Universidade Federal da Bahia - UFBA
Instituto de Matemática e Estatística - IME
Programa de Pós-Graduação em Matemática
Dissertação de Mestrado

A teoria do transporte ótimo aplicada ao estudo de sistemas de funções iteradas

Agábio Brasil dos Santos

Salvador-Bahia

A teoria do transporte ótimo aplicada ao estudo de sistemas de funções iteradas

Agábio Brasil dos Santos

Dissertação de Mestrado apresentada ao Colegiado da Pós-Graduação em Matemática da Universidade Federal da Bahia como requisito parcial para obtenção do título de Mestre em Matemática.

Orientador: Prof. Dr. Edgar Matias da Silva

Salvador-Bahia

Novembro de 2025

Ficha catalográfica elaborada pela Biblioteca Universitária de Ciências e Tecnologias Prof. Omar Catunda, SIBI – UFBA.

S237 Santos, Agábio Brasil dos

A teoria do transporte ótimo aplicada ao estudo de sistemas de funções iteradas / Agábio Brasil dos Santos. – 2025.

54 f.

Orientador: Prof. Dr. Edgar Matias da Silva.

Dissertação (mestrado) – Universidade Federal da Bahia, Instituto de Matemática e Estatística, 2025.

1. Transporte ótimo. 2. Sistemas de funções iteradas. 3. Medidas estacionárias. 4. Distância de Wasserstein. 5. Skew-products. I. Silva, Edgar Matias da. II. Universidade Federal da Bahia. III. Título.

CDU:517.98

A teoria do transporte ótimo aplicada ao estudo de sistemas de funções iteradas.

Agábio Brasil dos Santos

Dissertação apresentada ao Colegiado do Curso de Pós-graduação em Matemática da Universidade Federal da Bahia, como requisito parcial para obtenção do Título de Mestre em Matemática.

Banca examinadora

Edgar Matias da Silva.

Prof. Dr. Edgar Matias da Silva
(Orientador - UFBA)

Lizana

Profª. Dra. Cristina Lizana Araneda
(Interno à Instituição - UFBA)

Diego Daltro Conceição

Prof. Dr. Diego Daltro Conceição
(Externo à Instituição – UFRB)

Agradecimentos

Primeiramente, gostaria de agradecer a Deus pelo sopro de vida de cada dia.

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro, sem o qual não seria possível realizar o presente estudo.

Gostaria de agradecer a minha mãe, que me deu condições e encorajamento para cursar o mestrado, sem a qual, talvez não teria conseguido chegar onde cheguei, muito obrigado minha mãe querida. Também não teria como deixar de agradecer a minha noiva, que desde sempre esteve ao meu lado, me incentivando, encorajando, me dando forças, onde, em seus braços, pude encontrar o que precisava para seguir em frente no curso nos momentos mais difíceis, muito obrigado meu amor.

Gostaria de agradecer a meus amigos por tudo, realmente, por todas as provas que passamos juntos, por todos os trabalhos que estudamos, por todas as dificuldades que me ajudaram a passar, mas também agradeço pelos momentos de leveza, os quais tornavam o ambiente de estudo mais agradável e mais acolhedor para o dia a dia de estudo em nosso instituto.

Agradeço a minha família, àqueles que me ajudaram e me apoiaram de coração, muito obrigado por tudo.

Agradeço a todos os meus professores por tudo que ensinaram, pela paciência e pelos momentos de acolhimento.

"O que sou eu aos olhos da maioria das pessoas? Uma não entidade, ou um homem excêntrico e desagradável - alguém que não tem e nunca terá posição na vida, em suma, o menor dos menores. Muito bem, mesmo que isso fosse verdade, devo querer que o meu trabalho mostre o que vai no coração de um homem excêntrico e desse joão-ninguém"

Vincent Van Gogh

Resumo

Este trabalho consiste em um estudo sobre Sistemas de Funções Iteradas que contraem em média, utilizando conceitos e resultados da Teoria do Transporte Ótimo para essa análise. Para fundamentar a discussão, foi necessário revisitar noções como espaços métricos completos, topologia fraca-*, além de conceitos da Teoria da Medida, Teoria Ergódica e Teoria da Probabilidade. Dentro desta última, são exploradas as noções de cadeias de Markov homogêneas e de medidas estacionárias associadas a essas cadeias. Com base nesses conceitos preliminares, define-se o acoplamento ótimo, demonstrando-se que, sob certas condições, a existência de tal acoplamento é sempre garantida. Introduz-se também a distância de Wasserstein, a qual desempenha um papel essencial nos resultados subsequentes. Munido das ferramentas da Teoria do Transporte Ótimo, o trabalho investiga as medidas estacionárias de Sistemas de Funções Iteradas que contraem em média, provando que, nesse contexto, há existência e unicidade de uma medida estacionária, além de se obter estimativas para seus momentos de ordem q. Por fim, os conceitos desenvolvidos são aplicados ao estudo de skew-products que possuem fibras contrativas, mostrando-se que tais sistemas admitem uma única "medida estacionária" com suporte limitado.

Palavras-chave: Medidas estacionárias; Distância de Wasserstein; Skew-products.

Abstract

This work consists of a study of Iterated Function Systems that contract on average, using concepts and results from Optimal Transport Theory for this analysis. To support the discussion, it was necessary to revisit notions such as complete metric spaces, weak topology, as well as concepts from Measure Theory, Ergodic Theory, and Probability Theory. Within the latter, the notions of homogeneous Markov chains and the stationary measures associated with these chains are explored. Based on these preliminary concepts, optimal coupling is defined, demonstrating that, under certain conditions, the existence of such coupling is always guaranteed. The Wasserstein distance is also introduced, which plays an essential role in the subsequent results. Equipped with the tools of Optimal Transport Theory, this work investigates the stationary measures of Iterated Function Systems that contract on average, proving that, in this context, there is existence and uniqueness of a stationary measure, in addition to obtaining estimates for its moments of order q. Finally, the developed concepts are applied to the study of skew-products that have contractive fibers, showing that such systems admit a single "stationary measure" with limited support.

Keywords: Stationary measures; Wasserstein distance; Skew-products.

Sumário

Introdução

Sejam $f_1, \ldots, f_k \colon A \to A$ funções contínuas definidas em um espaço métrico completo A e $(\mathcal{X}_n)$ uma sequência de variáveis aleatórias independentes e identicamente distribuídas tomando valores em $\{1, \ldots, k\}$. Cada estado inicial $x \in A$ gera a órbita aleatória

$$\mathcal{Z}_n(x) = f_{\mathcal{X}_n} \circ \cdots \circ f_{\mathcal{X}_0}(x).$$

A sequência $\mathcal{Z}_n$ é uma cadeia de Markov homogênea.

Quando as funções $f_i's$ são contrações, pode-se mostrar que a distribuição de $\mathcal{Z}_n(x)$ converge para a única medida estacionária atratora μ suportada no atrator de Hutchinson, isto é, o único compacto K satisfazendo $K = \bigcup_{i=1}^k f_i(K)$. Mais ainda, vale o seguinte teorema ergódico: para $x \in A$, com probabilidade 1, tem-se

$$\frac{1}{n} \sum \delta_{\mathcal{Z}_n(x)} \to \mu.$$

Em particular, o conjunto $\{\mathcal{Z}_n(x), \mathcal{Z}_{n+1}(x), \ldots\}$ converge para o atrator de Hutchinson na distância de Hausdorff. Esse fato é difundido na literatura dizendo-se que o *jogo do caos* desenha o atrator.

O estudo desses objetos possuem aplicações em diversos ramos da ciência, tais como dinâmica hiperbólica, compressão de imagens, convoluções de Bernoulli, folheações holomorfas, teoria da dimensão de fractais e formalismo termodinâmico.

Uma pergunta natural é entender em que condições pode-se garantir a existência de uma medida estacionária atratora quando pelo menos uma das funções não é mais uma contração. Uma das primeiras classes consideradas é a classe de sistemas de funções iteradas que contraem na média. Essa classe está amplamente estudada na literatura, mas uma abordagem via a Teoria do Transporte Ótimo apareceu pela primeira vez em [10], Kloeckner mostrou como a existência do transporte ótimo para a função custo definida a partir da métrica pode ser usada para obter (e melhorar) uma série de resultados.

A Teoria do Transporte Ótimo foi concebida em problemas de otimização, os quais, foram trazidos por Gaspard Monge e, posteriormente, revisitados por Leonid Vitalyevich Kantorovich. O problema de otimização, que fora denominado como problema de Monge-Kantorovich, busca encontrar uma medida de probabilidade que otimiza a esperança de uma função custo dada. Tal medida representa a distribuição de massa no problema de otimização. Apesar de ter sido concebida em uma esfera mais limitada, nos tempos atuais, a Teoria do Transporte Ótimo possui aplicações em diversas áreas do conhecimento; no contexto dos Sistemas Dinâmicos uma utilização é para estudos relacionados a sistemas de funções iteradas, o qual é o objeto de pesquisa deste trabalho. Desta forma, busca-se entender como utilizar a Teoria do Transporte Ótimo para estudar sistemas de funções iteradas que contraem na média.

Estrutura-se o presente trabalho da seguinte forma. No Capítulo 1 traz-se definições e resultados que serão auxiliadores para os principais objetos a serem estudados, tais como conceitos elementares de espaços métricos e topologia, Teoria da Medida e Teoria da Probabilidade, onde vê-se o que é uma cadeia de Markov e também como define-se uma medida estacionária para uma cadeia de Markov homogênea.

No Capítulo 2 discorre-se sobre a Teoria do Transporte Ótimo, onde faz-se uma contextualização da área através de uma breve história do seu surgimento, além de uma explicação do problema de Monge-Kantorovich; define-se o que é um acoplamento, demonstra-se que no espaço das medidas de probabilidade sobre um espaço métrico, separável e completo pode-se garantir um acoplamento ótimo, além disso, prova-se que nessas condições, o espaço das medidas de probabilidade é métrico, separável e completo com a distância de Wasserstein.

Em sequência, no Capítulo 3 é feito uma explanação do que é um Sistema de Funções Iteradas, além de definir-se em que condições tal sistema é dito que contrai na média, posteriormente é abordado alguns resultados inerentes a esses sistemas, em seguida define-se o operador de Markov, com o qual será provado que, dado um sistema de funções iteradas que contrai na média existe uma única medida estacionária, além disso se tem uma estimativa para o q-ésimo momento da medida, desde que tenha-se certas condições sobre o sistema. Tal resultado é uma versão do trazido por [3], onde são utilizadas as ideias de [10]. Posteriormente é utilizado uma adaptação da distância de Wasserstein para provar que um skew-product, em certas condições, que contraem as fibras possui uma única medida estacionária com suporte limitado.

Capítulo 1

Conceitos preliminares

Neste capítulo serão abordados os conceitos iniciais do estudo principal do presente trabalho, onde será tratado as definições para a estruturação dos espaços em que se trabalhará nos próximos capítulos. Para tal, serão utilizados como referenciais teóricos [12], [14], [4], [17], [13], [7], [18], [1] e [19].

1.1 Elementos de Espaços Métricos e Topologia

Inicialmente, serão abordados os conceitos elementares de espaços métricos e topologia, os quais utiliza-se como base para o que será feito posteriormente.

Definição 1.1. *Dado um conjunto A, pode-se definir como* ***métrica*** *uma função $d : A\times A \to \mathbb{R}_+$ que associa a cada par ordenado de elementos $x, y \in A$ um número real $d(x, y)$, chamado de distância de x a y, de modo que sejam satisfeitas as seguintes propriedades para quaisquer $x, y, z \in A$:*

1. *$d(x, y) = 0$ se, e somente se, $x = y$*
2. *$d(x, y) = d(y, x)$;*
3. *$d(x, z) \leq d(x, y) + d(y, z)$.*

No espaço $\mathbb{R}^n$ tem-se a métrica Euclidiana. Dados $x, y \in \mathbb{R}^d$, $d : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ com $x = (x_1, x_2, \ldots, x_d)$ e $y = (y_1, y_2, \ldots, y_d)$, define-se

$$d(x, y) = \sqrt{\sum_{i=1}^{d}(x_i - y_i)^2}.$$

Assim, dado uma métrica d em um conjunto A, define-se como **espaço métrico** a dupla (A, d). Um exemplo clássico de espaço métrico é a dupla $(\mathbb{R}^n, d)$, onde d é a métrica euclidiana, o qual chama-se espaço euclidiano. Dado um conjunto, pode-se definir mais de uma métrica no espaço, dessa forma, pode-se ser perguntado se não há uma “equivalência” entre essas métricas, ou seja, ao trabalhar-se com uma, seria similarmente “a mesma coisa” que trabalhar com a outra, o que significaria que topologicamente se estaria trabalhando com os mesmos objetos, embora a “geometria” não seja a mesma. Nesse sentido, dez-se que duas métricas d_1 e d_2 definidas sobre um conjunto A, são **métricas equivalentes** se existe $C > 0$ tal que para todo $x, y \in A$, tem-se $C^{-1} \cdot d_2(x, y) \leq d_1(x, y) \leq C \cdot d_2(x, y)$.

Um espaço métrico que será recorrentemente utilizado neste trabalho será o chamado espaço das funções contínuas. Dado um espaço métrico A, denota-se como o conjunto de todas as funções contínuas $f : A \to A$ como $C(A)$. No conjunto $C(A)$ pode-se definir a métrica

$$d(f, g) = \sup\{|f(x) - g(x)|; x \in M\},$$

para qualquer $f, g \in C(A)$. Desta maneira o par $(C(A), d)$ é um espaço métrico.

Pode-se, assim como tem-se a distância entre dois pontos, definir a distância entre um ponto e um conjunto como sendo, dado (A, d) um espaço métrico e $X \subset A$ um subconjunto de A qualquer, a **distância do ponto y ao conjunto** X como o número real $d(y, X) = \inf\limits_{x \in X} d(y, x)$.

Agora, serão abordados alguns conceitos topológicos. Seja X um subconjunto de um espaço métrico A. Um ponto $a \in X$ diz-se um **ponto interior** a X quando é centro de uma bola aberta contida em X, ou seja, quando existe $r > 0$ tal que $d(x, a) < r \Rightarrow x \in X$. Denota-se como $int(X)$ o conjunto de todos os pontos interiores de X. Um subconjunto O de um espaço métrico A diz-se **aberto** em A quando todos os seus pontos são interiores, isto é, $int(O) = O$.

Será definido então um dos conceitos que é, historicamente, muito importante para a abstração de noções como continuidade, a Topologia. **Topologia** é uma família τ de subconjuntos de um conjunto X tal que:

(i) $\emptyset, X \in \tau$;

(ii) Se $A_1, \ldots, A_n \in \tau$ então $A_1 \cap \cdots \cap A_n \in \tau$;

(iii) Dada uma coleção arbitrária $(A_\lambda)_{\lambda \in L}$ com $A_\lambda \in \tau$ para cada $\lambda \in L$, tem-se $\bigcup\limits_{\lambda \in L} A_\lambda \in \tau$.

Chama-se o par (X, τ) de **espaço topológico**.

A exemplo, seja τ a coleção de todos os subconjuntos abertos de $\mathbb{R}^d$. τ é uma topologia em $\mathbb{R}^d$ e é chamada de topologia usual de $\mathbb{R}^d$, logo o par $(\mathbb{R}^d, \tau)$ é um espaço topológico. Assim, pode-se definir também como métricas equivalentes as métricas que induzem a mesma topologia.

Uma coleção de conjuntos abertos $B \subset \tau$, onde τ é uma topologia, com a propriedade de que qualquer elemento de τ pode ser escrito como uma união de elementos de B é definida como **base da topologia**. Um exemplo que pode ser observado é, considerando $X = \mathbb{R}^d$, o qual sabe-se que é um espaço topológico com uma topologia induzida pelas bolas abertas definidas pela métrica no espaço, considere o conjunto de todas as bolas de raio racional centradas em pontos com coordenadas racionais. Desta forma define-se uma base enumerável para a topologia como sendo essa coleção.

Uma sequência (x_n) num espaço métrico A chama-se uma **sequência de Cauchy** quando, para todo $\varepsilon > 0$ dado, existe $n_0 \in \mathbb{N}$ tal que $m, n > n_0 \Rightarrow d(x_m, x_n) < \varepsilon$. Define-se como **espaço métrico completo**, um espaço onde todas as sequências de Cauchy são convergentes. Neste trabalho, será considerado como axiomático que os espaços $\mathbb{R}^n$ são completos.

Tem-se, um teorema bastante utilizado quando estuda-se espaço métricos e contrações, o Teorema do ponto fixo de Banach:

Teorema 1.1. *(Ponto fixo de Banach para contrações). Se A é um espaço métrico completo, toda contração $f : A \to A$ possui um único ponto fixo atrator em A. Mais precisamente, se escolhermos um ponto qualquer $x_0 \in M$ e pusermos $x_1 = f(x_0), x_2 = f(x_1), \ldots, x_{n+1} = f(x_n), \ldots$ a sequência (x_n) converge em A e $a = \lim x_n$ é o único ponto fixo de f.*

Demonstração. Suponha que (x_n) convirja para um ponto $a \in A$. Desta maneira, como a função f é contínua, tem-se $f(a) = f(\lim x_n) = \lim f(x_n) = \lim x_{n+1} = a$, logo a é ponto fixo de f. Será mostrado agora que f não admite dois pontos fixos distintos. De fato, se $f(a) = a$, $f(b) = b$, e vale $d(f(x), f(y)) \leq c \cdot d(x, y)$, com $0 \leq c < 1$, para $x, y \in A$ quaisquer, então

$$d(a, b) = d(f(a), f(b)) \leq c \cdot d(a, b),$$

onde $(1 - c) \cdot d(a, b) \leq 0$. Como $1 - c > 0$, concluí-se que $d(a, b) = 0$, ou seja, $a = b$.

Agora, será mostrado que a sequência é Cauchy. Tem-se que,

$$d(x_1, x_2) = d(f(x_0), f(x_1)) \leq c \cdot d(x_0, x_1),$$

$$d(x_2, x_3) = d(f(x_1), f(x_2)) \leq c \cdot d(x_1, x_2) \leq c^2 \cdot d(x_0, x_1)$$

e, em geral, tem-se

$$d(x_n, x_{n+1}) \leq c^n \cdot d(x_0, x_1)$$

para todo $n \in \mathbb{N}$. Segue-se que, para $n, p \in \mathbb{N}$ quaisquer:

$$\begin{aligned} d(x_n, x_{n+p}) &\leq d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}) + \cdots + d(x_{n+p-1}, x_{n+p}) \\ &\leq [c^n + c^{n+1} + \cdots + c^{n+p-1}] \cdot d(x_0, x_1) \\ &= c^n[1 + c + \cdots + c^{p-1}] \cdot d(x_0, x_1) \\ &\leq \frac{c^n}{1-c} \cdot d(x_0, x_1). \end{aligned}$$

Como $\lim\limits_{n\to\infty} c^n = 0$, concluí-se que (x_n) é uma sequência de Cauchy em A, concluindo a demonstração. ■

1.2 Elementos de Teoria da Medida e Teoria Ergódica

Primeiramente, serão tratados dos conceitos iniciais referentes a Teoria da Medida, dando ênfase nas medidas de probabilidade. Antes de definir-se o que são medidas de probabilidade, precisa-se estar bem definido alguns outros conceitos.

Começa-se buscando definir e estruturar o que chama-se de σ-álgebras. Dado um conjunto X define-se como **σ-álgebra** uma família $\mathcal{F}$ de subconjuntos de X tal que:

(i) $X \in \mathcal{F}$;

(ii) se $A \in \mathcal{F}$, então $X \backslash A \in \mathcal{F}$;

(iii) Seja (A_n) uma sequência de conjuntos tais que $A_n \in \mathcal{F}$, $n = 1, 2, \ldots$, então $\bigcup\limits_n A_n \in \mathcal{F}$.

Observação 1.1. *Pode-se observar que, se $\mathcal{F}$ é uma σ-álgebra, então, por i) e ii), têm-se que $\emptyset \in \mathcal{F}$ diretamente. Além disso, se (A_n) é uma sequência de conjuntos tais que $A_n \in \mathcal{F}$, $n = 1, 2, \ldots$, então, por ii) e iii), $\bigcap\limits_n (X \backslash A_n) \in \mathcal{F}$.*

Assim, sendo X um conjunto e $\mathcal{F}$ uma σ-álgebra de subconjuntos de X, então chama-se $(X, \mathcal{F})$ de **espaço mensurável**. Os elementos de $\mathcal{F}$ são denominados de mensuráveis. Seja X um conjunto. Pode-se observar que o conjunto das partes 2^X é uma σ-álgebra de X.

Define-se como **σ-álgebra gerada** por uma família S de subconjuntos de X como sendo a menor σ-álgebra tal que contenha S. Denota-se σ-álgebra por $\sigma(S)$.

Um comentário que pode-se ser feito neste momento é a diferenciação entre σ-álgebras e topologias. Enquanto uma topologia pede que interseções finitas pertençam a τ, uma σ-álgebra não possui essa restrição, além disso, em uma topologia é garantido que uma união arbitrária pertence a τ, enquanto que em uma σ-álgebra as união precisam ser enumeráveis. Tal comparação é relevante para percebermos que topologias em geral são diferentes de σ-álgebras em essência.

Dados esses comentários, define-se então uma das sigma álgebras mais relevantes para teoria da medida, a σ-álgebra de Borel.

Definição 1.2. *A* ***σ-álgebra de Borel*** *(ou σ-álgebra boreliana) de um espaço topológico (X, τ) é a σ-álgebra $\sigma(\tau)$ gerada pela topologia τ. Neste caso, os conjuntos mensuráveis recebem o nome de borelianos.*

Se $S = \{(-\infty, x] \subset \mathbb{R}; x \in \mathbb{R}\}$, então também é possível concluir que $\sigma(S) = B(\mathbb{R})$.

Dados os conceitos já apresentados, pode-se definir então o que é uma medida.

Definição 1.3. *Define-se como* ***medida*** *em $(X, \mathcal{F})$ uma função $\mu : \mathcal{F} \to \mathbb{R}$ tal que:*

(1) $\mu(\emptyset) = 0$*;*

(2) $\mu(A) \geq 0$, $\forall A \in \mathcal{F}$*;*

(3) Se (A_n) é uma sequência enumerável de conjuntos disjuntos com $A_n \in \mathcal{F}$, $\forall n \in \mathbb{N}$, então

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n).$$

Se $\mu(X) < \infty$, então diz-se que μ é finita. Se $\mu(X) = 1$ diz-se que μ é uma ***medida de probabilidade****.*

Denota-se por $\mathcal{P}(X)$ o conjunto de todas as medidas de probabilidade definidas no espaço mensurável $(X, \mathcal{F})$. Chama-se a tripla $(X, \mathcal{F}, \mu)$ de **espaço de medida**. Se μ é uma medida de probabilidade, diz-se que $(X, \mathcal{F}, \mu)$ é uma **espaço de probabilidade**.

Exemplo 1.1. *Seja X um conjunto e será considerado a σ-álgebra $B = 2^X$. Fixado $p \in X$, seja a função $\delta_p : 2^X \to [0, +\infty]$ definida por:*

$$\delta_p(A) = \begin{cases} 1, & se\ p \in A \\ 0, & se\ p \in A. \end{cases}$$

Esta medida δ_p é usualmente designada medida de Dirac, ou delta de Dirac, no ponto p.

Proposição 1.1. *Se $(X, \mathcal{F}, \mu)$ é um espaço de medida e $A \in \mathcal{F}$, então a função $\mu_{|A}$ dada por $\mu_{|A}(B) = \mu(A \cap B)$ também é uma medida.*

Demonstração. Para provar que $\mu_{|A}$ é uma medida, precisa-se mostrar os itens (1), (2) e (3) da definição de medida, o que será feito a seguir

(1) Observa-se que é trivial, pois

$$\mu_{|A}(\emptyset) = \mu(A \cap \emptyset) = \mu(\emptyset) = 0.$$

(2) Seja $B \in \mathcal{F}$, logo $X \backslash B = B^c \in \mathcal{F}$, assim $A^c \cup B^c \in \mathcal{F}$, o que implica que $(A^c \cup B^c)^c = A \cap B \in \mathcal{F}$, desta forma

$$\mu_{|A}(B) = \mu(A \cap B) \geq 0.$$

(3) Seja (B_n) é uma sequência enumerável de conjuntos disjuntos com $B_n \in \mathcal{F}$, $\forall n \in \mathbb{N}$, logo

$$\mu_{|A}\left(\bigcup_n B_n\right) = \mu\left(A \cap \left(\bigcup_n B_n\right)\right) = \mu\left(\bigcup_n A \cap B_n\right).$$

Como (B_n) é uma sequência enumerável de conjuntos disjuntos, então $(A \cap B_n)_{n \in \mathbb{N}}$ será uma sequência enumerável de conjuntos disjuntos, desta forma

$$\mu\left(\bigcup_n A \cap B_n\right) = \sum_n \mu(A \cap B_n) = \sum_n \mu_{|A}(B_n).$$

■

Aqui, será tomado como axiomático a existência e unicidade de uma medida em $(\mathbb{R}, B(\mathbb{R}))$ tal que $m((a, b]) = b - a$. Tal medida é a que denomina-se **medida de Lebesgue**. Também será considerado como axiomático a existência da medida de Lebesgue para $\mathbb{R}^n$.

De forma equivalente as funções contínuas no estudo feito para os espaços topológicos, as funções mensuráveis adentram na teoria da medida com um papel fundamentalmente similar, o de preservar, agora, conjuntos mensuráveis. Desta forma, dados espaços mensuráveis $(X, \mathcal{B})$ e $(Y, \mathcal{C})$, diz-se que uma aplicação $f : X \to Y$ é **mensurável** se $f^{-1}(C) \in \mathcal{B}$ para todo $C \in \mathcal{C}$.

Exemplo 1.2. *Uma função $f : X \to [-\infty, \infty]$ é mensurável se, e somente se, $f^{-1}((c, +\infty])$ pertence a $\mathcal{B}$ para todo $c \in \mathbb{R}$. Pode-se notar que essa condição é suficiente uma vez que a σ-álgebra gerada pelos intervalos semi-infinitos é a sigma álgebra de Borel.*

Exemplo 1.3. *Se X é um espaço topológico e $\mathcal{B}$ é a sua σ-álgebra de Borel, então toda função contínua $f : X \to \mathbb{R}$ é mensurável.*

De fato, continuidade significa que a pré-imagem de qualquer aberto de $\mathbb{R}$ é um aberto de X e, portanto, está em $\mathcal{B}$. Como os abertos geram a σ-álgebra de Borel de $\mathbb{R}$, segue que a pré-imagem de qualquer boreliano da reta também está em $\mathcal{B}$.

Exemplo 1.4. *Dado um conjunto $A \subset X$ define-se a função característica $\mathcal{X}_A : X \to \mathbb{R}$ de A por:*

$$\mathcal{X}_A(x) = \begin{cases} 1, & \textit{se } x \in A; \\ 0, & \textit{caso contrário.} \end{cases}$$

Observa-se que a função $\mathcal{X}_A$ é mensurável se, e somente se, A é um subconjunto mensurável.

Diz-se que uma função $s : X \to \mathbb{R}$ é **simples** se existem constantes $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$ e conjuntos mensuráveis $A_1, \ldots, A_k \in \mathcal{B}$ disjuntos dois-a-dois tais que

$$s = \sum_{j=1}^{k} \alpha_k \, \mathcal{X}_{A_k},$$

onde $\mathcal{X}_A$ é a função característica do conjunto A.

Neste momento, será definido a **integral de Lebesgue**, a construção será feita por etapas. Seja $s = \displaystyle\sum_{j=1}^{k} \alpha_k \mathcal{X}_{A_k}$ uma função simples. Então a integral de s em relação à medida μ é dada por:

$$\int s \, d\mu = \sum_{j=1}^{k} \alpha_k \mu(A_k).$$

Em seguida, seja $f : X \to [0, \infty]$ uma função mensurável não negativa. Então

$$\int f \, d\mu = \limsup_n \int s_n \, d\mu,$$

onde $(s_n)_{n \geq 1}$ é uma sequência de funções simples limitadas por f. Sendo $f^+(x) = \max\{f(x), 0\}$ e $f^-(x) = \max\{-f(x), 0\}$ (as quais pode-se provar que são mensuráveis), tem-se então que, sendo $f : X \to [-\infty, +\infty]$ uma função mensurável, então,

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu,$$

desde que pelo menos uma das integrais do lado direito seja finita. Diz-se que uma função $f : X \to [-\infty, +\infty]$ é **integrável** se for mensurável e a sua integral for um número real. Denota-se o conjunto das funções integráveis por $L^1(\mu)$. Dada uma função mensurável $f : X \to [-\infty, \infty]$ e um conjunto mensurável E define-se a integral de f sobre E por

$$\int_E f \, d\mu = \int f \cdot \mathcal{X}_E \, d\mu.$$

Definição 1.4. *Diz-se que uma propriedade é válida μ-**quase certamente** (μ-q.c.), se é válida em todo o espaço X exceto, possivelmente, num conjunto de medida nula. À exemplo, diz-se 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 $x \in X \backslash N$.*

Sejam $(X, \mathcal{F}, \mu)$ um espaço de medida, $(Y, \mathcal{S})$ um espaço mensurável e $f : X \to Y$ uma função mensurável, a medida $\nu : \mathcal{S} \to \mathbb{R}$ definida como $\nu(A) = \mu(f^{-1}(A)) = f_*\mu(A)$, $\forall A \in \mathcal{S}$, é uma medida em Y, a qual é chamada de **medida imagem**. Neste caso, tem-se que a integral $\int g\, d\nu = \int g\, d(\mu \circ f^{-1}) = \int g \circ f\, d\mu$.

Os próximos resultados são considerados como *teoremas de convergência*, os quais trazem condições suficientes para se garantir a convergência, ou estimativas, de integrais de funções.

Teorema 1.2. *(Convergência monótona). Seja $f_n : X \to [-\infty, +\infty]$ uma sequência monótona de funções mensuráveis não negativas e seja f a função definida por $f(x) = \lim_n f_n(x)$. Então*

$$\lim_n \int f_n d\mu = \int f(x) d\mu.$$

Demonstração. Nota-se que a função f é mensurável. Como $f_n \leq f_{n+1} \leq f$, segue que

$$\int f_n d\mu \leq \int f_{n+1} d\mu \leq \int f\, d\mu$$

para todo $n \in \mathbb{N}$. Portanto, tem-se

$$\lim \int f_n\, d\mu \leq \int f\, d\mu.$$

Para estabelecer a desigualdade oposta, seja α um número real que satisfaz $0 < \alpha < 1$ e seja φ uma função mensurável simples que satisfaz $0 \leq \varphi \leq f$. Seja $A_n = \{x \in X : f_n(x) \geq \alpha\varphi(x)\}$ de modo que $A_n \in \mathcal{F}$, $A_n \subseteq A_{n+1}$, e $X = \bigcup A_n$. Assim

$$\int_{A_n} \alpha\varphi\, d\mu \leq \int_{A_n} f_n\, d\mu \leq \int f_n\, d\mu$$

Como a sequência (A_n) é monótona crescente e tem união X, segue-se que

$$\int \varphi\, d\mu = \lim \int_{A_n} \varphi\, d\mu.$$

Portanto, ao tomar o limite em relação a n, obtém-se

$$\alpha \int \varphi\, d\mu \leq \lim \int f_n\, d\mu.$$

Como isso vale para todo α com $0 < \alpha < 1$, infere-se que

$$\int \varphi\, d\mu \leq \lim \int f_n\, d\mu.$$

e como φ é uma função simples arbitrária no espaço que satisfaz $0 \leq \varphi \leq f$, concluí-se que

$$\int f\, d\mu = \sup_\varphi \int \varphi\, d\mu \leq \lim_n \int f_n\, d\mu$$

Ao combinar isso com a desigualdade oposta, obtém-se o que queríamos. ■

Teorema 1.3. *(Lema de Fatou). Seja $f_n : X \to [0, +\infty]$ uma sequência de funções mensuráveis não negativas. Então, a função $f : X \to [-\infty, +\infty]$ definida por $f(x) = \liminf_n f_n(x)$ é integrável e vale*

$$\int \liminf_n f_n(x) d\mu \leq \liminf_n \int f_n d\mu.$$

Demonstração. Seja $g_m = \inf\{f_m, f_{m+1}, \ldots\}$ de modo que $g_m \leq f_n$ sempre que $m \leq n$. Portanto,

$$\int g_m \, d\mu \leq \int f_n \, d\mu, \ m \leq n,$$

de modo que

$$\int g_m \, d\mu \leq \liminf_n \int f_n \, d\mu.$$

Como a sequência (g_m) é crescente e converge para $\liminf f_n$, o Teorema da Convergência Monótona implica que

$$\begin{aligned} \int (\liminf_n f_n) \, d\mu &= \lim \int g_m \, d\mu \\ &\leq \liminf_n \int f_n d\mu. \end{aligned}$$

■

Proposição 1.2. *(Convergência dominada). Seja $f_n : X \to \mathbb{R}$ uma sequência de funções mensuráveis e suponha que existe uma função integrável g tal que $|f_n(x)| \leq |g(x)|$ para μ-quase todo $x \in X$. Suponha também que a sequência $(f_n)_n$ converge em μ-quase todo ponto para uma função f. Então f é integrável e vale:*

$$\lim_n \int f_n d\mu = \int f d\mu.$$

Demonstração. Redefinindo as funções f_n, f em um conjunto de medida 0, pode-se assumir que a convergência ocorre em todo X. Segue-se que f é integrável. Como $g + f_n \geq 0$, pode-se aplicar o Lema de Fatou para obter

$$\begin{aligned} \int g \, d\mu + \int f \, d\mu &= \int (g + f) d\mu \\ &\leq \liminf \int (g + f_n) d\mu \\ &= \liminf \left(\int g \, d\mu + \int f_n \, d\mu \right) \\ &= \int g \, d\mu + \liminf \int f_n \, d\mu. \end{aligned}$$

Portanto, segue-se que

$$\int f \, d\mu \leq \liminf \int f_n d\mu.$$

Como $g - f_n \geq 0$, outra aplicação do Lema de Fatou produz

$$\begin{aligned}\int g\,d\mu - \int f\,d\mu &= \int (g-f)\,d\mu \\ &\leq \liminf \int (g - f_n)\,d\mu \\ &\leq \int g\,d\mu - \limsup \int f_n\,d\mu,\end{aligned}$$

do qual se segue que

$$\limsup \int f_n\,d\mu \leq \int f\,d\mu.$$

Infere-se então que

$$\int f\,d\mu = \lim \int f_n\,d\mu.$$

■

Neste momento, serão trazidos algumas desigualdades fundamentais.

Proposição 1.3. *(Desigualdade de Hölder). Sejam $p, q > 1$ tal que $\frac{1}{p} + \frac{1}{q} = 1$, então*

$$\int |f \cdot g|\,d\mu \leq \left(\int |f|^p\,d\mu\right)^{\frac{1}{p}} \cdot \left(\int |g|^q\,d\mu\right)^{\frac{1}{q}},$$

ou, definido $||f||_q := (\int |f|^q\,d\mu)^{\frac{1}{q}}$,

$$||f \cdot g||_1 \leq ||f||_p \cdot ||g||_q.$$

Demonstração. Supõem-se que $||f||_p < \infty$ e $||g||_q < \infty$.

Seja α um número real que satisfaz $0 < \alpha < 1$ e considere a função φ definida para $t \geq 0$ por

$$\varphi(t) = \alpha t - t^{\alpha}.$$

É fácil verificar que $\varphi'(t) < 0$ para $0 < t < 1$ e $\varphi'(t) > 0$ para $t > 1$.

Decorre do Teorema do Valor Médio que $\varphi(t) \geq \varphi(1)$ e que $\varphi(t) = \varphi(1)$ se, e somente se, $t = 1$. Portanto, tem-se

$$t^{\alpha} \leq \alpha \cdot t + (1 - \alpha),\ t \geq 0.$$

Se a, b forem não negativos, fazendo $t = a/b$ e multiplicando por b, obtém-se a desigualdade

$$a^{\alpha} b^{(1-\alpha)} \leq a \cdot \alpha + (1 - \alpha) b,$$

onde a igualdade é válida se e somente se $a = b$. Agora, sejam p e q satisfazendo $1 < p < \infty$ e $\frac{1}{p} + \frac{1}{q} = 1$ e tome $\alpha = 1/p$. Segue-se que, se A, B são quaisquer números reais não negativos, então

$$A \cdot B \leq \frac{A^p}{p} + \frac{B^q}{q}$$

e que a igualdade se mantém se, e somente se, $A^p = B^q$.

Suponha que $||f||_p \neq 0$ e $||g||_q \neq 0$. O produto $f \cdot g$ é mensurável e, sendo

$$A = \frac{|f(x)|}{||f||_p} \text{ e } B = \frac{|g(x)|}{||g||_q}$$

implica que

$$\frac{|f(x)g(x)|}{||f||_p||g||_q} \leq \frac{|f(x)|^p}{p||f||_p^p} + \frac{|g(x)|^q}{q||g||_q^q}$$

Como ambos os termos à direita são integráveis, segue-se que $f \cdot g$ é integrável. Além disso, ao integrar, obtém-se

$$\frac{||f \cdot g||_1}{||f||_p||g||_q} \leq \frac{1}{p} + \frac{1}{q} = 1.$$

■

Um caso particular da desigualdade de Hölder é a desigualdade de Cauchy-Schwarz, onde $p = q = 2$.

Proposição 1.4. *(Desigualdade de Minkowski). Seja $p \geq 1$ e suponha que $f + g$ é bem definida, logo*

$$\left(\int |f+g|^p \, d\mu\right)^{\frac{1}{p}} \leq \left(\int |f|^p \, d\mu\right)^{\frac{1}{p}} + \left(\int |g|^p \, d\mu\right)^{\frac{1}{p}}.$$

Demonstração. Supõem-se $p > 1$. $f + g$ é evidentemente mensurável. Como

$$|f+g|^p \leq [2\sup\{(|f|, |g|\}]^p \leq 2^p(|f|^p + |g|^p)$$

segue-se que, se $||f||_p < \infty$ e $||g||_p < \infty$, então $||f+g||_p < \infty$. Além disso,

$$\begin{aligned} |f+g|^p &= |f+g||f+g|^{p-1} \\ &\leq |f||f+g|^{p-1} + |g||f+g|^{p-1} \end{aligned}$$

Portanto, pode-se aplicar a Desigualdade de Hölder para inferir que

$$\begin{aligned} \int |f||f+g|^{p-1} \, d\mu &\leq ||f||_p \left\{\int (|f+g|^{p-1})^q \, d\mu\right\}^{\frac{1}{q}} \\ &= ||f||_p \cdot ||f+g||_p^{\frac{p}{q}} \end{aligned}$$

Se tratarmos o segundo termo à direita de forma semelhante, obtém-se

$$\begin{aligned} ||f+g||_p^p &\leq ||f||_p \cdot ||f+g||_p^{\frac{p}{q}} + ||g||_p \cdot ||f+g||_p^{\frac{p}{q}} \\ &= (||f||_p + ||g||_p) \cdot ||f+g||_p^{\frac{p}{q}}. \end{aligned}$$

Se $A = ||f+q||_p = 0$, então a equação é trivial. Se $A \neq 0$, pode-se dividir a desigualdade acima por $A^{\frac{p}{q}}$; como $p - \frac{p}{q} = 1$, conclui-se. ■

Observação 1.2. *A desigualdade de Minkowski pode ser generalizada de modo a obter-se*

$$\left(\int |\sum_{k=1}^{n} f_k|^p \, d\mu\right)^{\frac{1}{p}} \leq \sum_{k=1}^{n} \left(\int |f_k|^p \, d\mu\right)^{\frac{1}{p}},$$

ou então

$$\left(\int |\sum_{k=1}^{\infty} f_k|^p \, d\mu\right)^{\frac{1}{p}} \leq \sum_{k=1}^{\infty} \left(\int |f_k|^p \, d\mu\right)^{\frac{1}{p}}.$$

Sejam $(A_1, \mathcal{F}_1, \mu_1)$ e $(A_2, \mathcal{F}_2, \mu_2)$ dois espaços de medida, logo, o que define-se como **medida produto** no espaço $(A_1 \times A_2, \mathcal{F}_1 \otimes \mathcal{F}_1)$, onde $\mathcal{F}_1 \otimes \mathcal{F}_1 = \{B_1 \times B_2 : B_1 \in \mathcal{F}_1 \text{ e } B_2 \in \mathcal{F}_2\}$ é a medida $\mu_1 \otimes \mu_2$ que satisfaz $(\mu_1 \otimes \mu_2)(B_1 \times B_2) = \mu_1(B_1) \cdot \mu_2(B_2)$, para qualquer $B_1 \in \mathcal{F}_1$ e $B_2 \in \mathcal{F}_2$.

Teorema 1.4. *(Teorema de Fubini-Tonelli). Seja $f : A_1 \times A_2 \to \mathbb{R}$ não-negativa e mensurável. Então*

$$\int f \, d\mu_1 \otimes \mu_2 = \int \left(\int f \, d\mu_1\right) d\mu_2 = \int \left(\int f \, d\mu_2\right) d\mu_1.$$

Demonstração. A demonstração pode ser encontrada em [4]. ■

Uma medida λ em $\mathcal{F}$ é dita **absolutamente contínua** em relação a uma medida μ em $\mathcal{F}$ se $E \in \mathcal{F}$ e $\mu(E) = 0$ implicam que $\lambda(E) = 0$. Neste caso, escrevemos $\lambda << \mu$.

Teorema 1.5. *(Radon-Nikodym). Sejam λ e μ medidas σ-finitas definidas em $\mathcal{F}$ e suponha que λ seja absolutamente contínua em relação a μ. Então, existe uma função $f : \Omega \to \mathbb{R}$ não negativa tal que*

$$\lambda(E) = \int_E f \, d\mu, E \in \Omega.$$

Além disso, a função f é determinada unicamente μ quase todos os lugares.

Demonstração. A demonstração pode ser encontrada em [4]. ■

Observação 1.3. *O conceito usual de topologia que é visto frequentemente em cursos de graduação é utilizado para generalizar e desenvolver muitos conceitos importantes de maneira significativa, entretanto, algumas propriedades em certos espaços, como os de dimensão infinita, acabam não sendo possíveis de serem observadas nestas topologias, a exemplo, em espaços de dimensão infinita não é garantido que o fecho de uma bola unitária seja compacta, o que "limita" o estudo desses espaços. Por esse motivo, surge o que chama-se de topologia $fraca^*$.*

Definição 1.5. *Dada uma medida $\mu \in \mathcal{P}(X)$, um conjunto finito $\Phi = \{\phi_1, \ldots, \phi_N\}$ de funções contínuas limitadas $\phi_i : X \to \mathbb{R}$ e um número $\varepsilon > 0$, define-se*

$$V(\mu, \Phi, \varepsilon) = \left\{ \nu \in \mathcal{P}(X) : \left| \int \phi_i d\nu - \int \phi_i d\mu \right| < \varepsilon, \text{ para todo } i \right\}.$$

*Define-se a **topologia** $fraca^*$ como a topologia definida por estas bases de vizinhanças. Em outras palavras, os abertos da topologia $fraca^*$ são os conjuntos $A \subset \mathcal{P}(X)$ tais que para todo elemento $\mu \in A$ existe algum $V(\mu, \Phi, \varepsilon)$ contido em A.*

Tal topologia pode ser entendida como sendo uma topologia induzida no espaço das medidas de probabilidade através dos funcionais lineares limitados e da topologia do espaço X. É feita de tal modo que seja o mais "fina" possível, ou seja, de uma maneira que consegue-se as propriedades desejadas no espaço.

Diz-se que uma sequência é fracamente convergente (denota-se como $\mu_k \rightharpoonup \mu$) se, somente se, converge na topologia fraca*.

Lema 1.1. *Uma sequência $(\mu_n)_{n\in\mathbb{N}}$ converge para uma medida $\mu \in \mathcal{P}(M)$ na topologia $fraca^*$ se, e somente se,*

$$\int \phi d\mu_n \to \int \phi d\mu$$

para toda função contínua limitada $\phi : X \to \mathbb{R}$.

Demonstração. Para os fins desse trabalho, não será demonstrado o resultado, mas pode-se encontrar uma demonstração para um caso similar ao deste lema em [1]. ■

Define-se a métrica induzida no espaço $\mathcal{P}(A)$, como

$$d(\mu, \nu) = \sup_f | \int f d\mu - \int f d\nu|$$

sendo $f : A \to \mathbb{R}$ uma função lipschitz.

Proposição 1.5. *Se X é espaço métrico separável então a topologia induzida pela distância d coincide com a topologia $fraca^*$ em $\mathcal{P}(X)$*

Demonstração. Pode-se encontrar uma demonstração para a proposição em [14] ■

A distância d nos permite induzir a topologia fraca*, porém, não consegue-se que $\mathcal{P}(X)$ seja completo com a métrica, desta forma, no Capítulo 2 será introduzido uma métrica em $\mathcal{P}(X)$ que garantirá a completude do espaço.

1.3 Conceitos da Teoria da Probabilidade

Neste momento será discorrido a respeito dos objetos da teoria da probabilidade. Neste contexto, serão utilizados os espaços de medida $(\Omega, \mathcal{F})$ como espaços amostrais, o qual seguirá a formulação axiomática de Kolmogorov, ou seja, uma probabilidade $\mathbb{P}$ neste contexto será uma medida de probabilidade, como trazida anteriormente, logo, $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ é uma medida tal que $\mathbb{P}(\Omega) = 1$, onde $\mathcal{F}$ é uma σ-álgebra de Ω.

Em probabilidade são estudados os "eventos aleatórios", os quais pode-se considerar como sendo aqueles que, ao serem repetidos nas mesmas condições, geram resultados diferentes. Neste contexto, os espaços amostrais são aqueles conjuntos que possuem todas as possibilidades de resultados que pode-se obter ao realizar o experimento aleatório. E os elementos da sigma álgebra $\mathcal{F}$ são chamados de eventos.

Sejam $(\Omega, \mathcal{F}, \mathbb{P})$ espaço de probabilidade e $A, B \in \mathcal{F}$ eventos, com $\mathbb{P}(B) > 0$, define-se a **probabilidade condicional** de A dado B por

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Quando $\mathbb{P}(B) = 0$, define-se $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Proposição 1.6. *Dado $B \in \mathcal{F}$, a função $\mathbb{P}(\cdot|B)$ que leva A em $\mathbb{P}(A|B)$ é uma medida de probabilidade.*

Demonstração. A demonstração decorre diretamente da Proposição 1.1, uma vez que $\mathbb{P}(B)$ é constante e

$$\mathbb{P}_B(\Omega) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1.$$

■

Observação 1.4. *Pode-se observar que, dado a probabilidade condicional $\mathbb{P}(A|B)$, tem-se que*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B).$$

Juntamente com o conceito de probabilidade condicional tem-se a noção de eventos independentes. Os eventos aleatórios A e B são ditos **independentes** se

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Neste caso, diz-se que A é independente de B. Ou seja $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Definição 1.6. *Uma **variável aleatória** $\mathcal{X}$ em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ é uma função mensurável definida no espaço Ω. Ou seja, o conjunto $\{\omega \in \Omega : \mathcal{X}(\omega) \leq x\}$ é um evento aleatório para todo $x \in \mathbb{R}$, isto é, $\{\omega \in \Omega : \mathcal{X}(\omega) \leq x\} \in \mathcal{F}$ para todo $x \in \mathbb{R}$. Também escreve-se $\{\omega \in \Omega : \mathcal{X}(\omega) \leq x\} = \{\mathcal{X} \leq x\}$.*

Dado uma espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$, a **distribuição** de uma variável aleatória $X : \Omega \to Y$ é definida como sendo a medida de probabilidade em Y tal que, para qualquer $A \subset B(\mathbb{R})$,

$$m(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(X \in A) = \mathbb{P}_X(A).$$

Define-se o **espaço de probabilidade induzido** por $\mathcal{X}$ como $(Y, \sigma(\mathcal{X}), \mathbb{P}_{\mathcal{X}})$. Também denota-se, para $B \subset Y$, $\mathcal{X}^{-1}(B) = \{\omega \in \Omega : \mathcal{X}(\omega) \in B\} = [\mathcal{X} \in B]$.

Diz-se que uma variável aleatória $\mathcal{X}$ é **discreta** se existe um conjunto enumerável $A = \{x_1, x_2, x_3, \ldots\} \subseteq \mathbb{R}$ tal que $\mathbb{P}(\mathcal{X} \in A) = 1$.

Diz-se que $\mathcal{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_n)$ é um **vetor aleatório** se $\mathcal{X}_j$ for variável aleatória para todo $j = 1, \ldots, n$. Dado um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ e um vetor aleatório $\mathcal{X}$, define-se o espaço de probabilidade induzido por $\mathcal{X}$ como $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbb{P}_{\mathcal{X}})$, onde

$$\mathbb{P}_{\mathcal{X}}(B) = \mathbb{P}(\{\omega \in \Omega : \mathcal{X}(\omega) \in B\}), B \in \mathcal{B}(\mathbb{R}^n).$$

Diz-se que as variáveis aleatórias $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n$ são **independentes** se $\mathbb{P}(\mathcal{X}_1 \in B_1, \ldots, \mathcal{X}_n \in B_n) = \mathbb{P}(\mathcal{X}_1 \in B_1) \cdots \mathbb{P}(\mathcal{X}_n \in B_n)$ para todos $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$. Dito de forma mais sofisticada,

$$\mathbb{P}_{\mathcal{X}}(B_1 \times \cdots \times B_n) = \mathbb{P}_{\mathcal{X}_1}(B_1) \cdots \mathbb{P}_{\mathcal{X}_n}(B_n).$$

Dada uma variável aleatória $\mathcal{X}$, define-se a **esperança** de $\mathcal{X}$, ou média de $\mathcal{X}$, denotada por $\mathbb{E}\mathcal{X}$, por

$$\mathbb{E}\mathcal{X} = \int \mathcal{X}\, d\mathbb{P}$$

Observação 1.5. *Sejam $\mathcal{X}$ e $\mathcal{Y}$ variáveis aleatórias e $a, b \in \mathbb{R}$. Então pode-se obter que $\mathbb{E}[a\mathcal{X} + b\mathcal{Y}] = a\mathbb{E}\mathcal{X} + b\mathbb{E}\mathcal{Y}$.*

Proposição 1.7. *Sejam $\mathcal{X}$ e $\mathcal{Y}$ variáveis aleatórias e $a, b \in \mathbb{R}$. Então valem as seguintes propriedades:*

1. *Se $\mathcal{X} = 1_A$ para algum $A \in \mathcal{F}$, então $\mathbb{E}\mathcal{X} = \mathbb{P}(A)$;*

2. *Se $0 \leq \mathcal{X} \leq \mathcal{Y}$ para todo $\omega \in \Omega$, então $0 \leq \mathbb{E}\mathcal{X} \leq \mathbb{E}\mathcal{Y}$.*

Demonstração. Primeiramente, observa-se que, se $\mathcal{X} = 1_A$, então $\mathcal{X}$ é uma variável aleatória simples, onde

$$\mathbb{E}\mathcal{X} = 1 \cdot \mathbb{P}(\mathcal{X} = 1) = \mathbb{P}(A).$$

Para a segunda afirmação, basta observar que

$$\mathbb{E}\mathcal{X} = \int \mathbb{P}(\mathcal{X} > x)\, dx \leq \int \mathbb{P}(Y > x)\, dx = \mathbb{E}\mathcal{Y}.$$

∎

É possível mostrar que, dada uma variável aleatória $\mathcal{X}$ e $a, b, c \in \mathbb{R}$, valem que, se $\mathcal{X} = c$ para todo ω, então $\mathbb{E}\mathcal{X} = c$; se $a \leq \mathcal{X} \leq b$ para todo ω, então $a \leq \mathbb{E}\mathcal{X} \leq b$; se $\mathcal{X} > 0$ para todo ω, então $\mathbb{E}\mathcal{X} > 0$; e, se $\mathcal{X} > 0$ para todo ω e $\mathbb{E}\mathcal{X} = 0$, então $\mathbb{P}(\mathcal{X} = 0) = 1$.

Observação 1.6. *Pode-se provar*[1] *também que, se $\mathcal{X}$ e $\mathcal{Y}$ são independentes e não-negativas, então*

$$\mathbb{E}[\mathcal{X}\mathcal{Y}] = \mathbb{E}\mathcal{X} \cdot \mathbb{E}\mathcal{Y}.$$

Neste momento serão trazidos algumas desigualdades fundamentais relevantes para o trabalho.

Proposição 1.8. *(Desigualdade de Markov). Sejam $\mathcal{X}$ uma variável aleatória, $\lambda > 0$ e $t > 0$. Então*

$$\mathbb{P}(|\mathcal{X}| \geq \lambda) \leq \frac{\mathbb{E}|\mathcal{X}|^t}{\lambda^t}.$$

Demonstração. Como $|X|^t \geq \lambda^t 1_{|X|^t \geq \lambda^t}$, segue que

$$\mathbb{E}|X|^t \geq \lambda^t \cdot \mathbb{P}(|X|^t \geq \lambda^t).$$

Portanto,

$$\mathbb{P}(|X| \geq \lambda) = \mathbb{P}(|X|^t \geq \lambda^t) \leq \frac{\mathbb{E}|X|^t}{\lambda^t}.$$

■

Proposição 1.9. *(Desigualdade de Cauchy-Schwarz). Sejam $\mathcal{X}$ e $\mathcal{Y}$ tais que $\mathbb{E}\mathcal{X}^2 < \infty$ e $\mathbb{E}\mathcal{Y}^2 < \infty$, então $\mathcal{X}\mathcal{Y}$ é integrável e*

$$\mathbb{E}[\mathcal{X}\mathcal{Y}] \leq \sqrt{\mathbb{E}\mathcal{X}^2}\sqrt{\mathbb{E}\mathcal{Y}^2}.$$

Demonstração. A demonstração segue diretamente da desigualdade de Hölder. ■

Dado um intervalo aberto $I \subseteq \mathbb{R}$, diz-se que $g : I \to \mathbb{R}$ é uma função **convexa** se $g(a \cdot x + b \cdot y) \leq a \cdot g(x) + b \cdot g(y)$ para quaisquer $x, y \in I$ e $a, b \in [0, 1]$ com $a + b = 1$. Essa condição de convexidade pode ser reescrita da seguinte maneira: para todos $x < z < y$ em I, vale

$$\frac{g(z) - g(x)}{z - x} \leq \frac{g(y) - g(z)}{y - z}.$$

Pode-se obter outras caracterizações de convexidade explorando a possível diferenciabilidade de g. Se g' existe e é não-decrescente em todo I, então pelo Teorema do Valor Médio g satisfaz a equação acima, logo é convexa.

[1] O leitor pode encontrar uma demonstração dessa afirmação em [18].

Proposição 1.10. *(Desigualdade de Jensen). Sejam $I \subseteq \mathbb{R}$ um intervalo aberto, $g : I \to \mathbb{R}$ uma função convexa, e $\mathcal{X}$ uma variável aleatória integrável assumindo valores em I. Então*

$$\mathbb{E}[g(\mathcal{X})] \geq g(\mathbb{E}\mathcal{X}).$$

Demonstração. Preliminarmente, afirma-se que, para cada $z \in I$ fixo, existe $c \in \mathbb{R}$ tal que $g(w) \geq g(z) + c(w - z)$ para todo $w \in I$ (caso g seja diferenciável, pode-se tomar $c = g'(z)$, logo o gráfico de g está acima de suas retas tangentes). Com efeito, considerando os possíveis valores dos lados esquerdo e direito da desigualdade

$$\frac{g(z) - g(x)}{z - x} \leq \frac{g(y) - g(z)}{y - z}.$$

obtém-se $c \in \mathbb{R}$ tal que

$$\frac{g(z) - g(x)}{z - x} \leq c \leq \frac{g(y) - g(z)}{y - z}.$$

para todo $x < z$ e todo $y > z$, provando a afirmação.

Finalmente, tomando $z = \mathbb{E}\mathcal{X}$ e usando $\mathcal{X}$ no lugar de w, obtém-se

$$\mathbb{E}[g(\mathcal{X})] \geq \mathbb{E}[g(\mathbb{E}\mathcal{X}) + c(\mathcal{X} - \mathbb{E}\mathcal{X})] = g(\mathbb{E}\mathcal{X}) + c\mathbb{E}\mathcal{X} - c\mathbb{E}\mathcal{X} = g(\mathbb{E}\mathcal{X}),$$

o que conclui a demonstração. ■

Observação 1.7. *De maneira similar, pode-se encontrar a desigualdade de Jensen para funções côncavas, de modo que, se $g : I \to \mathbb{R}$ é uma função côncava, ou seja para todos $x < z < y$ em I, vale*

$$g(a \cdot x + b \cdot y) \geq a \cdot g(x) + b \cdot g(y),$$

e $\mathcal{X}$ uma variável aleatória integrável assumindo valores em I, então

$$\mathbb{E}[g(\mathcal{X})] \leq g(\mathbb{E}\mathcal{X}).$$

Definição 1.7. *Sejam $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade, $\mathcal{X}$ uma variável aleatória estendida integrável ou não-negativa, e $\mathcal{G} \subseteq \mathcal{F}$ uma σ-álgebra. Diz-se que uma variável aleatória estendida $\mathcal{Z}$ é a* ***esperança condicional*** *de $\mathcal{X}$ dado $\mathcal{G}$, denotada $\mathcal{Z} = \mathbb{E}[\mathcal{X}|\mathcal{G}]$, se $\mathcal{Z}$ é $\mathcal{G}$-mensurável e*

$$\int_A \mathcal{Z}\, d\mathbb{P} = \int_A \mathcal{X}\, d\mathbb{P}$$

para todo $A \in \mathcal{G}$.

Se pode definir também a esperança de $\mathcal{X}$ dado $\mathcal{Y}$, onde $\mathcal{X}$ e $\mathcal{Y}$ são duas variáveis aleatórias, como

$$\mathbb{E}[\mathcal{X}|\mathcal{Y}] = \mathbb{E}[\mathcal{X}|\sigma(\mathcal{Y})].$$

Pode-se observar também que

$$\mathbb{P}(A|\mathcal{F}) = \mathbb{E}[1_A|\mathcal{F}].$$

Teorema 1.6. *A esperança condicional $\mathcal{Z} = \mathbb{E}[\mathcal{X}|\mathcal{G}]$ acima está bem definida.*

Demonstração. Considera-se inicialmente o caso onde $\mathcal{X}$ é não-negativa e integrável. Define-se a medida ν no espaço mensurável $(\Omega, \mathcal{G})$ como

$$\nu(A) = \int_A \mathcal{X}\, d\mathbb{P} \text{ para cada } A \in \mathcal{G}.$$

Como $\nu(\Omega) = \mathbb{E}\mathcal{X} < \infty$, tem-se que é σ-finita. E como $\nu \ll \mathbb{P}$, pelo Teorema de Radon-Nikodým, a derivada $\frac{d\nu}{d\mathbb{P}}$ existe e é uma função boreliana definida no espaço $(\Omega, \mathcal{G})$. Observando que $Z = \frac{d\nu}{d\mathbb{P}}$ é $\mathcal{G}$-mensurável e satisfaz $\int_A \mathcal{Z}\, d\mathbb{P} = \int_A \mathcal{X}\, d\mathbb{P}$, conclui-se que $Z = \mathbb{E}[\mathcal{X}|\mathcal{G}]$.

Supõe-se agora que $\mathcal{X}$ seja integrável. Pelo caso anterior, existem variáveis aleatórias integráveis $\mathbb{E}[\mathcal{X}^+|\mathcal{G}]$ e $\mathbb{E}[\mathcal{X}^-|\mathcal{G}]$ tais que

$$\int_A \mathbb{E}[\mathcal{X}^\pm|\mathcal{G}]\, d\mathbb{P} = \int_A \mathcal{X}^\pm\, d\mathbb{P}$$

para todo $A \in \mathcal{G}$. Definindo $Z = \mathbb{E}[\mathcal{X}^+|\mathcal{G}] - \mathbb{E}[\mathcal{X}^-|\mathcal{G}]$, obtém-se,

$$\begin{aligned}
\int_A Z\, d\mathbb{P} &= \int_A \mathbb{E}[\mathcal{X}^+|\mathcal{G}] - \mathbb{E}[\mathcal{X}^-|\mathcal{G}]\, d\mathbb{P} \\
&= \int_A \mathbb{E}[\mathcal{X}^+|\mathcal{G}]\, d\mathbb{P} - \int_A \mathbb{E}[\mathcal{X}^-|\mathcal{G}]\, d\mathbb{P} \\
&= \int_A \mathcal{X}^+\, d\mathbb{P} - \int_A \mathcal{X}^-\, d\mathbb{P} = \int_A \mathcal{X}\, d\mathbb{P}
\end{aligned}$$

para todo $A \in \mathcal{G}$, o que está bem definido pois todas as variáveis aleatórias envolvidas são integráveis. Portanto Z cumpre a definição de $\mathbb{E}[\mathcal{X}|\mathcal{G}]$, concluindo a demostração deste caso.

Por último, supõe-se que $\mathcal{X}$ seja não-negativa mas não necessariamente integrável. Toma-se uma sequência $0 \leq \mathcal{X}_n \nearrow \mathcal{X}$ onde $\mathcal{X}_n$ são variáveis aleatórias simples não-negativas. Pelo caso anterior, existe $\mathbb{E}[\mathcal{X}_n|\mathcal{G}]$ para todo n. Tem-se, pelo teorema da convergência monótona, $\mathbb{E}[\mathcal{X}_n|\mathcal{G}] \nearrow Z$ para alguma variável aleatória estendida Z que é $\mathcal{G}$-mensurável e satisfaz

$$\int_A \mathcal{X}\, d\mathbb{P} = \int_A Z\, d\mathbb{P},$$

para todo $A \in \mathcal{G}$. Ou seja, Z cumpre a definição de $\mathbb{E}[\mathcal{X}|\mathcal{G}]$, concluindo a demostração do teorema. ∎

1.3.1 Cadeias de Markov

Considera-se o espaço de medida $(E, \mathcal{F})$ para o que será feito a seguir.

Uma sequência de variáveis aleatórias (X_n) é chamada **cadeia de Markov** se para todo conjunto mensurável A tem-se

$$\mathbb{P}(X_{n+1} \in A|X_0, \ldots, X_n) = \mathbb{P}(X_{n+1} \in A|X_n).$$

Desta forma, pode-se entender como cadeia de Markov uma sequência de "acontecimentos" na qual o que ocorreu no "passado" não é relevante para se entender o que vai acontecer no "futuro", apenas o "presente" o é.

Nesse texto será trabalhado apenas com cadeias Markov homogêneas. Para definir esse conceito, precisa-se dizer o que é uma *probabilidade de transição*.

Definição 1.8. *Uma **probabilidade de transição** em E é uma função de $E \times \mathcal{F}$ em $[0, 1]$ tal que*

(a) A função $\Lambda \mapsto p(x, \Lambda)$ é uma probabilidade para todo x;

(b) A função $x \mapsto p(x, \Lambda)$ é mensurável para todo Λ.

Diz-se que uma sequência de variáveis (X_n) é uma cadeia de Markov *homogênea* com probabilidade de transição p se

$$\mathbb{P}(X_{n+1} \in A|X_0, \ldots, X_n) = p(X_n, A)$$

para todo mensurável A.

A probabilidade de transição p determina completamente a distribuição das variáveis X_n. De fato, se X_n tem distribuição m, então X_{n+1} terá distribuição Tm onde T é um operador que age no espaço das probabilidades definido por

$$Tm(A) = \int p(x, A)\, dm(x).$$

Para essa afirmação, basta observar que

$$\begin{aligned} Tm(A) &= \int p(x, A)\, dm(x) = \int p(x, A)\, d\mathbb{P}_{X_n}(x) \\ &= \int p(x, A)\, d\mathbb{P}(X_n^{-1}(x)) = \int p(X_n(x), A)\, d\mathbb{P}(x) \\ &= \int \mathbb{P}(X_{n+1}(x) \in A|X_n(x)) d\mathbb{P}(x) \\ &= \mathbb{P}(X_{n+1} \in A). \end{aligned}$$

Pode-se observar que, se m for a distribuição de X_0, $T^n m$ será a distribuição da variável X_n, onde entende-se o inteiro n como um "tempo" na sequência.

Seja $X_0, X_1, \ldots,$ uma sequência de variáveis aleatórias independentes tomando valores em $\{1, \ldots, k\}$ com mesma distribuição $\nu = p_1\delta_1 + \cdots + p_k\delta_k$. Seja Z_0 uma variável aleatória com valores num espaço métrico M independente de $X = \{X_n : n \geq 0\}$.

Considere funções contínuas $f_1, \dots, f_k \colon M \to M$ e defina a sequência de variáveis aleatórias

$$Z_{n+1}(\omega) = f_{X_n(\omega)} \circ \cdots \circ f_{X_0(\omega)}(Z_0(\omega)) = f_{X_n}(Z_n). \tag{1.1}$$

Proposição 1.11. *A sequência Z_n, $n \geq 0$, é uma cadeia de Markov homogênea com probabilidade de transição dada por*

$$p(x, A) = \int 1_A(f_i(x))\, d\nu(i)$$

Demonstração. Precisa-se mostrar que para todo Boreliano A tem-se que

$$\mathbb{P}(Z_{n+1} \in A | Z_0, Z_1, \dots, Z_n)(\omega) = p(Z_n(\omega), A)$$

para $\mathbb{P}$-quase todo ponto ω. Para isso, seja $B \in \sigma(Z_0, \dots, Z_n)$. Tem-se que verificar

$$\int_B 1_A(Z_{n+1}(\omega))\, d\mathbb{P}(\omega) = \int_B p(Z_n(\omega), A)\, d\mathbb{P}(\omega).$$

Note que $\sigma(Z_0, \dots, Z_n) \subset \sigma(Z_0, X_0, \dots, X_{n-1})$, e $\sigma(Z_0, X_0, \dots, X_{n-1})$ é a σ-algebra gerada pela aplicação

$$\iota(\omega) = (X_0(\omega), \dots, X_{n-1}(\omega), Z_0(\omega)).$$

Assim, B é um conjunto da forma $\iota^{-1}(B')$ onde B' é um conjunto mensurável de $\{1, \dots, k\}^n \times M$. Então tem-se que

$$\begin{aligned}
\int_B 1_A(Z_{n+1})\, d\mathbb{P} &= \int 1_A(Z_{n+1}) 1_B\, d\mathbb{P} \\
&= \int 1_A(f_{X_n} \circ \cdots \circ f_{X_0}(Z_0)) 1_{B'}(X_0, \dots, X_{n-1}, Z_0)\, d\mathbb{P}
\end{aligned}$$

Defina $\Pi \colon \{1, \dots, k\}^{n+1} \times M \to \mathbb{R}$ por

$$\Pi(\alpha_0, \dots, \alpha_n, x) = 1_A(f_{\alpha_n} \circ \cdots \circ f_{\alpha_0}(x)) 1_{B'}(\alpha_0, \dots, \alpha_{n-1}, x).$$

Então

$$\int_B 1_A(Z_{n+1}(\omega))\, d\mathbb{P}(\omega) = \int \Pi(X_0, \dots, X_n, Z_0)\, \mathbb{P}.$$

Como $X_0, \dots, X_n, Z_0$ são independentes então a distribuição do vetor $(X_0, \dots, X_n, Z_0)$ é $\nu^{n+1} \times \mu$. Concluí-se que

$$\int_B 1_A(Z_{n+1}(\omega))\, d\mathbb{P}(\omega) = \iint 1_A(f_{\alpha_n} \circ \cdots \circ f_{\alpha_0}(x)) 1_{B'}(\alpha_0, \dots, \alpha_{n-1}, x)\, d\nu^{n+1}\, d\mu$$

Como

$$\int 1_A(f_{\alpha_n} \circ \cdots \circ f_{\alpha_0}(x)) d\nu(\alpha_n) = p(f_{\alpha_{n-1}} \circ \cdots \circ f_{\alpha_0}(x), A)$$

então conclui-se que

$$\begin{aligned}
\int_B 1_A(Z_{n+1}(\omega))\, d\mathbb{P}(\omega) &= \int p(f_{\alpha_{n-1}} \circ \cdots \circ f_{\alpha_0}(x), A) 1_{B'}(\alpha_0, \dots, \alpha_{n-1}, x)\, d\nu^n\, d\mu \\
&= \int_B p(f_{X_{n-1}} \circ \cdots \circ f_{X_0}(Z_0), A)\, d\mathbb{P}.
\end{aligned}$$

■

Medidas estacionárias

Chama-se **medida estacionária** de uma cadeia de Markov homogênea com probabilidade de transição p, uma medida $\pi \in \mathcal{P}(\Omega)$ tal que

$$\pi(A) = \int \pi(dx)p(x, A).$$

Ou seja, diz-se que uma medida é estacionária se $Tm = m$, em outras palavras, se X_0 tem distribuição m, então m também será a distribuição de X_n, para qualquer n.

Considerando $p(x, A) = \int 1_A(f_i(x))\, d\nu(i)$, tem-se, pelo teorema de Fubini-Tonelli, que

$$\begin{aligned} \pi(A) &= \iint 1_A(f_i(x))\, d\nu(i)\pi(dx) = \iint 1_A(f_i(x))\, \pi(dx)d\nu(i) \\ &= \int \pi(f_i^{-1}(A))\, d\nu(i) = \int f_{i\,*}\pi(A)\, d\nu. \end{aligned}$$

Capítulo 2

Transporte ótimo

Neste capítulo serão trabalhados os conceitos e resultados relacionados a teoria do transporte ótimo, ao problema de Monge-Kantorovich e a métrica de Wasserstein. Para tal, utiliza-se como referência [20] e [8].

2.1 O problema de Monge-Kantorovich

Neste momento, será discorrido sobre o problema que foi o "ponta pé" da teoria do transporte ótimo, posteriormente chamado de "O problema de Monge-Kantorovich".

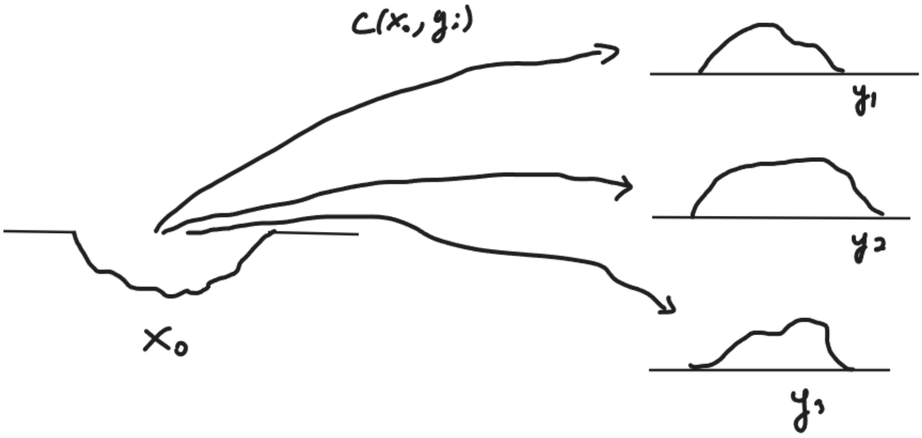
Inicialmente, o problema que foi trazido por Gaspard Monge (1746-1818) (Figura 2.1) era, basicamente, entender como transportar areia de um determinado local para vários outros locais, os quais são todos conhecidos (Uma ilustração pode ser visto na Figura 2.2, onde o ponto x_0 é onde será "retirado" a areia, os pontos y_1, y_2 e y_3 são onde será "despejado" e $c(x_0, y_i)$ é o custo de transportar do ponto x_0 ao ponto y_i). O que tem-se que observar nesse problema é como e em que parte do local será feito a extração da areia e para qual local deve ser levado. O porque se resume ao custo que esse transporte tem, devido, principalmente, ao peso e ao tempo que se leva para o fazer. Monge assumiu que o custo de transporte de uma unidade de massa ao longo de uma certa distância era dado pelo produto da massa pela distância.

Uma outra interpretação do problema proposto por Monge seria o seguinte: Em uma cidade da Bahia existe uma grande franquia de padarias que distribui os seus pães que são feitos na madrugada para estabelecimentos nas cidades vizinhas de modo que a entrega aconteça antes do horário de funcionamento do estabelecimento em si. A quantidade de pão que deve ser produzido em cada uma das padarias da franquia e quanto deve ser enviado para cada padaria das cidades vizinhas é conhecido de forma antecipada, sendo assim possível modelar matematicamente (o que será falado logo em breve). O problema é encontrar na prática onde cada unidade de pão deve ir, de forma a minimizar o custo total

Figura 2.1: Gaspard Monge

Fonte: https://mathshistory.st-andrews.ac.uk/Biographies/Monge/

Figura 2.2: Exemplificando o transporte de areia



Fonte: os autores

de transporte. Esta maneira de interpretar é um meio de observar que este problema pode ser utilizado para mais que transportar fardos de areia, mas também aplicar em outras esferas, como em economia.

Monge estudou o problema em três dimensões para uma distribuição contínua de massa. Guiado por sua intuição geométrica, ele fez a importante observação de que o transporte deveria seguir linhas retas ortogonais a uma família de superfícies. Este estudo o levou à descoberta de linhas de curvatura, um conceito que por si só foi uma grande contribuição para a geometria das superfícies. Suas ideias foram desenvolvidas por Charles Dupin e mais tarde por Paul Appell. Pelos padrões matemáticos atuais, todos esses argumentos eram falhos, mas certamente o problema trazido por Monge se fez relevante para ser estudado com ferramentas modernas. Posteriormente, o problema proposto por Monge geraria a principal ferramenta deste capítulo, o acoplamento ótimo.

Mais tarde o problema de Monge foi revisitado por Leonid Vitalyevich Kantorovich (Figura 2.3). No caso que é de interesse direto para o estudo do problema do acoplamento ótimo, Kantorovich enunciou e provou, por meio de ferramentas analíticas funcionais, um teorema de dualidade que desempenharia um papel crucial posteriormente.

Ele também concebeu uma noção conveniente de distância entre medidas de probabilidade que é hoje chamada de distância de Kantorovich–Rubinstein, ou, como será mencionada neste trabalho, distância de Wasserstein, que se provou ser particularmente flexível e útil.

Figura 2.3: Leonid Vitalyevich Kantorovich

Fonte: https://mathshistory.st-andrews.ac.uk/Biographies/Kantorovich/

Foi apenas vários anos após seus principais resultados que Kantorovich fez a conexão com o trabalho de Monge. Kantorovich também estabeleceu uma hipótese mais fraca do problema de Monge, enquanto Monge considerava que toda massa de um ponto x fosse transportada para um único ponto y, Kantorovich estabeleceu que a massa x pudesse ser transportada para mais de um ponto, o que gerou alguns resultados bastante relevantes. O problema do acoplamento ótimo desde então tem sido chamado de problema de Monge-Kantorovich.

Dada a contextualização acima, será observado agora como associar o problema de Monge-Kantorovich com medidas de probabilidade.

Pensa-se da seguinte maneira: dado o problema de transporte de pães, entende-se como X o espaço cujo os pontos representam as padarias da franquia citada e o espaço Y cujo os pontos representam os estabelecimentos das outras cidades que serão entregues os pães. Desta forma, pode-se observar que a função $c(x, y)$ que é a distância do ponto x ao ponto y será a função que representará o custo de transporte da padaria x da franquia para o estabelecimento y.

Assim, pensa-se então em medidas $\mu \in \mathcal{P}(X)$ e $\nu \in \mathcal{P}(Y)$, onde, como nesse caso X e Y são discretos, μ representa a forma como será a porcentagem de pães que sairão do ponto de partida uma padaria x_i e ν representa a quantidade de pães que chegarão no destino, um estabelecimento y_j. Ao pensar desta maneira, tem-se que, ao escolher uma padaria e um estabelecimento aleatoriamente, pode-se estar avaliando o custo do transporte com a distancia vezes a quantidade a ser distribuída pensando nessa quantidade através das medidas μ e ν separadamente, isto pode gerar algumas interpretações que não levariam na otimização desse custo. Para contornarmos isso, faz-se o seguinte, considera-se que, agora, pode ser feito entregas de pães de uma padaria x_i para mais de um estabelecimento y_j e uma padaria y_i pode receber pães de mais de um local; considera-se uma probabilidade γ no espaço $X \times Y$, onde associa um par $(x_i, y_j) \in X \times Y$ a distribuição a ser feita naquela escolha de entregar de x_i em y_j, assim, $d\gamma$ descreve a distribuição que foi feita dos x_i's para os estabelecimentos, ou seja, escolhe-se uma medida de probabilidade $\gamma \in \mathcal{P}(X \times Y)$ de tal modo que $\gamma(X \times \{y_j\}) = \nu(\{y_j\})$ e $\gamma(\{x_i\} \times Y) = \mu(\{x_i\})$, para $\{y_j\} \subset Y$ e $\{x_i\} \subset X$. Desta forma, para otimizarmos o custo da entrega, faz-se o ínfimo da média da função custo com relação a probabilidade γ, ou seja, resume-se em

$$\inf_{\gamma} \int_{X \times Y} c(x, y)\, d\gamma.$$

Toda a formalidade e construção de tal conceito será explanado na secção a seguir.

2.2 Existência de um Acoplamento Ótimo e Distância de Wasserstein

Neste momento, serão discutidos tópicos da teoria do transporte ótimo, nos quais busca-se a base para as ferramentas do estudo de sistemas de funções iteradas.

Definição 2.1. *Dadas as medidas $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$, um plano de transporte ou* ***acoplamento*** *de μ e ν é uma medida de probabilidade $\gamma \in \mathcal{P}(X \times Y)$ tal que $\gamma(A \times Y) = \mu(A)$ e $\gamma(X \times B) = \nu(B)$ para todos mensuráveis $A \subset X$ e $B \subset Y$. Denota-se o conjunto de todos os acoplamentos de μ, ν como $\Pi(\mu, \nu)$.*

Uma outra maneira que pode-se definir é: dadas as medidas $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$, um acoplamento de (μ, ν) é a construção de duas variáveis aleatórias $\mathcal{X}$ e $\mathcal{Y}$ no espaço de probabilidade $(X \times Y, \mathbb{P})$, tal que $\mathcal{X}_*\mathbb{P} = \mu$ e $\mathcal{Y}_*\mathbb{P} = \nu$. Observa-se que as duas formas de definir-se são equivalentes, neste trabalho, utiliza-se a maneira mais conveniente para o que se quer provar e estudar.

O que será buscado aqui é uma maneira de minimizar a esperança de uma função custo dada de um problema de transporte ótimo, ou seja, dadas μ, ν medidas de probabilidade de X e Y respectivamente, e uma função custo c, quer-se encontrar um acoplamento tal que minimize $\mathbb{E}c(\mathcal{X}, \mathcal{Y})$ (função custo total), onde chama-se tal acoplamento que realiza tal mínimo de acoplamento ótimo. Uma pergunta pertinente seria se tal mínimo existe. Ver-se então uma propriedade importante dos acoplamentos ótimos, eles existem.

Teorema 2.1. *(Existência de um acoplamento ótimo, Villani 2009). Sejam (X, μ) e (Y, ν) dois espaços de probabilidade metrizáveis, separáveis e completos; sejam $a : X \to \mathbb{R} \cup \{-\infty\}$ e $b : Y \to \mathbb{R} \cup \{-\infty\}$ duas funções semicontínuas superiores tais que $a \in L^1(\mu)$, $b \in L^1(\nu)$. Seja $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ uma função custo semicontínua inferior, tal que*

$c(x,y) \geq a(x) + b(y)$ para todo x, y. Então existe um acoplamento de (μ,ν) que minimiza o custo total $\mathbb{E}c(\mathcal{X},\mathcal{Y})$ entre todos os acoplamentos possíveis.

A suposição de limite inferior que é dada para c é utilizada para garantir que a função custo total $\mathbb{E}c(\mathcal{X},\mathcal{Y})$ esteja bem definido em $\mathbb{R} \cup \{+\infty\}$. Na maioria dos casos - embora não seja em todos - pode-se escolher $a \equiv 0$ e $b \equiv 0$.

Para a demonstração do Teorema 2.1 utiliza-se o Teorema de Prokhorov e os Lemas 2.1 e 2.2.

Teorema 2.2. *(Teorema de Prokhorov) Se X é um espaço métrico, completo e separável, então um conjunto $P \subset \mathcal{P}(X)$ é pré-compactado (ou seja, seu fecho é um conjunto compacto) em relação a topologia fraca se, e somente se, for rígido, ou seja, para qualquer $\varepsilon > 0$ existe um conjunto compacto K_ε tal que $\mu[X \backslash K_\varepsilon] \leq \varepsilon$ para todo $\mu \in P$.*

A demonstração do Teorema de Prokhorov pode ser encontrada em [16].

Lema 2.1. *(Semicontinuidade inferior da função custo). Sejam X e Y dois espaços métricos, completos e separáveis, e $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ uma função custo semicontínua inferior. Seja $h : X \times Y \to \mathbb{R} \cup \{-\infty\}$ uma função semicontínua superior tal que $c \geq h$. Seja $(\pi_k)_{k\in\mathbb{N}}$ uma sequência de medidas de probabilidade em $X \times Y$, convergindo na topologia fraca para algum $\pi \in \mathcal{P}(X \times Y)$, de tal forma que $h \in L^1(\pi_k) \cap L^1(\pi)$, e*

$$\int_{X\times Y} h \, d\pi_k \longrightarrow \int_{X\times Y} h \, d\pi,$$

quando $k \longrightarrow +\infty$. Então

$$\int_{X\times Y} c \, d\pi \leq \liminf_{k\to\infty} \int_{X\times Y} c \, d\pi_k.$$

Em particular, se c for não negativa, então $F : \pi \to \int c \, d\pi$ é semicontínua inferior em $\mathcal{P}(X \times Y)$, na convergência da topologia fraca.

Demonstração. Considerando c como $c - h$, pode-se assumir que c é uma função semicontínua inferior não negativa. Então c pode ser escrito como o limite pontual de uma família não decrescente $(c_l)_{l\in\mathbb{N}}$ de funções contínuas reais. Pelo teorema da convergência monótona,

$$\begin{aligned}\int c \, d\pi = \lim_{l\to\infty} \int c_l \, d\pi &= \lim_{l\to\infty} \lim_{k\to\infty} \int c_l \, d\pi_k \\ &\leq \liminf_{k\to\infty} \int c \, d\pi_k.\end{aligned}$$

∎

Lema 2.2. *(Rigidez dos acoplamentos). Sejam X e Y dois espaços métricos, completos e separáveis. Sejam $P \subset \mathcal{P}(X)$ e $Q \subset \mathcal{P}(Y)$ subconjuntos rígidos de $\mathcal{P}(X)$ e $\mathcal{P}(Y)$ respectivamente. Então o conjunto $\Pi(P, Q)$ de todos os acoplamentos cujas marginais estão em P e Q respectivamente, é ele próprio rígido em $\mathcal{P}(X \times Y)$.*

Demonstração. Sejam $\mu \in P$, $\nu \in Q$ e $\pi \in \Pi(\mu, \nu)$. Por suposição, para qualquer $\varepsilon > 0$ existe um conjunto compacto $K_\varepsilon \subset X$, independente da escolha de μ em P, tal que $\mu[X \backslash K_\varepsilon] \leq \varepsilon/2$ e, similarmente, há um conjunto compacto $L_\varepsilon \subset Y$, independente da escolha de ν em Q, tal que $\nu[Y \backslash L_\varepsilon] \leq \varepsilon/2$. Então para qualquer acoplamento de (μ, ν),

$$\mathbb{P}[(\mathcal{X}, \mathcal{Y}) \notin K_\varepsilon \times L_\varepsilon] \leq \mathbb{P}[\mathcal{X} \notin K_\varepsilon] + \mathbb{P}[\mathcal{Y} \notin L_\varepsilon] \leq 2\frac{\varepsilon}{2} = \varepsilon.$$

O resultado desejado segue uma vez que este limite é independente do acoplamento, e $K_\varepsilon \times L_\varepsilon$ é compacto em $X \times Y$. ■

Demonstração. (Teorema 2.1) Como X e Y são espaços métricos, completos e separáveis, $\{\mu\}$ é rígida em $\mathcal{P}(X)$ e, similarmente, $\{\nu\}$ é rígida em $\mathcal{P}(Y)$. Pelo Lema 2.2, $\Pi(\mu, \nu)$ é rígido em $\mathcal{P}(X \times Y)$, e, pelo teorema de Prokhorov, este conjunto é pré-compacto. Ao passar o limite na equação das marginais, vemos que $\Pi(\mu, \nu)$ é fechado, então é, de fato, compacto.

Seja $(\pi_k)_{k \in \mathbb{N}}$ uma sequência de medidas de probabilidade em $X \times Y$, tal que $\int c \, d\pi_k$ converge para o ínfimo do custo total. Extraindo uma subsequência se necessário, pode-se assumir que π_k converge para algum $\pi \in \Pi(\mu, \nu)$. A função $h : (x, y) \mapsto a(x) + b(y)$ pertence a $L^1(\pi_k) \cap L^1(\pi)$, e $c \geq h$ por hipótese; além disso,

$$\int h \, d\pi_k = \int h \, d\pi = \int a \, d\mu + \int b \, d\nu;$$

então o Lema 2.1 implica

$$\int c \, d\pi \leq \liminf_{k \to \infty} \int c \, d\pi_k.$$

Assim, π é minimizante. ■

Sejam $\mu_0, \mu_1 \in \mathcal{P}(X)$, logo, os acoplamentos de μ_0, μ_1 são as medidas de probabilidade $\gamma \in \mathcal{P}(X \times X)$ tais que $\gamma(A \times X) = \mu_0(A)$ e $\gamma(X \times B) = \mu_1(B)$ para todos mensuráveis $A, B \subset X$. Denota-se como $\Gamma(\mu_0, \mu_1)$ o conjunto de todos os acoplamentos de μ_0, μ_1. Assim, dadas essas noções, pode-se definir então uma das principais ferramentas do presente estudo, uma métrica especial induzida no espaço das medidas de probabilidade $\mathcal{P}(X)$, a distância de Wasserstein:

Definição 2.2. *A **distância de Wasserstein** de expoente q entre duas medidas de probabilidade é dada por:*

$$W_q(\nu_0, \nu_1) = C_q(\nu_0, \nu_1)^{\min(1, \frac{1}{q})}$$

onde $C_q(\nu_0, \nu_1)$ é o custo total $C_q(\nu_0, \nu_1) = \inf_{\gamma \in \Gamma(\nu_0, \nu_1)} \int d(x, y)^q d\gamma(x, y)$.

Considerando o abordado até o momento, será utilizado o Teorema 2.1 para concluir que, no principal contexto deste trabalho, sempre consegue-se a existência de um acoplamento ótimo.

Corolário 2.1. *Seja X um espaço de probabilidade metrizável, separável e completo. Seja $d : X \times X \to \mathbb{R}$ a métrica induzida no espaço X. Então existe um acoplamento $\gamma \in \Gamma(\mu_0, \mu_1)$ que realiza o ínfimo do custo total $C_q(\mu_0, \mu_1)$ entre todos os acoplamentos possíveis.*

Demonstração. Pode-se observar que, $\forall q$, $d(x, y)^q \geq 0$ e é contínua, assim, como já comentado, pode-se considerar a função custo como $c(x, y) = d(x, y)^q$, $a \equiv 0$ e $b \equiv 0$, o que irá satisfazer o Teorema 2.1, logo, existirá um acoplamento ótimo que realizará o ínfimo do custo total $C_q(\mu_0, \mu_1)$. ■

Desta forma, tem-se que, se X for um espaço de probabilidade metrizável, separável e completo, então

$$W_q(\nu_0, \nu_1) = \left(\int d(x, y)^q d\gamma(x, y) \right)^{\min(1, \frac{1}{q})}$$

então o espaço $\mathcal{P}(X)$ com a métrica W_q será um espaço métrico completo, o que será provado a seguir.

Observação 2.1. *Observa-se que, pela desigualdade de Hölder, para $p \leq q \Rightarrow W_p(\mu, \nu) \leq W_q(\mu, \nu)$. De fato, para qualquer $\gamma \in \Gamma(\mu, \nu)$, aplica-se a desigualdade de Hölder com $q/p \geq 1$ e obtém-se:*

$$\begin{aligned} \int_X d(x, y)^p \, d\gamma(x, y) &\leq \left(\int_X d(x, y)^q \, d\gamma(x, y) \right)^{\frac{p}{q}} \cdot \gamma(X \times X)^{\frac{q}{1-p}} \\ &= \left(\int_X d(x, y)^q \, d\gamma(x, y) \right)^{\frac{p}{q}} \end{aligned}$$

Por outro lado, se d é uma distância limitada, tem-se

$$\begin{aligned} \int_X d(x, y)^q \, d\gamma(x, y) &= \int_X d(x, y)^{p+(q-p)} \, d\gamma(x, y) \\ &\leq \mathsf{diam}(X)^{q-p} \cdot \int_X d(x, y)^p \, d\gamma(x, y) \end{aligned}$$

o que implica

$$W_q(\mu, \nu) \leq \mathsf{diam}(X)^{1-p/q} \cdot W_p(\mu, \nu)^{p/q}.$$

Em particular, as distâncias são equivalentes quando a distância é limitada.

Será focado em demonstrar a completude do espaço $\mathcal{P}(X)$ com a métrica de Wasserstein, logo, não será provado que W_q é uma métrica em $\mathcal{P}(X)$, porém, o leitor pode encontrar uma demonstração em [8].

Para o Lema 2.3, será utilizado o Fórmula de Kantorovich-Rubinstein:

Teorema 2.3. *(Fórmula de Kantorovich-Rubinstein). Para X um espaço metrizável, separável e completo e $c = d : X \times X \to \mathbb{R}$, tem-se*

$$\inf_{\gamma\in\Gamma(\mu,\nu)} \int_{X\times X} d(x,y)\, d\gamma(x,y) = \sup_{||\phi||_{Lip}\leq 1} \int_X \phi(x)\, d(\mu-\nu)(x).$$

Demonstração. Uma demonstração deste teorema pode ser encontrada em [6]. ■

Lema 2.3. *Seja X um espaço de probabilidade metrizável, separável e completo e $\{\mu_n\}_{n\in\mathbb{N}}$ uma sequência de Cauchy em $\mathcal{P}_p(X) = \{\mu \in \mathcal{P} | x_0 \in X, \int_X d(x,x_0)^p\, d\mu(x) < +\infty\}$ (o qual não depende de x_0) com a distância de Wasserstein W_p. Então, $\{\mu_n\}_{n\in\mathbb{N}}$ é rígida.*

Demonstração. De acordo com a Observação 2.1, $\{\mu_n\}_{n\in\mathbb{N}}$ também é de Cauchy com respeito à métrica W_1. Tem-se que, dado $\varepsilon > 0$, existe $N \in \mathbb{N}$ tal que

$$n \geq N \Rightarrow W_1(\mu_n, \mu_N) < \varepsilon^2.$$

Será provado agora que, dado $n \in \mathbb{N}$, existe $j \in \{1, 2, \dots, N\}$ tal que

$$W_1(\mu_n, \mu_j) < \varepsilon^2.$$

De fato, se $n \in \mathbb{N}$ escolhe-se $j = N$; se $n < N$ escolhe-se $j = n$. O conjunto $\{\mu_1, \mu_2, \dots, \mu_N\}$ é um conjunto finito de medidas de radon, logo, é rígido. Isto significa que existe um conjunto compacto K tal que

$$\mu_j(X \backslash K) < \varepsilon$$

para todo $j \in \{1, 2, \dots, N\}$. Sendo K compacto, pode-se cobrir K por um número finito de bolas de raio ε:

$$K \subset \bigcup_{i=1}^{m} B(x_i; \varepsilon) =: U$$

Aumentando U pelo fator ε:

$$U_\varepsilon := \{x \in U | d(x,U) < \varepsilon\}$$

pode-se considerar uma função contínua ϕ que satisfaça $X_U \leq \phi \leq X_{U_\varepsilon}$. Pode ser tomado como exemplo de função do tipo como

$$\phi(x) = \left(1 - \frac{d(x,U)}{\varepsilon}\right)_+,$$

que é $1/\varepsilon$-Lipschitz. Tem-se, pela fórmula de dualidade de Kantorovich-Rubinstein,

$$\begin{aligned} \mu_n(U_\varepsilon) &= \int X_{U_\varepsilon}\, d\mu_n \geq \int \phi\, d\mu_n \\ &= \int \phi\, d\mu_j + \int \phi\, d\mu_n - \int \phi\, d\mu_j \\ &\geq \mu_j(U) - \frac{W_1(\mu_n, \mu_j)}{\varepsilon}. \end{aligned}$$

Para $j \leq N$ tem-se que $\mu_j(U) \geq \mu_j(K) \geq 1 - \varepsilon$ e que, para cada n, pode-se usar a afirmação acima para limitar W_1. Logo, para todo $n \in \mathbb{N}$,

$$\mu_n(U_\varepsilon) \geq 1 - 2\varepsilon.$$

Como U_ε pode não ser compacto, faz-se então o seguinte: trocando ε por $\varepsilon/2^{l+1}$ acima, obtém-se

$$\mu_n\left(\bigcup_{i=1}^{m(l)} B(x_i; 2^l)\right) \geq 1 - \varepsilon/2^l$$

e define-se

$$K := \bigcap_{l=1}^{+\infty} \bigcup_{i=1}^{m(l)} \overline{B(x_i; 2^l)}.$$

Assim, por construção, K é totalmente limitado: dado qualquer $\varepsilon > 0$, existe uma cobertura finita por bolas abertas de raio δ. Sendo totalmente limitado e fechado em um espaço métrico completo, K é compacto. ■

Proposição 2.1. *Se X é um espaço de probabilidade metrizável, separável e completo, e $q \in \mathbb{R}_+^*$ fixado, então $(\mathcal{P}_q(X), W_q)$ é um espaço metrizável, separável e completo.*

Demonstração. Será provado inicialmente que $(\mathcal{P}_p(X), W_p)$ é um espaço separável: considera-se $\{x_i\}_{i \in \mathbb{N}} \subset X$ um subconjunto denso enumerável de X e define-se

$$D := \left\{\sum_{i=1}^{N} r_i \delta_{x_i} | r_i \in \mathbb{Q} \text{ e } N \in \mathbb{N}\right\}.$$

Em outras palavras, D é o conjunto de combinações lineares finitas e com coeficientes racionais de massas unitárias nos pontos de $\{x_i\}$. Já que $\mu \in \mathcal{P}_p(X)$, tem-se que, dado $\varepsilon > 0$, existe um compacto $K \subset X$ tal que

$$\int_{X \setminus K} d(x, x_0)^p \, d\mu(x) < \varepsilon^p.$$

Utiliza-se este $\varepsilon > 0$ para obter uma cobertura finita (com centro em alguns x_i) do compacto K:

$$K \subset \bigcup_{i=1}^{N} B\left(x_i; \frac{\varepsilon}{2}\right).$$

Além disso, obtém-se uma cobertura disjunta ao considerar:

$$K \subset \bigcup_{i=1}^{N} \overline{B_k}$$

onde

$$\overline{B_k} := B(x_k; \varepsilon) \setminus \bigcup_{j \leq k} B(x_j; \varepsilon).$$

Define-se

$$f(x) := \begin{cases} x_k, & \text{se } x \in \overline{B_k}\backslash K, k = 1, 2, \ldots, N; \\ x_0, & \text{se } x \in X\backslash K. \end{cases}$$

Nota-se que $f_*\mu \in D$, pois

$$\begin{aligned}(f_*\mu)(A) &= \mu(f^{-1}(A)) \\ &= \mu(X\backslash K)\delta_{x_0} + \sum_{i=1}^{N} \mu(\overline{B_i} \cap K)\delta_{x_i} \\ &=: \sum_{i=1}^{N} a_i \delta_{x_i}\end{aligned}$$

Além disso, tem-se $d(x, f(x)) < \varepsilon$ para todo $x \in K$, de modo que

$$\begin{aligned}\int_X d(x, f(X))^p \, d\mu(X) &= \int_K d(x, f(X))^p \, d\mu(X) + \int_{X\backslash K} d(x, x_0)^p \, d\mu(x) \\ &\leq \varepsilon^p \mu(K) + \varepsilon^p = 2\varepsilon^p.\end{aligned}$$

Utilizando que $(I, f)_*\mu \in \Gamma(\mu, f_*\mu)$, tem-se que

$$W_p(\mu, f_*\mu) \leq 2^{1/p}\varepsilon,$$

o que mostra que D é denso em $\mathcal{P}_p(X)$.

Para mostrar que $\mathcal{P}_p(X)$ é completo, basta supor que X é um espaço métrico completo. De fato, seja μ_k uma sequência de Cauchy em $\mathcal{P}_p(X)$. Então, pelo Lema 2.3, $\{\mu_k\}$ é um conjunto rígido. Segue do Teorema de Prokhorov que existe uma subsequência fracamente convergente $\mu_{k_j} \rightharpoonup \mu$. Por semicontinuidade inferior do custo total, tem-se que

$$\int_X d(x, x_0)^p \, d\mu(x) \leq \liminf_{j\to\infty} \int_X d(x, x_0)^p \, d\mu_{k_j}(x) < +\infty$$

e, logo, $\mu \in \mathcal{P}_p(X)$. Por semicontinuidade da distância W_p, tem-se

$$W_p(\mu, \mu_{k_i}) \leq \liminf_{i\to+\infty} W_p(\mu_{m_j}, \mu_{k_i})$$

de modo que

$$\limsup_{i\to+\infty} W_p(\mu, \mu_{k_i}) \leq \limsup_{i,j\to+\infty} W_p(\mu_{k_j}, \mu_{k_i}) = 0.$$

Segue que $\mu_{k_j} \to \mu \in \mathcal{P}_p(X)$ em distância. Sendo μ_k uma sequência de Cauchy com subsequência convergente, μ_k é convergente. ■

Com a métrica de Wasserstein pode-se induzir uma topologia no espaço $\mathcal{P}(X)$, a qual coincidirá com a topologia fraca* trazida no Capítulo 1, porém, agora, tem-se que o espaço é completo com a métrica, o que permitirá que sejam usados alguns resultados, tais como o Teorema do ponto fixo de Banach.

Capítulo 3

Sistemas de funções iteradas

Neste capítulo será trabalhado os conceitos e resultados relacionados a Medidas Estacionárias para Sistemas de Funções Iteradas (SFI) e o "elo" entre esses conceitos e a Teoria do Transporte Ótimo. Para tal, utiliza-se como principal referência [10].

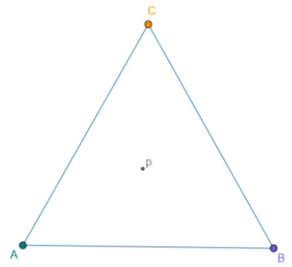
3.1 SFI's e fractais

Neste momento utilizam-se algumas ideias de [19] e [15]. Inicialmente, observa-se uma maneira de gerar fractais, desta forma, começa-se a partir deste ponto, assim, traz-se o procedimento abaixo: Considera-se inicialmente três pontos dispostos em um plano de modo que, se liga-los, formam um triângulo equilátero. Assim, escolhe-se um ponto aleatoriamente dentro da área do triângulo (como na Figura 3.1), após, escolhe-se um dos 3 vértices de maneira aleatória. Em sequência, marca-se o ponto médio entre o ponto escolhido e o vértice escolhido (como na Figura 3.2). Em seguida, escolhe-se novamente um dos 3 vértices aleatoriamente; após, marca-se o ponto médio entre o ponto marcado anteriormente e o vértice que foi escolhido. Repete-se o processo infinitamente, assim, gera-se como figura geométrica, o Triângulo de Sierpinski[1] (A Figura 3.3 representa a imagem gerada após 1012 iterações).

O processo descrito acima é um exemplo clássico de um fractal gerado por um processo regido por uma cadeia de Markov. Tal cadeia de Markov é definida da seguinte forma: Considere as funções $T_1, T_2, T_3 : \mathbb{R}^2 \to \mathbb{R}^2$, onde $T_1(x, y) = (\frac{x}{2}, \frac{y}{2})$, $T_2(x, y) = (\frac{x}{2}+\frac{1}{2}, \frac{y}{2})$, $T_3(x, y) = (\frac{x}{2}+\frac{1}{4}, \frac{y}{2}+\frac{1}{2})$. Sejam $X_0, X_1, \ldots,$ uma sequência de variáveis aleatórias independentes tomando valores em $\{1, 2, 3\}$ com mesma distribuição $\nu = \frac{1}{3}\delta_1 + \frac{1}{3}\delta_2 + \frac{1}{3}\delta_3$. Seja Z_0 uma variável aleatória com valores em $\mathbb{R}^2$ independente de $X = \{X_n : n \geq 0\}$.

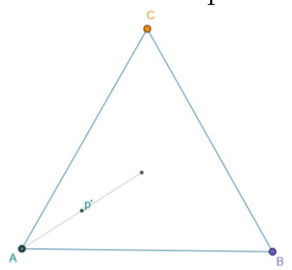
[1] O leitor pode encontrar uma simulação deste processo no site https://apps.univesp.br/jogo-do-caos/

Figura 3.1: Processo após 0 iterações



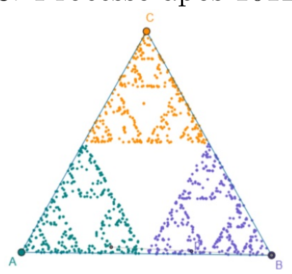
Fonte: imagem gerada através do site `https://apps.univesp.br/jogo-do-caos/`

Figura 3.2: Processo após 1 iteração



Fonte: imagem gerada através do site `https://apps.univesp.br/jogo-do-caos/`

Figura 3.3: Processo após 1012 iterações



Fonte: imagem gerada através do site `https://apps.univesp.br/jogo-do-caos/`

Define-se a sequência de variáveis aleatórias

$$Z_{n+1}(\omega) = T_{X_{n-1}(\omega)} \circ \cdots \circ T_{X_0(\omega)}(Z_0(\omega)) = T_{X_n}(Z_n).$$

Tal sequência de variáveis aleatórias é uma cadeia de Markov homogênea pela Proposição 1.11. Esta sequência é a que desenha a imagem que está sendo gerada, que é conhecida como Triângulo de Sierpinsk.

Sendo $T(A) = T_1(A) \cup T_2(A) \cup T_3(A)$, para $A \subset \mathbb{R}^2$, tem-se que T é chamado de operador de Hutchinson, o qual, possui um único ponto fixo atrator, que é chamado de atrator de Hutchinson. Neste caso, o Triângulo de Sierpinsk. Tal resultado, em um âmbito mais geral, é provado por [9].

Considerando $\Phi = \{T_1, T_2, T_3\}$ e ν definida acima, este é um dos primeiros exemplos de Sistemas de Funções Iteradas (SFI's) que se conhece. Onde será definido com toda a formalidade na próxima seção.

SFI's é um conceito que também é utilizado no processo de compressão de imagens. Tem-se um método de compressão pode reduzir consideravelmente o tamanho de uma imagem quando a comprimi, um desses métodos é chamado por [11] de *fractal image compression* (compressão de imagens fractais - tradução nossa). Esta compressão se faz muito útil no mundo digital, já que em alguns casos se quer realizar transferências de dados mais rápidas e de maneira mais eficiente.

3.2 SFI's que Contraem na Média

Agora serão trabalhados os conceitos relacionados a sistemas de funções iteradas, sua definição, definições auxiliares e resultados.

Seja (X, d) um espaço métrico completo, I um conjunto de indexações e $\Phi = \{\phi_i : i \in I\}$ uma família de funções contínuas $\phi_i : X \to X$. Chama-se de **Sistema de Funções Iteradas** o par (Φ, η), onde $\eta \in \mathcal{P}(I)$.

Um SFI pode ser estudado como uma cadeia de Markov, basta considerar o caso (1.1), assim, define-se a cadeia de Markov $(\mathcal{Z}_n)_{n\geq 0}$ tal que $\mathcal{Z}_n(x) = \phi_{\mathcal{X}_{n-1}(x)} \circ \cdots \circ \phi_{\mathcal{X}_0(x)}(\mathcal{Z}_0(x))$, onde $(\mathcal{X}_n)_{n\geq 0}$ é uma sequência de variáveis aleatórias assumindo valores em I.

Dado um SFI, sabe-se que há casos em que cada ϕ_i é uma contração, basta considerar o trazido na seção anterior à exemplo, porém, pode-se ter uma hipótese um pouco mais fraca sobre o sistema, como o de contração na média.

Definição 3.1. *Diz-se que um SFI **contrai na média** se existirem constantes $q > 0$ e $0 < \rho < 1$ tais que*

$$\int d(\phi_i(x), \phi_i(y))^q \, d\eta(i) \leq \rho d(x, y)^q$$

para todo $x, y \in X$.

Exemplo 3.1. *Sejam $a \in (0, 1)$ e $b \in (1, \frac{1}{a})$, $I = \{0, 1\}$, $\eta \in \mathcal{P}(I)$ e $\Phi = \{\phi_0, \phi_1\}$ tais que $\eta(\{0\}) = \eta(\{1\}) = \frac{1}{2}$ e $\phi_0(x) = ax, \phi_1(x) = bx + 1, \forall x \in \mathbb{R}$. Além disso, considere $q \in (0, q_0)$ onde q_0 é a solução única em $(0, +\infty)$ de $a^{q_0} + b^{q_0} = 2$. Então, (Φ, η) é um sistema de funções iteradas que contrai na média.*

Dado o Exemplo 3.1, se pode mostrar que (Φ, η) contrai na média de maneira simples. Primeiro observa-se que a função $q \mapsto a^q + b^q$ é convexa, onde também tem-se que é monótona decrescente em algum intervalo $(0, q_1)$ e $a^q + b^q \to +\infty$, quando $q \to +\infty$, de modo que esta função assume o valor 2 apenas em 0 e q_0. Além disso

$$\int |\phi_i(x) - \phi_i(y)|^q d\eta(i) = \frac{1}{2}(a^q + b^q)|x - y|^q,$$

assim, o SFI contrai na media quando $q < q_0$, com razão $\rho = \frac{1}{2}(a^q + b^q)$.

Neste capítulo serão utilizados as ferramentas de transporte ótimo trazidas no Capítulo 2 para se extrair resultados pertinentes referentes a SFI's que contraem na média. Antes disso, será apresentado algumas ferramentas, pequenos resultados consideravelmente úteis, referentes a SFI's que contraem na média.

Lema 3.1. *Se (Φ, η) satisfaz $\int d(x_0, \phi_i(x_0))^q d\eta(i) \leq A$, para $q > 0, A > 0$ e x_0 pré fixado, e contrai na média, então para todo $q' \in (0, q)$ ele também satisfaz as hipóteses para as contrantes q', $\rho' := \rho^{q'/q}$ e $A' := A^{q'/q}$.*

Demonstração. Primeiro, observa-se que a função $r \mapsto r^{q'/q}$ é côncava, logo, aplicando a desigualdade de Jensen, têm-se que

$$\begin{aligned}
\int d(\phi_i(x), \phi_i(y))^{q'}\, d\eta(i) &= \int (d(\phi_i(x), \phi_i(y))^q)^{q'/q}\, d\eta(i) \\
&\leq \left(\int d(\phi_i(x), \phi_i(y))^q\, d\eta(i) \right)^{q'/q} \\
&\leq (\rho d(x,y)^q)^{q'/q} \\
&= \rho^{q'/q} d(x,y)^{q'}.
\end{aligned}$$

Aplica-se novamente a desigualdade de Jensen e obtém-se

$$\begin{aligned}
\int d(x_0, \phi_i(x_0))^{q'} d\eta(i) &= \int (d(x_0, \phi_i(x_0))^q)^{q'/q} d\eta(i) \\
&\leq \left(\int d(x_0, \phi_i(x_0))^q d\eta(i) \right)^{q'/q} \\
&\leq (A)^{q'/q}.
\end{aligned}$$

■

Os próximos dois lemas permitem reduzir as condições de contração na média já apresentadas a "outras hipóteses de contração na média", incluindo as utilizadas por [3].

Lema 3.2. *Se (Φ, η) satisfaz*

$$\int \log Lip(\phi_i) d\eta(i) < 0 \text{ e } \exists p > 0, \int Lip(\phi_i)^p d\eta(i) < +\infty,$$

então existe $q > 0$, $\rho \in (0, 1)$ tal que o sistema contrai na média.

Demonstração. A ideia é simplesmente diferenciar $\int Lip(\phi_i)^t d\eta(i)$ com relação a t em $t = 0$; será usado truncamento para diferenciar sob o sinal integral.

Para todo $n \in \mathbb{N}$, considere as funções $f_n : I \to \mathbb{R}$ e $F_n : (0, p] \to \mathbb{R}$ definidas por

$$f_n(i) = \max(Lip(\phi_i), 1/n) \text{ e } F_n(t) = \int f_n(i)^t d\eta(i).$$

Como
$$-\log n \leq \log f_n(i) \leq \frac{1}{p} Lip(\phi_i)^p$$
para todo i, n, as funções $\log f_n$ são η-integráveis. O teorema da convergência monótona implica que
$$\int \log f_n(i) d\eta(i) \longrightarrow \int \log Lip(\phi_i) d\eta(i) \in [-\infty, 0)$$
de modo que para algum $n \in \mathbb{N}$ tém-se
$$\int \log f_n(i) d\eta(i) \in (-\infty, 0).$$
Agora $F_n(0) = 1$ e para todo $t \in [0, \frac{p}{2}]$:
$$-\log(n) \max\{1, f_n(i)^{\frac{p}{2}}\} \leq \frac{d}{dt} f_n(i)^t = \log f_n(i) \cdot f_n(i)^t \leq \frac{2}{p} f_n(i)^{t+\frac{p}{2}}$$
de modo que $\frac{d}{dt} f_n(i)^t$ é η-integrável, uniformemente em t (sendo n fixado acima). A função F_n é, portanto, diferenciável em $[0, \frac{p}{2}]$, com $F_n'(0) = \int \log f_n(i) d\eta(i) < 0$. Concluí-se que existe algum $q \in (0, \frac{p}{2})$ tal que $F_n(q) \in (0, 1)$. Agora
$$\int Lip(\phi_i)^q d\eta(i) \leq \int f_n(i)^q d\eta(i) = F_n(q) < 1$$
o que implica prontamente que o sistema contrai na média. ■

Lema 3.3. *Seja (Φ, η) um SFI. Suponha que existam $L \geq 1$ e $r \in (0, 1)$ tal que $Lip(\phi_i) \leq L$ para todo $i \in I$ e*
$$exp \int \log d(\phi_i(x), \phi_i(y)) d\eta(i) \leq r d(x, y)$$
para todo $x, y \in X$. Então existe $q > 0$ e $\rho \in (0, 1)$ tal que o sistema contai na média.

Demonstração. A aplicação da fórmula de Taylor de segunda ordem com resto de Lagrange a $q \mapsto a^q$ garante que $a^q \leq 1 + q\log(a) + Cq^2$ para todos $a \in (0, L)$ e todos $q \in (0, 1]$, onde $C = \sup\{\frac{a}{2}(\log a)^2 : a \in (0, L]\}$. Fixe $x \neq y \in X$ e para todos $i \in I$ aplique isso a
$$a = d(\phi_i(x), \phi_i(y))/d(x, y)$$
e integre em relação a η para obter:
$$\begin{aligned} \int \left(\frac{d(\phi_i(x), \phi_i(y))}{d(x, y)} \right)^q d\eta(i) &\leq 1 + q \int \log \left(\frac{d(\phi_i(x), \phi_i(y))}{d(x, y)} \right) d\eta(i) + Cq^2 \\ &\leq 1 + q\log(r) + Cq^2 \end{aligned}$$
onde C é independente de x, y. Como $\log(r) < 0$, pode-se encontrar q tal que $1 + q\log(r) + Cq^2 < 1$, assim
$$\int \left(\frac{d(\phi_i(x), \phi_i(y))}{d(x, y)} \right)^q d\eta(i) < 1 \Rightarrow \int d(\phi_i(x), \phi_i(y))^q d\eta(i) < d(x, y)^q,$$
assim, concluí-se a demonstração. ■

3.3 Unicidade da Medida Estacionária

Seja (Φ, η) um sistema de funções iteradas, considera-se a probabilidade de transição $p(x, A) = \int 1_A(\phi_i)\, d\eta$, com $\eta \in \mathcal{P}(I)$, onde I é um conjunto de indexações, e $(X, \mathcal{F}, \mu)$ um espaço de probabilidade. Neste contexto, considera-se $\mu \in \mathcal{P}(X)$ uma medida estacionária para (Φ, η), similarmente como foi feito no Capítulo 1, se satisfaz $\mu = \displaystyle\int (\phi_i)_*\mu\, d\eta(i)$.

Sendo o SFI do Exemplo 3.1, não é simples, nem mesmo direto, encontrar uma medida estacionária para esse sistema, nem mesmo garantir a existência de tal medida, para isso, serão introduzidos alguns conceitos para o estudo desses casos.

Fixado um ponto de referência $x_0 \in X$, para cada $q \in (0, +\infty)$, o **q-ésimo momento** de $\mu \in \mathcal{P}(X)$ é

$$m^q_{x_0}(\mu) := \int d(x, x_0)^q d\mu(x) \in [0, +\infty].$$

O conjunto de medidas de probabilidade μ com q-ésimo momento finito é denotado por $\mathcal{P}_q(X)$, o qual não depende do ponto x_0 fixado.

Definição 3.2. *Define-se o* ***operador de Markov*** *L agindo no espaço $\mathcal{P}(X)$ por*

$$L\mu = \int (\phi_i)_*\mu\, d\eta(i),$$

associado a uma probabilidade de transição que descreve a iteração aleatória das funções de acordo com a distribuição η. Ou seja, $L\mu$ é a lei de $\mathcal{X}_{n+1}$ se $(\mathcal{X}_n)_{n\in\mathbb{N}}$ é uma cadeia de Markov saltando de x para $\phi_i(x)$ com probabilidade $d\eta(i)$ e $\mathcal{X}_n \sim \mu$ (Com distribuição μ).

Desta forma, pode-se considerar que estudar medidas estacionária é, no presente contexto, estudar pontos fixos para o operador de Markov.

De maneira natural, define-se como **operador de Markov dual**, agindo, por exemplo, no espaço de funções mensuráveis limitadas $f : X \to \mathbb{R}$, como

$$Lf(x) = \int f \circ \phi_i(x) d\eta(i).$$

L é um operador positivo que fixa cada função constante, de modo que quando $a \leq f \leq b$ com $a, b \in \mathbb{R}$, então tem-se também $a \leq Lf \leq b$. A relação de dualidade é dada como

$$\int f d(L\mu) = \int Lf d\mu$$

e é uma consequência direta do teorema de Fubini.

Pode-se definir também a extensão do operador de Markov para o espaço de acoplamentos: dado $\gamma \in \Gamma(\mu_0, \mu_1)$, define-se

$$L^*\gamma(A \times B) = \int \gamma((\phi_i^{-1}(A) \times \phi_i^{-1}(B))d\eta(i).$$

Uma propriedade crucial dessa extensão é que L^* manda um acoplamento de μ_0 e μ_1 em um acoplamento de $L\mu_0$ e $L\mu_1$.

Pode-se mostrar isso observando que, para todos $A, B \in \mathcal{F}$:

$$\begin{aligned} L^*\gamma(A \times X) &= \int \gamma((\phi_i^{-1}(A) \times \phi_i^{-1}(X))d\eta(i) \\ &= \int \gamma((\phi_i^{-1}(A) \times X)d\eta(i) \\ &= \int \mu_0((\phi_i^{-1}(A))d\eta(i) = L\mu_0(A) \end{aligned}$$

e

$$\begin{aligned} L^*\gamma(X \times B) &= \int \gamma((\phi_i^{-1}(X) \times \phi_i^{-1}(B))d\eta(i) \\ &= \int \gamma(X \times \phi_i^{-1}(B))d\eta(i) \\ &= \int \mu_1((\phi_i^{-1}(B))d\eta(i) = L\mu_1(B). \end{aligned}$$

O resultado a seguir é uma variante do trazido por [3], o qual garante a existência de uma medida estacionária para sistemas de funções iteradas que contraem na média, onde não é necessário que a cardinalidade do conjunto de indexações I seja finito, e dá uma estimativa para o q-ésimo momento da medida estacionária.

Teorema 3.1. *(variante de Barnsley, Elton 1988). Seja (Φ, η) um SFI em um espaço métrico completo (X, d) e fixe qualquer $x_0 \in X$. Suponha que para algum $q > 0$, $A > 0$, $\rho \in (0, 1)$ o seguinte é válido:*

$$\int d(\phi_i(x), \phi_i(y))^q d\eta(i) \leq \rho d(x, y)^q, \ \forall x, y \in X \tag{3.1}$$

$$\int d(x_0, \phi_i(x_0))^q d\eta(i) \leq A. \tag{3.2}$$

Então (Φ, η) tem uma medida estacionária única $\mu \in \mathcal{P}(X)$; além disso μ tem q-ésimo momento finito:

$$m_{x_0}^q(\mu) \leq \begin{cases} \frac{A}{1-\rho} \text{ quando } q \leq 1, \\ \frac{A}{(1-\rho^{\frac{1}{q}})^q} \text{ quando } q \geq 1. \end{cases}$$

Para a demonstração do Teorema 3.1, será provado os cinco lemas a seguir:

Lema 3.4. *O operador de Markov preserva $\mathcal{P}_q(X)$ e é uma contração da razão não superior a $\overline{\rho} := \rho^{\min(1, \frac{1}{q})}$.*

Demonstração. Seja $\mu_0, \mu_1 \in \mathcal{P}_q(X)$ e escolhe-se um acoplamento ótimo $\gamma \in \Gamma(\mu_0, \mu_1)$ para W_q, o qual existe pelo Corolário 2.1. Então

$$\begin{aligned} W_q(L\mu_0, L\mu_1) &\leq \left(\int d(x,y)^q d(L^*\gamma)(x,y)\right)^{\min\{1,\frac{1}{q}\}} \\ &= \left(\iint d(\phi_i(x), \phi_i(y))^q d\eta(i) d\gamma(x,y)\right)^{\min\{1,\frac{1}{q}\}} \\ &\leq \rho^{\min\{1,\frac{1}{q}\}} \left(\int d(x,y)^q d\gamma(x,y)\right)^{\min\{1,\frac{1}{q}\}} \\ &= \overline{\rho} W_q(\mu_0, \mu_1). \end{aligned}$$

Ou seja, o operador de Markov é uma contração com relação a métrica de Wasserstein com razão não superior a $\overline{\rho}$.

Em particular, todos os elementos de $\mathcal{P}_q(X)$ são enviados a uma distância W_q finita de $L\delta_{x_0}$ e só nos resta provar que $L\delta_{x_0} \in \mathcal{P}_q(X)$, o que segue de (3.2):

$$C_q(\delta_{x_0}, L\delta_{x_0}) = \int d(x,y)^q d(\delta_{x_0} \otimes L\delta_{x_0})(x,y) = \int d(x_0, \phi_i(x_0)) d\eta(i) \leq A.$$

∎

Lema 3.5. *Existe uma medida estacionária única em $\mathcal{P}_q(X)$. Além disso, para todo $\nu \in \mathcal{P}_q(X)$, têm-se que $L^k\nu \to \mu$ na distância W_q, exponencialmente rápido.*

Demonstração. Pelo Lema 3.4, tem-se que L é uma contração, já que $\forall \nu \in \mathcal{P}_q(X)$, $W_q(L^k\nu, \mu) \leq (\rho^{\min\{1,\frac{1}{q}\}})^k W_q(\nu, \mu)$, assim, pelo teorema do ponto fixo de Banach, garante-se a existência e unicidade do ponto fixo $\mu \in \mathcal{P}_q(X)$ para o operador de Markov, o que conclui a prova. ∎

Através do Lema 3.5, consegue-se garantir que no espaço $\mathcal{P}_q(X)$ existe apenas uma única medida estacionária para o SFI, embora ainda não se possa afirmar a unicidade em $\mathcal{P}(X)$, isso é o que será feito a seguir.

Lema 3.6. *$L^k f(x) \to \int f d\mu$ para todas as funções limitadas contínuas $f : X \to \mathbb{R}$ e para todo $x \in X$.*

Demonstração. Como $\delta_x \in \mathcal{P}_q(X)$,

$$L^k f(x) = \int L^k f d\delta_x = \int f d(L^k \delta_x) \to \int f d\mu.$$

∎

Lema 3.7. *Toda medida estacionária tem q-ésimo momento finito, portanto existe uma única medida estacionária em $\mathcal{P}(X)$.*

Demonstração. Para cada $n \in \mathbb{N}$, define-se a função limitada e contínua $f_n : X \to \mathbb{R}$ por

$$f_n(x) = \min(d(x_0, x)^q, n).$$

Seja $\mu' \in \mathcal{P}(X)$ uma medida estacionária, cujo q-ésimo momento não é assumido ser finito. Para todo $n, k \in \mathbb{N}$,

$$\int f_n d\mu' = \int f_n d(L^k \mu') = \int L^k f_n d\mu'.$$

Como $L^k f_n$ é limitado entre 0 e n para todo k, como $\mu \in \mathcal{P}_q(X)$ é uma medida estacionária com q-ésimo momento finito, aplicando o teorema da convergência dominada com $k \to \infty$, obtém-se pelo Lema 3.6 que

$$\int f_n d\mu' = \int \lim_{k\to\infty} L^k f_n d\mu' = \int (\int f_n \, d\mu) d\mu' = \int f_n d\mu \leq \int d(x_0, x)^q d\mu < \infty.$$

Aplicando o teorema da convergência monótona a f_n com $n \to \infty$ mostra então que

$$\int d(x_0, x)^q d\mu' \leq \int d(x_0, x)^q d\mu < \infty,$$

de modo que $\mu' \in \mathcal{P}_q(X)$. Como μ e μ' são medidas estacionárias com q-ésimo momento finito, o Lema 3.5 mostra que $\mu' = \mu$. ■

Lema 3.8. *A única medida estacionária μ satisfaz $m^q_{x_0}(\mu) \leq \frac{A}{(1-\overline{\rho})^{\max(1,q)}}$.*

Demonstração. Seja $\overline{A} = A^{\min(1,\frac{1}{q})}$ e usa-se $L\mu = \mu$ se obtêm:

$$\begin{aligned} W_q(\delta_{x_0}, \mu) &\leq W_q(\delta_{x_0}, L\delta_{x_0}) + W_q(L\delta_{x_0}, L\mu) \\ &\leq \overline{A} + \overline{\rho} W_q(\delta_{x_0}, \mu). \end{aligned}$$

Logo, tem-se que

$$(1 - \overline{\rho}) W_q(\delta_{x_0}, \mu) \leq \overline{A} \Rightarrow W_q(\delta_{x_0}, \mu) \leq \frac{\overline{A}}{(1 - \overline{\rho})}.$$

Se $q \leq 1$, $\overline{A} = A$ e $\overline{\rho} = \rho$; obtém-se

$$\int d(x_0, x)^q d\mu \leq \frac{A}{1 - \rho}.$$

Se $q \geq 1$, $\overline{A} = A^{\frac{1}{q}}$ e $\overline{\rho} = \rho^{\frac{1}{q}}$ obtém-se que

$$\left(\int d(x_0, x)^q d\mu\right)^{\frac{1}{q}} \leq \frac{A^{\frac{1}{q}}}{(1 - \rho^{\frac{1}{q}})} \Rightarrow \int d(x_0, x)^q d\mu \leq \frac{A}{(1 - \rho^{\frac{1}{q}})^q}.$$

■

Demonstração. (Teorema 3.1). A demonstração segue dos Lemas 3.4, 3.5, 3.6, 3.7, 3.8. ■

Agora será utilizado o Teorema 3.1 para estudar um SFI que contrai na média que já foi visto anteriormente.

Corolário 3.1. *Seja o SFI dado no Exemplo 3.1, então o sistema* (Φ, η) *possui uma única medida estacionária* μ*, que possui momentos finitos de todas as ordens* $q \in (0, q_0)$*. Mais precisamente*

$$m_0^q(\mu) \leq \frac{2}{2 - (a^q + b^q)}.$$

Além disso, para todo $q \in (0, \min(1, q_0))$*, toda função contínua* q*-Hölder* $f : \mathbb{R}_+ \to \mathbb{R}$ *e todo* x_0*,*

$$|\frac{1}{2^n} \sum_{\omega \in \{0,1\}^n} f(\phi_\omega(x_0)) - \int f d\mu| \leq C\rho^n$$

onde $\rho = \frac{1}{2}(a^q + b^q) < 1$ *e* $C = Hol_q(f)/(1 - \rho)$*.*

Demonstração. Para todo $q > 0$, tem-se que (3.2) é satisfeita com $A = 1$. Desta maneira, a existência, unicidade e estimativas para os q-ésimos momentos de μ, a medida estacionária para o sistema, segue, portanto, do Teorema 3.1.

A convergência das médias empíricas de f em direção à sua integral em relação a μ segue do Lema 3.4, observando que

$$L^n f(x_0) = \frac{1}{2^n} \sum_{\omega \in \{0,1\}^n} f(\phi_\omega(x_0)).$$

Desta forma, como na prova do Lema 3.6, já que $q < 1$, têm-se que

$$\begin{aligned} |L^n f(x_0) - \int f d\mu| &= \left| \int f d(L^n \delta_{x_0}) - \int f d\mu \right| \\ &\leq Hol_q(f) \cdot W_q(L^n \delta_{x_0}, \mu) \\ &\leq Hol_q(f) \rho^n W_q(\delta_{x_0}, \mu) \\ &= \rho^n Hol_q(f) m_{x_0}^q(\mu) \\ &\leq \rho^n Hol_q(f) \frac{1}{1 - \rho}. \end{aligned}$$

■

Como mais uma aplicação direta do Teorema 3.1, e uma maneira de ilustrar a importância da suposição (3.2), tem-se a seguinte proposição.

Proposição 3.1. *Seja* $q > 0$*; se* $\sum n^q p_n < +\infty$ *então o seguinte SFI tem uma medida estacionária única* μ*, e* μ *tem* q*-ésimo momento finito:* $I = \mathbb{N}$*,* $\eta(n) = p_n$*,* $a \in (0, 1)$*,*

$$\forall n > 0, \forall x \in \mathbb{R} : \phi_n(x) = x + n, \phi_0(x) = ax;$$

onde $p_0 > 0$ *e, claro,* $p_n \geq 0$ *e* $\sum_{n\geq 0} p_n = 1$*. Se* $\sum n^q p_n = +\infty$*, então qualquer medida estacionária do SFI tem q-ésimo momento infinito.*

Demonstração. Tem-se que

$$\begin{aligned}
\int d(\phi_i(x), \phi_i(y))^q d\eta(i) &= \sum_{i\geq 0} p_i d(\phi_i(x)\phi_i(y))^q \\
&= p_0 d(\phi_0(x), \phi_0(y))^q + \sum_{i>0} p_i d(\phi_i(x), \phi_i(y))^q \\
&= p_0 \cdot |ax - ay|^q + \sum_{i>0} p_i |x + i - (y + i)|^q \\
&= p_0 \cdot a^q |x - y|^q + |x - y|^q \sum_{i>0} p_i \\
&= (p_0 a^q + \sum_{i>0} p_i)|x - y|^q \\
&= (p_0 a^q + \sum_{i\geq 0} p_i - p_0)|x - y|^q \\
&= (p_0 a^q + 1 - p_0)|x - y|^q \\
&= (1 + p_0(a^q - 1))|x - y|^q \\
&= (1 - p_0(1 - a^q))|x - y|^q.
\end{aligned}$$

Logo, como $\rho = 1 - p_0(1 - a^q) < 1$, tem-se que o sistema contrai na média, além disso, sendo $x_0 \in \mathbb{R}$, tem-se que:

$$\begin{aligned}
\int d(x_0, \phi_i(x_0))^q d\eta(i) &= \sum_{i\geq 0} p_i d(x_0, \phi_i(x_0))^q \\
&= p_0 |x_0 - ax_0|^q + \sum_{i>0} p_i |x_0 - x_0 + i|^q \\
&= p_0 |x_0|^q |1 - a|^q + \sum_{i>0} p_i \cdot i^q \\
&\leq p_0 |x_0|^q + \sum_{i\geq 0} p_i \cdot i^q.
\end{aligned}$$

Desta forma, se $\sum n^q p_n < +\infty$, tem-se que $\int d(x_0, \phi_i(x_0))^q d\eta(i) < +\infty$, para todo q, logo, o sistema satisfaz (3.2). Assim, utiliza-se o Teorema 3.1 e obtém-se a existência e unicidade da medida estacionária μ do sistema e estimativas para o seu q-ésimo momento. Se $\sum n^q p_n = +\infty$, então $\int d(x_0, \phi_i(x_0))^q d\eta(i) = +\infty$, ou seja, não pode-se utilizar o Teorema 3.1, porém, pelo Lema 3.8, pode-se inferir que, caso exista uma medida estacionária μ para o sistema, então $m^q_{x_0}(\mu) = +\infty$, para todo q.

∎

3.4 Medidas Estacionárias em Skew-products

O caso dos SFI's definidos e estudados acima, onde a aleatoriedade é uma consequência de uma sequência de variáveis aleatórias independentes de lei $\nu \in P(I)$ é um dos casos mais comumente estudados, porém, já há esforços para generalizar este conceito. Uma maneira seria substituir a sequência i.i.d. por uma cadeia de Markov estacionária, outra forma seria desenhar a palavra infinita $\omega = \omega_0 \ldots \omega_k \ldots$ aleatoriamente com a lei sendo uma medida arbitrária de deslocamento invariante $\nu \in P(I^{\mathbb{N}})$. Então pode-se considerar ainda como uma generalização o caso onde o deslocamento é substituído por um sistema dinâmico arbitrário que preserva medida. Este ultimo caso especificamente será o que será estudado a partir de agora. Para isso, precisa-se definir o que é uma função que preserva medida. Seja $(X, \mathcal{F}, \mu)$ um espaço de medida e seja $f : X \to X$ uma transformação mensurável.

Definição 3.3. *Diz-se que f **preserva a medida** μ se $\mu(A) = \mu(f^{-1}(A))$ para todo conjunto mensurável $A \subset X$. Também é dito que μ é invariante por f.*

Definição 3.4. *Considere (X, d) um espaço métrico completo e fixe um espaço de medida padrão (Y, A) (ou seja, é isomorfo a $[0, 1]$ com a σ-álgebra de Borel) equipado com uma medida de probabilidade e uma aplicação que preserva medida $S : Y \to Y$. Uma aplicação de **Skew-products** sobre S com fibra X é uma aplicação*

$$\begin{aligned} \Psi : X \times Y &\to X \times Y \\ (x, y) &\mapsto (\psi_y(x), S(y)) \end{aligned}$$

onde $(x, y) \mapsto \psi_y(x)$ é uma aplicação mensurável.

No caso visto anteriormente, um SFI pode ser estudado dinamicamente observando uma órbita aleatória $x_0, x_{n+1} = \phi_{\omega_n}(x_n)$ onde $(\omega_n)_{n\geq 1}$ são variáveis aleatórias de lei η independentes e identicamente distribuídas, no cenário atual a sequência aleatória correspondente de pontos é dada por $x_{n+1} = \psi_{S_n(y)}(x_n)$ onde y é um elemento aleatório de Y com lei ν, tomando o lugar de toda a sequência $(\omega_1, \omega_2, \ldots)$. Em outras palavras, um SFI corresponde ao caso particular quando $Y = \omega^{\mathbb{N}}$, $\nu = \eta^{\otimes \mathbb{N}}$, S é o deslocamento $y = (y_0, y_1, \ldots) \mapsto S(y) = (y_1, y_2, \ldots)$ e $\psi_y(x) = \phi_{y_0}(x)$.

Denota-se por π^X, π^Y as aplicações de projeção de $X \times Y$ para cada fator. Uma medida $\mu \in \mathcal{P}(X)$ é dita uma **medida estacionária do skew-product** (Ψ, ν) quando existe uma medida $\widehat{\nu} \in P(X \times Y)$ tal que:

$$\widehat{\nu} \text{ é } \Psi - invariante,$$

$$\pi_*^Y \widehat{\nu} = \nu,$$

e

$$\pi_*^X \widehat{\nu} = \mu.$$

Se considerar-se um SFI, a definição coincide com o que já conhece-se como medida estacionária das seções anteriores. A medida $\widehat{\nu}$ trazida acima será chamada de elevação de ν. O que se quer saber agora é se, nas devidas condições, existe uma medida estacionária única. Isso é o que será provado nesta seção.

Para tal, precisa-se do conceito de contração das fibras e de suporte de uma medida.

Definição 3.5. *Diz-se que Ψ* ***contrai as fibras*** *sempre que existir $\rho \in (0,1)$ tal que para todo $y \in Y$, a aplicação ψ_y é ρ-Lipschitz.*

Definição 3.6. *Diz-se que Ψ tem* ***deslocamento limitado*** *se para algum $x_0 \in X$, existe um $A > 0$ tal que o conjunto $d(x_0, \psi_y(x_0)) \leq A$ para todo $y \in Y$.*

O **suporte** *supp* μ da medida μ é o conjunto formado pelos pontos $x \in X$ tais que $\mu(V) > 0$ para qualquer vizinhança V de x.

Teorema 3.2. *Seja Ψ uma aplicação de skew-product em $X \times Y$ que contrai as fibras e tem deslocamento limitado. Cada S-invariante $\nu \in \mathcal{P}(Y)$ tem um levantamento único e, em particular, o skew-product (Ψ, ν) tem uma medida estacionária única μ, que além disso tem suporte limitado.*

Seja $(x_k)_{k\in\mathbb{N}}$ um processo estocástico associado a (Ψ, ν) como acima, com x_0 independente de y e de lei arbitrária $\widetilde{\mu}_0 \in P_q(X)$ para algum $q > 0$, e seja $\widetilde{\mu}_k \in \mathcal{P}(X)$ a lei de x_k. Então para todo $k \in \mathbb{N}$,

$$W_q(\widetilde{\mu}_k, \mu) \leq D\widetilde{\rho}^{-k}$$

onde

$$\widetilde{\rho} = \rho^{\min(q,1)} \in (0,1), D = m_{x_0}^q(\widetilde{\mu}_0)^{\min(1,\frac{1}{q})} + (\frac{A}{1-\rho})^{\min(1,q)},$$

e A,ρ são as constantes nas hipóteses de deslocamento limitado e contração da fibra.

Para a demonstração do teorema, precisa-se de algumas ferramentas, uma delas é uma variação da distância de Wasserstein.

Fixe qualquer $\nu \in \mathcal{P}(Y)$ e faz-se $\mathcal{P}^\nu := (\pi_*^Y)^{-1}(\nu) \subset \mathcal{P}(X \times Y)$ ser a fibra de ν, ou seja, o conjunto de medidas em $X \times Y$ com segunda marginal igual a ν. Faz-se também $\mathcal{P}_q^\nu$ como o subconjunto de P^ν consistindo de medidas de q-ésimo momento finito (o qual não depende de x_0).

O produto $X^2 \times Y$ identifica-se como os pares de pontos no espaço total que se projetam para o mesmo ponto na base Y. Considera-se as aplicações

$$\begin{aligned}\pi_{02} &: (x, x', y) \mapsto (x, y)\\ \pi_{12} &: (x, x', y) \mapsto (x', y)\\ \pi_2 &: (x, x', y) \mapsto y.\end{aligned}$$

Para todo $\sigma_0, \sigma_1 \in P^\nu$ seja $\Gamma^\nu(\sigma_0, \sigma_1) := \{\gamma \in P(X^2 \times Y) | (\pi_{02\,*})\gamma = \sigma_0 \text{ e } (\pi_{12\,*})\gamma = \sigma_1\}$ e então define-se

$$C_q^\nu(\sigma_0, \sigma_1) = \inf_{\gamma \in \Gamma^\nu(\sigma_0,\sigma_1)} \int d(x, x')^q \, d\gamma(x, x', y)$$
$$W_q^\nu(\sigma_0, \sigma_1) = (C_q^\nu(\sigma_0, \sigma_1))^{\min(1, \frac{1}{q})}.$$

Então, para demonstrar o Teorema 3.2, será provado primeiro a Proposição 3.2, para isso, será preciso a definição de desintegração de uma medida (vide [2]), do Teorema da seleção mensurável (retirado de [21]) e do Teorema de Riesz-Fischer.

Suponha que

$$\mu \in P_{\mathbb{P}}(\Omega \times X) := \{\mu \text{ probabilidade em } (\Omega \times X, \mathcal{F} \otimes \mathcal{B}) \text{ com marginal } \mathbb{P} \text{ em } (\Omega, \mathcal{F})\}.$$

Chama-se uma função $\mu(\cdot) : \Omega \times \mathcal{B} \to [0, 1]$ de uma **desintegração** (ou **fatoração**, ou **medida amostral**) de μ em relação a $\mathbb{P}$ se

1. para todo $B \in \mathcal{B}$, $\omega \mapsto \mu_\omega(B)$ é $\mathcal{F}$-mensurável,

2. para $\mathbb{P}$-q.t.p. $\omega \in \Omega$, $B \mapsto \mu_\omega(B)$ é uma medida de probabilidade em $(X, \mathcal{B})$,

3. para todo $A \in \mathcal{F} \otimes \mathcal{B}$
$$\mu(A) = \int_\Omega \int_X \mathbf{1}_A(\omega, x) \mu_\omega(dx) \mathbb{P}(d\omega).$$

Escreve-se

$$\mu(d\omega, dx) = \mu_\omega(dx)\mathbb{P}(d\omega).$$

Teorema 3.3. *(Teorema da seleção mensurável). Seja X um espaço separável, métrico e completo, $\mathcal{B}(X)$ a σ-álgebra de Borel de X, $(\Omega, \mathcal{F})$ um espaço mensurável e ψ uma multifunção em Ω que toma valores no conjunto de subconjuntos fechados não vazios de X.*

Suponha que ψ seja $\mathcal{F}$-fracamente mensurável, ou seja, para todo subconjunto aberto U de X, tem-se

$$\{\omega : \psi(\omega) \cap U \neq \emptyset\} \in \mathcal{F}.$$

Então ψ tem uma seleção que é $\mathcal{F}$-$\mathcal{B}(X)$-mensurável.

Demonstração. Uma demonstração para este teorema pode ser encontrado em [5]. ■

Proposição 3.2. *Para todo $\sigma_0, \sigma_1 \in P^\nu$, o conjunto $\Gamma^\nu(\sigma_0, \sigma_1)$ não é vazio. Se os momentos $m_{x_0}^q(\sigma_i)$ são finitos para $i \in \{0, 1\}$, então $W^\nu(\sigma_0, \sigma_1) < \infty$. Além disso, se $(\xi_y)_{y\in Y}$ e $(\zeta_y)_{y\in Y}$ são as desintegrações de σ_0 e σ_1 com relação a π^Y, então*

$$W_q^\nu(\sigma_0, \sigma_1) = \begin{cases} \left(\int W_q(\xi_y, \zeta_y)^q d\nu(y)\right)^{\frac{1}{q}} & \textit{quando } q \geq 1 \\ \int W_q(\xi_y, \zeta_y) d\nu(y) & \textit{quando } q \leq 1. \end{cases} \tag{3.3}$$

Finalmente, W_q^ν é uma métrica completa no conjunto P_q^ν.

Demonstração. Sejam $(\xi_y)_{y\in Y}$ e $(\zeta_y)_{y\in Y}$ as desintegrações de σ_0 e σ_1 com relação a π^Y, assim, ao identificar X com as fibras de π^Y, $(\xi_y)_{y\in Y}$ é, desta maneira, uma família de medidas em X caracterizada por

$$\int f(x, y) \mathrm{d}\xi_y(x) \mathrm{d}\nu(y) = \int f(x, y) \mathrm{d}\sigma_0(x, y)$$

para todas as funções limitadas contínuas $f : X \times Y \to \mathbb{R}$. De maneira similar, obtém-se, considerando agora em relação a $(\zeta_y)_{y\in Y}$,

$$\int f(x', y) \mathrm{d}\zeta_y(x') \mathrm{d}\nu(y) = \int f(x', y) \mathrm{d}\sigma_1(x', y).$$

De qualquer escolha mensurável de $y \mapsto \gamma_y \in \Gamma(\xi_y, \zeta_y)$, à exemplo, $\gamma_y = \xi_y \otimes \zeta_y$, pode-se construir um elemento $\gamma \in \Gamma^\nu(\sigma_0, \sigma_1)$ definindo

$$\int f(x, x', y) \mathrm{d}\gamma(x, x', y) = \iint f(x, x', y) \mathrm{d}\gamma_y(x, x') \mathrm{d}\nu(y).$$

Em particular $\Gamma^\nu(\sigma_0, \sigma_1)$ não é vazio.

Por outro lado, dado qualquer $\gamma \in \Gamma^\nu(\sigma_0, \sigma_1)$ sua desintegração com relação a π_2 é uma família $(\gamma_y)_{y\in Y}$ de medidas em $X \times X$, e testando contra integrandos da forma $f(x)g(y)$ e $f(x')g(y)$ vê-se que

$$\gamma_y \in \Gamma(\xi_y, \zeta_y)$$

para ν-quase todo y.

Como

$$\int d(x, x')^q \, \mathrm{d}\gamma(x, x', y) = \iint d(x, x')^q \, \mathrm{d}\gamma_y(x, x') \, \mathrm{d}\nu(y) \geq \int \mathrm{C}_q(\xi_y, \zeta_y) \, \mathrm{d}\hat{\mu}(y),$$

tomando um ínfimo obtém-se

$$\mathrm{C}_q^\nu(\mu_0, \mu_1) \geq \int \mathrm{C}_q(\xi_y, \zeta_y) \mathrm{d}\hat{\mu}.$$

Para cada y, o conjunto de planos de transporte ótimos de ξ_y para ζ_y é compacto (vide Teorema 2.1), portanto, pelo teorema da seleção mensurável, há uma família mensurável $(\gamma_y)y \in Y$ tal que para ν-quase todo $y \in Y$,

$$\int d(x,x')^q \mathrm{d}\gamma_y(x,x') = \mathrm{C}_q(\xi_y,\zeta_y).$$

Segue-se que

$$\mathrm{C}_q^\nu(\sigma_0,\sigma_1) \leq \int \mathrm{C}_q(\xi_y,\zeta_y)\mathrm{d}\nu$$

e (3.3) está provado.

Para completar a prova, resta ver que $\mathrm{C}_q^\nu(\sigma_0,\sigma_1) < \infty$ e que W_q^ν é uma métrica tornando $\mathcal{P}_q^\nu$ um espaço completo. A desigualdade triangular segue de (3.3), e então a finitude é obtida observando

$$\mathrm{W}_q^\nu(\sigma_0,\sigma_1) \leq \mathrm{W}_q^\nu(\sigma_0,\delta_{x_0}\otimes\nu) + \mathrm{W}_q^\nu(\delta_{x_0}\otimes\nu,\sigma_1) = m_{x_0}^q(\sigma_0) + m_{x_0}^q(\sigma_1).$$

Finalmente, o Teorema de Riesz-Fischer para funções com valores de espaço métrico garante que W_q^ν é uma métrica completa em $\mathcal{P}_q^\nu$, vista por meio da desintegração como um subconjunto fechado do espaço de aplicações $Y \to \mathcal{P}_q(X)$. ■

Demonstração. (demonstração do Teorema 3.2). Seja ν qualquer medida de probabilidade S-invariante em Y. Primeiro observa-se que as propriedades de contração da fibra e deslocamento limitado garantem que Ψ_* preserva $\mathcal{P}_q^\nu$ para todo q, para isso, basta notar o seguinte, sendo $A \subset Y$ e $\mu \in \mathcal{P}_q^\nu$:

$$\begin{aligned}\Psi_*\mu(X\times A) &= \mu(\Psi^{-1}(X\times A)) = \mu(\psi_y^{-1}(X)\times S^{-1}(A))\\ &= \mu(X\times S^{-1}(A)) = \mu(X\times A) = \nu(A).\end{aligned}$$

Essas suposições uniformes também garantem que para algum conjunto limitado $B \subset X$, o conjunto $B\times Y$ é um *conjunto invariante absorvente*, ou seja, $\Psi(B\times Y) \subset B\times Y$ e para todo $(x,y) \in X\times Y$ existe algum $k \in \mathbb{N}$ tal que $\Psi^k(x,y) \in B\times Y$. Seja de fato $A > 0$ tal que para todo y, $d(x_0,\psi_y(x_0)) \leq A$, fixe qualquer $\varepsilon > 0$, defina $R = \frac{(1+\varepsilon)A}{(1-\rho)}$ e seja $B = B(x_0;R)$ a bola de centro x_0 e raio R em X; então para todo $(x,y) \in X\times Y$

$$\begin{aligned}d(x_0,\psi_y(x)) &\leq d(x_0,\psi_y(x_0)) + d(\psi_y(x_0),\psi_y(x))\\ &\leq A + \rho d(x_0,x).\end{aligned}$$

Quando $x \in B$, o lado direito é no máximo

$$A + \rho R = \frac{1+\varepsilon\rho}{1-\rho}A < R,$$

provando que $B\times Y$ é Ψ-invariante. Quando $x \notin B$, tem-se

$$A < \frac{1-\rho}{1+\varepsilon}d(x_0,x)$$

e o lado direito é no máximo

$$\left(\frac{1-\rho}{1+\varepsilon}+\rho\right)d(x_0,x)=\frac{1+\varepsilon\rho}{1+\varepsilon}d(x_0,x)$$

onde $\frac{1+\varepsilon\rho}{1+\varepsilon}<1$, provando a propriedade absorvente com $k\simeq\log d(x_0,x)$.

Seja $\sigma_0,\sigma_1\in\mathcal{P}_q^\nu$. Consideramos a aplicação $X^2\times Y\to X^2\times Y$ definido por

$$\Psi_2(x,x',y)=(\psi_y(x),\psi_y(x'),S(y)).$$

Para $i\in\{0,1\}$ tem-se $\pi_{i2}\circ\Psi_2=\Psi\circ\pi_{i2}$, como consequência, para qualquer $\gamma\in\Gamma^\nu(\sigma_0,\sigma_1)$ tem-se $\Psi_{2*}\gamma\in\Gamma^\nu(\Psi_*\sigma_0,\Psi_*\sigma_1)$. Observando

$$\begin{aligned}\int d(x,x')^q\mathrm{d}\Psi_{2*}\gamma(x,x',y)&=\int d(\psi_y(x),\psi_y(x'))^q\mathrm{d}\gamma(x,x',y)\\&\leq\rho^q\int d(x,x')^q\mathrm{d}\gamma(x,x',y)\end{aligned}$$

e tomando um ínfimo, vê-se que

$$\mathrm{W}_q^\nu(\Psi_*\sigma_0,\Psi_*\sigma_1)\leq\rho^{\min(1,q)}\mathrm{W}_q^\nu(\sigma_0,\sigma_1),$$

em particular Ψ_* induz uma contração no espaço métrico completo $(\mathcal{P}_q^\nu,\mathrm{W}_q^\nu)$. Portanto, existe um único levantamento $\hat{\nu}$ de ν Ψ-invariante tendo q-ésimo momento finito. Ao considerar diferentes q, já vê-se que a medida $\hat{\nu}$ tem momentos finitos de todas as ordens, mas, como $B\times Y$ é absorvente, qualquer medida Ψ-invariante possui $B\times Y$ como suporte. Prova-se isso através da seguinte observação, se γ é uma medida Ψ-invariante e $k>\log d(x_0,x)$ com $k\in\mathbb{Z}$, então:

$$\begin{aligned}\gamma(B\times Y)&=\gamma(\Psi^{-k}(B\times Y))=\gamma(\psi_y^{-k}(B)\times S^{-k}(Y))\\&=\gamma(\psi_y^{-k}(B)\times Y)=\int_Y\mu(\psi_y^{-k}(B))\,d\nu(y)=1.\end{aligned}$$

Isso prova que $\hat{\nu}$ é o único levantamento Ψ-invariante de ν em todo $\mathcal{P}^\nu$, e que sua primeira marginal μ é suportado em um conjunto limitado. Explicitamente, ao deixar ε acima ir para 0, obtém-se que μ é concentrada em $B(x_0;A/(1-\rho))$.

Considere agora o processo estocástico $(\mathrm{x}_k)_{k\in\mathbb{N}}$. Seja $\sigma_0:=\bar{\mu}_0\otimes\nu$ a lei de $(\mathrm{x}_0,\mathrm{Y})$; então a lei de $(\mathrm{x}_k,S^k(\mathrm{Y}))=\Psi^k(\mathrm{x}_0,\mathrm{Y})$ é $\sigma_k=\Psi_*^k(\sigma_0)$, por definição tem primeira marginal $\bar{\mu}_k$, e por invariância tem segunda marginal ν. Como Ψ_* é uma contração em $\mathcal{P}_q^\nu\ni\sigma_0$, obtém-se que

$$\mathrm{W}_q^\nu(\sigma_k,\hat{\nu})\leq\rho^{k\min(1,q)}\mathrm{W}_q^\nu(\sigma_0,\hat{\nu}).\tag{3.4}$$

Por um lado, usando o plano de transporte obtido pela projeção de um ótimo $\gamma\in\Gamma^\nu(\sigma_k,\hat{\nu})$ nas duas primeiras variáveis, obtém-se

$$\mathrm{W}_q(\bar{\mu}_k,\mu)\leq\mathrm{W}_q^\nu(\sigma_k,\hat{\nu}).$$

Por outro lado,

$$\begin{aligned}\mathrm{W}_q^\nu(\sigma_0, \hat{\nu}) &\leq \mathrm{W}_q^\nu(\bar{\mu}_0 \otimes \nu, \delta_{x_0} \otimes \nu) + \mathrm{W}_q^\nu(\delta x_0 \otimes \nu, \hat{\nu}) \\ &\leq m_{x_0}^q(\bar{\mu}_0)^{\min(1,\frac{1}{q})} + (A/(1-\rho))^{\min(1,q)}\end{aligned}$$

já que $\hat{\nu}$ é concentrada em $B(x_0; A/(1-\rho)) \times Y$. Junto com (3.4), isso conclui a demonstração. ∎

Referências

[1] ARAÚJO, V. D. M. de. **Uma apresentação à Análise Funcional**. 1. ed. Joinville: Clube de Autores, 2023.

[2] ARNOLD, L. Random dynamical systems. In: **Dynamical Systems: Lectures Given at the 2nd Session of the Centro Internazionale Matematico Estivo (CIME) held in Montecatini Terme, Italy, June 13–22, 1994**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. p. 1-43.

[3] BARNSLEY, M. F.; ELTON, J. H. A new class of Markov processes for image encoding. **Advances in Applied Probability**, v. 20, n. 1, p. 14–32, 1988.

[4] BARTLE, R. G. **The elements of integration and Lebesgue measure**. New York: John Wiley & Sons, 2014.

[5] CRAUEL, H. **Random probability measures on Polish spaces**. CRC press, 2002.

[6] DUDLEY, R. M. **Real Analysis and Probability**. California: Wadsworth and Brook Cole, Pacific Grove, 1989.

[7] DURRETT, R. **Probability: theory and examples**. v. 49. Cambridge: Cambridge University Press, 2019.

[8] FARIAS, D. M. **Transporte Ótimo - Notas de Aula**. Notas de Aula, jun. 2018. Disponível em: http://www.mat.ufrgs.br/~dmarcon/transporte2018-1/notas.pdf. Acesso em: 28 jan. 2025.

[9] HUTCHINSON, J. E. **Fractals and self similarity.** Indiana University Mathematics Journal, JSTOR, v. 30, n. 5, p. 713–747, 1981.

[10] KLOECKNER, B. R. Optimal transportation and stationary measures for iterated function systems. In: **MATHEMATICAL Proceedings of the Cambridge Philosophical Society**, v. 173, n. 1, p. 163–187, 2022.

[11] LACROIX, Bruno. **Fractal Image Compression.** 1998. 62 f. Thesis (Master of Science) - Department of Mathematics and Statistics, Carleton University, Ottawa, 1998.

[12] LIMA, E. L. **Espaços Métricos**. Rio de Janeiro: IMPA, 2014.

[13] MEYN, S. P.; TWEEDIE, R. L. **Markov chains and stochastic stability**. 2. ed. Cambridge: Cambridge University Press, 2009.

[14] OLIVEIRA, K.; VIANA, M. **Fundamentos da Teoria Ergódica**. 1. ed. Rio de Janeiro: SBM, 2014.

[15] PEITGEN, H. O. et al. **Chaos and fractals: new frontiers of science**. 2. ed. New York: Springer, 2004.

[16] PROKHOROV, Y. V. Convergence of random processes and limit theorems in probability theory. **Theory of Probability & Its Applications**, v. 1, n. 2, p. 157–214, 1956.

[17] REVUZ, D. **Markov chains**. v. 11. Amsterdam: Elsevier, 1991.

[18] ROLLA, L. T.; LIMA, B. N. B. de. **Probabilidade**. jun. 2023. Disponível em: https://www.ime.usp.br/~leorolla/papers/probabilidade.pdf. Acesso em: 28 jan. 2025.

[19] dos SANTOS, A. B. **UM OLHAR SOBRE A TEORIA DA DIMENSÃO: A DIMENSÃO DE HAUSDORFF DO CONJUNTO DE CANTOR E OUTRAS APLICAÇÕES**. 2023. Trabalho de Conclusão de Curso (Graduação em Matemática) - Universidade Federal do Recôncavo da Bahia, Amargosa, 2023.

[20] VILLANI, C. **Optimal transport: old and new**. v. 338. Berlin: Springer-Verlag, 2009.

[21] **Wikipedia.** Kuratowski and Ryll-Nardzewski measurable selection theorem. Disponível em: https://en.wikipedia.org/wiki/Kuratowski_and_Ryll-Nardzewski_measurable_selection_theorem. Acessado em: 6 de Outubro de 2025.

Universidade Federal da Bahia - UFBA
Instituto de Matemática / Programa de pós-graduação em Matemática

Av. Adhemar de Barros, s/n, Campus Universitário de Ondina, Salvador - BA
CEP: 40170 -110
<http://www.pgmat.ufba.br>