ e } d(x, y) < \rho \right\}.$$

Analogamente, $\beta(g_1, g_2)$ é o supremo deste mesmo conjunto. \square

O fato crucial que faz a demonstração do Teorema 12.3.1 funcionar é que o operador de transferência tende a melhorar a regularidade das funções, mais precisamente, as suas constantes de Hölder. A próxima proposição é uma manifestação concreta desse fato:

Lema 12.3.9. *Para cada $\beta \in (0, \alpha]$ existe uma constante $\lambda_0 \in (0, 1)$ tal que $\mathcal{L}(C(b, \beta)) \subset C(\lambda_0 b, \beta)$ para todo b suficientemente grande (dependendo de β).*

Demonstração. Segue diretamente da expressão (12.1.2) do operador de transferência que $\mathcal{L}g$ é positiva sempre que g é positiva. Portanto, só temos que verificar a segunda condição na definição de $C(\lambda_0 b, \beta)$. Considere $y_1, y_2 \in M$ com $d(y_1, y_2) < \rho$. A expressão (12.1.3) dá que

$$\mathcal{L}g(y_i) = \sum_{j=1}^k e^{\varphi(x_{i,j})} g(x_{i,j})$$

para $i = 1, 2$, onde os pontos $x_{i,j} \in f^{-1}(y_i)$ satisfazem $d(x_{1i}, x_{2i}) \leq \sigma^{-1}d(y_1, y_2)$ para todo $1 \leq j \leq k$. Por hipótese, φ é (K_0, α) -Hölder. Como supomos que $\beta \leq \alpha$, segue que φ é (K, β) -Hölder, com $K = K_0(\text{diam } M)^{\alpha-\beta}$. Portanto, para todo $g \in C(b, \beta)$,

$$\begin{aligned} (\mathcal{L}g)(y_1) &= \sum_{i=1}^k e^{\varphi(x_{1,i})} g(x_{1,i}) = \sum_{i=1}^k e^{\varphi(x_{2,i})} g(x_{2,i}) \frac{g(x_{1,i})e^{\varphi(x_{1,i})}}{g(x_{2,i})e^{\varphi(x_{2,i})}} \\ &\leq \sum_{i=1}^k e^{\varphi(x_{2,i})} g(x_{2,i}) \exp(bd(x_{1,i}, x_{2,i})^\beta + Kd(x_{1,i}, x_{2,i})^\beta) \\ &\leq (\mathcal{L}g)(y_2) \exp((b+K)\sigma^{-\beta}d(y_1, y_2)^\beta). \end{aligned}$$

Fixe $\lambda_0 \in (\sigma^{-\beta}, 1)$. Para todo b suficientemente grande, $(b+K)\sigma^{-\beta} \leq b\lambda_0$. Então, a relação anterior dá que

$$(\mathcal{L}g)(y_1) \leq (\mathcal{L}g)(y_2) \exp(\lambda_0 b d(y_1, y_2)^\beta),$$

para quaisquer $y_1, y_2 \in M$ com $d(y_1, y_2) < \rho$. Permutando os papéis de y_1 e y_2 obtemos a outra desigualdade. \square

A seguir usaremos a família de cones $C(L)$ introduzida no Exemplo 12.3.7:

Lema 12.3.10. *Existe $N \geq 1$ e para todo $\beta > 0$ e todo $b > 0$ existe $L > 1$ satisfazendo $\mathcal{L}^N(C(b, \beta)) \subset C(L)$.*

Demonstração. Por hipótese, f é topologicamente exata. Logo, existe $N \geq 1$ tal que $f^N(B(z, \rho)) = M$ para todo $z \in M$. Fixe N de uma vez por todas. Dada $g \in C(b, \beta)$, considere qualquer ponto $z \in M$ tal que $g(z) = \sup g$. Considere $y_1, y_2 \in M$. Por um lado,

$$\mathcal{L}^N g(y_1) = \sum_{x \in f^{-N}(y_1)} e^{S_N \varphi(x)} g(x) \leq \text{grau}(f^N) e^{N \sup |\varphi|} g(z).$$

Por outro lado, pela escolha de N , existe algum $x \in B(z, \rho)$ tal que $f^N(x) = y_2$. Então,

$$\mathcal{L}^N g(y_2) \geq e^{S_N \varphi(x)} g(x) \geq e^{-N \sup |\varphi|} e^{-bd(x, z)^\beta} g(z) \geq e^{-N \sup |\varphi| - b\rho^\beta} g(z).$$

Como y_1, y_2 são arbitrários, isto prova que

$$\frac{\sup \mathcal{L}^N g}{\inf \mathcal{L}^N g} \leq \text{grau}(f^N) e^{2N \sup |\varphi| + b\rho^\beta}.$$

Agora basta tomar L igual à expressão do lado direito desta desigualdade. \square

Combinando os Lemas 12.3.9 e 12.3.10 obtemos que existe $N \geq 1$ e, dado $\beta \in (0, \alpha]$ existe $\lambda_0 \in (0, 1)$ tal que, para todo $b > 0$ suficientemente grande (dependendo de N e β) existe $L > 1$, satisfazendo

$$\mathcal{L}^N(C(b, \beta)) \subset C(\lambda_0^N b, \beta) \cap C(L). \quad (12.3.9)$$

No que segue, escreveremos $C(c, \beta, R) = C(c, \beta) \cap C(R)$ para quaisquer $c > 0$, $\beta > 0$ e $R > 1$.

Lema 12.3.11. *Para todo $c \in (0, b)$ e $R > 1$, o conjunto $C(c, \beta, R) \subset C(b, \beta)$ tem diâmetro finito relativamente à distância projetiva do cone $C(b, \beta)$.*

Demonstração. Usaremos a expressão de θ dada pelo Lema 12.3.8. Por um lado, a hipótese de que $g_1, g_2 \in C(c, \beta)$ garante que

$$\begin{aligned} \frac{\exp(bd(x, y)^\beta)g_2(x) - g_2(y)}{\exp(bd(x, y)^\beta)g_1(x) - g_1(y)} &= \frac{g_2(x)}{g_1(x)} \frac{1 - \exp(-bd(x, y)^\beta)(g_2(y)/g_2(x))}{1 - \exp(-bd(x, y)^\beta)(g_1(y)/g_1(x))} \\ &\geq \frac{g_2(x)}{g_1(x)} \frac{1 - \exp(-(b-c)d(x, y)^\beta)}{1 - \exp(-(b+c)d(x, y)^\beta)} \\ &\geq \frac{g_2(x)}{g_1(x)} \frac{1 - \exp(-(b-c)\rho^\beta)}{1 - \exp(-(b+c)\rho^\beta)} \end{aligned}$$

para quaisquer $x, y \in M$ com $d(x, y) < \rho$. Representemos por r o valor da última fração do lado direito. Então, observando que $r \in (0, 1)$,

$$\alpha(g_1, g_2) \geq \inf \left\{ \frac{g_2(x)}{g_1(x)}, r \frac{g_2(x)}{g_1(x)} : x \in M \right\} = r \inf \left\{ \frac{g_2(x)}{g_1(x)} : x \in M \right\} \geq r \frac{\inf g_2}{\sup g_1}.$$

Analogamente,

$$\beta(g_1, g_2) \leq \sup \left\{ \frac{g_2(x)}{g_1(x)}, \frac{1}{r} \frac{g_2(x)}{g_1(x)} : x \in M \right\} = \frac{1}{r} \sup \left\{ \frac{g_2(x)}{g_1(x)} : x \in M \right\} \leq \frac{1}{r} \frac{\sup g_2}{\inf g_1}.$$

Por outro lado, a hipótese de que $g_1, g_2 \in C(R)$ dá que

$$\frac{\sup g_2}{\inf g_1} \leq R^2 \frac{\inf g_2}{\sup g_1}.$$

Combinando estas três desigualdades, concluímos que $\theta(g_1, g_2) \leq \log(R^2/r^2)$ para quaisquer $g_1, g_2 \in C(c, \beta, R)$. \square

Corolário 12.3.12. *Existe $N \geq 1$ tal que para todo $\beta \in (0, \alpha]$ e todo $b > 0$ suficientemente grande existe $\Lambda_0 < 1$ tal que*

$$\theta(\mathcal{L}^N g_1, \mathcal{L}^N g_2) \leq \Lambda_0 \theta(g_1, g_2) \quad \text{para quaisquer } g_1, g_2 \in C(b, \beta).$$

Demonstração. Tome $N \geq 1$, $\lambda_0 \in (0, 1)$ e $L > 1$ como em (12.3.9) e considere

$$c = \lambda_0^N b \quad \text{e} \quad R = L. \quad (12.3.10)$$

Então $\mathcal{L}^N(C(b, \beta)) \subset C(c, \beta, R)$ e segue do Lema 12.3.11 que o diâmetro D da imagem $\mathcal{L}^N(C(b, \beta))$ relativamente à distância projetiva θ é finito. Tome $\Lambda_0 = \tanh(D/4)$. Agora a conclusão do corolário é uma aplicação imediata da Proposição 12.3.6. \square

12.3.3 Convergência exponencial

Fixe $N \geq 1$ e $\beta \in (0, \alpha]$ e $b > 0$ e $L > 1$, nas condições do Corolário 12.3.12, e considere $c > 0$ e $R > 1$ dados por (12.3.10). Continuamos representando por h a autofunção positiva (Lema 12.1.11) e por λ o raio espectral (Corolário 12.1.15) do operador de transferência \mathcal{L} . Lembre que h é α -Hölder e limitada de zero e infinito. Portanto, a menos de aumentar as constantes b e L se necessário, podemos considerar $h \in C(c, \beta, R)$.

O próximo lema segue diretamente das considerações anteriores e nos permitirá obter a estimativa no Teorema 12.3.1. Continuamos representando por $\|\cdot\|$ a norma definida em $C^0(M)$ por $\|\phi\| = \sup\{|\phi(x)| : x \in M\}$.

Lema 12.3.13. *Existe $C > 0$ e $\Lambda \in (0, 1)$ tal que*

$$\|\lambda^{-n} \mathcal{L}^n g - h \int g d\nu\| \leq C \Lambda^n \int g d\nu \quad \text{para } g \in C(c, \beta, R) \text{ e } n \geq 1.$$

Demonstração. Seja $g \in C(c, \beta, R)$. Em particular, $g > 0$ e portanto $\int g d\nu > 0$. A conclusão do lema não é afetada quando multiplicamos g por qualquer número positivo. Logo, não é restrição supor que $\int g d\nu = 1$. Então,

$$\int \lambda^{-n} \mathcal{L}^n g d\nu = \int \lambda^{-n} g d(\mathcal{L}^{*n} \nu) = \int g d\nu = 1 = \int h d\nu$$

e, portanto, $\inf(\lambda^{-n}\mathcal{L}^n g/h) \leq 1 \leq \sup(\lambda^{-n}\mathcal{L}^n g/h)$ para todo $n \geq 1$. Agora segue das expressões no Lema 12.3.8 que

$$\begin{aligned}\alpha(\lambda^{-jN}\mathcal{L}^{jN}g, h) &\leq \inf \frac{\lambda^{-jN}\mathcal{L}^{jN}g}{h} \leq 1 \\ \beta(\lambda^{-jN}\mathcal{L}^{jN}g, h) &\geq \sup \frac{\lambda^{-jN}\mathcal{L}^{jN}g}{h} \geq 1.\end{aligned}$$

Consequentemente,

$$\begin{aligned}\theta(\lambda^{-jN}\mathcal{L}^{jN}g, h) &\geq \log \beta(\lambda^{-jN}\mathcal{L}^{jN}g, h) \geq \log \sup \frac{\lambda^{-jN}\mathcal{L}^{jN}g}{h} \\ \theta(\lambda^{-jN}\mathcal{L}^{jN}g, h) &\geq -\log \alpha(\lambda^{-jN}\mathcal{L}^{jN}g, h) \geq -\log \inf \frac{\lambda^{-jN}\mathcal{L}^{jN}g}{h}\end{aligned}$$

para todo $j \geq 0$. Agora seja D o diâmetro de $C(c, \beta, R)$ relativamente à distância projetiva θ (Lema 12.3.11). Pela Proposição 12.3.3 e pelo Corolário 12.3.12,

$$\theta(\lambda^{-jN}\mathcal{L}^{jN}g, h) = \theta(\mathcal{L}^{jN}g, \mathcal{L}^{jN}h) \leq \Lambda_0^j \theta(g, h) \leq \Lambda_0^j D$$

para todo $j \geq 0$. Combinando isto com as duas desigualdades anteriores,

$$\exp(-\Lambda_0^j D) \leq \inf \frac{\lambda^{-jN}\mathcal{L}^{jN}g}{h} \leq \sup \frac{\lambda^{-jN}\mathcal{L}^{jN}g}{h} \leq \exp(\Lambda_0^j D)$$

para todo $j \geq 0$. Fixe $C_1 > 0$ tal que $|e^x - 1| \leq C_1|x|$ sempre que $|x| \leq D$. Então a relação anterior implica que

$$|\lambda^{-jN}\mathcal{L}^{jN}g(x) - h(x)| \leq h(x)C_1\Lambda_0^j D \quad \text{para todo } x \in M \text{ e } j \geq 0. \quad (12.3.11)$$

Tome $C_2 = C_1 D \sup h$ e $\Lambda = \Lambda_0^{1/N}$. A desigualdade (12.3.11) significa que

$$\|\lambda^{-jN}\mathcal{L}^{jN}g - h\| \leq C_2\Lambda^{jN} \quad \text{para todo } j \geq 1.$$

Dado qualquer $n \geq 1$, escreva $n = jN + r$ com $j \geq 0$ e $0 \leq r < N$. Como o operador $\mathcal{L} : C^0(M) \rightarrow C^0(M)$ é contínuo e $\mathcal{L}h = \lambda h$,

$$\|\lambda^{-n}\mathcal{L}^n g - h\| = \|\lambda^{-r}\mathcal{L}^r(\lambda^{-jN}\mathcal{L}^{jN}g - h)\| \leq (\|\mathcal{L}\|/\lambda)^r \|\lambda^{-jN}\mathcal{L}^{jN}g - h\|.$$

Combinando as duas últimas desigualdades,

$$\|\lambda^{-n}\mathcal{L}^n g - h\| \leq (\|\mathcal{L}\|/\lambda)^r C_2 \Lambda^{n-r}.$$

Isto prova a conclusão do lema, desde que tomemos $C \geq C_2(\|\mathcal{L}\|/(\lambda\Lambda))^r$ para todo $0 \leq r < N$. \square

Agora estamos prontos para provar o Teorema 12.3.1:

Demonstração. Comece por considerar $g_2 \in C(c, \beta, R)$. Usando a igualdade no Exercício 12.1.4 e lembrando que $\mu = h\nu$,

$$\begin{aligned} B_n(g_1, g_2) &= \left| \int g_1 \left(\lambda^{-n} \mathcal{L}^n g_2 - h \int g_2 d\nu \right) d\nu \right| \\ &\leq \left\| \lambda^{-n} \mathcal{L}^n g_2 - h \int g_2 d\nu \right\| \int |g_1| d\nu. \end{aligned}$$

Portanto, usando o Lema 12.3.13,

$$B_n(g_1, g_2) \leq C\Lambda^n \int |g_1| d\nu \int g_2 d\nu. \quad (12.3.12)$$

Agora seja $g_2 : M \rightarrow \mathbb{R}$ uma função β -Hölder qualquer e seja $H = H_\beta(g_2)$. Escrevemos $g_1 = g_2^+ - g_2^-$, onde

$$g_2^+ = \frac{1}{2}(|g_2| + g_2) + B \quad \text{e} \quad g_2^- = \frac{1}{2}(|g_2| - g_2) + B$$

e a constante positiva B é definida por $B = \max\{H/c, \sup |g_2|/(R-1)\}$. É claro que as funções g_2^\pm são positivas: $g_2^\pm \geq B > 0$. Além disso, elas são (H, β) -Hölder:

$$|g_2^\pm(x) - g_2^\pm(y)| \leq |g_2(x) - g_2(y)| \leq Hd(x, y)^\beta,$$

para $x, y \in M$. Logo, usando o teorema da média e o fato de que $B \geq H/c$,

$$\left| \log g_2^\pm(x) - \log g_2^\pm(y) \right| \leq \frac{|g_2^\pm(x) - g_2^\pm(y)|}{B} \leq \frac{Hd(x, y)^\beta}{B} \leq cd(x, y)^\beta.$$

Além disso, como $B \geq \sup |g_2|/(R-1)$,

$$\sup g_2^\pm \leq \sup |g_2| + B \leq RB \leq R \inf g_2^\pm$$

Juntamente com a relação anterior, isto significa que $g_2^\pm \in C(c, \beta, R)$, tal como afirmamos. Então podemos aplicar (12.3.12) às duas funções:

$$B_n(g_1, g_2^\pm) \leq C\Lambda^n \int |g_1| d\nu \int g_2^\pm d\nu$$

e, conseqüentemente,

$$\begin{aligned} B_n(g_1, g_2) &\leq B_n(g_1, g_2^+) + B_n(g_1, g_2^-) \\ &\leq C\Lambda^n \int |g_1| d\nu \int (g_2^+ + g_2^-) d\nu. \end{aligned} \quad (12.3.13)$$

Além disso, pela definição de g_2^\pm ,

$$\begin{aligned} \int (g_2^+ + g_2^-) d\nu &= \int |g_2| d\nu + 2B \leq \int |g_2| d\nu + \frac{2H}{c} + \frac{2 \sup |g_2|}{R-1} \\ &\leq \frac{2}{c} H_\beta(g_2) + \frac{R+1}{R-1} \sup |g_2|. \end{aligned} \quad (12.3.14)$$

Tome $C_1 = C \max\{2/c, (R+1)/(R-1)\}$ e defina

$$K_1(g_2) = 2C_1(\sup |g_2| + H_\beta(g_2)).$$

As relações (12.3.13) e (12.3.14) dão que

$$B_n(g_1, g_2) \leq C_1 \Lambda^n \int |g_1| d\nu (H_\beta(g_2) + \sup |g_2|) \leq \frac{1}{2} K_1(g_2) \Lambda^n \int |g_1| d\nu.$$

Isto termina a prova do teorema no caso em que g_2 é uma função real.

O caso geral (complexo) segue facilmente. Note que $K_1(\Re g_2) \leq K_1(g_2)$, pois $\sup |\Re g_2| \leq \sup |g_2|$ e $H_\beta(\Re g_2) \leq H_\beta(g_2)$. Analogamente, $K_1(\Im g_2) \leq K_1(g_2)$. Portanto, os argumentos anteriores dão que

$$\begin{aligned} B_n(g_1, g_2) &\leq B_n(g_1, \Re g_2) + B_n(g_1, \Im g_2) \leq \frac{1}{2} (K_1(\Re g_2) + K_1(\Im g_2)) \Lambda^n \int |g_1| d\nu \\ &\leq K_1(g_2) \Lambda^n \int |g_1| d\nu \end{aligned}$$

Isto completa a prova do teorema de convergência para o equilíbrio. \square

Vamos encerrar esta seção com alguns comentários sobre a propriedade da lacuna espectral. Seja $C^\beta(M)$ o espaço vetorial das funções β -Hölder $g : M \rightarrow \mathbb{C}$. Deixamos ao cuidado do leitor provar os seguintes fatos (Exercício 12.3.6):

- (i) A função $\|g\|_{\beta, \rho} = \sup |g| + H_{\beta, \rho}(g)$ é uma norma completa em $C^\beta(M)$.
- (ii) $C^\beta(M)$ é invariante pelo operador de transferência: $\mathcal{L}(C^\beta(M)) \subset C^\beta(M)$.
- (iii) A restrição $\mathcal{L} : C^\beta(M) \rightarrow C^\beta(M)$ é contínua relativamente à norma $\|\cdot\|_{\beta, \rho}$.

Note que $h \in C^\beta(M)$, já que $\beta \leq \alpha$. Defina $V = \{g \in C^\beta(M) : \int g d\nu = 0\}$. Então $C^\beta(M) = V \oplus \mathbb{C}h$, pois toda função $g \in C^\beta(M)$ pode ser decomposta, de maneira única,

$$g = (g - h \int g d\nu) + h \int g d\nu$$

como soma de uma função em V com um múltiplo de h . Além disso, a soma direta $C^\beta(M) = V \oplus \mathbb{C}h$ é invariante pelo operador de transferência. De fato, se $g \in V$ então

$$g \in V \quad \Rightarrow \quad \int \mathcal{L}g d\nu = \int g d\mathcal{L}^*\nu = \lambda \int g d\nu = 0 \quad \Rightarrow \quad \mathcal{L}g \in V.$$

Segue que o espectro de \mathcal{L} no espaço de Banach $C^\beta(M)$ é a união de $\{\lambda\}$ com o espectro da restrição \mathcal{L} ao hiperplano V . No Exercício 12.3.8 propomos ao leitor mostrar que o raio espectral de $\mathcal{L}|_V$ é estritamente menor que λ . Consequentemente, $\mathcal{L} : C^\beta(M) \rightarrow C^\beta(M)$ tem a propriedade da lacuna espectral.

No livro de Viviane Baladi [Bal00] o leitor encontrará uma apresentação aprofundada da teoria espectral dos operadores de transferência e suas conexões com o decaimento de correlações, para transformações expansoras diferenciáveis ou diferenciáveis por pedaços e também para difeomorfismos uniformemente hiperbólicos.

12.3.4 Exercícios

12.3.1. Mostre que a razão cruzada $R(a, x, y, b)$ é invariante por todo automorfismo de Möbius da reta, ou seja, $R(\phi(a), \phi(b), \phi(c), \phi(d)) = R(a, b, c, d)$ para quaisquer $a < b \leq c < d$ e toda transformação da forma $\phi(x) = (\alpha x + \beta)/(\gamma x + \delta)$ com $\alpha\delta - \beta\gamma \neq 0$.

12.3.2. Considere o cone $C = \{(z, s) \in \mathbb{C} \times \mathbb{R} : s > |z|\}$. O seu quociente projetivo pode ser identificado com o disco unitário $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$, por meio de $(z, 1) \mapsto z$. Seja d a distância induzida em \mathbb{D} , mediante esta identificação, pela distância projetiva de C . Mostre que d coincide com a distância de Cayley-Klein Δ , definida por

$$\Delta(p, q) = \log \frac{|aq| |pb|}{|ap| |bq|}, \quad \text{para } p, q \in \mathbb{D},$$

onde a e b são os pontos onde a reta que passa por p e q intersecta o bordo do disco, designados de tal modo que p esteja entre a e q e q esteja entre p e b . [Observação: A distância de Cayley-Klein relaciona-se com a distância de Poincaré do disco mediante a aplicação $z \mapsto (2z)/(1 + |z|^2)$.]

12.3.3. Mostre que a distância projetiva associada ao cone C_+ apresentado no Exemplo 12.3.7 é completa, no seguinte sentido: relativamente à distância projetiva, toda sequência de Cauchy $(g_n)_n$ converge para algum elemento de C_+ . Além disso, se normalizarmos as funções (por exemplo, fixando uma probabilidade qualquer η em M e exigindo que $\int g_n d\eta = 1 = \int g d\eta$ para todo n , então $(g_n)_n$ converge uniformemente para g .

12.3.4. Seja M uma variedade compacta e seja C_1 o cone das funções positivas diferenciáveis em M . Mostre que a respectiva distância projetiva θ_1 não é completa.

12.3.5. Verifique que se g_1 e g_2 são funções β -Hölder, $\theta : M \rightarrow M$ é transformação L -Lipschitz e η é uma probabilidade em M então

- (a) $H_\beta(g_1 g_2) \leq \sup |g_1| H_\beta(g_2) + \sup |g_2| H_\beta(g_1)$;
- (b) $\int |g_1| d\eta \leq \sup |g_1| \leq \int |g_1| d\eta + H_\beta(g_1) (\text{diam } M)^\beta$;
- (c) $H_\beta(g \circ \theta) \leq L^\beta H_\beta(g)$.

O item (a) permanece verdadeiro se substituirmos H_β por $H_{\beta, \rho}$. O mesmo vale para o item (c), se suposermos $L \leq 1$.

12.3.6. Seja $C^\beta(M)$ o espaço vetorial das funções β -Hölder num espaço métrico compacto M . Prove as propriedades (i), (ii), (iii) que foram enunciadas ao final da Seção 12.3.

12.3.7. Considere $C^\beta(M)$ munido da norma $\|\cdot\|_{\beta, \rho}$. Seja $\mathcal{L} : C^\beta(M) \rightarrow C^\beta(M)$ o operador de transferência associado a um potencial α -Hölder, com $\alpha \geq \beta$. Sejam λ o raio espectral, ν a medida de referência, h a auto-função e $\mu = h\nu$ o estado de equilíbrio do potencial φ . Considere o operador de transferência $\mathcal{P} : C^\beta(M) \rightarrow C^\beta(M)$ associado ao potencial $\psi = \varphi + \log h - \log h \circ f - \log \lambda$.

- (a) Mostre que \mathcal{L} é linearmente conjugado a $\lambda\mathcal{P}$, logo $\text{esp}(\mathcal{L}) = \lambda \text{esp}(\mathcal{P})$. Além disso, $\mathcal{P}1 = 1$ e $\mathcal{P}^*\mu = \mu$.
- (b) Mostre que $\int |\mathcal{P}^n g| d\mu \leq \int |g| d\mu$ e $\sup |\mathcal{P}^n g| \leq \sup |g|$ e existem constantes $C > 0$ e $\tau < 1$ tais que $H_{\beta,\rho}(\mathcal{P}^n g) \leq \tau^n H_{\beta,\rho}(g) + C \sup |g|$ para todo $g \in C^\beta(M)$ e todo $n \geq 1$.

12.3.8. Mantemos as notações do Exercício 12.3.7. O objetivo é mostrar que o raio espectral da restrição de \mathcal{L} ao hiperplano $V = \{g \in C^\beta(M) : \int g d\nu = 0\}$ é estritamente menor que λ . Pela parte (a) do exercício anterior, basta considerar o caso $\mathcal{L} = \mathcal{P}$ (com $\lambda = 1, \nu = \mu$ e $h = 1$). Fixe b, β, R como no Corolário 12.3.12.

- (a) Mostre que existem $K > 1$ e $r > 0$ tais que, para todo $v \in V$ com $\|v\|_{\beta,\rho} \leq r$, a função $g = 1 + v$ está no cone $C(b, \beta, R)$ e satisfaz

$$K^{-1}\|v\|_{\beta,\rho} \leq \theta(1, g) \leq K\|v\|_{\beta,\rho}.$$

- (b) Use o Corolário 12.3.12 e o item anterior para encontrar $C > 0$ e $\tau < 1$ tais que $\|\mathcal{P}^n v\|_{\beta,\rho} \leq C\tau^n \|v\|_{\beta,\rho}$ para todo $v \in V$.

12.4 Dimensão de repulsores conformes

Nesta seção apresentaremos uma aplicação da teoria desenvolvida anteriormente ao cálculo da dimensão de Hausdorff de certos conjuntos invariantes por transformações expansoras, que chamamos repulsores conformes. O principal resultado (Teorema 12.4.3) contém uma fórmula para o valor da dimensão de Hausdorff do repulsor em termos da pressão de certos potenciais.

O leitor interessado em aprofundar o estudo da dimensão de Hausdorff e suas diversas aplicações encontrará apresentações detalhadas nos livros de Falconer [Fal90], Palis, Takens [PT93, Capítulo 4], Pesin [Pes97] e Bonatti, Díaz, Viana [BDV05, Capítulo 3].

12.4.1 Dimensão de Hausdorff

Seja M um espaço métrico. Nesta seção entenderemos por *cobertura* de M qualquer família finita ou enumerável de subconjuntos de M cuja união é M . O *diâmetro* de uma cobertura \mathcal{U} é o supremo dos diâmetros dos seus elementos. Para cada $d > 0$ e $\delta > 0$, definimos

$$m_d(M, \delta) = \inf \left\{ \sum_{U \in \mathcal{U}} \text{diam}(U)^d : \mathcal{U} \text{ cobertura com } \text{diam} \mathcal{U} < \delta \right\}. \quad (12.4.1)$$

Ou seja, consideramos todas as possíveis coberturas de M por subconjuntos com diâmetro menor que δ e procuramos minimizar a soma dos diâmetros elevada a d . Este número varia com δ de maneira monótona: quando δ diminui, a classe de coberturas admissíveis diminui e portanto o ínfimo só pode aumentar. Chamamos *medida de Hausdorff de M em dimensão d* ao limite

$$m_d(M) = \lim_{\varepsilon \rightarrow 0} m_d(M, \varepsilon). \quad (12.4.2)$$

Note que $m_d(M) \in [0, \infty]$. Além disso, segue diretamente da definição que

$$m_{d_1}(M, \delta) \leq \delta^{d_1 - d_2} m_{d_2}(M, \delta) \quad \text{para todo } \delta > 0 \text{ e quaisquer } d_1 > d_2 > 0.$$

Fazendo $\delta \rightarrow 0$ segue que $m_{d_1}(M) = 0$ ou $m_{d_2}(M) = \infty$ ou ambos. Portanto, existe um único valor $d(M) \in [0, \infty]$ tal que $m_d(M) = \infty$ para todo $d < d(M)$ e $m_d(M) = 0$ para todo $d > d(M)$. Chamamos $d(M)$ de *dimensão de Hausdorff* do espaço métrico M .

A medida de Hausdorff de M na dimensão $d(M)$ pode tomar qualquer valor. Em muitos exemplos interessantes, tal como aquele que vamos apresentar a seguir, ela é positiva e finita. Mas existem muitos outros casos em que ela é zero ou infinito.

Exemplo 12.4.1. Consideremos o conjunto de Cantor usual K na reta. Isto é,

$$K = \bigcap_{n=0}^{\infty} K^n$$

onde $K^0 = [0, 1]$ e todo K^n , $n \geq 1$ é obtido retirando de cada componente conexa de K^{n-1} o subintervalo aberto central de comprimento relativo $1/3$. Seja $d_0 = \log 2 / \log 3$. Vamos mostrar que $m_{d_0}(M) = 1$. Isso implica $d(M) = d_0$.

Para provar a estimativa superior considere, para cada $n \geq 0$, a cobertura \mathcal{V}^n de K cujos elementos são as interseções de K com cada uma das componentes conexas de K^n . É claro que a sequência $(\mathcal{V}^n)_n$ é crescente: $\mathcal{V}^{n-1} \prec \mathcal{V}^n$ para todo $n \geq 1$. Note que \mathcal{V}^n tem exatamente 2^n elementos, todos com diâmetro igual a 3^{-n} . Portanto,

$$\sum_{V \in \mathcal{V}^n} (\text{diam } V)^{d_0} = 2^n 3^{-nd_0} = 1 \quad (12.4.3)$$

para todo n . Como $\text{diam } \mathcal{V}^n \rightarrow 0$ quando $n \rightarrow \infty$, segue que $m_{d_0}(M) \leq 1$.

A estimativa inferior para a medida e a dimensão de Hausdorff é um pouco mais difícil, porque é necessário lidar com coberturas arbitrárias. Vamos mostrar que, dada qualquer cobertura \mathcal{U} de M ,

$$\sum_{U \in \mathcal{U}} (\text{diam } U)^{d_0} \geq 1. \quad (12.4.4)$$

Chamamos intervalo aberto de K a qualquer interseção de K com um intervalo aberto da reta. É claro que todo subconjunto de K está contido num intervalo aberto com diâmetro tão próximo quanto se queira. Logo, sempre podemos formar coberturas \mathcal{U}' cujos elementos são intervalos abertos e tais que $\sum_{U' \in \mathcal{U}'} (\text{diam } U')^{d_0}$ está tão perto de $\sum_{U \in \mathcal{U}} (\text{diam } U)^{d_0}$ quanto se queira. Logo, não é restrição supor que os elementos de \mathcal{U} são intervalos abertos. Então, pela compacidade de K , também não é restrição supor que \mathcal{U} é finita. Considere qualquer $U \in \mathcal{U}$. Como U é aberto e $\text{diam } \mathcal{V}^n \rightarrow 0$, existe $m \geq 1$ tal que para todo $n \geq m$ todo elemento de \mathcal{V}^n que intersecta U está contido em U . Além

disso, como \mathcal{U} é finita, podemos fixar o mesmo m para todos os seus elementos. Afirmamos que

$$\sum_{U \in \mathcal{U}} (\text{diam } U)^{d_0} \geq \sum_{V \in \mathcal{V}^m} (\text{diam } V)^{d_0}. \quad (12.4.5)$$

A estratégia da prova é modificar a cobertura \mathcal{U} sucessivamente de tal forma que em cada passo a expressão no lado esquerdo de (12.4.5) não aumenta e após um número finito de passos se obtém a cobertura \mathcal{V}^m . Para cada $U \in \mathcal{U}$, considere $k \geq 0$ mínimo tal que U intersecta um único elemento V de \mathcal{V}^k . A escolha de m implica que $k \leq m$: para $k > m$, se U intersecta um elemento de \mathcal{V}^k então ele contém todos os elementos de \mathcal{V}^k dentro do mesmo elemento de \mathcal{V}^m . Suponha que $k < m$. Pela escolha de k , o conjunto U intersecta exatamente dois elementos V_1 e V_2 de \mathcal{V}^{k+1} . Sejam U_1 e U_2 as interseções. Então,

$$\text{diam } U_i \leq \text{diam } V_i = 3^{-n-1} \quad \text{e} \quad \text{diam } U = \text{diam } U_1 + 1/3 + \text{diam } U_2.$$

Logo (Exercício 12.4.1),

$$(\text{diam } U)^{d_0} \geq (\text{diam } U_1)^{d_0} + (\text{diam } U_2)^{d_0}$$

Isto quer dizer que o valor do lado esquerdo de (12.4.5) não aumenta quando substituímos U por U_1 e U_2 na cobertura \mathcal{U} . Por um lado, a nova cobertura satisfaz as mesmas condições que a original: U_1 e U_2 são intervalos abertos (porque V_1 , V_2 e U são intervalos abertos) e eles contêm todo elemento de \mathcal{V}^m que eles intersectam. Por outro lado, por construção, cada um deles intersecta um único elemento de \mathcal{V}^{k+1} . Portanto, ao cabo de um número finito de repetições deste procedimento reduzimos a situação inicial ao caso em que $k = m$ para todo $U \in \mathcal{U}$. Ora, a escolha de m implica que nesse caso cada $U \in \mathcal{U}$ contém o respectivo $V \in \mathcal{V}^m$. Aliás, como U não intersecta nenhum outro elemento de \mathcal{V}^m , segue que $U = V$. Em particular, quaisquer elementos de \mathcal{U} que correspondam a um mesmo $V \in \mathcal{V}^m$ devem coincidir. Eliminando tais repetições obtemos a cobertura \mathcal{V}^m . Isto completa a prova da afirmação (12.4.5) a qual, juntamente com (12.4.3), implica (12.4.4).

12.4.2 Repulsores conformes

Sejam D, D_1, \dots, D_N compactos convexos num espaço euclidiano \mathbb{R}^ℓ tais que $D_i \subset D$ para todo i e $D_i \cap D_j = \emptyset$ sempre que $i \neq j$. Denote $D_* = D_1 \cup \dots \cup D_N$ e suponha que

$$\text{vol}(D \setminus D_*) > 0 \quad (12.4.6)$$

onde vol denota a medida de volume em \mathbb{R}^ℓ . Suponha também que existe uma aplicação $f : D_* \rightarrow D$ tal que a restrição a cada D_i é um homeomorfismo sobre D . Veja a Figura 12.2. Note que a sequência das pré-imagens $f^{-n}(D)$ é não crescente. Chamamos *repulsor* de f à sua interseção:

$$\Lambda = \bigcap_{n=0}^{\infty} f^{-n}(D). \quad (12.4.7)$$

Em outras palavras, Λ é o conjunto dos pontos x cujos iterados $f^n(x)$ estão definidos para todo $n \geq 1$. É claro que Λ é compacto e que $f^{-1}(\Lambda) = \Lambda$.

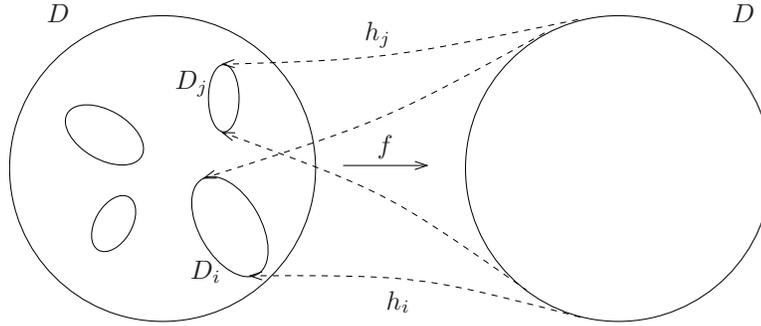


Figura 12.2: Repulsor

Exemplo 12.4.2. O conjunto de Cantor K no Exemplo 12.4.1 é o repulsor da transformação $f : [0, 1/3] \cup [2/3, 1] \rightarrow [0, 1]$ dada por $f(x) = 3x$ se $x \in [0, 1/3]$ e $f(x) = 3x - 2$ se $x \in [2/3, 1]$. Uma classe mais geral de exemplos em dimensão 1 foi introduzida no Exemplo 11.2.3.

No que segue suporemos que a aplicação $f : D_* \rightarrow D$ é de classe C^1 ; para pontos no bordo do domínio isto quer dizer que f se estende a uma aplicação C^1 numa vizinhança. Também fazemos as seguintes hipóteses adicionais.

A primeira hipótese é que a aplicação f é expansora: existe $\sigma > 1$ tal que

$$\|Df(x)v\| \geq \sigma \quad \text{para todo } x \in D_* \text{ e todo } v \in \mathbb{R}^\ell. \quad (12.4.8)$$

Não é difícil deduzir que a restrição $f : \Lambda \rightarrow \Lambda$ da aplicação ao repulsor é uma transformação expansora no sentido da Seção 11.2.

A segunda hipótese é que o logaritmo do jacobiano de f é Hölder: existem $C > 0$ e $\theta > 0$ tais que

$$\log \frac{|\det Df(x)|}{|\det Df(y)|} \leq C \|x - y\|^\theta \quad \text{para todo } x, y \in D_*. \quad (12.4.9)$$

A menos de escolher C suficientemente grande, a desigualdade é automaticamente satisfeita quando x e y pertencem a subdomínios D_i e D_j distintos, uma vez que $d(D_i, D_j) > 0$.

A terceira e última hipótese é que a aplicação f é conforme:

$$\|Df(x)\| \|Df(x)^{-1}\| = 1 \quad \text{para todo } x \in D_*. \quad (12.4.10)$$

É importante notar que esta condição é automática quando $\ell = 1$. Quando $\ell = 2$ ela vale se, e somente se, a transformação f é analítica.

Todas estas condições são satisfeitas no caso do conjunto de Cantor (lembre os Exemplos 12.4.1 e 12.4.2). Elas também são satisfeitas no Exemplo 11.2.3, desde que suponhamos que a derivada da respectiva aplicação f é Hölder.

Teorema 12.4.3 (Fórmula de Bowen-Manning). *Suponha que $f : D_* \rightarrow D$ satisfaz as condições (12.4.8), (12.4.9) e (12.4.10). Então a dimensão de Hausdorff do repulsor é dada por*

$$d(\Lambda) = d_0 \ell,$$

onde $d_0 \in (0, 1)$ é o único número tal que $P(f, -d_0 \log |\det Df|) = 0$.

Alertamos o leitor para um pequeno abuso de linguagem que cometeremos ao longo desta seção, para não sobrecarregar demasiado as notações: $P(f, \psi)$ sempre representará a pressão de um potencial $\psi : \Lambda \rightarrow \mathbb{R}$ relativamente à restrição $f : \Lambda \rightarrow \Lambda$ ao repulsor, embora em outros pontos dos argumentos consideremos a aplicação f definida em todo o domínio D_* .

Antes de passarmos à demonstração do teorema, mencionemos o seguinte caso particular interessante:

Exemplo 12.4.4. Seja $f : J \rightarrow [0, 1]$ uma aplicação como no Exemplo 11.2.3 e suponha que a restrição de f a cada componente conexa J_i de J é afim: o valor absoluto da derivada é constante, igual ao inverso do comprimento $|J_i|$. Então a dimensão de Hausdorff do repulsor K da aplicação f coincide com o único número τ tal que

$$\sum_i |J_i|^\tau = 1. \quad (12.4.11)$$

Para obter esta conclusão a partir do Teorema 12.4.3 basta observar que

$$P(f, -t \log |f'|) = \log \sum_i |J_i|^t \quad \text{para todo } t. \quad (12.4.12)$$

Deixamos ao cuidado do leitor verificar esta igualdade (Exercício 12.4.6).

12.4.3 Distorção e conformalidade

Chamamos ramo inverso de f à inversa $h_i : D \rightarrow D_i$ da restrição de f a cada um dos domínios D_i . Mais geralmente, chamamos ramo inverso de f^n a qualquer composição

$$h^n = h_{i_0} \circ \cdots \circ h_{i_{n-1}} \quad (12.4.13)$$

com $i_0, \dots, i_{n-1} \in \{1, \dots, N\}$. Para cada $n \geq 1$, denotamos por \mathcal{I}^n a família dos ramos inversos h^n de f^n . Por construção, as imagens $h^n(D)$, $h^n \in \mathcal{I}^n$ são disjuntas dois-a-dois e a sua união contém Λ .

O principal objetivo desta seção é provar a seguinte estimativa geométrica, que está base da demonstração do Teorema 12.4.3:

Proposição 12.4.5. *Existe $C_0 > 1$ tal que para todo $n \geq 1$, todo $h^n \in \mathcal{I}^n$, todo $E \subset h^n(D)$ e todo $x \in h^n(D)$, tem-se:*

$$\frac{1}{C_0} [\text{diam } f^n(E)]^\ell \leq [\text{diam } E]^\ell |\det Df^n(x)| \leq C_0 [\text{diam } f^n(E)]^\ell. \quad (12.4.14)$$

Iniciando a demonstração desta proposição, observe que as nossas hipóteses implicam que todo ramo inverso h_i de f é um difeomorfismo com $\|Dh_i\| \leq \sigma^{-1}$. Então, como D é convexo, podemos usar o teorema da média para concluir que

$$\|h_i(z) - h_i(w)\| \leq \sigma^{-1} \|z - w\| \quad \text{para todo } z, w \in D. \quad (12.4.15)$$

Para cada ramo inverso h^n como em (12.4.13), consideraremos a sequência de ramos inversos

$$h^{n-k} = h_{i_k} \circ \dots \circ h_{i_{n-1}}, \quad k = 0, \dots, n-1. \quad (12.4.16)$$

Note que $h^{n-k}(D) \subset D_{i_k}$ para cada k . Segue de (12.4.15) que cada h^{n-k} é uma σ^{k-n} -contração. Em particular,

$$\text{diam } h^{n-k}(D) \leq \sigma^{k-n} \text{diam } D \quad \text{para todo } k = 0, \dots, n-1. \quad (12.4.17)$$

No próximo resultado consideraremos as envolventes convexas dos conjuntos $h^{n-k}(D)$. Lembre que a *envolvente convexa* de um subconjunto de \mathbb{R}^ℓ é a união de todos os segmentos de reta cujos extremos estão no conjunto. É claro que a envolvente convexa tem o mesmo diâmetro que o próprio conjunto. Como D_i é convexo para todo i , a envolvente convexa de cada $h^{n-k}(D)$ está contida em D_{i_k} . Em particular, a derivada Df está definida em todo ponto da envolvente convexa.

Lema 12.4.6. *Existe $C_1 > 1$ tal que, para todo $n \geq 1$ e todo ramo inverso h^n de f^n , vale*

$$\log \prod_{k=0}^{n-1} \frac{|\det Df(z_k)|}{|\det Df(w_k)|} \leq C_1$$

para quaisquer z_k, w_k na envolvente convexa de $h^{n-k}(D)$ para $k = 0, \dots, n-1$.

Demonstração. A condição (12.4.9) dá que

$$\log \frac{|\det Df(z_k)|}{|\det Df(w_k)|} \leq C \|z_k - w_k\|^\theta \leq C [\text{diam } h^{n-k}(D)]^\theta$$

para cada $k = 0, \dots, n-1$. Então, usando (12.4.17),

$$\log \prod_{k=0}^{n-1} \frac{|\det Df(z_k)|}{|\det Df(w_k)|} \leq \sum_{k=0}^{n-1} C [\text{diam } h^{n-k}(D)]^\theta \leq C [\text{diam } D]^\theta \sum_{k=0}^{n-1} \sigma^{(k-n)\theta}$$

Portanto, basta tomar $C_1 = C [\text{diam } D]^\theta \sum_{j=1}^{\infty} \sigma^{-j\theta}$. □

Chegou a hora de explorarmos a hipótese de conformalidade (12.4.10). Dado qualquer isomorfismo linear $L : \mathbb{R}^\ell \rightarrow \mathbb{R}$, é imediato que $|\det L| \leq \|L\|^\ell$ e analogamente para a inversa. Portanto,

$$1 = |\det L| |\det L^{-1}| \leq (\|L\| \|L^{-1}\|)^\ell.$$

Logo, $\|L\| \|L^{-1}\| = 1$ implica que $|\det L| = \|L\|^\ell$ e analogamente para a inversa. Portanto, (12.4.10) implica que

$$|\det Df(y)| = \|Df(y)\|^\ell \quad \text{para todo } y \in D_*. \quad (12.4.18)$$

Agora estamos prontos para demonstrar a Proposição 12.4.5:

Demonstração da Proposição 12.4.5. Sejam n , h^n , E e x como no enunciado. Seja w um ponto de máximo para a norma de Dh^n no domínio D . Então, pelo teorema da média,

$$\|x_1 - x_2\| \leq \|Dh^n(w)\| \|f^n(x_1) - f^n(x_2)\| \quad (12.4.19)$$

para quaisquer x_1, x_2 em E . Note que $Dh^n(w)$ é a inversa de $Df^n(z)$, onde $z = h^n(w)$. Logo, por conformalidade, $\|Dh^n(w)\| = \|Df^n(z)\|^{-1}$. Além disso, usando o Lema 12.4.6 e a igualdade (12.4.18),

$$|\det Df^n(x)| \leq C_1 |\det Df^n(z)| = C_1 \|Df^n(z)\|^\ell. \quad (12.4.20)$$

Combinando (12.4.19) and (12.4.20), obtemos

$$\|x_1 - x_2\|^\ell \leq C_1 |\det Df^n(x)|^{-1} \|f^n(x_1) - f^n(x_2)\|^\ell.$$

Fazendo variar $x_1, x_2 \in E$, vem que

$$[\text{diam } E]^\ell \leq C_1 |\det Df^n(x)|^{-1} [\text{diam } f^n(E)]^\ell.$$

Isto prova a segunda desigualdade em (12.4.14), desde que tomemos $C_0 \geq C_1$.

A prova da outra desigualdade é semelhante. Para cada $k = 0, \dots, n-1$, seja z_k um ponto de máximo para a norma de Df restrita à envolvente convexa de $h^{n-k}(D)$. Então,

$$\begin{aligned} \|f^{k+1}(x_1) - f^{k+1}(x_2)\| &\leq \|Df(z_k)\| \|f^k(x_1) - f^k(x_2)\| \\ &= |\det Df(z_k)|^{1/\ell} \|f^k(x_1) - f^k(x_2)\| \end{aligned}$$

para todo k e quaisquer $x_1, x_2 \in E$. Logo,

$$\|f^n(x_1) - f^n(x_2)\|^\ell \leq \prod_{k=0}^{n-1} |\det Df(z_k)| \|x_1 - x_2\|^\ell. \quad (12.4.21)$$

Pelo Lema 12.4.6,

$$\prod_{k=0}^{n-1} |\det Df(z_k)| \leq C_1 |\det Df^n(x)|. \quad (12.4.22)$$

Combinando (12.4.21) e (12.4.22), obtemo

$$\|y_1 - y_2\|^\ell \leq C_1 |\det Df^n(x)| \|x_1 - x_2\|^\ell.$$

Fazendo variar y_1, y_2 , concluímos que

$$[\text{diam } f^n(E)]^\ell \leq C_1 |\det Df^n(x)| [\text{diam } E]^\ell.$$

Isto prova a primeira desigualdade em (12.4.14), para qualquer $C_0 \geq C_1$. \square

12.4.4 Existência e unicidade de d_0

Nesta seção provamos a existência e unicidade do número d_0 no enunciado do Teorema 12.4.3. Denote $\phi = -\log |\det Df|$ e considere a função

$$\Psi : \mathbb{R} \rightarrow \mathbb{R}, \quad \Psi(t) = P(f, t\phi).$$

Queremos mostrar que existe um único d_0 tal que $\Psi(d_0) = 0$.

A unicidade é fácil de obter. De fato, as hipóteses (12.4.8) e (12.4.10) implicam que

$$\phi = \log |\det Df^{-1} \circ f| = \ell \log \|Df^{-1} \circ f\| \leq -\ell \log \sigma$$

Então, dados quaisquer $s < t$, tem-se $t\phi \leq s\phi - (t-s)\ell \log \sigma$. Usando (10.3.5) e (10.3.6), segue que

$$P(f, t\phi) \leq P(f, s\phi) - (t-s)\ell \log \sigma < P(f, s\phi).$$

Isto prova que Ψ é estritamente decrescente e, portanto, existe um único $d_0 \in \mathbb{R}$ tal que $\Psi(d_0) = 0$.

Por outro lado, segue da Proposição 10.3.10 que Ψ é contínua. Logo, para provar a existência de d_0 basta mostrar que $\Psi(0) < 0 < \Psi(1)$. Isso pode ser feito do seguinte modo.

Seja \mathcal{L} a cobertura aberta de Λ cujos elementos são as imagens $h(\Lambda)$ pelos ramos inversos de f . Para cada $n \geq 1$, a cobertura \mathcal{L}^n está formada pelas imagens $h^n(\Lambda)$ do repulsor Λ pelos ramos inversos de f^n . Segue de (12.4.17) que $\text{diam } \mathcal{L}^n \leq \sigma^{-n} \text{diam } D$ para todo n e, portanto, $\text{diam } \mathcal{L}^n \rightarrow 0$. Então a Proposição 10.3.5 dá que

$$P(f, \psi) = P(f, \psi, \mathcal{L}) \quad \text{para todo potencial } \psi. \quad (12.4.23)$$

Em particular, $\Psi(0) = P(f, 0, \mathcal{L}) = h(f, \mathcal{L})$. Note que cada \mathcal{L}^n é uma cobertura minimal de Λ , isto é, nenhuma subfamília própria cobre Λ . Portanto, $H(\mathcal{L}^n) = \log \#\mathcal{L}^n = n \log N$ e, conseqüentemente, $h(f, \mathcal{L}) = \log N$. Isto prova que $\Psi(0)$ é positivo.

Proposição 12.4.7. $\Psi(1) = \lim_n \frac{1}{n} \log \text{vol}(f^{-n}(D)) < 0$.

Demonstração. Por (12.4.23), temos que $\Psi(1) = P(f, \phi, \mathcal{L})$. Em outras palavras

$$\Psi(1) = \lim_n \frac{1}{n} \log P_n(f, \phi, \mathcal{L}) = \lim_n \frac{1}{n} \log \sum_{h^n \in \mathcal{I}^n} e^{S_n \phi(h^n(\Lambda))}.$$

Como $\phi = -\log |\det Df|$, isto quer dizer que

$$\Psi(1) = \lim_n \frac{1}{n} \log \sum_{h^n \in \mathcal{I}^n} \sup_{h^n(D)} \frac{1}{|\det Df^n|}. \quad (12.4.24)$$

Por outro lado, pela fórmula de mudança de variáveis em integrais múltiplas,

$$\text{vol}(f^{-n}(D)) = \sum_{h^n \in \mathcal{I}^n} \text{vol}(h^n(D)) = \sum_{h^n \in \mathcal{I}^n} \int_D \frac{1}{|\det Df^n|} \circ h^n.$$

Segue do Lema 12.4.6 que

$$\inf_{h^n(D)} |\det Df^n| \leq |\det Df^n|(h^n(z)) \leq C_1 \inf_{h^n(D)} |\det Df^n|$$

para todo $z \in h^n(D)$ e todo $h^n \in \mathcal{I}^n$. Consequentemente,

$$\text{vol}(f^{-n}(D)) \leq \sum_{h^n \in \mathcal{I}^n} \sup_{h^n(D)} \frac{1}{|\det Df^n|} \leq C_1 \text{vol}(f^{-n}(D)).$$

Combinando estas desigualdades com (12.4.24), concluimos que

$$\limsup_n \frac{1}{n} \log \text{vol}(f^{-n}(D)) \leq \Psi(1) \leq \liminf_n \frac{1}{n} \log \text{vol}(f^{-n}(D)).$$

Isto prova a igualdade no enunciado da proposição.

Resta provar que o volume das pré-imagens $f^{-n}(D)$ decai exponencialmente. Par isso, observe que $f^{-(n+1)}(D) = f^{-n}(D_*)$ é a união disjunta das imagens $h^n(D_*)$, com $h^n \in \mathcal{I}^n$. Portanto,

$$\frac{\text{vol}(f^{-(n+1)}(D))}{\text{vol}(f^{-(n+1)}(D))} = \frac{\sum_{h^n \in \mathcal{I}^n} \text{vol}(h^n(D_*))}{\sum_{h^n \in \mathcal{I}^n} \text{vol}(h^n(D))} \leq \max_{h^n \in \mathcal{I}^n} \frac{\text{vol}(h^n(D_*))}{\text{vol}(h^n(D))}. \quad (12.4.25)$$

Pela fórmula de mudança de variável,

$$\text{vol}(h^n(D)) = \int_D \frac{1}{|\det Df^n|} \circ h^n \quad \text{e} \quad \text{vol}(h^n(D \setminus D_*)) = \int_{D \setminus D_*} \frac{1}{|\det Df^n|} \circ h^n.$$

Logo, usando o Lema 12.4.6,

$$\frac{\text{vol}(h^n(D \setminus D_*))}{\text{vol}(h^n(D))} \geq \frac{1}{C_1} \frac{\text{vol}(D_*)}{\text{vol}(D)} \quad (12.4.26)$$

para todo $h^n \in \mathcal{I}^n$. Pela hipótese (12.4.6), a expressão no lado direito de (12.4.26) é positiva. Fixe $\beta > 0$ suficientemente próximo de zero para que $1 - e^{-\beta}$ seja menor que essa expressão. Então temos

$$\frac{\text{vol}(h^n(D) \setminus h^n(D_*))}{\text{vol}(h^n(D))} \geq 1 - e^{-\beta}$$

para todo $h^n \in \mathcal{I}^n$. Combinando esta desigualdade com (12.4.25), obtemos que

$$\frac{\text{vol}(f^{-(n+1)}(D))}{\text{vol}(f^{-(n+1)}(D))} \leq e^{-\beta} \quad \text{para todo } n \geq 0$$

(o caso $n = 0$ segue diretamente da hipótese (12.4.6)). Logo,

$$\lim_n \frac{1}{n} \log \text{vol}(f^{-n}(D)) \leq \beta < 0.$$

Isto conclui a prova da proposição. \square

A Figura 12.3 resume as conclusões desta seção. Lembre que a função definida por $\Psi(t) = P(f, -t \log |\det Df|)$ é convexa, pela Proposição 10.3.9.

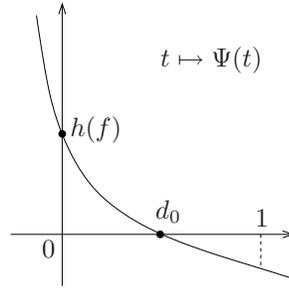


Figura 12.3: Pressão e dimensão de Hausdorff

12.4.5 Cota superior

Vamos mostrar que $d(\Lambda) \leq b\ell$ para todo b tal que $P(f, b\phi) < 0$. Em vista das observações na seção precedente, isto prova que $d(\Lambda) \leq d_0\ell$.

Seja \mathcal{L} a cobertura aberta de Λ cujos elementos são as imagens $h(\Lambda)$ pelos ramos inversos de f . Para cada $n \geq 1$, a cobertura \mathcal{L}^n está formada pelas imagens $h^n(\Lambda)$ do repulsor Λ pelos ramos inversos de f^n . Segue de (12.4.17) que $\text{diam } \mathcal{L}^n \leq \sigma^{-n} \text{diam } D$ para todo n e, portanto, $\text{diam } \mathcal{L}^n \rightarrow 0$. Pela Proposição 10.3.5, segue que $P(f, \psi) = P(f, \psi, \mathcal{L})$ para todo potencial ψ . Em particular,

$$P(f, b\phi, \mathcal{L}) = P(f, b\phi) < -\kappa$$

para algum $\kappa > 0$. Pela definição (10.3.3), isto implica que

$$P_n(f, b\phi, \mathcal{L}) \leq e^{-\kappa n} \quad \text{para todo } n \text{ suficientemente grande.} \quad (12.4.27)$$

Claro que \mathcal{L}^n é cobertura minimal de Λ : nenhuma subfamília própria cobre Λ . Logo, lembrando a definição (10.3.2), a desigualdade (12.4.27) implica que

$$\sum_{L \in \mathcal{L}^n} e^{bS_n\phi(L)} \leq e^{-\kappa n} \quad \text{para todo } n \text{ suficientemente grande.} \quad (12.4.28)$$

É claro que todo $L \in \mathcal{L}^n$ é compacto. Logo, por continuidade do jacobiano,

$$e^{S_n\phi(L)} = \sup_L |\det Df^n|^{-1} = |\det Df^n(x)|^{-1}$$

para algum $x \in L$. Também é claro que $f^n(L) = \Lambda$ para todo $L \in \mathcal{L}^n$. Então, tomando $E = L$ na Proposição 12.4.5, obtemos

$$[\text{diam } L]^\ell e^{-S_n\phi(L)} \leq C_0 [\text{diam } \Lambda]^\ell.$$

Combinando esta desigualdade com (12.4.28), obtemos que

$$\sum_{L \in \mathcal{L}^n} [\text{diam } L]^{b\ell} \leq C_0^b [\text{diam } \Lambda]^{b\ell} \sum_{L \in \mathcal{L}^n} e^{bS_n\phi(L)} \leq C_0^b [\text{diam } \Lambda]^{b\ell} e^{-\kappa n}$$

para todo n suficientemente grande. Como a expressão do lado direito converge para zero, e o diâmetro das coberturas \mathcal{L}^n também converge para zero, segue que $m_{b\ell}(M) = 0$. Portanto, $d(M) \leq b\ell$.

12.4.6 Cota inferior

Agora vamos mostrar que $d(\Lambda) \geq a$ para todo a tal que $P(f, a\phi) > 0$. Isso implica que $d(\Lambda) \geq d_0$, o que completa a demonstração do Teorema 12.4.3.

Como observamos na seção anterior, a cobertura \mathcal{L} realiza a pressão e todos os seus iterados \mathcal{L}^n são coberturas minimais de Λ . Logo, a escolha de a implica que existe $\kappa > 0$ tal que

$$P_n(f, a, \phi) = \sum_{L \in \mathcal{L}^n} e^{aS_n\phi(L)} \geq e^{\kappa n} \quad \text{para todo } n \text{ suficientemente grande.} \quad (12.4.29)$$

Fixe n nessas condições. Seja $\varepsilon > 0$ um minorante para a distância entre dois elementos quaisquer de \mathcal{L}^n : tal minorante existe porque os elementos de \mathcal{L}^n são compactos e disjuntos dois-a-dois. Fixe $\rho \in (0, \varepsilon^{a\ell})$. A razão de ser desta escolha ficará clara num instante. Afirmamos que

$$\sum_{U \in \mathcal{U}} [\text{diam } U]^{a\ell} \geq 2^{-a\ell} \rho \quad (12.4.30)$$

para toda cobertura \mathcal{U} de Λ . Por definição, isto implica que $m_{a\ell}(\Lambda) \geq 2^{-a\ell} \rho > 0$ e, conseqüentemente, $d(\Lambda) \geq a\ell$. Portanto, para terminar a prova do Teorema 12.4.3 basta provar esta afirmação.

Suponhamos que existe alguma cobertura de Λ que não satisfaz (12.4.30). Então, usando o Exercício 12.4.3, existe alguma cobertura aberta \mathcal{U} de Λ com

$$\sum_{U \in \mathcal{U}} [\text{diam } U]^{a\ell} < \rho < \varepsilon^{a\ell}. \quad (12.4.31)$$

Por compacidade, podemos supor que esta cobertura \mathcal{U} aberta é finita. A relação (12.4.31) que todo $U \in \mathcal{U}$ tem diâmetro menor que ε . Logo, cada $U \in \mathcal{U}$ intersecta no máximo um $L \in \mathcal{L}^n$. Como \mathcal{L}^n cobre Λ e U é subconjunto não vazio de Λ , também temos que U intersecta algum $L \in \mathcal{L}^n$. Isto quer dizer que \mathcal{U} é a união disjunta das famílias

$$\mathcal{U}_L = \{U \in \mathcal{U} : U \cap L \neq \emptyset\}, \quad L \in \mathcal{L}^n.$$

Se $U \in \mathcal{U}_L$ então $U \subset L$. Consideremos as famílias $f^n(\mathcal{U}_L) = \{f^n(U) : U \in \mathcal{U}_L\}$. Observe que cada uma delas é uma cobertura de Λ . Além disso, usando a Proposição 12.4.5,

$$\sum_{V \in f^n(\mathcal{U}_L)} [\text{diam } V]^{a\ell} = \sum_{U \in \mathcal{U}_L} [\text{diam } f^n(U)]^{a\ell} \leq C_0 e^{-aS_n\phi(L)} \sum_{U \in \mathcal{U}_L} [\text{diam } U]^{a\ell}. \quad (12.4.32)$$

Portanto,

$$\sum_{U \in \mathcal{U}} [\text{diam } U]^{a\ell} = \sum_{L \in \mathcal{L}^n} \sum_{U \in \mathcal{U}_L} [\text{diam } U]^{a\ell} \geq \sum_{L \in \mathcal{L}^n} C_0^{-1} e^{aS_n \phi(L)} \sum_{V \in f^n(\mathcal{U}_L)} [\text{diam } V]^{a\ell}.$$

Suponhamos que

$$\sum_{V \in f^n(\mathcal{U}_L)} [\text{diam } V]^{a\ell} \geq \sum_{U \in \mathcal{U}} [\text{diam } U]^{a\ell} \quad \text{para todo } L \in \mathcal{L}^n.$$

Então a desigualdade anterior implica

$$\sum_{U \in \mathcal{U}} [\text{diam } U]^{a\ell} \geq \sum_{L \in \mathcal{L}^n} C_0^{-1} e^{aS_n \phi(L)} \sum_{U \in \mathcal{U}} [\text{diam } U]^{a\ell} \geq C_0^{-1} e^{\kappa n} \sum_{U \in \mathcal{U}} [\text{diam } U]^{a\ell}.$$

Isto é uma contradição, uma vez que $e^{\kappa n} > C_0$. Logo, existe algum $L \in \mathcal{L}^n$ tal que

$$\sum_{V \in f^n(\mathcal{U}_L)} [\text{diam } V]^{a\ell} \leq \sum_{U \in \mathcal{U}} [\text{diam } U]^{a\ell} < \rho.$$

Portanto, podemos repetir o procedimento anterior com $f^n(\mathcal{U}_L)$ no lugar de \mathcal{U} . Observe, no entanto, que $\#f^n(\mathcal{U}_L) = \#\mathcal{U}_L$ é estritamente menor que $\#\mathcal{U}$. Portanto, este processo precisa parar após um número finito de etapas. Esta contradição prova a afirmação 12.4.30.

A demonstração do Teorema 12.4.3 está completa. No entanto, é possível provar um resultado mais forte: nas condições do teorema, a medida de Hausdorff de Λ na dimensão $d(M)$ é positiva e finita. Deixamos este enunciado como um desafio especial (Exercício 12.4.7) para o leitor que nos acompanhou até o final deste livro!

12.4.7 Exercícios

12.4.1. Seja $d = \log 2 / \log 3$. Mostre que $(x_1 + 1 + x_2)^d \geq x_1^d + x_2^d$ para todo $x_1, x_2 \in [0, 1]$. Além disso, vale a igualdade se, e somente se, $x_1 = x_2 = 1$.

12.4.2. Seja $f : M \rightarrow N$ uma aplicação Lipschitz, com constante de Lipschitz L . Mostre que

$$m_d(f(A)) \leq L^d m_d(A).$$

para qualquer $d \in (0, \infty)$ e qualquer $A \subset M$. Use esse fato para mostrar que $A \subset \mathbb{R}^n$ e $t \in \mathbb{R}$, então $m_d(tA) = t^d m_d(A)$, onde $tA = \{tx : x \in A\}$.

12.4.3. Represente por $m_d^a(M)$ e $m_d^f(M)$ os números definidos do mesmo modo que a medida de Hausdorff $m_d(M)$ mas considerando apenas coberturas por subconjuntos abertos e por subconjuntos fechados, respectivamente. Mostre que $m_d(M) = m_d^a(M)$ e $m_d^f(M) = m_d(M)$.

12.4.4. (Princípio da distribuição de massa) Seja μ uma medida finita em um espaço métrico compacto M e suponha que existem números $d, K, \rho > 0$ tais que $\mu(B) \leq K(\text{diam } U)^d$ para todo conjunto com diâmetro menor que ρ . Mostre que se $A \subset M$ é tal que $\mu(A) > 0$ então $m_d(A) > 0$ e, portanto, $d(A) \geq d$.

12.4.5. Use o princípio da distribuição de massa para mostrar que a dimensão de Hausdorff do *triângulo de Sierpinski* (Figura 12.4) é igual a $d_0 = \log 3 / \log 2$ e que a medida de Hausdorff em dimensão d_0 é positiva e finita.

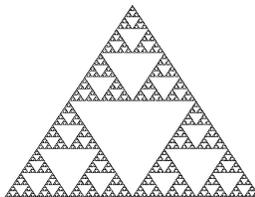


Figura 12.4: Triângulo de Sierpinski

12.4.6. Verifique a fórmula (12.4.12) para a pressão.

12.4.7. Adaptando argumentos do Exercício 12.4.5, mostre que nas condições do Teorema 12.4.3 vale que $0 < m_{d(\Lambda)}(\Lambda) < \infty$.

Apêndice A

Elementos de Medida, Topologia e Análise

Nesta série de apêndices 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, topologia geral e diferencial e teoria espectral, para uma leitura autocontida do livro.

De modo geral, omitiremos as demonstrações dos resultados apresentados. No que se refere aos Apêndices A.1, A.2 e A.5, o leitor poderá consultar os livros de Castro [Cas04], Fernandez [Fer02], Halmos [Hal50], Royden [Roy63] ou Rudin [Rud87]. A apresentação no Apêndice A.3 será mais completa, incluindo as demonstrações da maioria dos resultados, mas o leitor poderá encontrar informação adicional relevante nos livros de Billingsley [Bil68, Bil71]. Ao leitor interessado em aprofundar os temas do Apêndice A.4 recomendamos os livros de Hirsch [Hir94] e do Carmo [dC79]. Para informações mais detalhadas sobre o material dos Apêndices A.6 e A.7, incluindo demonstrações, consulte o livro de Halmos [Hal51] e o tratado de Dunford, Schwarz [DS57, DS63], particularmente, a Seção IV.4 do primeiro volume e as seções iniciais do segundo volume.

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 euclidianos. 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 $j = 1, \dots, n, \dots$ implica $\bigcup_{j=1}^{\infty} A_j \in \mathcal{B}$.

Observe que uma σ -álgebra \mathcal{B} também é fechada para as interseções enumeráveis: se $A_j \in \mathcal{B}$ para $j = 1, \dots, n, \dots$ então $\bigcap_{j=1}^{\infty} A_j = \left(\bigcup_{j=1}^{\infty} A_j^c\right)^c$ 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 $\mathcal{B} = \bigcap_{i \in \mathcal{I}} \mathcal{B}_i$ 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.

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,

$$\liminf_n E_n = \bigcup_{n \geq 1} \bigcap_{m \geq n} E_m \quad \text{e} \quad \limsup_n E_n = \bigcap_n \bigcup_{m \geq n} E_m$$

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 $\arctan : \mathbb{R} \rightarrow (-\pi/2, \pi/2)$ se estende imediatamente a um

homeomorfismo (ou seja, uma bijeção contínua cuja inversa também é contínua) entre \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 $\mu : \mathcal{B} \rightarrow [0, +\infty]$ tal que $\mu(\emptyset) = 0$ e

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu(A_j) \quad \text{para quaisquer } A_j \in \mathcal{B} \text{ disjuntos dois-a-dois.}$$

Esta última propriedade é chamada *σ -aditividade*. A tripla (X, \mathcal{B}, μ) 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}, μ) 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 $\delta_p : 2^X \rightarrow [0, +\infty]$ definida por:

$$\delta_p(A) = \begin{cases} 1 & \text{se } p \in A \\ 0 & \text{se } p \notin A. \end{cases}$$

Esta medida δ_p é usualmente designada *medida de Dirac* no ponto p .

Definição A.1.12. Dizemos que uma medida é *σ -finita* se existe uma seqüência de subconjuntos A_1, \dots, A_n, \dots 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 $\mu : \mathcal{B} \rightarrow [0, +\infty]$ é *finitamente aditiva* se

$$\mu\left(\bigcup_{j=1}^N A_j\right) = \sum_{j=1}^N \mu(A_j)$$

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 $\mu_0 : \mathcal{A} \rightarrow [0, +\infty]$ 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 $\mu : \mathcal{A} \rightarrow [0, +\infty)$ uma função finitamente aditiva com $\mu(X) < \infty$. Então μ é σ -aditiva se, e somente se,*

$$\lim_n \mu(A_n) = 0 \tag{A.1.1}$$

para toda sequência $A_1 \supset \dots \supset A_j \supset \dots$ de elementos de \mathcal{A} com $\bigcap_{j=1}^{\infty} A_j = \emptyset$.

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 $A_1 \supset \dots \supset A_n \supset \dots$ 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 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 $\bigcap_n K_n$ de qualquer sequência decrescente $K_1 \supset \dots \supset K_n \supset \dots$ 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 $\mu : \mathcal{A} \rightarrow [0, +\infty)$ 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 $A_1 \subset A_2 \subset \dots$ em \mathcal{C} , então $\cup_{n \geq 1} A_n \in \mathcal{C}$ e
- dados subconjuntos $A_1 \supset A_2 \supset \dots$ 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}, μ) 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}, μ) num espaço completo, do seguinte modo. A família $\tilde{\mathcal{B}}$ 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 $\bar{\mu} : \tilde{\mathcal{B}} \rightarrow [0, +\infty]$ 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 $\tilde{\mathcal{B}}$, cuja restrição a \mathcal{B} coincide com μ . Por construção, $(X, \tilde{\mathcal{B}}, \bar{\mu})$ é um espaço de medida completo. É frequente na literatura (embora seja um abuso de linguagem) chamar uma função em (X, \mathcal{B}, μ) de mensurável se ela for mensurável para a σ -álgebra completada $\tilde{\mathcal{B}}$.

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 serão úteis no que segue.

Chamaremos *medida com sinal* num espaço mensurável (X, \mathcal{B}) a toda função σ -aditiva $\mu : \mathcal{B} \rightarrow [-\infty, \infty]$ 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 uma função ϕ é integrável para μ^+ e para μ^- , dizemos que ela é integrável para μ e definimos:

$$\int \phi d\mu = \int \phi d\mu^+ - \int \phi d\mu^-.$$

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 $\mu : \mathcal{B} \rightarrow \mathbb{C}$. 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 euclidiano \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, \dots, 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 $m_0 : \mathcal{A} \rightarrow [0, 1]$ definida nesta álgebra por

$$m_0(I_1 \cup \cdots \cup I_N) = |I_1| + \cdots + |I_N|,$$

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, \dots, R_N são retângulos disjuntos dois-a-dois, e definimos

$$m_0(A) = m_0(R_1) + \cdots + m_0(R_N)$$

para todo B 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 , decomponemos 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 $T_{k_1, \dots, k_d}(x) = x - (k_1, \dots, k_d)$ que envia o ponto (k_1, k_2, \dots, k_d) na origem. Isso nos permite definir uma medida m_{k_1, k_2, \dots, k_d} em C , dada por

$$m_{k_1, k_2, \dots, k_d}(B) = m_0(T_{k_1, \dots, k_d}(B))$$

para todo o conjunto mensurável $B \subset C$. Finalmente, dado qualquer conjunto mensurável $B \subset \mathbb{R}^d$, definimos

$$m(B) = \sum_{k_1 \in \mathbb{Z}} \cdots \sum_{k_d \in \mathbb{Z}} m_{k_1, \dots, k_d}(B \cap [k_1, k_1 + 1) \times \cdots \times [k_d, k_d + 1)).$$

Note que m não é uma medida finita, mas é uma medida σ -finita.

Exemplo A.1.22. Lembremos 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 um conjunto $E \subset \mathbb{R}$ ao número

$$m^*(E) = \inf \left\{ \sum_k m_0(R_k) : \right. \\ \left. (R_k)_k \text{ é cobertura enumerável de } E \text{ por retângulos abertos} \right\}.$$

Esta função está definida para todo $E \subset \mathbb{R}$, 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 $\phi : [0, 1] \rightarrow \mathbb{R}$ uma função contínua e positiva. Dado qualquer intervalo I , com extremos $0 \leq a < b \leq 1$, defina

$$\mu_\phi(I) = \int_a^b \phi(x) dx \quad (\text{integral de Riemann}).$$

Em seguida, estenda a definição de μ_ϕ para a álgebra \mathcal{A} das uniões finitas $A = I_1 \cup \dots \cup I_k$ 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 será estudada com mais detalhe na Seção 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 em um 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 ν' de subconjuntos de X é uma *base de vizinhanças* de um ponto $x \in X$ se para toda vizinhança V existe algum $V' \in \nu'$ tal que $x \in V' \subset V$. Dizemos que x admite *base enumerável de vizinhanças* se ν' 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 $\text{supp } \mu$ da medida μ é o conjunto 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 $\text{supp } \mu$ é não vazio.*

Demonstração. Se $\text{supp } \mu$ 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

$$\mu(X) = \mu\left(\bigcup_{x \in X} A_{i(x)}\right) \leq \sum_{i=1}^{\infty} \mu(A_i) = 0.$$

Isto é uma contradição e, portanto, $\text{supp } \mu$ não pode ser vazio. \square

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 $f : X \rightarrow Y$ 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 $f : X \rightarrow [-\infty, \infty]$ é mensurável se, e somente se, $f^{-1}((c, +\infty])$ 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 $f : X \rightarrow \mathbb{R}$ 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* $\mathcal{X}_B : X \rightarrow \mathbb{R}$ 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 $f, g : X \rightarrow [-\infty, +\infty]$ 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 $f_n : X \rightarrow [-\infty, +\infty]$ 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 $s : X \rightarrow \mathbb{R}$ 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

$$s = \sum_{j=1}^k \alpha_j \mathcal{X}_{A_j}, \quad (\text{A.1.2})$$

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 $f : X \rightarrow [-\infty, +\infty]$ 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 toma valores em \mathbb{R} , podemos tomar s_n com valores em \mathbb{R} . Se f é limitada, a sequência pode ser escolhida tal que a convergência seja uniforme. Se f é não negativa, podemos tomar $0 \leq s_1 \leq s_2 \leq \dots \leq f$.

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 $f : X \rightarrow Y$, 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}, μ) um espaço de medida. Mostre que se A_1, A_2, \dots estão em \mathcal{B} então

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) \leq \sum_{j=1}^{\infty} \mu(A_j).$$

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, $F = \limsup_n E_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} E_n$. 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 $m_0 : \mathcal{A} \rightarrow [0, 1]$ a função definida nesta álgebra por

$$m_0(I_1 \cup \dots \cup I_N) = |I_1| + \dots + |I_N|,$$

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 $\mu : \mathcal{B} \rightarrow [0, +\infty)$ 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. $\lim_n \mu(A_n) = \mu(\bigcap_{j=1}^{\infty} A_j)$ para toda sequência $A_1 \supset \dots \supset A_j \supset \dots$ de elementos de \mathcal{B} ;
2. $\lim_n \mu(A_n) = \mu(\bigcup_{j=1}^{\infty} A_j)$ para toda sequência $A_1 \subset \dots \subset A_j \subset \dots$ 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 $X = \{1, \dots, d\}$ 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 em um 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 $\mu : \mathcal{A} \rightarrow [0, 1]$ 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, $K = \bigcap_{n=0}^{\infty} K_n$ 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 complemento 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 $g_n : M \rightarrow \mathbb{R}$, $n \geq 1$ 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 $f : X \rightarrow X$ 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 $\omega_5 : [0, 1] \rightarrow [0, 1]$ 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 $x = 0, a_0a_1a_2 \dots$ 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 seqüências. 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}, μ) 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 $s = \sum_{j=1}^k \alpha_k \mathcal{X}_{A_k}$ uma função simples. Então a *integral* de s em relação à medida μ é dada por:

$$\int s \, d\mu = \sum_{j=1}^k \alpha_k \mu(A_k).$$

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 $f : X \rightarrow [0, \infty]$ uma função mensurável não negativa. Então

$$\int f \, d\mu = \lim_n \int s_n \, d\mu,$$

onde $s_1 \leq s_2 \leq \dots$ é uma seqüência 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 seqüência de funções simples crescendo para f .

Agora, para estender a definição de integral a qualquer função mensurável, observemos que dada uma função $f : X \rightarrow [-\infty, +\infty]$ 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 $f : X \rightarrow [-\infty, +\infty]$ 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 $f : X \rightarrow [-\infty, +\infty]$ 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 $f : X \rightarrow [-\infty, \infty]$ e um conjunto mensurável E definimos a *integral de f sobre E* por

$$\int_E f \, d\mu = \int f \chi_E \, d\mu,$$

onde χ_E é a função característica do conjunto E .

Exemplo A.2.5. Considere $X = [0, 1]$ munido da medida de Lebesgue m . Seja $f = \chi_B$, 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 $p_1 + \dots + p_m = 1$. 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 $f : X \rightarrow [-\infty, +\infty]$,

$$\int f \, d\mu = \sum_{i=1}^m p_i f(x_i).$$

Proposição A.2.7. *O conjunto $\mathcal{L}^1(\mu)$ das funções reais integráveis é um espaço vetorial real. Além disso, a aplicação $I : \mathcal{L}^1(\mu) \rightarrow \mathbb{R}$ dada por $I(f) = \int f d\mu$ é um funcional linear positivo, ou seja:*

$$(1) \int af + bg d\mu = a \int f d\mu + b \int g d\mu \text{ e}$$

$$(2) \int f d\mu \geq \int g d\mu \text{ se } f(x) \geq g(x) \text{ para todo } x.$$

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, $f \in \mathcal{L}^1(\mu)$.

Vamos encerrar esta seção observando que 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 $f : X \rightarrow \mathbb{C}$. Nesse caso, dizemos que f é integrável se, e somente se, a parte real $\Re f$ e a parte imaginária $\Im f$ forem integráveis. Nesse caso, 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

$$\int f d\mu = \int g d\mu \text{ se } f = g \text{ em } \mu\text{-quase todo ponto.}$$

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 $\int f d\mu = \int g d\mu$.

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 $f_n : X \rightarrow [-\infty, +\infty]$ uma sequência monótona de funções mensuráveis não negativas e seja f 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 $f_n : X \rightarrow [0, +\infty]$ uma sequência de funções mensuráveis não negativas. Então, a função $f : X \rightarrow [-\infty, +\infty]$ 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 $f_n : X \rightarrow \mathbb{R}$ 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 . Então f é integrável e vale:*

$$\lim_n \int f_n d\mu = \int f d\mu.$$

A.2.3 Produto de medidas

Sejam $(X_j, \mathcal{A}_j, \mu_j)$, $j = 1, \dots, n$ espaços de medida finita, isto é, tais que $\mu_j(X_j) < \infty$. É possível tornar o produto cartesiano $X_1 \times \dots \times X_n$ um espaço de medida, da seguinte forma. Considere em $X_1 \times \dots \times X_n$ a σ -álgebra gerada pela família de todos os conjuntos da forma $A_1 \times \dots \times A_n$ com $A_j \in \mathcal{A}_j$. Ela é chamada σ -álgebra produto e é representada por $\mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n$.

Teorema A.2.12. *Existe uma única medida μ em $(X_1 \times \dots \times X_n, \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n)$ tal que $\mu(A_1 \times \dots \times A_n) = \mu_1(A_1) \dots \mu_n(A_n)$ para todo $A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n$. 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 $\mu_1 \times \dots \times \mu_n$ e é chamada *produto* das medidas μ_1, \dots, μ_n . Desta forma, fica definido o *espaço de medida produto*

$$(X_1 \times \dots \times X_n, \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n, \mu_1 \times \dots \times \mu_n).$$

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

$$\Sigma = \prod_{j \in \mathcal{I}} X_j = \{(x_j)_{j \in \mathcal{I}} : x_j \in X_j\}. \quad (\text{A.2.1})$$

Chamamos *cilindros* de Σ os subconjuntos da forma

$$[m; A_m, \dots, A_n] = \{(x_j)_{j \in \mathcal{I}} : x_j \in A_j \text{ para } m \leq j \leq n\} \quad (\text{A.2.2})$$

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 (Σ, \mathcal{B}) tal que*

$$\mu([m; A_m, \dots, A_n]) = \mu_m(A_m) \cdots \mu_n(A_n) \quad (\text{A.2.3})$$

para qualquer cilindro $[m; A_m, \dots, A_n]$. 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 da construção anterior ocorre quando os espaços $(X_i, \mathcal{B}_i, \mu_i)$ são todos iguais a um dado (X, \mathcal{C}, ν) . 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

$$\mu([m; A_m, \dots, A_n]) = \prod_{j=m}^n \nu(A_j), \quad (\text{A.2.4})$$

a qual pode ser lida nos seguintes termos: a probabilidade de todo evento composto $\{x_m \in A_m, \dots, x_n \in A_n\}$ 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 e 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*

$$[m; a_m, \dots, a_n] = \{(x_j)_{j \in \mathcal{I}} \in X : x_m = a_m, \dots, x_n = a_n\}, \quad (\text{A.2.5})$$

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

$$\mu([m; a_m, \dots, a_n]) = p_{a_m} \cdots p_{a_n} \quad \text{onde } p_a = \nu(\{a\}) \text{ para } a \in X. \quad (\text{A.2.6})$$

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

$$d((x_i)_{i \in \mathcal{I}}, (y_i)_{i \in \mathcal{I}}) = \theta^N, \quad (\text{A.2.7})$$

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 $a \in A$ é um *ponto de densidade* de A se este conjunto preenche a maior parte de qualquer pequena vizinhança de a , isto é,

$$\lim_{\delta \rightarrow 0} \frac{m(B(a, \delta) \cap A)}{m(B(a, \delta))} = 1. \quad (\text{A.2.8})$$

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 $f : \mathbb{R}^d \rightarrow \mathbb{R}$ é *localmente integrável* se o produto $f \chi_K$ 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 $f : X \rightarrow \mathbb{R}$ uma função localmente integrável. Então*

$$\lim_{r \rightarrow 0} \frac{1}{m(B(x, r))} \int_{B(x, r)} |f(y) - f(x)| dm = 0 \quad \text{em } m\text{-quase todo ponto.}$$

Em particular,

$$\lim_{r \rightarrow 0} \frac{1}{m(B(x, r))} \int_{B(x, r)} f(y) dm = f(x) \quad \text{em } m\text{-quase todo o ponto.}$$

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 $B(x_j, n_j)$ são disjuntas duas-a-duas;
2. $m(\cup_j B_{n_j}(x_j) \setminus A) < \varepsilon$ e $m(A \setminus \cup_j B_{n_j}(x_j)) = 0$.

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 natural:

Definição A.2.17. Sejam μ e ν duas medidas num 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 Radón-Nikodym, afirma que se $\nu \ll \mu$ então a medida ν pode ser vista como a integral com respeito a μ de uma certa função mensurável ρ :

Teorema A.2.18 (Radón-Nikodym). *Se μ e ν são medidas finitas tais que $\nu \ll \mu$ então existe uma função mensurável $\rho : X \rightarrow [0, +\infty]$ tal que $\nu = \rho\mu$, ou seja, tal que*

$$\int \phi d\nu = \int \phi \rho d\mu \quad \text{para toda função mensurável limitada } \phi : X \rightarrow \mathbb{R}. \quad (\text{A.2.9})$$

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 Radón-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 Radón-Nikodym, obtemos:

Teorema A.2.20 (Decomposição de Lebesgue). *Dadas medidas finitas μ e ν , existe uma função mensurável $\rho : X \rightarrow [0, +\infty]$ 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 $f : M \rightarrow [0, +\infty)$ 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 $f : M \rightarrow \mathbb{R}$ uma função positiva integrável com respeito a uma medida finita μ . Então, dado qualquer número real $a > 0$,

$$\mu(\{x \in M : f(x) \geq a\}) \leq \frac{1}{a} \int_X f d\mu.$$

Em particular, se $\int |f| d\mu = 0$, então $\mu(\{x \in X : f(x) \neq 0\}) = 0$.

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 $\psi_1, \dots, \psi_N : M \rightarrow \mathbb{R}$ funções mensuráveis limitadas num espaço de probabilidade (M, \mathcal{B}, μ) . 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

$$\left| \int \psi d\mu - \sum_{j=1}^s \alpha_j \psi_i(x_j) \right| < \varepsilon \quad \text{para todo } i = 1, \dots, N.$$

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 $M \rightarrow \mathbb{R}$ 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 $|f| \leq |g|$ 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,

$$\liminf_{\delta \rightarrow 0} \left\{ \frac{m(B \cap A)}{m(B)} : B \text{ bola contida em } B(a, \delta) \text{ e contendo } a \right\} = 1 \quad (\text{A.2.10})$$

A.2.10. Seja \mathcal{P}_n , $n \geq 1$ uma sequência de partições enumeráveis de \mathbb{R}^d em subconjuntos mensuráveis. Suponha que o diâmetro $\text{diam } \mathcal{P}_n = \sup\{\text{diam } P : P \in \mathcal{P}_n\}$ converge para zero quando $n \rightarrow \infty$. Mostre que, dado qualquer conjunto mensurável $A \subset \mathbb{R}^d$ com medida de Lebesgue positiva, é possível escolher conjuntos $P_n \in \mathcal{P}_n$, $n \geq 1$ de tal forma que $m(A \cap P_n)/m(P_n) \rightarrow 1$ quando $n \rightarrow \infty$.

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

$$\mu(A) = \sum_{x_i \in A} p_i, \quad \nu(A) = \sum_{x_i \in A} q_i,$$

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 Radón-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 $f : X \rightarrow X$ 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 invertível (ou seja, 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 $J_\eta : X \rightarrow [0, +\infty]$ tal que

$$\int_{f(B_n)} \psi d\eta = \int_{B_n} (\psi \circ f) J_\eta d\eta$$

para toda função mensurável limitada $\psi : X \rightarrow \mathbb{R}$ 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 ρ^\pm e τ^\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

$$\frac{d\mu}{d\nu} = \lim_n \frac{d\hat{\mu}_n}{d\hat{\nu}_n} \quad \text{em } \nu\text{-quase todo ponto.}$$

A.3 Medidas em espaços métricos

Ao longo desta seção, a menos de indicação em contrário, suporemos que μ é uma medida de probabilidade boreliana (ou seja, definida na σ -álgebra de Borel) de um espaço métrico M .

Lembre que um *espaço métrico* é uma dupla (M, d) em que M é um conjunto e d é uma *distância* em M , ou seja, uma função $d : M \times M \rightarrow \mathbb{R}$ satisfazendo

1. $d(x, y) \geq 0$ para quaisquer x, y e vale a igualdade se, e somente se, $x = y$;
2. $d(x, y) = d(y, x)$ para quaisquer x, y ;
3. $d(x, y) \leq d(x, z) + d(z, y)$ para quaisquer x, y, z .

Representamos por $B(x, r)$ a *bola* de centro $x \in M$ e raio $r > 0$. Todo espaço métrico possui uma estrutura natural de espaço topológico, em que as bolas são bases de vizinhanças ou, equivalentemente, os abertos são os subconjuntos que contêm alguma bola centrada em qualquer dos seus pontos. Na direção oposta, diremos que um espaço topológico é *metrizável* se a sua topologia puder ser definida dessa forma, a partir de alguma função distância.

A.3.1 Medidas regulares

Um primeiro fato interessante é que qualquer probabilidade boreliana em um espaço métrico fica completamente determinada pelos valores que ela toma nos subconjuntos abertos (ou nos subconjuntos fechados) do espaço.

Definição A.3.1. Uma medida boreliana μ num espaço topológico é *regular* se para todo subconjunto mensurável B e todo $\varepsilon > 0$ existe um conjunto fechado F e um conjunto aberto A tais que $F \subset B \subset A$ e $\mu(F \setminus A) < \varepsilon$.

Proposição A.3.2. *Toda medida de probabilidade num espaço métrico é regular.*

Demonstração. Seja \mathcal{B}_0 a família dos subconjuntos borelianos B tais que para todo $\varepsilon > 0$ existe um fechado F e um aberto A satisfazendo $F \subset B \subset A$ e $\mu(F \setminus A) < \varepsilon$. Note que \mathcal{B}_0 contém todos os fechados de M . De fato, seja B um fechado qualquer e seja B^δ o conjunto (aberto) dos pontos cuja distância a B é

menor que δ . Pelo Teorema A.1.14, temos que $\mu(B^\delta \setminus B) \rightarrow 0$ quando $\delta \rightarrow 0$ e, portanto, podemos tomar $F = B$ e $A = B^\delta$ para algum $\delta > 0$ pequeno.

É imediato que a família \mathcal{B}_0 é fechada para a passagem ao complementar, isto é, $B^c \in \mathcal{B}_0$ sempre que $B \in \mathcal{B}_0$. Além disso, considere uma família enumerável B_n , $n = 1, 2, \dots$ qualquer de elementos de \mathcal{B}_0 e denote $B = \bigcup_{n=1}^{\infty} B_n$. Por hipótese, para todo $n \in \mathbb{N}$ e $\varepsilon > 0$, existe um fechado F_n e um aberto A_n satisfazendo $F_n \subset B_n \subset A_n$ e $\mu(A_n \setminus F_n) < \varepsilon/2^{n+1}$. A união $A = \bigcup_{n=1}^{\infty} A_n$ é um aberto e qualquer união finita $F = \bigcup_{n=1}^m F_n$ é um fechado. Fixemos o inteiro m suficientemente grande para que

$$\mu\left(\bigcup_{n=1}^{\infty} F_n \setminus F\right) < \varepsilon/2$$

(lembre o Teorema A.1.14). Então $F \subset B \subset A$ e

$$\mu(A \setminus F) \leq \sum_{n=1}^{\infty} \mu(A_n \setminus F_n) + \mu\left(\bigcup_{n=1}^{\infty} F_n \setminus F\right) < \sum_{n=1}^{\infty} \frac{\varepsilon}{2^{n+1}} + \frac{\varepsilon}{2} = \varepsilon.$$

Isto mostra que $B \in \mathcal{B}_0$. Deste jeito mostramos que \mathcal{B}_0 é uma σ -álgebra. Logo \mathcal{B}_0 contém todos os subconjuntos borelianos de M . \square

Uma consequência do resultado acima é que os valores que uma probabilidade μ toma nos fechados de M determinam μ completamente: se ν é outra probabilidade tal que $\mu(F) = \nu(F)$ para todo fechado F então $\mu = \nu$, isto é, $\mu(B) = \nu(B)$ para todo boreliano B . Passando ao complementar, obtemos que o mesmo vale para os conjuntos abertos.

A proposição que provaremos a seguir implica que os valores das integrais das funções contínuas limitadas também determinam a probabilidade completamente. De fato, isso vale mesmo para o conjunto (menor) das funções Lipschitz limitadas.

Lembre que uma aplicação $h : M \rightarrow N$ é *Lipschitz* se existe alguma constante $C > 0$ tal que $d(h(x), h(y)) \leq Cd(x, y)$ para todo $x, y \in M$. Se for necessário especificar a constante diremos que h é função C -Lipschitz. Mais geralmente, dizemos que h é *Hölder* se existem $C, \theta > 0$ tais que $d(h(x), h(y)) \leq Cd(x, y)^\theta$ para todo $x, y \in M$. Nesse caso, também dizemos que h é função θ -Hölder ou mesmo (C, θ) -Hölder.

Proposição A.3.3. *Se μ e ν são probabilidades em M tais que $\int \varphi d\mu = \int \varphi d\nu$ para toda função $\varphi : M \rightarrow \mathbb{R}$ Lipschitz limitada, então $\mu = \nu$.*

Demonstração. Usaremos o seguinte fato simples de topologia:

Lema A.3.4. *Dado qualquer subconjunto fechado F de M e dado $\delta > 0$, existe uma função Lipschitz $g_\delta : M \rightarrow [0, 1]$ tal que $g_\delta(x) = 1$ para todo $x \in F$ e $g_\delta(x) = 0$ para todo $x \in M$ tal que $d(x, F) \geq \delta$.*

Demonstração. Considere a função $h : \mathbb{R} \rightarrow [0, 1]$ dada da seguinte forma: $h(s) = 1$ se $s \leq 0$, $h(s) = 0$ se $s \geq 1$ e $h(s) = 1 - s$ se $0 \leq s \leq 1$. Defina

$$g : M \rightarrow [0, 1], \quad g(x) = h\left(\frac{1}{\delta}d(x, F)\right).$$

Note que g é Lipschitz, por ser composição de funções Lipschitz. As demais propriedades no lema são imediatas. \square

Agora podemos terminar a prova da Proposição A.3.3. Seja F um subconjunto fechado qualquer de M e, para cada $\delta > 0$, seja $g_\delta : M \rightarrow [0, 1]$ uma função como no lema acima. Por hipótese,

$$\int g_\delta d\mu = \int g_\delta d\nu \quad \text{para todo } \delta > 0.$$

Além disso, pelo teorema da convergência dominada (Teorema A.2.11),

$$\lim_{\delta \rightarrow 0} \int g_\delta d\mu = \mu(F) \quad \text{e} \quad \lim_{\delta \rightarrow 0} \int g_\delta d\nu = \nu(F).$$

Isto mostra que $\mu(F) = \nu(F)$ para todo fechado F , donde resulta que $\mu = \nu$. \square

Como observamos no Exemplo A.1.29, funções contínuas são automaticamente mensuráveis, relativamente à σ -álgebra de Borel. O resultado que vamos provar a seguir afirma que, mediante uma hipótese simples sobre o espaço métrico, vale uma espécie de recíproca: funções mensuráveis são contínuas se restringirmos a certos subconjuntos com medida quase total.

Um subconjunto de um espaço métrico (ou topológico) M é *denso* se ele intersecta todo aberto de M . Dizemos que o espaço é *separável*, se ele admite algum subconjunto enumerável denso. Para espaços métricos isso equivale a dizer que a topologia admite uma base enumerável de abertos (Exercício A.3.1).

Teorema A.3.5 (Lusin). *Seja $\varphi : M \rightarrow N$ uma aplicação mensurável com valores num espaço métrico separável N . Dado qualquer $\varepsilon > 0$ existe um conjunto fechado $F \subset M$ tal que $\mu(M \setminus F) < \varepsilon$ e a restrição de φ a F é contínua.*

Demonstração. Seja $\{x_n : n \in \mathbb{N}\}$ um subconjunto enumerável denso de N e, para cada $k \geq 1$, seja $B_{n,k}$ a bola de centro x_n e raio $1/k$. Fixe $\varepsilon > 0$. Pela Proposição A.3.2, para cada (n, k) podemos encontrar um conjunto aberto $A_{n,k} \subset M$ contendo $\varphi^{-1}(B_{n,k})$ e satisfazendo $\mu(A_{n,k} \setminus \varphi^{-1}(B_{n,k})) < \varepsilon/2^{n+k+1}$. Defina

$$E = \bigcap_{n,k=1}^{\infty} (\varphi^{-1}(B_{n,k}) \cup A_{n,k}^c).$$

Por um lado,

$$\mu(M \setminus E) \leq \sum_{n,k=1}^{\infty} \mu(A_{n,k} \setminus \varphi^{-1}(B_{n,k})) < \sum_{n,k=1}^{\infty} \frac{\varepsilon}{2^{n+k+1}} = \frac{\varepsilon}{2}.$$

Por outro lado, todo $\varphi^{-1}(B_{n,k})$ é subconjunto aberto de $\varphi^{-1}(B_{n,k}) \cup A_{n,k}^c$, já que o complementar é o conjunto fechado $A_{n,k}^c$. Consequentemente, $\varphi^{-1}(B_{n,k})$ é aberto em E para todo (n, k) . Isto mostra que a restrição de φ ao conjunto E é contínua. Para terminar a demonstração basta usar a Proposição A.3.2 mais uma vez para encontrar um conjunto fechado $F \subset E$ tal que $\mu(E \setminus F) < \varepsilon/2$. \square

A.3.2 Espaços métricos separáveis completos

Vamos agora discutir outra propriedade importante de medidas em espaços métricos que são ao mesmo tempo separáveis e completos. Lembre que esta última condição significa que toda sequência de Cauchy é convergente.

Definição A.3.6. Uma medida μ num espaço topológico é *justa* se para todo $\varepsilon > 0$ existe um subconjunto compacto K tal que $\mu(K^c) < \varepsilon$.

Como todo subconjunto fechado de um espaço métrico compacto também é compacto, segue imediatamente da Proposição A.3.2 que toda medida de probabilidade num espaço métrico compacto é justa. Mas, esta conclusão é ainda mais geral:

Proposição A.3.7. *Toda medida de probabilidade num espaço métrico separável completo é justa.*

Demonstração. Seja $\{p_k : k \in \mathbb{N}\}$ um subconjunto enumerável denso de M . Então, para cada $n \geq 1$, as bolas fechadas $\bar{B}(p_k, 1/n)$, $k \in \mathbb{N}$ formam uma cobertura enumerável de M . Dado $\varepsilon > 0$ e $n \geq 1$, fixemos $k(n) \geq 1$ de tal forma que o conjunto (fechado)

$$L_n = \bigcup_{k=1}^{k(n)} \bar{B}(p_k, 1/n)$$

satisfaça $\mu(L_n) > 1 - \varepsilon/2^n$. Tomemos $K = \bigcap_{n=1}^{\infty} L_n$. Note que K é fechado e

$$\mu(K^c) \leq \mu\left(\bigcup_{n=1}^{\infty} L_n^c\right) < \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = \varepsilon.$$

Resta verificar que K é compacto. Para isso basta mostrar que toda sequência $(x_i)_i$ em K admite alguma subsequência de Cauchy (uma vez que M é completo, segue que esta subsequência é convergente). Tal subsequência pode ser construída da seguinte forma. Como $x_i \in L_1$ para todo i , existe $l(1) \leq k(1)$ tal que o conjunto de índices

$$\mathcal{I}_1 = \{i \in \mathbb{N} : x_i \in B(p_{l(1)}, 1)\}$$

é infinito. Seja $i(1)$ o menor elemento de \mathcal{I}_1 . Em seguida, como $x_i \in L_2$ para todo i , existe $l(2) \leq k(2)$ tal que

$$\mathcal{I}_2 = \{i \in \mathcal{I}_1 : x_i \in B(p_{l(2)}, 1/2)\}$$

é infinito. Seja $i(2)$ o menor elemento de $\mathcal{I}_2 \setminus \{i(1)\}$. Repetindo este procedimento, construímos uma sequência decrescente \mathcal{I}_n de subconjuntos infinitos de \mathbb{N} , e uma sequência crescente $i(1) < i(2) < \dots < i(n) < \dots$ de números tais que $i(n) \in \mathcal{I}_n$ e todos os x_i , $i \in \mathcal{I}_n$ estão contidos numa mesma bola fechada de raio $1/n$. Em particular,

$$d(x_{i(a)}, x_{i(b)}) \leq 2/n \quad \text{para todo } a, b \geq n.$$

Isto mostra que a subsequência $(x_{i(n)})_n$ é realmente de Cauchy. \square

Corolário A.3.8. *Suponha que M é um espaço métrico separável completo e μ é uma medida de probabilidade em M . Para todo $\varepsilon > 0$ e todo conjunto boreliano $B \subset M$ existe um compacto $L \subset B$ tal que $\mu(B \setminus L) < \varepsilon$.*

Demonstração. Pela Proposição A.3.2 podemos encontrar um conjunto fechado $F \subset B$ tal que $\mu(B \setminus F) < \varepsilon/2$. Pelo Teorema A.3.5, existe um subconjunto compacto $K \subset M$ tal que $\mu(M \setminus K) < \varepsilon/2$. Tome $L = F \cap K$. Então L é compacto e $\mu(B \setminus L) < \varepsilon$. \square

Analogamente, quando o espaço métrico M é separável completo podemos melhorar o enunciado do teorema de Lusin, substituindo ‘fechado’ por ‘compacto’ na conclusão:

Teorema A.3.9 (Lusin). *Suponha que M é um espaço métrico separável completo e que μ é uma medida de probabilidade em M . Seja $\varphi : M \rightarrow N$ uma transformação mensurável com valores num espaço métrico separável N . Então dado qualquer $\varepsilon > 0$ existe um conjunto compacto $K \subset M$ tal que $\mu(M \setminus K) < \varepsilon$ e a restrição de φ a K é contínua.*

Concluimos com outro fato importante sobre medidas em espaços métricos separáveis completos. Uma medida μ diz-se *atômica* se existe algum ponto x tal que $\mu(\{x\}) > 0$. Tais pontos são chamados de *átomos*. Caso contrário, μ diz-se não atômica.

O próximo teorema afirma que toda medida não atômica de probabilidade num espaço métrico separável completo é equivalente à medida de Lebesgue no intervalo. Ele será demonstrado na Seção 8.5.

Teorema A.3.10. *Seja M um espaço métrico separável completo e seja μ uma medida não atômica de probabilidade em M . Então existe uma aplicação mensurável $\psi : M \rightarrow [0, 1]$ tal que ψ é uma bijeção com inversa mensurável, restrita a um subconjunto com medida total, e $\psi_*\mu$ é a medida de Lebesgue em $[0, 1]$.*

A.3.3 Espaço das funções contínuas

No que resta desta seção descreveremos algumas propriedades importantes do espaço vetorial $C^0(M)$ das funções contínuas (reais ou complexas), quando M é um espaço métrico compacto. As conclusões das seções anteriores permanecem

válidas, uma vez que todo espaço métrico compacto é separável e completo. Outro fato útil sobre espaços métricos compactos é que toda cobertura aberta admite algum *número de Lebesgue*, ou seja, algum número $\rho > 0$ tal que para todo $x \in M$ existe algum elemento da cobertura que contém a bola $B(x, \rho)$.

Consideremos o espaço $C^0(M)$ munido da norma da convergência uniforme:

$$\|\phi\| = \sup\{|\phi(x)| : x \in M\}.$$

Esta norma é completa e, portanto, define em $C^0(M)$ uma estrutura de espaço de Banach. Um funcional linear $\Phi : C^0(M) \rightarrow \mathbb{C}$ diz-se *positivo* se $\Phi(\varphi) \geq 0$ para toda função $\varphi \in C^0(M)$ tal que $\varphi(x) \geq 0$ para todo x . O teorema de Riesz-Markov (veja o Teorema 6.19 no livro de Rudin [Rud87]) mostra que os únicos funcionais lineares positivos em $C^0(M)$ são as integrais:

Teorema A.3.11 (Riesz-Markov). *Seja M um espaço métrico compacto. Seja $\Phi : C^0(M) \rightarrow \mathbb{C}$ um funcional linear positivo. Então existe uma única medida boreliana finita μ em M tal que*

$$\Phi(\varphi) = \int \varphi d\mu \quad \text{para toda } \varphi \in C^0(M).$$

Além disso, μ é uma probabilidade se e somente se $\Phi(1) = 1$.

O próximo resultado, que também é conhecido como teorema de Riesz-Markov, dá uma representação análoga para funcionais lineares contínuos em $C^0(M)$. Lembre que a *norma* de um funcional linear $\Phi : C^0(M) \rightarrow \mathbb{C}$ é definida por

$$\|\Phi\| = \sup \left\{ \frac{|\Phi(\varphi)|}{\|\varphi\|} : \varphi \neq 0 \right\} \quad (\text{A.3.1})$$

e que Φ é contínuo se, e somente se, a sua norma é finita.

Teorema A.3.12 (Riesz-Markov). *Seja M um espaço métrico compacto. Seja $\Phi : C^0(M) \rightarrow \mathbb{R}$ um funcional linear contínuo. Então existe uma única medida boreliana complexa μ em M tal que*

$$\Phi(\varphi) = \int \varphi d\mu \quad \text{para toda } \varphi \in C^0(M).$$

A norma $\|\mu\| = |\mu|(X)$ da medida μ coincide com a norma $\|\Phi\|$ do funcional Φ . Além disso, μ toma valores em $[0, \infty)$ se, e somente se, o operador Φ é positivo e μ toma valores em \mathbb{R} se, e somente se, $\Phi(\varphi) \in \mathbb{R}$ para toda função real φ .

Em outras palavras, este resultado afirma que o *dual* de $C^0(M)$ é isometricamente isomorfo a $\mathcal{M}(M)$. Os Teoremas A.3.11 e A.3.12 se estendem para espaços topológicos localmente compactos, com hipóteses adequadas sobre o comportamento das funções no infinito. Nesse contexto a medida μ (ainda) é regular mas não necessariamente finita.

Também usaremos o fato de que o espaço $C^0(M)$ tem subconjuntos enumeráveis densos (o Exercício A.3.6 é um caso particular):

Teorema A.3.13. *Se M é espaço métrico compacto então $C^0(M)$ é separável.*

Demonstração. Seja $\{a_k : k \in \mathbb{N}\}$ um subconjunto enumerável denso de M (todo espaço métrico compacto é separável). Para cada $k \in \mathbb{N}$, considere a função $f_k : M \rightarrow \mathbb{R}$ definida por $f_k(x) = d(x, a_k)$. Represente por \mathcal{A} o conjunto de todas as funções $f : M \rightarrow \mathbb{R}$ da forma

$$f = c + \sum_{k_1, \dots, k_n} c_{k_1, \dots, k_n} f_{k_1} \cdots f_{k_n} \quad (\text{A.3.2})$$

com $c \in \mathbb{R}$ e $c_{k_1, \dots, k_n} \in \mathbb{R}$ para quaisquer $k_1, \dots, k_n \in \mathbb{N}$. Observe que \mathcal{A} é uma *álgebra de funções*, ou seja, é fechada pelas operações de adição, multiplicação e multiplicação por número real. É claro que esta álgebra contém a função constante igual a 1. Além disso, \mathcal{A} *separa* os pontos de M , ou seja, para quaisquer pontos $x \neq y$ existe $f \in \mathcal{A}$ tal que $f(x) \neq f(y)$. Para ver isso, basta fixar $\varepsilon > 0$ tal que $d(x, y) > 2\varepsilon$, considerar $k \in \mathbb{N}$ tal que $d(x, a_k) < \varepsilon$ e tomar $f = f_k$. Note que $f(x) = d(x, a_k) < \varepsilon$ enquanto que, pela desigualdade triangular, $f(y) = d(y, a_k) \geq d(x, y) - d(x, a_k) > \varepsilon$. Portanto, a álgebra \mathcal{A} é separadora, tal como afirmamos. O teorema de Stone-Weierstrass (veja [DS57, Teorema 4.6.16]) afirma que toda sub-álgebra separadora do espaço das funções contínuas que contém a função 1 é densa em $C^0(M)$. Pelo que acabamos de dizer, isto vale para \mathcal{A} . Segue que o conjunto (enumerável) das funções da forma (A.3.2), com $c \in \mathbb{Q}$ e $c_{k_1, \dots, k_n} \in \mathbb{Q}$ também é denso em $C^0(M)$. \square

A.3.4 Exercícios

A.3.1. Seja M um espaço topológico métrizável. Mostre que todo ponto de M admite base enumerável de vizinhanças. Mostre que M é separável se, e somente se, ele admite base enumerável de abertos. Dê exemplos de espaços métricos separáveis e não separáveis.

A.3.2. Seja μ uma medida finita num espaço métrico M . Mostre que para todo fechado $F \subset M$ existe um conjunto finito ou enumerável $E \subset (0, \infty)$ tal que

$$\mu(\{x \in M : d(x, F) = r\}) = 0 \quad \text{para todo } r \in (0, \infty) \setminus E.$$

A.3.3. Seja μ uma medida finita num espaço métrico separável M . Conclua que para todo $\varepsilon > 0$ é possível encontrar uma partição enumerável de M em conjuntos mensuráveis com diâmetro menor que ε e cujo bordo tem medida nula.

A.3.4. Seja μ uma probabilidade em $[0, 1]$ e seja $\phi : [0, 1] \rightarrow [0, 1]$ a função dada por $\phi(x) = \mu([0, x])$. Verifique que ϕ é contínua se, e somente se, μ é não atômica e que ϕ é absolutamente contínua se, e somente se, μ é absolutamente contínua com relação à medida de Lebesgue.

A.3.5. Seja M um espaço métrico. Mostre que para toda função mensurável $\psi : M \rightarrow \mathbb{R}$ existe uma sequência $\psi_n : M \rightarrow \mathbb{R}$, $n \geq 1$ de funções uniformemente contínuas convergindo para ψ em todo ponto. Além disso, se ψ é limitada podemos escolher a sequência de tal forma que $\sup |\psi_n| \leq \sup |\psi|$ para todo n .

A.3.6. Mostre que o espaço $C^0([0, 1]^d)$ das funções contínuas, reais ou complexas, no cubo unitário compacto é separável, para todo $d \geq 1$.

A.4 Variedades diferenciáveis

Aqui vamos rever algumas noções e fatos básicos de topologia diferenciável e geometria riemanniana.

A.4.1 Variedades e aplicações diferenciáveis

Um *variedade diferenciável* de dimensão d é um espaço topológico M munido de um *atlas diferenciável* de dimensão d , ou seja, uma família de homeomorfismos $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$ tais que

1. cada U_α é aberto de M e cada X_α é aberto de \mathbb{R}^d e $M = \cup_\alpha U_\alpha$;
2. a aplicação $\varphi_\beta \circ \varphi_\alpha^{-1} : \varphi_\alpha(U_\alpha \cap U_\beta) \rightarrow \varphi_\beta(U_\alpha \cap U_\beta)$ é diferenciável, para quaisquer α e β tais que $U_\alpha \cap U_\beta \neq \emptyset$.

Mais geralmente, no lugar de \mathbb{R}^d podemos considerar qualquer espaço de Banach E . Então dizemos que M é uma variedade diferenciável *modelada* no espaço E .

A menos de menção em contrário, consideraremos apenas variedades em que o espaço M admite base enumerável de abertos e é *conexo*. Esta última propriedade significa que não existem subconjuntos simultaneamente abertos e fechados, à parte M e o conjunto vazio.

Os homeomorfismos φ_α são chamados *cartas locais*, ou *coordenadas locais*, e as transformações $\varphi_\beta \circ \varphi_\alpha^{-1}$ são chamadas *mudanças de coordenadas*. Permutando os papéis de α e β , vemos que a inversa $(\varphi_\beta \circ \varphi_\alpha^{-1})^{-1} = \varphi_\alpha \circ \varphi_\beta^{-1}$ também é diferenciável. Portanto, a definição de variedade requer que as mudanças de coordenadas sejam difeomorfismos entre abertos do espaço euclidiano.

Seja $r \in \mathbb{N} \cup \{\infty\}$. Se toda mudança de coordenadas é de classe C^r (isto é, se todas as suas derivadas parciais até a ordem r existem e são contínuas), diremos que a variedade M (e o atlas $\{\varphi_\alpha : U_\alpha \rightarrow X_\alpha\}$) é de classe C^r . Claro que toda variedade de classe C^r também é de classe C^s para qualquer $s \leq r$.

Exemplo A.4.1. Os seguintes espaços são variedades C^∞ de dimensão d :

Espaço euclidiano \mathbb{R}^d : considere o atlas formado por uma única aplicação, a aplicação identidade $\mathbb{R}^d \rightarrow \mathbb{R}^d$;

Esfera $S^d = \{(x_0, x_1, \dots, x_d) \in \mathbb{R}^{d+1} : x_0^2 + x_1^2 + \dots + x_d^2 = 1\}$: considere o atlas formado pelas duas *projeções estereográficas*:

$$\begin{aligned} S^d \setminus \{(1, 0, \dots, 0)\} &\rightarrow \mathbb{R}^d, & (x_0, x_1, \dots, x_d) &\mapsto (x_1, \dots, x_d)/(1 - x_0) \\ S^d \setminus \{(-1, 0, \dots, 0)\} &\rightarrow \mathbb{R}^d, & (x_0, x_1, \dots, x_d) &\mapsto (x_1, \dots, x_d)/(1 + x_0). \end{aligned}$$

Toro $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$: considere o atlas formado pelas inversas das aplicações $g_z : (0, 1)^d \rightarrow \mathbb{T}^d$, definidas por $g_z(x) = z + x \pmod{\mathbb{Z}^d}$ para cada $z \in \mathbb{R}^d$.

Exemplo A.4.2 (Variedades de Grassmann). Dados $0 \leq k \leq d$, represente por $\text{Gr}(k, d)$ o conjunto dos subespaços vetoriais do espaço \mathbb{R}^d com dimensão k . Para cada $j_1 < \dots < j_k$, represente por $\text{Gr}(k, d, j_1, \dots, j_k)$ o conjunto dos elementos de $\text{Gr}(k, d)$ que são transversais a $\{(x_j)_j \in \mathbb{R}^d : x_{j_1} = \dots = x_{j_k} = 0\}$. Para todo $V \in \text{Gr}(k, d, j_1, \dots, j_k)$ existe uma única matriz $(u_{i,j})_{i,j}$ com $(d-k)$ linhas e k colunas tal que

$$V = \{(x_j)_j \in \mathbb{R}^d : x_i = u_{i,j_1}x_{j_1} + \dots + u_{i,j_k}x_{j_k} \text{ para todo } i \notin \{j_1, \dots, j_k\}\}.$$

As aplicações $\text{Gr}(k, d, j_1, \dots, j_k) \rightarrow \mathbb{R}^{(d-k)k}$, que a cada V no seu domínio associam a respectiva matriz $(u_{i,j})_{i,j}$ constituem um atlas de classe C^∞ em $\text{Gr}(k, d)$. Portanto, cada $\text{Gr}(k, d)$ é uma variedade C^∞ de dimensão $(d-k)k$.

Seja M uma variedade de dimensão d e seja $\mathcal{A} = \{\varphi_\alpha : U_\alpha \rightarrow X_\alpha\}$ o respectivo atlas. Seja S um subconjunto de M . Dizemos que S é uma *subvariedade* de dimensão $k < d$ se existir algum atlas $\mathcal{B} = \{\psi_\beta : V_\beta \rightarrow Y_\beta\}$ de M tal que

- (a) \mathcal{A} e \mathcal{B} são *compatíveis*: as mudanças de coordenadas $\psi_\beta \circ \varphi_\alpha^{-1}$ e $\varphi_\alpha \circ \psi_\beta^{-1}$ são diferenciáveis nos seus domínios, para todo α e todo β ;
- (b) para todo β , a carta local ψ_β envia $V'_\beta = V_\beta \cap S$ sobre um aberto Y'_β de $\mathbb{R}^k \times \{0^{d-k}\}$.

Identificando $\mathbb{R}^k \times \{0^{d-k}\} \simeq \mathbb{R}^k$, vem que a família das restrições $\psi_\beta : V'_\beta \rightarrow Y'_\beta$ constitui um atlas para S ; portanto, S é uma variedade de dimensão k . Se M é variedade de classe C^r e os atlas \mathcal{A} e \mathcal{B} são C^r -compatíveis, ou seja, se todas as mudanças de coordenadas em (a) são de classe C^r , então S é (sub)variedade de classe C^r .

Diremos que uma aplicação $f : M \rightarrow N$ entre duas variedades é *diferenciável* se

$$\psi_\beta \circ f \circ \varphi_\alpha^{-1} : \varphi_\alpha(U_\alpha \cap f^{-1}(V_\beta)) \rightarrow \psi_\beta(V_\beta \cap f(U_\alpha)) \quad (\text{A.4.1})$$

é aplicação diferenciável para toda carta local $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$ de M e toda carta local $\psi_\beta : V_\beta \rightarrow Y_\beta$ de N com $f(U_\alpha) \cap V_\beta \neq \emptyset$. Além disso, diremos que f é *de classe C^r* se M e N são variedades de classe C^r e toda aplicação $\psi_\beta \circ f \circ \varphi_\alpha^{-1}$ em (A.4.1) é de classe C^r . Chamamos *difeomorfismo* a toda bijeção $f : M \rightarrow N$ tal que tanto f quanto f^{-1} são diferenciáveis; se as duas aplicações forem de classe C^r , diremos que o difeomorfismo é *de classe C^r* .

Seja $C^r(M, N)$ o espaço das aplicações de classe C^r entre duas variedades M e N . Vamos introduzir neste espaço uma certa topologia, chamada *topologia C^r* , que expressa a ideia de que duas aplicações estão próximas se elas estão uniformemente próximas e o mesmo vale para as suas derivadas até a ordem r . A definição pode ser dada num contexto muito amplo (veja a Seção 2.1 de Hirsch [Hir94]), mas nós nos restringiremos ao caso em que M e N são compactas. Então a topologia C^r pode ser definida da seguinte forma.

Fixe famílias finitas de cartas locais $\varphi_i : U_i \rightarrow X_i$ de M e $\psi_j : V_j \rightarrow Y_j$ de N , tais que $\cup_i U_i = M$ e $\cup_j V_j = N$. Seja $\delta > 0$ um número de Lebesgue para a cobertura aberta $\{U_i \cap f^{-1}(V_j)\}$ de M . Para cada par (i, j) tal que

$U_i \cap f^{-1}(V_j) \neq \emptyset$, seja $K_{i,j}$ o conjunto dos pontos cuja distância ao complementar de $U_i \cap f^{-1}(V_j)$ é maior ou igual que δ . Então $K_{i,j}$ é um compacto contido em $U_i \cap f^{-1}(V_j)$ e a união $\cup_{i,j} K_{i,j}$ é todo o M . Considere

$$\mathcal{U}(f) = \{g \in C^r(M, N) : g(K_{i,j}) \subset V_j \text{ para quaisquer } i, j\}.$$

Claro que $f \in \mathcal{U}(f)$. Para cada $g \in \mathcal{U}(f)$ e cada par (i, j) tal que $K_{i,j}$ é não vazio, represente por $g_{i,j}$ a restrição de $\psi_j \circ g \circ \varphi_i^{-1}$ ao conjunto $\varphi_i(K_{i,j})$. Para cada $r \in \mathbb{N}$ e $\varepsilon > 0$, defina

$$\mathcal{U}^r(f, \varepsilon) = \{g \in \mathcal{U}(f) : \sup_{s,x,i,j} \|D^s f_{i,j}(x) - D^s g_{i,j}(x)\| < \varepsilon\}, \quad (\text{A.4.2})$$

onde o supremo é sobre todo $s \in \{1, \dots, r\}$, todo $x \in \varphi_i(K_{i,j})$ e todo par (i, j) tal que $K_{i,j} \neq \emptyset$. Por definição, a família $\{\mathcal{U}^r(f, \varepsilon) : \varepsilon > 0\}$ é uma base de vizinhanças de cada $f \in C^r(M, N)$ relativamente à topologia C^r . Também por definição, a família $\{\mathcal{U}^r(f, \varepsilon) : \varepsilon > 0 \text{ e } r \in \mathbb{N}\}$ é uma base de vizinhanças de $f \in C^\infty(M, N)$ relativamente à topologia C^∞ .

A topologia C^r tem muito boas propriedades: em particular, ela admite base enumerável de abertos e é *completamente metrizável*, ou seja, ela é gerada por alguma distância completa. Uma consequência interessante é que $C^r(M, N)$ é um *espaço de Baire*: toda interseção de uma família enumerável de subconjuntos abertos densos é densa no espaço. O conjunto $\text{Difeo}^r(M)$ dos difeomorfismos de classe C^r em M é um subconjunto aberto de $C^r(M, M)$ relativamente à topologia C^r .

A.4.2 Espaço tangente e aplicação derivada

Seja M uma variedade. Para cada $p \in M$, considere o conjunto $\mathcal{C}(p)$ de todas as curvas $c : I \rightarrow M$, onde I é um intervalo aberto contendo $0 \in \mathbb{R}$, tais que $c(0) = p$ e c é diferenciável no ponto 0, ou seja, a aplicação $\varphi_\alpha \circ c$ é diferenciável no ponto 0, para toda carta local $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$ com $p \in U_\alpha$. Dizemos que duas curvas $c_1, c_2 \in \mathcal{C}(p)$ são *equivalentes* se $(\varphi_\alpha \circ c_1)'(0) = (\varphi_\alpha \circ c_2)'(0)$ para toda carta local $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$ com $p \in U_\alpha$. De fato, se a igualdade vale para uma carta local então ela vale para qualquer outra. Representaremos por $[c]$ a classe de equivalência de qualquer curva $c \in \mathcal{C}(p)$.

O *espaço tangente* à variedade M no ponto p é o conjunto, que representaremos por $T_p M$, de tais classes de equivalência. Dada qualquer carta local $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$ com $p \in U_\alpha$, a aplicação

$$D\varphi_\alpha(p) : T_p M \rightarrow \mathbb{R}^d, \quad [c] \mapsto (\varphi_\alpha \circ c)'(0)$$

está bem definida e é um bijeção. Podemos usar esta bijeção para identificar $T_p M$ com \mathbb{R}^d . Desta forma, o espaço tangente adquire uma estrutura de espaço vetorial, transportada de \mathbb{R}^d por $D\varphi_\alpha(p)$. Embora a identificação $D\varphi_\alpha(p)$ depende da escolha da carta local, a estrutura de espaço vetorial em $T_p M$ não depende. De fato, dada qualquer outra carta local $\varphi_\beta : U_\beta \rightarrow X_\beta$ com $p \in U_\beta$, a respectiva aplicação $D\varphi_\beta(p)$ está dada por

$$D\varphi_\beta(p) = D(\varphi_\beta \circ \varphi_\alpha^{-1})(\varphi_\alpha(p)) \circ D\varphi_\alpha(p).$$

Como $D(\varphi_\beta \circ \varphi_\alpha^{-1})(\varphi_\alpha(p))$ é um isomorfismo linear, segue que as estruturas de espaço vetorial transportadas por $D\varphi_\alpha(p)$ e $D\varphi_\beta(p)$ do espaço euclidiano para T_pM coincidem, tal como afirmamos.

Se $f : M \rightarrow N$ é uma aplicação diferenciável, a sua *derivada* num ponto $p \in M$ é a aplicação linear $Df(p) : T_pM \rightarrow T_{f(p)}N$ definida por

$$Df(p) = D\psi_\beta(f(p))^{-1} \circ D(\psi_\beta \circ f \circ \varphi_\alpha^{-1})(\varphi_\alpha(p)) \circ D\varphi_\alpha(p),$$

onde $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$ é carta local de M com $p \in U_\alpha$ e $\psi_\beta : V_\beta \rightarrow Y_\beta$ é carta local de N com $f(p) \in V_\beta$. A definição não depende da escolha destas cartas locais.

O *fibrado tangente* a M é a união $TM = \cup_{p \in M} T_pM$ de todos os espaços tangentes a M . Observe que esta união é disjunta. Para cada carta local $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$, considere $TU_\alpha M = \cup_{p \in U_\alpha} T_pM$ e a aplicação

$$D\varphi_\alpha : TU_\alpha M \rightarrow X_\alpha \times \mathbb{R}^d$$

que associa a cada $[c] \in TU_\alpha M$ o par $((\varphi_\alpha \circ c)(0), (\varphi_\alpha \circ c)'(0)) \in X_\alpha \times \mathbb{R}^d$. Consideramos em TM a (única) topologia que torna cada $D\varphi_\alpha$ um homeomorfismo. Supondo que o atlas $\{\varphi_\alpha : U_\alpha \rightarrow X_\alpha\}$ de M é de classe C^r então, a mudança de coordenadas

$$D\varphi_\beta \circ D\varphi_\alpha^{-1} : \varphi_\alpha(U_\alpha \cap U_\beta) \times \mathbb{R}^d \rightarrow \varphi_\beta(U_\alpha \cap U_\beta) \times \mathbb{R}^d$$

é aplicação de classe C^{r-1} para quaisquer α e β tais que $U_\alpha \cap U_\beta \neq \emptyset$. Portanto, o fibrado tangente TM está munido com a estrutura de variedade de classe C^{r-1} e dimensão $2d$.

A *derivada* de uma transformação diferenciável $f : M \rightarrow N$ é a aplicação $Df : TM \rightarrow TN$ cuja restrição a cada espaço tangente T_pM está dada por $Df(p)$. Se f é de classe C^r então Df é de classe C^{r-1} , relativamente à estrutura de variedade nos fibrados tangentes TM e TN introduzida no parágrafo anterior. Por exemplo, a projeção canônica $\pi : TM \rightarrow M$ que associa a cada $v \in TM$ o único ponto $p \in M$ tal que $v \in T_pM$ é uma aplicação de classe C^{r-1} (Exercício A.4.8).

Um *campo de vetores* numa variedade M é uma aplicação que associa a cada ponto $p \in M$ um elemento $X(p)$ do espaço tangente T_pM , ou seja, uma aplicação $X : M \rightarrow TM$ tal que $\pi \circ X = \text{id}$. Dizemos que o campo de vetores é de classe C^k , com $k \leq r - 1$, se essa aplicação for de classe C^k .

Suponhamos que $k \geq 1$. Então o teorema de existência e unicidade de soluções de equações diferenciais garante que, para todo ponto $p \in M$, existe uma única curva $c_p : I_p \rightarrow M$ tal que $c_p(0) = p$ e $c_p'(t) = X(c_p(t))$ para todo $t \in I_p$ e I_p é intervalo aberto maximal com esta propriedade. Se M é compacta, tem-se $I_p = \mathbb{R}$ para qualquer $p \in M$. Além disso, as transformações $f^t : M \rightarrow M$ definidas por $f^t(p) = c_p(t)$ são difeomorfismos de classe C^k , com $f^0 = \text{id}$ e $f^s \circ f^t = f^{s+t}$ para quaisquer $s, t \in \mathbb{R}$. A família $\{f^t : t \in \mathbb{R}\}$ é chamada *fluxo* do campo de vetores X .

A.4.3 Transversalidade

O resultado que vamos enunciar a seguir é uma das principais ferramentas para construir novas variedades. Dizemos que $y \in N$ é *valor regular* de uma transformação diferenciável $f : M \rightarrow N$ se a derivada $Df(x) : T_xM \rightarrow T_yN$ é sobrejetiva para todo $x \in f^{-1}(y)$. Note que esta condição é satisfeita, automaticamente, se y não está na imagem de f . Por outro lado, para que algum ponto $y \in f(M)$ seja um valor regular de f é necessário que $\dim M \geq \dim N$.

Teorema A.4.3. *Seja $f : M \rightarrow N$ uma aplicação de classe C^r e seja $y \in f(M)$ um valor regular de f . Então $f^{-1}(y)$ é uma subvariedade (não necessariamente conexa) de classe C^r de M , com dimensão igual a $\dim M - \dim N$.*

Exemplo A.4.4. Para qualquer $d \geq 1$, o espaço das matrizes quadradas de dimensão d com coeficientes reais é isomorfo ao espaço euclidiano $\mathbb{R}^{(d^2)}$ e, portanto, é uma variedade de dimensão d^2 e classe C^∞ . O chamado *grupo linear* $GL(d, \mathbb{R})$ das matrizes invertíveis é um aberto desse espaço e, portanto, também é uma variedade de dimensão d^2 e classe C^∞ . A função $\det : GL(d, \mathbb{R}) \rightarrow \mathbb{R}$ que associa a cada matriz o seu determinante é de classe C^∞ e $y = 1$ é um valor regular (veja o Exercício A.4.5). Usando o Teorema A.4.3, segue que o *grupo linear especial* $SL(d, \mathbb{R})$, formado pelas matrizes com determinante igual a 1, é uma subvariedade C^∞ de $GL(d, \mathbb{R})$, com dimensão igual a $d^2 - 1$.

É possível generalizar bastante o Teorema A.4.3, usando a noção de transversalidade. Dizemos que uma subvariedade S de N é *transversal* a f se

$$Df(x)(T_xM) + T_{f(x)}S = T_{f(x)}N \quad \text{para todo } x \in f^{-1}(S). \quad (\text{A.4.3})$$

Por exemplo, se S é uma subvariedade de dimensão zero, ou seja, se ela está formada por um único ponto, então S é transversal a f se, e somente se, esse ponto é um valor regular de f . Portanto, o seguinte resultado generaliza o Teorema A.4.3:

Teorema A.4.5. *Seja $f : M \rightarrow N$ uma aplicação de classe C^r e seja S uma subvariedade de classe C^r de N transversal a f . Então $f^{-1}(S)$ é uma subvariedade (não necessariamente conexa) de classe C^r de M , com dimensão igual a $\dim M - \dim N + \dim S$.*

O próximo teorema afirma que, para toda aplicação diferenciável $f : M \rightarrow N$ de classe suficientemente elevada, "quase todos" os pontos $y \in N$ são valores regulares. Dizemos que um conjunto $X \subset N$ é *residual* se ele contém alguma interseção enumerável de subconjuntos abertos e densos. Todo conjunto residual uma variedade é denso na variedade (porque variedades são espaços de Baire). Dizemos que um conjunto $Z \subset N$ tem *volume zero* se para toda carta local $\psi_\beta : V_\beta \rightarrow Y_\beta$ a imagem $\psi_\beta(Z \cap V_\beta)$ é um subconjunto do espaço euclidiano com volume zero, ou seja, ele admite coberturas por bolas tais que a soma dos volumes dessas bolas é arbitrariamente pequena.

Teorema A.4.6 (Sard). *Suponha que $f : M \rightarrow N$ é aplicação de classe C^r com $r > \max\{0, \dim M - \dim N\}$. Então o conjunto dos pontos regulares é um subconjunto residual e o seu complementar tem volume zero.*

A.4.4 Variedades riemannianas

Uma *métrica riemanniana* numa variedade M é uma aplicação que associa a cada ponto $p \in M$ um produto interno no espaço tangente $T_p M$, ou seja, uma aplicação bilinear simétrica

$$\cdot_p : T_p M \times T_p M \rightarrow \mathbb{R}$$

tal que $v \cdot_p v > 0$ para todo vetor não nulo $v \in T_p M$. Como parte da definição, também pedimos que este produto interno varia diferenciavelmente com o ponto p , no seguinte sentido. Considere qualquer carta local $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$ de M . Conforme explicado anteriormente, para cada $p \in U_\alpha$ podemos identificar $T_p M$ com \mathbb{R}^d , via $D\varphi_\alpha(p)$, e então podemos considerar \cdot_p como sendo um produto interno no espaço euclidiano. Seja e_1, \dots, e_d uma base de \mathbb{R}^d . Então, pedimos que as funções $g_{\alpha,i,j}(p) = e_i \cdot_p e_j$ sejam diferenciáveis para todo par (i, j) e para qualquer escolha da carta local e da base.

Chamamos *variedade riemanniana* a qualquer variedade munida de uma métrica riemanniana. Toda subvariedade S de uma variedade riemanniana M herda uma estrutura de variedade riemanniana, dada pela restrição do produto interno \cdot_p de M ao subespaço tangente $T_p S$ de cada ponto $p \in S$. Toda variedade compacta admite (infinitas) métricas riemannianas. Isso segue do teorema de Whitney (veja a Seção 1.3 de Hirsch [Hir94]), o qual afirma que toda variedade compacta pode ser realizada como subvariedade de algum espaço euclidiano. Na verdade, isso permanece válido na classe bem mais ampla das variedades *paracompactas* (que não definiremos aqui): toda variedade paracompacta de dimensão d pode ser realizada como subvariedade de \mathbb{R}^{2d} . Em particular, tais variedades são sempre metrizáveis.

A partir da métrica riemanniana, podemos definir o *comprimento* de qualquer curva diferenciável $\gamma : [a, b] \rightarrow M$, mediante

$$\text{comp}(\gamma) = \int_a^b \|\gamma'(t)\|_{\gamma(t)} dt, \quad \text{onde } \|v\|_p = (v \cdot_p v)^{1/2}.$$

Por sua vez, isso nos permite definir na variedade M a seguinte *distância associada à métrica riemanniana*: a distância $d(p, q)$ entre dois pontos $p, q \in M$ é o ínfimo dos comprimentos de todas as curvas diferenciáveis ligando os dois pontos. Dizemos que uma curva diferenciável $\gamma : [a, b] \rightarrow M$ é *minimizante* se ela realiza a distância entre os seus pontos extremos, isto é,

$$\text{comp}(\gamma) = d(\gamma(a), \gamma(b)).$$

Quaisquer dois pontos $p, q \in M$ estão ligados por alguma curva minimizante; em outras palavras, o ínfimo na definição de $d(p, q)$ sempre é realizado.

Uma curva diferenciável $\gamma : I \rightarrow M$, definida em um intervalo aberto I , é chamada *geodésica* se ela é localmente minimizante, no seguinte sentido: para todo $c \in I$ existe $\delta > 0$ tal que a restrição de γ ao intervalo $[c - \delta, c + \delta]$ é minimizante. Toda curva minimizante é geodésica, mas a recíproca não é verdadeira: por exemplo, as geodésicas da esfera S^2 são os círculos máximos,

mas curvas fechadas nunca são minimizantes. Um fato importante é que se γ é geodésica então a norma $\|\gamma'(t)\|_{\gamma(t)}$ é constante em I . A teoria das equações diferenciais permite mostrar que para todo $p \in M$ e todo $v \in T_pM$ existe uma única geodésica $\gamma_{p,v} : I_{p,v} \rightarrow M$ tal que $\gamma_{p,v}(0) = p$ e $\gamma'_{p,v}(0) = v$ e $I_{p,v}$ é intervalo maximal tal que $\gamma_{p,v}$ é localmente minimizante.

Se a variedade M é compacta então $I_{p,v} = \mathbb{R}$ para todo $p \in M$ e todo $v \in T_pM$. Então, definimos a *aplicação exponencial* em cada ponto $p \in M$:

$$\exp_p : T_pM \rightarrow M, \quad v \mapsto \gamma_{p,v}(1).$$

Esta é uma aplicação diferenciável e a sua derivada em $v = 0$ é a aplicação identidade no espaço tangente T_pM . Também definimos o *fluxo geodésico* no fibrado tangente

$$f^t : TM \rightarrow TM, \quad (p, v) \mapsto (\gamma_{p,v}(t), \gamma'_{p,v}(t)).$$

Frequentemente, considera-se a restrição do fluxo geodésico ao *fibrado tangente unitário* $T^1M = \{(p, v) \in TM : \|v\|_p = 1\}$. Isto está bem definido pois, conforme mencionamos anteriormente, a norma do vetor velocidade de toda geodésica é constante.

A.4.5 Exercícios

A.4.1. Verifique que todo conjunto X com cardinalidade de \mathbb{R} pode ser munido com uma estrutura de variedade diferenciável de classe C^∞ e dimensão d , para qualquer $d \geq 1$.

A.4.2. Considere as variedades diferenciáveis $M = (\mathbb{R}, \mathcal{A})$ e $N = (\mathbb{R}, \mathcal{B})$, onde \mathcal{A} é o atlas formado pela aplicação $\phi(x) = x$ e \mathcal{B} é o atlas formado pela aplicação $\psi(x) = x^3$. A aplicação $f : M \rightarrow N$ definida por $f(x) = x$ é um difeomorfismo entre estas variedades?

A.4.3. Um espaço topológico é *conexo por caminhos* se quaisquer dois pontos são ligados por alguma curva contínua. Mostre que toda variedade (conexa) é conexa por caminhos.

A.4.4. Para cada $d \geq 2$, chamamos *espaço projetivo de dimensão d* ao conjunto \mathbb{P}^d dos subespaços de dimensão 1 de \mathbb{R}^{d+1} . Equivalentemente, \mathbb{P}^d é o espaço quociente de $\mathbb{R}^{d+1} \setminus \{0\}$ relativamente à relação de equivalência definida por:

$$(x_0, \dots, x_d) \sim (y_0, \dots, y_d) \Leftrightarrow \text{existe } c \neq 0 \text{ tal que } x_i = cy_i \text{ para todo } i.$$

Mostre que a família das transformações $\varphi_i : U_i \rightarrow \mathbb{R}^d$, $i = 0, \dots, d$ dadas por $U_i = \{[x_0 : \dots : x_d] \in \mathbb{P}^d : x_i \neq 0\}$ (onde $[x_0 : \dots : x_d]$ representa a classe de equivalência de (x_0, \dots, x_d)) e

$$\varphi_i([x_0 : \dots : x_d]) = \left(\frac{x_0}{x_i}, \dots, \frac{x_{i-1}}{x_i}, \frac{x_{i+1}}{x_i}, \dots, \frac{x_d}{x_i} \right),$$

constitui um atlas de classe C^∞ e dimensão d para \mathbb{P}^d .

A.4.5. Verifique as afirmações do Exemplo A.4.4.

A.4.6. Considere em $\mathbb{R}_+ = \{x \in \mathbb{R} : x > 0\}$ a métrica riemanniana definida por $u \cdot_x v = uv/x^2$. Calcule a distância $d(a, b)$ entre dois pontos $a, b \in \mathbb{R}_+$

A.4.7. Sejam M e N subvariedades de \mathbb{R}^{m+n} , com $\dim M = m$ e $\dim N = n$. Mostre que existe um conjunto $Z \subset \mathbb{R}^{m+n}$ com volume zero tal que, para todo v no complementar de Z , o translado $M + v$ é *transversal* a N :

$$T_x(M + v) + T_x N = \mathbb{R}^d \quad \text{para todo } x \in (M + v) \cap N.$$

A.4.8. Mostre que se M é variedade de classe C^r então a projeção canônica $\pi : TM \rightarrow M$ é uma aplicação de classe C^{r-1} .

A.5 Espaços $L^p(\mu)$

Nesta seção vamos definir espaços formados por funções que possuem propriedades de integrabilidade especiais. Trata-se de espaços de Banach, ou seja, espaços vetoriais normados que são completos com respeito a essa norma. Também enunciaremos algumas propriedades das normas destes espaços. Ao longo da seção (X, \mathcal{B}, μ) será sempre um espaço de medida.

A.5.1 Espaços $L^p(\mu)$ com $1 \leq p < \infty$

Dado qualquer $p \in [1, \infty)$, dizemos que $f : X \rightarrow \mathbb{C}$ é uma *função p -integrável* com relação a μ se a função $|f|^p$ é integrável com relação a μ .

Definição A.5.1. Denotamos por $L^p(\mu)$ o conjunto das funções complexas p -integráveis com relação a μ , módulo a relação de equivalência que identifica quaisquer funções que são iguais em μ -quase todo ponto.

Note que se μ é finita, o que ocorrerá frequentemente nos nossos exemplos, todas as funções mensuráveis limitadas estão em $L^p(\mu)$, uma vez que

$$\int |f|^p d\mu \leq (\sup |f|)^p m(X).$$

Em particular, se X é espaço topológico compacto então toda função contínua está em $L^p(\mu)$. Em outras palavras, o espaço $C^0(X)$ das funções contínuas está contido em $L^p(\mu)$ para todo p .

Para cada função $f \in L^p(\mu)$, definimos a *norma L^p* de f :

$$\|f\|_p = \left(\int |f|^p d\mu \right)^{\frac{1}{p}}.$$

O próximo teorema mostra que $L^p(\mu)$ fica munido da estrutura de um espaço de Banach:

Teorema A.5.2. *O conjunto $L^p(\mu)$ é um espaço vetorial complexo. Além disso, $\|f\|_p$ é uma norma em $L^p(\mu)$ e essa norma é completa.*

A parte mais interessante da demonstração deste teorema é provar a desigualdade triangular que, neste caso, é conhecida como *desigualdade de Minkowski*:

Teorema A.5.3 (Desigualdade de Minkowski). *Sejam $f, g \in L^p(\mu)$. Então:*

$$\left(\int |f + g|^p d\mu \right)^{\frac{1}{p}} \leq \left(\int |f|^p d\mu \right)^{\frac{1}{p}} + \left(\int |g|^p d\mu \right)^{\frac{1}{p}}.$$

Nos Exercícios A.5.2 e A.5.5 propomos que o leitor prove a desigualdade de Minkowski e, a partir daí, complete a prova do Teorema A.5.2.

A.5.2 Produto interno em $L^2(\mu)$

O caso $p = 2$ merece atenção especial da nossa parte. A razão é que neste caso a norma $\|\cdot\|_2$ definida anteriormente vem de um produto interno (hermitiano), a saber:

$$f \cdot g = \int f \bar{g} d\mu. \quad (\text{A.5.1})$$

Segue das propriedades da integral que esta expressão realmente define um produto interno em $L^2(\mu)$. Este produto se relaciona com a norma $\|\cdot\|_2$ por:

$$\|f\|_2 = (f \cdot f)^{1/2}.$$

Em particular, vale a desigualdade de Cauchy-Schwarz:

Teorema A.5.4 (Desigualdade de Cauchy-Schwarz). *Dadas $f, g \in L^2(\mu)$, então $f\bar{g} \in L^1(\mu)$ e vale a desigualdade:*

$$\left| \int f \bar{g} d\mu \right| \leq \left(\int |f|^2 d\mu \right)^{1/2} \left(\int |g|^2 d\mu \right)^{1/2}.$$

Esta desigualdade tem a seguinte consequência interessante. Suponha que a medida μ é finita e seja $f \in L^2(\mu)$ qualquer. Então, tomando $g \equiv 1$,

$$\int |f| d\mu = \int |f\bar{g}| d\mu \leq \left(\int |f|^2 d\mu \right)^{1/2} \left(\int 1 d\mu \right)^{1/2} < \infty. \quad (\text{A.5.2})$$

Isto mostra que toda função em $L^2(\mu)$ está em $L^1(\mu)$. De fato, quando a medida μ é finita tem-se $L^p(\mu) \subset L^q(\mu)$ sempre que $p \geq q$ (Exercício A.5.3).

Para valores de p quaisquer, vale a seguinte generalização da desigualdade de Cauchy-Schwarz:

Teorema A.5.5 (Desigualdade de Hölder). *Dado $1 < p < \infty$ considere q definido pela relação $\frac{1}{p} + \frac{1}{q} = 1$. Então, para toda $f \in L^p(\mu)$ e $g \in L^q(\mu)$ temos que $f\bar{g} \in L^1(\mu)$ e vale a desigualdade:*

$$\int |f\bar{g}| d\mu \leq \left(\int |f|^p d\mu \right)^{\frac{1}{p}} \left(\int |g|^q d\mu \right)^{\frac{1}{q}}.$$

A.5.3 Funções essencialmente limitadas

Em seguida vamos estender a definição dos espaços $L^p(\mu)$ para o caso $p = \infty$. Para isso precisamos da seguinte noção. Dizemos que uma função $f : X \rightarrow \mathbb{C}$ é *essencialmente limitada* com relação a μ se existe uma constante $K > 0$ tal que $|f(x)| \leq K$ em μ -quase todo ponto. Nesse caso, chamamos *supremo essencial* de f , e denotamos por $\text{supess}_\mu(f)$, o ínfimo dos valores de K satisfazendo essa condição.

Definição A.5.6. Denotamos por $L^\infty(\mu)$ o conjunto das funções complexas essencialmente limitadas com relação a μ , módulo a relação de equivalência que identifica quaisquer funções que são iguais em μ -quase todo ponto.

Podemos definir uma norma em $L^\infty(\mu)$ por

$$\|f\|_\infty = \text{supess}_\mu(f).$$

Deixamos como exercício para o leitor (veja o Exercício A.5.5) verificar que a conclusão da Proposição A.5.2 permanece válida no caso $p = \infty$. Mais precisamente, $L^\infty(\mu)$ munido da norma $\|\cdot\|_\infty$ é um espaço de Banach. Observe que se μ é uma medida finita, então $L^\infty(\mu) \subset L^p(\mu)$ para qualquer $p \geq 1$.

O *dual* de um espaço de Banach complexo E é o espaço E^* dos funcionais lineares contínuos $\phi : E \rightarrow \mathbb{C}$, munido da norma

$$\|\phi\| = \sup \left\{ \frac{|\phi(v)|}{\|v\|} : v \in E \setminus \{0\} \right\}. \quad (\text{A.5.3})$$

A desigualdade de Hölder (Teorema A.5.5) permite obter a seguinte caracterização explícita dos duais dos espaços $L^p(\mu)$:

Teorema A.5.7. Para cada $p \in [1, \infty)$ considere $q \in (1, \infty]$ caracterizado por $1/p + 1/q = 1$. A aplicação $L^q(\mu) \rightarrow L^p(\mu)^*$ definida por $g \mapsto [f \mapsto \int fg d\mu]$ é um isomorfismo isométrico entre $L^q(\mu)$ e o dual de $L^p(\mu)$.

Este enunciado é falso para $p = \infty$: em geral, o dual do espaço L^∞ não é isomorfo a $L^1(\mu)$.

A.5.4 Convexidade

Dizemos que uma função $\phi : I \rightarrow \mathbb{R}$ definida num intervalo I da reta é *convexa* se para cada $x, y \in I$ e $t \in [0, 1]$ vale

$$\phi(tx + (1-t)y) \leq t\phi(x) + (1-t)\phi(y).$$

Além disso, dizemos que ϕ é *côncava* se $-\phi$ for convexa. Para funções duas vezes diferenciáveis temos o seguinte critério prático (Exercício A.5.1): ϕ é convexa se $\phi''(x) \geq 0$ para todo $x \in I$ e é côncava se $\phi''(x) \leq 0$ para todo $x \in I$.

Teorema A.5.8 (Desigualdade de Jensen). *Seja $\phi : I \rightarrow \mathbb{R}$ uma função convexa. Se μ é uma probabilidade em X e $f \in L^1(\mu)$ é tal que $\int f d\mu \in I$, então:*

$$\phi\left(\int f d\mu\right) \leq \int \phi \circ f d\mu.$$

Exemplo A.5.9. Para toda probabilidade μ e toda função positiva integrável f temos

$$\log \int f d\mu \geq \int \log f d\mu.$$

De fato, isto corresponde à desigualdade de Jensen para a função $\phi : (0, \infty) \rightarrow \mathbb{R}$ dada por $\phi(x) = -\log x$. Observe que ϕ é convexa: $\phi''(x) = 1/x^2 > 0$ para todo x .

Exemplo A.5.10. Seja $\phi : \mathbb{R} \rightarrow \mathbb{R}$ uma função convexa, seja $(\lambda_i)_i$ uma sequência de números reais não negativos satisfazendo $\sum_{i=1}^{\infty} \lambda_i \leq 1$ e seja $(a_i)_i$ uma sequência limitada de números reais. Então,

$$\phi\left(\sum_{i=1}^{\infty} \lambda_i a_i\right) \leq \sum_{i=1}^{\infty} \lambda_i \phi(a_i). \quad (\text{A.5.4})$$

Isto pode ser visto da seguinte forma. Considere $X = [0, 1]$ munido da medida de Lebesgue μ . Seja $f : [0, 1] \rightarrow \mathbb{R}$ uma função da forma $f(x) = \sum_{i=1}^{\infty} a_i \chi_{E_i}$ onde os E_i são conjuntos mensuráveis disjuntos dois-a-dois tais que $\mu(E_i) = \lambda_i$. A desigualdade de Jensen aplicada à função f dá precisamente a relação (A.5.4).

A.5.5 Exercícios

A.5.1. Considere uma função $\varphi : (a, b) \rightarrow \mathbb{R}$. Mostre que se φ é duas vezes diferenciável e $\varphi'' > 0$, então φ é convexa. Mostre que se φ é convexa então ela é contínua.

A.5.2. Considere $p, q > 1$ tais que $1/p + 1/q = 1$. Prove

- (a) A desigualdade de Young: $ab \leq a^p/p + a^q/q$ para todo $a, b > 0$.
- (b) A desigualdade de Hölder (Teorema A.5.5).
- (c) A desigualdade de Minkowski (Teorema A.5.3).

A.5.3. Mostre que se μ é medida finita, $L^q(\mu) \subset L^p(\mu)$ para $1 \leq p < q \leq \infty$.

A.5.4. Seja μ uma medida finita e seja $f \in L^\infty(\mu)$ diferente de zero. Mostre que

$$\|f\|_\infty = \lim_n \frac{\int |f|^{n+1} d\mu}{\int |f|^n d\mu}.$$

A.5.5. Mostre que um espaço vetorial normado $(V, \|\cdot\|)$ é completo se, e somente se, toda série $\sum_k v_k$ absolutamente somável (ou seja, tal que $\sum_k \|v_k\|$ converge) é convergente. Use este fato para mostrar se μ é uma probabilidade, $\|\cdot\|_p$ é uma norma completa em $L^p(\mu)$ para todo $1 \leq p \leq \infty$.

A.5.6. Mostre que se μ é uma medida finita e $1/p + 1/q = 1$ com $1 \leq p < \infty$, a aplicação $\Phi : L^q(\mu) \rightarrow L^p(\mu)^*$, $\Phi(g)f = \int fg d\mu$ é um isomorfismo e uma isometria. [Observação: Isto não é verdade para $p = \infty$: em geral, o dual de $L^\infty(\mu)$ não é isomorfo a $L^1(\mu)$.]

A.5.7. Mostre que se X é um espaço métrico então, dada qualquer probabilidade boreliana μ , o conjunto $C^0(X)$ das funções contínuas é denso em $L^p(\mu)$ para todo $1 \leq p \leq \infty$. De fato, o mesmo vale para o subconjunto das funções uniformemente contínuas limitadas.

A.5.8. Sejam $f, g : X \rightarrow \mathbb{R}$ funções mensuráveis positivas tais que $f(x)g(x) \geq 1$ para todo x . Mostre que $\int f d\mu \int g d\mu \geq 1$ para toda probabilidade μ .

A.6 Espaços de Hilbert

Seja H um espaço vetorial, real ou complexo. Um *produto interno* em H é uma aplicação $(u, v) \mapsto u \cdot v$ de $H \times H$ no corpo dos escalares (\mathbb{R} ou \mathbb{C} , respectivamente) satisfazendo: para quaisquer $u, v, w \in H$ e qualquer escalar λ ,

1. $(u + w) \cdot v = u \cdot v + w \cdot v$ e $u \cdot (v + w) = u \cdot v + u \cdot w$;
2. $(\lambda u) \cdot v = \lambda(u \cdot v)$ e $u \cdot (\lambda v) = \bar{\lambda}(u \cdot v)$;
3. $u \cdot v = \overline{v \cdot u}$;
4. $u \cdot u \geq 0$ e $u \cdot u = 0$ se, e somente se, $u = 0$.

Então, definimos a *norma* de um vetor $u \in H$ como sendo $\|u\| = (u \cdot u)^{1/2}$.

Um *espaço de Hilbert* é um espaço vetorial munido de um produto interno cuja norma $\|\cdot\|$ é completa: relativamente a $\|\cdot\|$ toda sequência de Cauchy é convergente. Em outras palavras, H munido dessa norma é um espaço de Banach. Um exemplo padrão é o espaço $L^2(\mu)$ das funções com quadrado integrável num espaço de probabilidade (M, \mathcal{B}, μ) , munido do produto interno:

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

Dado $v \in H$ e uma família qualquer $(v_\alpha)_\alpha$ de vetores de H , dizemos que $v = \sum_\alpha v_\alpha$ se para todo $\varepsilon > 0$ existe um conjunto finito I tal que

$$\|v - \sum_{\beta \in J} v_\beta\| \leq \varepsilon \quad \text{para todo conjunto finito } J \supset I.$$

Dada qualquer família $(H_\alpha)_\alpha$ de subespaços de H , o conjunto dos vetores da forma $v = \sum_\alpha v_\alpha$ com $v_\alpha \in H_\alpha$ para todo α é um subespaço de H (veja o Exercício A.6.2). Ele é chamado *soma* da família $(H_\alpha)_\alpha$ e é representado por $\sum_\alpha H_\alpha$.

A.6.1 Ortogonalidade

Seja H um espaço de Hilbert. Dois vetores $u, v \in H$ dizem-se *ortogonais* se $u \cdot v = 0$. Um subconjunto de H diz-se *ortonormado* se os seus elementos têm norma 1 e são ortogonais dois-a-dois.

Uma *base de Hilbert* de H é um subconjunto ortonormado $B = \{v_\beta\}$ tal que o conjunto das combinações lineares (finitas) dos elementos de B é denso em H . Por exemplo, a *base de Fourier*

$$\{x \mapsto e^{2\pi i k x} : k \in \mathbb{Z}\} \quad (\text{A.6.1})$$

é uma base de Hilbert do espaço $L^2(m)$ das funções mensuráveis no círculo unitário C cujo quadrado é integrável para a medida de Lebesgue. Uma base de Hilbert não precisa ser uma base do espaço vetorial no sentido usual da Álgebra Linear (base de Hamel). Por outro lado, se $B = \{v_\beta\}$ é base de Hilbert então todo vetor $v \in H$ pode ser escrito como combinação linear infinita dos elementos de B :

$$v = \sum_{\beta} (v \cdot v_{\beta}) v_{\beta} \quad \text{e, além disso,} \quad \|v\|^2 = \sum_{\beta} |v \cdot v_{\beta}|^2.$$

Em particular, $v \cdot v_{\beta} = 0$ exceto, possivelmente, para um conjunto enumerável de valores de β .

Todo subconjunto ortonormado de H pode ser estendido a uma base de Hilbert. Em particular, bases de Hilbert sempre existem. Além disso, duas bases de Hilbert quaisquer têm o mesmo cardinal, que é chamado *dimensão de Hilbert* de H . A dimensão depende monotonamente do espaço: se H_1 é subespaço de H_2 então $\dim H_1 \leq \dim H_2$. Dizemos que dois espaços de Hilbert são *isometricamente isomorfos* se existe algum isomorfismo entre eles preservando o produto interno. Isso acontece se, e somente se, os dois espaços têm a mesma dimensão de Hilbert.

Um espaço de Hilbert diz-se *separável* se existe algum subconjunto enumerável denso (para a topologia definida pela norma). Isto acontece se, e somente se, a dimensão é enumerável ou finita. Em particular, todos os espaços de Hilbert separáveis de dimensão infinita são isometricamente isomorfos. Por essa razão, é usual encontrar na literatura (sobretudo na área de Física Matemática) menções ao espaço de Hilbert, no singular.

Dada uma família qualquer $(H_{\alpha})_{\alpha}$ de espaços de Hilbert, denotamos por $\oplus_{\alpha} H_{\alpha}$ a sua *soma direta ortogonal*, ou seja o espaço vetorial dos $(v_{\alpha})_{\alpha} \in \prod_{\alpha} H_{\alpha}$ tais que $\sum_{\alpha} \|v_{\alpha}\|_{\alpha}^2 < \infty$ (isto implica que $v_{\alpha} = 0$ exceto, possivelmente, para um conjunto enumerável de valores de α), munido do produto interno

$$(v_{\alpha})_{\alpha} \cdot (w_{\alpha})_{\alpha} = \sum_{\alpha} v_{\alpha} \bar{w}_{\alpha}.$$

O *complemento ortogonal* de um subconjunto S de um espaço de Hilbert é o conjunto S^{\perp} de todos os vetores de H que são ortogonais a todo vetor de S . É fácil ver que S^{\perp} é um subespaço fechado de H (Exercício A.6.7). Se S é um subespaço fechado de H então $S = (S^{\perp})^{\perp}$ e todo vetor $v \in H$ pode ser

decomposto, de forma única, como soma $v = s + s^\perp$ de algum $s \in S$ e algum $s^\perp \in S^\perp$. Além disso, essa decomposição é única e os vetores s e s^\perp são os elementos dos respectivos subespaços mais próximos de v .

A.6.2 Dualidade

Um *funcional linear* num espaço de Hilbert H (ou, mais geralmente, num espaço de Banach) é uma aplicação linear de H no corpo dos escalares (\mathbb{R} ou \mathbb{C}). O funcional linear é *limitado* se

$$\|\phi\| = \sup \left\{ \frac{|\phi(v)|}{\|v\|} : v \neq 0 \right\} < \infty$$

(isto é equivalente a dizer que o funcional é contínuo relativamente à topologia definida pela norma de H ; veja o Exercício A.6.3). O *dual* de um espaço de Hilbert H é o espaço vetorial H^* dos funcionais lineares contínuos. A função $\phi \mapsto \|\phi\|$ é uma norma completa em H^* e, portanto, mune o dual de uma estrutura de espaço de Banach. A aplicação

$$h : H \rightarrow H^*, \quad w \mapsto [v \mapsto v \cdot w] \quad (\text{A.6.2})$$

é uma bijeção entre os dois espaços e preserva as normas. Além disso, ela satisfaz $h(w_1 + w_2) = h(w_1) + h(w_2)$ e $h(\lambda w) = \lambda h(w)$.

A topologia *fraca* em H é a topologia menos fina relativamente à qual todos os funcionais lineares $v \mapsto v \cdot w$ são contínuos. Em termos de sequências, ela tem a propriedade de que

$$(w_n)_n \rightarrow w \text{ fracamente} \quad \Leftrightarrow \quad (v \cdot w_n)_n \rightarrow v \cdot w \text{ para todo } v \in H.$$

A topologia *fraca** no dual H^* é a topologia menos fina relativamente à qual $\phi \mapsto \phi(v)$ é contínua para todo $v \in H$.

É sabido da teoria de espaços de Banach (teorema de Banach-Alaoglu) que todo subconjunto fechado limitado (para a norma) do dual é compacto para a topologia *fraca**. No caso dos espaços de Hilbert, a topologia *fraca* no espaço H é homeomorfa à topologia *fraca** no dual H^* . Mais ainda, conforme mostra (A.6.2), o homeomorfismo pode ser tomado preservando a classe dos conjuntos limitados. Segue, portanto, que a topologia *fraca* no próprio espaço H disfruta da propriedade no teorema de Banach-Alaoglu:

Teorema A.6.1 (Banach-Alaoglu). *Todo subconjunto fechado limitado de um espaço de Hilbert H é compacto para a topologia *fraca* em H .*

Um operador linear $L : H_1 \rightarrow H_2$ entre dois espaços de Hilbert é *contínuo* (ou *limitado*) se

$$\|L\| = \sup \left\{ \frac{\|L(v)\|}{\|v\|} : v \neq 0 \right\} < \infty.$$

O *adjunto* de um operador linear contínuo é o operador linear $L^* : H_2 \rightarrow H_1$ definido por

$$v \cdot Lw = L^*v \cdot w \quad \text{para todo } v, w \in H.$$

O operador adjunto é contínuo, com $\|L^*\| = \|L\|$ e $\|L^*L\| = \|LL^*\| = \|L\|^2$. Além disso, $(L^*)^* = L$ e $(L_1 + L_2)^* = L_1^* + L_2^*$ e $(\lambda L)^* = \bar{\lambda}L^*$ (propomos, no Exercício A.6.5, que o leitor prove estes fatos).

Um operador linear contínuo $L : H \rightarrow H$ é *autoadjunto* se $L = L^*$. Mais geralmente, um operador linear contínuo é *normal* se ele satisfaz $L^*L = LL^*$. Ao longo do texto, estaremos especialmente interessados no caso em que L é *unitário*, isto é, $L^*L = \text{id} = LL^*$. Chamamos *isometria* a todo operador linear $L : H \rightarrow H$ tal que $L^*L = \text{id}$. Portanto, os operadores unitários são as isometrias lineares que também são operadores normais.

A.6.3 Exercícios

A.6.1. Seja H um espaço de Hilbert. Prove que:

1. Toda bola (fechada ou aberta) num espaço de Hilbert é um conjunto convexo.
2. A *identidade do paralelogramo*: $\|v + w\|^2 + \|v - w\|^2 = \|v\|^2 + \|w\|^2$ para quaisquer $v, w \in H$.
3. A *identidade de polarização*: $4(v \cdot w) = \|v + w\|^2 - \|v - w\|^2$ (caso real) ou $4(v \cdot w) = (\|v + w\|^2 - \|v - w\|^2) + i(\|v + iw\|^2 - \|v - iw\|^2)$ (caso complexo).

A.6.2. Mostre que, dada uma família $(H_\alpha)_\alpha$ de subespaços de H , o conjunto dos vetores da forma $v = \sum_\alpha v_\alpha$ com $v_\alpha \in H_\alpha$ para todo α é um subespaço vetorial de H .

A.6.3. Mostre que um operador linear $L : E_1 \rightarrow E_2$ entre dois espaços de Banach é contínuo se, e somente se, existe $C > 0$ tal que $\|L(v)\|_2 \leq C\|v\|_1$ para todo $v \in E_1$, onde $\|\cdot\|_i$ representa a norma no espaço E_i (dizemos que L é um operador *limitado*).

A.6.4. Considere o espaço de Hilbert $L^2(\mu)$. Seja V o subespaço das funções constantes. Quando uma função $f \in L^2(\mu)$ pertence ao complemento ortogonal de V ? Determine a projeção em V de uma função $g \in L^2(\mu)$.

A.6.5. Prove que se $L : H \rightarrow H$ é um operador limitado em um espaço de Hilbert H então o operador adjunto L^* também é limitado e $\|L^*\| = \|L\|$ e $\|L^*L\| = \|LL^*\| = \|L\|^2$ e $(L^*)^* = L$.

A.6.6. Mostre que se K é um subconjunto fechado e convexo de um espaço de Hilbert, para todo $z \in H$ existe um único $v \in K$ tal que $\|z - v\| = d(z, K)$.

A.6.7. Seja S um subespaço de um espaço de Hilbert H . Prove que

1. O complementar ortogonal S^\perp de S é um subespaço fechado de H e coincide com o complementar ortogonal do fecho \bar{S} . Além disso, $(S^\perp)^\perp = \bar{S}$.

2. Todo $v \in H$ pode ser escrito, de forma única, como soma $v = s + s^\perp$ de algum $s \in \bar{S}$ e algum $s^\perp \in S^\perp$. Os vetores s e s^\perp são os elementos dos respectivos subespaços mais próximos de v .

A.6.8. Seja E um subespaço fechado de um espaço de Hilbert H . Mostre que E também é fechado na topologia fraca. Além disso, $U(E)$ é um subespaço fechado de H , para toda isometria $U : H \rightarrow H$.

A.6.9. Mostre que um operador linear $L : H \rightarrow H$ num espaço de Hilbert H é uma isometria se, e somente se, $\|L(v)\| = \|v\|$ para todo $v \in H$. Além disso, L é um operador unitário se, e somente se, ele é uma isometria e é invertível.

A.7 Teoremas espectrais

Seja H um espaço de Hilbert complexo. O *espectro* de um operador linear contínuo $L : H \rightarrow H$ é o conjunto $\text{esp}(L)$ dos números $\lambda \in \mathbb{C}$ tais que $L - \lambda \text{id}$ não é um isomorfismo. O espectro é fechado e está contido no disco fechado com centro em zero e raio igual a $\|L\|$. Em particular, $\text{esp}(L)$ é um subconjunto compacto do plano complexo. Quando H tem dimensão finita, $\text{esp}(L)$ está formado pelos lores de L , ou seja, os números complexos λ tais que $L - \lambda \text{id}$ não é injetivo. Em geral, o espectro é estritamente maior que o conjunto dos autovalores (veja o Exercício A.7.2).

A.7.1 Medidas espectrais

Por definição, uma *projeção* em H é um operador linear contínuo $P : H \rightarrow H$ que é idempotente ($P^2 = P$) e autoadjunto ($P^* = P$). A imagem e o núcleo de P são subespaços fechados de H e são complementares ortogonais um do outro. De fato, a imagem coincide com o conjunto dos pontos fixos de P .

Considere uma aplicação E associando a cada subconjunto mensurável do plano \mathbb{C} uma projeção em H . Essa aplicação é chamada de *medida espectral* se ela satisfaz $E(\mathbb{C}) = \text{id}$ e é σ -aditiva:

$$E\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \sum_{n \in \mathbb{N}} E(B_n)$$

sempre que os B_n são disjuntos dois-a-dois. Então, dados quaisquer $v, w \in H$, a função

$$Ev \cdot w : B \mapsto E(B)v \cdot w \tag{A.7.1}$$

é uma medida complexa em \mathbb{C} . Claramente, ela depende de modo bilinear da dupla (v, w) .

Chamamos *suporte* de uma medida espectral E o conjunto $\text{supp } E$ dos pontos $z \in \mathbb{C}$ tal que $E(V) \neq 0$ para toda vizinhança V de z . Observe que o suporte é sempre um conjunto fechado. Além disso, o suporte da medida complexa $Ev \cdot w$ está contido em $\text{supp } E$ para todo v, w .

Exemplo A.7.1. Considere $\{\lambda_1, \dots, \lambda_s\} \subset \mathbb{C}$ e seja V_1, \dots, V_s uma família finita de subespaços de \mathbb{C}^d , ortogonais dois-a-dois e tais que $\mathbb{C}^d = V_1 \oplus \dots \oplus V_s$. Para cada conjunto $J \subset \{1, \dots, s\}$, denote por P_J a projeção em \mathbb{C}^d cuja imagem é $\bigoplus_{j \in J} V_j$. Para cada conjunto mensurável $B \subset \mathbb{C}$ defina

$$E(B) : \mathbb{C}^d \rightarrow \mathbb{C}^d, \quad E(B) = P_{J(B)}$$

onde $J(B)$ é o conjunto dos $j \in \{1, \dots, s\}$ tais que $\lambda_j \in B$. A função E é uma medida espectral.

Exemplo A.7.2. Seja μ uma probabilidade em \mathbb{C} e seja $H = L^2(\mu)$ o espaço das funções complexas com quadrado integrável para μ . Para cada conjunto mensurável $B \subset \mathbb{C}$ defina

$$E(B) : L^2(\mu) \rightarrow L^2(\mu), \quad \varphi \mapsto \chi_B \varphi.$$

Cada $E(B)$ é uma projeção e a função E é uma medida espectral.

O lema a seguir reúne algumas propriedades simples das medidas espectrais:

Lema A.7.3. *Seja E uma medida espectral e sejam A, B subconjuntos mensuráveis de \mathbb{C} . Então:*

1. $E(\emptyset) = 0$ e $E(\text{supp } E) = \text{id}$;
2. se $A \subset B$ então $E(A) \leq E(B)$ e $E(B \setminus A) = E(B) - E(A)$;
3. $E(A \cup B) + E(A \cap B) = E(A) + E(B)$;
4. $E(A)E(B) = E(A \cap B) = E(B)E(A)$.

No que segue suporemos que E é uma medida espectral com suporte compacto. Então o suporte de toda medida complexa $E v \cdot w$ também é compacto. Consequentemente, a integral $\int z d(E(z)v \cdot w)$ está bem definida e é uma função bilinear de (v, w) . Então existe um operador linear limitado $L : H \rightarrow H$ tal que

$$L v \cdot w = \int z d(E(z)v \cdot w) \quad \text{para todo } v, w \in H. \quad (\text{A.7.2})$$

Escrevemos, numa forma mais sucinta:

$$L = \int z dE(z). \quad (\text{A.7.3})$$

Mais geralmente, dada qualquer função mensurável limitada ψ no suporte da medida espectral E , existe um operador linear limitado $\psi(L) : H \rightarrow H$ caracterizado por

$$\psi(L)v \cdot w = \int \psi(z) d(E(z)v \cdot w) \quad \text{para todo } v, w \in H. \quad (\text{A.7.4})$$

Escrevemos

$$\psi(L) = \int \psi(z) dE(z). \quad (\text{A.7.5})$$

Lema A.7.4. *Seja E uma medida espectral. Dadas funções mensuráveis limitadas φ, ψ e números $\alpha, \beta \in \mathbb{C}$,*

$$(1) \int (\alpha\varphi + \beta\psi)(z) dE(z) = \alpha \int \varphi dE(z) + \beta \int \psi dE(z);$$

$$(2) \int \bar{\varphi}(z) dE(z) = \left(\int \varphi(z) dE(z) \right)^*;$$

$$(3) \int (\varphi\psi)(z) dE(z) = \left(\int \varphi(z) dE(z) \right) \left(\int \psi(z) dE(z) \right)$$

Em particular, pelo item (3) do lema,

$$L^j = \left(\int z dE(z) \right)^j = \int z^j dE(z) \quad \text{para todo } j \in \mathbb{N}. \quad (\text{A.7.6})$$

Analogamente, usando também o item (2),

$$\begin{aligned} LL^* &= \left(\int z dE(z) \right) \left(\int \bar{z} dE(z) \right) = \int |z|^2 dE(z) \\ &= \left(\int \bar{z} dE(z) \right) \left(\int z dE(z) \right) = L^*L. \end{aligned} \quad (\text{A.7.7})$$

Consequentemente, o operador linear definido por (A.7.3) é normal. O teorema espectral, que enunciamos a seguir, afirma que todo operador normal pode ser escrito desta forma:

Teorema A.7.5 (Espectral). *Para todo operador normal $L : H \rightarrow H$ existe uma medida espectral E tal que $L = \int z dE(z)$. Essa medida é única e o seu suporte coincide com o espectro de L . Em particular, L é unitário se, e somente se, $\text{supp } E$ está contido no círculo unitário $\{z \in \mathbb{C} : |z| = 1\}$.*

Exemplo A.7.6 (Teorema espectral em dimensão finita). O teorema espectral em dimensão finita afirma que para todo operador normal complexo L existe uma base de Hilbert formada por autovetores de L . Sejam $\lambda_1, \dots, \lambda_s$ os autovalores de L . Os autoespaços $V_j = \ker(L - \lambda_j \text{id})$ são ortogonais dois-a-dois, porque L é normal. Além disso, pelo Teorema A.7.5, a soma direta $\bigoplus_{j=1}^s V_j$ é todo o \mathbb{C}^d . Então

$$L = \sum_{j=1}^s \lambda_j \pi_j$$

onde π_j representa a projeção sobre V_j . Em outras palavras, a medida espectral E do operador L é dada por $E(\{\lambda_j\}) = \pi_j$ para cada $j = 1, \dots, s$ e $E(B) = 0$ se B não contém nenhum autovalor de L .

Exemplo A.7.7. Seja $(\sigma_\alpha)_{\alpha \in \mathcal{A}}$ uma família qualquer de medidas finitas no círculo unitário $\{z \in \mathbb{C} : |z| = 1\}$. Considere $H = \bigoplus_{\alpha \in \mathcal{A}} L^2(\sigma_\alpha)$ e o operador linear

$$L : H \rightarrow H, \quad (\varphi_\alpha)_\alpha \mapsto (z \mapsto z\varphi_\alpha(z))_\alpha.$$

Considere a medida espectral dada por

$$E(B) : H \rightarrow H, \quad (\varphi_\alpha)_\alpha \mapsto (\mathcal{X}_B \varphi_\alpha)_\alpha$$

(compare com o Exemplo A.7.2). Então $L = \int z dE(z)$. De fato, a definição de E dá que $E\varphi \cdot \psi = \sum_{\alpha} \varphi_{\alpha} \bar{\psi}_{\alpha} \sigma_{\alpha}$ para todo $\varphi = (\varphi_{\alpha})_{\alpha}$ e $\psi = (\psi_{\alpha})_{\alpha}$ no espaço H . Então,

$$L\varphi \cdot \psi = \sum_{\alpha} \int z \varphi_{\alpha}(z) \bar{\psi}_{\alpha}(z) d\sigma_{\alpha}(z) = \int z d(E(z)\varphi \cdot \psi) \quad (\text{A.7.8})$$

para todo φ, ψ .

Dizemos que λ é um *átomo* da medida espectral se $E(\{\lambda\}) \neq 0$ ou, equivalentemente, se existe algum vetor não nulo $\omega \in H$ tal que $E(\{\lambda\})\omega \neq 0$. A demonstração da proposição a seguir está delineada no Exercício A.7.4.

Proposição A.7.8. *Todo autovalor de L é átomo da medida espectral E . Reciprocamente, se λ é átomo de E então ele é autovalor do operador L , tendo qualquer vetor não nulo da forma $v = E(\{\lambda\})\omega$ como autovetor.*

A.7.2 Representação espectral

O Teorema A.7.5 mostra que os operadores lineares normais num espaço de Hilbert são essencialmente a mesma coisa que as medidas espectrais nesse espaço. Resultados deste tipo, que estabelecem um ‘dicionário’ entre duas classes de objetos que *a priori* não estão relacionadas, estão entre os mais fascinantes da Matemática. Claro que a sua utilidade para o estudo de uma dessas classes (os operadores lineares normais, digamos) depende de até que ponto somos capazes de compreender a outra classe (neste caso, as medidas espectrais). Na presente situação isso é atendido, de modo inteiramente satisfatório, pelo próximo resultado, que exhibe uma forma canônica (inspirada no Exemplo A.7.2) na qual todo operador normal pode ser escrito.

Como antes, usamos \oplus para representar a soma direta ortogonal de espaços de Hilbert. Dado um cardinal χ , finito ou infinito, e um espaço de Hilbert V , representamos por V^{χ} a soma direta ortogonal de χ cópias de V .

Teorema A.7.9 (Representação espectral). *Seja $L : H \rightarrow H$ um operador linear normal. Então existem medidas finitas mutuamente singulares $(\sigma_j)_j$ com suporte no espectro de L , existem cardinais $(\chi_j)_j$ e existe um operador unitário $U : H \rightarrow \bigoplus_j L^2(\sigma_j)^{\chi_j}$, tal que o operador conjugado $ULU^{-1} = T$ é dado por:*

$$T : \bigoplus_j L^2(\sigma_j)^{\chi_j} \rightarrow \bigoplus_j L^2(\sigma_j)^{\chi_j}, \quad (\varphi_{j,l})_{j,l} \mapsto (z \mapsto z\varphi_{j,l}(z))_{j,l}. \quad (\text{A.7.9})$$

Chamaremos (A.7.9) de *representação espectral* do operador normal L . Deve ser observado que as medidas σ_j no Teorema A.7.9 não são unicamente determinadas. Mas a representação espectral é única, no seguinte sentido.

Chamamos *função de multiplicidades* do operador L à função que a cada medida finita θ em \mathbb{C} associa o menor cardinal χ_j tal que as medidas θ e σ_j não são mutuamente singulares. Mostra-se que esta função está unicamente determinada pelo operador L , ou seja, ela não depende da escolha das medidas

σ_j no enunciado. Além disso, dois operadores normais são conjugados por um operador unitário se, e somente se, eles têm a mesma função de multiplicidades.

Exemplo A.7.10 (Representação espectral em dimensão finita). Voltamos ao contexto do Exemplo A.7.6. Para cada $j = 1, \dots, s$, seja σ_j a medida de Dirac no autovalor λ_j e seja χ_j a dimensão do autoespaço V_j . Note que o espaço $L^2(\sigma_j)$ tem dimensão 1. Portanto, podemos escolher um operador unitário $U_j : V_j \rightarrow L^2(\sigma_j)$, para cada $j = 1, \dots, s$. Como $L = \lambda_j \text{id}$ restrito a V_j , temos que $T_j = U_j L U_j^{-1} = \lambda_j \text{id}$, ou seja,

$$T_j : ((\varphi_\alpha)_\alpha) \mapsto (z \mapsto \lambda_j \varphi_\alpha(z))_\alpha = (z \mapsto z \varphi_\alpha(z))_\alpha.$$

Desta forma fica definido um operador unitário

$$U : \mathbb{C}^d \rightarrow \bigoplus_{j=1}^s L^2(\sigma_j)^{\chi_j}.$$

tal que $T = U L U^{-1}$ é a representação espectral de L .

A.7.3 Exercícios

A.7.1. Seja $T : E \rightarrow E$ um isomorfismo de espaço de Banach, ou seja, uma bijeção linear contínua com inversa contínua. Mostre que $T+H$ é um isomorfismo de espaço de Banach para toda aplicação linear tal que $\|H\| \|T^{-1}\| < 1$. Use este fato para provar que o espectro de todo operador linear limitado $L : E \rightarrow E$ é fechado e está contido no disco fechado de raio igual a $\|L\|$.

A.7.2. Mostre que se $L : H \rightarrow H$ é um operador linear num espaço de Hilbert de dimensão finita então $\text{esp}(L)$ está formado pelos autovalores de L , ou seja, pelos números complexos λ para os quais $L - \lambda \text{id}$ não é injetivo. Dê um exemplo, em dimensão infinita, tal que o espectro é estritamente maior que o conjunto dos autovalores.

A.7.3. Prove o Lema A.7.3.

A.7.4. Prove a Proposição A.7.8, usando o seguinte roteiro:

1. Suponha que $Lv = \lambda v$ para algum $v \neq 0$. Considere as funções

$$\varphi_n(z) = \begin{cases} (z - \lambda)^{-1} & \text{se } |z - \lambda| > 1/n \\ 0 & \text{caso contrário.} \end{cases}$$

Mostre que $\varphi_n(L)(L - \lambda \text{id}) = E(\{z : |z - \lambda| > 1/n\})$ para todo n . Conclua que $E(\{\lambda\})v = v$ e, portanto, λ é átomo de E .

2. Suponha que existe $w \in H$ tal que $v = E(\{\lambda\})w$ é não nulo. Mostre que, dado qualquer conjunto mensurável $B \subset \mathbb{C}$,

$$E(B)v = \begin{cases} v & \text{se } \lambda \in B \\ 0 & \text{se } \lambda \notin B. \end{cases}$$

Conclua que $Lv = \lambda v$ e, portanto, λ é autovalor de L .

A.7.5. Seja (σ_j) a família de medidas dada no Teorema A.7.9. Dado um conjunto mensurável $B \subset \mathbb{C}$, verifique que $E(B) = 0$ se, e somente se, $\sigma_j(B) = 0$ para todo j . Portanto, dada qualquer medida η em \mathbb{C} , temos que $E \ll \eta$ se, e somente se, $\sigma_j \ll \eta$ para todo j .

Dicas e soluções dos exercícios

1.1.1. Seja $P = \{p, \dots, f^{-k+1}(p)\}$. A medida $\delta_{p,k}$ é invariante se, e somente se, $\#A \cap P = \#f^{-1}(A) \cap P$ qualquer que seja o conjunto mensurável $A \subset M$. Considerando $A = P$, mostre que isso acontece se, e somente se, $f^k(p) = p$.

1.1.2. Use o Exercício A.3.5 para aproximar funções características por funções contínuas.

1.1.3. Argumente, por indução em k , que se $\mu(f^{-1}(A)) = \mu(A)$ para todo conjunto mensurável $A \subset M$ então $\mu(f^{-k}(A)) = \mu(A)$ para todo conjunto mensurável $A \subset M$.

1.1.4. Como $\mu(B) = \mu(f^{-1}(B))$, as condições (1) e (2) são equivalentes. Segue que qualquer uma delas é equivalente a (3). Além disso, é imediato que (4) implica (1). Supondo que vale (1), tome $C = \bigcap_{k=1}^{\infty} \bigcup_{j=k}^{\infty} f^{-j}(B)$.

1.1.5. Use a fórmula de mudança de variáveis em integrais.

1.2.1. Para cada $k \geq 1$, seja E_k o conjunto dos pontos $x \in E$ que regressam a E , pela primeira vez, em exatamente k iterados. O teorema de Poincaré afirma que $\mu(\bigcup_k E_k) = \mu(E) > 0$. Logo $\mu(E_N) > 0$ para algum N . Reciprocamente, se o teorema de Poincaré fosse falso, existiriam $E \subset F$ tais que $\mu(E) > 0$ e nenhum ponto de E regressa a F . Tomando $D = E$, isso contradiria a hipótese.

1.2.2. Seja F o conjunto dos pontos $x \in E$ com número finito de iterados em B e seja E_0 o conjunto dos pontos $x \in E \setminus F$ que nunca regressam a E . Mostre que os iterados $f^n(E_0)$, $n \geq 1$ são disjuntos dois-a-dois e conclua que $\sum_n \mu(E_0 \cap f^{-n}(B)) \leq \mu(B) < \infty$. Por outro lado, $\sum_n \mathcal{X}_{f^{-n}(B)}(x) = \infty$ para todo $x \in E_0$. Deduza que $\mu(E_0) = 0$. Para cada $k \geq 1$, seja E_k o conjunto dos pontos $x \in E \setminus F$ que regressam a E exatamente k vezes. Mostre, por indução, que $\mu(E_k) = 0$ para todo k .

1.2.3. Utilize o exercício anterior com $B = M_k$ e, em seguida, tome a interseção sobre todos os valores de k .

1.2.4. Seja D_0 o conjunto dos $x \in D$ onde o limite é zero. Então $\lim_n (1/n) \#\{0 \leq j \leq n-1 : f^j(x) \in D_0\}$ também é zero, para todo $x \in D_0$. Deduza que o mesmo vale em todo ponto, ou seja, que a sequência $(1/n) \sum_{j=0}^{n-1} \mathcal{X}_{D_0}(f^j(x))$ converge para zero para todo $x \in M$. Conclua, usando o teorema da convergência dominada, que $\mu(D_0) = 0$.

1.2.5. Mostre que se $N > 1/\mu(A)$, então existe $j \in V_A$ com $0 \leq j \leq N$. Adaptando a prova da afirmação anterior, conclua que se K é um conjunto de inteiros não-negativos com $\#K > 1/\mu(A)$, então podemos achar $k_1, k_2 \in K$ e $n \in V_A$ tais que $n = k_1 - k_2$. Ou seja, o conjunto $K - K = \{k_1 - k_2; k_1, k_2 \in K\}$ intersecta V_A . Para concluir que S é sindético assuma, por absurdo, que para cada $n \in \mathbb{N}$ existe uma número l_n tal que $\{l_n, l_n+1, \dots, l_n+n\} \cap V_A = \emptyset$. Considere um elemento $k_1 \notin V_A$ e construa recursivamente a sequência $k_{j+1} = l_{k_j} + k_j$. Prove que o conjunto $K = \{k_1, \dots, k_N\}$ é tal que $(K - K) \cap V_A = \emptyset$.

1.2.6. Se a conclusão não é verdadeira, existe $k \geq 1$ e $b > 1$ tal que o conjunto $B = \{x \in [0, 1] : n|f^n(x) - x| > b \text{ para todo } n \geq k\}$ tem medida positiva. Tome $a \in B$ um ponto de densidade de B . Considere $E = B \cap B(a, r)$, para r pequeno. Obtenha uma estimativa inferior

para o tempo de retorno a E de qualquer ponto $x \in E$ e use o teorema de Kač para chegar a uma contradição.

1.2.7. Dado $\varepsilon > 0$, temos $|\omega - q_n| < \varepsilon$ para todo n suficientemente grande. Conclua que $|R(q_n)| \leq |\omega - q_n|(\sqrt{5} + \varepsilon)$. Justifique que $n^2|R(q_n)| \geq 1$ e conclua que $n^2|\omega - q_n| \geq 1/(\sqrt{5} + \varepsilon)$ para todo n suficientemente grande.

1.3.1. Particione o intervalo em $I_i = [i/10, (i+1)/10)$ com $i = 0, \dots, 9$ e tome $\rho \equiv 1$.

1.3.2. Observe que $E = [0, 617, 0, 618)$ tem medida positiva e aplique o teorema de recorrência de Poincaré. Note que todo intervalo $J = [j/10^k, (j+1)/10^k)$ contém subintervalo J' tal que $m(J')/m(J) = 10^{-3}$ e $f^k(x) \in E$ para todo $x \in J'$. Usando o Teorema A.2.14, conclua quase todo $x \in [0, 1]$ tem pelo menos um iterado em E . Agora a segunda afirmação no exercício é uma consequência da primeira.

1.3.3. Observe que $E = (1/618, 1/617)$ tem medida positiva, para a medida invariante μ da transformação de Gauss e aplique o teorema de recorrência de Poincaré. Lembre que a medida μ é equivalente à medida de Lebesgue.

1.3.4. Verifique que se $x = p/q$ com $0 < p < q$ então $G(x) = q'/p$ com $0 \leq q' < p$. Deduza que $G^n(x) = 0$ para algum n . Reciprocamente, se $x_n = G^n(x) = 0$ para algum n então a expressão (1.3.3) mostra que x é um número racional.

1.3.5. Considere a sequência $\log_{10} a_n$ onde \log_{10} representa o logaritmo na base 10 e observe que $\log_{10} 2$ é um número irracional.

1.3.6. Seja $d \geq 1$ o grau de f , ou seja, $d = \#f^{-1}(y)$ para todo y . Para cada $y \in M$ e toda vizinhança suficientemente pequena V , a pré-imagem $f^{-1}(V)$ tem d componentes conexas, V_1, \dots, V_d , tais que cada restrição $f_j = f|_{V_j}$ é um difeomorfismo sobre V . Suponha que f preserva a medida $\mu = \rho \text{ vol}$. Então, por mudança de variáveis,

$$\int_V \rho d \text{ vol} = \sum_{j=1}^d \int_{V_j} \rho d \text{ vol} = \sum_{j=1}^d \int_V (\rho \circ f_j^{-1}) |\det Df_j^{-1}| d \text{ vol}.$$

Fazendo o raio de V ir para zero, deduza que $\rho(y) = \sum_{j=1}^d \rho(f_j^{-1}(y)) |\det Df_j^{-1}(y)|$. Reciprocamente, se esta igualdade vale em todo ponto então $\mu(V) = \mu(f^{-1}(V))$ para todo conjunto V com diâmetro suficientemente pequeno. Deduza que f preserva μ .

1.3.7. Verifique que $\#f_A^{-1}(y) = |\det A|$ para todo y e utilize o Exercício 1.3.6 com $\rho \equiv 1$.

1.3.8. Considere os intervalos da forma $I_{k,j} = [(j-1)2^{-k}, j2^{-k})$ com $k \geq 1$ e $1 \leq j \leq 2^k$. Usando que μ é invariante pela rotação $R_{2^{-k}}$, mostre que $\mu(I_{k,j}) = 2^{-k} = m(I_{k,j})$ para todo j e todo k . Conclua que $\mu = m$.

1.3.9. Seja $k = (k_1, \dots, k_d) \in \mathbb{Z}^d \setminus \{0\}$ tal que $k \cdot \theta \in \mathbb{Z}$. Defina $\varphi([x]) = e^{2\pi i k \cdot x}$ e verifique que φ é invariante. Como φ é não constante, podemos encontrar conjuntos abertos disjuntos $A, B \subset \mathbb{C}$ tais que $U = \varphi^{-1}(A)$ e $V = \varphi^{-1}(B)$ são não vazios. Toda órbita que intersecta U não intersecta V , logo não existe nenhuma órbita densa.

1.3.10. (Katok, Hasselblatt [KH95]) Considere a série de Fourier $\mathcal{X}_V = \sum_{k \in \mathbb{Z}^d} a_k e^{2\pi i k \cdot x}$ da função característica de V . Usando a invariância de V e a hipótese de que θ é racionalmente independente, conclua que $a_k = 0$ para todo $k \neq 0$. Deduza que V tem medida de Lebesgue total e, portanto, é denso em \mathbb{T}^d . Dado qualquer aberto U , observe que $\cup_{n \in \mathbb{Z}} R_\theta^n(U)$ é invariante pela rotação e aplique a conclusão anterior. Seja $\{U_k : k \geq 1\}$ uma base enumerável de abertos de \mathbb{T}^d . Mostre que $X = \cap_k \cup_n R_\theta^n(U_k)$ é não vazio e todo $[x] \in X$ tem órbita densa. Fixe qualquer $[x] \in X$. Dado qualquer $[y] \in \mathbb{T}^d$, considere $h = R_{y-x}$. Note que $h([x]) = [y]$ e $h \circ R_\theta \circ h^{-1} = R_\theta$. Conclua que $\mathcal{O}([y])$ é densa.

1.3.11. Calcule o divergente do campo de vetores hamiltoniano e use o Teorema 1.3.7.

1.3.12. Considere bases ortonormadas $\{v_1, v_2, \dots, v_d\}$, em x , e $\{w_1, w_2, \dots, w_d\}$, em $f(x)$, tais que v_1 e w_1 são ortogonais a H_c . Verifique que $\text{grad } H(f(x)) \cdot Df(x)v = \text{grad } H(x) \cdot v$ para todo v . Deduza que a matriz de $Df(x)$ relativamente a essas bases tem a forma

$$Df(x) = \begin{pmatrix} \alpha & 0 & \cdots & 0 \\ \beta_2 & \gamma_{2,2} & \cdots & \gamma_{2,d} \\ \cdots & \cdots & \cdots & \cdots \\ \beta_d & \gamma_{d,2} & \cdots & \gamma_{d,d} \end{pmatrix}$$

com $\|\text{grad } H(f(x))\| |\alpha| = \|\text{grad } H(x)\|$. Note que $\Gamma = (\gamma_{i,j})_{i,j}$ é a matriz de $D(f | H_c)$ e observe que $|\det \Gamma| = \|\text{grad } H(x)\| / \|\text{grad } H(f(x))\|$. Usando a fórmula de mudança de variáveis, conclua que $f | H_c$ preserva a medida $ds / \|\text{grad } H\|$.

1.4.1. Considere os intervalos $I_k = (1/2^{k+1}, 1/2^k)$ e observe que $f(I_k) = I_{k-1}$ para todo $k \geq 1$ e, portanto, $\tau(x) = k$ para $x \in I_k$. Observe que $f(1) = 1$ e $f(1/2^k) = 1/2^{k-1}$. Mostre que se g está C^1 próxima de f , existe z_0 próximo de 1 tal que $g(z_0) = z_0$ e existem $z_k, k \geq 1$ tais que $g(z_k) = z_{k-1}$ e para todo $k \geq 1$ e $|z_{k-1} - z - k|$ decresce exponencialmente.

1.4.2. Verifique que existem $\alpha_1, \alpha_2, \alpha_3 > 0$ tais que $\alpha_1 x^d \leq Df(x) - 1 \leq \alpha_2 x^d$ e $|D^2 f(x)| \leq \alpha_3 x^{d-1}$ para $x > 0$ perto de zero. Seja h a inversa de $f | (0, a)$. Deduza que $Dh(y) - 1 \approx -y^d$ para y perto de zero e, portanto, existem $\beta_2 > \beta_1 > 0$ tais que $y - \beta_1 y^{d+1} \geq h(y) \geq y - \beta_2 y^{d+1}$ para y numa vizinhança de zero. Fixe $k \geq 1$ tal que a_j está nessa vizinhança para todo $j \geq k$. Fixe $K_2 > 1 > K_1 > 0$ tais que $1 + K_1 x \leq (1+x)^{1/d} \leq 1 + K_2 x$ para $0 \leq x \leq 1$. Escolha $c_2 > c_1 > 0$ tais que $c_1 \leq a_j j^{1/d} \leq c_2$ para todo $1 \leq j \leq k$. Conclua, por indução, que

$$a_{j+1} = h(a_j) \geq \left(\frac{c_1}{j^{1/d}} \right) - \beta_2 \left(\frac{c_1}{j^{1/d}} \right)^{d+1} \geq \frac{c_1}{(j+1)^{1/d}} \left(1 + K_1 \frac{1}{j} \right) \left(1 - \beta_2 c_1^d \frac{1}{j} \right)$$

Tome c_1 tal que $2\beta_2 c_1^d < K_1$ e conclua que $a_{j+1} \geq c_1 / (j+1)^{1/d}$. A cota superior é análoga, tomando $\beta_1 c_2^d > K_2$. Deduza a segunda parte do exercício, com $c_3 = \beta_1 c_1^{d+1}$ e $c_4 = \beta_2 c_2^{d+1}$. Finalmente, (1.4.5) dá que $\nu_\rho((0, 1)) = \sum_{n=0}^{\infty} \nu((0, a_n))$. Usando o fato de que a densidade de ν é limitada, conclua que $\nu_\rho((0, 1)) < \infty$ se, e somente se, $d < 1$.

1.4.3. $g(x) = f^k(x)$ para todo $x \in B_k = \{(x_n)_n : x_0 = x_k = 1 \text{ e } x_i \neq 1 \text{ para } 0 < i < k\}$.

1.4.4. Escolha um conjunto E com medida menor que ε/n e, para cada $k \geq 1$, seja E_k dos pontos $x \in E$ que retornam a E em exatamente k iterados. Tome para B a união dos conjuntos E_k com $k \geq n$, dos n -ésimos iterados dos conjuntos E_k com $k \geq 2n$, e assim sucessivamente. Para a segunda parte, observe que se (f, μ) é aperiódico então μ não pode ter átomos.

1.4.5. Pela hipótese, $f^\tau(y) \in H_{n-\tau(y)}$ sempre que $y \in H_n$ com $n > \tau(y)$. Portanto $T(y) \in H$ se $y \in H$. Considere $A_n = \{1 \leq j \leq n : x \in H_j\}$ e $B_n = \{l \geq 1 : \sum_{i=0}^l \tau(T^i(x)) \leq n\}$. Mostre, por indução, que $\#A_n \leq \#B_n$ e deduza que $\limsup_n \#B_n/n \geq \theta$. Agora suponha que $\liminf_k (1/k) \sum_{i=0}^{k-1} \tau(T^i(x)) > (1/\theta)$. Mostre que existe $\theta_0 < \theta$ tal que $\#B_n < \theta_0 n$, para todo n suficientemente grande. Esta conclusão contradiz a anterior.

1.4.6. Como $\tau_{n+1}(x) - \tau_n(x) = \tau(f^{\tau_n}(x))$, temos que $\tau_{n+1}(x)/\tau_n(x) \rightarrow 1$ se, e somente se, $\tau(f^{\tau_n}(x))/\tau_n(x) \rightarrow 0$. Para provar que esta última relação é válida em quase todo ponto, seja D o conjunto dos pontos x tais que $\limsup_n \tau(f^{\tau_n(x)}(x))/\tau_n(x) > 0$. Verifique que, para cada ρ fixado, os conjuntos $L_\rho(n) = \{x \in M : \tau(x) > \rho n\}$ satisfazem $\sum_n \mu(L_\rho(n)) < \infty$. Mostre que $D \subset \cup_{\rho \in \mathbb{Q}} \cap_{m \geq 1} \cup_{n \geq m} f^{-n}(L_\rho(n))$. Em seguida, use o lema de Borel-Cantelli para concluir que $\mu(D) = 0$.

1.5.1. Considere $f_1(x) = x/2$ e $f_2(x) = (x+1)/2$.

1.5.2. Aplique o Lema de Zorn no conjunto \mathcal{G} dos subconjuntos fechados de M , munido da relação de ordem dada pelo inverso da inclusão.

1.5.3. Considere qualquer sequência $(p_n)_n$ em M tal que $(\varphi(p_n))_n$ converge para $\sup \varphi$. Seja p um ponto de acumulação.

1.5.4. Suponha que φ é semicontínua superiormente. Para cada $k \geq 1$, seja F_k o conjunto dos pontos $x \in X$ tais que toda vizinhança contém algum y tal que $\varphi(y) \leq \varphi(x) - (1/k)$. Verifique que F_k é fechado e tem interior vazio. Além disso, o conjunto dos pontos de continuidade de φ é $\cap_k F_k^c$. Para a parte final do exercício, lembre que todo espaço métrico compacto é um espaço de Baire.

1.5.5. Observe que as aplicações f, f^2, \dots, f^k comutam e, em seguida, use o teorema de recorrência múltipla de Poincaré.

1.5.6. Por definição, o complementar de $\Omega(f_1, \dots, f_q)^c$ é um aberto. O teorema de recorrência múltipla de Birkhoff garante que o conjunto não errante é não vazio.

2.1.1. As condições (1), (2), (3), (4) e (5) significam que a sequência $(\mu_n)_n$ converge nas topologias (2.1.1), (2.1.3), (2.1.4), (2.1.5) e (2.1.2), respectivamente. Como estas topologias coincidem (Teorema 2.1.2), elas têm as mesmas sequências convergentes.

2.1.2. Suponha que $\int \phi d\mu_n \rightarrow \int \phi d\mu$ para todo $\phi \in \mathcal{F}$. Dada qualquer função contínua limitada $\psi : M \rightarrow \mathbb{R}$ com $\|\psi\| \leq 1$ e dado qualquer $\varepsilon > 0$, tome $\phi \in \mathcal{F}$ com $\|\psi - \phi\| \leq \varepsilon$. Então $\int \psi d\mu - \varepsilon \leq \liminf_n \int \psi d\mu_n \leq \limsup_n \int \psi d\mu_n \leq \int \psi d\mu + \varepsilon$. Como ε é arbitrário, isto implica que $\lim_n \int \psi d\mu_n = \int \psi d\mu$. Se $\|\psi\| > 1$, aplique o argumento anterior a $\psi/\|\psi\|$.

2.1.3. Tome Z subconjunto enumerável denso de M . O conjunto das medidas da forma $\sum_{i=1}^s q_i \delta_{z_i}$ com $q_i \in \mathbb{Q}$, $z_i \in Z$ e $\sum_{i=1}^s q_i$ é denso em $\mathcal{M}_1(M)$. Para ver isso, considere $\varepsilon > 0$ e funções contínuas limitadas ϕ_1, \dots, ϕ_k . Seja $K > 0$ um majorante para todo $|\phi_i|$ e seja I_j , $j = 1, \dots, m$ uma partição de $[-K, K]$ em intervalos com comprimento menor que ε . Seja A_l , $l = 1, \dots, m$ uma partição de M em conjuntos mensuráveis tais que cada $\phi_i(A_l)$ está contido em algum I_j . Escolha algum x_l em cada A_l . Então $|\int \phi_i d\mu - \sum_{l=1}^m \mu(A_l) \phi_i(x_l)| < \varepsilon$. Esta desigualdade não é afetada se substituirmos cada $\mu(A_l)$ por um número racional q_l e cada x_l por um ponto $z_l \in Z$, suficientemente próximos.

2.1.4. $V_u(\mu, \varepsilon) \subset V_p(\mu, \mathcal{B}, \varepsilon)$ para todo $\mathcal{B} = \{B_1, \dots, B_N\}$, logo a topologia uniforme é mais forte que a topologia pontual. $V_p(\mu, \mathcal{B}, \varepsilon) = V_c(\mu, \mathcal{B}, \varepsilon)$ se os B_j são conjuntos de continuidade de μ , logo a topologia pontual é mais forte que a topologia (2.1.5). Seja $\phi : \mathbb{R} \rightarrow [0, 1]$ com $\text{supp } \phi \subset [-1, 1]$ e $\int_{-1}^1 \phi(x) dx = 1$. Considere a sequência de probabilidades $\mu_n = n \phi(nx) dx$. Então $(\mu_n)_n$ converge para δ_0 na topologia fraca* mas não converge pontualmente: $\mu_n(\{0\}) = 0$ para todo n , enquanto que $\delta_0(\{0\}) = 1$. Considere $\psi : \mathbb{R} \rightarrow \mathbb{R}$ dada por $\psi(x) = 0$ para $x \in [k, k + 1/2)$ e $\psi(x) = 2$ para $x \in [k + 1/2, k + 1)$, com $k \in \mathbb{Z}$. Seja μ a medida de Lebesgue em $[0, 1]$ e seja $\mu_n = \varphi(2^n x) \mu$ para cada n . Verifique que para todo n existe um conjunto mensurável B_n tal que $\mu_n(B_n) = 0$ e $\mu(B_n) = 1/2$. Seja \mathcal{A} a família dos subconjuntos mensuráveis $E \subset [0, 1]$ tais que $(\mu_n(E))_n$ converge para $\mu(E)$. Mostre que \mathcal{A} contém todo intervalo, é uma álgebra e uma classe monótona. Logo, $(\mu_n)_n$ converge pontualmente para μ .

2.1.5. Segue diretamente da definição que $D \geq 0$ e $D(\mu, \nu) = D(\nu, \mu)$ para quaisquer μ e ν . Para mostrar que $D(\mu, \nu) \leq D(\mu, \eta) + D(\eta, \nu)$, note que se

$$\mu(B) < \eta(B^{\delta_1}) + \delta_1, \quad \eta(B) < \mu(B^{\delta_1}) + \delta_1 \quad \text{e} \quad \eta(B) < \nu(B^{\delta_2}) + \delta_2, \quad \nu(B) < \eta(B^{\delta_2}) + \delta_2$$

para todo conjunto mensurável B , então também temos

$$\mu(B) < \nu(B^{\delta_1 + \delta_2}) + \delta_1 + \delta_2 \quad \text{e} \quad \nu(B) < \mu(B^{\delta_1 + \delta_2}) + \delta_1 + \delta_2.$$

Logo $D(\mu, \nu) \leq \delta_1 + \delta_2$ e agora basta passar ao ínfimo em δ_1 e δ_2 .

2.1.6. Considere a imagem $V_*\mu$ da medida μ pela função V . Verifique que $V_*\mu((a, b]) = F(b) - F(a)$ para todo $a < b$. Consequentemente, $V_*\mu(\{b\}) = F(b) - \lim_{a \rightarrow b^-} F(a)$. Portanto, $(-\infty, b]$ é conjunto de continuidade para $V_*\mu$ se, e somente se, b é ponto de continuidade para F . Usando o Teorema 2.1.2, segue que se $(V_{k*}\mu)_k$ converge para $V_*\mu$ na topologia fraca* então $(V_k)_k$ converge para V em distribuição. Reciprocamente, se $(V_k)_k$ converge em distribuição para V então $V_{k*}\mu((a, b]) = F_k(b) - F_k(a)$ converge para $F(b) - F(a) = V_*\mu((a, b])$, quaisquer que sejam os pontos de continuidade $a < b$ de F . Observando que tais intervalos $(a, b]$ geram a σ -álgebra de Borel da reta, conclua que $(V_{k*}\mu)_k$ converge para $V_*\mu$ na topologia fraca*.

2.1.7. Pelo Exercício 2.1.1, $\mu(B) = \lim_n \mu_n(B)$ e portanto $\mu_n(B) > 0$ para todo n suficientemente grande. Seja $B' \subset B$ qualquer conjunto de continuidade para $(\mu|B)/\mu(B)$. Então $\mu(\partial B' \cap B) = 0$ e, como $\partial B' \subset (\partial B' \cap B) \cup \partial B$, segue que B' é conjunto de continuidade para μ . Segue que $\mu_n(B')/\mu_n(B) \rightarrow \mu(B')/\mu(B)$. Como B' é arbitrário, o Exercício 2.1.1 dá que $(\mu_n|B)/\mu_n(B)$ converge para $(\mu|B)/\mu(B)$ na topologia fraca*. Este argumento se estende aos casos em que B é um subconjunto fechado ou aberto, lembrando que se B' é subconjunto fechado (aberto) de B então também é um subconjunto fechado (aberto) do ambiente M . No entanto, nestes casos precisamos supor que $\mu(B) = \lim_n \mu_n(B)$; sem a hipótese adicional, o enunciado é falso em geral. [Observação: Dê exemplos!]

2.1.8. (Billingsley [Bil68]) Use a hipótese para mostrar que se $(U_n)_n$ é uma sequência crescente de abertos de M tais que $\cup_n U_n = M$ então para todo $\varepsilon > 0$ existe n tal que $\mu(U_n) \geq 1 - \varepsilon$ para todo $\mu \in \mathcal{K}$. Em seguida, imite a demonstração da Proposição A.3.7.

2.2.1. Por compacidade, passando a uma subsequência se necessário, podemos supor que $(\mu_n)_n$ converge para algum $\mu \in \mathcal{M}_1(M)$. Como $f_* : \mathcal{M}_1(M) \rightarrow \mathcal{M}_1(M)$ é contínua, $(f_*\mu_n)_n$ converge para $f_*\mu$. Então, dado qualquer $\varepsilon > 0$ e qualquer função contínuas limitada ϕ ,

$$\left| \frac{1}{\#I_n} \sum_{j \in I_n} \int \phi \circ f^j d\nu - \int \phi d\mu \right| < \varepsilon \quad \text{e} \quad \left| \frac{1}{\#I_n} \sum_{j \in I_n} \int \phi \circ f^{j+1} d\nu - \int (\phi \circ f) d\mu \right| < \varepsilon$$

para todo n suficientemente grande. Por outro lado,

$$\left| \frac{1}{\#I_n} \sum_{j \in I_n} \int \phi \circ f^{j+1} d\nu - \frac{1}{\#I_n} \sum_{j \in I_n} \int \phi f^j d\nu \right| \leq \frac{2}{\#I_n} \sup |\phi| < \varepsilon$$

para todo n suficientemente grande. Logo, $|\int(\phi \circ f) d\mu - \int \phi d\mu| < 3\varepsilon$ para todo ϕ e todo ε . Agora faça $\varepsilon \rightarrow 0$.

2.2.2. Para a primeira parte do enunciado use indução em q . O caso $q = 1$ corresponde ao Teorema 2.1. Considere transformações contínuas $f_i : M \rightarrow M$, $1 \leq i \leq q + 1$ que comutam entre si. Por hipótese de indução, existe probabilidade ν invariante por f_i , $1 \leq i \leq q$. Defina $\mu_n = (1/n) \sum_{j=0}^{n-1} (f_{q+1})^j_*(\nu)$. Note que $(f_i)_*\mu_n = \mu_n$ para todo $1 \leq i \leq q$ e todo n . Logo, todo ponto de acumulação de $(\mu_n)_n$ é invariante por todo f_i , $1 \leq i \leq q$. Por compacidade, existe algum ponto de acumulação $\mu \in \mathcal{M}_1(M)$. Verifique que μ é invariante por f_{q+1} . Para a segunda parte, represente por $M_q \subset \mathcal{M}_1(M)$ o conjunto das probabilidades invariantes por f_i , $1 \leq i \leq q$. Então $(M_q)_q$ é uma sequência não crescente de subconjuntos fechados não vazios de $\mathcal{M}_1(M)$. Por compacidade, a interseção $\cap_q M_q$ é não vazia.

2.2.3. Para cada $k \geq 1$ tome $a_k = 1/(10^k - 1)$. Note que $f^k(a_k) = a_k$ e portanto a probabilidade $\mu_k = (1/k)(\delta_{a_k} + \delta_{f(a_k)} + \dots + \delta_{f^{k-1}(a_k)})$ é invariante. (a) Considere $\mu = \sum_{k=1}^{\infty} 2^{-k} \mu_k$, onde μ_k é a probabilidade suportada na órbita do ponto periódico p_k introduzida anteriormente. Verifique que o fecho de $A = \{f^j(a_k) : 0 \leq j < k \text{ e } k \geq 1\}$ é enumerável. (b) Considere o conjunto de Cantor $K = \cap_{n=0}^{\infty} K_n$ definido da seguinte maneira: $K_0 = [0, 1]$ e cada K_{n+1} é obtido removendo de cada componente conexa C de K_n o subintervalo aberto com o mesmo centro que C e com comprimento igual a $8m(C)/10$. Tome a probabilidade μ suportada em K tal que $\mu(K \cap C) = 2^{-n}$ para toda componente conexas C de K_n . (c) Seja B o conjunto dos pontos $x \in [0, 1]$ cuja expansão decimal é periódica. Então B é subconjunto enumerável denso de $[0, 1]$. Além disso, todo $x \in B$ é periódico para f . Seja $\{b_k : k \geq 1\}$ um conjunto formado por exatamente um ponto em cada uma destas órbitas periódicas. Considere $\mu = \sum_{k=1}^{\infty} 2^{-k} \nu_k$ onde ν_k é a probabilidade invariante suportada na órbita de b_k . Verifique que $\mu(B) = 1$ e, portanto, $\text{supp } \mu = [0, 1]$. No entanto, $m(B) = 0$.

2.2.4. Pelo Exercício 2.2.2, existe alguma probabilidade μ invariante por $f^{1/q}$, para todo $q \geq 1$. Observe que μ é invariante por f^t para todo $t \in \mathbb{Q}$. Deduza que o mesmo vale para todo $t \in \mathbb{R}$.

2.2.5. Mostre que $\rho(x) = 1/\sqrt{1-x^2}$ satisfaz $\rho(y) = \sum_{x \in f^{-1}(y)} \rho(x)/|f'(x)|$ para todo y .

2.2.6. Defina μ em cada iterado $f^j(W)$, $j \in \mathbb{Z}$ colocando $\mu(A) = m(f^{-j}(A))$ para cada conjunto mensurável $A \subset f^j(W)$.

2.2.7. Seja W o conjunto dos $x \in X$ tais que $f^n(x) \notin X$ para todo $n \geq 1$ e seja $Y = X \setminus \cup_{k=0}^{\infty} f^{-k}(W)$. Os pontos de Y regressam infinitas vezes a X . Note que $f^i(W) \cap f^j(W) = \emptyset$ para todo $i < j$.

2.2.8. Por hipótese, f é invertível. Seja μ medida invariante por f absolutamente contínua com relação a m . Se μ é σ -finita μ então existem $X_1 \subset \dots \subset X_n \subset \dots$ com $\mu(X_n) < \infty$. A restrição μ_n de μ a cada X_n é invariante e absolutamente contínua relativamente à restrição de m a X_n (Seção 1.4.1). Reciprocamente, dados X_n e μ_n como no enunciado, considere $Y_1 = X_1$ e $Y_n = X_n \setminus (X_1 \cup \dots \cup X_{n-1})$, para $n > 1$, e seja ν_n a restrição e μ_n a Y_n . Use a construção na Seção 1.4.2 para obter, a partir de cada ν_n , uma medida $\tilde{\nu}_n$ invariante por f e tal que o complementar de $Z_n = \cup_j f^j(Y_n)$ tem medida nula. Observando que os Z_n são disjuntos dois-a-dois, considere a medida $\mu = \sum_n \tilde{\nu}_n$.

2.2.9. Considere a transformação $f : [0, 1] \rightarrow [0, 1]$ dada no Exemplo 1.4.5, com $\alpha \geq 1$. Usando o Exercício 1.4.2, mostre que f não admite medida invariante finita absolutamente contínua com relação à medida de Lebesgue m no intervalo $[0, 1]$.

2.3.1. (a) Seja $(a^k)_k$ uma sequência de Cauchy em ℓ^1 . Escreva $a^k = (a_n^k)_n$. Verifique que $(a_n^k)_k$ é sequência de Cauchy em \mathbb{C} para todo n . Seja b_n o seu limite e $b = (b_n)_n$. Mostre que $(\|b - a^k\|_1)_k \rightarrow 0$. Isto implica que $b \in \ell^1$ e que $(a^k)_k \rightarrow b$ em ℓ^1 . Seja $(a^k)_k$ uma sequência de Cauchy em ℓ^∞ . Como no caso anterior, cada $(a_n^k)_k$ é sequência de Cauchy em \mathbb{C} . Seja b_n o seu limite e seja $b = (b_n)_n$. Mostre que $(\|b - a^k\|_\infty)_k \rightarrow 0$. Isto implica que $b \in \ell^\infty$ e que $(a^k)_k \rightarrow b$ em ℓ^∞ . No caso particular em que $(a^k)_k$ é sequência em c_0 ,

ou seja, quando $\lim_n a_n^k = 0$ para todo k , a relação $\|b - a^k\|_\infty < \varepsilon$ para k grande implica $|b_n - a_n^k| < \varepsilon$ para k grande e todo n e isso implica $|b_n| < 2\varepsilon$ para n grande. Portanto, $(b_n)_n \rightarrow 0$ nesse caso e, então, o argumento anterior mostra que c_0 é completo. (b) Verifique que se $(a_n)_n \in \ell^\infty$ então $\phi : (b_n)_n \mapsto \sum_n a_n b_n$ é um funcional linear contínuo em ℓ^1 , com $\|\phi\| \leq \|(a_n)_n\|_\infty$. Reciprocamente, dado um funcional linear contínuo $\phi : \ell^1 \rightarrow \mathbb{C}$, considere $(a_n)_n$ definida por $a_n = \phi(0, \dots, 0, 1, 0, \dots)$, onde a coordenada igual a 1 está na n ésima posição. Mostre que $\|(a_n)_n\|_\infty \leq \|\phi\|$ e que $\phi((b_n)_n) = \sum_n a_n b_n$ para todo $(b_n)_n \in \ell^1$. Verifique que se $(a_n)_n \in \ell^1$ então $\phi : (b_n)_n \mapsto \sum_n a_n b_n$ é um funcional linear contínuo em c_0 , com $\|\phi\| \leq \|(a_n)_n\|_1$. Reciprocamente, dado um funcional linear contínuo $\phi : c_0 \rightarrow \mathbb{C}$, considere $(a_n)_n$ definida por $a_n = \phi(0, \dots, 0, 1, 0, \dots)$. Mostre que $\|(a_n)_n\|_1 \leq \|\phi\|$ e que $\phi((b_n)_n) = \sum_n a_n b_n$ para todo $(b_n)_n \in \ell^1$. [Observação: Não é verdade que o dual $(\ell^\infty)^*$ seja isomorfo a ℓ^1 . Por que o argumento que acabamos de dar não é válido nesse contexto?]

2.3.2. Claro que convergência em norma implica convergência fraca. Para provar a recíproca, suponha que $(x^k)_k$ converge para zero na topologia fraca mas não na topologia da norma. A primeira condição implica que, para todo N fixo, $\sum_{n=0}^N |x_n^k|$ converge para zero quando $k \rightarrow \infty$. A segunda condição significa que, a menos de restringir a uma subsequência, existe $\delta > 0$ tal que $\|x^k\| > \delta$ para todo k . Então, existe alguma sequência crescente $(l_k)_k$, tal que

$$\sum_{n=0}^{l_{k-1}} |x_n^k| \leq \frac{1}{k} \quad \text{mas} \quad \sum_{n=0}^{l_k} |x_n^k| \geq \|x^k\| - \frac{1}{k} \geq \delta - \frac{1}{k} \quad \text{para todo } k.$$

Tome $a_n = \overline{x_n^k} / |x_n^k|$ para cada $l_{k-1} < n \leq l_k$. Então, para todo k ,

$$\left| \sum_{n=0}^{\infty} a_n x_n^k \right| \geq \sum_{l_{k-1} < n \leq l_k} |x_n^k| - \sum_{n \leq l_{k-1}} |x_n^k| - \sum_{n > l_k} |x_n^k| \geq \|x^k\| - \frac{4}{k} \geq \delta - \frac{4}{k}.$$

Isto contradiz as hipóteses. Agora tome $x_n^k = 1$ se $k = n$ e $x_n^k = 0$ caso contrário. Dada qualquer $(a_n)_n \in c_0$, temos que $\sum_n a_n x_n^k = a_k$ converge para zero quando $k \rightarrow \infty$. Portanto, $(x^k)_k$ converge para zero na topologia fraca*. Mas $\|x^k\| = 1$ para todo k , logo $(x^k)_k$ não converge para zero na topologia da norma.

2.3.3. É claro que vale $\lambda C_0^+ \subset C_0^+$ para todo $\lambda \geq 0$ e $C_0^+ \cap (-C_0^+) = 0$. Além disso, dados $\phi, \psi \in C_0^+$ com $\|\phi\| = \|\psi\| = 1$, podemos tomar $x, y \in M$ tal que $\phi(x) = \psi(y) = 1$. Então $\|\phi + \psi\| \geq \phi(x) + \psi(x) \geq 1$. Assim, $\inf\{\|\phi + \psi\| : \phi, \psi \in C_0^+, \|\phi\| = \|\psi\| = 1\} \geq 1$.

2.3.4. Use a base de Fourier (A.6.1) de $L^2(m)$ para calcular autovalores e autovetores do operador de Koopman.

2.3.5. Considere a transformação expansão decimal $f : [0, 1] \rightarrow [0, 1]$ e seja m a medida de Lebesgue. Como $f(x + 1/10) = f(x)$ para todo $x \in [0, 1]$, temos que toda $\phi \in L^1(m)$ na imagem do operador de Koopman satisfaz $\psi(x + 1/10) = \psi(x)$ para m -quase todo x . Conclua que U_f não é sobrejetiva.

2.3.6. Tome $W = U(H)^\perp$ e $V = (\bigoplus_{n=0}^{\infty} U^n(W))^\perp$.

2.3.7. Suponha que existem funcionais tangentes T_1 e T_2 com $T_1(v) > T_2(v)$ para algum $v \in E$. Mostre que $\phi(u + tv) + \phi(u - tv) - 2\phi(u) \geq t(T_1(u) - T_2(u))$ para todo t e deduza que ϕ não é diferenciável na direção de v .

2.4.1. Considere o conjunto \mathcal{P} das probabilidades em $X \times M$ da forma $\nu^z \times \eta$. Note que \mathcal{P} é compacto na topologia fraca* e é invariante pelo operador F_* .

2.4.2. A condição $\hat{p} \circ g = \hat{f} \circ \hat{p}$ acarreta $\hat{f}^n \circ \hat{p} = \hat{p} \circ g^n$ para todo $n \in \mathbb{Z}$. Usando $\pi \circ \hat{p} = p$, segue que $\pi \circ \hat{f}^n \circ \hat{p} = p \circ g^n$ para todo $n \leq 0$. Portanto, $\hat{p}(y) = (p(g^n(y)))_{n \leq 0}$. Isto prova a existência e unicidade de \hat{p} . Agora suponha que p é sobrejetiva. As hipóteses de compacidade e continuidade garantem que

$$(g^{-n}(p^{-1}(\{x_n\})))_{n \leq 0}$$

é uma sequência encaixada de compactos, para todo $(x_n)_{n \leq 0} \in \hat{M}$. Tome y na interseção e note que $\hat{p}(y) = (x_n)_{n \leq 0}$.

2.4.3. Por definição, dado qualquer $y \in M_g$ e qualquer $n \geq 0$, existe $z_n \in M$ tal que $g^n(z_n) = y$. Em outras palavras, $g^{n-1}(z_n) \in g^{-1}(y)$ para todo n . No caso (a), seja z um

ponto de acumulação da sequência $(g^{n-1}(z_n))_n$. Mostre que $z \in M_g$ e $z \in g^{-1}(y)$. No caso (b) observe que, como $g^{-1}(y)$ é um conjunto finito, existe $z \in g^{-1}(y)$ tal que $g^{n-1}(z_n) = z$ para infinitos valores de n . Verifique que $z \in M_g$. Se μ é probabilidade g -invariante então $\mu(g^{n-1}(M) \setminus g^n(M)) = 0$ para todo $n \geq 1$ e, portanto, $\mu(M_g) = 1$. Logo μ é probabilidade f -invariante. Se $x \notin M_g$, existe $n \geq 1$ tal que $x \notin f^n(M)$. Usando que $f^n(M)$ é fechado, conclua que x não pode ser recorrente.

2.4.4. Suponha que M é separável e seja $\{x_k : k \geq 1\}$ um subconjunto enumerável denso. Para cada $k \geq 1$, escolha $\hat{x}_k \in \hat{M}$ tal que $\pi(\hat{x}_k) = x_k$; isto é possível porque f é sobrejetiva. Usando a hipótese de que f é contínua, e a definição (2.4.7) da distância \hat{d} , verifique que o conjunto $\{f^n(\hat{x}_k) : n \geq 0, k \geq 1\}$ é denso em \hat{M} . Logo \hat{M} é separável. Agora suponha que M é completo. Seja $(\hat{x}_k)_n$ uma sequência de Cauchy em \hat{M} . Escrevendo, $\hat{x}_k = (x_{k,n})_{n \leq 0}$, verifique que $(x_{k,n})_k$ é uma sequência de Cauchy em M e, portanto, existe $y_n = \lim_k x_{k,n}$, para cada $n \leq 0$. Lembrando que f é contínua, mostre que $f(y_n) = y_{n+1}$ para todo $n < 0$. Em outras palavras, a sequência $\hat{y} = (y_n)_{n \leq 0}$ está em \hat{M} . Mostre que $d(\hat{x}_k, \hat{y})$ vai para zero quando $k \rightarrow \infty$. Isto prova que \hat{M} é completo.

2.4.5. (a) Observe que \hat{M} é um subconjunto fechado de $M^{\mathbb{Z}_q}$ e conclua que \hat{M} é compacto. Além disso, se d é uma distância em M então a seguinte função (escrevemos $\bar{n} = (n_1, \dots, n_q)$) é distância em \hat{M} :

$$\hat{d}((x_{\bar{n}})_{\bar{n}}, (y_{\bar{n}})_{\bar{n}}) = \sum_{n_1, \dots, n_q \leq 0} 2^{n_1 + \dots + n_q} \min\{d(x_{\bar{n}}, y_{\bar{n}}), 1\}.$$

(b) A inversa de \hat{f}_i envia $(x_{n_1, \dots, n_i, \dots, n_q})_{\bar{n}}$ em $(x_{n_1, \dots, n_i-1, \dots, n_q})_{\bar{n}}$. Tanto \hat{f}_i quanto a inversa são contínuas porque cada coordenada da imagem é função contínua de um número finito de coordenadas da variável. Por definição,

$$\pi \hat{f}_i((x_{n_1, \dots, n_i, \dots, n_q})_{\bar{n}}) = x_0, \dots, 1, \dots, 0 = f_i \circ \pi((x_{n_1, \dots, n_i, \dots, n_q})_{\bar{n}})$$

em todo ponto. De modo semelhante, $\hat{f}_i \circ \hat{f}_j = \hat{f}_j \circ \hat{f}_i$. (c) É claro que π é contínua para a topologia herdada do espaço produto. Para mostrar que ela é sobrejetiva, considere $x \in M$. Como supomos que f_1, \dots, f_q são sobrejetivas, podemos encontrar uma sequência $(x_n)_{n \leq 0}$ em M tal que $x_0 = 0$ e $f_1 \cdots f_n(x_n) = x_{n+1}$ para cada $n < 0$. Defina $x_{n_1, \dots, n_q} = f_1^{n_1-n} \cdots f_q^{n_q-n}(x_n)$ para qualquer $n \leq \min\{n_1, \dots, n_q\}$; verifique que a definição não depende da escolha de n . Justifique que a sequência $(x_{n_1, \dots, n_q})_{\bar{n}}$ definida deste modo está em \hat{M} e a sua imagem por π é o ponto x .

2.4.6. (a) A inclusão \supset é trivial. Para provar \subset , verifique que se $n = \max\{n_1, \dots, n_q\}$ então $g_1^n \cdots g_q^n(M)$ está contido em $g_1^{n_1} \cdots g_q^{n_q}(M)$. (b) A inclusão $g_i(M_g) \subset M_g$ segue diretamente do item (a). Para provar a inclusão reversa, considere $y \in M_g$. Usando o item (a), para cada $n \geq 1$ existe $z_n \in M$ tal que $g_i g_1^n \cdots g_q^n(z_n) = y$. Seja z um ponto de acumulação da sequência $(g_1^n \cdots g_q^n(z_n))_n$. Verifique que $z \in M_g$ e $z \in g_i^{-1}(y)$.

2.4.7. Pelo Exercício 2.4.6, não é restrição supor que as transformações f_1, \dots, f_2 são sobrejetivas. Sejam $\hat{f}_1, \dots, \hat{f}_q : \hat{M} \rightarrow \hat{M}$ as extensões naturais, no sentido do Exercício 2.4.5. Pelo caso invertível do Teorema 1.5.1, existe algum $\hat{x} \in \hat{M}$ que é simultaneamente recorrente para $\hat{f}_1, \dots, \hat{f}_q$. Verifique que $x = \pi(\hat{x})$ é simultaneamente recorrente para f_1, \dots, f_q .

2.5.1. Basta notar que $\liminf_n x_n = 1 - \limsup_n (1 - x_n)$ e $\liminf_n (x_n) + \liminf_n (y_n) \leq \liminf_n (x_n + y_n)$ e $\limsup_n (x_n) + \limsup_n (y_n) \leq \limsup_n (x_n + y_n)$.

2.5.2. Fixe q e l . Suponha que para todo $n \geq 1$ existe uma partição $\{S_1^n, \dots, S_l^n\}$ do conjunto $\{1, \dots, n\}$ tal que nenhum dos subconjuntos S_j^n contém uma progressão aritmética com comprimento q . Considere a função $\phi_n : \mathbb{N} \rightarrow \{1, \dots, l\}$ dada por $\phi_n(i) = j$ se $i \in S_j^n$ e $\phi_n(i) = l$ se $i > n$. Tome $(n_k)_k \rightarrow \infty$ tal que a subsequência $(\phi_{n_k})_k$ converge em todo ponto para alguma função $\phi : \mathbb{N} \rightarrow \{1, \dots, l\}$. Considere $S_j = \phi^{-1}(j)$ para $j = 1, \dots, l$. Algum S_j contém alguma progressão aritmética de comprimento q . Então $S_j^{n_k}$ contém essa progressão aritmética para todo k suficientemente grande.

2.5.3. Tome uma sequência de partições $\mathcal{P}_\ell = \{P_1^\ell, \dots, P_{k_\ell}^\ell\}$ de M com $\text{diam } \mathcal{P}_\ell \rightarrow 0$ e seja $x \in M$ um ponto arbitrário. Para cada ℓ , associe a x a sequência $(a_i)_{i \in \mathbb{Z}}$ de elementos

de $\{1, 2, \dots, k_\ell\}$ dada por $f^i(x) \in P_j^\ell$. Isso determina uma partição $\{S_1^\ell, \dots, S_{k_\ell}^\ell\}$ de \mathbb{Z} , definida por $i \in S_j^\ell \Leftrightarrow a_i = j$. Pelo Teorema de van der Waerden, existe $m_\ell \in \{1, 2, \dots, k_\ell\}$ tal que $S_{m_\ell}^\ell$ contém progressões aritméticas $c_k, c_k + n, \dots, c_k + kn \in S_{m_\ell}^\ell$ de comprimento $k + 1$ arbitrariamente grande. Note que $f^{c_k}(x) \in \cap_{j=0}^k f^{-jn}(P_{m_\ell}^\ell)$. Escolha qualquer ponto $y_\ell \in P_{m_\ell}^\ell$. Por compacidade, podemos supor que $(y_\ell)_\ell$ converge para $y \in M$. Afirmamos que tal y é super não errante. De fato, dado qualquer vizinhança U de y , seja $r > 0$ tal que $B(y, r) \subset U$. Considere ℓ é suficientemente grande para que $P_{m_\ell}^\ell \subset B_{r/2}(y_\ell) \subset B_r(y) \subset U$. Tome k e n como anteriormente. Então $\cap_{j=0}^k f^{-jn}(U) \supset \cap_{j=0}^k f^{-jn}(P_{m_\ell}^\ell) \neq \emptyset$, tal como queríamos demonstrar. Em seguida explicaremos como obter o teorema de van der Waerden a partir da propriedade no enunciado. Seja $\{S_1, \dots, S_\ell\}$ uma partição de \mathbb{Z} . Considere $a = (a_j)_{j \in \mathbb{Z}}$ com $a_j \in \{1, \dots, \ell\}$ definido por $j \in S_{a_j}$. Seja M o fecho da órbita de a pelo deslocamento bilateral σ . Por hipótese, σ possui algum ponto super não errante $y = (y_j)_{j \in \mathbb{Z}}$. Seja $U = [0; y_0]$ o cilindro de comprimento 1 que contém y . Dado qualquer $k \geq 1$, existe $n \geq 1$ e existe $z = (z_j)_{j \in \mathbb{Z}}$ tal que $z \in \cap_{j=0}^k \sigma^{-jn}(U)$. Isso significa que $z_{jn} = y_0$ para $j = 0, \dots, k$. Como z é acumulado por iterados de a , concluímos que existe $m \in \mathbb{Z}$ tal que $m + jn \in S_{y_0}$ para $j = 0, \dots, k$ (tal como afirma o teorema de van der Waerden). [Observação: O mesmo argumento mostra que toda aplicação invertível, não necessariamente contínua, num espaço compacto possui algum ponto não errante.]

2.5.4. Considere $\Sigma = \{1, \dots, l\}^{\mathbb{N}^k}$, munido da distância $d(\omega, \omega') = 2^{-N}$ onde $N \geq 0$ é máximo tal que $\omega(i_1, \dots, i_k) = \omega'(i_1, \dots, i_k)$ para todo $i_1, \dots, i_k < N$. Note que Σ é um espaço métrico compacto. Dado $q \geq 1$, seja $F_q = \{(a_1, \dots, a_k) : 1 \leq a_i \leq q \text{ e } 1 \leq i \leq k\}$. Seja e_1, \dots, e_m uma enumeração dos elementos de F_q . Para cada $j = 1, \dots, m$, considere o deslocamento $\sigma_j : \Sigma \rightarrow \Sigma$ dado por $(\sigma_j \omega)(n) = \omega(n + e_j)$ para $n \in \mathbb{N}^k$. Considere o ponto $\omega \in \Sigma$ definido por $\omega(n) = i \Leftrightarrow n \in S_i$. Seja Z o fecho de $\{\sigma_1^{l_1} \dots \sigma_m^{l_m}(\omega) : l_1, \dots, l_m \in \mathbb{N}\}$. Note que Z é invariante pelos deslocamentos σ_j . Pelo teorema de recorrência múltipla de Birkhoff, existem $\zeta \in Z$ e $s \geq 1$ tais que $d(\sigma_j^s(\zeta), \zeta) < 1$ para todo $j = 1, \dots, m$. Seja $e = (1, \dots, 1) \in \mathbb{N}^k$. Então $\zeta(e) = \zeta(e + se_1) = \dots = \zeta(e + se_m)$. Considere $\sigma_1^{l_1} \dots \sigma_m^{l_m}(\omega)$ suficientemente próximo de ζ para que $\omega(b) = \omega(b + se_1) = \dots = \omega(b + se_m)$, onde $b = e + l_1 e_1 + \dots + l_m e_m$. Segue que se $i = \omega(b)$, então $b + sF_q \subset S_i$. Dado que temos apenas uma quantidade finita de conjuntos S_i , algum deles deve conter infinitos conjuntos do tipo $b + sF_q$, com q arbitrariamente grande, e isto termina a prova.

3.1.1. Imite a prova do Teorema 3.1.5.

3.1.2. Suponha o contrário, ou seja, que para todo $k \in \mathbb{N}$ existe $n_k \in \mathbb{N}$ tal que $\mu(A \cap f^{-j}(A)) = 0$ para todo $n_k + 1 \leq j \leq n_k + k$. Não é restrição supor que $(n_k)_k \rightarrow \infty$. Tome $\varphi = \chi_A$. Pelo Exercício 3.1.1, temos que $(1/k) \sum_{j=n_k+1}^{n_k+k} \varphi \circ f^j \rightarrow \varphi \cdot P(\varphi)$. O lado esquerdo é identicamente nulo e o lado direito é igual a $\|P(\varphi)\|^2$. Logo a média temporal $P(\varphi) = 0$ e, portanto, $\mu(A) = \int P(\varphi) d\mu = 0$.

3.1.3. Claro que F é um subespaço. Suponha que $(\varphi_n)_n \rightarrow \varphi$ com $\varphi_n \circ f = \varphi_n$ em μ -quase todo ponto. Então $\varphi \circ f = \varphi$ em μ -quase todo ponto:

$$\int |\varphi \circ f - \varphi| d\mu = \lim_n \int |\varphi_n \circ f - \varphi| d\mu = \lim_n \int |\varphi_n - \varphi| d\mu = 0.$$

3.1.4. O teorema afirma que $(1/T) \int_0^T \varphi(f^t(x)) dt$ converge para $P(\varphi)$ em $L^2(\mu)$ quando $T \rightarrow \infty$, onde $P(\cdot)$ é a projeção no espaço das funções invariantes. Para provar, seja $\psi(x) = \int_0^1 \varphi(f^t(x)) dt$ e $\Psi(x) = (\int_0^1 |\varphi|^2(f^t(x)) dt)^{1/2}$. Usando a desigualdade de Cauchy-Schwarz, $|\psi|^2 \leq \Psi^2$. Usando o teorema de Fubini e o fato de que a medida μ é invariante, $\int \Psi^2 d\mu = \int |\varphi|^2 d\mu$. Logo, $\psi, \Psi \in L^2(\mu)$. Escreva

$$\frac{1}{T} \int_0^T \varphi(f^t(x)) dt = \frac{1}{T} \sum_{j=0}^{[T]} \psi(f^j(x)) + \frac{1}{T} \int_{[T]}^T \varphi(f^j(x)) dt.$$

Pelo Teorema 3.1.5 aplicado a ψ , o primeiro termo do lado direito converge para $P(\psi)$ em $L^2(\mu)$. Observe que $P(\psi) = \int_0^1 P(\varphi) \circ f^t dt = P(\varphi)$, uma vez que $P(\varphi)$ é função invariante. O último termo é limitado por $(1/T)\Psi(f^{[T]}(x))$ e, portanto, converge para zero em $L^2(\mu)$.

3.1.5. Basta mostrar que $\ker A = \bigcap_{t \in \mathbb{R}} \ker(U_t - \text{id})$. Se $A\varphi = 0$ então $e^{2\pi itA}\varphi = \varphi$ para todo $t \in \mathbb{R}$. Reciprocamente, suponha que $e^{2\pi itA}\varphi = \varphi$ para todo $t \in \mathbb{R}$. Derivando em relação a t , obtemos que $e^{2\pi itA}(2\pi iA\varphi) = 0$. Como U_t é isomorfismo, isso implica que $A\varphi = 0$.

3.2.1. Verificação direta a partir das definições.

3.2.2. Observe que a função $\psi = \varphi - \varphi \circ f$ é integrável e satisfaz $\int \psi d\mu = 0$. Aplique o teorema de Birkhoff a ψ .

3.2.3. (a) Considere $\varepsilon = 1$ e seja $C = \sup\{|\varphi(l)| : |l| \leq L(1)\}$. Dado $n \in \mathbb{Z}$, fixe $s \in \mathbb{Z}$ tal que $sL(1) < n \leq (s+1)L(1)$. Por hipótese, existe $\tau \in \{sL(1)+1, \dots, (s+1)L(1)\}$ tal que $|\varphi(k+\tau) - \varphi(k)| < 1$ para todo $k \in \mathbb{Z}$. Tome $k = n - \tau$ e observe que $|k| \leq L(1)$. Segue que $|\varphi(n)| < 1 + C$. (b) Tome $\rho\varepsilon > 2L(\varepsilon) \sup|\varphi|$. Para todo $n \in \mathbb{Z}$ existe algum ε -quase período $\tau = n\rho + r$ com $1 \leq r \leq L(\varepsilon)$. Então,

$$\left| \sum_{j=n\rho+1}^{(n+1)\rho} \varphi(j) - \sum_{j=1-r}^{\rho-r} \varphi(j) \right| < \rho\varepsilon \quad \text{e} \quad \left| \sum_{j=1-r}^{\rho-r} \varphi(j) - \sum_{j=1}^{\rho} \varphi(j) \right| \leq 2r \sup|\varphi| < \rho\varepsilon.$$

(c) Dado $\varepsilon > 0$, tome ρ como no item (b). Para cada $n \geq 1$, escreva $n = s\rho + r$, com $1 \leq r \leq \rho$. Então,

$$\frac{1}{n} \sum_{j=1}^n \varphi(j) = \frac{\rho}{s\rho + r} \sum_{i=0}^{s-1} \frac{1}{\rho} \sum_{l=i\rho+1}^{(i+1)\rho} \varphi(l) + \frac{1}{n} \sum_{l=s\rho+1}^{s\rho+r} \varphi(l).$$

Para s grande, o primeiro termo do lado direito está próximo de $(1/\rho) \sum_{j=0}^{\rho-1} \varphi(j)$ (item (b)) e o último termo está próximo de zero (item (a)). Conclua que o lado esquerdo da igualdade constitui uma sequência de Cauchy. (d) Observe que

$$\left| \frac{1}{n} \sum_{j=1}^n \varphi(x+k) - \frac{1}{n} \sum_{j=1}^n \varphi(j) \right| \leq \frac{2|x|}{n} \sup|\varphi|$$

e use os itens (a) e (c).

3.2.4. Observe que a função $\phi = \sum_{k=1}^{\infty} \log k \mathcal{X}_{I_k}$, onde $I_k = (1/(k+1), 1/k]$, é integrável com relação à medida de Gauss μ . Aplique o teorema de Birkhoff a ϕ .

3.2.5. O caso $p = 2$ foi feito na Seção 3.2.3.

3.2.6. Dado $\varphi \in L^1(\mu)$, defina $\psi(x) = \int_0^1 \varphi(f^t(x)) dt$ e $\Psi(x) = \int_0^1 |\varphi(f^t(x))| dt$. Então $\int |\psi| d\mu \leq \int \Psi d\mu \leq \int |\varphi| d\mu$, logo $\psi, \Psi \in L^1(\mu)$. Além disso

$$\frac{1}{T} \int_0^T \varphi(f^t(x)) dt = \frac{1}{T} \sum_{j=0}^{[T]-1} \psi(f^j(x)) + \frac{1}{T} \int_{[T]}^T \varphi(f^t(x)) dt.$$

Pelo teorema de Birkhoff aplicado a (f^1, μ) e à função ψ , o primeiro termo do lado direito converge em quase todo ponto quando $T \rightarrow \infty$. O último termo é limitado por $(1/T)\Psi(f^{[T]}(x))$. Usando o Lema 3.2.5, segue que ele converge para zero quando $T \rightarrow \infty$.

3.2.7. O exercício está resolvido na Proposição 6.2.1.

3.3.1. Como $\varphi_{n+1} \leq \varphi_n \circ f + \varphi_1$ para todo n , temos que $\varphi_- \leq \varphi_- \circ f$. Logo, a primeira afirmação segue de $\int_M (\varphi_- \circ f - \varphi_-) d\mu = 0$. A segunda é análoga.

3.3.2. Seja μ uma probabilidade invariante por um fluxo $f^t : M \rightarrow M$, $t \in \mathbb{R}$ e seja $(\varphi_s)_{s>0}$ uma família de funções, indexada pelos reais positivos, tal que $\varphi_{s+t} \leq \varphi_t + \varphi_s \circ f^t$ e a função $\Phi = \sup_{0 < s < 1} \varphi_s^+$ está em $L^1(\mu)$. Então $(1/T)\varphi_T$ converge em μ -quase todo ponto para uma função φ tal que $\varphi^+ \in L^1(\mu)$ e $\int \varphi d\mu = \lim_{T \rightarrow \infty} (1/T) \int \varphi_T d\mu$. Para provar este enunciado, tome $\varphi = \lim_n (1/n)\varphi_n$ (Teorema 3.3.3). Para $T > 0$ não inteiro, escreva $T = n + s$ com $N \in \mathbb{N}$ e $s \in (0, 1)$. Então

$$\varphi_T \leq \varphi_n + \varphi_s \circ f^n \leq \varphi_n + \Phi \circ f^n \quad \text{e} \quad \varphi_T \geq \varphi_{n+1} - \varphi_{1-s} \circ f^T \geq \varphi_n - \Phi \circ f^T.$$

Usando o Lema 3.2.5, a primeira desigualdade mostra que $\limsup_{T \rightarrow \infty} (1/T)\varphi_T \leq \varphi$. Analogamente, a segunda desigualdade, usando a versão do Lema 3.2.5 para tempo contínuo, dá que $\liminf_{T \rightarrow \infty} (1/T)\varphi_T \geq \varphi$. Também segue que $\lim_{T \rightarrow \infty} (1/T) \int \varphi_T d\mu$ coincide com $\lim_n (1/n) \int \varphi_n d\mu$. Pelo Teorema 3.3.3, este último limite é igual a $\int \varphi d\mu$.

3.3.3. Combine o Lema 1.3.5 com a propriedade (c2) na Seção 3.3.5.

3.3.4. Defina $\psi_n = n\tau$ e $\gamma_n = \varphi_n - n\tau$, onde τ é a constante temporal.

3.3.5. Como $\log^+ \|\phi\| \in L^1(\mu)$, para todo $\varepsilon > 0$ existe $\delta > 0$ tal que $\mu(B) < \delta$ implica $\int_B \log^+ \|\theta\| d\mu < \varepsilon$. Usando que $\log^+ \|\phi^n\| \leq \sum_{j=0}^{n-1} \log^+ \|\theta\| \circ f^j$, vem que

$$\mu(E) < \delta \quad \Rightarrow \quad \frac{1}{n} \int_E \log^+ \|\phi^n\| d\mu \leq \frac{1}{n} \sum_{j=0}^{n-1} \int_{f^{-j}(E)} \log^+ \|\theta\| d\mu \leq \varepsilon.$$

3.3.6. Fixe $k \geq 1$. Por subaditividade,

$$\lambda_{\max}(x) = \lim_n \frac{1}{nk} \log \|\phi^{nk}(x)\| \leq \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \frac{1}{k} \log \|\phi^k(f^{jk}(x))\| = \Psi_k(x).$$

Dados $\rho > 0$ e $k \geq 1$, defina $\Delta_k = \{x : \psi_k(x) > \lambda_{\max}(x) + \rho/2\}$. Represente por \mathcal{X}_k a função característica de Δ_k . Então,

$$\Psi_k(x) \leq (\lambda_{\max}(x) + \frac{\rho}{2}) + \lim_n \frac{1}{n} \sum_{j=0}^{n-1} (\psi_k^+ \mathcal{X}_k)(f^{jk}(x)).$$

Seja Γ_k o conjunto dos pontos x onde este último termo é maior que $\rho/2$. Se $x \notin \Gamma_k$ então $\Psi_k(x) \leq \lambda_{\max}(x) + \rho$. Portanto, basta mostrar que μ -quase todo x pertence a Γ_k^c para algum k . Usando o lema de Fatou, vemos que

$$\frac{\rho}{2} \mu(\Gamma_k) \leq \int \lim_n \frac{1}{n} \sum_{j=0}^{n-1} (\psi_k^+ \mathcal{X}_k) \circ f^{jk} d\mu \leq \lim_n \int \frac{1}{n} \sum_{j=0}^{n-1} (\psi_k^+ \mathcal{X}_k) \circ f^{jk} d\mu = \int_{\Delta_k} \psi_k^+ d\mu.$$

Como $\lambda_{\max} = \lim_k \psi_k$, temos que $\mu(\Delta_k)$ converge para zero. Por integrabilidade uniforme, segue que $\int_{\Delta_k} \psi_k d\mu$ converge para zero. Logo $\mu(\Gamma_k) \rightarrow 0$ e isso implica o que dissemos.

3.4.1. Se $A, B \subset \Sigma_\rho$ são disjuntos então A_δ e B_δ também são, para $\delta < \rho$. Segue que (3.4.6) é σ -aditiva. Quanto a (3.4.7): se $A_n \subset \Sigma$ são disjuntos dois-a-dois,

$$\sum_n \mu(A_n) = \sum_n \sup_\rho \mu(A_n \cap \Sigma_\rho) = \sup_\rho \sum_n \mu(A_n \cap \Sigma_\rho) = \sup_\rho \mu(\cup_n A_n \cap \Sigma_\rho) = \mu(\cup_n A_n).$$

3.4.2. (a) Suponha que $f^{n_j}(x) \rightarrow x$. Então $\pi(f^{n_j}(x), s) \rightarrow \pi(x, s)$ para todo s . Pela condição (3.4.1), existe $(t_j)_j \rightarrow \infty$ tal que $\pi(f^{n_j}(x), 0) = \pi(x, t_j)$ para cada j . Logo, $g^{t_j}(\pi(x, s)) = \pi(x, s + t_j) = \pi(f^{n_j}(x), s)$ para todo s . Em outras palavras, $g^{t_j}(\pi(x, s)) \rightarrow \pi(x, s)$. (b) Suponha que $g^{t_j}(\pi(x, s)) \rightarrow \pi(x, s)$ com $t_j \rightarrow \infty$. Fixe qualquer $0 < \sigma < \tau(x)$. Como $g^{\sigma-s}$ é contínuo, temos que $g^{t_j}(\pi(x, \sigma)) \rightarrow \pi(x, \sigma)$. Para cada j , tome n_j inteiro e $0 \leq \sigma_j < \tau(f^{n_j}(x))$ tais que $g^{t_j}(\pi(x, \sigma)) = \pi(f^{n_j}(x), \sigma_j)$. Então $\pi(f^{n_j}(x), \sigma_j) \rightarrow \pi(x, \sigma)$ e, pela escolha de σ e σ_j , isso implica que $f^{n_j}(x) \rightarrow x$ e $\sigma_j \rightarrow \sigma$. (c) A primeira afirmação é consequência direta dos itens (a) e (b) e da definição da medida ν . A segunda segue usando o teorema de recorrência de Poincaré.

3.4.3. Considere coordenadas locais $x = (x_1, x_2, \dots, x_d)$ tais que $\Sigma \subset \{x_1 = 0\}$. Escreva $\nu = \psi(x) dx_1 dx_2 \dots dx_d$. Então $\nu_\Sigma = \psi(y) dx_2 \dots dx_d$ com $y = (0, x_2, \dots, x_d)$. Dados $A \subset \Sigma$ e $\delta > 0$, a aplicação $\xi : (t, y) \mapsto g^t(y)$ é um difeomorfismo de $[0, \delta] \times A$ em A_δ . Portanto, $\nu(A_\delta) = \int_{[0, \delta] \times A} (\psi \circ \xi) |\det D\xi| dt dx_2 \dots dx_d$ e, consequentemente,

$$\lim_{\delta \rightarrow 0} \frac{\nu(A_\delta)}{\delta} = \int_A \psi(y) |\det D\xi|(y) dx_2 \dots dx_d.$$

Em seguida, note $|\det D\xi|(y) = |X(y) \cdot (\partial/\partial t)| = \phi(y)$ para todo $y \in \Sigma$. Segue que o transporte de ν coincide com a medida $\eta = \phi\nu_\Sigma$. Em particular, η é invariante pela transformação de Poincaré.

3.4.4. (a) As afirmações $\hat{a} \circ F \equiv \hat{a}$ e $G^t \circ F = F \circ G^t$ e $\hat{a} \circ G^t \equiv \hat{a}$ seguem diretamente das definições. As duas primeiras garantem a existência de a e de $(g^t)_t$ e a terceira implica que o funcional é invariante pelo fluxo. Também temos que $\det DF \equiv 1$ e $\det DG^t \equiv 1$, logo $\hat{\nu}$ é invariante por F e por G^t para todo t . Note que $\text{grad } \hat{a} \neq 0$ em todo ponto. Logo, podemos usar o Exercício 1.3.12 para concluir que η_c é invariante por F e por G^t para todo t e todo c . (b) Dado $(\lambda_1, \lambda_2, h_1, h_2)$, escreva $(\lambda_1^n, \lambda_2^n, h_1^n, h_2^n) = F^n(\lambda_1, \lambda_2, h_1, h_2)$ para $n \in \mathbb{Z}$. Por definição, $\lambda_1^n + \lambda_2^n = \max\{\lambda_1^{n-1}, \lambda_2^{n-1}\}$ para todo n . Verifique que $(\lambda_1^n)_n$ e $(\lambda_2^n)_n$ são não crescentes e convergem para zero quando $n \rightarrow +\infty$ e para infinito quando $n \rightarrow -\infty$. Conclua que existe um único n tal que $\lambda_1^n + \lambda_2^n \geq 1 > \max\{\lambda_1^n, \lambda_2^n\}$. Isto mostra que D é um domínio fundamental para \sim . A medida ν é infinita porque $\hat{\nu}(D) = \infty$. Dado outro domínio fundamental D' , podemos escrever D e D' como uniões enumeráveis disjuntas $D = \cup D_k$ e $D' = \cup D'_k$ de tal modo que $D'_k = F^k(D_k)$ para todo k . Como $\hat{\nu}$ é invariante por F , segue que $\pi(\hat{\nu} | D') = \sum_k \pi(\hat{\nu} | D'_k) = \sum_k \pi(\hat{\nu} | D_k) = \pi(\hat{\nu} | D)$. Para verificar que ν é invariante por $(g^t)_t$ imite a Proposição 3.4.1. (c) O campo de vetores $X = (\partial G^t / \partial t) |_{t=0}$ está dado por $X(\lambda_1, \lambda_2, h_1, h_2) = (\lambda_1, \lambda_2, -h_1, -h_2)$. É claro que ele é transversal à seção $\Sigma = \{\lambda_1 + \lambda_2 = 1\}$. O tempo de primeiro retorno a Σ é a solução de $\max\{e^\tau \lambda_1, e^\tau \lambda_2\} = 1$, ou seja, $\tau = \min\{-\log \lambda_1, -\log \lambda_2\}$. Isto dá:

$$f(\lambda_1, \lambda_2, h_1, h_2) = \begin{cases} (1 - \lambda_2/\lambda_1, \lambda_2/\lambda_1, \lambda_1 h_1, \lambda_1(h_1 + h_2)) & \text{se } \lambda_1 > \lambda_2 \\ (\lambda_1/\lambda_2, 1 - \lambda_1/\lambda_2, \lambda_2(h_1 + h_2), \lambda_2 h_2) & \text{se } \lambda_1 < \lambda_2. \end{cases}$$

Seja ds o comprimento de arco no segmento $\{(\lambda_1, \lambda_2) \in \mathbb{R}_+^2 : \lambda_1 + \lambda_2 = 1\}$. A medida de volume induzida pela métrica Riemanianna em Σ está dada por $dsdh_1dh_2$. O campo de vetores $n(\lambda_1, \lambda_2, h_1, h_2) = (1/\sqrt{2}, 1/\sqrt{2}, 0, 0)$ é unitário e ortogonal a Σ . Note que $X \cdot n \equiv 1/\sqrt{2}$ em Σ . Pelo Exercício 3.4.3, segue que a medida transporte está dada por $\mu = (1/\sqrt{2}) dsdh_1dh_2$. Note que $\mu(\Sigma) = \infty$. (d) O item (a) implica que N_c é invariante e, imitando a Proposição 3.4.1, que η_c é invariante. Para cada (λ_1, λ_2) , o segmento $\{(h_1, h_2) : a(\lambda_1, \lambda_2, h_1, h_2) = c\}$ pode ser parametrizado por $h_1 = cs/\lambda_1$ e $h_2 = c(1-s)/\lambda_2$, com $s \in (0, 1)$. Como $\text{grad } \hat{a}(\lambda_1, \lambda_2, h_1, h_2) = (h_1, h_2, \lambda_1, \lambda_2)$, segue que

$$\eta_c(N_c) = \int \int_D d\lambda_1 d\lambda_2 \int_0^1 \frac{\sqrt{(c/\lambda_1)^2 + (c/\lambda_2)^2} ds}{\sqrt{(cs/\lambda_1)^2 + (c(1-s)/\lambda_2)^2 + \lambda_1^2 + \lambda_2^2}}.$$

A função integranda é menor do que $2 \max\{\lambda_1^{-1}, \lambda_2^{-1}\}$. Portanto,

$$\eta_c(N_c) \leq \int \int_D 2 \max\{\lambda_1^{-1}, \lambda_2^{-1}\} d\lambda_1 d\lambda_2 < \infty.$$

Segue do Teorema 1.2.4 que η_c -quase todo ponto é recorrente para $(g^t)_t$. Como η_c é equivalente a $\nu_c = \pi_*(\hat{\nu}_c \cap D)$, segue que ν_c -quase todo ponto é recorrente para $(g^t)_t$, para todo c . Logo, ν -quase todo ponto é recorrente.

4.1.1. Se A é um conjunto invariante então a órbita de p está contida ou em A ou no seu complementar.

4.1.2. Use o teorema de Birkhoff e o teorema da convergência dominada.

4.1.3. (a) \Rightarrow : Se A é invariante então $\cup_n f^{-n}(A) = A$. (a) \Leftarrow : $B = \cup_n f^{-n}(A)$ satisfaz $f^{-1}(B) \subset B$. (b) \Rightarrow : Se A é invariante então $\mu(f^{-n}(A) \cap A^c) = 0$, logo $\mu(A)\mu(A^c) = 0$. (b) \Leftarrow : Se μ é ergódica, $\cup_n f^{-n}(A)$ tem medida 1, logo intersecta todo conjunto com medida positiva. (c) \Rightarrow : Para cada $\varphi \in L^p(\mu)$ os dois lados de (4.1.2) são funções lineares contínuas de $\psi \in L^q(\mu)$. Logo, o conjunto onde vale a igualdade é fechado. Esta observação permanece válida se permutarmos os papéis de φ e ψ . \Rightarrow (c): Está contido na Proposição 4.1.4. (d) \Rightarrow : Funções características estão em $L^p(\mu)$. \Rightarrow (d): Use (4.1.2) com $\psi \equiv 1$. (e) \Rightarrow : As funções características de conjuntos invariantes satisfazem a hipótese. \Rightarrow (e): $\int \varphi \circ f d\mu = \int \varphi d\mu$, logo a hipótese implica que φ é função invariante.

4.1.4. Seja A um conjunto invariante. Pelo Exercício A.5.7, existem funções uniformemente contínuas limitadas $(\varphi_k)_k$ convergindo para \mathcal{X}_A em μ -quase todo ponto e em $L^2(\mu)$. Use o teorema de von Neumann para verificar que as médias temporais $(\tilde{\varphi}_k)_k$ convergem para

$\tilde{\mathcal{X}}_A = \mathcal{X}_A$. Então a hipótese implica que \mathcal{X}_A é constante em μ -quase todo ponto, ou seja, $\mu(A) = 0$ ou $\mu(A^c) = 0$.

4.1.5. Pelo teorema de Birkhoff, a média temporal é uma função invariante. Isso dá a primeira afirmação. A segunda está contida na Proposição 4.1.3.

4.1.6. Use o Exercício 4.1.5.

4.1.7. Este exercício está resolvido no Exemplo 7.1.10.

4.1.8. Suponha que $U_f \varphi = \lambda \varphi$. Como U_f é isometria, $|\lambda| = 1$. Se $\lambda^n = 1$ para algum n então $\varphi \circ f^n = \varphi$ e, por ergodicidade, φ é constante em quase todo ponto. Caso contrário, dado qualquer $c \neq 0$, os conjuntos $\varphi^{-1}(\lambda^{-k}c)$, $k \geq 0$ são disjuntos dois-a-dois. Como a sua medida é independente de k , ela precisa ser nula. Finalmente, $\varphi^{-1}(c)$ é um conjunto invariante por f e, conseqüentemente, a sua medida é zero ou total.

4.2.1. Escreva $\varphi(x) = \sum_{k \in \mathbb{Z}^d} a_k e^{2\pi i k \cdot x}$. Verifique que $\varphi \circ R_\theta$ se, e somente se, $a_k e^{2\pi i k \cdot \theta} = a_k$ para todo $k \in \mathbb{Z}^d$. para θ racionalmente independente, isto implica que $a_k = 0$ para todo $k \neq 0$.

4.2.2. Considere $\Sigma = X^{\mathbb{N}}$; o outro caso é análogo. Para ver que σ é contínua, observe que $\sigma^{-1}([m; A_m, \dots, A_n]) = [m+1; A_m, \dots, A_n]$. Escrevendo $X = \{1, \dots, d\}$, considere o ponto $x = (1, 2, \dots, d, 1, 1, 1, 2, \dots, d, 1, 1, 1, 1, 2, \dots) \in \Sigma$ e verifique que $\{\sigma^n(x) : n \geq 1\}$ intersecta todo cilindro de Σ . Dado qualquer cilindro $[m; A_m, \dots, A_{m+p-1}]$, considere a seqüência $x = (x_k)_n$ tal que $x_k \in A_k$ para $m \leq k < m+p$ e $x_k = x_{k+p}$ para todo $k \geq 0$. Verifique que x é ponto periódico de σ .

4.2.3. Adaptando a demonstração da Proposição 4.2.4, mostre que se A é subconjunto invariante com medida positiva então $A \cap [(i-1)/3, i/3]$ tem medida total em $(i-1)/3, i/3]$ para algum $i \in \{1, 2, 3\}$. Deduza que A tem medida total em $[0, 1]$.

4.2.4. Podemos supor que X contém uma seqüência infinita de órbitas periódicas $(\mathcal{O}_n)_n$ com período indo para infinito. Seja $Y \subset X$ o conjunto de pontos de acumulação dessa seqüência. Mostre que Y não pode consistir de um único ponto. Sejam $p \neq q$ pontos periódicos em Y e seja z um ponto heteroclínico, ou seja, tal que $f^n(z)$ converge para a órbita de p quando $n \rightarrow -\infty$ e para a órbita de q quando $n \rightarrow +\infty$. Mostre que $z \in Y$ e deduza a conclusão do exercício.

4.2.5. Comece por verificar que se $\{U_{n,\alpha} : \alpha\}$ é base de uma topologia \mathcal{T}_n (respectivamente, gerador de uma σ -álgebra \mathcal{B}_n), para cada $n \geq 1$, então os cilindros $[1; U_{1,\alpha_1}, \dots, U_{n,\alpha_n}]$ formam uma base da topologia produto $\prod_n \mathcal{T}_n$ (respectivamente, um gerador da σ -álgebra produto $\prod_n \mathcal{B}_n$). Mostre também que se $(U_k)_k$ é base enumerável de abertos de um espaço topológico então ela gera a σ -álgebra de Borel do espaço. Combinando estas observações mostre que se $(V_k)_k$ é base enumerável de abertos de X então os cilindros $[1; V_{k_1}, \dots, V_{k_n}]$ geram tanto a σ -álgebra de Borel $X^{\mathbb{N}}$ quanto o produto $\mathcal{B}^{\mathbb{N}}$ da σ -álgebra de Borel \mathcal{B} de X . Logo, estas σ -álgebras coincidem.

4.2.6. Considere um ponto de densidade a de A e use o fato de que a órbita de a é densa (Exercício 1.3.10) para concluir que o complementar não tem pontos de densidade.

4.2.7. Para a primeira parte, verifique que a seqüência do lado direito de (4.2.19) é equicontínua e use o teorema de Ascoli-Arzelá. Na segunda parte, use o fato de que $\tilde{\varphi}$ é constante em cada órbita de R_θ e as órbita são densas em S^1 . Represente por $c(\varphi)$ o valor de $\tilde{\varphi}(x)$ para qualquer x . Dada qualquer probabilidade invariante μ e qualquer função contínua, o teorema de Birkhoff dá que $\int \varphi d\mu = c(\varphi)$. Pela Proposição A.3.3, isto mostra que μ está unicamente determinada.

4.2.8. A relação $Ax = \lambda x$ corresponde a um sistema homogêneo de d equações lineares com coeficientes racionais e, portanto, admite alguma solução racional $x \neq 0$. Por homogeneidade, podemos multiplicar as entradas de x por uma mesma constante, de modo torná-las inteiras.

4.2.9. Seja $J_k = (0, 1/k)$, para cada $k \geq 1$. Verifique que x tem expansão de tipo limitado se, e somente se, existe $k \geq 1$ tal que $G^n(x) \notin J_k$ para todo n . Observe que $\mu(J_k) > 0$ para todo k . Deduza que para todo k e μ -quase todo x existe $n \geq 1$ tal que $G^n(x) \in J_k$. Conclua que \mathcal{L} tem medida de Lebesgue nula.

4.2.10. Para cada $L \in \mathbb{N}$, considere $\varphi_L(x) = \min\{\phi(x), L\}$. Então $\varphi_L \in L^1(\mu)$ e, por ergodicidade, $\bar{\varphi}_L = \int \varphi_L d\mu$ em μ -quase todo ponto. Para concluir, observe que $\bar{\phi} \geq \bar{\varphi}_L$ para todo L e que $\int \varphi_L d\mu \rightarrow +\infty$.

4.2.11. A primeira parte segue do fato de que (G, μ) é ergódica (Proposição 4.2.10). Para a segunda parte, observe que a função $\psi = \sum_{k=1}^{\infty} k \mathcal{X}_{I_k}$ tem $\int \psi d\mu = +\infty$ e use o Exercício 4.2.10.

4.3.1. Seja $(U_k)_k$ uma base enumerável de abertos. Por ergodicidade, se U_k tem medida positiva então ele intersecta $(f^n(x))_n$ para todo x num conjunto M_k com medida total. Intersecte os conjuntos M_k obtidos deste modo.

4.3.2. Observe que a função $\eta \rightarrow \int \varphi d\eta$ é contínua e que $\mathcal{M}_1(f)$ é compacto (Teorema 2.1.5), relativamente à topologia fraca*

4.3.3. Observe que se $A \subset E$ é invariante por g e tem medida ν positiva então o conjunto $B = \cup_k \cup_{n < k} f^{-n}(A) \cap E_k = \cup_n \cup_{k > n} f^{-n}(A) \cap E_k$ é invariante por f e tem medida ν_ρ positiva. Supondo que (f, μ) é ergódica, segue que B tem medida ν_ρ total, o qual implica que $\nu(A) = 1$. Portanto, (g, ν) é ergódica. Para provar a recíproca, observe que se B é invariante por f e tem medida ν_ρ positiva então $A = B \cap E$ é invariante por g e tem medida ν positiva. Supondo que (g, ν) é ergódica, segue que $\nu(A) = 1$. Então B tem medida ν_ρ total, ou seja, $\mu(B) = 1$. Isto mostra que (f, μ) é ergódica.

4.3.4. Observe que se $A \subset M$ é invariante por f e tem medida μ positiva então $\pi^{-1}(A)$ é invariante por \hat{f} e tem medida $\hat{\mu}$ positiva. Isto mostra que (f, μ) é ergódico se $(\hat{f}, \hat{\mu})$ é ergódico. Para provar a recíproca, considere qualquer função uniformemente contínua limitada $\varphi : \hat{M} \rightarrow \mathbb{R}$ e seja $\bar{\varphi}$ a sua média temporal. Note que consideramos em \hat{M} a distância \hat{d} definida por (2.4.7). Verifique que se \hat{x}, \hat{y} são tais que $\pi(\hat{x}) = \pi(\hat{y})$ então $d(\hat{f}^n(\hat{x}), \hat{f}^n(\hat{y})) \rightarrow 0$ quando $n \rightarrow \infty$. Usando a continuidade uniforme, conclua que $\bar{\varphi}(\hat{x}) = \bar{\varphi}(\hat{y})$. Isto mostra que $\bar{\varphi} = \psi \circ \pi$ para alguma função $\psi : M \rightarrow \mathbb{R}$. Esta função ψ é mensurável a menos de medida nula. Isso pode ser mostrado da seguinte forma. Pelo Exercício 2.4.4, dado qualquer $\varepsilon > 0$ podemos aplicar o teorema de Lusin para encontrar um compacto $\hat{K} \subset \hat{M}$ tal que a restrição de $\bar{\varphi}$ a \hat{K} é contínua e $\hat{\mu}(\hat{K}) > 1 - \varepsilon$. Então, $K = \pi(\hat{K})$ é um compacto com $\mu(K) \geq \hat{\mu}(\hat{K}) > 1 - \varepsilon$. Mostre que a restrição de ψ a K é contínua. Fazendo $\varepsilon \rightarrow 0$, segue que ψ é mensurável a menos de medida nula, conforme afirmamos. Note também que ψ é f -invariante, porque $\bar{\varphi}$ é \hat{f} -invariante. Então, pela hipótese de que (f, μ) é ergódico, ψ é constante em μ -quase todo ponto. Isto significa que $\bar{\varphi}$ é constante em $\hat{\mu}$ -quase todo ponto. Pelo Exercício 4.1.4, segue que $(\hat{f}, \hat{\mu})$ é ergódico.

4.3.5. Observe que se $A \subset M$ é invariante por f e tem medida μ positiva então $\pi(A \times \mathbb{R})$ é invariante pelo fluxo e tem medida ν positiva. Para a recíproca, observe que se B é invariante pelo fluxo e tem medida ν positiva então $\pi^{-1}(B)$ tem a forma $A \times \mathbb{R}$, onde A é invariante por f e tem medida μ positiva.

4.3.6. Fixe $j \neq k$ em \mathcal{I} . Usando o Lema 4.3.1, mostre que existe conjunto mensurável $A_{j,k}$ tal que $\mu_j(A_{j,k}) > 0$ mas $\mu_k(A_{j,k}) = 0$. Seja $B_{j,k} = \cup_{j=0}^{\infty} f^{-j}(A_{j,k})$. Verifique que $\mu_j(B_{j,k}) > 0$ e $\mu_k(B_{j,k}) = 0$ e $f^{-1}(B_{j,k}) \subset B_{j,k}$. Defina $C_{j,k} = \cap_{n=0}^{\infty} f^{-n}(B_{j,k})$ e verifique que $f^{-1}(C_{j,k}) = C_{j,k}$. Mostre que $\mu_j(C_{j,k})$ é positiva e, portanto, $\mu_j(C_{j,k}) = 1$. Por outro lado, mostre que $\mu_k(C_{j,k}) = 0$. Agora defina $D_j = \cap_{k \neq j} C_{j,k}$ e $P_j = D_j \setminus \cup_{k \neq j} D_k$. Verifique que $f^{-1}(D_j) = D_j$ e $\mu_j(D_j) = 1$ e $\mu_k(D_j) = 0$ para todo $k \neq j$. Logo, $f^{-1}(P_j) = P_j$ para todo j e $\mu_j(P_j) = 1$ e $\mu_k(P_j) = 0$ para todo $k \neq j$. Além disso, os P_j são disjuntos dois-a-dois.

4.3.7. Seja $M = \{0, 1\}^{\mathbb{N}}$ e, para cada n , seja μ_n a medida invariante atômica suportada na órbita periódica $\alpha^n = (\alpha_k^n)_k$, de período $2n$, definida por $\alpha_k^n = 0$, se $0 \leq k < n$ e $\alpha_k^n = 1$ se $n \leq k < 2n$. Mostre que $(\mu_n)_n$ converge para $(\delta_0 + \delta_1)/2$, onde 0 e 1 são os pontos fixos do deslocamento.

4.3.8. Considere uma função contínua limitada qualquer $\varphi : M \rightarrow \mathbb{R}$. Por ergodicidade, $(1/n) \sum_{j=0}^{n-1} \varphi(f^j(x))$ converge para $\int \varphi d\mu$ em μ -quase todo ponto e, consequentemente, em ν -quase todo ponto Aplique o teorema da convergência dominada.

4.3.9. (1) Tome $k \geq 1$ tal que todo cilindro de comprimento k tem diâmetro menor que δ . Tome $y = (y_j)$ definido por $y_{j+n_i} = x_j^i$ para cada $0 \leq j < m_i + k$. (2) Tome $\delta > 0$ tal que $d(z, w) < \delta$ implica $|\varphi(z) - \varphi(w)| < \varepsilon$ e considere $k \geq 1$ dado pelo item 1. Escolha m_i , $i = 1, \dots, s$ tais que $m_i/n_s \approx \alpha_i$ para todo i . Então, tome y como no item 1. (3) Pelo teorema

ergódico, $\int \varphi d\mu = \int \tilde{\varphi} d\mu$. Tome $x^1, \dots, x^s \in \Sigma$ e $\alpha^1, \dots, \alpha^s$ tais que $\int \tilde{\varphi} d\mu \approx \sum_i \alpha_i \tilde{\varphi}(x^i)$. Note que $\tilde{\varphi}(y) = \int \varphi d\nu_y$, onde ν_y é a medida (atômica) invariante suportada na órbita de y . Lembre do Exercício 4.1.1.

4.4.1. (a) Discutimos apenas o caso $d = 2$. Se o vetor $\omega = (\omega_1, \omega_2)$ é (c, τ) -diofantino então o número $\theta = \omega_1/\omega_2$ é tal que $|\theta - (p/q)| \geq c/q^{\tau+1}$, para quaisquer $p, q \in \mathbb{Z}$ com $q \neq 0$. Usando o princípio da casa dos pombos, prove que dado qualquer irracional ξ existem inteiros k, l com $l \neq 0$ tais que $|\xi - k/l| < 1/l^2$ (este resultado é chamado teorema de Dirichlet). Conclua que $\tau + 1 \geq 2$. (b) Podemos supor que o vetor está no cubo unitário. Seja $V_{c,\tau}$ o conjunto dos vetores (c, τ) -diofantinos. Verifique que o complemento de $V_{c,\tau}$ é uma união de ‘placas’ com espessura $2c/\|k\|^{\tau+1}$ em torno do hiperplano ortogonal a cada vetor $k \in \mathbb{Z}^d$. Mostre que quando $\tau + 1 > d$ a medida do complementar de $V_{c,\tau}$ no cubo unitário vai para zero quando $c \rightarrow 0$.

4.4.2. Seja $B = [0, a] \times [0, b]$ a mesa do bilhar. Considere a aplicação $\pi : \mathbb{T}^2 \rightarrow B$ dada por $B([x, y]) = B([1-x, y]) = B([x, 1-y]) = B([1-x, 1-y]) = (2ax, 2by)$ para $(x, y) \in [0, 1/2]^2$.

Verifique que toda órbita do bilhar é imagem por π de uma reta no toro \mathbb{T}^2 . Logo a órbita é periódica ou densa dependendo da inclinação da reta ser racional ou irracional.

4.4.3. Considere em cada lado do triângulo o pé da respectiva altura (ou seja, o ponto onde esse lado intersecta a reta ortogonal a ele que passa pelo vértice oposto). Mostre que o triângulo definido por esses três pontos é uma órbita periódica do bilhar.

4.4.4. Mostre que o vetor $p_i - p_j$ é ortogonal à hipersuperfície $\{p \in \partial\Omega : \|p_i - p_j\| = 2\rho\}$ no ponto $p = (p_1, \dots, p_d)$. Deduza que as condições no Exemplo 4.4.4 implicam que (i) as componentes de $v_i - v_j$ e $v'_i - v'_j$ na direção ortogonal à hipersuperfície são simétricas e (ii) as componentes desses vetores na direção tangente ao bordo são iguais. Verifique que isto significa que a reflexão no bordo do bilhar é elástica.

4.4.5. Usando a Equação (4.4.8) e a condição de torção temos que para cada $\theta \in \mathbb{R}$ existe exatamente um número $\rho_\theta \in (a, b)$ tal que $\Theta(\theta, \rho_\theta) = \theta$. A função $\theta \mapsto \rho_\theta$ é contínua e periódica, com período 1. Considere o seu gráfico $\Gamma = \{(\theta, \rho_\theta) : \theta \in S^1\}$. Qualquer ponto em $\Gamma \cap f(\Gamma)$ é ponto fixo para f , uma vez que se $(\theta, \rho_\theta) = f(\gamma, \rho_\gamma) = (\Theta(\gamma, \rho_\gamma), R(\gamma, \rho_\gamma))$, como $\Theta(\gamma, \rho_\gamma) = \gamma$, temos $\theta = \gamma$ e, portanto, $\rho_\theta = \rho_\gamma$.

Como f preserva área, nenhuma das componentes conexas de $A \setminus \Gamma$ pode ser enviada dentro de si mesma. Isso implica que $f(\Gamma)$ intersecta Γ em, pelo menos, dois pontos.

4.4.6. Defina $G(\theta, \rho) = F^q(\theta, \rho) - (p, 0)$. Como F é levantamento de f , temos que G é um levantamento de f^q , já que se $\pi : \mathbb{R} \times [a, b] \rightarrow S^1 \times [a, b]$ é a projeção $\pi(\theta, \rho) = (\theta \bmod \mathbb{Z}, \rho)$ então $\pi G = \pi F^q = f^q \pi$. Além disso, G satisfaz (4.4.8). Logo, segue do Teorema 4.4.7 que f^q tem pelo menos dois pontos fixos no interior de A .

4.4.7. Inspirando-se no Exemplo 4.4.10, mostre que a transformação do bilhar em Ω é (ou, mais precisamente, se estende a) um *transformação de Dehn* no anel $A = S^1 \times [-\pi/2, \pi/2]$, ou seja, um homeomorfismo $f : A \rightarrow A$ que coincide com a identidade em ambas as componentes conexas do bordo mas é homotopicamente não trivial: na verdade, f admite um levantamento $F : \mathbb{R} \times [-\pi/2, \pi/2] \rightarrow \mathbb{R} \times [-\pi/2, \pi/2]$ tal que $F(s, -\pi/2) = (s - 2\pi, -\pi/2)$ e $F(s, \pi/2) = (s, \pi/2)$ para todo s . Considere números racionais $p_n/q_n \in (-2\pi, 0)$ com $q_n \rightarrow \infty$. Use o Exercício 4.4.6 para mostrar que g tem pontos periódicos de período q_n . Observe que, se tomarmos os q_n primos entre si, estes pontos periódicos são todos distintos.

5.1.1. Se $\mathcal{P} = \bigvee_n \{E_n, M \setminus E_n\}$, tome $\mathcal{P}_n = \bigvee_{j=1}^n \{E_j, M \setminus E_j\}$ e observe que $\mathcal{P} = \bigvee_n \mathcal{P}_n$. Reciprocamente, se $\mathcal{P} = \bigvee_n \mathcal{P}_n$ com \mathcal{P}_n finita ou enumerável, seja $\{E_k : k \in \mathbb{N}\}$ uma enumeração de $\bigcup_n \mathcal{P}_n$; verifique que $\mathcal{P} = \bigvee_n \{E_n, M \setminus E_n\}$.

5.1.2. Seja α o ínfimo das medidas dos conjuntos f^k -invariantes com medida positiva. Mostre que $\alpha = 1/l$ para algum inteiro $l \geq 1$ que divide k . Além disso, existe conjunto f^l -invariante A tal que $\mu(A) = 1/l$. Seja ν a restrição normalizada de μ a tal conjunto A . Verifique que a decomposição ergódica de μ para f^k é $(1/l) \sum_{j=0}^{l-1} f_*^j \nu$.

5.1.3. (a) \Rightarrow (b) Note que a aplicação $\mathcal{M}_1(M) \rightarrow \mathbb{R}$, $\nu \mapsto \int \varphi d\nu$ é contínua, logo mensurável. (b) \Rightarrow (a): Dados $\mu \in \mathcal{M}_1(M)$, $\varepsilon > 0$ e funções contínuas limitadas ϕ_1, \dots, ϕ_N , escreva

$a_j = \int \varphi_j d\mu$. Por hipótese, cada $\{x : \int \phi_j d\nu_x \in (a_j - \varepsilon, a_j + \varepsilon)\}$ é um conjunto mensurável. Intersectando em j , obtemos que $\{x : \nu_x \in V(\mu, \{\phi_1, \dots, \phi_N\}, \varepsilon)\}$ é mensurável. (b) \Rightarrow (c) Use o Exercício A.3.5 para aproximar ψ por funções contínuas, e use o teorema da convergência dominada. (c) \Rightarrow (d) é trivial. (d) \Rightarrow (b) Aproxime φ , uniformemente, por funções simples.

5.1.4. Seja $(\psi_k)_k$ uma sequência de funções simples, convergindo uniformemente para ψ . Então $P \mapsto \int \psi_k d\mu_P$ converge uniformemente para $P \mapsto \int \psi d\mu_P$. A conclusão do exercício segue facilmente.

5.1.5. Lembre (Exercício 4.3.4) que um sistema é ergódico se, e somente se, a sua extensão natural é ergódica. Conclua que, em geral, a decomposição ergódica da extensão natural é dada pelo levantamento da decomposição ergódica do sistema.

5.1.6. O Teorema 3.2.6 garante que M_0 tem medida total. Para mostrar que \mathcal{P} é partição mensurável, considere a sequência de partições $\mathcal{P}_n = \bigvee_{i,j=1}^n \{E_{i,j}, E_{i,j}^c\}$, com $E_{i,j} = \{\phi_i < q_j\}$, onde $\{\phi_i : i \geq 1\}$ é um subconjunto enumerável denso da bola unitária de $C^0(M, \mathbb{R})$ (Teorema A.3.13) e $\{q_j : j \geq 1\}$ é uma enumeração dos números racionais. Considere uma desintegração $\{\mu_P : P \in \mathcal{P}\}$ da medida μ relativamente a \mathcal{P} . Para ver que as medidas μ_P são ergódicas use o Exercício 4.1.4.

5.1.7. O enunciado não depende da escolha da decomposição ergódica, uma vez que ela é essencialmente única. Considere a construção no Exercício 5.1.6. O conjunto M_0 é saturado pela partição \mathcal{W}^s , ou seja, se $x \in M_0$ então $\mathcal{W}^s(x) \subset M_0$. Além disso, a aplicação $y \mapsto \mu_y$ é constante em cada $\mathcal{W}^s(x)$. Como a partição \mathcal{P} é caracterizada por $\mathcal{P}(x) = \mathcal{P}(y) \Leftrightarrow \mu_x = \mu_y$, segue que $\mathcal{P} \prec \mathcal{W}^s$ restrito a M_0 .

5.2.1. Considere as projeções canônicas $\pi_{\mathcal{P}} : M \rightarrow \mathcal{P}$ e $\pi_{\mathcal{Q}} : M \rightarrow \mathcal{Q}$, as medidas quociente $\hat{\mu}_{\mathcal{P}} = (\pi_{\mathcal{P}})_* \mu$ e $\hat{\mu}_{\mathcal{Q}} = (\pi_{\mathcal{Q}})_* \mu$ e as desintegrações $\mu = \int \mu_P d\hat{\mu}_{\mathcal{P}}(P)$ e $\mu = \int \mu_Q d\hat{\mu}_{\mathcal{Q}}(Q)$. Além disso, para cada $P \in \mathcal{P}$, considere $\hat{\mu}_{P,\mathcal{Q}} = (\pi_{\mathcal{Q}})_* \mu_P$ e a desintegração $\mu_P = \int \mu_{P,Q} d\hat{\mu}_{P,\mathcal{Q}}(Q)$. Observe que $\int \hat{\mu}_{P,\mathcal{Q}} d\hat{\mu}_{\mathcal{P}}(P) = \hat{\mu}_{\mathcal{Q}}$: dado qualquer $B \subset \mathcal{Q}$,

$$\int \hat{\mu}_{P,\mathcal{Q}}(B) d\hat{\mu}_{\mathcal{P}}(P) = \int \mu_P(\pi_{\mathcal{Q}}^{-1}(B)) d\hat{\mu}_{\mathcal{P}}(P) = \mu(\pi_{\mathcal{Q}}^{-1}(B)) = \hat{\mu}_{\mathcal{Q}}(B).$$

Para verificar que $\mu_{\pi(Q),Q}$ é desintegração de μ relativamente a \mathcal{Q} : (a) $\mu_{P,Q}(Q) = 1$ para $\hat{\mu}_{P,\mathcal{Q}}$ -quase todo Q e $\hat{\mu}_{\mathcal{P}}$ -quase todo P . Além disso, $\mu_{P,Q} = \mu_{\pi(Q),Q}$ para $\hat{\mu}_{P,\mathcal{Q}}$ -quase todo Q e $\hat{\mu}_{\mathcal{P}}$ -quase todo P , porque $\mu_P(P) = 1$ para $\hat{\mu}_{\mathcal{P}}$ -quase todo P . Pela observação anterior, segue que $\mu_{\pi(Q),Q}(Q) = 1$ para $\hat{\mu}_{\mathcal{Q}}$ -quase todo Q . (b) $P \mapsto \mu_P(E)$ é mensurável, a menos de medida nula, para todo conjunto boreliano $E \subset M$. Por construção (Seção 5.2.3), existe uma álgebra enumerável geradora \mathcal{A} tal que $\mu_{P,Q}(E) = \lim_n \mu_P(E \cap Q_n) / \mu_P(Q_n)$ para todo $E \in \mathcal{A}$ (onde Q_n é o elemento de \mathcal{Q}_n que contém Q). Deduza que $P \mapsto \mu_{\pi(Q),Q}(E)$ é mensurável, a menos de medida nula, para todo $E \in \mathcal{A}$. Estenda esta conclusão para todo boreliano E , usando o argumento de classes monótonas na Seção 5.2.3. (c) Temos $\mu = \int \mu_P d\hat{\mu}_{\mathcal{P}}(P) = \int \int \mu_{P,Q} d\hat{\mu}_{P,\mathcal{Q}}(Q) d\hat{\mu}_{\mathcal{P}}(P) = \int \int \mu_{\pi(Q),Q} d\hat{\mu}_{P,\mathcal{Q}}(Q) d\hat{\mu}_{\mathcal{P}}(P) = \int \mu_{\pi(Q),Q} d\hat{\mu}_{\mathcal{Q}}(Q)$.

5.2.2. Argumente que a partição \mathcal{Q} do espaço $\mathcal{M}_1(M)$ em pontos é mensurável. Dada uma desintegração $\{\mu_P : P \in \mathcal{P}\}$, considere a aplicação mensurável $M \mapsto \mathcal{M}_1(M)$, $x \mapsto \mu_{P(x)}$. A pré-imagem de \mathcal{Q} por esta aplicação é uma partição mensurável. Verifique que essa pré-imagem coincide com \mathcal{P} num subconjunto com medida total.

5.2.3. Observe que a família dos conjuntos $B \in \mathcal{B}$ tais que $\int_B e(\psi) d\mu = \int_B \psi d\mu$ é uma σ -álgebra completa e contém $\cup_n \mathcal{P}_n$, logo contém \mathcal{B} . Deduza que $e(\psi) = \psi$ em μ -quase todo ponto, tal como afirmado.

5.2.4. As hipóteses sobre M implicam que \hat{M} é espaço métrico e admite base enumerável de abertos $\hat{\mu}$ formada por cilindros (lembre da Seção 2.4.2). Usando (2.4.7) e a hipótese de que f é contínua, verifique que \hat{M} também é completo. Inicialmente, defina $\hat{\mu}$ no conjunto dos cilindros, por meio de (2.4.8). Mostre que $\hat{\mu}$ está bem definida: se $[A_k, \dots, A_0] = [B_l, \dots, B_0]$ com $k \leq l \leq 0$, então $A_k \cap \dots \cap f^{-k}(A_0) = f^{k-l}(B_l \cap \dots \cap f^{-l}(B_0))$ e, portanto, $\hat{\mu}([A_k, \dots, A_0]) = \hat{\mu}([B_l, \dots, B_0])$. Além disso, se $[A_k, \dots, A_0]$ coincide com uma união disjunta de cilindros $[A_k^i, \dots, A_0^i]$, $i = 1, \dots, N$ então $A_k \cap \dots \cap f^{-k}(A_0)$ coincide com a união disjunta dos conjuntos $A_k^i \cap \dots \cap f^{-k}(A_0^i)$ e, portanto, $\hat{\mu}([A_k, \dots, A_0]) = \sum_i \hat{\mu}([A_k^i, \dots, A_0^i])$. Conclua que podemos estender μ a uma função aditiva na álgebra gerada pelos cilindros, a qual coincide com a família das uniões finitas disjuntas de cilindros. Seja $\hat{\mathcal{A}}$ a álgebra gerada pela base

enumerável de abertos \hat{U} . Claro que $\hat{\mathcal{A}}$ está contida na álgebra gerada por todos os cilindros. Pela Proposição 2.4.4, a restrição de $\hat{\mu}$ a $\hat{\mathcal{A}}$ se estende a uma medida de probabilidade em \hat{M} . Conclua que esta probabilidade é um levantamento de μ .

6.1.1. Considere f sendo uma rotação racional em \mathbb{S}^1 .

6.1.2. Observe que se f é transitiva e ψ é uma função contínua satisfazendo $\psi \circ f = \psi$ então ψ é constante.

6.1.3. A função φ é invariante.

6.2.1. A transformação h é invertível e preserva a medida m . A relação de conjugação $h \circ f_0 = f \circ h$ é equivalente à equação cohomológica.

6.2.2. A transformação h é um homeomorfismo e a relação de conjugação $h \circ f_0 = f \circ h$ é equivalente à equação cohomológica.

6.2.3. Substitua $\phi(x) = \sum_n a_n e^{2\pi i n x}$ e $u(x) = \sum_n b_n e^{2\pi i n x}$ na equação cohomológica e resolva em ordem a b_n para obter (6.2.9). Observe que $b_0 = 0$ e que $b_{-n} = \bar{b}_n$ para todo $n \geq 1$. Portanto, $\sum_{n=1}^{\infty} |b_n|^2 < \infty \Leftrightarrow \sum_{n \in \mathbb{Z}} |b_n|^2 < \infty \Leftrightarrow u \in L^2(m)$.

6.2.4. Se α é diofantino então existe $C > 1$ tal que $|e^{2\pi i q \alpha} - 1| \geq C^{-1} |q|^{-\tau}$ para todo $q \geq 1$. Então $|b_n| \leq C \rho^n |n|^\tau$ para todo n suficientemente grande.

6.2.5. Represente por X o fecho da órbita do ponto x . Se X é minimal, para cada $y \in X$ existe $n(y) \geq 1$ tal que $d(f^{n(y)}(y), x) < \varepsilon$. Então, por continuidade, y admite uma vizinhança aberta $V(y)$ tal que $d(f^{n(y)}(z), x) < \varepsilon$ para todo $z \in V(y)$. Tome y_1, \dots, y_s tais que $X \subset \cup_i V(y_i)$ e tome $m = \max_i n(y_i)$. Dado qualquer $k \geq 1$, tome i tal que $f^k(x) \in V(y_i)$. Então $d(f^{k+n_i}(x), x) < \varepsilon$, ou seja, $k + n_i \in R_\varepsilon$. Isto prova que, dados quaisquer $m + 1$ inteiros consecutivos, pelo menos um deles está em R_ε . Logo R_ε é sindético. Agora suponha que X não é minimal. Então existe um conjunto invariante, fechado e não vazio F contido propriamente em X . Note que $x \notin F$ e, portanto, para todo ε suficientemente pequeno, existe um aberto U que contém F e não intersecta $B(x, \varepsilon)$. Por outro lado, como R_ε é sindético, existe $m \geq 1$ tal que, para qualquer $k \geq 1$ existe $n \in \{k, \dots, k + m\}$ tal que $f^n(x) \in B(x, \varepsilon)$. Tome k tal $f^k(x) \in U_1$, onde $U_1 = U \cap f^{-1}(U) \cap \dots \cap f^{-m}(U)$, e encontre uma contradição.

6.2.6. Pelo Exercício 6.2.5, o conjunto $R_\varepsilon = \{n \in \mathbb{N} : d(x, f^n(x)) < \varepsilon\}$ é sindético para todo $\varepsilon > 0$. Se y está próximo de x então $\{n \in \mathbb{N} : d(f^n(x), f^n(y)) < \varepsilon\}$ contém blocos de inteiros consecutivos com comprimento arbitrário, qualquer que seja $\varepsilon > 0$. Seja U_1 uma vizinhança qualquer de x . Segue das observações anteriores que existem infinitos valores de $n \in \mathbb{N}$ tais que $f^n(x), f^n(y)$ estão em U_1 . Fixe n_1 com essa propriedade. Em seguida, considere $U_2 = U_1 \cap f^{-n_1}(U_1)$. Pelo passo anterior, existe $n_2 > n_1$ tal que $f^{n_2}(x), f^{n_2}(y) \in U_2$. Continuando deste modo, construa uma sequência não crescente de abertos U_k e uma sequência crescente de números naturais n_k tais que $f^{n_k}(U_{k+1}) \subset U_k$ e $f^{n_k}(x), f^{n_k}(y) \in U_k$. Verifique que $f^{n_{i_1} + \dots + n_{i_k}}(x)$ e $f^{n_{i_1} + \dots + n_{i_k}}(y)$ estão em U_1 para quaisquer $i_1 < \dots < i_k, k \geq 1$.

6.2.7. Considere o deslocamento $\sigma : \Sigma \rightarrow \Sigma$ em $\Sigma = \{1, 2, \dots, q\}^{\mathbb{N}}$. A partição $\mathbb{N} = S_1 \cup \dots \cup S_q$ define um certo elemento $\alpha = (\alpha_n) \in \Sigma$, dado por $\alpha_n = i$ se e somente $n \in S_i$. Considere β no fecho da órbita de α tal que α e β estão próximos e o fecho da órbita de β é um conjunto minimal. Aplique o Exercício 6.2.6 com $x = \beta, y = \alpha$ e $U = [0; \alpha_0]$ para obter o resultado.

6.3.1. Escreva $\phi(g, h) = g \cdot h$ e sejam $\partial_1 \phi$ e $\partial_2 \phi$ as suas derivadas parciais. Verifique que $\partial_2 \phi(g_0, h_0)$ é um isomorfismo, para quaisquer $g_0, h_0 \in G$. Tome $h_0 = g_0^{-1}$ e use o teorema da função implícita.

6.3.2. Dado $g_0 \in G$ e dada qualquer sequência $(g_n)_n \rightarrow g_0$, seja h_0 qualquer ponto de acumulação da sequência $(g_n^{-1})_n$. Por continuidade, $g_0 \cdot h_0 = e$, ou seja, $h_0 = g_0^{-1}$. Conclua usando a hipótese de compacidade.

6.3.3. Verifique que $d(\alpha, \beta) \neq d(\alpha\gamma, \beta\gamma)$, onde

$$\alpha = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \beta = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \quad \gamma = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

6.3.4. Como G é localmente compacto, existe algum compacto $K \subset G$ com medida positiva. Supondo que G não é compacto, construa uma sequência $(g_n)_n$ em G tal que os conjuntos

$g_n K = \{g_n g : g \in K\}$ são disjuntos dois-a-dois. Conclua que a medida de Haar é infinita. Para a recíproca, tome um aberto $U \subset G$ com medida finita e considere a cobertura $\{g^{-1}U : g \in U\}$ de G . Por compacidade, existe subcobertura finita.

6.3.5. Dado qualquer $g \in \mathbb{R} \setminus \{0\}$,

$$\int_{\text{GL}(1, \mathbb{R})} \varphi d\mu = \int \frac{\varphi(x)}{|x|} dx = \int \frac{\varphi(gs)}{|gs|} |g| ds = \int \frac{\varphi(gs)}{|s|} ds = \int_{\text{GL}(1, \mathbb{R})} (\varphi \circ E_g) d\mu.$$

Para a segunda parte do enunciado, identifique $\text{GL}(1, \mathbb{C})$ com $\mathbb{C} \setminus \{0\}$ e considere a medida μ definida por

$$\int_{\text{GL}(1, \mathbb{C})} \varphi d\mu = \int_{\mathbb{C} \setminus \{0\}} \frac{\varphi(x+yi)}{|x+yi|^2} dx dy.$$

Use a fórmula de mudança de variável como no caso anterior, observando que se $g = a + bi$ então $E_g(x + yi) = (ax - by) + (ay + bx)i$ e o jacobiano de E_g é constante igual a $|a + bi|^2$.

6.3.6. Escreva $g = (a_{11}, a_{12}, a_2, a_{22})$. Então,

$$E_g(x_{11}, x_{12}, x_{21}, x_{22}) = (a_{11}x_{11} + a_{12}x_{21}, a_{11}x_{12} + a_{12}x_{22}, a_{21}x_{11} + a_{22}x_{21}, a_{21}x_{12} + a_{22}x_{22}).$$

Escreva o lado direito como $(y_{11}, y_{12}, y_{21}, y_{22})$. Use a fórmula de mudança de variáveis na integral, observando que $\det(y_{11}, y_{12}, y_{21}, y_{22}) = (\det g) \det(x_{11}, x_{12}, x_{21}, x_{22})$ e

$$dy_{11} dy_{12} dy_{21} dy_{22} = (\det g)^2 dx_{11} dx_{12} dx_{21} dx_{22}.$$

No caso complexo, tome

$$\int_{\text{GL}(2, \mathbb{R})} \varphi d\mu = \int \frac{\varphi(z_{11}, z_{12}, z_{21}, z_{22})}{|\det(z_{11}, z_{12}, z_{21}, z_{22})|^4} dx_{11} dy_{11} dx_{12} dy_{12} dx_{21} dy_{21} dx_{22} dy_{22},$$

onde $z_{jk} = x_{jk} + y_{jk}i$. [Observação: Generalize estas construções para dimensão qualquer!]

6.3.7. (3) \Rightarrow (2) é trivial. Para (2) \Rightarrow (3) use a distância invariante dada pelo Lema 6.3.6. (3) \Rightarrow (1) segue do Teorema 6.3.8. (1) \Rightarrow (3) segue da Proposição 6.2.1.

6.3.8. Dados $\alpha = (\alpha_n)_n$ e $\beta = (\beta_n)_n$ e $N \geq 1$, as primeiras N coordenadas de $\alpha + \beta$ dependem apenas das primeiras N coordenadas de α e de β . Analogamente, as N primeiras coordenadas de $-\alpha$ só dependem das primeiras N coordenadas de α . Isto garante que $+e$ e $x \mapsto -x$ são contínuas. A mesma observação mostra que as translações preservam a distância (6.3.7).

6.3.9. Dado $x \in M$, existe um único $0 \leq r < 10^k$ tal que $f^r(x) \in [b_0, \dots, b_{k-1}]$. Além disso, $f^n(x) \in [b_0, \dots, b_{k-1}]$ se, e somente se, $n - r$ é múltiplo de 10^k . Use esta observação para concluir que $\tau([b_0, \dots, b_{k-1}], x) = 10^{-k}$ para todo $x \in M$. Conclua que se f admite uma probabilidade invariante ergódica μ então $\mu([b_0, \dots, b_{k-1}]) = 10^{-k}$ para todo b_0, \dots, b_{k-1} . Isto determina μ de maneira única. Para terminar mostre que μ está bem definida e é invariante.

6.3.10. Note que $f_{n+1}(x) = f_n(x)$ exceto se x está no topo de \mathcal{S}_n . Os topos formam uma sequência decrescente cuja interseção é vazia (porque os intervalos são abertos à direita). Logo, $f(x) = f_n(x)$ para todo x e todo n suficientemente grande. Segue que $f^{-1}(A) = \cup_k \cap_{n>k} f_n^{-1}(A)$ para todo A e, portanto, f preserva a medida de Lebesgue. Para mostrar a unicidade ergódica, relacione f com o odômetro.

6.3.11. Considere a sequência de palavras w_n definida indutivamente por $w_1 = \alpha$ e $s(w_{n+1}) = w_n$, para $n \geq 1$. Decomponha a palavra $s(\alpha) = w_2 = \alpha r_1$ e prove, por indução, que w_{n+1} pode ser decomposto como $w_{n+1} = w_n r_n$, para alguma palavra r_n com comprimento maior ou igual a n , tal que $s(r_n) = r_{n+1}$. Defina $w = \alpha r_1 r_2 \dots$ e note que $s(w) = s(\alpha) s(r_1) s(r_2) \dots = \alpha r_1 r_2 r_3 \dots = w$. Isto prova a existência. Para obter a unicidade, seja $\gamma \in \Sigma$ uma sequência começando por α tal que $S(\gamma) = \gamma$. Decomponha γ como $\gamma = \alpha \gamma_1 \gamma_2 \gamma_3 \dots$, de tal modo que γ_i e r_i tenham o mesmo comprimento. Note que $S(\alpha) = \alpha \gamma_1 = \alpha r_1$, pelo que $\gamma_1 = r_1$. Conclua a prova por indução.

6.4.1. Suponha que $(z_j)_j$ é equidistribuída. Se $\limsup_n (1/n) \sum_{j=1}^n \chi_I(z_j) > m(I)$ então existem $\varepsilon > 0$ e infinitos valores de n para os quais $(1/n) \sum_{j=1}^n \chi_I(z_j) \geq m(I) + \varepsilon$. Fixe um intervalo aberto U com $\bar{I} \subset U$ e $m(U/I) < \varepsilon/2$. Considere uma função contínua $\varphi : S^1 \rightarrow [0, 1]$ tal que $\chi_I \leq \varphi \leq \chi_U$, para chegar a uma contradição. Logo, o limite superior é maior ou

igual que $m(I)$. Analogamente, o limite inferior é menor ou igual a $m(I)$. Para a recíproca, dada uma função contínua $\varphi : S^1 \rightarrow \mathbb{R}$ e dado $\varepsilon > 0$, considere funções simples g e h tais que $g \leq \varphi \leq h$ e $\int(\varphi - g) dm < \varepsilon$ e $\int(h - \varphi) dm < \varepsilon$.

6.4.2. Dados quaisquer $0 \leq \alpha < \beta \leq 1$, temos que $\sqrt{n} \in (\alpha, \beta)$ no círculo se, e somente se, existe algum inteiro $k \geq 1$ tal que $k^2 + 2k\alpha + \alpha^2 < n < k^2 + 2k\beta + \beta^2$. Para cada k o número de valores de n que satisfazem esta desigualdade é igual à parte inteira de $2k(\beta - \alpha) + (\beta^2 - \alpha^2)$. Portanto,

$$\#\{1 \leq n < N^2 : \sqrt{n} \in (\alpha, \beta)\} \leq \sum_{k=1}^{N-1} 2k(\beta - \alpha) + (\beta^2 - \alpha^2)$$

e a diferença entre o termo da direita e o da esquerda é menor que N . Logo

$$\lim \frac{1}{N^2} \#\{1 \leq n < N^2 : \sqrt{n} \in (\alpha, \beta)\} = \beta - \alpha.$$

Um cálculo semelhante mostra que a sequência $(\log n \bmod \mathbb{Z})_n$ não é equidistribuída no círculo. [Observação: Mas ela admite uma densidade limite contínua (não-constante). Calcule essa densidade!]

6.4.3. Defina $\phi_n = a^n + (-1/a)^n$. Verifique que $(\phi_n)_n$ é a sequência de Fibonacci e, em particular, $\phi_n \in \mathbb{N}$ para todo $n \geq 1$. Agora observe que $(-1/a)^n$ converge para zero. Logo, $\{n \geq 1 : a^n \bmod \mathbb{Z} \in I\}$ é finito, para qualquer intervalo $I \subset S^1$ cujo fecho não contém zero.

7.1.1. É claro que a condição é necessária. Para ver que é suficiente: Dado A , considere o subespaço fechado \mathcal{V} de $L^2(\mu)$ gerado pelas funções 1 e $\mathcal{X}_{f^{-k}(A)}$, $k \in \mathbb{N}$. A hipótese garante que $\lim_n U_f^n(\mathcal{X}_A) \cdot \mathcal{X}_{f^{-k}(A)} = (\mathcal{X}_A \cdot 1)(\mathcal{X}_{f^{-k}(A)} \cdot 1)$ para todo k . Conclua que $\lim_n U_f^n(\mathcal{X}_A) \cdot \phi = (\mathcal{X}_A \cdot 1)(\phi \cdot 1)$ para todo $\phi \in \mathcal{V}$. Dado um conjunto mensurável B , escreva $\mathcal{X}_B = \phi + \phi^\perp$ com $\phi \in \mathcal{V}$ e $\phi^\perp \in \mathcal{V}^\perp$ para concluir que $\lim_n U_f^n(\mathcal{X}_A) \cdot \mathcal{X}_B = (\mathcal{X}_A \cdot 1)(\mathcal{X}_B \cdot 1)$

7.1.2. Supondo que E existe, decomponha $(1/n) \sum_{j=0}^{n-1} |a_j|$ em duas parcelas, uma com $j \in E$ e a outra com $j \notin E$. As hipóteses sobre E implicam que as duas parcelas convergem para zero. Reciprocamente, suponha que $(1/n) \sum_{j=0}^{n-1} |a_j|$ converge para zero. Defina $E_m = \{j \geq 0 : |a_j| \geq (1/m)\}$ para cada $m \geq 1$. A sequência $(E_m)_m$ é crescente e cada E_m tem densidade zero; em particular, existe $\ell_m \geq 1$ tal que $(1/n) \#(E_m \cap \{0, \dots, n-1\}) < (1/m)$ para todo $n \geq \ell_m$. Escolha $(\ell_m)_m$ crescente e defina $E = \cup_m (E_m \cap \{\ell_m, \dots, \ell_{m+1} - 1\})$. Para a segunda parte do exercício, aplique a primeira parte às duas sequências, $(a_n)_n$ e $(a_n^2)_n$.

7.1.3. Segue do critério que foi apresentado no Exercício 7.1.2 que a sequência $(1/n) \sum_{j=0}^{n-1} |a_j|$ converge para zero se, e somente se, $(1/n) \sum_{j=0}^{n-1} |a_{k_j}|$ converge para zero para qualquer $k \geq 1$. [Observação: Alternativamente, podemos usar o Teorema 8.2.1.]

7.1.4. Seja m a medida de Haar. Considere a base de Fourier $\varphi_k(x) = e^{2\pi i k \cdot x}$, $k \in \mathbb{Z}$ do espaço $L^2(m)$. Note que $\varphi_k \circ f_A = \varphi_{A(k)}$. A hipótese implica que a sequência dos iterados $A^n(k)$, $n \in \mathbb{Z}$ é injetiva, para todo $k \neq 0$. Em particular, para quaisquer $k, l \in \mathbb{Z}^d$ (considere os casos $k \neq 0$ e $k = 0$ separadamente) $\varphi_k \circ f_A^n \cdot \varphi_l = (\varphi_k \cdot 1)(\varphi_l \cdot 1)$ para todo n suficientemente grande. Por linearidade, vale propriedade análoga para qualquer par de combinações finitas de elementos da base de Fourier. Conclua o argumento usando a Proposição 7.1.12.

7.1.5. Seja η uma probabilidade absolutamente contínua e $\psi = d\eta/d\mu$. Suponha que (f, μ) é misturador. Então $\int \varphi d(f_*^n \eta) = \int (\varphi \circ f^n) \psi d\mu \rightarrow \int \varphi d\mu$ para toda função contínua φ (note que $\int \psi d\mu = 1$ porque η é probabilidade). Isto mostra que $f_*^n \eta \rightarrow \mu$. Para a recíproca, considere $\varphi \in C^0(M)$ e $\psi \in L^1(\mu)$. Escreva $\psi = \psi^+ - \psi^-$ com $\psi^\pm > 0$: então $\eta^\pm = \psi^\pm \mu$ são medidas absolutamente contínuas relativamente a μ . Aplicando a hipótese às suas normalizações, $\int (\varphi \circ f^n) \psi^\pm d\mu = \int (\varphi \circ f^n) d\eta^\pm \rightarrow \nu^\pm(M) \int \varphi d\mu = \int \varphi d\mu \int \psi^\pm d\mu$. Agora use a Proposição 7.1.12 e o fato de que $C^0(M)$ é denso em $L^\infty(\mu)$ (Exercício A.5.7).

7.1.6. (Pollicott, Yuri [PY98]) Basta tratar o caso em que $\int \varphi_j d\mu = 0$ para todo j . Use indução no número k de funções. O caso $k = 1$ está contido no Teorema 3.1.5. Use as

desigualdades

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^n a_n &\leq \frac{1}{N} \sum_{n=1}^{N-m+1} \left(\frac{1}{m} \sum_{j=0}^{m-1} a_{n+j} \right) + \frac{m}{N} \left(\max_{1 \leq i \leq m} |a_i| + \max_{N-m \leq i \leq N} |a_i| \right) \\ \left(\frac{1}{N} \sum_{n=1}^N b_n \right)^2 &\leq (1/N) \sum_{n=1}^N |b_n|^2 \end{aligned}$$

para concluir que $\int |(1/N) \sum_{j=0}^{N-1} (\varphi_1 \circ f^j) \cdots (\varphi_k \circ f^{kn})|^2 d\mu$ é majorado por

$$\frac{1}{N} \sum_{n=1}^N \left(\int \left| \frac{1}{m} \sum_{j=0}^{m-1} (\varphi_1 \circ f^{n+j}) \cdots (\varphi_k \circ f^{k(n+j)}) \right|^2 d\mu + \left(\frac{2m}{N} + \frac{m^2}{N^2} \right) \left(\max_{1 \leq i \leq k} \text{supess } |\varphi_i| \right)^2 \right).$$

A integral é igual

$$\sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \int \prod_{l=1}^k (\varphi_l(\varphi_l \circ f^{l(j-i)})) \circ f^{l(n+i)} d\mu.$$

Pela hipótese de indução,

$$\frac{1}{N} \sum_{n=1}^N \prod_{l=2}^k (\varphi_l(\varphi_l \circ f^{l(j-i)})) \circ f^{l(n+i)} \rightarrow \prod_{l=2}^k \int \varphi_l(\varphi_l \circ f^{l(j-i)}) d\mu$$

em $L^2(\mu)$, quando $N \rightarrow \infty$. Portanto,

$$\frac{1}{N} \sum_{n=1}^N \int \prod_{l=1}^k (\varphi_l(\varphi_l \circ f^{l(j-i)})) \circ f^{l(n+i)} d\mu \rightarrow \prod_{l=1}^k \int \varphi_l(\varphi_l \circ f^{l(j-i)}) d\mu$$

em $L^2(\mu)$, quando $N \rightarrow \infty$. Juntando estas estimativas,

$$\limsup_N \int \left| \frac{1}{N} \sum_{n=1}^N (\varphi_1 \circ f^n) \cdots (\varphi_k \circ f^{kn}) \right|^2 d\mu \leq \frac{1}{m^2} \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \prod_{l=1}^k \int \varphi_l(\varphi_l \circ f^{l(j-i)}) d\mu.$$

Como (f, μ) é fracamente misturadora, $\int \varphi_l(\varphi_l \circ f^{lr}) d\mu$ converge para 0 quando $r \rightarrow \infty$, restrito a um conjunto de valores com densidade 1 no infinito (lembre do Exercício 7.1.2). Portanto, a expressão no lado direito está próxima de zero quando m é grande.

7.2.1. Estenda μ à álgebra \mathcal{A} das uniões finitas disjuntas de cilindros, por aditividade. Isso é possível porque $\sum_j P_{i,j} = 1$ para todo i e, portanto, a definição da medida de cada cilindro é $[m; a_m, \dots, a_n]$ é compatível com as definições das medidas dos seus subcilindros, no sentido de que $\mu([m; a_m, \dots, a_n]) = \sum_b \mu([m; a_m, \dots, a_n, b])$. Em seguida, use o fato de que a álgebra \mathcal{A} é compacta (porque todos os elementos são compactos), para concluir que μ satisfaz a hipótese do Teorema A.1.14.

7.2.2. [Pet83] Suporemos X finito. Sejam $P^k = (p_{i,j}^k)_{i,j}$ $k \geq 1$ os iterados da matriz de transição. Mostre que se o deslocamento é fracamente misturador então $(1/n) \sum_{k=0}^{n-1} |P_{i,j}^k - p_j| \rightarrow 0$. Conclua (usando o Exercício 7.1.2) que para cada i, j existe um conjunto $K = K(i, j)$ de densidade 0 tal que $\lim_{k \notin K} p_{i,j}^k = p_j$. Deduza que a matriz P é aperiódica.

7.2.3. Sim. Defina $(q_i)_i$ e $(Q_{i,j})_{i,j}$ por $q_i = \mu([0; i])$ e $q_i Q_{i,j} = \mu([0; j, i])$.

7.2.4. A definição de memória zero significa que

$$\frac{\mu([m-l; a_{m-l}, \dots, a_{m-1}, a_m])}{\mu([m-l; a_{m-l}, \dots, a_{m-1}])} = \mu([m; a_m]) = \mu([0; a_m])$$

para todo $l \geq 0$, todo m e todo $(a_n)_n \in \Sigma$. Escreva $p_i = \mu([0; i])$ para concluir que μ é medida de Bernoulli. A definição de memória um significa que

$$\frac{\mu([m-l; a_{m-l}, \dots, a_{m-1}, a_m])}{\mu([m-l; a_{m-l}, \dots, a_{m-1}])} = \frac{\mu([m-1; a_{m-1}, a_m])}{\mu([m-1; a_{m-1}])}$$

para todo $l \geq 0$, todo m e todo $(a_n)_n \in \Sigma$. Escreva $p_i = \mu([0; i])$ e $P_{i,j} = \mu([0; i, j]) / \mu([0; i])$ para concluir que μ é medida de Markov. Para $k \geq 2$, considere a aplicação $h : \Sigma \rightarrow \tilde{\Sigma}$ dada por $h((x_n)_n) = (\tilde{x}_n)_n$ onde $\tilde{x}_n = (x_n, \dots, x_{n+k-1})$. Verifique que $h_*\mu$ é medida de Markov para o deslocamento em $\tilde{\Sigma}$ e que a restrição de h ao suporte de μ é um homeomorfismo sobre o suporte de $\tilde{\mu}$. Para exemplificar, esboçamos o caso $k = 2$ e $\sigma = X^{\mathbb{N}}$. Dados $\tilde{a} = (a_1, a_2)$ e $\tilde{b} = (b_1, b_2)$ em \tilde{X} , defina

$$\tilde{P}_{\tilde{a}, \tilde{b}} = \frac{\mu([1; a_1, a_2, b_2])}{\mu([1; a_1, a_2])} \text{ se } a_2 = b_1 \quad (\text{diremos que a dupla } (\tilde{a}, \tilde{b}) \text{ é compatível})$$

e $\tilde{P}_{\tilde{a}, \tilde{b}} = 0$ caso contrário. Verifique que $\tilde{P}_{\tilde{a}, \tilde{b}} \geq 0$ e que $\sum_{\tilde{b} \in \tilde{X}} \tilde{P}_{\tilde{a}, \tilde{b}} = 1$ para todo $\tilde{a} \in \tilde{X}$. Logo, $\tilde{P}_{\tilde{a}, \tilde{b}}$ é matriz estocástica. Defina $p_{\tilde{a}} = \mu([0; a_1, a_2])$. Mostre que $\sum_{\tilde{a} \in \tilde{X}} p_{\tilde{a}} \tilde{P}_{\tilde{a}, \tilde{b}} = p_{\tilde{b}}$ para todo $\tilde{b} \in \tilde{X}$. Considere a medida de Markov ν definida por \tilde{P} e \tilde{p} . Verifique que

$$p_{\tilde{a}_m} \tilde{P}_{\tilde{a}_m, \tilde{a}_{m+1}} \cdots \tilde{P}_{\tilde{a}_{n-1}, \tilde{a}_n} = p_{a_m} P_{a_m, a_{m+1}} \cdots P_{a_n, a_{n+1}}$$

se as duplas $(\tilde{a}_j, \tilde{a}_{j+1})$ são todas compatíveis e é igual a zero caso contrário. Verifique que $h^{-1}([m; \tilde{a}_m, \dots, \tilde{a}_n]) = [m; a_m, \dots, a_{n+1}]$ se se as duplas $(\tilde{a}_j, \tilde{a}_{j+1})$ são todas compatíveis e é o conjunto vazio caso contrário. Conclua que $\nu = h_*\mu$.

7.2.5. O primeiro enunciado é análogo ao Exercício 7.2.1. A definição garante que μ_k tem memória k . Dado $\varepsilon > 0$ e qualquer função (uniformemente) contínua $\varphi : \Sigma \rightarrow \mathbb{R}$, existe $\kappa \geq 1$ tal que $|\int_C \varphi d\eta - \varphi(x)\eta(C)| \leq \varepsilon\eta(C)$ para todo $x \in C$, todo cilindro C de comprimento $l \geq \kappa$ e toda probabilidade η . Como $\mu = \mu_k$ para cilindros de comprimento k , segue que $|\int \varphi d\mu_k - \int \varphi d\mu| \leq \varepsilon$ para todo $k \geq \kappa$. Isto prova que $(\mu_k)_k$ converge para μ na topologia fraca*.

7.2.6. (1) Use que $P_{i,i}^{n_1+n_2} = \sum_j P_{i,j}^{n_1} P_{j,i}^{n_2}$. Todos os termos nesta expressão são não negativos e o termo correspondente a $j = i$ é positivo. (2) A menos de substituir R por R/κ , podemos supor $\kappa = 1$. Comece por mostrar que se $S \subset \mathbb{Z}$ é fechado por adição e subtração então $S = a\mathbb{Z}$, onde a é o menor elemento positivo de S . Use esse fato para mostrar que se a_1, \dots, a_s são inteiros positivos com máximo divisor comum igual a 1 então existem inteiros b_1, \dots, b_s tais que $b_1 a_1 + \dots + b_s a_s = 1$. Agora tome $a_1, \dots, a_s \in R$ tais que o seu máximo divisor comum seja igual a 1. Usando a observação anterior, e a hipótese de que R é fechado por adição, concluímos que existe $p, q \in R$ tais que $p - q = 1$. Para terminar, mostre que R contém todo inteiro $n \geq pq$. (3) Considere $i, j \in X$ quaisquer e sejam κ_i, κ_j os máximos divisores comuns de $R(i), R(j)$, respectivamente. Por irredutibilidade, existem $k, l \geq 1$ tais que $P_{i,j}^k > 0$ e $P_{j,i}^l > 0$. Deduza que se $n \in R(i)$ então $n + k + l \in R(j)$. Em vista de (2), isso só é possível se $\kappa_i \geq \kappa_j$. Por simetria dos papéis de i e j , segue que também temos $\kappa_i \leq \kappa_j$. Se $\kappa \geq 2$ então, para qualquer i , temos $P_{i,i}^n = 0$ para valores de n arbitrariamente grandes e, conseqüentemente, P não pode ser aperiódica. Agora suponha $\kappa = 1$. Então, usando (2) e a hipótese de que X é finito, existe $m \geq 1$ tal que $P_{i,i}^m > 0$ para todo $i \in X$ e todo $n \geq m$. Usando a irredutibilidade de P e a finitude de X , existe $k \geq 1$ tal que para quaisquer i, j existe $l \leq k$ tal que $P_{i,j}^l > 0$. Deduza que $P_{i,j}^{m+k} > 0$ para todo i, j e, portanto, P é aperiódica. (4) Fixe qualquer $i \in X$ e, para cada $r \in \{0, \dots, \kappa - 1\}$, defina $X_r = \{j \in X : \text{existe } n \equiv r \pmod{\kappa} \text{ tal que } P_{i,j}^n > 0\}$. Verifique que estes conjuntos X_r cobrem X e são disjuntos dois-a-dois. Mostre que a restrição de P^κ a cada um deles é aperiódica.

7.3.1. Pelo teorema de Darboux, existem coordenadas (x_1, x_2) na vizinhança de qualquer ponto de S tais que $\omega = dx_1 \wedge dx_2$. Considere a expressão do campo de vetores nessas coordenadas: $X = X_1(\partial/\partial x_1) + X_2(\partial/\partial x_2)$. Mostre que $\beta = X_1 dx - 2 - X_2 dx_1$ e, portanto, $d\beta = (\text{div } X) dx_1 \wedge dx_2$. Logo, β é fechada se, e somente se, o divergente de X é nulo.

7.3.2. (1) Suponha que π é redutível. Então existe uma posição k tal que os rótulos nas posições 1 a $k - 1$ nas duas linhas de π são os mesmos. Sejam α e β os rótulos na posição k na primeira linha e na segunda linha, respectivamente. Verifique que $\partial I_\alpha = \partial f(I_\beta)$ e deduza que (π, λ) não satisfaz a condição de Keane. (2) Suponha que (π, λ) não satisfaz a condição de Keane: existem $m \geq 1$, α e β tais que $f^m(\partial I_\alpha) = \partial I_\beta \neq 0$. Então $\partial I_\beta = \partial I_\alpha + w_{\gamma(1)} + \dots + w_{\gamma(m)}$, onde os $w_{\gamma(j)}$ são componentes do vetor de translações w . Deduza que o vetor de comprimentos λ é racionalmente dependente, usando que tanto ∂I_α e ∂I_β quanto as componentes de w são combinações lineares inteiras das componentes de λ .

7.3.3. Suponha que $f^m(x) = x$ para algum $m \geq 1$ e algum $x \in I$. Considere o conjunto J dos pontos y tais que $f^j(x)$ e $f^j(y)$ pertencem a um mesmo intervalo I_{β_j} para $j = 0, \dots, m-1$. Note que J é um intervalo e que f^m restrita a J é uma translação. Deduza que $f^m|_J = \text{id}$ e, em particular, $f^m(\partial J) = \partial J$. A definição de J implica que existem $1 \leq k \leq m$ e $\beta \in \mathcal{A}$ tais que $f^k(\partial J) = \partial I_\beta$. Conclua que $f^m(\partial I_\beta) = \partial I_\beta$ e que isto contradiz a condição de Keane, a menos que $\partial I_\beta = 0$, ou seja, a menos que β ocupe a primeira posição na linha de cima de π . Para obter uma contradição também neste segundo caso, argumente que existe $\alpha \in \mathcal{A}$ tal que $f(\partial I_\alpha) = 0 = \partial I_\beta$. Mostre que $f^m(\partial I_\alpha) = \partial I_\alpha$ e que $\partial I_\alpha \neq 0$, uma vez que π é irredutível.

7.3.4. Escreva $\pi = \begin{pmatrix} A_1 & \cdots & A_d \\ B_1 & \cdots & B_d \end{pmatrix}$ e $\lambda = (\lambda_\alpha)_{\alpha \in \mathcal{A}}$. Note que $A_d \neq B_d$, por irreducibilidade, e que $a = \max\{\partial_{A_d}, \partial_{B_d}\}$. Considere k e l tais que $A_k = B_d$ e $B_l = A_d$. Há três casos a considerar, dependendo do tamanho relativo dos intervalos I_{A_d} e I_{B_d} : (i) $|I_{A_d}| < |I_{B_d}|$: Então $a = \partial I_{A_d}$. Decomponha I_{A_d} em dois subintervalos, com rótulos A'_k e A'_d , de tal modo que $|I_{A'_d}| = |I_{A_d}|$. Em outras palavras, $I_{A'_d} = I_{A_k} \cap f^{-1}(I_{A_d})$. Mostre que $g(x) = f^2(x)$ se $x \in I_{A'_d}$ e $g(x) = f(x)$ em todos os demais pontos. Conclua que $R(f)$ está descrita por (π', λ') , com (lembre que $A'_k = B'_d$ e $B'_l = A'_d$)

$$\pi' = \begin{pmatrix} A_1 & \cdots & A'_k & A'_d & \cdots & \cdots & \cdots & A_{d-1} \\ B_1 & \cdots & \cdots & \cdots & \cdots & B'_l & \cdots & B'_d \end{pmatrix}$$

e $\lambda'_\alpha = (\lambda_{A_k} - \lambda_{A_d})/a$ se $\alpha = A'_k$ e $\lambda'_\alpha = \lambda_{A_d}/a$ se $\alpha = A'_d$ e $\lambda'_\alpha = \lambda_\alpha/a$ em todos os demais casos. (ii) $|I_{A_d}| > |I_{B_d}|$: Este caso pode ser reduzido ao anterior, substituindo f pela sua inversa (as duas linhas de π são permutadas) e observando que $R(f^{-1}) = R(f)^{-1}$. (iii) $|I_{A_d}| = |I_{B_d}|$: Neste caso $f(I_{A_k}) = I_{A_d}$ e $g(x) = f^2(x)$ se $x \in A_k$ e $g(x) = f(x)$ em todos os outros casos. A transformação pode ser descrita usando o alfabeto $\mathcal{A}' = \mathcal{A} \setminus \{A_d\}$: elimina o rótulo A_d na linha de cima de π e substitua $B_l = A_d$ por $A_k = B_d$ na linha de baixo.

7.3.5. Observe que f é invertível e que se A é um intervalo d -ádico A de nível $r \geq 1$, (ou seja, um intervalo da forma $A = [id^{-r}, (i+1)d^{-r}]$), existe $s \geq r$ tal que $f(A)$ está formado por d^{s-r} intervalos d -ádicos de nível s . Deduza que f preserva a medida de Lebesgue. Mostre também que se A e B são intervalos d -ádicos então, como σ não tem pontos periódicos, $m(f^k(A) \cap B) = m(A)m(B)$ para todo k suficientemente grande.

7.4.1. Se $(\varphi_n)_n$ é sequência de Cauchy relativamente à norma, dado $\varepsilon > 0$ temos $|\varphi_m(x) - \varphi_n(x)| \leq \varepsilon$ e $|(\varphi_m - \varphi_n)(x) - (\varphi_m - \varphi_n)(y)| \leq \varepsilon d(x, y)^\theta$ para quaisquer x, y e quaisquer m, n suficientemente grandes. A primeira condição implica que $(\varphi_n)_n$ converge uniformemente para alguma função contínua φ . Mostre que $\|\varphi - \varphi_n\| \leq \varepsilon$ para todo n suficientemente grande e, em particular, $\varphi \in \mathcal{E}$.

7.4.2. (1) Dados $y^1, y^2 \in M$, escreva $f^{-1}(y^i) = \{x_1^i, \dots, x_d^i\}$ com $d(x_j^1, x_j^2) \leq \sigma^{-1}d(y^1, y^2)$. Então

$$|\mathcal{L}\varphi(y^1) - \mathcal{L}\varphi(y^2)| = \frac{1}{d} \sum_{j=1}^d |\varphi(x_j^1) - \varphi(x_j^2)| \leq K_\theta(\varphi)\sigma^{-\theta}d(y^1, y^2)^\theta.$$

(2) Segue que $\|\mathcal{L}\varphi\| \leq \sup|\varphi| + \sigma^{-\theta}K_\theta(\varphi) \leq \|\varphi\|$ para todo $\varphi \in \mathcal{E}$, valendo a igualdade se, e somente se, φ é constante. Logo $\|\mathcal{L}\| = 1$. (3) Seja $J_n = [\inf \mathcal{L}^n \varphi, \sup \mathcal{L}^n \varphi]$. Pelo primeiro item, a sequência $(J_n)_n$ é decrescente e o diâmetro de J_n converge exponencialmente para zero. Tome para ν_φ o ponto na interseção e note que $\|\mathcal{L}^n \varphi - \nu_\varphi\| = \sup|\mathcal{L}^n \varphi - \nu_\varphi| + K_\theta(\mathcal{L}^n \varphi)$. (4) As funções constantes são autovetores de \mathcal{L} , associados ao autovalor $\lambda = 1$. Segue que $\nu_{\varphi+c} = \nu_\varphi + c$ para todo $\varphi \in \mathcal{E}$ e todo $c \in \mathbb{R}$. Então $H = \{\varphi : \nu_\varphi = 0\}$ é um hiperplano de \mathcal{E} transversal à reta das funções constantes. Este hiperplano é invariante por \mathcal{L} e, pelo terceiro item, o raio espectral de $\mathcal{L}|_H$ é menor ou igual que $\sigma^{-\theta} < 1$. (5) Pelo segundo item $\|\mathcal{L}^n \varphi - \mathcal{L}^n \psi\| \leq \|\mathcal{L}^k \varphi - \mathcal{L}^k \psi\|$ para todo $n \geq k \geq 1$. Fazendo $n \rightarrow \infty$, vem que $|\nu_\varphi - \nu_\psi| \leq \|\mathcal{L}^k \varphi - \mathcal{L}^k \psi\|$ para todo $k \geq 1$. Usando o primeiro item e fazendo $k \rightarrow \infty$, vem que $|\nu_\varphi - \nu_\psi| \leq \sup|\varphi - \psi|$. Portanto, o operador linear $\psi \mapsto \nu_\psi$ é contínuo relativamente à norma do espaço $C^0(M)$.

8.1.1. Tome $g(x) = 3x - [3x]$ e $\nu = h_*\mu$, onde h é definido da seguinte forma. Escreva $x = \sum_{n=1}^{\infty} a_n 2^{-n}$ com $a_n \in \{0, 1\}$ e $a_n \neq 1$ para infinitos valores de n . Então tome $h(x) =$

$\sum_{n=1}^{\infty} b_n 3^{-n}$, onde $b_n = 0$ se $a_n = 0$ e $b_n = 2$ se $a_n = 1$. O suporte de ν é o conjunto de Cantor usual.

8.1.2. Denote $X_i = X \cap [0; i]$ e $p_i = \mu([0; i])$, para $i = 1, \dots, k$. Como μ é uma medida de Bernoulli, $\mu(X_i) = p_i \mu(f(X_i))$. Logo, $\sum_i p_i \mu(f(X_i)) = 1$. Como $\sum_i p_i = 1$, segue que $\mu(f(X_i)) = 1$ para todo i . Consequentemente, $\cap_i f(X_i)$ tem medida total. Tome x nessa interseção. Se (f, μ) e (g, ν) são ergodicamente equivalentes então existe uma bijeção $\phi : X \rightarrow Y$ entre subconjuntos invariantes com medida total tal que $\phi \circ f = g \circ \phi$. Tome $x \in X$ com k pré-imagens x_1, \dots, x_k em X . Os pontos $\phi(x_i)$ são pré-imagens de $\phi(x)$ para a transformação g . Logo $k \leq l$; por simetria, também temos que $l \leq k$.

8.1.3. Considere todos os cilindros da forma

$$C(m, a_1, \dots, a_{m-1}) = [0; c_0, \dots, c_l, a_1, \dots, a_{m-1}, a_m, \dots, a_{m+l}]$$

com $m > 0$ e $(a_m, \dots, a_{m+l}) = (c_0, \dots, c_l) \neq (a_j, \dots, a_{j+l})$ para $m > j > 0$. Mostre que C é a união disjunta destes cilindros, a menos de medida nula. Defina $\sigma_C = \sigma^{l+m}$ restrito a cada cilindro $C(m, a_1, \dots, a_{m-1})$. Seja $\{C_k : k \in \mathbb{N}\}$ uma enumeração destes cilindros. Observe que σ_C envia cada C_k bijectivamente em C . A cada x em C associe $(k_i)_i \in \mathbb{N}^{\mathbb{N}}$ dada por $\sigma_C^i(x) \in C_{k_i}$. Mostre que a aplicação $h : C \rightarrow \mathbb{N}^{\mathbb{N}}$ assim definida é mensurável invertível e conjuga σ_C ao deslocamento $\sigma_{\mathbb{N}}$ em $\mathbb{N}^{\mathbb{N}}$. Supondo que μ é uma medida de Markov, observe que $\mu_C(E) = \mu_C(\sigma_C(E))\mu_C(C_k)$ para todo $E \subset C_k$ e utilize este fato para mostrar que σ_C preserva μ_C . Mostre que $\nu = h_*\mu_C$ é a medida de Bernoulli associada ao vetor de probabilidade $(p_k)_k$ com $p_k = \mu_C(C_k)$.

8.2.1. Pelo lema de Kakutani-Rokhlin (Exercício 1.4.4), dado n existe B tal que os iterados $f^i(B)$, $i = 0, \dots, n-1$ são disjuntos dois a dois e a sua união tem medida maior que $1 - 1/n$. Tome $\varphi_n = \psi_n / \|\psi_n\|$, onde $\psi_n = \sum_{i=0}^{n-1} \lambda^i \mathcal{X}_{f^i(B)}$. Verifique que $\|U_f \varphi_n - \lambda \varphi_n\| \rightarrow 0$. É claro que todo autovalor aproximado λ está no espectro de U_f : a definição significa que a inversa de $(U_f - \lambda \text{id})$, caso exista, não é um operador limitado.

8.2.2. Usando expansão em série de Fourier, os autovalores de U_α são as potências $e^{2\pi i n \alpha}$, $n \in \mathbb{Z}$ e o autoespaço associado a cada um deles é a reta gerada por $\phi_n(x) = e^{2\pi i n x}$. Para que U_α e U_β tenham os mesmos autovalores é necessário que $\alpha = \pm\beta$. Por outro lado, se $\alpha = -\beta$ então $x \mapsto -x$ define uma equivalência ergódica entre (R_α, m) e (R_β, m) .

8.2.3. A primeira parte é análoga ao Exercício 8.1.2. Para provar a decomposição em soma direta, observe que $U_k(\sum_n a_n e^{2\pi i n x}) = \sum_n a_n e^{2\pi i k n x}$. Logo, basta escrever

$$\sum_{n \in \mathbb{Z}} a_n e^{2\pi i n x} = a_0 + \sum_{j=0}^{\infty} \sum_{n=1}^j a_n e^{2\pi i n x}$$

onde a última soma é sobre os valores de n tais que k^j divide n mas k^{j+1} não divide n . Como ambos os subespaços H_p e H_q têm dimensão de Hilbert enumerável, existe isomorfismo isométrico $h : H_p \rightarrow H_q$. Defina $h \circ U_p^j = U_q^j \circ h$ para cada $j \geq 0$ e $h = \text{id}$ restrito ao subespaço das funções constantes. Em seguida, estenda h a todo o $L^2(m)$, por linearidade. Essa extensão é uma equivalência espectral.

8.2.4. Use o critério do Teorema 8.2.1. Use expansão em série de Fourier para encontrar as soluções de $U_f \varphi = \lambda \varphi$.

8.2.5. Suponha que (f, μ) não é fracamente mixing. Pelo Teorema 8.2.1, existe uma função não constante φ tal que $U_f \varphi = \lambda \varphi$ para algum $\lambda = e^{2\pi i \theta}$. Por ergodicidade, o valor absoluto de φ é constante em μ -quase todo ponto. Usando que f^n é érgódica para todo n (Exercício 4.1.8), θ é irracional e qualquer conjunto onde φ é constante tem medida nula. Dados $\alpha < \beta$ em $[0, 2\pi]$, considere $A = \{x \in \mathbb{C} : \alpha \leq \arg(\varphi(x)) \leq \beta\}$. Mostre que para todo $\varepsilon > 0$ existe n tal que $\mu(f^{-n}(A) \setminus A) < \varepsilon$. Mostre que, escolhendo $|\beta - \alpha|$ pequeno, é possível contradizer a desigualdade no enunciado.

8.2.6. Temos $|z^j - 1| \geq 1$ sempre que $\Re z^j \leq 0$. Supondo que $z \neq 1$, isso acontece para metade dos valores de j , pelo menos. Portanto, nesse caso o primeiro limite é maior ou igual que $1/2$. Quanto ao segundo limite, note que se $z \neq w$ então

$$\left| \frac{1}{n} \sum_{j=0}^{n-1} (z\bar{w})^j \right| = \frac{1}{n} \left| \frac{(z\bar{w})^n - 1}{z\bar{w} - 1} \right| \leq \frac{2}{n |z\bar{w} - 1|}.$$

8.2.7. Temos $f_{n+1}(x) = f_n(x)$ para todo $x \in J_n$ que não está no topo de \mathcal{S}_n . Logo (pelo argumento usado no Exercício 6.1), $f(x) = f_n(x)$ para todo $x \in [0, 1]$ e todo n suficientemente grande; além disso, f preserva a medida de Lebesgue. Seja $a_n = \#\mathcal{S}_n$ a altura de cada pilha \mathcal{S}_n . Representamos por $\{I^e, I^c, I^d\}$ a partição de cada $I \in \mathcal{S}_n$ em subintervalos de igual comprimento, ordenados da esquerda para a direita. (a) Se A é conjunto com $m(A) > 0$, então para todo $\varepsilon > 0$ existe $n \geq 1$ e algum intervalo $I \in \mathcal{S}_n$ tal que $m(A \cap I) \geq (1 - \varepsilon)m(I)$. Se A for invariante, segue que $m(A \cap J) \geq (1 - \varepsilon)m(J)$ para todo $J \in \mathcal{S}_n$. (b) Suponha que $U_f \varphi = \lambda \varphi$. Como U_f é isometria, $|\lambda| = 1$. Por ergodicidade, $|\varphi|$ é constante em quase todo ponto; podemos supor $|\varphi| \equiv 1$. Inicialmente, suponha que existe n e algum intervalo $I \in \mathcal{S}_n$ tal que a restrição de φ a I é constante. Tome $x \in I^e$ e $y \in I^c$ e $z \in I^d$. Então $\varphi(x) = \varphi(y) = \varphi(z)$ e $\varphi(y) = \lambda^{a_n} \varphi(x)$ e $\varphi(z) = \lambda^{a_n+1} \varphi(y)$. Logo $\lambda = 1$ e, por ergodicidade, φ é constante. No caso geral, use o teorema de Lusin (Teoremas A.3.5-A.3.9) para chegar à mesma conclusão. (c) A é uma união de intervalos I_j na pilha \mathcal{S}_n para cada $n \geq 2$. Então, $f^{a_n}(I_j^e) = I_j^c$ para todo j . Logo, $m(f^{a_n}(A) \cap A) \geq m(A)/3 = 2/27$.

8.3.1. Seja $\{v_j : j \in \mathcal{I}\}$ uma base de H formada por autovetores com norma 1 e seja λ_j o autovalor associado a cada autovetor v_j . A hipótese garante que podemos considerar $\mathcal{I} = \mathbb{N}$. Mostre que para todo $\delta > 0$ e todo $k \geq 1$ existe $n \geq 1$ tal que $|\lambda_j^n - 1| \leq \delta$ para todo $j \in \{1, \dots, k\}$ (isto pode ser feito usando o princípio das casas dos pombos). Decomponha $\varphi = \sum_j c_j v_j$, com $c_j \in \mathbb{C}$. Observe que $U_f^n \varphi = \sum_{j \in \mathbb{N}} c_j \lambda_j^n v_j$ e, portanto,

$$\|U_f^n \varphi - \varphi\|_2^2 \leq \sum_{j=1}^k |c_j(\lambda_j^n - 1)|^2 + \sum_{j=k+1}^{\infty} 2|c_j|^2 \leq \delta^2 \|\varphi\|_2^2 + \sum_{j=k+1}^{\infty} 2|c_j|^2.$$

Dado $\varepsilon > 0$, podemos escolher δ e k de tal modo que cada uma das parcelas do lado direito seja menor que $\varepsilon/2$.

8.3.2. Já sabemos que (a) implica (b). Para mostrar que (b) implica (c): pela Proposição 8.3.4 e por (8.3.4), temos que $\{e^{2\pi i k \cdot \theta} : k \in \mathbb{Z}^d\} = \{e^{2\pi i l \cdot \tau} : l \in \mathbb{Z}^d\}$. Em particular, cada grupo de autovalores contém os geradores do outro grupo: $e^{2\pi i \theta_j} \in G_\tau$ para todo $j = 1, \dots, d$ e $e^{2\pi i \tau_j} \in G_\theta$ para todo $j = 1, \dots, d$. A primeira condição quer dizer que para cada j existe $L_j = (L_{j,1}, \dots, L_{j,d}) \in \mathbb{Z}^d$ e existe $l_j \in \mathbb{Z}$ tais que $\theta_j = L_j \cdot \tau + l_j$. Seja L a matriz $(L_{j,k})_{j,k}$ e seja l o vetor $(l_j)_j$. Então $\theta = L\tau + l$. Analogamente, a segunda condição significa que existem $M \in \text{GL}(d, \mathbb{Z})$ e $m \in \mathbb{Z}^d$ tais que $\tau = M\theta + m$. Então, $\theta = LM\theta + (Lm + l)$ e $\tau = ML\tau + (Ml + m)$. Como θ e τ são racionalmente independentes, segue que $LM = \text{id} = ML$ e $Lm + l = 0 = Ml + m$. Segue que $\det L = \det M = \pm 1$, ou seja, as duas matrizes estão em $\text{SL}(d, \mathbb{Z})$. Para mostrar que (c) implica (a): represente por $\hat{R}_\theta : \mathbb{R}^d \rightarrow \mathbb{R}^d$ a translação $x \mapsto x + \theta$ e analogamente para \hat{R}_τ . A hipótese de que $\theta = L\tau$ módulo \mathbb{Z}^d dá que $L \circ \hat{R}_\tau \circ L^{-1}(x) = L(L^{-1}(x) + \tau) = x + \theta = \hat{R}_\theta(x)$ módulo \mathbb{Z}^d . Seja $h_L : \mathbb{T}^d \rightarrow \mathbb{T}^d$ o automorfismo linear induzido por L . Projetando a relação anterior para o toro, obtemos que $h_L \circ R_\tau \circ h_L^{-1} = R_\theta$. Como h_L preserva a medida de Lebesgue m , temos que as duas rotações são ergodicamente equivalentes.

8.3.3. Se R_θ é equivalente a R_τ então R_θ^s é equivalente a R_τ^s , ou seja, $R_{ps/q}$ é equivalente à identidade. Deduza que $s \geq q$; por simetria, $q \geq s$. Se $q = s$, defina $h = \text{id}$ no intervalo $[0, 1/q]$ e estenda h a S^1 de tal modo que $h \circ R_\theta = R_\tau \circ h$. Verifique que h fica bem definida e é uma equivalência ergódica.

8.4.1. Adapte os argumentos usados no caso dos deslocamentos de Bernoulli (Exemplos 8.4.2 e 8.4.3), da seguinte forma. No caso unilateral, as condições (a) e (c) na definição são evidentes. Para provar (b), observe que A_c pertence à σ -álgebra gerada pelos cilindros da forma $[n; c_n, \dots, c_m]$ para quaisquer $m \geq n$. Então, dado qualquer $C = [0; c_0, \dots, c_k]$ com $k < n$,

$$\mu(A_c \cap C) = \sum_i \mu(A_c \cap [n; i] \cap C) = \sum_i \mu(A \cap [n; i]) \frac{P^{n-k}_{c_k, i}}{p_i} \mu(C).$$

Por hipótese, $P^{n-k}_{c_k, i} p_i \rightarrow 1$ quando $n \rightarrow \infty$. Segue que $\mu(A_c \cap C) = \mu(A_c) \mu(C)$ e agora o argumento é análogo ao caso Bernoulli. No caso bilateral, a condição (a) é imediata, (b) segue dos argumentos que acabamos de usar e (c) é análoga ao caso Bernoulli.

8.4.2. Se a órbita de algum $v \neq 0$ é finita, existem $m < n$ tais que $(A^t)^m v = (A^t)^n v$. Mostre que algum autovalor de A^t é raiz da unidade e, portanto, o mesmo vale para A . Seja $\varphi_k(x) = e^{2\pi i k \cdot x}$, $k \in \mathbb{Z}^d$ a base de Fourier de $L^2(\mu)$. Observe que $U_f(\varphi_k) = \varphi_{A^t(k)}$. Tome $S \subset \mathbb{Z}^d \setminus \{0\}$ contendo exatamente um ponto em cada órbita de A^t . Tome $F =$ subespaço gerado por $\{\varphi_k : k \in S\}$ para mostrar que (f, μ) tem espectro de Lebesgue. Reciprocamente, se (f, μ) tem espectro de Lebesgue então (f^n, μ) é ergódico para todo n . Pelo Exercício 4.1.8, segue que nenhum autovalor de U_f é raiz da unidade.

8.4.3. Seja $U : H \rightarrow H$ uma isometria não invertível. Lembrando do Exercício 2.3.6: Mostre que existem subespaços fechados V e W de H tais que $U : H \rightarrow H$ é conjugado, por um operador unitário, ao operador $U_1 : V \oplus W^{\mathbb{N}} \rightarrow V \oplus W^{\mathbb{N}}$ dado por $U_1|_V = U|_V$ e $U_1|_{W^{\mathbb{N}}} = \text{id}$. Seja $U_2 : V \oplus W^{\mathbb{Z}} \rightarrow V \oplus W^{\mathbb{Z}}$ o operador linear definido por $U_1|_V = U|_V$ e $U_1|_{W^{\mathbb{Z}}} = \text{id}$. Verifique que U_2 é um operador unitário tal que $U_2 \circ j = j \circ U_1$, onde $j : V \oplus W^{\mathbb{N}} \rightarrow V \oplus W^{\mathbb{Z}}$ é a inclusão natural. Mostre que se $E \subset V \oplus W^{\mathbb{N}}$ satisfaz as condições da definição de espectro de Lebesgue para U_1 então $j(E)$ satisfaz essas mesmas condições para U_2 . Conclua que o posto de U_1 está bem definido.

8.4.4. Por um lado, pelo teorema de Fubini,

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(x^2+y^2)/2} dx dy = \left(\int_{-\infty}^{+\infty} e^{-x^2/2} dx \right)^2.$$

Por outro lado, usando coordenadas polares,

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(x^2+y^2)/2} dx dy = \int_0^{2\pi} \int_0^{\infty} e^{-\rho^2/2} \rho d\rho d\theta = 2\pi \int_0^{\infty} e^{-s} ds = 2\pi.$$

Para a segunda parte do enunciado, tome B tal que $BB^* = A$ e em seguida faça a mudança de variável $z = B(w)$.

8.4.5. Use a Proposição 8.4.10.

8.4.6. O lema de Riemann-Lebesgue garante que F toma valores em c_0 . O operador F é contínuo: $\|F(\varphi)\| \leq \|\varphi\|$ para todo $\varphi \in L^1(\lambda)$. Além disso, F é injetivo: se $F(\varphi) = 0$ então $\int \varphi(z)\psi(z) d\lambda(z) = 0$ para toda combinação linear $\psi(z) = \sum_{|j| \leq l} a_j z^j$, $a_j \in \mathbb{C}$. Dado qualquer intervalo $I \subset S^1$, a sequência $\psi_N = \sum_{|n| \leq N} c_n z^n$, $c_n = \int_I z^{-n} d\lambda(z)$ das somas parciais da série de Fourier da função característica χ_I é limitada (veja [Zyg68, página 90]). Usando o teorema da convergência dominada, segue que $F(\varphi) = 0$ implica $\int_I \varphi(z) d\lambda(z) = 0$, qualquer que seja o intervalo I . Logo $\varphi = 0$. Se F fosse bijetivo então, pelo teorema da aplicação aberta, o seu inverso seria um operador linear contínuo. Então existiria $c > 0$ tal que $\|F(\varphi)\| \geq c\|\varphi\|$ para todo $\varphi \in L^1(\lambda)$. Mas isso é falso: considere $D_N(z) = \sum_{|n| \leq N} z^n$ para $N \geq 0$. Verifique que $F(D_N) = (a_n^N)_n$ com $a_n^N = 1$ se $|n| \leq N$ e $a_n^N = 0$ caso contrário. Logo $\|F(D_N)\| = 1$ para todo N . Escrevendo $z = e^{2\pi i t}$, verifique que $D_N(z) = \text{sen}((2N+1)\pi t) / \text{sen}(\pi t)$. Conclua que $\|D_N\| = \int |D_N(z)| d\lambda(z)$ converge para infinito quando $N \rightarrow \infty$. [Observação: Também é possível dar exemplos explícitos. Por exemplo, se $(a_n)_n$ é uma sequência convergindo para zero tal que $\sum_{n=1}^{\infty} a_n/n = \infty$ então a sequência $(\alpha_n)_n$ dada por $\alpha_n = a_n/(2i)$ para $n \geq 1$ e $\alpha_n + \alpha_{-n} = 0$ para todo $n \geq 0$ não pode ser escrita na forma $\alpha_n = \int z^n d\nu(z)$. Veja a Seção 7.3.4 do livro de Edwards [Edw79].]

8.5.1. Inicialmente, suponha que $\{B_i : 1 \leq i \leq l\}$ e $\{C_j : 1 \leq j \leq k\}$ são partições. Mostre que $\sum_i b_i \mathcal{X}_{B_i} = \sum_j c_j \mathcal{X}_{C_j}$ se, e somente se, $b_i = c_j$ sempre que B_i intersecta C_j . Isso acontece sempre que $H(B_i)$ intersecta $H(C_j)$. Conclua que $\sum_i b_i \mathcal{X}_{H(B_i)} = \sum_j c_j \mathcal{X}_{H(C_j)}$. Reduza o caso geral ao caso de partições, substituindo $\{B_i : 1 \leq i \leq l\}$ e $\{C_j : 1 \leq j \leq k\}$ por $\{B'_i : 0 \leq i \leq l\}$ e $\{C'_j : 0 \leq j \leq k\}$ definidos por $B'_i = B_i \setminus (B_1 \cup \dots \cup B_{i-1})$ para $i = 1, \dots, s$ e $B_0 = (B_1 \cup \dots \cup B_s)^c$ e $b_0 = 0$ e analogamente para C'_j .

8.5.2. Para mostrar que \tilde{g} é injetivo, note que $\mu(g^{-1}(A)\Delta g^{-1}(B)) = \mu(A\Delta B)$. Se $g = h$ em quase todo ponto, é claro que $\tilde{g} = \tilde{h}$. Para a recíproca, comece por supor que N não tem átomos. Então (Proposição 8.5.5), podemos considerar $N = [0, 1]$. Se $\{x : g(x) \neq h(x)\}$ tem medida positiva, existe $r \in \mathbb{Q}$ tal que $\{x : g(x) < r < h(x)\} \cup \{x : g(x) > r > h(x)\}$ tem medida positiva. Seja $I = [0, r]$. Conclua que $g^{-1}(I)\Delta h^{-1}(I)$ tem medida positiva, ou seja, $\tilde{g}(I) \neq \tilde{h}(I)$. Estenda este argumento para o caso em que N tem átomos.

8.5.3. Pelo Exercício 8.5.2, \tilde{f} é sempre injetivo. Conclua que se \tilde{f} é sobrejetivo então ele é invertível: existe homomorfismo de álgebras de medida $h : \tilde{\mathcal{B}} \rightarrow \tilde{\mathcal{B}}$ tal que $h \circ \tilde{f} = \tilde{f} \circ h = \text{id}$. Use a Proposição 8.5.6 para encontrar $g : M \rightarrow M$ tal que $g \circ f = f \circ g$ em μ -quase todo ponto. A recíproca é fácil: se (f, μ) é invertível em quase todo ponto então o homomorfismo de álgebras de medida \tilde{g} associado a $g = f^{-1}$ satisfaz $\tilde{g} \circ \tilde{f} = \tilde{f} \circ \tilde{g} = \text{id}$; em particular, \tilde{f} é sobrejetivo.

8.5.4. Se U_f é sobrejetivo, para todo conjunto mensurável B existe função mensurável φ tal que $\mathcal{X}_B = \varphi \circ f$ em quase todo ponto. Mostre que φ é função característica de algum conjunto mensurável, a menos de medida nula. Deduza que \tilde{f} é sobrejetivo. Se \tilde{f} é sobrejetivo então todo conjunto mensurável é da forma $f^{-1}(B)$ para algum $B \in \mathcal{B}$. Conclua que a imagem de U_f contém a função característica de todo conjunto mensurável. Deduza que U_f é sobrejetivo. A última parte do enunciado segue do Exercício 8.5.3.

8.5.5. Verifique que se (f, μ) tem espectro discreto então U_f é invertível. Use o Exercício 8.5.4 para concluir.

8.5.6. Verifique que as uniões de elementos de $\cup_n \mathcal{P}_n$ são pré-imagens, pela inclusão ι , de abertos de K . Use esse fato para mostrar que se as cadeias têm medida nula então para cada $\delta > 0$ existe um aberto $A \subset K$ tal que $m(A) < \delta$ e todo ponto fora de A está na imagem da inclusão: em outras palavras, $K \setminus \iota(M\mathcal{P}) \subset A$. Conclua que $\iota(M\mathcal{P})$ é conjunto mensurável de Lebesgue e o seu complementar em K tem medida nula. Para a recíproca, use o fato de que (a) implica (c) no Exercício A.1.13.

8.5.7. Inicialmente, suponha que M não tem átomos. Pela Proposição 8.5.5, podemos supor que $M = [0, 1]$. Então, basta usar o Exercício 5.2.4. O caso geral, com átomos, é análogo.

8.5.8. Seja \mathcal{B} a σ -álgebra dos subconjuntos mensuráveis de Lebesgue $[0, 1]$. Para mostrar que μ está bem definida, considere $A, B \in \mathcal{B}$ tais que $M \cap A = M \cap B$. Verifique que $m(A \Delta B) = 0$ e, portanto, $m(A) = m(B)$. Para provar aditividade, considere $A, B \in \mathcal{B}$ tais que $M \cap A$ e $M \cap B$ são disjuntos. Verifique que $m(A \cap B) = 0$ e deduza que $m(A \cup B) = m(A) + m(B)$. Adapte este argumento, para provar que μ é σ -aditiva. Note que $\mu(M) = 1$. Logo μ é uma probabilidade em (M, \mathcal{M}) . Em seguida, verifique que a aplicação $H : \mathcal{B} \rightarrow \mathcal{M}$ definida por $H(B) = M \cap B$ é um isomorfismo de álgebras de medida. Mostre que este isomorfismo não é induzido por nenhuma aplicação mensurável $h : M \rightarrow [0, 1]$: a aplicação h teria que ser a identidade em quase todo ponto de M . Conclua que (M, \mathcal{M}, μ) não pode ser espaço de Lebesgue.

9.1.1. Dada qualquer partição infinita enumerável $\mathcal{P} = \{P_k : k \geq 1\}$ com entropia finita, considere a sequência de partições finitas $\mathcal{P}_k = \{P_1, \dots, P_k, Q_k = \cup_{j>k} P_j\}$. Então $h_\mu(f, \mathcal{P}_k) \leq h_\mu(f, \mathcal{P}) \leq h_\mu(f, \mathcal{P}_k) + H_\mu(\mathcal{P}/\mathcal{P}_k)$ para todo k . Usando que $H_\mu(\mathcal{P}) < \infty$, verifique que

$$H_\mu(\mathcal{P}/\mathcal{P}_k) = \sum_{n>k} -\mu(P_n) \log \frac{\mu(P_n)}{\mu(Q_k)} \rightarrow 0 \quad \text{quando } k \rightarrow \infty.$$

Portanto, $h_\mu(f, \mathcal{P}) = \sup_k h_\mu(f, \mathcal{P}_k)$.

9.1.2. Seja $g = f^k$. Então, $H_\mu(\cup_{i=0}^{k-1} f^{-i}(\mathcal{P}) / \cup_{j=k}^n f^{-j}(\mathcal{P})) = H_\mu(\mathcal{P}^k / \cup_{i=1} g^{-i}(\mathcal{P}^k))$. Pelo Lema 9.1.12, esta expressão converge para $h_\mu(g, \mathcal{P}_k)$. Agora basta usar o Lema 9.1.13.

9.1.3. (a) Dadas partições \mathcal{P} e \mathcal{Q} de A e B , respectivamente, considere a partição $\mathcal{R} = \mathcal{P} \cup \mathcal{Q}$ de M . Note que $\mathcal{R}^n = \mathcal{P}^n \cup \mathcal{Q}^n$ para todo n e deduza que

$$h_\mu(f, \mathcal{R}) = \mu(A)h_{\mu_A}(f, \mathcal{P}) + \mu(B)h_{\mu_B}(f, \mathcal{Q}).$$

Variando \mathcal{P} e \mathcal{Q} , conclua que $h_\mu(f) \geq \mu(A)h_{\mu_A}(f) + \mu(B)h_{\mu_B}(f)$. Reciprocamente, dada uma partição \mathcal{R} de M , observe que $\mathcal{R}' = \mathcal{R} \vee \{A, B\}$ pode ser escrita $\mathcal{R}' = \mathcal{P} \cup \mathcal{Q}$, onde \mathcal{P} e \mathcal{Q} são partições de A e B . Logo, $h_\mu(f, \mathcal{R}) \leq h_\mu(f, \mathcal{R}') \leq \mu(A)h_{\mu_A}(f) + \mu(B)h_{\mu_B}(f)$. (b) Pelo Lema 4.3.3, existem conjuntos invariantes A_1, \dots, A_n disjuntos dois-a-dois tais que $\mu_i(A_i) = 1$ para todo i . Agora prove o enunciado por indução em n , usando o item (a).

9.1.4. O conjunto \mathcal{A} das uniões finitas disjuntas de retângulos $A_i \times B_i$, com $A_i \subset M$ e $B_i \subset N$ é uma álgebra; por definição, ela gera a σ -álgebra de $M \times N$. Verifique que $f \times g$

preserva a restrição de $\mu \times \nu$ a esta álgebra e use o Lema 1.3.1 para concluir que $\mu \times \nu$ é invariante por $f \times g$. Dadas partições finitas \mathcal{P} e \mathcal{Q} de M e N , respectivamente, seja $\mathcal{P} \times \mathcal{Q} = \{P \times Q : P \in \mathcal{P} \text{ e } Q \in \mathcal{Q}\}$. Verifique que $h_{\mu \times \nu}(f \times g, \mathcal{P} \times \mathcal{Q}) = h_\mu(f, \mathcal{P}) + h_\nu(g, \mathcal{Q})$. Isto implica que $h_{\mu \times \nu}(f \times g) \geq h_\mu(f) + h_\nu(g)$. Para provar a desigualdade recíproca, observe que dada qualquer partição finita \mathcal{R} de $M \times N$ existem \mathcal{P} e \mathcal{Q} tais que $\mathcal{R} \prec \mathcal{P} \times \mathcal{Q}$.

9.1.5. Dada qualquer partição finita \mathcal{Q} de N , considere $\mathcal{P} = \{\phi^{-1}(Q) : Q \in \mathcal{Q}\}$. Verifique que $h_\mu(f, \mathcal{P}) = h_\nu(g, \mathcal{Q})$ e deduza que $h_\mu(f) \geq h_\nu(g)$.

9.2.1. Tome $\varepsilon = 1/20$. Considere $x \neq y$ e considere as respectivas expansões decimais $x = 0, x_1 \dots x_n \dots$ e $y = 0, y_1 \dots y_n \dots$. Se existe $k \geq 1$ tal que $|x_k - y_k| \geq 2$ então $|f^k(x) - f^k(y)| \geq 1/10 > \varepsilon$. Suponha que $|x_n - y_n| \leq 1$ para todo n . Tome $k \geq 1$ mínimo tal que $x_k \neq y_k$. Como $|x_{k+1} - y_{k+1}| \leq 1$, temos que $|f^k(x) - f^k(y)| > \varepsilon$.

9.2.2. A entropia de uma medida de Bernoulli em $\{1, \dots, d\}^{\mathbb{Z}}$ é dada por $\sum_{i=1}^d -p_i \log p_i$. Variando os pesos p_i esta expressão toma qualquer valor entre 0 e $\log d$. O número d de símbolos pode ser tomado arbitrariamente grande.

9.2.3. (a) Para todo $k \geq 1$, as órbitas de 0 e de $x_k = (1/k, \dots, 1/k, \dots)$ permanecem a distância $\leq 1/k$ para todos os iterados. (b) O sistema (σ, μ_k) é ergodicamente equivalente ao deslocamento de Bernoulli em $\{1, 2\}^{\mathbb{N}}$ com pesos $p_1 = p_2 = 1/2$. Logo $h_{\mu_k}(\sigma) = \log 2$. O suporte de μ_k está contido na bola fechada de raio $1/k$ em torno de $0 = (0, \dots, 0, \dots)$. Logo $\mu_k \rightarrow 0$ na topologia fraca*. Claro que $h_{\delta_0}(\sigma) = 0$. (c) Para cada k , seja \mathcal{P}_k a partição (finita) de Σ cujos elementos são os cilindros $[0; i]$ com $i = 1, \dots, k$ e $[0; X \setminus \{1, \dots, k\}]$. Mostre que $h_\mu(\sigma, \mathcal{P}_k) \geq \sum_{i=1}^k -p_{1/i} \log p_{1/i}$ e conclua que $h_\mu(\sigma) = \infty$.

9.2.4. Seja \mathcal{P} uma partição finita de S^1 em intervalos. Para cada $j \geq 1$ e $P \in \mathcal{P}$, a pré-imagem $f^{-j}(P)$ está formada por d^j intervalos, cujos extremos são as pré-imagens dos pontos extremos de P . Deduza que $\#\mathcal{P}^n \leq d^n \#\mathcal{P}$ para todo n e conclua que $h_\mu(f, \mathcal{P}) \leq \log d$. Termine usando o Corolário 9.2.6.

9.2.5. Escreva $\mathcal{Q}^n = \bigvee_{j=0}^{n-1} f^{-j}(\mathcal{Q})$ para cada n e seja \mathcal{A} a σ -álgebra gerada por $\bigcup_n \mathcal{Q}^n$. Verifique que f é mensurável relativamente à σ -álgebra \mathcal{A} . Mostre que a hipótese implica que $\mathcal{P} \subset \mathcal{A}$. Pelo Lema 9.2.2, segue que $H_\mu(\mathcal{P}/\mathcal{Q}^n)$ converge para zero. Pelos Lemas 9.1.11 e 9.1.13, temos $h_\mu(f, \mathcal{P}) \leq h_\mu(f, \mathcal{Q}) + H_\mu(\mathcal{P}/\mathcal{Q}^n)$ para todo n .

9.3.1. Para cada $n \geq 1$, defina $\varphi_n : M \rightarrow [0, \infty)$ e $\rho_n : (0, \infty) \rightarrow [0, 1]$ por

1. $\varphi_n(x) = (i-1)/n$ se $\varphi(x) \in ((i-1)/n, i/n]$ com $1 \leq i \leq n^2$ e $\varphi_n(x) = n$ se $\varphi(x) > n$;
2. $\rho_n(s) = \mu(\{\varphi > i/n\})$ se $s \in ((i-1)/n, i/n]$ com $1 \leq i \leq n^2$ e $\rho(s) = 0$ se $s > n$.

Verifique que, para todo n ,

$$\int \varphi_n d\mu = \sum_{i=1}^{n^2} \frac{1}{n} \mu(\{\varphi > \frac{i}{n}\}) = \int_0^\infty \rho_n(s) ds.$$

Use o teorema da convergência dominada para concluir o enunciado do exercício.

9.3.2. Seja μ a medida de Bernoulli associada a um vetor de probabilidade (p_1, \dots, p_d) . Defina $\eta_i(x, n) = \#\{0 \leq j \leq n-1 : x_j = i\}$. Seja \mathcal{P} a partição (geradora) de Σ em cilindros $[0; i]$. Note que

$$\mu(\mathcal{P}^n(x)) = \prod_{i=1}^d p_i^{\eta_i(x, n)}.$$

Por ergodicidade, $\eta_i(x, n)/n \rightarrow p_i$ em μ -quase todo x . Logo, $-(1/n) \log \mu(\mathcal{P}^n(x))$ converge para $\sum_{i=1}^d -p_i \log p_i$ em μ -quase todo ponto.

9.3.3. É claro que $B(x, n, \varepsilon) \subset B(f(x), n-1, \varepsilon)$. Logo, $h_\mu(f, x) \geq h_\mu(f, f(x))$ para μ -quase todo x . Por outro lado, $\int h_\mu(f, x) d\mu(x) = \int h_\mu(f, f(x)) d\mu(x)$ pois a medida μ é invariante por f .

9.3.4. Pelo Teorema 9.3.1 e por ergodicidade, $-(1/n) \log \mu(\mathcal{P}^n(x)) \rightarrow h_\mu(f, \mathcal{P})$ para μ -quase todo x . Logo, dado $\varepsilon > 0$,

$$\{x \in M : h_\mu(f, \mathcal{P}) - \varepsilon < -\frac{1}{n} \log \mu(\mathcal{P}^n(x)) < h_\mu(f, \mathcal{P}) + \varepsilon\}$$

tem medida maior que $1 - \varepsilon$ para todo n suficientemente grande. Este conjunto é uma união de átomos de $\mathcal{P}^n(x)$.

9.4.1. Escrevendo $\theta = (\theta_1, \dots, \theta_d)$, vemos que $R_\theta = R_{\theta_1} \times \dots \times R_{\theta_d}$ onde R_{θ_i} é uma rotação em S^1 . Além disso, a medida de Haar em \mathbb{T}^d é o produto das medidas de Haar em d cópias de S^1 . Logo, pelo Exercício 9.1.4, basta tratar o caso $d = 1$. Use um argumento semelhante ao do Exercício 9.2.4: para toda partição finita \mathcal{P} temos que $\#\mathcal{P}^n \leq n\#\mathcal{P}$ para todo n e, portanto, $h_\mu(f, \mathcal{P}) = 0$ para toda probabilidade invariante μ .

9.4.2. Use o seguinte fato, que é a seguinte consequência da forma canônica de Jordan: existem números $\rho_1, \dots, \rho_l > 0$, existe uma decomposição A -invariante $\mathbb{R}^d = E_1 \oplus \dots \oplus E_l$ e, dado $\alpha > 0$, existe um produto interno em \mathbb{R}^d relativamente ao qual os subespaços E_j são ortogonais e satisfazem $e^{-\alpha}\rho_j\|v\| \leq \|Av\| \leq e^\alpha\rho_j\|v\|$ para todo $v \in E_j$. Além disso, os ρ_i são os valores absolutos dos autovalores de A e eles satisfazem $\sum_{i=1}^d \log^+ |\lambda_i| = \sum_{j=1}^l \dim E_j \log^+ \rho_j$.

9.4.3. Considere qualquer partição finita ou enumerável \mathcal{P} com $\{B, B^c\} \prec \mathcal{P}$. Seja \mathcal{Q} a restrição de \mathcal{P} ao conjunto B . Escreva $\mathcal{P}^n = \bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P})$ e $\mathcal{Q}^k = \bigvee_{j=0}^{k-1} g^{-j}(\mathcal{Q})$. Verifique que, para todo $x \in \mathcal{B}$ e $k \geq 1$ existe $n_k \geq 1$ tal que $\mathcal{Q}^k(x) = \mathcal{P}^{n_k}(x)$. Além disso, por ergodicidade $\lim_k k/n_k = \tau(B, x) = \mu(B)$ para quase todo x . Pelo teorema de Shannon-McMillan-Breiman,

$$h_\nu(g, \mathcal{Q}, x) = \lim_k -\frac{1}{k} \log \nu(\mathcal{Q}^k(x)) \quad \text{e} \quad h_\mu(f, \mathcal{P}, x) = \lim_k -\frac{1}{n_k} \log \mu(\mathcal{P}^{n_k}(x)).$$

Conclua que $h_\nu(f, \mathcal{Q}, x) = \mu(B)h_\nu(g, \mathcal{Q}, x)$ para quase todo $x \in B$. Variando \mathcal{P} , deduza que $h_\nu(f) = \mu(B)h_\nu(g)$.

9.4.4. Dada uma partição \mathcal{P} de M , considere $\hat{\mathcal{P}} = \pi^{-1}(\mathcal{P})$, ou seja $\hat{\mathcal{P}} = \{\pi^{-1}(P) : P \in \mathcal{P}\}$. Claro que $\hat{\mathcal{P}}$ é uma partição de \hat{M} . Mostre que $\hat{\mathcal{P}}^n = \pi^{-1}(\mathcal{P}^n)$ e, logo, $H_\mu(\mathcal{P}^n) = H_{\hat{\mu}}(\hat{\mathcal{P}}^n)$ para todo n . Portanto, $h_\mu(f, \mathcal{P}) = h_{\hat{\mu}}(\hat{f}, \hat{\mathcal{P}})$ para toda partição \mathcal{P} de M . Suponha que existe gerador unilateral \mathcal{P} para (f, μ) . Mostre que $\hat{\mathcal{P}}$ é gerador bilateral para $(\hat{f}, \hat{\mu})$ e conclua que $h_\mu(f) = h_{\hat{\mu}}(\hat{f})$. Em geral, como (M, μ) é espaço de Lebesgue, sempre existem partições finitas $\mathcal{P}_1 \prec \dots \prec \mathcal{P}_k \prec \dots$ tais que $\cup_k \mathcal{P}_k$ gera a σ -álgebra de M . Verifique que $\cup_k \hat{\mathcal{P}}_k^{\pm k}$ gera a σ -álgebra de \hat{M} . Use o Lema 9.1.13 e o teorema de Kolmogorov-Sinai para concluir que $h_\mu(f) = \lim_k h_\mu(f, \mathcal{P}_k) = \lim_k h_{\hat{\mu}}(\hat{f}, \hat{\mathcal{P}}_k) = h_{\hat{\mu}}(\hat{f})$.

9.5.1. Como $h_\mu(f) = h_\mu(f^{-1})$, segue do teorema de Ornstein (Teorema 9.5.2) que (f, μ) e (f^{-1}, μ) são ergodicamente equivalentes. Dado qualquer $k \geq 1$, o Exercício 9.2.2 garante que existe algum automorfismo de Bernoulli (g, ν) tal que $h_\nu(g) = h_\mu(f)/k$. Então, pela Proposição 9.1.14, $h_\nu(g^k) = h_\mu(f)$ e segue do teorema de Ornstein que (f, μ) e (g^k, ν) são ergodicamente equivalentes.

9.5.2. Considere $A \in \cap_n f^{-n}(\mathcal{B})$ com $m(A) > 0$. Então, para cada n existe $A_n \in \mathcal{B}$ tal que $A = f^{-n}(A_n)$. Considere os intervalos $I_{j,n} = ((j-1)/10^n, j/10^n)$. Então

$$\frac{m(A \cap I_{j,n})}{m(I_{j,n})} = \frac{m(A_n)}{m((0, 1))} = m(A) \quad \text{para todo } 1 \leq j \leq 10^n.$$

Fazendo $n \rightarrow \infty$, concluímos que A^c não tem pontos de densidade. Logo $m(A^c) = 0$.

9.5.3. Análogo ao Exercício 9.5.2, usando a propriedade de distorção limitada dada pela Proposição 4.2.11.

9.5.4. Seja \mathcal{A} a σ -álgebra gerada pelos cilindros $[0; b_0, \dots, b_{n-1}]$ com $n \geq 1$. É claro que \mathcal{A} satisfaz as condições (a) e (c) na Definição 9.5.8. Para provar a condição (b), ou seja, que todo elemento de $\mathcal{Z} = \cap_k \sigma^{-k}(\mathcal{A})$ tem medida zero ou um, imite a demonstração do Teorema 7.2.11, da seguinte forma. Dados cilindros $A = [p; a_0, \dots, a_{m-1}]$ e $B = [0; b_0, \dots, b_{n-1}]$,

$$\min_{i,j} \frac{P_{i,j}^s}{p_j} \leq \frac{\mu(A \cap \sigma^{-k}(B))}{\mu(A)\mu(B)} \leq \max_{i,j} \frac{P_{i,j}^s}{p_j},$$

para todo $k \geq p+m$, onde $s = k - p - m + 1$. A hipótese implica que $P_{i,j}^s \rightarrow p_j$ quando $s \rightarrow \infty$, para todo i, j . Conclua que, dados $\varepsilon > 0$ e $A = [p; a_0, \dots, a_{m-1}]$, existe $k_0 \geq 1$ tal que a

família $\mathcal{C}(\varepsilon, A)$ dos conjuntos $B \in \mathcal{A}$ que satisfazem $|\mu(A \cap \sigma^{-k}(B)) - \mu(A)\mu(B)| \leq \varepsilon\mu(A)\mu(B)$ para todo $k \geq k_0$ é uma classe monótona e contém toda união finita disjunta de cilindros da forma $[0; b_0, \dots, b_{n-1}]$. Logo, $\mathcal{C}(\varepsilon, A) = \mathcal{A}$. Conclua que $|\mu(A \cap Z) - \mu(A)\mu(Z)| \leq \varepsilon\mu(A)\mu(Z)$ para todo $Z \in \mathcal{Z}$ e todo cilindro $A = [p; a_0, \dots, a_{m-1}]$. Verifique que a família \mathcal{D}_ε dos conjuntos $A \in \mathcal{B}$ tais que $|\mu(A \cap Z) - \mu(A)\mu(Z)| \leq \varepsilon\mu(A)\mu(Z)$ para todo $Z \in \mathcal{Z}$ é uma classe monótona e contém toda união finita disjunta de cilindros. Logo, $\mathcal{D}_\varepsilon = \mathcal{B}$. Em particular, $|\mu(Z) - \mu(z)^2| \leq \varepsilon\mu(Z)^2$ para todo $z \in \mathcal{Z}$ e todo $\varepsilon > 0$. Deduza que a σ -álgebra \mathcal{Z} é trivial.

9.5.5. Seja \mathcal{A} a σ -álgebra gerada pelos cilindros $[0; b_0, \dots, b_{n-1}]$ com $n \geq 1$. É claro que \mathcal{A} satisfaz as condições (a) e (c) na Definição 9.5.8. Para provar a condição (b), use o argumento do Exercício 9.5.4.

9.5.6. Suponha que $h_\mu(f, \mathcal{P}) = 0$. Use o Lema 9.5.4 para mostrar que $\mathcal{P} \prec \bigvee_{j=1}^{\infty} f^{-jk}(\mathcal{P})$ para todo $k \geq 1$. Deduza que, a menos de medida nula, \mathcal{P} está contida em $f^{-k}(\mathcal{B})$ para todo $k \geq 1$. Conclua que a partição \mathcal{P} é trivial.

9.6.1. A unicidade é imediata. Para provar a existência, considere o funcional Ψ definido por $\Psi(\psi) = \int (\int \psi d\eta) dW(\eta)$ no espaço das funções mensuráveis limitadas $\psi : M \rightarrow \mathbb{R}$. Note que Ψ é linear e não negativo e satisfaz $\Psi(1) = 1$. Use o teorema da convergência monótona para mostrar que se B_n , $n \geq 1$ são subconjuntos mensuráveis de M disjuntos dois-a-dois então $\Psi(\mathcal{X}_{\bigcup_n B_n}) = \sum_n \Psi(\mathcal{X}_{B_n})$. Conclua que $\xi(B) = \Psi(\mathcal{X}_B)$ define uma probabilidade na σ -álgebra dos subconjuntos mensuráveis de M . Mostre que $\int \psi d\xi = \Psi(\psi)$ para toda função mensurável limitada. Tome $\text{bar}(W) = \xi$.

9.6.2. Escreva $W = \sum_{i=1}^{\infty} t_i W_i$. Por definição,

$$\begin{aligned} \int \psi d\text{bar}(W) &= \int \left(\int \psi d\eta \right) dW(\eta) = \sum_{i=1}^{\infty} t_i \int \left(\int \psi d\eta \right) dW_i(\eta) \\ &= \sum_{i=1}^{\infty} t_i \int \psi d\text{bar}(W_i) = \int \psi d\left(\sum_{i=1}^{\infty} t_i \text{bar}(W_i) \right) \end{aligned}$$

para toda função mensurável limitada $\psi : M \rightarrow \mathbb{R}$. Logo, $\text{bar}(W) = \sum_{i=1}^{\infty} t_i \text{bar}(W_i)$.

9.6.3. Seja $\mu = \sum_{i=1}^{\infty} t_i \nu_i$. Para cada n , defina $s_n = \sum_{i=1}^n t_i$ e $\mu_n = \sum_{i=1}^n t_i / s_n \nu_i$ e $R_n = \sum_{i>n} t_i / (1 - s_n) \nu_i$. Então, $\mu = s_n \mu_n + (1 - s_n) R_n$. Como \mathcal{M} é convexo, μ_n está em \mathcal{M} para todo n . Mostre que $(\mu_n)_n$ converge para μ na topologia fraca*, para obter o primeiro enunciado. Em seguida, dado qualquer $\Phi = \{\phi_1, \dots, \phi_N\}$ e $\varepsilon > 0$, considere uma partição finita ou enumerável \mathcal{P} de \mathcal{M} tal que cada $P_n \in \mathcal{P}$ está contido na vizinhança $V(\nu_n, \Phi, \varepsilon)$ de algum $\nu_n \in \mathcal{M}$. Seja W_n a restrição normalizada de W a cada P_n . Então, $W = \sum_n W(P_n) W_n$ e $\text{bar}(W) = \sum_n W(P_n) \text{bar}(W_n)$ (pelo Exercício 9.6.2). Pelo Lema 9.6.7, $\text{bar}(W_n) \in V(\nu_n, \Phi, \varepsilon)$. Mostre que $\nu = \sum_n W(P_n) \nu_n$ está em \mathcal{M} e que $\text{bar}(W) \in V(\nu, \Phi, \varepsilon)$. Para concluir o segundo enunciado, faça $\varepsilon \rightarrow 0$.

9.6.4. Use a desigualdade de Jensen no item (1), o lema de Fatou no item (2) e o teorema de Kolmogorov-Sinai no item (3).

9.6.5. Para a penúltima igualdade seria necessário saber que a sequência $n^{-1} \log \mu_P(\mathcal{Q}^n(x))$ é dominada, por exemplo.

9.7.1. Sejam $\{U_k : k \geq 1\}$ e $\{V_l : l \geq 1\}$ duas coberturas. Todo subconjunto mensurável B de algum V_l pode ser escrito como união disjunta de conjuntos mensuráveis $A_k \subset U_k$, $k \geq 1$. Por σ -aditividade, $\eta(f(A_k)) = \int_{A_k} \xi d\eta$ para todo $k \geq 1$ implica $\eta(f(B)) = \int_B \xi d\eta$.

9.7.2. Seja (p_1, \dots, p_d) é o vetor de probabilidades associado a P . Para cada $x \in [0; a_0, a_1]$ com $P_{a_0, a_1} > 0$, tome $J_\mu \sigma(x) = p_{a_1} / (p_{a_0} P_{a_0, a_1})$. A união dos demais cilindros tem medida nula.

9.7.3. (a) A definição (9.7.2) dá que a fórmula vale para toda função característica $\psi = \mathcal{X}_{f(A)}$, onde A é um domínio de invertibilidade de f . Por linearidade, a igualdade se estende para funções simples. Use o teorema da convergência monótona para estendê-la para qualquer função mensurável não negativa. Use linearidade uma vez mais, para obter o caso geral. (b) Aplique o item (a) à função $\varphi = (\psi / J_\eta f) \circ (f|_A)^{-1}$.

9.7.4. Seja \mathcal{P} uma partição enumerável em domínios de invertibilidade. Imitando a prova de (9.7.8), mostre que

$$\int_P \hat{\psi} d\eta = \int_{f^{-1}(P)} \psi d\eta \quad \text{para todo } P \in \mathcal{P}, \quad \text{onde } \hat{\psi}(x) = \sum_{z \in f^{-1}(x)} \frac{\psi}{J_\eta f}(z)$$

Some sobre todo $P \in \mathcal{P}$, para obter a igualdade no enunciado.

9.7.5. Pelo Exercício 9.7.4, dada qualquer função mensurável limitada ψ ,

$$\int (\psi \circ f) d\eta = \int \psi(x) \left(\sum_{z \in f^{-1}(x)} \frac{1}{J_\eta f}(z) \right) d\eta(x).$$

Dedua a primeira parte do enunciado. Para a segunda parte, note que se η é invariante então $\eta(f(A)) = \eta(f^{-1}(f(A))) \geq \eta(A)$ para todo domínio de invertibilidade A .

9.7.6. Verifique que se f é localmente invertível então f^k também é. Se A é domínio de invertibilidade de f^k e $\psi : f^k(A) \rightarrow \mathbb{R}$ é mensurável e limitada,

$$m(f^k(A)) = \int_{f(A)} J_\eta f^{k-1} d\eta = \int_A (J_\eta f^{k-1} \circ f) J_\eta f d\eta.$$

Portanto, $J_\eta f^k = (J_\eta f^{k-1} \circ f) J_\eta f$ para todo k . A primeira parte do enunciado segue por indução. Quanto à segunda parte, se f é invertível e η é singular também em relação a f^{-1} , então $J_\eta f^{-1} = (1/J_\eta f) \circ f^{-1}$ em η -quase todo ponto.

9.7.7. Verifique que se A é domínio de invertibilidade para f então $B = \phi(A)$ é domínio de invertibilidade para g . Além disso, $\mu(f(A)) = \nu(g(B)) = \int_B J_\nu g d\nu = \int_A J_\nu g \circ \phi d\mu$. Use o fato de que o jacobiano é essencialmente único.

9.7.8. A parte 'se' do enunciado é fácil: podemos exibir a equivalência ergódica explicitamente. Suponha que os dois sistemas são ergodicamente equivalentes. O fato de que $k = l$ segue do Exercício 8.1.2. Para provar que p e q são permutações um do outro, use o fato de que o jacobiano é invariante por equivalência ergódica (Exercício 9.7.7), juntamente com as expressões dos jacobianos dadas pelo Exemplo 9.7.1.

10.1.1. Note que $N(\alpha) \leq N(\beta)$, uma vez que todo elemento de β está contido em algum elemento de α .

10.1.2. $H(f^{-1}(\beta)) \leq H(\beta)$. Se f é sobrejetiva, vale a igualdade $H(f^{-1}(\beta)) = H(\beta)$. Observe que se $\mathcal{A} = \{A_1, \dots, A_r\}$ é uma subcobertura aberta finita de α e $\mathcal{B} = \{B_1, \dots, B_s\}$ é subcobertura aberta finita de β , então $\mathcal{U} = \{A_n \cap B_m : 1 \leq n \leq r \text{ e } 1 \leq m \leq s\}$ é subcobertura aberta finita de $\alpha \vee \beta$, com no máximo nm elementos. Isso garante $N(\alpha \vee \beta) \leq N(\alpha)N(\beta)$. Para provar que $H(f^{-1}(\beta)) \leq H(\beta)$, observe que como f é contínua, dada \mathcal{B} subcobertura aberta finita de β , o conjunto $f^{-1}(\mathcal{B})$ é uma cobertura aberta com no máximo $\#\mathcal{B}$ elementos. Assim, $N(f^{-1}(\alpha)) \leq N(\alpha)$. Para a desigualdade reversa, basta observar que $f^{-1}(B) \neq \emptyset$, para cada $B \in \beta$.

10.1.3. Para mostrar que a entropia topológica de f com respeito à distância usual é $\log 2$, dado $\varepsilon > 0$ tome $k \in \mathbb{N}$ tal que $(1/2)^k < \varepsilon$. Verifique que o conjunto $E = \{m/2^{n+k} : m \in \mathbb{Z}\}$ é um conjunto (n, ε) -separado. Considerando agora a distância $d(x, y) = |\log x - \log y|$, mostre que f é uma isometria. Use o Exemplo 10.1.7 para concluir que a entropia de f é zero.

10.1.4. Para $i = 1, 2$ considere a bola dinâmica $B^i(a, n, \varepsilon) = \{x \in M : d_i(f^i(x), f^i(a)) < \varepsilon \text{ para } i = 0, \dots, n-1\}$. Da hipótese de que d_1 e d_2 são uniformemente equivalentes, conclua que para todo $\varepsilon > 0$ existe $\delta > 0$ tal que $B^2(a, n, \delta) \subset B^1(a, n, \varepsilon)$. Observe que isto garante que $g^1(f, \varepsilon, K) \leq g^2(f, \delta, K)$ para qualquer compacto K . Obtenha a desigualdade análoga $g^2(f, \varepsilon, K) \leq g^1(f, \delta, K)$ e conclua o exercício.

10.1.5. Seja $K \subset N$ um compacto e seja $K' = \psi(K)$. Observe que K' é compacto e a restrição $\psi : K \rightarrow K'$ é um homeomorfismo. Em particular, a sua inversa é uniformemente contínua. Use esse fato para mostrar que dado $\varepsilon > 0$, existe $\delta > 0$ tal que dado qualquer conjunto $E' \subset M$ que seja (n, δ) -gerador de K' , o conjunto $E = \psi^{-1}(E') \subset N$ é (n, ε) -gerador de K . Conclua que $g(g, \varepsilon, K) \leq g(f, \delta, \psi(K))$ e que $h(g) \leq h(f)$. Para a segunda parte,

considere um conjunto discreto $A \subset [0, 1]^{\mathbb{Z}}$ restrita a $A^{\mathbb{Z}}$ é uniformemente equivalente à distância definida em (9.2.13). Conclua, usando o Exemplo 10.1.2, que a entropia de σ é pelo menos $\log n$, para qualquer n .

10.1.6. Mimetize as provas do Lema 10.3.6 e do Lema 10.3.7.

10.1.7. Fixados $n \in \mathbb{N}$ e $\varepsilon > 0$, tome para cada $i = 1, 2, \dots, l$ um conjunto E_i que seja (n, ε) -gerador de K_i . Prove que $E = \bigcup_{1 \leq i \leq l} E_i$ é (n, ε) -gerador de K . Conclua que $g_n(f, \varepsilon, K) \leq \sum_{1 \leq i \leq l} g_n(f, \varepsilon, K_i)$. Deduza que $g(f, \varepsilon, K) \leq \max_{1 \leq i \leq l} g(f, \varepsilon, K_i)$. Conclua fazendo $\varepsilon \rightarrow 0$.

10.1.8. Considere $h(x) = \text{sen}(\pi x/2)$. Verifique que h é um homeomorfismo de $[-1, 1]$. Prove que $h(f(x)) = g(h(x))$, isto é, $\text{sen}(\pi/2(1 - 2x^2)) = 1 - 2|\text{sen}(\pi x/2)|$. Conclua que a entropia de f é $\log 2$.

10.1.9. Observe que cada palavra w de comprimento $n + m$ pode ser decomposta de modo único como $w = uv$, onde u é uma palavra de comprimento m e v é uma palavra de comprimento n . Isto define uma função injetiva do conjunto das palavras de comprimento $n + m$ no produto cartesiano do conjunto das palavras de comprimento m pelo conjunto das palavras de comprimento n . Segue que $c_{n+m}(x) \leq c_n(x)c_m(x)$ ou, em outras palavras, a sequência $\log c_n(x)$ é subaditiva. Conclua que $c(x)$ está bem-definido. Para mostrar que $c(x)$ coincide com a entropia topológica de σ restrita ao fecho de \mathcal{X} , mostre que $c_{n+k}(x)$ é igual a $g_n(f, (1/2)^k, \mathcal{X})$, observando que cada palavra de comprimento $n + k$ que aparece em x começando na posição i corresponde a uma bola dinâmica de comprimento n e raio $(1/2)^k$ centrada em $\sigma^i(x)$.

10.1.10. (Carlos Gustavo Moreira) Seja $\theta_1 = 0, \theta_2 = 01$ e, para $n \geq 2$, $\theta_{n+1} = \theta_n \theta_{n-1}$. Vamos mostrar que, para todo $n \geq 1$, existe uma palavra τ_n tal que $\theta_n \theta_{n+1} = \tau_n \alpha_n$ e $\theta_{n+1} \theta_n = \tau_n \beta_n$, onde $\alpha_n = 10$ e $\beta_n = 01$ se n é par e $\alpha_n = 01$ e $\beta_n = 10$ se n é ímpar. De fato, isso vale para $n = 1$ com $\tau_1 = 0$ e para $n = 2$ com $\tau_2 = 010$. Se vale para n , temos $\theta_{n+1} \theta_{n+2} = \theta_{n+1} \theta_{n+1} \theta_n = \theta_{n+1} \tau_n \beta_n$ e também $\theta_{n+2} \theta_{n+1} = \theta_{n+1} \theta_n \theta_{n+1} = \theta_{n+1} \tau_n \alpha_n$, desde que tomemos $\tau_{n+1} = \theta_{n+1} \tau_n$. Isto prova a nossa afirmação. Segue que as últimas letras de θ_n e de θ_{n+1} são distintas. Vamos deduzir que $\theta = \lim \theta_n$ não é pré-periódica. De fato, suponha que θ fosse periódica e seja m o tamanho de seu período. Como o tamanho de θ_n é F_{n+1} (onde F_k é o k -ésimo número de Fibonacci), podemos tomar n inteiro positivo tal que $m | F_{n+1}$ e tal que n é grande o suficiente para que o tamanho da eventual parte não-periódica de θ seja menor que F_{n+2} . Então θ começa com $\theta_{n+3} = \theta_{n+2} \theta_{n+1} = \theta_{n+1} \theta_n \theta_{n+1}$. Mas, como o tamanho F_{n+1} de θ_n é múltiplo do período m , a F_{m+2} -ésima letra de θ , que é a última letra de θ_{n+1} , deveria coincidir com a $(F_{m+2} + F_{n+1})$ -ésima letra de θ , que é a última letra de θ_n . Isto contradiria a conclusão do parágrafo anterior. Em seguida, afirmamos que $c_{k+1}(\theta) > c_k(\theta)$ para todo k . De fato, suponha que $c_{k+1}(\theta) = c_k(\theta)$ para algum k . Então cada subpalavra de tamanho k só pode ter uma continuação de tamanho $k + 1$. Logo, temos uma dinâmica no conjunto finito das subpalavras de tamanho k (que a cada subpalavra associa sua única continuação sem a primeira letra), cujas órbitas são todas pré-periódicas. Em particular, θ também é pré-periódica, o que contradiz a conclusão no parágrafo anterior. Isto prova a nossa afirmação. Como $c_1(\theta) = 2$, segue que $c_k(\theta) \geq k + 1$ para todo k . Agora afirmamos que $c_{F_{n+1}}(\theta) \leq F_{n+1} + 1$ para todo $n > 1$. Para provar esse fato, note que θ pode ser escrito como uma concatenação de palavras pertencentes a $\{\theta_n, \theta_{n+1}\}$, pois (por indução), para todo $r \geq n$, θ_r pode ser escrito como uma concatenação de palavras pertencentes a $\{\theta_n, \theta_{n+1}\}$. Assim, qualquer subpalavra de θ de tamanho F_{n+1} (que é o tamanho de θ_n) é uma subpalavra de $\theta_n \theta_{n+1}$ ou de $\theta_{n+1} \theta_n$. Como $\theta_n \theta_{n+1} = \theta_n \theta_n \theta_{n-1}$ é uma subpalavra de $\theta_n \theta_n \theta_{n-1} \theta_{n-2} = \theta_n \theta_n \theta_n$, há no máximo $|\theta_n| = F_{n+1}$ subpalavras de tamanho $|\theta_n| = F_{n+1}$ de $\theta_n \theta_n \theta_n$ e, logo, de $\theta_n \theta_{n+1}$. Como $\theta_n \theta_{n+1} = \tau_n \alpha_n$ e $\theta_{n+1} \theta_n = \tau_n \beta_n$ e $\theta_{n+1} \theta_n$ termina com θ_n e $|\beta_n| = 2$, a única subpalavra de $\theta_{n+1} \theta_n$ de tamanho $|\theta_n| = F_{n+1}$ que pode não ser subpalavra de $\theta_n \theta_{n+1}$ é a subpalavra que termina com a primeira letra de β_n (ou seja, uma posição antes do fim de $\theta_{n+1} \theta_n$). Logo, $c_{F_{n+1}}(\theta) \leq F_{n+1} + 1$ tal como foi afirmado. Estamos prontos para obter o enunciado do exercício. Suponha que $c_k(\theta) > k + 1$ para algum k . Tomando n tal que $F_{n+1} > k$, teríamos $c_{F_{n+1}}(\theta) - c_k(\theta) < F_{n+1} + 1 - (k + 1) = F_{n+1} - k$, donde, para algum m com $k \leq m < F_{n+1}$, deveríamos ter $c_{m+1}(\theta) \leq c_m(\theta)$. Isto contradiria fato provado anteriormente.

10.2.1. Considere as projeções canônicas $\pi_1 : M_1 \times M_2 \rightarrow M_1$ e $\pi_2 : M_1 \times M_2 \rightarrow M_2$. Dado qualquer compacto $K \subset M$, considere $K_i = \pi_i(K)$ para $i = 1, 2$. Note que $K \subset K_1 \times K_2$.

Mostre que se E_i é conjunto (n, ε) -gerador para f_i em K_i , com $i = 1, 2$, então $E = E_1 \times E_2$ é conjunto (n, ε) -gerador para f em K . Logo, $g_n(f, \varepsilon, K) \leq g_n(f_1, \varepsilon, K_1)g_n(f_2, \varepsilon, K_2)$. Deduza que $g(f, \varepsilon, K) \leq g(f_1, \varepsilon, K_1) + g(f_2, \varepsilon, K_2)$ para todo ε e conclua que $h(f) \leq h(f_1) + h(f_2)$. De modo semelhante, mostre que se $K_i \subset M_i$ é compacto e $E_i \subset K_i$ é conjunto (n, ε) -separado para f_i , com $i = 1, 2$, então $E_1 \times E_2 \subset K_1 \times K_2$ é conjunto (n, ε) -separado para f . Logo, $s_n(f, \varepsilon, K_1 \times K_2) \geq s_n(f_1, \varepsilon, K_1)s_n(f_2, \varepsilon, K_2)$ para todo ε . Supondo que M_2 é compacto, tome $K_2 = M_2$. Deduza que $s(f, \varepsilon, K_1 \times K_2) \geq s(f_1, \varepsilon, K_1) + \liminf_n \frac{1}{n} s_n(f_2, \varepsilon, M_2)$. Usando o Corolário 10.1.8, deduza que $h(f) \geq h(f_1) + h(f_2)$.

10.2.2. (a) Por hipótese, dada qualquer palavra finita A -admissível (x_0, \dots, x_k) , existe $n = n(x_0, x_k)$ tal que $A_{x_k, x_0}^n > 0$, ou seja, $(x_k, y_1, \dots, y_{n-1}, x_0)$ é A -admissível para alguma escolha de y_1, \dots, y_{n-1} . Logo, a seqüência periódica $(x_0, \dots, x_k, y_1, \dots, y_{n-1}, x_0, \dots)$ é admissível. Para os itens (b) e (c), use que $[0; i] \cap \sigma^{-k}([0; j]) \neq \emptyset$ se, e somente se, $A_{i,j}^k \geq 1$.

10.2.3. Pela Proposição 10.2.1, $h(f) = g(f, \delta, M)$ sempre que f é ε -expansiva e $\delta < \varepsilon/2$. Mostre que se $d(f, h) < \delta/3$ então $g(h, \delta/3, M) \leq g(f, \delta, M)$. Deduza que se $(f_k)_k$ converge para f então $\limsup_k h(f_k) = \limsup_k g(f_k, \delta/3, M) \leq g(f, \delta, M) = h(f)$.

10.2.4. Seja $S_i = A_{i,1} + A_{i,2} + \dots + A_{i,n}$ a soma das entradas da linha i da matriz A . Por hipótese $S_i > 0$ para todo i . Defina $P_{i,j} = A_{i,j}/S_i$. Note que P é uma matriz estocástica e que $P_{i,j}^n = 0$ se, e somente se, $A_{i,j}^n = 0$. Pelo Lema 7.2.7, temos que $\sum_{j=1}^d P_{i,j}^n = 1$ para todo i e todo n . Conclua que a linha i de A^n é não nula, para todo i .

10.2.5. (a) Por hipótese, existe $\varepsilon > 0$ tal que a restrição de f a toda bola de raio ε é injetiva. Mostre que $f^{-n}(y)$ é (n, δ) -separado para todo y , todo n e todo $\delta < \varepsilon$. A hipótese também garante que $\#f^{-n}(y) \geq d^n$. (b) Não é restrição supor que $d = |\deg f| \geq 2$. Mostre que existem intervalos I_j , $j = 1, \dots, d$ disjuntos dois-a-dois, ordenados ciclicamente e tais que $f(I_j) = S^1$ para todo j . Considere $\delta < \min_j |I_j|$. Para cada $y \in S^1$ e $n \geq 1$, seja $f_*^{-n}(y)$ o conjunto dos $x \in f^{-n}(y)$ tais que $f^i(x) \in J_2 \cup \dots \cup J_{2[d/2]}$ para $i = 0, \dots, n-1$. Mostre que $f_*^{-n}(y)$ é conjunto (n, δ) -separado e contém $[d/2]^n$ pontos. Deduza que $h(f) \geq \log[d/2]$. Usando a Proposição 10.1.11, conclua que $kh(f) \geq \log[d^k/2]$ para todo k . Fazendo $k \rightarrow \infty$, obtenha que $h(f) \geq \log d$.

10.2.6. Note que $|f(z)| = |z|/2$ e, portanto, $(f^n(z))_n \rightarrow 0$ para todo $z \neq \infty$. Logo, 0 e ∞ são os únicos pontos recorrentes. Deduza que toda probabilidade invariante tem entropia nula e conclua que $h(f) = 0$. Não há contradição, pois f não é homeomorfismo local (nem é diferenciável) na vizinhança de zero ou infinito.

10.2.7. (Bowen [Bow72]) Escreva $a = g_*(f, \varepsilon)$. Observe que se E é conjunto (n, δ) -gerador de M , com $\delta < \varepsilon$, então $M = \cup_{x \in E} B(x, n, \varepsilon)$. Combinando este fato com o resultado de Bowen, mostre que $g_n(f, \delta, M) \leq \#Ee^{c+(\alpha+b)n}$. Conclua a desigualdade, tomando $b \rightarrow 0$.

10.3.1. Dada γ uma subcobertura finita de α^{n+m} , observe que

$$\sum_{U \in \gamma} e^{\phi_{n+m}(U)} \leq \sum_{U \in \gamma} e^{\phi_n(U)} e^{\phi_m(f^n(U))} \leq \sum_{U \in \gamma} e^{\phi_n(U)} \sum_{V \in f^m(\gamma)} e^{\phi_m(V)}.$$

Conclua que $P_{n+m}(f, \phi, \alpha) \leq P_n(f, \phi, \alpha)P_m(f, \phi, \alpha)$.

10.3.2. Reflexiva: tome $u \equiv 0$. Simetria: troque u por $-u$. Transitividade: se $\psi = \phi + u \circ f - u$ e $\phi = \theta + v \circ f - v$ então $\psi = \theta + (u + v) \circ f - (u + v)$.

10.3.3. Estenda o argumento do Exercício 10.2.1 do seguinte modo. Seja $f = f_1 \times f_2$ e $\phi = \phi_1 \times \phi_2$. Lembrando a definição (10.3.9), comece por mostrar que $G_n(f, \phi, \varepsilon) \leq G_n(f_1, \phi_1, \varepsilon)G_n(f_2, \phi_2, \varepsilon)$. Deduza que $G(f_1 \times f_2, \phi_1 \times \phi_2, \varepsilon) \leq G(f_1, \phi_1, \varepsilon) + G(f_2, \phi_2, \varepsilon)$ para todo ε e conclua que $P(f, \phi) \leq P(f_1, \phi_1) + P(f_2, \phi_2)$. Analogamente, mostre que $S_n(f, \phi, \varepsilon) \geq S_n(f_1, \phi_1, \varepsilon)S_n(f_2, \phi_2, \varepsilon)$ para todo ε . Deduza que

$$S(f, \phi, \varepsilon) \geq S(f_1, \phi_1, \varepsilon) + \liminf_n \frac{1}{n} S_n(f_2, \phi_2, \varepsilon).$$

Usando (10.3.16), deduza que $P(f, \phi) \geq P(f_1, \phi_1) + P(f_2, \phi_2)$.

10.3.4. Se $f : M \rightarrow M$ é ε_0 -expansiva, $P(f, \phi) = P(f, \phi, \alpha) = G(f, \phi, \varepsilon) = S(f, \phi, \varepsilon)$ para toda cobertura aberta α com diâmetro menor que ε_0 e para todo $\varepsilon < \varepsilon_0/2$. Em particular,

$P(f, \phi) < \infty$. Para a demonstração, siga os passos da Proposição 10.2.1, substituindo a Proposição 10.1.9 pelo Lema 10.3.1 na prova da primeira igualdade. Ainda usando o roteiro da Proposição 10.1.9, mostre que $P_n(f, \phi, \alpha) \leq G_n(f, \phi, \varepsilon) \leq S_n(f, \phi, \varepsilon)$ para todo n e todo $\varepsilon > 0$, onde α é a cobertura de M pelas bolas de raio ε ; a partir deste fato, deduza as demais igualdades no enunciado.

10.3.5. Fixe uma cobertura aberta finita α com diâmetro menor que ε_0 . Usando as definições (10.3.3) e (10.3.2) e os Exercícios 10.3.4 e 10.3.1,

$$P(f, \phi) = P(f, \phi, \alpha) = \inf_n \frac{1}{n} \log P_n(f, \phi, \alpha) = \inf_n \inf_\gamma \frac{1}{n} \log \sum_{U \in \gamma} e^{\phi_n(U)}$$

(o segundo ínfimo é sobre todas as subcoberturas γ de α^n). Isto permanece válido para toda transformação ε_0 -expansiva g próxima de f , módulo substituírmos α^n e ϕ_n pelos objetos correspondentes, α_g^n e $\phi_{g,n}$, relativos a g . Mostre que dados n, γ e $\varepsilon > 0$, existe uma vizinhança \mathcal{V} de f tal que para cada $g \in \mathcal{V}$ existe uma subcobertura γ_g de α_g^n e existe uma bijeção $\iota : \gamma \rightarrow \gamma_g$ tal que $\phi_{g,n}(\iota(U)) \leq \phi_n(U) + n\varepsilon$ para todo $U \in \gamma$. Deduza que $P(g, \phi) \leq P(f, \phi) + 2\varepsilon$ para todo $g \in \mathcal{V}$ que seja ε_0 -expansivo.

10.3.6. Mostre que dado $\varepsilon > 0$ existe $\kappa \geq 1$ tal que toda bola dinâmica $B(x, n, \varepsilon)$ tem diâmetro igual a $\varepsilon 2^{-n}$ e contém algum ponto periódico p_x^n de período $n + \kappa$. Mostre que dados $C, \theta > 0$ existe $K > 0$ tal que $|\phi_n(y) - \phi_n(p_x^n)| \leq K$ para todo $y \in B(x, n, \varepsilon)$, todo $n \geq 1$ e toda função (C, θ) -Hölder $\phi : S^1 \rightarrow \mathbb{R}$. Use este fato para substituir conjuntos geradores (ou separados) por conjuntos de pontos periódicos na definição da pressão.

10.4.1. Consequência direta do Princípio Variacional.

10.4.2. Use o Princípio Variacional e o fato que toda medida invariante de $f|_X$ é também medida invariante de f .

10.4.3. Observe que $\int \phi d\mu = \int \psi d\mu$ para qualquer medida invariante μ . Em seguida, aplique o Princípio Variacional.

10.4.4. Considere o deslocamento σ no espaço $\Sigma = \{0, 1\}^{\mathbb{N}}$. Considere a função $\phi : \Sigma \rightarrow \mathbb{R}$ definida por $\phi(x) = 0$ se $x_0 = 0$ e $\phi(x) = 1$ se $x_0 = 1$. Seja N o conjunto dos pontos $x \in \Sigma$ tais que a média temporal na órbita de x não converge. Verifique que N é invariante por σ e é não vazio: para cada sequência finita (z_0, \dots, z_k) é possível encontrar $x \in N$ com $x_i = z_i$ para $i = 0, \dots, k$. Deduza que a entropia topológica da restrição $f|_N$ é igual a $\log 2$. Justifique que N não suporta nenhuma medida invariante por f .

10.4.5. Considere a cobertura aberta ξ de K cujos elementos são $K \cap [0, \alpha]$ e $K \cap [1 - \beta, 1]$. Verifique que $P(f, \phi) = P(f, \phi, \xi)$ para todo potencial ϕ . Além disso,

$$P_n(f, -t \log g', \xi) = \sum_{U \in \alpha^n} [(g^n)']^{-t}(U) = (\alpha^t + \beta^t)^n.$$

Conclua que $\psi(t) = \log(\alpha^t + \beta^t)$. Verifique que $\psi' < 0$ e $\psi'' > 0$ (convexidade também segue da Proposição 10.3.11). Além disso, $\psi(0) > 0 > \psi(1)$. Pelo Princípio Variacional, a última desigualdade também implica que $h_\mu(f) - \int \log g' d\mu < 0$.

10.4.6. Pelo teorema de Jacobs (Teorema 9.6.2), o funcional $\mathcal{M}_1(f) \rightarrow \mathbb{R}, \nu \mapsto h_\nu(f) + \int \varphi d\nu$ é afim. Segue que o seu supremo no domínio $\mathcal{M}_1(f)$ coincide com o supremo no conjunto $\mathcal{M}_e(f)$ dos pontos extremais. Como $\mathcal{M}_e(f)$ é finito, o supremo é alcançado.

10.5.1. Como vimos no Exemplo 9.1.10, $h_\mu(f) = \sum_{i=1}^d d - p_i \log p_i$. Use o Lema 10.4.4 com $a_1 = \dots = a_d = 0$ para mostrar que a entropia é máxima quando $p_i = 1/d$ para todo i .

10.5.2. Seja α a cobertura de Σ pelos cilindros $[0; i]$ e seja a_i o valor de ϕ em cada cilindro. Justifique que $P(f, \phi) = P(f, \phi, \alpha)$ e verifique que

$$P_n(f, \phi, \alpha) = \sum_{i_0, \dots, i_{n-1}} e^{a_{i_0} + \dots + a_{i_{n-1}}} = \left(\sum_i e^{a_i} \right)^n.$$

Conclua que $P(f, \phi) = \log(\sum_i e^{a_i})$. Use o Lema 10.4.4 para mostrar que existe um único vetor de probabilidade $p = (p_1, \dots, p_d)$ cuja medida de Bernoulli μ satisfaz $h_\mu(\sigma) + \int \phi d\mu = P(f, \phi)$.

10.5.3. A propriedade de Gibbs dá que $\lim_n(1/n) \log \mu(C^n(x)) = \tilde{\varphi}(x) - P$, onde $C^n(x)$ é o cilindro de comprimento n que contém x . Combine esta igualdade com o teorema de Brin-Katok (Teorema 9.3.3) e o teorema de Birkhoff para obter a primeira afirmação. Agora suponha que μ_1 e μ_2 sejam dois estados de Gibbs ergódicos com a mesma constante P . Observe que existe C tal que $C^{-1}\mu_1(A) \leq \mu_2(A) \leq C\mu_1(A)$ para todo A na álgebra das uniões finitas disjuntas de cilindros. Deduza, usando o teorema das classes monótonas (Teorema A.1.18), deduza que $C^{-1}\mu_1(A) \leq \mu_2(A) \leq C\mu_1(A)$ para qualquer conjunto mensurável A . Isto implica que μ_1 e μ_2 são equivalentes. Usando o Lema 4.3.1, segue que $\mu_1 = \mu_2$.

10.5.4. Pelo princípio variacional e a hipótese de que μ é um estado de equilíbrio para ϕ ,

$$F_\mu(\psi) = \int \psi d\mu = h_\mu(f) + \int (\psi + \phi) d\mu - (h_\mu(f) + \int \phi d\mu) \leq P(f, \phi + \psi) - P(f, \phi).$$

Isto mostra que F_μ é funcional linear tangente a ϕ . Logo, a segunda parte do enunciado segue da primeira e do Exercício 2.3.7.

10.5.5. Pela Proposição 10.3.11, a função pressão é convexa (pelo Exercício A.5.1, segue que ela também é contínua). Pelo teorema de suavidade de Mazur (lembre do Exercício 2.3.7), existe um subconjunto residual $\mathcal{R} \subset C^0(M)$ tal que a função pressão é derivável em cada $\varphi \in \mathcal{R}$. Aplique o Exercício 10.5.4.

10.5.6. Chamamos cilindro em \mathcal{H} a todo $[p; x_p, \dots, x_q]_{\mathcal{H}} = [p; x_p, \dots, x_q] \cap \mathcal{H}$ que seja *não vazio*. (a) Dados $A = [p; x_p, \dots, x_{p+q}]_{\mathcal{H}}$ e $B = [r; y_r, \dots, y_{r+s}]_{\mathcal{H}}$ considere

$$C = [p; x_p, \dots, x_{p+q}, 0, \dots, 0, y_r, \dots, y_{r+s}]_{\mathcal{H}},$$

onde inserimos $n > \max\{q, s\}$ zeros. Observe que C é não vazio e $C \subset A \cap f^{r-p-q-n-1}(B)$. Isto mostra que o deslocamento é topologicamente misturador em \mathcal{H} . (b) Seja α a cobertura aberta de \mathcal{H} por cilindros $[0; x]_{\mathcal{H}}$. Então $N(\alpha^n)$ é igual ao número de cilindros de comprimento n em \mathcal{H} . Estes são de 3 tipos: (i) Cilindros monocromáticos contidos em $V^{\mathbb{Z}}$ ou $A^{\mathbb{Z}}$. Em cada um dos dois casos, são 2^n cilindros. (ii) Cilindros monocromáticos da forma $[0; 0, \dots, 0, x_0, \dots, x_{k-1}]$ ou $[0; x_0, \dots, x_{k-1}, 0, \dots, 0]$ com $\{x_0, \dots, x_{k-1}\}$ contido em V ou em A . Verifique que existem menos de 2^{n+1} cilindros deste tipo. (iii) Cilindros bicromáticos $[0; x_0, \dots, x_{n-1}]$, envolvendo símbolos tanto verdes quanto amarelos. Usando que o número s de índices j tais que $x_j \neq 0$ não ultrapassa $n/2$, e que estes símbolos se agrupam em $2 \leq k \leq s$ blocos monocromáticos de cores alternadas, mostre que o número total de cilindros deste tipo não ultrapassa

$$2 \sum_{s=1}^{n/2} \sum_{k=2}^s 2^s \binom{s-1}{k-1} \leq 2 \sum_{s=1}^{n/2} 2^s (2^{s-1} - 1) \leq 2^{n+1}.$$

Isto prova que $2^{n+1} \leq N(\alpha^n) \leq 3 \cdot 2^{n+1}$ para todo n . Logo $h(\sigma) = h(\sigma, \alpha) = \log 2$.

11.1.1. Seja $\{A_i : i\}$ uma cobertura finita de M por bolas de raio $\rho > 0$ (Lema 11.1.3). Sejam $h_{i,j} : A_i \rightarrow M$ os ramos inversos de f e denote $V_{i,j} = h_{i,j}(A_i)$. (1) Escreva B como união de conjuntos mensuráveis $B_i \subset A_i$. Então $f^{-1}(B) = \cup_i f^{-1}(B_i) = \cup_{i,j} h_{i,j}(B_i)$. Além disso, $m(h_{i,j}(B_i)) = \int_{B_i} |\det Dh_{i,j}(x)| dm(x) = 0$ para todo i e j . Logo $m(f^{-1}(B)) = 0$. (2) e (3) Escreva B como união de conjuntos mensuráveis $B_{i,j} \subset V_{i,j}$. Então $f(B) = \cup_{i,j} f(B_{i,j}) = \cup_{i,j} h_{i,k}^{-1}(B_{i,j})$ é mensurável. Tal como no item (1), o teorema de mudança de variáveis implica que $m(f(B)) = 0$. (4) É consequência simples de (1) e (3). (5) Se $f^{-1}(A) = A$ a menos de medida nula então, por (4), $f(A) = f(f^{-1}(A))$ a menos de medida nula. Todo difeomorfismo local numa variedade compacta (conexa) é sobrejetivo; logo, $f(f^{-1}(A)) = A$.

11.1.2. Considere $\theta = \sigma^{1/k}$ e $\langle v \rangle^2 = \sum_{j=0}^{k-1} \theta^{-2j} \|Df^j(x)v\|^2$. A norma $\langle \cdot \rangle$ provém de uma métrica Riemanniana (produto interno) e é equivalente à norma dada $\|\cdot\|$. Ela também satisfaz a conclusão do exercício.

11.1.3. Adapte os argumentos que usamos para a transformação de Gauss nas Seção 9.4.2, do seguinte modo. Comece por verificar que os iterados de f têm distorção limitada: existe $K > 1$ tal que

$$\frac{1}{K} \leq \frac{|Df^n(x)|}{|Df^n(y)|} \leq K,$$

para todo $n \geq 1$ e quaisquer pontos x, y com $\mathcal{P}^n(x) = \mathcal{P}^n(y)$. Considere a sequência das médias dos iterados da medida de Lebesgue m , isto é, $\mu_n = (1/n) \sum_{j=0}^{n-1} f_j^* m$. Mostre que as derivadas de Radon-Nikodym $d\mu_n/dm$ são uniformemente limitadas e são Hölder, com constantes de Hölder uniformes. Deduza que todo ponto de acumulação μ desta sequência é uma probabilidade invariante absolutamente contínua com relação à medida de Lebesgue. Mostre que a derivada de Radon-Nikodym $\rho = d\mu/dm$ é limitada de zero e de infinito (em outras palavras, $\log \rho$ é limitado). Mostre que ρ e $\log \rho$ são Hölder.

11.1.4. Seja B um conjunto com $\mu(B) > 0$ tal que, para cada $n \geq 1$, existe B algum conjunto mensurável B_n satisfazendo $B = f^{-n}(B_n)$. Note que $\mu(B_n) = \mu(B)$ para todo n . Para cada n , considere a partição \mathcal{P}^n formada pelas imagens $h^n(P)$ dos conjuntos $P \in \mathcal{P}$ pelos ramos inversos h^n do iterado f^n . Mostre que para todo $\varepsilon > 0$ e todo $n \geq 1$ suficientemente grande existe algum $h^n(P_n) \in \mathcal{P}^n$ tal que $\mu(h^n(P_n) \setminus B) \leq \varepsilon \mu(h^n(P_n))$. Usando que a distorção é limitada e $f(P_n) = (0, 1)$, deduza que $1 - \mu(B) = \mu((0, 1) \setminus B_{n+1}) \leq \text{const} \varepsilon$. Fazendo $\varepsilon \rightarrow 0$, conclua que μ é exata. Se ν é probabilidade invariante absolutamente contínua então $\nu \ll \mu$ e, como μ é ergódica, isso implica $\mu = \nu$.

11.1.5. Verifique que $J_\mu f = (\rho \circ f)|f'|/\rho$ e use a fórmula de Rokhlin (Teorema 9.7.3).

11.1.6. Os argumentos do Exercício 11.1.3 para provar distorção limitada e existência de probabilidades invariantes absolutamente contínuas permanecem válidos: a hipótese implica que todo $f^n(\mathcal{P}^n(x))$ tem medida de Lebesgue maior que δ e isto basta para concluir que as densidades $d\mu_n/dm$ são uniformemente limitadas. Os argumentos do Exercício 11.1.4 mostram que se B é um conjunto mensurável invariante com $m(B) > 0$ então $m(B) \geq \delta$. Deduza que existem não mais que $1/\delta$ conjuntos invariantes com medida positiva minimais (isto é, que não podem ser decompostos em subconjuntos invariantes com medida de Lebesgue positiva) e eles são disjuntos dois-a-dois. Conclua que toda probabilidade invariante absolutamente contínua é combinação linear convexa de suas restrições normalizadas a esses conjuntos invariantes minimais, as quais são ergódicas. A prova de que $\mu(B) \geq \delta$ dá um pouco mais: existe algum intervalo aberto I com comprimento δ tal que $I \setminus B$ tem medida de Lebesgue nula. Suponha que f é transitivo. Dados dois conjuntos invariantes minimais com medida positiva, B_1 e B_2 , considere intervalos abertos I_1 e I_2 nessas condições. Por transitividade, $f^n(I_1)$ intersecta I_2 para algum n . Deduza que $m(B_1 \cap B_2) > 0$ e, portanto, $B_1 = B_2$. Usando as observações anteriores, conclua que a medida invariante absolutamente contínua é única.

11.2.1. Idêntico ao Lema 11.1.3.

11.2.2. Use o lema de combreamento (Proposição 11.2.9) para mostrar que se $f : M \rightarrow M$ é topologicamente misturadora então o conjunto dos pontos periódicos é denso. Isso permite usar o Teorema 11.2.15. Como f é topologicamente misturadora e M_1 é aberto invariante, temos que $M_1 = M$, ou seja $k = 1$. De modo semelhante, mostre que para que f seja topologicamente misturadora é necessário que $m(1) = 1$. Então, pelo Teorema 11.2.15, a transformação $f : M \rightarrow M$ é topologicamente exata.

11.2.3. Por hipótese, para cada $x \in M$ existe $N(x)$ tal que $f^{N(x)}(B(x, r/3)) = M$. Considere uma cobertura finita $B(x_i, r/3)$, $i = 1, \dots, n$ de M por bolas de raio $r/3$ e tome $N = \max_i N(x_i)$. Basta observar que toda bola $B(x, r)$ contém alguma $B(x_i, r/3)$.

11.2.4. Tome $\Lambda = \{2^{-n} : n \geq 0\} \bmod \mathbb{Z}$. Para ver que a restrição $f : \Lambda \rightarrow \Lambda$ não pode ser uma transformação expansora, basta observar que $1/2$ é ponto isolado em Λ mas $1 = f(1/2)$ não é. [Observação: Note que $\Lambda = S^1 \setminus \bigcup_{n=0}^{\infty} f^{-n}(I)$, onde $I = (1/2, 1) \bmod \mathbb{Z}$. Modificando adequadamente a escolha de I obtemos muitos outros exemplos, inclusive com Λ não enumerável.]

11.2.5. Note que $M = \bigcup_{k=0}^{\infty} f^{-k}(\Lambda) = \Lambda \cup \bigcup_{k=1}^{\infty} (f^{-k}(\Lambda) \setminus f^{-k+1}(\Lambda))$, pelo Teorema 11.2.14. Para qualquer $X \subset (f^{-k}(\Lambda) \setminus f^{-k+1}(\Lambda))$, as pré-imagens $f^{-n}(X)$, $n \geq 1$ são disjuntas duas-a-duas. Deduza que $\mu(\Lambda) = 1$ para toda probabilidade invariante e use o princípio variacional (Teorema 10.1) para concluir.

11.2.6. Toda probabilidade invariante ergódica μ satisfaz $\mu(M_i) = 1$ para algum i . Use esse fato e o princípio variacional (Corolário 10.4.2) para concluir que $h(f) = \max_i h(f|_{M_i})$. Se μ é uma probabilidade em M_i invariante por $f|_{M_i}$, a sua restrição normalizada μ_j ao subconjunto $M_{i,j}$ é invariante por $f^{m(i)}$. De fato, $\mu \mapsto \mu_j$ é uma bijeção do espaço das

probabilidades em M_i invariantes por $f | M_i$ no espaço das probabilidades em $M_{i,j}$ invariantes por $f^{m(i)} | M_{i,j}$. Usando a Proposição 10.1.11, mostre que

$$h_{\mu_j}(f^{m(i)} | M_{i,j}) = h_{\mu}(f^{m(i)} | M_i) = m(i)h_{\mu}(f | M_i).$$

Use o princípio variacional (Teorema 10.1) para obter $h(f^{m(i)} | M_{i,j}) = m(i)h(f | M_i)$.

11.2.7. Lembre do Exercício 10.2.2. Dados i, j no espaço dos símbolos $X = \{1, \dots, d\}$, defina $i \sim j \Leftrightarrow$ existe n tal que $A_{i,j}^n > 0$. A partição M_1, \dots, M_k em conjuntos invariantes transitivos (Teorema 11.2.15) corresponde à decomposição de X nas classes de equivalência X_1, \dots, X_k da relação \sim : a restrição da matriz A a cada X_i é irreduzível. Em seguida, a partição de cada M_i em subconjuntos $M_{i,j}$ (Teorema 11.2.15) corresponde à decomposição de X_i dada no Exercício 7.2.6: existe $\kappa = \kappa(i)$ tal que a restrição de A^κ a cada um desses subconjuntos de X_i é aperiódica.

11.3.1. Como f é uniformemente contínua, dado $\varepsilon > 0$ existe $\varepsilon' > 0$ tal que $d(x, y) < \varepsilon'$ implica $d(f^i(x), f^i(y)) < \varepsilon$ para todo $i = 0, \dots, l-1$.

11.3.2. Sejam U_1 e U_2 abertos não-vazios de M . Considere $x_1 \in U_1$ e $x_2 \in U_2$. Seja $\delta > 0$ de modo que $B(x_i, \delta) \subset U_i$ para $i = 1, 2$. Pela propriedade de especificação, existe $\kappa \geq 1$ tal que para todo $n > \kappa$ existe p tal que $d(p, x_1) < \delta$ e $d(f^n(p), x_2) < \delta$. Logo, $f^{-n}(U_2) \cap U_1 \neq \emptyset$.

11.3.3. Sejam $a = \int \varphi d\mu_1$ e $b = \int \varphi d\mu_2$. Suponha que $a < b$ e escreva $r = (b - a)/5$. Pelo teorema de decomposição ergódica, podemos supor que μ_1 e μ_2 são ergódicas. Então existem x_1 e x_2 tais que $\varphi(x_1) = a$ e $\varphi(x_2) = b$. Usando a hipótese de que f é topologicamente exata, construa uma pseudo-órbita $(z_n)_{n \geq 0}$ alternando segmentos (longos) das órbitas de x_1 e x_2 , de tal modo que a seqüência das médias temporais de φ ao longo da pseudo-órbita $(z_n)_n$ oscile de $a + r$ a $b - r$ (isto é, $\liminf \leq a + r$ e $\limsup \geq b - r$). Em seguida, use o lema de sombreamento para encontrar $x \in M$ cuja órbita sombreia essa pseudo-órbita. Usando que φ é uniformemente contínua, conclua que a seqüência das médias temporais de φ ao longo da órbita de x oscila de $a + 2r$ a $b - 2r$.

11.3.4. Use a propriedade de contração uniforme nos ramos inversos para mostrar que, dado $\varepsilon > 0$ existe $K \geq 1$ tal que $e^{\phi_n(y_1)} \leq Ke^{\phi_n(y_2)}$ para quaisquer $y_1, y_2 \in B(x, n, \varepsilon)$, todo $x \in M$ e todo $n \geq 1$. Combinando esta observação com os argumentos na prova da Proposição 10.2.2, mostre que $\sum_{p \in \text{Fix}(f^k)} e^{\phi_k(p)} \leq KP_k(f, \phi, \alpha)$ para toda cobertura α com diâmetro menor que a constante de expansividade. Deduza que $\limsup_k (1/k) \log \sum_{p \in \text{Fix}(f^k)} e^{\phi_k(p)} \leq P(f, \phi)$. De modo semelhante, combinando a observação acima com os argumentos na prova da Proposição 11.3.2, mostre que

$$S_n(f, \phi, \varepsilon) \leq \sum_{p \in \text{Fix}(f^{n+\kappa})} Ke^{\phi_n(p)} \leq Ke^{\kappa \sup |\phi|} \sum_{p \in \text{Fix}(f^{n+\kappa})} e^{\phi_{n+\kappa}(p)}.$$

Usando (10.3.16), conclua que $P(f, \phi) \leq \liminf_k (1/k) \sum_{p \in \text{Fix}(f^k)} e^{\phi_k(p)}$.

11.3.5. (a) Fixe $\varepsilon > 0$ tal que 2ε seja uma constante de expansividade para f e então tome $\delta > 0$ dado pelo lema de sombreamento (Proposição 11.2.9). Tome para \mathcal{U}_0 a δ -vizinhança de f . Seja $x_n = g^n(x)$, $n \in \mathbb{Z}$ a órbita por $g \in \mathcal{U}_0$ de um ponto $x \in M$. Então $d(f(x_n), x_{n+1}) = d(f(x_n), g(x_n)) < \delta$ para todo n , ou seja, $(x_n)_{n \in \mathbb{Z}}$ é uma δ -pseudo-órbita para f . Seja y o único ponto que ε -sombreia esta pseudo-órbita. Defina $\phi(x) = y$. Usando a unicidade do ponto sombreado, mostre que $\phi(g(x)) = f(\phi(x))$ para todo $x \in M$. Para verificar que ϕ é contínua, lembre (Proposição 11.2.9) que o ponto sombreador foi obtido como $y = \lim h^n(x_n)$, para uma escolha conveniente de ramos inversos h^n . É claro da construção que cada $x \mapsto h^n(x_n) = h^n(g^n(x))$ é contínua e também que a convergência é uniforme: $d(h^n(x_n), y) \leq \sigma^{-n}\varepsilon$ para todo x e todo n . Portanto, ϕ é realmente contínua. Além disso, ϕ está uniformemente próxima da identidade: $d(\phi(x), x) < \varepsilon$ para todo x . Estes fatos implicam que ϕ é sobrejetiva (a prova usa métodos de topologia algébrica e depende do fato de que M é uma variedade compacta) e, portanto, f é um fator topológico de g . Pela Proposição 10.1.3, segue que $h(f) \leq h(g)$. (b) Tome $\mathcal{U}_1 \subset \mathcal{U}_0$ suficientemente pequena para garantir que toda $g \in \mathcal{U}_1$ é expansora, com constante de expansão σ uniforme em \mathcal{U}_1 . Então, invertendo os papéis de f e g na construção do item (a), obtemos uma aplicação contínua $\psi : M \rightarrow M$ tal que a g -órbita de $\psi(x)$ sombreia

a f -órbita de x e $\psi(f(x)) = g(\psi(x))$ para todo x . Usando mais uma vez a unicidade do ponto sombreador, conclua que $\phi \circ \psi = \psi \circ \phi = \text{id}$, logo h é um homeomorfismo.

12.1.1. Suponha que existe uma sequência $(y_n)_n$ no suporte de η tal que $\eta(B(y_n, \varepsilon)) \rightarrow 0$. Por compacidade, podemos supor que $(y_n)_n$ convergindo para algum y . Como $B(y, \varepsilon/2) \subset B(y_n, \varepsilon)$ para todo n suficientemente grande, segue que $\eta(B(y, \varepsilon/2)) = 0$, contradizendo o fato que $y \in \text{supp } \eta$.

12.1.2. É imediato da definição que $\mathcal{M}_1(M)$ é convexo e o Teorema 2.1.5 dá que $\mathcal{M}_1(M)$ é compacto, para a topologia fraca*. Verifique que o operador $\mathcal{L} : C^0(M) \rightarrow C^0(M)$ é contínuo e deduza que o seu dual $\mathcal{L}^* : \mathcal{M}(M) \rightarrow \mathcal{M}(M)$ também é contínuo. Se $(\eta_n)_n \rightarrow \eta$ na topologia fraca* então $(\int \mathcal{L}1 d\eta_n)_n \rightarrow \int \mathcal{L}1 d\eta$. Conclua que o operador $G : \mathcal{M}_1(M) \rightarrow \mathcal{M}_1(M)$ é contínuo. Logo, pelo teorema de Tychonoff-Schauder, G possui algum ponto fixo ν . Isto quer dizer que $\mathcal{L}^*\nu = \lambda\nu$, onde $\lambda = \int \mathcal{L}1 d\nu$. Como $\lambda > 0$, isto prova que ν é uma medida de referência. Usando o Corolário 12.1.9, verifique que $\lambda = \limsup_n \sqrt[n]{\|\mathcal{L}^n 1\|}$ e deduza que λ é o raio espectral de \mathcal{L} .

12.1.3. Verifique que $\mu(\sigma(C)) = \mu(C)p_{i_1}/(p_{i_0}P_{i_0, i_1})$ para todo cilindro $C \subset [0; i_0, i_1]$. Logo, $J_{\mu\sigma} = p_{i_1}/(p_{i_0}P_{i_0, i_1})$ (isto corresponde ao Exercício 9.7.2). Represente por σ_i^{-1} o ramo inverso de σ com valores em $[0; i]$. Então,

$$\int \varphi d\mathcal{L}^*\mu = \int \mathcal{L}\varphi d\mu = \int \sum_i \frac{\varphi}{J_{\mu\sigma}} \circ \sigma_i^{-1} d\mu = \sum_i \int_{[0; i]} \varphi d\mu = \int \varphi d\mu$$

para toda função contínua φ (a penúltima igualdade usa o Exercício 9.7.3).

12.1.4. Dada uma função contínua $v : M \rightarrow \mathbb{R}$ qualquer, considere os funcionais lineares T_v e L_v definidos em L^1 por $T_v(u) = \int u(\lambda^{-1}\mathcal{L}v) d\nu$ e $L_v(u) = \int (u \circ f)v d\nu$. Verifique que T_v e L_v são contínuos. Observe que se $u \in C^0(M)$ então $\mathcal{L}((u \circ f)v) = u\mathcal{L}v$ e, portanto,

$$\lambda T_v(u) = \int u\mathcal{L}v d\nu = \int \mathcal{L}((u \circ f)v) d\nu = \int (u \circ f)v d\mathcal{L}^*\nu = \lambda \int (u \circ f)v d\nu = \lambda L_v(u).$$

Use a continuidade de T_v e L_v para concluir que a igualdade $T_v(u) = L_v(u)$ se estende para todo $u \in L^1(\nu)$.

12.2.1. Fixe uma sequência $(a_i)_{i < 0}$. Para cada $x = (x_i)_{i \in \mathbb{Z}}$ defina $r(x) = (r_i)_{i \in \mathbb{Z}}$ por $r_i = a_i$ se $i < 0$ e $r_i = x_i$ se $i \geq 0$. Dada uma função Hölder φ , considere $u : \Sigma \rightarrow \mathbb{R}$ dada por

$$u(x) = \sum_{i=0}^{\infty} \varphi(\sigma^i(x)) - \varphi(\sigma^i(r(x))).$$

Note que $|\varphi(\sigma^i(x)) - \varphi(\sigma^i(r(x)))| \leq Cd(\sigma^i(x), \sigma^i(r(x)))^\alpha \leq C(1/2)^{\alpha i}$, uma vez que $\sigma^i(x)$ e $\sigma^i(r(x))$ coincidem em cada posição $n \geq -i$. Deduza que u está bem definida e é Hölder. Seja $\varphi^+ = \varphi + u \circ \sigma - u$. Como φ e u são Hölder, segue que φ^+ é Hölder. Verifique que φ^+ satisfaz a condição no enunciado do exercício.

12.2.2. Segue da hipótese e da definição (10.3.9) que $|G_n(f, \varphi, \varepsilon) - G_n(f, \psi, \varepsilon) - C| < L$ para todo n e todo ε . Logo, $G(f, \varphi, \varepsilon) = G(f, \psi, \varepsilon) + C$ para todo ε e $P(f, \varphi) = P(f, \psi) + C$. A hipótese também implica que $\tilde{\varphi}(x) = \psi(x)$ para η -quase todo ponto e qualquer probabilidade invariante η . Consequentemente, $\int \varphi d\eta = \int \tilde{\varphi} d\eta = \int \tilde{\psi} d\eta + C = \int \psi d\eta + C$. Sejam μ_φ e μ_ψ os estados de equilíbrio para φ e ψ , respectivamente. Deduza das observações anteriores que μ_φ também é estado de equilíbrio para ψ . Então, por unicidade, $\mu_\varphi = \mu_\psi$. Use o Teorema 12.2.3 para obter que φ é cohomólogo a $\psi + c$ para algum $c \in \mathbb{R}$ e mostre que $c = C$.

12.2.3. Fixe em S^1 a orientação induzida por \mathbb{R} . Considere o ponto fixo $p_0 = 0$ de f e sejam p_1, \dots, p_d as suas pré-imagens, ordenadas ciclicamente, com $p_d = p_0$. Analogamente, seja q_0 um ponto fixo de g e sejam q_1, \dots, q_d as suas pré-imagens, ordenadas ciclicamente, com $q_d = q_0$. Note que f envia cada $[p_{i-1}, p_i]$ e g envia cada $[q_{i-1}, q_i]$ sobre S^1 . Então, para cada sequência $(i_n)_n \in \{1, \dots, d\}^{\mathbb{N}}$ existe exatamente um ponto $x \in S^1$ e um ponto $y \in S^1$ tais que $f^n(x) \in [p, p]$ e $g^n(y) \in [q, q]$ para todo n . Claro que as aplicações $(i_n)_n \mapsto x$ e $(i_n)_n \mapsto y$ são sobrejetivas. Considere que duas sequências $(i_n)_n$ e $(j_n)_n$ são equivalentes se existe $N \in \mathbb{N} \cup \{\infty\}$ tal que (1) $i_n = j_n$ para $n \leq N$ e (2a) $i_n = 1$ e $j_n = d$ para todo

$n > N$ ou (2b) $i_n = d$ e $j_n = 1$ para todo $n > N$. Mostre que os pontos x correspondentes a $(i_n)_n$ e $(j_n)_n$ coincidem se, e somente se, as duas seqüências são equivalentes e que vale fato análogo para os pontos y correspondentes às duas seqüências. Conclua que a aplicação $\phi : x \mapsto y$ está bem definida e é uma bjeção em S^1 tal que $\phi(f(x)) = g(\phi(x))$ para todo x . Observe que ϕ é monótona crescente (isto é, ela preserva a orientação de S^1) e, portanto, é um homeomorfismo.

12.2.4. (a) \Rightarrow (b): Trivial. (b) \Rightarrow (c): Seja μ_a a medida invariante absolutamente contínua e μ_m a medida invariante de máxima entropia de f ; sejam ν_a e ν_m as medidas correspondentes para g . Mostre que $\mu_a = \mu_m$. Seja $\phi : S^1 \rightarrow S^1$ uma conjugação topológica. Mostre que $\nu_m = \phi_*\mu_m$ e que se ϕ é absolutamente contínua então $\nu_a = \phi_*\mu_a$. Use o Corolário 12.2.4 para concluir que nesse caso $|(g^n)'(x)| = k^n$ para todo $x \in \text{Fix}(f^n)$. (c) \Rightarrow (a): A hipótese implica que $\nu_a = \nu_m$ e, portanto, $\nu_a = \phi_*\mu_a$. Lembre (Proposição 12.1.20) que as densidades $d\mu_a/dm$ e $d\nu_a/dm$ são contínuas e limitadas de zero e infinito. Conclua que ϕ é diferenciável, com $\phi' = (d\mu/dm)/(d\nu/dm) \circ \phi$.

12.3.1. A afirmação pode ser provada diretamente, substituindo a expressão de ϕ na definição da razão cruzada. Segue outra demonstração, baseada nas ideias da Seção 12.3.1. É imediato que a razão cruzada é invariante por translações e por homotetias. Logo, não é restrição supor que $a = -1 = \phi(a)$ e $d = 1 = \phi(d)$. Em outras palavras (lembre do Exemplo 12.3.4), basta provar que se ϕ preserva o intervalo I então ele também preserva a distância induzida em I pela distância projetiva do cone C . Verifique que todo automorfismo de Möbius $\phi(x) = (\alpha x + \beta)/(\gamma x + \delta)$ é induzido por algum isomorfismo linear $L : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, a saber

$$L = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix},$$

no sentido de que o vetor $(\phi(x), 1)$ é colinear a $L(x, 1)$ para todo x . Além disso, se ϕ preserva o intervalo I então L preserva o cone C . Então, pela Proposição 12.3.6 aplicada a L e à sua inversa, L preserva a distância projetiva de C . Isto significa que ϕ preserva a distância induzida d .

12.3.2. Considere $A = (a, 1)$, $P = (p, 1)$, $Q = (q, 1)$, $B = (b, 1)$, $O = (0, 0) \in \mathbb{C} \times \mathbb{R}$. Seja A' (respectivamente, B') o ponto onde a reta paralela a OQ (respectivamente, OP) que passa por P (respectivamente, Q) intersecta o bordo de C . Note que todos estes pontos pertencem ao plano determinado por P , Q e O , e que $A' \in OA$ e $B' \in OB$. Por definição, $\alpha(P, Q) = |B'Q|/|OP|$ e $\beta(P, Q) = |OQ|/|A'P|$. Mostre, usando semelhança de triângulos, que $|AP|/|AQ| = |A'P|/|OQ|$ e $|BQ|/|BP| = |B'Q|/|OP|$. Logo,

$$\theta(P, Q) = \log \frac{\beta(P, Q)}{\alpha(P, Q)} = \log \frac{|OQ||OP|}{|A'P||B'Q|} = \log \frac{|AQ||BP|}{|AP||BQ|}.$$

Em outras palavras, $d(p, q) = \log(|aq| |bp|)/(|ap| |bq|) = \Delta(p, q)$, para quaisquer $p, q \in \mathbb{D}$.

12.3.3. Seja $(g_n)_n$ uma seqüência de Cauchy com respeito à distância θ :

(A) dado qualquer $\varepsilon > 0$, existe $N \geq 1$ tal que $\theta(g_n, g_m) = \log \sup(g_n/g_m)/\inf(g_n/g_m) \leq \varepsilon$ para todo $m, n \geq N$.

Desprezando uma quantidade finita de termos, se necessário, podemos supor que $\theta(g_n, g_m) \leq 1$ para todo $m, n \geq 1$. Em particular, $1/e \leq (g_n(x)g_1(y))/(g_n(y)g_1(x)) \leq e$ para todo $x, y \in X$ and $n \geq 1$. Consequentemente,

(B) $1/R \leq g_n(x)/g_n(y) \leq R$ para todo $x, y \in X$ e $n \geq 1$,

onde $R = e \sup\{g_1(s)/g_1(t) : s, t \in X\}$. Seja ν uma probabilidade em X . Substituindo cada g_n por um múltiplo $t_n g_n$ com $t_n > 0$, se necessário, podemos supor que $\int g_n d\nu = 1$ para todo $n \geq 1$. Isto implica que

(C) $\inf g_n \leq 1 \leq \sup g_n$ e $\inf g_n/g_m \leq 1 \leq \sup g_n/g_m$ para todo $n, m \geq 1$.

Então, por um lado, (B) implica que $1/R \leq g_n(x) \leq R$ para todo $x \in X$ e $n \geq 1$. Por outro lado, (A) e a segunda parte de (C) nos dão que

(D) $e^{-\varepsilon} \leq \inf g_n/g_m \leq 1 \leq \sup g_n/g_m \leq e^\varepsilon$ para todo $m, n \geq N$.

Assim, $\sup |g_m - g_n| \leq \sup g_m \sup |g_n/g_m - 1| \leq R(e^\varepsilon - 1)$. Isto mostra que $(g_n)_n$ é uma sequência de Cauchy com respeito à norma do supremo e, portanto, é uniformemente convergente. Seja g_0 o limite. Então $R^{-1} \leq g_0(x) \leq R$ para todo $x \in X$; logo, $g_0 \in C$. Passando o limite quando $m \rightarrow \infty$ em (D), obtemos que $e^{-\varepsilon} \leq \inf g_n/g_0 \leq 1 \leq \sup g_n/g_0 \leq e^\varepsilon$ para todo $n \geq N$. Isto prova que tanto $\sup(g_n/g_0)$ quanto $\inf(g_n/g_0)$ convergem para 1. Segue que $\theta(g_n, g_0)$ converge para zero quando $n \rightarrow \infty$. Em outras palavras, $(g_n)_n$ converge para g_0 com respeito à distância θ .

12.3.4. Considere o cone C_0 das funções contínuas positivas em M . A respectiva distância projetiva θ_0 está dada no Exemplo 12.3.5. Verifique que θ_1 é a restrição de θ_0 ao cone C_1 . Considere uma sequência de funções positivas diferenciáveis convergindo uniformemente para uma função g_0 (contínua mas) não diferenciável. Mostre que $(g_n)_n$ converge para g_0 relativamente à distância θ_0 e, portanto, é sequência de Cauchy para θ_0 e θ_1 . Argumente que $(g_n)_n$ não pode ser convergente para θ_1 .

12.3.5. (a) Fixe $H_1 > H_\beta(g_1)$ e $H_2 > H_\beta(g_2)$. Dados quaisquer $x, y \in M$,

$$\begin{aligned} |g_1 g_2(x) - g_1 g_2(y)| &\leq |g_1(x) - g_1(y)| |g_2(x)| + |g_1(y)| |g_2(x) - g_2(y)| \\ &\leq H_1 d(x, y)^\beta \sup |g_2| + \sup |g_1| H_2 d(x, y)^\beta. \end{aligned}$$

Agora faça $H_i \rightarrow H_\beta(g_i)$, para cada $i = 1, 2$. (b) Como η é uma probabilidade, existe $x_\eta \in M$ tal que $|g_1(x_\eta)| \leq \int |g_1| d\eta$. Note também que $H_1 > H_\beta(g_1) \geq H_\beta(|g_1|)$. Então, para todo $x \in M$,

$$|g_1(x)| \leq |g_1(x_\eta)| + H_1 d(x, x_\eta)^\beta \leq \int |g_1| d\eta + H_1 (\text{diam } M)^\beta.$$

Isto mostra que $\sup |g_1|$ é menor ou igual que $\int |g_1| d\eta + H_1 (\text{diam } M)^\beta$. Faça $H_1 \rightarrow H_\beta(g_1)$. (c) Se $H > H_\beta(g)$ então $|g(\theta(x)) - g(\theta(y))| \leq H d(\theta(x), \theta(y))^\beta \leq H L^\beta d(x, y)^\beta$. Faça $H \rightarrow H_\beta(g)$ para concluir. As afirmações sobre $H_{\beta, \rho}(\cdot)$ são análogas.

12.3.6. (i) A única parte interessante para provar que $\|\cdot\|_{\beta, \rho}$ é uma norma é a desigualdade triangular. Observe que se $H_i > H_{\beta, \rho}(g_i)$ para $i = 1, 2$ então, sempre que $d(x, y) < \rho$,

$$|g_1(x) + g_2(x) - g_1(y) - g_2(y)| \leq |g_1(x) - g_1(y)| + |g_2(x) - g_2(y)| \leq (H_1 + H_2) d(x, y)^\beta.$$

Fazendo $H_i \rightarrow H_{\beta, \rho}(g_i)$, obtemos que $H_{\beta, \rho}(g_1 + g_2) \leq H_{\beta, \rho}(g_1) + H_{\beta, \rho}(g_2)$. Como vale uma desigualdade análoga para o supremo, segue que $\|g_1 + g_2\|_{\beta, \rho} \leq \|g_1\|_{\beta, \rho} + \|g_2\|_{\beta, \rho}$. Agora considere uma sequência $(g_n)_n$ de Cauchy relativamente à norma $\|\cdot\|_{\beta, \rho}$. Então $(g_n)_n$ também é sequência de Cauchy relativamente à norma do supremo. Como esta última norma é completa, segue que existe alguma função contínua g tal que $\sup_{x \in M} |g_n(x) - g(x)|$ converge para zero. Note que, sempre que $d(x, y) < \rho$,

$$\frac{|g(x) - g(y)|}{d(x, y)^\beta} = \lim_n \frac{|g_n(x) - g_n(y)|}{d(x, y)^\beta} \leq \sup_n H_{\beta, \rho}(g_n) < \infty$$

(porque toda sequência de Cauchy é limitada). Logo, $g \in C^\beta(M)$. Além disso,

$$\frac{|(g - g_n)(x) - (g - g_n)(y)|}{d(x, y)^\beta} = \lim_m \frac{|(g_m - g_n)(x) - (g_m - g_n)(y)|}{d(x, y)^\beta} \leq \lim_m H_{\beta, \rho}(g_m - g_n).$$

Como $H_{\beta, \rho}(g_m - g_n) \leq \|g_m - g_n\|_{\beta, \rho} \leq \varepsilon$ se m e n são grandes, segue que $H_{\beta, \rho}(g - g_n) \rightarrow 0$ e, portanto, $\|g - g_n\|_{\beta, \rho} \rightarrow 0$ quando $n \rightarrow \infty$. (ii) Seja $g \in C^\beta(M)$. Tal como em (12.1.4),

$$(1) \sup |\mathcal{L}g| \leq \text{grau}(f) e^{\sup \varphi} \sup |g| < \infty \text{ (lembre que } \text{grau}(f) = \max_y \#f^{-1}(y)\text{)}.$$

Dado $y \in M$, sejam $h_j : B(y, \rho) \rightarrow M$, $j = 1, \dots, d$ os ramos inversos de f na bola $B(y, \rho)$. Então, $\mathcal{L}g(z) = \sum_j (e^\varphi g)(h_j(z))$ para todo $z \in B(y, \rho)$. Como estes ramos inversos são σ^{-1} -Lipschitz, usando os itens (a) e (c) do Exercício 12.3.5) obtemos que

$$H_{\beta, \rho}(\mathcal{L}g) \leq \sum_j H_{\beta, \rho}((e^\varphi g) \circ h_j) \leq \sum_j (H_{\beta, \rho}(e^\varphi) \sup |g| + e^{\sup \varphi} H_{\beta, \rho}(g)) \sigma^{-\beta}$$

e, portanto,

$$(2) H_{\beta, \rho}(\mathcal{L}g) \leq \text{grau}(f) H_{\beta, \rho}(e^\varphi) \sigma^{-\beta} \sup |g| + \text{grau}(f) e^{\sup \varphi} \sigma^{-\beta} H_{\beta, \rho}(g) < \infty.$$

(iii) As desigualdades (1) e (2) acima dão que $\|\mathcal{L}\|_{\beta,\rho} \leq \text{grau}(f)(e^{\sup \varphi} + H_{\beta,\rho}(e^\varphi)\sigma^{-\beta})$.

12.3.7. (a) Tome $H : C^\beta(M) \rightarrow C^\beta(M)$, $g \mapsto gh$. Verifique que $\lambda\mathcal{P} = H^{-1} \circ \mathcal{L} \circ H$. Em particular, $\mathcal{P}(1) = \mathcal{L}h/(\lambda h) = 1$. Além disso, dada qualquer função mensurável limitada ξ ,

$$\int \xi d\mathcal{P}^* \mu = \int \mathcal{P}(\xi) d\mu = \int \mathcal{L}(\xi h) \frac{h}{\lambda h} d\nu = \int \frac{\xi h}{\lambda} d\mathcal{L}^* \nu = \int \xi h d\nu = \int \xi d\mu$$

Isto mostra que $\tilde{\mathcal{L}}^* \mu = \lambda\mu$. (b) Use o fato de que \mathcal{P} é não negativo, juntamente com a propriedade $\mathcal{P}^* \mu = \mu$, para verificar que $\int |\mathcal{P}g| d\mu \leq \int \mathcal{P}(|g|) d\mu = \int |g| d\mu$ para todo $g \in C^\beta(M)$. Logo, $\int |\mathcal{P}^n g| d\mu \leq \int |g| d\mu$ para todo n e todo g . Analogamente, usando a propriedade $\mathcal{P}1 = 1$, verifique que $\sup |\mathcal{P}g| \leq \mathcal{P}(\sup |g|) = \sup |g|$ para todo $g \in C^\beta(M)$. Logo, $\sup |\mathcal{P}^n g| \leq \sup |g|$ para todo n e todo g . Para quaisquer $y, z \in M$ com $d(y, z) < \rho$,

$$\begin{aligned} \frac{|\mathcal{P}g(y) - \mathcal{P}g(z)|}{d(y, z)^\beta} &\leq \sum_j \frac{(ge^\psi)(h_j(y)) - (ge^\psi)(h_j(z))}{d(y, z)^\beta} \\ &\leq \sum_j e^{\psi(h_j(y))} \frac{g(h_j(y)) - g(h_j(z))}{d(y, z)^\beta} + \sum_j g(h_j(y)) \frac{e^{\psi(h_j(y))} - e^{\psi(h_j(z))}}{d(y, z)^\beta} \end{aligned}$$

Verifique que a primeira parcela é limitada por $\mathcal{P}(H_{\beta,\rho}(g)\sigma^{-\beta}) = H_{\beta,\rho}(g)\sigma^{-\beta}$ e a segunda é limitada por $\text{grau}(f) \sup |g| H_{\beta,\rho}(e^\psi)\sigma^{-\beta}$. Conclua que $H_{\beta,\rho}(\mathcal{P}g) \leq \tau H_{\beta,\rho}(g) + C_0 \sup |g|$ para todo g , com $\tau = \sigma^{-\beta}$ e $C_0 = \text{grau}(f)H_{\beta,\rho}(e^\psi)\sigma^{-\beta}$. Então, para cada n ,

$$H_{\beta,\rho}(\mathcal{P}^n g) \leq \tau H_{\beta,\rho}(\mathcal{P}^{n-1} g) + C_0 \sup |\mathcal{P}^{n-1} g| \leq \tau H_{\beta,\rho}(\mathcal{P}^{n-1} g) + C_0 \sup |g|.$$

Deduzo que $H_{\beta,\rho}(\mathcal{P}^n g) \leq \tau^n H_{\beta,\rho}(g) + C \sup |g|$, com $C = C_0 \sum_{i=0}^{\infty} \tau^i$.

12.3.8. (a) É claro que $\log g$ é (b, β) -Hölder e $\sup g / \inf g$ está próximo de 1 se a norma $\|v\|_{\beta,\rho}$ é pequena (isto estará implícito em tudo que segue). Então $g \in C(b, \beta, R)$. Para estimar $\theta(1, g)$, usaremos a expressão dada pelo Lema 12.3.8. Observe que

$$\beta(1, g) = \sup \left\{ g(x), \frac{\exp(b\delta)g(x) - g(y)}{\exp(b\delta) - 1} : x \neq y, d(x, y) < \rho \right\} \quad \text{onde } \delta = d(x, y)^\beta.$$

Claro que $g(x) \leq 1 + \sup |v|$. Além disso,

$$\frac{\exp(b\delta)g(x) - g(y)}{\exp(b\delta) - 1} \leq \frac{\exp(b\delta)g(y) + \exp(b\delta)H_{\beta,\rho}(v)\delta - g(y)}{\exp(b\delta) - 1} = g(y) + \frac{\delta \exp(b\delta)}{\exp(b\delta) - 1} H_{\beta,\rho}(v).$$

Tome $K_1 > K_2 > 0$, dependendo apenas de b, β, ρ , tais que $K_1 \geq \exp(bs)/(\exp(bs) - 1) \geq K_2$ para todo $s \in [0, \rho^\beta]$. Então o termo do lado direito da desigualdade anterior é limitado por $1 + \sup |v| + K_1 H_{\beta,\rho}(v)$. Logo, $\log \beta(1, g) \leq \log(1 + \sup |v| + K_1 H_{\beta,\rho}(v)) \leq K'_1 \|v\|_{\beta,\rho}$, onde $K'_1 = \max\{K_1, 1\}$. Os argumentos anteriores, fazendo variar x e y , também mostram que $\beta(1, g) \geq 1 + \sup |v|$ e $\beta(1, g) \geq 1 - \sup |v| + K_2 H_{\beta,\rho}(v)$. Deduza que

$$\log \beta(1, g) \geq \max \{ \log(1 + \sup |v|), \log(1 - \sup |v| + K_2 H_{\beta,\rho}(v)) \} \geq K'_2 \|v\|_{\beta,\rho},$$

onde a constante K'_2 só depende K_2, β e ρ . Analogamente, existem constantes $K'_3 > K'_4 > 0$ tais que $-K'_3 \|v\|_{\beta,\rho} \leq \log \alpha(1, g) \leq -K'_4 \|v\|_{\beta,\rho}$. Fixando $K \geq \max\{K_1 + K_3, 1/(K_2 + K_4)\}$, segue que $K^{-1} \|v\|_{\beta,\rho} \leq \theta(1, g) \leq K \|v\|_{\beta,\rho}$. (b) Não é restrição supor $\|v\|_{\beta,\rho} < r$. Note que $\mathcal{P}^n g = 1 + \mathcal{P}^n v$ para todo n . O Corolário 12.3.12 dá que

$$\theta(\mathcal{P}^{kN} g, 1) \leq \Lambda_0^k \theta(1, g) \quad \text{para todo } k,$$

com $\Lambda_0 < 1$. Pelo item (a), segue que $\|\mathcal{P}^{kN} v\| \leq K^2 \Lambda_0^k$ para todo k . Isto dá o enunciado, com $\tau = \Lambda_0^{1/k}$ e $C = K^2 \|\mathcal{P}\|^N \Lambda_0^{-1}$.

12.4.1. Considere $f(x_1, x_2) = (x_1 + 1 + x_2)^d - (x_1^d + x_2^d)$. Observe que, como $d < 1$, as derivadas parciais $\partial_1 f(x_1, x_2)$ e $\partial_2 f(x_1, x_2)$ são negativas em $[0, 1]^2$. Logo, o mínimo da função f é atingido no ponto $(1, 1)$. A escolha de d dá que $f(1, 1) = 0$.

12.4.2. Dado $\delta > 0$, seja \mathcal{U} uma cobertura de A com diâmetro menor que δ . Então $f(\mathcal{U})$ é cobertura de $f(A)$ com diâmetro menor que $L\delta$. Além disso, $\sum_{U \in \mathcal{U}} \text{diam}(f(U))^d \leq$

$\sum_{U \in \mathcal{U}} L^d \text{diam}(U)^d$. Fazendo variar \mathcal{U} , concluímos que $m_d(f(A), L\delta) \leq L^d m_d(A, \delta)$. A primeira parte do enunciado segue, fazendo $\delta \rightarrow 0$. A segunda parte é trivial no caso $t = 0$. Para $t \neq 0$ basta aplicar a primeira parte à homotetia $f(x) = tx$ e à sua inversa (com $L = t$ e $L = t^{-1}$, respectivamente).

12.4.3. É claro que $m_d(M) \leq m_d^a(M)$ e $m_d(M) \geq m_d^f(M)$. Para provar a recíproca da primeira desigualdade, note que dado $\varepsilon > 0$ e dado um conjunto U com diâmetro positivo podemos encontrar um aberto $A \supset U$ tal que $\text{diam } A \leq \text{diam } U(1 + 2\varepsilon)$: basta tomar $A =$ união das bolas de raio $\varepsilon \text{diam } U$ centradas em pontos de U . Em particular, dado $\delta > 0$ e dada qualquer cobertura \mathcal{U} com diâmetro menor que δ , podemos encontrar uma cobertura aberta \mathcal{A} com diâmetro menor que $\delta(1 + 2\varepsilon)$ e tal que $\sum_{A \in \mathcal{A}} (\text{diam } A)^d \leq (1 + 2\varepsilon)^d \sum_{U \in \mathcal{U}} (\text{diam } U)^d$. Isto mostra que $m_d^a(M, \delta(1 + 2\varepsilon)) \leq (1 + 2\varepsilon)^d m_d(M, \delta)$. Fazendo $\delta \rightarrow 0$ e depois $\varepsilon \rightarrow 0$, obtemos a desigualdade pretendida. A recíproca da segunda desigualdade é análoga e um pouco mais simples, porque o fecho F de qualquer conjunto U tem o mesmo diâmetro que U .

12.4.4. Considere $0 < \delta \leq \rho$. Para toda cobertura \mathcal{U} de A com diâmetro menor que δ , temos $\sum_{U \in \mathcal{U}} (\text{diam } U)^d \geq \sum_{U \in \mathcal{U}} K^{-1} \mu(U) \geq K^{-1} \mu(A)$. Tomando o ínfimo sobre \mathcal{U} , vem que $m_d(A, \delta) \geq K^{-1} \mu(A)$. Fazendo $\delta \rightarrow 0$, obtemos que $m_d(A) > K^{-1} \mu(A)$; logo, $d(A) \geq d$.

12.4.5. O triângulo de Sierpinski T pode ser escrito como interseção de uma sequência decrescente de compactos T_n , cada um deles formado por 3^n triângulos equiláteros com diâmetro igual a 2^{-n} . Deduza que $m_d(T) \leq \lim_n 3^n 2^{-nd}$ para todo $d > 0$. Em particular, $m_{d_0}(T) \leq 1$. Considere a probabilidade μ definida em T pela condição de que cada um dos triângulos que formam T_n tem peso 3^{-n} (justifique que esta regra define realmente uma única probabilidade na σ -álgebra de T). Verifique que qualquer conjunto com diâmetro menor que 2^{-n} intersecta, no máximo, 2 dos triângulos de lado 2^{-n+1} que formam T_{n-1} ; logo, a sua medida é menor ou igual que $2 \cdot 3^{-n+1}$. Deduza que $\mu(U) \leq 18 (\text{diam } U)^{d_0}$ para todo U com diâmetro menor que $1/2$. Pelo Exercício 12.4.4, segue que $m_{d_0}(T) \geq 1/18$. As duas desigualdades obtidas para $m_{d_0}(T)$ implicam que $d(T) = d_0$.

12.4.6. Seja $\Sigma = \{1, \dots, d\}^{\mathbb{N}}$ e $h : \Sigma \rightarrow K$ a aplicação que associa a cada sequência $(i_n)_{n \geq 0}$ o único ponto $x \in K$ tal que $f^n(x) \in J_{i_n}$ para todo n . Então h é um homeomorfismo e conjuga $f : K \rightarrow K$ ao deslocamento $\sigma : \Sigma \rightarrow \Sigma$. Seja $\psi : K \rightarrow \mathbb{R}$ a função definida por $\psi(x) = -t \log |f'(x)|$ e seja $\varphi = \psi \circ f$. Pela Proposição 10.3.2, $P(f, \psi) = P(\sigma, \varphi)$. Como $|f'|$ é constante em cada intervalo J_i , a função φ é constante em cada cilindro $[0; i]$. Portanto o cálculo de $P(\sigma, \varphi)$ é um caso particular do Exercício 10.5.2.

12.4.7. Considere $\ell = 1$. Então D, D_1, \dots, D_N (Seção 12.4.3) são intervalos compactos. Não é restrição supor que $D = [0, 1]$. Escreveremos $D_{i^n} = h_{i_0} \circ \dots \circ h_{i_{n-1}}(D)$ para cada $i^n = (i_0, \dots, i_{n-1})$ em $\{1, \dots, N\}^n$. A partir da propriedade de distorção limitada (Proposição 12.4.5), prove que existe $c > 0$ tal que, para todo i^n e todo n ,

- (a) $c \leq |(f^n)'(x)| \text{diam } D_{i^n} \leq c^{-1}$ para todo $x \in D_{i^n}$;
- (b) $d(D_{i^n}, D_{j^n}) \geq c \text{diam } D_{i^n}$ para todo $j^n \neq i^n$;
- (c) $\text{diam } D_{i^{n+1}} \geq c \text{diam } D_{i^n}$ para todo i^n , onde $i^{n+1} = (i_0, \dots, i_{n-1}, i_n)$.

Seja ν a medida de referência do potencial $\varphi = -d_0 \log |f'|$. Como $P(f, \varphi) = 0$, o Lema 12.1.3 e o Corolário 12.1.15 dão que $J_\nu f = |f'|^{d_0}$. Deduza que $c \leq |(f^n)'(x)|^{d_0} \nu(D_{i^n}) \leq c^{-1}$ para qualquer $x \in D_{i^n}$ e, usando (a) uma vez mais, conclua que

$$c^2 \leq \frac{\text{diam}(D_{i^n})^{d_0}}{\nu(D_{i^n})} \leq c^{-2} \quad \text{para todo } i^n \text{ e todo } n.$$

Segue que $\sum_{i^n} \text{diam}(D_{i^n})^{d_0} \leq c^{-2} \sum_{i^n} \nu(D_{i^n}) = c^{-2}$. Como o diâmetro dos D_{i^n} converge uniformemente para zero quando $n \rightarrow \infty$, isto implica que $m_{d_0}(\Lambda) \leq c^{-2}$. Para a estimativa inferior, vamos provar que ν satisfaz a hipótese do princípio da distribuição de massa (Exercício 12.4.4). Dado qualquer U com $\text{diam } U < c \min\{\text{diam } D_1, \dots, \text{diam } D_N\}$, existem $n \geq 1$ e i^n tais que D_{i^n} intersecta U e $c \text{diam } D_{i^n} > \text{diam } U$. Por (b), temos que $\nu(U) \leq \nu(D_{i^n}) \leq c^{-2} \text{diam } D_{i^n}^{d_0}$. Tome n máximo. Então, usando (c), $\text{diam } U \geq c \text{diam } D_{i^{n+1}} \geq c^2 \text{diam } D_{i^n}$ para alguma escolha de i^n . Combinando as duas desigualdades, $\nu(U) \leq c^{-2-2d_0} (\text{diam } U)^{d_0}$. Pelo princípio da distribuição de massa, segue que $m_{d_0}(\Lambda) \geq c^{2+2d_0}$. Finalmente, estenda estes argumentos para qualquer dimensão $\ell \geq 1$.

A.1.1. Tome $\mathcal{B} = \{f^{-1}(C) : C \in \mathcal{C}\}$. Para ver que \mathcal{B} é σ -álgebra, note que a pré-imagem preserva interseções e a passagem ao complementar. Além disso, a pré-imagem do conjunto vazio é o conjunto vazio.

A.1.2. \mathcal{B}_0 é fechada por uniões finitas: a união de dois conjuntos finitos é um conjunto finito; se A^c é finito então $(A \cup B)^c$ é finito, qualquer que seja B . Se X é finito, \mathcal{B}_0 contém todos os subconjuntos, logo é σ -álgebra. Caso contrário, X admite algum subconjunto infinito enumerável Y tal que Y^c é infinito. Então Y está na σ -álgebra gerada por \mathcal{B}_0 mas não está em \mathcal{B}_0 . \mathcal{B}_1 é fechada por uniões enumeráveis: a união enumerável de conjuntos numeráveis é um conjunto enumerável; se A_i^c é enumerável então $(\cup_n A_n)^c$ é enumerável. Todo conjunto enumerável pode ser escrito como união enumerável de conjuntos finitos; logo, a σ -álgebra gerada por \mathcal{B}_0 contém (e, portanto, coincide com) \mathcal{B}_1 .

A.1.3. \mathcal{B} é fechada por passagem ao complementar: se $A \in \mathcal{B}_i$ para todo i então $A^c \in \mathcal{B}_i$ para todo i . \mathcal{B} é fechada por uniões enumeráveis: se $A_n \in \mathcal{B}_i$ para todo i e todo $n \in \mathbb{N}$ então $\cup_n A_n \in \mathcal{B}_i$ para todo i .

A.1.4. Toda classe de equivalência de \sim é invariante pela translação $x \mapsto x + 1$ e, portanto, intersecta $[0, 1)$. Para cada $x \in [0, 1)$, denote $E_x = \{(e+x) - [e+x] : e \in E_0\}$. Verifique que se E_0 é mensurável então E_x é mensurável e $m(E_x) = m(E_0)$. Mostre que $\cup_{n \in \mathbb{Z}} E_{n\alpha} = [0, 1)$ e os conjuntos $E_{n\alpha}$, $n \in \mathbb{Z}$ são disjuntos dois-a-dois. Usando o fato de que a medida de Lebesgue é σ -aditiva, conclua que E_0 não pode ser mensurável.

A.1.5. Defina $B_1 = A_1$ e $B_k = A_k \setminus (A_1 \cup A_2 \cup \dots \cup A_{k-1})$ para $k \geq 2$. Observe que $\cup_{k=1}^{\infty} A_k = \cup_{k=1}^{\infty} B_k$ e $\mu(B_k) \subset \mu(A_k)$.

A.1.6. Dado $\varepsilon > 0$, temos $\mu(F) \leq \sum_{n \geq k} \mu(E_n) < \varepsilon$ para todo k suficientemente grande.

A.1.7. Dados quaisquer B_1, \dots, B_n, \dots em \mathcal{A} , disjuntos dois-a-dois e tais que $B = \cup_n B_n$ também está em \mathcal{A} , defina $C_n = B_1 \cup \dots \cup B_n$. Verifique que os conjuntos $A_n = B \setminus C_n$ satisfazem a condição (A.1.1) no Teorema A.1.14. Logo $\mu(B) - \mu(C_n) = \mu(A_n)$ converge para zero. Como μ é aditiva, segue que $\mu(B) = \lim_n \mu(C_n) = \sum_{j=1}^{\infty} \mu(B_j)$. Reciprocamente, suponha que μ é σ -aditiva. Dados $A_1 \supset \dots \supset A_n \supset \dots$ em \mathcal{A} tais que $\cap_j A_j = \emptyset$, defina $B_j = A_j \setminus A_{j+1}$. Como A_1 é a união disjunta dos B_n , temos $\mu(A_1) = \sum_{j=1}^{\infty} \mu(B_j) = \lim_n \sum_{j=1}^n \mu(A_j) - \mu(A_{j+1}) = \mu(A_1) - \lim_n \mu(A_n)$. Portanto, $\mu(A_n) \rightarrow 0$.

A.1.8. \mathcal{B}_0 é fechada por uniões finitas e por passagem ao complementar e $\mathbb{R} \in \mathcal{B}$. Verifique que m_0 está bem definida: $m_0(B)$ não depende da escolha da representação de $B \in \mathcal{B}_0$ como união disjunta de intervalos. Note que m_0 é aditiva. Seja $(A_n)_n$ sequência decrescente de elementos de \mathcal{B}_0 tal que $\cap_n A_n = \emptyset$. Dado qualquer $\varepsilon > 0$, construa uma sequência decrescente $(K_n)_n$ de compactos com $K_n \subset A_n$ e $m_0(A_n \setminus K_n) < \varepsilon$ para todo n . Então $K_n = \emptyset$ e, portanto, $m_0(A_n) < \varepsilon$, para todo n suficientemente grande. Isto mostra que m_0 é contínua no vazio (Teorema A.1.14).

A.1.9. Dados $A_1 \supset \dots \supset A_i \supset \dots$ tome $A = \cap_{i=1}^{\infty} A_i$. Para $j \geq 1$, considere $A'_j = A_j \setminus A$. Pelo Teorema A.1.14, temos que $\mu(A'_j) \rightarrow 0$ e, portanto, $\mu(A_j) \rightarrow \mu(A)$. Dados $A_1 \subset \dots \subset A_i \subset \dots$ tome $A = \cup_{i=1}^{\infty} A_i$. Para cada j , considere $A'_j = A \setminus A_j$. Pelo Teorema A.1.14, temos que $\mu(A'_j) \rightarrow 0$, ou seja, $\mu(A_j) \rightarrow \mu(A)$.

A.1.10. A parte mais sutil é mostrar que a norma $\|\cdot\|$ é completa. Seja $(\mu_n)_n$ uma sequência de Cauchy. Dado $\varepsilon > 0$, temos $|\mu_m(A) - \mu_n(A)| \leq \|\mu_m - \mu_n\| \leq \varepsilon$ para todo conjunto mensurável $A \subset X$ e quaisquer m, n suficientemente grandes. Segue que $\mu_n(A)$ converge (uniformemente) para algum $\mu(A) \in \mathbb{C}$. Mais geralmente, se m, n são suficientemente grandes então $\sum_k |\mu_m(A_k) - \mu_n(A_k)| \leq \|\mu_m - \mu_n\| \leq \varepsilon$ para qualquer família enumerável $\{A_k : k\}$ de conjuntos mensuráveis, disjuntos dois-a-dois. Tomando o limite quando $m \rightarrow \infty$, obtemos que se n é suficientemente grande então $\sum_{k=1}^{\infty} |\mu(A_k) - \mu_n(A_k)| \leq \varepsilon$ para quaisquer conjuntos mensuráveis A_k disjuntos dois-a-dois. Combinando este fato com $|\mu(\cup_k A_k) - \mu_n(\cup_k A_k)| \leq \varepsilon$ e o fato de que cada μ_n é σ -aditiva, conclua que μ é σ -aditiva. Portanto, $\mu \in \mathcal{M}(X)$. Além disso, tomando o supremo sobre todas as escolhas da família $\{A_k : k\}$, obtemos que $\|\mu - \mu_n\| \leq \varepsilon$ para todo n suficientemente grande.

A.1.11. Para fixar ideias, considere $\mathcal{I} = \mathbb{N}$ (o outro caso é análogo). (a) Note que $d(x, y) = d(y, x)$ e que $d(x, y) > 0$ se, e somente se, existe algum índice i para o qual $x_i \neq y_i$. Conclua que $d(x, y) = 0$ se, e somente se $x = y$. Observe que $N(x, z) \geq \min\{N(x, y), N(x, z)\}$ e,

portanto, $d(x, z) \leq \max\{d(x, y), d(y, z)\} \leq d(x, y) + d(y, z)$. Se $\theta^{m+1} < \rho \leq \theta^m$, a bola aberta de centro x e raio ρ é o cilindro $[0; a_0, \dots, a_m]$. Este cilindro também coincide com a bola fechada, se $\theta^{m+1} \leq \rho < \theta^m$. Para concluir que as duas topologias coincidem basta notar que todo cilindro é união finita deste tipo de cilindros. (b) Como $M = X^{\mathbb{I}}$ é um espaço métrico, basta mostrar que toda sequência $(x_k)_k$ em M admite alguma subsequência convergente. Escreva $x_k = (x_{k,i})_i$. Use um argumento diagonal para encontrar $(k_j)_j \rightarrow \infty$ tal que $x_{k_j,i} = x_{k_i,i}$ para todo $j \geq i$. Logo, a subsequência $(x_{k_j})_j$ é convergente. (c) Vimos no item (a) que todo cilindro $[0; a_0, \dots, a_m]$ é fechado em M . Como M é compacto, segue que todo elemento de \mathcal{A} é compacto; em particular, a álgebra \mathcal{A} é compacta.

A.1.12. Por construção, $m(K_n) = (2/3)m(K_{n+1})$ para todo n . Logo, $m(K) = \lim_n m(K_n) = 0$.

A.1.13. (Royden [Roy63]) (b) \Rightarrow (a) Suponha que existem conjuntos borelianos B_1, B_2 tais que $B_1 \subset E \subset B_2$ e $m(B_2 \setminus B_1) = 0$. Deduza que $m^*(E \setminus B_1) = 0$, logo $E \setminus B_1$ é conjunto mensurável de Lebesgue. Conclua que E é conjunto mensurável de Lebesgue. (a) \Rightarrow (c) Seja E um conjunto mensurável de Lebesgue tal que $m^*(E) < \infty$. Dado $\varepsilon > 0$, existe uma cobertura por retângulos abertos $(R_k)_k$ tais que $\sum_k m^*(R_k) < m^*(E) + \varepsilon$. Então $A = \cup_k R_k$ é aberto contendo E e tal que $m^*(A) - m^*(E) < \varepsilon$. Usando que E é conjunto mensurável de Lebesgue, deduza que $m^*(A \setminus E) < \varepsilon$. Para tratar o caso geral, escreva E como união disjunta de conjuntos mensuráveis de Lebesgue com medida exterior finita. (c) \Leftrightarrow (d) É claro que E é conjunto mensurável de Lebesgue se, e somente se, o seu complementar é conjunto mensurável de Lebesgue. (c) e (d) \Rightarrow (b) Para cada $k \geq 1$, considere fechado $F_k \subset E$ e aberto $A_k \supset E$ tais que $m^*(E \setminus F_k)$ e $m^*(A_k \setminus E)$ sejam menores que $1/k$. Então $B_1 = \cup_k F_k$ e $B_2 = \cap_k A_k$ são borelianos tais que $B_1 \subset E \subset B_2$ e que $m^*(E \setminus B_1) = m^*(B_2 \setminus E) = 0$. Conclua que $m(B_2 \setminus B_1) = m^*(B_2 \setminus B_1) = 0$.

A.1.14. Note que $\{x : (f + g)(x) < c\} = \cup_{q \in \mathbb{Q}} \{x : f(x) < q \text{ e } g(x) < c - q\}$. Logo $f + g$ é mensurável se f e g são. Use um argumento análogo para $f \cdot g$, considerando separadamente os domínios onde f e g têm sinais constantes. Em seguida, $\{x : i(x) < c\} = \cup_n \{x : f_n(x) < c\}$. Portanto, o ínfimo de funções mensuráveis é mensurável. Para o supremo, use o fato de que $\sup_n f_n = -\inf_n(-f_n)$. Como consequência, $f_* = \sup_k \inf_{n > k} f_n$ e $f^* = \inf_k \sup_{n > k} f_n$ são funções mensuráveis.

A.1.15. Segue do Exercício A.1.14, uma vez que $\sum_{n=1}^{\infty} g_n = \lim_k \sum_{n=1}^k g_n$.

A.1.16. Para cada $n \geq 1$ e $k \in \mathbb{Z}$, defina $A_k^n := f^{-1}([k/n, (k+1)/n])$. Defina também $A_-^n = f^{-1}([-\infty, n])$ e $A_+^n = f^{-1}([n, +\infty])$. Mostre que a seguinte sequência de funções simples tem as propriedades desejadas (se f toma valores em \mathbb{R} , substitua ∞ por n no primeiro e no último termo do lado direito da igualdade):

$$s_n = -\infty \mathcal{X}_{A_-} + \sum_{k=-n^2}^{n^2-1} \frac{k}{2^n} \mathcal{X}_{A_k} + \infty \mathcal{X}_{A_+}$$

A.1.17. Note que $f^{-1}(\emptyset) = \emptyset$ e que se $A_k, k \geq 1$ são disjuntos dois-a-dois então o mesmo vale $f^{-1}(A_k), k \geq 1$. Além disso, $f^{-1}(\cup_k A_k) = \cup_k f^{-1}(A_k)$. Segue que $f_* \nu$ é σ -aditiva. Note que $f_* \nu(X) = \nu(X)$.

A.1.18. Mostre que $x \mapsto \frac{1}{n} \#\{0 \leq j \leq n-1 : a_j = 5\}$ é uma função simples para cada $n \geq 1$. Pela Proposição A.1.31, segue que ω_5 é mensurável.

A.2.1. Sejam A_1, \dots, A_m e B_1, \dots, B_n conjuntos mensuráveis e sejam a_1, \dots, a_m e b_1, \dots, b_n números reais tais que $\sum_{i=1}^m a_i \mathcal{X}_{A_i} = \sum_{j=1}^n b_j \mathcal{X}_{B_j}$ em μ -quase todo ponto. Considere conjuntos mensuráveis C_1, \dots, C_p disjuntos dois-a-dois tais que cada A_i é uma união de C_k 's e o mesmo vale para cada B_j . Então $\sum_{i=1}^m a_i \mathcal{X}_{A_i} = \sum_{k=1}^p c_k \mathcal{X}_{C_k}$ onde cada c_k é a soma dos a_i tais que A_i contém C_k . Logo,

$$\sum_{i=1}^m a_i \mu(A_i) = \sum_{i=1}^m a_i \sum_{C_k \subset A_i} \mu(C_k) = \sum_{k=1}^p \left(\sum_{C_k \subset A_i} a_i \right) \mu(C_k) = \sum_{k=1}^p c_k \mu(C_k).$$

Analogamente, $\sum_{j=1}^n b_j \mu(B_j) = \sum_{k=1}^p c_k \mu(C_k) = \sum_{i=1}^m a_i \mu(A_i)$.

A.2.2. Escreva $r_n = \sum_{i=1}^{p(n)} a_i^n \mathcal{X}_{A_i^n}$ e $s_n = \sum_{j=1}^{q(n)} b_j^n \mathcal{X}_{B_j^n}$, com $A_i^n, i = 1, \dots, p(n)$ disjuntos dois-a-dois e $B_j^n, j = 1, \dots, q(n)$ disjuntos dois-a-dois. Note que

$$\lim_n \int r_n d\mu = \sup_n \sum_{i=1}^{p(n)} a_i^n \mu(A_i^n) \quad \text{e} \quad \lim_n \int s_n d\mu = \sup_n \sum_{j=1}^{q(n)} b_j^n \mu(B_j^n).$$

Fixe $n \geq 1$. As hipóteses implicam que dado $\varepsilon > 0$ existe $X_\varepsilon \subset X$ com $\mu(X_\varepsilon) < \varepsilon$ e existe $m_\varepsilon \geq 1$ tal que $r_m(x) \geq s_n(x) - \varepsilon$ para todo $x \notin X_\varepsilon$ e todo $m \geq m_\varepsilon$. Fixe $m = m_\varepsilon$. Seja $\{C_k : k = 1, \dots, p\}$ a família das interseções $A_i^m \cap B_j^n$ não vazias. Defina $a_k = a_i^m$ se $C_k \subset A_i^m$ e $b_k = b_j^n$ se $C_k \subset B_j^n$. Então $a_k \geq b_k - \varepsilon$ a menos que C_k esteja contido em X_ε . Portanto,

$$\begin{aligned} \sum_i a_i^m \mu(A_i^m) &= \sum_k a_k \mu(C_k) \geq \sum_{k: C_k \not\subset X_\varepsilon} a_k \mu(C_k) \geq \sum_{k: C_k \not\subset X_\varepsilon} b_k \mu(C_k) - \varepsilon \\ &\geq \sum_k b_k \mu(C_k) - \varepsilon(1 + \max_k b_k) = \sum_j b_j^n \mu(B_j^n) - \varepsilon(1 + \max_j b_j^n). \end{aligned}$$

Fazendo $\varepsilon \rightarrow 0$ e $m \rightarrow \infty$, obtemos que $\lim_m \sum_i a_i^m \mu(A_i^m) \geq \sum_j b_j^n \mu(B_j^n)$ para todo n e, portanto, $\lim_m \sum_i a_i^m \mu(A_i^m) \geq \lim_n \sum_j b_j^n \mu(B_j^n)$. A desigualdade recíproca é análoga.

A.2.3. A Definição A.2.1 dá a linearidade para funções simples. O caso geral de (1) segue, usando as Definições A.2.2 e A.2.3. As Definições A.2.1 e A.2.2 dão que $\int f d\mu \geq 0$ sempre que $f \geq 0$. Combinado com a linearidade, isto implica o enunciado em (2). Como $-|f| \leq f \leq |f|$, segue que $|\int f d\mu| \leq \int |f| d\mu$. Em particular, f é integrável se $|f|$ é integrável. Para a recíproca, use $|f| = f^+ + f^-$.

A.2.4. Considere $A = \{x : f(x) \geq a\}$ e $s : X \rightarrow \mathbb{R}$ dada por $s(x) = a \mathcal{X}_A$. Argumente que $\int f d\mu \geq \int s d\mu$. Em seguida, aplicando a desigualdade a $|f|$, obtemos que $\mu(\{x : |f(x)| \geq 1/k\}) \leq k \int |f| d\mu = 0$ para todo $k \geq 1$. Conclua que $\{x : f(x) \neq 0\}$ tem medida nula.

A.2.5. Considere a medida ν definida por $\nu(E) = \int_E |f| d\mu$. Suponha que a conclusão do enunciado fosse falsa. Então existiriam $\varepsilon > 0$ e conjuntos mensuráveis $E_n, n \geq 1$ tais que $\nu(E_n) \geq \varepsilon$ para todo n mas $\mu(E_n) \rightarrow 0$. A menos de restringir a uma subsequência, podemos supor $\mu(E_n) \leq 2^{-n}$. Seja $E = \limsup_n E_n = \bigcap_k \bigcup_{n \geq k} E_n$. Conclua que $\nu(E) \geq \varepsilon$ mas $\mu(E) = 0$, contradizendo o fato de que ν é absolutamente contínua relativamente a μ .

A.2.6. Considere funções simples $\xi_i = \sum_l a_{i,l} \mathcal{X}_{A_{i,l}}$ com $\sup |\psi_i - \xi_i| < \varepsilon/2$ para todo i . Então $|\int \psi_i d\mu - \int \xi_i d\mu| < \varepsilon/2$. Mostre que podemos reescrever $\xi_i = \sum_{j=1}^s b_{i,j} \mathcal{X}_{B_j}$ onde os conjuntos mensuráveis B_1, \dots, B_s não dependem de i . Então $\int \xi_i d\mu = \sum_{j=1}^s b_{i,j} \mu(B_j)$. Tome $x_j \in B_j$ e $\alpha_j = \mu(B_j)$ para $j = 1, \dots, s$. Então $|\sum_{j=1}^s b_{i,j} \mu(B_j) - \sum_{j=1}^s \alpha_j \psi_i(x_j)| < \varepsilon/2$ para todo i . O enunciado segue das duas desigualdades.

A.2.7. Aplicando o lema de Fatou às duas sequências $|g| - f_n$ e $|g| + f_n$, conclua que $\limsup_n \int f_n d\mu \leq \int f d\mu \leq \liminf_n \int f_n d\mu$.

A.2.8. (a) Suponha que \mathcal{F} é uniformemente integrável. Considere $C > 0$ correspondente a $\alpha = 1$ e tome $L = C + 1$. Verifique que $\int |f| d\mu < L$ para todo $f \in \mathcal{F}$. Dado $\varepsilon > 0$, considere $C > 0$ correspondente a $\alpha = \varepsilon/2$ e tome $\delta = \varepsilon/(2C)$. Verifique que $\int_A |f| d\mu < \varepsilon$ para todo $f \in \mathcal{F}$ e todo conjunto com $\mu(A) < \delta$. Reciprocamente, dado $\alpha > 0$, tome $\delta > 0$ correspondente a $\varepsilon = \alpha$ e tome $C = L/\delta$. Mostre que $\int_{|f| > C} |f| d\mu < \alpha$. (b) Aplicando o Exercício A.2.5 à função $|g|$, mostre que \mathcal{F} satisfaz o critério em (a). (c) Vamos provar três fatos sobre $f = \lim_n f_n$. (i) f é finita em quase todo ponto: Considere L como em (a). Note que $\mu(\{x : |f_n(x)| \geq k\}) \leq L/k$ para todo $n, k \geq 1$ (Exercício A.2.4) e deduza que $\mu(\{x : |f(x)| \geq k\}) \leq L/k$ para todo $k \geq 1$. (ii) f é integrável: Fixe $K > 0$. Dado qualquer $\varepsilon > 0$, tome δ como em (a). Considere n suficientemente grande para que $\mu(\{x : |f_n(x) - f(x)| > \varepsilon\}) < \delta$. Por um lado,

$$\int_{|f| \leq K} |f| d\mu \leq \int_{|f_n - f| \leq \varepsilon} |f| d\mu + \int_{|f| \leq K, |f_n - f| > \varepsilon} |f| d\mu \leq (L + \varepsilon) + K\delta.$$

Deduza que $\int_{|f| \leq K} |f| d\mu \leq L$ para todo K e $\int |f| d\mu \leq L$. (iii) $(f_n)_n$ converge para f em $L^1(\mu)$: Mostre que, dado $\varepsilon > 0$ existe $K > 0$ tal que $\int_{|f| > K} |f| d\mu < \varepsilon$ e $\int_{|f| > K} |f_n| d\mu < \varepsilon$

para todo n . Tome δ como no item (a) e n grande tal que $\mu(\{x : |f_n(x) - f(x)| > \varepsilon\}) < \delta$. Então,

$$\int_{|f| \leq K} |f_n - f| d\mu \leq \int_{|f_n - f| \leq \varepsilon} |f_n - f| d\mu + \int_{|f_n - f| > \varepsilon} |f_n| d\mu + \int_{|f_n| \leq K, |f_n - f| > \varepsilon} |f| d\mu.$$

O lado direito é majorado por $2\varepsilon + K\delta$. Juntando estas desigualdades, $\int |f_n - f| d\mu < 4\varepsilon + K\delta$ para todo n suficientemente grande.

A.2.9. A condição (A.2.10) implica que $\lim_{\varepsilon} m(B(a, \varepsilon) \cap A)/m(B(a, \varepsilon)) = 1$, ou seja, a é ponto de densidade de A . Para a recíproca, suponha que a é ponto de densidade. Então, dado $\varepsilon > 0$ existe $\delta > 0$ tal que $m(B(a, r) \setminus A^c)/m(B(a, r)) < \varepsilon$ para todo $0 < r \leq \delta$. Mostre que se B é uma bola contida em $B(a, \delta)$ e contendo a , existe $r \leq \delta$ tal que $B \subset B(a, r)$ mas $m(B) \geq 2^{-d}m(B(a, r))$. Logo $m(B \cap A^c)/m(B) < 2^d\varepsilon$. Isto prova (A.2.10).

A.2.10. Se a conclusão do exercício fosse falsa, existiriam $\theta < 1$ e $n_j \rightarrow \infty$ tais que $m(A \cap P) \leq \theta m(P)$ para todo $P \in \mathcal{P}_{n_j}$. Considere qualquer fechado $K \subset A$ com volume positivo e finito e denote $K^\delta = \{x : d(x, K) < \delta\}$. Então,

$$m(K) = \sum_P m(K \cap P) \leq \theta \sum_P m(P) \leq \theta m(K^\delta)$$

onde as somas são sobre todos os $P \in \mathcal{P}_{n_j}$ que intersectam K e supõe-se que $\text{diam } \mathcal{P}_{n_j} < \delta$. Faça $\delta \rightarrow 0$ para chegar a uma contradição.

A.2.11. Suponha que existe $\theta < 1$ e existe $A_0 \subset A$ com medida positiva tal que todo ponto $a \in A_0$ admite bolas $B(a, r_a)$ com raio arbitrariamente pequeno, satisfazendo $m(B(a, r_a) \setminus A_0) \geq (1 - \theta)m(B(a, r_a))$. Use o lema de Vitali para concluir que $m(A_0^\delta \setminus A_0) \geq (1 - \theta)m(A_0)$. Fazendo $\delta \rightarrow 0$, chegamos a uma contradição.

A.2.12. Note que $\mu(A) = 0 \Leftrightarrow A \cap \{x_1, x_2\} = \emptyset \Leftrightarrow \nu(A)$. Portanto $\mu \sim \nu$. Mostre que $d\mu/d\nu = (p_1/q_1)\mathcal{X}_{\{x_1\}} + (p_2/q_2)\mathcal{X}_{\{x_2\}}$ e $d\nu/d\mu$ tem uma expressão análoga, permutando os papéis de p_i e q_i .

A.2.13. Seja $U \subset [0, 1]$ uma união disjunta de intervalos abertos, exatamente 2^{k-1} intervalos de comprimento 2^{-k-1} , para cada $k \geq 1$. Tome $K = [0, 1] \setminus U$ e defina $\mu = 2(m \upharpoonright U)$. Note que $\mu \ll m$ e que $m(K) = m(U) = 1/2$. Não é possível pedir $m(K) = 1$ pois então $m(K^c) = 0$, logo $\mu(K^c) = 0$ e, portanto, μ seria identicamente nula.

A.2.14. Não é restrição supor que os B_n são disjuntos dois-a-dois. Para cada n considere a medida η_n definida em B_n por $\eta_n(A) = \eta(f(A))$. Então $\eta_n \ll (\eta \upharpoonright B_n)$ e, pelo teorema de Radón-Nikodym, existe $\rho_n : B_n \rightarrow [0, +\infty]$ tal que $\int_{B_n} \phi d\eta_n = \int_{B_n} \phi \rho_n d\eta$ para toda $\phi : B_n \rightarrow \mathbb{R}$ mensurável limitada. Defina $J_\eta \upharpoonright B_n = \rho_n$. A unicidade essencial de J_η é consequência da unicidade essencial da derivada de Radón-Nikodym.

A.2.15. Usando a notação do Teorema A.1.21: $\mu^+(A) = \mu(A) = |\mu|(A)$ e $\mu^-(A) = 0$ para todo $A \subset P$ e $\mu^-(B) = -\mu(B) = |\mu|(B)$ e $\mu^+(B) = 0$ para todo $B \subset N$. Logo $\mu^+ = \mathcal{X}_P \mu = \mathcal{X}_P |\mu|$ e $\mu^- = -\mathcal{X}_N \mu = \mathcal{X}_N |\mu|$.

A.2.16. Por definição, $\mu(A) = \lim_n \hat{\mu}_n(A) = \sup_n \hat{\mu}_n(A)$ e valem relações análogas para $\nu(A)$. Logo $\nu(A) = 0 \Rightarrow \hat{\nu}_n(A) = 0$ para todo $n \Rightarrow \hat{\mu}_n(A) = 0$ para todo $n \Rightarrow \mu(A) = 0$. Portanto, $\mu \ll \nu$ e $\hat{\mu}_n \ll \hat{\nu}_n$ para todo n . Observe que $(d\hat{\mu}_n/d\nu) \leq (d\hat{\mu}_n/d\hat{\nu}_n)$ em ν -quase todo ponto e a sequência no lado esquerdo desta desigualdade é não decrescente. Logo, basta mostrar que $\sup_n (d\hat{\mu}_n/d\nu) = (d\mu/d\nu)$ em ν -quase todo ponto. Para isso, suponha que existem $\varepsilon > 0$ e um conjunto E com medida positiva tal que $(d\hat{\mu}_n/d\nu)(x) \leq (d\mu/d\nu)(x) - \varepsilon$ para todo $x \in E$ e todo n . Mostre que $\hat{\mu}_n(E) \leq \mu(E) - \varepsilon\nu(E)$ para todo n , contradizendo a hipótese.

A.3.1. As bolas $B(x, r)$ com $r \in \mathbb{Q}$ formam uma base enumerável de vizinhanças de cada ponto x . Se que M admite algum subconjunto X enumerável e denso então a família \mathcal{B} das bolas $B(x, r)$ com $x \in X$ e $r \in \mathbb{Q}$ é uma base enumerável de abertos: observe que todo subconjunto aberto U coincide com a união dos elementos de \mathcal{B} que ele contém. Recíproca, se \mathcal{B} é uma base enumerável de abertos então tomando um ponto em cada elemento de \mathcal{B} obtemos um subconjunto enumerável denso de M (isto vale mesmo que M seja apenas um espaço topológico). Todo espaço euclidiano \mathbb{R}^d , munido da topologia usual, é separável. Qualquer

espaço M munido com a distância discreta (definida por $d(x, y) = 0$ se $x = y$ e $d(x, y) = 1$ se $x \neq y$) é não separável se M não for enumerável.

A.3.2. As esferas $\{x \in M : d(x, F) = r\}$, $r > 0$ são disjuntas duas-a-duas.

A.3.3. Seja $\{x_k : k \in \mathbb{N}\}$ um subconjunto enumerável denso. Para cada k , considere $r_k \in (\varepsilon/4, \varepsilon/2)$ tal que o bordo de $B_k = \{x \in M : d(x, x_k) < r_k\}$ tem medida nula (Exercício A.3.2). Defina $P_1 = B_1$ e $P_{k+1} = B_{k+1} \setminus (P_1 \cup \dots \cup P_k)$ para $k \geq 1$. Observe que $\{B_k\}$ é uma cobertura e, portanto, $\{P_k\}$ é uma partição de M . Além disso, $\partial P_k \subset \partial B_1 \cup \dots \cup \partial B_k$ tem medida nula, para todo k .

A.3.4. Verifique que ϕ é não decrescente e $\lim_{x \rightarrow a^+} \phi(x) - \lim_{x \rightarrow a^-} \phi(x) = \mu(\{a\})$. Suponha que ϕ é absolutamente contínua, ou seja, para todo $\varepsilon > 0$ existe $\delta > 0$ tal que $\sum_j |a_j - b_j| < \delta$ implica $\sum_j |\phi(a_j) - \phi(b_j)| < \varepsilon$, para qualquer família $\{(a_j, b_j) : j\}$ de intervalos disjuntos dois-a-duos. Observando que $|\phi(a_j) - \phi(b_j)| = \mu((a_j, b_j])$, mostre que se $m(E) = 0$ então $\mu(E) = 0$. Reciprocamente, suponha que $\mu \ll m$. Pelo Exercício A.2.5, dado $\varepsilon > 0$ existe $\delta > 0$ tal que $m(E) < \delta$ implica $\mu(E) < \varepsilon$. Deduza que ϕ é absolutamente contínua.

A.3.5. Seja \mathcal{A} a família dos borelianos cuja função característica é limite de funções uniformemente contínuas $\psi_n : M \rightarrow [0, 1]$. Use o Lema A.3.4 para mostrar que \mathcal{A} contém todos os subconjuntos fechados de M . Note que \mathcal{A} é uma álgebra. Observando que a função característica de $\cup_{j=1}^n A_j$ converge em todo ponto para a função característica de $\cup_{j=1}^{\infty} A_j$, use um argumento diagonal para mostrar que \mathcal{A} é uma σ -álgebra e, portanto, contém todos os borelianos. Conclua que a afirmação no exercício vale para toda função simples. Dada uma função mensurável ψ qualquer, use a Proposição A.1.33 para obter uma sequência de funções simples $(s_n)_n$ tal que $|s_n(x)| \leq |\psi(x)|$ e $s_n(x) \rightarrow \psi(x)$ em todo ponto. Use, mais uma vez, um argumento diagonal para concluir que a afirmação no exercício vale para ψ .

A.3.6. Basta tratar o caso real. Para cada $n \geq 1$, considere o reticulado regular F_n de lado n^{-1} no cubo unitário, isto é, $F_n = \{(i_1/n, \dots, i_d/n) : 0 \leq i_1, \dots, i_d \leq n\}$. Observe que o conjunto \mathcal{F}_n das funções $\varphi : F_n \rightarrow \mathbb{Q}$ é enumerável. Para cada $n \geq 1$ e $\varphi \in \mathcal{F}_n$, represente por $\Phi_{n,\varphi}$ a função que coincide com φ nos pontos de F_n e é afim em cada subcubo $[i_1/n, (i_1+1)/n] \times \dots \times [i_d/n, (i_d+1)/n]$. Verifique que $\Phi_{n,\varphi}$ está bem definida e é contínua. Mostre que para toda função contínua $f : [0, 1]^d \rightarrow \mathbb{R}$ e todo $\varepsilon > 0$ existe (n, φ) tal que $\|f - \Phi_{n,\varphi}\| < \varepsilon$. Portanto, o espaço $C^0([0, 1]^d, \mathbb{R})$ é separável.

A.4.1. Como o cardinal de X é igual ao cardinal de \mathbb{R} , o qual é igual ao cardinal de \mathbb{R}^d , existe alguma bijeção $\varphi : X \rightarrow \mathbb{R}^d$. Defina a topologia em X de modo que φ seja um homeomorfismo. A aplicação φ constitui um atlas de classe C^∞ e dimensão d .

A.4.2. Não. Caso f fosse um difeomorfismo, a aplicação $\psi \circ f \circ \phi^{-1}(x) = x^3$ seria um difeomorfismo do espaço euclidiano \mathbb{R} , o que não é verdade.

A.4.3. Considere a relação de equivalência $x \sim y \Leftrightarrow$ existe alguma curva contínua ligando x a y . Usando cartas locais, observe que as classes de equivalência são conjuntos abertos. Conclua que existe uma única classe de equivalência.

A.4.4. Comece por observar que as aplicações φ_i são bijeções, com inversas $\varphi_i^{-1} : \mathbb{R}^d \rightarrow U_i$ dadas por $\varphi_i^{-1}(y_1, \dots, y_d) = [y_1 : \dots : y_{i-1} : 1 : y_{i+1} : \dots : y_d]$. Defina a topologia de \mathbb{P}^d de modo que cada φ_i seja um homeomorfismo. Observe que as mudanças de coordenadas $\varphi_i \circ \varphi_j^{-1} : \varphi_j(U_i \cap U_j) \rightarrow \varphi_i(U_i \cap U_j)$ são C^∞ :

$$\begin{aligned} \varphi_i \circ \varphi_j^{-1}((y_1, \dots, y_d)) &= \varphi_i([y_1 : \dots : y_{j-1} : 1 : y_{j+1} : \dots : y_d]) \\ &= \left(\frac{y_1}{y_i}, \dots, \frac{y_{i-1}}{y_i}, \frac{y_{i+1}}{y_i}, \dots, \frac{y_{j-1}}{y_i}, \frac{1}{y_i}, \frac{y_{j+1}}{y_i}, \dots, \frac{y_d}{y_i} \right). \end{aligned}$$

A.4.5. Defina $f : GL(d) \rightarrow \mathbb{R}$ por $f(A) = \det A$. Como o determinante é uma função polinomial dos coeficientes da matriz, a aplicação f é C^∞ . Escreva $A = (A_{i,j})_{i,j}$ e, para cada par (i, j) , represente por $A_{i,j}$ a matriz $(d-1) \times (d-1)$ obtida quando removemos a i -ésima linha e a j -ésima coluna da matriz A . Observe que $f(A) = (-1)^{i+j} a_{i,j} \det A_{11} + R_{i,j}$ onde $R_{i,j}$ não depende do coeficiente $a_{i,j}$. Desta forma, $(\partial f / \partial x_{i,j}) = (-1)^{i+j} \det A_{i,j}$. Se $\det A$ é não nulo, existe algum (i, j) tal que $\det A_{i,j} \neq 0$ e portanto $Df(A)$ é sobrejetiva. Isto mostra que todo $y \neq 0$ é valor regular de f .

A.4.6. Dados pontos $b > a$ em \mathbb{R}_{++} , considere uma curva diferenciável $\gamma : [0, 1] \rightarrow \mathbb{R}_+$ que realiza a distância entre os dois pontos, ou seja, tal que $\gamma(0) = a$ e $\gamma(1) = b$ e

$$d(a, b) = \text{comp}(\gamma) = \int_0^1 \left\| \frac{d\gamma}{dt}(t) \right\|_{\gamma(t)} dt = \int_0^1 \frac{1}{\gamma(t)} \frac{d\gamma}{dt}(t) dt.$$

É claro que γ é monótona crescente: caso contrário, existiria alguma restrição ligando a a b , contradizendo a hipótese de que a curva γ é minimizante. Então podemos fazer a mudança de variável $s = \gamma(t)$ na última integral. Desta forma obtemos $d(a, b) = \int_a^b ds/s = \log b - \log a$.

A.4.7. Considere $h : M \times N \rightarrow \mathbb{R}^{m+n}$ definida por $h(x, y) = x - y$. Esta aplicação é C^∞ . Pelo teorema de Sard (Teorema A.4.6), o complementar do conjunto dos valores não regulares tem volume zero em \mathbb{R}^{m+n} . Agora basta observar que se $v \in \mathbb{R}^{m+n}$ é valor regular então $M + v$ é transversal a N .

A.4.8. Considere cartas locais $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$, $x \mapsto \varphi_\alpha(x)$ de M e $\varphi_\alpha : TU_\alpha M \rightarrow X_\alpha \times \mathbb{R}^d$, $(x, v) \mapsto (\varphi_\alpha, D\varphi_\alpha(x)v)$ de TM . Note que $\varphi_\alpha \circ \pi \circ D\varphi_\alpha^{-1}$ é a projeção canônica $X_\alpha \times \mathbb{R}^d \rightarrow X_\alpha$, a qual é infinitamente diferenciável. Como M é de classe C^r e TM é de classe C^{r-1} , segue que π é de classe C^{r-1} .

A.5.1. Dados quaisquer $x, y \in (a, b)$, considere a função $\psi : [0, 1] \rightarrow \mathbb{R}$ dada por $\psi(t) = \phi(tx + (1-t)y) - t\phi(x) + (1-t)\phi(y)$. Verifique que $\psi(0) = \psi(1) = 0$ e $\psi'' \geq 0$. Conclua que $\psi(t) \leq 0$ para todo t . Para provar a segunda parte, fixe $a < \alpha < x < \beta < b$. Seja $f(t) = \phi(x) + m(t-x)$ a função afim cujo gráfico passa por $(s, \phi(s))$ e $(x, \phi(x))$ e seja $g(t) = \phi(x) + n(x-t)$ a função afim cujo gráfico passa por $(x, \phi(x))$ e $(t, \phi(t))$. Por convexidade, $f(t) \leq \phi(t) \leq g(t)$ para $t > x$ e $g(t) \leq \phi(t) \leq f(t)$ para $t < x$. Logo ϕ é contínua no ponto x .

A.5.2. (a) Use o fato de que a função exponencial é convexa. (b) A partir da desigualdade de Young, mostre que $\int |f\bar{g}| d\mu \leq 1$ sempre que $\|f\|_p = \|g\|_q = 1$. Deduza o caso geral da desigualdade de Hölder. (c) Comece por notar que $|f+g|^p \leq |f| |f+g|^{p-1} + |g| |f+g|^{p-1}$. Aplique a desigualdade de Hölder a cada uma das parcelas do lado direito desta desigualdade para obter a desigualdade Minkowski.

A.5.3. Usando a desigualdade de Hölder mostre que $\|f\|_p \leq \mu(X)^{(1/p)-(1/q)} \|f\|_q$ sempre que $1 \leq p < q < \infty$. Mostre que esta desigualdade ainda vale quando $q = \infty$: $\|f\|_p \leq \mu(X)^{(1/p)} \|f\|_\infty$. Portanto $L^p(\mu) \supset L^q(\mu)$ para todo $1 \leq p < q \leq \infty$.

A.5.4. A menos de substituir f por $f/\|f\|_\infty$, podemos supor que $\|f\|_\infty = 1$. Então $\int |f|^{n+1} d\mu \leq \int |f|^n d\mu$ para todo n . Considere $\alpha < \beta < 1$ e $n \geq 1$. Escrevendo $\int |f|^n d\mu$ como soma das integrais sobre os domínios $X_\beta = \{|f| \geq \beta\}$, $Y_{\alpha,\beta} = \{\alpha \leq |f| < \beta\}$ e $Z_\alpha = \{|f| < \alpha\}$, mostre que $\int |f|^n d\mu / \int_{X_\beta} |f|^n d\mu \leq 1 + \mu(Y_{\alpha,\beta})/\mu(X_\beta) + (\alpha/\beta)^n \mu(Z_\alpha)/\mu(X_\beta)$. Tome α próximo de β e em seguida tome n grande, de tal forma que a expressão do lado direito da desigualdade seja menor que $1 + (1-\beta)$. Deste modo,

$$1 \geq \frac{\int |f|^{n+1} d\mu}{\int |f|^n d\mu} \geq \frac{\int_{X_\beta} |f|^{n+1} d\mu}{\int_{X_\beta} |f|^n d\mu} \frac{1}{1 + (1-\beta)} \geq \frac{\beta}{2-\beta}$$

para todo $\beta < 1$ e todo n suficientemente grande.

A.5.5. (i) $\|\cdot\|_p$ é uma norma. A única parte interessante é a desigualdade triangular: para $1 \leq p < \infty$ ela é dada pelo Teorema A.5.3; para $p = \infty$ ela segue imediatamente de $|f+g| \leq |f| + |g|$. (ii) Critério de completude. Suponha que o espaço é completo. Se $\sum_k v_k$ é absolutamente somável então $s_n = \sum_{k=1}^n v_k$ é sequência de Cauchy e, portanto, converge. Reciprocamente, se $(s_n)_n$ é sequência de Cauchy, existe $(n_k)_k \rightarrow \infty$ tal que $\|s_m - s_n\| \leq 2^{-k}$ para quaisquer $m, n \geq n_k$. Defina $v_k = s_{n_{k+1}} - s_{n_k}$. A série $\sum_k v_k$ é absolutamente somável, logo convergente. Isto quer dizer que a subsequência $(s_{n_k})_k$ converge. Conclua, usando o fato de que toda sequência de Cauchy que admite subsequência convergente é convergente. (iii) $L^p(\mu)$ é completo. Considere qualquer série $\sum_k f_k$ tal que $A = \sum_k \|f_k\|_p$ é finito. Seja $g = \sum_{k=1}^\infty |f_k|$ e $g_n = \sum_{k=1}^n |f_k|$, para cada $n \geq 1$. Pela desigualdade triangular, $\|g_n\|_p \leq A$ para todo n . Então $\|g\|_p \leq A$: para $p < \infty$ use o teorema da convergência monótona; o caso $p = \infty$ é imediato. Em particular, $g(x)$ é finito e, portanto, a série complexa $\sum_{k=1}^\infty f_k(x)$ é absolutamente somável, para μ -quase todo x . Seja $f(x)$ o limite. Resta mostrar que $f \in$

$L^p(\mu)$ e $\sum_{k=1}^n f_k$ converge para f em $L^p(\mu)$. Dado $\varepsilon > 0$, temos $\|\sum_{k=1}^m f_k - \sum_{k=1}^n f_k\|_p \leq \sum_{k>n} \|f_k\|_p \leq \varepsilon$ para $m > n$ e todo n suficientemente. Passando ao limite quando $m \rightarrow \infty$, obtemos que $\|f - \sum_{k=1}^n f_k\|_p \leq \varepsilon$: para $p < \infty$ use o teorema da convergência; o caso $p = \infty$ é, mais uma vez, imediato. Em particular, $\|f\|_p < \infty$.

A.5.6. (Rudin [Rud87, Teorema 6.16]) Note que $\Phi(g) \in L^p(\mu)^*$ e $\|\Phi(g)\| \leq \|g\|_q$: para $q < \infty$ isso segue da desigualdade de Hölder; o caso $q = \infty$ é imediato. É claro que Φ é linear. Para ver que é injetivo, dada g tal que $\Phi(g) = 0$, considere uma função β com valores no círculo unitário tal que $\beta g = |g|$. Então $\phi(\mathcal{X}_B)\beta = \int |g| d\mu = 0$, logo $g = 0$. Resta mostrar que para todo $\phi \in L^p(\mu)^*$ existe $g \in L^q(\mu)$ tal que $\phi = \Phi(g)$ e $\|g\|_q = \|\phi\|$. Para cada conjunto mensurável $B \subset M$, defina $\eta(B) = \phi(\mathcal{X}_B)$. Verifique que η é uma medida complexa (para provar σ -aditividade precisará usar $p < \infty$) e observe que $\eta \ll \mu$. Considere a derivada de Radón-Nikodym $g = (d\eta/d\mu)$. Então $\phi(\mathcal{X}_B) = \int_B g d\mu$ para todo B ; conclua que $\phi(f) = \int f g d\mu$ para toda $f \in L^\infty(\mu)$. No caso $p = 1$ esta construção dá $|\int_B g d\mu| \leq \|\phi\| \mu(B)$ para todo conjunto mensurável. Deduza que $\|g\|_\infty \leq \|\phi\|$. Agora suponha $1 < p < \infty$. Tome $f_n = \mathcal{X}_{B_n} \beta |g|^{q-1}$, onde $B_n = \{x : |g(x)| \leq n\}$. Observe que $f_n \in L^\infty(\mu)$ e $|f_n|^p = |g|^q$ no conjunto B_n e

$$\int_{B_n} |g|^q d\mu = \int f_n g d\mu = \phi(f_n) \leq \|\phi\| \left(\int |f_n|^p d\mu \right)^{1/p} \leq \|\phi\| \left(\int_{B_n} |g|^q d\mu \right)^{1/p}.$$

Isto dá $\int_{B_n} |g|^q d\mu \leq \|\phi\|^q$ para todo n e, portanto, $\|g\|_q \leq \|\phi\|$. Finalmente, $\phi(f) = \int f g d\mu$ para todo $f \in L^p(\mu)$, pois os dois lados da igualdade são funcionais contínuos que coincidem no subconjunto denso L^∞ .

A.5.7. Use o Exercício A.3.5 e o teorema da convergência dominada.

A.5.8. Lembre do Exemplo A.5.9.

A.6.1. (1) Verifique que $\|tx + (1-t)y\|^2 \leq \max\{\|x\|^2, \|y\|^2\}$ para quaisquer $x, y \in H$ e $t \in [0, 1]$. Deduza que toda bola, fechada ou aberta é convexa. (2) Por definição, $\|v \pm w\|^2 = \|v\|^2 \pm (v \cdot w + w \cdot v) + 2\|w\|^2$. (3) Análogo a (2).

A.6.2. Suponha que $v = \sum_\alpha v_\alpha$ e $v' = \sum_\alpha v'_\alpha$, ou seja, para todo $\varepsilon > 0$ existem conjuntos finitos I e I' tais que $\|v - \sum_{\beta \in J} v_\beta\| < \varepsilon$ para todo conjunto finito $J \supset I$ e $\|v' - \sum_{\beta \in J} v'_\beta\| < \varepsilon$ para todo conjunto finito $J \supset I'$. Então $\|(v + v') - \sum_{\beta \in J} (v_\beta + v'_\beta)\| < 2\varepsilon$ para todo conjunto finito $J \supset I \cup I'$.

A.6.3. Se L é contínuo na origem, dado $\varepsilon > 0$ existe $\delta > 0$ tal que $\|v\| \leq \delta$ implica $\|L(v)\| \leq \varepsilon$. Segue que $\|L(v)\| \leq (\varepsilon/\delta)\|v\|$ para todo v . Portanto, L é limitado. Reciprocamente, é claro que se L é limitado então L é contínuo na origem. Deduza que é contínuo em todo ponto.

A.6.4. Observe que V^\perp é o espaço das funções f tais que $\int f d\mu = 0$. Conclua que $P_H : L^2 \rightarrow H$ é dada por $P_H(g) = \int g d\mu$.

A.6.5. Por definição, $u \cdot Lv = L^*u \cdot v$ e $u \cdot L^*v = (L^*)^*u \cdot v$ para quaisquer u e v . Logo $v \cdot (L^*)^*u = L^*v \cdot u$ para quaisquer u e v . Invertendo os papéis de u e v , vemos que $L = (L^*)^*$. Note que $\|L^*u \cdot v\| \leq \|L\| \|u\| \|v\|$ para todo u e v . Tomando $v = L^*u$, segue que $\|L^*u\| \leq \|L\| \|u\|$ para todo u , logo $\|L^*\| \leq \|L\|$. Como $L = (L^*)^*$, segue que $\|L\| \leq \|L^*\|$, logo as duas normas coincidem. Como a norma de operador é submultiplicativa, $\|L^*L\| \leq \|L\|^2$. Por outro lado, $u \cdot L^*Lu = \|Lu\|^2$ e, portanto, $\|L^*L\| \|u\|^2 \geq \|Lu\|^2$, para todo u . Deduza que $\|L^*L\| \geq \|L\|^2$ e, portanto, as duas expressões coincidem. Analogamente, $\|LL^*\| = \|L\|^2$.

A.6.6. Podemos supor $z = 0$. Deduza da igualdade do paralelogramo (Exercício A.6.1) que $2(\|x\|^2 + \|y\|^2) = \|x + y\|^2 + \|x - y\|^2 \geq 4d(z, K)^2 + \|x - y\|^2$, para quaisquer $x, y \in K$. Considere $(v_n)_n$ em K tal que $d(z, v_n) \rightarrow d(z, K)$. Conclua da desigualdade anterior que $(v_n)_n$ é sequência de Cauchy. Tome $v = \lim_n v_n$. Para provar a unicidade: se v' é outro ponto minimizador, $4d(z, K)^2 = 2(\|v\|^2 + \|v'\|^2) \geq 4d(z, K)^2 + \|v - v'\|^2$ e, portanto, $v = v'$.

A.6.7. A continuidade do produto interno dá que $S^\perp = \bar{S}^\perp$ e este subespaço é fechado. Dado $v \in H$, seja s o elemento de \bar{S} mais próximo de v (Exercício A.6.6) e seja $s^\perp = v - s$. Verifique que $s^\perp \in S^\perp$. A unicidade da decomposição é consequência de $\bar{S} \cap S^\perp = \{0\}$. É imediato da definição e das observações anteriores que $\bar{S} \subset (S^\perp)^\perp = (S^\perp)^\perp$. Para a recíproca, considere qualquer $v \in (S^\perp)^\perp$ e analise a respectiva decomposição $v = s + s^\perp$.

A.6.8. Suponha que $v \in H$ e $(u_n)_n$ é uma sequência em E tal que $u_n \cdot v \rightarrow u \cdot v$ para todo $v \in H$. Considerando $v \in E^\perp$, conclua que $v \in (E^\perp)^\perp$. Pelo Exercício A.6.7, segue que $u \in E$. Portanto E é fechado na topologia fraca. Agora considere qualquer sequência $(v_n)_n$ em $U(E)$ convergindo para algum $v \in H$. Para cada n , tome $u_n = h^{-1}(v_n) \in E$. Como h é uma isometria, $\|u_m - u_n\| = \|v_m - v_n\|$ para quaisquer m, n . Segue que $(u_n)_n$ é uma sequência de Cauchy em E e, portanto, admite limite $u \in E$. Logo, $v = h(u)$ está em $U(E)$.

A.6.9. Observe que L é isometria se, e somente se, preserva o produto interno: $L^*L = \text{id} \Leftrightarrow L^*Lu \cdot v = u \cdot v \forall u, v \Leftrightarrow Lu \cdot Lv = u \cdot v \forall u, v$. Então L preserva a norma. Reciprocamente, a identidade de polarização (Exercício A.6.1) implica que se L preserva a norma então L preserva o produto interno. É claro da definição que se L é unitário então L é isometria e $L^{-1} = L^*$. Reciprocamente, suponha que $L^*L = \text{id}$ e L é invertível. Então $LL^*L(u) = L(u)$ para todo u e, como L é sobrejetivo, isto implica que $LL^* = \text{id}$.

A.7.1. A inversa de $T + H$ está dada pela equação $(T + H)(T^{-1} + J) = \text{id}$, a qual pode ser reescrita como uma equação de ponto fixo $J = -L^{-1}HL^{-1} + L^{-1}HJ$. Use a hipótese para mostrar que esta equação admite uma (única) solução. Logo $T + H$ é um isomorfismo. Deduza que $L - \lambda \text{id}$ sempre que $\lambda > \|L\|$. Portanto, o espectro de L está contido no disco de raio $\|L\|$. Também segue da observação anterior que se $L - \lambda \text{id}$ é um isomorfismo então o mesmo vale para $L - \lambda' \text{id}$ se λ' está suficientemente próximo de λ .

A.7.2. Como $\dim \ker L + \dim L(H) = \dim H$, o operador L é injetivo se, e somente se, é sobrejetivo e, nesse caso. Além disso, L é necessariamente contínuo. Seja ℓ^2 o espaço das seqüências $x = (x_n)_{n \in \mathbb{N}}$ de números complexos tais que $\|x\| = \sum_n |x_n|^2$ é finita e seja $L: \ell^2 \rightarrow \ell^2$ o deslocamento $L((x_n)_n) = (x_{n+1})_n$. Verifique que L é um operador unitário, o conjunto dos autovalores é o disco aberto e o espectro de L é o disco fechado.

A.7.3. (1) Aditividade implica $E(\emptyset) = 0$. O complementar de $\text{supp } E$ está coberto por uma família enumerável de conjuntos com medida nula. Logo $E((\text{supp } E)^c) = 0$ e, por aditividade, $E(\text{supp } E) = \text{id}$. (2) Se $A \subset B$ então $E(B \setminus A) = E(B) - E(A)$, por aditividade. Consequentemente,

$$E(B \setminus A) = E(B) - E(B)E(A) - E(A)E(B) + E(A) \Rightarrow$$

$$2E(A) = E(B)E(A) + E(A)E(B) \Rightarrow \begin{cases} 2E(A) = E(A)E(B)E(A) + E(A)E(B) \\ 2E(A) = E(B)E(A) + E(A)E(B)E(A). \end{cases}$$

Segue que $E(A)E(B) = E(B)E(A)$. Substituindo esta conclusão com a igualdade no meio, vem que $E(A) = E(A)E(B)$, ou seja, $E(A) \leq E(B)$. (3) Decomponha $A \cup B$ como união disjunta $A \cup B = (A \setminus (A \cap B)) \cup (B \setminus (A \cap B)) \cup (A \cap B)$ e use a aditividade. (4) Multiplique a identidade em (3) por (A) e use (2) para concluir que $E(A)E(B) = E(A \cap B)$ e, analogamente, $E(B)E(A) = E(A \cap B)$.

A.7.4. (1) Observe que $L - \lambda \text{id} = \int (z - \lambda) dE(z)$ e use o Lema A.7.4. Pela propriedade de continuidade inferior (Exercício A.1.9), $E(\{\lambda\}) = \lim_n E(\{z : |z - \lambda| \leq 1/n\})$. Segue que $E(\{\lambda\})v = v$. (2) Segue do Exercício A.7.3 que $E(B)E(\{\lambda\}) = E(\{\lambda\})$ se $\lambda \in B$ e $E(B)E(\{\lambda\}) = E(\emptyset) = 0$ caso contrário. Como $L = \int z dE(z)$, vem que $Lv = \lambda E(\{\lambda\})v = \lambda v$.

A.7.5. Lembre do Exemplo A.7.7. A medida espectral de T está dada por $E_T(B)(\varphi_{j,l})_{j,l} = (\mathcal{X}_B \varphi_{j,l})_{j,l}$. Logo $E_T(B) = 0$ se, e somente se, $\sigma_j(B) = 0$ para todo j . Então o mesmo vale para a medida espectral de L , dada por $E(B) = U^{-1}E_T(B)U$.

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/.
- [AF07] A. Avila and G. Forni. Weak mixing for interval exchange transformations and translation flows. *Ann. of Math.*, 165:637–664, 2007.
- [AKM65] R. Adler, A. Konheim, and M. McAndrew. Topological entropy. *Trans. Amer. Math. Soc.*, 114:309–319, 1965.
- [AKN06] V. Arnold, V. Kozlov, and A. Neishtadt. *Mathematical aspects of classical and celestial mechanics*, volume 3 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, third edition, 2006. [Dynamical systems. III], Translated from the Russian original by E. Khukhro.
- [Ano67] D. V. Anosov. Geodesic flows on closed Riemannian manifolds of negative curvature. *Proc. Steklov Math. Inst.*, 90:1–235, 1967.
- [Arn78] V. I. Arnold. *Mathematical methods of classical mechanics*. Springer Verlag, 1978.
- [AS67] D. V. Anosov and Ya. G. Sinai. Certain smooth ergodic systems. *Russian Math. Surveys*, 22:103–167, 1967.
- [Bal00] V. Baladi. *Positive transfer operators and decay of correlations*. World Scientific Publishing Co. Inc., 2000.
- [BDV05] C. Bonatti, L. J. Díaz, and M. Viana. *Dynamics beyond uniform hyperbolicity*, volume 102 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, 2005.
- [Bil68] P. Billingsley. *Convergence of probability measures*. John Wiley & Sons Inc., New York, 1968.
- [Bil71] P. Billingsley. *Weak convergence of measures: Applications in probability*. Society for Industrial and Applied Mathematics, 1971. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 5.
- [Bir13] G. D. Birkhoff. Proof of Poincaré’s last Geometric Theorem. *Trans. Amer. Math. Soc.*, 14:14–22, 1913.
- [Bir67] G. Birkhoff. *Lattice theory*, volume 25. A.M.S. Colloq. Publ., 1967.
- [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, 1983.
- [Bos86] J.-B. Bost. Tores invariants des systèmes hamiltoniens. *Astérisque*, 133–134:113–157, 1986.
- [Bos93] M. Boshernitzan. Quantitative recurrence results. *Invent. Math.*, 113(3):617–631, 1993.
- [Bow71] R. Bowen. Entropy for group endomorphisms and homogeneous spaces. *Trans. Amer. Math. Soc.*, 153:401–414, 1971.

- [Bow72] R. Bowen. Entropy expansive maps. *Trans. Am. Math. Soc.*, 164:323–331, 1972.
- [Bow75a] R. Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, volume 470 of *Lect. Notes in Math.* Springer Verlag, 1975.
- [Bow75b] R. Bowen. A horseshoe with positive measure. *Invent. Math.*, 29:203–204, 1975.
- [Bow78] R. Bowen. Entropy and the fundamental group. In *The Structure of Attractors in Dynamical Systems*, volume 668 of *Lecture Notes in Math.*, pages 21–29. Springer-Verlag, 1978.
- [BS00] L. Barreira and J. Schmeling. Sets of “non-typical” points have full topological entropy and full Hausdorff dimension. *Israel J. Math.*, 116:29–70, 2000.
- [Cas04] A. A. Castro. *Teoria da medida*. Projeto Euclides. IMPA, 2004.
- [Cla72] J. Clark. *A Kolmogorov shift with no roots*. ProQuest LLC, Ann Arbor, MI, 1972. Thesis (Ph.D.)—Stanford University.
- [dC79] M. do Carmo. *Geometria riemanniana*, volume 10 of *Projeto Euclides*. Instituto de Matemática Pura e Aplicada, 1979.
- [Dei85] K. Deimling. *Nonlinear functional analysis*. Springer Verlag, 1985.
- [Din70] E. Dinaburg. A correlation between topological entropy and metric entropy. *Dokl. Akad. Nauk SSSR*, 190:19–22, 1970.
- [Din71] E. Dinaburg. A connection between various entropy characterizations of dynamical systems. *Izv. Akad. Nauk SSSR Ser. Mat.*, 35:324–366, 1971.
- [dlL93] R. de la Llave. Introduction to K.A.M. theory. In *Computational physics (Almuñécar, 1992)*, pages 73–105. World Sci. Publ., 1993.
- [DS57] N. Dunford and J. Schwarz. *Linear operators I: General theory*. Wiley & Sons, 1957.
- [DS63] N. Dunford and J. Schwarz. *Linear operators II: Spectral theory*. Wiley & Sons, 1963.
- [Dug66] J. Dugundji. *Topology*. Allyn and Bacon Inc., 1966.
- [Edw79] R. E. Edwards. *Fourier series. A modern introduction. Vol. 1*, volume 64 of *Graduate Texts in Mathematics*. Springer-Verlag, second edition, 1979.
- [ET36] P. Erdős and P. Turán. On some sequences of integers. *J. London. Math. Soc.*, 11:261–264, 1936.
- [Fal90] K. Falconer. *Fractal geometry*. John Wiley & Sons Ltd., 1990. Mathematical foundations and applications.
- [Fer02] R. Fernandez. *Introdução à teoria da medida*. Projeto Euclides. IMPA, 2002.
- [FFT09] S. Ferenczi, A. Fisher, and M. Talet. Minimality and unique ergodicity for adic transformations. *J. Anal. Math.*, 109:1–31, 2009.
- [FO70] N. Friedman and D. Ornstein. On isomorphism of weak Bernoulli transformations. *Advances in Math.*, 5:365–394, 1970.
- [Fri69] N. Friedman. *Introduction to ergodic theory*. Van Nostrand, 1969.
- [Fur61] H. Furstenberg. Strict ergodicity and transformation of the torus. *Amer. J. Math.*, 83:573–601, 1961.
- [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 University Press, 1981.
- [Goo71a] T. Goodman. Relating topological entropy and measure entropy. *Bull. London Math. Soc.*, 3:176–180, 1971.
- [Goo71b] G. Goodwin. Optimal input signals for nonlinear-system identification. *Proc. Inst. Elec. Engrs.*, 118:922–926, 1971.

- [GT08] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math.*, 167:481–547, 2008.
- [Gur61] B. M. Gurevič. The entropy of horocycle flows. *Dokl. Akad. Nauk SSSR*, 136:768–770, 1961.
- [Hal50] P. Halmos. *Measure Theory*. D. Van Nostrand Company, 1950.
- [Hal51] P. Halmos. *Introduction to Hilbert Space and the theory of Spectral Multiplicity*. Chelsea Publishing Company, New York, 1951.
- [Hay] N. Haydn. Multiple measures of maximal entropy and equilibrium states for one-dimensional subshifts. Preprint Penn State University.
- [Hir94] M. Hirsch. *Differential topology*, volume 33 of *Graduate Texts in Mathematics*. Springer-Verlag, 1994. Corrected reprint of the 1976 original.
- [Hof77] F. Hofbauer. Examples for the nonuniqueness of the equilibrium state. *Trans. Amer. Math. Soc.*, 228:223–241, 1977.
- [Hop39] E. Hopf. Statistik der geodätischen Linien in Mannigfaltigkeiten negativer Krümmung. *Ber. Verh. Sächs. Akad. Wiss. Leipzig*, 91:261–304, 1939.
- [HvN42] P. Halmos and J. von Neumann. Operator methods in classical mechanics. II. *Ann. of Math.*, 43:332–350, 1942.
- [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.
- [Kal82] S. Kalikow. T, T^{-1} transformation is not loosely Bernoulli. *Ann. of Math.*, 115:393–409, 1982.
- [Kat71] Yi. Katznelson. Ergodic automorphisms of T^m are Bernoulli shifts. *Israel J. Math.*, 10:186–195, 1971.
- [Kat80] A. Katok. Lyapunov exponents, entropy and periodic points of diffeomorphisms. *Publ. Math. IHES*, 51:137–173, 1980.
- [Kea75] M. Keane. Interval exchange transformations. *Math. Zeit.*, 141:25–31, 1975.
- [KH95] A. Katok and B. Hasselblatt. *Introduction to the modern theory of dynamical systems*, volume 54 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 1995. With a supplementary chapter by Katok and Leonardo Mendoza.
- [KM10] S. Kalikow and R. McCutcheon. *An outline of ergodic theory*, volume 122 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2010.
- [Kok35] J. F. Koksma. Ein mengentheoretischer Satz über die Gleichverteilung modulo Eins. *Compositio Math.*, 2:250–258, 1935.
- [KR80] M. Keane and G. Rauzy. Stricte ergodicité des échanges d’intervalles. *Math. Zeit.*, 174:203–212, 1980.
- [Kri70] W. Krieger. On entropy and generators of measure-preserving transformations. *Trans. Amer. Math. Soc.*, 149:453–464, 1970.
- [Kri75] W. Krieger. On the uniqueness of the equilibrium state. *Math. Systems Theory*, 8:97–104, 1974/75.
- [KSS91] A. Krámli, N. Simányi, and D. Szász. The K -property of three billiard balls. *Ann. of Math.*, 133:37–72, 1991.
- [KSS92] A. Krámli, N. Simányi, and D. Szász. The K -property of four billiard balls. *Comm. Math. Phys.*, 144:107–148, 1992.
- [KW82] Y. Katznelson and B. Weiss. A simple proof of some ergodic theorems. *Israel J. Math.*, 42:291–296, 1982.

- [Led84] F. Ledrappier. Propriétés ergodiques des mesures de Sinai. *Publ. Math. I.H.E.S.*, 59:163–188, 1984.
- [Lin77] D. Lind. The structure of skew products with ergodic group actions. *Israel J. Math.*, 28:205–248, 1977.
- [LVY] G. Liao, M. Viana, and J. Yang. Entropy conjecture for diffeomorphisms away from tangencies. *J. European Math. Soc.*
- [LY85] F. Ledrappier and L.-S. Young. The metric entropy of diffeomorphisms. I. Characterization of measures satisfying Pesin’s entropy formula. *Ann. of Math.*, 122:509–539, 1985.
- [Man75] A. Manning. Topological entropy and the first homology group. In *Dynamical Systems, Warwick, 1974*, volume 468 of *Lecture Notes in Math.*, pages 185–190. Springer-Verlag, 1975.
- [Mañ81] R. Mañé. A proof of Pesin’s formula. *Ergod. Th. & Dynam. Sys.*, 1:95–101, 1981.
- [Mañ85] R. Mañé. Hyperbolicity, sinks and measure in one-dimensional dynamics. *Comm. Math. Phys.*, 100:495–524, 1985.
- [Mañ87] R. Mañé. *Ergodic theory and differentiable dynamics*. Springer Verlag, 1987.
- [Mas82] H. Masur. Interval exchange transformations and measured foliations. *Ann. of Math.*, 115:169–200, 1982.
- [Mey00] C. Meyer. *Matrix analysis and applied linear algebra*. Society for Industrial and Applied Mathematics (SIAM), 2000.
- [Mis76] M. Misiurewicz. A short proof of the variational principle for a Z_+^N action on a compact space. *Asterisque*, 40:147–187, 1976.
- [MP77a] M. Misiurewicz and F. Przytycki. Entropy conjecture for tori. *Bull. Pol. Acad. Sci. Math.*, 25:575–578, 1977.
- [MP77b] M. Misiurewicz and F. Przytycki. Topological entropy and degree of smooth mappings. *Bull. Pol. Acad. Sci. Math.*, 25:573–574, 1977.
- [MP08] W. Marzantowicz and F. Przytycki. Estimates of the topological entropy from below for continuous self-maps on some compact manifolds. *Discrete Contin. Dyn. Syst. Ser.*, 21:501–512, 2008.
- [MT78] G. Miles and R. Thomas. Generalized torus automorphisms are Bernoullian. *Advances in Math. Supplementary Studies*, 2:231–249, 1978.
- [New88] S. Newhouse. Entropy and volume. *Ergodic Theory Dynam. Systems*, 8*(Charles Conley Memorial Issue):283–299, 1988.
- [NP66] D. Newton and W. Parry. On a factor automorphism of a normal dynamical system. *Ann. Math. Statist*, 37:1528–1533, 1966.
- [NR97] A. Nogueira and D. Rudolph. Topological weak-mixing of interval exchange maps. *Ergod. Th. & Dynam. Sys.*, 17:1183–1209, 1997.
- [Orn60] D. Ornstein. On invariant measures. *Bull. Amer. Math. Soc.*, 66:297–300, 1960.
- [Orn70] D. Ornstein. Bernoulli shifts with the same entropy are isomorphic. *Advances in Math.*, 4:337–352 (1970), 1970.
- [Orn72] Donald S. Ornstein. On the root problem in ergodic theory. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971)*, Vol. II: *Probability theory*, pages 347–356. Univ. California Press, 1972.
- [Orn74] D. Ornstein. *Ergodic theory, randomness, and dynamical systems*. Yale University Press, 1974. James K. Whittemore Lectures in Mathematics given at Yale University, Yale Mathematical Monographs, No. 5.
- [OS73] D. Ornstein and P. Shields. An uncountable family of K -automorphisms. *Advances in Math.*, 10:63–88, 1973.

- [OU41] J. C. Oxtoby and S. M. Ulam. Measure-preserving homeomorphisms and metrical transitivity. *Ann. of Math.*, 42:874–920, 1941.
- [Par53] O. S. Parasyuk. Flows of horocycles on surfaces of constant negative curvature. *Uspehi Matem. Nauk (N.S.)*, 8:125–126, 1953.
- [Pes77] Ya. B. Pesin. Characteristic Lyapunov exponents and smooth ergodic theory. *Russian Math. Surveys*, 324:55–114, 1977.
- [Pes97] Ya. Pesin. *Dimension theory in dynamical systems*. University of Chicago Press, 1997. Contemporary views and applications.
- [Pet83] K. Petersen. *Ergodic theory*. Cambridge Univ. Press, 1983.
- [Phe93] R. Phelps. *Convex functions, monotone operators and differentiability*, volume 1364 of *Lecture Notes in Mathematics*. Springer-Verlag, second edition, 1993.
- [Pin60] M. S. Pinsker. *Informatsiya i informatsionnaya ustoychivostsluchainykh velichin i protsessov*. Problemy Peredači Informacii, Vyp. 7. Izdat. Akad. Nauk SSSR, 1960.
- [PT93] J. Palis and F. Takens. *Hyperbolicity and sensitive-chaotic dynamics at homoclinic bifurcations*. Cambridge University Press, 1993.
- [PU10] F. Przytycki and M. Urbański. *Conformal fractals: ergodic theory methods*, volume 371 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, 2010.
- [PW72a] W. Parry and P. Walters. Errata: “Endomorphisms of a Lebesgue space”. *Bull. Amer. Math. Soc.*, 78:628, 1972.
- [PW72b] W. Parry and P. Walters. Endomorphisms of a Lebesgue space. *Bull. Amer. Math. Soc.*, 78:272–276, 1972.
- [PY98] M. Pollicott and M. Yuri. *Dynamical systems and ergodic theory*, volume 40 of *London Mathematical Society Student Texts*. Cambridge University Press, 1998.
- [Que87] M. Queffélec. *Substitution dynamical systems—spectral analysis*, volume 1294 of *Lecture Notes in Mathematics*. Springer-Verlag, 1987.
- [Rok61] V. A. Rokhlin. Exact endomorphisms of a Lebesgue space. *Izv. Akad. Nauk SSSR Ser. Mat.*, 25:499–530, 1961.
- [Rok62] V. A. Rokhlin. On the fundamental ideas of measure theory. *A. M. S. Transl.*, 10:1–54, 1962. Transl. from *Mat. Sbornik* 25 (1949), 107–150. First published by the A. M. S. in 1952 as Translation Number 71.
- [Rok67a] 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.
- [Rok67b] V. A. Rokhlin. Metric properties of endomorphisms of compact commutative groups. *Amer. Math. Soc. Transl.*, 64:244–252, 1967.
- [Roy63] H. L. Royden. *Real analysis*. The Macmillan Co., 1963.
- [RS61] V. A. Rokhlin and Ja. G. Sinaĭ. The structure and properties of invariant measurable partitions. *Dokl. Akad. Nauk SSSR*, 141:1038–1041, 1961.
- [Rud87] W. Rudin. *Real and complex analysis*. McGraw-Hill, 3 edition, 1987.
- [Rue73] D. Ruelle. Statistical Mechanics on a compact set with Z^ν action satisfying expansiveness and specification. *Trans. Amer. Math. Soc.*, 186:237–251, 1973.
- [Rue04] D. Ruelle. *Thermodynamic formalism*. Cambridge Mathematical Library. Cambridge University Press, second edition, 2004. The mathematical structures of equilibrium statistical mechanics.
- [RY80] C. Robinson and L. S. Young. Nonabsolutely continuous foliations for an Anosov diffeomorphism. *Invent. Math.*, 61:159–176, 1980.

- [SC87] Ya. Sinaĭ and Nikolay Chernov. Ergodic properties of some systems of two-dimensional disks and three-dimensional balls. *Uspekhi Mat. Nauk*, 42:153–174, 256, 1987.
- [Shu69] M. Shub. Endomorphisms of compact differentiable manifolds. *Amer. Journal of Math.*, 91:129–155, 1969.
- [Shu74] M. Shub. Dynamical systems, filtrations and entropy. *Bull. Amer. Math. Soc.*, 80:27–41, 1974.
- [Sim02] N. Simányi. The complete hyperbolicity of cylindric billiards. *Ergodic Theory Dynam. Systems*, 22:281–302, 2002.
- [Sin63] Ya. Sinaĭ. On the foundations of the ergodic hypothesis for a dynamical system of Statistical Mechanics. *Soviet. Math. Dokl.*, 4:1818–1822, 1963.
- [Sin70] Ya. Sinaĭ. Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards. *Uspekhi Mat. Nauk*, 25:141–192, 1970.
- [Ste58] E. Sternberg. On the structure of local homeomorphisms of Euclidean n -space - II. *Amer. J. Math.*, 80:623–631, 1958.
- [SW75] M. Shub and R. Williams. Entropy and stability. *Topology*, 14:329–338, 1975.
- [SX10] R. Saghin and Z. Xia. The entropy conjecture for partially hyperbolic diffeomorphisms with 1-D center. *Topology Appl.*, 157:29–34, 2010.
- [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.
- [Vee82] W. Veech. Gauss measures for transformations on the space of interval exchange maps. *Ann. of Math.*, 115:201–242, 1982.
- [Ver99] Alberto Verjovsky. *Sistemas de Anosov*, volume 9 of *Monographs of the Institute of Mathematics and Related Sciences*. Instituto de Matemática y Ciencias Afines, IMCA, Lima, 1999.
- [Wal73] P. Walters. Some results on the classification of non-invertible measure preserving transformations. In *Recent advances in topological dynamics (Proc. Conf. Topological Dynamics, Yale Univ., New Haven, Conn., 1972; in honor of Gustav Arnold Hedlund)*, pages 266–276. Lecture Notes in Math., Vol. 318. Springer, 1973.
- [Wal75] P. Walters. A variational principle for the pressure of continuous transformations. *Amer. J. Math.*, 97:937–971, 1975.
- [Wal82] P. Walters. *An introduction to ergodic theory*. Springer Verlag, 1982.
- [Wey16] H. Weyl. Über die Gleichverteilungen von Zahlen mod Eins. *Math. Ann.*, 77:313–352, 1916.
- [Yoc92] J.-C. Yoccoz. Travaux de Herman sur les tores invariants. *Astérisque*, 206:Exp. No. 754, 4, 311–344, 1992. Séminaire Bourbaki, Vol. 1991/92.
- [Yom87] Y. Yomdin. Volume growth and entropy. *Israel J. Math.*, 57:285–300, 1987.
- [Yos68] K. Yosida. *Functional analysis*. Second edition. Die Grundlehren der mathematischen Wissenschaften, Band 123. Springer-Verlag, 1968.
- [Yuz68] S. A. Yuzvinskii. Metric properties of endomorphisms of compact groups. *Amer. Math. Soc. Transl.*, 66:63–98, 1968.
- [Zyg68] A. Zygmund. *Trigonometric series: Vols. I, II*. Second edition, reprinted with corrections and some additions. Cambridge University Press, 1968.

Índice

- 2^X família de todos os subconjuntos, 428
- $A\Delta B$ diferença simétrica de conjuntos, 432
- $B(x, T, \varepsilon)$ bola dinâmica de um fluxo, 317
- $B(x, \infty, \varepsilon)$ bola dinâmica infinita, 324
- $B(x, n, \varepsilon)$ bola dinâmica, 266
- $B(x, r)$ bola de centro x e raio r , 450
- B^δ δ -vizinhança de um conjunto, 36, 522
- C^* cone dual, 51
- $C^0(M)$ espaço das funções contínuas, 50, 433, 454
- $C_+^0(M)$ cone das funções positivas, 51
- $C^\beta(M)$ espaço de funções Hölder, 412
- $C^r(M, N)$ espaço das aplicações C^r , 458
- $C_n(\varphi, \psi)$ sequência de correlações, 184
- D_i densidade inferior, 58
- D_s densidade superior, 58
- Df aplicação derivada, 460
- $E(A, P)$ esperança condicional, 153
- E^* dual de um espaço de Banach, 49
- $G(f, \phi)$ pressão via conjuntos geradores, 327
- $H(\alpha)$ entropia de uma cobertura, 302
- $H_\beta(g)$ constante de Hölder, 399
- $H_\mu(\mathcal{P})$ entropia de uma partição, 250
- $H_\mu(\mathcal{P}/\mathcal{Q})$ entropia condicional, 252
- $H_{\beta, \rho}(g)$ constante de Hölder local, 399
- $I(U)$ conjunto dos vetores invariantes, 67
- $I(\mathcal{A})$ informação média de um alfabeto, 249
- $I(a)$ quantidade de informação de um caracter, 249
- $I_{\mathcal{P}}$ função de informação de uma partição, 250
- $L^\infty(\mu)$ espaço das funções essencialmente limitadas, 466
- $L^p(\mu)$ espaço das funções p -integráveis, 464
- $P(f, \phi)$ pressão, 325
- $P(f, \phi, \alpha)$ pressão com respeito a uma cobertura, 324
- $P(x, \cdot)$ probabilidade de transição, 193
- $P_{i,j}$ probabilidade de transição, 193
- R_θ rotação no círculo ou no toro, 16
- $S(f, \phi)$ pressão via conjuntos separados, 327
- S^1 círculo, 16
- S^\perp complemento ortogonal, 469
- S^d esfera de dimensão d , 457
- $S_n\varphi$ soma orbital de uma função, 378
- TM fibrado tangente, 460
- T^1M fibrado tangente unitário, 463
- T_pM

- espaço tangente a uma variedade, 459
 U_f operador de Koopman, 50, 51
 U_f^* dual do operador de Koopman, 51
 $V(\mu, \Phi, \varepsilon)$ base de vizinhanças da topologia fraca*, 36
 $V(v, \{g_1, \dots, g_N\}, \varepsilon)$ base de vizinhanças da topologia fraca, 49
 $V^*(g, \{v_1, \dots, v_N\}, \varepsilon)$ base de vizinhanças da topologia fraca*, 49
 $V_a(\mu, \mathcal{A}, \varepsilon)$ base de vizinhanças da topologia fraca*, 37
 $V_c(\mu, \mathcal{B}, \varepsilon)$ base de vizinhanças da topologia fraca*, 37
 $V_f(\mu, \mathcal{F}, \varepsilon)$ base de vizinhanças da topologia fraca*, 37
 $V_p(\mu, \mathcal{B}, \varepsilon)$ base de vizinhanças da topologia pontual, 44
 $V_u(\mu, \varepsilon)$ base de vizinhanças da topologia uniforme, 44
 \mathcal{X}_B função característica de conjunto, 437
 $\text{Difeo}^r(M)$ espaço dos difeomorfismos C^r , 459
 $\text{Fix}(f)$ conjunto dos pontos fixos, 313
 $\text{GL}(d, \mathbb{R})$ grupo linear, 166, 461
 $\text{O}(d, \mathbb{R})$ grupo ortogonal, 166
 $\text{SL}(d, \mathbb{R})$ grupo linear especial, 166, 461
 Σ_A deslocamento de tipo finito, 314
 Σ_P deslocamento de tipo finito, 195
 $\alpha \vee \beta$ soma de coberturas, 302
 $\alpha \prec \beta$ cobertura menos fina, 302
 β^n iterado unilateral de uma cobertura, 309
 $\beta^{\pm n}$ iterado bilateral de uma cobertura, 309
 $\mathcal{L}^1(\mu)$ espaço das funções integráveis, 442
 $\mathcal{M}(X)$ espaço das medidas, 50, 433
 $\mathcal{M}_1(M)$ espaço das medidas de probabilidade, 36
 $\mathcal{M}_1(f)$ espaço das medidas invariantes de probabilidade, 118
 $\mathcal{M}_e(f)$ espaço das medidas ergódicas de probabilidade, 118
 $\mathcal{P} \prec \mathcal{Q}$ partição menos fina, 252
 $\mathcal{P} \vee \mathcal{Q}$ soma de partições, 250
 \mathcal{P}^n iterado de uma partição, 254
 $\mathcal{P}^{\pm n}$ iterado bilateral de uma partição, 257
 $\mathcal{U}^r(f, \varepsilon)$ vizinhança C^r , 459
 δ_p medida de Dirac, 47
 $\text{div } F$ divergente de um campo de vetores, 20
 $\lambda = (\lambda_\alpha)_\alpha$ vetor de translações, 204
 λ_{\max} expoente de Lyapunov máximo, 85
 λ_{\min} expoente de Lyapunov mínimo, 85
 $\mu \perp \nu$ medidas mutuamente singulares, 447
 $\nu \ll \mu$ relação de continuidade absoluta, 447
 $\text{GL}(d)$ grupo das matrizes invertíveis de dimensão d , 79
 ∂I extremo inferior de intervalo, 205
 $\partial \mathcal{P}$ bordo de uma partição, 263
 \mathbb{P}^d espaço projetivo, 463
 $\rho(B)$ raio espectral, 316
 $\text{esp}(L)$ espectro de um operador linear, 472
 supess supremo essencial, 466
 supp suporte de uma medida, 436, 472
 \tanh tangente hiperbólica, 404
 $\tau(E, x)$ tempo médio de visita, 65
 $\theta(g_1, g_2)$ distância projetiva, 401
 $\tilde{\varphi}$ média temporal de uma função, 72

- \mathbb{T}^d toro de dimensão d , 18, 457
 φ^+ parte positiva de uma função, 442
 φ^- parte negativa de uma função, 442
 $\vee_\alpha \mathcal{M}_\alpha$ σ -álgebra gerada, 281
 $d(M)$ dimensão de Hausdorff, 415
 $e(\psi, x)$ esperança condicional, 151
 $f_*\mu$ imagem de uma medida, 44, 50
 f_A endomorfismo linear, 112
 $g(\phi)$ entropia topológica de fluxos via conjuntos geradores, 318
 $g(f)$ entropia topológica via conjuntos geradores, 305
 $h(f)$ entropia topológica, 303
 $h(f, \alpha)$ entropia com respeito a uma cobertura, 303
 $h^\pm(f, \varepsilon, x)$ entropia local, 266
 $h_\mu(f)$ entropia de um sistema dinâmico, 255
 $h_\mu(f, \mathcal{P})$ entropia com respeito a uma partição, 255
 $h_\mu(f, \mathcal{P}, x)$ entropia local, 266
 $m_d(M)$ medida de Hausdorff d -dimensional, 414
 $s(\phi)$ entropia topológica de fluxos via conjuntos separados, 318
 $s(f)$ entropia topológica via conjuntos separados, 305
 $w = (w_\alpha)_\alpha$ vetor de comprimentos, 204
a menos de medida nula, 432
alfabeto, 204
álgebra, 428
compacta, 431
de funções, 456
separadora, 456
de medida, 237
 σ -álgebra, 428
boreliana, 429
de Borel, 429
gerada, 281, 429
gerada a menos de medida nula, 432
produto, 105, 444, 445
aplicação
conforme, 417
contínua, 437
de classe C^r , 458
de primeiro retorno, 5
derivada, 460
diferenciável, 458
exponencial, 463
Hölder, 451
Lipschitz, 451
mensurável, 437
atlas
de classe C^r , 457
diferenciável, 457
atlas compatíveis, 458
átomo, 454, 475
automorfismo
de Bernoulli, 277
de grupo, 166
de Kolmogorov, 281, 284
de Möbius, 413
autovalor, 221
aproximado, 225
bacia de uma medida, 101, 350
baricentro
de uma medida, 289
base
da topologia, 436
de abertos, 436
de Fourier, 469
de Hammel, 469
de Hilbert, 469
de vizinhanças, 36, 37, 436
enumerável de abertos, 436
enumerável de vizinhanças, 436
base de uma pilha, 172
bilhar, 135
canto, 135
dispersivo, 139
mesa, 135
semidispersivo, 140
bola dinâmica, 266, 305
de um fluxos, 317
bordo de uma partição, 263
círculo, 16
unitário, 16
cadeia, 245
campo de vetores, 460
carta local, 457
cilindro, 445
aberto, 55
elementar, 445

- mensurável, 54
- classe monótona, 431
- cobertura, 414
 - aberta, 302, 431
 - menos fina, 302
- cociclo, 85
- cogumelo de Bunimovich, 140
- complementar ortogonal, 67, 469
- completamento de espaço de medida, 432
- comprimento
 - de progressão aritmética, 58
 - de uma curva, 462
- condição
 - de Keane, 205
 - de torção, 126–128, 131
- cone, 51, 400
 - dual, 51, 380
 - normal, 51
- conjectura da entropia, 320
- conjugação
 - topológica, 219, 304
- conjunto
 - de continuidade de uma medida, 37
 - boreliano, 429
 - com volume zero, 461
 - convexo, 45
 - de Aubry -Mather, 129
 - de Cantor, 415
 - dos vetores invariantes, 67
 - fortemente convexo, 289
 - gerador, 305
 - gerador para fluxos, 317
 - invariante, 55, 96, 355
 - justo de medidas, 42
 - mensurável, 428
 - mensurável de Lebesgue, 241, 435, 440
 - minimal, 161
 - não errante, 34
 - ortonormado, 469
 - residual, 122, 459, 461
 - separado, 305
 - separado para fluxos, 318
 - sindético, 9, 164
 - transitivo, 120
- constante
 - de Avogadro, 332
 - de Boltzmann, 333
 - de expansividade, 264, 312
 - temporal, 87
- continuidade absoluta, 116, 435, 447
- contração, 308
- convergência
 - em $L^2(\mu)$, 69
 - em distribuição, 44
 - em quase todo ponto, 443
 - para equilíbrio, 399
- convexidade, 119, 287
- coordenada local, 457
- coordenadas ação-ângulo, 125
- correlação, 183
- cristal reticulado, 332
- curva minimizante, 462
- curvatura negativa, 132
- decaimento de correlações, 210
- decomposição
 - de Hahn, 433
 - de Lebesgue, 448
 - de Oseledets, 86
- densidade
 - de uma medida, 351, 447
 - superior, 58, 61
 - zero no infinito, 192
- derivada de Radón-Nikodym, 447
- desigualdade
 - de Cauchy-Schwartz, 465
 - de Hölder, 465, 467
 - de Jensen, 467
 - de Minkowski, 465, 467
 - de Tchebysheff-Markov, 448
 - de Young, 467
- desintegração de uma medida, 145
- deslocamento, 60
 - de Bernoulli, 105, 106
 - de Markov, 193
 - de Markov ergódico, 196
 - de Markov finito, 194
 - de Markov misturador, 199
 - de tipo enumerável, 232
 - de tipo finito, 195, 314
 - gaussiano, 235, 284
- diâmetro
 - de uma cobertura, 309, 414
- diagonal, 30
- difeomorfismo, 457, 458
 - de class C^r , 458
- diferença
 - ortogonal, 231
 - simétrica, 432
- dimensão
 - de Hausdorff, 415
 - de Hilbert, 469
- diofantino
 - número, 164
 - vetor, 126
- distância, 450
 - associada uma métrica riemanniana, 462
 - de Cayley-Klein, 413
 - de Poincaré, 413
 - hiperbólica, 413
 - invariante, 170
 - projetiva, 401
- distorção limitada, 104, 105, 109, 110
- distribuição de Gibbs, 333

- divergente de um campo de vetores, 20
- domínio
 - de invertibilidade, 296
 - fundamental, 88
- dual de um espaço de Banach, 49, 466
- dualidade, 50, 211, 378, 465
- endomorfismo
 - de grupo, 166
 - linear, 112
- energia livre de Gibbs, 333
- entropia, 332
 - com respeito a uma cobertura, 303
 - com respeito a uma partição, 255
 - condicional, 252
 - da transformação de Gauss, 273
 - de um canal de comunicação, 249
 - de um sistema dinâmico, 255
 - de uma cobertura, 302
 - de uma partição, 250
 - dos deslocamentos de Markov, 272
 - dos endomorfismos lineares do toro, 275
 - local, 266
- entropia topológica, 301, 303, 307
 - de um fluxo, 318
- envolvente convexa, 419
- equação cohomológica, 162
- equações de Hamilton-Jacobi, 123
- equivalência
 - ergódica, 186, 217, 218, 238
 - espectral, 217, 220, 238
- ergodicidade
 - da expansão decimal, 103
 - da rotação irracional, 102, 103
 - dos deslocamentos de Markov, 196
- esfera de dimensão d , 457
- espaço
 - compacto, 431
 - completamente metrizável, 459
 - conexo, 457
 - conexo por caminhos, 463
 - de Baire, 120, 122, 459, 461
 - de Banach, 49, 464, 468
 - de configurações, 332
 - de Hausdorff, 429
 - de Hilbert, 468
 - de Hilbert separável, 469
 - de Lebesgue, 237, 241, 243
 - de medida, 430
 - de medida completo, 432
 - de probabilidade, 430
 - não trivial, 281
 - padrão, 237
 - dual, 470
 - euclidiano, 457
 - métrico, 450
 - completo, 453
 - mensurável, 428
 - metrizável, 39, 450
 - produto, 444, 445
 - projetivo, 463
 - separável, 39, 452, 455, 456
 - tangente a uma variedade, 459
 - topológico, 429
 - vetorial topológico, 45
- espaços
 - isometricamente isomorfos, 455, 469
- especificação, 371
- espectro
 - de Lebesgue, 217, 229
 - de um operador linear, 472
 - de uma rotação, 228
 - de uma transformação, 221
 - discreto, 217, 227
- esperança condicional, 151, 153, 268
- estádio de Bunimovich, 140
- estado
 - de equilíbrio, 302, 332, 341
 - de Gibbs, 332, 346, 377
- exemplo
 - de Chacon, 224
 - de Furstenberg, 162
- expansão
 - de Fourier, 102, 113
 - em fração contínua, 12
- expansividade, 264
- expoentes de Lyapunov, 86
- extensão
 - de uma transformação, 54
 - natural, 54, 57
 - natural múltipla, 57
- fórmula
 - da entropia de Pesin, 394
 - de Bowen-Manning, 378, 418
 - de Liouville, 20
 - de mudança de variáveis, 300
 - de Rokhlin, 295, 297, 357
- fator, 259
 - topológico, 304, 375
- fibrado tangente, 132, 460
 - unitário, 132, 463
- filtração de Oseledets, 86
- fluxo, 2, 460
 - conservativo, 19
 - de Anosov, 133
 - geodésico, 132, 463
 - hamiltoniano, 123
 - horocíclico, 284
 - suspensão, 87
 - teorema de existência de medidas invariantes, 48
 - teorema de recorrência, 5
 - teorema ergódico de Birkhoff, 78

- teorema ergódico subaditivo, 87
- uniformemente contínuo, 318
- uniformemente hiperbólico, 132
- folhas
 - estáveis, 114
 - instáveis, 114
- folheação
 - estável, 114
 - instável, 114
- forma
 - de área, 203
 - de volume, 20, 126
 - simplética, 126
- fração contínua, 12
 - de tipo limitado, 118
- função
 - σ -aditiva, 430
 - p -integrável, 464
 - afim, 287
 - côncava, 466
 - característica, 437
 - contínua, 437
 - convexa, 466
 - de distribuição, 44
 - de Hölder, 213
 - de informação de uma partição, 250
 - de multiplicidades, 476
 - entropia, 262
 - essencialmente limitada, 466
 - finitamente aditiva, 430
 - fortemente afim, 294
 - Hölder, 399
 - integrável, 442
 - invariante, 69, 96
 - localmente constante, 211
 - localmente integrável, 446
 - quase periódica, 125
 - semicontínua, 33
 - simples, 438
- funcional linear
 - contínuo, 455, 470
 - limitado, 470
 - norma, 455
 - positivo, 443, 455
 - positivo sobre um cone, 51
 - tangente, 53, 347
- gás
 - ideal, 137
 - reticulado, 332
- gerador
 - bilateral, 261
 - unilateral, 261
- Grassmanniana, 458
- grau de uma aplicação, 352, 362
- grupo
 - compacto, 169
 - de Lie, 165
 - linear, 166, 461
 - linear especial, 166, 461
 - localmente compacto, 166
 - metrizável, 169
 - ortogonal, 166
 - topológico, 165
- hamiltoniano, 123
 - não degenerado, 125
- hipótese ergódica
 - de Boltzmann, 65, 122
 - de Boltzmann-Sinai, 135
- hipersuperfície de energia, 123
- homeomorfismo, 430
 - de torção, 129
- homomorfismo
 - de álgebras de medida, 238
- identidade do paralelogramo, 471
- imagem de uma medida, 44, 50
- informação média de um alfabeto, 249
- integrabilidade
 - uniforme, 448
- integral
 - de Lebesgue, 442
 - de uma função simples, 441
 - primeira, 22, 123
- intercâmbio de intervalos, 202
 - redutível, 205
- intervalo
 - d -ádico, 499
 - em \mathbb{Z} , 58
- invariante de equivalência
 - ergódica, 219
 - espectral, 221
- isometria, 308
 - linear, 50, 471, 472
- isomorfismo
 - de álgebras de medida, 238
 - ergódico, 218, 237, 238
- iterado
 - de uma medida, 44, 50
 - de uma partição, 254, 257
- jacobiano, 296
- lacuna espectral, 211, 412
- laminação, 133
- lema
 - de Borel-Cantelli, 439
 - de distorção, 352, 353
 - de Fatou, 444
 - de Kakutani-Rokhlin, 28
 - de Riemann-Lebesgue, 236
 - de Vitali, 447, 522
 - de Zorn, 31

- do sombreamento, 363
- levantamento
 - de um conjunto invariante, 55
 - de uma medida invariante, 56
- limite
 - inferior, 429
 - superior, 429
- máquina de somar, 171
- média
 - orbital, 72
 - temporal, 72
- método de empilhamento, 173
- métrica
 - de Levy-Prohorov, 39
 - riemanniana, 462
- matriz
 - de covariâncias, 236
 - de transição, 314
 - estocástica, 194
 - aperiódica, 199
 - irreduzível, 196
 - hiperbólica, 114
 - positiva definida, 235
 - simétrica, 235
- medida, 430
 - σ -finita, 8, 74, 430
 - absolutamente contínua, 118
 - atômica, 454
 - boreliana, 450
 - com memória finita, 201
 - com sinal, 50, 432
 - com sinal finita, 433
 - completa, 432
 - complexa, 433, 472
 - de Bernoulli, 193, 445
 - de Dirac, 47, 430
 - de Haar, 168
 - de Hausdorff, 414
 - de Lebesgue, 434
 - de Lebesgue no círculo, 16
 - de Liouville, 123
 - de máxima entropia, 341
 - de Markov, 193
 - de probabilidade, 430
 - de referência, 379, 380
 - ergódica, 74
 - espectral, 472
 - estacionária, 57, 193
 - exterior de Lebesgue, 435
 - física, 357
 - finita, 430
 - gaussiana, 235
 - infinita, 7
 - invariante, 2, 44
 - justa, 453
 - não atômica, 454
 - não singular, 296, 449
 - positiva, 432
 - produto, 106, 444, 445
 - quociente, 144
 - regular, 450
 - suspensão, 92
 - transporte, 91, 92
- medidas
 - equivalentes, 14, 447
 - mutuamente singulares, 119, 447
- memória finita, 192, 201
- minimalidade, 18, 159, 206
- mistura fraca, 222
- mudança de coordenadas, 457
- multiplicidade
 - de um autovalor, 221
 - de um expoente de Lyapunov, 86
- número
 - balanceado, 105
 - de Lebesgue, 455
 - diofantino, 164
 - normal, 11, 105
- número de rotação, 128
- norma, 465, 468
 - L^∞ , 466
 - L^p , 464
 - da convergência uniforme, 455
 - de funcional linear, 49
 - de operador, 170, 316, 319
 - de um funcional linear, 455
 - de uma matriz, 78
 - de uma medida, 433
- odômetro, 171
- operador
 - de Koopman, 50, 51
 - de Ruelle-Perron-Frobenius, 378
 - de transferência, 211, 378
 - linear
 - adjunto, 470
 - autoadjunto, 471, 472
 - contínuo, 470
 - dual, 51, 378
 - idempotente, 472
 - limitado, 470, 471
 - normal, 471, 474, 475
 - positivo, 41, 50, 378
 - positivo sobre um cone, 51
 - unitário, 471, 474
- paradoxo do macaco, 108
- partição, 249, 449
 - de \mathbb{Z} , 58
 - geradora, 260, 262
 - menos fina, 147, 252
 - mensurável, 147

- trivial, 284
- partições
 - independentes, 250
- pilha simples, 172
- ponto
 - de densidade, 446
 - fixo
 - elítico, 128–131
 - elítico genérico, 128
 - hiperbólico, 129
 - não degenerado, 128, 130
 - heteroclínico, 490
 - homoclínico transversal, 129
 - recorrente, 7
 - super não errante, 63
- posto, 232, 234
- potenciais cohomólogos, 331, 395
- potencial, 302
- pré-órbita, 54, 364
- pressão, 301, 324, 325
- primeiro retorno, 22
- princípio
 - da distribuição de massa, 425
 - variacional, 332, 334
- probabilidade, 430
 - condicional, 145
 - de transição, 192
- produto
 - de medidas, 444, 445
 - enumerável, 445
 - finito, 444
 - interno, 465, 468
 - semi-direto, 53
- progressão aritmética, 58
- projeção, 472
 - estereográfica, 457
 - ortogonal, 67
- propriedade da lacuna espectral, 211, 412
- pseudo-órbita, 363
 - periódica, 363
- quantidade de informação de um caracter, 249
- quase todo ponto, 443
- quociente projetivo, 401
- raio espectral, 51, 316
- raiz de um sistema, 287
- ramo
 - contrativo, 362
 - inverso, 352, 362, 418
- razão
 - áurea, 9, 182
 - cruzada, 403
- relação de cohomologia, 331
- renormalização de Rauzy-Veech, 210
- representação espectral, 475
- repulsor, 416
 - conforme, 377, 417
- restrição normalizada, 159
- retângulo, 434
- reta estendida, 429, 437
- retorno
 - primeiro, 22
 - simultâneo, 29
- rotação, 16
 - irracional, 17
 - no círculo, 16
 - no toro, 18
 - racional, 17
- série
 - absolutamente somável, 468
 - de Fourier, 113
- seção transversal, 90
- semicontinuidade da entropia, 262
- sequência
 - aditiva, 78
 - admissível, 314
 - de correlações, 184
 - equidistribuída, 175
 - não lacunar, 29
 - separadora, 239
 - subaditiva, 78, 79
 - uniformemente integrável, 87
- sistema
 - aperiódico, 28, 262
 - com espectro de Lebesgue, 229, 284
 - com espectro discreto, 227
 - conservativo, 48, 122
 - de Kolmogorov, 281, 284
 - de spin, 332
 - enumeravelmente gerado, 232
 - ergódico, 95, 96
 - fracamente misturador, 187, 222
 - hamiltoniano, 123
 - integrável, 124
 - invertível em quase todo ponto, 244
 - minimal, 159, 161, 206
 - misturador, 184
 - totalmente dissipativo, 48
 - unicamente ergódico, 159
- soma
 - de coberturas, 302
 - de partições, 250
 - de Riemann, 442
 - de uma família de subespaços, 468
 - de uma família de vetores, 468
 - direta ortogonal, 469
 - orbital, 378
- sombreamento, 363
- subcobertura, 302, 431
- substituição, 173, 175, 312
 - de Cantor, 173
 - de Chacon, 173

- de Feigenbaum, 173
- de Fibonacci, 173, 312
- de Thue-Morse, 173
- primitiva, 173
- subvariedade, 458
 - riemanniana, 462
- suporte
 - de medida espectral, 472
 - de uma medida, 436
- supremo essencial, 466
- suspensão
 - de uma medida, 89
 - de uma transformação, 87
- tempo
 - de primeiro retorno, 5, 22, 90
 - de retorno, 87, 89
 - médio de retorno, 6
 - médio de visita, 65, 71
- teorema
 - da continuidade
 - inferior, 439
 - no vazio, 431
 - superior, 439
 - da convergência
 - dominada, 444
 - monótona, 444
 - da decomposição
 - de Hahn, 433
 - de Lebesgue, 448
 - dinâmica, 367
 - ergódica, 144
 - da desintegração, 148
 - da forma normal de Birkhoff, 128, 130
 - da representação espectral, 475
 - das classes monótonas, 432
 - de Anosov, 132
 - de aproximação, 432
 - de Banach-Alaoglu, 50, 470
 - de Banach-Mazur, 52
 - de Brin-Katok, 267
 - de continuidade absoluta, 134
 - de derivação de Lebesgue, 446
 - de Dirichlet, 492
 - de ergodicidade de Sinai, 140
 - de existência de medidas invariantes, 35
 - para fluxos, 48
 - de extensão de medidas, 431
 - de Friedman-Ornstein, 285
 - de Furstenberg, 162
 - de Furstenberg-Kesten, 85
 - de Gottschalk, 164
 - de Grünwald, 63
 - de Green-Tao, 60
 - de Haar, 166
 - de Halmos-von Neumann, 239
 - de Hindman, 164
 - de Jacobs, 288, 289
 - de Kač, 5
 - de Kakutani-Rokhlin, 28
 - de Keane, 206
 - de Kingman, 66, 79
 - de Kolmogorov-Arnold-Moser, 126, 127
 - de Kolmogorov-Sinai, 259
 - de Liouville, 20
 - de Livsič, 377, 397
 - de Lusin, 452, 454
 - de Masur-Veech, 207
 - de Mazur, 53
 - de Oseledets, 85
 - de Oxtoby-Ulam, 122
 - de Perron-Fröbenius, 194
 - de Prohorov, 42
 - de Radón-Nikodym, 447
 - de recorrência
 - de Birkhoff, 7, 8, 47
 - de Poincaré, 4, 7
 - para fluxos, 5
 - de recorrência múltipla
 - de Birkhoff, 29
 - de Poincaré, 29
 - de Riesz-Markov, 433, 455
 - de Rokhlin, 148, 157
 - de Ruelle, 377
 - de Sard, 461
 - de Schauder-Tychonoff, 45
 - de Shannon-McMillan-Breiman, 266
 - de Stone, 68
 - de Stone-Weierstrass, 456
 - de Szemerédi, 59, 61
 - de van der Waerden, 58, 60
 - de Vitali, 447
 - de Weyl, 176
 - de Whitney, 462
 - do espectro discreto, 239
 - do ponto fixo de Poincaré-Birkhoff, 129
 - ergódico
 - de Birkhoff, 66, 71, 72, 75
 - de Birkhoff para fluxos, 78
 - de von Neumann, 66, 68, 75
 - multiplicativo, 85
 - subaditivo, 66, 79
 - subaditivo para fluxos, 87
 - espectral, 474
 - normal de Borel, 105
 - Portmanteau, 37
 - topo de uma pilha, 172
 - topologia, 429
 - C^r , 458
 - definida por
 - bases de vizinhanças, 36
 - uma distância, 450
 - discreta, 108
 - fraca, 49, 470

- fraca*, 36, 49, 470
- mais forte, 37
- mais fraca, 37
- pontual, 44
- produto, 108, 446
- uniforme, 44
- topologias
 - equivalentes, 37
- toro de dimensão d , 18, 457
- torre, 27
 - de Kakutani-Rokhlin, 28
- transformação
 - h -expansiva, 324
 - conservativa, 19
 - de Dehn, 492
 - de Gauss, 13, 23
 - de Manneville-Pomeau, 26
 - de Poincaré, 88, 90
 - de primeiro retorno, 22, 88, 90
 - expansão decimal, 10
 - expansiva, 264, 312, 352, 363
 - expansora, 26, 359
 - do intervalo, 358
 - numa variedade, 350
 - induzida, 24
 - integrável, 127
 - intrinsecamente ergódica, 346
 - localmente invertível, 295
 - logística, 312
 - minimal, 18, 103
 - não degenerada, 127
 - simplética, 126
 - tempo 1, 5
 - topologicamente
 - exata, 356
 - fracamente misturadora, 224
 - misturadora, 186
 - transitiva, 108
- transformações
 - que comutam entre si, 29
 - topologicamente conjugadas, 304
 - topologicamente equivalentes, 304
- transição de fase, 332
- transitividade, 120
- translação
 - à direita, 166
 - à esquerda, 166
 - num grupo compacto, 308
- transporte de uma medida, 91
- transversalidade, 461, 464
- triângulo de Sierpinski, 426

- último teorema de Poincaré, 129
- unicidade ergódica, 159

- valor regular de uma aplicação, 461
- variável aleatória, 44

- variação total, 433
- variedade
 - de classe C^r , 457
 - de Grassmann, 458
 - diferenciável, 457
 - modelada em espaço de Banach, 457
 - riemanniana, 462
- vetor
 - de comprimentos, 204
 - de translações, 204
 - diofantino, 126
 - frequências, 125
 - racionalmente independente, 18, 205
- vetores ortogonais, 469
- vizinhança
 - de um conjunto, 36
 - de um ponto, 436
- volume associado a métrica riemanniana, 167
- volume zero, 461