Pontifícia Universidade Católica
do Rio de Janeiro

Marcelo Durães Capeleiro Pinto

Hölder continuity for Lyapunov exponents of random linear cocycles

Dissertação de Mestrado

Dissertation presented to the Programa de Pós–graduação em Matemática, do Departamento de Matemática da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Matemática.

Advisor: Prof. Silvius Klein

Rio de Janeiro
March 2021

Pontifícia Universidade Católica
do Rio de Janeiro

Marcelo Durães Capeleiro Pinto

Hölder continuity for Lyapunov exponents of random linear cocycles

Dissertation presented to the Programa de Pós–graduação em Matemática da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Matemática. Approved by the Examination Committee:

Prof. Silvius Klein
Advisor
Departamento de Matemática – PUC-Rio

Prof. Alexander Eduardo Arbieto Mendoza
Instituto de Matemática – UFRJ

Prof. Lorenzo Justiniano Díaz Casado
Departamento de Matemática – PUC-Rio

Prof. Lucas Henrique Backes
Departamento de Matemática – UFRGS

Prof. Mauricio José Poletti Merlo
Departamento de Matemática – UFC

Rio de Janeiro, March the 25th, 2021

Marcelo Durães Capeleiro Pinto

Holds a Bachelor's degree in Mathematics from Pontifícia Universidade Católica, Rio de Janeiro-RJ, Brasil.

Bibliographic data

Durães Capeleiro Pinto, Marcelo

Hölder continuity for Lyapunov exponents of random linear cocycles / Marcelo Durães Capeleiro Pinto; advisor: Silvius Klein. – 2021.

101 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Matemática, 2021.

Inclui bibliografia

1. Matemática – Teses. 2. Sistemas Dinâmicos. 3. Teoria Ergódica. 4. Expoentes de Lyapunov. 5. Teorema de Oseledets. 6. Medidas estacionárias. 7. Fórmula de Furstenberg. 8. Métrica de Wasserstein. 9. Continuidade Hölder dos expoentes de Lyapunov. I. Klein, Silvius. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Matemática. III. Título.

CDD: 510

To my parents, for their support and encouragement.
To my teachers and professors, for their valuable lessons.
To my students, who keep the motivation despite all the difficulties.

Acknowledgments

First I would like to first thank my parents, Márcia and Antonio, for supporting me during all my life. I would also like to thank Sebastiana and Bob, who are part of my family.

Then, I wish to thank my advisor Silvius Klein for working with me even from distance during the Coronavirus pandemic and also for choosing an interesting topic for my research that made me learn a lot of beautiful Mathematics.

I would also like to thank Alexander Arbieto, Lorenzo Justiniano Díaz Casado, Lucas Backes and Mauricio Poletti for accepting to be part of the jury.

Moreover, I would like to thank all my teachers and professors, who taught me everything that I know.

Also, I thank Creuza and all the administrative staff of the Mathematics Department for helping me with all the bureaucratic requirements.

I would like to thank my colleagues from ELOS and also my students, who inspire me every week.

I wish to thank all my friends from Colégio pH and PUC-Rio, who shared with me a lot of great moments.

I also wish to thank my colleagues from the math department of PUC-Rio, specially Daniel Tolomei, who taught me a lot of Mathematics during my undergrad.

A special thanks to Aline Melo, who helped me a lot throughout my masters.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Finally, I also thank CNPq for all the financial support for this work.

Abstract

Durães Capeleiro Pinto, Marcelo; Klein, Silvius (Advisor). **Hölder continuity for Lyapunov exponents of random linear cocycles**. Rio de Janeiro, 2021. 101p. Dissertação de Mestrado – Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

A compactly supported probability measure on a group of matrices determines a sequence of i.i.d. random matrices. Consider the corresponding multiplicative process and its geometric averages. Furstenberg-Kesten's theorem, the analogue of the law of large numbers in this setting, ensures that the geometric averages of this multiplicative process converge almost surely to a constant, called the maximal Lyapunov exponent of the given measure. This concept can be reformulated in the more general context of ergodic theory using random linear cocycles over the Bernoulli shift.

A natural question concerns the regularity properties of the Lyapunov exponent as a function of the data. Under an irreducibility condition and in a specific setting (which was later generalized by various authors) Le Page established the Hölder continuity of the Lyapunov exponent. Recently, Baraviera and Duarte obtained a direct and elegant proof of this type of result. Their argument uses Furstenberg's formula and the regularity properties of the stationary measure.

Following their approach, in this work we obtain a new result showing that under the same irreducibility hypothesis, the Lyapunov exponent depends Hölder continuously on the measure, relative to the Wasserstein metric, thus generalizing the result of Baraviera and Duarte.

Keywords

Dynamical Systems; Ergodic Theory; Lyapunov exponents; Oseledets Theorem; Stationary measures; Furstenberg's Formula; Wasserstein metric; Hölder Continuity of the Lyapunov Exponents .

Resumo

Durães Capeleiro Pinto, Marcelo; Klein, Silvius. **Continuidade Hölder para os expoentes de Lyapunov de cociclos lineares aleatórios**. Rio de Janeiro, 2021. 101p. Dissertação de Mestrado – Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

Uma medida de probabilidade com suporte compacto em um grupo de matrizes determina uma sequência de matrizes aleatórias i.i.d. Considere o processo multiplicativo correspondente e suas médias geométricas. O teorema de Furstenberg-Kesten, análogo da lei dos grandes números neste cenário, garante que as médias geométricas desse processo multiplicativo convergem quase certamente para uma constante, chamada de expoente de Lyapunov maximal da medida dada. Este conceito pode ser reformulado no contexto mais geral da teoria ergódica usando cociclos lineares aleatórios sobre o shift de Bernoulli.

Uma questão natural diz respeito às propriedades de regularidade do expoente de Lyapunov como uma função dos seus dados. Sob uma condição de irredutibilidade e em um cenário específico (que foi posteriormente generalizado por vários autores) Le Page estabeleceu a continuidade de Hölder do expoente de Lyapunov. Recentemente, Baraviera e Duarte obtiveram uma prova direta e elegante deste tipo de resultado. Seu argumento usa a fórmula de Furstenberg e as propriedades de regularidade da medida estacionária.

Seguindo sua abordagem, neste trabalho obtemos um novo resultado mostrando que, sob a mesma hipótese de irredutibilidade, o expoente de Lyapunov depende Hölder continuamente da medida, relativamente à métrica de Wasserstein, generalizando assim o resultado de Baraviera e Duarte.

Palavras-chave

Sistemas Dinâmicos; Teoria Ergódica; Expoentes de Lyapunov; Teorema de Oseledets; Medidas estacionárias; Fórmula de Furstenberg; Métrica de Wasserstein; Continuidade Hölder dos expoentes de Lyapunov.

Table of contents

1 Introduction 10

2 The Multiplicative Ergodic Theorem 15
2.1 The Subadditive Ergodic Theorem 15
2.2 Linear Cocycles and Furstenberg-Kesten's Theorem 17
2.3 Multiplicative Ergodic Theorem in Dimension 2 20
2.4 Invertible case of Multiplicative Ergodic Theorem in dimension 2 28

3 Stationary measures and Furstenberg's Formula 34
3.1 Stationary measures 34
3.2 u-states and s-states 38
3.3 Furstenberg-Ledrappier's Formula 44
3.4 Coninuity of Lyapunov exponents for irreducible cocycles 53

4 The Wasserstein Metric 57
4.1 A useful metric in the space of probability measures 57
4.2 Kantorovich-Rubinstein's duality Theorem 62
4.3 Convolution of measures 68

5 Hölder Continuity of the Lyapunov exponents 71
5.1 The Markov Operator 72
5.2 Existence and uniqueness of the stationary measure 80
5.3 Hölder continuous dependence on the measure 87
5.4 Generalization of Le Page's Theorem 91

A Elements of Functional Analysis and Ergodic Theory 96

Bibliography 100

It is not knowledge, but the act of learning, not possession but the act of getting there, which grants the greatest enjoyment.

Carl Friedrich Gauss, *Gauss: Titan of Science.*

1
Introduction

Dynamical Systems is a branch of mathematics that studies how processes evolve over time. It is a very large and interesting field that has connections with many different areas of mathematics. Moreover, its concepts and results have a wide range of applications, from the study of celestial mechanics to the prediction of the effects of the construction of a new road on the traffic jam of a city.

Formally, a dynamical system is a pair (X, f) where X is a set and $f\colon X \to X$ is a transformation that acts on X. Therefore, given a point $x \in X$, it is natural to study the behaviour of the iterates $f^n(x)$, which can be thought of as the position of the point x at time n. The sequence described by the iterates

$$x, f(x), f^2(x), ..., f^n(x), ...$$

is called the orbit of x.

It is possible to study the orbit of a point from different points of view. For example, from the topological perspective, one can ask if the orbit has fixed or periodic points, if there are accumulation points or if the orbit is dense in the space or not.

Ergodic Theory is an area that studies Dynamical Systems from a measure theoretical point of view and our work is based on this perspective. Therefore, we consider a quadruple $(X, \mathcal{B}, \mu, f)$, where now $f\colon X \to X$ is a measurable transformation and the new elements are a σ-algebra $\mathcal{B}$ and an f-invariant probability measure μ, which means that $\mu(E) = \mu(f^{-1}E)$ for every $E \in \mathcal{B}$. We say that $(X, \mathcal{B}, \mu, f)$ is a measure preserving dynamical system. A measure preserving dynamical system is called ergodic if every set E that satisfies $f^{-1}(E) = E$ has full measure or zero measure, which means that we cannot split the system into two unconnected sub-systems. Ergodic Systems represent a very important class of dynamical systems.

In this work we consider a process which is the multiplicative analogue of the sum of i.i.d random variables. Consider the group of invertible $d \times d$ matrices $\mathrm{GL}(d)$ and a probability measure μ with compact support $\Sigma \subset \mathrm{GL}(d)$. Let $\{g_n\}_{n\geq 1}$ be an i.i.d multiplicative process with law μ. That is, we choose a matrix g_0 according to the law μ. Then we choose another matrix g_1 according to the same law μ, independently of the previous choice and multiply both matrices $g_1 g_0$. Therefore, after n repetitions we get the product $g_{n-1} \dots g_1 g_0$. We are interested to see how the norm of this product grows, so we study the limit

$$\lambda_+(\mu) := \lim_{n\to\infty} \frac{1}{n} \log \|g_{n-1} \dots g_1 g_0\|.$$

This limit exists μ-almost surely, by a more general result of Furstenberg and Kesten (the multiplicative analogue of the law of large numbers) and it is called the maximal *Lyapunov exponent* of the process.

The previous problem, described in probabilistic terms, can also be described in the more general setting of dynamical systems. One important concept in ergodic theory is that of *linear cocycles*. A linear cocycle is a skew product transformation $F\colon M \times \mathbb{R}^d \to M \times \mathbb{R}^d$ given by a pair (f, A), where $f\colon M \to M$ is our original base transformation (usually ergodic) and $A\colon M \to \mathrm{GL}(d)$ is a measurable map, such that:

$$\begin{aligned} F\colon M \times \mathbb{R}^d &\to M \times \mathbb{R}^d \\ (x, v) &\mapsto (f(x), A(x)v). \end{aligned}$$

The n-th iterate of F is defined as $F^n(x, v) = (f^n(x), A^n(x)v)$, where

$$A^n(x) \;=\; A(f^{n-1}(x))...A(f(x))A(x).$$

As an application of this concept, we can consider a measure ρ on $\mathrm{GL}(d)$ with compact support Σ and let M be the space of sequences $\Sigma^{\mathbb{Z}}$, equipped with the product measure $\mu = \rho^{\mathbb{Z}}$. Let $f\colon M \to M$ be the Bernoulli shift over M and $A\colon M \to \mathrm{GL}(d)$ be the projection of the sequence to its first coordinate. This is an example of what is called a *random linear cocycle*. Its iterates model exactly the previous multiplicative random process.

Moreover, in this more general approach, we can use many tools already developed from ergodic theory to study this process. One of the main tools to study linear cocycles are their associated extremal Lyapunov exponents λ_+ and λ_-. They are functions that depend on the cocycle F, hence on A and f, the point x and also on the measure μ. Since the base transformation is usually fixed and ergodic, we eliminate the dependence on x and f. They are defined by the Furstenberg-Kesten Theorem, which says that if A satisfies some integrability condition, then

$$\lambda_+(A, \mu) = \lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)\| \quad \text{and} \quad \lambda_-(A, \mu) = \lim_{n\to\infty} \frac{1}{n} \log \|(A^n(x))^{-1}\|^{-1}$$

exist μ-almost everywhere.

The continuity of the Lyapunov exponents is a very rich and challenging topic of research since, in general, they are not continuous functions. In order to illustrate this, we introduce the following example.

Example 1.0.1 *Consider two matrices g_1 and g_2 defined by:*

$$g_1 = \begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \quad \textit{and} \quad g_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{1.1}$$

Consider the probability measure $\mu = p_1\delta_{g_1} + p_2\delta g_2$, where p_1 and p_2 are non negative coefficients satisfying $p_1 + p_2 = 1$. Note that when $p_1 > 0$ and $p_2 > 0$ we always have $\lambda_\pm(\mu) = 0$. However, when $p_1 = 1$ and $p_2 = 0$, the Lyapunov exponents are given by $\lambda_\pm(\mu) = \pm \log 2$. Therefore one can observe that Lyapunov exponents are not continuous functions relative to the measure.

Although Lyapunov exponents are not continuous functions in a general setting, one can establish their continuity for some classes of linear cocycles under certain generic hypotheses. In 2017, Pedro Duarte and Alexandre Baraviera proved in [1] (see also [3] for a different approach), a result in this direction, extending the more classical result of Emille Le Page from [5].

In their setting, f is the Bernoulli shift, $A\colon M \to \mathrm{SL}(2)$ depends only on one coordinate of x and the measure μ is fixed. They proved that if A satisfies some generic property of irreducibility and the cocycle has positive Lyapunov exponent, then λ_+ is a locally Hölder continuous function.

Since in their work, the transformation on the base f is fixed (the Bernoulli shift) and the measure μ is fixed, the Lyapunov exponent λ_+ depends only on the map A. Therefore they prove that the map $A \mapsto \lambda_+(A)$ is locally Hölder continuous, when A satisfies some generic hypothesis of irreducibility and $\lambda_+(A) > 0$. Intuitively, irreducibility refers to the non existence of proper invariant subspaces for the cocycle. Precise definitions will be given later.

While studying their work, a natural question arose: what happens if we also let the measure μ vary? Is there any similar result of local Hölder continuity for the Lyapunov exponents in this wider setting?

In this work, we prove a new result that answers this questions. More precisely, we prove the following theorem:

Theorem 1.0.1 *Let μ be a probability measure in $\mathrm{SL}(2)$ with compact support. Suppose that μ is quasi irreducible and $\lambda_+(\mu) > 0$. Then, there exists a neighbourhood V of μ in the weak star topology, such that on V, the map $\nu \mapsto \lambda_+(\nu)$ is Hölder continuous relative to the Wasserstein's metric.*

It is also worth to mention two other results in the same direction and compare all these results. The first one is due to El Hadji Yaya Tall and Marcelo Viana [10], who proved in 2019 that, in the same setting, without the irreducibility hypothesis, the Lyapunov exponents are pointwise Hölder continuous functions with respect to the probability measure. In our work we consider an extra irreducibility hypothesis and we obtain a local result. Also in 2019, Silvius Klein and Pedro Duarte studied the problem without the irreducibility hypothesis, but assuming that the measure is fixed and finitely supported. They proved in [4] that the Lyapunov exponents are locally Hölder continuous functions with respect to the map A. Therefore, they obtained the same conclusion of Baraviera and Duarte by substituting the irreducibility hypothesis by the hypothesis of finite support.

In chapters 2 and 3 we introduce results related to linear cocycles and Lyapunov exponents following the book of Marcelo Viana [11]. In chapter 2 we give a complete and detailed proof of Oseledets Theorem, which is a refined version of Furstenberg-Kesten Theorem, that will be used many times throughout the text.

In order to prove Hölder continuity, we need to estimate the difference between Lyapunov exponents $|\lambda_+(\mu) - \lambda_+(\nu)|$ when the measures μ and ν are close relative to an appropriate metric. Thus, it would be useful to have a formula that describes them. In order to obtain this formula, we dedicate some work in chapter 3 to the study of *stationary measures*, which are a weaker version of invariant measures and also to the study of *u-states* and *s-states* which are special types of measures.

The main goal of chapter 3 is to prove the Furstenberg-Ledrappier Formula. It describes the Lyapunov exponents of random cocycles in terms of stationary measures associated to their projective cocycle, that we denote by $\mathbb{P}F$. Consider the map Φ given by

$$\Phi : M \times \mathbb{PR}^2 \to \mathbb{R}$$
$$\Phi(x, \hat{v}) = \log \frac{\|A(x)v\|}{\|v\|}$$

where the vector $v \in \mathbb{R}^2 \backslash \{0\}$ is a representative of the projective point $\hat{v} \in \mathbb{PR}^2$. Then, the formula is given by

Theorem 1.0.2 *(Furstenberg-Ledrappier's Formula) Let F be a random cocycle over the Bernoulli shift (M, β, μ, f). Then*

$$\lambda_+(F, \mu) = \max \left\{ \int_{M \times \mathbb{PR}^2} \Phi \, d(\mu \times \eta) : \ \eta \text{ is a stationary measure for } \mathbb{P}F \right\}.$$

With this new machinery in hand, a reader that once was only familiar with basic concepts of ergodic theory is now able to understand the proof of Baraviera-Duarte. However, in order to obtain such a quantitative result of Hölder continuity varying the measure, we should also choose a useful metric in the space of probability measures.

Chapter 4 is dedicated to the study of Wasserstein's metric, which is a very interesting metric by itself, but also very useful for our purposes and many others. One of the main topics of the chapter is a duality theorem of Kantorovich-Rubinstein that gives a characterization of Wasserstein's metric. In this chapter we follow the book of Cédric Villani [13] and lecture notes from an advanced measure theory course that took place at PUC-Rio in 2020.

In chapter 5, using the approach of Baraviera and Duarte in [3] and tools developed in chapter 4, we establish out main result, theorem 1.0.1.

The idea of the proof is to use Furstenberg-Ledrappier's Formula to estimate the difference $|\lambda_+(\mu_1) - \lambda_+(\mu_2)|$ when μ_1 and μ_2 are close to μ relative to Wasserstein's metric. The first step will be to study the Markov operator, in order to prove that, under the hypotheses of the theorem, there exist unique stationary measures η_{μ_1} and η_{μ_2} for the cocycles associated to μ_1 and μ_2, respectively. They will be the ones used in Furstenberg-Ledrappier's Formula.

The next step is to use the triangle inequality where, in one of the terms the measure μ_1 is fixed and we need to prove that the stationary measure η_{μ_1} associated to it varies in a Hölder continuous way. In order to estimate the other term, where one of the stationary measures is fixed, we use the Kantorovich-Rubinstein's duality theorem. With both estimates together, we are able to conclude the proof of the theorem.

This dissertation is written aiming at a public of graduate students and researchers that are already familiar with concepts of measure theory, ergodic theory and functional analysis. We also include an appendix with some results from these areas that are used throughout the text, which we hope will help the reader.

2
The Multiplicative Ergodic Theorem

In this chapter we present a detailed proof of the Multiplicative Ergodic Theorem of Oseledets in dimension 2, following chapter 3 of [11] (see also [6]). Moreover, in the course of the proof we describe some technical results from [3] and [2] that will be needed in chapter 5. We provide complete arguments for the relevant technical steps formulated in [11] and [3].

This chapter is organized as follows: in section 2.1 we state Kingman's Theorem and look at some of its consequences. Then, in section 2.2, we define one of the most important concepts in this work: the linear cocycles. In the same section we present Furstenberg-Kesten Theorem as a consequence of Kingman's Theorem and define the concept of Lyapunov Exponents; sections 2.3 and 2.4 are devoted to a detailed proof of Oseledets Theorem for both the one-sided and the two-sided cases.

2.1
The Subadditive Ergodic Theorem

Throughout this chapter $(M, \mathcal{B}, \mu)$ will be a probability space and $f: M \to M$ a measure-preserving transformation.

In this section we present Kingman's Subadditive Theorem, which generalizes Birkhoff's Ergodic Theorem.

We begin by presenting Birkhoff's Ergodic Theorem, which is a classical result in ergodic theory and has several applications in different areas. Then, we prove a corollary that will be used in the proof of the Oseledets Theorem.

Theorem 2.1.1 *(Birkhoff Ergodic Theorem) Consider $\phi : M \to \mathbb{R}$ such that $\phi \in L^1(\mu)$. Then*

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

exists μ almost everywhere. Moreover, $\tilde{\phi} \in L^1(\mu)$ and it is f-invariant with

$$\int \tilde{\phi}\, d\mu = \int \phi\, d\mu.$$

Corollary 2.1.1 *Let $\phi : M \to \mathbb{R}$ be a μ integrable function. Then:*

$$\lim_{n\to\infty} \frac{1}{n}\phi(f^n(x)) = 0 \text{ for } \mu\text{-almost every } x \in M.$$

Proof. We can write $\phi(f^n(x)) = \phi(x) + \sum_{j=0}^{n-1}(\phi \circ f - \phi)(f^j(x))$ for every x and every n. Then, by Birkhoff's Ergodic Theorem applied to the observable $\phi \circ f - \phi$, the limit

$$\lim_{n\to\infty} \frac{1}{n}\phi(f^n(x)) = \lim_{n\to\infty} \frac{1}{n}\phi(x) + \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1}(\phi \circ f - \phi)(f^j(x))$$

exists μ-almost everywhere. Notice that, since μ is $f-$invariant,

$$\mu\left\{x : \left|\frac{1}{n}\phi(f^n(x))\right| \geq c\right\} = \mu\left\{x : |\phi(x)| \geq nc\right\} \to 0 \quad \text{when } n \to \infty.$$

So the sequence $\frac{1}{n}\phi \circ f^n$ converges in measure to zero. Hence there is a subsequence that converges to zero μ-almost everywhere. Since the limit exists, it must be zero for μ-almost every x. ∎

Remark 1 *Corollary* 2.1.1 *also holds if we assume just $\phi \circ f - \phi$ to be integrable.*

Now, we state Kingman's Ergodic Theorem and observe that Birkhoff's Ergodic Theorem can be obtained from it.

Definition 2.1.1 *A sequence $\{\varphi_n\}_{n\geq 1}$ of measurable functions $\varphi_n :\ M \to\ [-\infty, +\infty)$ is called subadditive (relative to f) if*

$$\varphi_{m+n} \leq \varphi_m + \varphi_n \circ f^m \quad \forall m, n \geq 1.$$

Theorem 2.1.2 *(Kingman) Let $\varphi_n : M \to [-\infty, +\infty), n \geq 1$ be a subadditive sequence of measurable functions such that $\varphi_1^+ \in L^1(\mu)$. Then $\left\{\frac{\varphi_n}{n}\right\}_n$ converges $\mu-$almost everywhere to some invariant function φ: $M \to [-\infty, +\infty)$. Moreover, the positive part φ^+ is integrable and*

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

Remark 2 *(Kingman generalizes Birkhoff). Consider $\phi_n = \sum_{j=0}^{n-1} \phi \circ f^j$. Then $\phi_{m+n} = \phi_m + \phi_n \circ f^m$ for every m and n. Hence $(\phi_n)_n$ is an additive sequence, in particular subadditive. Also, by the fact that $\phi : M \to \mathbb{R}$ is μ integrable, we are in the conditions of Kingman's Theorem. Therefore, we are able to apply it and obtain Birkhoff.*

2.2
Linear Cocycles and Furstenberg-Kesten's Theorem

We begin by recalling the concepts of operator norm and conorm of a matrix:

Definition 2.2.1 *Given a matrix $g \in \mathrm{GL}(d)$, the quantities*

$$\|g\| = \sup_{v \in \mathbb{R}^d} \frac{\|gv\|}{\|v\|} \quad \text{and} \quad \|g^{-1}\|^{-1} = \inf_{v \in \mathbb{R}^d} \frac{\|gv\|}{\|v\|}$$

are called respectively the operator norm and the operator conorm of the matrix.

Remark 3 *Note that $\|g\| \geq \|g^{-1}\|^{-1}$ for all $g \in \mathrm{GL}(d)$.*

We now introduce one of the main objects of this work, the linear cocycles.

Let $A : M \to \mathrm{GL}(d)$ be a measurable function.The linear cocycle defined by A over the base transformation f is the skew product map:

$$F : M \times \mathbb{R}^d \to M \times \mathbb{R}^d$$
$$(x, v) \mapsto (f(x), A(x)v).$$

Moreover, the iterates of the linear cocycle F are given by $F^n(x, v) = (f^n(x), A^n(x)v)$ for every $n \geq 1$, where:

$$A^n(x) = A(f^{n-1}(x))...A(f(x))A(x).$$

If f is invertible, so is F. Its inverse is the map $F^{-1} : M \times \mathbb{R}^d \to M \times \mathbb{R}^d$, $F^{-1}(x, v) = (f^{-1}(x), A^{-1}(x)v)$ and the iterates of the inverse cocycle $A^{-1} \colon M \to \mathrm{GL}(d)$ satisfy that for every $n \in \mathbb{N}$ and $x \in M$:

$$A^{-n}(x) = A^n(f^{-n}(x))^{-1} = (A^{-1})^n(x).$$

We also define the adjoint cocycle $F^* : M \times \mathbb{R}^d \to M \times \mathbb{R}^d$, such that $F^*(x, v) = (f^{-1}(x), A(f^{-1}(x))^* v)$. Its iterates are defined as $F^n(x, v) = (f^{-n}(x), (A^*)^n(x)v)$, where $A^* : M \to \mathrm{GL}(d)$ is given by:

$$(A^*)^n(x) = A^n(f^{-n}(x))^* \quad \text{for all } n \geq 1, \ x \in M.$$

Theorem 2.2.1 *(Furstenberg-Kesten) If $\log^+ \|A^{\pm 1}\| \in L^1(\mu)$, then*

$$\lambda_+(x) = \lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)\| \quad \text{and} \quad \lambda_-(x) = \lim_{n\to\infty} \frac{1}{n} \log \|(A^n(x))^{-1}\|^{-1}$$

exist for μ almost every $x \in M$. Moreover, the functions $\lambda_\pm \in L^1(\mu)$ and are f-invariant with

$$\int \lambda_+ \, d\mu = \lim_{n\to\infty} \frac{1}{n} \int \log \|A^n(x)\| \, d\mu \qquad \textit{and}$$
$$\int \lambda_- \, d\mu = \lim_{n\to\infty} \frac{1}{n} \int \log \|(A^n(x))^{-1}\|^{-1} \, d\mu.$$

Proof. This follows immediately from theorem 2.2.1 (although historically it was obtained before). Define

$$\varphi_n(x) = \log \|A^n(x)\| \quad \text{and} \quad \psi_n(x) = \log \|(A^n(x))^{-1}\|.$$

For every g_1 and g_2 in $\mathrm{GL}(d)$, $\|g_1 g_2\| \le \|g_1\|\|g_2\|$. Hence $(\varphi_n)_n$ is subadditive. Similarly, the conorm is supper multiplicative, so $-\psi_n$ is supper additive. Thus, the sequence $(\psi_n)_n$ is subadditive. By hypothesis, φ_1^+ and ψ_1^+ are integrable. Now apply Theorem 2.1.2 to these two sequences and we conclude the proof. ∎

The functions $\lambda_\pm$ are called extremal Lyapunov exponents. Moreover, $\lambda_+ \ge \lambda_-$ because $\|g\| \ge \|g^{-1}\|^{-1}$ for every invertible matrix g.

We can also define other Lyapunov exponents in higher dimensions in a similar way, using the singular values of a matrix. Note that the norm $\|g\|$ and conorm $\|g^{-1}\|^{-1}$ of a matrix g are just particular cases of the singular values (the greatest and the smallest). However, we will work mainly in dimension 2.

The Multiplicative Ergodic Theorem of Oseledets, which is the main topic of this chapter gives us a more precise statement than 2.2.1, because it gives information not only on the growth of the norm of the matrix but also on the growth of columns and linear combinations of them.

Before proceeding to Oseledets Theorem, we present some facts about the Lyapunov exponents when A takes values in $\mathrm{SL}(2)$ and the relation between the Lyapunov exponent of A and its adjoint. We start with the following lemma from linear algebra:

Lemma 2.2.2 *For every $g \in \mathrm{SL}(2)$, it holds that $\|g^{-1}\| = \|g\|$.*

Proof. We will give two proofs of this fact. The first one is more algebraic and the second one more geometric. Let g be a matrix in $\mathrm{SL}(2)$. Then

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad g^* = \begin{pmatrix} a & c \\ b & d \end{pmatrix}, \quad g^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \quad (g^{-1})^* = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}.$$

The norm of g is equal to the greatest singular value of g. Hence it is equal to the greatest eigenvalue of g^*g, which is the greatest root of the characteristic polynomial $P_{g^*g}(\lambda)$. In the same way, the norm of g^{-1} is the greatest root of $P_{(g^{-1})^*g^{-1}}(\lambda)$. A straightforward calculation shows that

$$\begin{aligned} P_{(g^{-1})^*g^{-1}}(\lambda) &= \lambda^2 - \lambda Tr\left((g^{-1})^* g^{-1}\right) + \det\left((g^{-1})^* g^{-1}\right) \\ &= \lambda^2 - \lambda Tr\left(g^* g\right) + \det\left(g^* g\right) \\ &= P_{g^*g}(\lambda). \end{aligned}$$

Since both characteristic polynomials are the same, $\|g^{-1}\| = \|g\|$ for every $g \in \mathrm{SL}(2)$. This finishes the first proof.

There is also a geometric and intuitive way to think about this lemma. A matrix g in SL(2) preserves area and sends the unit circle to an ellipse that must have area equal to π. The area of the ellipse is given by πab where a and b are the semi axes. The semi axes are exactly the norm and the conorm of the matrix. Hence we must have $\|g\|\|g^{-1}\|^{-1} = 1$ and we conclude that $\|g\| = \|g^{-1}\|$. ■

Proposition 2.2.3 *Consider* $A : M \to \mathrm{SL}(2)$*, then* $\lambda_+ \geq 0 \geq \lambda_-$*.*

Proof. Remember that the norm of a matrix is equal to its greatest singular value, which is larger than or equal to its eigenvalues. Since the determinant is the product of the eigenvalues, if $|\det(g)| = 1$, then there exists an eigenvalue greater than or equal to 1. Therefore $\|g\| \geq 1$. Furthermore, by proposition 2.2.2, the conorm $\|g^{-1}\|^{-1} = \|g\|^{-1}$ so $\|g^{-1}\|^{-1} \leq 1$. Applying these observations to the iterates $A^n(x) \in \mathrm{SL}(2)$ of the cocycle, we conclude that $\lambda_+ \geq 0 \geq \lambda_-$. ■

Proposition 2.2.4 *If* $A : M \to \mathrm{SL}(2)$*, then* $\lambda_+ + \lambda_- = 0$*.*

Proof. By Theorem 2.2.1, we know that:

$$\lambda_+(x) = \lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)\| \quad \text{and} \quad \lambda_-(x) = \lim_{n\to\infty} \frac{1}{n} \log \|(A^n(x))^{-1}\|^{-1}.$$

So, by adding these two terms,

$$\begin{aligned} \lambda_+ + \lambda_- &= \lim_{n\to\infty} \frac{1}{n} \left(\log \|A^n(x)\| + \log \|(A^n(x))^{-1}\|^{-1}\right) \\ &= \lim_{n\to\infty} \frac{1}{n} \log\left(\|A^n(x)\| \|(A^n(x))^{-1}\|^{-1}\right) \\ &= \lim_{n\to\infty} \frac{1}{n} \log\left(\|A^n(x)\| \|A^n(x)\|^{-1}\right) \qquad \text{(by lemma 2.2.2)} \\ &= 0. \end{aligned}$$

■

Proposition 2.2.5 *If $\log^+ \|A\| \in L^1(\mu)$, then $\log^+ \|A^*\| \in L^1(\mu)$. Furthermore, if f is ergodic, then a cocycle A and its adjoint A^* have the same Lyapunov exponents: $\lambda_\pm(A) = \lambda_\pm(A^*)$.*

Proof. The first claim follows from the fact that a matrix and its adjoint have the same norm and from the f-invariance of the measure μ:

$$\begin{aligned}
\int_M \log^+ \|A^*(x)\| \, d\mu(x) &= \int_M \log^+ \|A\left(f^{-1}(x)\right)^* \| \, d\mu(x) \\
&= \int_M \log^+ \|A\left(f^{-1}(x)\right) \| \, d\mu(x) \\
&= \int_M \log^+ \|A(x)\| \, d\mu(x).
\end{aligned}$$

The proof of the second statement is similar. By theorem 2.1.2 and using the fact that f is ergodic:

$$\begin{aligned}
\lambda_+(A^*) &= \lim_{n\to\infty} \int_M \frac{1}{n} \log \|A^{*(n)}(x)\| \, d\mu(x) \\
&= \lim_{n\to\infty} \int_M \frac{1}{n} \log \|A^{(n)}\left(f^{-n}(x)\right)^* \| \, d\mu(x) \\
&= \lim_{n\to\infty} \int_M \frac{1}{n} \log \|A^{(n)}\left(f^{-n}(x)\right) \| \, d\mu(x) \\
&= \lim_{n\to\infty} \int_M \frac{1}{n} \log \|A^{(n)}(x)\| \, d\mu(x) \\
&= \lambda_+(A).
\end{aligned}$$

A similar proof also holds for λ_-, by the fact that $\|A^{-1}\|^{-1} = \|A^{*-1}\|^{-1}$. More generally, we can obtain this equality not only for the extremal Lyapunov exponents, but also for every other Lyapunov exponents, by a similar procedure, using the fact that the singular values of A and A^* are the same. ■

2.3 Multiplicative Ergodic Theorem in Dimension 2

In this section we state and prove the Multiplicative Ergodic Theorem of Oseledets in dimension 2 for the one sided case. The invertible case is considered in the next section.

Let $F : M \times \mathbb{R}^2 \to M \times \mathbb{R}^2$ be given by $F(x, v) = (f(x), A(x)v)$, for some measurable function $A : M \to \mathrm{GL}(2)$ satisfying $\log^+ \|A^{\pm 1}\| \in L^1(\mu)$.

Theorem 2.3.1 *(Oseledets) For μ-almost every $x \in M$,*

(1) either $\lambda_-(x) = \lambda_+(x)$ and

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)v\| = \lambda_\pm(x), \quad \textit{for all} \quad v \in \mathbb{R}^2 \backslash \{0\};$$

(2) or $\lambda_+(x) > \lambda_-(x)$ and there exists a vector line $E^s_x \subset \mathbb{R}^2$ such that

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)v\| = \begin{cases} \lambda_-(x), & \text{if } \; v \in E^s_x \backslash \{0\} \\ \lambda_+(x), & \text{if } \; v \in \mathbb{R}^2 \backslash E^s_x. \end{cases}$$

Moreover $A(x)E^s_x = E^s_{f(x)}$ for every x as in (2).

Proof. We begin with the case where A takes values in SL(2) and then we extend the conclusions to the GL(2) setting. Thus, $\lambda_+(x) + \lambda_-(x) = 0$ for μ-almost every $x \in M$. Consider x as in the conclusion of theorem 2.2.1 and $\lambda(x) = \lambda_+(x) = -\lambda_-(x)$.

Assume $\lambda(x) = 0$. Then, for all $v \in \mathbb{R}^2 \backslash \{0\}$:

$$\|A^n(x)^{-1}\|^{-1}\|v\| \leq \|A^n(x)v\| \leq \|A^n(x)\|\|v\|.$$

Hence,

$$\frac{1}{n} \log \left(\|A^n(x)^{-1}\|^{-1}\|v\| \right) \leq \frac{1}{n} \log \left(\|A^n(x)v\| \right) \leq \frac{1}{n} \log \left(\|A^n(x)\|\|v\| \right).$$

Now we take the limit as $n \to \infty$. The left hand side goes to $-\lambda(x)$ and the right hand side goes to $\lambda(x)$ by theorem 2.2.1. So we conclude the first case: if $\lambda_+ = \lambda_-$, then for $\mu-$almost every $x \in M$:

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)v\| = \lambda_\pm(x), \quad \text{for all} \quad v \in \mathbb{R}^2 \backslash \{0\}.$$

Before we proceed to the case $x \in M$ such that $\lambda(x) > 0$, we review an important concept from linear algebra.

Proposition 2.3.2 *Given $g \in$ SL(2) such that $\|g\| \neq 1$, there exist unit vectors s and u such that $\|gu\| = \|g\|$ and $\|gs\| = \|g^{-1}\|^{-1} = \|g\|^{-1}$. These vectors are unique, up to multiplication by -1, they are orthogonal, and their images gs and gu are also orthogonal.*

The pair of vectors u and s are the singular vectors of the matrix g and represent respectively the most and less expanded vectors by g. The singular vectors form an orthogonal basis and their image is also orthogonal because they are eigenvectors associated to different eigenvalues of g^*g.

The geometric meaning of the previous proposition is that the unit circle is transformed into an ellipse by g (not a circle because $\|g\| \neq 1$). Moreover, the semi-axes of the ellipse are exactly s and u.

The unit vectors $s(A^n(x))$ and $u(A^n(x))$, respectively, most contracted and most expanded under $A^n(x)$ will play a key role in the proof of Oseledets

Theorem. When it is clear to what cocycle we are referring to, we will use the notation $s_n(x)$ for $s(A^n(x))$ and $u_n(x)$ for $u(A^n(x))$.

Now consider the case $x \in M$ with $\lambda(x) > 0$. Fix any such x. Since $\|A^n(x)\|$ is approximately $e^{n\lambda(x)}$ for n sufficiently large, in this case there exists some N such that, for every $n > N$, we have that $\|A^n(x)\| > 1$.

Therefore, by proposition 2.3.2, there exist unit vectors $s_n(x)$ and $u_n(x)$, respectively, most contracted and most expanded under $A^n(x)$.

Lemma 2.3.3 *The angle $\angle(s_n(x), s_{n+1}(x))$ decreases exponentially:*

$$\limsup_{n\to\infty} \frac{1}{n} \log(|\sin \angle(s_n(x), s_{n+1}(x))|) \le -2\lambda(x).$$

Proof. We denote the angle $\angle(s_n(x), s_{n+1}(x))$ by α_n. Since $u_{n+1}(x)$ and $s_{n+1}(x)$ are orthogonal, we can write

$$s_n(x) = u_{n+1}(x)\sin(\alpha_n) + s_{n+1}(x)\cos(\alpha_n).$$

Hence, applying $A^{n+1}(x)$ to both sides and using the linearity, we have:

$$A^{n+1}(x)s_n(x) = A^{n+1}(x)u_{n+1}(x)\sin(\alpha_n) + A^{n+1}(x)s_{n+1}(x)\cos(\alpha_n).$$

Moreover, using proposition 2.3.2, $A^{n+1}(x)u_{n+1}(x)$ and $A^{n+1}(x)s_{n+1}(x)$ are orthogonal. So, by Pythagoras Theorem and the fact that u_{n+1} is the most expanded unit vector under A^{n+1}, it follows that

$$\|A^{n+1}(x)s_n(x)\| \ge \|A^{n+1}(x)u_{n+1}(x)\sin(\alpha_n)\| = |\sin(\alpha_n)|\|A^{n+1}(x)\|.$$

On the other hand,

$$\begin{aligned}\|A^{n+1}(x)s_n(x)\| = \|A(f^n(x))A^n(x)s_n(x)\| &\le \|A(f^n(x))\|\|A^n(x)s_n(x)\| \\ &= \|A(f^n(x))\|\|A^n(x)\|^{-1}.\end{aligned}$$

Then, by the previous inequalities we can conclude that:

$$|\sin(\alpha_n)| \le \frac{\|A(f^n(x))\|}{\|A^{n+1}(x)\|\|A^n(x)\|}.$$

Therefore,

$$\limsup_{n\to\infty} \frac{1}{n} \log|\sin(\alpha_n)| \le \limsup_{n\to\infty} \frac{1}{n} \log\left(\frac{\|A(f^n(x))\|}{\|A^{n+1}(x)\|\|A^n(x)\|}\right).$$

We can rewrite the right hand side of the previous inequality as

$$\limsup_{n\to\infty} \frac{1}{n}\Big[\log\left(\|A(f^n(x))\|\right) - \log\Big(\|A^{n+1}(x)\|\Big) - \log\left(\|A^n(x)\|\right)\Big].$$

By corollary 2.1.1 with ϕ equal to $\log\|A\|$, the first term is zero. Moreover, by Furstenberg-Kesten's Theorem, each of the last two terms is equal to $-\lambda(x)$. We conclude that

$$\limsup_{n\to\infty} \frac{1}{n}\log\left(|\sin\angle(s_n(x), s_{n+1}(x))|\right) \leq -2\lambda(x).$$

■

Observe that the corollary used in the previous lemma asks $\log\|A\|$ to be in $L^1(\mu)$ however, by hypothesis, we only know that $\log^+\|A^{\pm 1}\| \in L^1(\mu)$. This is not a problem, because given a measurable function $A: M \to \mathrm{SL}(2)$, these two conditions are equivalent, since $\|A(x)\| \geq 1$ for every $x \in M$.

Definition 2.3.1 *Given two vectors v_1 and v_2 in $\mathbb{R}^d$, we denote their projections into the projective plane by $\hat{v}_1$ and $\hat{v}_2$. The projective distance $\delta\colon \mathbb{P}(\mathbb{R}^d) \times \mathbb{P}(\mathbb{R}^d) \to [0,\infty)$ is given by*

$$\delta(\hat{v}_1, \hat{v}_2) := |\sin\angle(v_1, v_2)| = \frac{\|v_1 \wedge v_2\|}{\|v_1\|\|v_2\|},$$

where the symbol $\wedge$ is the exterior product (wedge product).

Lemma 2.3.4 *The sequence $\{s_n(x)\}_n$ is Cauchy in the projective space.*

Proof. We estimate $\|s_n(x) - s_{n+1}(x)\|$ in the projective space, so we are able to replace $s_j(x)$ by $-s_j(x)$ when necessary and the same for $u_n(x)$. Remember that $u_{n+1}(x)$ and $s_{n+1}(x)$ are unitary, orthogonal and α_n is the angle between s_n and s_{n+1}. Then

$$s_n(x) - s_{n+1}(x) = u_{n+1}(x)\sin(\alpha_n) + s_{n+1}(x)(\cos(\alpha_n) - 1).$$

Hence,

$$\|s_n(x) - s_{n+1}(x)\| \leq \|u_{n+1}(x)\sin(\alpha_n)\| + \|s_{n+1}(x)(\cos(\alpha_n) - 1)\|.$$

Now we are going to bound $|\sin(\alpha_n)|$ and $|\cos(\alpha_n) - 1|$.

First consider $\varepsilon > 0$ such that $-2\lambda(x) + \varepsilon < 0$. Note that we can choose such an ε because $\lambda(x) > 0$. Hence, by the previous lemma,

$$|\sin(\alpha_n)| \leq e^{n(-2\lambda(x)+\epsilon)}$$

for n large enough. The other term, $|\cos(\alpha_n)-1|$ is of order $|\alpha_n^2|$, while $|\sin(\alpha_n)|$ is of order $|\alpha_n|$ when n goes to infinite. Therefore, $|\cos(\alpha_n)-1|$ goes to zero at an exponential rate of order $e^{-4\lambda(x)n}$. We conclude that for some constant $C<\infty$ and for n large enough,

$$\|s_n(x)-s_{n+1}(x)\| \le Ce^{n(-2\lambda(x)+\epsilon)}.$$

Now we are able to estimate $\|s_{n+k}(x)-s_n(x)\|$:

$$\begin{aligned}\|s_{n+k}(x)-s_n(x)\| &\le \|s_{n+k}(x)-s_{n+k-1}(x)\| + ... + \|s_{n+1}(x)-s_n(x)\| \\ &\le Ce^{(n+k-1)(-2\lambda(x)+\epsilon)} + ... + Ce^{n(-2\lambda(x)+\epsilon)} \\ &\le Ce^{n(-2\lambda(x)+\epsilon)}\left[1+e^{(-2\lambda(x)+\epsilon)} + ... + e^{(k-1)(-2\lambda(x)+\epsilon)}\right] \\ &\le \overline{C}e^{n(-2\lambda(x)+\epsilon)}.\end{aligned}$$

We used the triangle inequality and the fact that we had a geometric progression with common ratio $e^{-2\lambda(x)+\epsilon}<1$.

In particular we conclude that the sequence is Cauchy in the projective space.

■

Proposition 2.3.5 *If a_n, $b_n>0$ for every n, then*

(a) $\limsup\limits_{n\to\infty}\dfrac{1}{n}\log(a_n+b_n) = \max\left\{\limsup\limits_{n\to\infty}\dfrac{1}{n}\log a_n, \limsup\limits_{n\to\infty}\dfrac{1}{n}\log b_n\right\}$

(b) $\liminf\limits_{n\to\infty}\dfrac{1}{n}\log(a_n+b_n) \ge \max\left\{\liminf\limits_{n\to\infty}\dfrac{1}{n}\log a_n, \liminf\limits_{n\to\infty}\dfrac{1}{n}\log b_n\right\}$

Proof.

(a) Consider $a_n'=\min\{a_n,b_n\}$ and $b_n'=\max\{a_n,b_n\}$. Observe that

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log(a_n+b_n) &= \limsup_{n\to\infty}\frac{1}{n}\log(a_n'+b_n') \\ &= \limsup_{n\to\infty}\frac{1}{n}\left(\log(1+\frac{a_n'}{b_n'})+\log b_n'\right).\end{aligned}$$

We can also see that

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\left(\log b_n'\right) &\le \limsup_{n\to\infty}\frac{1}{n}\left(\log(1+\frac{a_n'}{b_n'})+\log b_n'\right) \\ &\le \limsup_{n\to\infty}\frac{1}{n}\left(\log(2)+\log b_n'\right).\end{aligned}$$

By these previous inequalities we conclude that

$$\limsup_{n\to\infty} \frac{1}{n} \log(a_n + b_n) = \limsup_{n\to\infty} \frac{1}{n} \log(\max\{a_n, b_n\}).$$

Since log is an increasing function, $\log(\max\{a_n, b_n\}) = \max\{\log a_n, \log b_n\}$. It remains to show that

$$\limsup_{n\to\infty} \max\{x_n, y_n\} = \max\{\limsup_{n\to\infty} x_n, \limsup_{n\to\infty} y_n\},$$

where $x_n = \log a_n$ and $y_n = \log b_n$.

By the definition of lim sup, there exists a subsequence $(\max\{x_{n_k}, y_{n_k}\})_{n_k}$ such that $\limsup\limits_{n\to\infty} \max\{x_n, y_n\} = \lim\limits_{k\to\infty} \max\{x_{n_k}, y_{n_k}\}$. This subsequence must have infinite elements from x_{n_k} or y_{n_k}. Without loss of generality suppose it has infinite elements from x_{n_k}. Then, there exists a subsequence $\left(\max\{x_{n_{k_l}}, y_{n_{k_l}}\}\right)_{n_{k_l}}$ such that for all l, $\max\{x_{n_{k_l}}, y_{n_{k_l}}\} = x_{n_{k_l}}$. Hence,

$$\limsup_{n\to\infty} \max\{x_n, y_n\} = \lim_{l\to\infty} \max\{x_{n_{k_l}}, y_{n_{k_l}}\} = \lim_{l\to\infty} x_{n_{k_l}}.$$

Moreover,

$$\lim_{l\to\infty} x_{n_{k_l}} \leq \limsup x_n \leq \max\{\limsup_{n\to\infty} x_n, \limsup_{n\to\infty} y_n\},$$

which implies that $\limsup\limits_{n\to\infty} \max\{x_n, y_n\} \leq \max\{\limsup\limits_{n\to\infty} x_n, \limsup\limits_{n\to\infty} y_n\}$. On the other hand, it is clear that

$$\limsup_{n\to\infty} \max\{x_n, y_n\} \geq \max\{\limsup_{n\to\infty} x_n, \limsup_{n\to\infty} y_n\},$$

which concludes the first part of the proposition.

(b) This proof is analogous to the previous item. Notice that

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{n} \log(a_n + b_n) &= \liminf_{n\to\infty} \frac{1}{n} \log(\max\{a_n, b_n\}) \\ &= \liminf_{n\to\infty} \frac{1}{n} \max\{\log a_n, \log b_n\}. \end{aligned}$$

Therefore, we conclude that

$$\liminf_{n\to\infty} \frac{1}{n} \max\{\log a_n, \log b_n\} \geq \max\{\liminf_{n\to\infty} \frac{1}{n} \log a_n, \liminf_{n\to\infty} \frac{1}{n} \log b_n\},$$

which finishes the proof.

■

Define $s(x) = \lim_{n\to\infty} s_n(x)$. As $\{s_n(x)\}_n$ is a cauchy sequence and the projective space is compact, hence complete, this limit exists.

Lemma 2.3.6 *The vector $s(x)$ is contracted at the rate $-\lambda(x)$:*

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)s(x)\| = -\lambda(x).$$

Proof. Let $\beta_n = \angle\left(s(x), s_n(x)\right)$. We can write $s(x)$ in terms of $u_n(x)$ and $s_n(x)$, as follows:

$$s(x) = u_n(x)\sin(\beta_n) + s_n(x)\cos(\beta_n).$$

Applying $A^n(x)$, using its linearity and then using the triangle inequality, we have:

$$\|A^n(x)s(x)\| \leq |\cos\beta_n|\|A^n(x)s_n(x)\| + |\sin\beta_n|\|A^n(x)u_n(x)\|$$

Now we are going to take the logarithm, multiply by $\frac{1}{n}$, apply the lim sup to both sides of the equation and use the first item of the previous proposition:

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log \|A^n(x)s(x)\| \leq \max \Big\{ & \limsup_{n\to\infty} \frac{1}{n} \log\left(|\cos\beta_n|\|A^n(x)s_n(x)\|\right), \\ & \limsup_{n\to\infty} \frac{1}{n} \log\left(|\sin\beta_n|\|A^n(x)u_n(x)\|\right) \Big\}.\end{aligned}$$

Since $s_n(x) \to s(x)$, we have $\cos\beta_n = \cos\angle\left(s(x), s_n(x)\right) \to \cos 0 = 1$. Then:

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} |\cos\beta_n|\|A^n(x)s_n(x)\| &= \limsup_{n\to\infty} \frac{1}{n} \|A^n(x)s_n(x)\| \\ &= \limsup_{n\to\infty} \frac{1}{n} \|A^n(x)^{-1}\|^{-1} \\ &= -\lambda(x).\end{aligned}$$

We already know how to estimate the other term:

$$\begin{aligned}&\limsup_{n\to\infty} \frac{1}{n} \log\left(|\sin\beta_n|\|A^n(x)u_n(x)\|\right) \\ &= \limsup_{n\to\infty} \frac{1}{n} \log\left(|\sin\beta_n|\right) + \limsup_{n\to\infty} \frac{1}{n} \log \|A^n(x)u_n(x)\| \\ &\leq -2\lambda(x) + \lambda(x) = -\lambda(x).\end{aligned}$$

By this previous calculation we showed that

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)s(x)\| \leq -\lambda(x).$$

To prove the other inequality, we use the second part of the previous proposition:

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n}\log\|A^n(x)s(x)\| \geq \max\Big\{&\liminf_{n\to\infty}\frac{1}{n}\log\left(|\cos\beta_n|\|A^n(x)s_n(x)\|\right),\\ &\liminf_{n\to\infty}\frac{1}{n}\log\left(|\sin\beta_n|\|A^n(x)u_n(x)\|\right)\Big\}\\ \geq{}& -\lambda(x).\end{aligned}$$

Together, both inequalities imply the claim of the lemma. ■

Note that we can take the line $\mathbb{R}s(x)$ generated by $s(x)$ to be E^s_x in Oseledets Theorem. Thus if a vector v belongs to $E^s_x\backslash\{0\}$, then $\lim_{n\to\infty}\frac{1}{n}\log\|A^n(x)v\| = -\lambda(x)$. In the next lemma we consider the case $v \notin E^s_x$.

Lemma 2.3.7 *If $v \in \mathbb{R}^2$ is not colinear with $s(x)$ then*

$$\lim_{n\to\infty}\frac{1}{n}\log\|A^n(x)v\| = \lambda(x).$$

Proof. Denote $\gamma_n = \angle(v, s_n(x))$. Then we can decompose v as: $v = u_n(x)\sin(\gamma_n) + s_n(x)\cos(\gamma_n)$. Applying $A^n(x)$ to both sides and taking the norm,

$$\|A^n(x)v\| = \|A^n(x)u_n(x)\sin(\gamma_n) + A^n(x)s_n(x)\cos(\gamma_n)\|.$$

Since $A^n(x)u_n(x)$ and $A^n(x)s_n(x)$ are orthogonal,

$$\|A^n(x)v\|^2 = |\sin(\gamma_n)|^2\|A^n(x)u_n(x)\|^2 + |\cos(\gamma_n)|^2\|A^n(x)s_n(x)\|^2,$$

by the Pythagorean Theorem. Hence,

$$\begin{aligned}\|A^n(x)v\| &\geq |\sin(\gamma_n)|\|A^n(x)u_n(x)\|\\ \frac{1}{n}\log\|A^n(x)v\| &\geq \frac{1}{n}\log|\sin(\gamma_n)| + \frac{1}{n}\log\|A^n(x)u_n(x)\|\\ \liminf_{n\to\infty}\frac{1}{n}\log\|A^n(x)v\| &\geq \lambda(x).\end{aligned}$$

In the last inequality we used the fact that v is not collinear to $s(x)$. For the other inequality, we use $\|A^n(x)v\| \leq \|A^n(x)\|\|v\|$. So,

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{n}\log\|A^n(x)v\| \leq \limsup_{n\to\infty}\frac{1}{n}\log\|A^n(x)\| + \limsup_{n\to\infty}\frac{1}{n}\log\|v\|\\ &\limsup_{n\to\infty}\frac{1}{n}\log\|A^n(x)v\| \leq \lambda(x).\end{aligned}$$

By putting both inequalities together we finish the proof of the lemma. ■

To finish the proof of the theorem, it remains to show that $A(x)E^s_x = E^s_{f(x)}$. This will be the purpose of the next lemma.

Lemma 2.3.8 *$A(x)s(x)$ is collinear to $s(f(x))$.*

Proof. By lemma 2.3.6, we know that

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n(f(x))A(x)s(x)\| = \lim_{n\to\infty} \frac{1}{n+1} \log \|A^{n+1}(x)s(x)\| = -\lambda(x).$$

By lemma 2.3.7, for every v non collinear to $s(f(x))$,

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n(f(x))v\| = \lambda(f(x)).$$

Remember that, by theorem 2.2.1, λ is f-invariant. Therefore,

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n(f(x))v\| = \lambda(x)$$

for every v not collinear to $s(f(x))$. So, $A(x)s(x)$ is collinear to $s(f(x))$. ∎

Together these lemmas complete the proof of Theorem 2.3.1. ∎

2.4 Invertible case of Multiplicative Ergodic Theorem in dimension 2

In this section, complete the proof of Oseledets Theorem in the invertible case.

Theorem 2.4.1 *If $f : M \to M$ is invertible, then for μ almost every $x \in M$, we have*

(1) either $\lambda_-(x) = \lambda_+(x)$ and

$$\lim_{n\to\pm\infty} \frac{1}{n} \log \|A^n(x)v\| = \lambda_\pm(x), \quad \textit{for all} \quad v \in \mathbb{R}^2;$$

(2) or $\lambda_+(x) > \lambda_-(x)$ and there exists a direct sum decomposition $\mathbb{R}^2 = E^s_x \oplus E^u_x$ such that

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)v\| = \begin{cases} \lambda_-(x), & \textit{if} \quad v \in E^s_x \backslash \{0\} \\ \lambda_+(x), & \textit{if} \quad v \in \mathbb{R}^2 \backslash E^s_x \end{cases}$$

$$\lim_{n\to-\infty} \frac{1}{n} \log \|A^n(x)v\| = \begin{cases} \lambda_+(x), & \textit{if} \quad v \in E^u_x \backslash \{0\} \\ \lambda_-(x), & \textit{if} \quad v \in \mathbb{R}^2 \backslash E^u_x \end{cases}$$

Moreover, in the second case, $A(x)E^s_x = E^s_{f(x)}$ *and* $A(x)E^u_x = E^u_{f(x)}$ *and the angle between the two lines decays subexponentially fast along the orbit:*

$$\lim_{n\to\pm\infty} \frac{1}{n} \log|\sin\angle(E^u_{f^n(x)}, E^s_{f^n(x)})| = 0.$$

Proof. The case $\lambda(x) = 0$ follows from Theorem 2.3.1 applied to F and to F^{-1}.

Now we deal with the case $\lambda(x) > 0$. Let $E^s_x = \mathbb{R}s(x)$ and $E^u_x = \mathbb{R}u(x)$ be the subspaces given by Theorem 2.3.1 for F and F^{-1}, respectively. Before we continue with the proof, it is worth to state some remarks that will be useful later in chapter 5.

As $E^s_x = \lim_{n\to\infty} s_n(x)$, one should expect that $\lim_{n\to\infty} u_n(x) = E^u_x$. However, this is false. Indeed, the true statement is:

Proposition 2.4.2 *If* $\lambda_+(A) > 0$ *and* f *is ergodic, then for* μ*-almost every* $x \in M$*,*

$$\lim_{n\to\infty} u_n(x) = E^u_x(A^*).$$

Proof. First note that since $\lambda_+(A^*) = \lambda_+(A) > 0$, the most and least expanded directions are well defined for the cocycle A^*.

Also, for every $g \in \mathrm{GL}_2(\mathbb{R})$, a direct calculation shows that the singular vector of g^{-1} associated to the greatest singular value of g^{-1} is equal to the singular vector of g^* associated to the smaller singular value of g^* and vice versa. That is,

$$u(g^{-1}) = s(g^*) \quad \text{and} \quad s(g^{-1}) = u(g^*).$$

Therefore, since we defined $E^u_x(A) = \lim_{n\to\infty} s(A^{-n}(x))$, it holds for μ-almost every $x \in M$ that

$$\begin{aligned} E^u_x(A^*) &= \lim_{n\to\infty} s(A^{*(-n)}(x)) \\ &= \lim_{n\to\infty} u((A^{(-1)(-n)})(x)) \\ &= \lim_{n\to\infty} u(A^{(n)}(x)) \\ &= \lim_{n\to\infty} u_n(x). \end{aligned}$$

■

Lemma 2.4.3 *The vectors* $s(x)$ *and* $u(x)$ *are non-collinear, for* $\mu-$*almost every point in* $\{x : \lambda(x) > 0\}$*.*

Proof. By assumption, E^u_x is the line generated by the less expanded direction for $A^{-n}(x)$. Hence, $\lim_{n\to\infty} \frac{1}{n}\log\|A^{-n}(x) \mid E^u_x\| = -\lambda(x)$ and then $\lim_{n\to-\infty} \frac{1}{n}\log\|A^n(x) \mid E^u_x\| = \lambda(x)$.

So, in order to prove the lemma, we need to show that:

$$\lim_{n\to-\infty}\frac{1}{n}\log\|A^n(x)\mid E^s_x\| = -\lambda(x).$$

By Oseledets Theorem, the previous limit exist for almost every $x\in M$. Denote this limit by $\psi(x)$ and consider the sequences of functions:

$$\psi_n(x)=\frac{1}{-n}\log\|A^{-n}(x)\mid E^s_x\| \quad\text{and}\quad \phi_n(y)=\frac{1}{-n}\log\left\|\left(A^n(y)\mid E^s_y\right)^{-1}\right\|.$$

Our goal is to show that ψ_n converges in measure to $-\lambda$. This will be sufficient for proving the lemma, because we already know that ψ_n converges almost everywhere to ψ, hence it converges in measure to the same limit. Finally, by the uniqueness of the limit, it will follow that $\psi=-\lambda$.

From the definition of A^{-n}, we can see that

$$\psi_n(x)=\frac{1}{-n}\log\|(A^n(f^{-n}(x)\mid E^s_x)^{-1}\| = \phi_n(f^{-n}(x)) \quad\text{for every } n\geq 1.$$

Since E^s is one dimensional,

$$\phi_n(y)=\frac{1}{n}\log\|\left(A^n(y)\mid E^s_y\right)\|.$$

So, $\lim_{n\to\infty}\phi_n(y)=-\lambda(y)$ for $\mu-$almost every $y\in M$. This implies that ϕ_n converges in measure to $-\lambda(y)$:

$$\lim_{n\to\infty}\mu\left(\{y:|\phi_n(y)+\lambda(y)|>\delta\}\right)=0 \quad\text{for every } \delta>0.$$

Since f is a measure preserving transformation:

$$\lim_{n\to\infty}\mu\left(\left\{y:|\phi_n(f^{-n}(y))+\lambda(f^{-n}(y))|>\delta\right\}\right)=0 \quad\text{for every } \delta>0.$$

Now, using the fact that $\psi_n(x)=\phi_n(f^{-n}(x))$ and the fact that the Lyapunov Exponent is an invariant function:

$$\lim_{n\to\infty}\mu\left(\{y:|\psi_n(y)+\lambda(y)|>\delta\}\right)=0 \quad\text{for every } \delta>0.$$

So we concluded that ψ_n converges in measure to $-\lambda$ and, by the previous arguments, $\psi=-\lambda$. Consequently the vectors $s(x)$ and $u(x)$ are non-collinear, for $\mu-$almost every point in $\{x:\lambda(x)>0\}$.

■

In fact, there is a more general statement for an ergodic f:

Proposition 2.4.4 *If a cocycle A has Lyapunov exponent $\lambda_+ > 0$ and f is ergodic, then for μ-almost everywhere:*

$$E^u_x(A) = E^s_x(A^*)^\perp \quad \textit{and} \quad E^s_x(A) = E^u_x(A^*)^\perp.$$

In particular, we conclude that for μ-almost everywhere:

$$E^u_x(A) = \lim_{n\to\infty} u\left(A^{*(n)}(x)\right) \quad \textit{and} \quad E^s_x(A) = \left(\lim_{n\to\infty} u\left(A^{(n)}(x)\right)\right)^\perp.$$

Proof. Note that if we prove $E^u_x(A) = E^s_x(A^*)^\perp$ μ-a.e, then by considering the cocycle A^* instead of A and the fact that $A(x)^{**} = A(x)$, we can conclude that $E^s_x(A) = E^u_x(A^*)^\perp$ μ-a.e. Hence it is sufficient to prove the first equality.

By proposition 2.4.2, the first equality says that

$$\lim_{n\to\infty} u\left(A^{*(n)}(x)\right) = \lim_{n\to\infty} s\left(A^{*(n)}(x)\right)^\perp \quad \mu\text{-a.e.}$$

For every $n \geq 1$ and μ-a.e, we know that $\left\langle u\left(A^{*(n)}(x)\right), s\left(A^{*(n)}(x)\right)\right\rangle = 0$. Since the inner product is a continuous function, the statement holds.

Also by proposition 2.4.2, the fact that $A(x)^{**} = A(x)$ and the previous results in this proposition, we conclude that for μ-almost everywhere $E^u_x(A) = \lim_{n\to\infty} u\left(A^{*(n)}(x)\right)$ and $E^s_x(A) = \left(\lim_{n\to\infty} u\left(A^{(n)}(x)\right)\right)^\perp$. ■

Proposition 2.4.5 *For every $g \in \mathrm{SL}(2)$ and non-zero vectors $u, v \in \mathbb{R}^2$, the angle between gu and gv must satisfy the following relation:*

$$\|g\|^{-2} \leq \frac{|\sin\angle(g(u), g(v))|}{|\sin\angle(u, v))|} \leq \|g\|^2.$$

Proof. Since $g \in \mathrm{SL}_2(\mathbb{R})$, it preserves area, so $\|gp \wedge gq\| = \|p \wedge q\|$. Then,

$$\begin{aligned} \frac{|\sin\angle(g(u), g(v))|}{|\sin\angle(u, v))|} &= \frac{\|gp \wedge gq\|}{\|gp\|\|gq\|}\frac{\|p\|\|q\|}{\|p \wedge q\|} \\ &= \frac{1}{\|gp\|}\frac{1}{\|gq\|}. \end{aligned}$$

By definition 2.2.1,

$$\frac{1}{\|g\|^2} \leq \frac{1}{\|gp\|}\frac{1}{\|gq\|} \leq \|g\|^2,$$

which concludes the proof. ■

Lemma 2.4.6 *Let $\theta(y) = \angle(E^s_y, E^u_y)$. For μ-almost every x with $\lambda(x) > 0$,*

$$\lim_{n\to\infty} \frac{1}{n} \log|\sin\theta(f^n(x))| = 0.$$

Proof. We prove that $\log|\sin\theta|\circ f-\log|\sin\theta|\in L^1(\mu)$. Hence, by the remark of 2.1.1 we conclude the result.

By proposition 2.4.5, applied to our setting:

$$\|A(x)\|^{-2}\le\frac{|\sin\theta(f(x))|}{|\sin\theta(x)|}\le\|A(x)\|^2,$$

where

$$\sin\theta(f(x)) \;=\; \sin\left(\angle(E^s_{f(x)},E^u_{f(x)})\right)=\sin\left(\angle(A(x)E^s_x,A(x)E^u_x)\right).$$

Then, since log is an increasing function:

$$\Big|\log|\sin\theta(f(x))|-\log|\sin\theta(x)|\Big|\le 2\log\|A(x)\|,$$

which means that $\log|\sin\theta|\circ f-\log|\sin\theta|\in L^1(\mu)$. ■

This finishes the proof for the case where A takes values in SL(2). Now we are going to extend it to the general case.

Lemma 2.4.7 *Oseledets theorem can be reduced to the case in which A takes values in* SL(2).

Proof. Given a linear cocycle $A: M\to \mathrm{GL}(2)$, define $c(x)=|\det A(x)|^{\frac{1}{2}}$ and let $B:M\to\mathrm{SL}(2)$ be given by $A(x)=c(x)B(x)$.

First we claim that, if $\log^+\|A^{\pm1}\|$ is in $L^1(\mu)$, then $\log c$ and $\log^+\|B^{\pm1}\|$ must also be in $L^1(\mu)$.

By definition, $\log^+\|A(x)\| = \max\{\log c(x)+\log\|B(x)\|,0\}$. Since $B(x)\in\mathrm{SL}(2)$, we know that $\|B(x)\|\ge 1$, hence $0\le\log\|B(x)\|\le\log\|B\|$. So, $\log^+\|B(x)\|=\log\|B(x)\|\in L^1(\mu)$ because that M is a probability space. Together with the fact that $\log^+\|A\|\in L^1(\mu)$, this implies that $\log^+ c\in L^1(\mu)$. Also, $\log^+\|A(x)^{-1}\|=\max\{\log c^{-1}(x)+\log\|B(x)^{-1}\|,0\}$. Since $B(x)\in\mathrm{SL}(2)$, $\|B(x)\|=\|B(x)^{-1}\|$. Then $\log^+\|B(x)^{-1}\|\in L^1(\mu)$ by the same argument as before and $\log^- c\in L^1(\mu)$. This concludes the proof of the claim.

Now we are going to check that for $\mu-$almost every $x\in M$,

$$\lim_{n\to\infty}\frac{1}{n}\log\|A^n(x)v\|=\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\log c(f^j(x))+\lim_{n\to\infty}\frac{1}{n}\log\|B^n(x)v\|\ \ \forall v\neq 0.$$

By definition,

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)v\| \\
&= \lim_{n\to\infty} \frac{1}{n} \log \|(cB)^n(x)v\| \\
&= \lim_{n\to\infty} \frac{1}{n} \log \|(cB)(f^{n-1}x)...(cB)(f(x))(cB)(x)v\| \\
&= \lim_{n\to\infty} \frac{1}{n} \log \|c(f^{n-1}x)...c(f(x))c(x)B^n(x)v\| \\
&= \lim_{n\to\infty} \frac{1}{n} \log \left\{|c(f^{n-1}x)...c(f(x))c(x)|\|B^n(x)v\|\right\} \\
&= \lim_{n\to\infty} \frac{1}{n} \log \left\{c(f^{n-1}x)...c(f(x))c(x)\right\} + \lim_{n\to\infty} \frac{1}{n} \log \|B^n(x)v\| \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \log c(f^j(x)) + \lim_{n\to\infty} \frac{1}{n} \log \|B^n(x)v\|.
\end{aligned}$$

The left term is the Birkhoff average of the function $\log c$, which is in $L^1(\mu)$, so this limit exists for $\mu-$almost every $x \in M$. The right term is the limit that appears at Oseledets Theorem with $B(x) \in \mathrm{SL}(2)$, so it also exists for $\mu-$almost every $x \in M$ and every $v \neq 0$.

By the previous expression, we conclude that the associated cocycles $F(x,v) = (f(x), A(x)v)$ and $G(x,v) = (f(x), B(x)v)$ have the same Oseledets decomposition at almost every point. Moreover, the Lyapunov spectrum of the cocycle F is a translation of the cocycle B. ■

This concludes the proof of the Multiplicative Ergodic Theorem of Oseledets. ■

3
Stationary measures and Furstenberg's Formula

The goal of this chapter is to formulate and present a detailed proof of Furstenberg-Ledrappier Formula, describing the Lyapunov exponents of random linear cocycles. This formula plays a central role in the proof of Le Page's Theorem in [1] and also in our result, that extends this theorem in chapter 5. We follow chapters 5 and 6 in [11].

In section 3.1 we present a class of cocycles called random cocycles and we introduce basic concepts of stationary measures. In section 3.2 we go through the theory of u-states and s-states and relate these concepts with the previous ones in order to prepare the reader for the next section. Section 3.3 is the heart of this chapter, where we provide a detailed proof of Furstenberg-Ledrappier Formula. In section 3.4, we derive a particular version of the previous result and use it to prove a theorem of Furstenberg and Kifer on the continuity of the Lyapunov exponents.

3.1
Stationary measures

In this section we present two fundamental concepts: the random cocycles and the stationary measures. They will play a central role in this work.

Consider a probability space $(X, \mathcal{B}, \rho)$. Let $M = X^{\mathbb{Z}}$ be the space of biinfinite sequences in X endowed with the product σ-algebra $\mathcal{A} = \mathcal{B}^{\mathbb{Z}}$ and the product measure $\mu = \rho^{\mathbb{Z}}$. Let $f : M \to M$ be the shift map on M.

Let $(N, \mathcal{C})$ be a measurable space and consider $M \times N$ endowed with the product σ-algebra $\mathcal{A} \otimes \mathcal{C}$. For most of our applications, specially in chapter 5, we will consider X to be a compact metric space and N to be $\mathbb{R}^2$ or $\mathbb{PR}^2$. For the time being, we maintain this generality.

Definition 3.1.1 *A random transformation (also called locally constant skew product) over f is a measurable transformation of the form:*

$$F : M \times N \to M \times N$$
$$F(x, v) = (f(x), F_x(v)) \, ,$$

where $F_x \colon N \to N$ depends only on the zeroth coordinate of $x \in M$.

In order to illustrate this concept we present the following examples:

Example 3.1.1 *Let X be a compact subset of* $\mathrm{GL}(d)$*. Consider a probability measure ρ defined on X. Let $M = X^{\mathbb{Z}}$, so that an element of M is given by $\{g_n\}_{n\in\mathbb{Z}}$, and $\mu = \rho^{\mathbb{Z}}$. For the transformation on the base we consider the Bernoulli shift. Now we present two random transformations: the random linear cocycle: $F : M \times \mathbb{R}^d \to M \times \mathbb{R}^d$,*

$$(\{g_n\}_n, v) \mapsto (\{g_{n+1}\}_n, g_0(v)),$$

and its associated projective cocycle: $\mathbb{P}F : M \times \mathbb{PR}^d \to M \times \mathbb{PR}^d$,

$$(\{g_n\}_n, \hat{v}) \mapsto \left(\{g_{n+1}\}_n, \widehat{g_0(v)}\right).$$

Now we are going to work toward the definition of a stationary measure. For this we introduce the transition operator associated to a random transformation and its adjoint.

Definition 3.1.2 *The transition operator associated to a random transformation F is the linear map $\mathcal{P} : L^\infty(N) \to L^\infty(N)$, defined by*

$$\mathcal{P}\varphi(v) = \int_M \varphi(F_x(v))\, d\mu(x).$$

Note that the transition operator is bounded: $\|\mathcal{P}\varphi\| \leq \|\varphi\|$.

Definition 3.1.3 *The adjoint transition operator $\mathcal{P}^*$ associated to a random transformation F acts on the space of probability measures η on N by:*

$$\mathcal{P}^*\eta(B) = \int_M (F_x)_*\eta(B)\, d\mu(x) = \int_M \eta(F_x^{-1}(B))\, d\mu(x),$$

for every measurable set $B \subset N$.

Consider the space $\mathcal{T}(N)$ of measurable transformations $T\colon N \to N$, endowed with some σ-algebra such that the map $\mathcal{F} : M \to \mathcal{T}(N)$, $\mathcal{F}(x) = F_x$ is measurable, for example the push forward of the σ-algebra $\mathcal{A}$ under $\mathcal{F}$. Let ν be the push forward $\mathcal{F}_*\mu$ of the probability measure μ by the map $\mathcal{F}$. Then we can characterize the transition operators by this new measure ν:

$$\mathcal{P}\varphi(v) = \int_{\mathcal{T}(N)} \varphi\left(g(v)\right)\, d\nu(g) \quad \text{and} \quad \mathcal{P}^*\eta(B) = \int_{\mathcal{T}(N)} \eta\left(g^{-1}(B)\right)\, d\nu(g).$$

The next lemma relates the transition and adjoint transition operators.

Lemma 3.1.1 *Let $\varphi \in L^\infty(N)$. Then*

$$\int_N \varphi \; d(\mathcal{P}^*\eta) = \int_N (\mathcal{P}\varphi) \; d\eta.$$

Proof. Our strategy is to use a standard argument in measure theory: first we prove the statement for indicator functions. Let 1_B be the indicator function of a measurable set $B \subset N$. Then

$$\begin{aligned}
\int_N 1_B \; d(\mathcal{P}^*\eta) &= \mathcal{P}^*\eta(B) \\
&= \int_M \eta(F_x^{-1}(B)) \; d\mu(x) \\
&= \int_M \eta\left(\left\{v \in N : (x,v) \in F^{-1}(M \times B)\right\}\right) \; d\mu(x) \\
&= (\mu \times \eta)\left(F^{-1}(M \times B)\right).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\int_N \mathcal{P}1_B \; d\eta &= \int_N \int_M 1_B(F_x(v)) \; d\mu(x) \; d\eta(v) \\
&= \int_N \mu\left(x \in M : F_x(v) \in B\right) \; d\eta(v) \\
&= \int_N \mu\left(\left\{x \in M : (x,v) \in F^{-1}(M \times B)\right\}\right) \; d\eta(v) \\
&= (\mu \times \eta)\left(F^{-1}(M \times B)\right).
\end{aligned}$$

Then, we can extend this result by linearity to simple functions. Finally, any given function $\varphi \in L^\infty(N)$ is an uniform limit of simple functions $\{f_n\}_n$, hence we have:

$$\begin{aligned}
\int_N \varphi \; d(\mathcal{P}^*\eta) &= \int_N \lim_{n\to\infty} f_n \; d(\mathcal{P}^*\eta) \\
&= \lim_{n\to\infty} \int_N f_n \; d(\mathcal{P}^*\eta) \\
&= \lim_{n\to\infty} \int_N (\mathcal{P}f_n) \; d\eta \\
&= \int_N \lim_{n\to\infty} (\mathcal{P}f_n) \; d\eta \\
&= \int_N (\mathcal{P}\varphi) \; d\eta \qquad \text{since } \mathcal{P} \text{ is continuous.}
\end{aligned}$$

This concludes the lemma. ■

We now introduce one of the most important concepts of this work, the notion of a *stationary measure*. In ergodic theory, usually consider *f-invariant measures*, that is, measures that satisfy $\eta(B) = \eta(f^{-1}(B))$ for any measurable set B. The following concept, as we are going to see, is a weaker version of thereof.

Definition 3.1.4 *A probability measure η on N is called stationary for the random transformation F if $\mathcal{P}^*\eta = \eta$; that is, if*

$$\eta(B) = \int_M \eta(F_x^{-1}(B)\ d\mu(x) = \int_{\mathcal{T}(N)} \eta(g^{-1}(B))\ d\mathcal{F}_*\mu(g)$$

for every measurable set $B \subset N$.

This definition means that the measure η is invariant on average (with respect to μ), over all transformations F_x. Therefore, if a probability measure is F_x-invariant for μ-almost every x, then it is also stationary for F. However, the converse statement is not true in general.

Remark 4 *The stationary measures are fixed points of the operator $\mathcal{P}$. When the set N is a compact metric space and the transformation $F_x : N \to N$ is continuous, the operator $\mathcal{P}^*$ is continuous with respect to the weak star topology in the space of probability measures on N. This implies that the set of stationary measures for F is closed, hence compact for the weak star topology.*

Now we introduce a characterization of the stationary measures for one sided random transformations. Here we consider $M = X^{\mathbb{N}}$ and $f : M \to M$ to be the one sided shift.

Proposition 3.1.2 *Let $F\colon M \times N \to M \times N$ be a one sided random transformation. A probability measure η on N is stationary for F if and only if the probability measure $\mu \times \eta$ on $M \times N$ is F-invariant.*

Proof. Suppose $(\mu \times \eta)$ is F-invariant. Let $\varphi \in L^\infty(N)$ and define the function $\psi\colon\ M \times\ N \to \mathbb{R}$ by $\psi(x, v) = \varphi(v)$. Then,

$$\begin{aligned}
\int_N \varphi(v)\ (d\mathcal{P}^*\eta)(v) &= \int_N \mathcal{P}\varphi(v)\ d\eta(v) && \text{(by lemma 3.1.1)} \\
&= \int_N \int_M \varphi(F_x(v))\ d\mu(x) d\eta(v) && \text{(by definition of } \mathcal{P}) \\
&= \int_N \int_M \psi(f(x), F_x(v))\ d\mu(x) d\eta(v) && \text{(by definition of } \psi) \\
&= \int_N \int_M \psi(x, v)\ d\mu(x) d\eta(v) && \text{(by the invariance of } (\mu \times \eta)) \\
&= \int_N \varphi(v)\ d\eta(v) && \text{(since M is a probability space)}
\end{aligned}$$

and this is sufficient to prove that $\mathcal{P}^*\eta = \eta$, since φ was arbitrary. Therefore, η is F-stationary.

For the converse statement, assume that η is F-stationary. Let $\psi \in\ L^\infty(M \times\ N)$ and consider $\varphi\colon N \to \mathbb{R}$ defined by $\varphi(v) = \int_M \psi(x, v)\ d\mu(x)$.

Then,

$$\begin{aligned}
\int_N \int_M \psi(x,v)\, d\mu(x) d\eta(v) &= \int_N \varphi(v)\, d\eta(v) && \text{(by definition of } \varphi) \\
&= \int_N \varphi(v)\, d(\mathcal{P}^*\eta)(v) && \text{(since } \eta \text{ is F-stationary)} \\
&= \int_N \mathcal{P}\varphi(v)\, d\eta(v) && \text{(by lemma 3.1.1)} \\
&= \int_N \int_M \varphi(F_x(v))\, d\mu(x) d\eta(v) && \text{(by definition of } \mathcal{P}) \\
&= \int_N \int_M \int_M \psi(y, F_x(v))\, d\mu(y) d\mu(x) d\eta(v)
\end{aligned}$$

where the last line follows from the definition of φ. Now we are going to do a smart change of variables. Write $x = (x_0, x_1, ...)$ and $y = (y_0, y_1, ...)$. Remember that F_x only depends on x_0, hence the previous expression can be rewritten as

$$\int_N \int_X \int_M \psi(y, F_{x_0}(v))\, d\rho^{\mathbb{N}}(y) d\rho(x_0) d\eta(v).$$

Let $z = (x_0, y_0, y_1, ...)$. Since f is the unilateral shift, $f(z) = y$. Then we can rewrite again the previous expression as:

$$\int_N \int_M \psi(f(z), F_z(v))\, d\mu(z) d\eta(v).$$

Hence, we concluded that

$$\int_N \int_M \psi\, d\mu d\eta = \int_N \int_M (\psi \circ F)\, d\mu d\eta.$$

Since ψ was arbitrary, this proves that $(\mu \times \eta)$ is F-invariant. ■

3.2
u-states and s-states

In the previous section we discussed the one-sided case. Now we present the case when F is invertible. Consider $M = X^{\mathbb{Z}}$ and $\mu = \rho^{\mathbb{Z}}$ an f-invariant probability measure, for some probability measure ρ in X. Also assume that X is complete and separable, hence so is M. Moreover, let $f : M \to M$ be the full shift.

Before we continue, let us introduce some notations:

$$\mathbb{Z}^+ = \{n \in \mathbb{Z} : n \geq 0\} \quad \text{and} \quad \mathbb{Z}^- = \{n \in \mathbb{Z} : n < 0\}.$$

Let $M^\pm = X^{\mathbb{Z}^\pm}$ and $\mu^\pm = \rho^{\mathbb{Z}^\pm}$. Moreover, consider $\pi^\pm : M \to M^\pm$ the canonical projections and define $f^\pm : M^\pm \to M^\pm$ to be the one sided Bernoulli shifts associated to μ^+ and μ^- respectively:

$$f^+ \circ \pi^+ = \pi^+ \circ f \quad \text{and} \quad f^- \circ \pi^- = \pi^- \circ f^{-1}.$$

Then, every invertible random transformation $F : M \times N \to M \times N$ over $f : M \to M$ induces two one-sided transformations as follows. Given $x \in M$, let $x^{\pm} = \pi^{\pm}(x)$ and for $v \in N$ let

$$F^+_{x^+}(v) = F_x(v) \quad \text{and} \quad F^-_{x^-}(v) = \left(F_{f^{-1}(x)}(v)\right)^{-1}.$$

Then we may define

$$\begin{aligned} F^+ : M^+ \times N \to M^+ \times N, &\quad F^+(x^+, v) = (f^+(x^+), F^+_{x^+}v) \\ F^- : M^- \times N \to M^- \times N, &\quad F^-(x^-, v) = (f^-(x^-), F^-_{x^-}v). \end{aligned}$$

Moreover, F_x depends only on the zeroth coordinate of x, hence both F^+ and F^- are random transformations since $F^+_{x^+}$ depends only on x^+_0 and $F^-_{x^-}$ depends only on x^-_{-1}.

Definition 3.2.1 *A probability measure η on N is forward stationary with respect to F, if it is stationary with respect to F^+:*

$$\eta(B) = \int_{M^+} \eta\left((F^+_{x^+})^{-1}(B)\right)\, d\mu^+(x^+) = \int_M \eta\left((F_x)^{-1}(B)\right)\, d\mu(x),$$

for every measurable set $B \subset N$.

Moreover, we say that a probability measure η on N is backward stationary with respect to F, if it is stationary with respect to F^-:

$$\begin{aligned} \eta(B) &= \int_{M^-} \eta\left((F^-_{x^-})^{-1}(B)\right)\, d\mu^-(x^-) \\ &= \int_M \eta\left(F_{f^{-1}(x)}(B)\right)\, d\mu(x) \\ &= \int_M \eta\left(F_x(B)\right)\, d\mu(x) \end{aligned}$$

for every measurable set $B \subset N$. (The last equality uses the fact that μ is f-invariant.)

Now we are going to state a proposition that says that the invariant probability measures of F over μ are in a one to one correspondence with the invariant probabilities of F^+ over μ^+ and also with F^- over μ^-. Consider the canonical projections:

$$\begin{aligned} \mathrm{proj}_1 : M \times N \to M \quad &\text{and} \quad \mathrm{proj}_2 : M \times N \to N \\ \mathrm{proj}_1^{\pm} : M^{\pm} \times N \to M^{\pm} \quad &\text{and} \quad \mathrm{proj}_2^{\pm} : M^{\pm} \times N \to N^{\pm}, \end{aligned}$$

and let $\Pi^{\pm} : M \times N \to M^{\pm} \times N$ be given by $\Pi^{\pm}(x, v) = (\pi^{\pm}(x), v)$, where $\pi^{\pm} : M \to M^{\pm}$ are the canonical projections between the shift spaces.

Proposition 3.2.1 *There is a one to one correspondence between the invariant probability measures of F, F^+ and F^-. More precisely,*

(i) If m is an F-invariant probability measure with $(\mathrm{proj}_1)_* m = \mu$, then $m^+ = \Pi^+_* m$ is an F^+-invariant probability measure with $(\mathrm{proj}_1^+)_* m^+ = \mu^+$.

(ii) Given any F^+-invariant probability measure m^+ with $(\mathrm{proj}_1^+)_* m^+ = \mu^+$, there exists a unique F-invariant probability measure m such that $\Pi^+_* m = m^+$ and $(\mathrm{proj}_1)_* m = \mu$.

Moreover, the previous statements remain true if we replace the $+$ signs by $-$ signs.

Proof. This is a classical result, which can be found at chapter 5 of [9] (see also chapter 2 of [10] for a similar result with a more detailed proof). ■

Remark 5 *The measure m is usually called the lift of m^+ and m^-.*

We are now ready to introduce and study the u-states and s-states, which will play an important role in the proof of the Furstenberg-Ledrappier Formula.

Definition 3.2.2 *Let m be an F-invariant probability measure on $M \times N$ with $(\mathrm{proj}_1)_* m = \mu$. We say that m is an s-state if, for any measurable sets $A^- \subset M^-$, $A^+ \subset M^+$ and $B \subset N$,*

$$\frac{m(A^- \times A^+ \times B)}{\mu(A^- \times A^+)} \quad \textit{does not depend on } A^-.$$

We say that m is a u-state if, for measurable sets $A^- \subset M^-$, $A^+ \subset M^+$ and $B \subset N$,

$$\frac{m(A^- \times A^+ \times B)}{\mu(A^- \times A^+)} \quad \textit{does not depend on } A^+.$$

Moreover, we say that m is an su-state if m is both a u-state and an s-state.

Proposition 3.2.2 *Let m be an F-invariant probability measure on $M \times N$ such that $(\mathrm{proj}_1)_* m = \mu$. If m is a u-state, then $F_* m$ is also a u-state.*

Proof. Let $A^- \subset M^-$, $A^+ \subset M^+$ and $B \subset N$ be measurable sets. Since m is F-invariant, $F_* m(A^- \times A^+ \times B) = m(F^{-1}(A^- \times A^+ \times B)) = m(A^- \times A^+ \times B)$. Therefore,

$$\frac{F_* m(A^- \times A^+ \times B)}{\mu(A^- \times A^+)} = \frac{m(A^- \times A^+ \times B)}{\mu(A^- \times A^+)},$$

so it does not depend on A^+. ∎

Remark 6 *Analogously, if m satisfies the s-state condition, then the same holds for $F_*^{-1} m$.*

Proposition 3.2.3 *An F-invariant probability measure m on $M \times N$ is an s-state if and only if $m = \mu^- \times m^+$ for some probability measure m^+ on $M^+ \times N$. In this case $m^+ = \Pi^+_* m$ and it is F^+-invariant.*

Similarly, m is an u-state if and only if $m = \mu^+ \times m^-$ for some probability measure m^- on $M^- \times N$, in which case $m^- = \Pi^-_* m$ and it is F^--invariant.

Proof. We prove the first statement, the second one is similar. Suppose m is an s-state. Then, for any measurable sets $A^- \subset M^-$, $A^+ \subset M^+$ and $B \subset N$:

$$\begin{aligned}
\frac{m(A^- \times A^+ \times B)}{\mu(A^- \times A^+)} &= \frac{m(M^- \times A^+ \times B)}{\mu(M^- \times A^+)} \\
&= \frac{m\left((\Pi^+)^{-1}(A^+ \times B)\right)}{\mu\left((\pi^+)^{-1}(A^+)\right)} \\
&= \frac{\Pi^+_* m(A^+ \times B)}{\pi^+_* \mu(A^+)}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
m^+(A^+ \times B) &= \Pi^+_* m(A^+ \times B) \\
&= \frac{m(A^- \times A^+ \times B)}{\mu(A^- \times A^+)} \mu^+(A^+) \\
&= \frac{m(A^- \times A^+ \times B)}{\mu^-(A^-)}.
\end{aligned}$$

This proves that $m = \mu^- \times m^+$. Now we are going to verify that m^+ is F^+-invariant:

$$
\begin{aligned}
m^+\left((F^+)^{-1}(A^+ \times B)\right) &= \Pi^+_* m\left((F^+)^{-1}(A^+ \times B)\right) \\
&= m\left((\Pi^+)^{-1} \circ (F^+)^{-1}(A^+ \times B)\right) \\
&= m\left((F^+ \circ \Pi^+)^{-1}(A^+ \times B)\right) \\
&= m\left((\Pi^+ \circ F)^{-1}(A^+ \times B)\right) \\
&= m\left(F^{-1} \circ (\Pi^+)^{-1}(A^+ \times B)\right) \\
&= m\left(F^{-1}(M^- \times A^+ \times B)\right) \\
&= m(M^- \times A^+ \times B) \\
&= m\left((\Pi^+)^{-1}(A^+ \times B)\right) \\
&= m^+(A^+ \times B).
\end{aligned}
$$

To obtain the forth line we used that $F^+ \circ \Pi^+ = \Pi^+ \circ F$ and in the seventh line we used the fact that m is F-invariant. For the converse, suppose that $m = \mu^- \times m^+$. Then, for any measurable sets $A^- \subset M^-$, $A^+ \subset M^+$ and $B \subset N$:

$$
\frac{m(A^- \times A^+ \times B)}{\mu(A^- \times A^+)} = \frac{\mu^-(A^-) \times m^+(A^+ \times B)}{\mu^-(A^-) \times \mu^+(A^+)} = \frac{m^+(A^+ \times B)}{\mu^+(A^+)},
$$

that does not depend on A^-, which means that m is an s-state. ∎

Proposition 3.2.4 *If an F-invariant probability measure m on $M \times N$ is a u-state (respectively an s-state), then $\eta = (\mathrm{proj}_2)_* m$ is a forward (respectively backward) stationary measure.*

Proof. We prove the case where m is an u-state. The s-state case is analogous. Let m be a u-state and $m^+ = \Pi^+_* m$. For every measurable sets $A^+ \subset M^+$ and $B \subset N$ define

$$
\eta(B) = \frac{m^+(A^+ \times B)}{\mu^+(A^+)} = \frac{m(M^- \times A^+ \times B)}{\mu(M^- \times A^+)}.
$$

Note that since m is a u-state, the previous expression depends only on B. By its definition, η is a probability measure on N. Moreover, by proposition 3.2.3, we can write $m^+ = \mu^+ \times \eta$. Since m^+ is F^+-invariant, then by proposition 3.1.2, η is a forward stationary measure for F. Also note that $(\mathrm{proj}_2)_* m = (\mathrm{proj}_2)_* m^+ = \eta$, which concludes the proof. ∎

Remark 7 *The converse is also true: if η is a forward (respectively backward) stationary measure on N, then its lift is a u-state (respectively an s-state).*

The next proposition says that the accumulation points of a sequence of s-states (u-states) are also s-states (u-states). We propose an alternative proof to the one given in [11].

Proposition 3.2.5 *Let $\{m_n\}_n$ be a sequence of probability measures on $M \times N$ projecting to Bernoulli measures $\{\mu_n\}_n$ on M and satisfying the s-state (respectively u-state) condition. If $\{m_n\}_n$ converges to some probability measure m and $\{\mu_n\}_n$ converges to some Bernoulli measure μ in the weak* topology, then m satisfies the s-state (u-state) condition.*

Proof. Since m_n satisfies the s-state condition for every n, we can write $m_n = \mu_n^- \times m_n^+$ by proposition 3.2.3. By hypothesis, $\{m_n\}_n$ converges to m and $\{\mu_n^-\}_n$ converges to μ^- in the weak* topology. We claim that $\{m_n^+\}_n$ converges to $m^+ = \Pi_*^+ m$. Consider any $\varphi \in C(M^+ \times N)$, then:

$$\int_{M^+\times N} \varphi \, d(m_n^+) = \int_{M^+\times N} \varphi \, d(\Pi_*^+ m_n) = \int_{M\times N} \varphi \circ \Pi^+ \, dm_n.$$

Since $\varphi \circ \Pi^+$ is a continuous function and $\{m_n\}_n$ converges to m in the weak* topology, we know that the last term converges to $\int_{M\times N} \varphi \circ \Pi^+ \, dm$. Hence, applying again the change of variables, we can conclude that

$$\int_{M^+\times N} \varphi \, d(m_n^+) \to \int_{M^+\times N} \varphi \, d(m^+).$$

This completes the proof of the claim. Moreover, since we know that $m_n = \mu_n^- \times m_n^+ \to \mu^- \times m^+$ and $m_n \to m$, we conclude that $m = \mu^- \times m^+$. This means that m is also an s-state, by proposition 3.2.3. ■

Before we continue to the next proposition, we need to introduce the concept of disintegration of a measure.

Definition 3.2.3 *Let m be a probability measure on $M \times N$ that projects down to μ. A disintegration of m along vertical fibers is a measurable family $\{m_x : x \in M\}$ of probability measures on N satisfying*

$$m(E) = \int_M m_x\left(\{v : (x, v) \in E\}\right) \, d\mu(x)$$

for every measurable set $E \subset M \times N$.

Remark 8 *The probability measures m_x are called conditional probabilities.*

Rokhlin's Disintegration Theorem A.0.8 implies that a disintegration along vertical fibers does exist and that it is essentially unique, in the sense that given two different disintegrations they coincide on μ-almost everywhere (see A.0.7).

Proposition 3.2.6 *If an F-invariant probability measure on $M \times N$ with $(\text{proj}_1)_* m = \mu$ admits a disintegration such that each m_x depends only on x^-, then it is an u-state. Analogously if it admits a disintegration such that m_x depends only on x^+, then it is an s-state.*

Proof. Suppose that m admits a disintegration such that each m_x depends only on x^-. Then, for every measurable sets $A^- \subset M^-$, $A^+ \subset M^+$ and $B \subset N$:

$$\begin{aligned}\frac{m(A^- \times A^+ \times B)}{\mu(A^- \times A^+)} &= \frac{\int_{A^+} d\mu^+(x^+) \int_{A^-} m_x(B)\, d\mu^-(x^-)}{\mu^-(A^-) \times (\mu^+(A^+))} \\ &= \frac{\int_{A^-} m_x(B)\, d\mu^-(x^-)}{\mu^-(A^-)} \\ &= \frac{m(A^- \times M^+ \times B)}{\mu(A^- \times M^+)}.\end{aligned}$$

Therefore, it does not depend on A^+, so it is an u-state. An analogous argument shows that if m_x depends only on x^+, then it is an s-state. ■

Remark 9 *In fact, the converse of this proposition is also true: if an F-invariant probability measure on $M \times N$ with $(\text{proj}_1)_* m = \mu$ is a u-state (s-state), then it admits a disintegration such that each m_x depends only on x^- (respectively x^+).*

3.3 Furstenberg-Ledrappier's Formula

In this section our goal is to derive a formula that describes the Lyapunov exponents of random cocycles using the machinery that we have already developed in the previous sections.

Proposition 3.3.1 *Let m be an F-invariant probability measure in $M \times N$. Then it follows that:*

(i) Its ergodic components are also F-invariant.

(ii) If m projects down to μ, then its ergodic components also project down to μ.

(iii) If m gives full weight to some subspace, then its ergodic components also give full weight to the same subspace.

Proof. Let m be an F-invariant probability measure in $M \times N$. By the Ergodic Decomposition Theorem A.0.6, there exist a partition $\mathcal{P}$ of $M \times N$ and a familiy of probabilities $\{m_P : P \in \mathcal{P}\}$ such that for every measurable set E,

$$m(E) = \int_{M \times N} m_P(E)\, d\pi_*(m)(P),$$

where $\pi : M\times N \to \mathcal{P}$ maps every point x to the element $\mathcal{P}(x)$ of the partitions that contains x.

Note that the support of each measure m_P is disjoint from the others. Hence item (iii) follows immediately. Also by the same observation, given an element P^* of the partition, the corresponding measure μ_{P^*} must satisfy items (i) and (ii). Suppose, by contradiction that it does not satisfy item (i). Then there exists some measurable set $E \subset P^*$, such that $m_{P^*}(F^{-1}(E)) \neq m_{P^*}(E)$. However,

$$m_{P^*}(F^{-1}(E)) = \frac{m(F^{-1}(E))}{m(P^*)} = \frac{m(E)}{m(P^*)} = m_{P^*}(E),$$

which is a contradiction. The proof of item (ii) is analogous. ■

Before we state the next theorem, we introduce the concept of a flag. A flag in $\mathbb{R}^d$ is a decreasing family $\mathbb{R}^d = V^1 \supsetneq ... \supsetneq V^k \supsetneq \{0\}$ of vector subspaces of $\mathbb{R}^d$. Since we work in $\mathbb{R}^2$, there are two possible types of flag: $V^1 = \mathbb{R}^2$ and $V^2 = \{0\}$, or $V^1 = \mathbb{R}^2$, V^2 is a subspace of dimension 1 and $V^3 = \{0\}$. When we are in the second case we say that the flag is complete.

Moreover, we can relate this concept with the Oseledets Theorem. In the non-invertible case of Oseledets Theorem, there are two possibilities for the Oseledets flag: for μ-almost every x we have $V^1_x = \mathbb{R}^2$ and $V^2_x = \{0\}$, or $V^1_x = \mathbb{R}^2$, $V^2_x = E^s_x$ and $V^3_x = \{0\}$. The invertible case is the same and we also know that $\mathbb{R}^2 = E^s_x \oplus E^u_x$. Moreover, we can also consider the Oseledets flag for F^{-1}, where there are also both possibilities but, in the second one, we exchange E^s_x by E^u_x.

For convenience of notation, for the rest of this section we will denote E^u_x by E^1_x,E^s_x by E^2_x, λ_+ by λ_1 and λ_- by λ_2.

Consider the map

$$\Phi : M \times \mathbb{PR}^2 \to \mathbb{R}$$
$$\Phi(x, \hat{v}) = \log \frac{\|A(x)v\|}{\|v\|},$$

where $v \in \mathbb{R}^2\backslash\{0\}$ is a representative of the projective point $\hat{v} \in \mathbb{PR}^2$ and $A\colon M \to \mathrm{SL}(2)$ is a measurable function such that $\log\| A^{\pm 1}\|$ is integrable.

Theorem 3.3.2 *(Ledrappier) Given any $\mathbb{P}F$-invariant ergodic probability measure m on $M \times \mathbb{PR}^2$ that projects down to μ, there exists $j \in \{1, 2\}$ such that*

$$\int_{M\times\mathbb{PR}^2} \Phi \, dm = \lambda_j \quad \textit{and} \quad m\left(\{(x, \hat{v}) : v \in V^j_x\backslash V^{j+1}_x\}\right) = 1. \tag{3.1}$$

Conversely, given $j \in \{1, 2\}$, there is a $\mathbb{P}F-$invariant ergodic probability measure m projecting down to μ and satisfying (3.1). *When F is invertible, one may replace $V_x^j \backslash V_x^{j+1}$ by E_x^j in* (3.1).

Proof. Let m be a $\mathbb{P}F$-invariant ergodic probability measure. By Birkhoff's ergodic theorem, for m-almost every $(x, \hat{v})$:

$$\begin{aligned}
\int_{M \times \mathbb{P}\mathbb{R}^d} \Phi \, dm &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \Phi \circ \mathbb{P}F^i(x, \hat{v}) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \Phi \circ (f^i(x), \widehat{A^i(x)v}) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \log \frac{\|A(f^i x)A^i(x)v\|}{\|A^i(x)v\|} \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \log \Big(\|A^{i+1}(x)v\| - \|A^i(x)v\| \Big) \\
&= \lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)v\| - \frac{1}{n} \log \|v\| \\
&= \lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)v\|.
\end{aligned}$$

By Oseledets theorem, the last limit above is λ_j for μ-almost every $x \in M$ and for $v \in V_x^j \backslash V_x^{j+1}$. Since m projects to μ, we conclude that $\int_{M \times \mathbb{P}\mathbb{R}^d} \Phi \, dm = \lambda_j$ for some Lyapunov exponent λ_j and

$$m\Big(\{(x, \hat{v}) : v \in V_x^j \backslash V_x^{j+1}\}\Big) = 1.$$

For the invertible case, we can take the limit on both directions $n \to \pm\infty$. Thus, m gives full weight to the set

$$\{(x, \hat{v}) : v \in E_x^j\}$$

of pairs for which the limit is λ_j for both $n \to +\infty$ and $n \to -\infty$.

This completes the direct statement of the theorem. To prove the converse, we need some lemmas.

Consider the projection on the first coordinate $\pi : M \times \mathbb{P}\mathbb{R}^2 \to M$ that sends $(x, \hat{v}) \mapsto x$. We say that a measure m projects to μ if $\mu = \pi_* m$. Let $\mathcal{M}(\mu)$ denote the space of probability measures on $M \times \mathbb{P}\mathbb{R}^2$ that project down to μ.

Lemma 3.3.3 *The push-forward $F_* \colon \mathcal{M}(\mu) \to \mathcal{M}(\mu)$ is well defined and continuous relative to the weak star topology.*

Proof. First we are going to show that given a probability measure $\nu \in \mathcal{M}(\mu)$, then the push forward $F_* \nu$ is also in $\mathcal{M}(\mu)$. Note that $\pi \circ F = f \circ \pi$. Moreover,

$\pi_* \circ F_* = f_* \circ \pi_*$. Then, $\pi_*(F_*\nu) = f_*(\pi_*\nu) = f_*\mu = \mu$ since μ is f-invariant. Thus, the push forward is well defined.

Now we are going to show it is continuous relative to the weak star topology. Let $(m_n)_n$ be a sequence in $\mathcal{M}(\mu)$, such that it converges in the weak star topology to a measure m. We need to prove that F_*m_n converges in the weak star topology to F_*m, that is the same as

$$\int_{M\times\mathbb{PR}^2} \varphi \, dF_*m_n \to \int_{M\times\mathbb{PR}^2} \varphi \, dF_*m \qquad \text{for every } \varphi \in C_c(M\times\mathbb{PR}^2). \quad (3.2)$$

By the change of variables, (3.2) is equivalent to

$$\int_{M\times\mathbb{PR}^2} \varphi\circ F \, dm_n \to \int_{M\times\mathbb{PR}^2} \varphi\circ F \, dm, \qquad \text{for every } \varphi \in C_c(M\times\mathbb{PR}^2). \quad (3.3)$$

Since $f\colon M \to M$ and $A\colon M \to \mathrm{SL}_2(\mathbb{R})$ are measurable, by Lusin's Theorem, for every $\varepsilon > 0$ there exists a compact set $K \subset M$ such that $\mu(K^c) < \varepsilon$ and the restriction $F\big|_L$, where $L = K \times \mathbb{PR}^2$ is continuous.

Hence $\varphi \circ F$ is continuous restricted to L. Then, by Tietze extension theorem, there exists some continuous function $\psi : M \times \mathbb{PR}^2 \to \mathbb{R}$ that is an extension of $\varphi \circ F$.

Therefore:

$$\begin{aligned}
&\left|\int_{M\times\mathbb{PR}^2}(\varphi\circ F)\, dm_n - \int_{M\times\mathbb{PR}^2}(\varphi\circ F)\, dm\right| \\
&\le \left|\int_{L}(\varphi\circ F)\, dm_n - \int_{L}(\varphi\circ F)\, dm\right| + \left|\int_{L^c}(\varphi\circ F)\, dm_n - \int_{L^c}(\varphi\circ F)\, dm\right| \\
&\le \left|\int_{M\times\mathbb{PR}^2}\psi\, dm_n - \int_{M\times\mathbb{PR}^2}\psi\, dm\right| + \left|\int_{L^c}\psi\, dm_n - \int_{L^c}\psi\, dm\right| \\
&+ \left|\int_{L^c}(\varphi\circ F)\, dm_n - \int_{L^c}(\varphi\circ F)\, dm\right|.
\end{aligned}$$

The first term is smaller than ε for n sufficiently large, since ψ is continuous and m_n converges to m in the weak star topology. The second and third terms are bounded by $\|\varphi\|_\infty [m_n(L^c) + m(L^c)]$ each. Thus, each term is smaller than $\|\varphi\|_\infty 2\varepsilon$, which concludes the proof. ∎

Lemma 3.3.4 *$\mathcal{M}(\mu)$ is sequentially compact, relative to the weak* topology.*

Proof. Since every Borel measure in a separable complete metric space is tight, for every $\varepsilon > 0$ there exists a compact set $K \subset M$ such that $\mu(K) > 1-\varepsilon$. Then $K \times \mathbb{PR}^2$ is compact and $m(K\times\mathbb{PR}^2) = \mu(K) > 1-\varepsilon$. Therefore $\mathcal{M}(\mu)$ is tight and by Prohorov's theorem it is sequentially compact. ∎

Lemma 3.3.5 *Let $x \mapsto V_x$ be a measurable sub-bundle of $M\times\mathbb{R}^2$. Then the subset of probability measures $m \in \mathcal{M}(\mu)$ such that $m\left(\{(x,\hat{v}) : v \in V_x\}\right) = 1$ is closed in the weak star topology.*

Proof. Let $\{m_n\}_n$ be a sequence in $\mathcal{M}(\mu)$ such that

$$m_n\left(\{(x,\hat{v}) : v \in V_x\}\right) = 1 \quad \text{for all } n,$$

and which converges to some m in the weak star topology. Since $x \mapsto V_x$ is measurable, then, by Lusin's theorem, for every $\varepsilon > 0$ we can find a compact set $K \subset M$ with $\mu(K) > 1 - \varepsilon$ such that the map $x \mapsto V_x$ restricted to K is continuous. Therefore, $\{(x,\hat{v}) \in K \times \mathbb{PR}^2 : v \in V_x\}$ is the graph of a continuous function, hence it is closed. By Portmanteau's theorem,

$$\begin{aligned} m\{(x,\hat{v}) \in K \times \mathbb{PR}^2 : v \in V_x\} &\geq \limsup_{n\to\infty} m_n\{(x,\hat{v}) \in K \times \mathbb{PR}^2 : v \in V_x\} \\ &\geq 1 - \mu(K^c) \\ &> 1 - \varepsilon. \end{aligned}$$

Since ε is arbitrary, $m\left(\{(x,[v]) : v \in V_x\}\right) = 1$. ■

Furthermore, the set of probability measures $\nu \in \mathcal{M}(\mu)$ such that $\nu\left(\{(x,\hat{v}) : v \in V_x\}\right) = 1$ is $\mathbb{P}F$-invariant, since the Oseledets sub-bundle is invariant under the cocycle. Given $\nu \in \mathcal{M}(\mu)$ such that $\nu\left(\{(x,\hat{v}) : v \in V_x\}\right) = 1$,

$$\begin{aligned} (\mathbb{P}F)_*\nu(\{(x,\hat{v}) : v \in V_x\}) &= \nu\left((\mathbb{P}F)^{-1}\{(x,\hat{v}) : v \in V_x\}\right) \\ &= \nu\left(\{(x,\hat{v}) : v \in V_x\}\right) \\ &= 1. \end{aligned}$$

Now we are ready to finish the proof of theorem 3.3.2. The idea of the proof uses a very common method. We prove the invertible case and then deduce the general case by using the invertible extension of the cocycle $F\colon M \times \mathbb{R}^2 \to M \times \mathbb{R}^2$.

Let j be fixed. Note that the Oseledets sub bundle $x \mapsto E_x^j$ is measurable, since it can only be the map $s : M \to \mathbb{PR}^2$ such that $s(x) = \lim_{n\to\infty} s_n(x)$ or $u : M \to \mathbb{PR}^2$ such that $u(x) = \lim_{n\to\infty} u_n(x)$, where both are limits of measurable functions. Thus, it admits a measurable section. Hence, there exists a measurable vector field $x \mapsto \sigma(x)$ such that $\sigma(x) \in E_x^j$ for every x. Consider m_0 the probability measure in $\mathcal{M}(\mu)$ that admits $\delta_{\sigma(x)}$ as a disintegration:

$$m_0(B) = \mu\left(\{x \in M : (x,\sigma(x)) \in B\}\right)$$

for every measurable $B \subset M \times \mathbb{PR}^d$. Now define, for $n \geq 1$:

$$m_n = \frac{1}{n}\sum_{i=0}^{n-1} (\mathbb{P}F)_*^i m_0.$$

Notice that $m_n \in \mathcal{M}(\mu)$ by lemma 3.3.3 and $m_n\left(\{(x,[v]) : v \in V_x\}\right) = 1$, because the set of probability measures $\nu \in \mathcal{M}(\mu)$ such that $\nu\left(\{(x,[v]) : v \in V_x\}\right) = 1$ is $\mathbb{P}F$-invariant. Then, by lemma 3.3.4, there exists a subsequence $\{n_k\}_k$ such that m_{n_k} converges to some measure $m \in \mathcal{M}(\mu)$. Then, by lemma 3.3.5, $m\left(\{(x,[v]) : v \in V_x\}\right) = 1$. Moreover, m is $\mathbb{P}F$-invariant:

$$
\begin{aligned}
(\mathbb{P}F)_* m &= (\mathbb{P}F)_*(\lim_{k\to\infty} m_{n_k}) \\
&= \lim_{k\to\infty} (\mathbb{P}F)_* m_{n_k} && \text{(by lemma 3.3.3)} \\
&= \lim_{k\to\infty} (\mathbb{P}F)_* \frac{1}{n_k} \sum_{i=0}^{n_k-1} (\mathbb{P}F)_*^i m_0 \\
&= \lim_{k\to\infty} \frac{1}{n_k} \sum_{i=0}^{n_k-1} (\mathbb{P}F)_*(\mathbb{P}F)_*^i m_0 \\
&= \lim_{k\to\infty} \frac{1}{n_k} \sum_{i=1}^{n_k} (\mathbb{P}F)_*^i m_0 \\
&= \lim_{k\to\infty} \left(\frac{1}{n_k} \sum_{i=0}^{n_k-1} (\mathbb{P}F)_*^i m_0 + \frac{1}{n_k}(\mathbb{P}F)_*^{n_k} m_0 - \frac{1}{n_k} m_0 \right) \\
&= m.
\end{aligned}
$$

If we consider the ergodic components of m, we conclude that there exists some $\mathbb{P}F$-invariant ergodic probability measure $\overline{m}$ such that $\overline{m}\left(\{(x,[v]) : v \in V_x\}\right) = 1$. Finally, by Birkhoff's theorem, for $\overline{m}$-almost every $(x, \hat{v})$:

$$
\int_{M\times\mathbb{PR}^d} \Phi \, d\overline{m} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \Phi \circ \mathbb{P}F^i(x,\hat{v}) \, d\overline{m} = \lambda_j.
$$

This finishes the proof of the invertible case.

As we have already claimed before, we are now going to consider the invertible extension of $F : M \times \mathbb{PR}^2 \to M \times \mathbb{PR}^2$. Consider $\hat{M}$ the set of all pre-orbits of f, that is, the set of all sequences $\{x_n\}_{n\leq 0}$ such that $f(x_n) = x_{n+1}$ for every $n < 0$. Consider $\pi : \hat{M} \to M$ the projection on the first coordinate. Note that $\pi(\hat{M}) = M$. Let $\hat{f} : \hat{M} \to \hat{M}$ be the left shift:

$$
\hat{f}(\ldots, x_n, \ldots x_0) = (\ldots, x_n, \ldots x_0, f(x_0)).
$$

By proposition 3.2.1 there exists a unique $\hat{f}$-invariant measure on $\hat{M}$ such that $\pi_*\hat{\mu} = \mu$. Define $\hat{A} = A \circ \pi : \hat{M} \to \mathrm{GL}(d)$ and let $\hat{F} : \hat{M} \times \mathbb{R}^d \to \hat{M} \times \mathbb{R}^d$ be the linear cocycle defined by $\hat{A}$ over $\hat{f}$.

Note that $\hat{V}^i_{\hat{x}} = V^i_{\pi(\hat{x})}$ and also: $\lambda_+(F,\mu) = \lambda_+(\hat{F},\mu)$:

$$
\begin{aligned}
\lambda_+(F,\mu) &= \lim_{n\to\infty}\frac{1}{n}\int_M \log\|A^n(x)\|\, d\mu(x)\\
&= \lim_{n\to\infty}\frac{1}{n}\int_M \log\|A^n(x)\|\, d\pi_*\hat{\mu}(x)\\
&= \lim_{n\to\infty}\frac{1}{n}\int_{\hat{M}} \log\|A^n(\pi(\hat{x}))\|\, d\hat{\mu}(\hat{x})\\
&= \lim_{n\to\infty}\frac{1}{n}\int_{\hat{M}} \log\|\hat{A}^n(\hat{x})\|\, d\hat{\mu}(\hat{x})\\
&= \lambda_+(\hat{F},\mu).
\end{aligned}
$$

Therefore, there exists some $\mathbb{P}\hat{F}$-invariant probability measure $\hat{m}$ on $\hat{M}\times\mathbb{R}^2$ that projects down to μ and satisfies $\hat{m}\left(\{(x,\hat{v}) : v\in E^j_x\}\right) = 1$. Consider the image of $\hat{m}$ by the map $\pi\times id : \hat{M}\times\mathbb{PR}^2 \to M\times\mathbb{PR}^2$, $m' = (\pi\times id)_*\hat{m}$. We can see that m' is a $\mathbb{P}F$-invariant probability measure. Given $A\subset M$ and $B\subset\mathbb{PR}^2$ measurable sets we have:

$$
\begin{aligned}
m'\left(\mathbb{P}F^{-1}(A\times B)\right) &= (\pi\times id)_*\hat{m}\left(\mathbb{P}F^{-1}(A\times B)\right)\\
&= \hat{m}\left((\pi\times id)^{-1}\circ\mathbb{P}F^{-1}(A\times B)\right)\\
&= \hat{m}\left((\mathbb{P}F\circ(\pi\times id))^{-1}(A\times B)\right)\\
&= \hat{m}\left(\left((\pi\times id)\circ\mathbb{P}\hat{F}\right)^{-1}(A\times B)\right)\\
&= \hat{m}\left(\mathbb{P}\hat{F}^{-1}\circ(\pi\times id)^{-1}(A\times B)\right)\\
&= \hat{m}\left((\pi\times id)^{-1}(A\times B)\right)\\
&= m'(A\times B).
\end{aligned}
$$

Moreover, we claim that $m'\left(\{(x,\hat{v})\colon v\in V^j_x\backslash V^{j+1}_x\}\right) = 1$ because $\hat{E}^j_{\hat{x}} \subset V^j_{\pi(\hat{x})}\backslash V^{j+1}_{\pi(\hat{x})}$. To illustrate the last claim, observe that $E^u_x\subset\mathbb{R}^2\backslash E^s_x$ and also $E^s_x\subset E^s_x\backslash\{0\}$ for the case with a complete Oseledets splitting. When $E^u_x = E^s_x$, we have $E^u_x = E^s_x\subset\mathbb{R}^2\backslash\{0\}$, so it is true.

Then, by considering ergodic components according to proposition 3.3.1, we conclude that there exists some $\mathbb{P}F$-invariant ergodic probability measure m such that $m\left(\{(x,\hat{v}) : v\in V^j_x\backslash V^{j+1}_x\}\right) = 1$. The other claim, that $\int_{M\times\mathbb{PR}^d}\Phi\, dm = \lambda_j$, follows in the same way as we already did for the invertible case. ■

Proposition 3.3.6 *Let $F\colon M\times\mathbb{R}^2\to M\times\mathbb{R}^2$ be a random cocycle. Then,*

(1) $\lambda_+(F,\mu) = \max\left\{\int_{M\times\mathbb{PR}^2}\Phi\, dm :\ m \text{ is a u-state for } \mathbb{P}F\right\}$,

(2) $\lambda_-(F,\mu) = \min\left\{\int_{M\times\mathbb{PR}^2}\Phi\, dm :\ m \text{ is a s-state for } \mathbb{P}F\right\}$.

Proof. We are going to prove item (i) and a dual argument can be done to prove item (ii). By Ledrappier's Theorem (3.1):

$$\lambda_- \leq \int_{M\times\mathbb{PR}^2} \Phi \, dm \leq \lambda_+ \tag{3.4}$$

for every $\mathbb{P}F$-invariant ergodic probability measure m projecting to μ. By the Ergodic Decomposition Theorem, we can decompose every $\mathbb{P}F$-invariant measure into ergodic components, which, by proposition 3.3.1, also project down to μ. Therefore equation (3.4) holds for every $\mathbb{P}F$-invariant probability measure m projecting to μ. We show that the maximum is realised by some u-state.

Consider the Oseledets sub-bundle E^u corresponding to λ_+. Note that the subspace E^u_x depends only on the negative part x^- of x. Hence we may find a measurable section $x \mapsto \sigma(x)$ of the sub-bundle E^u such that $\sigma(x)$ depends only on x^-. Now we consider again, as in Ledrappier's Theorem, m_0 the measure in $\mathcal{M}(\mu)$ that admits $\delta_{\sigma(x)}$ as a disintegration:

$$m_0(B) = \mu\left(\{x \in M : (x, \sigma(x)) \in B\}\right)$$

for every measurable $B \subset M \times \mathbb{PR}^d$. Also, for $n \geq 1$:

$$m_n = \frac{1}{n}\sum_{i=0}^{n-1} (\mathbb{P}F)^i_* m_0.$$

We already saw in the proof of Ledrappier's Theorem, that every accumulation point of this sequence is an $\mathbb{P}F$-invariant probability measure that projects down to μ and that accumulation points do exist. Moreover, by proposition 3.2.6, m_0 is a u-state. Then, by proposition 3.2.2, every m_n is also a u-state. Finally, according to proposition 3.2.5 every accumulation point m is a u-state.

Notice that, similarly to what we did before: $m\left(\{(x, [v]) : v \in E^u_x\}\right) = 1$. This implies that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \Phi \circ \mathbb{P}F(x, \hat{v}) = \int_{M\times\mathbb{PR}^2} \Phi \, dm = \lambda_+$$

for m-almost every $(x, \hat{v})$. This concludes that the maximum in item (i) is realized by some u-state. A dual proof can be done in order to show that the minimum is realized by some s-state. ■

Corollary 3.3.1 *Let $F\colon M \times \mathbb{R}^2 \to M \times \mathbb{R}^2$ be an invertible random cocycle. Then it follows that:*

(1) $\lambda_+(F,\mu) = \max\left\{\int_{M\times\mathbb{P}\mathbb{R}^2} \Phi\, d(\mu\times\eta) : \ \eta \textit{ is forward stationary for } \mathbb{P}F\right\}$,

(2) $\lambda_-(F,\mu) = \min\left\{\int_{M\times\mathbb{P}\mathbb{R}^2} \Phi\, d(\mu\times\eta) : \ \eta \textit{ is backwad stationary for } \mathbb{P}F\right\}$.

Proof. First note that, by proposition 3.1.2, if η is a forward stationary measure for $\mathbb{P}F$, then $m = \mu\times\eta$ is $\mathbb{P}F$-invariant. We already know that for every $\mathbb{P}F$-invariant probability measure projecting down to μ,

$$\int_{M\times\mathbb{P}\mathbb{R}^2} \Phi\, dm \leq \lambda_+.$$

Therefore, for every forward stationary measure η on $\mathbb{P}\mathbb{R}^2$,

$$\int \Phi\, d(\mu\times\eta) \leq \lambda_+.$$

So we just need to show the other inequality. By proposition 3.2.1, every $\mathbb{P}F$-invariant probability measure m is the lift of some $\mathbb{P}F^+$-invariant probability measure m^+. Moreover

$$\int_{M\times\mathbb{P}\mathbb{R}^2} \Phi\, dm = \int_{M^+\times\mathbb{P}\mathbb{R}^2} \Phi\, dm^+,$$

since Φ depends only on the coordinate zero. By proposition 3.2.4, if m is a u-state, then $m^+ = \mu^+\times\eta$ for some stationary measure η. Hence,

$$\begin{aligned}\int_{M\times\mathbb{P}\mathbb{R}^2} \Phi\, dm &= \int_{M^+\times\mathbb{P}\mathbb{R}^2} \Phi\, dm^+ \\ &= \int_{M^+\times\mathbb{P}\mathbb{R}^2} \Phi\, d(\mu^+\times\eta) \\ &= \int_{M\times\mathbb{P}\mathbb{R}^2} \Phi\, d(\mu\times\eta).\end{aligned}$$

Therefore, $\max\left\{\int_{M\times\mathbb{P}\mathbb{R}^2} \Phi\, d(\mu\times\eta) : \ \eta \text{ is forward stationary for } \mathbb{P}F\right\}$ is greater than or equal to $\max\left\{\int_{M\times\mathbb{P}\mathbb{R}^2} \Phi\, dm : \ \text{m is a u-state for } \mathbb{P}F\right\}$, which is equal to λ_+.

Note that we could obtain the equality immediately from the statement (left unproven) in remark 7. Our proof, however, is overall more direct.

This concludes the proof of the first item. The second one is analogous, using dual arguments for $\mathbb{P}F$-invariant probability measures, backward stationary measures and s-states. ∎

Theorem 3.3.7 *(Furstenberg-Ledrappier's Formula)*
Let $F\colon M\times\mathbb{R}^2 \to M\times\mathbb{R}^2$ be a general random cocycle. Then,

$$\lambda_+(F,\mu) = \max\left\{\int_{M\times\mathbb{P}\mathbb{R}^2} \Phi\, d(\mu\times\eta) : \ \eta \textit{ is a stationary for } \mathbb{P}F\right\}.$$

Proof. Consider the invertible extension $\hat{F} : \hat{M} \times \mathbb{R}^d \to \hat{M} \times \mathbb{R}^d$ of the linear cocycle $F : M \times \mathbb{R}^d \to M \times \mathbb{R}^d$. Let $\hat{f} : \hat{M} \to \hat{M}$ be the two sided shift on $\hat{M} = X^{\mathbb{Z}}$ and let $\hat{A} : \hat{M} \to \mathrm{GL}(d)$ be given by $\hat{A}(\hat{x}) = A(\pi(\hat{x}))$. Define $\hat{F}(\hat{x}, v) = (\hat{f}(\hat{x}), \hat{A}(\hat{x})v)$.

Remember that we have already proved that $\lambda_+(F,\mu) = \lambda_+(\hat{F},\mu)$. Also note that a probability measure η on $\mathbb{PR}^d$ is stationary for $\mathbb{P}F$ if and only if it is forward stationary for $\mathbb{P}\hat{F}$. Indeed let $B \subset \mathbb{PR}^d$ be a measurable set. If η is stationary for $\mathbb{P}F$:

$$\begin{aligned}
\eta(B) &= \int_M \eta(A(x)^{-1}(B))\, d\mu(x) \\
&= \int_M \eta(A(x)^{-1}(B))\, d\pi_*\hat{\mu}(x) \\
&= \int_{\hat{M}} \eta(A(\pi(\hat{x}))^{-1}(B))\, d\hat{\mu}(x) \\
&= \int_{\hat{M}} \eta(\hat{A}(\hat{x})^{-1}(B))\, d\hat{\mu}(x).
\end{aligned}$$

Hence η is also stationary for $\mathbb{P}\hat{F}$. For the converse statement we just need to follow the equalities from bottom to top.

Now we can conclude that

$$\begin{aligned}
\lambda_+(F,\mu) &= \lambda_+(\hat{F},\mu) \\
&= \max\left\{ \int_{M\times\mathbb{PR}^2} \Phi\, d(\mu\times\eta) : \ \eta \text{ is forward stationary for } \mathbb{P}\hat{F} \right\} \\
&= \max\left\{ \int_{M\times\mathbb{PR}^2} \Phi\, d(\mu\times\eta) : \ \eta \text{ is a stationary for } \mathbb{P}F \right\}.
\end{aligned}$$

■

3.4
Coninuity of Lyapunov exponents for irreducible cocycles

In the previous section, we saw that the largest Lyapunov exponent is equal to the integral of the function Φ with respect to *some* $\mathbb{P}F$-invariant probability measure $m = \mu \times \eta$, where η is stationary. In this section, we will prove that under some irreducibility hypothesis, the Lyapunov exponent of random cocycles can be expressed as

$$\lambda_+(F,\mu) = \int_{M\times\mathbb{PR}^2} \Phi\, d(\mu\times\eta),$$

for *every* $\mathbb{P}F$-invariant probability measure of the form $\mu \times \eta$, where η is stationary. We will use this result to prove the continuity of the Lyapunov exponents of locally constant cocyles.

Definition 3.4.1 *A linear cocycle $F\colon M \times \mathbb{R}^2 \to M \times \mathbb{R}^2$, such that $F(x,v) = (f(x), A(x)v)$, is called strongly irreducible if there is no finite family of proper subspace of $\mathbb{R}^2$ which is invariant under $A(x)$ for μ-almost every $x \in M$.*

Remark 10 *Strong irreducibility is an open property.*

Theorem 3.4.1 *(Furstenberg's Formula) If $F : M \times \mathbb{R}^2 \to M \times \mathbb{R}^2$ is strongly irreducible, then*

$$\lambda_+(F,\mu) = \int_{M\times\mathbb{PR}^2} \Phi \; d(\mu \times \eta)$$

for every stationary measure η of the associated projective cocycle $\mathbb{P}F$.

Proof. Let η be a stationary measure of the associated projective cocycle $\mathbb{P}F$ and let $m = \mu \times \eta$. By the computation that we did in the proof of Ledrappier's Theorem and by Birkhoff's ergodic Theorem, we know that

$$\Psi(x,\hat{v}) = \lim_{n\to\infty} \frac{1}{n} \log \frac{\|A^n(x)v\|}{\|v\|} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \Phi \circ \mathbb{P}F^i(x,\hat{v})$$

exists at m-almost everywhere and it satisfies $\int_{M\times\mathbb{PR}^2} \Psi \; dm = \int_{M\times\mathbb{PR}^2} \Phi \; dm$.

By Oseledets Theorem, $\Phi(x,\hat{v}) = \lambda_+$ for μ-almost every $x \in M$ and every $v \notin V_x^2$. Remember $V_x^2 = E_x^s$ if $\lambda_+ > 0$ and $V_x^2 = \{0\}$ if $\lambda_+ = 0$. So, it suffices to prove that the set of pairs $(x,\hat{v})$ with $v \in V_x^2$ has zero m-measure. This will be the topic of the next lemma.

Lemma 3.4.2 *If the cocycle F is strongly irreducible, then $\eta(\hat{V}) = 0$ for any proper projective subspace $\hat{V}$ of $\mathbb{PR}^2$ and any $\mathbb{P}F$-stationary measure η.*

Proof. Note that a proper subspace V of $\mathbb{R}^2$ must be of dimension one and its projectivization $\hat{V}$ represents a point, so we are going to show that the measure η is non atomic.

Suppose by contradiction that there exists some proper projective subspace $\hat{V}$ with $\eta(\hat{V}) > 0$. Let c be the maximum value of $\eta(\hat{V})$ over all subspaces V of dimension one and let $\mathcal{V}$ be the family of all subspaces $\hat{V}$ such that $\eta(\hat{V}) = c$.

Since points are closed subsets of $\mathbb{PR}^2$, we have $\eta(\hat{V}) = c$ for every accumulation point $\hat{V}$ of any sequence $\{\hat{V}_n\}_n$ such that $\eta(\hat{V}_n) \to c$. By the compactness of the Grassmannian manifold $G(1,2)$, accumulation points do exist, hence the family $\mathcal{V}$ is non empty.

Moreover, $\mathcal{V}$ is finite because η is a probability measure and the elements of $\mathcal{V}$ are disjoint. We can also say that $\#\mathcal{V} \leq \frac{1}{c}$ and write: $\mathcal{V} = \{V_1, ..., V_n\}$.

Also, since η is $\mathbb{P}F$-stationary,

$$c = \eta(V_i) = \int_M \eta\left(A(x)^{-1}(V_i)\right)\, d\mu(x)$$

and $\eta\left(A(x)^{-1}(V_i)\right) \leq c$ for every $x \in M$. Therefore, we must have $\eta\left(A(x)^{-1}(V_i)\right) = c$ for μ-almost every $x \in M$, which means that $A(x)^{-1}(V_i) \in \mathcal{V}$ for μ-almost every $x \in M$ and contradicts the strong irreducibility hypothesis. ■

Thus, $\eta(V_x^2) = 0$ for every $x \in M$ and so $m(\{(x,\hat{v}) : v \in V_x^2\}) = 0$. Therefore, for any $\mathbb{P}F$-invariant probability measure $m = \mu \times \eta$, we have that

$$\int_{M\times\mathbb{PR}^2} \Phi\, dm = \int_{M\times\mathbb{PR}^2} \Psi\, dm = \lambda_+.$$

■

Finally, as an application of the previous formula, we are going to prove the continuity of the Lyapunov exponents for strongly irreducible cocycles.

Similarly to a cocycle being strongly irreducible, we can define an analogous concept for measures. Given a group of matrices, for example SL(2) and a compact subset Σ, consider a probability measure μ such that $\operatorname{supp}(\mu) \subset \Sigma$. We say that μ is strongly irreducible if there is no finite family of proper subspaces of $\mathbb{R}^2$ which is invariant under g for μ-almost every $g \in \mathrm{SL}(2)$.

Remember that we saw in example 3.1.1 that a probability measure μ with compact support Σ in SL(2) determines a random linear cocycle over the Bernoulli shift. Therefore, when we ask μ to be strongly irreducible, the induced random linear cocycle is also strongly irreducible, hence Furstenberg's Formula holds. Moreover, it holds for a neighbourhood of μ, since it is an open property.

In this context, consider $M = \Sigma^{\mathbb{N}}$, $\mu = \nu^{\mathbb{N}}$ and f to be the shift on the space (M,μ) and let μ be strongly irreducible.

Theorem 3.4.3 *Consider a sequence $\{\nu_k\}_k$ of probability measures in* SL(2) *with compact support $\Sigma_k \subset \Sigma \subset \mathrm{SL}(d)$ and $\mu_k = \nu_k^{\mathbb{N}}$. Suppose that ν_k converges to ν in the weak star topology. Then $\lambda_+(\mu_k) \to \lambda_+(\mu)$ when $k \to \infty$.*

Proof. For each $k \geq 1$, consider the cocycle F_k induced by μ_k and let η_k be a $\mathbb{P}F_k$-stationary measure related to it. By Furstenberg's Formula:

$$\lambda_+(\mu_k) = \int_{M\times\mathbb{PR}^2} \Phi(\{g_n\}_n, \hat{v})\, d(\mu_k(\{g_n\}_n) \times \eta_k(\hat{v})), \tag{3.5}$$

where

$$\Phi : M \times \mathbb{PR}^2 \to \mathbb{R}$$
$$\Phi(\{g_n\}_n, \hat{v}) = \log \frac{\|g_0 v\|}{\|v\|}.$$

Since we are in the case of remark 4, the space of stationary measures is compact in the weak star topology. Therefore, it has a converging subsequence η_{kl} and we may suppose that it converges to some $\mathbb{P}F$-stationary measure η.

Then, also by Furstenberg's Formula:

$$\lambda_+(\mu) = \int_{M \times \mathbb{PR}^2} \Phi(\{g_n\}_n, [v])\, d(\mu(\{g_n\}_n) \times \eta(\hat{v})), \tag{3.6}$$

So, we need to show that the integral in (3.5) converges to the integral in (3.6). First observe that

$$\int_{M \times \mathbb{PR}^2} \Phi(\{g_n\}_n, \hat{v})\, d(\mu_k(\{g_n\}_n) \times \eta_k(\hat{v})) = \int_{\Sigma \times \mathbb{PR}^2} \Phi(g_0, \hat{v})\, d(\nu_k(g_0) \times \eta_k(\hat{v}))$$

Since ν_k and η_{kl} converge to ν and η, respectively, both in the weak star topology, it follows that $\nu_{kl} \times \eta_{kl}$ converges to $\nu \times \eta$ in the weak star topology. Moreover, since the function Φ is continuous, for every $\varepsilon > 0$, there exists a sufficiently large integer l such that

$$\left| \int_{\Sigma \times \mathbb{PR}^2} \Phi(g_0, \hat{v})\, d(\nu_{kl}(g_0) \times \eta_{kl}(\hat{v})) - \int_{\Sigma \times \mathbb{PR}^2} \Phi(g_0, \hat{v})\, d(\nu(g_0) \times \eta(\hat{v})) \right| < \varepsilon.$$

Combining the previous results we get that $|\lambda_+(\mu_{kl}) - \lambda_+(\mu)| < \varepsilon$ for every l sufficiently large. We proved that there exists a subsequence $\{\lambda_+(\mu_{kl})\}_l$ that converges to $\lambda_+(\mu)$. However we can prove that the sequence $\{\lambda_+(\mu_k)\}_k$ in fact converges to $\lambda_+(\mu)$.

Instead of considering a subsequence η_{kl}, we should consider a subsequence η_{kl_j} of this subsequence. Hence, using the same arguments, we will conclude every subsequence of $\{\lambda_+(\mu_k)\}_k$ has a further subsequence $\{\lambda_+(\mu_{kl_j})\}_j$ that converges to $\lambda_+(\mu)$. This implies that the sequence $\{\lambda_+(\mu_k)\}_k$ itself converges to $\lambda_+(\mu)$. ■

4
The Wasserstein Metric

The weak-star topology on the set of probability measures is metrizable in various ways, of which, the Wasserstein metric is one of the most useful. Our result on the Hölder continuity of the Lyapunov exponent is formulated relative to this metric. In this chapter we introduce the Wasserstein metric and study some of its main properties.

In section 4.1 we introduce the concept of coupling measures, define Wasserstein's metric and prove some of its basic properties. We devote section 4.2 to the proof of a duality theorem of Kantorovich-Rubinstein which gives a characterization of the Wasserstein metric. It will be an important tool in the proof of our main theorem. Finally, in section 4.3, we introduce the concept of convolution of measures and its relations with the Wasserstein's metric, which will be used many times in chapter 5.

Most of the results in this chapter follow the lecture notes and exercises of a measure theoretic course held at PUC-Rio in 2020, which were adapted from Villani's monograph [13].

4.1
A useful metric in the space of probability measures

In this chapter we will always consider (X, d) to be a compact metric space, which will be enough for our purposes. There are also similar results when X is a Polish space. The maps $\mathrm{proj}_i : X \times X \to X$, with $i \in \{1, 2\}$, such that $\mathrm{proj}_1(x, y) = x$ and $\mathrm{proj}_2(x, y) = y$, denote the canonical projections on the first and second coordinate respectively. Moreover, $\mathrm{Prob}(X)$ denote the set of probability measures on X and we always consider convergence in $\mathrm{Prob}(X)$ with respect to the weak star topology.

Definition 4.1.1 *Let $(X, \mathcal{B})$ be a measurable space. Given two measures $\mu, \nu \in \mathrm{Prob}(X)$, a coupling between μ and ν is a measure $\pi \in \mathrm{Prob}(X \times X)$ with marginals μ and ν: $(\mathrm{proj}_1)_*\pi = \mu$ and $(\mathrm{proj}_2)_*\pi = \nu$.*

A trivial example of a coupling between two measures is the product of two probability measures:

Example 4.1.1 *Consider the product measure $\mu \times \nu \in \mathrm{Prob}(X_1 \times X_2)$, where μ and ν are probability measures for X_1 and X_2 respectively. Then for every measurable set $E \subset X_1$:*

$$(\mathrm{proj}_1)_*(\mu \times \nu)(E) = (\mu \times \nu)\left[(\mathrm{proj}_1)^{-1}(E)\right] = (\mu \times \nu)(E \times X_2) = \mu(E).$$

Therefore, $(\mathrm{proj}_1)_*(\mu \times \nu) = \mu$. Analogously, $(\mathrm{proj}_2)_*(\mu \times \nu) = \nu$.

Given two measures $\mu, \nu \in \mathrm{Prob}(X)$ we can also consider the set of all possible couplings between μ and ν:

$$\Pi(\mu, \nu) := \{\pi \in \mathrm{Prob}(X \times X) : (\mathrm{proj}_1)_*\pi = \mu \text{ and } (\mathrm{proj}_2)_*\pi = \nu\}.$$

Remark 11 *Note that the product measure $\mu \times \nu \in \Pi(\mu, \nu)$. Hence the space $\Pi(\mu, \nu)$ is not empty.*

Lemma 4.1.1 *Let X and Y be metric spaces and $f : X \to Y$ a continuous function. If $\mu_n \to \mu$ in $\mathrm{Prob}(X)$, then $f_*\mu_n \to f_*\mu$ in $\mathrm{Prob}(Y)$.*

Proof. Let $\varphi : Y \to \mathbb{R}$ be a continuous function. By hypothesis, $\mu_n \to \mu$, hence, using the exchange of variables formula:

$$\int_Y \varphi \, d(f_*\mu_n) = \int_X \varphi \circ f \, d\mu_n \to \int_X \varphi \circ f \, d\mu = \int_Y \varphi \, d(f_*\mu).$$

Note that φ and f are continuous functions, hence so is $\varphi \circ f$. Therefore, we concluded that $f_*\mu_n \to f_*\mu$. Moreover, the function $\mu \mapsto f_*\mu$ is continuous with respect to the weak star topology. ■

Proposition 4.1.2 $\Pi(\mu, \nu)$ *is closed in the weak star topology.*

Proof. Let $\{\pi_n\}_n$ be a sequence in $\Pi(\mu, \nu)$ that converges to π. Since $\pi_n \in \Pi(\mu, \nu)$ for every $n \in \mathbb{N}$, we have that

$$(\mathrm{proj}_1)_*\pi_n = \mu \quad \text{and} \quad (\mathrm{proj}_2)_*\pi_n = \nu$$

for every $n \in \mathbb{N}$. Moreover, proj_1 and proj_2 are continuous functions, then

$$\mu = (\mathrm{proj}_1)_*\pi_n \to (\mathrm{proj}_1)_*\pi \quad \text{and} \quad \nu = (\mathrm{proj}_2)_*\pi_n \to (\mathrm{proj}_2)_*\pi,$$

by lemma 4.1.1. Therefore $\pi \in \Pi(\mu, \nu)$. ■

Definition 4.1.2 *Given $\mu, \nu \in \mathrm{Prob}(X)$ and $p \geq 1$,*

$$W_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times X} d(x, y)^p \, d\pi(x, y) \right)^{\frac{1}{p}}$$

is called the Wasserstein distance of order p.

An immediate consequence is that, if $p_1 \leq p_2$, then $W_{p_1} \leq W_{p_2}$ by Hölder's inequality.

In our text, we are going to work mainly with the distance W_1.

Proposition 4.1.3 *There exists $\pi^* \in \Pi(\mu, \nu)$ such that*

$$\int_{X\times X} d(x,y)^p \; d\pi^*(x,y) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{X\times X} d(x,y)^p \; d\pi(x,y).$$

Thus, the infimun is attained.

Proof. There exists a sequence $\{\pi_n\}_n \subset \Pi(\mu, \nu) \subset \text{Prob}(X \times X)$ such that

$$\int_{X\times X} d(x,y)^p \; d\pi_n(x,y) \to \inf_{\pi\in\Pi(\mu,\nu)} \int_{X\times X} d(x,y)^p \; d\pi(x,y),$$

by the definition of infimun. Moreover, since $X \times X$ is compact, by Prohorov's Theorem there exists a converging subsequence $\pi_{nk} \to \pi^*$. Note that not only $\pi^* \in \text{Prob}(X \times X)$ but also $\pi^* \in \Pi(\mu, \nu)$ by proposition (4.1.2). By the fact that the distance is a continuous function,

$$\int_{X\times X} d(x,y)^p \; d\pi_{nk}(x,y) \to \int_{X\times X} d(x,y)^p \; d\pi^*(x,y).$$

Therefore,

$$\int_{X\times X} d(x,y)^p \; d\pi^* v(x,y) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{X\times X} d(x,y)^p \; d\pi(x,y).$$

■

The next lemma gives two new equivalent definitions for $\Pi(\mu, \nu)$.

Lemma 4.1.4 *For a probability measure $\pi \in \text{Prob}(X \times X)$, the following are equivalent:*

(i) $\pi \in \Pi(\mu, \nu)$.

(ii) For every measurable sets $A \subset X$ and $B \subset X$,

$$\pi(A \times X) = \mu(A) \quad \textit{and} \quad \pi(X \times B) = \nu(B).$$

(iii) For every positive measurable functions φ, $\psi : X \to [0, +\infty)$,

$$\int_{X\times X} [\varphi(x) + \psi(y)] \; d\pi(x,y) = \int_X \varphi(x) \; d\mu(x) + \int_X \psi(y) \; d\nu(y).$$

Proof. First we prove $(i) \Leftrightarrow (ii)$. Consider two measurable sets $A \subset X$ and $B \subset X$. Assume π is a coupling between μ and ν:

$$(\mathrm{proj}_1)_*\pi = \mu \quad \text{and} \quad (\mathrm{proj}_2)_*\pi = \nu.$$

Therefore,

$$\begin{aligned}\mu(A) &= \pi\left(\mathrm{proj}_1^{-1}(A)\right) = \pi(A \times X) \quad \text{and}\\ \nu(B) &= \pi\left(\mathrm{proj}_2^{-1}(B)\right) = \pi(X \times B).\end{aligned}$$

For the converse statement, let A and B be arbitrary measurable sets. By (ii),

$$\begin{aligned}(\mathrm{proj}_1)_*\pi(A) &= \pi\left(\mathrm{proj}_1^{-1}(A)\right) = \pi(A \times X) = \mu(A) \quad \text{and}\\ (\mathrm{proj}_2)_*\pi(B) &= \pi\left(\mathrm{proj}_1^{-1}(B)\right) = \pi(X \times B) = \nu(B).\end{aligned}$$

Thus, π is a coupling between μ and ν.

Now we are going to show that $(i) \Leftrightarrow (iii)$. First, assume that $\pi \in \Pi(\mu,\nu)$. Consider $\varphi, \psi : X \to [0,+\infty)$ measurable functions. Then,

$$\begin{aligned}&\int_{X\times X} [\varphi(x) + \psi(y)] \; d\pi(x,y)\\ &= \int_{X\times X} \varphi \circ \mathrm{proj}_1(x,y) \; d\pi(x,y) + \int_{X\times X} \psi \circ \mathrm{proj}_2(x,y) \; d\pi(x,y)\\ &= \int_X \varphi(x) \; d(\mathrm{proj}_1)_*\pi(x) + \int_X \psi(y) \; d(\mathrm{proj}_2)_*\pi(y)\\ &= \int_X \varphi(x) \; d\mu(x) + \int_X \psi(y) \; d\nu(y)\end{aligned}$$

where the last line uses the hypothesis that π is a coupling between μ and ν.

Finally, assume that (iii) holds. Consider a measurable set A. Then

$$\mu(A) = \int_X \mathcal{X}_A(x) \; d\mu(x) = \int_{X\times X} \mathcal{X}_A(x) \; d\pi(x,y) = \int_{A\times X} 1 \; d\pi(x,y) = \pi(A \times X),$$

where we used (iii) in the second equality. Similarly, the same holds for ν and we conclude that $\nu(B) = \pi(X \times B)$, which finishes the proof. ■

Proposition 4.1.5 *W_1 is a metric in* $\mathrm{Prob}(X)$.

Proof. Note that, from the definition of W_1 and the compactness of X, it follows that it is positive and finite $0 \leq W_1 < \infty$ and also that it is symmetric $W_1(\mu,\nu) = W_1(\nu,\mu)$.

Now, we are going to show that $W_1(\mu,\nu) = 0 \iff \mu = \nu$. Suppose $W_1(\mu,\nu) = 0$. Then, there exists $\pi \in \Pi(\mu,\nu)$ such that $\int_{X\times X} d(x,y) \; d\pi(x,y) = 0$. Since $d(x,y) \geq 0$ for every $x, y \in X$, we must have $\mathrm{supp}(\pi) \subset \{(x,x) : x \in X\}$. This implies that $\mathrm{proj}_1 = \mathrm{proj}_2$. Moreover, since $\pi \in \Pi(\mu,\nu)$, $\mu = (\mathrm{proj}_1)_*\pi$ and $\nu = (\mathrm{proj}_2)_*\pi$. By the previous observation, we conclude that $\mu = \nu$. In order to prove the converse we can follow the proof backwards.

Finally, we are going to prove the triangle inequality. Let $X_1 = X_2 = X_3 = X$ and $\mu_1, \mu_2, \mu_3 \in \mathrm{Prob}(X)$. Let $\pi_{12} \in \Pi(\mu_1,\mu_2)$ and $\pi_{23} \in \Pi(\mu_2,\mu_3)$ such that

$$\begin{aligned} W_1(\mu_1,\mu_2) &= \int_{X_1\times X_2} d(x_1,x_2)\, d\pi_{12}(x_1,x_2) \quad \text{and} \\ W_1(\mu_2,\mu_3) &= \int_{X_2\times X_3} d(x_2,x_3)\, d\pi_{23}(x_2,x_3). \end{aligned}$$

We claim that there exists $\mu \in \mathrm{Prob}(X_1\times X_2\times X_3)$ such that

$$(\mathrm{proj}_{12})_*\mu = \pi_{12} \quad \text{and} \quad (\mathrm{proj}_{23})_*\mu = \pi_{23}.$$

In order to prove the claim, consider the disintegrations $(\pi_{12})_{x_2} \in \mathrm{Prob}(X)$ of π_{12} with respect to μ_2 and $(\pi_{23})_{x_2} \in \mathrm{Prob}(X)$ of π_{23} with respect to μ_2:

$$\pi_{12} = \int_{X_2} (\pi_{12})_{x_2}\, d\mu_2(x_2) \quad \text{and} \quad \pi_{23} = \int_{X_2} (\pi_{23})_{x_2}\, d\mu_2(x_2).$$

Consider $\mu \in \mathrm{Prob}(X_1\times X_2\times X_3)$ such that $\mu = (\pi_{12})_{x_2} \times \mu_2 \times (\pi_{23})_{x_2}$. Given $E \in \mathcal{B}(X_1\times X_2\times X_3)$,

$$\mu(E) := \int_{X_2} (\pi_{12})_{x_2} \times (\pi_{23})_{x_2}(E_{x_2})\, d\mu_2(x_2),$$

where $E_{x_2} = \{(x_1,x_3) \in X_1\times X_3 : (x_1,x_2,x_3) \in E\}$. We are going to show that μ satisfies the claim.

Let $A \subset X_1 \times X_2$ be a measurable set. Then

$$\begin{aligned} (\mathrm{proj}_{12})_*\mu(A) &= \mu_*(\mathrm{proj}_{12}^{-1}(A)) \\ &= \mu(A\times X_3) \\ &= \int_{X_2} (\pi_{12})_{x_2} \times (\pi_{23})_{x_2}(A\times X_3)\, d\mu_2(x_2) \\ &= \int_{X_2} (\pi_{12})_{x_2}(A)\, d\mu_2(x_2) \\ &= \pi_{12}(A). \end{aligned}$$

An analogous computation shows that every measurable set $B \subset X_2\times X_3$ satisfies $(\mathrm{proj}_{23})_*\mu(B) = \pi_{23}(B)$, which concludes the proof of the claim.

Now define $\pi_{13} := (\mathrm{proj}_{13})_*\mu$. We claim that $\pi_{13} \in \Pi(\mu_1, \mu_3)$. Let $A \subset X_1$ be a measurable set. Then

$$\begin{aligned}
(\mathrm{proj}_1)_*\pi_{13}(A) &= (\mathrm{proj}_1)_*(\mathrm{proj}_{13})_*\mu(A) \\
&= (\mathrm{proj}_{13})_*\mu(\mathrm{proj}_1^{-1}(A)) \\
&= (\mathrm{proj}_{13})_*\mu(A \times X_3) \\
&= \mu\left(\mathrm{proj}_{13}^{-1}(A \times X_3)\right) \\
&= \mu(A \times X_2 \times X_3) \\
&= \mu_1(A).
\end{aligned}$$

An analogous computation shows that every measurable set $B \subset X_3$ satisfies $(\mathrm{proj}_3)_*\pi_{13}(B) = \mu_3(B)$.

Finally, consider the triangle inequality $d(x_1, x_3) \leq d(x_1, x_2) + d(x_2, x_3)$ and integrate it with respect to μ:

$$\begin{aligned}
\int_{X_1\times X_2\times X_3} d(x_1, x_3)\ d\mu(x_1, x_2, x_3) \leq &\int_{X_1\times X_2\times X_3} d(x_1, x_2)\ d\mu(x_1, x_2, x_3) \\
&+ \int_{X_1\times X_2\times X_3} d(x_2, x_3)\ d\mu(x_1, x_2, x_3).
\end{aligned}$$

Notice that we can rewrite the previous inequality as

$$\begin{aligned}
\int_{X_1\times X_3} d(x_1, x_3)\ d\pi_{13}(x_1, x_3) \leq &\int_{X_1\times X_2} d(x_1, x_2)\ d\pi_{12}(x_1, x_2) \\
&+ \int_{X_2\times X_3} d(x_2, x_3)\ d\pi_{23}(x_2, x_3) \\
&= W_1(\mu_1, \mu_2) + W_1(\mu_2, \mu_3).
\end{aligned}$$

By definition, $W_1(\mu, \nu) \leq \int_{X_1\times X_3} d(x_1, x_3)\ d\pi_{13}(x_1, x_3)$. Hence, we conclude that

$$W_1(\mu_1, \mu_3) \leq W_1(\mu_1, \mu_2) + W_1(\mu_2, \mu_3),$$

which finishes the proof. ■

Remark 12 *W_p is also a metric for every $p \geq 1$. The proof of this fact follows from Minkowsi's Inequality.*

4.2 Kantorovich-Rubinstein's duality Theorem

In this section we prove the main result of the chapter, the Kantorovich-Rubinstein's duality Theorem.

Definition 4.2.1 *Let (X, d) be a compact metric space. Given $L < \infty$, we can define the space of Lipschitz functions with constant less or equal to L:*

$$\mathrm{Lip}_L(X) := \{\varphi : X \to \mathbb{R} : |\varphi(x) - \varphi(y)| \leq L\, d(x, y),\ \forall x, y \in X\}.$$

Definition 4.2.2 *Given $\mu, \nu \in \mathrm{Prob}(X)$, we define*

$$\gamma(\mu, \nu) := \sup_{\varphi \in \mathrm{Lip}_1(X)} \left| \int_X \varphi\, d\mu - \int_X \varphi\, d\nu \right|.$$

Remark 13 *It follows from Arzelà-Ascoli theorem that there exists $\varphi \in \mathrm{Lip}_1(X)$ such that*

$$\int_X \varphi\, d(\mu - \nu) = \gamma(\mu, \nu).$$

Thus, the supremum is attained.

Theorem 4.2.1 *(Kantorovich-Rubinstein) Let (X, d) be a compact metric space. Then*

$$W_1(\mu, \nu) = \gamma(\mu, \nu) \quad \forall \mu, \nu \in \mathrm{Prob}(X).$$

Proof. In order to prove this theorem, we need an intermediate result. Given $\varphi, \psi \in C(X)$, let $\varphi \oplus \psi \in C(X \times X)$ be the function $\varphi \oplus \psi(x, y) := \varphi(x) + \psi(y)$. Define

$$W^*(\mu, \nu) := \sup \left\{ \int_X \varphi\, d\mu + \int_X \psi\, d\nu : \varphi, \psi \in C(X),\ \varphi \oplus \psi \leq d \right\}.$$

Note that by lemma (4.1.4) we have that for every $\pi \in \Pi(\mu, \nu)$:

$$W^*(\mu, \nu) = \sup \left\{ \int_{X \times X} \varphi \oplus \psi\, d\pi : \varphi, \psi \in C(X), \quad \varphi \oplus \psi \leq d \right\}.$$

We will prove that $\gamma(\mu, \nu) \leq W^*(\mu, \nu) \leq W_1(\mu, \nu) \leq W^*(\mu, \nu) \leq \gamma(\mu, \nu)$. Thus, we split the proof into four lemmas, one for each inequality.

Lemma 4.2.2 *For every $\mu, \nu \in \mathrm{Prob}(X)$, $\gamma(\mu, \nu) \leq W^*(\mu, \nu)$.*

Proof. Given $\varphi \in \mathrm{Lip}_1(X)$, we know that $\varphi, -\varphi \in C(X)$ and also

$$\int_X \varphi\, d\mu - \int_X \varphi\, d\nu = \int_X \varphi\, d\mu + \int_X (-\varphi)\, d\nu = \int_X \varphi \oplus (-\varphi)\, d\pi$$

for every $\pi \in \Pi(\mu, \nu)$. Furthermore, $\varphi \oplus (-\varphi)(x, y) = \varphi(x) - \varphi(y) \leq d(x, y)$ because $\varphi \in \mathrm{Lip}_1(X)$. Thus, for every $\varphi \in \mathrm{Lip}_1(X)$,

$$\int_X \varphi\, d\mu - \int_X \varphi\, d\nu \leq W^*(\mu, \nu).$$

Therefore, taking the supremum over all $\varphi \in \mathrm{Lip}_1(X)$, we conclude that $\gamma(\mu,\nu) \le W^*(\mu,\nu)$. ■

Lemma 4.2.3 *For every $\mu,\nu \in \mathrm{Prob}(X)$, $W^*(\mu,\nu) \le W_1(\mu,\nu)$.*

Proof. Notice that for every $\varphi,\psi \in C(X)$ such that $\varphi \oplus \psi \le d$ and every probability measure $\pi \in \Pi(\mu,\nu)$,

$$\int_X \varphi \, d\mu + \int_X \psi \, d\nu = \int_{X\times X} \varphi \oplus \psi \, d\pi \le \int_{X\times X} d(x,y) \, d\pi(x,y).$$

Hence, by taking the supremumm over the functions $\varphi,\psi \in C(X)$: $\varphi \oplus \psi \le d$, it follows that

$$W^*(\mu,\nu) \le \int_{X\times X} d(x,y) \, d\pi(x,y)$$

for every $\pi \in \Pi(\mu,\nu)$. Finally, taking the infimum over the coupling measures $\pi \in \Pi(\mu,\nu)$, we conclude that $W^*(\mu,\nu) \le W_1(\mu,\nu)$. ■

Lemma 4.2.4 *For every $\mu,\nu \in \mathrm{Prob}(X)$, $W_1(\mu,\nu) \le W^*(\mu,\nu)$.*

Proof. Let $V = C(X\times X)$ be endowed with the uniform norm. Consider

$$E := \{f \in C(X\times X) : \exists\, \varphi,\psi \in C(X) \text{ such that } f = \varphi \oplus \psi\},$$

which is a subspace of V. Also consider the following set that we claim to be open and convex

$$U := \{f \in C(X\times X) : f < d\}.$$

Indeed, since X is compact, if $f(x,y) < d(x,y)$ $\forall$ $(x,y) \in X\times X$, then $\|f-d\|_0 := \varepsilon_0 > 0$. Thus, if $\|g-f\|_0 \le \frac{\varepsilon_0}{2}$, we conclude that $g < d$, which implies that U is open. Moreover, given $f_1, f_2 \in U$ and $\lambda \in (0,1)$, notice that $\lambda f_1 + (1-\lambda f_2) < \lambda d + (1-\lambda)d < d$. Hence, $\lambda f_1 + (1-\lambda f_2) \in U$, which means U is convex.

Define the linear operator $I : E \to \mathbb{R}$, such that

$$\begin{aligned} I(f) :&= \int_X \varphi \, d\mu + \int_X \psi \, d\nu \\ &= \int_{X\times X} \varphi \oplus \psi \, d\pi \quad \text{for every } \pi \in \Pi(\mu,\nu) \\ &= \int_{X\times X} f \, d\pi \qquad\quad \text{for every } \pi \in \Pi(\mu,\nu). \end{aligned}$$

Note that I is bounded, because $\|I(f)\| \le \|f\|\pi(X\times X) = \|f\|$. Moreover, $E \cap U = \{\varphi\oplus\psi : \varphi,\psi \in C(X) \text{ and } \varphi\oplus\psi < d\}$. Therefore,

$$W^*(\mu,\nu) = \sup_{E\cap U} I(f) := \alpha.$$

Hence, by the special version of Hahn-Banach's Theorem A.0.1, there exists a linear functional $\tilde{I} : C(X \times X) \to \mathbb{R}$ that satisfies $\tilde{I}(f) = I(f)$ for every $f \in E$ and $\sup\limits_{f \in U} \tilde{I}(f) = \sup\limits_{f \in U \cap E} I(f) = W^*(\mu, \nu)$.

We are going to show that $W^*(\mu, \nu) = \sup\limits_{f \in U} \tilde{I}(f) \geq W_1(\mu, \nu)$. Our plan is to prove that $\tilde{I}$ is monotone, hence $\sup\limits_{f \in U} \tilde{I}(f) = \tilde{I}(d)$. Moreover, we are going to prove that there exists $\pi_0 \in \Pi(\mu, \nu)$ such that

$$\tilde{I}(f) = \int_{X \times X} f \, d\pi_0 \quad \text{for every } f \in C(X \times X),$$

using Riesz-Markov Theorem A.0.5. Therefore, we can conclude that

$$\begin{aligned} W^*(\mu, \nu) = \tilde{I}(d) = \int_{X \times X} d(x, y) \, d\pi_0(x, y) &\geq \inf_{\pi \in \Pi(\mu, \nu)} \int_{X \times X} d(x, y) \, d\pi(x, y) \\ &= W_1(\mu, \nu), \end{aligned}$$

which will finish the proof.

First note that $1 = 1 \oplus 0$. Hence $\tilde{I}(1) = \int_X 1 \, d\mu + \int_X 0 \, d\nu = 1$.

Now we are going to show that $\tilde{I}$ is positive (see definition A.0.3). Let $f \in C(X \times X)$, $f \geq 0$. For every $t \geq 0$ and $\varepsilon > 0$, consider the function

$$d - tf - \varepsilon \in C(X \times X).$$

Notice that $d - tf - \varepsilon \in C(X \times X) < d$, hence it is in U. Thus,

$$\tilde{I}(d - tf - \varepsilon) \leq \sup_{U \cap E} I = W^*(\mu, \nu) < \infty.$$

Moreover, by linearity,

$$\tilde{I}(d - tf - \varepsilon) = \tilde{I}(d) - t\tilde{I}(f) - \tilde{I}(\varepsilon) < \infty \quad \forall t > 0.$$

Therefore, we must have $\tilde{I}(f) \geq 0$. If it was not positive, then we would have $\tilde{I}(d) - t\tilde{I}(f) - \tilde{I}(\varepsilon) \to \infty$ when $t \to \infty$, which contradicts the fact that $\tilde{I}(d - tf - \varepsilon) \leq W^*(\mu, \nu)$.

We claim that since $\tilde{I}$ is positive, it is also monotone. Let $f_1 \leq f_2$, so that $f_2 - f_1 \geq 0$. Then $\tilde{I}(f_2 - f_1) \geq 0$. Finally, by linearity, $\tilde{I}(f_2) - \tilde{I}(f_1) \geq 0$, hence $\tilde{I}(f_2) \geq \tilde{I}(f_1)$.

Hence, we are in the conditions of Riesz-Markov Theorem, so there exists $\pi_0 \in \text{Prob}(X \times X)$ such that

$$\tilde{I}(f) = \int_{X \times X} f \, d\pi_0 \quad \text{for every } f \in C(X \times X).$$

It remains to show that $\pi_0 \in \Pi(\mu, \nu)$. Notice that $\tilde{I}$ is an extension of I, so we know that for every $\varphi, \psi \in C(X)$

$$\tilde{I}(\varphi + \psi) = I(\varphi + \psi) = \int_X \varphi \, d\mu + \int_X \psi \, d\nu = \int_{X \times X} \varphi \oplus \psi \, d\pi_0.$$

Therefore, $\pi_0 \in \Pi(\mu, \nu)$, which, by the previous observations, completes the proof.

■

Lemma 4.2.5 *For every $\mu, \nu \in \mathrm{Prob}(X)$, $W^*(\mu, \nu) \leq \gamma(\mu, \nu)$.*

Proof. We are going to prove that for every $\varepsilon > 0$,

$$W^*(\mu, \nu) - \ \varepsilon \ < \ \gamma(\mu, \nu).$$

Fix $\varepsilon > 0$. Then, by the definition of W^*, there exist $\varphi, \psi \in C(X)$, with $\varphi \oplus \psi \leq d$ such that

$$W^*(\mu, \nu) - \varepsilon < \int_X \varphi \, d\mu + \int_X \psi \, d\nu.$$

We are going to construct a function $\tau \in \mathrm{Lip}_1(X)$, using φ and ψ such that $\varphi \leq \tau$ and $\psi \leq -\tau$. Suppose that τ exists. Then

$$\begin{aligned} \int_X \varphi \, d\mu + \int_X \psi \, d\nu &\leq \int_X \tau \, d\mu + \int_X -\tau \, d\nu \\ &= \int_X \tau \, d(\mu - \nu) \leq \gamma(\mu, \nu). \end{aligned}$$

Hence, for every $\varepsilon > 0$, we have that $W^*(\mu, \nu) - \varepsilon < \gamma(\mu, \nu)$, which concludes the proof. It remains to construct τ.

Define

$$\tau(x) := \inf_{y \in X} d(x, y) - \psi(y).$$

Since $\varphi \oplus \psi \leq d$, we have that $\varphi(x) \leq d(x, y) - \psi(y)$ for every $x, y \in X$. Hence, $\varphi(x) \leq \tau(x)$. Moreover, $\tau(x) \leq d(x, x) - \psi(x) = -\psi(x)$.

Finally, we are going to verify that $\tau \in \mathrm{Lip}_1(X)$. Fix $x_1, x_2 \in X$, then

$$\begin{aligned} \tau(x_1) &= \inf_{y \in X} [d(x_1, y) - \psi(y)] \\ &\leq \inf_{y \in X} [d(x_1, x_2) + d(x_2, y) - \psi(y)] \\ &= d(x_1, x_2) + \tau(x_2). \end{aligned}$$

This means that $\tau(x_1) - \tau(x_2) \leq d(x_1, x_2)$ and $\tau \in \mathrm{Lip}_1(X)$, which finishes the proof. ■

Therefore, by lemmas 4.2.2 to 4.2.5 we conclude the proof of Kantorovich-Rubinstein's duality Theorem.

■

Theorem 4.2.6 *Let* (X, d) *be a compact metric space,* $\{\mu_n\}_{n\geq 1} \subset \mathrm{Prob}(X)$*. Then*

$$\mu_n \to \mu \text{ (in the weak star topology) } \iff W_1(\mu_n, \mu) \to 0,$$

which means that convergence in the weak star topology is equivalent to the convergence of W_1 *distance in* $\mathrm{Prob}(X)$*. In particular,* W_1 *metrizes the weak star topology. (In fact the same is true for* W_p*,* $p \geq 1$*).*

Proof. First suppose $W_1(\mu_n, \mu) \to 0$. Let $\varphi \in \mathrm{Lip}(X)$ with Lipschitz constant $0 < L < \infty$, so that $\varphi \in \mathrm{Lip}_L(X)$ and $\frac{1}{L}\varphi \in \mathrm{Lip}_1(X)$. Then, by Kantorovich-Rubinstain duality Theorem (4.2.1):

$$\int_X \frac{1}{L}\varphi \, d(\mu_n - \mu) \leq \gamma(\mu_n, \mu) = W_1(\mu_n, \mu) \to 0.$$

Therefore, for every $\varphi \in \mathrm{Lip}(X)$:

$$\int_X \varphi \, d\mu_n \to \int_X \varphi \, d\mu,$$

which by Portmanteau's Theorem means that $\mu_n \to \mu$ in the weak star topology.

For the converse statement, we assume that μ_n converges to μ in the weak star topology and it is sufficient to prove that $\limsup_{n\to\infty} W_1(\mu_n, \mu) = 0$. Suppose by contradiction that there exists a subsequence $\{\mu_{nk}\}_k$ such that $\lim_{k\to\infty} W_1(\mu_{n_k}, \mu) = \alpha \geq 0$. Then, again by the Kantorovich Rubinstein duality Theorem,

$$\lim_{k\to\infty} \max_{\varphi\in \mathrm{Lip}_1(X)} \int_X \varphi \, d(\mu_{N_k} - \mu) = \alpha.$$

Hence, there exist some sequence $\{\varphi_{n_k}\}_k$ of functions in $\mathrm{Lip}_1(X)$ such that

$$\int_X \varphi_{n_k} \, d(\mu_{n_k} - \mu) = \gamma(\mu_{n_k} - \mu) \to \alpha.$$

By Arzelà-Ascoli's Theorem, there exists a subsequence of $\{\varphi_{n_k}\}_k$ that converges uniformly to a function $\varphi \in \mathrm{Lip}_1(X)$. So, for every $\varepsilon > 0$, there exists $K_0 > 0$ such that for every $K \geq K_0$, $\|\varphi_{n_{k_j}} - \varphi\|_0 \leq \varepsilon$.

Thus, for every $K \geq K_0$,

$$\begin{aligned}\left|\int_X \varphi_{n_{kj}}\, d(\mu_{n_{kj}} - \mu)\right| &\leq \left|\int_X (\varphi_{n_{kj}} - \varphi)\, d(\mu_{n_{kj}} - \mu)\right| + \left|\int_X \varphi\, d(\mu_{n_{kj}} - \mu)\right| \\ &\leq \left|\int_X (\varphi_{n_{kj}} - \varphi)\, d(\mu_{n_{kj}})\right| + \left|\int_X (\varphi_{n_{kj}} - \varphi)\, d(\mu)\right| + \varepsilon \\ &\leq 3\varepsilon.\end{aligned}$$

We concluded that

$$\int_X \varphi_{n_{kj}}\, d(\mu_{n_{kj}} - \mu) \to 0,$$

which is a contradiction. Therefore, we must have $\alpha = \limsup_{n\to\infty} W_1(\mu_n, \mu) = 0$, hence

$$W_1(\mu_n, \mu) \to 0.$$

■

Moreover, we conclude that convergence in weak star topology is equivalent to convergence in the Wasserstein's metric W_1. This means that both topologies, the weak star and the one induced by W_1 are the same.

4.3 Convolution of measures

We begin with the definition of the convolution of measures in a more general setting for measures on groups. Afterwards we are going to work mainly in the particular case where measures are supported in a compact set of the group.

Definition 4.3.1 *Let G be a group that acts on a set M. Let μ be a measure in G and ν a be measure in M. Then we define the convolution of μ and ν as the measure $\mu * \nu$ on M such that:*

$$(\mu * \nu)(E) = \int_G \int_M 1_E(gx)\, d\nu(x) d\mu(g)$$

for every measurable set $E \subset M$.

Then, by standard arguments of measure theory, we conclude that, in the same context,

$$\int_M f(x)\, d(\mu * \nu)(x) = \int_G \int_M f(gx)\, d\nu(x) d\mu(g)$$

for every $f \in L^1(M)$.

Given a measure $\mu \in \mathrm{Prob}(G)$, and $k \geq 2$ we define

$$\mu^{*k} := \mu * ... * \mu \qquad (k \text{ times})$$

the k-th convolution of μ with itself.

We can also define $\mu^{*1} := \mu$ or think that μ^{*1} is the convolution of μ with a Dirac measure centered at the identity element of the group G.

Proposition 4.3.1 *Let $\Sigma \subset \mathrm{SL}(2)$ be a compact set. Given $\mu \in \mathrm{Prob}(\Sigma)$ and n fixed, the map $\mu \mapsto \mu^{*n}$ is Lipschitz with respect to the Wasserstein metric.*

Proof. We split the proof into the three following lemmas:

Lemma 4.3.2 *Fix $n \in \mathbb{N}$, then the map $\mu \mapsto \mu \times ... \times \mu$ (n times), is Lipschitz with respect to the Wasserstein metric, with Lipschitz constant n.*

Proof. Let $\varphi \in \mathrm{Lip}_1\, \Sigma \times \Sigma$. Observe that

$$\int_{\Sigma\times\Sigma} \varphi(g,h)\ d\mu(g)d\mu(h) - \int_{\Sigma\times\Sigma} \varphi(g,h\ )d\nu(g)d\nu(h) =$$
$$\int_{\Sigma\times\Sigma} \varphi(g,h)\ d(\mu-\nu)(g)d\mu(h) + \int_{\Sigma\times\Sigma} \varphi(g,h)\ d(\mu-\nu)(h)d\nu(g).$$

Now fix h. The map $g \mapsto \varphi(g,h)$ is 1-Lipschitz. Then

$$\int_{\Sigma\times\Sigma} \varphi(g,h)\ d(\mu-\nu)(g)d\mu(h) \leq \int_{\Sigma\times\Sigma} W_1(\mu,\nu)\ d\mu(h) \leq W_1(\mu,\nu),$$

since $\mu \in \mathrm{Prob}(\Sigma)$. The same result is true for the other term:

$$\int_{\Sigma\times\Sigma} \varphi(g,h)\ d(\mu-\nu)(h)d\nu(g) \leq W_1(\mu,\nu).$$

Therefore we conclude that $W_1(\mu\times\mu, \nu\times\nu) \leq 2W_1(\mu,\nu)$ because φ was chosen arbitrarily. By induction, we conclude the lemma. ■

Lemma 4.3.3 *Let $\mu \in \mathrm{Prob}(\Sigma)$ and φ be the group action of $\mathrm{SL}(2)$ on itself $\varphi : \mathrm{SL}(2) \times \mathrm{SL}(2) \to \mathrm{SL}(2)$, $\varphi(g_1, g_2) = g_1 g_2$. Then $\mu * \mu = \varphi_*(\mu\times\mu)$ and φ is Lipschitz.*

Proof. Given a measurable set $E \subset \mathrm{SL}(2)$, by the definition of convolution of measures:

$$\begin{aligned}
\mu * \mu(E) &= \int_{\mathrm{SL}(2)\times\mathrm{SL}(2)} 1_E(g_1 g_2)\ d\mu(g_1)d\mu(g_2) \\
&= \int_{\mathrm{SL}(2)\times\mathrm{SL}_2(\mathbb{R})} 1_E(\varphi(g_1,g_2))\ d\mu(g_1)d\mu(g_2) \\
&= \int_{\mathrm{SL}(2)} 1_E(g)\ d\varphi_*(\mu\times\mu)(g) \\
&= \varphi_*(\mu\times\mu)(E).
\end{aligned}$$

Since E was arbitrary, we conclude that $\mu * \mu = \varphi_*(\mu \times \mu)$.

It remains to show that φ is Lipschitz. We consider the distance $\overline{d}((g_1g_2),(h_1h_2)) := d(g_1,h_1) + d(g_2,h_2)$ on $\mathrm{SL}(2) \times \mathrm{SL}(2)$. Hence,

$$\begin{aligned}\overline{d}((g_1g_2),(h_1h_2)) &\leq \overline{d}\left((g_1g_2),(h_1g_2)\right) + \overline{d}\left((h_1g_2),(h_1h_2)\right)\\ &\leq \|g_2\| d(g_1,h_1) + \|h_1\| d(g_2,h_2).\end{aligned}$$

Since μ has compact support, there exist a uniform constant $C > 0$ such that $\|g\| \leq C$ for all $g \in \mathrm{supp}(\mu)$. Therefore,

$$\|g_2\| d(g_1,h_1) + \|h_1\| d(g_2,h_2) \leq C[d(g_1,h_1) + d(g_2,h_2)] = C\overline{d}\left((g_1g_2),(h_1h_2)\right).$$

This proves that $\overline{d}\left((g_1g_2),(h_1h_2)\right) \leq C\overline{d}\left((g_1g_2),(h_1h_2)\right)$, so φ is Lipschitz continuous, and its Lipschitz constant depends only on the compact support Σ.

■

Lemma 4.3.4 *If $\varphi : X_1 \to X_2$ is Lipschitz with Lipschitz constant C, then the map $\mu \mapsto \varphi_*\mu$ is Lipschitz with the same Lipschitz constant.*

Proof. Consider an arbitrary function $f \in \mathrm{Lip}_1(X_2)$. Observe that

$$\frac{1}{C} W_1(\varphi_*\mu, \varphi_*\nu) = \frac{1}{C}\int_{X_2} f \; d(\varphi_*\mu - \varphi_*\nu) = \frac{1}{C}\int_{X_1} f \circ \varphi \; d(\mu - \nu).$$

Since φ has Lipschitz constant C, then $\frac{\varphi}{C} \in \mathrm{Lip}_1(X_1)$. Also the composition $\frac{1}{C} f \circ \varphi \in \mathrm{Lip}_1(X_1)$. Therefore $\frac{1}{C}\int_{X_1} f \circ \varphi \; d(\mu - \nu) \leq W_1(\mu,\nu)$ and we conclude that

$$W_1(\varphi_*\mu, \varphi_*\nu) \leq C W_1(\mu,\nu),$$

which proves the lemma. ■

Finally, by lemma 4.3.3, $\mu * \mu = \varphi_*(\mu \times \mu)$ with φ Lipschitz. By lemmas 4.3.2 and 4.3.4, the maps $\mu \mapsto \mu \times \mu$ and $\mu \times \mu \mapsto \varphi_*(\mu \times \mu)$ are also Lipschitz, therefore so is their composition. This concludes the proof.

■

5
Hölder Continuity of the Lyapunov exponents

Given a compact metric space (Σ, d), called the space of symbols, we consider the space of sequences $M = \Sigma^{\mathbb{Z}}$, endowed with the product topology. Then we can define the full shift map: $f : M \to M$ that shifts the sequence to the left, i.e $f(\{w_i\}_{i\in\mathbb{Z}}) = \{w_{i+1}\}_{i\in\mathbb{Z}}$.

Let $\text{Prob}(\Sigma)$ be the space of Borel probability measures on Σ. Given a measure $\mu \in \text{Prob}(\Sigma)$, we can define the product measure $\mu^{\mathbb{Z}}$ in M. Then, the triple $(M, \mu^{\mathbb{Z}}, f)$ is an ergodic transformation called the full (Bernoulli) shift.

Let $L^\infty(\Sigma, \text{SL}(2))$, be the space of bounded Borel measurable transformations $A : \Sigma \to \text{SL}_2(\mathbb{R})$ endowed with the uniform distance:

$$d(A, B) := \sup_{x\in\Sigma} \|A(x) - B(x)\|.$$

Given $A \in L^\infty(\Sigma, \text{SL}(2))$ and $\mu \in \text{Prob}(\Sigma)$, the corresponding random linear cocycle $F = F(A, \mu)$ is defined as follows:

$$\begin{aligned} F\colon M \times \mathbb{R}^2 &\to M \times \mathbb{R}^2 \\ (x, v) &\mapsto (f(x), \tilde{A}(x)v), \end{aligned}$$

where $\tilde{A}\colon M \to \text{SL}(2)$ satisfies $\tilde{A}(x) = x_0$, that is, $\tilde{A}$ depends only on the zeroth coordinate of $x \in M$.

Therefore, the linear cocycle F is a random transformation (a locally constant skew product over f). In this context, the maximal Lyapunov exponent of the linear cocycle is usually denoted by $\lambda_+(A)$, since the measure μ is fixed.

A very important result about the continuity of the Lyapunov exponents is E. Le Page's Theorem. It states that, locally near an irreducible cocycle $F(A, \mu)$ with $\lambda_+(A) > 0$, the Lyapunov exponent λ_+ is a Hölder continuous function of A.

A natural question is then what happens if we let the measure μ vary, can we still obtain locally Hölder continuity for the maximal Lyapunov exponent? Our goal in this chapter is to answer this questions and propose a generalized version of E. Le Page's Theorem.

In section 5.1 we introduce the Markov operator, a useful tool employed in section 5.2 to prove the existence and uniqueness of the stationary measure associated to a random linear cocycle that satisfies the hypothesis of the theorem. In section 5.3 we prove that this stationary measure depends Hölder continuously on the measure that determines the cocycle. Finally, in section 5.4 we prove a generalization of Le Page's Theorem using the results developed in the previous sections. We also show that the version of Baraviera-Duarte follows from this one and propose a new problem that may be solved using an analogous approach.

5.1 The Markov Operator

In this section we prepare the proof of our main result. We present the Markov operator and study some of its properties. The main result of this section is to show that, under some hypotheses, the Markov operator is quasi compact and simple: its spectrum (see A.0.2) admits a decomposition in disjoint closed sets $K \cup \{\lambda_0\}$ such that $\lambda_0 \in \mathbb{C}$ is a simple eigenvalue of the operator and $|\lambda| < |\lambda_0|$ for all $\lambda \in K$.

Definition 5.1.1 *Let X be a vector space. A semi-norm on X is a real valued function $p : X \to \mathbb{R}$ such that:*

a) $p(x+y) \leq p(x) + p(y), \quad x, y \in X.$

b) $p(\alpha x) = |\alpha| p(x), \quad x \in X,\ \alpha \in \mathbb{C}.$

We should observe that a semi-norm p satisfies $p(0) = 0$ and also that $0 = p(0) \leq p(-x) + p(x) = p(2x)$. Hence $p(x) \geq 0$, $\forall x \in X$. So, we conclude that a semi-norm is a norm if and only if $p(x) = 0$ implies $x = 0$.

Throughout this chapter, we denote by $\delta \colon \mathbb{PR}^d \times \mathbb{PR}^d \to [0, \infty)$ the projective distance, definition 2.3.1, given in chapter 2.

Given $\phi \in L^\infty(\mathbb{PR}^2)$ and $0 < \alpha \leq 1$, we define:

$$\|\phi\|_\infty := \sup_{\hat{p} \in \mathbb{PR}^2} |\phi(\hat{p})|$$
$$v_\alpha(\phi) := \sup_{\hat{p} \neq \hat{q}} \frac{|\phi(\hat{p}) - \phi(\hat{q})|}{\delta(\hat{p}, \hat{q})^\alpha}.$$

Proposition 5.1.1 *v_α is a semi-norm.*

Proof. We show that v_α satisfies the two properties of a semi-norm:

a)

$$\begin{aligned}
v_\alpha(\phi+\varphi) &= \sup_{\hat{p}\neq\hat{q}} \frac{|(\phi+\varphi)(\hat{p})-(\phi+\varphi)(\hat{q})|}{\delta(\hat{p},\hat{q})^\alpha} \\
&= \sup_{\hat{p}\neq\hat{q}} \frac{|\phi(\hat{p})-\phi(\hat{q})+\varphi(\hat{p})-\varphi(\hat{q})|}{\delta(\hat{q},\hat{q})^\alpha} \\
&\leq \sup_{\hat{p}\neq\hat{q}} \left(\frac{|\phi(\hat{p})-\phi(\hat{q})|}{\delta(\hat{p},\hat{q})^\alpha} + \frac{|\varphi(\hat{p})-\varphi(\hat{q})|}{\delta(\hat{p},\hat{q})^\alpha} \right) \\
&\leq v_\alpha(\phi)+v_\alpha(\varphi).
\end{aligned}$$

b)

$$\begin{aligned}
v_\alpha(\beta\phi) &= \sup_{\hat{p}\neq\hat{q}} \frac{|\beta\phi(\hat{p})-\beta\phi(\hat{q})|}{\delta(\hat{p},\hat{q})^\alpha} \\
&= |\beta| v_\alpha(\phi).
\end{aligned}$$

■

Remark 14 v_α *is not a norm.*

Proof. Consider a constant function $\phi=\lambda$, with $\lambda\in\mathbb{C}$. Then

$$v_\alpha(\phi) = \sup_{\hat{p}\neq\hat{q}} \frac{|\lambda-\lambda|}{\delta(\hat{p},\hat{q})^\alpha} = 0.$$

So $v_\alpha(\phi)=0$ does not imply $\phi=0$, that's why v_α is not a norm. ■

Definition 5.1.2 *The space of $\alpha-$Hölder continuous functions on $\mathbb{PR}^2$ is given by*

$$\mathcal{H}_\alpha(\mathbb{PR}^2) := \{\phi\in L^\infty(\mathbb{PR}^2) : \|\phi\|_\alpha<\infty\},$$

where the norm $\|\phi\|_\alpha$ is given by

$$\|\phi\|_\alpha = \|\phi\|_\infty + v_\alpha(\phi).$$

From now on, we fix a compact set $\Sigma\subset \mathrm{SL}(2)$.

Definition 5.1.3 *Let $\mu\in\mathrm{Prob}(\Sigma)$. We define the average Hölder constant of the projective action $\hat{g}(\hat{x}) : \mathbb{PR}^2\to\mathbb{PR}^2$ by*

$$k_\alpha(\mu) := \sup_{\hat{p}\neq\hat{q}} \int_\Sigma \left(\frac{\delta(\hat{g}\hat{p},\hat{g}\hat{q})}{\delta(\hat{p},\hat{q})}\right)^\alpha d\mu(g).$$

Proposition 5.1.2 *The sequence $k_\alpha(\mu^{*n})$ is sub-multiplicative:*

$$k_\alpha(\mu^{*(n+m)}) \leq k_\alpha(\mu^{*n})k_\alpha(\mu^{*m}).$$

Proof.

$$
\begin{aligned}
k_\alpha(\mu^{*(n+m)}) &= \sup_{\hat{p}\neq\hat{q}} \int_\Sigma \left(\frac{\delta(\hat{g}\hat{p},\hat{g}\hat{q})}{\delta(\hat{p},\hat{q})}\right)^\alpha d\mu^{*(n+m)}(g) \\
&= \sup_{\hat{p}\neq\hat{q}} \int_{\Sigma^{n+m}} \left(\frac{\delta(\hat{g}_{n+m-1}...\hat{g}_0(\hat{p}),\hat{g}_{n+m-1}...\hat{g}_0(\hat{q}))}{\delta(\hat{p},\hat{q})}\right)^\alpha d\mu(g_{n+m-1})...d\mu(g_0) \\
&= \sup_{\hat{p}\neq\hat{q}} \int_{\Sigma^{n+m}} \left(\frac{\delta(\hat{g}_{n+m-1}...\hat{g}_0(\hat{p}),\hat{g}_{n+m-1}...\hat{g}_0(\hat{q}))}{\delta(\hat{g}_{n-1}...\hat{g}_0(\hat{p}),\hat{g}_{n-1}...\hat{g}_0(\hat{q}))}\right)^\alpha \times \\
&\qquad \left(\frac{\delta(\hat{g}_{n-1}...\hat{g}_0(\hat{p}),\hat{g}_{n-1}...\hat{g}_0(\hat{q}))}{\delta(\hat{p},\hat{q})}\right)^\alpha d\mu(g_{n+m-1})...d\mu(g_0) \\
&\leq \sup_{\hat{p}\neq\hat{q}} \int_{\Sigma^{m}} \left(\frac{\delta(\hat{g}_{n+m-1}...\hat{g}_{n-1}...\hat{g}_0(\hat{p}),\hat{g}_{n+m-1}...\hat{g}_{n-1}...\hat{g}_0(\hat{q}))}{\delta(\hat{g}_{n-1}...\hat{g}_0(\hat{p}),\hat{g}_{n-1}...\hat{g}_0(\hat{q}))}\right)^\alpha d\mu(g_{n+m-1})...d\mu(g_n) \\
&\times \sup_{\hat{p'}\neq\hat{q'}} \int_{\Sigma^{n}} \left(\frac{\delta(\hat{g}_{n-1}...\hat{g}_0(\hat{p'}),\hat{g}_{n-1}...\hat{g}_0(\hat{q'}))}{\delta(\hat{p'},\hat{q'})}\right)^\alpha d\mu(g_{n-1})...d\mu(g_0) \\
&= k_\alpha(\mu^{*m})k_\alpha(\mu^{*n}).
\end{aligned}
$$

■

Definition 5.1.4 *Given a measure $\mu \in \mathrm{Prob}(\Sigma)$, we define its Markov operator $Q_\mu : L^\infty(\mathbb{PR}^2) \to L^\infty(\mathbb{PR}^2)$ by*

$$Q_\mu(\phi)(\hat{p}) := \int_\Sigma \phi(\hat{g}\hat{p})\, d\mu(g),$$

where $\hat{g} : \mathbb{PR}^2 \to \mathbb{PR}^2$ is the projective action of g.

A remark about the previous definition is that we can define the Markov operator for functions from smaller spaces, for example from $\mathcal{H}_\alpha(\mathbb{PR}^2)$ to itself, in the same way that we did before.

Lemma 5.1.3 *The Markov operator $Q_\mu : L^\infty(\mathbb{PR}^2) \to L^\infty(\mathbb{PR}^2)$ associated to a measure $\mu \in \mathrm{Prob}(\Sigma)$ is bounded and $\|Q_\mu\|_\infty \leq 1$.*

Proof. By a direct calculation:

$$
\begin{aligned}
\|Q_\mu(\varphi)\|_\infty &= \sup_{\hat{p}\in\mathbb{PR}^2} |Q_\mu(\varphi)\hat{p}| \\
&= \sup_{\hat{p}\in\mathbb{PR}^2} \left|\int_\Sigma \varphi(\hat{g}\hat{p})\, d\mu(g)\right| \\
&\leq \|\varphi\|_\infty \mu(\Sigma) \\
&\leq \|\varphi\|_\infty
\end{aligned}
$$

Hence $\|Q_\mu\|_\infty \leq 1$. ■

If we consider the Markov operator $Q_\mu : \mathcal{H}_\alpha(\mathbb{PR}^2) \to \mathcal{H}_\alpha(\mathbb{PR}^2)$ associated to a measure $\mu \in \mathrm{Prob}(\Sigma)$, we also obtain that Q_μ is bounded:

$$\begin{aligned}
\|Q_\mu(\varphi)\|_\alpha &= \sup_{\hat{p}\in\mathbb{PR}^2} |Q_\mu(\varphi)\hat{p}| + \sup_{\hat{p}\neq\hat{q}} \left| \frac{Q_\mu(\varphi)\hat{p} - Q_\mu(\varphi)\hat{q}}{\delta(\hat{p},\hat{q})^\alpha} \right| \\
&= \sup_{\hat{p}\in\mathbb{PR}^2} \left| \int_\Sigma \varphi(\hat{g}\hat{p})\, d\mu(g) \right| + \sup_{\hat{p}\neq\hat{q}} \left| \frac{\int_\Sigma \varphi(\hat{g}\hat{p})\, d\mu(g) - \int_\Sigma \varphi(\hat{g}\hat{q})\, d\mu(g)}{\delta(\hat{p},\hat{q})^\alpha} \right| \\
&\leq \|\varphi\|_\alpha + k_\alpha(\mu) v_\alpha(\varphi).
\end{aligned}$$

Hence $\|Q_\mu\|_\alpha \leq \|\varphi\|_\alpha + k_\alpha(\mu) v_\alpha(\varphi)$.

Proposition 5.1.4 *For all $\phi \in \mathcal{H}_\alpha(\mathbb{PR}^2)$,*

$$v_\alpha\left(Q_\mu(\phi)\right) \leq k_\alpha(\mu) v_\alpha(\phi).$$

Proof. Given $\phi \in \mathcal{H}_\alpha(\mathbb{PR}^2)$ and $\hat{p}, \hat{q} \in \mathbb{P}(\mathbb{R}^2)$,

$$\begin{aligned}
\frac{|Q_\mu(\phi)(\hat{p}) - Q_\mu(\phi)(\hat{q})|}{\delta(\hat{p},\hat{q})^\alpha} &= \frac{|\int_\Sigma \phi(\hat{g}\hat{p})\, d\mu(g) - \int_\Sigma \phi(\hat{g}\hat{q})\, d\mu(g)|}{\delta(\hat{p},\hat{q})^\alpha} \\
&= \left| \int_\Sigma \frac{\phi(\hat{g}\hat{p}) - \phi(\hat{g}\hat{q})}{\delta(\hat{p},\hat{q})^\alpha}\, d\mu(g) \right| \\
&\leq \int_\Sigma \left| \frac{\phi(\hat{g}\hat{p}) - \phi(\hat{g}\hat{q})}{\delta(\hat{p},\hat{q})^\alpha} \right| d\mu(g) \\
&\leq \int_\Sigma \left| \frac{\phi(\hat{g}\hat{p}) - \phi(\hat{g}\hat{q})}{\delta(\hat{g}\hat{p},\hat{g}\hat{q})^\alpha} \right| \cdot \left| \frac{\delta(\hat{g}\hat{p},\hat{g}\hat{q})^\alpha}{\delta(\hat{p},\hat{q})^\alpha} \right| d\mu(g) \\
&\leq v_\alpha(\phi) \int_\Sigma \left| \frac{\delta(\hat{g}\hat{p},\hat{g}\hat{q})^\alpha}{\delta(\hat{p},\hat{q})^\alpha} \right| d\mu(g).
\end{aligned}$$

Now, applying $\sup_{\hat{p}\neq\hat{q}}$ to both sides we conclude that

$$v_\alpha\left(Q_\mu(\phi)\right) \leq k_\alpha(\mu) v_\alpha(\phi).$$

■

Proposition 5.1.5 *For all $n \in \mathbb{N}$:*

$$(Q_\mu)^n = Q_{\mu^{*n}}.$$

Proof. We proceed by induction. Let $\phi \in L^\infty(\mathbb{PR}^2)$ and $\hat{p} \in \mathbb{PR}^2$. The case

$n = 1$ is trivial. For $n = 2$ we have:

$$\begin{aligned}(Q_\mu)^2(\phi)(\hat{p}) &= \int_\Sigma \int_\Sigma \phi(\hat{g}_1\hat{g}_0\hat{p})d\mu(g_1)d\mu(g_0)\\ &= \int_\Sigma \phi(\hat{g}\hat{p})d(\mu * \mu)(g)\\ &= \int_\Sigma \phi(\hat{g}\hat{p})d(\mu^{*2})(g)\\ &= (Q_{\mu^{*2}})(\phi)(\hat{p}).\end{aligned}$$

Now suppose that our statment is true for every $k \leq n-1$. We are going to show that it also holds for the case $k = n$.

$$\begin{aligned}(Q_\mu)^n(\phi)(\hat{p}) &= \int_\Sigma \cdots \int_\Sigma \phi(\hat{g_{n-1}}\hat{g_{n-2}}...\hat{g}_0\hat{p})\ d\mu(g_{n-1})d\mu(g_{n-2})...d\mu(g_0)\\ &= \int_\Sigma \phi(\hat{g}\hat{p})\ d(\mu * ... * \mu)(g) \qquad \text{(n-th convolution of } \mu \text{ with itself)}\\ &= \int_\Sigma \phi(\hat{g}\hat{p})\ d(\mu^{*n})(g)\\ &= (Q_{\mu^{*n}})(\phi)(\hat{p}).\end{aligned}$$

■

In this section we will need to generalize our notion of stationary measures. In the previous chapter we defined a stationary measure relative to a cocycle F. Since every measure μ defines a cocycle, it is natural to extend the previous concept to a stationary measure η relative to some measure μ.

Definition 5.1.5 *Given a group G and a measure $\mu \in \text{Prob}(G)$, with compact support Σ, we say that a measure η is stationary with respect to μ if it satisfies:*

$$\eta(B) = \int_\Sigma \eta(\hat{g}^{-1}(B))\ d\mu(g)$$

for every measurable set $B \subset \mathbb{PR}^2$, where $\hat{g}$ is the projective action of g.

Proposition 5.1.6 *Let $Q_\mu : C(\mathbb{PR}^2) \to C(\mathbb{PR}^2)$ be the Markov operator. A measure η is stationary with respect to μ if and only if for every continuous function $\varphi \in C(\mathbb{PR}^2)$,*

$$\int_{\mathbb{PR}^2} Q_\mu(\varphi)(\hat{p})\ d\eta(p) = \int_{\mathbb{PR}^2} \varphi(\hat{p})\ d\eta(p).$$

Proof. Let η be a stationary measure with respect to $\mu \in \text{Prob}(\text{SL}(2))$, such that $\text{supp}(\mu) \subset \Sigma$. Then, for an arbitrary measurable set $B \subset \mathbb{PR}^2$, consider its indicator function 1_B,

$$\begin{aligned}\int_{\mathbb{P}\mathbb{R}^2} Q_\mu(1_B)(\hat{p})\, d\eta(p) &= \int_{\mathbb{P}\mathbb{R}^2}\int_\Sigma 1_B(\hat{g})(\hat{p})\, d\mu(g)d\eta(p)\\ &= \int_\Sigma \int_{g^{-1}(B)} 1(\hat{p})\, d\eta(p)d\mu(g)\\ &= \int_\Sigma \eta(g^{-1}(B)d\mu(g)\\ &= \eta(B)\\ &= \int_{\mathbb{P}\mathbb{R}^2} 1_B(\hat{p})\, d\eta(p).\end{aligned}$$

Since the statement is true for indicator functions it is also true for simple functions by linearity. Hence it is also true for every positive measurable function, by the monotone convergence theorem. Finally, by writing $\varphi = \varphi^+ - \varphi^-$, where φ^+ and φ^- are the positive and negative parts of φ, the property holds for every function in $L^1(\mathbb{P}\mathbb{R}^2)$. Since $C(\mathbb{P}\mathbb{R}^2) \subset L^1(\mathbb{P}\mathbb{R}^2)$, we conclude the first part of the proof.

Conversely, assume that for every continuous function $\varphi \in C(\mathbb{P}\mathbb{R}^2)$, it holds that $\int_{\mathbb{P}\mathbb{R}^2} Q_\mu(\varphi)(\hat{p})\, d\eta(p) = \int_{\mathbb{P}\mathbb{R}^2} \varphi(\hat{p})\, d\eta(p)$. Then, by Lusin's Theorem, for every $\varepsilon > 0$ there exists a closed set F with measure $1-\varepsilon$ and a continuous function φ such that $\varphi = 1_B$ in F. Therefore,

$$\begin{aligned}\eta(B) = \int_{\mathbb{P}\mathbb{R}^2} 1_B(\hat{p})\, d\eta(p) &= \int_F 1_B(\hat{p})\, d\eta(p) + \int_{F^c} 1_B(\hat{p})\, d\eta(p)\\ &= \int_F 1_B(\hat{p})\, d\eta(p) + \int_{F^c} 1_B(\hat{p})\, d\eta(p)\\ &= \int_F Q_\mu(1_B)(\hat{p})d\eta(p) + \int_{F^c} 1_B(\hat{p})\, d\eta(p)\\ &= \int_F \int_\Sigma 1_B(\hat{g})(\hat{p})\, d\mu(g)d\eta(p) + \int_{F^c} 1_B(\hat{p})\, d\eta(p)\\ &= \int_\Sigma \int_{(g^{-1}(B))\cap F} 1(\hat{p})\, d\eta(p)d\mu(g) + \int_{F^c} 1_B(\hat{p})\, d\eta(p)\\ &= \int_\Sigma \eta(g^{-1}(B)\cap F)d\mu(g) + \int_{F^c} 1_B(\hat{p})\, d\eta(p).\end{aligned}$$

Then,

$$\begin{aligned}\left|\eta(B) - \int_\Sigma \eta(g^{-1}(B))d\mu(g)\right| &\le \left|\eta(B) - \int_\Sigma \eta(g^{-1}(B)\cap F)d\mu(g)\right| +\\ &+ \left|\int_\Sigma \eta(g^{-1}(B)\cap F)d\mu(g) - \int_\Sigma \eta(g^{-1}(B))d\mu(g)\right|\\ &\le 2\varepsilon.\end{aligned}$$

Since ε was arbitrary we conclude that η is stationary with respect to μ.

■

From now on, we write **1** in bold to represent the constant function 1.

Proposition 5.1.7 *Let $\mu \in \text{Prob}(\Sigma)$ be a measure such that for some $0 < \alpha < 1$ and $n > 1$:*

$$k_\alpha(\mu^{*n})^{\frac{1}{n}} \leq \sigma < 1.$$

Then the Markov operator $Q_\mu : \mathcal{H}_\alpha(\mathbb{PR}^2) \to \mathcal{H}_\alpha(\mathbb{PR}^2)$ is quasi-compact and simple. More precisely, there exists a unique stationary measure $\nu \in \text{Prob}(\mathbb{PR}^2)$ with respect to the cocycle determined by μ such that defining the subspace

$$\mathcal{N}_\alpha(\nu) := \left\{\varphi \in \mathcal{H}_\alpha(\mathbb{PR}^2) : \int_{\mathbb{PR}^2} \varphi \, d\nu = 0\right\}$$

the operator Q_μ has the following properties:

1. *spec*$(Q_\mu : \mathcal{H}_\alpha(\mathbb{PR}^2) \to \mathcal{H}_\alpha(\mathbb{PR}^2)) \subset \{1\} \cup \overline{\mathcal{D}_\sigma(0)}$,
2. $\mathcal{H}_\alpha(\mathbb{PR}^2) = \mathbb{C}\mathbf{1} \oplus \mathcal{N}_\alpha(\nu)$ *is a Q_μ-invariant decomposition,*
3. *Q_μ fixes every function in $\mathbb{C}\mathbf{1}$ and acts as a contraction with spectral radius less than or equal to σ on $\mathcal{N}_\alpha(\nu)$.*

Proof. By proposition 5.1.1 and remark 14, v_α is a semi-norm for the space $\mathcal{H}_\alpha(\mathbb{PR}^2)$ but not a norm. However, it induces a norm on the quotient space $\mathcal{H}_\alpha(\mathbb{PR}^2)/\mathbb{C}\mathbf{1}$. Note that we do not have the problem that we had in remark 14, because we are identifying the constant functions. Also, if $\phi \neq 0$ and ϕ is not a constant, there exist $\hat{p}$ and $\hat{q} \in \mathbb{PR}^2$ such that $\phi(\hat{p}) \neq \phi(\hat{q})$, hence $v_\alpha(\phi) > 0$. So it is, in fact, a norm on the space $\mathcal{H}_\alpha(\mathbb{PR}^2)/\mathbb{C}\mathbf{1}$.

By hypothesis, there exist some $0 < \alpha < 1$ and $n > 1$ such that $k_\alpha(\mu^{*n}) \leq \sigma^n < 1$. So, by propositions 5.1.4 and 5.1.5, Q_μ^n acts on $\mathcal{H}_\alpha(\mathbb{PR}^2)/\mathbb{C}\mathbf{1}$ as a σ^n contraction:

$$v_\alpha\left(Q_\mu^n(\phi)\right) = v_\alpha(Q_{\mu^{*n}}) \leq k_\alpha(\mu^{*n}) v_\alpha(\phi) \leq \sigma^n v_\alpha(\phi), \tag{5.1}$$

since v_α is a norm for $\mathcal{H}_\alpha(\mathbb{PR}^2)/\mathbb{C}\mathbf{1}$.

Also, the relation (5.1) means that the norm of the operator Q_μ^n is bounded by $\sigma^n < 1$. Therefore, by corollary $A.0.2$, the spectrum of Q_μ^n on $\mathcal{H}_\alpha(\mathbb{PR}^2)/\mathbb{C}\mathbf{1}$ is contained in the closed disc of radius σ^n. Since Q also fixes the constant functions in $\mathbb{C}\mathbf{1}$, it is a quasi-compact operator with a simple eigenvalue 1 (associated to the eigen-space $\mathbb{C}\mathbf{1}$). Thus, *spec*$(Q_\mu \colon \mathcal{H}_\alpha(\mathbb{PR}^2) \to \mathcal{H}_\alpha(\mathbb{PR}^2)) \subset 1 \cup \overline{D_\sigma(0)}$.

By the previous conclusion and $(A.0.4)$ there exists a Q_μ-invariant decomposition $\mathcal{H}_\alpha(\mathbb{PR}^2) = \mathbb{C}\mathbf{1} \oplus \mathcal{N}_\alpha$ such that Q_μ acts as a contraction with spectral radius $\leq \sigma$ on $\mathcal{N}_\alpha$. Now it remains to show that there exists a unique stationary measure $\nu \in \text{Prob}(\mathbb{PR}^2)$ with respect to the cocycle determined by μ such that $\mathcal{N}_\alpha = \mathcal{N}_\alpha(\nu)$.

We start by defining a linear functional $\Lambda : \mathcal{H}_\alpha(\mathbb{PR}^2) \to \mathbb{C}$ setting $\Lambda(c1+\psi) := c$ for $\psi \in \mathcal{N}_\alpha$. Our goal will be to extend this functional and apply Riesz-Markov Theorem A.0.5 to show that $\mathcal{N}_\alpha = \mathcal{N}_\alpha(\nu)$ and then conclude that the measure ν is, in fact, a stationary measure with respect to μ. First, let us check some properties of Λ:

- $\Lambda(1) = 1$ by its definition.
- Λ is linear. Given $c_1, c_2, \beta \in \mathbb{C}$ and $\psi_1, \psi_2 \in \mathcal{N}_\alpha$,

$$\Lambda(c_1 + \psi_1 + c_2 + \psi_2) = c_1 + c_2 = \Lambda(c_1) + \Lambda(c_2),$$
$$\Lambda(\beta(c + \psi)) = \Lambda(\beta c + \beta\psi) = \beta c = \beta\Lambda(c).$$

- Λ is positive. Let $\varphi = c\mathbf{1} + \psi \geq 0$ with $\psi \in \mathcal{N}_\alpha$. Since Q_μ is positive (the integral of a positive function is positive),

$$Q_\mu^n(0) = 0 \leq Q_\mu^n(c1 + \psi) = Q_\mu^n(c) + Q_\mu^n(\psi) = c + Q_\mu^n(\psi) \quad \forall n \geq 0.$$

  Moreover, $\psi \in \mathcal{N}_\alpha$, where Q_μ^n acts as a contraction, so we must have $\lim\limits_{n\to\infty} Q_\mu^n(\psi) = 0$. This fact, together with the the previous result that $c + Q_\mu^n(\psi) \geq 0$ for every $n \geq 0$, imply that $c \geq 0$. Since $\Lambda(\varphi) = c$, we conclude that Λ is a positive operator.
- Λ is continuous with respect to the norm $\|.\|_\infty$. For every $\varphi \in \mathcal{H}_\alpha(\mathbb{PR}^2)$,

$$-\|\varphi\|_\infty \mathbf{1} \leq \varphi \leq \|\varphi\|_\infty \mathbf{1}.$$

  Since Λ is positive:

$$|\Lambda(\varphi)| \leq \Lambda(\|\varphi\|_\infty \mathbf{1}) = \|\varphi\|_\infty.$$

- Λ has an extension $\overline{\Lambda} : C(\mathbb{PR}^2) \to \mathbb{C}$ that is positive, continuous and linear by proposition $A.0.2$.

Therefore, we are in the conditions to apply Riesz-Markov Theorem A.0.5 to $\overline{\Lambda}$, so there exists a unique probability measure $\nu \in \text{Prob}(\mathbb{PR}^2)$ such that

$$\overline{\Lambda}(\varphi) = \int_{\mathbb{PR}^2} \varphi \, d\nu \quad \forall \varphi \in C(\mathbb{PR}^2).$$

Moreover, since $\mathcal{N}_\alpha$ is the kernel of Λ, we have that $\mathcal{N}_\alpha = \mathcal{N}_\alpha(\nu)$. In order to conclude this proposition we are going to observe that ν is stationary with respect to μ. By proposition 5.1.6, it is sufficient to show that $\forall \varphi \in C(\mathbb{PR}^2)$,

$$\int_{\mathbb{PR}^2} Q_\mu(\varphi) \, d\nu = \int_{\mathbb{PR}^2} \varphi \, d\nu.$$

Let $\varphi \in C(\mathbb{P}\mathbb{R}^2)$ and decompose it as:

$$\varphi = \left(\mathbf{1}\int_{\mathbb{P}\mathbb{R}^2} \varphi \, d\nu\right) + \left(\varphi - \mathbf{1}\int_{\mathbb{P}\mathbb{R}^2} \varphi \, d\nu\right). \tag{5.2}$$

Now apply the Markov operator Q_μ on both sides of (5.2). Observe that Q_μ is linear and

$$Q_\mu\left(\mathbf{1}\int_{\mathbb{P}\mathbb{R}^2} \varphi \, d\nu\right) = \mathbf{1}\int_{\mathbb{P}\mathbb{R}^2} \varphi \, d\nu,$$

since it fixes constant functions. Then we arrive at

$$Q_\mu(\varphi) = \mathbf{1}\int_{\mathbb{P}\mathbb{R}^2} \varphi \, d\nu + Q_\mu\left(\varphi - \mathbf{1}\int_{\mathbb{P}\mathbb{R}^2} \varphi \, d\nu\right).$$

Note that $(\varphi - \mathbf{1}\int_{\mathbb{P}\mathbb{R}^2} \varphi \, d\nu) \in \mathcal{N}_\alpha(\nu)$. Hence, since the decomposition $\mathcal{H}_\alpha(\mathbb{P}\mathbb{R}^2) = \mathbb{C}\mathbf{1} \oplus \mathcal{N}_\alpha(\nu)$ is Q_μ-invariant, $Q_\mu\left(\varphi - \mathbf{1}\int_{\mathbb{P}\mathbb{R}^2} \varphi \, d\nu\right) \in \mathcal{N}_\alpha(\nu)$. Thus, integrating both sides, we conclude the proposition.

■

5.2 Existence and uniqueness of the stationary measure

In this section we prove that under some hypotheses on the measure μ which generates the linear random cocycle, the conditions of proposition 5.1.7 are met. Therefore, there exists a unique stationary measure η_μ associated to the cocycle generated by μ.

Lemma 5.2.1 *Given $g \in \mathrm{SL}_2(\mathbb{R})$ with $\|g\| > 1$, $\hat{p} \in \mathbb{P}\mathbb{R}^2$ and $\hat{u}(g)$ the most expanding direction of g, we have*

$$\alpha := \cos\left(p, u(g)\right) \leq \frac{\|gp\|}{\|g\|}.$$

Proof. Choose p and u unitary representatives of $\hat{p}$ and $\hat{u}(g)$ such that $\angle(p, u)$ is not obtuse. We can write $p = \alpha u + w$, with $w \perp u$. Thus, by linearity, $gp = \alpha g u + g w$. Now observe that

$$\begin{aligned} \alpha^2 \|g\|^2 &\leq \alpha^2 \|g\|^2 + \|gw\|^2 \\ &= \alpha^2 \|gu\|^2 + \|gw\|^2 \\ &= \|gp\|^2. \end{aligned}$$

The last line comes from Pythagora's theorem, since $\|gu\|$ and $\|gw\|$ are perpendicular: $\langle gw, gu\rangle = \langle w, g^* g u\rangle = \lambda\langle w, u\rangle = 0$ because u is an eigenvector of $g^* g$. Therefore,

$$\alpha \leq \frac{\|gp\|}{\|g\|}.$$

■

Next we introduce the main condition on the measure μ which will guarantee the Hölder continuity of the Lyapunov exponent. This condition, quasi irreducibility, is weaker than the irreducibility condition previously defined, because it allows the existence of an invariant line, as long as the *maximal* Lyapunov exponent λ_+ is reached along that line.

Given $\mu \in \mathrm{Prob}(\Sigma)$, where $\Sigma \in \mathrm{SL}(2)$ is compact, let $\{g_n\}_{n\geq 0}$ be an i.i.d. multiplicative process with law μ. Recall that by the Oseledets Theorem, for every $v \in \mathbb{R}^2\backslash\{0\}$ and μ-almost everywhere,

$$\lim_{n\to\infty} \frac{1}{n} \log \|g_{n-1}\dots g_1 g_0 v\|$$

exists and equals $\lambda_+(\mu)$ or $\lambda_-(\mu)$. If $l \subset \mathbb{R}^2$ is an invariant line, that is, if $gl = l$ for μ-almost every $g \in \mathrm{SL}(2)$, then it is easy to see (by Birkhoff's Ergodic Theorem) that for a vector $v \in l$, $v \neq 0$, we have μ-almost everywhere,

$$\lim_{n\to\infty} \frac{1}{n} \log \|g_{n-1}\dots g_1 g_0 v\| = \int_\Sigma \log \frac{\|gv\|}{\|v\|}\, d\mu(g).$$

By abuse of notation and for convenience we denote it by $\lambda_+(\mu\big|_l)$. We are now ready to formally define quasi irreducibility.

Definition 5.2.1 *A measure $\mu \in \mathrm{Prob}(\Sigma)$ is called quasi irreducible if there is no (invariant) line $l \in \mathbb{R}^2$ such that $gl = l$ for $\mu-$almost every $g \in \mathrm{SL}(2)$ and $\lambda_+(\mu\big|_l) < \lambda_+(\mu)$.*

Therefore, if μ is quasi irreducible, either it admits no invariant lines or else, if $l \subset \mathbb{R}^2$ is invariant, then $\lambda_+(\mu\big|_l) = \lambda_+(\mu)$.

Proposition 5.2.2 *Let $\mu \in \mathrm{Prob}(\Sigma)$ be a quasi irreducible measure with $\lambda_+(\mu) > 0$. Then*

$$\lim_{n\to\infty} \frac{1}{n} \int_\Sigma [\log \|gp\|]\, d\mu^{*n}(g) = \lambda_+(\mu),$$

with uniform convergence in $p \in \mathbb{S}^1 = \{v \in \mathbb{R}^2 \colon \|v\| = 1\}$.

Proof. We split this proof into two steps. First, we prove the pointwise convergence, and then the fact that the convergence is uniform in $p \in \mathbb{S}^1$. Let $F \subset \Sigma^{\mathbb{Z}}$ be a $T-$invariant set of full measure, consisting of Oseledets regular points. For every $\{g_n\}_n \in F$ we have the Oseledets decomposition: $\mathbb{R}^2 = E^+(\{g_n\}_n) \oplus E^-(\{g_n\}_n)$ which is invariant under the cocycle action.

Moreover, given $\{g_n\}_n \in F$ and a unit vector $p \in \mathbb{R}^2$, either $p \in E^-(\{g_n\}_n)$ or

$$\lim_{n\to\infty} \frac{1}{n} \log \|g_{n-1}...g_0 p\| = \lambda_+(\mu).$$

Now, consider the linear subspace

$$S := \left\{ p \in \mathbb{R}^2 : p \in E^-(\{g_n\}_n), \text{ for } \mu^{\mathbb{Z}}\text{almost every } \{g_n\}_n \right\}.$$

Since $g_0 E^{\pm}(\{g_n\}_n) = E^{\pm}(T(\{g_n\}_n))$, for all $\{g_n\}_n \in F$, it follows that $gS = S$ for $\mu-$almost every $g \in \Sigma$. We also know that $\lambda_+(\mu) > 0$, hence $\dim S \leq 1$. Suppose that $\dim S = 1$. Since the measure is quasi irreducible and it satisfies $gS = S$ for μ-almost every $g \in \Sigma$, we cannot have $\lambda_+(\mu\big|_S) < \lambda_+(\mu)$, because, by the definition of S, $\lambda_+(\mu\big|_S) = \lambda_-(\mu) < \lambda_+(\mu)$. Therefore $S = \{0\}$.

This implies that given any unit vector $p \in \mathbb{R}^2$,

$$\lim_{n\to\infty} \frac{1}{n} \log \|g_{n-1}...g_0 p\| = \lambda_+(\mu), \quad \text{for } \mu - \text{almost every } \{g_n\}_n.$$

Observe that since Σ is compact, there exists $C > 0$ such that $\|g\| < C$ for all $g \in \Sigma$. Then

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \log \|g_{n-1}...g_0 p\| &\leq \lim_{n\to\infty} \frac{1}{n} \log \|g_{n-1}\|...\|g_0\|\|p\| \\
&\leq \lim_{n\to\infty} \frac{1}{n} \log C^n \\
&\leq C.
\end{aligned}$$

Now, we are able to apply the dominated convergence theorem:

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \int_{\Sigma} [\log \|gp\|] \; d\mu^{*n} &= \lim_{n\to\infty} \frac{1}{n} \int_{\Sigma^n} \log \|g_{n-1}...g_0 p\| d\mu(g_{n-1})...d\mu(g_0) \\
&= \int_{\Sigma^n} \lim_{n\to\infty} \frac{1}{n} \log \|g_{n-1}...g_0 p\| d\mu(g_{n-1})...d\mu(g_0) \\
&= \lambda_+(\mu).
\end{aligned}$$

This shows the pointwise convergence and concludes the first part of the proof. Now it remains to prove that this convergence is uniform. Suppose it is not uniformly convergent in $p \in \mathbb{S}^1$. Then, there exists a sequence of unitary vectors $\{p_n\}_n \in \mathbb{R}^2$ and $\delta > 0$ such that for every large n,

$$\frac{1}{n} \int_{\Sigma} [\log \|gp_n\|] \; d\mu^{*n} \leq \lambda_+(\mu) - \delta.$$

By the compactness of the unit circle, there exists a subsequence $\{p_{nk}\}_k$

that converges to a unit vector $p \in \mathbb{R}^2$. We claim that $\frac{1}{n_k}\int_\Sigma [\log \|gp_{n_k}\|] \ d\mu^{*n_k}$ converges to $\lambda_+(\mu)$, which contradicts the previous assumption. Note that by lemma 5.2.1,

$$\frac{\|g_{n_k-1}..g_{n_0}p_{n_k}\|}{\|g_{n_k-1}...g_{n_0}\|} \geq \cos\left(p_{n_k}, u(g_{n_k-1}...g_{n_0})\right) = |p_{n_k} \cdot u(g_{n_k-1}...g_{n_0})|,$$

where $u(g_{n_k-1}...g_{n_0})$ is the most expanded unit vector by $g_{n_k-1}...g_{n_0}$, as we saw in chapter 2. Moreover, we know already by proposition 2.4.2 that μ-almost everywhere,

$$|p_{n_k} \cdot u(g_{n_k-1}...g_{n_0})| \to |p \cdot u(\mu)|,$$

where $u(\mu) = \lim_{k\to\infty} u(g_{n_k-1}...g_{n_0})$.

Suppose $|p \cdot u(\mu)| = 0$, so $p \in u(\mu)^\perp$. By proposition 2.4.4, $u(\mu)^\perp = E^s(\mu)$. Since $S = \{0\}$,

$$\liminf_{k\to\infty} \frac{\|g_{n_k-1}...g_{n_0}p_{n_k}\|}{\|g_{n_k-1}...g_{n_0}\|} > 0 \quad \mu\text{-almost everywhere.}$$

Therefore, $\frac{1}{n_k}\log\frac{\|g_{n_k-1}...g_{n_0}p_{n_k}\|}{\|g_{n_k-1}...g_{n_0}\|}$ converges to zero $\mu-$almost everywhere. Then, using again the dominated convergence theorem:

$$\begin{aligned}\lim_{k\to\infty}\frac{1}{n_k}\int_\Sigma [\log\|gp_{n_k}\|] \ d\mu^{*n_k}(g) &= \lim_{k\to\infty}\frac{1}{n_k}\int_\Sigma [\log\|g\|] \ d\mu^{*n_k}(g) \\ &+ \lim_{k\to\infty}\frac{1}{n_k}\int_\Sigma \left[\log\frac{\|gp_{n_k}\|}{\|g\|}\right] \ d\mu^{*n_k}(g) \\ &= \lambda_+(\mu).\end{aligned}$$

This proves the claim and concludes the proof. ■

Proposition 5.2.3 *Given $\alpha > 0$ and unit vectors $p, q \in \mathbb{R}^2$,*

$$\left[\frac{\delta(\hat{g}\hat{p}, \hat{g}\hat{q})}{\delta(\hat{p},\hat{q})}\right]^\alpha \leq \frac{1}{2}\left\{\frac{1}{\|gp\|^{2\alpha}} + \frac{1}{\|gq\|^{2\alpha}}\right\}.$$

Proof. Since $g \in \mathrm{SL}_2(\mathbb{R})$, it preserves area, so $\|gp \wedge gq\| = \|p \wedge q\|$. Then,

$$\begin{aligned}\left[\frac{\delta(\hat{g}\hat{p}, \hat{g}\hat{q})}{\delta(\hat{p},\hat{q})}\right]^\alpha &= \left[\frac{\|gp\wedge gq\|}{\|gp\|\|gq\|}\frac{\|p\|\|q\|}{\|p\wedge q\|}\right]^\alpha \\ &= \frac{1}{\|gp\|^\alpha}\frac{1}{\|gq\|^\alpha} \\ &\leq \frac{1}{2}\left\{\frac{1}{\|gp\|^{2\alpha}} + \frac{1}{\|gq\|^{2\alpha}}\right\}.\end{aligned}$$

In the last inequality we used the fact that the geometric mean is less than or equal to the arithmetic mean. ■

Proposition 5.2.4 *Given $g \in \mathrm{SL}(2)$, a unit vector $p \in \mathbb{R}^2$ and $v \in T_p\mathbb{R}^2$, the map $\varphi_g : \mathbb{S}^1 \subset \mathbb{R}^2 \to \mathbb{R}^2$ that takes x to $\frac{g(x)}{\|g(x)\|}$ has derivative*

$$D(\varphi_g)_p(v) = \frac{1}{\|gp\|^2}.$$

Proof. Given a unitary vector $x \in \mathbb{R}^2$, define the orthogonal projection map $\pi_x^\perp : \mathbb{R}^2 \to \mathbb{R}^2$ such that $\pi_x^\perp(v) = v - (v \cdot x)x$.

Observe that given two unitary vectors $p, v \in \mathbb{R}^2$ with $p \perp v$ and $g \in \mathrm{SL}(2)$, one has:

$$1 = \|v \wedge p\| = \|(gv) \wedge (gp)\| = \|gp\| \|\pi^\perp_{\frac{gp}{\|gp\|}}(gv)\|.$$

Let h be the map $x \to \frac{1}{\|gx\|}$ and f be the map $x \mapsto gx$ so that $\varphi_g = hf$. Then for $v \in T_x\mathbb{S}^1 \subset \mathbb{R}^2$,

$$\begin{aligned} D(\varphi_g)_x(v) &= D(h)_x(v)gx + \frac{1}{\|gx\|}D(f)_x(v) \\ &= D(\frac{1}{\langle gx, gx\rangle^{\frac{1}{2}}})_x(v)gx + \frac{1}{\|gx\|}gv \\ &= -\langle gx, gv\rangle \frac{1}{\|gx\|^3}gx + \frac{1}{\|gx\|}gv \\ &= \frac{1}{\|gx\|}\pi^\perp_{\frac{gx}{\|gx\|}}(gv). \end{aligned}$$

Hence,

$$D(\varphi_g)_p(v) = \frac{\pi^\perp_{\frac{gp}{\|gp\|}}(gv)}{\|gp\|} = \frac{1}{\|gp\|^2}.$$

■

We can also consider the projective map $\hat{g} : \mathbb{PR}^2 \to \mathbb{PR}^2$ such that $\hat{g}(\hat{p}) = \widehat{gp}$. By making the correct identifications and using the last proposition, we can conclude that

$$D(\hat{g})_p(v) = \frac{1}{\|gp\|^2}.$$

Proposition 5.2.5 *Given a measure $\mu \in \mathrm{Prob}(\mathrm{SL}_2(\mathbb{R}))$ with $\mathrm{supp}(\mu) \subset \Sigma$, a compact set,*

$$k_\alpha(\mu) = \sup_{\hat{p} \in \mathbb{PR}^d} \int_\Sigma \left[\|gp\|^{-2\alpha}\right] \, d\mu(g) \quad \textit{for all } \alpha > 0,$$

where p is a unit representative of $\hat{p} \in \mathbb{PR}^2$.

Proof. By proposition 5.2.3, we know that:

$$\left[\frac{\delta(\hat{g}\hat{p},\hat{g}\hat{q})}{\delta(\hat{p},\hat{q})}\right]^{\alpha} \leq \frac{1}{2}\left\{\frac{1}{\|gp\|^{2\alpha}}+\frac{1}{\|gq\|^{2\alpha}}\right\}.$$

So, if we integrate both sides then apply the supremum, we conclude that

$$k_\alpha(\mu) \leq \sup_{\hat{p},\hat{q}\in\mathbb{PR}^2}\int_\Sigma \frac{1}{2}\left[\|gp\|^{-2\alpha}+\|gq\|^{-2\alpha}\right]\,d\mu(g) = \sup_{\hat{p}\in\mathbb{PR}^2}\int_\Sigma\left[\|gp\|^{-2\alpha}\right]\,d\mu(g),$$

for every $\alpha > 0$.

For the other inequality, we need to observe that

$$\frac{1}{\|gx\|^2} = D(\hat{g}(\hat{x}))v = \lim_{\hat{y}\to\hat{x}}\left[\frac{\delta(\hat{g}\hat{x},\hat{g}\hat{y})}{\delta(\hat{x},\hat{y})}\right],$$

where the limit is taken over the projective line $\operatorname{span}\{x,v\}\subset\mathbb{PR}^2$. Hence,

$$\begin{aligned} k_\alpha(\mu) = \sup_{\hat{p}\neq\hat{q}}\int_\Sigma\left(\frac{\delta(\hat{g}\hat{p},\hat{g}\hat{q})}{\delta(\hat{p},\hat{q})}\right)^{\alpha}\,d\mu(g) &\geq \sup_{\hat{p}\neq\hat{q}}\int_\Sigma\lim_{\hat{q}\to\hat{p}}\left[\frac{\delta(\hat{g}\hat{p},\hat{g}\hat{q})}{\delta(\hat{p},\hat{q})}\right]^{\alpha}\,d\mu(g)\\ &= \sup_{\hat{p}\in\mathbb{PR}^2}\int_\Sigma\left[\|gp\|^{-2}\right]^{\alpha}\,d\mu(g). \end{aligned}$$

■

Proposition 5.2.6 *Let $\mu \in \operatorname{Prob}(\Sigma)$ be a quasi irreducible measure with $\lambda_+(\mu) > 0$. There are numbers $\delta > 0$, $0 < \alpha < 1$, $0 < k < 1$ and $n \in \mathbb{N}$ such that for all $\nu \in \operatorname{Prob}(\Sigma)$ with $W_1(\mu,\nu) < \delta$, one has $k_\alpha(\nu^{*n}) \leq k$.*

Proof.

Our strategy to prove this proposition will be to show that if μ satisfies the hypothesis, then there exist $0 < \alpha < 1$, $0 < k < 1$ and $n \in \mathbb{N}$ such that $k_\alpha(\mu^{*n}) \leq k$. Then, we are going to extend this result to a neighbourhood of μ.

In order to prove the proposition for $k_\alpha(\mu^{*n})$, it is sufficient to bound $\int_\Sigma\left[\|gp\|^{-2\alpha}\right]\,d\mu^{*n}(g)$ by a constant smaller then 1 that does not depend on p because of the result in proposition 5.2.5.

First we are going to state some inequalities. Note that

$$\lim_{n\to\infty}\frac{1}{n}\int_\Sigma\left[\log\|gp\|^{-2}\right]\,d\mu^{*n}(g) = -2\lambda_+(\mu) < 0 \quad \text{(by proposition 5.2.2)}$$

with uniform convergence in $p \in \mathbb{S}^1$.

Thus, for every $\epsilon > 0$ and every $p \in \mathbb{S}^1$, there exists some $N \in \mathbb{N}$ (that does not depend on p) such that for every $n > N$ we have that

$$-2\lambda_+(\mu) - \epsilon \leq \frac{1}{n}\int_\Sigma \left[\log \|gp\|^{-2}\right] \, d\mu^{*n}(g) \leq -2\lambda_+(\mu) + \epsilon.$$

Hence, by choosing ϵ small and n sufficiently large, we conclude that

$$\int_\Sigma \left[\log \|gp\|^{-2}\right] \, d\mu^{*n}(g) \leq n\left(-2\lambda_+(\mu) + \epsilon\right) \leq -1. \tag{5.3}$$

Since $g \in \mathrm{SL}(2)$, we have that:

$$\log \|g^{-1}\|^{-1} = \log \|g\|^{-1} \leq \log \|gp\| \leq \log \|g\|.$$

Hence,

$$\left|\log \|gp\|\right| \leq \log \|g\|. \tag{5.4}$$

Consider also the following classical inequality:

$$e^x \leq 1 + x + \frac{x^2}{2}e^{|x|}. \tag{5.5}$$

Now we are ready to estimate $\int_\Sigma \left[\|gp\|^{-2\alpha}\right] \, d\mu^{*n}(g)$ for every unit vector $p \in \mathbb{R}^2$, using (5.3), (5.4) and (5.5):

$$\begin{aligned}
&\int_\Sigma \left[\|gp\|^{-2\alpha}\right] \, d\mu^{*n}(g) = \int_\Sigma \left[e^{\log \|gp\|^{-2\alpha}}\right] \, d\mu^{*n}(g) \\
&\leq \int_\Sigma \left[1 + \alpha \log \|gp\|^{-2} + \frac{\alpha^2 \log^2 \left(\|gp\|^{-2}\right)}{2} e^{\left|\alpha \log \|gp\|^{-2}\right|}\right] \, d\mu^{*n}(g) \\
&= 1 + \alpha \int_\Sigma \left[\log \|gp\|^{-2}\right] \, d\mu^{*n}(g) + \frac{\alpha^2}{2}\int_\Sigma \left[\log^2\left(\|gp\|^{-2}\right) e^{\left|\alpha \log \|gp\|^{-2}\right|}\right] \, d\mu^{*n}(g) \\
&\leq 1 + \alpha(-1) + \frac{\alpha^2}{2}\int_\Sigma \left[4 \log^2 \left(\|g\|\right) e^{2\alpha \log \|g\|}\right] \, d\mu^{*n}(g).
\end{aligned}$$

Note that the last term is a constant K that depends only on μ and n. Then we conclude that for every unit vector $p \in \mathbb{R}^2$:

$$\int_\Sigma \left[\|gp\|^{-2\alpha}\right] \, d\mu^{*n}(g) \leq 1 - \alpha + \frac{\alpha^2}{2}K(\mu, n).$$

Thus, by taking α sufficiently small we conclude that

$$\int_\Sigma \left[\|gp\|^{-2\alpha}\right] \, d\mu^{*n}(g) \leq 1.$$

Hence, by proposition 5.2.5,

$$k_\alpha(\mu^{*n}) = \sup_{\hat{p} \in \mathbb{P}\mathbb{R}^d} \int_\Sigma \left[\|gp\|^{-2\alpha}\right] \, d\mu^{*n}(g) < 1.$$

Finally, note that $k_\alpha(\mu^{*n})$ depends continuously on μ^{*n} and remember that the map $\mu \mapsto \mu^{*n}$ is Lipschitz by proposition 4.3.1.
So, $\forall \epsilon > 0,\ \exists \delta > 0$ such that

$$W_1(\nu,\mu) < \delta \implies |k_\alpha(\mu^{*n}) - k_\alpha(\nu^{*n})| \leq \epsilon.$$

Since $k_\alpha(\mu^{*n}) < 1$, we can choose $\epsilon = \frac{1-k_\alpha(\mu^{*n})}{2}$. Then, there exists a δ-neighbourhood of μ, such that for every $\nu \in \mathrm{Prob}(\Sigma)$ with $W_1(\nu,\mu) < \delta$, we have that $k_\alpha(\nu^{*n}) \leq k_\alpha(\mu^{*n}) + \epsilon := k < 1$. ■

5.3 Hölder continuous dependence on the measure

The goal of this section is to prove that the stationary measure associated to the linear random cocycle depends Hölder continuously on the measure μ that generates the cocycle.

First we start with some preliminary results that will help us to obtain the desired estimate.

Lemma 5.3.1 *Given two different points $p, q \in \mathbb{R}^2 \backslash \{0\}$,*

$$\left\| \frac{p}{\|p\|} - \frac{q}{\|q\|} \right\| \leq \|p-q\| \max\left\{ \frac{1}{\|p\|}, \frac{1}{\|q\|} \right\}.$$

Proof. Consider $v_1, v_2 \in \mathbb{R}^2$, such that $\|v_1\| \geq \|v_2\| = 1$. We are give the analytic proof with an argument of plane geometry. It follows from the law of cosines that $\left\| \frac{v_1}{\|v_1\|} - \frac{v_2}{\|v_2\|} \right\|^2 = 2 - 2\cos\alpha$, where α is the angle between v_1 and v_2. Consider $y = v_1 - \frac{v_1}{\|v_1\|}$, so that $y \geq 0$. Again, by the law of cosines, $\|v_1 - v_2\|^2 = 2 + 2y + y^2 - 2\cos\alpha - 2y\cos\alpha$. Since $2y + y^2 - 2y\cos\alpha \geq 0$ for every α, we have that

$$\left\| \frac{v_1}{\|v_1\|} - \frac{v_2}{\|v_2\|} \right\| \leq \|v_1 - v_2\|. \tag{5.6}$$

Now, suppose that $\|p\| \leq \|q\|$. Set $v_1 = \frac{q}{\|p\|}$ and $v_2 = \frac{p}{\|p\|}$. Then, substitute this values in (5.6) to get:

$$\left\| \frac{q}{\|q\|} - \frac{p}{\|p\|} \right\| \leq \left\| \frac{q}{\|p\|} - \frac{p}{\|p\|} \right\| = \frac{1}{\|p\|}\|p-q\|.$$

It is clear that if $\|q\| \leq \|p\|$, then obtain an analogous result:

$$\left\| \frac{q}{\|q\|} - \frac{p}{\|p\|} \right\| \leq \frac{1}{\|q\|}\|p-q\|.$$

Therefore, we conclude that

$$\left\| \frac{q}{\|q\|} - \frac{p}{\|p\|} \right\| \leq \max\left\{ \frac{1}{\|p\|}, \frac{1}{\|q\|} \right\} \|p-q\|.$$

∎

Corollary 5.3.1 *Let $g_1, g_2 \in \mathrm{SL}_2(\mathbb{R})$ and consider their projective actions, $\hat{g}_i : \mathbb{PR}^2 \to \mathbb{PR}^2$ taking $\hat{p}$ to $\frac{g_i(p)}{\|g_i(p)\|}$ for $i = 1, 2$.*

Then

$$\delta(\hat{g}_1\hat{p}, \hat{g}_2\hat{p}) \leq \|g_1 - g_2\| \max\left\{ \frac{1}{\|g_1(p)\|}, \frac{1}{\|g_2(p)\|} \right\}.$$

Proof. Observe that

$$\delta(\hat{g}_1\hat{p}, \hat{g}_2\hat{p}) \leq \left\| \frac{g_1(p)}{\|g_1(p)\|} - \frac{g_2(p)}{\|g_2(p)\|} \right\|.$$

Then, by lemma 5.3.1,

$$\begin{aligned}\left\| \frac{g_1(p)}{\|g_1(p)\|} - \frac{g_2(p)}{\|g_2(p)\|} \right\| &\leq \|g_1(p) - g_2(p)\| \max\left\{ \frac{1}{\|g_1(p)\|}, \frac{1}{\|g_2(p)\|} \right\} \\ &\leq \|g_1 - g_2\| \max\left\{ \frac{1}{\|g_1(p)\|}, \frac{1}{\|g_2(p)\|} \right\}.\end{aligned}$$

∎

Proposition 5.3.2 *Let $\mu_1, \mu_2 \in \mathrm{Prob}(\Sigma)$. Assume that $k := k_\alpha(\mu_1) < 1$ for some $0 < \alpha \leq 1$. Then for all $n \in \mathbb{N}$ and $\varphi \in \mathcal{H}_\alpha(\mathbb{PR}^2)$,*

$$\|Q^n_{\mu_1}(\varphi) - Q^n_{\mu_2}(\varphi)\|_\infty \leq \frac{W_1(\mu_1, \mu_2)^\alpha}{1-k} v_\alpha(\varphi).$$

Moreover, if also $k_\alpha(\mu_2) < 1$ then for all $\varphi \in \mathcal{H}_\alpha(\mathbb{PR}^2)$,

$$\left| \int_{\mathbb{PR}^2} \varphi \, d\nu_{\mu_1} - \int_{\mathbb{PR}^2} \varphi \, d\nu_{\mu_2} \right| \leq \frac{W_1(\mu_1, \mu_2)^\alpha}{1-k} v_\alpha(\varphi),$$

where ν_{μ_1} and ν_{μ_2} are the unique stationary measures with respect to the cocycle generated by the measures μ_1 and μ_2, respectively.

Proof. For $n = 1$ we have that:

$$
\begin{aligned}
\|Q_{\mu_1}(\varphi) - Q_{\mu_2}(\varphi)\|_\infty &= \sup_{\hat{p}\in\mathbb{PR}^2} \left| \int_\Sigma \varphi(\hat{g}_1(\hat{p}))d\mu_1(g_1) - \int_\Sigma \varphi(\hat{g}_2(\hat{p}))\, d\mu_2(g_2) \right| \\
&= \sup_{\hat{p}\in\mathbb{PR}^2} \left| \int_{\Sigma\times\Sigma} \varphi(\hat{g}_1(\hat{p})) - \varphi(\hat{g}_2(\hat{p}))\, d\pi(g_1,g_2) \right| \qquad \forall \pi \in \Pi(\mu_1,\mu_2) \\
&\leq \sup_{\hat{p}\in\mathbb{PR}^2} \int_{\Sigma\times\Sigma} |\varphi(\hat{g}_1(\hat{p})) - \varphi(\hat{g}_2(\hat{p}))|\ d\pi(g_1,g_2) \qquad \forall \pi \in \Pi(\mu_1,\mu_2) \\
&\leq \sup_{\hat{p}\in\mathbb{PR}^2} \int_{\Sigma\times\Sigma} |\varphi(\hat{g}_1(\hat{p})) - \varphi(\hat{g}_2(\hat{p}))| \frac{\delta\left(\hat{g}_1\hat{p},\hat{g}_2\hat{p}\right)^\alpha}{\delta\left(\hat{g}_1\hat{p},\hat{g}_2\hat{p}\right)^\alpha}\, d\pi(g_1,g_2) \quad \forall \pi \in \Pi(\mu_1,\mu_2) \\
&\leq v_\alpha(\varphi) \sup_{\hat{p}\in\mathbb{PR}^2} \int_{\Sigma\times\Sigma} \delta\left(\hat{g}_1\hat{p},\hat{g}_2\hat{p}\right)^\alpha\ d\pi(g_1,g_2) \qquad \forall \pi \in \Pi(\mu_1,\mu_2) \\
&\leq v_\alpha(\varphi) \sup_{\hat{p}\in\mathbb{PR}^2} \int_{\Sigma\times\Sigma} \|g_1-g_2\|^\alpha \max\left\{\frac{1}{\|g_1(p)\|}, \frac{1}{\|g_2(p)\|}\right\}^\alpha\ d\pi(g_1,g_2),
\end{aligned}
$$

for every $\pi \in \Pi(\mu_1,\mu_2)$, by corollary 5.3.1.

Since Σ is compact, $\max\left\{\frac{1}{\|g_1(p)\|}, \frac{1}{\|g_2(p)\|}\right\} \leq \max\{\|g_1\|,\|g_2\|\} = C$.

Then, for every $\pi \in \Pi(\mu_1,\mu_2)$,

$$
\begin{aligned}
\|Q_{\mu_1}(\varphi) - Q_{\mu_2}(\varphi)\|_\infty &\leq C^\alpha v_\alpha(\varphi) \sup_{\hat{p}\in\mathbb{PR}^2} \int_{\Sigma\times\Sigma} \|g_1-g_2\|^\alpha\ d\pi(g_1,g_2) \\
&\leq C^\alpha v_\alpha(\varphi) \sup_{\hat{p}\in\mathbb{PR}^2} \left(\int_{\Sigma\times\Sigma} \|g_1-g_2\|\ d\pi(g_1,g_2) \right)^\alpha \\
&\leq C^\alpha v_\alpha(\varphi) W_1(\mu_1,\mu_2)^\alpha,
\end{aligned}
$$

where on the second line we used Jensen's inequality and the concavity of the function $t \mapsto t^\alpha$, which holds when $t \in [0,\infty)$ and $\alpha \in (0,1]$.

Now observe that the difference $Q^n_{\mu_1} - Q^n_{\mu_2}$ can be written as a telescopic sum as follows:

$$
\begin{aligned}
Q^n_{\mu_1} - Q^n_{\mu_2} &= Q^n_{\mu_1} - Q_{\mu_2} \circ Q^{n-1}_{\mu_1} + Q_{\mu_2} \circ Q^{n-1}_{\mu_1} - \ldots + Q^{n-1}_{\mu_2} \circ Q_{\mu_1} - Q^n_{\mu_2} \\
&= \sum_{i=0}^{n-1} Q^i_{\mu_2} \circ (Q_{\mu_1} - Q_{\mu_2}) \circ Q^{n-i-1}_{\mu_1}.
\end{aligned}
$$

Now we use the previous relation to prove the desired estimate:

$$
\begin{aligned}
\|Q^n_{\mu_1}(\varphi) - Q^n_{\mu_2}(\varphi)\|_\infty &= \| \sum_{i=0}^{n-1} Q^i_{\mu_2} \circ (Q_{\mu_1} - Q_{\mu_2}) \circ \left(Q^{n-i-1}_{\mu_1}(\varphi)\right) \|_\infty \\
&\leq \sum_{i=0}^{n-1} \|Q^i_{\mu_2} \circ (Q_{\mu_1} - Q_{\mu_2}) \circ \left(Q^{n-i-1}_{\mu_1}(\varphi)\right) \|_\infty \\
&\leq \sum_{i=0}^{n-1} \| (Q_{\mu_1} - Q_{\mu_2}) \circ \left(Q^{n-i-1}_{\mu_1}(\varphi)\right) \|_\infty \ \text{(by Lemma 5.1.3)}
\end{aligned}
$$

By the case $n = 1$ we have that

$$
\begin{aligned}
&\sum_{i=0}^{n-1} \| \left(Q_{\mu_1} - Q_{\mu_2}\right) \circ \left(Q_{\mu_1}^{n-i-1}(\varphi)\right) \|_\infty \leq \sum_{i=0}^{n-1} C^\alpha v_\alpha \left(Q_{\mu_1}^{n-i-1}(\varphi)\right) W_1(\mu_1,\mu_2)^\alpha \\
&= C^\alpha W_1(\mu_1,\mu_2)^\alpha \sum_{i=0}^{n-1} v_\alpha \left(Q_{\mu_1^{n-i-1}}(\varphi)\right) \\
&\leq C^\alpha W_1(\mu_1,\mu_2)^\alpha \sum_{i=0}^{n-1} k_\alpha \left(\mu_1^{n-i-1}\right) v_\alpha(\varphi) \qquad \text{(by proposition 5.1.4)} \\
&\leq v_\alpha(\varphi) C^\alpha W_1(\mu_1,\mu_2)^\alpha \sum_{i=0}^{n-1} \left(k_\alpha(\mu_1)\right)^{n-i-1} \qquad \text{(by proposition 5.1.2)} \\
&\leq v_\alpha(\varphi) C^\alpha W_1(\mu_1,\mu_2)^\alpha \sum_{i=0}^{\infty} \left(k_\alpha(\mu_1)\right)^{i} \\
&\leq \frac{C^\alpha W_1(\mu_1,\mu_2)^\alpha}{1-k} v_\alpha(\varphi). \qquad \text{(since } k := k_\alpha < 1)
\end{aligned}
$$

This concludes the fisrt part of the proof.

Now, consider that also $k_\alpha(\mu_2) < 1$. By proposition 5.1.7, there exist stationary measures $\nu_{\mu_1}, \nu_{\mu_2} \in \mathrm{Prob}(\mathbb{PR}^2)$ with respect to μ_1 and μ_2, respectively, such that the Markov operators associated to Q_{μ_1} and Q_{μ_2} satisfy the three properties of that proposition.

Then we can write

$$\varphi = \left(\int_{\mathbb{PR}^2} \varphi \, d\nu_{\mu_1}\right) \mathbf{1} + \varphi - \int_{\mathbb{PR}^2} \varphi \, d\nu_{\mu_1} \mathbf{1}.$$

Note that $\left(\int_{\mathbb{PR}^2} \varphi \, d\nu_{\mu_1}\right) \mathbf{1} \in \mathcal{C}\mathbf{1}$ and $\Psi := \varphi - \int_{\mathbb{PR}^2} \varphi \, d\nu_{\mu_1} \mathbf{1} \in \mathcal{N}_\alpha(\nu_{\mu_1})$. Hence,

$$Q_{\mu_1}^n(\varphi) = Q_{\mu_1}^n\left(\left(\int_{\mathbb{PR}^2} \varphi \, d\nu_{\mu_1}\right) \mathbf{1}\right) + Q_{\mu_1}^n(\Psi).$$

Since $\Psi \in \mathcal{N}_\alpha(\nu_{\mu_1})$, the second term goes to zero when n goes to infinity, because Q_{μ_1} acts as a contraction on $\mathcal{N}_\alpha(\nu_{\mu_1})$ for large n. Moreover,

$$Q_{\mu_1}^n\left(\left(\int_{\mathbb{PR}^2} \varphi \, d\nu_{\mu_1}\right) \mathbf{1}\right) = \int_{\mathbb{PR}^2} \varphi \, d\nu_{\mu_1} \, \mathbf{1}$$

for every n, because Q_{μ_1} fixes every function in $\mathbb{C}\mathbf{1}$. So,

$$\lim_{n\to\infty} Q_{\mu_1}^n(\varphi) = \int_{\mathbb{PR}^2} \varphi \, d\nu_{\mu_1} \, \mathbf{1}.$$

Similarly we can do the same argument for Q_{μ_2} and conclude that

$$\lim_{n\to\infty} Q_{\mu_2}^n(\varphi) = \int_{\mathbb{PR}^2} \varphi \, d\nu_{\mu_2} \, \mathbf{1}.$$

Finally, we have that

$$\left|\int_{\mathbb{PR}^2} \varphi\, d\nu_{\mu_1} - \int_{\mathbb{PR}^2} \varphi\, d\nu_{\mu_2}\right| \leq \sup_{n\to\infty} \|Q^n_{\mu_1}(\varphi) - Q^n_{\mu_2}(\varphi)\|_\infty \leq \frac{C^\alpha W_1(\mu_1,\mu_2)^\alpha}{1-k} v_\alpha(\varphi),$$

which concludes the proof of the proposition. ■

5.4
Generalization of Le Page's Theorem

In this section we propose a generalization to the result from Baraviera-Duarte and show that their version of Le Page's Theorem can be derived from ours.

Proposition 5.4.1 *For every $n_0 \in \mathbb{N}$ and $\mu \in \Sigma$,*

$$\lambda_+(\mu^{*n_0}) = n_0 \lambda_+(\mu).$$

Proof. By Furstenberg-Kersten Theorem 2.2.1, since f is ergodic:

$$\begin{aligned}
\lambda_+(\mu^{*n_0}) &= \lim_{n\to\infty} \frac{1}{n} \int_{\Sigma^n} \log \|g_{n-1}...g_0\|\, d\mu^{*n_0}(g_{n-1})...d\mu^{*n_0}(g_0) \\
&= \lim_{n\to\infty} \frac{1}{n} \int_{\Sigma^{nn_0}} \log \|g_{nn_0-1}...g_0\|\, d\mu(g_{nn_0-1})...d\mu(g_0) \\
&= n_0 \lim_{n\to\infty} \frac{1}{nn_0} \int_{\Sigma^{nn_0}} \log \|g_{nn_0-1}...g_0\|\, d\mu(g_{nn_0-1})...d\mu(g_0) \\
&= n_0 \lambda_+(\mu).
\end{aligned}$$

■

Theorem 5.4.2 *(M. Durães and S. Klein). Let $\Sigma \in \mathrm{SL}(2)$ be a compact set, $\mu \in \mathrm{Prob}(\mathrm{SL}(2))$ such that $\mathrm{supp}(\mu) \subset \Sigma$. Suppose that μ is quasi irreducible and $\lambda_+(\mu) > 0$. Hence there exist $\delta > 0$, $C < \infty$ and $\alpha \in (0,1)$ such that given any μ_i, $i \in \{1,2\}$, satisfying $W_1(\mu_i, \mu) < \delta$, we have*

$$|\lambda_+(\mu_1) - \lambda_+(\mu_2)| \leq C W_1(\mu_1,\mu_2)^\alpha.$$

Proof. By the hypothesis, we are in the conditions of proposition 5.2.6. Hence, there exist $\overline{\delta} > 0$, $\alpha \in (0,1)$, $k \in (0,1)$ and $n_0 \in \mathbb{N}$ such that for all $\nu \in \mathrm{Prob}(\Sigma)$ with $W_1(\mu,\nu) < \overline{\delta}$, one has $k_\alpha(\nu^{*n_0}) \leq k$.

By proposition 4.3.1, the map $\nu \to \nu^{*n}$ is Lipschitz with some constant K. Let $\mu_1, \mu_2 \in \mathrm{Prob}(\Sigma)$ satisfy $W_1(\mu_i,\mu) < \frac{\overline{\delta}}{K} := \delta$ for $i \in \{1,2\}$. Hence, $\mu_1^{*n_0}, \mu_2^{*n_0}$ satisfy $W_1(\mu_i^{*n0}, \mu) < \overline{\delta}$ for $i \in \{1,2\}$.

Thus, we can apply proposition 5.1.7 to $\mu_1^{*n_0}$ and $\mu_2^{*n_0}$. Therefore, there exist unique stationary measures $\eta_{\mu_1^{*n_0}}$ and $\eta_{\mu_2^{*n_0}}$ with respect to $\mu_1^{*n_0}$ and $\mu_2^{*n_0}$ respectively.

Hence, by Furstenberg-Ledrappier's Formula :

$$\begin{aligned}\lambda_+(\mu_1^{*n_0}) &= \max\left\{\int_{\mathbb{PR}^2}\int_\Sigma \varphi \, d\mu_1^{*n_0} d\nu_{\mu_1^{*n_0}} : \nu_{\mu_1^{*n_0}} \text{ is a stationary measure for } \mathbb{PR}^2\right\}\\ &= \int_{\mathbb{PR}^2}\int_\Sigma \varphi \, d\mu_1^{*n_0} d\eta_{\mu_1^{*n_0}}\end{aligned}$$

and

$$\begin{aligned}\lambda_+(\mu_2^{*n_0}) &= \max\left\{\int_{\mathbb{PR}^2}\int_\Sigma \varphi \, d\mu_2^{*n_0} d\nu_{\mu_2^{*n_0}} : \nu_{\mu_2^{*n_0}} \text{ is a stationary measure for } \mathbb{PR}^2\right\}\\ &= \int_{\mathbb{PR}^2}\int_\Sigma \varphi \, d\mu_2^{*n_0} d\eta_{\mu_2^{*n_0}},\end{aligned}$$

where $\varphi(g,\hat{p}) = \log \|gp\|$ and p is a unitary representative of $\hat{p}$. Now we are able to estimate:

$$\begin{aligned}&|\lambda_+(\mu_1^{*n_0}) - \lambda_+(\mu_2^{*n_0})| = \left|\int_{\mathbb{PR}^2}\int_\Sigma \varphi \, d\mu_1^{*n_0} d\eta_{\mu_1^{*n_0}} - \int_{\mathbb{PR}^2}\int_\Sigma \varphi \, d\mu_2^{*n_0} d\eta_{\mu_2^{*n_0}}\right|\\ &\leq \left|\int_{\mathbb{PR}^2}\int_\Sigma \varphi \, d\mu_1^{*n_0} d\eta_{\mu_1^{*n_0}} - \int_{\mathbb{PR}^2}\int_\Sigma \varphi \, d\mu_1^{*n_0} d\eta_{\mu_2^{*n_0}}\right| +\\ &+ \left|\int_{\mathbb{PR}^2}\int_\Sigma \varphi \, d\mu_1^{*n_0} d\eta_{\mu_2^{*n_0}} - \int_{\mathbb{PR}^2}\int_\Sigma \varphi \, d\mu_2^{*n_0} d\eta_{\mu_2^{*n_0}}\right|\\ &= \left|\int_\Sigma\int_{\mathbb{PR}^2} \varphi \, d\left(\eta_{\mu_1^{*n_0}} - \eta_{\mu_2^{*n_0}}\right) d\mu_1^{*n_0}\right| + \left|\int_{\mathbb{PR}^2}\int_\Sigma \varphi \, d(\mu_1^{*n_0} - \mu_2^{*n_0}) d\eta_{\mu_2^{*n_0}}\right|.\end{aligned}$$

Since $k_\alpha(\mu_1^{*n_0}) \leq k < 1$ and $k_\alpha(\mu_2^{*n_0}) \leq k < 1$, we know by proposition 5.3.2 that the first term is bounded by

$$\frac{W_1(\mu_1^{*n_0}, \mu_2^{*n_0})^\alpha}{1 - k_\alpha(\mu_1^{*n_0})} v_\alpha(\varphi).$$

Note that φ is Lipschitz on Σ with respect to the first coordinate with some Lipschitz constant L. Then, using Kantorovich-Rubinstein Theorem, the second term is bounded by $L \cdot W_1(\mu_1^{*n_0}, \mu_2^{*n_0})$.

Hence, there exists some C_0 such that

$$|\lambda_+(\mu_1^{*n_0}) - \lambda_+(\mu_2^{*n_0})| \leq C_0 W_1(\mu_1^{*n_0}, \mu_2^{*n_0})^\alpha.$$

Now let $C = \frac{C_0 K^\alpha}{n_0}$. By proposition 5.4.1, the previous result and proposition 4.3.1:

$$\begin{aligned}|\lambda_+(\mu_1) - \lambda_+(\mu_2)| &= \frac{|\lambda_+(\mu_1^{*n_0}) - \lambda_+(\mu_2^{*n_0})|}{n_0} \\ &\le CW_1(\mu_1, \mu_2)^\alpha.\end{aligned}$$

This proves that the maximal Lyapunov exponent is Hölder continuous in a neighbourhood of μ. ∎

Now we show that our theorem implies the result of Baraviera and Duarte.

Definition 5.4.1 *Given a probability space* $(X, \mathcal{B}, \mu)$ *we can define the support of the measure* μ *by:*

$$\operatorname{supp}(\mu) = \bigcap_{E \text{ closed and } \mu(E^c)=0} E.$$

Lemma 5.4.3 *Consider a probability space* $(X, \mathcal{F}, \rho)$ *and* $A : X \to \mathrm{SL}_2(\mathbb{R})$ *a measurable function. Then* $\operatorname{supp}(A_*\rho) \subset \overline{\operatorname{Im}(A)}$.

Proof. Observe that $\overline{\operatorname{Im}(A)}^c$ is disjoint from $\operatorname{Im}(A)$. Hence $A^{-1}\left(\overline{\operatorname{Im}(A)}^c\right) = \emptyset$. Then

$$A_*\rho(\overline{\operatorname{Im}(A)}^c) = \rho(A^{-1}(\overline{\operatorname{Im}(A)}^c)) = \rho(\emptyset) = 0.$$

So, $\overline{\operatorname{Im}(A)}$ is closed and $A_*\rho(\overline{\operatorname{Im}(A)}^c) = 0$. Since $\operatorname{supp}(A_*\rho)$ is the smallest set with this properties, we conclude that

$$\operatorname{supp}(A_*\rho) \subset \overline{\operatorname{Im}(A)}.$$

∎

Fix a measure ρ and let the base dynamics be the Bernoulli Shift. We claim that the random linear cocycle defined by $A \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$ is generated by a push-forward measure $A_*\rho$.

Remember that the probability measure $A_*\rho$ generates a random linear cocycle with the Bernoulli Shift on the base and the projection on the first coordinate on the fiber. Hence,

$$\lambda_+(A_*\rho) = \lim_{n\to\infty} \frac{1}{n} \log \|g_{n-1} \dots g_1 g_0\|,$$

where each matrix g_j, with $j \in \{1, ..., n-1\}$, is chosen according to the probability law $A_*\rho$.

Note that the random linear cocycle defined by $A \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$ satisfies

$$\lambda_+(A,\rho) = \lim_{n\to\infty} \frac{1}{n} \log \|A^{n-1}(x)...A(f(x))A(x)\|,$$

where each x is chosen according to the probability law ρ. Therefore each element $A(f^j(x))$, with $j \in \{1,...,n-1\}$, is chosen accordingly to the probability law $A_*\rho$. Thus the two cocycles are the same.

Consider the following set of probability measures on SL(2):

$$\mathcal{F} = \{A_*\rho : A \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))\}.$$

Proposition 5.4.4 *There exists a compact set K in the weak star topology such that $\mathcal{F} \subset K$.*

Proof. By lemma 5.4.3, $\mathrm{supp}(A_*\rho) \subset \overline{\mathrm{Im}(A)}$. Since $A \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$, there exists some $C > 0$ such that $\overline{\mathrm{Im}(A)} \subset \overline{B_{\mathrm{SL}_2(\mathbb{R})}(0,C)}$, where $B_{\mathrm{SL}_2(\mathbb{R})}(0,C)$ is the ball in SL(2) with center 0 and radius C. This implies that

$$\mathcal{F} \subset \mathrm{Prob}(\overline{\mathrm{Im}(A)}) \subset \mathrm{Prob}(\overline{B_{\mathrm{SL}_2(\mathbb{R})}(0,C)}).$$

Since $\overline{B_{\mathrm{SL}_2(\mathbb{R})}(0,C)}$ is compact, we know, by Prohorov's Theorem, that $\mathrm{Prob}(\overline{B_{\mathrm{SL}_2(\mathbb{R})}(0,C)})$ is compact in the weak star topology. Hence, there exists a compact set, namely $K = \mathrm{Prob}(\overline{B_{\mathrm{SL}_2(\mathbb{R})}(0,C)})$ that contains all the measures in $\mathcal{F}$. ∎

Corollary 5.4.1 *(A. Baraviera and P. Duarte).*
Let $(A,\mu) \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R})) \times \mathrm{Prob}(\Sigma)$ be quasi-irreducible with $L(A,\mu) > 0$. Then the cocycle is locally Hölder continuous: there are positive constants $\alpha > 0$, $C < \infty$ and $\delta > 0$ such that for all $B_1, B_2 \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$ if $\|B_j - A\|_\infty < \delta$, $j = 1,2$, then

$$|\lambda_+(B_1) - \lambda_+(B_2)| \leq C\,(\|B_1 - B_2\|_\infty)^\alpha.$$

Proof.

Observe that the version of Le Page's Theorem in Baraviera-Duarte assumes the cocycles to be in $L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$. Hence the cocycles are not uniformly bounded as in the previous results. However, this is not a problem, since our theorem is a local result. Given $A \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$, all the cocycles B such that $\|B - A\|_\infty < \delta$ are uniformly bounded by $\|A\|_\infty + \delta$. Hence all the measures supported in this neighbourhood of A (that generate all of these cocycles) are contained in a compact set. This means that we are in the conditions of our generalized version of Le Page's Theorem, thus concluding the result of Baraviera-Duarte.

■

The method described in this work for establishing the Hölder continuity of Lyapunov exponents of a linear cocycle can likely be applied to some other models, with different ergodic dynamics on the base. We intend to consider the case of a subshift of finite type (or, more generally of a Markov system) in a future project.

Two of the main elements of the proof are the Furstenberg-Ledrappier Formula in theorem 3.3.7 and the uniform convergence result in proposition 5.2.2. Since the two ingredients mentioned above are essentially already available, this project appears feasible.

A
Elements of Functional Analysis and Ergodic Theory

In this appendix, we review some results from functional analysis and ergodic theory that will be useful during the text. We are going to give references for some concepts used during the text in order to keep it self-contained.

Definition A.0.1 *Let X be a real vector space. A sublinear functional on X is a function $p : X \to \mathbb{R}$ such that*

(i) $p(x+y) \leq p(x) + p(y) \quad \forall x, y \in X.$

(ii) $p(\alpha x) = \alpha p(x) \quad \forall x \in X \quad$ *and* $\alpha \geq 0.$

Remark 15 *Every sublinear functional $p : X \to \mathbb{R}$ satisfy:*

$$0 = p(0) = p(u + (-u)) \leq p(u) + p(-u).$$

Therefore $-p(-u) \leq p(u)$.

Theorem A.0.1 *(Hahn-Banach) Let X be a real vector space and let p be a sublinear functional on X. Suppose that W is a linear subspace of X and f_W is a linear functional in W satisfying*

$$f_W(w) \leq p(w) \quad \forall w \in W.$$

Then f_W has an extension f_X on X such that

$$f_X(x) \leq p(x) \quad \forall x \in X.$$

Proof. See chapter 5 of [9]. ■

Corollary A.0.1 *(Special version of Hahn-Banach) Let X be a real vector space, W a linear subspace of X and U an open convex set in X containing the origin. If $f_W : W \to \mathbb{R}$ is a linear functional on W, then there exists an extension $f_X : X \to \mathbb{R}$ such that*

$$\sup_U f_X(x) = \sup_{U \cap W} f_W(x).$$

Proof. Consider the sublinear functional $q : X \to [0, +\infty)$ such that

$$q(v) := \inf\{t > 0 : \frac{v}{t} \in U\}.$$

Now let $p : X \to \mathbb{R}$ be such that $p(v) = \alpha q(v)$, where $\alpha = \sup_{U \cap W} f_W$. Notice that p is also a sublinear functional on X. Moreover, since $f_W(v) \leq \alpha$ for every $v \in W \cap U$, then for every $w \in W$,

$$f_W(w) = f_W\left(\frac{w}{(1+\varepsilon)q(w)}\right)(1+\varepsilon)\, q(w) \leq \alpha(1+\varepsilon)\, q(w)$$

for every $\varepsilon > 0$. Hence, $f_W \leq p$ and we can apply theorem (A.0.1) to obtain an extension f_X of f_W such that $f_X(x) \leq p(x)$ for every $x \in X$. Furthermore, $\sup_U f_X \leq \alpha = \sup_{U \cap W} f_X$. The other inequality is trivial since $U \cap W \subset U$. ∎

Proposition A.0.2 *Let X be a normed linear space and let W be a dense subspace of X. Let Y be a Banach space and let S be a positive, linear and bounded transformation from W to Y.*

(a) If $x \in X$ and $\{x_n\}$, $\{y_n\}$ are sequences in W such that $\lim_{n\to\infty} x_n = \lim_{n\to\infty} y_n = x$, then $\{S(x_n)\}_n$ and $\{S(y_n)\}_n$ both converge and $\lim_{n\to\infty} S(x_n) = \lim_{n\to\infty} S(y_n)$.

(b) There exists T positive, linear and bounded such that $\|T\| = \|S\|$ and $Tx = Sx$ for all $x \in W$.

Proof. This proposition is proved for a linear and bounded transformation S in chapter 4 of [9] with a linear and bounded extension T in item (b). It is straightforward to see that if we add the hypothesis of S being positive, then T will also be positive by its construction. ∎

Let X be a normed linear space. The set of all bounded linear transformations from X to X is denoted by B(X).

Definition A.0.2 *Let X be a Banach space, let $I \in B(X)$ be the identity operator and let $T \in B(X)$. The spectrum of T, denoted by $\sigma(T)$, is defined to be*

$$\sigma(T) = \{\lambda \in \mathbb{C} : T - \lambda I \text{ is not invertible}\}.$$

Proposition A.0.3 *Let X be a Banach space. If $T \in B(X)$ is an operator with $\|T\| < 1$, then $I - T$ is invertible and the inverse is given by*

$$(I - T)^{-1} = \sum_{n=0}^{\infty} T^n.$$

Proof. See chapter 4 of [9]. ■

Corollary A.0.2 *Let X be a Banach space and $T \in B(X)$. Let λ be a complex number. If $|\lambda| > \|T\|$, then $\lambda \notin \sigma(T)$.*

Proof. If $|\lambda| > \|T\|$, then $\|\lambda^{-1}T\| < 1$. So, by $(A.0.3)$ the operator $I - \lambda^{-1}T$ is invertible. Hence $\lambda I - T$ is invertible, which means that $\lambda \notin \sigma(T)$. ■

Theorem A.0.4 *Let X be a Banach space, $T \in B(X)$ and σ_1, σ_2 be two complementary isolated parts of its spectrum (σ_1 or σ_2 can be empty). We can then decompose the space X in a vector sum of two linearly independent subspaces M_1 and M_2 each of which invariant by T and with the property that $\sigma\left(T\Big|_{M_i}\right) = \sigma_i$, for $i \in \{1, 2\}$.*

Proof. See chapter 11 of [7]. ■

Definition A.0.3 *Let M be a compact metric space. We say that a linear operator $\Phi : C^0(M) \to \mathbb{C}$ is positive if $\Phi(\varphi) \geq 0$ for all $\varphi \in C^0(M)$ such that $\varphi(x) \geq 0$ for every $x \in M$.*

Theorem A.0.5 *(Riesz-Markov). Let M be a compact metric space. Consider any bounded linear functional $\Phi : C^0(M) \to \mathbb{C}$. There exists a unique borel measure μ in M such that*

$$\Phi(\varphi) = \int \varphi \, d\mu \quad \forall \varphi \in C^0(M).$$

The norm $\|\mu\| = |\mu|(M)$ of the measure μ coincides with the norm $\|\Phi\|$ of the functional Φ. Also, μ is a probability measure if and only if $\Phi(1) = 1$. Moreover, μ takes values on $[0, \infty)$ if and only if the operator Φ is positive.

Proof. See chapter 6 of [8]. ■

For this last part of the appendix, we are going to assume that $(M, \mathcal{B}, \mu)$ is a probability space and $\mathcal{P}$ a partition of M into measurable subsets. We will denote by $\pi : M \to \mathcal{P}$ the projection of an element x to the element $\mathcal{P}(x)$ of the partition that contains x.

The proof of the last three results can be found at chapter 5 of [12].

Theorem A.0.6 *(Ergodic Decomposition). Let M be a complete and separable metric space and consider $f : M \to M$ a measurable transformation with an invariant probability measure μ. Then, there exists a measurable set $M_0 \subset M$ of full measure, a partition $\mathcal{P}$ of M_0 into measurable subsets and a family of probability measures $\{\mu_P : P \in \mathcal{P}\}$ in M satisfying:*

(a) $\mu_P(P) = 1$ *for* $\pi_*\mu$*-almost every* $P \in \mathcal{P}$*;*

(b) $P \to \mu_P(E)$ *is measurable for every measurable set* $E \subset M$*;*

(c) μ_P *is invariant and ergodic for* $\pi_*\mu$*-almost every* $P \in \mathcal{P}$*;*

(d) $\mu(E) = \int \mu_P(E) d\pi_*\mu(P)$ *for every measurable set* $E \subset M$*.*

Proposition A.0.7 *Let* M *be a topological space with countable basis of open sets and consider* $\mathcal{B}$ *its Borel* σ*-algebra. Suppose* $\mathcal{B}$ *admits a countable generator. If* $\{\mu_P^1 : P \in \mathcal{P}\}$ *and* $\{\mu_P^2 : P \in \mathcal{P}\}$ *are desintegrations of* μ *with respect to* $\mathcal{P}$*, then* $\mu_P^1 = \mu_P^2$ *for* $\pi_*\mu$*-almost every* $P \in \mathcal{P}$*.*

Theorem A.0.8 *(Rokhlin's Disintegration Theorem). Suppose that* M *is a complete and separable metric space,* μ *is a probability measure and* $\mathcal{P}$ *is a measurable partition. Then, the probability* μ *admits some disintegration relatively to* $\mathcal{P}$*.*

Bibliography

[1] BARAVIERA, A.; DUARTE, P. **Approximating Lyapunov exponents and stationary measures**. Journal of Dynamics and Differential Equations, Jan 2019.

[2] DUARTE, P.; KLEIN, S. **Lyapunov exponents of linear cocycles; continuity via large deviations**, volume 3 **Atlantis Studies in Dynamical Systems**. Atlantis Press, 2016.

[3] DUARTE, P.; KLEIN, S. **Continuity of the Lyapunov exponents of linear cocycles**. Publicações Matemáticas, 31° Colóquio Brasileiro de Matemática. IMPA, 2017. available at `https://impa.br/wp-content/uploads/2017/08/31CBM_02.pdf`.

[4] DUARTE, P.; KLEIN, S.. **Large deviations for products of random two dimensional matrices**. Comm. Math. Phys., 375(3):2191–2257, 2020.

[5] LE PAGE, E. **Régularité du plus grand exposant caractéristique des produits de matrices aléatoires indépendantes et applications**. Ann. Inst. H. Poincaré Probab. Statist., 25(2):109–142, 1989.

[6] MAQUERA, H. M. C. **Teorema de Furstenberg sobre o produto aleatório de matrizes**. Dissertação de mestrado, Instituto de Ciências Matemáticas de Computação, Universidade de São Paulo, São Carlos, São Paulo, 2018.

[7] RIESZ, F.; SZ.-NAGY, B. **Functional analysis**. Dover Books on Advanced Mathematics. Dover Publications, Inc., New York, 1990. Translated from the second French edition by Leo F. Boron, Reprint of the 1955 original.

[8] RUDIN, W. **Real and complex analysis**. McGraw-Hill Book Co., New York, third edition, 1987.

[9] RYNNE, B. P.; YOUNGSON, M. A. **Linear functional analysis**. Springer Undergraduate Mathematics Series. Springer-Verlag London, Ltd., London, second edition, 2008.

[10] TALL, E. H. Y.; VIANA, M.. **Moduli of continuity for the Lyapunov exponents of random $\mathrm{GL}(2)$-cocycles**. Trans. Amer. Math. Soc., 373(2):1343–1383, 2020.

[11] VIANA, M. **Lectures on Lyapunov Exponents**. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2014.

[12] VIANA, M.; OLIVEIRA, K. **Fundamentos da Teoria Ergódica**. Coleção Fronteiras da Matemática. SBM, 2019.

[13] VILLANI, C. **Optimal Transport: Old and New**, volume 338 **Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]**. Springer-Verlag, Berlin, 2009.