Pontifícia Universidade Católica
do Rio de Janeiro

**Marcelo Durães Capeleiro Pinto**

# Lyapunov Exponents of Random Linear Cocycles: Regularity and Statistical Properties

**Tese de Doutorado**

Thesis 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 Doutor em Matemática.

Advisor: Prof. Silvius Klein

Rio de Janeiro
April 2025

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

**Marcelo Durães Capeleiro Pinto**

# Lyapunov Exponents of Random Linear Cocycles: Regularity and Statistical Properties

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

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

**Prof. Jérôme François Alain Jean Rousseau**
Instituto de Matemática e Estatística - UFBA

**Prof. José Ferreira Alves**
Departamento de Matemática - FCUP

**Prof. Lucas Henrique Backes**
Departamento de Matemática - UFRGS

**Prof. Marcelo Miranda Viana da Silva**
Instituto de Matemática Pura e Aplicada - IMPA

**Prof. Maria Joana da Costa Cruz de Oliveira Torres**
Departamento de Matemática - UMINHO

**Prof. Rafael Oswaldo Ruggiero Rodriguez**
Departamento de Matemática – PUC-Rio

Rio de Janeiro, April the 15th, 2025

**Marcelo Durães Capeleiro Pinto**

Bachelor's degree in Mathematics by Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) in 2018. Master's degree in Mathematics by Pontifícia Universidade Católica do Rio de Janeiro (PUC Rio) in 2021 under the guidance of professor Silvius Klein. During the Ph.D. at PUC-Rio, also studied at Faculdade de Ciências da Universidade de Lisboa (FCUL). Research interests include Dynamical Systems, Ergodic Theory, Linear Cocycles and Regularity of the Lyapunov exponents.

Bibliographic data

Pinto, Marcelo Durães Capeleiro

Lyapunov Exponents of Random Linear Cocycles: Regularity and Statistical Properties / Marcelo Durães Capeleiro Pinto; advisor: Silvius Klein. – 2025.

149 f: il. color. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Matemática, 2025.

Inclui bibliografia

1. Matemática – Teses. 2. Sistemas Dinâmicos. 3. Teoria Ergódica. 4. Cociclos Lineares. 5. Expoentes de Lyapunov. 6. Operadores de Markov. 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 inspiring lessons.
To my students, who keep the motivation despite all the difficulties.

## Acknowledgments

Firstly, I would like to thank my parents Marcia and Antonio for their support through all my life, especially during the period of my doctorate.

I would like to thank my advisor Silvius Klein for his guidance during the last years. He introduced me to a lot of beautiful Mathematics and I am very grateful to have had the opportunity to work with him.

I would also like to thank professors Jérôme Rousseau, José Ferreira Alves, Lucas Backes, Marcelo Viana, Maria Joana Torres and Rafael Oswaldo Ruggiero for accepting to be part of this jury.

Moreover, I would like to thank Pedro Duarte and Tomé Graxinha, who made me feel very welcome during my stay in Lisbon. I have learned a lot from our talks at FCUL.

I would also like to thank my collaborators Ao Cai and Artur Amorim for the many discussions we had that ended up in fruitful joint works.

I would like to thank all my friends and professors from the math department of PUC-Rio, in particular, the dynamical systems group: Aline Melo, Anselmo Pontes, Anthony Guillen, Camila Crispim, Edhin Mamani, Guilherme Guglielmo, Kauane Araujo, Odylo Costa and Raul Rodriguez and professors Lorenzo Díaz, Rafael Ruggiero and Silvius Klein.

I would like to thank Aline Melo. I am very grateful for all the help and support during the last years.

I wish to thank the administrative staff of the mathematics department: Carlos, Kátia, Mariana and especially Creuza, who helped me a lot during my doctorate to fulfil the administrative and bureaucratic requirements.

Finally, I acknowledge CNPq for the financial support for this work.

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

## Abstract

Pinto, Marcelo Durães Capeleiro; Klein, Silvius (Advisor). **Lyapunov Exponents of Random Linear Cocycles: Regularity and Statistical Properties**. Rio de Janeiro, 2025. 149p. Tese de Doutorado – Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

This work is concerned with the study of the regularity and the statistical properties of Lyapunov exponents of random locally constant linear cocycles. We investigate both the case when the support of the underlying measure consists of only invertible matrices, as well as the case when it also contains non-invertible matrices. It turns out that these two settings exhibit strikingly different behaviors.

In the invertible case we study the regularity of the Lyapunov exponent as a function of the underlying measure relative to two different topologies. We establish its Hölder continuity in the generic setting with respect to the Wasserstein distance and its analyticity with respect to the total variation norm. In the non-invertible case, under appropriate assumptions, we obtain a characterization of uniform hyperbolicity via multi-cones and use it to establish a dichotomy between the analyticity and the discontinuity of the Lyapunov exponent. We also derive large deviations estimates and a central limit theorem for all of these models.

While there are many interesting remaining open problems, our results attempt to provide an almost complete picture in the context of two-dimensional random locally constant cocycles with finitely supported measures.



## Keywords

Dynamical Systems; Ergodic Theory; Linear Cocycles; Lyapunov Exponents; Markov Operators.

## Resumo

Pinto, Marcelo Durães Capeleiro; Klein, Silvius. **Expoentes de Lyapunov de Cociclos Lineares Aleatórios: Regularidade e Propriedades Estatísticas**. Rio de Janeiro, 2025. 149p. Tese de Doutorado – Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

Este trabalho estuda a regularidade e as propriedades estatísticas dos expoentes de Lyapunov de cociclos lineares aleatórios localmente constantes. Investigamos tanto o caso em que o suporte da medida subjacente consiste apenas em matrizes invertíveis, quanto o caso em que também contém matrizes não invertíveis. Esses dois cenários exibem comportamentos notavelmente diferentes.

No caso invertível, estudamos a regularidade do expoente de Lyapunov como função da medida subjacente em relação a duas topologias diferentes. Estabelecemos sua continuidade de Hölder no caso genérico em relação à distância de Wasserstein e sua analiticidade em relação à norma de variação total. No caso não invertível, sob hipóteses apropriadas, obtemos uma caracterização da hiperbolicidade uniforme por meio de multicones e a usamos para estabelecer uma dicotomia entre a analiticidade e a descontinuidade do expoente de Lyapunov. Também provamos estimativas de grandes desvios e um teorema central do limite para todos esses modelos.

Embora existam muitos problemas interessantes ainda em aberto, nossos resultados tentam fornecer uma imagem quase completa no contexto de cociclos aleatórios bidimensionais localmente constantes com medidas com suporte finito.

## Palavras-chave

Sistemas Dinâmicos; Teoria Ergódica; Cociclos Lineares; Expoentes de Lyapunov; Operadores de Markov.

# Table of contents

*The beauty of mathematics only shows itself to more patient followers.*

**Maryam Mirzakhani**, *Interview to The Guardian.*

# 1
# Introduction and Statements

## 1.1
## Lyapunov exponents and linear cocycles

Dynamical systems are an incredibly vast and active area of research in mathematics.

A dynamical system is a pair $(X, f)$, which consists of a (usually compact) metric space $X$ and a (usually continuous) transformation $f\colon X \to X$. Given any point $x \in X$, we can consider its iterates

$$x,\ f(x),\ f^2(x) = f(f(x)), \ldots, f^n(x) = f(f^{n-1}(x)),\ \ldots,$$

which we call the (forward) orbit of $x$. There are many different points of view from which one can study a dynamical system.

One way is trying to understand it from a topological point of view, by answering questions such as: is the orbit of a point $x$ periodic; is the orbit dense; are there any fixed points for $f$?

Another interesting approach is to try to understand the relation between the geometry of the space $X$ and the transformation $f$. If $X$ is a manifold, how does its curvature influence the behavior of the orbits? Does the geometry generate any rigidity property on the dynamics?

Furthermore, one can also look at the same problem from a probabilistic point of view, trying to understand the behavior of the orbits on average. Do most of them share similar properties? Do they satisfy any limit laws? This probabilistic approach is what we call ergodic theory and it lies at the heart of this thesis.

Let us take a detour and recall some probabilistic results about random additive processes. Consider a sequence of i.i.d. random variables $\{X_n\}_{n\geq 0}$. Let $S_n = X_0 + \cdots + X_{n-1}$ denote their partial sums. An important question concerns the behavior of the (arithmetic) average process $\frac{1}{n}S_n$ as $n \to \infty$.

The law of large numbers (LLN) says that this sequence of averages converges to the expected value of $X_0$ almost surely. In particular, it also converges in probability.

That is, for all $\varepsilon > 0$,

$$\mathbb{P}\left\{\left|\frac{1}{n}S_n - \mathbb{E}X_0\right| > \varepsilon\right\} \to 0 \quad \text{as } n \to \infty. \tag{1.1}$$

The next natural question is whether there is an explicit rate of convergence to 0 in equation (1.1). The large deviations principle (LDP) of Cramér says that under general conditions,

$$\mathbb{P}\left\{\left|\frac{1}{n}S_n - \mathbb{E}X_0\right| > \varepsilon\right\} \asymp e^{-c(\varepsilon)n} \quad \text{as } n \to \infty, \tag{1.2}$$

where $c(\varepsilon) \approx c_0\varepsilon^2$, for some $c_0 > 0$.

Note that the previous result is of an asymptotic type. Let us now describe Hoeffding's inequality, an effective, non-asymptotic (or finitary) large deviations type (LDT) estimate. If $\left|X_0\right| \leq c$ almost surely, then the inequality

$$\mathbb{P}\left\{\left|\frac{1}{n}S_n - \mathbb{E}X_0\right| > \varepsilon\right\} \leq 2e^{-(2c)^{-2}\varepsilon^2 n} \tag{1.3}$$

holds for all $n \in \mathbb{N}$.

Finally, one can ask what is the typical size of the difference $S_n - n\mathbb{E}X_0$. The central limit theorem (CLT) says that in a certain sense, $S_n - n\mathbb{E}X_0 \asymp \sqrt{n}$. More precisely, if $\mathbb{E}X_0^2 < \infty$ and if $X_0$ is not almost surely constant, then we have the following convergence in law to a Gaussian distribution:

$$\frac{S_n - n\mathbb{E}X_0}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma^2). \tag{1.4}$$

Let us describe in probabilistic terms the central element of this work, which is the multiplicative analogue of the law of large numbers. Consider a multiplicative (semi)group of $d \times d$ matrices $G$, for instance $\mathrm{SL}_d(\mathbb{R})$, or $\mathrm{GL}_{\mathrm{d}}(\mathbb{R})$ or $\mathrm{Mat}_2^+(\mathbb{R})$, the set of $d \times d$ matrices with nonnegative determinant. Let $\mu$ be a probability measure over $G$, which we will assume to have compact support.

Let $g_0, g_1, \ldots, g_n, \ldots$ be an i.i.d. sequence of matrices in $G$, chosen according to the distribution $\mu$. Let $\Pi_n := g_{n-1} \cdots g_1\, g_0$ denote the corresponding multiplicative process and consider the geometric average $\|\Pi_n\|^{1/n}$. This product of random matrices typically grows exponentially, therefore we consider the logarithm of this expression.

By the celebrated theorem of Furstenberg and Kesten [28], under a general integrability condition, this average process converges almost surely to a constant. More precisely, with full probability,

$$L_1(\mu) = \lim_{n\to\infty} \frac{1}{n} \log\|g_{n-1} \cdots g_1 g_0\| \tag{1.5}$$

where the almost sure limit $L_1(\mu)$ is called the first or the top *Lyapunov exponent* of the process.

Replacing the norm (that is, the first singular value) of the matrices $\Pi_n$ by their second, their third and so on singular values, we obtain, respectively, the second Lyapunov exponent $L_2(\mu)$, the third Lyapunov exponent $L_3(\mu)$ and so on (until de dimension $d$).

An interesting question is whether there are also multiplicative analogues of the other statistical properties, such as large deviations estimates and a central limit theorem. Although the theory of additive processes is well developed, our understanding of the theory of multiplicative random processes is still much more limited. Part of this work is dedicated to exploring some new settings where we can establish LDT estimates and central limit theorems.

A difficult and important problem in the study of multiplicative processes concerns the regularity of the Lyapunov exponent, the limit quantity in (1.5), as a function of the input data. More precisely, the first Lyapunov exponent is a function of the probability measure $\mu$, so the question is how would a small perturbation of $\mu$ affect the limit $L_1(\mu)$. This is a very subtle question that has motivated a lot of research from renowned mathematicians over the last decades. It turns out that the type of perturbation considered, the corresponding topology, will greatly influence the answer.

Some classical results in this direction are due to Furstenberg and Kifer [29] who proved the continuity of the Lyapunov exponents under some generic conditions, to Le Page [38], who proved a Hölder modulus of continuity, also in the generic setting and to Peres [39], who established the analyticity of the Lyapunov exponent with respect to the transition probabilities when the support of $\mu$ is a finite set.

A recent result, considered to be a breakthrough in this theory, is due to Avila, Eskin and Viana [3] and it establishes the continuity of the Lyapunov exponents of random matrix products in any dimension and without any further genericity conditions compared to the classical result of Furstenberg and Kifer.

The concepts and problems mentioned previously were described in probabilistic terms. It turns out that they can be studied in a much more general, dynamical systems framework, that of *linear cocycles*. A linear cocycle is a skew product transformation $F \colon M \times \mathbb{R}^d \to M \times \mathbb{R}^d$ determined by a pair $(f, A)$, where $f \colon M \to M$ is a base transformation (usually assumed to be ergodic), $A \colon M \to G$ is a measurable fiber map and

$$F(x, v) = (f(x), A(x)v).$$

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

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

Moreover, by Furstenberg-Kesten's theorem (or by a more general result, the Kingman subadditive ergodic theorem), the limit

$$L_1(F) = \lim_{n\to\infty} \frac{1}{n} \log\|A^n(x)\|$$

exists (and it is constant) for $\mu$-a.e. $x \in M$ and it is called the first Lyapunov exponent of the cocycle $F$.

Random products of matrices fit this general framework as follows. Consider a measure $\mu$ on a semigroup $G$ of $d \times d$ matrices. Assume that the support $\Sigma$ of this measure is compact and let $M$ be the space of sequences $\Sigma^{\mathbb{Z}}$, equipped with the product measure $\mu^{\mathbb{Z}}$. Let $A\colon M \to G$ be the projection of a sequence $\underline{g} = \{g_n\}_{n\in\mathbb{Z}}$ to its zeroth coordinate $g_0$, let $\sigma\colon M \to M$ be the Bernoulli shift $\sigma(\underline{g}) = \{g_{n+1}\}_{n\in\mathbb{Z}}$, and let $F$ be the linear cocycle corresponding to the pair $(\sigma, A)$. Then the $n$-th iterate of the fiber map $A^n(\underline{g})$ is exactly the random product of matrices $g_{n-1} \dots g_1\, g_0$ and the limiting quantity $L_1(F) = L_1(\mu)$. Such a cocycle will be referred to as a *random linear cocycle* as its fiber iterates encode products of i.i.d. matrices.

Note that, a priori, the map $A$ could be chosen to be much more general. However, a celebrated theorem of Mañé-Bochi [9] says that given any measure preserving dynamical system $(M, f, \mu)$ where $f$ is an aperiodic (meaning its set of periodic points has measure zero) homeomorphism, for any fiber map $A\colon M \to \mathrm{SL}_2(\mathbb{R})$, either the cocycle associated with $A$ is hyperbolic over the support of $\mu$ or else $A$ is approximated in the $C^0$ topology by fiber maps with zero Lyapunov exponent. Regarding the Lyapunov exponent as a function of the fiber map, it follows that if $L_1(A) > 0$, then the first Lyapunov exponent is either analytic (this holds by a theorem of Ruelle [40] when $A$ is hyperbolic) or it is discontinuous at $A$ in the $C^0$ topology of $\mathrm{SL}_2(\mathbb{R})$-valued cocycles. In other words, the regularity of the first Lyapunov exponent exhibits a strongly dichotomic behavior, analyticity versus discontinuity.

Therefore, in order to establish regularity properties of the Lyapunov exponent (when the cocycle is not uniformly hyperbolic), the map $A$ is often assumed to be highly regular. That is also why we will restrict our attention to the case when $A$ is the projection of a sequence of matrices to the zeroth coordinate, in other words $A$ is a locally constant map and the corresponding cocycle is called a random, *locally constant* linear cocycle. All of the results in this manuscript are concerned with this model, which for simplicity will be referred to only as a random cocycle.

Another commonly assumed hypothesis is that the semigroup $G$ where the fiber map takes values consists only of invertible matrices. All of the three classical results that we cited before, as well as [3], fit into this setting, that of locally constant linear cocycles determined by invertible matrices. The reason for this assumption is made more apparent by example 6.1, which also appears in both [20] and [3]. In this example the support of the measure $\mu$ consists of two matrices, one hyperbolic, hence invertible and the other a projection, which, in particular, is singular (it has zero determinant). A straightforward computation shows that the Lyapunov exponent is discontinuous at $\mu$, when regarded as a function of the matrices determining the fiber map. This example will be explained in more details in chapter 6. Moreover, in [3], Avila, Eskin and Viana stated that “there is no hope to obtain any general regularity result for Lyapunov exponents in this more general setting” (of non invertible matrices).

Although the last paragraph suggests that we should restrict ourselves to the classical setting of invertible matrices, this manuscript also explores the case in which the cocycle admits singular matrices. Surprisingly, it turns out that this setting admits a very rich theory, while at the same time we can confirm that the previous quote is indeed correct, as it will be explained later.

## 1.2 Main results

We are now ready to formulate some of the main results of this manuscript. We start with the study of the regularity of the map $\mu \mapsto L_1(\mu)$ for $\mathrm{GL_d}(\mathbb{R})$-valued cocycles, relative to two different topologies: the weak star and the one induced by the total variation norm.

We will use a certain concept of irreducibility for random linear cocycles. Irreducibility usually refers to the *non-existence* of a proper, invariant sub-bundle for the skew-product dynamics. We will need a slightly weaker property, the quasi-irreducibility, which may allow the existence of such proper, invariant sub-bundle, as long as the first Lyapunov exponent along it coincides with the first Lyapunov exponent $L_1(\mu)$ on the entire space.

Relative to the weak star topology, we prove a more general version of Le Page’s theorem, based on the author’s Master’s thesis [27] and on [4]. In what follows, $W_1$ refers to the Wasserstein’s metric, a distance in the space of measures that metrizes the weak star topology. Precise definitions will be given in chapter 4.

**Theorem 1.1** *Let* $\Sigma \subset \mathrm{GL_d}(\mathbb{R})$ *be a compact set and let* $\mu \in \mathrm{Prob}(\Sigma)$. *Assume that* $\mu$ *is quasi-irreducible and that* $L_1(\mu) > L_2(\mu)$. *Then there exist* $\delta > 0$, $C > 0$ *and* $\alpha \in (0, 1]$ *such that given* $\mu_1, \mu_2 \in \mathrm{Prob}(\Sigma)$ *satisfying* $W_1(\mu_i, \mu) < \delta$ *for* $i \in \{1, 2\}$, *we have that*

$$|L_1(\mu_1) - L_1(\mu_2)| \leq CW_1(\mu_1, \mu_2)^\alpha.$$

Relative to the total variation norm, we generalize the result of Peres [39] proving the analyticity of the Lyapunov exponent, from probability measures with finite support to probability measures with compact, possibly infinite support. Precise definitions of the total variation norm and the concept of analyticity in this setting are given in chapter 5. This theorem is part of a joint work with Amorim and Melo [1]. We obtained two results, one that assumes the quasi-irreducibility of the measure/cocycle, and one without this assumption, but instead the measure has to have full support.

**Theorem 1.2** *Let* $\Sigma \subset \mathrm{GL}_d(\mathbb{R})$ *be a compact set, let* $\mu_0 \in \mathrm{Prob}(\Sigma)$ *and assume that* $L_1(\mu_0) > L_2(\mu_0)$.

*(1) If* $\mu_0$ *is quasi-irreducible, then the map* $\mathrm{Prob}(\Sigma) \ni \mu \mapsto L_1(\mu)$ *is real analytic with respect to the total variation norm in a neighborhood of* $\mu_0$.

*(2) If* $\mathrm{supp}(\mu_0) = \Sigma$, *then the map* $\mathrm{Prob}(\Sigma) \ni \mu \mapsto L_1(\mu)$ *is real analytic with respect to the total variation norm in a neighborhood of* $\mu_0$.

The next results are part of two projects with Duarte, Graxinha and Klein: [18] and [19]. We study cocycles in $\mathrm{Mat}_2^+(\mathbb{R})^k$, the semigroup of two dimensional matrices with non-negative determinant. The starting point of these projects is a generalization of the works of Yoccoz [44] and Avila, Bochi and Yoccoz [2] providing a characterization of the dominated splitting property (also referred to as projective uniform hyperbolicity) in terms of multi-cones. We extend the scope of the results in [44], [2] from $\mathrm{SL}_2(\mathbb{R})$-valued cocycles to $\mathrm{Mat}_2^+(\mathbb{R})$-valued cocycles.

We fix a probability vector $p = (p_1, \dots, p_k)$ with $\sum_{i=1}^k p_i = 1$ and $p_i > 0$ for all indices $i$ and consider random linear cocycles with finitely supported measures, thus determined by a $k$-tuple $\underline{A} = (A_1, \dots, A_k) \in \mathrm{Mat}_2^+(\mathbb{R})^k$ and by the probability vector $p$. We identify the cocycle with the tuple $\underline{A}$ that determines it.

Let $\mathbb{P}(\mathbb{R}^2)$ denote the real projective line. An invariant multi-cone for such a cocycle $\underline{A}$ is an open subset of $M \subset \mathbb{P}(\mathbb{R}^2)$ such that $\bar{M} \neq \mathbb{P}(\mathbb{R}^2)$ and $\overline{A_i M} \subset M$ for every index $1 \leq i \leq k$.

**Theorem 1.3** *Given a random linear cocycle* $\underline{A} = (A_1, \ldots, A_k) \in \mathrm{Mat}_2^+(\mathbb{R})^k$, $\underline{A}$ *is projectively uniformly hyperbolic if and only if* $\underline{A}$ *admits an invariant multi-cone.*

This characterization is one of the main tools in the study of rank one cocycles. We say that $\underline{A} = (A_1, \ldots, A_k) \in \mathrm{Mat}_2^+(\mathbb{R})^k$ has rank 1 if there is at least one index $j$ such that $A_j$ is singular (i.e. non-invertible) and there are no null words (i.e. finite products of matrices).

The next theorem has the flavor of Mañé-Bochi's dichotomy. Surprisingly, this dichotomy holds in the high regularity setting of *locally constant* random linear cocycles, rather than the $C^0$-topology of Mañé-Bochi's celebrated result.

**Theorem 1.4** *Let* $\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k$ *be a locally constant random linear cocycle of rank* 1*. Then either* $\underline{A}$ *is projectively uniformly hyperbolic or else there exists a sequence of random linear cocycles* $\{\underline{A}_n\}_n \to \underline{A}$ *such that* $\underline{A}_n$ *has a null word and, in particular,* $L_1(\underline{A}_n) = -\infty$ *for every* $n \in \mathbb{N}$.

Moreover, using Ruelle's theorem [40], we conclude the following.

**Corollary 1.1** *Given* $\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k$ *with at least one singular and one invertible component, if* $L_1(\underline{A}) > -\infty$ *then the following dichotomy holds: either the Lyapunov exponent* $L_1$ *is analytic at* $\underline{A}$ *or it is discontinuous at* $\underline{A}$.

An argument involving subharmonic functions (see Corollary 6.3) shows that in fact $L_1(\underline{A}) > -\infty$ holds for Lebesgue almost every $\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k$ with at least one singular and one invertible component. Therefore, the previous corollary applies to almost every cocycle and the aforementioned quote from [3] proves to be very accurate: if the cocycle is not projectively uniformly hyperbolic, then it is a discontinuity point of the Lyapunov exponent.

Furthermore, using a topological argument, the set of continuity points of $L_1$ is a Baire residual subset of $\mathrm{Mat}_2^+(\mathbb{R})^k$.

Putting together other recent results on random two dimensional cocycles in finite symbols, we obtain the following almost complete picture on the regularity of their first Lyapunov exponent $L_1$.

At any invertible cocycle $\underline{A}$, $L_1$ is a continuous function on $\mathrm{GL_d(\mathbb{R})^k}$ (an open set in $\mathrm{Mat}_d(\mathbb{R})^k$) and, moreover, its regularity varies from log-Hölder continuous[1] to analytic (see [40], [11], [35], [41], [25], [22], [24]).

In the algebraic variety

$$\mathcal{R}_1 = \left\{\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k \colon \mathrm{rank}(A_i) = 1 \ \forall\, 1 \leq i \leq k\right\}$$

[1]Given a metric space $(M, d)$, a function $\phi\colon M \to \mathbb{R}$ is said to be log-Hölder continuous if $\left|\phi(x) - \phi(y)\right| \leq C \left(\log \frac{1}{d(x,y)}\right)^{-1}$ for some $C < \infty$ and all $x, y \in M$.

the map $L_1$ is always continuous. Moreover, every $\underline{A} \in \mathcal{R}_1$ is a continuity point of the Lyapunov exponent $L_1$ in $\mathrm{Mat}_2^+(\mathbb{R})^k$. Furthermore, (see Theorem 6.4) a cocycle $\underline{A} \in \mathcal{R}_1$ is projectively uniformly hyperbolic if and only if $L_1(\underline{A}) > -\infty$, which is equivalent to the absence of null words (i.e. vanishing finite products of components of $\underline{A}$). This latter condition holds for almost every such cocycle. Therefore $L_1$ is analytic at Lebesgue almost every $\underline{A} \in \mathcal{R}_1$.

In the remaining case, given a cocycle $\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k$ with at least one invertible and one non-invertible component, either $L_1$ is analytic or it is discontinuous at $\underline{A}$.

Consider now the problem of statistical properties (large deviations and the central limit theorem) in the singular setting. We obtain the following results.

**Theorem 1.5** *For Lebesgue almost every cocycle $\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k$ with at least one singular and one invertible component and for every $\varepsilon > 0$ it holds that*

$$\mathbb{P}\left\{\left|\frac{1}{n}\log\|A^n\| - L_1(\underline{A})\right| > \varepsilon\right\} \leq C\, e^{-c_0(\varepsilon)\, n^{1/3}}$$

*where $C < \infty$, $c_0(\varepsilon) > 0$ is an explicit function of $\varepsilon$ and $\mathbb{P} = p^{\mathbb{Z}}$ refers to the Bernoulli measure on the space of sequences of matrices.*

**Theorem 1.6** *For Lebesgue almost every cocycle $\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k$ with at least one singular and one invertible component, there exists $\sigma > 0$ such that the following convergence in distribution to the normalized Gaussian holds:*

$$\frac{\log\|A^n\| - n\, L_1(\underline{A})}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)\,.$$

We note that these statistical properties are sensitive to perturbations of the cocycle, that is, the parameters appearing in these estimates are not uniform in a neighborhood of $\underline{A}$. This is unlike the case of invertible matrices, where the LDT estimates are uniform in the data, something that can be used directly to deduce a modulus of continuity of the Lyapunov exponent, see [21, 22].

The table below summarizes what it is known regarding the minimal regularity of the Lyapunov exponent (R-LE), namely its modulus of continuity[2]

[2]Given a metric space $(M,d)$, a function $\phi\colon M \to \mathbb{R}$ is said to be weak-Hölder continuous if $\left|\phi(x) - \phi(y)\right| \leq C\exp\left(-\alpha\,\log^b \frac{1}{d(x,y)}\right)$ for some $C < \infty, \alpha, b \in (0,1]$ and all $x, y \in M$. When $b = 1$, this corresponds to $\alpha$-Hölder continuity.
Moreover, if $\left|\phi(a) - \phi(x)\right| \leq C\, d(a,x)^\alpha$ holds for a given point $a$ and all $x$, we call $\phi$ pointwise Hölder at $a$.

or whether it is discontinuous (Disc.) as well as the availability of large deviations type (LDT) estimates and of a central limit theorem (CLT) for invertible or $\mathrm{Mat}_2^+(\mathbb{R})$-valued Bernoulli cocycles $\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k$, $k \geq 2$. For the purpose of this table, we assume that $L_1 > L_2$ (so $L_1 > -\infty$ as well). There are three possibilities for such a given cocycle $\underline{A}$: rank $= 2$, meaning its components are all invertible; rank $= 1$, meaning its components are all singular; rank $= 1\&2$, where some components are singular and some are invertible with positive determinant. The last two cases are treated in chapters 6 and 7.

| | R-LE | LDT | CLT |
|---|---|---|---|
| rank $= 2$ | (Weak) Hölder[(a)] | Yes[(b)] | Yes[(c)] |
| rank $= 1$ | $C^\omega$ (Cor. 6.4) | Yes (Rmk. 7.6) | Yes (Rmk. 7.6) |
| rank $= 1\&2$ | Disc. (Cor. 1.1) | Yes (Thm. 1.5) | Yes (Thm. 1.6) |

[(a)] Locally Hölder for quasi-irreducible cocycles [38], [21]; locally weak-Hölder in the remaining case [24]; pointwise Hölder always [41].

[(b)] Locally uniform LDT of exponential type in the quasi-irreducible case [21]; locally uniform LDT of sub-exponential type in the remaining case [24]; non-uniform LDT of exponential type holds always [24].

[(c)] See [37] and [6].

Table 1.1: Random two dimensional cocycles.

Recall that in the remaining cases, if the cocycle is projectively uniformly hyperbolic then the Lyapunov exponent is analytic, while when $L_1 = L_2$ the Lyapunov exponent is automatically continuous and in fact pointwise log-Hölder continuous (see [41]) in the invertible case.

Moreover, in chapter 8, we state that the results on the second and third lines of the table also hold for the more general setting of mixing Markov cocycles. Furthermore, in this setting, the results on the first line are available only in the generic (irreducible) case (see [21, Chapter 5]), but we expect their analogues from the Bernoulli setting to still hold without the generic assumption.

Finally, it is also worth mentioning theorems 8.1 and 8.3, which are part of a joint work with Cai, Klein and Melo, where we prove a Markovian analogue of both Furstenberg-Kifer's multiplicative ergodic theorem and Le Page's theorem. In the interest of the readability of the manuscript, we chose to present all the results of this manuscript and their proofs in the i.i.d.

(rather than the Markovian) setting, therefore the proofs of these theorems were omitted, but they can be found in [15].

The work presented in this manuscript uses concepts, methods and tools from many different fields, including hyperbolic dynamics, ergodic theory, probabilities, spectral theory, holomorphic functions in Banach spaces, potential theory (the theory of subharmonic functions), as well as parameter elimination arguments.

The rest of this manuscript is organized as follows. In chapter 2 we introduce some of the main elements present in the text: the Markov operators and the stationary measures, as well as classical results that will be used all throughout. In chapter 3 we give a detailed proof of Furstenberg-Kifer's multiplicative ergodic theorem, also known as Furstenberg-Kifer's non-random filtration. This is a key result in the study of random linear cocycles, which appears in essentially every work related to the regularity of the Lyapunov exponents of random linear cocycles. In chapter 4 we prove theorem 1.1. For completion, we also included a section with the proof of statistical properties for this model. In chapter 5 we prove theorem 1.2. In chapter 6 we study the regularity of the Lyapunov exponents for cocycles in $\mathrm{Mat}_2^+(\mathbb{R})$ and we prove theorems 1.3 and 1.4. In chapter 7 we study the statistical properties for cocycles in $\mathrm{Mat}_2^+(\mathbb{R})$ and we prove theorems 1.5 and 1.6. In chapter 8 we discuss how we can generalize the results in the previous chapters to Markov cocycles and we also formulate some further problems and questions related to the singular setting.

# 2
# Markov Operators and Stationary Measures

In this chapter we introduce the main elements that will appear throughout the text: the Markov operator and the stationary measure. They are the foundation for all the results in this thesis.

In section 2.1 we define the Markov operator and explain why it is an essential tool in the study of the Lyapunov exponents of random linear cocycles. In section 2.2 we study three types of dynamics related to linear cocycles, one deterministic and two stochastic, and explain the relation between the invariant measure (for the deterministic dynamics) and the stationary measure (for the stochastic dynamics). This will be very useful later, since we will be able to switch between these settings and choose which one is more adequate to work with, according to the specific problem. In section 2.3 we present two abstract theorems that are used to establish statistical properties (large deviations and central limit theorem) for Lyapunov exponents. Therefore, whenever we plan to prove such properties, we have to verify that the hypotheses of those theorems are fulfilled.

## 2.1
## The Markov operator

### 2.1.1
### Stochastic dynamical systems and Markov kernels

A *deterministic dynamical system* (DDS) is defined by a pair $(M, f)$, where $M$ is a compact metric space and $f\colon M \to M$ is a continuous transformation that acts on $M$. In this setting, for each point $x \in M$, the law $f$ determines its trajectory:

$$x \to f(x) \to f^2(x) \to \cdots \to f^n(x) \to \ldots$$

We denote by $\text{Prob}(M)$ the set of all probability measures over $M$. Endowed with the weak-star topology, $\text{Prob}(M)$ is a compact, metrizable topological space.

A probability measure $\mu \in \text{Prob}(M)$ is called $f$-invariant if $f_*\mu = \mu$, where $f_*\mu$ is the push-forward measure defined by $f_*\mu(E) = \mu(f^{-1}(E))$ for all Borel sets $E \subset M$. It is easy to verify that $\mu$ is $f$-invariant if and only if $\mu = \int_M \delta_{f(x)}\, d\mu(x)$.

A *stochastic dynamical system* (SDS) is a pair $(M, K)$, where $M$ is still a compact metric space and $K\colon M \to \mathrm{Prob}(M)$ is a continuous map which associates to each point $x \in M$ a probability measure $K_x \in \mathrm{Prob}(M)$. This map is called a Markov (or a transition) kernel. Given any Borel set $E \subset M$, the number $K_x(E)$ denotes the probability that $x$ transitions to $E$.

One can also consider the iterated kernels $K^n\colon M \to \mathrm{Prob}(M)$ defined inductively for every $n \in \mathbb{N}$ as follows. $K_x^1 = K_x$, $K_x^2 = \int_M K_y\, dK_x(y)$ and $K_x^{n+1} = \int K_y^n\, dK_x(y)$. An intuitive meaning of $K_x^n(E)$ is that it represents the probability that $x$ transitions to $E$ in $n$ steps.

**Remark 2.1** *Note that given any DDS $(M, f)$ one can define an SDS via the following Markov kernel: $K_x = \delta_{f(x)}$. Thus every deterministic dynamical system is also a stochastic dynamical system.*

**Definition 2.1** *A measure $\mu \in \mathrm{Prob}(M)$ is called $K$-stationary if it satisfies*

$$\mu = \int_M K_x\, d\mu(x).$$

*We also write the equation above as $\mu = K * \mu$.*

It is important to note that stationary measures do exist. Let $\pi \in \mathrm{Prob}(M)$ and consider

$$\pi_n\colon = \frac{1}{n}\sum_{j=1}^{n} K_x^j * \pi$$

which belongs to $\mathrm{Prob}(M)$. Since we assume $M$ to be compact, $\mathrm{Prob}(M)$ is weak star compact, thus there exists a subsequence $\{\pi_{n_j}\}_j$ that converges to some measure $\mu \in \mathrm{Prob}(M)$ which, by construction, is $K$-stationary.

The triplet $(M, K, \mu)$, where $M$ is a compact metric space, $K$ a continuous Markov kernel and $\mu$ a $K$-stationary measure is called a *Markov system*.

### 2.1.2 Markov operators

Let $L^\infty(M)$ denote the set of measurable and bounded functions. Given a Markov kernel $K$, we associate to it a Markov operator $Q = Q_K$, defined as

$$Q\colon L^\infty(M) \to L^\infty(M)$$
$$Q\varphi(x) = \int_M \varphi(y)\, dK_x(y).$$

Note that $Q$ satisfies the following properties:

(i) Constant functions are invariant under $Q$.

(ii) $Q$ is a bounded linear operator.

(iii) If $\varphi \geq 0$, then $Q\varphi \geq 0$.

Operators that satisfy the third property are called *positive*, so the Markov operator is a positive operator taking constant functions to constant functions.

Moreover, note that $\mu$ is a $K$-stationary measure if and only if

$$\int_M Q\varphi \, d\mu = \int_M \varphi \, d\mu \quad \text{for every } \varphi \in L^\infty(M). \tag{2.1}$$

This characterization is very useful and will be used many times throughout the text.

**Lemma 2.1** *Given a Markov kernel* $K\colon M \to \mathrm{Prob}(M)$*, the following relation holds*

$$Q_{K^n} = Q_K^n.$$

*Proof.* Let $\varphi \in L^\infty(M)$. The relation is trivial when $n = 1$. For $n = 2$

$$\begin{aligned} Q_K^2\varphi(x) &= \int_M \int_M \varphi(z) \, dK_y(z) dK_x(y) \\ &= \int_M \varphi(z) \, dK_x^2(z) = Q_{K^2}(x) \end{aligned}$$

The proof follows by induction. ■

### 2.1.3 Random linear cocycles

Consider the quadruple $(M, f, \mathcal{B}, \mu)$, where $M$ is a compact metric space, $\mathcal{B}$ is a sigma algebra, $f\colon M \to M$ is a measurable transformation and $\mu$ is a probability measure. We say that $(M, f, \mathcal{B}, \mu)$ is a measure preserving dynamical system if $\mu(f^{-1}(E)) = \mu(E)$ holds for every $E \in \mathcal{B}$.

Moreover, an ergodic dynamical system is a measure preserving dynamical system such that if $E$ is an $f$-invariant set (meaning that $f^{-1}(E) = E$) then $\mu(E) = 0$ or $\mu(E) = 1$.

Given an ergodic dynamical system $(M, f, \mathcal{B}, \mu)$, we recall the concept of *Birkhoff sums*. Let $\varphi\colon M \to \mathbb{R}$ be a bounded observable and let $n \geq 1$. The $n$-th Birkhoff sum of $\varphi$, which we denote by $S_n\varphi$ is defined as follows:

$$\begin{aligned} &S_n\varphi\colon M \to \mathbb{R} \\ &S_n\varphi(x) = \varphi(x) + \varphi(f(x)) + \cdots + \varphi(f^{n-1}(x)). \end{aligned}$$

Furthermore, by Birkhoff's ergodic theorem,

$$\frac{1}{n}S_n\varphi \to \int_M \varphi \, d\mu \quad \mu-\text{a.e.}$$

We will introduce the stochastic analogue of Birkhoff sums, but in order to present this concept, we recall the notion of Markov chains. Let $(\Omega, \mathcal{B}, \mathbb{P})$ be a probability space. A *Markov chain* is a sequence $\{Z_n\}_{n\geq 0}$ of random variables with values in $M$ that satisfy the Markov property. More precisely, $Z_n : \Omega \to M$ is such that for every $E \in \mathcal{B}$,

$$\mathbb{P}(Z_{n+1} \in E | Z_n, \dots \; Z_0) = \mathbb{P}(Z_{n+1} \in E | Z_n).$$

Given a Markov kernel $K\colon M \to \text{Prob}(M)$ and a probability measure $\pi \in \text{Prob}(M)$, we say that $\{Z_n\}_{n\geq 0}$ is a Markov chain with transition $K$ and initial distribution $\pi$ if $\forall E \in \mathcal{B}$, $\forall x \in M$ and $\forall n \in \mathbb{N}$ the following hold:

(i) $\mathbb{P}(Z_0 \in E) = \pi(E)$,

(ii) $\mathbb{P}(Z_{n+1} \in E | Z_n = x) = K_x(E)$.

Given a Markov chain $\{Z_n\}_{n\geq 0}$ and an observable $\varphi\colon M \to \mathbb{R}$, we define the $n$-th *stochastic Birkhoff sum* of $\varphi$ as

$$S_n\varphi = \varphi \circ Z_0 + \varphi \circ Z_1 + \cdots + \varphi \circ Z_{n-1}.$$

An important example that illustrates the previous concept is the study of the stochastic dynamical system associated to a linear cocycle.

Let $\mu \in \text{Prob}(\text{GL}_\text{d}(\mathbb{R}))$ be a probability measure, where $\text{supp}(\mu) = \Sigma$ is a compact set. The measure $\mu$ determines the random linear cocycle (that is, the locally constant linear cocycle over the Bernoulli shift) $F\colon \Sigma^{\mathbb{N}} \times \mathbb{R}^d \to \Sigma^{\mathbb{N}} \times \mathbb{R}^d$ given by

$$F(\underline{g}, v) \colon = (\sigma \underline{g}, g_0 v),$$

where $\underline{g} = \{g_n\}_{n\in\mathbb{N}}$ is a sequence in $\Sigma^{\mathbb{N}}$ and $\sigma\colon \Sigma^{\mathbb{N}} \to \Sigma^{\mathbb{N}}$ is the Bernoulli shift. The iterates of $F$ are given by $F^n(\underline{g}, v) = (\sigma^n \underline{g}, g_{n-1} \dots g_0 v)$.

A natural SDS associated to this linear cocycle is obtained by choosing $M = \Sigma \times \mathbb{S}^{d-1}$ and $K\colon \Sigma \times \mathbb{S}^{d-1} \to \text{Prob}(\Sigma \times \mathbb{S}^{d-1})$ such that $K_{(g_0,v)} = \mu \times \delta_{\frac{g_0 v}{\|g_0 v\|}}$. This Markov kernel induces the following Markov chain:

$$Z_0(\underline{g}, v) = (g_0, v) \to Z_1(\underline{g}, v) = \left(g_1, \frac{g_0 v}{\|g_0 v\|}\right) \to Z_2(\underline{g}, v) = \left(g_2, \frac{g_1 g_0 v}{\|g_1 g_0 v\|}\right) \to \dots$$

Moreover, once we have the Markov kernel, we can explicitly write the associated Markov operator $Q\colon L^\infty(\Sigma \times \mathbb{S}^{d-1}) \to L^\infty(\Sigma \times \mathbb{S}^{d-1})$

$$Q\varphi(g_0, v) = \int_M \varphi\left(g_1, \frac{g_0 v}{\|g_0 v\|}\right) \, d\mu(g_1).$$

An important observable in this setting is $\Phi\colon \Sigma \times \mathbb{S}^{d-1} \to \mathbb{R}$ given by

$$\Phi(g_0, v) = \log\|g_0 v\|.$$

Let us describe the stochastic Birkhoff sum $S_n\Phi(\underline{g}, v)$:

$$\begin{aligned} S_n\Phi(\underline{g}, v) &= \Phi(g_0, v) + \Phi\left(g_1, \frac{g_0 v}{\|g_0 v\|}\right) + \cdots + \Phi\left(g_{n-1}, \frac{g_{n-1}\cdots g_0 v}{\|g_{n-1\ldots g_0 v}\|}\right) \\ &= \log\|g_0 v\| + \log\frac{\|g_1 g_0 v\|}{\|g_0 v\|} + \cdots + \log\frac{\|g_{n-1}\cdots g_0 v\|}{\|g_{n-2}\cdots g_0 v\|} \\ &= \log\|g_{n-1}\cdots g_0 v\|. \end{aligned}$$

Therefore, we conclude that the study of the properties of random linear cocycles and their Lyapunov exponents can be reduced to the study of an associated Markov system.

## 2.2 Stationary measures

We denote by $\mathrm{Prob}_K(M)$ the set of $K$-stationary measures on $M$. Since to each Markov kernel $K$ we can also associate a Markov operator $Q$, we also use the notation $\mathrm{Prob}_Q(M)$ to denote this set.

### 2.2.1 The three levels of dynamics

Consider the linear cocycle that was introduced in the previous section $F\colon \Sigma^{\mathbb{N}} \times \mathbb{R}^d \to \Sigma^{\mathbb{N}} \times \mathbb{R}^d$ such that

$$F(\underline{g}, v) := (\sigma\underline{g}, g_0 v),$$

where $\underline{g} = \{g_n\}_{n\in\mathbb{N}}$ is a sequence in $\Sigma^{\mathbb{N}}$ and $\sigma : \Sigma^{\mathbb{N}} \to \Sigma^{\mathbb{N}}$ is the Bernoulli shift.

This is an example of a DDS. We already discussed an example of a natural SDS associated to it, that will be very useful in the study of Lyapunov exponents. We will introduce other natural dynamical systems associated to this linear cocycle, which we divide into three different levels.

(i) In the first level, more similar to the original cocycle, we have the projective (or projectivized) cocycle, which is also a deterministic dynamical system (DDS). It is important to note that it acts on $\Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^d)$, a compact space, where $\mathbb{P}(\mathbb{R}^d)$ is the projective space over $\mathbb{R}^d$, and we denote by $\hat{v}$ the projective point corresponding to the nonzero vector $v \in \mathbb{R}^d$. This skew-product dynamical system is defined as

$$\hat{F}\colon \Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^d) \to \Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^d)$$
$$(\underline{g}, \hat{v}) \mapsto (\sigma \underline{g}, \hat{g}_0 \hat{v}),$$

where given a matrix (linear map) $g \in \mathrm{GL_d}(\mathbb{R})$, $\hat{g}$ denotes the induced projective map on $\mathbb{P}(\mathbb{R}^d)$, namely $\hat{g}\hat{v} = \widehat{g\,v}$.

(ii) In the second level we define a stochastic dynamical system (SDS) as follows. The ambient compact space is $\Sigma \times \mathbb{P}(\mathbb{R}^d)$ and the transition kernel is

$$\bar{K}\colon \Sigma \times \mathbb{P}(\mathbb{R}^d) \to \mathrm{Prob}(\Sigma \times \mathbb{P}(\mathbb{R}^d))$$
$$(g_0, \hat{v}) \mapsto \mu \times \delta_{\hat{g}_0 \hat{v}}.$$

Its corresponding Markov operator is

$$\bar{Q}\colon L^\infty(\Sigma \times \mathbb{P}(\mathbb{R}^d)) \to L^\infty(\Sigma \times \mathbb{P}(\mathbb{R}^d))$$
$$\bar{Q}\varphi(g_0, \hat{v}) = \int_{\Sigma \times \mathbb{P}(\mathbb{R}^d)} \varphi(\hat{g}_1, \hat{g}_0 \hat{v})\, d\mu(g_1).$$

Moreover, consider the Markov chain $\{\bar{Z}_n\}_{n \geq 0}$ given by:

$$\bar{Z}_n\colon \Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^d) \to \Sigma \times \mathbb{P}(\mathbb{R}^d)$$
$$\bar{Z}_n(\underline{g}, \hat{v}) = (\hat{g}_n, \hat{g}_{n-1} \dots \hat{g}_1 \hat{g}_0 \hat{v}).$$

It is easy to verify that the transition kernel of this Markov chain is precisely the kernel $\bar{K}$ defined above. As an initial distribution of this chain one may choose the (not necessarily stationary) measure $\mu \times \delta_{\hat{v}}$, in order to begin in a specific direction $\hat{v}$, or a stationary measure, which turns out to always be of the form $\mu \times \eta$, for some measure $\eta$ on the projective space.

(iii) The third level is also an SDS, but it acts on the smaller space $\mathbb{P}(\mathbb{R}^d)$, therefore it is usually easier to deal with. It is defined by

$$\begin{aligned} K\colon \mathbb{P}(\mathbb{R}^d) &\to \mathrm{Prob}(\mathbb{P}(\mathbb{R}^d)) \\ \hat{v} &\mapsto \int_\Sigma \delta_{\hat{g}_0\hat{v}}\, d\mu(g_0). \end{aligned}$$

Its corresponding Markov operator is

$$\begin{aligned} Q\colon L^\infty(\mathbb{P}(\mathbb{R}^d)) &\to L^\infty(\mathbb{P}(\mathbb{R}^d)) \\ Q\varphi(\hat{v}) &= \int_{\mathbb{P}(\mathbb{R}^d)} \varphi(\hat{g}_0\hat{v})\, d\mu(g_0). \end{aligned}$$

Furthermore, the associated $K$-Markov chain $\{Z_n\}_{n\geq 0}$ is given by

$$\begin{aligned} Z_n\colon \Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^d) &\to \mathbb{P}(\mathbb{R}^d) \\ Z_n(\underline{g}, \hat{v}) &= \hat{g}_{n-1}\dots\hat{g}_1\hat{g}_0\hat{v}. \end{aligned}$$

Similarly to the second level, natural choices for the initial distribution are $\delta_{\hat{v}}$ or a stationary measure $\eta$.

Moreover, this three levels of dynamics are closely related in regards to their Markov operators, Markov kernels, and stationary measures. In what follows we will describe some of these relations.

**Proposition 2.1** *Given $\eta \in \mathrm{Prob}_Q(\mathbb{P}(\mathbb{R}^d))$, the following are equivalent:*

*(i) $\eta$ is a $Q$-stationary measure.*

*(ii) $\mu \times \eta$ is a $\bar{Q}$-stationary measure.*

*(iii) $\mu^{\mathbb{N}} \times \eta$ is $\hat{F}$-invariant.*

*Proof.* We start by proving that $(i) \iff (ii)$. Note that $\eta$ is $Q$-stationary if and only if $\eta = K*\eta$. Moreover, $\mu\times\eta$ is $\bar{Q}$-stationary if and only if $\mu\times\eta = \bar{K}*(\mu\times\eta)$. Hence, it is enough to prove that $\bar{K} * (\mu \times \eta) = \mu \times (K * \eta)$. In other words, it is sufficient to prove that for every $\varphi \in C^0(\Sigma \times \mathbb{P}(\mathbb{R}^d))$

$$\int \varphi\, d(\bar{K} * (\mu \times \eta)) = \int \varphi\, d\mu d(K * \eta). \tag{2.2}$$

Therefore, let us expand both sides using the definitions of $K$ and $\bar{K}$. First note that

$$\bar{K} * (\mu \times \eta) = \int K_{(g_0,\hat{v})}\, d\mu(g_0)d\eta(\hat{v}) = \int \mu \times \delta_{\hat{g}_0\hat{v}}\, d\mu(g_0)d\eta(\hat{v}) \quad \text{and}$$
$$K * \eta = \int K_{\hat{v}} d\eta(\hat{v}) = \int \delta_{\hat{g}_0\hat{v}}\, d\mu(g_0)d\eta(\hat{v}).$$

Now we substitute the previous computation in equation (2.2) as follows.

$$\begin{aligned}\int \varphi \, d(\bar{K} * (\mu \times \eta)) &= \int \varphi(g_1, \hat{g}_0 \hat{v}) \, d\mu(g_1) d\mu(g_0) d\eta(\hat{v}) \\ &= \int \varphi \, d\mu d(K * \eta).\end{aligned}$$

Now we prove that $(i) \iff (iii)$. Note that $\mu^{\mathbb{N}} \times \eta$ is $\hat{F}$-invariant if and only if for every $\varphi \in C^0(\Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^d))$

$$\int \varphi \, d(\mu^{\mathbb{N}} \times \eta) = \int \varphi \circ \hat{F} \, d(\mu^{\mathbb{N}} \times \eta) = \int \varphi(\sigma g_0, \hat{g}_0 \hat{v}) \, d\mu^{\mathbb{N}}(g_0, g_1, \dots) d\eta(\hat{v}). \quad (2.3)$$

Consider the map $\Psi \colon \mathbb{P}(\mathbb{R}^d) \to \mathbb{R}$, such that $\Psi(\hat{v}) = \int \varphi(\underline{g}', \hat{v}) \, d\mu^{\mathbb{N}}(\underline{g}')$. Since $\varphi$ is arbitrary, so is $\Psi$. Note that the left-hand side of equation (2.3) is equal to $\int \Psi \, d\eta$, while its right hand side is equal to $\int \Psi(\hat{g}_0 \hat{v}) \, d\mu(g_0) d\eta(\hat{v})$, which can be written as $\int Q\Psi(\hat{v}) \, d\eta(\hat{v})$. Therefore $\mu^{\mathbb{N}} \times \eta$ is $\hat{F}$-invariant if and only if $\int \Psi \, d\eta = \int Q\Psi(\hat{v}) \, d\eta(\hat{v})$, which is equivalent to $\eta$ being $Q$-stationary. ■

### 2.2.2 Ergodic stationary measures

By equation (2.1), stationary measures are fixed points of $Q^*$, the dual of the Markov operator $Q$. Therefore $\mathrm{Prob}_Q(M)$ is closed, thus compact in the weak-star topology. Moreover, $\mathrm{Prob}_Q(M)$ is also convex. Indeed, given two different $Q$-stationary measures $\eta_1$ and $\eta_2$ and $t \in (0,1)$,

$$\begin{aligned}\int_M Q\varphi \, d(t\eta_1 + (1-t)\eta_2) &= t \int_M Q\varphi \, d\eta_1 + (1-t) \int_M Q\varphi \, d\eta_2 \\ &= t \int_M \varphi \, d\eta_1 + (1-t) \int_M \varphi \, d\eta_2 \\ &= \int_M \varphi \, d(t\eta_1 + (1-t)\eta_2),\end{aligned}$$

showing that $t\eta_1 + (1-t)\eta_2$ is also $Q$-stationary.

Let $\mathfrak{X}$ be a topological vector space which is Hausdorff and locally convex. Given a set $V \subset \mathfrak{X}$, we say that $p$ is an extremal point of $V$ if whenever $x, y \in V$ and $t \in (0,1)$ are such that $p = tx + (1-t)y$, we necessarily have that $x = y$.

**Theorem 2.1** *(Krein-Milman) If $K \subset \mathfrak{X}$ is compact, convex and non empty, then $K$ has at least an extremal point. Moreover, the closed convex hull of the extreme points of $K$ is equal to $K$.*

Let $\mathfrak{X}$ be the topological vector space of signed measures on $M$ and let $K = \mathrm{Prob}_Q(M)$. Since $\mathrm{Prob}_Q(M)$ is compact, convex and non empty, the Krein-Milman theorem is applicable, therefore there exist extremal stationary measures.

**Definition 2.2** *Let $(M, K, \eta)$ be a Markov system. A function $\phi \in L^\infty(M)$ is called $Q$-stationary if $Q\phi = \phi$ $\eta$-a.e. Moreover, a Borel set $F \subset M$ is called $Q$-stationary if its indicator function $\mathbb{1}_F$ is $Q$-stationary.*

The next result (which should be well known but we could only find a version thereof in a more particular setting, see [42, Proposition 5.11]) characterizes the ergodicity of a Markov system.

**Proposition 2.2** *Let $\eta \in \mathrm{Prob}_K(\mathbb{P}(\mathbb{R}^d))$. The following are equivalent:*

*(i)* *$\eta$ is an extremal point of $\mathrm{Prob}_K(\mathbb{P}(\mathbb{R}^d))$.*

*(ii)* *If a Borel set $F \subset \mathbb{P}(\mathbb{R}^d)$ is $Q$-stationary, then $\eta(F) = 0$ or $\eta(F) = 1$.*

*(iii)* *If $\varphi \in L^\infty(\mathbb{P}(\mathbb{R}^d))$ is $Q$-stationary, then $\varphi$ is constant $\eta$-a.e.*

*(iv)* *The projective cocycle $\hat{F} \colon \Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^d) \to \Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^d)$ is an ergodic dynamical system when endowed with the invariant measure $\mu^{\mathbb{N}} \times \eta$.*

*Proof.* We start with (i) $\Rightarrow$ (ii). Assume that there exists a $Q$-stationary Borel set $F \subset \mathbb{P}(\mathbb{R}^d)$ such that $\eta(F) = t \in (0, 1)$. Then $F^c := \mathbb{P}(\mathbb{R}^d) \setminus F$ is also $Q$-stationary and $\eta(F^c) = 1 - t \in (0, 1)$. Let $\eta_F$ and $\eta_{F^c}$ be the probability measures on $\mathbb{P}(\mathbb{R}^d)$ given by $\eta_F(E) = \frac{\eta(E \cap F)}{\nu(F)}$ and $\eta_{F^c}(E) = \frac{\eta(E \cap F^c)}{\eta(F^c)}$. Note that $\eta_F \neq \eta_{F^c}$ and $\eta = t\eta_F + (1 - t)\eta_{F^c}$.

We show that $\eta_F$ and $\eta_{F^c}$ are $Q$-stationary, which will imply that $\eta$ is not an extremal point of $\mathrm{Prob}_Q(\mathbb{P}(\mathbb{R}^d))$. Indeed, since the indicator function $\mathbb{1}_F$ is stationary, $\forall \hat{v} \in \mathbb{P}(\mathbb{R}^d)$, it holds that

$$\mathbb{1}_F(\hat{v}) = Q\mathbb{1}_F(\hat{v}) = \int \mathbb{1}_F(\hat{g}_0 \hat{v})\, d\mu(g_0)$$

so $\mathbb{1}_F(\hat{v}) = \mathbb{1}_F(\hat{g}_0 \hat{v})$ for $\mu$-a.e. $g_0 \in \Sigma$. This combined with the fact that $\eta$ is $Q$-stationary shows that $\forall \phi \in L^\infty(\mathbb{P}(\mathbb{R}^d))$,

$$\begin{aligned}
\int_{\mathbb{P}(\mathbb{R}^d)} Q\phi(\hat{v})\, d\eta_F(\hat{v}) &= \frac{1}{\eta(F)} \int_F Q\phi(\hat{v})\, d\eta(\hat{v}) \\
&= \frac{1}{\eta(F)} \int_{\mathbb{P}(\mathbb{R}^d)} \int_{\mathbb{P}(\mathbb{R}^d)} \phi\left(\hat{g}_0\hat{v}\right) \mathbb{1}_F(\hat{v})\, d\mu(g_0)\, d\eta(\hat{v}) \\
&= \frac{1}{\eta(F)} \int_{\mathbb{P}(\mathbb{R}^d)} \int_{\mathbb{P}(\mathbb{R}^d)} \phi\left(\hat{g}_0\hat{v}\right) \mathbb{1}_F(\hat{g}_0\hat{v})\, d\mu(g_0)\, d\eta(\hat{v}) \\
&= \frac{1}{\eta(F)} \int_{\mathbb{P}(\mathbb{R}^d)} \int_{\mathbb{P}(\mathbb{R}^d)} \left(\phi\, \mathbb{1}_F\right)\, (\hat{g}_0\hat{v})\, d\mu(g_0)\, d\eta(\hat{v}) \\
&= \frac{1}{\eta(F)} \int_{\mathbb{P}(\mathbb{R}^d)} Q(\phi\, \mathbb{1}_F)(\hat{v})\, d\eta(\hat{v}) \\
&= \frac{1}{\eta(F)} \int_{\mathbb{P}(\mathbb{R}^d)} (\phi\, \mathbb{1}_F)(\hat{v})\, d\eta(\hat{v}) \\
&= \frac{1}{\eta(F)} \int_F \phi(\hat{v})\, d\eta(\hat{v}) = \int_{\mathbb{P}(\mathbb{R}^d)} \phi\, d\eta_F\,.
\end{aligned}$$

Therefore $\eta_F$ is a $Q$-stationary probability measure. Similarly, $\eta_{F^c}$ is also $Q$-stationary, therefore $\eta$ is not an extremal point of $\mathrm{Prob}_Q(\mathbb{P}(\mathbb{R}^d))$.

Now we prove that (ii) implies (iii). Let $\varphi \in L^\infty(\mathbb{P}(\mathbb{R}^d))$ be a $Q$ stationary function. Fix $c \in \mathbb{R}$ and consider the set $E = \{\hat{v} \colon \varphi(\hat{v}) < c\}$. It is enough to prove that $E$ is $Q$ stationary, because then, by (ii), either $\eta(E) = 0$ or $\eta(E) = 1$; since $c$ is arbitrary, this would imply the existence of a constant $c^*$ such that $\varphi = c^*$ $\eta$-a.e. Let

$$S = \left\{\psi \in L^\infty(\mathbb{P}(\mathbb{R}^d)) \colon \psi \text{ is } Q\text{-stationary}\right\}.$$

It is clear that $S$ is a linear space. We show that $S$ is a lattice, i.e if $\phi \in S$, $|\phi| \in S$ and if $\phi, \psi \in S$, then $\min\{\phi,\psi\} \in S$ and $\max\{\phi,\psi\} \in S$. Let $\varphi \in S$. Then $\varphi(\hat{v}) = Q\varphi(\hat{v})$ $\eta$-a.e $\hat{v}$. Hence for $\eta$ almost every $\hat{v}$,

$$|\varphi(\hat{v})| = |Q\varphi(\hat{v})| = \left|\int \varphi(\hat{g}_0\hat{v})\, d\mu(g_0)\right| \leq \int |\varphi(\hat{g}_0\hat{v})|\, d\mu(g_0) = Q|\varphi|(\hat{v}).$$

Thus $|\varphi| \leq Q|\varphi|$ for $\eta$-a.e. $\hat{v} \in \mathbb{P}(\mathbb{R}^d)$. Since $\eta \in \mathrm{Prob}_Q(\mathbb{P}(\mathbb{R}^d))$, it follows that $\int Q|\varphi|\ d\eta = \int |\varphi|\ d\eta$ and we conclude that $|\varphi| = Q|\varphi|$ $\eta$ almost everywhere, hence $|\varphi| \in S$. Moreover, since $\min\{\phi,\psi\} = \frac{\phi+\psi}{2} - \frac{|\phi-\psi|}{2}$ and $\max\{\phi,\psi\} = \frac{\phi+\psi}{2} + \frac{|\phi-\psi|}{2}$, using the linearity of $S$ we conclude that $S$ is a lattice.

Now consider $\varphi_n(\hat{v}) = \min\{1, n\max\{c - \varphi(\hat{v}), 0\}\}$. Note that for every $\varphi_n \in S$ and $\hat{v} \in \mathbb{P}(\mathbb{R}^d)$, $\varphi_n \to \mathbb{1}_E$ as $n \to \infty$. Hence $Q\varphi_n \to Q\mathbb{1}_E$ and $Q\varphi_n = \varphi_n \to \mathbb{1}_E$. We conclude that $Q\mathbb{1}_E = \mathbb{1}_E$ which means that $E$ is $Q$-stationary.

We now suppose that (iii) is true and prove that $\mu^{\mathbb{N}} \times \eta$ is $\hat{F}$-ergodic. Let

$\psi \in L^\infty(\Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^d))$ be a $\hat{F}$-invariant function, so $\psi(\hat{F}(\underline{g}, \hat{v})) = \psi(\underline{g}, \hat{v})$ for $\mu^{\mathbb{N}} \times \eta$ almost every $(\underline{g}, \hat{v})$. We have to show that $\psi$ is constant $\mu^{\mathbb{N}} \times \eta$-a.e.

Let $\phi\colon \mathbb{P}(\mathbb{R}^d) \to \mathbb{R}$, $\phi(\hat{v}) = \int \psi(\underline{g}, \hat{v}))\, d\mu^{\mathbb{N}}(\underline{g})$. We start by showing that it is constant $\eta$-a.e. By $(iii)$, it is enough to prove that $\phi$ is $Q$-stationary.

For all $\hat{v} \in \mathbb{P}(\mathbb{R}^d)$,

$$\begin{aligned}\phi(\hat{v}) &= \int \psi(\underline{g}, \hat{v}))\, d\mu^{\mathbb{N}}(\underline{g}) = \int \psi(\hat{F}(\underline{g}, \hat{v}))\, d\mu^{\mathbb{N}}(\underline{g}) \\ &= \int \psi(\{g_n\}_{n\geq 1}, \hat{g}_0 \hat{v})\, d\mu^{\mathbb{N}}(\underline{g}) \\ &= \int \psi(\underline{g}', \hat{g}_0 \hat{v})\, d\mu^{\mathbb{N}}(\underline{g}') d\mu(g_0) \\ &= \int \phi(\hat{g}_0 \hat{v})\, d\mu(g_0) = (Q\phi)(\hat{v})\,.\end{aligned}$$

It follows that $\phi$ is constant $\eta$-a.e., that is, there are $c \in \mathbb{R}$ and $E \subset \mathbb{P}(\mathbb{R}^d)$ with $\eta(E) = 0$ such that

$$\phi(\hat{v}) = \int \psi(\underline{g}, \hat{v})\, d\mu^{\mathbb{N}}(\underline{g}) = c \quad \forall \hat{v} \notin E\,.$$

This shows that $\psi$ does not depend $\eta$-a.s. on the variable $\hat{v}$. We will show that it also does not depend $\mu^{\mathbb{N}}$-a.s. on $\underline{g}$.

For every $k \in \mathbb{N}$ let $\mathcal{F}_k$ be the $\sigma$-algebra generated by cylinders $C_k \subset \Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^d)$, where $C_k$ is a cylinder in the coordinates $(g_0, g_1, \ldots, g_{k-1}, \hat{v})$. This sequence of $\sigma$-algebras forms a filtration which generates the Borel $\sigma$-algebra of $\Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^d)$. We show that the conditional expectation (with respect to $\mu^{\mathbb{N}} \times \eta$) $\mathbb{E}(\psi|\mathcal{F}_k)$ is constant $\mu^{\mathbb{N}} \times \eta$-a.e. for all $k \in \mathbb{N}$. Indeed, given any $k \in \mathbb{N}$,

$$\begin{aligned}\mathbb{E}(\psi|\mathcal{F}_k) &= \int \psi(\underline{g}, \hat{v})\, d\mu^{\mathbb{N}} \times \eta(\{g_n\}_{n\geq k}, \hat{v}) \\ &= \int \psi(\hat{F}^k(\underline{g}, \hat{v}))\, d\mu^{\mathbb{N}} \times \eta(\{g_n\}_{n\geq k}, \hat{v}) = \phi(\hat{g}_0 \hat{g}_1 \ldots \hat{g}_{k-1} \hat{v}) = c.\end{aligned}$$

provided that $\hat{g}_0 \hat{g}_1 \ldots \hat{g}_{k-1} \hat{v} \notin E$.

Let $\pi$ denote the projection $\pi : \mathrm{Prob}(\Sigma \times \mathbb{P}(\mathbb{R}^d)) \to \mathrm{Prob}(\mathbb{P}(\mathbb{R}^d))$. Since $\hat{F}$ is $\mu^{\mathbb{N}} \times \eta$-invariant and $\mu^{\mathbb{N}} \times \eta$ projects to $\eta$, $(\pi)_*(\mu^{\mathbb{N}} \times \eta) = \eta$, it holds that

$$\begin{aligned}\mu^{\mathbb{N}} \times \eta \left\{(\underline{g}, \hat{v})\colon \hat{g}_0 \hat{g}_1 \ldots \hat{g}_{k-1} \hat{v} \in E\right\} &= \mu^{\mathbb{N}} \times \eta \left((\hat{F}^k)^{-1}(\pi^{-1} E)\right) \\ &= \mu^{\mathbb{N}} \times \eta \left(\pi^{-1} E\right) = \eta(E) = 0,\end{aligned}$$

hence $\mathbb{E}(\psi|\mathcal{F}_k) = c$ holds $\mu^{\mathbb{N}} \times \eta$-a.e. Therefore, given any $k \in \mathbb{N}$ and any $F \in \mathcal{F}_k$, by the definition of the conditional expectation we conclude that

$$\int_F \psi\, d\mu^{\mathbb{N}} \times \eta = \int_F \mathbb{E}(\psi|\mathcal{F}_k)\, d\mu^{\mathbb{N}} \times \eta = \int_F c\, d\mu^{\mathbb{N}} \times \eta.$$

Since $\bigcup_{k\in\mathbb{N}} \mathcal{F}_k$ generates the Borel $\sigma$-algebra of $\Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^d)$, we then conclude that $\psi = c$ holds $\mu^{\mathbb{N}} \times \eta$-a.e.

It remains to prove that (iv) implies (i). Assume by contradiction that $\eta$ is not an extremal point of $\mathrm{Prob}_Q(\mathbb{P}(\mathbb{R}^d))$, so there are $t \in (0,1)$ and $\eta_1 \neq \eta_2 \in \mathrm{Prob}_Q(\mathbb{P}(\mathbb{R}^d))$ such that $\eta = t\eta_1 + (1-t)\eta_2$. Since $\eta_1, \eta_2 \in \mathrm{Prob}_Q(\mathbb{P}(\mathbb{R}^d))$, both $\mu^{\mathbb{N}} \times \eta_1$ and $\mu^{\mathbb{N}} \times \eta_2$ are $\hat{F}$-invariant. Then the $\hat{F}$-invariant measure $\mu^{\mathbb{N}} \times \eta = t(\mu^{\mathbb{N}} \times \eta_1) + (1-t)(\mu^{\mathbb{N}} \times \eta_2)$ is not an extremal point for the set of invariant measures over $\Sigma \times \mathbb{P}(\mathbb{R}^d)$, hence $\mu^{\mathbb{N}} \times \eta$ is not $\hat{F}$-ergodic. ■

## 2.3 Mixing and statistical properties

### 2.3.1 Abstract LDT and CLT

Let $L^\infty(M)$ denote the Banach space of all measurable and bounded functions from $M$ to $\mathbb{R}$, which is endowed with the norm $\|\varphi\|_\infty = \sup_{x\in M} |\varphi(x)|$. Let $(\mathcal{E}, \|\cdot\|_{\mathcal{E}})$ be a Banach subspace of $(L^\infty, \|\cdot\|_\infty)$ that satisfies the following properties:

(i) $\|\varphi\|_\infty \leq \|\varphi\|_{\mathcal{E}}$.

(ii) It is invariant under the Markov operator: $Q(\mathcal{E}) \subset \mathcal{E}$.

(iii) $Q|_{\mathcal{E}}$ is a bounded operator.

(iv) The constant function $\mathbf{1}$ belongs to $\mathcal{E}$.

**Definition 2.3** *A Markov system $(M, K, \nu)$ is called strongly mixing in $(\mathcal{E}, \|\cdot\|_{\mathcal{E}})$ if there exist constants $c \in \mathbb{R}$ and $\sigma \in (0,1)$ such that*

$$\|Q^n\varphi - \int \varphi \, d\nu\|_\infty \leq c\sigma^n \|\varphi\|_{\mathcal{E}} \quad \forall\, \varphi \in \mathcal{E}.$$

This property will play a central role throughout the text, because it is closely related to the statistical properties of the Lyapunov exponents, as we will show in the remaining of this section. It turns out that it is also related to the regularity of the Lyapunov exponents.

The next theorem is due to Cai, Duarte, Klein [13]. It is an abstract result that guarantees large deviations type estimates under much more general assumptions than the commonly used spectral gap property of the transfer/transition operators. The exponential rate of convergence of $Q^n\varphi$ to $\int \varphi \, d\nu$ in definition 2.3 is not strictly necessary, any rate will do.

**Theorem 2.2** *Let $(M, K, \nu)$ be a strongly mixing Markov system on $(\mathcal{E}, \|\cdot\|_{\mathcal{E}})$. Then, for all $\varphi \in \mathcal{E}$ and $\varepsilon > 0$ there are $c(\varepsilon) > 0$ and $n(\varepsilon) \in \mathbb{N}$ such that for all $n \geq n(\varepsilon)$ ,*

$$\mathbb{P}\left\{\left|\frac{1}{n}S_n\varphi - \int_M \varphi \, d\nu\right| > \varepsilon\right\} \leq 8e^{-c(\varepsilon)n}. \tag{2.4}$$

*The rate function $c(\varepsilon)$ is essentially of order $\varepsilon^2 \|\varphi\|_{\mathcal{E}}^{-2}$.*

Next we state an abstract central limit theorem (CLT) theorem due to Gordin-Livšic [31].

Let $\varphi \in L^2(M, \nu)$ be an observable with $\int \varphi \, d\nu = 0$. If $\sum_{n=0}^{\infty}\|Q^n\varphi\|_2 < \infty$ then we can define $g \in L^2(M, \nu)$ by

$$g := \sum_{n=0}^{\infty} Q^n\varphi.$$

Then $\varphi = g - Qg$. Let $\sigma^2(\varphi) = \|g\|_2^2 - \|Qg\|_2^2$.

**Theorem 2.3** *(Gordin-Livsic) Let $(M, Q, \nu)$ be an ergodic Markov system and let $\varphi \in L^2(M)$ with $\int \varphi \, d\nu = 0$. Assume that*

$$\sum_{n=0}^{\infty}\|Q^n\varphi\|_2 < \infty \quad \text{and } \sigma^2(\varphi) \in (0, \infty).$$

*Then the following central limit theorem holds:*

$$\frac{S_n\varphi}{\sigma\sqrt{n}} \to \mathcal{N}(0, 1).$$

*We note that this result holds not only relative to the Markov probability with initial distribution $\eta$, but also with initial distribution $\delta_x$ for $\nu$-a.e. $x \in M$.*

A version of this abstract result, more immediately applicable to dynamical systems was derived in [13]. We formulate it below.

**Proposition 2.3** *Let $(M, K, \nu)$ be a strongly mixing Markov system in $(\mathcal{E}, \|\cdot\|_{\mathcal{E}})$, where $\mathcal{E}$ is a dense subspace of $C^0(M)$. Assume that for every open set $U \subset M$ with $\nu(U) > 0$, there exists $\Phi \in \mathcal{E}$ such that $0 \leq \Phi \leq 1_U$ and $\int \Phi \, d\nu > 0$.*

*Then given any observable $\varphi \in \mathcal{E}$ such that it is not $\nu$-a.e constant, it follows that $\sigma^2(\varphi) > 0$ and theorem 2.3 holds:*

$$\frac{S_n\varphi}{\sigma\sqrt{n}} \to \mathcal{N}(0, 1).$$

Theorems 2.2 and 2.3 are abstract in the sense that they hold for very general Markov chains and so they are applicable in principle to a wide class of dynamical systems. In order to apply them to our setting, that of random linear cocycles, and prove statistical properties for the Lyapunov exponent, we need to consider the special observable

$$\Phi\colon \Sigma \times \mathbb{P}(\mathbb{R}^d), \quad \Phi(g_0, \hat{v}) = \log \frac{\|g_0 v\|}{\|v\|}. \tag{2.5}$$

This observable is relevant because its stochastic Birkhoff average associated to the $\bar{K}$-Markov chain define above is (easily show to be) given by

$$\frac{1}{n} S_n \Phi(\underline{g}, \hat{v}) = \frac{1}{n} \log\|g_{n-1} \dots g_1 g_0 v\| - \frac{1}{n} \log\|v\|$$

which converges almost surely to the top Lyapunov exponent $L_1(\mu)$.

In particular, in order to prove statistical properties for the Lyapunov exponent we will need to prove that under appropriate hypotheses, the Markov system $(\Sigma \times \mathbb{P}(\mathbb{R}^d), \bar{K})$ is strongly mixing in a suitable Banach space $\mathcal{E}$ such that $\Phi \in \mathcal{E}$. This is how we will proceed in chapter 4. Later, in chapter 7, we will deal with non invertible matrices, hence this observable becomes *unbounded* and the approach is much more delicate.

### 2.3.2 The reduction from the second to the third level

Now we show that in order to prove that the Markov system on the second level $(\Sigma \times \mathbb{P}(\mathbb{R}^d), \bar{K}, \mu \times \eta)$ is strongly mixing, it suffices to prove that the Markov system on the third level $(\mathbb{P}(\mathbb{R}^d), K, \eta)$ is strongly mixing. Here $\eta$ is a $K$-stationary measure, that always exists, and under the various additional assumptions we will eventually impose, it will be unique. Since the third level is simpler to deal with, we will thus work with a smaller degree of complexity.

**Lemma 2.2** *Consider the following projection map*

$$\pi\colon L^\infty(\Sigma \times \mathbb{P}(\mathbb{R}^d)) \to L^\infty(\mathbb{P}(\mathbb{R}^d)), \quad \pi\varphi(\hat{v}) = \int \varphi(g_0, \hat{v})\, d\mu(g_0).$$

*Then for every $\varphi \in L^\infty(\Sigma \times \mathbb{P}(\mathbb{R}^d))$, the following hold:*

*(i)* $\pi \circ \bar{Q} = Q \circ \pi$.

*(ii)* $\bar{Q}\varphi(g_0, \hat{v}) = (\pi\varphi)(\hat{g}_0 \hat{v})$.

*(iii)* $\bar{Q}^n \varphi(g_0, \hat{v}) = Q^{n-1}(\pi\varphi)(\hat{g}_0 \hat{v})$ *for every* $n \in \mathbb{N}$.

*Proof.* Items (i) and (ii) are straightforward computations. Note that

$$\bar{Q}\varphi(g_0, \hat{v}) = \int \varphi(g_1, \hat{g}_0 \hat{v})\, d\mu(g_1) = \pi\varphi(\hat{g}_0 \hat{v}).$$

Item (iii) follows by induction, where the base case is item (ii). For the induction step, assume that item (iii) holds for $n = k$. Then

$$\bar{Q}^{k+1}\varphi(g_0, \hat{v}) = \bar{Q}^k\bar{Q}\varphi(g_0, \hat{v}) = Q^{k-1}(\pi\bar{Q}\varphi)(\hat{g}_0\hat{v}) = Q^k(\pi\varphi)(\hat{g}_0\hat{v}).$$

■

**Proposition 2.4** *The Markov system in the second level* $(\Sigma\times\mathbb{P}(\mathbb{R}^d), \bar{K}, \mu\times\eta)$ *is strongly mixing in a Banach space* $\mathcal{E}$ *if and only if the Markov system in the third level* $(\mathbb{P}(\mathbb{R}^d), K, \eta)$ *is strongly mixing in* $\pi(\mathcal{E})$.

*Proof.* Consider an arbitrary $\varphi \in L^\infty(\mathbb{P}(\mathbb{R}^d))$. Let $\bar{\varphi} \in L^\infty(\Sigma \times \mathbb{P}(\mathbb{R}^d))$ be such that $\pi\bar{\varphi} = \varphi$ and $\|\bar{\varphi}\|_{\mathcal{E}} = \|\pi\bar{\varphi}\|_{\pi(\mathcal{E})}$. Then

$$\int_{\mathbb{P}(\mathbb{R}^d)} \varphi \, d\eta = \int_{\mathbb{P}(\mathbb{R}^d)} \pi\bar{\varphi} \, d\eta = \int_{\Sigma\times\mathbb{P}(\mathbb{R}^d)} \bar{\varphi}. \, d\mu \times \eta$$

Therefore, by lemma 2.2, we conclude that

$$\|\bar{Q}^n\bar{\varphi} - \int_{\Sigma\times\mathbb{P}(\mathbb{R}^d)} \bar{\varphi} \, d\mu \times \eta\|_\infty \leq c\sigma^n\|\bar{\varphi}\|_{\mathcal{E}} \iff$$
$$\|Q^{n-1}(\pi\varphi) - \int_{\mathbb{P}(\mathbb{R}^d)} \pi\bar{\varphi} \, d\eta\|_\infty \leq c\sigma^n\|\pi\bar{\varphi}\|_{\pi(\mathcal{E})}$$

which concludes the proof. ■

# 3
# Furstenberg-Kifer's Multiplicative Ergodic Theorem

The goal of this chapter is to introduce Furstenberg-Kifer's Multiplicative Ergodic Theorem, which is one of the main tools in the study of random linear cocycles. This theorem was first proved in [29], then generalized by Kifer in his monograph [34]. In this chapter we provide a detailed proof of this classical result, which we believe to benefit the reader, as the argument in Kifer's monograph may be found rather difficult to read. In a joint work [15] with Cai, Klein and Melo we further extend this result to a more general context which we will discuss in chapter 8.

This type of result, also referred to as Furstenberg-Kifer's non-random filtration, is used in essentially every proof of continuity of the Lyapunov exponents of random linear cocycles, e.g.: in the original result in the generic, i.i.d. setting of Furstenberg and Kifer [29]; in the more recent results, eliminating this generic condition, by Avila, Eskin and Viana, see [43] and [3]; in the quantitative results like the ones in [12], [21, Chapter 5] and chapters 4 and 5 of this manuscript; in the study of other types of random-type linear cocycles (e.g. mixed random-quasiperiodic) as in [14] and related works.

This chapter is organized as follows. In section 3.1 we prove the classical Furstenberg's Formula and we also introduce some important elements of the proof of Furstenberg-Kifer's multiplicative ergodic theorem, which is then derived in section 3.2. In section 3.3 we formally introduce the concept of irreducible random cocycle and relate it to the main result of the chapter.

## 3.1
## Furstenberg's formula

### 3.1.1
### Martingale construction

Consider a probability space $(\Omega, \mathcal{F}, \mu)$, where $\Omega$ is a compact set, $\mathcal{F}$ is a $\sigma$-algebra and $\mu$ is a probability measure on $\Omega$. For every $n \in \mathbb{N}$, let $\xi_n \colon \Omega \to \mathbb{R}$ be a random variable. A martingale is a sequence $\{(\xi_n, \mathcal{F}_n)\}_{n \geq 1}$ that satisfies

(i) $\{\mathcal{F}_n\}_{n \geq 1}$ is a filtration of $\sigma$-algebras with $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for every $n \in \mathbb{N}$.

(ii) $\mathbb{E}|\xi_n| < \infty$ for all $n$.

(iii) $\xi_n$ is $\mathcal{F}_n$-measurable for all $n$.

(iv) $\mathbb{E}(\xi_{n+1}|\mathcal{F}_n) = \xi_n$ for all $n$.

Recall that given any random variable $\xi\colon \Omega \to \mathbb{R}$, the sigma algebra generated by $\xi$ is defined as $\sigma(\xi) = \{\xi^{-1}(B)\colon B \in \mathcal{F}\}$. This is the smallest sigma algebra relative to which $\xi$ is a measurable function.

**Example 3.1** *(The random walk) Let $\{\xi_n\}_n$ be a sequence of i.i.d real random variables satisfying $\mathbb{E}(\xi_1) = 0$ and $\mathbb{E}|\xi_1| < \infty$. Let $S_n = \xi_1 + \cdots + \xi_n$ and let $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$, the sigma algebra generated by $\{\xi_i\colon 1 \le i \le n\}$. Then $\{S_n, \mathcal{F}_n\}_n$ is a martingale.*

*Indeed, properties (i)-(iii) are clearly satisfied and we only verify (iv):*

$$\mathbb{E}(S_{n+1}|\mathcal{F}_n) = \mathbb{E}(\xi_{n+1} + S_n|\mathcal{F}_n) = \mathbb{E}(\xi_{n+1}|\mathcal{F}_n) + \mathbb{E}(S_n|\mathcal{F}_n) = 0 + \mathbb{E}(S_n).$$

**Theorem 3.1** *(Doob's martingale convergence theorem) Let $\{(\xi_n, \mathcal{F}_n)\}_n$ be a martingale in $(\Omega, \mathcal{F}, \mu)$ such that $\sup_n \mathbb{E}(|\xi_n|) < \infty$. Then there exists $\xi_\infty \in L^1(\mu)$ such that*

*(i)* $\xi_n \to \xi_\infty$ *a.s.*

*(ii)* $\mathbb{E}(\xi_\infty|\mathcal{F}_n) = \xi_n$

*(iii)* $\xi_\infty$ *is* $\mathcal{F}_\infty$ *measurable, where* $\mathcal{F}_\infty = \sigma(\cup_{n\ge1}\mathcal{F}_n)$.

A proof of this theorem can be found at [26].

The following result is due to Furstenberg and Kifer, see [29].

**Theorem 3.2** *Let $(\Omega, \mathcal{F}, \mu)$ be a probability space. Let $M$ be a compact metric space and let $K\colon M \to \mathrm{Prob}(M)$ be a continuous Markov kernel. Given a $K$-Markov chain $\{Z_n\colon \Omega \to M\}_{n\ge0}$, for any observable $\varphi \in C^0(M)$, the following hold for $\mu$-almost every $\omega \in \Omega$:*

*(i)* $\displaystyle\limsup_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \varphi(Z_j) \le \max\Big\{\int_M \varphi\, d\eta : \eta \in \mathrm{Prob}_K(M)\Big\}.$

*(ii) If for some* $\beta \in \mathbb{R}$, $\int_M \varphi\, d\eta = \beta$ *for all* $\eta \in \mathrm{Prob}_K(M)$*, then*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \varphi(Z_j) = \beta.$$

*Proof.* Note that since $\text{Prob}_K(M)$ is compact and the map $\eta \mapsto \displaystyle\int_M \varphi \, d\eta$ is continuous, there exists the maximum of the set $\{\int_M \varphi \, d\eta : \eta \in \text{Prob}_K(M)\}$, which we will denote by $\beta$.

Let $Q$ be the Markov operator of the transition kernel $K$, that is,

$$Q\varphi(x) = \int_M \varphi(y) \, dK_x(y).$$

We split the proof of item $(i)$ into two steps. In the first step, we prove the claim for co-boundary observables, which are observables $\varphi$ that satisfy the hypothesis that $\varphi = Qg - g$ for some $g \in C^0(M)$. Then we extend the result to any observable $\varphi \in C^0(M)$.

Note that when $\varphi$ is a co-boundary, $\int_M \varphi \, d\eta = \int_M Qg \, d\eta - \int_M g \, d\eta = 0$ for all $\eta \in \text{Prob}_K(M)$. We will then have to prove that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(Z_j) = 0 \quad \text{a.s.}$$

Consider the sequence of random variables $\xi_n : \Omega \to M$,

$$\xi_n = \sum_{j=1}^{n} \frac{Qg(Z_{j-1}) - g(Z_j)}{j},$$

which depends on $Z_0, Z_1, \ldots, Z_n$. Let $\mathcal{F}_n \colon = \sigma(Z_0, Z_1, \ldots, Z_n)$.

We claim that $\{(\xi_n, \mathcal{F}_n)\}$ is a martingale which, moreover, satisfies the assumptions of Doob's martingale convergence theorem above.

We begin with the proof of the fourth item of the martingale definition (the other items are obvious). Note that

$$\xi_{n+1} = \xi_n + \frac{1}{n} \left[ Qg(Z_n) - g(Z_{n+1}) \right].$$

Therefore, applying the conditional expectation to both sides,

$$\mathbb{E}(\xi_{n+1} | \mathcal{F}_n) = \xi_n + \frac{1}{n+1} \left[ \mathbb{E}(Qg(Z_n) | \mathcal{F}_n) - \mathbb{E}(g(Z_{n+1}) | \mathcal{F}_n) \right].$$

Note that $\mathbb{E}(Qg(Z_n) | \mathcal{F}_n) = Qg(Z_n)$. Moreover,

$$\mathbb{E}(g(Z_{n+1}) | \mathcal{F}_n) = \mathbb{E}(g(Z_{n+1}) | Z_0, Z_1, \ldots, Z_n) = \mathbb{E}(g(Z_{n+1}) | Z_n),$$

where the last equality is due to the Markov property. By the definition of a $K$-Markov chain, given any $E \in \mathcal{F}$, we get $\mu(Z_{n+1} \in E | Z_n = x) = K_x(E)$.

Thus,

$$\mathbb{E}(g(Z_{n+1})|Z_n) = \int_M g\, dK_{Z_n} = Qg(Z_n).$$

We conclude that $\mathbb{E}(\xi_{n+1}|\mathcal{F}_n) = \xi_n$, hence $\{(\xi_n, \mathcal{F}_n)\}$ is a martingale.

Let us now check that $\sup_n \mathbb{E}(|\xi_n|) < \infty$. In order to prove this, we start with the following lemma which was explained to the author by P. Duarte.

**Lemma 3.1** *Let $f \in C^0(M)$. Consider the random variables*

$$\begin{aligned}\Delta_n: &= (Qf)(Z_{n-1}) - f(Z_n) = \mathbb{E}(f(Z_n)|Z_{n-1}) - f(Z_n)\\ &= \mathbb{E}(f(Z_n)|\mathcal{F}_n) - f(Z_n),\end{aligned}$$

*where $\mathcal{F}_n = \sigma(Z_0, Z_1, \ldots, Z_n)$. Then these random variables are pairwise uncorrelated, that is,*

$$\mathbb{E}(\Delta_k \Delta_n) = 0 \quad \textit{for all } 0 \leq k < n.$$

*Proof.* Let $\eta$ be a $K$-stationary measure and let $\mathcal{B}$ be the Borel sigma algebra of $M$. Note that

$$\begin{aligned}\mathbb{E}(\Delta_n|\mathcal{F}_n) &= \mathbb{E}(\mathbb{E}(f(Z_n)|\mathcal{F}_n) - f(Z_n)|\mathcal{F}_n)\\ &= \mathbb{E}(f(Z_n)|\mathcal{F}_n) - \mathbb{E}(f(Z_n)|\mathcal{F}_n)) = 0.\end{aligned}$$

Let us recall that given sub sigma algebras $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{B}$, we have that $\mathbb{E}(f|\mathcal{G}_1) = \mathbb{E}(\mathbb{E}(f|\mathcal{G}_2)|\mathcal{G}_1)$. Moreover, if $\mathcal{G}_0 = \{\emptyset, M\}$ is the trivial sigma algebra, then $\mathbb{E}(f) = \mathbb{E}(f|\mathcal{G}_0)$. Thus

$$\mathbb{E}(\Delta_n) = \mathbb{E}(\Delta_n|\mathcal{G}_0) = \mathbb{E}(\mathbb{E}(\Delta_n|\mathcal{F}_n)|\mathcal{G}_0) = 0.$$

Given $k < n$, it follows that

$$\Delta_n \Delta_k = \Delta_n(\mathbb{E}(f(Z_k)|\mathcal{F}_k) - f(Z_k)) = \Delta_n(g_k - h_k),$$

where $g_k = \mathbb{E}(f(Z_k)|\mathcal{F}_k)$ and $h_k = f(Z_k)$.

Since $g_k$ is $\mathcal{F}_k$ measurable,

$$\mathbb{E}(\Delta_n g_k|\mathcal{F}_k) = g_k \mathbb{E}(\Delta_n|\mathcal{F}_k) = \mathbb{E}(\mathbb{E}(\Delta_n|\mathcal{F}_n)|\mathcal{F}_k) = 0.$$

The first equality in the previous expression holds because given any $f, g \in L^2(M, \mathcal{B}, \eta)$ with $g$ being $\mathcal{F}$ measurable, we have that $\mathbb{E}(fg|\mathcal{F}) = g\mathbb{E}(f|\mathcal{F})$.

Therefore,

$$\mathbb{E}(\Delta_n g_k) = \mathbb{E}(\Delta_n g_k | \mathcal{F}_0) = \mathbb{E}(\mathbb{E}(\Delta_n g_k | \mathcal{F}_k) | \mathcal{F}_0) = 0.$$

Similarly, since $h_k$ is $\mathcal{F}_k$-measurable,

$$\mathbb{E}(\Delta_n h_k | \mathcal{F}_k) = h_k \mathbb{E}(\Delta_n | \mathcal{F}_k) = h_k \mathbb{E}(\mathbb{E}(\Delta_n | \mathcal{F}_n) | \mathcal{F}_k) = 0.$$

Therefore, it follows that

$$\mathbb{E}(\Delta_n h_k) = \mathbb{E}(\Delta_n h_k | \mathcal{F}_0) = \mathbb{E}(\mathbb{E}(\Delta_n h_k | \mathcal{F}_k) | \mathcal{F}_0) = 0.$$

Together, these facts imply that $\mathbb{E}(\Delta_n \Delta_k) = 0$, which proves the lemma. ■

Using this orthogonality property of the random variables, the Pythagorean theorem implies that the martingale random variables

$$\xi_{n+1} := \sum_{k=0}^{n} \frac{1}{k}(Qg)(Z_k) - g(Z_{k+1}) = \sum_{k=0}^{n} \frac{1}{k} \Delta_{k+1}$$

satisfy the following inequality

$$\|\xi_n\|_{L^2}^2 \leq 4\|g\|_\infty \sum_{k=1}^{\infty} \frac{1}{k^2}.$$

Therefore, by the Cauchy-Schwarz inequality, we conclude that $\sup_n \mathbb{E}(|\xi_n|) < \infty$, so Doob's martingale convergence theorem is indeed applicable to the martingale process $\{(\xi_n, \mathcal{F}_n)\}$ and we conclude that

$$\lim_{n\to\infty} \sum_{j=1}^{n} \frac{Qg(Z_{j-1}) - g(Z_j)}{j}$$

exists and it is finite for almost every $\omega \in \Omega$.

Recall the following Kronecker lemma: if $\{a_n\}_n \subset \mathbb{R}$ is such that $\sum_{n=1}^{\infty} a_n$ converges, then $\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{\infty} j a_j = 0$. Hence

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} [Qg(Z_{j-1}) - g(Z_j)] = 0 \quad \text{a.e.}$$

By expanding the terms in this telescopic sum, we conclude that for $\mu$-a.e. $\omega \in \Omega$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \varphi(Z_j) = 0$$

which completes the first step of the proof of item (i).

Now we proceed to proving that (i) holds for every $\varphi \in C^0(M)$. Since $M$ is compact, $C^0(M)$ is separable, hence there exists a countable set $\{g_1, \dots, g_k, \dots\}$ which is dense in $C^0(M)$. Therefore, we may apply the previous result to every $\varphi_k\colon = Qg_k - g_k$. Thus, for every $k \in \mathbb{N}$, there exists $\Omega_k \subset \Omega$, with $\mu(\Omega_k) = 1$ and such that for every $\omega \in \Omega_k$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi_k(Z_j(\omega)) = 0.$$

Let $\Omega_* = \cap_k \Omega_k$ and note that $\mu(\Omega_*) = 1$. Fix an arbitrary $\omega \in \Omega_*$ and for every $n \in \mathbb{N}$, define the following measure

$$\eta_n\colon = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{Z_j(\omega)} \in \operatorname{Prob}(M).$$

Note that $\int_M Qg_k - g_k \, d\eta_n \to 0$ for every $k \in \mathbb{N}$. Since $\{g_1, \dots, g_k, \dots\}$ is dense in $C^0(M)$, it also holds that $\int_M Qg - g \, d\eta_n \to 0$ for every $g \in C^0(M)$. Therefore, every limit point $\eta_\infty$ of $\{\eta_n\}_n$ is $Q$-stationary.

Note that there exists a subsequence $\{n_k\}_k$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \sum_{j=0}^{n-1} \varphi(Z_j) = \lim_{k\to\infty} \frac{1}{n_k} \log \sum_{j=0}^{n_k-1} \varphi(Z_j) = \lim_{k\to\infty} \int_M \varphi \, d\eta_{n_k} .$$

Finally, since $\operatorname{Prob}(M)$ is compact, there exists a further subsequence $\{\eta_{n_{k_i}}\}_i$ that converges to some $\eta_\infty$, which, by the previous argument belongs to $\operatorname{Prob}_Q(M)$. Thus, we conclude that

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log \sum_{j=0}^{n-1} \varphi(Z_j) &= \lim_{i\to\infty} \int_M \varphi \, d\eta_{n_{k_i}} = \int_M \varphi \, d\eta_\infty \\ &\leq \max\left\{ \int_M \varphi; d\eta : \eta \in \operatorname{Prob}_Q(M) \right\}.\end{aligned}$$

Next we prove item (ii). We apply item (i) to $-\varphi$ and conclude that for a.e $\omega \in \Omega$, we have that $\limsup_{n\to\infty} \frac{1}{n} \log \sum_{j=0}^{n-1} -\varphi(Z_j) \leq -\beta$. Thus $-\liminf_{n\to\infty} \frac{1}{n} \log \sum_{j=0}^{n-1} \varphi(Z_j) \leq -\beta$, hence $\liminf_{n\to\infty} \frac{1}{n} \log \sum_{j=0}^{n-1} \varphi(Z_j) \geq \beta$. Thuswe conclude that

$$\limsup_{n\to\infty} \frac{1}{n} \log \sum_{j=0}^{n-1} \varphi(Z_j) = \liminf_{n\to\infty} \frac{1}{n} \log \sum_{j=0}^{n-1} \varphi(Z_j) = \beta$$

which completes the proof. ■

### 3.1.2 Application to random linear cocycles

Consider the random linear cocycle $F$ generated by a probability measure $\mu \in \text{Prob}(\text{GL}_\text{d}(\mathbb{R}))$ with compact support $\Sigma = \text{supp}(\mu)$. The cocycle is defined as the skew-product map $F\colon \Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^d) \to \Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^d)$ such that $F(\underline{g}, \hat{v}) = (\sigma \underline{g}, \hat{g_0}\hat{v})$, where $\sigma\colon \Sigma^{\mathbb{N}} \to \Sigma^{\mathbb{N}}$ is the Bernoulli shift.

Consider the Markov kernel $K$ from the third level, defined in section 2.2.1, namely

$$\begin{aligned} K\colon \mathbb{P}(\mathbb{R}^d) &\to \text{Prob}(\mathbb{P}(\mathbb{R}^d)) \\ \hat{v} &\mapsto \int_\Sigma \delta_{\hat{g}_0\hat{v}}\, d\mu(g_0)\,. \end{aligned}$$

Consider the special observable defined in (2.5),

$$\Phi(g_0, \hat{v}) = \log\|g_0 v\|$$

where $v$ is a unitary representative of the projective point $\hat{v} \in \mathbb{P}(\mathbb{R}^d)$.

We define a linear functional $\alpha\colon \text{Prob}_K(\mathbb{P}(\mathbb{R}^d)) \to \mathbb{R}$ as follows:

$$\alpha(\eta) = \int_{\Sigma\times\mathbb{P}(\mathbb{R}^d)} \Phi(g_0, \hat{v})\, d\mu(g_0)d\eta(\hat{v}). \tag{3.1}$$

Moreover, since $\text{Prob}_K(\mathbb{P}(\mathbb{R}^d))$ is compact and $\alpha$ is continuous, its maximum is attained. Then let

$$\beta := \max\left\{\alpha(\eta)\colon \eta \in \text{Prob}_K(\mathbb{P}(\mathbb{R}^d))\right\}.$$

**Theorem 3.3** *For every vector $v \in \mathbb{R}^d\backslash\{0\}$ the following hold.*

*(i) For $\mu^{\mathbb{N}}$-almost every $\underline{g} \in \Sigma^{\mathbb{N}}$,*

$$\limsup_{n\to\infty} \frac{1}{n} \log\|g_{n-1}\dots g_1 g_0 v\| \le \beta.$$

*(ii) If $\alpha(\eta) = \beta$ for every $\eta \in \text{Prob}_K(\mathbb{P}(\mathbb{R}^d))$, then for $\mu^{\mathbb{N}}$-a.e $\underline{g} \in \Sigma^{\mathbb{N}}$,*

$$\lim_{n\to\infty} \frac{1}{n} \log\|g_{n-1}\dots g_1 g_0 v\| = \beta.$$

*Proof.* Consider the Markov kernel $\bar{K}$ from the second level, defined in section 2.2.1 by $\bar{K}\colon \Sigma \times \mathbb{P}(\mathbb{R}^d) \to \text{Prob}(\Sigma \times \mathbb{P}(\mathbb{R}^d))$, $\bar{K}(g_0, \hat{v}) = \mu \times \delta_{\hat{g}_0\hat{v}}$.

Fix any $\hat{v} \in \mathbb{P}(\mathbb{R}^d)$ and consider the $\bar{K}$-Markov chain

$$\bar{Z}_n \colon \Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^d) \to \Sigma \times \mathbb{P}(\mathbb{R}^d)$$
$$\bar{Z}_n(\underline{g}, \hat{v}) = (g_n,\, \hat{g}_{n-1} \dots \hat{g}_1 \hat{g}_0 \hat{v}) \,,$$

where the space $\Omega = \Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^d)$ is equipped with the probability $\mu^{\mathbb{Z}} \times \delta_{\hat{v}}$.

Recall that for all $(\underline{g}, \hat{v})$,

$$\frac{1}{n} \sum_{j=0}^{n-1} \Phi(Z_j(\underline{g}, \hat{v})) = \frac{1}{n} \log\| g_{n-1} \dots g_0 v \| \,.$$

By theorem 3.2, we conclude that for every $v \in \mathbb{R}^d \backslash \{0\}$ and $\mu^{\mathbb{N}}$-a.e $\underline{g} \in \Sigma^{\mathbb{N}}$,

$$\limsup_{n\to\infty} \frac{1}{n} \log\| g_{n-1} \dots g_1 g_0 v \| \leq \max \Big\{ \int_{\Sigma\times\mathbb{P}(\mathbb{R}^d)} \Phi \, dm \colon m \in \mathrm{Prob}_{\bar{K}}(\Sigma \times \mathbb{P}(\mathbb{R}^d)) \Big\}.$$

We claim that the right hand side of the previous equation is equal to

$$\max \Big\{ \int \Phi \, d\mu d\eta \colon \eta \in \mathrm{Prob}_K(\mathbb{P}(\mathbb{R}^d)) \Big\} = \beta \,,$$

where, again, $K$ is the Markov kernel from the third level. In fact, the following lemma implies the claim which will then conclude the proof of item (i).

**Lemma 3.2** *If $m \in \mathrm{Prob}_{\bar{K}}(\Sigma \times \mathbb{P}(\mathbb{R}^d))$ then there exists $\eta \in \mathrm{Prob}_K(\mathbb{P}(\mathbb{R}^d))$ such that $m = \mu \times \eta$. Conversely, if $\eta \in \mathrm{Prob}_K(\mathbb{P}(\mathbb{R}^d))$ then the product $\mu \times \eta \in \mathrm{Prob}_{\bar{K}}(\Sigma \times \mathbb{P}(\mathbb{R}^d))$.*

*Proof.* Consider the following projection nmap:

$$\Pi \colon C(\Sigma \times \mathbb{P}(\mathbb{R}^d)) \to C(\mathbb{P}(\mathbb{R}^d)), \quad \Pi\varphi(\hat{v}) = \int_{\Sigma} \varphi(g_0, \hat{v}) \, d\mu(g_0).$$

Note that if $m \in \mathrm{Prob}_{\bar{K}}(\Sigma \times \mathbb{P}(\mathbb{R}^d))$, then its associated Markov operator is given by

$$\bar{Q}\varphi(g_0, \hat{v}) = \int_{\Sigma} \varphi(g_1, \hat{g}_0 \hat{v}) \, d\mu(g_1) = \Pi\varphi(\hat{g}_0 \hat{v}).$$

If $\Pi\varphi_1 = \Pi\varphi_2$, then $\bar{Q}\varphi_1 = \bar{Q}\varphi_2$. Therefore,

$$\int_{\Sigma\times\mathbb{P}(\mathbb{R}^d)} \bar{Q}\varphi_1 \, dm = \int_{\Sigma\times\mathbb{P}(\mathbb{R}^d)} \bar{Q}\varphi_2 \, dm.$$

Since $m$ is a $\bar{K}$-stationary measure,

$$\int_{\Sigma\times\mathbb{P}(\mathbb{R}^d)} \varphi_1 \, dm = \int_{\Sigma\times\mathbb{P}(\mathbb{R}^d)} \varphi_2 \, dm.$$

For any $\Psi \in C(\mathbb{P}(\mathbb{R}^d))$, consider

$$I(\Psi)\colon = \int_{\Sigma\times\mathbb{P}(\mathbb{R}^d)} \varphi \, dm,$$

where $\Pi\varphi = \Psi$. Note that $I$ is well defined, since $\Pi\varphi_1 = \Pi\varphi_2$ implies that $\int \varphi_1 \; dm = \int \varphi_2 \; dm$. Moreover, $I$ is a positive linear functional, thus by the Riesz-Markov-Kakutani representation theorem there exists a measure $\eta \in \mathrm{Prob}(\mathbb{P}(\mathbb{R}^d))$ such that $I(\Psi) = \int \Psi \; d\eta$. Therefore, $\forall \varphi \in C(\Sigma \times \mathbb{P}(\mathbb{R}^d))$,

$$\int \varphi \; dm = I(\Psi) = \int \Psi \; d\eta = \int \Pi\varphi \; d\eta = \int\int \varphi \; d\mu d\eta$$

showing that $m = \mu \times \eta$.

It is easy to derive the $K$-stationarity of $\eta$ from that of $m$. Moreover, the reverse statement is a simple calculation. ■

Returning to the proof of the theorem, by the previous lemma we have that if $\alpha(\eta) = \beta$ for every $\eta \in \mathrm{Prob}_K(\mathbb{P}(\mathbb{R}^d))$ then $\int \Phi \; dm = \beta$ for every $m \in \mathrm{Prob}_{\bar{K}}(\Sigma \times \mathbb{P}(\mathbb{R}^d))$. Therefore, by item (ii) of theorem 3.2 we have

$$\lim_{n\to\infty} \frac{1}{n} \log\|g_{n-1} \dots g_1 g_0 v\| = \lim_{n} \frac{1}{n} \sum_{j=0}^{n-1} \Phi(Z_j) = \beta$$

which establishes item (ii) of the theorem and completes its proof. ■

**Theorem 3.4** *(Furstenberg's formula) Given any compactly supported probability measure $\mu$ on $\mathrm{GL_d}(\mathbb{R})$,*

$$L_1(\mu) = \max\left\{\int_{\Sigma\times\mathbb{P}(\mathbb{R}^d)} \Phi \; d\mu d\eta\colon \eta \in \mathrm{Prob}_K(\mathbb{P}(\mathbb{R}^d))\right\},$$

*where $\Phi\colon \Sigma \times \mathbb{P}(\mathbb{R}^d) \to \mathbb{R}$ is the observable given by $\Phi(g, \hat{v}) = \log \frac{\|gv\|}{\|v\|}$.*

*Proof.* Let $\{e_1, \dots, e_m\}$ be the canonical basis of $\mathbb{R}^d$ and let $g \in \mathrm{GL_d}(\mathbb{R})$. Consider the norm $\|g\|' = \max_j \|ge_j\|$. Note that for $\mu^{\mathbb{N}}$ almost every $\{g_n\}_n$

$$\begin{aligned} L_1(\mu) &= \lim_{n\to\infty} \frac{1}{n} \log\|g_{n-1} \dots g_0\| \\ &= \lim_{n\to\infty} \frac{1}{n} \log\|g_{n-1} \dots g_0\|' \\ &= \max_{1\le j\le m} \limsup_{n\to\infty} \frac{1}{n} \log\|g_{n-1} \dots g_0 e_j\| \le \beta, \end{aligned}$$

by item (i) of the previous theorem, so $L_1(\mu) \le \beta$.

We proceed to proving that $\beta \le L_1(\mu)$. Let

$$\mathcal{B} := \left\{\eta \in \mathrm{Prob}_K(\mathbb{P}(\mathbb{R}^d))\colon \alpha(\eta) = \beta\right\}.$$

Then $\mathcal{B}$ is a non empty, convex and closed set, therefore it is also compact. By Krein-Milman's theorem, there exists $\eta_0 \in \mathcal{B}$ which is an extremal point of $\mathcal{B}$.

Let us show that $\eta_0$ is also an extremal point of the whole set $\mathrm{Prob}_K(\mathbb{P}(\mathbb{R}^d))$. Indeed, if $\eta_0 = t\eta_1 + (1-t)\eta_2$ for some $\eta_1,\ \eta_2 \in \mathrm{Prob}_K(\mathbb{P}(\mathbb{R}^d))$ and $t \in (0,1)$, then $\alpha(\eta_0) = \beta = t\alpha(\eta_1) + (1-t)\alpha(\eta_2)$. Since $\alpha(\eta_1)$ and $\alpha(\eta_2) \le \beta$ and $\alpha(\eta_2) \le \beta$, both of them must be equal to $\beta$. Hence $\eta_1, \eta_2 \in \mathcal{B}$. Since we assumed $\eta_0$ to be an extremal point of $\mathcal{B}$, we conclude that $\eta_1 = \eta_2$, therefore $\eta_0$ is an extremal point of $\mathrm{Prob}_K(\mathbb{P}(\mathbb{R}^d))$.

Let $\tilde{\Phi} \colon \Sigma^{\mathbb{Z}} \times \mathbb{P}(\mathbb{R}^d) \to \mathbb{R}$, $\tilde{\Phi}(\underline{g}, \hat{v}) := \Phi(g_0, \hat{v})$. We then clearly have that the stochastic Birkhoff sums of the observable $\Phi$ corresponding to the Markov chain $\{Z_n\}$ are equal to the Birkhoff sums of the observable $\tilde{\Phi}$ relative to the projective cocycle dynamics $\hat{F} \colon \Sigma^{\mathbb{Z}} \times \mathbb{P}(\mathbb{R}^d) \to \Sigma^{\mathbb{Z}} \times \mathbb{P}(\mathbb{R}^d)$, $\hat{F}(\underline{g}, \hat{v}) = (\sigma \underline{g}, \hat{g}_0 \hat{v})$, namely $S_n\Phi = S_n\tilde{\Phi}$ for all $n \ge 1$.

By proposition 2.2, $\mu^{\mathbb{N}} \times \eta_0$ is ergodic for the projective cocycle $\hat{F}$. Therefore, by Birkhoff's ergodic theorem, for $\mu^{\mathbb{N}} \times \eta_0$-a.e $(\underline{g}, \hat{v})$ we have

$$\frac{1}{n} \log\|g_{n-1} \dots g_0 v\| = \frac{1}{n} S_n\Phi(\underline{g}, \hat{v}) = \frac{1}{n} S_n\tilde{\Phi}(\underline{g}, \hat{v}) \to \int \tilde{\Phi}\, d\mu^{\mathbb{N}} d\eta_0.$$

But

$$\int \tilde{\Phi}\, d\mu^{\mathbb{N}} d\eta_0 = \int \Phi\, d\mu d\eta_0 = \alpha(\eta_0) = \beta.$$

We conclude that for $\mu^{\mathbb{N}}$-a.e. $\underline{g}$ and for $\eta_0$-a.e $\hat{v}$,

$$\beta = \lim_{n\to\infty} \frac{1}{n} \log\|g_{n-1} \dots g_0\, v\| \le \limsup_{n\to\infty} \frac{1}{n} \log\|g_{n-1} \dots g_0\| = L_1(\mu)$$

which completes the proof of the theorem. ■

## 3.2 Furstenberg-Kifer's multiplicative ergodic theorem

### 3.2.1 The construction of the filtration

Let $\mu$ be a compactly supported probability measure on $\mathrm{GL_d}(\mathbb{R})$ and let $K$ be the corresponding Markov kernel as above. Consider the set of values at the extremal points of the linear functional $\alpha$ defined above, that is, let

$$\mathcal{S}(\mu) := \left\{ \alpha(\eta) : \eta \text{ is an extremal point of } \mathrm{Prob}_K(\mathbb{P}(\mathbb{R}^d)) \right\}.$$

**Lemma 3.3** $\mathcal{S}(\mu) \subset \{L_1(\mu), \dots, L_d(\mu)\}$. *In particular, $\mathcal{S}(\mu)$ is finite.*

*Proof.* Let $\tilde{\Phi}\colon \Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^d) \to \mathbb{R}$ given by $\tilde{\Phi}(\underline{g}, \hat{v}) = \log \frac{\|g_0 v\|}{\|v\|}$ denote the natural extension of $\Phi$ as in the previous proof. Let $\eta$ be an extremal point of $\mathrm{Prob}_K(\mathbb{P}(\mathbb{R}^d))$. Then by proposition 2.2, the measure $\mu^{\mathbb{N}} \times \eta$ is $\hat{F}$-ergodic, so, by Birkhoff's ergodic theorem,

$$\begin{aligned}\alpha(\eta) &= \int \Phi(g_0, \hat{v})\, d\mu(g_0) d\eta(\hat{v}) = \int \tilde{\Phi}(\underline{g}, \hat{v})\, d\mu^{\mathbb{N}}(\underline{g}) d\eta(\hat{v}) \\ &= \lim_{n\to\infty} \frac{1}{n} \log\|g_{n-1} \dots g_0 v\|.\end{aligned}$$

By the Oseledets multiplicative ergodic theorem, this limit is equal to one of the Lyapunov exponents, thus we conclude that $\mathcal{S}(\mu)$ is contained in the Lyapunov spectrum of $\mu$. ■

Let us introduce some notation. In what follows, we often write $A^n(\underline{g})$ to denote the product $g_{n-1} \dots g_0$. Let $\beta_0 > \beta_1 > \dots > \beta_r$ denote the elements of $\mathcal{S}(\mu)$. We will call them the *Furstenberg-Kifer exponents*. Note that, by Furstenberg's formula, we already know that $L_1(\mu) = \beta = \beta_0$.

**Theorem 3.5** *(Furstenberg-Kifer's nonrandom filtration) Given a measure $\mu \in \mathrm{Prob}(\mathrm{GL_d}(\mathbb{R}))$ with $\Sigma := \mathrm{supp}(\mu)$ a compact set, there exists a filtration of $\mathbb{R}^d$*

$$\{0\} = \mathcal{L}_{r+1} \subsetneq \mathcal{L}_r \subsetneq \dots \subsetneq \mathcal{L}_1 \subsetneq \mathcal{L}_0 = \mathbb{R}^d,$$

*such that the following hold.*

*(i) $\forall\ 0 \le j \le r$, $\mathcal{L}_j$ is $\mu$-invariant, i.e. $g\mathcal{L}_j = \mathcal{L}_j$ for $\mu$-a.e. $g \in \Sigma$.*

*(ii) $\forall\ 0 \le j \le r$ and $\forall v \in \mathcal{L}_j \backslash \mathcal{L}_{j+1}$ we have that*

$$\lim_{n\to\infty} \frac{1}{n} \log\|A^n(\underline{g}) v\| = \beta_j \qquad \text{for } \mu^{\mathbb{N}}\text{-a.e } \underline{g} \in \Sigma^{\mathbb{N}}.$$

*(iii) $\forall\ 0 \le j \le r$, if $\eta$ is an extremal point of $\mathrm{Prob}_K(\mathbb{P}(\mathbb{R}^d))$ such that $\alpha(\eta) = \beta_j$, then $\eta(\hat{\mathcal{L}}_j) = 1$ and $\eta(\hat{\mathcal{L}}_{j+1}) = 0$, where $\hat{\mathcal{L}}_j = \{\hat{v}\colon v \in \mathcal{L}_j \setminus \{0\}\}$.*

*Moreover, each subspace $\mathcal{L}_j$ of the filtration is given explicitly by*

$$\mathcal{L}_j = \left\{ v \in \mathbb{R}^m\colon \limsup_{n\to\infty} \frac{1}{n} \log\|A^n(\underline{g}) v\| \le \beta_j \quad \text{for } \mu^{\mathbb{N}}\text{-a.e. } \underline{g} \in \Sigma^{\mathbb{N}} \right\}.$$

The argument is quite elaborate and consists of various lemmas and other technical results and concepts which will be described in this and in the next subsection.

**Lemma 3.4** *If the linear functional $\alpha$ is constant on $Prob_K(\mathbb{P}(\mathbb{R}^d))$ then the theorem holds for the trivial filtration $\mathcal{L}_0 = \mathbb{R}^d$ and $\mathcal{L}_1 = \{0\}$.*

*Proof.* Since $\alpha(\eta) = \beta$ for every $\eta \in \mathrm{Prob}_K(\mathbb{P}(\mathbb{R}^d))$, by theorem 3.3, for all $v \neq 0$ and $\mu^{\mathbb{N}}$-a.e. $\underline{g}$,

$$\lim_{n\to\infty} \frac{1}{n} \log\|g_{n-1}\dots g_0 v\| = \beta.$$

This immediately shows that in this case the trivial filtration $\mathcal{L}_0 = \mathbb{R}^d$ and $\mathcal{L}_1 = \{0\}$ satisfies items (i), (ii) and (iii) and the result follows. ∎

We will assume from now on that the linear functional $\alpha$ is not constant in $\mathrm{Prob}_K(\mathbb{P}(\mathbb{R}^d))$. By Krein-Milman's theorem, since there exist extremal points in $\mathrm{Prob}_K(\mathbb{P}(\mathbb{R}^d))$, the restriction of $\alpha$ to the extremal points of $\mathrm{Prob}_K(\mathbb{P}(\mathbb{R}^d))$ is also not constant. Therefore $\#\mathcal{S}(\mu) > 1$.

In the next lemma we define the first non trivial subspace of the filtration.

**Lemma 3.5** *Consider the following set:*

$$\mathcal{L}_1 = \left\{ v \in \mathbb{R}^d \colon \limsup_{n\to\infty} \frac{1}{n} \log\|g_{n-1}\dots g_0 v\| \leq \beta_1 \quad \text{for } \mu^{\mathbb{N}}\text{-a.e. } \underline{g} \in \Sigma^{\mathbb{N}} \right\}.$$

*The following statements hold.*

*(i) $\mathcal{L}_1$ is a vector subspace of $\mathbb{R}^d$.*

*(ii) $\mathcal{L}_1$ is $\mu$-invariant.*

*(iii) If $\eta$ is an extremal point of $\mathrm{Prob}_Q(\mathbb{P}(\mathbb{R}^d))$ such that $\alpha(\eta) = \beta_1$, then $\eta(\hat{\mathcal{L}}_1) = 1$.*

*(iv) If $\eta$ is an extremal point of $\mathrm{Prob}_Q(\mathbb{P}(\mathbb{R}^d))$ such that $\alpha(\eta) = \beta_0$, then $\eta(\hat{\mathcal{L}}_1) = 0$.*

*(v) $\mathcal{L}_1$ is proper.*

*Proof.* If $v \in \mathcal{L}_1$, then for any scalar $\lambda \neq 0$,

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log\|A^n(\underline{g})\lambda v\| &= \limsup_{n\to\infty} \frac{1}{n} \log\|A^n(\underline{g})v\| + \limsup_{n\to\infty} \frac{1}{n} \log|\lambda| \\ &= \limsup_{n\to\infty} \frac{1}{n} \log\|A^n(\underline{g})v\| \leq \beta_1.\end{aligned}$$

If $\lambda = 0$ then the lim sup above is equal to $-\infty$ and the inequality still holds. Hence $\lambda v \in \mathcal{L}_1$. Furthermore, given $v_1, v_2 \in \mathcal{L}_1$, then

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log\|A^n(\underline{g})(v_1+v_2)\| &\leq \limsup_{n\to\infty} \frac{1}{n} \log\left[2\max\{\|A^n(\underline{g})v_1\|, \|A^n(\underline{g})v_2\|\}\right] \\ &\leq \limsup_{n\to\infty} \max\left\{\frac{1}{n}\log\|A^n(\underline{g}v_1)\|, \frac{1}{n}\log\|A^n(\underline{g}v_2)\|\right\} + \log\frac{2}{n} \leq \beta_1.\end{aligned}$$

This implies that $v_1 + v_2 \in \mathcal{L}_1$ and that $\mathcal{L}_1$ is a vector subspace of $\mathbb{R}^d$.

By definition of $\mathcal{L}_1$, if $v \in \mathcal{L}_1$, then

$$\limsup_{n\to\infty} \frac{1}{n} \log\|A^n(\underline{g})v\| \leq \beta_1$$

for $\mu^{\mathbb{N}}$-a.e. $(g_0,\ g_1, \dots)$. Thus $g_0 v \in \mathcal{L}_1$ for $\mu$-a.e $g_0$. Therefore $g_0\mathcal{L}_1 \subset \mathcal{L}_1$. Since $g_0$ is invertible, $\dim g_0\mathcal{L}_1 = \dim \mathcal{L}_1$. Hence $g_0\mathcal{L}_1 = \mathcal{L}_1$ for $\mu$-a.e $g_0$ and we conclude that $\mathcal{L}_1$ is $\mu$-invariant.

Let $\eta_0$ be an extremal point of $\mathrm{Prob}_Q(\mathbb{P}(\mathbb{R}^d))$ such that $\alpha(\eta_0) = \beta_0$. Hence, $\mu^{\mathbb{N}} \times \eta_0$ is $\hat{F}$-ergodic and by Birkhoff's theorem

$$\lim_{n\to\infty} \frac{1}{n} \log\|A^n(\underline{g})v\| = \alpha(\eta_0) = \beta_0 > \beta_1$$

for $\mu^{\mathbb{N}} \times \eta_0$-a.e $(\underline{g}, \hat{v})$. Hence, for $\eta_0$-a.e $\hat{v}$ the previous limit holds for $\mu^{\mathbb{N}}$-a.e $\underline{g}$ and we conclude that $\eta_0(\hat{\mathcal{L}}_1) = 0$.

Similarly, let $\eta_1$ be an extremal point of $\mathrm{Prob}_Q(\mathbb{P}(\mathbb{R}^d))$ and assume that $\alpha(\eta_1) = \beta_1$. Then the measure $\mu^{\mathbb{N}} \times \eta_1$ is ergodic and by Birkhoff's ergodic theorem

$$\lim_{n\to\infty} \frac{1}{n} \log\|A^n(\underline{g})v\| = \alpha(\eta_1) = \beta_1$$

for $\mu^{\mathbb{N}} \times \eta_1$-a.e $(\underline{g}, \hat{v})$. Moreover, for $\eta_1$-a.e $\hat{v}$ the previous limit holds for $\mu^{\mathbb{N}}$-a.e $\underline{g}$. Hence $\eta_1(\hat{\mathcal{L}}_1) = 1$.

By the previous two arguments, $\mathcal{L}_1$ is a proper subspace of $\mathbb{R}^d$, which concludes the proof of the lemma. ■

### 3.2.2 The induced cocycle

Let $\mathcal{L} \subsetneq \mathbb{R}^d$ be a proper $\mu$-invariant subspace, i.e $g_0\mathcal{L} = \mathcal{L}$ for $\mu$-a.e $g_0 \in \Sigma$ and $\dim \mathcal{L} = k < d$. One can then consider the induced (or restricted) cocycle

$$F_{\mathcal{L}} \colon \Sigma^{\mathbb{N}} \times \mathcal{L} \to \Sigma^{\mathbb{N}} \times \mathcal{L}, \qquad F_{\mathcal{L}}(\underline{g}, v) = (\sigma\underline{g}, g_0 v).$$

Furthermore, we also consider its projective version:

$$\hat{F}_{\mathcal{L}} : \Sigma^{\mathbb{N}} \times \mathbb{P}(\mathcal{L}) \to \Sigma^{\mathbb{N}} \times \mathbb{P}(\mathcal{L}), \qquad \hat{F}_{\mathcal{L}}(\underline{g}, \hat{v}) = (\sigma\underline{g}, \hat{g}_0\hat{v}).$$

In fact, all the theory presented in the previous sections applies in the same way to the cocycle restricted to the invariant subspace $\mathcal{L}$. Let us summarize the main objects and introduce the relevant notations.

Define the observable $\Phi_{\mathcal{L}} = \Phi|_{\Sigma\times\mathcal{L}}$ and the induced Markov operator $Q_{\mathcal{L}} = Q|_{\mathbb{P}(\mathcal{L})}$. Similarly one may consider the induced linear functional

$\alpha_{\mathcal{L}}\colon \mathrm{Prob}_{Q_{\mathcal{L}}}(\mathbb{P}(\mathcal{L})) \to \mathbb{R}$ and its maximum value $\beta_{\mathcal{L}}$.

We can also consider the quotient vector space $\mathbb{R}^d/\mathcal{L} = \{[v]\colon v \in \mathbb{R}^d\}$, where $v$ is in the same equivalence class as $w$ if $v - w \in \mathcal{L}$. Analogously to what we did before, we can also induce a cocycle in the quotient vector space. Therefore it is also possible to define $F_{\mathbb{R}^d/\mathcal{L}}$, $\hat{F}_{\mathbb{R}^d/\mathcal{L}}$, $\Phi_{\mathbb{R}^d/\mathcal{L}}$, $Q_{\mathbb{R}^d/\mathcal{L}}$, $\alpha_{\mathbb{R}^d/\mathcal{L}}$, $\beta_{\mathbb{R}^d/\mathcal{L}}$.

Remember that $\dim \mathcal{L} = k < d$. Let $\{e_1, \dots, e_k\}$ be a basis in $\mathcal{L}$ and $\{E_{k+1}, \dots, E_d\}$ be a basis in $\mathbb{R}^d/\mathcal{L}$. Choose the representatives $e_{k+1} \in E_{k+1}, \dots, e_d \in E_d$. Then $\{e_1, \dots, e_d\}$ is a basis in $\mathbb{R}^d$. Hence, by a change of coordinates, we identify $\mathcal{L} = \mathbb{R}^k \times (0) \subset \mathbb{R}^d$ and $\mathbb{R}^d/\mathcal{L} = (0) \times \mathbb{R}^{d-k} \subset \mathbb{R}^d$. Thus, in this basis, any matrix $g \in \Sigma$ is written in blocks as $g = \begin{bmatrix} b & c \\ 0 & d \end{bmatrix}$, where $b$ is a $k \times k$ block, $c$ is a $k \times (m-k)$ block and $d$ is an $(m-k) \times (m-k)$ block. Therefore, the cocycle $A : \Sigma^{\mathbb{N}} \to \mathrm{GL_d}(\mathbb{R})$ and its iterates are written as

$$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}, \quad \text{and} \quad A^n = \begin{bmatrix} B^n & C_n \\ 0 & D^n \end{bmatrix}, \tag{3.2}$$

where $B$ is the cocycle $F_{\mathcal{L}}\colon \Sigma^{\mathbb{N}} \to \mathrm{GL_k}(\mathbb{R})$ and $D$ is the cocycle $F_{\mathbb{R}^d/\mathcal{L}}\colon \Sigma^{\mathbb{N}} \to \mathrm{GL_{d-k}}$. Moreover, $B^n$, $D^n$ are the iterates of the induced cocycles and $C_n = \sum_{i=0}^{n-1} B^{n-i-1} C D^i$.

**Lemma 3.6** *If $\mathcal{L}$ is a $\mu$-invariant subspace of $\mathbb{R}^d$, then $\beta = \max\{\beta_{\mathcal{L}}, \beta_{\mathbb{R}^d/\mathcal{L}}\}$.*

*Proof.* By Furstenberg's formula, $\beta = L_1(F)$. Applying Furstenberg's formula to the induced cocycles $F_{\mathcal{L}}$ and $F_{\mathbb{R}^d/\mathcal{L}}$, we conclude that $\beta_{\mathcal{L}} = L_1(F_{\mathcal{L}})$ and $\beta_{\mathbb{R}^d/\mathcal{L}} = L_1(F_{\mathbb{R}^d/\mathcal{L}})$. Note that by the decomposition above,

$$\begin{aligned} \beta = L_1(F) = \lim_{n\to\infty} \frac{1}{n} \log\|A^n\| &= \max\left\{ \lim_{n\to\infty} \frac{1}{n} \log\|B^n\|, \lim_{n\to\infty} \frac{1}{n} \log\|D^n\| \right\} \\ &= \max\left\{ L_1(F_{\mathcal{L}}), L_1(F_{\mathbb{R}^d/\mathcal{L}}) \right\} \\ &= \max\{\beta_{\mathcal{L}}, \beta_{\mathbb{R}^d/\mathcal{L}}\}. \end{aligned}$$

■

**Lemma 3.7** *Let $\mathcal{L}_1$ be the $\mu$-invariant subspace defined in lemma 3.5. Then $\beta_{\mathcal{L}_1} = \beta_1$.*

*Proof.* By Furstenberg's formula applied to the induced cocycle $F_{\mathcal{L}_1}$, for $\mu^{\mathbb{N}}$-a.e. $\underline{g} \in \Sigma^{\mathbb{N}}$ we have that

$$\begin{aligned} \beta_{\mathcal{L}_1} = L_1(F_{\mathcal{L}_1}) &= \lim_{n\to\infty} \frac{1}{n} \log\|A^n_{\mathcal{L}_1}(\underline{g})\| \\ &= \max_{e_j \in \mathcal{L}_1} \lim_{n\to\infty} \frac{1}{n} \log\|A^n(\underline{g}) e_j\| \leq \beta_1. \end{aligned}$$

On the other hand, $\beta_1 = \alpha(\eta_1)$ for some $\eta_1$ which is an extremal point of $\text{Prob}_Q(\mathbb{P}(\mathbb{R}^d))$. By Birkhoff's ergodic theorem,

$$\beta_1 = \alpha(\eta_1) = \lim_{n\to\infty} \frac{1}{n} \log\|A^n(\underline{g})v\|$$

for $\mu^n \times \eta_1$-a.e. $(\underline{g}, \hat{v})$. Hence there exists $\hat{w} \in \mathbb{P}(\mathbb{R}^d)$ such that

$$\frac{1}{n} \log\|A^n(\underline{g})w\| \to \beta_1$$

for almost every $\underline{g} \in \Sigma^{\mathbb{N}}$, thus $w \in \mathcal{L}_1$. Furthermore,

$$\beta_1 = \lim_{n\to\infty} \frac{1}{n} \log\|A^n(\underline{g})w\| \leq \lim_{n\to\infty} \frac{1}{n} \log\|A^n_{\mathcal{L}_1}(\underline{g})\| = \beta_{\mathcal{L}_1},$$

thus establishing the equality of the two quantities. ■

**Corollary 3.1** *$\beta_{\mathbb{R}^d/\mathcal{L}_1} = \beta$.*

*Proof.* By lemma 3.6, $\beta = \max\left\{\beta_{\mathcal{L}_1}, \beta_{\mathbb{R}^d/\mathcal{L}_1}\right\}$. By lemma 3.7 we have $\beta_{\mathcal{L}_1} = \beta_1 < \beta_0 = \beta$. It follows that $\beta = \beta_{\mathbb{R}^d/\mathcal{L}_1}$. ■

**Lemma 3.8** *If $\mathcal{L} \subset \mathbb{R}^d$ is a $\mu$-invariant subspace and $\beta_{\mathcal{L}} < \beta$, then $\beta_{\mathcal{L}} \leq \beta_1$.*

*Proof.* Recall that

$$\beta_{\mathcal{L}} = \max\left\{\alpha_{\mathcal{L}}(\eta_{\mathcal{L}})\colon \eta_{\mathcal{L}} \text{ is an extremal point of } \text{Prob}_{Q_{\mathcal{L}}}(\mathbb{P}(\mathcal{L}))\right\}.$$

Let $\eta_{\mathcal{L}}$ be such that $\beta_{\mathcal{L}} = \alpha_{\mathcal{L}}(\eta_{\mathcal{L}})$. Consider the extension $\eta \in \text{Prob}(\mathbb{P}(\mathbb{R}^d))$ of $\eta_{\mathcal{L}}$ such that $\eta(\mathbb{P}(\mathcal{L})^c) = 0$. We claim that $\eta$ is $Q$ stationary and it is an extremal point of $\text{Prob}_Q(\mathbb{P}(\mathbb{R}^d))$. For every $\varphi \in L^\infty(\mathbb{P}(\mathbb{R}^d))$,

$$\int Q\varphi \, d\eta = \int_{\mathbb{P}(\mathcal{L})} Q_{\mathcal{L}}\varphi_{\mathcal{L}} \, d\eta_{\mathcal{L}} = \int \varphi_{\mathcal{L}} \, d\eta_{\mathcal{L}} = \int \varphi \, d\eta.$$

Moreover, if $\eta$ is not extremal, then $\eta = t\eta_1 + (1-t)\eta_2$ for some $\eta_1 \neq \eta_2$. Since $\eta(\mathbb{P}(\mathcal{L})^c) = 0$, both $\eta_i(\mathbb{P}(\mathcal{L})^c) = 0$ for $i = 1, 2$. However $\eta_{\mathcal{L}}$ is extremal, therefore $\eta_1$ must be equal $\eta_2$ and we get a contradiction. Thus, we conclude that

$$\beta_{\mathcal{L}} = \alpha_{\mathcal{L}}(\eta_{\mathcal{L}}) = \int \Phi_{\mathcal{L}} \, d\eta_{\mathcal{L}} = \int \Phi \, d\eta,$$

where $\eta$ is an extremal point of $\text{Prob}_Q(\mathbb{P}(\mathbb{R}^d))$. Hence, $\beta_{\mathcal{L}} = \int \Phi \; d\eta \in \{\beta_0, \beta_1, \ldots, \beta_r\}$. By hypothesis, $\beta_{\mathcal{L}} < \beta_0$, so it must be that $\beta_{\mathcal{L}} \leq \beta_1$. ■

**Lemma 3.9** *The linear functional $\alpha_{\mathbb{R}^d/\mathcal{L}_1}\colon \text{Prob}_{Q_{\mathbb{R}^d/\mathcal{L}_1}}(\mathbb{P}(\mathbb{R}^d/\mathcal{L}_1)) \to \mathbb{R}$ is constant.*

*Proof.* Assume by contradiction that $\alpha_{\mathbb{R}^d/\mathcal{L}_1}$ is not constant. Therefore, there exists a proper subspace $V_1 \subset \mathbb{R}^d/\mathcal{L}_1$ which is invariant. Hence, $V_1 = \mathcal{L}/\mathcal{L}_1$, where $\mathcal{L} \subset \mathbb{R}^d$ is an invariant subspace such that $\mathcal{L} \supset \mathcal{L}_1$. Note that $\beta_{V_1} < \beta_{\mathbb{R}^d/\mathcal{L}_1} = \beta$. Moreover,

$$\beta_{\mathcal{L}} = \max\left\{\beta_{\mathcal{L}_1}, \beta_{\mathcal{L}/\mathcal{L}_1}\right\} = \max\left\{\beta_{\mathcal{L}_1}, \beta_{V_1}\right\} < \beta.$$

Thus, by lemma 3.8, $\beta_{\mathcal{L}} \leq \beta_1$. By the definition of $\mathcal{L}_1$, this implies that $\mathcal{L} \subset \mathcal{L}_1$. We conclude that $\mathcal{L} = \mathcal{L}_1$, so $V_1$ is a trivial subspace, contradicting the fact of being proper. ∎

**Lemma 3.10** *If $v \in \mathbb{R}^d/\mathcal{L}_1$ then for $\mu^{\mathbb{N}}$-a.e. $\underline{g} \in \Sigma^{\mathbb{N}}$ we have that*

$$\lim_{n\to\infty} \frac{1}{n} \log\|A^n(\underline{g})v\| = \beta_0.$$

*Proof.* We already know that for every $v \neq 0$ and for $\mu^{\mathbb{N}}$-a.e $\underline{g} \in \Sigma^{\mathbb{N}}$,

$$\limsup_{n\to\infty} \frac{1}{n} \log\|A^n(\underline{g})v\| \leq \beta_0.$$

Therefore it suffices to prove that for every $v \in \mathbb{R}^d/\mathcal{L}_1$ and for $\mu^{\mathbb{N}}$-a.e $\underline{g}$,

$$\liminf_{n\to\infty} \frac{1}{n} \log\|A^n(\underline{g})v\| \geq \beta_0.$$

By the change of coordinates described before, we can write the cocycle $A$ and its iterates as in (3.2), where $B$ represents the cocycle induced in the invariant subspace $\mathcal{L}_1$ and $D$ the cocycle induced in the quotient $\mathbb{R}^d/\mathcal{L}_1$. Then, a vector $v \in \mathbb{R}^d$ can be written as $v = (v_1, v_2)$ with $v_1 \in \mathbb{R}^k$ and $v_2 \in \mathbb{R}^{d-k}$. Let $v \in \mathbb{R}^d/\mathcal{L}_1$, then $v = (v_1, v_2)$ with $v_2 \neq 0$. Thus

$$A^n(\underline{g})v = \begin{pmatrix} B^n(\underline{g}) & C_n(\underline{g}) \\ 0 & D^n(\underline{g}) \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} B^n(\underline{g})v_1 + C_n v_2 \\ D^n(\underline{g})v_2 \end{pmatrix}.$$

Hence $\|A^n(\underline{g})v\| \geq \|D^n(\underline{g})v\|$ for every $n \in \mathbb{N}$ and for every $\underline{g} \in \Sigma^{\mathbb{N}}$. Moreover, it holds that

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \log\|A^n(\underline{g})v\| &\geq \liminf_{n\to\infty} \frac{1}{n} \log\|D^n(\underline{g})v_2\| \\ &= \lim_{n\to\infty} \frac{1}{n} \log\|D^n(\underline{g})v_2\| = \beta_{\mathbb{R}^d/\mathcal{L}_1} = \beta_0\end{aligned}$$

which establishes the claim. ∎

We are finally ready to complete the proof of theorem 3.5.

*Proof.* Recall that $r+1$ is the cardinality of the set of values of the linear functional $\alpha(\eta)$ at extremal points of $\text{Prob}_K(\mathbb{P}(\mathbb{R}^d))$. The case $r = 0$ was treated in lemma 3.4. Combining all the previous results, we conclude that if $r = 1$, then theorem 3.5 holds. Otherwise, $r \geq 2$ and we can apply the same procedure to the induced cocycle $F_{\mathcal{L}_1}$ to get another invariant subspace $\mathcal{L}_2$, which is a proper invariant subspace of $\mathcal{L}_1$ such that for every $v \in \mathcal{L}_1 \backslash \mathcal{L}_2$ and $\mu^{\mathbb{N}}$-a.e $\underline{g}$,

$$\lim_{n\to\infty} \frac{1}{n} \log\|A^n(\underline{g})v\| = \beta_1.$$

The completion of the proof follows by induction. ■

**Remark 3.1** *Although both Oseledets and Furstenberg-Kifer's multiplicative ergodic theorems produce filtrations, they do not need to coincide, see example 3.2 in [3]. As a consequence, Furstenberg-Kifer's exponents also do not need to coincide with the Lyapunov exponents. What holds, however, is that the set of Furstenberg-Kifer exponents is contained in the Lyapunov spectrum (the set of all Lyapunov exponents given by the Furstenberg-Kesten or the Oseledets theorem), as we saw in lemma 3.3. Moreover, by Furstenberg's formula, the first (largest) exponents in each set coincide.*

**Remark 3.2** *Another interesting fact is that the Oseledets filtration is random in the sense that for each point $\underline{g} \in \Sigma^{\mathbb{N}}$, the filtration possibly depends on all of its coordinates. On the other hand, Furstenberg-Kifer's filtration is non-random, in the sense that it does not depend on the point $\underline{g} \in \Sigma^{\mathbb{N}}$. In chapter 8 we will introduce a version of Furstenberg-Kifer's filtration for Markovian cocycles, which turns out to be only slightly random, since it depends just on the zeroth coordinate of $\underline{g}$.*

## 3.3 Irreducibility

In this section we discuss the concept of irreducibility, which is very important in the study of the Lyapunov exponents.

### 3.3.1 Introduction to the concept

There are many different versions of this concept in the literature, but we introduce only three of them. For a more detailed study of the topic, see [8] and the references therein.

Let $\mu$ be a probability measure on $\mathrm{GL_d}(\mathbb{R})$. Then $\mu$ is

(i) *strongly irreducible* if there is no finite family of proper subspaces of $\mathbb{R}^d$ which is invariant under $\mu$-a.e. $g \in \mathrm{GL_d}(\mathbb{R})$;

(ii) *irreducible* if there is no proper subspace of $\mathbb{R}^d$ which is invariant under $\mu$-a.e. $g \in \mathrm{GL_d}(\mathbb{R})$;

(iii) *quasi-irreducible* if it is irreducible or else there exists a proper subspace $V$ of $\mathbb{R}^d$ invariant under $\mu$-a.e. $g \in \mathrm{GL_d}(\mathbb{R})$ but the Lyapunov exponent of the cocycle restricted to $V$ is equal to the top Lyapunov exponent of the cocycle, that is, $L_1(F|_V) = L_1(F)$.

**Remark 3.3** *When a random linear cocycles satisfies one of the previous irreducibility conditions, it is often referred to as being in a generic setting. This is due to [33, Theorem 1.1], where Kifer proved that irreducibility is an open and dense property in* $\mathrm{Prob}(\mathrm{GL_d}(\mathbb{R}))$ *with respect to the weak* topology.*

**Remark 3.4** *If the Lyapunov exponent of the cocycle restricted to a maximal proper invariant subspace is strictly smaller than the top Lyapunov exponent of the cocycle, then that subspace is referred to as the "equator" in [3].*

It is clear that strong irreducibility implies irreducibility which implies quasi-irreducibility. We provide some examples of cocycles that satisfy the definitions above.

**Example 3.2** *A strongly irreducible cocycle:*

$$g_1 = \begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \qquad g_2 = \begin{pmatrix} \cos(2\pi\theta) & -\sin(2\pi\theta) \\ \sin(2\pi\theta) & \cos(2\pi\theta) \end{pmatrix}$$

In this example, we consider $\theta \in \mathbb{R}\backslash\mathbb{Q}$ and both matrices with positive weights: $p_1 > 0$, $p_2 > 0$ and $p_1 + p_2 = 1$. Note that $g_1$ is a hyperbolic matrix that fixes both the $x$ and the $y$ axes and $g_2$ is an irrational rotation that does not leave any finite family of subspaces invariant.

Note also that by a theorem of Furstenberg, the Lyapunov exponent of this cocycle is positive. Moreover, if the $p_1 = 0$, the cocycle generated by $g_2$ is still strongly irreducible, but has zero Lyapunov exponent.

**Example 3.3** *An irreducible cocycle which is not strongly irreducible:*

$$g_1 = \begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \qquad g_2 = \begin{pmatrix} \cos(\frac{\pi}{2}) & -\sin(\frac{\pi}{2}) \\ \sin(\frac{\pi}{2}) & \cos(\frac{\pi}{2}) \end{pmatrix}$$

This is a slight variation of the previous example. In this case, we chose $\theta$ to be rational ($\theta = \frac{1}{4}$) and the situation is completely different. In this case, $g_2$ still does not leave any subspace invariant, but it keeps the finite family of $x$ and $y$ axes invariant. Therefore, this cocycle is strongly irreducible but not irreducible.

Moreover, note that if $p_1 > 0$ and $p_2 > 0$, then the cocycle has zero Lyapunov exponent. On the other hand, if $p_2 = 0$, the Lyapunov exponent is positive. This is known is the "Kifer's example" of discontinuity of the Lyapunov exponents.

**Example 3.4** *A quasi-irreducible cocycle which is not irreducible:*

$$g_1 = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$$

The cocycle generated by this triangular matrix preserves the $x$-axis. Thus it is not irreducible. However, the Lyapunov exponent along the $x$-axis is equal to $\log 2$, which is the same as the top Lyapunov exponent generated by $g_1$, hence it is a quasi-irreducible cocycle.

In [8], the authors provide a general criterion for the quasi-irreducibility of an $\mathrm{SL}_2(\mathbb{R})$-valued triangular cocycle, depending on the matrix entries and the weights (the probability vector).

**Example 3.5** *A (completely) reducible cocycle (or not quasi-irreducible):*

$$g_1 = \begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

This is a diagonal cocycle, so it preserves both axes. Along the $x$-axis the Lyapunov exponent is maximal, however, along the $y$-axis the Lyapunov exponent is smaller than the maximal. Therefore, this cocycle is not quasi-irreducible.

### 3.3.2 Relation with Furstenberg-Kifer's filtration

In this manuscript we work with the third, hence weakest concept of irreducibility, the quasi-irreducibility. Furstenberg-Kifer's non-random filtration in theorem 3.5 provides the following characterization of this concept.

**Corollary 3.2** *A random cocycle (or equivalently, the measure generating it) is quasi-irreducible if and only if its corresponding Furstenberg-Kifer filtration is trivial, that is, $\mathcal{L}_0 = \mathbb{R}^d$ and $\mathcal{L}_1 = \{0\}$.*

*Proof.* If a cocycle is quasi-irreducible, then either it does not have any proper invariant subspace, in which case Furstenberg-Kifer's filtration is trivial, or else it has an invariant subspace $V$ and

$$\frac{1}{n}\log\|g_{n-1}\dots g_1 g_0 v\| \to L_1(\mu) = \beta_0 = \beta \tag{3.3}$$

for $\mu^{\mathbb{N}}$-a.e. $\underline{g} \in \Sigma^{\mathbb{N}}$ and every nonzero $v \in V$. In this case the filtration is also trivial.

If we assume that the filtration is trivial, then equation (3.3) holds for $\mu^{\mathbb{N}}$-a.e. $\underline{g} \in \Sigma^{\mathbb{N}}$ and for every $v \in \mathbb{R}^d$. Thus if the cocycle admits an invariant subspace, its top Lyapunov exponent must be attained along it, as it is attained in any direction. ■

# 4
# Hölder Continuity of the First Lyapunov Exponent in the Generic Setting

We study the regularity of the Lyapunov exponent of locally constant random linear cocycles, as a function of the measure $\mu$ driving the multiplicative process, under the quasi-irreducibility hypothesis (which represents the generic setting). As we will see, the regularity can be extremely different according to the topology that we choose. In this chapter we consider the weak-star topology, which is metrizable by the Wasserstein's metric.

The main result of this chapter is a general version of Le Page's theorem, that proves the Hölder continuity of the Lyapunov exponent in the generic setting. The proof we present in this chapter is based on the author's Master's thesis [27], which generalizes a result in [4, Theorem 1, item1].

One of the main elements in this proof is the so called strong mixing property of the Markov operator (see definition 2.3) associated to a measure $\mu$ on the group of matrices. By strong mixing we mean the uniform convergence of the powers of the Markov operator to its (a posteriori) unique stationary measure. This property plays a central role in the proof because one can reduce the study of the regularity of the Lyapunov exponent to the regularity of the stationary measure, via Furstenberg's formula. By the strong mixing property, the latter can then be deduced from the fact that the Markov operator and its powers depend nicely on the measure $\mu$.

The strong mixing property also implies, by general principles, statistical properties (such as large deviations and a central limit theorem) for the Lyapunov exponent. These results are not new in this specific setting, but we include them for completeness.

In section 4.1 we introduce the Wasserstein's metric, some technical facts about the convolution of measures and the relation between them. In section 4.2 we prove the strong mixing of the Markov operator, which is then used to prove the Hölder continuity of the Lyapunov exponent in section 4.3 and the statistical properties in section 4.4.

## 4.1 Convolution of measures

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.

**Definition 4.1** *Let* $(X, d)$ *be a compact metric space and let* $\mathrm{Prob}(X)$ *denote the set of Borel probability measures on* $X$*. Given two measures* $\mu, \nu \in \mathrm{Prob}(X)$*, a coupling between* $\mu$ *and* $\nu$ *is a measure* $\pi \in \mathrm{Prob}(X \times X)$ *with marginals* $\mu$ *and* $\nu$*.*

*More precisely, the push forward of* $\pi$ *via the projections* $\mathrm{proj}_1, \mathrm{proj}_2$ *in the first and second coordinates are* $\mu$ *and* $\nu$*, that is,* $(\mathrm{proj}_1)_*\pi = \mu$ *and* $(\mathrm{proj}_2)_*\pi = \nu$*.*

Given two measures $\mu, \nu \in \mathrm{Prob}(X)$, let $\Pi(\mu, \nu)$ denote the set of all possible couplings between $\mu$ and $\nu$. Note that the product measure $\mu \times \nu \in \Pi(\mu, \nu)$, hence $\Pi(\mu, \nu)$ is not empty.

**Definition 4.2** *Given* $\mu, \nu \in \mathrm{Prob}(X)$*,*

$$W_1(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times X} d(x, y)\, d\pi(x, y)$$

*is called the Wasserstein distance (of order* 1*).*

Kantorovich-Rubinstein's duality theorem gives a characterization of $W_1$ in terms of Lipschitz functions. It says that for all $\mu, \nu \in \mathrm{Prob}(X)$,

$$W_1(\mu, \nu) = \sup_{\varphi \in \mathrm{Lip}_1(X)} \left| \int_X \varphi\, d\mu - \int_X \varphi\, d\nu \right|,$$

where $\mathrm{Lip}_1(X)$ denotes the space of Lipschitz continuous functions with Lipschitz constant less than or equal to 1.

**Remark 4.1** *It turns out that there exists* $\pi^* \in \Pi(\mu, \nu)$ *such that the infimum in definition 4.2 is attained. Moreover, there also exists* $\varphi^* \in \mathrm{Lip}_1(X)$ *such that the supremum Kantorovich-Rubinstein's duality theorem is attained.*

**Definition 4.3** *Let* $G$ *be a group that acts on a set* $M$*. Let* $\mu$ *be a measure in* $G$ *and let* $\nu$ *a be measure in* $M$*. Then we define the convolution of* $\mu$ *and* $\nu$ *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* $E \subset M$*.*

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

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

to be the k-th convolution power of $\mu$. We also set $\mu^{*1} := \mu$ or think of $\mu^{*1}$ as the convolution of $\mu$ with a Dirac measure centered at the identity element of the group G.

Let $G$ be a multiplicative subgroup of $\mathrm{Mat}_d(\mathbb{R})$. Given $\Sigma \subset G$ a compact set, $\mu \in \mathrm{Prob}(\Sigma)$ and $n \geq 1$, the next proposition shows that the convolution power map $\mu \mapsto \mu^{*n}$ is Lipschitz (with Lipschitz constant depending on $n$ and $\Sigma$) with respect to the Wasserstein metric.

**Proposition 4.1** *Let $\Sigma \subset G$ be a compact set and let $n \in \mathbb{N}$. Then the map $\mathrm{Prob}(\Sigma) \ni \mu \mapsto \mu^{*n}$ is Lipschitz with respect to the Wasserstein metric.*

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

**Lemma 4.1** *Given $n \in \mathbb{N}$, 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 $\varphi$ was chosen arbitrarily. By induction, we conclude the lemma. ■

**Lemma 4.2** *Let $\mu \in \mathrm{Prob}(\Sigma)$ and let $\{\varphi_n\}_{n\in\mathbb{N}}$ be the group action of $G$ on itself, $\varphi_n\colon G \times G \times \cdots \times G \to G$, $\varphi_n(g_1, g_2, \ldots, g_n) = g_1 g_2 \cdots g_n$. Then $\mu^{*n} = (\varphi_n)_*(\mu \times \mu \times \cdots \times \mu)$ and $\varphi_n$ is Lipschitz.*

*Proof.* We only prove the case $n = 2$, the general case following by induction. For simplicity, we denote $\varphi_2$ by $\varphi$. Given a measurable set $E \subset G$, by the definition of the convolution of measures,

$$\begin{aligned}\mu * \mu(E) &= \int_{G\times G} 1_E(g_1 g_2)\, d\mu(g_1)d\mu(g_2)\\ &= \int_{G\times G} 1_E(\varphi(g_1, g_2))\, d\mu(g_1)d\mu(g_2)\\ &= \int_G 1_E(g)\, d\varphi_*(\mu\times\mu)(g)\\ &= \varphi_*(\mu\times\mu)(E).\end{aligned}$$

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

It remains to show that $\varphi$ is Lipschitz. We consider the distance $\overline{d}((g_1, g_2), (h_1, h_2)) := d(g_1, h_1) + d(g_2, h_2)$ on $G\times G$ . Hence,

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

Since $\mu$ has support in the compact set $\Sigma$, there exists a uniform constant $C > 0$ (depending on $\Sigma$) such that $\|g\| \le C$ for all $g \in \operatorname{supp}(\mu)$. Therefore,

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

This proves that $d\left(\varphi(g_1g_2), \varphi(h_1h_2)\right) \le C\overline{d}\left((g_1, g_2), (h_1, h_2)\right)$, so $\varphi$ is Lipschitz continuous, and its Lipschitz constant depends only on the compact support $\Sigma$. ■

**Lemma 4.3** *If $\varphi\colon X \to Y$ is Lipschitz with Lipschitz constant C, then the map $\mu \mapsto \varphi_*\mu$ is Lipschitz with the same Lipschitz constant.*

*Proof.* By remark 4.1, there exists $f \in \mathrm{Lip}_1(Y)$ such that

$$W_1(\varphi_*\mu, \varphi_*\nu) = \int_Y f\, d(\varphi_*\mu - \varphi_*\nu) = \int_X f\circ\varphi\, d(\mu-\nu).$$

Since $\varphi$ has Lipschitz constant C, then $\frac{\varphi}{C} \in \mathrm{Lip}_1(X)$. Also the composition $\frac{1}{C} f\circ\varphi \in \mathrm{Lip}_1(X)$. Therefore $\int_X f\circ\varphi\, d(\mu-\nu) \le CW_1(\mu,\nu)$ and we conclude that

$$W_1(\varphi_*\mu, \varphi_*\nu) \le CW_1(\mu,\nu),$$

which proves the lemma. ■

Finally, by lemma 4.2, $\mu^{*n} = (\varphi_n)*(\mu\times\mu\times\cdots\times\mu)$ where $\varphi_n$ is Lipschitz. By lemmas 4.1 and 4.3, the map $\mu \mapsto (\varphi_n)_*(\mu\times\mu\times\cdots\times\mu)$ is also Lipschitz, which concludes the proof of the proposition. ■

## 4.2 Mixing of the Markov operator

Let $\Sigma \subset \mathrm{GL_d}(\mathbb{R})$ be a compact set and let $\mu$ be a probability measure on $\Sigma$, that is, $\mu \in \mathrm{Prob}(\Sigma)$. We can define the Markov operator $Q_\mu \colon L^\infty(\mathbb{P}(\mathbb{R}^d)) \to L^\infty(\mathbb{P}(\mathbb{R}^d))$ associated to $\mu$ as follows:

$$Q_\mu(\varphi)(\hat{v}) = \int_\Sigma \varphi(\hat{g}\hat{v})\, d\mu(g), \tag{4.1}$$

where $\hat{g} \colon \mathbb{P}(\mathbb{R}^d) \to \mathbb{P}(\mathbb{R}^d)$ is the projective action of $g$ defined by $\widehat{gv} = \hat{g}\hat{v}$.

The goal of this section is to prove that if $L_1(\mu) > L_2(\mu)$ and if $\mu$ satisfies the quasi-irreducibility condition then the associated Markov operator $Q_\mu$ is strongly mixing. Note that the Markov operator defined above is the one from the third level (already defined in section 2.2.1) and by proposition 2.4, if it is indeed strongly mixing, then the related Markov operator from the second level also satisfies this strong mixing property. A general reference for this and related concepts in this section is [22].

### 4.2.1 Uniform convergence to the top Lyapunov exponent

We start with an important consequence of Furstenberg-Kifer's multiplicative ergodic theorem. Together with the hypothesis above, the finite time averages of the special observable $\log\|gp\|$ converge uniformly in $\hat{p} \in \mathbb{S}^{d-1}$ to the top Lyapunov exponent. This is a key property needed in order to establish the strong mixing property.

**Proposition 4.2** *Let $\mu \in \mathrm{Prob}(\Sigma)$ be a quasi irreducible measure with $L_1(\mu) > L_2(\mu)$. Then*

$$\lim_{n\to\infty} \frac{1}{n} \int_\Sigma \log\|gp\|\, d\mu^{*n}(g) = L_1(\mu),$$

*with uniform convergence in $p \in \mathbb{S}^{d-1} = \{v \in \mathbb{R}^d \colon \|v\| = 1\}$.*

*Proof.* Since $\mu$ is quasi irreducible, by remark 3.2, the Furstenberg-Kifer's non random filtration is trivial. Therefore, for every $p \in \mathbb{R}^d \backslash \{0\}$, we have that almost everywhere:

$$\lim_{n\to\infty} \frac{1}{n} \log\|g_{n-1}\cdots g_0 p\| = L_1(\mu).$$

By the definition of the $n$-th convolution power and dominated convergence we get that for every $p \in \mathbb{S}^{d-1}$,

$$\lim_{n\to\infty} \frac{1}{n} \int_{\Sigma} \log \|gp\| \, d\mu^{*n}(g) = \lim_{n\to\infty} \frac{1}{n} \int_{\Sigma^{\mathbb{N}}} \log \|g_{n-1} \cdots g_0 p\| \, d\mu^n(g_{n-1}, \ldots, g_0)$$
$$= L_1(\mu).$$

Hence we establish the pointwise convergence in $p$. Assume by contradiction that the convergence above is not uniform in $p \in \mathbb{S}^{d-1}$. Then there exists a sequence of unitary vectors $\{p_n\}_n \in \mathbb{R}^d$ and $\delta > 0$ such that for all large $n$,

$$\frac{1}{n} \int_{\Sigma} \log \|gp_n\| \, d\mu^{*n}(g) \leq L_1(\mu) - \delta.$$

By the compactness of the unitary circle, there exists a subsequence $\{p_{nk}\}_k$ that converges to a unit vector $p \in \mathbb{R}^d$. We claim that

$$\lim_{k\to\infty} \frac{1}{n_k} \int_{\Sigma} \log \|gp_{nk}\| \, d\mu^{*n_k}(g) = L_1(\mu),$$

which would contradict the previous assumption.

Note that by proposition 2.22 of [21],

$$\frac{\|g_{n_k-1} \cdots g_{n_0} \, p_{nk}\|}{\|g_{n_k-1} \cdots g_{n_0}\|} \geq \cos \angle \left(p_{n_k} \, u(g_{n_k-1} \cdots g_{n_0})\right) = |p_{n_k} \cdot u(g_{n_k-1} \cdots g_{n_0})|,$$

where $u(g_{n_k-1} \cdots g_{n_0})$ is the most expanded unit vector of the matrix $g_{n_k-1} \cdots g_{n_0}$. Moreover, by proposition 4.4 of [21], the limit

$$u(\mu) := \lim_{k\to\infty} u(g_{n_k-1} \cdots g_{n_0})$$

exists $\mu$-almost everywhere, so $|p_{n_k} \cdot u(g_{n_k-1} \cdots g_{n_0})| \to |p \cdot u(\mu)|$.

We claim that for $\mu$-almost every sequence $\underline{g}$, it holds that $|p \cdot u(\mu)| > 0$. Note that $u(\mu)$ is the most expanded direction of the *adjoint* cocycle (see Proposition 2.4.2 from [27]), which is, for almost every sequence $\underline{g}$, orthogonal to all the less expanding Oseledets directions $E^2(\underline{g}), \ldots, E^d(\underline{g})$ (see the beginning of the proof of Theorem 4.4 from [21]).

Since $\mu$ is quasi-irreducible, we know that

$$\lim_n \frac{1}{n} \log\|g_{n-1} \ldots g_0 v\| = L_1(\mu)$$

for every $v$ and $\mu$-almost every $\underline{g}$. Hence, if $p \cdot u(\mu) = 0$, then $p$ belongs to an Oseledets direction different from the most expanded one, which happens with zero probability. Therefore, the claim holds.

Hence,

$$\liminf_{k\to\infty} \frac{\|g_{n_k-1}\cdots g_{n_0}p_{n_k}\|}{\|g_{n_k-1}\cdots g_{n_0}\|} \geq |p\cdot u(\mu)| > 0$$

$\mu^{\mathbb{N}}$-almost everywhere. Then

$$\liminf_{k\to\infty} \frac{1}{n_k}\log \frac{\|g_{n_k-1}\cdots g_{n_0}p_{n_k}\|}{\|g_{n_k-1}\cdots g_{n_0}\|} = 0$$

$\mu^{\mathbb{N}}$-almost everywhere. Finally, using the definition of the convolution power and the dominated convergence theorem,

$$\begin{aligned}\lim_{k\to\infty} \frac{1}{n_k}\int_\Sigma \log\|gp_{n_k}\|\, d\mu^{*n_k}(g) &= \lim_{k\to\infty} \frac{1}{n_k}\int_\Sigma \log\|g\|\, d\mu^{*n_k}(g)\\ &+ \lim_{k\to\infty} \frac{1}{n_k}\int_\Sigma \log\frac{\|gp_{n_k}\|}{\|g\|}\, d\mu^{*n_k}(g)\\ &= L_1(\mu).\end{aligned}$$

This proves the claim and concludes the proof of the proposition. ■

### 4.2.2 The contracting property of the Hölder seminorm

We start by showing the relation between the iterates $Q_\mu^n$ of the Markov operator and the Markov operator $Q_{\mu^{*n}}$ of the measure $\mu^{*n}$.

**Lemma 4.4** *Let* $\mu \in \mathrm{Prob}(\Sigma)$. *Then*

$$Q_{\mu^{*n}} = Q_\mu^n.$$

*Proof.* The proof proceeds by induction. Let $\varphi \in L^\infty(\mathbb{P}(\mathbb{R}^d))$ and $\hat{v} \in \mathbb{P}(\mathbb{R}^d)$. The statement is trivial when $n = 1$. For $n = 2$ we have

$$\begin{aligned}(Q_\mu)^2(\varphi)(\hat{v}) &= \int_\Sigma\int_\Sigma \varphi(\hat{g}_1\hat{g}_0\hat{v})\, d\mu(g_1)d\mu(g_0)\\ &= \int_\Sigma\int_\Sigma \varphi(\widehat{g_1 g_0}\hat{v})\, d\mu(g_1)d\mu(g_0)\\ &= \int_\Sigma \varphi(\hat{g}\hat{v})\, d\mu^{*2}(g) = Q_{\mu^{*2}}(\varphi)(\hat{v}).\end{aligned}$$

Now suppose that it is true for every $k \leq n-1$. We are going to prove that is also true when $k = n$.

$$\begin{aligned}(Q_\mu)^n(\varphi)(\hat{v}) &= \int_\Sigma \cdots \int_\Sigma \varphi(\hat{g}_{n-1}\cdots\hat{g}_0\hat{v})\, d\mu(g_{n-1})\cdots d\mu(g_0)\\ &= \int_\Sigma\int_\Sigma \varphi(\hat{g}(\hat{g}_0\hat{v}))\, d\mu^{*(n-1)}(g)d\mu(g_0)\\ &= \int_\Sigma \varphi(\hat{g}\hat{v})\, d\mu^{*n}(g) = Q_{\mu^{*n}}(\varphi)(\hat{v}).\end{aligned}$$

■

Given $\hat{p},\hat{q}\in\mathbb{P}(\mathbb{R}^d)$, denote by $\delta\colon \mathbb{P}(\mathbb{R}^d)\times\mathbb{P}(\mathbb{R}^d)\to[0,\infty)$ the projective distance on $\mathbb{P}(\mathbb{R}^d)$:

$$\delta(\hat{p},\hat{q}) := \frac{\|p\wedge q\|}{\|p\|\,\|q\|}\,. \tag{4.2}$$

**Definition 4.4** *Let $\mu\in\mathrm{Prob}(\Sigma)$ and let $\alpha\in(0,1]$. We define the average $\alpha$-Hölder constant of the projective action $\hat{g}\colon\mathbb{PR}^d\to\mathbb{PR}^d$ by*

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

**Lemma 4.5** *$k_\alpha(\mu^{*n})$ is sub-multiplicative:*

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

*Proof.* By a straightforward computation:

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

■

Given $\varphi\in L^\infty(\mathbb{P}(\mathbb{R}^d))$ and $0\le\alpha\le1$ define the Hölder semi-norm

$$v_\alpha(\varphi) := \sup_{\hat{v}_1\neq\hat{v}_2}\frac{|\varphi(\hat{v}_1)-\varphi(\hat{v}_2)|}{\delta(\hat{v}_1,\hat{v}_2)^\alpha}.$$

If $v_\alpha(\varphi)<\infty$ then $\varphi$ is $\alpha$-Hölder continuous. Let $C^\alpha(\mathbb{P}(\mathbb{R}^d))$ be the space of all Hölder continuous functions, which we endow with its natural norm $\|\cdot\|_\alpha = \|\cdot\|_\infty + v_\alpha(\cdot)$.

**Lemma 4.6** *For all $\phi \in C^\alpha(\mathbb{PR}^d)$,*

$$v_\alpha\left(Q_\mu \phi\right) \leq k_\alpha(\mu)\, v_\alpha(\phi).$$

*Proof.* Given $\phi \in C^\alpha(\mathbb{PR}^d)$ and $\hat{p}, \hat{q} \in \mathbb{P}(\mathbb{R}^d)$,

$$\begin{aligned}\frac{|(Q_\mu\phi)(\hat{p}) - (Q_\mu\phi)(\hat{q})|}{\delta(\hat{p},\hat{q})^\alpha} &\leq \int_\Sigma \left| \frac{\phi(\hat{g}\hat{p}) - \phi(\hat{g}\hat{q})}{\delta(\hat{p},\hat{q})^\alpha} \right| d\mu(g) \\ &\leq v_\alpha(\phi) \int_\Sigma \left| \frac{\delta(\hat{g}\hat{p}, \hat{g}\hat{q})^\alpha}{\delta(\hat{p},\hat{q})^\alpha} \right| d\mu(g).\end{aligned}$$

Take the supremum in $\hat{p} \neq \hat{q}$ on both sides and conclude the proof. ■

**Lemma 4.7** *Given a measure $\mu \in \mathrm{Prob}(\mathrm{GL_d}(\mathbb{R}))$ with $\mathrm{supp}(\mu) \subseteq \Sigma$, a compact set, then for every $\alpha > 0$,*

$$k_\alpha(\mu) \leq \sup_{\hat{p} \in \mathbb{PR}^d} \int_\Sigma \left( \frac{s_1(g) s_2(g)}{\|gp\|^2} \right)^\alpha d\mu(g),$$

*where $s_1(g)$ and $s_2(g)$ are the first and second singular values of $g$ and $p$ is a unit representative of $\hat{p} \in \mathbb{PR}^d$.*

*Proof.* By the properties of the exterior algebra,

$$\|gp \wedge gq\| = \|(\wedge_2 g)(p \wedge q)\| = s_1(g)s_2(g)\|p \wedge q\|.$$

Hence, by the definition of the projective distance and the fact that the geometric mean is less than or equal to the arithmetic mean,

$$\begin{aligned}k_\alpha(\mu) &= \sup_{\hat{p} \neq \hat{q}} \int_\Sigma \left( \frac{\|gp \wedge gq\|}{\|gp\|\|gq\|} \frac{\|p\|\|q\|}{\|p \wedge q\|} \right)^\alpha d\mu(g) \\ &= \sup_{\hat{p} \neq \hat{q}} \int_\Sigma \left( \frac{s_1(g)s_2(g)}{\|gp\|\|gq\|} \right)^\alpha d\mu(g) \\ &\leq \sup_{\hat{p} \neq \hat{q}} \int_\Sigma \left( \frac{s_1(g)s_2(g)}{2} \right)^\alpha \left\{ \frac{1}{\|gp\|^{2\alpha}} + \frac{1}{\|gq\|^{2\alpha}} \right\} d\mu(g).\end{aligned}$$

■

**Proposition 4.3** *Let $\mu \in \mathrm{Prob}(\Sigma)$ be a quasi-irreducible measure with $L_1(\mu) > L_2(\mu)$. There are numbers $0 < \alpha \leq 1$, $\theta > 1$, $C > 0$ and $\delta > 0$ such that for all $n \in \mathbb{N}$ and for all $\nu \in \mathrm{Prob}(\Sigma)$ satisfying $W_1(\mu,\nu) < \delta$ we have that*

$$k_\alpha(\nu^{*n}) \leq C\theta^{-n}. \tag{4.3}$$

*Proof.* Note that

$$\lim_{n\to\infty} \frac{1}{n} \int_{\Sigma} \log \|gp\|^{-2} \, d\mu^{*n}(g) = -2L_1(\mu) < 0 \quad \text{(by proposition 4.2)}$$

with uniform convergence in $p \in \mathbb{S}^{d-1}$.

Thus for every $\epsilon > 0$ and every $p \in \mathbb{S}^{d-1}$, there exists some $N \in \mathbb{N}$ (that does not depend on $p$) such that for every $n > N$ we have that

$$-2L_1(\mu) - \epsilon \leq \frac{1}{n} \int_{\Sigma} \log \|gp\|^{-2} \, d\mu^{*n}(g) \leq -2L_1(\mu) + \epsilon.$$

Hence, by choosing $\epsilon$ small enough, e.g $\epsilon < \frac{1}{4}(L_1(\mu) - L_2(\mu))$, and $n$ sufficiently large, we conclude that

$$\int_{\Sigma} \log \|gp\|^{-2} \, d\mu^{*n}(g) \leq n\left(-2L_1(\mu) + \epsilon\right). \tag{4.4}$$

Moreover, since for a cocycle $A$, the first Lyapunov exponent $L_1(\wedge_2 A)$ of its second exterior power $\wedge_2 A$ is equal to $L_1(A) + L_2(A)$, for large enough $n$ we get

$$\int_{\Sigma} \log |s_1(g)\, s_2(g)| \, d\mu^{*n}(g) = \int_{\Sigma} \log |\wedge_2 \, g| \, d\mu^{*n}(g) \leq n(L_1(\mu) + L_2(\mu) + \epsilon).$$

We combine these two estimates to conclude that, for $n$ sufficiently large,

$$\lim_{n\to\infty} \frac{1}{n} \int_{\Sigma} \log \left[ \frac{s_1(g)s_2(g)}{\|gp\|^2} \right] \, d\mu^{*n}(g) \leq -1,$$

since $L_1(\mu) > L_2(\mu)$.

By the classical inequality $e^x \leq 1 + x + \frac{x^2}{2} e^{|x|}$, we conclude that for every $p \in \mathbb{R}^2$,

$$\begin{aligned} k_\alpha(\mu^{*n}) &\leq \int_{\Sigma} \left[ \frac{s_1(g)s_2(g)}{\|gp\|^{2\alpha}} \right] \, d\mu^{*n}(g) \leq \\ &\leq \int_{\Sigma} \left[ 1 + \alpha \log \frac{s_1(g)s_2(g)}{\|gp\|^2} + \frac{\alpha^2}{2} \log^2 \left( \frac{s_1(g)s_2(g)}{\|gp\|^2} \right) e^{\left|\alpha \log \frac{s_1(g)s_2(g)}{\|gp\|^2}\right|} \right] \, d\mu^{*n}(g) \\ &\leq 1 - \alpha + \frac{\alpha^2}{2} C(\mu, n). \end{aligned}$$

Note that $C$ is a constant that depends only on $\mu$ and $n$. Thus, by taking $\alpha$ sufficiently small we conclude that

$$k_\alpha(\mu^{*n}) \leq \sup_{\hat{p} \in \mathbb{PR}^d} \int_{\Sigma} \left[ \frac{s_1(g)s_2(g)}{\|gp\|^{2\alpha}} \right] \, d\mu^{*n}(g) < 1.$$

Finally, note that $k_\alpha(\mu^{*n})$ depends continuously on $\mu^{*n}$ and that the map

$\mu \mapsto \mu^{*n}$ is Lipschitz by proposition 4.1. Therefore, there exists $\delta > 0$ such that for every $\nu$ satisfying $W_1(\nu, \mu) < \delta$ we have $k_\alpha(\nu^{*n}) < 1$. By the sub-multiplicative property of $k_\alpha$ we conclude that there exists $C > 0$ and $\theta > 1$ such that the inequality 4.3 holds for every $n \in \mathbb{N}$. ■

**Corollary 4.1** *Let $\mu \in \text{Prob}(\Sigma)$ be quasi-irreducible and $L_1(\mu) > L_2(\mu)$. Then there exists $0 < \alpha \leq 1$, $\theta > 1$, $C > 0$ and a neighborhood $V \subset \text{Prob}(\Sigma)$ of $\mu$ with respect to the Wasserstein distance, such that for every $n \in \mathbb{N}$, every $\nu \in V$ and every $\varphi \in C^\alpha(\mathbb{P}(\mathbb{R}^d))$ we have*

$$v_\alpha(Q_\nu^n \varphi) \leq C\,\theta^{-n}\,v_\alpha(\varphi).$$

*Proof.* By the previous proposition, there exists $0 < \alpha \leq 1$, $\theta > 1$, $C > 0$ and a neighborhood $V \subset \text{Prob}(\Sigma)$ of $\mu$ with respect to the Wasserstein distance, such that for every $n \in \mathbb{N}$ and every $\nu \in V$, $k_\alpha(\nu^{*n}) \leq C\theta^{-n}$. Together with proposition 4.6, we conclude that for every $\varphi \in C^\alpha(\mathbb{P}(\mathbb{R}^d))$,

$$v_\alpha(Q_\nu^n \varphi) \leq k_\alpha(\nu^{*n})\,v_\alpha(\varphi) \leq C\theta^{-n}\,v_\alpha(\varphi).$$

■

### 4.2.3 The strong mixing property

A consequence of corollary 4.1 is that for every $\nu$ in the neighborhood $V \subset \text{Prob}(\Sigma)$ of $\mu$, the associated Markov operator $Q_\nu$ satisfies the strong mixing property. More precisely we get the following.

**Proposition 4.4** *Let $\mu \in \text{Prob}(\Sigma)$ be quasi-irreducible with $L_1(\mu) > L_2(\mu)$. There exist $\alpha \in (0, 1]$, $\theta \in (0, 1)$ and $C < \infty$ such that for every $\nu \in V \subset \text{Prob}(\Sigma)$, every $n \in \mathbb{N}$ and every $\varphi \in C^\alpha(\mathbb{P}(\mathbb{R}^d))$,*

$$\|Q_\nu^n \varphi - \int_{\mathbb{P}(\mathbb{R}^d)} \varphi\, d\eta_\nu\|_\alpha \leq C\theta^n \|\varphi\|_\alpha, \tag{4.5}$$

*where $\eta_\nu$ is a $Q_\nu$-stationary measure on $\mathbb{P}(\mathbb{R}^d)$.*

*Proof.* Note that

$$\begin{aligned}\|\varphi - \int \varphi\, d\eta_\nu\|_\infty &= \left|\varphi(\hat{v}) - \int \varphi(\hat{p})\, d\eta_\nu(\hat{p})\right| \\ &\leq \int |\varphi(\hat{v}) - \varphi(\hat{p})|\, d\eta_\nu(\hat{p}) \leq v_0(\varphi) \leq v_\alpha(\varphi).\end{aligned}$$

Therefore, since $\eta_\nu$ is $Q_\nu$ stationary,

$$\|Q^n\varphi - \int \varphi \, d\eta_\nu\|_\infty = \|Q^n\varphi - \int Q_\nu^n\varphi \, d\eta_\nu\|_\infty \leq v_\alpha(Q_\nu^n\varphi) \leq C\theta^{-n} \|\varphi\|_\alpha.$$

Moroever,

$$v_\alpha\left(Q^n\varphi - \int \varphi \, d\eta_\nu\right) = v_\alpha\left(Q^n\varphi\right) \leq C\theta^{-n} \|\varphi\|_\alpha,$$

which completes the argument. ■

We claim that $\eta_\nu$ is the unique $Q_\nu$-stationary measure of the cocycle. Indeed, if there exists another $Q_\nu$-stationary measure $\eta'_\nu$, then by integrating the inequality (4.5) with respect to $\eta'_\nu$, we conclude that

$$\int_{\mathbb{P}(\mathbb{R}^d)} \varphi \, d\eta'_\nu = \int_{\mathbb{P}(\mathbb{R}^d)} Q_\nu^n\varphi \, d\eta'_\nu \to \int_{\mathbb{P}(\mathbb{R}^d)} \varphi \, d\eta_\nu,$$

where the last convergence holds because $\eta'_\nu$ is a probability measure. Therefore it must hold that $\eta'_\nu = \eta_\nu$.

For each measure $\nu \in V$, consider the observable $\Phi_\nu \colon \mathbb{P}(\mathbb{R}^d) \to \mathbb{R}$ given by

$$\Phi_\nu(\hat{v}) = \int_\Sigma \log \frac{\|gv\|}{\|v\|} d\nu(g), \tag{4.6}$$

which belongs to $C^\alpha(\mathbb{P}(\mathbb{R}^d))$.

Since every $\nu \in V \subset \mathrm{Prob}(\Sigma)$ admits a unique stationary measure, the Furstenberg's formula implies that

$$L_1(\nu) = \int_{\mathbb{P}(\mathbb{R}^d)} \Phi_\nu(\hat{v}) \, d\eta_\nu \, .$$

**Remark 4.2** *Note that for every probability measure $\nu$ it is possible to associate a linear functional $\alpha_\nu$ as we did in equation* (3.1) *for the measure $\mu$. The uniqueness of the stationary measure $\eta_\nu$ implies that $\alpha_\nu$ is constant, hence $\nu$ admits a trivial Furstenberg-Kifer's filtration. This means that every $\nu \in V \subset \mathrm{Prob}(\Sigma)$ is also quasi-irreducible.*

## 4.3
## Hölder continuity of the Lyapunov exponent

**Proposition 4.5** *Let $\mu \in \mathrm{Prob}(\Sigma)$. Assume that $\mu$ is quasi-irreducible and $L_1(\mu) > L_2(\mu)$. Then there exist $\delta > 0$, $C < \infty$ and $0 < \alpha \leq 1$ such that for all $\mu_1, \mu_2 \in \mathrm{Prob}(\Sigma)$ satisfying $W_1(\mu_i, \mu) < \delta$, for all $n \in \mathbb{N}$ and for all $\varphi \in C^\alpha(\mathbb{P}\mathbb{R}^d)$, we have*

$$\|Q_{\mu_1}^n(\varphi) - Q_{\mu_2}^n(\varphi)\|_\infty \leq C W_1(\mu_1, \mu_2)^\alpha. \tag{4.7}$$

*Proof.* For $n = 1$ we have that

$$\begin{aligned}
\|Q_{\mu_1}(\varphi) - Q_{\mu_2}(\varphi)\|_\infty &= \sup_{\hat{p}\in\mathbb{PR}^d} \left| \int_\Sigma \varphi(\hat{g}_1(\hat{p}))d\mu_1(g_1) - \int_\Sigma \varphi(\hat{g}_2(\hat{p}))\, d\mu_2(g_2) \right| \\
&= \sup_{\hat{p}\in\mathbb{PR}^d} \left| \int_{\Sigma\times\Sigma} \varphi(\hat{g}_1(\hat{p})) - \varphi(\hat{g}_2(\hat{p}))\, d\pi(g_1, g_2) \right| \qquad \forall \pi \in \Pi(\mu_1, \mu_2) \\
&\leq \sup_{\hat{p}\in\mathbb{PR}^d} \int_{\Sigma\times\Sigma} |\varphi(\hat{g}_1(\hat{p})) - \varphi(\hat{g}_2(\hat{p}))|\; d\pi(g_1, g_2) \qquad \forall \pi \in \Pi(\mu_1, \mu_2) \\
&\leq v_\alpha(\varphi) \sup_{\hat{p}\in\mathbb{PR}^d} \int_{\Sigma\times\Sigma} \delta\left(\hat{g}_1\hat{p}, \hat{g}_2\hat{p}\right)^\alpha\; d\pi(g_1, g_2) \qquad \forall \pi \in \Pi(\mu_1, \mu_2) \\
&\leq v_\alpha(\varphi) \sup_{\hat{p}\in\mathbb{PR}^d} \int_{\Sigma\times\Sigma} \|g_1 - g_2\|^\alpha \max\left\{ \frac{1}{\|g_1(p)\|}, \frac{1}{\|g_2(p)\|} \right\}^\alpha\; d\pi(g_1, g_2),
\end{aligned}$$

for every $\pi \in \Pi(\mu_1, \mu_2)$. The last inequality follows from lemma 2.9 of [21].

Since $\Sigma$ is compact, there exists $C_1 > 0$ such that for every $\hat{p} \in \mathbb{PR}^d$, $\max\left\{\frac{1}{\|g_1(p)\|}, \frac{1}{\|g_2(p)\|}\right\} \leq C_1$.

Then, for every $\pi \in \Pi(\mu_1, \mu_2)$,

$$\begin{aligned}
\|Q_{\mu_1}(\varphi) - Q_{\mu_2}(\varphi)\|_\infty &\leq C_1^\alpha v_\alpha(\varphi) \int_{\Sigma\times\Sigma} \|g_1 - g_2\|^\alpha\; d\pi(g_1, g_2) \\
&\leq C_1^\alpha v_\alpha(\varphi) \left( \int_{\Sigma\times\Sigma} \|g_1 - g_2\|\; d\pi(g_1, g_2) \right)^\alpha \\
&\leq C_1^\alpha v_\alpha(\varphi) W_1(\mu_1, \mu_2)^\alpha,
\end{aligned}$$

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 $\alpha \in (0, 1]$. Hence we conclude that inequality (4.7) holds for $n = 1$. For now, we keep explicitly the term $v_\alpha(\varphi)$ in the conclusion instead of saying it is a constant, because it will play an important role in the induction process.

Now observe that the difference $Q_{\mu_1}^n - Q_{\mu_2}^n$ can be written as a telescopic sum as follows:

$$Q_{\mu_1}^n - Q_{\mu_2}^n = \sum_{i=0}^{n-1} Q_{\mu_2}^i \circ (Q_{\mu_1} - Q_{\mu_2}) \circ Q_{\mu_1}^{n-i-1}.$$

We use the previous relation to prove the desired estimate:

$$
\begin{aligned}
\|Q^n_{\mu_1}(\varphi) - Q^n_{\mu_2}(\varphi)\|_\infty &= \| \sum_{i=0}^{n-1} Q^i_{\mu_2} \circ (Q_{\mu_1} - Q_{\mu_2}) \circ \left(Q^{n-i-1}_{\mu_1}(\varphi)\right) \|_\infty \\
&\leq \sum_{i=0}^{n-1} \|Q^i_{\mu_2} \circ (Q_{\mu_1} - Q_{\mu_2}) \circ \left(Q^{n-i-1}_{\mu_1}(\varphi)\right) \|_\infty \\
&\leq \sum_{i=0}^{n-1} \| (Q_{\mu_1} - Q_{\mu_2}) \circ \left(Q^{n-i-1}_{\mu_1}(\varphi)\right) \|_\infty \\
&\leq \sum_{i=0}^{n-1} C_1^\alpha v_\alpha \left(Q^{n-i-1}_{\mu_1}(\varphi)\right) W_1(\mu_1,\mu_2)^\alpha \\
&\leq C_1^\alpha W_1(\mu_1,\mu_2)^\alpha \sum_{i=0}^{n-1} v_\alpha \left(Q^{i}_{\mu_1}(\varphi)\right) \\
&\leq C_1^\alpha W_1(\mu_1,\mu_2)^\alpha \sum_{i=0}^{\infty} C_2 \theta^{-i} \\
&\leq C W_1(\mu_1,\mu_2)^\alpha.
\end{aligned}
$$

In the previous estimate, $C_2$ and $\theta > 1$ are the constants from corollary 4.1. ■

**Corollary 4.2** *Let* $\mu \in \text{Prob}(\Sigma)$. *Assume that* $\mu$ *is quasi-irreducible and* $L_1(\mu) > L_2(\mu)$. *There exist* $\delta > 0$, $C < \infty$ *and* $0 < \alpha \leq 1$ *such that for every* $\mu_1$, $\mu_2$ *satisfying* $W_1(\mu_i, \mu) < \delta$, *for all* $n \in \mathbb{N}$ *and* $\varphi \in C^\alpha(\mathbb{PR}^d)$ *we have that*

$$\left| \int_{\mathbb{PR}^d} \varphi \, d\eta_{\mu_1} - \int_{\mathbb{PR}^d} \varphi \, d\eta_{\mu_2} \right| \leq C W_1(\mu_1,\mu_2)^\alpha,$$

*where* $\eta_{\mu_1}$ *and* $\eta_{\mu_2}$ *are the unique stationary measures with respect to* $Q_{\mu_1}$ *and* $Q_{\mu_2}$, *respectively.*

*Proof.* By proposition 4.4,

$$\lim_{n\to\infty} Q^n_{\mu_1}(\varphi) = \int_{\mathbb{PR}^d} \varphi \, d\eta_{\mu_1} \quad \text{and} \quad \lim_{n\to\infty} Q^n_{\mu_2}(\varphi) = \int_{\mathbb{PR}^d} \varphi \, d\eta_{\mu_2}.$$

Hence, by proposition 4.5, we conclude that

$$\left| \int_{\mathbb{PR}^d} \varphi \, d\eta_{\mu_1} - \int_{\mathbb{PR}^d} \varphi \, d\eta_{\mu_2} \right| = \lim_{n\to\infty} \|Q^n_{\mu_1}(\varphi) - Q^n_{\mu_2}(\varphi)\|_\infty \leq C W_1(\mu_1,\mu_2)^\alpha.$$

■

**Theorem 4.1** *Let* $\mu \in \text{Prob}(\Sigma)$. *Assume that* $\mu$ *is quasi-irreducible and* $L_1(\mu) > L_2(\mu)$. *There exist* $\delta > 0$, $C > 0$ *and* $\alpha \in (0,1]$ *such that given any* $\mu_i$, $i \in \{1,2\}$, *satisfying* $W_1(\mu_i, \mu) < \delta$, *we have that*

$$|L_1(\mu_1) - L_1(\mu_2)| \leq C W_1(\mu_1,\mu_2)^\alpha.$$

*Proof.* By Furstenberg's formula,

$$L_1(\mu_i) = \int_{\mathbb{P}\mathbb{R}^d} \int_{\Sigma} \varphi \; d\mu_i d\eta_{\mu_i},$$

where $\varphi(g, \hat{p}) = \log \frac{\|gp\|}{\|p\|}$ and $\eta_{\mu_i}$ is the unique stationary measure with respect to $Q_{\mu_i}$, for $i = 1, 2$. Then

$$\begin{aligned}
&|L_1(\mu_1) - L_1(\mu_2)| = \left| \int_{\mathbb{P}\mathbb{R}^d} \int_{\Sigma} \varphi \; d\mu_1 d\eta_{\mu_1} - \int_{\mathbb{P}\mathbb{R}^d} \int_{\Sigma} \varphi \; d\mu_2 d\eta_{\mu_2} \right| \\
&\leq \left| \int_{\mathbb{P}\mathbb{R}^d} \int_{\Sigma} \varphi \; d\mu_1 d\eta_{\mu_1} - \int_{\mathbb{P}\mathbb{R}^d} \int_{\Sigma} \varphi \; d\mu_1 d\eta_{\mu_2} \right| + \\
&+ \left| \int_{\mathbb{P}\mathbb{R}^d} \int_{\Sigma} \varphi \; d\mu_1 d\eta_{\mu_2} - \int_{\mathbb{P}\mathbb{R}^d} \int_{\Sigma} \varphi \; d\mu_2 d\eta_{\mu_2} \right| \\
&= \left| \int_{\Sigma} \int_{\mathbb{P}\mathbb{R}^d} \varphi \; d\left(\eta_{\mu_1} - \eta_{\mu_2}\right) d\mu_1 \right| + \left| \int_{\mathbb{P}\mathbb{R}^d} \int_{\Sigma} \varphi \; d(\mu_1 - \mu_2) d\eta_{\mu_2} \right|.
\end{aligned}$$

By corollary 4.2, the first term is bounded by $CW_1(\mu_1, \mu_2)^\alpha$. Note that $\varphi$ is Lipschitz on $\Sigma$ with respect to the first coordinate with some Lipschitz constant $C'$. Then, using Kantorovich-Rubinstein's theorem, the second term is bounded by $C' \, W_1(\mu_1, \mu_2)$.

Hence, there exists a constant $C > 0$ and $0 < \alpha \leq 1$ such that

$$|L_1(\mu_1) - L_1(\mu_2)| \leq CW_1(\mu_1, \mu_2)^\alpha.$$

This proves that the maximal Lyapunov exponent is Hölder continuous in a neighborhood of $\mu$. ■

**Remark 4.3** *Recently, Barrientos and Malicet obtained a similar result in [5], using a moment condition instead of the compactness. Moreover, in private communication with the authors, we were made aware of the independent work of Duarte and Graxinha on a similar problem, but in the more general non compact and not necessarily invertible setting.*

## 4.4 Statistical Properties: LDT & CLT

In this section we explore statistical properties of the Lyapunov exponents. We will use the machinery from section 2.3.

In proposition 4.4 we proved that given $\mu \in \text{Prob}(\Sigma)$, which is quasi-irreducible and satisfies $L_1(\mu) > L_2(\mu)$, for every $\nu$ in a neighborhood of $\mu$, the associated Markov system $(\mathbb{P}(\mathbb{R}^d), K_\nu, \eta_\nu)$ is strongly mixing in $C^\alpha(\mathbb{P}(\mathbb{R}^d))$, where $K_\nu(\hat{v}) = \int_\Sigma \delta_{\hat{g}_0 \hat{v}} \; d\nu(g_0)$.

Moreover, by proposition 2.4, this implies that the Markov system on the second level $(\Sigma \times \mathbb{P}(\mathbb{R}^d), \bar{K}_\nu, \nu \times \eta_\nu)$ is strongly mixing in $C^\alpha(\Sigma \times \mathbb{P})$, where $\bar{K}_\nu(g_0, \hat{v}) = \mu \times \delta_{\hat{g}_0\hat{v}}$.

**Remark 4.4** *The Markov system $(\Sigma \times \mathbb{P}(\mathbb{R}^d), \bar{K}_\nu, \nu \times \eta_\nu)$ is also strongly mixing in a slightly larger space, which we define as follows. First consider the seminorm $v_\alpha^{\mathbb{P}}$ which is the Hölder seminorm in the second coordinate:*

$$v_\alpha^{\mathbb{P}}(\varphi) = \sup_{\substack{p \neq q \\ g \in \Sigma}} \frac{|\varphi(g,p) - \varphi(g,q)|}{\delta(p,q)^\alpha}.$$

*Define the norm $\|\varphi\|_\alpha^{\mathbb{P}} = \|\varphi\|_\infty + v_\alpha^{\mathbb{P}}(\varphi)$ and consider the space*

$$\mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^d)) = \{\varphi \in L^\infty(\Sigma \times \mathbb{P}(\mathbb{R}^d)) : v_\alpha^{\mathbb{P}}(\varphi) < \infty\}.$$

Moreover, the special observable $\varphi(g_0\hat{v}) = \log \frac{\|g_0 v\|}{\|v\|}$ belongs to both $C^\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^d))$ and $\mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^d))$. Therefore, theorem 2.2 is applicable and we conclude that the following theorem holds.

**Theorem 4.2** *If $\mu \in \mathrm{Prob}(\Sigma)$ is quasi-irreducible and $L_1(\mu) > L_2(\mu)$, then there exists $\delta > 0$ such that for every $\nu$ satisfying $W_1(\mu,\nu) < \delta$, for every $v \neq 0$ and for all $\varepsilon > 0$*

$$\nu^{\mathbb{N}} \left\{ \underline{g} \in \Sigma^{\mathbb{N}} : \left| \frac{1}{n} \log\|g_{n-1} \dots g_1 g_0 v\| - L_1(\nu) \right| > \varepsilon \right\} \leq 8e^{-c(\varepsilon)n}$$

*where $c(\varepsilon) > 0$ depends explicitly on the data.*

We now prove a central limit theorem for this random linear cocycle. We already showed that the Markov system $(\Sigma \times \mathbb{P}(\mathbb{R}^d), \bar{K}_\nu, \nu \times \eta_\nu)$ is strongly mixing in $C^\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^d))$ for some $\alpha \in (0,1)$. Thus, in order to apply proposition 2.3 and conclude that the central limit theorem holds, we just need to verify the following extra condition.

**Lemma 4.8** *For every open set $U \subset \Sigma \times \mathbb{P}(\mathbb{R}^d)$ with $\nu(U) > 0$, there exists $\Phi \in C^\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^d))$ such that $0 \leq \Phi \leq \mathbb{1}_U$ and $\int \Phi \, d\nu > 0$.*

*Proof.* Let $U \subset \Sigma \times \mathbb{P}(\mathbb{R}^d)$ be an open set such that $\nu \times \eta_\nu(U) > 0$. Then there exist $U_1 \subset \Sigma$ and $U_2 \subset \mathbb{P}(\mathbb{R}^d)$ open sets such that $U_1 \times U_2 \subset U$ and $\nu(U_1) > 0$, $\eta_\nu(U_2) > 0$. The statement then follows by applying a version of Uryshon's lemma. ■

We conclude that the following CLT holds.

**Theorem 4.3** *If $\mu \in \operatorname{Prob}(\Sigma)$ is quasi-irreducible and $L_1(\mu) > L_2(\mu)$, then there exists $\delta > 0$ such that for every $\nu$ satisfying $W_1(\mu, \nu) < \delta$ and for every $v \neq 0$, we have that*

$$\frac{\log\|g_{n-1} \dots g_1 g_0 v\| - nL_1(\nu)}{\sigma\sqrt{n}} \to \mathcal{N}(0, 1)$$

*for some constant $\sigma = \sigma(\nu) \in (0, \infty)$.*

# 5
# Analyticity of the Lyapunov Exponent

In the previous chapter we studied the regularity of the Lyapunov exponent with respect to the weak-star topology. Now we study its regularity with respect to the total variation norm. Although there are similar ideas involved, it turns out that the Lyapunov exponent is much more regular in this setting.

The main result of this chapter is the analyticity of the maximal Lyapunov exponent as a function of the transition probabilities, which extends the results and methods of Peres from a finite to an infinite (but compact) space of symbols. This chapter is based on a joint work with Amorim and Melo [1]. Our approach combines the strong mixing property of the associated Markov operator with the theory of holomorphic functions in Banach spaces.

By putting together ideas of Peres, Baraviera and Duarte and tools of complex analysis in Banach spaces, we establish the analyticity of the top Lyapunov exponent with respect to the total variation norm in two different settings. Precise definitions of analyticity and total variation will be given in section 5.1.1.

In the first setting, we assume a quasi-irreducibility hypothesis. In the second one, instead of quasi-irreducibility, we assume that the probability measure has full support (which is an analogue of the assumption that each matrix has a positive probability in the finite support case of Peres [39]).

**Remark 5.1** *Similar results hold for absolutely continuous measures and for random locally constant linear cocycles whose domain* $\Sigma$ *is an arbitrary compact set mapped to* $\mathrm{GL_d}(\mathbb{R})$ *by a measurable and bounded function (see Section 5.2.3).*

**Remark 5.2** *In section 5.2.3 we include an example where* $\Sigma$ *is not compact and the Lyapunov exponent is not even continuous.*

## 5.1 Holomorphic functions in Banach spaces and complex Markov operators

This section is divided into two parts, both of which serve the purpose of constructing a setting that permits a generalization of Peres' arguments. The first part recalls the concept of analyticity in infinite dimensional Banach spaces as well as a useful criteria thereof. The second part is devoted to the study of complex Markov operators. We generalize the tools introduced in chapter 4 to the broader scenario of complex measures and we show that the analogue properties still hold.

### 5.1.1 Holomorphic functions in Banach spaces

Throughout this section, $M$ and $N$ will denote Banach spaces over $\mathbb{C}$ and $U$ will denote an open subset of $M$.

A function $f\colon U \to N$ is said to be *holomorphic at a point* $a \in U$ if for all $n \in \mathbb{N}$ there is an $n$-linear symmetric continuous map $T_n\colon M \times \cdots \times M \to N$ ($T_0$ is identically equal to a vector) such that:

$$f(x) = \sum_{n=0}^{\infty} T_n(x-a)^n,$$

for every $x \in B(a,r) \subset U$ (for some $r > 0$), where $T_n y^n$ denotes $T_n(y, y, \ldots, y)$. If $f$ is holomorphic at every point of $U$ then $f$ is said to be *holomorphic* on $U$.

We introduce the following notation:

$$U(a,b) := \{z \in \mathbb{C}\colon a + zb \in U\}.$$

**Definition 5.1** *A map $f\colon U \to N$ is said to be* Gâteux holomorphic *(or* G-holomorphic*) if for every $a \in U$ and for every $b \in M$, the map*

$$z \mapsto f(a+zb)$$

*is holomorphic on $U(a,b) \subset \mathbb{C}$.*

It is clear that every holomorphic map is also $G$-holomorphic. However the converse in general is not true when $M$ is infinite dimensional. The following theorem, see [16, Chapter 14], provides a criterion for when the converse holds.

**Theorem 5.1** *Let $U$ be an open subset of a Banach space and let $f\colon U \to N$. The following are equivalent:*

*(i) $f$ is holomorphic on $U$.*

*(ii) $f$ is G-holomorphic and continous on $U$.*

The *variation of a complex measure* $\mu$ is defined as

$$|\mu| := \sup_{\pi} \sum_{A\in\pi} |\mu(A)|$$

where the supremum is taken over all partitions $\pi$ of a measurable set $E$ into a countable number of disjoint measurable sets.

Another characterization of the variation of a complex measure is the following:

$$|\mu|(E) = \sup\left\{\left|\int_E f(g)\; d\mu(g)\right| : f \in L^{\infty}(\mu) \text{ and } \|f\|_{\infty} \leq 1\right\}.$$

Note that if $f \in L^1(\mu)$, then:

$$\left|\int_{\Sigma} f(x)\; d\mu(x)\right| \leq \int_{\Sigma} |f(x)|\; d|\mu|(x).$$

Let $\Sigma$ be a compact metric space. The total variation of a complex measure is defined as $\|\mu\| :\, = |\mu|(\Sigma)$. If a measure satisfies $\|\mu\| < \infty$, then we say that $\mu$ is finite or that it is of bounded variation.

We will consider $\Sigma$ to be a compact (but possibly infinite) space of symbols. We denote by $\mathcal{M}(\Sigma)$ the set of complex valued measures over $\Sigma$ with bounded variation. The set $\mathcal{M}(\Sigma)$, endowed with the total variation norm, will play the role of the Banach space $M$.

In this work, we consider slightly more general definitions of holomorphy and G-holomorphy, in which the domain could also be a translation of a Banach subspace. The following construction shows how one can transfer the holomorphic structure from Banach spaces to affine subspaces via translation.

Let $V \subset M$ be a closed subspace, $v_0 \in M$ and consider a closed affine subspace $V_0 = V + v_0$ of $M$. Let $U_0 \subset V_0$ be an open set of $V_0$. We consider a function $f_0 : U_0 \to N$ to be holomorphic (G-holomorphic) at $x_0 \in U_0$ if there exists a function $f\colon U = U_0 - v_0 \to N$ which is holomorphic (G-holomorphic) at $x_0 - v_0$, such that $f(x) = f_0(x + v_0)$ for every $x \in U$. Moreover, if $f$ is holomorphic (G-holomorphic) at every point of its domain, then so is $f_0$.

It is then immediate that Theorem 5.1 also holds in this context.

### 5.1.2 Complex Markov operators

Recall from the last chapter that given $\mu \in \mathrm{Prob}(\Sigma)$, one can consider the associated Markov operator $Q_\mu \colon L^\infty(\mathbb{P}) \to L^\infty(\mathbb{P})$

$$Q_\mu(\varphi)(\hat{v}) = \int_\Sigma \varphi(\hat{g}\hat{v})\, d\mu(g).$$

In this chapter we consider small perturbations of $\mu$ given by complex measures $\nu \in \mathcal{M}(\Sigma)$ and their associated operators $Q_\nu$. Although $Q_\mu$ is a Markov operator, $Q_\nu$ may not be a Markov operator. This can happen because when $\nu$ is not a probability measure, it does not fix constants. Although the operator $Q_\nu$ associated to a complex measure may not be a Markov operator, we will show that it satisfies similar properties.

First, note that the relation $Q_\mu^n = Q_{\mu^{*n}}$ described in lemma 4.4 still holds when $\mu \in \mathcal{M}(\Sigma)$ is a complex measure with bounded variation. The proof is exactly the same.

Now we proceed to introduce an analogous version of the average Hölder constant of the projective action $k_\alpha$. Note that when $\mu$ is a complex measure, the image of $k_\alpha$ (from definition 4.4) may not be a real number, thus, instead of working with $\mu$, we consider its variation $|\mu|$, see below.

**Definition 5.2** *Let $\mu \in \mathcal{M}(\Sigma)$ be a complex measure of bounded variation and let $\alpha \in (0,1]$. We define the average $\alpha$-Hölder constant of the projective action $\hat{g}_0 \colon \mathbb{P}(\mathbb{R}^d) \to \mathbb{P}(\mathbb{R}^d)$ by*

$$k_\alpha(\mu) := \sup_{v_1 \neq v_2} \int_\Sigma \left( \frac{\delta(\hat{g}_0\hat{v}_1, \hat{g}_0\hat{v}_2)}{\delta(\hat{v}_1, \hat{v}_2)} \right)^\alpha d|\mu|(g_0).$$

Throughout this chapter, whenever we refer to $k_\alpha$ one should consider the one from the previous definition 5.2.

**Lemma 5.1** *For every two complex measures $\mu, \nu \in \mathcal{M}(\Sigma)$, it holds that $|\mu * \nu| \leq |\mu||\nu|$. In particular, for every $\varphi \in L^\infty(\mu^{*n})$, it holds that*

$$\left| \int \varphi \, d\mu^{*n} \right| \leq \int |\varphi| \, d|\mu^{*n}| \leq \int |\varphi| \, d|\mu|^{*n}.$$

*Proof.* Let $\varphi \in L^\infty(\mu * \nu)$. Then:

$$\begin{aligned}
\left|\int_\Sigma \varphi(x)\, d\mu * \nu(x)\right| &= \left|\int_\Sigma \int_\Sigma \varphi(gx)\, d\nu(x) d\mu(g)\right| \\
&\leq \int_\Sigma \int_\Sigma |\varphi(gx)|\, d|\nu|(x) d|\mu|(g) \\
&= \int_\Sigma |\varphi(x)|\, d|\mu| * |\nu|(x)
\end{aligned}$$

By restricting it to $\|\varphi\|_\infty \leq 1$ and taking the supremum on both sides, it follows that $|\mu * \nu| \leq |\mu||\nu|$. Moreover, applying the inequality above multiple times with $\nu = \mu$ concludes the result. ■

Now, we prove the analogues of lemmas 4.5 and 4.6 for the complex analogue of the Markov operator.

**Lemma 5.2** *The sequence $k_\alpha(\mu^{*n})$ is sub-multiplicative:*

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

*Proof.*

$$\begin{aligned}
k_\alpha(\mu^{(m+n)}) &= \sup_{v_1 \neq v_2} \int_\Sigma \left(\frac{\delta(\hat{g}\hat{v}_1, \hat{g}\hat{v}_2)}{\delta(\hat{v}_1, \hat{v}_2)}\right)^\alpha d|\mu^{*(m+n)}|(g) \\
&\leq \sup_{v_1 \neq v_2} \int_\Sigma \left(\frac{\delta(\hat{g}\hat{v}_1, \hat{g}\hat{v}_2)}{\delta(\hat{v}_1, \hat{v}_2)}\right)^\alpha d|\mu^{*n}| * |\mu^{*m}|(g) \\
&= \sup_{v_1 \neq v_2} \int_\Sigma \int_\Sigma \left(\frac{\delta(\hat{g}_2\hat{g}_1\hat{v}_1, \hat{g}_2\hat{g}_1\hat{v}_2)}{\delta(\hat{v}_1, \hat{v}_2)}\right)^\alpha d|\mu^{*m}|(g_1) d|\mu^{*n}|(g_2) \\
&= \sup_{v_1 \neq v_2} \int_\Sigma \int_\Sigma \left(\frac{\delta(\hat{g}_2\hat{g}_1\hat{v}_1, \hat{g}_2\hat{g}_1\hat{v}_2)}{\delta(\hat{g}_1\hat{v}_1, \hat{g}_1\hat{v}_2)} \cdot \frac{\delta(\hat{g}_1\hat{v}_1, \hat{g}_1\hat{v}_2)}{\delta(\hat{v}_1, \hat{v}_2)}\right)^\alpha d|\mu^{*m}|(g_1) d|\mu^{*n}|(g_2) \\
&\leq \sup_{v_1 \neq v_2} \int_\Sigma \left(\frac{\delta(\hat{g}_1\hat{v}_1, \hat{g}_1\hat{v}_2)}{\delta(\hat{v}_1, \hat{v}_2)}\right)^\alpha k_\alpha(\mu^{*n}) d|\mu^{*m}|(g_1) = k_\alpha(\mu^{*m}) k_\alpha(\mu^{*n}).
\end{aligned}$$

■

**Lemma 5.3** *For every $n \geq 1$, $\mu \in \mathcal{M}(\Sigma)$ and $\varphi \in C^\alpha(\mathbb{P}(\mathbb{R}^d))$, the following inequality holds:*

$$v_\alpha(Q_\mu^n(\varphi)) \leq k_\alpha(\mu^{*n})\, v_\alpha(\varphi).$$

*Proof.*

$$\begin{aligned}
\frac{|Q_\mu^n\varphi(\hat{v}_1)-Q_\mu^n\varphi(\hat{v}_2)|}{\delta(\hat{v}_1,\hat{v}_2)^\alpha} &= \frac{|\int_\Sigma \varphi(\hat{g}\hat{v}_1)\,d\mu^{*n}(g)-\int_\Sigma \varphi(\hat{g}\hat{v}_2)\,d\mu^{*n}(g)|}{\delta(\hat{v}_1,\hat{v}_2)^\alpha} \\
&= \left|\int_\Sigma \frac{\varphi(\hat{g}\hat{v}_1)-\varphi(\hat{g}\hat{v}_2)}{\delta(\hat{v}_1,\hat{v}_2)^\alpha}\,d\mu^{*n}(g)\right| \\
&\le \int_\Sigma \frac{|\varphi(\hat{g}\hat{v}_1)-\varphi(\hat{g}\hat{v}_2)|}{\delta(\hat{v}_1,\hat{v}_2)^\alpha}\,d|\mu^{*n}|(g) \\
&\le v_\alpha(\varphi)\int_\Sigma \frac{\delta(\hat{g}\hat{v}_1,\hat{g}\hat{v}_2)^\alpha}{\delta(\hat{v}_1,\hat{v}_2)^\alpha}\,d|\mu^{*n}|(g)
\end{aligned}$$

We conclude the lemma by applying the supremum in $\hat{v}_1 \neq \hat{v}_2$ to both sides. ■

## 5.2 The holomorphic extension of the Lyapunov exponent

This section is divided into three parts. The first one is devoted to proving that the ideas in chapter 4 still hold for complex measures. We show that the powers of the Markov operator $Q_\mu$ converge to a number which, when $\mu$ is a probability measure, is the Lyapunov exponent. In the second part we use the concept of Gâteux holomorphy to write the Markov operator as a polynomial. Therefore, using ideas of [39], we show that the Lyapunov exponent of a probability measure has a holomorphic extension, which in turn implies its analyticity, the main result of the chapter. In the last part of this section we present some consequences of this result and we include one example that shows the importance of the compactness of the support of the measure.

### 5.2.1 The convergence of the iterates of the Markov operator

Let $\mu_0$ be a probability measure on $\mathrm{GL_d}(\mathbb{R})$ with support in the compact set $\Sigma$. Assume that $\mu_0$ is quasi-irreducible. This implies that its Furstenberg-Kifer's filtration (see theorem 3.5) is trivial, that is, for every $v \in \mathbb{R}^d \backslash \{0\}$ and $\mu_0^{\mathbb{N}}$-almost every $\{g_n\}_n \in \Sigma^{\mathbb{N}}$,

$$\lim_{n\to\infty}\frac{1}{n}\log\|g_{n-1}\dots g_1 g_0 v\| = L_1(\mu_0).$$

Moreover, together with the hypothesis that $L_1(\mu_0) > L_2(\mu_0)$, a consequence of the previous fact is proposition 4.2, which says that

$$\lim_{n\to\infty}\frac{1}{n}\int_\Sigma \log\|gv\|d\mu_0^{*n}(g) = L_1(\mu_0), \tag{5.1}$$

with uniform convergence in $\hat{v} \in \mathbb{P}(\mathbb{R}^d)$.

Now, similarly to what we have done in chapter 4, we proceed to prove that the complex analogue of the Markov operator associated to any complex measure in a neighborhood of $\mu_0$ contracts the $v_\alpha$ seminorm.

**Lemma 5.4** *For every $\mu \in \mathcal{M}(\Sigma)$ and every $\alpha > 0$,*

$$k_\alpha(\mu) \leq \sup_{\hat{v} \in \mathbb{P}} \int_\Sigma \left( \frac{s_1(g)s_2(g)}{\|gv\|^2} \right)^\alpha d|\mu|(g),$$

*where $s_1(g)$ and $s_2(g)$ are the first and second singular values of a matrix $g \in \mathrm{GL_d}(\mathbb{R})$.*

*Proof.* Recall that

$$\|gp \wedge gq\| = s_1(g)s_2(g)\|p \wedge q\|.$$

Hence, by (4.2), given $\alpha > 0$, two points $\hat{p}, \hat{q} \in \mathbb{P}$ and any $g \in \Sigma$, it holds that

$$\begin{aligned} \left[ \frac{\delta(\hat{g}\hat{p}, \hat{g}\hat{q})}{\delta(\hat{p}, \hat{q})} \right]^\alpha &= \left[ (s_1(g)s_2(g)) \frac{\|p\|\|q\|}{\|gp\|\|gq\|} \right]^\alpha \\ &\leq \frac{[s_1(g)s_2(g)]^\alpha}{2} \left[ \frac{1}{\|gp\|^{2\alpha}} + \frac{1}{\|gq\|^{2\alpha}} \right] \end{aligned}$$

since the geometric mean is less or equal the arithmetic mean.

Note that if we integrate with respect to the measure $|\mu|$ and take the supremum in $\hat{p} \neq \hat{q}$ on both sides of this inequality, we conclude the lemma. ■

**Proposition 5.1** *Assume that $\mu_0 \in \mathrm{Prob}(\Sigma)$ is quasi-irreducible and $L_1(\mu_0) > L_2(\mu_0)$. Then there exist $0 < \alpha \leq 1$, $\theta > 1$, $C > 0$ and a neighborhood $V \subset \mathcal{M}(\Sigma)$ of $\mu_0$ with respect to the total variation distance, such that for every $n \in \mathbb{N}$ and for every $\mu \in V$,*

$$k_\alpha(\mu^{*n}) \leq C\theta^{-n}. \tag{5.2}$$

*Proof.* We start the proof for $\mu_0$ by following the argument in proposition 4.3. Since $\mu_0 \in \mathrm{Prob}(\Sigma)$, the same estimates from chapter 4 hold and we conclude that

$$k_\alpha(\mu_0^{*n}) \leq \int_\Sigma \left( \frac{s_1(g)s_2(g)}{\|gv\|^2} \right)^\alpha d|\mu_0|^{*n} = \int_\Sigma \left( \frac{s_1(g)s_2(g)}{\|gv\|^2} \right)^\alpha d\mu_0^{*n} \leq 1 - \alpha + C\frac{\alpha^2}{2}$$

for some finite constant $C$ that depends only on $g, \mu_0$ and $n$.

Thus, fixing $n_0$ sufficiently large and considering $\alpha$ small enough, we conclude by Lemma 5.4 that

$$k_\alpha(\mu_0^{*n_0}) \leq 1 - \alpha + C\frac{\alpha^2}{2} < 1.$$

Note that for a fixed $n$, the quantity $\int \left(\frac{s_1(g)s_2(g)}{\|gv\|^2}\right)^\alpha d|\mu_0|^{*n}$, which bounds from above $k_\alpha(\mu_0^{*n})$, depends continuously on the measure $\mu_0$. Then the previous inequality extends to a neighborhood of $\mu_0$. There exists $\kappa < 1$ and a neighborhood $V \subset \mathcal{M}(\Sigma)$ of $\mu_0$ with respect to the total variation distance, such that $k_\alpha(\mu^{*n_0}) \leq \kappa < 1$ for every $\mu \in V$.

Because of the sub-multiplicative property of $k_\alpha$, we conclude that there exists $C > 0$ and $\theta > 1$, such that inequality 5.2 holds for every $n \in \mathbb{N}$. ■

**Corollary 5.1** *Assume that $\mu_0 \in \mathrm{Prob}(\Sigma)$ is quasi-irreducible and $L_1(\mu_0) > L_2(\mu_0)$. Then there exist $0 < \alpha \leq 1$, $\theta > 1$, $C > 0$ and a neighborhood $V \subset \mathcal{M}(\Sigma)$ of $\mu_0$ with respect to the total variation distance, such that for every $n \in \mathbb{N}$, for every $\mu \in V$ and every $\varphi \in C^\alpha(\mathbb{P}(\mathbb{R}^d))$,*

$$v_\alpha(Q_\mu^n \varphi) \leq C\theta^{-n}\, v_\alpha(\varphi).$$

*Proof.* By proposition 5.1, there exist $0 < \alpha \leq 1$, $\theta > 1$, $C > 0$ and a neighbourhood $V \subset \mathcal{M}(\Sigma)$ of $\mu_0$ with respect to the total variation distance, such that for every $n \in \mathbb{N}$ and for every $\mu \in V$, we have that $k_\alpha(\mu^{*n}) \leq C\theta^{-n}$. Together with lemma 5.3, we conclude that

$$v_\alpha(Q_\mu^n(\varphi)) \leq k_\alpha(\mu^{*n})\, v_\alpha(\varphi) \leq C\theta^{-n}.$$

■

We already showed in proposition 4.4 that when $\mu$ is a probability measure on $\Sigma$, a consequence of the previous corollary is that $Q_\mu$ is strongly mixing. There exist $\alpha \in (0,1]$, $\theta \in (0,1)$ and $C < \infty$ such that for every $n \in \mathbb{N}$ and every $\varphi \in C^\alpha(\mathbb{P}(\mathbb{R}^d))$,

$$\|Q_\mu^n \varphi - \int \varphi\, d\eta_\mu\|_\alpha \leq C\theta^n \|\varphi\|_\alpha, \tag{5.3}$$

where $\eta_\mu$ is the unique $\mu$-stationary measure on $\mathbb{P}(\mathbb{R}^d)$.

Consider the observable $\varphi\colon \mathbb{P}(\mathbb{R}^d) \to \mathbb{R}$ given by

$$\varphi(\hat{v}) = \int_\Sigma \log \frac{\|gv\|}{\|v\|}\, d\mu(g). \tag{5.4}$$

If $\mu$ is a probability measure, then, by Furstenberg's formula,

$$\int_{\mathbb{P}(\mathbb{R}^d)} \varphi(\hat{v})\, d\eta_\mu = L_1(\mu).$$

**Remark 5.3** *When $\mu$ is a probability measure, for a fixed $\hat{v} \in \mathbb{P}(\mathbb{R}^d)$ the iterates $Q_\mu^n \varphi(\hat{v})$ converge uniformly to the top Lyapunov exponent $L_1(\mu)$.*

### 5.2.2 The domain of holomorphy

Now we establish a holomorphic extension of the Lyapunov exponent $L_1$. We start by defining the domain where $L_1$ will be shown to be analytic.

Let $\mathcal{M}_0(\Sigma)$ be the set of finite complex measures that give measure zero to $\Sigma$. Therefore, every $\mu \in \mathcal{M}_0(\Sigma)$ satisfies $\mu(\Sigma) = 0$.

**Lemma 5.5** *$\mathcal{M}_0(\Sigma)$ is a Banach space.*

*Proof.* First note that $\mathcal{M}_0(\Sigma)$ is a vector subspace of $\mathcal{M}(\Sigma)$. Moreover, it is the kernel of the linear functional that assigns to each finite measure, its measure of the whole space: $\nu \mapsto \nu(\Sigma)$. Therefore, it is a closed subspace of $\mathcal{M}(\Sigma)$, hence it is also a Banach space. ■

Let $\mathcal{M}_1(\Sigma)$ denote the set of finite complex measures that give measure one to $\Sigma$. Note that $\mathcal{M}_1(\Sigma)$ is an affine subspace of $\mathcal{M}(\Sigma)$, namely $\mathcal{M}_1(\Sigma) = \mathcal{M}_0(\Sigma) + \mu_0$, for some $\mu_0 \in \text{Prob}(\Sigma)$. Therefore $\mathcal{M}_1(\Sigma)$ can be endowed with an analytic structure as seen in section 2.1.

We are going to prove that $L_1$ admits a holomorphic extension to $V \cap \mathcal{M}_1(\Sigma)$, where $V$ is the neighborhood of $\mu_0$ from proposition 5.1. In fact, all the proofs from the previous sections were done in $\mathcal{M}(\Sigma)$, but could have been done directly in $\mathcal{M}_1(\Sigma)$. Thus, from now on, we are going to consider the neighborhood $V$ to be in $\mathcal{M}_1(\Sigma)$ and we will write just $V$ instead of $V \cap \mathcal{M}_1(\Sigma)$.

**Lemma 5.6** *For every $\mu \in V \subset \mathcal{M}_1(\Sigma)$, $\hat{v}_1 \neq \hat{v}_2 \in \mathbb{P}(\mathbb{R}^d)$ and $\varphi \in C^\alpha(\mathbb{P}(\mathbb{R}^d))$ it holds that*

$$\left| Q_\mu^n \varphi(\hat{v}_1) - Q_\mu^n \varphi(\hat{v}_2) \right| \leq C\theta^{-n}. \tag{5.5}$$

*Proof.* By lemma 5.3, for every $\mu \in \mathcal{M}(\Sigma)$ and every $\varphi \in C^\alpha(\mathbb{P}(\mathbb{R}^d))$,

$$v_\alpha(Q_\mu^n \varphi) \leq v_\alpha(\varphi) k_\alpha(\mu^{*n}).$$

Therefore, it also holds that for every $\hat{v}_1 \neq \hat{v}_2 \in \mathbb{P}(\mathbb{R}^d)$,

$$\left| Q_\mu^n \varphi(\hat{v}_1) - Q_\mu^n \varphi(\hat{v}_2) \right| \leq v_\alpha(\varphi) k_\alpha(\mu^{*n}).$$

Using proposition 5.1 we conclude the proof. ■

**Proposition 5.2** *For every $\mu \in V \subset \mathcal{M}_1(\Sigma)$ and $\hat{v}_1 \neq \hat{v}_2 \in \mathbb{P}(\mathbb{R}^d)$,*

$$\left| Q_\mu^{n+1} \varphi(\hat{v}) - Q_\mu^n \varphi(\hat{v}) \right| \leq C\theta^{-n}.$$

*Proof.* Note that for every $\hat{v}_1 \neq \hat{v}_2 \in \mathbb{P}(\mathbb{R}^d)$,

$$\left| Q_\mu^{n+1}\varphi(\hat{v}) - Q_\mu^n\varphi(\hat{v}) \right| = \left| \int_\Sigma Q_\mu^n \varphi(\hat{g}\hat{v})\, d\mu - Q_\mu^n\varphi(\hat{v}) \right|.$$

Moreover, for every $\mu \in V \subset \mathcal{M}_1(\Sigma)$ and $\hat{v_1} \neq \hat{v_2} \in \mathbb{P}(\mathbb{R}^d)$,

$$\left| \int_\Sigma Q_\mu^n \varphi(\hat{g}\hat{v})\, d\mu - Q_\mu^n\varphi(\hat{v}) \right| = \left| \int_\Sigma Q_\mu^n \varphi(\hat{g}\hat{v}) - Q_\mu^n\varphi(\hat{v})\, d\mu \right| \leq Ce^{-n}.$$

■

We are ready to state and prove the main result of this chapter.

**Theorem 5.2** *Let* $\Sigma \subset \mathrm{GL}_d(\mathbb{R})$ *be a compact subset,* $\mu_0 \in \mathrm{Prob}(\Sigma)$ *and assume that* $L_1(\mu_0) > L_2(\mu_0)$.

*(1) If* $\mu_0$ *is quasi-irreducible, then* $\mu \mapsto L_1(\mu)$ *is real analytic with respect to the total variation norm in a neighbourhood of* $\mu_0$.

*(2) If* $\mathrm{supp}(\mu_0) = \Sigma$*, then* $\mu \mapsto L_1(\mu)$ *is real analytic with respect to the total variation norm in a neighbourhood of* $\mu_0$.

*Proof.* For a fixed $\hat{v}$ and $\varphi$ given in equation 5.4, the sequence $\{Q_\mu^n\varphi(\hat{v})\}_n$ is Cauchy. Therefore its limit, denoted by $Q_\mu^\infty\varphi(\hat{v})$, exists. Moreover, note that $\mu \mapsto Q_\mu^n\varphi(\hat{v})$ is continuous. Since $\mu \mapsto Q_\mu^\infty\varphi(\hat{v})$ is a uniform limit of continuous functions, it is also continuous. Furthermore, when $\mu$ is a probability measure, $Q_\mu^\infty\varphi(\hat{v}) = L_1(\mu)$, the top Lyapunov exponent (as shown in remark 5.3).

We want to prove that $\mu \mapsto Q_\mu^\infty\varphi(\hat{v})$ is holomorphic. For this, we are going to use theorem 5.1. Since we already know that the limit is continuous, it suffices to prove that it is also G-holomorphic.

As we stated, $\mathcal{M}_1(\Sigma)$ is not a Banach space, therefore, we need to transfer the holormphic structure from $\mathcal{M}_0(\Sigma)$ to it. Intuitively, G-holomorphy means to be holomorphic along complex lines, hence to say that the map $\mu \mapsto Q_\mu^\infty\varphi(\hat{v})$ from $V \subset \mathcal{M}_1(\Sigma)$ to $\mathbb{C}$ is Gâteaux holomorphic means that $\forall \mu \in V$, $\forall \nu \in \mathcal{M}_0(\Sigma)$, the map $z \mapsto Q_{\mu+z\nu}^\infty\varphi(\hat{v})$ is holomorphic on $V(\mu,\nu) = \{z \in \mathbb{C} \colon \mu + z\nu \in V\}$.

Consider measures $\mu_z$ of the form $\mu_z = \mu + z\nu$, where $\mu \in V$ and $\nu$ is any finite complex measure with $\nu(\Sigma) = 0$. Note that, since $\mu \in \mathcal{M}_1(\Sigma)$, we have that $\mu_z \in \mathcal{M}_1(\Sigma)$ for every $z \in \mathbb{C}$. Consider small perturbations of the Markov operator in the following sense: for each $z \in \mathbb{C}$, let the operator

$Q_{\mu+z\nu}\colon L^\infty(\mathbb{P},\mathbb{C}) \to L^\infty(\mathbb{P},\mathbb{C})$ be defined by

$$\begin{aligned} Q_{\mu+z\nu}(\varphi)(\hat{v}) &= \int_\Sigma \varphi(\hat{g}\hat{v})\, d(\mu+z\nu)(g) \\ &= \int_\Sigma \varphi(\hat{g}\hat{v})\, d(\mu) + z\int_\Sigma \varphi(\hat{g}\hat{v})\, d(\nu) \\ &= Q_\mu(\varphi)(\hat{v}) + zQ_\nu(\varphi)(\hat{v}). \end{aligned}$$

Note that, for a fixed vector $\hat{v}$, each $Q^n_{\mu_z}(\varphi)(\hat{v})$ is a polynomial of degree smaller or equal to $n$, in particular, the map $z \mapsto Q^n_{\mu_z}(\varphi)(\hat{v})$ is holomorphic for $\mu_z \in V$. Therefore, for every $z \in V(\mu,\nu)$, the limit function is a uniform limit of holomorphic functions, hence $z \mapsto Q^\infty_{\mu_z}\varphi(\hat{v})$ is holomorphic. In other words, the Lyapunov exponent is G-holomorphic in the neighbourhood $V \subset \mathcal{M}_1(\Sigma)$ of $\mu_0$. Together with the continuity, we conclude that it is indeed holomorphic. This concludes the proof of item (1) the theorem.

Let us now prove item (2). We drop the assumption of irreducibility of $\mu_0$ and instead we assume that $\mathrm{supp}\mu_0 = \Sigma$. Let $W$ be a non trivial vector subspace of $\mathbb{R}^d$ that is invariant for $\mu_0$ almost every matrix $g$. We saw in section 3.2.2 that the measure $\mu_0$ defines the measures $\mu_{0,W}$ and $\mu_{0,\mathbb{R}^d/W}$ in $\mathrm{GL_d(W)}$ and $\mathrm{GL_d(\mathbb{R}^d/W)}$. Moreover, they induce the linear cocycles restricted to $W$ and to $\mathbb{R}^d/W$, with Lyapunov exponents $L_1(\mu_{0,W})$ and $L_1(\mu_{0,\mathbb{R}^d/W})$.

By lemma 3.6, $L_1(\mu_0) = \max\{L_1(\mu_{0,W}), L_1(\mu_{0,\mathbb{R}^d/W})\}$. Without loss of generality, we may suppose that $L_1(\mu_0) = L_1(\mu_{0,W})$. The other case is similar. The fact that we consider $\mathrm{supp}\mu_0 = \Sigma$ implies that $gW = W\ \forall g \in \Sigma$. Therefore, for every $\mu \in \mathcal{M}_1(\Sigma)$ also satisfies $gW = W$ for $\mu$-a.e $g$.

By corollary $B$ of [39], if $\mu_n \to \mu_0$ in the weak star topology and $\mathrm{supp}\mu_n \subset \mathrm{supp}\mu_0$ for every $n$, then $L_1(\mu_n) \to L_1(\mu_0)$. Therefore, the continuity of the Lyapunov exponents imply that for every $\mu$ sufficiently close to $\mu_0$, it holds that $L_1(\mu) = L_1(\mu_W)$. This happens because $L_1(\mu) = \max\{L_1(\mu_W), L_1(\mu_{\mathbb{R}^d/W})\}$.

If $\mu_{0,W}$ is irreducible, then the map $\mu \mapsto L_1(\mu_W)$ is holomorphic. Since $L_1(\mu) = L_1(\mu_W)$ in a neighborhood of $\mu_0$, we conclude that $\mu \mapsto L_1(\mu)$ is also holomorphic in a neighborhood of $\mu_0$.

If $\mu_{0,W}$ is not irreducible, there exists another non trivial invariant subspace $W' \subset W$. The measure $\mu_{0,V}$ defines measures $\mu_{0,W'}$ and $\mu_{0,W/W'}$. Then we do the same procedure. Since the invariant subspaces are of decreasing dimension, this process must stop after a finite number of steps. Therefore, we conclude the proof of the theorem. ■

**Remark 5.4** *Note that under the quasi-irreducibility hypothesis the analyticity holds in a neighborhood of $\mu_0$, even if its support is not total. In order to exclude the irreducibility hypothesis, we assume in item (2) that $\mu_0$ has full* *support. When the measure $\mu_0$ is not quasi irreducible, is the full support hypothesis strictly necessary?*

*The answer is yes, and it is based on Kifer's example. Recall that in example 3.3, if $p_2 > 0$, the top Lyapunov exponent of this cocycle is zero. In the limit case, with $p_2 = 0$, the cocycle is generated only by $g_1$, which is a not quasi irreducible cocycle, since it preserves both axis and one of them is an equator, i.e the top Lyapunov exponent in this direction is not maximum. Moreover, the top Lyapunov exponent of the limit cocycle is equal to* $\log 2$, *which makes this cocycle a discontinuity point. Therefore, if the cocycle is not quasi irreducible and also does not have full support, it can be discontinuity point of the Lyapunov exponent.*

### 5.2.3
### Corollaries and remarks

Let $\Sigma$ be an abstract compact space, $X = \Sigma^{\mathbb{N}}$ and $\sigma\colon X \to X$ be the forward shift on $X$. We fix a measurable and bounded function $A\colon \Sigma \to \mathrm{GL_d}(\mathbb{R})$ and denote also by $A$ the locally constant (fiber) map $A\colon X \to \mathrm{GL_d}(\mathbb{R})$ given by $A((x_n)_{n\in\mathbb{N}}) = A(x_0)$.

Given $\mu \in \mathrm{Prob}(\Sigma)$, let $\mu^{\mathbb{N}}$ be the product (Bernoulli) measure on $X$. A random (Bernoulli) locally constant linear cocycle $F_A\colon X \times \mathbb{R}^d \to X \times \mathbb{R}^d$ relative to the product measure $\mu^{\mathbb{N}}$ is a skew product transformation such that

$$F_A(\omega, v) = (\sigma(\omega), A(x)v).$$

Its iterates are given by

$$F_A^n(\omega, v) = (\sigma^n(\omega), A^n(\omega)v),$$

where $A^n(\omega) := A(\omega_{n-1}) \dots A(\omega_1)A(\omega_0)$.

A seminal result from Furstenberg and Kesten states that under the integrability condition $\log^+\|A^{\pm}\| \in L^1(\mu)$, the limit

$$L_1(A, \mu) = \lim_n \frac{1}{n} \log\|A^n(\omega)\|$$

exists $\mu^{\mathbb{N}}$ a.e. and it is called the top Lyapunov exponent of this cocycle.

**Corollary 5.2** *In this context, let $\Sigma \subset \mathrm{GL}_d(\mathbb{R})$ be a compact subset and $X = \Sigma^{\mathbb{N}}$. Fix any locally constant fiber map $A : X \to \mathbb{R}^d$. Let $\mu_0 \in$* $\mathrm{Prob}(\Sigma)$ *and assume that $L_1(A, \mu_0) > L_2(A, \mu_0)$.*

*(1) If $\mu_0$ is quasi-irreducible, then $\mu \mapsto L_1(A, \mu)$ is real analytic with respect to the total variation norm in a neighborhood of $\mu_0$.*

(2) *If* $\operatorname{supp}(\mu_0) = \Sigma$*, then* $\mu \mapsto L_1(A, \mu)$ *is real analytic with respect to the total variation norm in a neighborhood of* $\mu_0$*.*

*Proof.* Consider the push forward measure on $\operatorname{Prob}(\mathrm{GL_d}(\mathbb{R}))$ given by $A_*\mu$. By the boundedness of $A$, the support of $A_*\mu$ remains compact, and therefore its Lyapunov exponent is well defined. A change of variables shows that the Lyapunov exponent $L_1(A_*\mu)$ associated to the measure $A_*\mu$ is equal to $L_1(A, \mu)$. Moreover, the application $A_* : \mathcal{M}(\Sigma) \to \mathcal{M}(\mathrm{GL_d}(\mathbb{R}))$ is a linear continuous (and, therefore, analytic) mapping that preserves probabilities. Since the composition on analytic maps is analytic, it follows by Theorem 5.2 that the map $L_1(A, \mu) = L_1(A_*\mu)$ is analytic with respect to $\mu$, which guarantees that the result in theorem 5.2 also holds for arbitrary locally constant linear cocyles. ■

A second corollary is an analogue of the finite case, in which the support is of the measure is a fixed compact set on $\mathrm{GL_d}(\mathbb{R})$ and we look at the dependence on the probability weights. Let $\mu_0 \in \operatorname{Prob}(\Sigma)$ be a reference measure of full support. We restrict to the measures in $\Sigma$ which are absolutely continuous with respect to $\mu_0$.

By the Radon-Nikodym Theorem, this space is identified with the space $L^1(\mu_0)$ of integrable complex functions with respect to $\mu_0$ through the map $I : L^1(\mu_0) \to \mathcal{M}(\Sigma)$ given by

$$I(f)(E) = \int_E f(x) d\mu_0(x)$$

for every measurable set $E$. The map $I$ is an isomorphism between $L^1(\mu_0)$ and the measures on $\mathcal{M}(\Sigma)$ which are absolutely continuous with respect to $\mu_0$. Observe that, given $f, g \in L^1(\mu_0)$, it follows that

$$\|I(f) - I(g)\|_{TV} \leq \|f - g\|_1 \leq \|f - g\|_p,$$

where $\|.\|_{TV}$ denotes the total variation norm, $\|.\|_1$ denotes the $L^1$ norm and $\|.\|_p$ denotes the $L^p$ norm, with $p \in [1, +\infty]$. This fact guarantees that, given $r > 0$, it follows that $B_p(f, r) \subset B_1(f, r) \subset B_{TV}(I(f), r)$, where each of the previous sets denotes an open ball on its respective norm.

This observation, together with the Theorem 5.2, proves the following.

**Corollary 5.3** *Let* $\mu_0 \in \operatorname{Prob}(\Sigma)$ *have full support and assume that* $L_1(\mu_0) > L_2(\mu_0)$*. For* $p \in [1, +\infty]$*, define*

$$L_1^p(\mu_0) := \left\{ f \colon \Omega \to \mathbb{R} \colon f \in L^p(\mu_0) \ \text{and} \int f(x) d\mu_0(x) = 1 \right\}.$$

*Then the Lyapunov exponent $L_1 \colon L_1^p(\mu_0) \to \mathbb{R}$ is locally a real analytic function of $\mu_0$ with respect to the $L^p$ norm.*

We now consider the set in which $L_1$ is analytical. We say that a measure $\mu \in \mathcal{M}(\mathrm{GL_d}(\mathbb{R}))$ is *irreducible* if there is no proper subspace $V \subset \mathrm{GL_d}(\mathbb{R})$ such that $gV = V$ for $\mu$-a.e.$g$. Notice that every irreducible measure is quasi-irreducible.

Observe that irreducibility is a dense property with respect to the total variation norm. Indeed, let $\mu_0$ be a probability in $\mathrm{GL_d}(\mathbb{R})$, and let $\nu$ be another probability in $\mathrm{GL_d}(\mathbb{R})$, with compact support. If $\nu$ is irreducible and given $\varepsilon > 0$, then $\mu_\varepsilon = (1-\varepsilon)\mu + \varepsilon\nu$ is an irreducible probability in $\mathrm{GL_d}(\mathbb{R})$. To see this, let $V$ be a proper subspace of $\mathrm{GL_d}(\mathbb{R})$. Since $\nu$ is irreducible, there exists a borelian set $B \subset \mathrm{GL_d}(\mathbb{R})$ such that $\nu(B) > 0$ and $gV \neq V$ for every $g \in B$. Notice that $\mu_\varepsilon(B) = (1-\varepsilon)\mu(B) + \varepsilon\nu(B) \geq \varepsilon\nu(B) > 0$, so it follows that $\mu_\varepsilon$ is irreducible.

Notice also that $|\mu_\varepsilon - \mu| = |\varepsilon\nu - \varepsilon\mu| \leq 2\varepsilon$, so we can choose $\varepsilon$ sufficiently small such that $\mu_\varepsilon$ is arbitrarily close to $\mu$. Moreover, supp $\mu_\epsilon =$ supp $\mu \cup$ supp $\nu$, so if $\mu$ has compact support, $\mu_\epsilon$ also has compact support, and if both supp $\mu \subset \Sigma$ and supp $\nu \subset \Sigma$, then supp $\mu_\epsilon \subset \Sigma$.

We also observe that in [33], Kifer proved that being irreducible is an open property on $\mathrm{Prob}(\mathrm{GL_d}(\mathbb{R}))$ with respect to the weak* topology. Since the total variation norm generates a finer topology than the weak* topology, it follows that being irreducible is also an open property with respect to the total variation norm. Therefore, by the first result of 5.2, we can conclude that $L_1$ is analytical on the set of compactly supported irreducible measures on $\mathrm{GL_d}(\mathbb{R})$, which is a dense open set on the space $\mathrm{Prob}(\mathrm{GL_d}(\mathbb{R}))$ with respect to the total variation norm.

To conclude this section, we show that the restriction of the probabilities to a compact set $\Sigma$ in 5.2 cannot be removed without any other extra hypothesis to replace it.

Indeed, let $\mu \in \mathrm{Prob}(\mathrm{GL(d)})$ be a compactly supported measure with $L_1(\mu) > L_2(\mu)$. Let $a, \varepsilon > 0$ and consider the measure $\mu_{a,\varepsilon} := (1-\varepsilon)\mu + \varepsilon\delta_{aI}$, where $I$ is the identity matrix. This measure is $2\varepsilon$ close to $\mu$ in the total variation distance, however, depending on the choice of $a$, their supports may be very distant from each other.

By a previous comment, the measure $\mu_{a,\varepsilon}$ is compactly supported. Moreover, the identity:

$$L_1(\nu) + ... + L_d(\nu) = \int \log|\det g| \, d\nu(g),$$

which is true for any compactly supported measure on GL(d), gives us that:

$$L_1(\mu_{a,\varepsilon}) \geq \frac{L_1(\mu_{a,\varepsilon}) + ... + L_d(\mu_{a,\varepsilon})}{d} = \frac{1}{d}\int \log|\det g|\; d\mu_{a,\varepsilon}(g)$$

Hence, it is possible to choose $a$ sufficiently large such that the previous term is much bigger than the Lyapunov exponent $L_1(\mu)$. Since there is no restriction on the support of the measures, for every $\varepsilon > 0$, one can find infinitely measures of the form $\mu_{a,\varepsilon}$ which are close to $\mu$ in the total variation norm, but that have Lyapunov exponents very far from it. Thus, $L_1$ cannot even be continuous.

Although it is clear that compactness plays an important role in this result, in recent conversations with Duarte and Graxinha, they explained that the techniques from section 4 are being extended by them to the non-compact setting assuming, however, a certain growth condition, namely the finiteness of an exponential moment. Independently, Barrientos and Malicet [5] were also able to adapt some of the techniques presented in section 4 to the non-compact setting and obtain the strong mixing of the Markov operator. Therefore, we believe that the same kind of hypothesis should be sufficient to prove theorem 5.2 in this more general setting.

# 6
# Dichotomic Behavior in the Singular Setting: Analyticity vs Discontinuity

A version of the Bochi-Mañé dichotomy theorem in the context of linear cocycles states that given any measure preserving dynamical system $(X, f, \mu)$ where $f$ is an aperiodic (meaning its set of periodic points has measure zero) homeomorphism, for any fiber map $A\colon X \to \mathrm{SL}_2(\mathbb{R})$, either the cocycle associated with $A$ is hyperbolic over the support of $\mu$ or else $A$ is approximated in the $C^0$ topology by fiber maps with zero Lyapunov exponent.

By Ruelle's theorem, hyperbolic cocycles are points of analyticity of the Lyapunov exponent; moreover, since the Lyapunov exponent is always upper semicontinuous, it must be continuous at any $\mathrm{SL}_2(\mathbb{R})$-valued cocycle with zero Lyapunov exponent. Therefore, a fiber map $A\colon X \to \mathrm{SL}_2(\mathbb{R})$ is a continuity point for the first Lyapunov exponent in $C^0(X, \mathrm{SL}_2(\mathbb{R}))$ if and only if the corresponding linear cocycle is either hyperbolic over the support of $\mu$ (in which case the Lyapunov exponent is in fact analytic) or $L_1(A) = 0$.

In other words, if $L_1(A) > 0$, the first Lyapunov exponent is either analytic or discontinuous at $A$ in the $C^0$ topology of $\mathrm{SL}_2(\mathbb{R})$ cocycles. However, this behavior changes dramatically when the fiber map is highly regular and it varies in a space endowed with an appropriate, strong topology.

Indeed, consider a locally constant linear cocycle over a Bernoulli shift. More precisely, let $\mathcal{A} := \{1, \dots, k\}$ be a finite alphabet and let $p = (p_1, \dots, p_k)$ be a probability vector with $p_i > 0$ for all $i$. Denote by $X := \mathcal{A}^{\mathbb{Z}}$ the space of bi-infinite sequences $\omega = \{\omega_n\}_{n\in\mathbb{Z}}$ on this alphabet, which we endow with the product measure $\mu = p^{\mathbb{Z}}$. Let $\sigma\colon X \to X$ be the corresponding forward shift $\sigma\omega = \{\omega_{n+1}\}_{n\in\mathbb{Z}}$. Then $(X, \mu, \sigma)$ is a measure preserving, ergodic dynamical system called a Bernoulli shift (in finite symbols).

Let $\mathrm{Mat}_2(\mathbb{R})$ denote the semigroup of $2 \times 2$ matrices. A $k$-tuple $\underline{A} = (A_1, \dots, A_k) \in \mathrm{Mat}_2(\mathbb{R})^k$ determines the locally constant fiber map $A\colon X \to \mathrm{Mat}_2(\mathbb{R})$, $A(\omega) = A_{\omega_0}$, which in turns determines a linear cocycle over the Bernoulli shift, referred to as a (random) Bernoulli cocycle. We identify this cocycle with the fiber map $A$ and with the tuple $\underline{A}$ and denote by $L_1(\underline{A}) = L_1(\underline{A}, p)$ its first Lyapunov exponent.

When we restrict to *invertible* random linear cocycles, that is, when $\underline{A} = (A_1, \ldots, A_k) \in \mathrm{GL}_2(\mathbb{R})^{\mathrm{k}}$, the Lyapunov exponent is always continuous; moreover, it has a good modulus of continuity as we saw in the previous chapters. This has been the subject of intense research throughout the years, starting with the celebrated work of Furstenberg and Kifer [29].

However, as it turns out, the behavior of the first Lyapunov exponent at random cocycles with *both singular and invertible* components is strikingly different, showing a dichotomy in the spirit of Mañé-Bochi's (see theorem 6.3).

This chapter is based on the joint work [18] with Duarte, Graxinha and Klein. In section 6.1 we generalize Avila, Bochi, Yoccoz [2] theory of projective uniform hyperbolicity in terms of multi-cones, from $\mathrm{SL}_2(\mathbb{R})$-valued cocycles to $\mathrm{Mat}_2^+(\mathbb{R})$-valued cocycles. Recall that $\mathrm{Mat}_2^+(\mathbb{R})$ is the semigroup of matrices $g \in \mathrm{Mat}_2(\mathbb{R})$ with $\det(g) \geq 0$. In section 6.2 we use this result to prove a Mañé-Bochi type of dichotomy for singular cocycles and also derive other interesting corollaries from it. In particular, we conclude that Lebesgue almost every cocycle $\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k$ with both singular and invertible components satisfies a sharp regularity dichotomy: it is either a point of analyticity or a point of discontinuity of the Lyapunov exponent.

## 6.1 Projective uniform hyperbolicity

### 6.1.1 Extension of the multi-cone theory to non-invertible cocycles

Consider a random linear cocycle $F\colon X \times \mathbb{R}^2 \to X \times \mathbb{R}^2$ determined by the data $(\underline{A}, p)$ where $\underline{A} := (A_i)_{i \in \mathcal{A}} \in \mathrm{Mat}_2^+(\mathbb{R})^k$.

**Definition 6.1** *We say that a linear cocycle $F$ is projectively uniformly hyperbolic if there exists an $F$-invariant decomposition into one-dimensional subspaces $\mathbb{R}^2 = E_0(\omega) \oplus E_1(\omega)$ where the sub-bundles $X \ni \omega \mapsto E_i(\omega)$ are continuous functions and there exists $n \in \mathbb{N}$ such that $\|A^n|_{E_0}(\omega)\| < \|A^n|_{E_1}(\omega)\|$ for all $\omega \in X$.*

Since $X$ is compact and $A\colon X \to \mathrm{Mat}_2^+(\mathbb{R})$ is continuous, the last condition is equivalent to

$$\sup_{\omega \in X} \frac{\|A^n|_{E_0}(\omega)\|}{\|A^n|_{E_1}(\omega)\|} \leq \lambda^{-1} < 1,$$

for some $\lambda > 1$. This is further equivalent to the existence of $c > 0$ and $\lambda > 1$ such that for all $\omega \in X$ and all $n \in \mathbb{N}$,

$$\|A^n|_{E_1}(\omega)\| \geq c\lambda^n \, \|A^n|_{E_0}(\omega)\|.$$

Note that projective uniform hyperbolicity is also sometimes referred to as *dominated splitting.*

Given an *invertible* matrix $A \in \mathrm{Mat}_2(\mathbb{R})$, its induced projection action $\hat{A}\colon \mathbb{P}(\mathbb{R}^2) \to \mathbb{P}(\mathbb{R}^2)$ is given by $\hat{A}\hat{v} := \widehat{Av}$. If $A$ has rank 1 (that is, if it is nonzero and noninvertible), we define its projection action as the constant map $\hat{A}\hat{v} := \hat{r}$, where $r = \mathrm{Range}(A)$.

Moreover, let $\mathrm{m}(A)$ denote its co-norm (its smallest singular value). If $A$ is invertible, $\mathrm{m}(A) = \|A^{-1}\|^{-1}$, otherwise $\mathrm{m}(A) = 0$. An immediate consequence of it is that, if a cocycle $\underline{A}$ admits a singular component, then $L_2(\underline{A}) = -\infty$.

**Definition 6.2** *An invariant multi-cone for* $\underline{A} = (A_i)_{i\in\mathcal{A}}$ *is a nonempty set* $M$ *such that:*

1. *$M$ is an open subset of* $\mathbb{P}(\mathbb{R}^2)$*,*
2. *its closure* $\bar{M} \neq \mathbb{P}(\mathbb{R}^2)$*,*
3. $A_i M \Subset M$*, i.e.,* $\overline{A_i M} \subset M$ *for every* $i \in \mathcal{A}$ *.*

The next theorem extends to $\mathrm{Mat}_2^+(\mathbb{R})$-valued cocycles the results of Avila, Bochi, Yoccoz [2, Theorem 2.3] and Yoccoz [44, Proposition 2] established for $\mathrm{SL}_2(\mathbb{R})$-cocycles over sub-shifts of finite type.

**Theorem 6.1** *Given a random linear cocycle* $\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k$*, the following are equivalent:*

*(1)* $\underline{A}$ *is projectively uniformly hyperbolic.*

*(2) There exist* $c > 0$ *and* $\lambda > 1$ *such that for all* $n \in \mathbb{N}$ *and* $\omega \in X$*,* $\|A^n(\omega)\| \geq c\,\lambda^n\, \mathrm{m}(A^n(\omega))$*.*

*(3)* $\underline{A}$ *admits an invariant multi-cone.*

*Proof.* (1) $\Rightarrow$ (2): Suppose that $\underline{A}$ is projectively uniformly hyperbolic (it admits a dominated splitting). Then for some $c > 0$ and $\lambda > 1$ we have that for all $\omega \in X$ and $n \in \mathbb{N}$,

$$\frac{\|A^n(\omega)\|}{\mathrm{m}(A^n(\omega))} \geq \frac{\|A^n|_{E_1}(\omega)\|}{\|A^n|_{E_0}(\omega)\|} \geq c\,\lambda^n.$$

(2) $\Rightarrow$ (1): This is a straightforward adaptation of [42, Proposition 2.1] to the case where $\underline{A}$ takes values in $\mathrm{Mat}_2^+(\mathbb{R})$. The idea of this proof is to define the Oseledets invariant directions $E_0(\omega)$ and $E_1(\omega)$ as uniform limits of continuous functions, by exploiting the contracting behavior of the cocycle action on fibers due to the cocycle's hyperbolicity. This implies that the Oseledets splitting is continuous. In our setting, when singular matrices appear, their actions contract the whole projective space to points. This helps the convergence and poses no problem regarding the continuity of the approximations, which are defined as singular directions of the the matrices $A^n(\omega)$ and $A^n(\sigma^{-n}\omega)$.

For the remaining implications we need a *desingularization construction* that associates a family of invertible cocycles $\underline{A}^*(\mu) \in \mathrm{SL}_2(\mathbb{R})$ to every singular cocycle $\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k$. For each $A_i \in \mathrm{Mat}_2^+(\mathbb{R})$ with $i \in \mathcal{A}$, consider its singular value decomposition: $A_i = R_i \Sigma_i R_i^*$. If $A_i$ is not invertible, consider a small perturbation of $\Sigma_i$ that transforms its zero singular value into a small constant $\mu^{-2}$. In other words, if $A_i = R_i \begin{bmatrix} \|A_i\| & 0 \\ 0 & 0 \end{bmatrix} R_i^*$, let

$$\tilde{A}_{\mu,i} := \|A_i\| R_i \begin{bmatrix} 1 & 0 \\ 0 & \mu^{-2} \end{bmatrix} R_i^* = \frac{\|A_i\|}{\mu} R_i \begin{bmatrix} \mu & 0 \\ 0 & \mu^{-1} \end{bmatrix} R_i^* .$$

If $A_i$ is invertible put $\tilde{A}_{\mu,i} := A_i$. Then set $A^*_{\mu,i} := \frac{1}{\sqrt{\det(\tilde{A}_{\mu,i})}} \tilde{A}_{\mu,i}$. The cocycle $\underline{A}^*_\mu = (A^*_{\mu,i})_{i\in\mathcal{A}} \in \mathrm{SL}_2(\mathbb{R})^k$ has the same projective action as $(\tilde{A}_{\mu,i})_{i\in\mathcal{A}}$, which for large $\mu$ approximates that of $\underline{A}$.

**Lemma 6.1** *Let* $\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k$.

*(1) If* $(M_i)_{i\in\mathcal{A}}$ *is an invariant multi-cone of* $\underline{A}^*_\mu$*, for some* $\mu > 0$*, then it is also an invariant multi-cone for* $\underline{A}$.

*(2) If* $(M_i)_{i\in\mathcal{A}}$ *is an invariant multi-cone of* $\underline{A}$ *then it is also an invariant multi-cone of* $\underline{A}^*_\mu$ *for all sufficiently large* $\mu$.

*Proof.* (1) Since $M$ is an invariant multi-cone for $\underline{A}^*_\mu$, $\tilde{A}_{\mu,i} M \Subset M$ for every $i \in \mathcal{A}$. Moreover, the pair of matrices $A_i$ and $\tilde{A}_{\mu,i}$ share the same singular directions, but their contraction strengths are different (infinite in the case of non-invertible matrices vs. finite for invertible ones). Thus since the contraction is stronger for non-invertible matrices, we conclude that $A_i M \subseteq A_{\mu,i} M \Subset M$ for every $i \in \mathcal{A}$.

(2) This holds because multi-cones are stable under perturbations and $\underline{A} = \lim_{\mu\to\infty} \underline{\tilde{A}}_\mu$, while $\underline{A}^*_\mu$ and $\underline{\tilde{A}}_\mu$ share their invariant multi-cones, because they induce the same action on $\mathbb{P}(\mathbb{R}^2)$. ∎

We now return to the proof of the theorem.

(1) $\Rightarrow$ (3): Since projective uniform hyperbolicity is an open property, if $\underline{A}$ is projectively uniformly hyperbolic then the approximating cocycles $\underline{\tilde{A}}_\mu$ and $\underline{A}^*_\mu$ are also projectively uniformly hyperbolic for large $\mu$. By [2, Theorem 2.3], the cocycle $\underline{A}^*_\mu$ admits an invariant multi-cone $M$. Therefore, by (1) of Lemma 6.1, $M$ is also an invariant multi-cone for $\underline{A}$.

(3) $\Rightarrow$ (2): Suppose $\underline{A}$ admits an invariant multi-cone $M = (M_i)_{i\in\mathcal{A}}$. By (2) of Lemma 6.1, for some large $\mu$, $M$ is also an invariant multi-cone for $\underline{A}^*_\mu$. Therefore, by [2, Theorem 2.3], the cocycle $\underline{A}^*_\mu$ is uniformly hyperbolic and by [2, Proposition 2.1] there exists $c > 0$ and $\lambda > 1$ such that $\|(A^*_\mu)^n(\omega)\| \geq c\,\lambda^n$ for all $\omega \in X$, which in turn implies that

$$\frac{\|A^n(\omega)\|}{\mathrm{m}(A^n(\omega))} \geq \frac{\|\tilde{A}^n_\mu(\omega)\|}{\mathrm{m}(\tilde{A}^n_\mu(\omega))} = \|(A^*_\mu)^n(\omega)\|^2 \geq c\,\lambda^{2n} \quad \forall\, \omega \in X.$$

Note that either $A^n(\omega) = A_{\omega_{n-1}} \cdots A_{\omega_0}$ does not include singular matrices so $A^n(\omega) = \tilde{A}^n_\mu(\omega)$, or else it does and then $\mathrm{m}(A^n(\omega)) = 0$, which implies that the left-hand side of the above inequality is $\infty$. ■

**Definition 6.3** *We say that $\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k$ has rank 1 if*

$$\lim_{n\to\infty} \mathrm{rank}(A^n(\omega)) = 1 \quad \text{for a.e. } \omega \in X.$$

We do not consider cocycles of rank 0 because in this case the first Lyapunov exponent is equal to $-\infty$. Cocycles of rank 2, that is, in $\mathrm{GL}_2(\mathbb{R})^{\mathrm{k}}$, are also not considered here because, as mentioned in the introduction, they have already been extensively studied.

**Remark 6.1** *A cocycle $\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k$ has rank 1 if and only if*

*(1)* $\mathrm{rank}(A_i) = 1$ *for some $i \in \mathcal{A}$ and*

*(2) $\underline{A}$ has no null word, i.e. $A^n(\omega) := A_{\omega_{n-1}} \cdots A_{\omega_1} A_{\omega_0} \neq 0$ for every $\omega \in X$.*

From now on we only consider cocycles $\underline{A}$ of rank 1. We split the alphabet $\mathcal{A}$ into two parts:

$$\mathcal{A}_{\mathrm{inv}} := \{i \in \mathcal{A} \colon \det A_i \neq 0\} \quad \text{and} \quad \mathcal{A}_{\mathrm{sing}} := \{i \in \mathcal{A} \colon \mathrm{rank} A_i = 1\}.$$

For $i \in \mathcal{A}_{\mathrm{sing}}$, write $r_i := \mathrm{Range}(A_i)$ and $k_i := \mathrm{Ker}(A_i)$ and set

$$\mathcal{K}(\underline{A}) := \{k_i \colon i \in \mathcal{A}_{\mathrm{sing}}\},$$
$$\mathcal{R}(\underline{A}) := \{r_i \colon i \in \mathcal{A}_{\mathrm{sing}}\}.$$

Moreover, we define the following sets:

$$\mathcal{W}^+ := \overline{\bigcup_{i\in\mathcal{A}_{\rm sing}} \bigcup_{n\geq 0} \{A^n(\underline{\omega})\, r_i\}} \quad \text{and} \quad \mathcal{W}^- := \overline{\bigcup_{i\in\mathcal{A}_{\rm sing}} \bigcup_{n\geq 0} \{A^{-n}(\underline{\omega})\, \kappa_i\}}.$$

These sets represent the forward (backward) iterates of the ranges (kernels) and their accumulation points. They play a central role in the theory that we develop to study singular cocycles.

**Definition 6.4** *If $M$ is an invariant multi-cone for the invertible cocycle $(A_i)_{i\in\mathcal{A}_{\rm inv}}$, we define the sets $K^u_{\rm inv}$ and $K^s_{\rm inv}$ as in [2, Subsection 2.3] by*

$$K^u_{\rm inv} = \bigcap_{n=0}^{\infty} \bigcup_{i_1,\cdots,i_n\in\mathcal{A}_{\rm inv}} A_{i_n}\cdots A_{i_1}(M)$$

*and*

$$K^s_{\rm inv} = \bigcap_{n=0}^{\infty} \bigcup_{i_1,\cdots,i_n\in\mathcal{A}_{\rm inv}} (A_{i_n}\cdots A_{i_1})^{-1}(\mathbb{P}^1\setminus \overline{M}).$$

**Proposition 6.1** *If $M$ is an invariant multi-cone for the invertible cocycle $(A_i)_{i\in\mathcal{A}_{\rm inv}}$ then:*

*(1) $K^u_{\rm inv}$ is the set of unstable Oseledets directions $E^u(\omega)$ of the cocycle $(A_i)_{i\in\mathcal{A}_{\rm inv}}$ over the set of points $\omega\in\mathcal{A}^{\mathbb{Z}}_{\rm inv}$,*

*(2) $K^s_{\rm inv}$ is the set of stable Oseledets directions $E^s(\omega)$ of the cocycle $(A_i)_{i\in\mathcal{A}_{\rm inv}}$ over the set of points $\omega\in\mathcal{A}^{\mathbb{Z}}_{\rm inv}$.*

*Proof.* Fix $\hat{w}\in M$. Then by dominated splitting, $\hat{v}\in K^u_{\rm inv}$ is the limit

$$\hat{v} = \lim_{n\to\infty} A_{\omega_{-1}}\,\cdots\, A_{\omega_{-n}}\,\hat{w} = \lim_{n\to\infty} A^n(\sigma^{-n}\omega)\,\hat{w} = E^u(\omega),$$

for some sequence $\omega\in\mathcal{A}^{\mathbb{Z}}_{\rm inv}$. Similarly, $\hat{v}\in K^s_{\rm inv}$ is the limit

$$\hat{v} = \lim_{n\to\infty} (A_{\omega_{n-1}}\,\cdots\, A_{\omega_0})^{-1}\,\hat{w} = \lim_{n\to\infty} A^{-n}(\sigma^n\omega)\,\hat{w} = E^s(\omega),$$

for some sequence $\omega\in\mathcal{A}^{\mathbb{Z}}_{\rm inv}$. ■

**Proposition 6.2** *Let $M$ be an invariant multi-cone of the invertible cocycle $(A_i)_{i\in\mathcal{A}_{\rm inv}}$ and take $v\in M\setminus K^u_{\rm inv}$. Then there exists an invariant multi-cone $\tilde{M}$ of $(A_i)_{i\in\mathcal{A}_{\rm inv}}$ such that $\tilde{M}\Subset M$ and $v\notin\tilde{M}$.*

*Proof.* Define the set

$$M_n := \bigcup_{\substack{|\omega|=n \\ \omega\in\mathcal{A}^n_{inv}}} A^n(\underline{\omega})M$$

where the union is taken over all admissible invertible words of length $n$. We claim that $M_n$ is an invariant multi-cone of of $(A_i)_{i\in\mathcal{A}_{\text{inv}}}$, for every $n\in\mathbb{N}$ and that there exists a sufficiently large $N\in\mathbb{N}$ such that $v\notin M_N$. Let us prove, by induction, that $M_{n+1}\Subset M_n\Subset\cdots\Subset M$. Since $M$ is an invariant multi-cone associated to $(A_i)_{i\in\mathcal{A}_{\text{inv}}}$, then $A(\omega)M\Subset M$ for every invertible word $\omega$ such that $|\omega|=1$. Thus $\bigcup_{\substack{|\omega|=1 \\ \omega\in\mathcal{A}^1_{inv}}} A(\omega)M\Subset M$ as the union is finite. Then

$$\begin{aligned}M_{n+1} &= \bigcup_{\substack{|\omega|=n+1 \\ \omega\in\mathcal{A}^{n+1}_{inv}}} A^{n+1}(\underline{\omega})M = \bigcup_{\substack{|z|=1 \\ z\in\mathcal{A}^1_{inv}}} A(\underline{z}) \bigcup_{\substack{|\omega|=n \\ \omega\in\mathcal{A}^n_{inv}}} A^n(\underline{\omega})M \\ &= \bigcup_{\substack{|z|=1 \\ z\in\mathcal{A}^1_{inv}}} A(\underline{z})M_n \Subset M_n.\end{aligned}$$

In particular, as $M\neq\mathbb{P}(\mathbb{R}^2)$ we have that $M_n\neq\mathbb{P}(\mathbb{R}^2)$ $\forall n\in\mathbb{N}$. Notice that as $M$ is open and the cocycle $(A_i)_{i\in\mathcal{A}_{\text{inv}}}$ is invertible, then $M_n$ is open for every $n\in\mathbb{N}$. Therefore $M_n$ is an invariant multi-cone $\forall n\in\mathbb{N}$. To prove that for any $v\in M\setminus K^u_{\text{inv}}$ there exists $N\in\mathbb{N}$ such that $v\notin M_N$ we simply notice that $K^u_{\text{inv}}$ is closed and that

$$\lim_n M_n := \lim_n \bigcup_{\substack{|\omega|=n \\ \omega\in\mathcal{A}^n_{inv}}} A^n(\omega)M = \bigcap_{n=0}^{\infty} \bigcup_{\substack{|\omega|=n \\ \omega\in\mathcal{A}^n_{inv}}} A^n(\omega)M =: K^u_{\text{inv}},$$

because $M_n$ is a monotonous sequence. ■

### 6.1.2 A criteria for projective uniform hyperbolicity in the singular setting

**Proposition 6.3** *Consider a cocycle* $\underline{A}\in\text{Mat}^+_2(\mathbb{R})^k$ *of rank* 1 *such that* $\underline{A}_{\text{inv}} := (A_i)_{i\in\mathcal{A}_{\text{inv}}}$ *is projectively uniformly hyperbolic,* $\underline{A}$ *is not diagonalizable and* $\mathcal{W}^+\cap\mathcal{W}^-=\emptyset$*. Then*

$$\mathcal{K}(\underline{A})\cap K^u_{\text{inv}}=\emptyset \quad \textit{and} \quad \mathcal{R}(\underline{A})\cap K^s_{\text{inv}}=\emptyset.$$

*Proof.* We only prove that $\mathcal{K}(\underline{A})\cap K^u_{\text{inv}}=\emptyset$, the other proof being analogous. Suppose by contradiction that there exists $\hat{k}\in\mathcal{K}(\underline{A})\cap K^u_{\text{inv}}$. We split the proof into two cases:

- $\underline{A}_{\text{inv}}$ is not diagonalizable, and
- $\underline{A}_{\text{inv}}$ is diagonalizable but $\underline{A}$ is not.

Let us start with the assumption that $\underline{A}_{\text{inv}}$ is not diagonalizable. We will say that $\hat{v} \in \mathbb{P}(\mathbb{R}^2)$ is $\underline{A}_{\text{inv}}$-invariant if $A_i\,\hat{v} = \hat{v}$ for all $i \in \mathcal{A}_{\text{inv}}$.

**Lemma 6.2** *There exists an $\underline{A}_{\text{inv}}$-invariant element in $K^u_{\text{inv}}$ if and only if $\#K^u_{\text{inv}} = 1$. Analogously, there exists an $\underline{A}_{\text{inv}}$-invariant element in $K^s_{\text{inv}}$ if and only if $\#K^s_{\text{inv}} = 1$.*

*Proof.* Suppose $\#K^u_{\text{inv}} = 1$. By Proposition 6.1, there exists $\hat{v} \in K^u_{\text{inv}}$ such that for almost every $\omega \in X$, $E^u(\omega) = \hat{v}$. Moreover, by the invariance of the Oseledets subspaces, for almost every $\omega \in X$, $E^u(f(\omega)) = A(\omega)E^u(\omega)$. Hence, $\hat{v} = \hat{A}(\omega)\hat{v}$ for almost every $\omega \in X$. In particular, $A_i\,\hat{v} = \hat{v}$ for all $i \in \mathcal{A}_{\text{inv}}$.

Conversely, if $A_i\,\hat{v} = \hat{v} \in K^u_{\text{inv}}$ for all $i \in \mathcal{A}_{\text{inv}}$, since $\hat{v} \in K^u_{\text{inv}} \subset M$,

$$\hat{v} = \lim_{n\to\infty} A_{\omega_{-1}} \cdots A_{\omega_{-n}}\,\hat{v} = \lim_{n\to\infty} A^n(\sigma^{-n}\omega)\,\hat{v} = E^u(\omega),$$

and $E^u(\omega) = \hat{v}$ for all $\omega \in \mathcal{A}^{\mathbb{Z}}_{\text{inv}}$. Thus by Proposition 6.1 $K^u_{\text{inv}} = \{\hat{v}\}$.

The conclusion for $K^s_{\text{inv}}$ follows under a similar argument. ∎

By Lemma 6.2, $\#K^u_{\text{inv}} > 1$ or $\#K^s_{\text{inv}} > 1$, for otherwise $\underline{A}_{\text{inv}}$ would be diagonalizable. We treat each of these cases separately.

- First assume that $\#K^u_{\text{inv}} > 1$.

If there exists $\hat{r} \in \mathcal{R}(\underline{A})$ such that $\hat{r} \notin K^s_{\text{inv}}$, then for every $\varepsilon > 0$, we can choose $\omega \in \mathcal{A}^{\mathbb{Z}}_{\text{inv}}$ and $n \in \mathbb{N}$ such that $d(A^n(\omega)\,\hat{r}, \hat{k}) < \varepsilon$. This contradicts the fact that $\mathcal{W}^+ \cap \mathcal{W}^- = \emptyset$.

On the other hand, if $\hat{r} \in \mathcal{R}(\underline{A}) \cap K^s_{\text{inv}}$, since $K^u_{\text{inv}}$ has at least two elements one of which is $\hat{k} \in K^u_{\text{inv}}$, by Lemma 6.2 there exists $i \in \mathcal{A}_{\text{inv}}$ such that $A_i^{-1}\hat{k} \neq \hat{k}$. Hence there exists a word $\omega \in \mathcal{A}^{\mathbb{Z}}_{\text{inv}}$ and $n_0 \in \mathbb{N}$ such that $A^{-n_0}(\omega)\,\hat{k} \notin K^u_{\text{inv}}$. Choosing the coordinates of $\omega$ appropriately we can force the convergence of $A^{-n}(\omega)\,\hat{k}$ to $\hat{r} \in K^s_{\text{inv}}$, which also contradicts the fact that $\mathcal{W}^+ \cap \mathcal{W}^- = \emptyset$.

- Now assume $\#K^s_{\text{inv}} > 1$.

Suppose that there exists $\hat{r} \in \mathcal{R}(\underline{A})$ such that $\hat{r} \notin K^s_{\text{inv}}$. Then it is possible to iterate $\hat{r}$ forward by a suitable invertible word in a way that it converges to $\hat{k} \in K^u_{\text{inv}}$. This contradicts $\mathcal{W}^+ \cap \mathcal{W}^- = \emptyset$.

If, on the other hand, there exists $\hat{r} \in \mathcal{R}(\underline{A}) \cap K^s_{\text{inv}}$, then by Lemma 6.2 there is $i \in \mathcal{A}_{\text{inv}}$ such that $A_i\hat{r} \neq \hat{r}$. Hence there exists a word $\omega \in \mathcal{A}^{\mathbb{Z}}_{\text{inv}}$ and $n_0 \in \mathbb{N}$ such that $A^{n_0}(\omega)\,\hat{r} \notin K^s_{\text{inv}}$. Finally, choosing the coordinates of $\omega$ appropriately we can force the convergence of $A^n(\omega)\,\hat{r}$ to $\hat{k} \in K^u_{\text{inv}}$, which also contradicts the fact that $\mathcal{W}^+ \cap \mathcal{W}^- = \emptyset$.

We now proceed to the case that $\underline{A}_{\text{inv}}$ is diagonalizable but $\underline{A}$ is not. Since $\underline{A}_{\text{inv}}$ is projectively uniformly hyperbolic and $\underline{A}_{\text{inv}}$ is diagonalizable, there are exactly two invariant directions, $K^s_{\text{inv}}$ and $K^u_{\text{inv}}$.

**Lemma 6.3** *Suppose $\underline{A}_{\text{inv}}$ is diagonalizable but $\underline{A}$ is not. Then either*

*(i) there exists $\hat{r} \in \mathcal{R}(\underline{A})$ such that $\hat{r} \notin K^s_{\text{inv}} \cup K^u_{\text{inv}}$, or else*

*(ii) there exists $\hat{k} \in \mathcal{K}(\underline{A})$ such that $\hat{k} \notin K^s_{\text{inv}} \cup K^u_{\text{inv}}$.*

*Proof.* Since $\underline{A}_{\text{inv}}$ is diagonalizable, $K^u_{\text{inv}} = \{\hat{e}^u\}$ and $K^s_{\text{inv}} = \{\hat{e}^s\}$ are singletons where $\hat{e}^u$ and $\hat{e}^s$ are respectively the unstable and stable directions of Oseledets. If (i) and (ii) were both false, then for every $i \in \mathcal{A}_{\text{sing}}$, the matrix $A_i$ would preserve both directions $\hat{e}^u$ and $\hat{e}^s$, which would imply that $\underline{A}$ is diagonalizable. ■

Next we analyze the two cases given by Lemma 6.3.

If there exists $\hat{r} \in \mathcal{R}(\underline{A})$ such that $\hat{r} \notin K^s_{\text{inv}}$ then iterating $\hat{r}$ by any invertible word, it converges to $K^u_{\text{inv}} = \{\hat{k}\}$. This contradicts $\mathcal{W}^+ \cap \mathcal{W}^- = \emptyset$. Otherwise, every $\hat{r} \in \mathcal{R}(\underline{A})$ satisfies $K^s_{\text{inv}} = \{\hat{r}\}$ and there exists $\hat{k}' \in \mathcal{K}(\underline{A})$ such that $\hat{k}' \notin K^s_{\text{inv}} \cup K^u_{\text{inv}}$. Hence iterating $\hat{k}'$ backwards by any invertible word, it converges to $\hat{r} = K^s_{\text{inv}}$, which contradicts $\mathcal{W}^+ \cap \mathcal{W}^- = \emptyset$. This concludes the proof. ■

**Remark 6.2** *Note that in the previous proposition, the assumption that $\mathcal{W}^+ \cap \mathcal{W}^- = \emptyset$ can be replaced by $\mathcal{W}^+ \cap \mathcal{K}(\underline{A}) = \emptyset$ and $\mathcal{R}(\underline{A}) \cap \mathcal{W}^- = \emptyset$.*

**Theorem 6.2** *Given a random cocycle $\underline{A} \in \text{Mat}_2^+(\mathbb{R})^k$ of rank 1 such that $\underline{A}_{\text{inv}} := (A_i)_{i \in \mathcal{A}_{\text{inv}}}$ is projectively uniformly hyperbolic and $\underline{A}$ is not diagonalizable, the following are equivalent:*

*(1) $\underline{A}$ is projectively uniformly hyperbolic,*

*(2) $\mathcal{W}^+ \cap \mathcal{W}^- = \emptyset$,*

*(3) $\mathcal{W}^+ \cap \mathcal{K}(\underline{A}) = \emptyset$ and $\mathcal{R}(\underline{A}) \cap \mathcal{W}^- = \emptyset$.*

*Proof.* (1) $\Rightarrow$ (2): Assume by contradiction that $\underline{A}$ is projectively uniformly hyperbolic but $\mathcal{W}^+ \cap \mathcal{W}^- \neq \emptyset$. Then it is possible to produce null words under arbitrarily small perturbations of the cocycle's matrices. This implies the loss of the projective uniform hyperbolicity for the perturbed cocycle. However, this leads to a contradiction, as the projective uniform hyperbolicity is an open property.

The fact that (2) $\Rightarrow$ (3) is trivial, since $\mathcal{K}(\underline{A}) \subset \mathcal{W}^-$ and $\mathcal{R}(\underline{A}) \subset \mathcal{W}^+$.

(3) $\Rightarrow$ (1): We know that $\underline{A}_{\text{inv}}$ is projectively uniformly hyperbolic so by Theorem 6.1, there is an invariant multi-cone $M$ associated to $\underline{A}_{\text{inv}}$. Note that by Proposition 6.3 and Remark 6.2 there are no ranges of singular matrices in $K^s_{\text{inv}}$ and there are no kernels of singular matrices in $K^u_{\text{inv}}$. Also, by Proposition 6.2, we can shrink the multi-cone $M$ so that it does not contain kernels of singular matrices $A_i$, $i \in \mathcal{A}_{\text{sing}}$. Now, because $\underline{A}_{\text{inv}}$ is projectively uniformly hyperbolic, there exists $N \in \mathbb{N}$ such that for every $\omega \in \mathcal{A}^{\mathbb{Z}}_{\text{inv}}$ and every range $\hat{r} \in \mathcal{R}(\underline{A})$, which as we have previously seen is not in $K^s_{\text{inv}}$, we have $A^N(\omega)\,\hat{r} \in M$. Thus, because there are only finitely many ranges in $\mathcal{R}(\underline{A})$ and finitely many words of length $N$, using (3), we can find sufficiently small numbers

$$0 < \epsilon_1 < \cdots < \epsilon_{N-1}$$

independent of the words $\omega$ and of the ranges $\hat{r}$ such that the following inclusions of balls in the projective space hold:

$$A_{\omega_N} B_{\epsilon_{N-1}}(A^{N-1}(\omega)\,\hat{r}) \Subset M$$

and for every $1 \leq j \leq N-1$

$$A_{\omega_j} B_{\epsilon_j}(A^j(\omega)\,\hat{r}) \Subset B_{\epsilon_{j+1}}(A^{j+1}(\omega)\,\hat{r})).$$

The union of $M$ with all these balls, for every word $\omega \in \mathcal{A}^N_{\text{inv}}$ and every range $\hat{r} \in \mathcal{R}(\underline{A})$ is an invariant multi-cone associated to $\underline{A}$. ∎

**Remark 6.3** *If $\underline{A}_{\text{inv}}$ is projectively uniformly hyperbolic and the cocycle $\underline{A}$ is diagonalizable with*

$$K^u_{\text{inv}} = \mathcal{R}(\underline{A}) = \{\hat{r}\} \;\; \textit{and} \;\; K^s_{\text{inv}} = \mathcal{K}(\underline{A}) = \{\hat{k}\},$$

*then the equivalences in Theorem 6.2 still hold. The non diagonalizable hypothesis aims to exclude the case when*

$$K^u_{\text{inv}} = \mathcal{K}(\underline{A}) = \{\hat{k}\} \;\; \textit{and} \;\; K^s_{\text{inv}} = \mathcal{R}(\underline{A}) = \{\hat{r}\}. \tag{6.1}$$

In the view of theorem 6.2 and remark 6.3, we provide an example of a cocycle $\underline{A}$ with rank 1, such that $\underline{A}_{\text{inv}}$ is uniformly hyperbolic, but $\underline{A}$ is diagonalizable and satisfies (6.1). Moreover, we show that $\mathcal{W}^+ \cap \mathcal{W}^- = \emptyset$ but $\underline{A}$ is not uniformly hyperbolic.

**Example 6.1** *Let $\underline{A} = (A_0, B_0)$ be the cocycle taking values*

$$A_0 = \begin{bmatrix} 2 & 0 \\ 0 & 2^{-1} \end{bmatrix} \quad \textit{and} \quad B_0 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

*with probability measure* $\mu = \frac{1}{2}\,\delta_{A_0} + \frac{1}{2}\,\delta_{B_0}$. *Notice that* $L_1(\underline{A}) = \frac{-\log 2}{2}$.

*Note that* $A_0$ *preserves both the* $x$ *and the* $y$ *axes. Moreover,* $\hat{r}(B_0) = (0,1)$ *and* $\hat{k}(B_0) = (1,0)$. *Therefore, both of them are preserved under the invertible part of the cocycle* $\underline{A}$. *Thus* $\mathcal{W}^+ = \hat{r}(B_0) = (0,1)$ *and* $\mathcal{W}^- = \hat{k}(B_0) = (1,0)$, *so that* $\mathcal{W}^+ \cap \mathcal{W}^- = \emptyset$.

*Next we show that* $\underline{A}$ *is not projectively uniformly hyperbolic. Consider the sets*

$$\Sigma_n = \{(A,B)\colon\ A \textit{ is hyperbolic}, \operatorname{rank}(B) = 1,\ A^n\,\hat{r}(B) = \hat{k}(B)\,\},$$

*and the one-parameter family of cocycles* $\underline{A}_t := \left(A_0, B_t := \begin{bmatrix} -t^2 & t \\ -t & 1 \end{bmatrix}\right)$, $t \in \mathbb{R}$, *such that* $\underline{A} = \underline{A}_0$. *The cocycle* $\underline{A} = (A_0, B_0)$ *is an accumulation point of the cocycles* $\underline{A}_{t_n}$, *where* $t_n = 2^{-n}$, *and so it is also an accumulation point of the hypersurfaces* $\Sigma_n$ *as* $\underline{A}_{t_n} \in \Sigma_n$. *Then* $L_1(\underline{A}_{t_n}) = -\infty$, *for all* $n$. *In particular, since projective uniform hyperbolicity is an open property, it follows that* $\underline{A} = \underline{A}_0$ *cannot be projectively uniformly hyperbolic.*

Note that in the previous example, the Lyapunov exponent of $\underline{A}$ is equal to $\frac{-\log 2}{2}$ and we were able to approximate it by cocycles with null words, in particular, with Lyapunov exponent equal to $-\infty$. Therefore, $\underline{A}$ is a discontinuity point of the Lyapunov exponent. This already gives a flavor of what is yet to come in the next section.

## 6.2 Dichotomy: analyticity vs discontinuity

In this section we prove a result in the spirit of Mañé-Bochi's theorem but for non-invertible random, locally constant cocycles. As a corollary, we obtain a dichotomy in the regularity of the Lyapunov exponent, between analyticity and discontinuity.

We start with an example that appears in the introduction of [3] as well as in [20, Section 3]. It was the starting point of this work and it was also responsible for much of the intuition behind the results in chapters 6 and 7.

**Example 6.2** *Consider the family of cocycles* $\underline{A}_t := (A, B_t)$, *where*

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \textit{and} \quad B_t = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix}$$

*are chosen with probabilities* $p = (\frac{1}{2}, \frac{1}{2})$. *By [3, Introduction] or [20, Proposition 3.1]:*

1. $L_1(\underline{A_t}) = \sum_{j=0}^{\infty} \frac{1}{2^{j+1}} \log\left|\cos(jt)\right|.$

2. *If* $t \in \pi\,\mathbb{Q}$ *there exists* $n \in \mathbb{N}$ *such that* $AB_t^n A = 0$ *and so* $L_1(\underline{A_t}) = -\infty$.

3. *The set* $\{t \in \mathbb{R} : L_1(\underline{A_t}) > -\infty\}$ *has full Lebesgue measure.*

4. *If* $t \in \pi\,(\mathbb{R} \setminus \mathbb{Q})$ *then* $\mathcal{W}^+ \cap \mathcal{W}^- = \mathbb{P}(\mathbb{R}^2)$.

*Hence, by* (2) *and* (3) *we conclude that the function* $t \mapsto L_1(\underline{A_t})$ *is discontinuous for almost every* $t \in \mathbb{R}$.

This example shows that for singular cocycles, it is not expected any type of continuity for the Lyapunov exponents. In what follows, we will prove that all cocycles which are not projectively uniformly hyperbolic present the same type of behavior.

### 6.2.1 A Mañé-Bochi type dichotomy

**Lemma 6.4** *Let* $\underline{A}$ *be a random linear cocycle of rank* 1 *such that* $\underline{A}_{inv}$ *is projectively uniformly hyperbolic and* $\mathcal{W}^+ \cap \mathcal{W}^- \neq \emptyset$. *Then either*

$$\mathcal{K}(\underline{A}) \cap K^u_{inv} \neq \emptyset \quad \text{or} \quad \mathcal{R}(\underline{A}) \cap K^s_{inv} \neq \emptyset.$$

*Proof.* The assumption that $\mathcal{W}^+ \cap \mathcal{W}^- \neq \emptyset$ implies that an accumulation point of the forward iterates by invertible words of the ranges is equal to an accumulation point of the backward iterates by invertible words of the kernels.

Since $\underline{A}$ has rank 1, so there are no null words, there cannot be a finite time matching, more precisely, $A^{n_1}(\omega_1)r_i \neq A^{-n_2}(\omega_2)k_j$ for every $\omega_1 \in \mathcal{A}^{n_1}_{\text{inv}}$, $\omega_2 \in \mathcal{A}^{n_2}_{\text{inv}}$, $i, j \in \mathcal{A}_{\text{sing}}$ and $n_1, n_2 \in \mathbb{N}$.

Moreover, since $\underline{A}_{inv}$ is projectively uniformly hyperbolic, the sets $K^u_{inv}$ and $K^s_{inv}$ of respectively unstable and stable Oseledets directions are well defined. If there is a range in $K^s_{inv}$ or a kernel in $K^u_{inv}$, then we conclude the result. If there is no range in $K^s_{inv}$ and no kernel in $K^u_{inv}$, then the forward iterates of the ranges converge by the invertible dynamics to $K^u_{inv}$ and the backwards iterates of the kernels converge to $K^s_{inv}$. Since $\underline{A}_{inv}$ is projectively uniformly hyperbolic, $K^s_{inv} \cap K^u_{inv} = \emptyset$. Hence, in order to satisfy $\mathcal{W}^+ \cap \mathcal{W}^- \neq \emptyset$, it must be that some iterate $A^n(\omega)r_i \in K^s_{inv}$ or $A^{-n}(\omega)k_j \in K^u_{inv}$. Since $K^s_{inv}$ is backward invariant and $K^u_{inv}$ is forward invariant, we conclude the result. ■

**Theorem 6.3** *Let* $\underline{A} \in \text{Mat}_2^+(\mathbb{R})^k$ *be a random linear cocycle of rank* 1. *Then either* $\underline{A}$ *is projectively uniformly hyperbolic or else there exists a sequence* *of random linear cocycles* $\{\underline{A}_n\}_n \to \underline{A}$ *such that* $\underline{A}_n$ *has a null word and, in particular,* $L_1(\underline{A}_n) = -\infty$ *for every* $n \in \mathbb{N}$.

*Proof.* Suppose that $\underline{A}$ is not projectively uniformly hyperbolic. We are going to show that there exist a sequence $\{\underline{A}_n\}_n \to \underline{A}$ such that $\underline{A}_n$ has a null word for every $n \in \mathbb{N}$. We divide the proof into three cases, according to theorem 6.2.

**Case 1**: Suppose that $\underline{A}_{\text{inv}}$ is projectively uniformly hyperbolic and $\underline{A}$ is not diagonalizable. Then by theorem 6.2, there exists a heteroclinic connection: $\mathcal{W}^+ \cap \mathcal{W}^- \neq \emptyset$. By lemma 6.4, either $\mathcal{K}(\underline{A}) \cap K^u_{inv} \neq \emptyset$ or $\mathcal{R}(\underline{A}) \cap K^s_{inv} \neq \emptyset$. Without loss of generality, assume that $\mathcal{K}(\underline{A}) \cap K^u_{inv} \neq \emptyset$, the other case is analogous. Since $\underline{A}_{\text{inv}}$ is projectively uniformly hyperbolic, the iterates $A^n(\omega)r_i$ converge uniformly to $K^u_{inv}$. Therefore, one can consider arbitrarily small perturbations of the kernel that belongs to $K^u_{inv}$ to generate null words in finite time. That is, consider the sequence of cocycles $\underline{A}_n$ whose entries are equal to those of $\underline{A}$ except for the singular matrix whose kernel is in $K^u_{inv}$, for which we perform a progressively smaller perturbation. Then $\underline{A}_n$ converges to $\underline{A}$ and each $\underline{A}_n$ has a null word.

**Case 2**: Suppose that $\underline{A}_{inv}$ is projectively uniformly hyperbolic and $\underline{A}$ is diagonalizable. Since $\underline{A}$ is diagonalizable, either $K^u_{inv} = \mathcal{R}(\underline{A})$ and $K^s_{inv} = \mathcal{K}(\underline{A})$ or $K^u_{inv} = \mathcal{K}(\underline{A})$ and $K^s_{inv} = \mathcal{R}(\underline{A})$.

Assume that $K^u_{\text{inv}} = \mathcal{R}(\underline{A})$ and $K^s_{\text{inv}} = \mathcal{K}(\underline{A})$, then, by Remark 6.3, Theorem 6.2 still holds. Therefore $\mathcal{K}(\underline{A}) \cap \mathcal{W}^+ = \emptyset$ and $\mathcal{R}(\underline{A}) \cap \mathcal{W}^- = \emptyset$, which by this theorem implies that $\underline{A}$ is projectively uniformly hyperbolic, contradicting our assumption. Therefore, this cannot happen.

Hence it suffices to consider the case in which $K^u_{\text{inv}} = \mathcal{K}(\underline{A})$ and $K^s_{\text{inv}} = \mathcal{R}(\underline{A})$. Note that with a small perturbation of the range, which moves it out of $K^s_{\text{inv}}$, its iterates will converge to $K^u_{\text{inv}}$, where lies a kernel. Hence, we return to the first case, where a small perturbation of the kernel can create a null word.

**Case 3**: Suppose that $\underline{A}_{\text{inv}}$ is not projectively uniformly hyperbolic. If $\underline{A}_{\text{inv}} \in \mathrm{GL}_2^+(\mathbb{R})^{\mathrm{k}}$ is not in $\mathrm{SL}_2(\mathbb{R})^k$, we consider the normalized cocycle $\underline{A}^*_{inv}$, which belongs to $\mathrm{SL}_2(\mathbb{R})^k$, and it is obtained by simply dividing each invertible matrix $A_i$ by $\frac{1}{\sqrt{\det(A_i)}}$. Moreover, $\underline{A}^*_{inv}$ has the same projective action as $\underline{A}_{\text{inv}}$.

Then, by proposition 6 of [44], $\underline{A}^*_{inv}$ can be approximated by elliptic cocycles (cocycles that admit elliptic matrices). Moreover, almost every elliptic matrix admits an irrational rotation number, thus we can assume that the approximation is by elliptic cocycles that admit elliptic matrices with irrational rotation number. Therefore, the original cocycle $\underline{A}_{inv}$ is also approximated by elliptic cocycles with elliptic matrix components with irrational rotation number by rescaling the matrices. Then both $\mathcal{W}^+$ and $\mathcal{W}^-$ are equal to $\mathbb{P}(\mathbb{R}^2)$ (see Example 6.2). We can then perform an arbitrarily small perturbation on the kernel (or range), thus creating null words. ■

**Corollary 6.1** *Let $\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k$ be a random linear cocycle of rank 1. If $L_1(\underline{A}) > -\infty$, then either $L_1(\underline{A})$ is analytic around $\underline{A}$ or else it is a discontinuity point of $\underline{B} \mapsto L_1(\underline{B})$.*

*Proof.* By theorem 6.3, either $\underline{A}$ is projectively uniformly hyperbolic or there exists a sequence $\underline{A}_n \to \underline{A}$ such that $L_1(\underline{A}_n) = -\infty$ for every $n$. In the first case, by a theorem of Ruelle [40], $L_1$ is analytic in a neighborhood of $\underline{A}$. In the second case, since we assume $L_1(\underline{A}) > -\infty$ and there is a sequence $\underline{A}_n \to t_0$ such that $L_1(\underline{A}_n) = -\infty$, the Lyapunov exponent is discontinuous at $\underline{A}$. ■

We say that a set is *residual* if it is a countable intersection of open and dense sets. The next corollary is an adaptation of [42, Corollary 9.6] and it shows that the set of continuity points of the Lyapunov exponent is a residual subset of $\mathrm{Mat}_2^+(\mathbb{R})^k$.

Let $\mathcal{PUH}$ denote the set of cocycles $\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k$ such that $\underline{A}$ is projectively uniformly hyperbolic.

**Corollary 6.2** *The set $\{\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k \colon L_1(\underline{A}) = -\infty\}$ is a residual subset of $\mathrm{Mat}_2^+(\mathbb{R})^k \backslash \mathcal{PUH}$.*

*Proof.* Given $a \in \mathbb{R}$ let $\mathcal{L}(a) := \left\{\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k \colon L_1(\underline{A}) < a\right\}$. Since the top Lyapunov exponent is upper semi-continuous, $\mathcal{L}(a)$ is open in $\mathrm{Mat}_2^+(\mathbb{R})^k$. Consider a sequence $a_n \to -\infty$. The set in the statement coincides with $\cap_n \mathcal{L}(a_n)$. By Theorem 6.3 this intersection is dense in $\mathrm{Mat}_2^+(\mathbb{R})^k \backslash \mathcal{PUH}$, hence so is each of the open sets $\mathcal{L}(a_n)$. Thus $\mathcal{L}(a) = \cap_n \mathcal{L}(a_n)$ is a residual subset of $\mathrm{Mat}_2^+(\mathbb{R})^k \setminus \mathcal{PUH}$. ■

Moreover, by the previous corollary, we conclude that the set of continuity points of the Lyapunov exponent is a residual subset of $\mathrm{Mat}_2^+(\mathbb{R})^k$. That is because the set of continuity points is exactly

$$\{\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k \colon L_1(\underline{A}) = -\infty\} \cup \mathcal{PUH}.$$

Since $\mathcal{PUH}$ is open, it is clear that $\cap_n (\mathcal{L}(a_n) \cup \mathcal{PUH})$ is a countable intersection of open and dense sets.

**Remark 6.4** *Since the map $A \mapsto L_1(A)$ is also upper semi-continuous for bounded and continuous cocycles with respect to the $C^0$ norm, it is also true that the set of cocycles $A \in C^0(X, \mathrm{Mat}_2^+(\mathbb{R}))$ which are either projectively uniformly hyperbolic, or else satisfy $L_1(A) = -\infty$, is a residual subset of $C^0(X, \mathrm{Mat}_2^+(\mathbb{R}))$.*

### 6.2.2 Regularity dichotomy for the Lyapunov exponent

**Definition 6.5** *A function $u\colon \Omega \to [-\infty, \infty)$ is called subharmonic in the domain $\Omega \subset \mathbb{C}$ if $u$ is upper semi-continuous and for every $z \in \Omega$,*

$$u(z) \le \int_0^1 u(z + re^{2\pi i\theta})d\theta\,,$$

*for some $r_0(z) > 0$ and for all $r \le r_0(z)$.*

Basic examples of subharmonic functions are $\log|z - z_0|$ or more generally, $\log|f(z)|$ for some analytic function $f(z)$ or $\int \log|z - \zeta|\, d\mu(\zeta)$ for some positive measure with compact support in $\mathbb{C}$.

The maximum of a finite collection of subharmonic functions is subharmonic, while the supremum of a collection (not necessarily finite) of subharmonic functions is subharmonic provided it is upper semi-continuous. In particular, this implies that if $A\colon \Omega \to \mathrm{Mat}_2^+(\mathbb{R})$ is a matrix valued analytic function on some open set $\Omega \subset \mathbb{C}$, then

$$u(t) := \log\|A(t)\| = \sup_{\|v\|, \|w\| \le 1} \log\Big|\langle A(t)\, v, w\rangle\Big|$$

is subharmonic in $\Omega$.

Moreover, the infimum of a decreasing sequence of subharmonic functions is subharmonic. This shows that if $\underline{A}\colon \Omega \to \mathrm{Mat}_2^+(\mathbb{R})^k$ is analytic, then the map $t \mapsto L_1(\underline{A}(t))$ is subharmonic on $\Omega$.

**Lemma 6.5** *Let $\Lambda \subset \mathbb{R}$ be a compact interval, let $\Omega \subset \mathbb{C}$ be an open, complex strip that contains $\Lambda$ and let $u\colon \Omega \to [-\infty, \infty)$ be a subharmonic function such that for some constant $C_0 < \infty$ we have $u(z) \le C_0$ for all $z \in \Omega$ and $u(t_0) \ge -C_0$ for some $t_0 \in \Omega$. Then for all $N \in \mathbb{N}$,*

$$\mathrm{Leb}\,\{t \in \Lambda\colon u(t) < -N\} \le C\, e^{-N^\gamma},$$

*where $\gamma > 0$ and $C$ is a finite constant depending on $C_0$, $\Lambda$ and $\Omega$.*

*In particular,* $\mathrm{Leb}\,\{t \in \Lambda\colon u(t) = -\infty\} = 0$.

*Proof.* The statement is essentially the one-dimensional version of [23, Lemma 3.1]. In [23] the set $\Omega$ is an annulus and $\Lambda$ is the torus, which are transformed into our setting via the complex logarithmic function. The proof of this result is a consequence of a quantitative version of the Riesz representation theorem for subharmonic functions and of Cartan's estimate for logarithmic potentials, see [30, Lemma 2.2 and Lemma 2.4]. ■

**Corollary 6.3** *For Lebesgue almost every cocycle* $\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k$ *with rank* 1 *we have* $L_1(\underline{A}) > -\infty$. *In particular, Lebesgue almost every cocycle* $\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k$ *with rank* 1 *satisfies the regularity dichotomy: its Lyapunov exponent is either analytic or discontinuous.*

*Proof.* For each $\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k$, consider the 1-parameter family $\underline{A}(t)$ such that for every $i \in \mathcal{A}_{\mathrm{inv}}$, $A_i(t) = A_i R_t$, where $R_t$ is a rotation and for $i \in \mathcal{A}_{\mathrm{sing}}$, $A_i(t) = A_i$. Consider $t \in [0, 2\pi]$ and note that this family gives a foliation of $\mathrm{Mat}_2^+(\mathbb{R})^k$ by closed curves.

Let $\Omega$ be a thin and open complex strip around $[0, 2\pi]$. Since the map $t \mapsto \underline{A}_t$ is real analytic, it extends to a holomorphic map $\Omega \ni t \mapsto \underline{A}_t \in \mathrm{Mat}_d(\mathbb{C})^k$. Then $t \mapsto L_1(\underline{A}(t))$ is subharmonic on $\Omega$.

Moreover, since $[0, 2\pi]$ is a compact set and the Lyapunov exponent is an upper semicontinuous function, there exists a global upper bound $C \in \mathbb{R}$ such that $L_1(\underline{A}(t)) \leq C$ for every $t \in \Omega$.

Now, we adapt Yoccoz's argument from [44, Lemma 2] to prove the existence of a parameter $t_0$ such that $\underline{A}(t_0)$ is projectively uniformly hyperbolic, hence $L_1(\underline{A}(t_0)) > -\infty$. Note that the real projective space is the equator of the complex projective space (the Riemann sphere) $\mathbb{CP}^1$ and splits it in two hemi-spheres $\mathbb{H}^-$ and $\mathbb{H}^+$, each of which can be identified as a hyperbolic plane. Given $1 \leq i \leq k$ and $\hat{v} \in \mathbb{P}(\mathbb{R}^2)$, when we complexify the curve $\mathbb{R} \ni t \mapsto \hat{A}_i(t)\,\hat{v} \in \mathbb{P}(\mathbb{R}^2)$, the map $\mathbb{C} \ni t \mapsto \hat{A}_i(t)\,\hat{v} \in \mathbb{CP}^1$, defined on a small strip $\mathcal{S}$ around $I$, takes the interval $I$ to the equator of $\mathbb{CP}^1$ and it maps the two semi-strips $\mathcal{S}^\pm := \{t \in \mathcal{S} \colon \mathrm{Im}(t) \gtrless 0\}$ away from the equator. This behavior is uniform in $\hat{v} \in \mathbb{P}(\mathbb{R}^2)$. Hence, if $t \in \mathcal{S}^+$, resp. $t \in \mathcal{S}^-$, then the projective map $\hat{A}_i(t) \colon \mathbb{CP}^1 \to \mathbb{CP}^1$ contracts $\mathbb{H}^+$, resp. $\mathbb{H}^-$, by a factor of order $\exp(-c\,|\mathrm{Im}(t)|)$. Note that the uniform behavior ensures that $\hat{A}_i(t)(\partial\mathbb{H}^+) = \hat{A}_i(t)(\mathbb{P}(\mathbb{R}^2)) \Subset \mathbb{H}^+$. The rest of the proof follows [44, Lemma 2].

We conclude that there are points $t_0 \in \Omega$ where $\underline{A}(t_0)$ is projectively uniformly hyperbolic, and in particular $L_1(\underline{A}(t_0)) > -\infty$. Since, on the other hand, $L_1(\underline{A}(t))$ is bounded from above on $\Omega$, lemma 6.5 is applicable to this subharmonic function and we conclude that $L_1(\underline{A}(t)) > -\infty$ for Lebesgue almost every parameter $t \in [0, 2\pi]$.

We proved that given any cocycle $\underline{A}$ of rank one, there is a curve (i.e. a smooth, one-parameter family of cocycles) passing through $\underline{A}$ such that the first Lyapunov exponent is finite for almost every cocycle along this curve. By Fubini, we obtain that Lebesgue almost every cocycle $\underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k$ with at least one invertible and one singular component satisfies $L_1(\underline{A}) > -\infty$ and, in particular, also satisfies the regularity dichotomy. ∎

### 6.2.3 Constant rank 1 cocycles

Consider the set of cocycles whose matrix entries have constant rank 1, that is,

$$\mathcal{R}_1 := \left\{ \underline{A} \in \mathrm{Mat}_2^+(\mathbb{R})^k \colon \mathrm{rank}(A_i) = 1 \; \forall \, 1 \leq i \leq k \right\},$$

which is an analytic sub-manifold of $\mathrm{Mat}_2^+(\mathbb{R})^k$ with co-dimension $k$.

**Theorem 6.4** *Given $\underline{A} \in \mathcal{R}_1$, either $d(\mathcal{K}(\underline{A}), \mathcal{R}(\underline{A})) > 0$ and $\underline{A}$ is projectively uniformly hyperbolic or $d(\mathcal{K}(\underline{A}), \mathcal{R}(\underline{A})) = 0$ and $\underline{A}$ admits a null word hence $L_1(\underline{A}) = -\infty$.*

*Proof.* If $d(\mathcal{K}(\underline{A}), \mathcal{R}(\underline{A})) > 0$, there are two disjoint open subsets of $\mathbb{P}(\mathbb{R}^2)$: one forward invariant containing the ranges and another backward invariant containing the kernels. The first is an invariant multi-cone. Therefore, $\underline{A}$ is projectively uniformly hyperbolic. If $d(\mathcal{K}(\underline{A}), \mathcal{R}(\underline{A})) = 0$, there exists a null word and $L_1(\underline{A}) = -\infty$. ∎

**Corollary 6.4** *The Lyapunov exponent $L_1 \colon \mathcal{R}_1 \to [-\infty, +\infty)$ is always continuous. Moreover, it is analytic on $\mathcal{R}_1 \setminus \{\underline{A} \colon L_1(\underline{A}) = -\infty\}$. Furthermore, for Lebesgue almost every $\underline{A} \in \mathcal{R}_1$, $L_1(\underline{A}) > -\infty$, therefore, the analyticity set has full Lebesgue measure.*

*Proof.* By Theorem 6.4, either $\underline{A}$ is projectively uniformly hyperbolic or $L_1(\underline{A}) = -\infty$. By Ruelle [40, Theorem 3.1], if $\underline{A}$ is projectively uniformly hyperbolic then the Lyapunov exponent is an analytic function. Moreover, the first Lyapunov exponent is continuous at points with $L_1(\underline{A}) = -\infty$ (since it is an upper semi-continuous function). The set

$$\begin{aligned} \{\underline{A} \in \mathcal{R}_1 \colon L_1(\underline{A}) = -\infty\} &= \{\underline{A} \in \mathcal{R}_1 \colon \mathcal{K}(\underline{A}) \cap \mathcal{R}(\underline{A}) \neq \emptyset\} \\ &= \bigcup_{i,j=1}^{k} \{\underline{A} \in \mathcal{R}_1 \colon k(A_i) = r(A_j)\} \in \mathcal{R}_1 \end{aligned}$$

is an algebraic sub-variety of $\mathcal{R}_1$ with positive co-dimension in $\mathcal{R}_1$. Hence it has zero Lebesgue measure in $\mathcal{R}_1$. ∎

# 7
# Statistical Properties

The main goal of this chapter is to prove statistical properties (large deviations estimates and a central limit theorem) for the Lyapunov exponent of random linear cocycles in the singular setting. To this end, we study the corresponding Markov operator and stationary measure which, surprisingly, behave quite differently from their analogues in the classical invertible setting.

This different behavior happens due to the fact that whenever a singular matrix $A_s$ appears at the end of any random product of matrices, every projective point $\hat{v}$ is mapped by the corresponding projective map to the range $\hat{r}_s$ of $\hat{A}_s$. In particular, this gives rise to a phenomena called a renewal process.

More precisely, if we regard the action of a random matrix product as a random walk on the projective space, we have that starting at a range, the random walk follows the action of the invertible matrices up until a singular matrix appears; then the walk returns to the range of a singular matrix and the process restarts.

This type of phenomena is a peculiarity of the singular setting and it produces many interesting consequences. This observation lies at the heart of all of the explicit formulas presented in this chapter.

This chapter is based on the joint work [19] with Duarte, Graxinha and Klein. In section 7.1 we establish an explicit formula for the stationary measure and investigate the particularities of the Markov operator in the singular setting. Specifically, we prove that it is strongly mixing on its largest domain, that of bounded, measurable observables (rather than on a properly chosen space of Hölder-type observables as in the invertible setting). Moreover, using the explicit formula of the (unique) stationary measure, we establish an analogue of Furstenberg's formula, the classical one not being available here. In section 7.2, using a parameter elimination argument, we establish large deviations type estimates and a central limit theorem for Lebesgue almost every such cocycle.

## 7.1 Furstenberg's formula in the singular setting

Given the alphabet $\mathcal{A} = \{1, \ldots, k\}$ and a partition $\mathcal{A} = \mathcal{A}_{\text{sing}} \sqcup \mathcal{A}_{\text{inv}}$ into two nonempty sets, consider the space $\mathcal{M}$ of all $k$-tuples $\underline{A} = (A_1, \ldots, A_k) \in \text{Mat}_2^+(\mathbb{R})^k$ such that $\text{rank} A_i = 1$ if $i \in \mathcal{A}_{\text{sing}}$ and $\det A_j > 0$ if $j \in \mathcal{A}_{\text{inv}}$. Moreover, let $\mathcal{M}^* \subset \mathcal{M}$ denote the set of such cocycles, which additionally satisfy $\hat{r}_i \neq \hat{k}_j$ for all $i, j \in \mathcal{A}_{\text{sing}}$. Recall that we identify such a $k$-tuple $\underline{A}$ with the locally constant map $A\colon X \to \text{Mat}_2^+(\mathbb{R})$, $A(\omega) = A_{\omega_0}$, where $\omega = \{\omega_n\}_{n\in\mathbb{Z}} \in X = \mathcal{A}^{\mathbb{Z}}$.

Given a probability vector $p = (p_1, \ldots, p_k)$ with $p_i > 0$ for all $i$, the $k$-tuple $\underline{A} = (A_1, \ldots, A_k)$ determines the random linear cocycle $F\colon X \times \mathbb{R}^2 \to X \times \mathbb{R}^2$, $F(\omega, v) = (\sigma\omega, A(\omega)v) = (\sigma\omega, A_{\omega_0}\, v)$ where $X$ is endowed with the product probability $\mu = p^{\mathbb{Z}}$. As before, we also identify the cocycle $F$ with the tuple $\underline{A} = (A_1, \ldots, A_k)$.

Let $L^\infty(\mathbb{P}(\mathbb{R}^2))$ be the Banach space of bounded and measurable functions $\varphi\colon \mathbb{P}(\mathbb{R}^2) \to \mathbb{R}$ endowed with the usual sup norm, denoted by $\|\cdot\|_\infty$. Recall that the projective action of a nonzero matrix $A \in \text{Mat}_2(\mathbb{R})$ is given by $\hat{A}\hat{v} = \widehat{Av}$ if $A$ is invertible and by $\hat{A}\hat{v} = \hat{r}$, where $r = \text{Range}(A)$ otherwise.

For the rank 1 random cocycle $\underline{A} = (A_i)_{i\in\mathcal{A}} \in \text{Mat}_2(\mathbb{R})^k$ we will use the notations $r_i$ and $k_i$ to represent, respectively, the range and the kernel of $A_i$, as well as, when convenient, a unit vector belonging to these one-dimensional subspaces.

### 7.1.1 Markov operators and stationary measures in the singular setting

The random linear cocycle $(\underline{A}, p)$ determines the Markov operator $\mathcal{Q}\colon L^\infty(\mathbb{P}(\mathbb{R}^2)) \to L^\infty(\mathbb{P}(\mathbb{R}^2))$ defined by

$$\begin{aligned}(\mathcal{Q}\varphi)(\hat{v}) &:= \sum_{i\in\mathcal{A}} \varphi(\hat{A}_i\, \hat{v})\, p_i \\ &= \sum_{i\in\mathcal{A}_{\text{inv}}} \varphi(\hat{A}_i\, \hat{v})\, p_i + \sum_{i\in\mathcal{A}_{\text{sing}}} \varphi(\hat{r}_i)\, p_i.\end{aligned}$$

Moreover, we write $\mathcal{Q} = \mathcal{Q}_{\text{inv}} + \mathcal{Q}_{\text{sing}}$, where the operators $\mathcal{Q}_{\text{inv}}$ and $\mathcal{Q}_{\text{sing}}$ are given respectively by the two terms above. Note that $\mathcal{Q}_{sing}$ is a projection to a constant function.

The operator $\mathcal{Q}$ is clearly linear, positive and it takes the constant function $\mathbb{1}$ to itself. Using the Riesz-Markov-Kakutani representation theorem, one can deduce the existence of a corresponding transition probability kernel; in other words, $\mathcal{Q}$ is a *Markov operator*.

Recall that a measure $\eta \in \mathrm{Prob}(\mathbb{P}(\mathbb{R}^2))$ is $\mathcal{Q}$-stationary if for all observables $\varphi \in L^\infty(\mathbb{P}(\mathbb{R}^2))$,

$$\int \mathcal{Q}\varphi \, d\eta = \int \varphi \, d\eta.$$

In this case we also call $\eta$ stationary relative to the cocycle $(\underline{A}, P)$.

**Lemma 7.1** *For every $n \in \mathbb{N}$,*

$$\mathcal{Q}^n = \sum_{i=0}^{n-1} \mathcal{Q}_{sing} \circ \mathcal{Q}_{inv}^i + \mathcal{Q}_{inv}^n.$$

*Proof.* We proceed by induction. For $n = 1$ the formula holds trivially, since $\mathcal{Q} = \mathcal{Q}_{sing} + \mathcal{Q}_{inv}$. Suppose that it also holds for $n = k$. Then for every $\varphi \in L^\infty(\mathbb{P}(\mathbb{R}^2))$ and every $\hat{v} \in \mathbb{P}(\mathbb{R}^2)$,

$$\begin{aligned} Q^{k+1}(\varphi)(\hat{v}) &= Q \circ \left[\sum_{i=0}^{k-1} \mathcal{Q}_{sing} \circ \mathcal{Q}_{inv}^i + \mathcal{Q}_{inv}^k\right](\varphi)(\hat{v}) \\ &= Q \circ \left(\sum_{i=0}^{k-1} \mathcal{Q}_{sing} \circ \mathcal{Q}_{inv}^i(\varphi)(\hat{v})\right) + \mathcal{Q} \circ \mathcal{Q}_{inv}^k(\varphi)(\hat{v}). \end{aligned}$$

Note that the first term is a constant function, hence it is preserved by the Markov operator $\mathcal{Q}$. Then

$$\begin{aligned} Q^{k+1}(\varphi)(\hat{v}) &= \sum_{i=0}^{k-1} \mathcal{Q}_{sing} \circ \mathcal{Q}_{inv}^i(\varphi)(\hat{v}) + \mathcal{Q} \circ \mathcal{Q}_{inv}^k(\varphi)(\hat{v}) \\ &= \sum_{i=0}^{k} \mathcal{Q}_{sing} \circ \mathcal{Q}_{inv}^i(\varphi)(\hat{v}) + \mathcal{Q}_{inv}^{k+1}(\varphi)(\hat{v}) \end{aligned}$$

which proves the identity for $k+1$. ■

Let us introduce some notations. Given $n \geq 1$, consider the word $\underline{\omega} = (\omega_1, \ldots, \omega_n) \in \mathcal{A}_{\mathrm{inv}}^n$ of length $n$. For such a *finite* word we denote $A^n(\underline{\omega}) = A_{\omega_n} \cdots A_{\omega_1}$, with the convention that $A^0(\underline{\omega})$ is the identity matrix. We denote the quantity $p(\underline{\omega}) := p_{\omega_1} \cdots p_{\omega_n}$. Moreover, in order to (visually) distinguish the weight of a singular symbol $i \in \mathcal{A}_{\mathrm{sing}}$ from that of an invertible one, we will re-denote it by $q_i$ (instead of $p_i$, which will be reserved for the weights of invertible symbols). Moreover, we define the sum of the weights of the singular matrices by $q := \sum_{i \in \mathcal{A}_{\mathrm{sing}}} q_i < 1$.

With these notations, we have the following explicit formula for a (or, a-posteriori, *the*) $\mathcal{Q}$-stationary measure.

**Proposition 7.1** *Let $\eta \in \mathrm{Prob}(\mathbb{P}(\mathbb{R}^2))$ be given by*

$$\eta = \sum_{s\in\mathcal{A}_{sing}} q_s \sum_{n=0}^{\infty} \sum_{\substack{|\underline{\omega}|=n \\ \underline{\omega}\in\mathcal{A}_{inv}^n}} p(\underline{\omega})\delta_{A^n(\underline{\omega})r_s}.$$

*Then $\eta$ is stationary with respect to the operator $\mathcal{Q}$.*

*Proof.* Note that $\eta$ is a probability measure. Indeed,

$$\sum_{s\in\mathcal{A}_{\mathrm{sing}}} q_s \sum_{n=0}^{\infty} \sum_{\substack{|\underline{\omega}|=n \\ \underline{\omega}\in\mathcal{A}_{inv}^n}} p(\underline{\omega}) = \sum_{s\in\mathcal{A}_{\mathrm{sing}}} q_s \sum_{n=0}^{\infty}(1-q)^n = q\,\frac{1}{q} = 1. \tag{7.1}$$

Let $\varphi \in L^\infty(\mathbb{P}(\mathbb{R}^2))$. A straightforward computation shows that

$$\int \varphi \, d\eta = \sum_{s\in\mathcal{A}_{\mathrm{sing}}} q_s \sum_{n=0}^{\infty} \sum_{\substack{|\underline{\omega}|=n \\ \underline{\omega}\in\mathcal{A}_{inv}^n}} p(\underline{\omega})\varphi(A^n(\underline{\omega})r_s).$$

Now show that the integral of $\mathcal{Q}\varphi$ with respect to $\eta$ has the same result.

$$\int \mathcal{Q}\varphi \, d\eta = \sum_{s\in\mathcal{A}_{\mathrm{sing}}} q_s \sum_{n=0}^{\infty} \sum_{\substack{|\underline{\omega}|=n \\ \underline{\omega}\in\mathcal{A}_{inv}^n}} p(\underline{\omega}) \sum_{j\in\mathcal{A}} p_j\varphi(A_jA^n(\underline{\omega})r_s)$$

$$= \sum_{s\in\mathcal{A}_{\mathrm{sing}}} q_s \sum_{n=0}^{\infty} \sum_{\substack{|\underline{\omega}|=n \\ \underline{\omega}\in\mathcal{A}_{inv}^n}} p(\underline{\omega}) \left( \sum_{j\in\mathcal{A}_{\mathrm{inv}}} p_j\varphi(A_jA^n(\underline{\omega})r_s) + \sum_{k\in\mathcal{A}_{\mathrm{sing}}} q_k\varphi(r_k) \right)$$

$$= \sum_{s\in\mathcal{A}_{\mathrm{sing}}} q_s \left( \sum_{n=0}^{\infty} \sum_{\substack{|\underline{\omega}|=n+1 \\ \underline{\omega}\in\mathcal{A}_{inv}^{n+1}}} p(\underline{\omega})\varphi(A^{n+1}(\underline{\omega})r_s) + \sum_{n=0}^{\infty} \sum_{|\underline{\omega}|=n} p(\underline{\omega}) \sum_{k\in\mathcal{A}_{\mathrm{sing}}} q_k\varphi(r_k) \right)$$

$$= \sum_{s\in\mathcal{A}_{\mathrm{sing}}} q_s \sum_{n=1}^{\infty} \sum_{\substack{|\underline{\omega}|=n \\ \underline{\omega}\in\mathcal{A}_{inv}^n}} p(\underline{\omega})\varphi(A^n(\underline{\omega})r_s) + q\sum_{n=0}^{\infty}(1-q)^n \sum_{k\in\mathcal{A}_{\mathrm{sing}}} q_k\varphi(r_k)$$

$$= \sum_{s\in\mathcal{A}_{\mathrm{sing}}} q_s \sum_{n=0}^{\infty} \sum_{\substack{|\underline{\omega}|=n \\ \underline{\omega}\in\mathcal{A}_{inv}^n}} p(\underline{\omega})\varphi(A^n(\underline{\omega})r_s)$$

which equals $\int \varphi \, d\eta$ and completes the proof. ∎

**Theorem 7.1** *If the cocycle $(\underline{A}, p)$ has at least one singular and one invertible component, then $\mathcal{Q}$ is strongly mixing on $L^\infty(\mathbb{P}(\mathbb{R}^2))$ endowed with the uniform norm. That is, there exist constants $C < \infty$ and $a > 0$ such that*

$$\|\mathcal{Q}^n\varphi - \int \varphi \, d\eta\|_\infty \leq C\, e^{-a\,n}\, \|\varphi\|_\infty$$

*for all $n \in \mathbb{N}$ and $\varphi \in L^\infty(\mathbb{P}(\mathbb{R}^2))$, where $\eta$ is any $\mathcal{Q}$-stationary measure.*

*Proof.* Let $\varphi \in L^\infty(\mathbb{P}(\mathbb{R}^2))$. By lemma 7.1, for every $n \in \mathbb{N}$, it follows that

$$\begin{aligned}
\mathcal{Q}^n\varphi - \int \varphi\, d\eta &= \sum_{i=0}^{n-1} \mathcal{Q}_{sing} \circ \mathcal{Q}^i_{inv}(\varphi)(v) + \mathcal{Q}^n_{inv}(\varphi)(v) - \int \varphi\, d\eta \\
&= \sum_{i=0}^{n-1} \mathcal{Q}_{sing} \circ \mathcal{Q}^i_{inv}(\varphi)(v) + \mathcal{Q}^n_{inv}(\varphi)(v) - \sum_{i\in\mathcal{A}_{sing}} q_i \sum_{n=0}^{\infty} \sum_{\substack{|\underline{\omega}|=n \\ \underline{\omega}\in\mathcal{A}^n_{inv}}} p(\underline{\omega})\varphi(A^n(\underline{\omega})r_i) \\
&= \sum_{\substack{|\underline{\omega}|=n \\ \underline{\omega}\in\mathcal{A}^n_{inv}}} p(\underline{\omega})\varphi(A^n(\underline{\omega})v) - \sum_{i\in\mathcal{A}_{sing}} q_i \sum_{j=n}^{\infty} \sum_{\substack{|\underline{\omega}|=j \\ \underline{\omega}\in\mathcal{A}^j_{inv}}} p(\underline{\omega})\varphi(A^j(\underline{\omega})r_i).
\end{aligned}$$

Then

$$\begin{aligned}
\|\mathcal{Q}^n\varphi - \int \varphi\, d\eta\|_\infty &\le \sum_{\substack{|\underline{\omega}|=n \\ \underline{\omega}\in\mathcal{A}^n_{inv}}} p(\underline{\omega})\|\varphi\|_\infty + \sum_{i\in\mathcal{A}_{sing}} q_i \sum_{j=n}^{\infty} \sum_{\substack{|\underline{\omega}|=j \\ \underline{\omega}\in\mathcal{A}^j_{inv}}} p(\underline{\omega})\|\varphi\|_\infty \\
&\le \|\varphi\|_\infty \left[(1-q)^n + q\sum_{j=n}^{\infty}(1-q)^j\right] \\
&= \|\varphi\|_\infty \left[(1-q)^n + q\left(\frac{1}{q} - \frac{1-(1-q)^{n-1}}{q}\right)\right] \\
&\le 2\|\varphi\|_\infty (1-q)^n,
\end{aligned}$$

which proves the statement with $C = 2$ and $\sigma = 1 - q$. ■

**Remark 7.1** *When a Markov operator is strongly mixing on the whole space of measurable, bounded functions relative to the* $L^\infty$ *norm (as it was shown to be the case with* $\mathcal{Q}$*), it is also called uniformly ergodic. Note that uniform ergodicity is the strongest form of strong mixing.*

**Corollary 7.1** *If the cocycle* $(\underline{A}, p)$ *has at least one singular and one invertible matrix component, then it admits a unique stationary measure.*

*Proof.* If $\eta_1$ and $\eta_2$ are $\mathcal{Q}$-stationary, then Theorem 7.1 applies to each of them, so for any $\varphi \in L^\infty(\mathbb{P}(\mathbb{R}^2))$, $\mathcal{Q}^n\varphi$ converges uniformly to $\int \varphi\, d\eta_1$ and to $\int \varphi\, d\eta_2$. Thus $\int \varphi\, d\eta_1 = \int \varphi\, d\eta_2$ for all observables $\varphi$, showing that $\eta_1 = \eta_2$. ■

### 7.1.2
### Description of the Lyapunov exponent via induced cocycles

**Lemma 7.2** *Given rank one matrices* $B_1, B_2, \ldots, B_n$ *and a unit vector* $r_0$*, it follows that*

$$\|B_n \cdots B_1\, r_0\| = \prod_{l=1}^{n} \|B_l\, r_{l-1}\|\,,$$

*where* $r_{l-1}$ *is a unit vector in the range of* $B_{l-1}$*.*

*Proof.* We write $B_1 r_0 = \lambda_1 r_1$, thus $|\lambda_1| = \|B_1 r_0\|$. It follows that $B_2 B_1 r_0 = \lambda_1 B_2 r_1$ so

$$\|B_2 B_1 r_0\| = |\lambda_1| \, \|B_2 r_1\| = \|B_2 r_1\| \|B_1 r_0\| \, .$$

From here,

$$\|B_{n+1} B_n \Big(B_{n-1} \cdots B_1 \, r_0\Big)\| = \|B_{n+1} r_n\| \|B_n \Big(B_{n-1} \cdots B_1 \, r_0\Big)\|$$

and the conclusion follows by induction. ■

Next we establish a preliminary formula for the first Lyapunov exponent for singular cocycles. It will be then used to derive a Furstenberg-type formula.

**Lemma 7.3** *Let $(\underline{A}, p)$ be a random cocycle with both singular and invertible components. Then*

$$L_1(F) = \sum_{i \in \mathcal{A}_{sing}} \sum_{j \in \mathcal{A}_{sing}} \sum_{n=0}^{\infty} \sum_{\substack{|\underline{\omega}|=n \\ \underline{\omega} \in \mathcal{A}_{inv}^n}} q_i q_j p(\underline{\omega}) \log \|A_j A^n(\underline{\omega}) r_i\|.$$

*Proof.* Consider the cylinders $C_i := [0; i]$ with $i \in \mathcal{A}_{\text{sing}}$ and their union $C := \bigcup_{i \in \mathcal{A}_{\text{sing}}} C_i$, the set of all (bi-infinite) words with a singular symbol at the zeroth position. Remember that $q = \mu(C) = \sum_{i \in \mathcal{A}_{\text{sing}}} q_i$.

Moreover, let $C_{i,j} := \bigcup_{n=0}^{\infty} \bigcup_{\substack{|\underline{\omega}|=n \\ \underline{\omega} \in \mathcal{A}_{inv}^n}} [0; i\omega j]$, with the convention that for $n = 0$, $\underline{\omega}$ is the null word. These sets give rise to the following partitions modulo a zero measure set (mod 0)

$$C_i = \bigcup_{j \in \mathcal{A}_{\text{sing}}} C_{i,j} = \bigcup_{j \in \mathcal{A}_{\text{sing}}} \bigcup_{n=0}^{\infty} \bigcup_{\substack{|\underline{\omega}|=n \\ \underline{\omega} \in \mathcal{A}_{inv}^n}} [0; i\omega j] \quad \text{and}$$

$$C = \bigcup_{j \in \mathcal{A}_{\text{sing}}} \bigcup_{i \in \mathcal{A}_{\text{sing}}} C_{i,j} = \bigcup_{i \in \mathcal{A}_{\text{sing}}} \bigcup_{j \in \mathcal{A}_{\text{sing}}} \bigcup_{n=0}^{\infty} \bigcup_{\substack{|\underline{\omega}|=n \\ \underline{\omega} \in \mathcal{A}_{inv}^n}} [0; i\omega j] \, .$$

Let $g \colon C \to C$ be the first return map to the cylinder $C$, given by $g(\omega) = \sigma^{\tau(\omega)}(\omega)$, where $\tau(\omega) = \min\{k \geq 1 \colon \sigma^k(\omega) \in C\}$. The map $g$ preserves the induced measure $\bar{\mu}_C = \mu(C)^{-1} \, \mu|_C = \frac{1}{q} \, \mu|_C$.

We define the induced cocycle $F_C \colon C \times \mathbb{R}^2 \to C \times \mathbb{R}^2$, given by $F_C(\omega, v) := (g(\omega), \mathcal{C}(\omega) \, v)$, where $\mathcal{C}(\omega) := A^{\tau(\omega)}(\omega)$.

By [42, Proposition 4.18 and Exercise 4.8] (these statements also hold for $\mathrm{Mat}_2^+(\mathbb{R})$-valued fiber maps) its Lyapunov exponent is related to that of the original cocycle $(\underline{A}, p)$ via the expression $L_1(F_C) = \frac{1}{q} L_1(\underline{A})$; thus it is enough to compute $L_1(F_C)$.

The induced cocycle $F_C$ leaves invariant the 1-dimensional sub-bundle $X \ni \omega \mapsto R(\omega) := \text{Range}(A_{\omega_0})$. Then using Oseledets' theorem, its first Lyapunov exponent is the rate of exponential growth of the fiber iterates of $F_C$ along this sub-bundle. Thus for $\mu$-a.e. $\omega \in C$ and for a unit vector $r_0 \in R(\omega) = \text{Range}(A_{\omega_0})$,

$$L_1(F_C) = \lim_{n\to\infty} \frac{1}{n} \log\|\mathcal{C}^n(\omega)\, r_0\| = \lim_{n\to\infty} \frac{1}{n} \log \prod_{l=1}^{n} \|\mathcal{C}(g^l\omega)\, r_0\| \,. \tag{7.2}$$

Given $\omega \in C$, let $0 = k_1 < k_2 < \cdots$ denote all future entries to the singular part of the alphabet, that is, $k \in \mathbb{N}$ is such that $\omega_k \in \mathcal{A}_{\text{sing}}$ if and only if $k = k_l$ for some $l \in \mathbb{N}$. Then clearly $g^l(\omega) = \sigma^{k_l}(\omega)$ (whose zeroth entry is $\omega_{k_l}$) and $\tau(g^l\omega) = k_{l+1} - k_l$ for all $l \in \mathbb{N}$. Moreover,

$$B_l := \mathcal{C}(g^l\omega) = A^{\tau(g^l\omega)}(g^l\omega) = A^{k_{l+1}-k_l}(\sigma^{k_l}\omega) = A_{\omega_{k_{l+1}}} \cdots A_{\omega_{k_l+1}} \,,$$

which is a rank one matrix whose range $r_l := \text{Range}(B_l) = r_{\omega_{k_{l+1}}}$.

By Lemma 7.2,

$$\begin{aligned}\|\mathcal{C}^n(\omega)\, r_0\| &= \|\prod_{l=1}^{n} \mathcal{C}(g^l\omega)\, r_0\| = \|\prod_{l=1}^{n} B_l\, r_0\| = \prod_{l=1}^{n} \|B_l\, r_{l-1}\| \\ &= \prod_{l=1}^{n} \|\mathcal{C}(g^l\omega)\, r_{\omega_{k_l}}\| \,.\end{aligned} \tag{7.3}$$

Consider the observable $\varphi\colon C \to \mathbb{R}$,

$$\varphi(\omega) := \log\|\mathcal{C}(\omega)\, r_{\omega_0}\| = \log\|A^{\tau(\omega)}(\omega)\, r_{\omega_0}\|.$$

By (7.2), (7.3) and Birkhoff's ergodic theorem,

$$\begin{aligned}L_1(F_C) &= \lim_{n\to\infty} \frac{1}{n} \sum_{l=1}^{n} \log\|\mathcal{C}(g^l\omega)\, r_{\omega_{k_l}}\| = \lim_{n\to\infty} \frac{1}{n} \sum_{l=1}^{n} \varphi(g^l\omega) \\ &= \int_C \varphi(\omega)\, d\bar{\mu}|_C(\omega) = \frac{1}{q} \int_C \log\|\mathcal{C}(\omega)\, r_{\omega_0}\|\, d\mu(\omega) \,.\end{aligned}$$

Note that on each given cylinder $[0; i\omega j]$ with $|\omega| = n$, $\omega \in \mathcal{A}^n_{\text{inv}}$, $i, j \in \mathcal{A}_{\text{sing}}$ and $n \geq 0$, the first return map $\tau$ is constant and equal to $n$, while the observable $\varphi$ is equal to $\log\|A_j A^n(\omega)\, r_i\|$, thus it is constant. Moreover, the Bernoulli measure of this cylinder is $\mu[0; i\omega j] = q_i q_j\, p(\omega)$. Since these cylinders partition (mod 0) the set $C$, we conclude that

$$\int_C \varphi(\omega)\, d\mu(\omega) = \sum_{i,j\in\mathcal{A}_{\text{sing}}} \sum_{n=0}^{\infty} \sum_{\substack{|\omega|=n \\ \omega\in\mathcal{A}^n_{inv}}} q_i q_j\, p(\omega)\, \log\|A_j A^n(\omega)\, r_i\| \,,$$

which completes the proof of the lemma. ∎

Consider the following observable $\Psi\colon \mathbb{P}(\mathbb{R}^2) \to [-\infty, \infty)$,

$$\Psi(v) = \sum_{i\in\mathcal{A}} p_i \log \frac{\|A_i v\|}{\|v\|}. \tag{7.4}$$

**Theorem 7.2 (Furstenberg's Formula)** *If the cocycle $(\underline{A}, p)$ has both singular and invertible components, then*

$$L_1(\underline{A}) = \int \Psi \, d\eta.$$

*Proof.* Using the explicit formula of the stationary measure $\eta$ derived in Proposition 7.1,

$$\int \Psi \, d\eta = \sum_{i\in\mathcal{A}_{inv}} p_i \sum_{j\in\mathcal{A}_{sing}} q_j \sum_{n=0}^{\infty} \sum_{\substack{|\omega|=n\\ \omega\in\mathcal{A}_{inv}^n}} p(\omega) \log \frac{\|A_i A^n(\omega) r_j\|}{\|A^n(\omega) r_j\|} +$$

$$+ \sum_{i\in\mathcal{A}_{sing}} q_i \sum_{j\in\mathcal{A}_{sing}} q_j \sum_{n=0}^{\infty} \sum_{\substack{|\omega|=n\\ \omega\in\mathcal{A}_{inv}^n}} p(\omega) \log \frac{\|A_i A^n(\omega) r_j\|}{\|A^n(\omega) r_j\|}.$$

Note that the first term is equal to

$$\sum_{i\in\mathcal{A}_{inv}} p_i \sum_{j\in\mathcal{A}_{sing}} q_j \sum_{n=0}^{\infty} \sum_{\substack{|\omega|=n\\ \omega\in\mathcal{A}_{inv}^n}} p(\omega) \log\|A_i A^n(\omega) r_j\| -$$

$$\sum_{i\in\mathcal{A}_{inv}} p_i \sum_{j\in\mathcal{A}_{sing}} q_j \sum_{n=0}^{\infty} \sum_{\substack{|\omega|=n\\ \omega\in\mathcal{A}_{inv}^n}} p(\omega) \log\|A^n(\omega) r_j\|.$$

Since $\sum_{i\in\mathcal{A}_{inv}} p_i = 1 - q$ and since $i \in \mathcal{A}_{\text{inv}}$, we note that the previous difference is a telescopic sum. Moreover, when $n = 0$, $A^0(\omega)$ is the identity matrix and since $\|r_i\| = 1$ for every $i$, we conclude that the first term of the integral is equal to

$$q \sum_{j\in\mathcal{A}_{sing}} q_j \sum_{n=0}^{\infty} \sum_{\substack{|\omega|=n\\ \omega\in\mathcal{A}_{inv}^n}} p(\omega) \log\|A^n(\omega) r_j\|.$$

Now expand the logarithm of the second term into a difference. Since $\sum_{i\in\mathcal{A}_{sing}} q_i = q$, the previous computation corresponds exactly to the term arising from the denominator. Hence, we conclude that

$$\int \Psi \, d\eta = \sum_{i\in\mathcal{A}_{sing}} \sum_{j\in\mathcal{A}_{sing}} \sum_{n=0}^{\infty} \sum_{\substack{|\omega|=n\\ \omega\in\mathcal{A}_{inv}^n}} q_i q_j p(\omega) \log\|A_j A^n(\omega) r_i\|$$

thus completing the proof. ∎

**Remark 7.2** *We note that the classical Furstenberg's formula (theorem 3.4) for random cocycles in* $\mathrm{GL_d}(\mathbb{R})$ *is not applicable to this singular setting. Moreover, the probabilistic approach used by Furstenberg and Kifer to establish such results (see e.g. [29, Theorem 1.4]) is not immediately applicable either, since the observable* $\Psi$ *is not continuous, not even bounded. That is why we had to employ an ad-hoc argument which uses the explicit formula of the stationary measure.*

## 7.2 Statistical properties

The goal of this section is to establish a large deviations type (LDT) estimate and a central limit theorem (CLT) for Lebesgue almost every cocycle $\underline{A} \in \mathcal{M}$. This will be obtained as a consequence of a stronger result, which essentially says that if $t \mapsto \underline{A}_t \in \mathcal{M}^*$ is a one-parameter family of such cocycles satisfying a positive winding condition, then an LDT estimate and a CLT hold for Lebesgue almost every parameter $t$. Recall that, by theorems 2.2 and 2.3, such limit laws hold as soon as the associated Markov operator satisfies a strong mixing condition on an appropriate space of observables. The Markov operator is, as we have seen, strongly mixing on $L^\infty(\mathbb{P}(\mathbb{R}^2))$. The problem is that the relevant observable $\Psi$ defined by (7.4) is *not* bounded, so the abstract LDT theorem 2.2 is not immediately applicable. Instead, we will truncate the observable at a level depending on the scale (number of iterates of the dynamics) and apply the abstract LDT to the truncated observable. The case of the CLT is slightly different, as the abstract theorem 2.3 requires a certain type of mixing on average. We will use a parameter elimination argument to show that for Lebesgue a.e. parameter $t$, these challenges in applying the abstract results can be overcome.

In order to make use of theorems 2.2 and 2.3 we will actually need to associate to a given cocycle $\underline{A} \in \mathcal{M}^*$ a Markov operator on a slightly larger space, $\bar{\mathcal{Q}} \colon L^\infty(\mathcal{A} \times \mathbb{P}(\mathbb{R}^2)) \to L^\infty(\mathcal{A} \times \mathbb{P}(\mathbb{R}^2))$ defined by

$$\bar{\mathcal{Q}}\varphi(j, \hat{v}) = \sum_{i=1}^{k} p_i \varphi(i, \hat{A}_j \hat{v}) \,.$$

Consider the projection $\pi \colon L^\infty(\mathcal{A} \times \mathbb{P}(\mathbb{R}^2)) \to L^\infty(\mathcal{A})$ given by

$$\pi\varphi(\hat{v}) = \sum_{i=1}^{k} p_i \varphi(i, \hat{v}) = \int_{\mathcal{A}} \varphi(i, \hat{v})\, dp(i) \,.$$

Then the Markov operators $\bar{\mathcal{Q}}$ and $\mathcal{Q}$ are related by lemma 2.2. Moreover, since $\eta$ is $\mathcal{Q}$-stationary and $\mathcal{Q}$ is uniformly ergodic, we conclude by proposition 2.4, that the measure $p \times \eta$ is $\bar{\mathcal{Q}}$-stationary and $\bar{Q}$ is also uniformly ergodic.

Let $\Lambda \subset \mathbb{R}$ be a compact interval (normalized to have length 1) and let $A\colon \Lambda \to \mathcal{M}^*$ be a smooth map. We think of this map as a one-parameter family of random cocycles so we use the subscript notation $t \mapsto A_t$. For its components we write $A_t(i)$ instead of $(A_t)_i$, while the fiber iterates are denoted by

$$A_t^n(\omega) = A_t(\omega_n) \cdots A_t(\omega_1)\,.$$

For every parameter $t \in \Lambda$, denote by $\mathcal{Q}_t$ and $\bar{\mathcal{Q}}_t$ the Markov operators corresponding to the cocycle $A_t$ (defined as above). Moreover, let $\eta_t \in \text{Prob}(\mathbb{P}(\mathbb{R}^2))$ be the unique $\mathcal{Q}_t$-stationary measure. Note that all the results proven in section 7.1 for a given cocycle $A \in \mathcal{M}$, namely the explicit formula of the stationary measure, Furstenberg's formula and the strong mixing of the Markov operator, apply to $A_t$ for all $t \in \Lambda$. Furthermore, since the probability vector $p = (p_1, \ldots, p_k)$ and the singular/invertible symbols do not change, all the mixing parameters are uniform in $t$. Finally, recall that the stationary measure is given by:

$$\eta_t = \sum_{s \in \mathcal{A}_{\text{sing}}} q_s \sum_{n=0}^{\infty} \sum_{\underline{\omega} \in \mathcal{A}_{\text{inv}}^n} p(\underline{\omega}) \delta_{\hat{A}_t^n(\underline{\omega})\, \hat{r}_s}\,.$$

### 7.2.1 The winding property

We assume that the family of cocycles $\Lambda \ni t \mapsto A_t$ is positively winding and that its singular components are constant. More precisely, we impose the following conditions on the smooth map $t \mapsto A_t \in \mathcal{M}^*$:

(A1) For all $i \in \mathcal{A}_{\text{sing}}$, $A_t(i) \equiv A_i$.

(A2) There is $c_0 > 0$ such that for all $t \in \Lambda$, $j \in \mathcal{A}_{\text{inv}}$ and $\hat{v} \in \mathbb{P}(\mathbb{R}^2)$ we have

$$\frac{A_t(j)\, v \wedge \frac{d}{dt} A_t(j) v}{\|A_t(j)\, v\|^2} \geq c_0\,.$$

**Remark 7.3** *By [7, Proposition 3.1], the quantity in item (A2) above, which we refer to as the rotation speed of the map $t \mapsto A_t$, can be characterized by*

$$\frac{d}{dt} \hat{A}_t(j)\hat{v} = \frac{d}{dt} \frac{A_t(j)v}{\|A_t(j)v\|} = \frac{A_t(j)\, v \wedge \frac{d}{dt} A_t(j)v}{\|A_t(j)\, v\|^2}\,.$$

**Remark 7.4** *A more general version of the winding condition requires that the inequality in assumption (A2) above holds for some iterate $\underline{A}_t^{n_0}$ of the cocycle. For simplicity we assume that $n_0 = 1$.*

*Moreover, if the assumption (A2) holds, then there exists $c_1 > 0$ that only depends on the map $A$, such that $\forall n \in \mathbb{N}$, $\forall \omega \in \mathcal{A}^n_{\text{inv}}$, $\forall \hat{v} \in \mathbb{P}(\mathbb{R}^2)$,*

$$\frac{d}{dt} \frac{A_t^n(\omega)v}{\|A_t^n(\omega)v\|} \geq c_1. \tag{7.5}$$

*In other words, if $A_t$ is a family of positively winding cocycles, then so is $A_t^n$ for every $n \in \mathbb{N}$, with rotation speed uniformly (in $n$) bounded from below. For a proof of this statement see [7, Section 3.1], specifically Propositions 3.3 and 3.4.*

**Example 7.1** *Given any tuple $\underline{A} = (A_1, \ldots, A_k) \in \mathcal{M}^*$, the one-parameter family $\left[-\frac{1}{2}, \frac{1}{2}\right] \mapsto A_t \in \text{Mat}_2^+(\mathbb{R})^k$, $A_t(i) = A_i$ if $i \in \mathcal{A}_{\text{sing}}$ and*

$$A_t(j) = A_j\, R_{2\pi t} = \begin{bmatrix} \cos 2\pi t & -\sin 2\pi t \\ \sin 2\pi t & \cos 2\pi t \end{bmatrix} \quad \text{if } j \in \mathcal{A}_{\text{inv}}$$

*satisfies the assumptions (A1) and (A2) above and passes through $\underline{A}$.*

In order to simplify the exposition in the estimates to follow, we will write $a \lesssim b$ if there is some absolute constant $C < \infty$ such that $a \leq C\, b$. Moreover, for an arc $\hat{I} \subset \mathbb{P}(\mathbb{R}^2)$, $m(\hat{I})$ denotes its length, while $\text{Leb}(E)$ stands for the Lebesgue measure of a subset $E$ of the real line.

**Lemma 7.4** *Given any $\varepsilon > 0$, the set*

$$\hat{I}_\varepsilon \colon = \left\{ \hat{v} \in \mathbb{P}(\mathbb{R}^2) \colon \|A_i \frac{v}{\|v\|}\| < \varepsilon \text{ for some } i \in \mathcal{A}_{\text{sing}} \right\}$$

*is a finite union of arcs with $m(\hat{I}_\varepsilon) \lesssim \varepsilon$.*

*Proof.* Let $i \in \mathcal{A}_{\text{sing}}$ and consider the normalized directions $r_i$ and $k_i$ of the range and kernel of $A_i$, respectively, such that $\|k_i\| = \|r_i\| = 1$. Given any vector $v \in \mathbb{R}^2$, we can write $v = \alpha_1 r_i + \alpha_2 k_i$.

Hence,

$$\|A_i \frac{v}{\|v\|}\| = \frac{|\alpha_1|}{\|v\|}.$$

Therefore, $\|A_i \frac{v}{\|v\|}\| < \varepsilon$ in a small interval around each kernel $k_i$. Moreover $\hat{I}_\varepsilon$ is a finite union of arcs of length approximately $2\varepsilon$ each. Hence $m(\hat{I}_\varepsilon) \lesssim \varepsilon$. ∎

**Lemma 7.5** *Given any arc $\hat{I} \subset \mathbb{P}(\mathbb{R}^2)$, $n \in \mathbb{N}$, $\omega \in \mathcal{A}_{\text{inv}}^n$ and $\hat{v} \in \mathbb{P}(\mathbb{R}^2)$, it holds that*

$$\text{Leb}\left\{t \in \Lambda \colon \hat{A}_t^n(\omega)\hat{v} \in \hat{I}\right\} \leq \frac{m(\hat{I})}{c_1}.$$

*Proof.* Since the projective line $\mathbb{P}(\mathbb{R}^2)$ is one dimensional, we may regard the map $\Lambda \ni t \mapsto \hat{A}_t^n(\omega)\hat{v}$ as a real valued map, whose derivative (because of the winding condition) is bounded from below by $c_1$. The result then follows by the mean value theorem. ■

The main point of introducing the winding property was to prove the previous lemma. Moreover, the assumption that the invertible matrices have positive determinant was made to guarantee the uniform lower bound $c_1$ on the rotation speed of $A_t^n$. If it is possible to prove lemma 7.5 using another method, then we believe that the main results of this chapter could be extended to $\text{Mat}_2(\mathbb{R})$ cocycles.

### 7.2.2
### Preparing the proofs

Consider the maps $\varphi_t \colon \mathcal{A} \times \mathbb{P}(\mathbb{R}^2) \to [-\infty, \infty)$ and $\psi_t \colon \mathbb{P}(\mathbb{R}^2) \to [-\infty, \infty)$ given by

$$\varphi_t(i, \hat{v}) := \log\|A_t(i)\frac{v}{\|v\|}\| \text{ and } \psi_t(\hat{v}) = \sum_{i \in \mathcal{A}} p_i \varphi_t(i, \hat{v}) = \int_{\mathcal{A}} \varphi_t(i, \hat{v}) dp(i).$$

Let $c = c(A)$ be a constant that satisfies

(i) $e^{-c} \leq \|A_t(j)\| \leq e^c \quad \forall t \in \Lambda$ and $\forall j \in \mathcal{A}_{\text{inv}}$.

(ii) $\|A_i\| \leq e^c \quad \forall i \in \mathcal{A}_{\text{sing}}$.

(iii) $\|A_i r_l\| \geq e^{-c} \quad \forall i, l \in \mathcal{A}_{\text{sing}}$.

Such a constant exists by the compactness of $\Lambda$ and the fact that $\underline{A} \in \mathcal{M}^*$, thus $\hat{r}_i \neq \hat{k}_j$ and $A_j r_i \neq 0$ for all $i, j \in \mathcal{A}_{\text{sing}}$. It follows that for all $j \in \mathcal{A}_{\text{inv}}$ and $\hat{v} \in \mathbb{P}(\mathbb{R}^2)$ we have $\left|\varphi_t(j, \hat{v})\right| \leq c$. Moreover, the upper bound $\varphi_t(j, \hat{v}) \leq c$ holds for every $j \in \mathcal{A}$ and $\hat{v} \in \mathbb{P}(\mathbb{R}^2)$.

**Lemma 7.6** *There is $C < \infty$ which depends on the constants $c, c_1$, such that the following hold.*

*(i) For all $n \in \mathbb{N}$, $\omega \in \mathcal{A}_{\text{inv}}^n$, $i \in \mathcal{A}_{\text{sing}}$, $\hat{v} \in \mathbb{P}(\mathbb{R}^2)$ and $N \geq 0$,*

$$\text{Leb}\{t \in \Lambda \colon \varphi_t(i, \hat{A}_t^n(\omega)\hat{v}) < -N\} \leq C\, e^{-N}.$$

*Moreover,*

$$\int_\Lambda \varphi_t^2(i, \hat{A}_t^n(\omega)\hat{v})\, dt \leq C.$$

*(ii) For all $n \in \mathbb{N}$, $\omega \in \mathcal{A}^n$, $i, s \in \mathcal{A}_{\rm sing}$ and $N \geq 0$,*

$$\mathrm{Leb}\{t \in \Lambda \colon \varphi_t(i, \hat{A}_t^n(\omega)\hat{r}_s) < -N\} \leq C\, e^{-N}\,.$$

*Moreover,*

$$\int_\Lambda \varphi_t^2(i, \hat{A}_t^n(\omega)\hat{r}_s)\, dt \leq C.$$

*(iii) Furthermore, for all $i \in \mathcal{A}$, $\omega \in X$ and $n \in \mathbb{N}$,*

$$\int_\Lambda \int_{\mathbb{P}(\mathbb{R}^2)} \varphi_t^2(i, \hat{A}_t^n(\omega)\hat{v})\, d\eta_t(\hat{v})\, dt \leq C.$$

*In particular,*

$$\int_\Lambda \int_{\mathbb{P}(\mathbb{R}^2)} \psi_t^2(\hat{A}_t^n(\omega)\hat{v})\, d\eta_t(\hat{v})\, dt \leq C.$$

*Proof.* (i) If $i \in \mathcal{A}_{\rm sing}$ then $\varphi_t(i, \hat{w}) < -N$ holds if and only if $\|A_i \frac{w}{\|w\|}\| < e^{-N}$. By Lemma 7.4 the set of such points $\hat{w}$ is a finite union of arcs with total measure of order $\varepsilon := e^{-N}$, and by Lemma 7.5 the measure of the set of parameters $t$ for which $\hat{A}_t^n(\omega)\hat{v}$ belongs to these arcs is $\lesssim \varepsilon$. This proves the first statement in item (i).

For the second statement, note that

$$\begin{aligned}
&\int_\Lambda \varphi_t^2(i, \hat{A}_t^n(\omega)\hat{v})\, dt = \int_0^\infty \mathrm{Leb}\{t \in \Lambda \colon \varphi_t^2(i, \hat{A}_t^n(\omega)\hat{v}) \geq x\}\, dx \\
&= \int_0^{2c^2} \mathrm{Leb}\{t \in \Lambda \colon \varphi_t^2(i, \hat{A}_t^n(\omega)\hat{v}) \geq x\}\, dx \\
&+ \int_{2c^2}^\infty \mathrm{Leb}\{t \in \Lambda \colon \varphi_t^2(i, \hat{A}_t^n(\omega)\hat{v}) \geq x\}\, dx \\
&\leq 2c^2 + \int_{2c^2}^\infty \mathrm{Leb}\{t \in \Lambda \colon \varphi_t(i, \hat{A}_t^n(\omega)\hat{v}) \leq -\sqrt{x}\}\, dx \\
&\lesssim c^2 + \int_{2c^2}^\infty e^{-\sqrt{x}}\, dx \lesssim 1.
\end{aligned}$$

The bound in the penultimate line above holds because $\Lambda$ has length one and, moreover, the upper bound $\varphi_t(i, \hat{v}) \leq c$ is valid for all $\hat{v} \in \mathbb{P}(\mathbb{R}^2)$, so that $\varphi_t(i, \hat{A}_t^n(\omega)\hat{v}) \leq c < \sqrt{x}$ when $x \geq 2c^2$.

(ii) Note that if $\omega \in \mathcal{A}_{\rm inv}^n$, then the statement follows from item (i). Now we consider any word $\omega$, invertible or not, which begins with a singular vector, namely $r_s$. Let $\omega = (\omega_0, \ldots, \omega_{n-1}) \notin \mathcal{A}_{\rm inv}^n$. We split the argument into two cases: either $\omega_{n-1} \in \mathcal{A}_{\rm sing}$ or else $\omega_{n-1} \in \mathcal{A}_{\rm inv}$ but there is $0 \leq j \leq n-2$ such that $\omega_j \in \mathcal{A}_{\rm sing}$.

If $\omega_{n-1} \in \mathcal{A}_{\text{sing}}$, then

$$\hat{A}_t^n(\omega)\hat{r}_s = \hat{A}_t(\omega_{n-1})\hat{A}_t^{n-1}(\omega)\hat{r}_s = \hat{A}_{\omega_{n-1}}\,\hat{A}_t^{n-1}(\omega)\hat{r}_s = \hat{r}_{\omega_{n-1}}.$$

Then by the choice of the constant $c$, for all parameters $t \in \Lambda$,

$$\varphi_t(i, \hat{A}_t^n(\omega)\hat{r}_s) = \log\|\hat{A}_i\hat{r}_{\omega_{n-1}}\| \geq -c,$$

hence the set $\{t \in \Lambda \colon \varphi_t(i, \hat{A}_t^n(\omega)\hat{r}_s) < -N\}$ becomes empty for $N \geq c$ and the statement follows.

If $\omega_{n-1} \in \mathcal{A}_{\text{inv}}$ and $\omega_j \in \mathcal{A}_{\text{sing}}$ for some index $0 \leq j \leq n-2$, let $n'$ be the largest such index and let $\omega' := (\omega_{n'+1}, \ldots, \omega_{n-1}) \in \mathcal{A}_{\text{inv}}^{n-n'-1}$. Then

$$\hat{A}_t^n(\omega)\hat{r}_s = \hat{A}_t^{n-n'-1}(\omega')\hat{A}_{\omega_{n'}}\hat{A}_t^{n'}(\omega)\hat{r}_s = \hat{A}_t^{n-n'-1}(\omega')\hat{r}_{\omega_{n'}}.$$

Thus $\varphi_t(i, \hat{A}_t^n(\omega)\hat{r}_s) = \varphi_t(i, \hat{A}_t^{n-n'-1}(\omega')\hat{r}_{\omega_{n'}})$, with $\omega' \in \mathcal{A}_{\text{inv}}^{n-n'-1}$, hence item (i) is applicable and the conclusion follows.

(iii) If $i \in \mathcal{A}_{\text{inv}}$, then the statement holds immediately since $\varphi_t^2(i, \hat{w}) \leq c^2$ for all $(t, \hat{w}) \in \Lambda \times \mathbb{P}(\mathbb{R}^2)$. Let us then fix $i \in \mathcal{A}_{\text{sing}}$. Using the explicit formula of $\eta_t$, note that

$$\begin{aligned}
&\int_\Lambda \int_{\mathbb{P}(\mathbb{R}^2)} \varphi_t^2(i, \hat{A}_t^n(\omega)\hat{v})\, d\eta_t(\hat{v})\, dt \\
&= \sum_{s \in \mathcal{A}_{\text{sing}}} q_s \sum_{j=0}^{\infty} \sum_{\underline{\omega} \in \mathcal{A}_{\text{inv}}^j} p(\underline{\omega}) \int_\Lambda \varphi_t^2(i, \hat{A}_t^n(\omega)\, \hat{A}_t^j(\underline{\omega})\hat{r}_s)\, dt \\
&= \sum_{s \in \mathcal{A}_{\text{sing}}} q_s \sum_{j=0}^{\infty} \sum_{\underline{\omega} \in \mathcal{A}_{\text{inv}}^j} p(\underline{\omega}) \int_\Lambda \varphi_t^2(i, \hat{A}_t^{n+j}(\underline{\omega}\omega)\hat{r}_s)\, dt \\
&\leq \sum_{s \in \mathcal{A}_{\text{sing}}} q_s \sum_{j=0}^{\infty} \sum_{\underline{\omega} \in \mathcal{A}_{\text{inv}}^n} p(\underline{\omega})\, C = C,
\end{aligned}$$

where the inequality in the last line follows from item (ii) above and the last equality comes from equation (7.1). ■

Let $\tilde{\eta} \in \text{Prob}(\Lambda \times \mathbb{P}(\mathbb{R}^2))$ be the probability measure whose disintegration relative to the Lebesgue measure on $\Lambda$ is $\{\eta_t\}$, that is, the measure characterized by

$$\int \phi(t, \hat{v})\, d\tilde{\eta}(t, \hat{v}) = \int_\Lambda \int_{\mathbb{P}(\mathbb{R}^2)} \phi(t, \hat{v})\, d\eta_t(\hat{v})\, dt \quad \forall \phi \in L^\infty(\Lambda \times \mathbb{P}(\mathbb{R}^2))\,.$$

**Lemma 7.7** *Given $i \in \mathcal{A}$, $n \in \mathbb{N}$, $N > c$ and $\omega \in X$, let*

$$P_{i,n,N}(\omega) := \left\{(t, \hat{v}) \colon \varphi_t(i, \hat{A}_t^n(\omega)\hat{v}) < -N\right\}.$$

*Then $\tilde{\eta}(P_{i,n,N}(\omega)) \le Ce^{-N}$.*

*Proof.* If $i \in \mathcal{A}_{\mathrm{inv}}$ then $\varphi_t(i, \hat{v}) \ge -c > -N$ for all $\hat{v} \in \mathbb{P}(\mathbb{R}^2)$, so $P_{i,n,N}(\omega)$ is empty. Hence, consider $i \in \mathcal{A}_{\mathrm{sing}}$ and note that

$$\begin{aligned}\tilde{\eta}(P_{i,n,N}(\omega)) &= \int_\Lambda \int_{\mathbb{P}(\mathbb{R}^2)} \mathbb{1}_{P_{i,n,N}(\omega)}(t, \hat{v})\, d\eta_t(\hat{v})dt \\ &= \sum_{s \in \mathcal{A}_{\mathrm{sing}}} q_s \sum_{j=0}^{\infty} \sum_{\underline{\omega} \in \mathcal{A}_{\mathrm{inv}}^j} p(\underline{\omega}) \int_\Lambda \mathbb{1}_{P_{i,n,N}(\omega)}(t, \hat{A}_t^j(\underline{\omega})\hat{r}_s)\, dt.\end{aligned}$$

Moreover,

$$\begin{aligned}(t, \hat{A}_t^j(\underline{\omega})\hat{r}_s) \in P_{i,n,N}(\omega) &\iff \varphi_t(i, \hat{A}_t^n(\omega)\hat{A}_t^j(\underline{\omega})\hat{r}_s) < -N \\ &\iff \varphi_t(i, \hat{A}_t^{n+j}(\underline{\omega}\omega)\hat{r}_s) < -N.\end{aligned}$$

By Lemma 7.6 item (ii), for all $j \in \mathbb{N}$,

$$\int_\Lambda \mathbb{1}_{P_{i,n,N}(\omega)}(t, \hat{A}_t^j(\underline{\omega})\hat{r}_s)\, dt = \mathrm{Leb}\{t \colon \varphi_t(i, \hat{A}_t^{n+j}(\underline{\omega}\omega)\hat{r}_s < -N\} \lesssim e^{-N}$$

and the conclusion follows from equation (7.1). ■

Given any $N > c$, consider the truncation

$$\varphi_{t,N} := \max\{\varphi_t, -N\}.$$

Note that $\varphi_{t,N} \in L^\infty(\mathcal{A} \times \mathbb{P}(\mathbb{R}^2))$ (a property that does not hold for $\varphi_t$) and $\|\varphi_{t,N}\|_\infty \le N$. Moreover, $\varphi_t(i, \hat{v}) = \varphi_{t,N}(i, \hat{v})$ if and only if $i \in \mathcal{A}_{\mathrm{inv}}$ or $i \in \mathcal{A}_{\mathrm{sing}}$ and $\varphi_t(i, \hat{v}) > -N$.

**Lemma 7.8** *For all $i \in \mathcal{A}$, $n \in \mathbb{N}$, $N > c$ and $\omega \in X$ we have*

$$\int_\Lambda \int_{\mathbb{P}(\mathbb{R}^2)} \left|\varphi_t(i, \hat{A}_t^n(\omega)\hat{v}) - \varphi_{t,N}(i, \hat{A}_t^n(\omega)\hat{v})\right| d\eta_t(\hat{v})\, dt \lesssim e^{-N/3}.$$

*Proof.* The statement follows immediately when $i \in \mathcal{A}_{\mathrm{inv}}$, since in this case $\varphi_t(i, \hat{v}) = \varphi_{t,N}(i, \hat{v})$ for all $\hat{v} \in \mathbb{P}(\mathbb{R}^2)$.

We fix $i \in \mathcal{A}_{\mathrm{sing}}$ and recall that $\varphi_t(i, \hat{A}_t^n(\omega)\hat{v}) \ne \varphi_{t,N}(i, \hat{A}_t^n(\omega)\hat{v})$ if and only if $\varphi_t(i, \hat{A}_t^n(\omega)\hat{v}) < -N$, that is, if and only if $(t, \hat{v}) \in P_{i,n,N}(\omega) =: P_N$, which, by Lemma 7.7, is a set of $\tilde{\eta}$-measure $\lesssim e^{-N}$. Then

$$\begin{aligned}&\int_\Lambda \int_{\mathbb{P}(\mathbb{R}^2)} \left|\varphi_t(i, \hat{A}_t^n(\omega)\hat{v}) - \varphi_{t,N}(i, \hat{A}_t^n(\omega)\hat{v})\right| d\eta_t(\hat{v})\, dt \\ &= \int_{\Lambda\times\mathbb{P}(\mathbb{R}^2)} \left|\varphi_t(i, \hat{A}_t^n(\omega)\hat{v}) - \varphi_{t,N}(i, \hat{A}_t^n(\omega)\hat{v})\right| \mathbb{1}_{P_N}(t,\hat{v})\, d\tilde{\eta}(t,\hat{v}) \\ &\leq \|g(t,\hat{v})\|_{L^2(\tilde{\eta})}\ \tilde{\eta}(P_N)^{1/2}\end{aligned}$$

where $g(t,\hat{v}) := \varphi_t(i, \hat{A}_t^n(\omega)\hat{v}) - \varphi_{t,N}(i, \hat{A}_t^n(\omega)\hat{v})$ and we used Cauchy-Schwarz in the last inequality above.

Moreover, by Lemma 7.6 item (iii),

$$\|g(t,\hat{v})\|_{L^2(\tilde{\eta})} \leq \|\varphi_t(i, \hat{A}_t^n(\omega)\hat{v})\|_{L^2(\tilde{\eta})} + \|\varphi_{t,N}(i, \hat{A}_t^n(\omega)\hat{v})\|_{L^2(\tilde{\eta})} \leq \sqrt{C} + N.$$

Thus, we conclude that

$$\int_\Lambda \int_{\mathbb{P}(\mathbb{R}^2)} \left|\varphi_t(i, \hat{A}_t^n(\omega)\hat{v}) - \varphi_{t,N}(i, \hat{A}_t^n(\omega)\hat{v})\right| d\eta_t(\hat{v}) dt \lesssim (\sqrt{C}+N)e^{-N/2} \lesssim e^{-N/3}$$

which completes the proof. ■

### 7.2.3 Large deviations

In order to prove large deviations estimates for cocycles in $\mathcal{M}^*$, we will use theorem 2.2. For the remaining part of the argument, it is very important to note the precise dependence of the large deviations estimate on the norm (in our case the $L^\infty$-norm) of the observable.

We are ready to state and prove a stronger, parametric version of the large deviations type (LDT) estimate for cocycles in $\mathcal{M}^*$, which will also imply the result stated in the introduction.

**Theorem 7.3** *Let $A\colon \Lambda \to \mathcal{M}^*$ be a smooth family of cocycles satisfying Assumptions A1 and A2 in subsection 7.2.1. Then for every $\epsilon > 0$ and $\tilde{\eta}$-a.e $(t,\hat{v})$, there exist $c_0(\epsilon) > 0$ and $n_0(\epsilon, t, \hat{v}) \in \mathbb{N}$ such that for every $n \geq n_0(\epsilon, t, \hat{v})$,*

$$\mu\left\{\omega \in X\colon \left|\frac{1}{n}\log\|A_t^n(\omega)v\| - L_1(A_t)\right| > \epsilon\right\} < e^{-c_0(\epsilon)n^{1/3}}.$$

*Moreover, for Lebesgue a.e. $t \in \Lambda$, given any $\epsilon > 0$ there are $c_0(\epsilon) > 0$ and $n_0(\epsilon, t) \in \mathbb{N}$ such that for all $n \geq n_0(\epsilon, t)$.*

$$\mu\left\{\omega \in X\colon \left|\frac{1}{n}\log\|A_t^n(\omega)\| - L_1(A_t)\right| > \epsilon\right\} < e^{-c_0(\epsilon)n^{1/3}},$$

*that is, a (sub-exponential) large deviations type estimate holds for Lebesgue almost every cocycle along the curve $t \mapsto A_t$.*

*Proof.* Given a parameter $t \in \Lambda$, consider the Markov chain $Z_n \colon X \times \mathbb{P}(\mathbb{R}^2) \to \mathcal{A} \times \mathbb{P}(\mathbb{R}^2)$,

$$Z_n(\omega, \hat{v}) := (\omega_n, \hat{A}_t^n(\omega)\hat{v}) .$$

Note that the associated Markov kernel is given by

$$\mathcal{A} \times \mathbb{P}(\mathbb{R}^2) \ni (\omega_0, \hat{v}) \mapsto p \times \delta_{\hat{A}_t(\omega_0)\hat{v}},$$

so that its corresponding Markov operator is precisely the operator $\bar{\mathcal{Q}}_t$ defined by $\bar{\mathcal{Q}}_t \varphi(j, \hat{v}) = \sum_{i=1}^k p_i \varphi(i, \hat{A}_t(j)\hat{v})$, whose stationary measure is $p \times \eta_t$. Recall that by theorem 7.1, $\mathcal{Q}_t$ is uniformly ergodic on $(L^\infty(\mathbb{P}(\mathbb{R}^2)), \|\cdot\|_\infty)$ and by proposition 2.4, $\bar{\mathcal{Q}}_t$ is uniformly ergodic on $(L^\infty(\mathcal{A} \times \mathbb{P}(\mathbb{R}^2)), \|\cdot\|_\infty)$.

For the observable $\varphi_t(i, \hat{v}) := \log\|A_t(i) \frac{v}{\|v\|}\|$ it holds that

$$\begin{aligned} \varphi_t \circ Z_{n-1}(\omega, \hat{v}) = \varphi_t(\omega_{n-1}, \hat{A}_t^{n-1}(\omega)\hat{v}) &= \log\|A_t(\omega_{n-1}) \frac{\hat{A}_t^{n-1}(\omega)\hat{v}}{\|\hat{A}_t^{n-1}(\omega)\hat{v}\|}\| \\ &= \log\|\hat{A}_t^n(\omega)\hat{v}\| - \log\|\hat{A}_t^{n-1}(\omega)\hat{v}\| . \end{aligned}$$

Thus, by the definition of stochastic Birkhoff sums (recall section 2.1.3),

$$\frac{1}{n} S_n \varphi_t(\omega, \hat{v}) = \frac{1}{n} \log\|A_t^n(\omega) v\|$$

where $v$ is a unit vector representative of the point $\hat{v}$.

Moreover, by Furstenberg's formula (7.2),

$$L_1(A_t) = \int \psi_t(\hat{v})\, d\eta_t(\hat{v}) = \int \varphi_t(i, \hat{v})\, d(p \times \eta_t)(i, \hat{v}).$$

Furthermore, note that Lemma 7.6 item (iii) implies that for Lebesgue a.e. $t \in \Lambda$, $L_1(A_t) > -\infty$.

Our statement (in fact, a stronger version thereof) would then immediately follow from theorem 2.2 if the observable $\varphi_t$ was bounded. That is not the case, precisely because of the singular matrices. The idea is then to use the truncations $\varphi_{t,N}$ (for which theorem 2.2 is applicable) as a substitute for $\varphi_t$, where the order $N$ of the truncation is adapted to the scale $n$ at which we prove the LDT estimate.

More precisely, given any large enough scale $n \in \mathbb{N}$, let $N_n \in \mathbb{N}$ be a truncation order to be chosen later (anticipating, the choice that optimizes the estimates will turn out to be $N_n \sim n^{1/3}$). We will transfer the LDT estimate at scale $n$ from $\varphi_{t,N_n}$ to $\varphi_t$ by eliminating an $\tilde{\eta}$-negligible set of parameters $(t, \hat{v})$.

For all $n \in \mathbb{N}$, define the following real-valued functions on $X \times \Lambda \times \mathbb{P}(\mathbb{R}^2)$:

$$\Delta_n(\omega, t, \hat{v}) := \frac{1}{n} S_n\varphi_t(\omega, \hat{v}) - \int \varphi_t \, dp \times \eta_t$$
$$E_n(\omega, t, \hat{v}) := \frac{1}{n} S_n\varphi_{t,N_n}(\omega, \hat{v}) - \int \varphi_{t,N_n} \, dp \times \eta_t$$

and

$$g_n(\omega, t, \hat{v}) := \frac{1}{n} S_n\varphi_t(\omega, \hat{v}) - \frac{1}{n} S_n\varphi_{t,N_n}(\omega, \hat{v}).$$

Then

$$\left|\Delta_n(\omega, t, \hat{v}) - E_n(\omega, t, \hat{v})\right| \le \left|g_n(\omega, t, \hat{v})\right| + \int \left|\varphi_t - \varphi_{t,N_n}\right| dp \times \eta_t. \tag{7.6}$$

Note that

$$\Delta_n(\omega, t, \hat{v}) = \frac{1}{n} \log\|A_t^n(\omega)v\| - L_1(A_t)\,.$$

Moreover, given $\epsilon > 0$, the general LDT estimate (2.4) shows that *for every* $(t, \hat{v}) \in \Lambda \times \mathbb{P}(\mathbb{R}^2)$,

$$\mu\left\{\omega \in X \colon \left|E_n(\omega, t, \hat{v})\right| > \frac{\epsilon}{2}\right\} \le 8\, e^{-c_0(\epsilon) N_n^{-2}\, n} \tag{7.7}$$

where $c_0(\epsilon)$ is essentially of order $\epsilon^2$.

By Lemma 7.8, it holds that

$$\int \left(\int \left|\varphi_t - \varphi_{t,N_n}\right| dp \times \eta_t\right) dt$$
$$= \sum_{i\in\mathcal{A}} p_i \int_\Lambda \int_{\mathbb{P}(\mathbb{R}^2)} \left|\varphi_t(i, \hat{v}) - \varphi_{t,N_n}(i, \hat{v})\right| d\eta_t(\hat{v})\, dt \lesssim e^{-N_n/3}\,.$$

Moreover, we will also show that

$$\iint \left|g_n(\omega, t, \hat{v})\right| d\mu(\omega)\, d\tilde{\eta}(t, \hat{v}) \lesssim \sqrt{n}\, e^{-N_n/3}\,. \tag{7.8}$$

Indeed, note that

$$g_n(\omega, t, \hat{v}) = \frac{1}{n} \sum_{j=0}^{n-1} \left(\varphi_t(\omega_j, \hat{A}_t^j(\omega)\hat{v}) - \varphi_{t,N_n}(\omega_j, \hat{A}_t^j(\omega)\hat{v})\right)\,.$$

Given $\omega \in X$, if $g_n(\omega, t, \hat{v}) \neq 0$ then there is $0 \le j \le n-1$ such that $\varphi_t(\omega_j, \hat{A}_t^j(\omega)\hat{v}) \neq \varphi_{t,N_n}(\omega_j, \hat{A}_t^j(\omega)\hat{v})$, that is, $\varphi_t(\omega_j, \hat{A}_t^j(\omega)\hat{v}) < -N_n$, or $(t, \hat{v}) \in P_{\omega_j, j, N_n}(\omega)$.

Thus for $(t, \hat{v}) \notin B_n(\omega) := \bigcup_{j=0}^{n-1} P_{\omega_j, j, N_n}(\omega)$, where $\tilde{\eta}(B_n(\omega) \lesssim n e^{-N_n}$, we have that $g_n(\omega, t, \hat{v}) = 0$.

Using Fubini and Cauchy-Schwarz it follows that

$$\begin{aligned}
&\iint \left|g_n(\omega,t,\hat v)\right| d\mu(\omega)\, d\tilde\eta(t,\hat v) = \iint \left|g_n(\omega,t,\hat v)\right| d\tilde\eta(t,\hat v)\, d\mu(\omega)\\
&= \iint \left|g_n(\omega,t,\hat v)\right| \mathbb{1}_{B_n(\omega)}(t,\hat v)\, d\tilde\eta(t,\hat v)\, d\mu(\omega)\\
&\le \int \|g_n(\omega,\cdot)\|_{L^2(\tilde\eta)}\; \tilde\eta(B_n(\omega))^{1/2}\, d\mu(\omega)\\
&\lesssim \sqrt{n}\, e^{-N_n/2} \int \|g_n(\omega,\cdot)\|_{L^2(\tilde\eta)}\, d\mu(\omega)\,.
\end{aligned}$$

By Lemma 7.6 item (iii), for all $\omega \in X$,

$$\int \varphi_t^2(\omega_j, \hat A_t^j(\omega)\hat v)\, d\eta_t(\hat v)\, dt \le C,$$

hence

$$\|\frac{1}{n} S_n\varphi_t(\omega,\hat v)\|_{L^2(\tilde\eta)} \le \sqrt{C},$$

while $\left|\varphi_{t,N_n}(i,\hat w)\right| \le N_n$. Therefore, we conclude that

$$\|\frac{1}{n} S_n\varphi_{t,N_n}(\omega,\hat v)\|_{L^2(\tilde\eta)} \le N_n.$$

Thus for all $\omega \in X$, $\|g_n(\omega,\cdot)\|_{L^2(\tilde\eta)} \le \sqrt{C} + N_n$, which then implies (7.8). Integrating the inequality 7.6 with respect to $\mu$ and $\tilde\eta$ we obtain

$$\iint \left|\Delta_n(\omega,t,\hat v) - E_n(\omega,t,\hat v)\right| d\mu(\omega)\, d\tilde\eta(t,\hat v) \lesssim \sqrt{n}\, e^{-N_n/3}\,.$$

Applying Chebyshev's inequality to the function

$$(t,\hat v) \mapsto \int \left|\Delta_n(\omega,t,\hat v) - E_n(\omega,t,\hat v)\right| d\mu(\omega),$$

there is a set $\mathcal{B}_n \subset \Lambda \times \mathbb{P}(\mathbb{R}^2)$ such that $\tilde\eta(\mathcal{B}_n) \lesssim \sqrt{n}\, e^{-N_n/6}$ and if $(t,\hat v) \notin \mathcal{B}_n$ then

$$\int \left|\Delta_n(\omega,t,\hat v) - E_n(\omega,t,\hat v)\right| d\mu(\omega) < e^{-N_n/6}\,.$$

Since $\sum_{n\ge 1} \tilde\eta(\mathcal{B}_n) < \infty$, by Borel-Cantelli, $\tilde\eta$-almost every $(t,\hat v)$ belongs to a finite number of sets $\mathcal{B}_n$. That is, for $\tilde\eta$-a.e. $(t,\hat v)$ there is $n_0(t,\hat v) \in \mathbb{N}$ such that for all $n \ge n_0(t,\hat v)$ we have $(t,\hat v) \notin \mathcal{B}_n$, so

$$\int \left|\Delta_n(\omega,t,\hat v) - E_n(\omega,t,\hat v)\right| d\mu(\omega) < e^{-N_n/6}\,.$$

Applying again Chebyshev we have that

$$\mu\left\{\omega \in X \colon \left|\Delta_n(\omega,t,\hat v) - E_n(\omega,t,\hat v)\right| > e^{-N_n/12}\right\} \le e^{-N_n/12}.$$

We have shown that there is a set $\Omega_n \subset X$ with $\mu(\Omega_n) < e^{-N_n/12}$ such that if $\omega \notin \Omega_n$ then

$$\left|\Delta_n(\omega, t, \hat{v}) - E_n(\omega, t, \hat{v})\right| < e^{-N_n/12} \leq \frac{\epsilon}{2}$$

provided $n$ is large enough depending on $(t, \hat{v})$ as above and on $\epsilon$.

Thus if $\left|\Delta_n(\omega, t, \hat{v})\right| > \epsilon$ then either $\omega \in \Omega_n$ or $\left|E_n(\omega, t, \hat{v})\right| > \frac{\epsilon}{2}$, the later holding for $\omega$ in a set of measure $\leq 8\, e^{-c_0(\epsilon) N_n^{-2}\, n}$ by (7.7).

We conclude that for $\tilde{\eta}$-a.e.$(t, \hat{v})$ and for all $\epsilon > 0$ there is $n_0 = n_0(\epsilon, t, \hat{v}) \in \mathbb{N}$ such that for all $n \geq n_0$ we have

$$\mu\left\{\omega \colon \left|\Delta_n(\omega, t, \hat{v})\right| > \epsilon\right\} < e^{-N_n/12} + 8\, e^{-c_0(\epsilon) N_n^{-2} n} < e^{-c_0(\epsilon)\, n^{1/3}}$$

provided we choose $N_n \sim n^{1/3}$. This establishes the first statement of the theorem.

In particular, by the definition of the measure $\tilde{\eta}$, for Lebesgue almost every parameter $t \in \Lambda$ and for $\eta_t$-a.e. point $\hat{v}$ this first statement of the theorem holds, namely

$$\mu\left\{\omega \in X \colon \left|\frac{1}{n} \log\|A_t^n(\omega) v\| - L_1(A_t)\right| > \epsilon\right\} < e^{-c_0(\epsilon) n^{1/3}}\,. \tag{7.9}$$

Recall that support of the measure $\eta_t$ is the set

$$\left\{\hat{A}_t^n(\underline{\omega})\hat{r}_s \colon s \in \mathcal{A}_{\text{sing}},\ \underline{\omega} \in \mathcal{A}_{\text{inv}}^n,\ n \in \mathbb{N}\right\}.$$

There is at least one singular symbol, $s \in \mathcal{A}_{\text{sing}}$ and at least one invertible one $i \in \mathcal{A}_{\text{inv}}$. By the positive winding condition,

$$\frac{d}{dt} \hat{A}_t(i)\, \hat{r}_s \geq c_0 > 0,$$

so $t \mapsto \hat{A}_t(i)\, \hat{r}_s$ cannot be constant on a set of positive Lebesgue measure. Therefore $\hat{A}_t(i)\, \hat{r}_s \neq \hat{r}_s$ for Lebesgue almost every $t$.

It follows that for Lebesgue almost every $t \in \Lambda$, the support of the measure $\eta_t$ has at least two points, $\hat{e}_t^1$ and $\hat{e}_t^2$. Thus (7.9) holds for two linearly independent vectors $e_t^1$ and $e_t^2$, hence it must hold with the matrix norm $\|A_t^n(\omega)\|$, which establishes the second statement of the theorem. ■

We now derive the version of the LDT estimate stated in the introduction of this manuscript, namely that such an estimate holds for Lebesgue almost every singular cocycle.

*Proof.* [of Theorem 1.5] Note that the set $\mathcal{M}\backslash\mathcal{M}^*$, which consist on cocycles for which a kernel $\hat{k}_i$ coincides with a range $\hat{r}_j$ for some $i, j \in \mathcal{A}_{\text{sing}}$, has zero Lebesgue measure in $\mathcal{M}$. Moreover, by 6.3, $L_1(\underline{A}) > -\infty$ for Lebesgue almost every $\underline{A} \in \mathcal{M}$. We may neglect these zero measure sets of cocycles.

Given any $\underline{A} \in \mathcal{M}$, define $\underline{A}(t) := (A_1(t), \dots, A_k(t))$ for every $t \in [-\pi, \pi]$, where

$$A_i(t) := \begin{cases} A_i\, R_t & \text{if } i \in I \\ A_i & \text{if } i \notin I \end{cases} \quad \text{and} \quad R_t := \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix}.$$

This is an analytic family taking values in $\mathcal{M}^*$ that satisfies the assumptions (A1) and (A2) of Section 7.2.1. By Theorem 7.3, $\underline{A}(t)$ satisfies large deviations of sub-exponential type for Lebesgue almost every $t \in \mathbb{R}$. Hence, since the map $\mathcal{M} \times [-\pi, \pi] \to \mathcal{M}_I$, $(\underline{A}, t) \mapsto \underline{A}(t)$, is a submersion, Lebesgue almost every $\underline{A} \in \mathcal{M}$ satisfies similar large deviations estimates. ■

### 7.2.4 Central limit theorem

We now establish a central limit theorem for singular cocycles. The proof uses the abstract central limit theorem 2.3 due to Gordin and Livšic.

We consider as in the previous subsection a smooth family of cocycles $\Lambda \ni t \mapsto \underline{A}_t \in \mathcal{M}^*$ satisfying the positive winding property. We also consider the special observables $\psi_t$ and $\varphi_t$ from before.

We introduced two related (families of) Markov operators acting on measurable bounded observables. They can actually be defined for arbitrary observables, namely if $\psi\colon \mathbb{P}(\mathbb{R}^2) \to [-\infty, \infty)$ we put

$$\mathcal{Q}_t\psi(\hat{v}) = \sum_{i\in\mathcal{A}} p_i \psi(\hat{A}_t(i)\hat{v})$$

and if $\varphi\colon \mathcal{A} \times \mathbb{P}(\mathbb{R}^2) \to [-\infty, \infty)$ we put

$$\bar{\mathcal{Q}}_t\varphi(j, \hat{v}) = \sum_{i\in\mathcal{A}} p_i \varphi(i, \hat{A}_t(j)\hat{v}).$$

Consider the projection:

$$\begin{aligned} &\pi\colon L^2(\mathcal{A} \times \mathbb{P}(\mathbb{R}^2)) \to L^2(\mathcal{A}) \\ &\pi\varphi(\hat{v}) := \sum_{i\in\mathcal{A}} p_i \varphi(i, \hat{v}). \end{aligned}$$

Using this notation, $\pi\varphi_t = \psi_t$. Moreover, $\bar{\mathcal{Q}}_t^{n+1}\varphi(i, \hat{v}) = \mathcal{Q}_t^n(\pi\varphi)(\hat{A}_t(i)\hat{v})$.

**Lemma 7.9** *For Lebesgue almost every parameter $t \in \Lambda$ we have that*

$$\sum_{n=0}^{\infty} \|\bar{\mathcal{Q}}_t^{n+1}\varphi_t - \int \varphi_t \, dp d\eta_t\|_{L^2(p\times\eta_t)} < \infty.$$

*Proof.* By the previous observations,

$$\sum_{n=0}^{\infty} \|\bar{\mathcal{Q}}_t^{n+1}\varphi_t - \int \varphi_t \, dp d\eta_t\|_{L^2} = \sum_{n=0}^{\infty} \|\mathcal{Q}_t^{n}\psi_t(\hat{A}_t(i)\hat{v}) - \int \psi_t \, d\eta_t\|_{L^2}$$
$$= \sum_{n=0}^{\infty} \left( \sum_{i\in\mathcal{A}} p_i \int \left| \mathcal{Q}_t^{n}\psi_t(\hat{A}_t(i)\hat{v}) - \int \psi_t \, d\eta_t \right|^2 \, d\eta_t(\hat{v}) \right)^{\frac{1}{2}}.$$

We start by showing that for every $i \in \mathcal{A}$ there exists $\sigma \in (0,1)$ such that for every $n \in \mathbb{N}$

$$\int_\Lambda \int_{\mathbb{P}(\mathbb{R}^2)} \left| \mathcal{Q}_t^{n}\psi_t(\hat{A}_t(i)\hat{v}) - \int \psi_t \, d\eta_t \right|^2 \, d\eta_t(\hat{v}) \, dt \lesssim \sigma^n.$$

Fix $i \in \mathcal{A}$ and note that by lemma 7.6 item (iii) and Jensen's inequality,

$$\begin{aligned}
&\int_\Lambda \int_{\mathbb{P}(\mathbb{R}^2)} \left| \mathcal{Q}_t^{n}\psi_t(\hat{A}_t(i)\hat{v}) - \int \psi_t \, d\eta_t \right|^2 \, d\eta_t(\hat{v}) \, dt \\
&= \int_\Lambda \int_{\mathbb{P}(\mathbb{R}^2)} \Bigg| \sum_{\underline{\omega}\in\mathcal{A}^n_{\text{inv}}} p(\underline{\omega})\psi_t(\hat{A}_t^n(\underline{\omega})\hat{A}_t(i)\hat{v}) \\
&- \sum_{s\in\mathcal{A}_{\text{sing}}} p_s \sum_{j=n}^{\infty} \sum_{\underline{\omega}\in\mathcal{A}^j_{\text{inv}}} p(\underline{\omega})\psi_t(\hat{A}_t^j(\underline{\omega})r_s) \Bigg|^2 d\eta_t(\hat{v})dt \\
&\leq 2 \int_\Lambda \int_{\mathbb{P}(\mathbb{R}^2)} \sum_{\underline{\omega}\in\mathcal{A}^n_{\text{inv}}} p(\underline{\omega}) \left| \psi_t(\hat{A}_t^n(\underline{\omega})\hat{A}_t(i)\hat{v}) \right|^2 \, d\tilde{\eta}(t,\hat{v}) \\
&+ 2 \int_\Lambda \int_{\mathbb{P}(\mathbb{R}^2)} \sum_{s\in\mathcal{A}_{\text{sing}}} p_s \sum_{j=n}^{\infty} \sum_{\underline{\omega}\in\mathcal{A}^j_{\text{inv}}} p(\underline{\omega}) \left| \psi_t(\hat{A}_t^j(\underline{\omega})r_s) \right|^2 \, d\tilde{\eta}(t,\hat{v}) \\
&\lesssim \sum_{\underline{\omega}\in\mathcal{A}^n_{\text{inv}}} p(\underline{\omega}) + \sum_{s\in\mathcal{A}_{\text{sing}}} p_s \sum_{j=n}^{\infty} \sum_{\underline{\omega}\in\mathcal{A}^j_{\text{inv}}} p(\underline{\omega}) \lesssim (1-q)^n
\end{aligned}$$

so the claim holds with $\sigma = 1-q$.

By Chebyshev's inequality,

$$\text{Leb}\left\{ t\in\Lambda\colon \int_{\mathbb{P}(\mathbb{R}^2)} \left| \mathcal{Q}_t^{n}\psi_t(\hat{A}_t(i)\hat{v}) - \int \psi_t \, d\eta_t \right|^2 \, d\eta_t(\hat{v}) > \sigma^{\frac{n}{2}} \right\} \lesssim \sigma^{\frac{n}{2}}.$$

Therefore, for every $i \in \mathcal{A}$ and for each $n \in \mathbb{N}$, there exists a set $B_n(i) \subset \Lambda$, with $\text{Leb}(B_n(i)) \lesssim \sigma^{\frac{n}{2}}$ such that for every $t \notin B_n(i)$,

$$\|\bar{\mathcal{Q}}_t^{n+1}\varphi_t - \int_{\mathcal{A}\times\mathbb{P}(\mathbb{R}^2)} \varphi_t \, dp d\eta_t\|_{L^2(p\times\eta_t)} \lesssim \sigma^{\frac{n}{2}}. \tag{7.10}$$

Moreover, since there is a finite number of symbols, there exists a set $B_n$ that satisfies the same properties for all symbols $i \in \mathcal{A}$ simultaneously. Furthermore, since $\sum_{n=0}^{\infty} \mathrm{Leb}(B_n) < \infty$, by the Borel-Cantelli lemma, for almost every $t \in \Lambda$, there exists $n_0(t) \in \mathbb{N}$ such that for every $n \geq n_0(t)$, $t \notin B_n$. Thus for almost every $t$, the inequality (7.10) holds, hence

$$\sum_{n=0}^{\infty} \| \bar{\mathcal{Q}}_t^{n+1} \varphi_t - \int \varphi_t \, dp d\eta_t \|_{L^2(p \times \eta_t)} < \infty.$$

which proves the lemma. ■

We are ready to formulate and to prove a parametric version of the CLT for singular cocycles.

**Theorem 7.4** *Let $\underline{A} : \Lambda \to \mathcal{M}^*$ be a smooth family of cocycles satisfying (A1) and (A2). Then, for almost every $t \in \Lambda$ there exists $\sigma = \sigma(t) > 0$ such that the following convergence in distribution*

$$\frac{\log \| A_t^n v \| - n \, L_1(A_t)}{\sigma \sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$$

*holds for $\eta_t$-a.e $\hat{v} \in \mathbb{P}(\mathbb{R}^2)$. Moreover,*

$$\frac{\log \| A_t^n \| - n \, L_1(A_t)}{\sigma \sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1).$$

*Proof.* In order to apply theorem 2.3 and establish the central limit theorem, we first note that for every parameter $t$, the Markov system $(\mathcal{A} \times \mathbb{P}(\mathbb{R}^2), \bar{\mathcal{Q}}_t)$ is ergodic, since it has a unique stationary measure $p \times \eta_t$. That is because $\bar{\mathcal{Q}}_t^n \varphi \to \int \varphi \, dp d\eta_t$ uniformly for every $\varphi \in L^\infty(\mathcal{A} \times \mathbb{P}(\mathbb{R}^2))$.

Consider the special observable $\varphi_t$ defined before and recall from the previous lemma that for Lebesgue almost every parameter $t \in \Lambda$ we have

$$\sum_{n=0}^{\infty} \| \bar{\mathcal{Q}}_t^{n} \varphi_t - \int \varphi_t \, dp d\eta_t \|_{L^2(p \times \eta_t)} < \infty.$$

This in particular allows us to define

$$g_t := \sum_{n=0}^{\infty} \bar{\mathcal{Q}}_t^n \left( \varphi_t - \int \varphi_t \, dp d\eta_t \right) \in L^2(p \times \eta_t).$$

Then for every parameter $t$,

$$\varphi_t - \int \varphi_t \, dp d\eta_t = g_t - \bar{\mathcal{Q}}_t g_t.$$

Let $\sigma^2(t) = \sigma_t^2(\varphi_t) := \|g_t\|_2^2 - \|\bar{\mathcal{Q}}_t g_t\|_2^2$.

In order to apply Theorem 2.3 it remains to prove that $\sigma^2(t) > 0$ for almost every $t \in \Lambda$. We accomplish this in several steps. The result will be proved in Lemma 7.12 and the next two lemmas prepare the proof.

**Lemma 7.10** *Let $\mathcal{PUH}$ be the set of parameters $t \in \Lambda$ such that the corresponding cocycle $\underline{A}_t$ is projectively uniformly hyperbolic. For Lebesgue almost every $t \in \mathcal{PUH}$, the corresponding stationary measure $p \times \eta_t$ has infinite support.*

*Proof.* Firstly note that if the parameter $t \in \mathcal{PUH}$ is such that $\underline{A}_t$ is diagonalizable, then the ranges of the singular matrices and the unstable directions of the invertible matrices are aligned. Therefore, the support of the corresponding stationary measure has only one point. However, this case only happens with zero measure, since this alignment will be undone by the winding property as the parameter $t$ varies. Note that singular matrices remain constant in the process.

Hence we only need to consider the case in which the cocycle $\underline{A}_t$ is not diagonalizable. Note that if $\underline{A}_t$ is projectively uniformly hyperbolic, then so is its invertible part $\underline{A}_{\text{inv}}(t) := (A_t(i))_{i \in \mathcal{A}_{\text{inv}}}$. Hence we can define $K_{\text{inv}}^u$, the set of unstable Oseledets directions $E^u(\omega)$ of the cocycle $(A_i)_{i \in \mathcal{A}_{\text{inv}}}$ over the set of points $\omega \in \mathcal{A}_{\text{inv}}^{\mathbb{Z}}$. We divide the proof in two cases: $\#K_{\text{inv}}^u = 1$ and $\#K_{\text{inv}}^u > 1$.

First, we suppose that $\#K_{\text{inv}}^u = 1$. Since we assume that $\underline{A}_t$ is not diagonalizable, there is no range in $K_{\text{inv}}^u$. Therefore, the iterates of the ranges are going to spread into infinitely many different points in the projective space by the projective uniform hyperbolic dynamics (in fact, the iterates will converge to $K_{\text{inv}}^u$). Since the corresponding stationary measure $p \times \eta_t$ is discrete and gives positive weight to every pair of the form $(j, \hat{A}_t^n(\underline{\omega})\hat{r}_s)$, we conclude that the support of $\eta_t$ is infinite.

Now, suppose that $\#K_{\text{inv}}^u > 1$. In fact, in this case the set $K_{\text{inv}}^u$ is infinite, since it is a Cantor set. Note that if there is any range outside of $K_{\text{inv}}^u$, the proof is the same as in the previous case, hence we assume that every range $r_i$ is contained in $K_{\text{inv}}^u$. Furthermore, since $\underline{A}_{inv}$ is projectively uniformly hyperbolic, the distance between $K_{\text{inv}}^u$ and $K_{\text{inv}}^s$ (the set of stable Oseledets directions $E^s(\omega)$ of the cocycle $(A_i)_{i \in \mathcal{A}_{\text{inv}}}$ over the set of points $\omega \in \mathcal{A}_{\text{inv}}^{\mathbb{Z}}$) is positive. Hence, by an appropriate choice of word $\omega$, one can iterate any range $r_i$ by $A_t^n(\omega)$ to converge to any desired element of $K_{\text{inv}}^u$. (Remember that by Lemma 6.2, there is no invariant element in $K_{\text{inv}}^u$ once $\#K_{\text{inv}}^u > 1$.) Since the corresponding stationary measure $p \times \eta_t$ is discrete and gives positive weight to every pair of the form $(j, \hat{A}_t^n(\underline{\omega})\hat{r}_s)$, we conclude that the support of $\eta_t$ is infinite. ∎

**Lemma 7.11** *For Lebesgue almost every parameter $t \in \Lambda$, the observable $\varphi_t$ cannot be $p \times \eta_t$ constant.*

*Proof.* By Theorem 6.3, for every $t \in \Lambda$, either $\underline{A}_t$ is projectively uniformly hyperbolic or it can be approximated by cocycles that admit null words. The proof is then divided into two cases.

First let us consider the set $\mathcal{B}$ of parameters $t \in \Lambda$ such that $\underline{A}_t$ is not projectively uniformly hyperbolic. Note that there exists $c > 0$ such that for every $t \in \mathcal{B}$ and all $j \in \mathcal{A}$ and $i \in \mathcal{A}_{\text{sing}}$, it holds that $|\varphi_t(j, \hat{r}_i)| \leq c$. Moreover, by Theorem 6.3 there exists $t'$ arbitrarily close to $t$ such that $\underline{A}_{t'}$ has a null word. Hence, by continuity, there exists some range $r_s$ and a finite word $\underline{\omega}$ such that $\|A^n_t(\underline{\omega})r_s\|$ is arbitrarily small, thus $|\varphi_t| \gg c$. Therefore, we conclude that for every $t \in \mathcal{B}$, $\varphi_t$ is not constant.

By Lemma 7.10 we have that for Lebesgue almost every $t \in \mathcal{PUH}$, there are infinitely many points in the support of the corresponding stationary measure $p \times \eta_t$. Moreover, $\varphi_t$ is $p \times \eta_t$-constant if and only if it takes the same value at every point in the support of $p \times \eta_t$. The only way for this to happen is if $\underline{A}_t$ had a conformal word (with a pair of non real eigenvalues). In fact, the presence of a conformal word implies that this word is projectively elliptic, hence not projectively uniformly hyperbolic. Therefore, we conclude that for almost every $t \in \mathcal{PUH}$, the observable $\varphi_t$ is not $p \times \eta_t$ constant, which finishes the proof. ■

The next lemma establishes the positivity of the variance type quantity $\sigma^2(t)$ for almost every $t \in \Lambda$ by an adaptation of the proof of [13, Proposition 2.2]. We note that this proposition is a version of the abstract CLT of Gordin and Livšic, which is more applicable to dynamical systems.

**Lemma 7.12** *The variance-type quantity defined above satisfies $\sigma^2(t) > 0$ for every $t \in \Lambda$.*

*Proof.* Consider the family of Markov operators $\bar{\mathcal{Q}}_t \colon L^\infty(\mathcal{A} \times \mathbb{P}(\mathbb{R}^2)) \to L^\infty(\mathcal{A} \times \mathbb{P}(\mathbb{R}^2))$, given by $(\bar{\mathcal{Q}}_t\varphi)(j, \hat{v}) = \sum_{i \in \mathcal{A}} \varphi(i, \hat{A}_t(j)\, \hat{v})\, p_i$ and the corresponding family of Markov kernels $\bar{K}_t \colon (\mathcal{A} \times \mathbb{P}(\mathbb{R}^2)) \to \text{Prob}(\mathcal{A} \times \mathbb{P}(\mathbb{R}^2))$. Recall that the measure $p \times \eta_t$ is $\bar{K}_t$-stationary, in the sense that $(\bar{K}_t)_* p \times \eta_t = p \times \eta_t$.

Fix any $t \in \Lambda$. Assume by contradiction that $\sigma^2_t(\varphi) = 0$. Let $x, y \in$

$\mathcal{A} \times \mathbb{P}(\mathbb{R}^2)$ with $x = (i, \hat{v}_1)$ and $y = (j, \hat{v}_2)$. Then

$$\begin{aligned}
0 &\leq \int (\bar{\mathcal{Q}}_t\, g_t(x) - g_t(y))^2 \, d\bar{K}_{t,x}(y) dp \times \eta_t(x) \\
&= \int \left\{ (\bar{\mathcal{Q}}_t\, g_t(x))^2 + (g_t(y))^2 - 2g_t(y)\bar{\mathcal{Q}}_t g_t(x) \right\} \, d\bar{K}_{t,x}(y) dp \times \eta_t(x) \\
&= \int \left\{ (g_t(y))^2 - (\bar{\mathcal{Q}}_t\, g_t(x))^2 \right\} \, d\bar{K}_{t,x}(y) dp \times \eta_t(x) \\
&= \|g_t\|_2^2 - \|\bar{\mathcal{Q}}_t\, g_t\|_2^2 = \sigma^2(t) = 0.
\end{aligned}$$

We then conclude that $g_t(y) = \bar{\mathcal{Q}}_t g_t(x)$ for $p \times \eta_t$-a.e $x$ and $\bar{K}_{t,x}$-a.e $y$. By induction, for every $n \geq 1$ $\bar{\mathcal{Q}}_t^n g_t(x) = g_t(y)$ for $p \times \eta_t$-a.e $x$ and $\bar{K}_{t,x}^n$-a.e $y$. Hence for all $n \geq 1$ and $p \times \eta_t$-a.e $x$, $g_t$ is $\bar{K}_{t,x}^n$-a.e constant.

We claim that in fact $g_t$ is $p \times \eta_t$-a.e constant. Assume by contradiction that $g_t$ is not constant for $p \times \eta_t$-a.e. Then there exist two points $x_1$ and $x_2$ in the support of the stationary measure $p \times \eta_t$ such that $p \times \eta_t(\{x_1\}) > 0$, $p \times \eta_t(\{x_2\}) > 0$ and $g_t|_{\{x_1\}} < g_t|_{\{x_2\}}$. Define $\phi_1 = \mathbb{1}_{\{x_1\}}$ and $\phi_2 = \mathbb{1}_{\{x_2\}}$. Note that $0 \leq \phi_i = \mathbb{1}_{\{x_i\}}$ and $\int \phi_i \, dp \times \eta_t > 0$ for $i = 1, 2$. Moreover, for every $x \in \mathcal{A} \times \mathbb{P}(\mathbb{R}^2)$ and every $n \geq 1$, uniformly in $x$ we have

$$\bar{K}_{t,x}^n(\{x_i\}) = (\bar{\mathcal{Q}}_t^n \mathbb{1}_{\{x_i\}})(x) = (\bar{\mathcal{Q}}_t^n \phi_i)(x) \to \int \phi_i \, dp \times \eta_t > 0.$$

Then for $n$ sufficiently large and for every $x \in \mathcal{A} \times \mathbb{P}(\mathbb{R}^2)$, $\{x_1\}$ and $\{x_2\}$ have positive $\bar{K}_{t,x}^n$ measure. However, $g_t|_{\{x_1\}} < g_t|_{\{x_2\}}$, contradicting the fact that $g_t$ is $\bar{K}_{t,x}^n$-a.e constant for $p \times \eta_t$-a.e $x$. Thus, $g_t$ is $p \times \eta_t$-a.e constant. Since $\bar{\mathcal{Q}}_t$ preserves constants, it follows that $\varphi_t = g_t - \bar{\mathcal{Q}}_t g_t = 0 \;\; p \times \eta_t$-a.e. which, by Lemma 7.11, cannot hold. Since we obtained a contradiction, we conclude that $\sigma^2(t) > 0$, as stated. ■

Theorem 2.3 is then applicable to the Markov system $(\mathcal{A} \times \mathbb{P}(\mathbb{R}^2), \bar{K}_t)$ and the observable $\varphi_t$, for Lebesgue almost every $t \in \Lambda$. We then conclude that for $\eta_t$-a.e. point $\hat{v} \in \mathbb{P}(\mathbb{R}^2)$,

$$\frac{\log\|A_t^n\, v\| - n\, L_1(A_t)}{\sigma\, \sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

In order to prove the second statement we choose a unit vector $v$ for which the CLT holds and note that

$$\frac{\log\|A_t^n\| - n\, L_1(A_t)}{\sigma\, \sqrt{n}} = \frac{\log\|A_t^n\, v\| - n\, L_1(A_t)}{\sigma\, \sqrt{n}} + \frac{\log\|A_t^n\| - \log\|A_t^n\, v\|}{\sigma\, \sqrt{n}}.$$

We claim that the sequence $\log\|A_t^n\| - \log\|A_t^n\, v\|$ is almost surely bounded, hence the last term above converges to 0 almost surely, which would then conclude the argument via Slutsky's theorem.

The claim clearly follows from the a.s. boundedness from below by a positive constant of the sequence $\frac{\|A_t^n v\|}{\|A_t^n\|}$. Let $u(A)$ denote the most expanding singular direction of the matrix $A$. By the Pythagorean theorem,

$$\frac{\|A_t^n(\omega)\, v\|}{\|A_t^n(\omega)\|} \geq \left| v \cdot u(A_t^n(\omega)) \right| \to \left| v \cdot u^\infty(A_t)(\omega) \right|.$$

Remember that the limit direction $u^\infty(A_t)(\omega)$ exists almost surely by arguments in the proof of Oseledets' theorem (see [21, Proposition 4.4]) and $u^\infty(A_t)(\omega)^\perp = E^s(\omega)$, the stable subspace in the Oseledets theorem (see the beginning of the proof of [21, Theorem 4.4]).

We must have that $\left| v \cdot u^\infty(A_t)(\omega) \right| > 0$ almost surely (which then establishes the claim). Otherwise, on a set of positive measure we would have that $v \in u^\infty(A_t)(\omega)^\perp = E^s(\omega)$, which is not possible. Indeed, $E^s(\omega)$ consists of the pre-images of the *kernel* $k_{\omega_n}$ via matrix products $A_t^n(\omega)$, for some $n \in \mathbb{N}$ such that $\omega_n \in \mathcal{A}_{\text{sing}}$ and $\omega_0, \ldots, \omega_{n-1} \in \mathcal{A}_{\text{inv}}$. On the other hand, $\hat{v}$ is in the support of $\eta_t$, which consists of images of *ranges* $\hat{r}_s$ via matrix products $A_t^n(y)$. Thus $v \in E^s(\omega)$ would imply the existence of null words, and in particular the fact that $L_1(\underline{A_t}) = -\infty$. But those parameters $t$ form a zero measure set. ∎

We note that the non-parametric version of the CLT stated in the introduction, namely theorem 1.6, can be derived form the parametric version exactly the same way the non-parametric LDT theorem 1.5 was derived from the parametric LDT theorem 7.3. Thus theorem 1.6 holds as well.

**Remark 7.5** *The statistical properties derived above are sensitive to perturbations of the cocycle, that is, the parameters that appear in these estimates are not locally uniform.*

**Remark 7.6** *Projectively uniformly hyperbolic cocycles automatically satisfy uniform LDT estimates and a CLT, since these properties can be immediately reduced to the corresponding limit laws for i.i.d. additive processes in classical probabilities. In particular, by Theorem 6.4, if all components of* $\underline{A} \in \mathrm{Mat}_2(\mathbb{R})^k$ *are singular and* $L_1(\underline{A}) > -\infty$ *then* $\underline{A}$ *satisfies uniform LDT estimates and a CLT.*

# 8
# Extensions and Further Problems

All of the results in this manuscript were presented in the Bernoulli setting for the sake of readability of the text. They are, however, available in a more general setting, that of locally constant linear cocycles over a mixing Markov shift, which we refer to as the Markov setting.

In this chapter we intend to offer a glimpse of these other available results as well as of other possible extensions, ongoing projects and some open questions.

The chapter is divided into seven sections. In the first we formally introduce the Markov setting, or the Markov cocycles. Each subsequent section, from 8.2 to 8.6, corresponds to a previous chapter of the manuscript, from chapter 3 to chapter 7. We describe how the results in these chapters can be adapted to the more general setting of Markov cocycles. Finally, in section 8.7 we explore further extensions and describe some open questions that we think are worth being explored.

Sections 8.2 and 8.3 are based on two joint works with Cai, Klein and Melo. In 8.2 we describe a Markovian analogue of Furstenberg-Kifer's multiplicative ergodic theorem with an explicit filtration. This can be used to prove the continuity of the Lyapunov exponents for Markovian compositions of cocycles (precise definitions will be given in that section). In 8.3 we establish a joint Hölder continuity in both the cocycle and the Markov kernel for the Lyapunov exponent with respect to a suitable topology. Section 8.4 is based on a joint work with Amorim and Melo and we establish the analyticity of the Lyapunov exponent with respect to Markov kernels. Sections 8.5, 8.6 and 8.7 are based on two joint works with Duarte, Graxinha and Klein. We discuss not only the extension to the Markovian setting but also include some thoughts on how to derive similar results in $\mathrm{Mat}_2(\mathbb{R})^k$ (that is, for cocycles whose matrix components may have negative determinants), which was the original goal of this project. We conclude with other relevant related problems.

## 8.1 Markov cocycles

Consider a Markov system $(\Sigma, K, \mu)$ where $\Sigma$ is a compact metric space, $K$ is a Markov kernel and $\mu$ is a $K$-stationary measure. Let $\mathbb{P}_{K,\mu}$ denote the Markov measure on $X^+ = \Sigma^{\mathbb{N}}$ with initial distribution $\mu$ and transition kernel $K$. We use the same notation for its extension to the space $X = \Sigma^{\mathbb{Z}}$ of double sided sequences. Let $\sigma$ be the forward shift on $X^+$ and on $X$. Then $(X^+, \mathbb{P}_{K,\mu}, \sigma)$ is a measure preserving, ergodic (non invertible) dynamical system, $(X, \mathbb{P}_{K,\mu}, \sigma)$ is its natural invertible extension, both of which we call Markov shifts.

A measurable function $A\colon \Sigma \times \Sigma \to \mathrm{GL_d}(\mathbb{R})$ induces the skew-product dynamical system $F = F_{A,K}\colon X \times \mathbb{R}^d \to X \times \mathbb{R}^d$,

$$F(\omega, v) = (\sigma\omega, A(\omega_1, \omega_0)v)$$

for $\omega = \{\omega_n\}_{n\in\mathbb{Z}} \in X$ and $v \in \mathbb{R}^d$. That is, $F_{A,K}$ is a linear cocycle over the base dynamics $(X, \mathbb{P}_{K,\mu_K}, \sigma)$, where the fiber dynamics is induced by the map $A$ (which depends on two consecutive symbols). We refer to such a dynamical system as a *Markov linear cocycle*. Its iterates are given by $F^n(\omega, v) = (\sigma^n\omega, A^n(\omega)v)$, where for $\omega = \{\omega_n\}_{n\in\mathbb{Z}} \in X$,

$$A^n(\omega) = A(\omega_n, \omega_{n-1}) \cdots A(\omega_2, \omega_1)\, A(\omega_1, \omega_0)\,.$$

By Kingman's ergodic theorem, if it satisfies an integrability condition, for example $A, A^{-1}$ are bounded, then the geometric averages of the fiber iterates of the cocycle $F_{A,K}$ converge $\mathbb{P}_{K,\mu}$-a.s.

$$\frac{1}{n}\,\log\|A^n(\omega)\| \to L_1(A, K)$$

and the limit $L_1(A, K)$ is called the maximal Lyapunov exponent of the system. Replacing the norm (or largest singular value) of the iterates $A^n(\omega)$ by the other singular values, we obtain all the other Lyapunov exponents $L_2(A, K), \ldots, L_m(A, K)$ of the cocycle $F_{A,K}$.

**Remark 8.1** *Although we only consider Markov cocycles that depend on two symbols, the same results will hold for cocycles that depend on any fixed, finite number of symbols. The idea is to do an analogous argument to the one developed in chapter 2 and establish relations between the stationary measures and Markov operators on different levels.*

## 8.2 More on Furstenberg-Kifer's multiplicative ergodic theorem

Recall that Furstenberg-Kifer's filtration is "non random" in the sense that it does not depend on the point $\omega$. Oseledets filtration is completely random, since it depends on possibly all the coordinates of $\omega$. In what follows, we show a Markovian analogue of Furstenberg-Kifer's filtration that is only slightly random, since it depends only on the zeroth coordinate of $\omega$.

Let $(\Sigma, K, \mu)$ be an ergodic Markov system, that is, $\mu \in \mathrm{Prob}(\Sigma)$ is an extremal point of the set of $K$-stationary measures. Note that $\mu$ is not necessarily unique, but we fix $\mu$ and everything will depend on it.

Let $A\colon \Sigma \times \Sigma \to \mathrm{GL_d}(\mathbb{R})$ be a Borel measurable fiber map with $A, A^{-1}$ bounded. Together with $K$, it determines the Markov cocycle $F = F_{A,K}$ over the ergodic Markov shift $(\Sigma^{\mathbb{N}}, \mathbb{P}_{K,\mu}, \sigma)$. Note that the corresponding Lyapunov exponents also depend on $\mu$, so we denote them by $L_j(A, K, \mu)$, $1 \le j \le m$. Consider the observable $\psi = \psi_A \colon \Sigma \times \Sigma \times \mathbb{P} \to \mathbb{R}$ given by

$$\psi_A(\omega_1, \omega_0, \hat{v}) := \log \left\| A(\omega_1, \omega_0) \frac{v}{\|v\|} \right\| \tag{8.1}$$

where $v$ is any vector representing the projective point $\hat{v}$.

Let $\mathrm{Prob}^{\mu}_{Q_A}(\Sigma \times \mathbb{P}(\mathbb{R}^d))$ denote the set of all probability measures $\eta \in \mathrm{Prob}_{Q_A}(\Sigma \times \mathbb{P}(\mathbb{R}^d))$ that project down to $\mu$.

Define the continuous linear functional $\alpha\colon \mathrm{Prob}^{\mu}_{Q_A}(\Sigma \times \mathbb{P}(\mathbb{R}^d)) \to \mathbb{R}$

$$\alpha(\eta) := \int_{\Sigma\times\Sigma\times\mathbb{P}} \psi(\omega_1, \omega_0, \hat{v}) dK_{\omega_0}(\omega_1) d\eta(\omega_0, \hat{v})$$

and let

$$\beta(\mu) := \max\{\alpha(\eta) \colon \eta \in \mathrm{Prob}^{\mu}_{Q_A}(\Sigma \times \mathbb{P}(\mathbb{R}^d))\}\,.$$

Denote by $\Sigma^{\mu}(\alpha)$ the set of all values of the linear functional $\alpha$ over the extremal points of $\mathrm{Prob}^{\mu}_{Q_A}(\Sigma \times \mathbb{P}(\mathbb{R}^d))$. By the Krein-Milman theorem this set is nonempty. Moreover, $\beta(\mu) = \max \Sigma^{\mu}(\alpha)$.

**Theorem 8.1** *Let $(\Sigma, K, \mu)$ be an ergodic Markov system and let $A\colon \Sigma \times \Sigma \to \mathrm{GL_m}(\mathbb{R})$ be a measurable fiber map with $A, A^{-1}$ bounded. There exists a filtration $\mathcal{L} = (\mathcal{L}_0, \mathcal{L}_1, \ldots, \mathcal{L}_r)$, that is, for $0 \le j \le r$, $\mathcal{L}_j \colon \Sigma \to \mathrm{Gr}(\mathbb{R}^m)$ is a measurable section and for all $\omega_0 \in \Sigma$*

$$\{0\} = \mathcal{L}_{r+1}(\omega_0) \subsetneq \mathcal{L}_r(\omega_0) \subsetneq \ldots \subsetneq \mathcal{L}_1(\omega_0) \subsetneq \mathcal{L}_0(\omega_0) = \mathbb{R}^m$$

*such that for all indices $0 \le j \le r$,*

*(i) the section $\mathcal{L}_j$ is A-invariant,*

*(ii) for $\mu$-a.e. $\omega_0 \in \Sigma$, $\forall v \in \mathcal{L}_j(\omega_0) \setminus \mathcal{L}_{j+1}(\omega_0)$ and $\mathbb{P}_{\omega_0}$-a.e. $\omega \in \Sigma^{\mathbb{N}}$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)v\| = \beta_j(\mu) \,,$$

*(iii) if $\eta$ is an extremal point of $\mathrm{Prob}^{\mu}_{Q_A}(\Sigma \times \mathbb{P})$ with $\alpha(\eta) = \beta_j$ then $\eta\{(\omega_0, \hat{v}) \colon \omega_0 \in \Sigma,\, v \in \mathcal{L}_j(\omega_0)\} = 1$ and $\eta\{(\omega_0, \hat{v}) \colon \omega_0 \in \Sigma, v \in \mathcal{L}_{j+1}(\omega_0)\} = 0$.*

*Moreover, the filtration can be explicitly defined as*

$$\mathcal{L}_j(\omega_0) := \left\{ v \in \mathbb{R}^m \colon \limsup_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)v\| \leq \beta_j(\mu) \quad \mathbb{P}_{\omega_0}\text{-a.e. } \omega \right\} \tag{8.2}$$

*for all $\omega_0 \in \Sigma$ and $0 \leq j \leq r$.*

In [15] we were able to provide a direct proof of the Markovian filtration and describe explicitly each of its components. The existence of such filtration was already known by work of Bougerol, as a consequence of Kifer's theory of random i.i.d composition of cocycles [34, Theorem 1.2 in Chapter 3], which we will briefly describe bellow. Moreover, we will prove in 8.2 that the Markovian filtration also implies a filtration for random composition of cocycles, therefore we conclude that the theorems are in fact equivalent.

Let $M$ be a compact metric space, let $d \geq 1$ and denote by $\mathcal{G}$ the set of linear cocycles $\omega \colon M \times \mathbb{R}^d \to M \times \mathbb{R}^d$, $\omega(x, v) = (\tau_\omega(x), A_\omega(x)v)$, where the base dynamics of $\omega$ is a continuous function $\tau_\omega \colon M \to M$ and the fiber map of $\omega$ is a measurable, bounded, with bounded inverse function $A_\omega \colon M \to \mathrm{GL_d}(\mathbb{R})$.

Let $\Omega \subset \mathcal{G}$ be a compact set. Given a measure $\nu \in \mathrm{Prob}(\mathcal{G})$ with $\mathrm{supp}\nu \subset \Omega$, consider the multiplicative process in $\mathcal{G}$ (i.e. the random composition of linear cocycles)

$$\Pi^n_\omega := \omega_{n-1} \circ \dots \omega_1 \circ \omega_0, \quad n \in \mathbb{N}$$

where $\omega = \{\omega_n\}_{n\geq 0}$ is an i.i.d. sequence of random variables in $\mathcal{G}$ with common distribution $\nu$. Note that for all $(x, v) \in M \times \mathbb{R}^d$ and $n \in \mathbb{N}$,

$$\Pi^n_\omega(x, v) = (\tau^n_\omega(x), A^n_\omega(x)v) \,,$$

where

$$\begin{aligned} \tau^n_\omega &= \tau_{\omega_{n-1}\circ\dots\circ\omega_1\circ\omega_0} = \tau_{\omega_{n-1}} \circ \dots \tau_{\omega_1} \circ \tau_{\omega_0} \quad \text{and} \\ A^n_\omega(x) &= A_{\omega_{n-1}}(\tau^{n-1}_\omega x) \cdots A_{\omega_1}(\tau_{\omega_0} x)\, A_{\omega_0}(x) \,. \end{aligned}$$

Evidently $(\mathcal{G}, \circ)$ is a monoid that acts naturally on the set $M$ by $\omega_0 \cdot x := \tau_{\omega_0}(x)$. This action induces a convolution operation (or an action of

Prob($\mathcal{G}$) on Prob($M$)). More precisely, given $\nu \in \mathrm{Prob}(\mathcal{G})$ and $\mathrm{m} \in \mathrm{Prob}(M)$, their convolution $\nu * \mathrm{m} \in \mathrm{Prob}(M)$ is the push-forward of the product measure $\nu \times \mathrm{m}$ via the map $(\omega_0, x) \mapsto \omega_0 \cdot x$.

A measure $\mathrm{m} \in \mathrm{Prob}(M)$ is $\nu$-stationary if and only if $\nu * \mathrm{m} = \mathrm{m}$, which means that for all $\phi \in C(M)$

$$\int \phi(\tau_{\omega_0} x)\, d\nu(\omega_0)\, d\mathrm{m}(x) = \int \phi(x)\, d\mathrm{m}(x)\,.$$

Moreover, if m is an extremal point in the set of $\nu$-stationary measures on $M$, we say that the pair $(\nu, \mathrm{m})$ is ergodic. This can be shown to be equivalent to the ergodicity of the skew-product dynamical system $(\Omega^{\mathbb{N}} \times M, \nu^{\mathbb{N}} \times \mathrm{m}, f)$, where $f(\omega, x) := (\sigma\omega, \tau_{\omega_0}(x))$.

Next we show that Theorem 8.1 implies the existence of a non-random filtration for i.i.d. compositions of linear cocycles, which provides a different argument for [34, Theorem 1.2 in Chapter 3].

**Theorem 8.2** *Let $\mathcal{G}$ be a set of linear cocycles on $M \times \mathbb{R}^d$ as above, let $\nu \in \mathrm{Prob}_c(\mathcal{G})$, let $\mathrm{m} \in \mathrm{Prob}(M)$ and assume that $(\nu, \mathrm{m})$ is ergodic. There are an integer $0 \leq r \leq d$, real numbers $\beta_0 > \beta_1 > \ldots > \beta_r$ that depend on $\nu$ and* m, *and a measurable filtration $\mathcal{L} = (\mathcal{L}_0, \mathcal{L}_1, \ldots, \mathcal{L}_r)$, that is, $\mathcal{L}_j \colon M \to \mathrm{Gr}(\mathbb{R}^d)$ are measurable maps with*

$$\{0\} = \mathcal{L}_{r+1}(x) \subsetneq \mathcal{L}_r(x) \subsetneq \ldots \subsetneq \mathcal{L}_1(x) \subsetneq \mathcal{L}_0(x) = \mathbb{R}^d$$

*such that for all $0 \leq j \leq r$ and for* m*-a.e. $x \in M$, the following hold.*

*(i) $A_{\omega_0}(x)\, \mathcal{L}_j(x) = \mathcal{L}_j(\tau_{\omega_0}(x))$ for $\nu$-a.e. $\omega_0 \in \mathcal{G}$.*

*(ii) For all $v \in \mathcal{L}_j(x) \setminus \mathcal{L}_{j+1}(x)$ and $\nu^{\mathbb{N}}$-a.e. $\omega \in \Sigma^{\mathbb{N}}$,*

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

*(iii) For $\nu^{\mathbb{N}}$-a.e. $\omega \in \Sigma^{\mathbb{N}}$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n_\omega(x)\| = \beta_0\,.$$

*Proof.* We associate to the multiplicative i.i.d. process $\Pi^n_\omega$ a Markov linear cocycle as follows.

Let $\Sigma := \Omega \times M$ and consider the Markov chain in $\Sigma$

$$(\omega_0, x) \to (\omega_1, \tau_{\omega_0} x) \to \ldots \to (\omega_n, \tau^n_\omega x) \to \ldots$$

with transition kernel $K\colon \Sigma \to \mathrm{Prob}(\Sigma)$, $K_{(\omega_0,x)} := \nu \times \delta_{\tau_{\omega_0}x}$.

Since m is $\nu$-stationary, it follows that $\mu := \nu \times \mathrm{m}$ is $K$-stationary. Moreover, since $(\nu, \mathrm{m})$ is ergodic, namely m is an extremal point in the set of $\nu$-stationary measures on $M$, then $\mu$ is extremal also which shows that $(K, \mu)$ is an ergodic Markov system on $\Sigma$.

Note that the "admissible" sequences in $(\Omega \times M)^{\mathbb{N}}$ are the elements of the set

$$X := \Big\{\{(\omega_n, \tau_{\omega_{n-1}} \circ \dots \tau_{\omega_1} \circ \tau_{\omega_0} x)\}_{n\in\mathbb{N}} \colon \omega_n \in \Omega \ \forall n \in \mathbb{N},\, x \in M\Big\}$$

in the sense that $\mathbb{P}_{K,\mu}(X) = 1$ and consequently $\mathbb{P}_{(\omega_0,x)}(X) = 1$ for $\nu$-a.e. $\omega_0 \in \mathcal{G}$ and m-a.e. $x \in M$.

Define the fiber map (depending on one variable) $A\colon \Sigma \to \mathrm{GL_d}(\mathbb{R})$, $A(\omega_0, x) := A_{\omega_0}(x)$. Then for all points $(\omega, \underline{x}) = \{(\omega_n, \tau^n_\omega x)\}_{n\geq 0} \in X$, the iterates of the fiber map are

$$A^n(\omega, \underline{x}) = A_{\omega_{n-1}}(\tau_\omega^{n-1} x) \cdots A_{\omega_1}(\tau_{\omega_0} x)\, A_{\omega_0}(x) = A^n_\omega(x)\,.$$

Theorem 8.1 applied to the corresponding Markov linear cocycle $F_{A,K}$ implies the existence of $r \in \mathbb{N}$, $\beta_0 > \beta_1 > \ldots > \beta_r$ and, for every index $0 \leq j \leq r$, of a measurable section $\mathcal{L}_j \colon \Omega \times M \to \mathrm{Gr}(\mathbb{R}^d)$ with the stated properties, given (see (8.2)) by

$$\begin{aligned}\mathcal{L}_j(\omega_0, x) &= \left\{v \in \mathbb{R}^m \colon \limsup_{n\to\infty} \frac{1}{n} \log\|A^n(\omega, \underline{x})v\| \leq \beta_j \ \mathbb{P}_{(\omega_0,\underline{x})} \text{ a.e.}(\omega, \underline{x})\right\}\\ &= \left\{v \in \mathbb{R}^m \colon \limsup_{n\to\infty} \frac{1}{n} \log\|A^n_\omega(x)v\| \leq \beta_j \ \ \mathbb{P}_{\omega_0} \text{ a.e. } \omega\right\}.\end{aligned}$$

Since $\omega_0, \omega_1, \dots \omega_{n-1}, \dots$ are chosen independently and according to the same distribution $\nu$, it follows that $\mathcal{L}_j(\omega_0, x)$ does not depend on $\omega_0$ $\nu$-almost surely, that is, $\mathcal{L}_j(\omega_0, x) = \mathcal{L}_j(x)$ for $\nu$-almost every $\omega_0$. Items (i) and (ii) are consequences of the corresponding statements in Theorem 8.1, while item (iii) follows from the Markovian analogue of Theorem 3.3. ■

In an ongoing project, we are investigating the Markovian analogue of the theory of i.i.d random composition of cocycles: the Markovian composition of cocycles. Therefore, we consider an ergodic Markov system $(\Omega, K, \mu)$ and the points $\omega_0, \omega_1, \dots$ are chosen accordingly to it. One of the goals is to establish the continuity of the Lyapunov exponents under some irreducibility hypothesis to Markovian composition of cocycles.

## 8.3 Hölder continuity in the generic setting

The goal of this section is to present a Markovian analogue of theorem 4.1. We study the continuity of the maximal Lyapunov exponent of Markov cocycles $F_{A,K}$ as a function of the fiber map $A$ and the transition kernel $K$.

Let $\Sigma$ be a compact metric space. We assume that the Markov kernel $K$ is *uniformly ergodic*, meaning that $K_x^n$ converges to its stationary measure $\mu$ uniformly (in $x \in \Sigma$) relative to the total variation distance. In this case the convergence is necessarily exponential and the $K$-stationary measure $\mu = \mu_K$ is unique. Let $Q_K \colon L^\infty(\Sigma) \to L^\infty(\Sigma)$,

$$Q_K \phi(x) := \int_\Sigma \phi(y)\, dK_x(y)$$

be the corresponding Markov operator. It turns out that the uniform ergodicity of the transition kernel $K$ is equivalent to the existence of constants $C < \infty$ and $c > 0$ such that

$$\left\| Q_K^n \varphi - \int \varphi d\mu \right\|_\infty \leq C\, e^{-cn} \|\varphi\|_\infty \quad \forall \varphi \in L^\infty(\Sigma), \forall n \in \mathbb{N}\,,$$

in which case we refer to $K$ as being $(C, c)$-uniformly ergodic.[1]

In the Markov setting, irreducibility refers to the non-existence of a proper, $A$-invariant section, that is, of a measurable function $\mathcal{L} \colon \Sigma \to \mathrm{Gr}(\mathbb{R}^d)$ (here $\mathrm{Gr}(\mathbb{R}^d)$ denotes the Grassmannian of $\mathbb{R}^d$) such that $0 < \dim(V) < d$ and

$$A(\omega_{n+1}, \omega_n)\mathcal{L}(\omega_n) = \mathcal{L}(\omega_{n+1}), \text{ for } \mathbb{P}_{K,\mu_K}\text{-a.e. } \omega = \{\omega_n\}_n\,.$$

Quasi-irreducibility is a weaker version of this property, where such a proper $A$-invariant section $\mathcal{L}$ may exist, but in this case, the maximal Lyapunov exponent of the fiber restriction of the cocycle $F_{A,K}$ to $\mathcal{L}$ must equal $L_1(A, K)$.

Given $M < \infty$, a fiber map $A \colon \Sigma \times \Sigma \to \mathrm{GL_d}(\mathbb{R})$ is $M$-Lipschitz continuous if for all $(\omega_1, \omega_0), (z_1, z_0) \in \Sigma \times \Sigma$,

$$\left| A(\omega_1, \omega_0) - A(z_1, z_0) \right| \leq M\, (d(\omega_1, z_1) + d(\omega_0, z_0))\,,$$

while a transition kernel $K \colon \Sigma \to \mathrm{Prob}(\Sigma)$ is $M$-Lipschitz continuous with respect to the Wasserstein distance $W_1$ if for all $x, y \in \Sigma$,

$$W_1(K_x, K_y) \leq M d(x, y).$$

[1]For these and other characterizations of uniform ergodicity see [36, Theorem 16.0.2].

Fix some constants $M, C < \infty$ and $c > 0$ and consider the set $\mathcal{C} = \mathcal{C}(M, C, c)$ of Markov linear cocycles with $M$-Lipschitz data and $(C, c)$-uniformly ergodic transition kernel, that is,

$$\mathcal{C} := \Big\{(A, K)\colon A \text{ is } M\text{-Lipschitz}, K \text{ is } M\text{-Lipschitz}, \\ K \text{ is } (C, c)\text{-uniformly ergodic}\Big\}.$$

We endow this set with the distance

$$d\left((A, K), (B, T)\right) := d_\infty(A, B) + d_{W_1}(K, T),$$

where $d_{W_1}$ is the distance between kernels induced by the Wasserstein distance, namely

$$d_{W_1}(K, T) := \sup_{x \in \Sigma} W_1(K_x, T_x).$$

**Theorem 8.3** *Let $(A, K) \in \mathcal{C}$ be a Markov cocycle. Assume that $(A, K)$ is quasi-irreducible and that $L_1(A, K) > L_2(A, K)$. Then locally near $(A, K)$, the map $\mathcal{C} \ni (B, T) \mapsto L_1(B, T)$ is Hölder continuous.*

A proof of this result is given in [15].

## 8.4 Analyticity of the Lyapunov exponent with respect to Markov kernels

In the previous section we presented a result about the joint regularity of $(A, K) \mapsto L_1(A, K)$. During that work, it became clear that the main obstruction to obtain a good modulus of continuity was to control $A$. In this section we introduce the Markovian analogue of theorem 5.2. That theorem shows that the Lyapunov exponent admits a high regularity with respect to the measure in the total variation norm. We now present a result that shows that it also admits a high regularity with respect to the Markov kernel in a suitable norm. Similarly to chapter 5, where we considered complex valued measures, we consider complex Markov kernels $K\colon \Sigma \to \mathcal{M}(\Sigma)$.

Denote by $\mathcal{K}(\Sigma)$ the set of continuous and complex Markov kernels over $\Sigma$ such that for every $\omega \in \Sigma$, $K_\omega(\Sigma)$ has bounded variation. Consider the following norm in $\mathcal{K}(\Sigma)$:

$$\|K\| := \sup_{\omega \in \Sigma} \|K_\omega\|_{T.V}.$$

Given two kernels $K$ and $L$, we may consider a partial order relation between their supports. We say that $\operatorname{supp}(L) \leq \operatorname{supp}(K)$ if $\operatorname{supp}(L_\omega) \subset \operatorname{supp}(K_\omega)$ for every $\omega \in \Sigma$. For any fixed kernel $K$, we may define the set of continuous kernels whose supports are contained in the support of $K$ as the following:

$$\mathcal{S}(K) := \{L \in \mathcal{K}(\Sigma) \colon \operatorname{supp}(L) \leq \operatorname{supp}(K)\}.$$

Moreover, we call a transition kernel $K$ quasi-irreducible if the corresponding Markov cocycle $F_{A,K}$ is quasi-irreducible.

**Theorem 8.4** *Let $K_0$ be a uniformly ergodic kernel on $\Sigma$ such that $L_1(K_0) > L_2(K_0)$.*

*(1) If $K_0$ is quasi-irreducible, then $K \mapsto L_1(K)$ is real analytic in a neighbourhood of $K_0$.*

*(2) The map $K \mapsto L_1(K)$ is real analytic in a neighbourhood of $K_0$ in $\mathcal{S}(K_0)$.*

The main ideas in the proof of the Markov setting are analogous to the ones that appear in the i.i.d case, but we need to use tools from [15] and [21] to deal with it. A full proof can be found on [1].

## 8.5 Regularity dichotomy in the singular setting

In this section we consider linear cocycles over a Markov shift and we explain how to extend the results in chapter 6 to this setting.

Let $P \in \operatorname{Mat}_k(\mathbb{R})$ be a (left) stochastic matrix, $P = (p_{ij})_{1 \leq i,j \leq k}$ with $p_{ij} \geq 0$ and $\sum_{i=1}^{k} p_{ij} = 1$ for all $1 \leq j \leq k$. Given a $P$-stationary probability vector $q = P\,q$, the pair $(P, q)$ determines a probability measure $\mu$ on $X$ for which the process $\xi_n \colon X \to \mathcal{A}$, $\xi_n(\omega) := \omega_n$ is a stationary Markov chain in $\mathcal{A}$ with probability transition matrix $P$ and initial distribution law $q$. Then $(X, \sigma, \mu)$ is a measure preserving dynamical system.

The stochastic matrix $P$ determines a directed weighted graph on the vertex set $\mathcal{A}$ with an edge $j \mapsto i$, from $j$ to $i$, whenever $p_{ij} > 0$. Sequences in $X$ describing paths in this graph are called $P$-admissible. The set $X_P$ of all $P$-admissible sequences is the support of the Markov measure $\mu$. Given a finite admissible word $(i_0, i_1, \ldots, i_n) \in \mathcal{A}^{n+1}$ and $k \in \mathbb{Z}$, the set

$$[k; i_0, i_1, \ldots, i_n] := \{\omega \in X_P \colon \omega_{k+l} = i_l \quad \text{for all } 0 \leq l \leq n\}$$

is called a cylinder of length $n+1$ in $X_P$. Its (Markov) measure is then

$$\mu\left([k; i_0, i_1, \ldots, i_n]\right) = q_{i_0}\, p_{i_1, i_0} \cdots p_{i_n, i_{n-1}}\,.$$

We will assume that the matrix $P$ is primitive, i.e. there exists $n \geq 1$ such that $p_{ij}^n > 0$ for all entries of the power matrix $P^n$. Then $\lim_{n\to\infty} p_{ij}^n = q_i > 0$ for all $1 \leq i, j \leq k$ and the corresponding Markov shift $(X, \mu, \sigma)$ is ergodic and mixing. As before, a $k$-tuple $\underline{A} = (A_1, \ldots, A_k) \in \mathrm{Mat}_2(\mathbb{R})^k$ determines a locally constant linear cocycle over this base dynamics, which we refer to as a random Markov cocycle. Its first Lyapunov exponent is denoted by $L_1(\underline{A}) = L_1(\underline{A}, P, q)$.

Consider a random Markov linear cocycle $F\colon X_P \times \mathbb{R}^2 \to X_P \times \mathbb{R}^2$ determined by the data $(P, q, \underline{A})$ where $\underline{A} := (A_i)_{i\in\mathcal{A}} \in \mathrm{Mat}_2^+(\mathbb{R})^k$.

By a slight abuse of notation, in the Markov setting we also write $\hat{A}_i\colon \mathcal{A} \times \mathbb{P}(\mathbb{R}^2) \to \mathcal{A} \times \mathbb{P}(\mathbb{R}^2)$ to denote, for each $i \in \mathcal{A}$, the non-invertible map $\hat{A}_i(j, \hat{v}) := (i, \hat{A}_i\, \hat{v})$.

In what follows we introduce a series of analogous definitions to the main concepts from chapter 6. All results proven in the Bernoulli setting in chapter 6 extend to the Markov setting, using the definitions below and ideas similar to the ones presented above, properly adapted to this more general setting.

**Definition 8.1** *An invariant multi-cone for $\underline{A}$ is a set $M \subset \mathcal{A} \times \mathbb{P}(\mathbb{R}^2)$ that satisfies the following properties*

*(1) $M$ is an open subset of $\mathcal{A} \times \mathbb{P}(\mathbb{R}^2)$,*

*(2) the closure of each fiber of $M$ is a proper subset of $\{i\} \times \mathbb{P}(\mathbb{R}^2)$,*

*(3) $\hat{A}_i M \Subset M$ for every $i \in \mathcal{A}$ .*

**Definition 8.2** *For $i \in \mathcal{A}_{\mathrm{sing}}$, consider $r_i := \mathrm{Range}(A_i)$ and $k_i := \mathrm{Ker}(A_i)$ and set*

$$\mathcal{K}(\underline{A}) := \{(j, k_i)\colon i \in \mathcal{A}_{\mathrm{sing}}, j \in \mathcal{A}\}\ ,$$
$$\mathcal{R}(\underline{A}) := \{(i, r_i)\colon i \in \mathcal{A}_{\mathrm{sing}}\}.$$

The complement $\mathcal{A} \times \mathbb{P}(\mathbb{R}^2) \setminus \mathcal{K}(\underline{A})$ is the common domain of all maps $\hat{A}_i$ with $i \in \mathcal{A}_{\mathrm{sing}}$, while $\mathcal{R}(\underline{A})$ is the union of the ranges of these same maps.

**Definition 8.3** *Given $i \in \mathcal{A}_{\mathrm{sing}}$, a branch departing from $i$ is any $P$-admissible word $\underline{\omega} = (\omega_0, \omega_1, \cdots, \omega_n) \in \mathcal{A}^{n+1}$ with $n \geq 0$ such that $\omega_0 = i$ and $\omega_l \in \mathcal{A}_{\mathrm{inv}}$ for all $l = 1, \ldots, n$. Similarly, a branch arriving at $i$ is a $P$-admissible word $\underline{\omega} = (\omega_0, \omega_1, \cdots, \omega_n) \in \mathcal{A}^{n+1}$ with $n \geq 0$ such that $\omega_n = i$ and $\omega_l \in \mathcal{A}_{\mathrm{inv}}$ for all $l = 0, \ldots, n-1$.*

Denote by $\mathcal{B}_n^+(i)$ the set of all branches $\underline{\omega} = (\omega_0, \omega_1, \cdots, \omega_n) \in \mathcal{A}^{n+1}$ departing from $i$ and by $\mathcal{B}_n^-(i)$ the set of all branches $\underline{\omega} = (\omega_0, \omega_1, \cdots, \omega_n) \in \mathcal{A}^{n+1}$ arriving at $i$. For $\underline{\omega} \in \mathcal{B}_n^+(i)$ we write $A^{n-1}(\underline{\omega}) := A_{\omega_{n-1}} \ldots A_{\omega_2} A_{\omega_1}$ while for $\underline{\omega} \in \mathcal{B}_n^-(i)$ we write $A^{-(n-1)}(\underline{\omega}) := A_{\omega_1}^{-1} A_{\omega_2}^{-1} \ldots A_{\omega_{n-1}}^{-1} = (A_{\omega_{n-1}} \ldots A_{\omega_2} A_{\omega_1})^{-1}$. These matrices are invertible by definition of a branch.

**Definition 8.4** *The subset $\mathcal{W}^+ \subset \mathcal{A} \times \mathbb{P}(\mathbb{R}^2)$ is defined to be the closure of the union of the ranges of compositions of the partial maps $\hat{A}_{\omega_l}$ along all branches departing from $i$ while the subset $\mathcal{W}^- \subset \mathcal{A} \times \mathbb{P}(\mathbb{R}^2)$ is the closure of the union of the pre-images of $0$ under compositions of the partial maps $\hat{A}_{\omega_l}$ along all branches arriving at $i$.*

In the Markov case, $K^u_{\text{inv}}$ and $K^s_{\text{inv}}$ are subsets of $\mathcal{A} \times \mathbb{P}(\mathbb{R}^2)$. Firstly, the multi-cone is to be interpreted according to definition 8.1 and $\mathbb{P}(\mathbb{R}^2) \setminus \overline{M}$ replaced by $\mathcal{A} \times \mathbb{P}(\mathbb{R}^2) \setminus \overline{M}$. Secondly, the matrix products $A_{i_n} \cdots A_{i_1}$ are to be substituted by composition of the maps $\hat{A}_{i_n} \circ \cdots \circ \hat{A}_{i_1}$ along admissible invertible words.

## 8.6
## Statistical properties in the singular setting

The cocycle $(P, q, \underline{A})$ determines the operator $\mathcal{Q}\colon L^\infty(\mathcal{A} \times \mathbb{P}(\mathbb{R}^2)) \to L^\infty(\mathcal{A} \times \mathbb{P}(\mathbb{R}^2))$ defined by

$$\begin{aligned}(\mathcal{Q}\varphi)(j, \hat{v}) &:= \sum_{i\in\mathcal{A}} \varphi(i, \hat{A}_i\, \hat{v})\, p_{ij} \\ &= \sum_{i\in\mathcal{A}_{\text{inv}}} \varphi(i, \hat{A}_i\, \hat{v})\, p_{ij} + \sum_{i\in\mathcal{A}_{\text{sing}}} \varphi(i, \hat{r}_i)\, p_{ij}.\end{aligned}$$

Let $\pi\colon \mathcal{A} \times \mathbb{P}(\mathbb{R}^2) \to \mathcal{A}$ denote the canonical projection in the first coordinate. If $\eta \in \text{Prob}(\mathcal{A} \times \mathbb{P}(\mathbb{R}^2))$ is $\mathcal{Q}$-stationary, then its push-forward measure via $\pi$ is the $P$-stationary measure $q$ on $\mathcal{A}$, that is, $\pi_*\eta = q$.

Since $\eta$ projects down via $\pi$ to $q$, we can consider its disintegration $\{\eta_i\}_{i\in\mathcal{A}} \subset \text{Prob}(\mathbb{P}(\mathbb{R}^2))$, which is characterized by

$$\int_{\mathcal{A}\times\mathbb{P}(\mathbb{R}^2)} \varphi(i, \hat{v})\, d\eta(i, \hat{v}) = \sum_{i\in\mathcal{A}} q_i \int_{\mathbb{P}(\mathbb{R}^2)} \varphi(i, \hat{v})\, d\eta_i(\hat{v}) \quad \forall \varphi \in L^\infty(\mathcal{A} \times \mathbb{P}(\mathbb{R}^2))\,.$$

Then

$$\eta = \sum_{i\in\mathcal{A}} q_i\, \delta_i \times \eta_i \quad \text{and} \quad \eta_i(E) = \frac{1}{q_i}\, \eta(\{i\} \times E)\ \forall i \in \mathcal{A}, E \subset \mathbb{P}(\mathbb{R}^2) \text{ Borel.}$$

Given $n \geq 1$, $s \in \mathcal{A}_{\text{sing}}$ and $l \in \mathcal{A}$, let $\mathcal{B}_n(s, l)$ denote the set of admissible words $\underline{\omega} = (\omega_0, \ldots, \omega_n)$ of length $n+1$ such that $\omega_0 = s$, $\omega_n = l$ and $\omega_i \in \mathcal{A}_{\text{inv}}$ for all $i = 1, \ldots, n-1$. For such a word we write $A^n(\underline{\omega}) := A_{\omega_n} \ldots A_{\omega_2} A_{\omega_1}$ and also $p(\underline{\omega}) := p_{\omega_n \omega_{n-1}} \cdots p_{\omega_1 \omega_0}$.

With these notations, we have the following explicit formula for a (or, a-posteriori, *the unique*) $\mathcal{Q}$-stationary measure.

**Proposition 8.1** *Let* $\eta = \sum_{j\in\mathcal{A}} q_j \delta_j \times \eta_j$ *where for all* $j \in \mathcal{A}$,

$$\eta_j := \frac{1}{q_j} \sum_{s\in\mathcal{A}_{\text{sing}}} q_s \sum_{n=1}^{\infty} \sum_{\underline{\omega}\in\mathcal{B}_n(s,j)} p(\underline{\omega})\delta_{\hat{A}^n(\underline{\omega})\hat{r}_s}. \tag{8.3}$$

*Then* $\eta$ *is a* $(P, q, \underline{A})$*-stationary probability measure on* $\mathcal{A} \times \mathbb{P}(\mathbb{R}^2)$.

Moreover, with this stationary measure, we are able to establish analogous theorems to all of the results in chapter 7. In particular, we describe the Furstenberg's Formula for the Markov setting.

Consider the following observable $\Psi\colon \mathcal{A} \times \mathbb{P}(\mathbb{R}^2) \to [-\infty, \infty)$,

$$\Psi(j, \hat{v}) = \sum_{i\in\mathcal{A}} p_{ij} \log \frac{\|A_i\, v\|}{\|v\|}.$$

**Theorem 8.5 (Furstenberg's Formula)** *If the cocycle* $(P, q, \underline{A})$ *has rank one then*

$$L_1(\underline{A}) = \int \Psi \, d\eta.$$

All the proofs follow in a generally similar manner to the ones presented in chapter 7. However, some of them are much more technical, which is why we preferred to write the text in the Bernoulli context. In particular, the proofs of the LDT estimates and of the CLT in the singular setting for Markov cocycles also follow a similar scheme to the one shown above, but using the stationary measure described by the more complex expression (8.3).

## 8.7
## More about the singular setting

In this section we address other extensions and questions related to the study of linear cocycles in the singular setting.

We start by explaining some technical difficulties that made us chose to work with matrices in $\text{Mat}_2^+(\mathbb{R})$ instead of $\text{Mat}_2(\mathbb{R})$ in chapters 6 and 7. This assumption was made for different reasons in each chapter. Regarding chapter 7, we only introduced this hypothesis in order to use the winding property so that we could estimate how much time the orbit of a range stays close to a kernel. Therefore, it is merely a technical assumption which we believe that it could be removed by using a different approach. The one we plan to use is based more heavily on the properties of subharmonic functions (arising from considering analytic curves of cocycles) rather than on the positive winding property, which limits us to $\text{Mat}_2^+(\mathbb{R})$-valued cocycles.

On the other hand, a great part of the work in chapter 6 was to extend Avila, Bochi, Yoccoz (ABY) theory from $\mathrm{SL}_2(\mathbb{R})$ to $\mathrm{Mat}_2^+(\mathbb{R})$. In order to extend this theory to $\mathrm{Mat}_2(\mathbb{R})$-valued cocycles, the strategy would be to first extend the results of Avila, Bochi, Yoccoz to cocycles with matrices in $\mathrm{SL}_2'(\mathbb{R}) := \{A \in \mathrm{GL}_2(\mathbb{R})\colon \det \mathrm{A} = \pm 1\}$ and then use our approach to further extend from $\mathrm{SL}_2'(\mathbb{R})$-valued to $\mathrm{Mat}_2(\mathbb{R})$-valued cocycles.

We believe that some of the results from this text, especially the local ones, such as the multi-cone criteria, should not be difficult to be extended to $\mathrm{SL}_2'(\mathbb{R})$-valued cocycles. However, there are other results that rely on a global understanding of the boundaries of the set of uniformly hyperbolic ($\mathcal{UH}$) cocycles and of the set of elliptic ($\mathcal{E}$) cocycles. These type of problems are still far from being completely understood.

One example of such a result proven in [44] for $\mathrm{SL}_2(\mathbb{R})$-valued cocycles, says that the complement of the uniformly hyperbolic cocycles is exactly the closure of the elliptic cocycles: $\mathcal{UH}^c = \bar{\mathcal{E}}$. It is an open question whether this result still holds in $\mathrm{SL}_2'(\mathbb{R})$ or not. Therefore we propose the following questions.

**Question 1.** Is every random linear cocycle with rank 1 in $\mathrm{Mat}_2(\mathbb{R})$ either projectively uniformly hyperbolic or approximated by cocycles with null words (in particular with Lyapunov exponent $-\infty$)?

A weaker version of this question is the following.

**Question 1'.** Is Lebesgue *almost every* random linear cocycle with rank 1 in $\mathrm{Mat}_2(\mathbb{R})$ either projectively uniformly hyperbolic or approximated by cocycles with null words (in particular with Lyapunov exponent $-\infty$)?

Some related, potentially useful questions are the following.

**Question 2.** Is it true that in the space of $\mathrm{SL}_2'(\mathbb{R})$-valued cocycles we have $\mathcal{UH}^c = \bar{\mathcal{E}}$?

**Question 2'.** Is it true that in the space of $\mathrm{SL}_2'(\mathbb{R})$-valued cocycles, the identity $\mathcal{UH}^c = \bar{\mathcal{E}}$ holds at least modulo a Lebesgue zero measure set of such cocycles?

Moreover, in [44] Yoccoz asks whether $\partial\mathcal{UH} = \partial\mathcal{E}$ or not. This is proved to be true in the case of the full shift in two symbols, see [2, Theorem 3.3]. However, the answer to this question is negative for the full shift with $N \geq 3$ symbols, see [32] and [17]. The type of counter example provided in both papers shows that there are elements from $\partial\mathcal{E}$ which are not in $\partial\mathcal{UH}$. An interesting question that could lead to some progress regarding questions 2 and 2' is the following.

**Question 3.** Consider the set of random $\mathrm{SL}_2'(\mathbb{R})$-valued cocycles. Is there any cocycle that belongs to $\partial\mathcal{UH}$ but not to $\partial\mathcal{E}$?

A further extension would be regarding the support of the measure. In this work, we only consider linear cocycles over a finitely supported measure in order to use ABY theory of multi-cones. Since that result is available only for a finite number of matrices, we followed this direction. However, we believe that if one could extend the ABY theory to infinitely supported measures, then it would also be possible to obtain a similar theory to the one in this manuscript for singular cocycles with infinitely supported measure.

**Question 4.** Does theorem 6.3 still hold for infinitely supported measures?

Another direction that we would like to investigate in the future is the study of singular cocycles in higher dimensions. Some steps in this direction are already available, since in [10], Bochi and Gourmelon developed an analogous theory of multi-cones in higher dimensions. Moreover, recently, Avila, Eskin and Viana proved the continuity of the Lyapunov exponents for random linear cocycles of dimension $d$ for invertible matrices, i.e., with full rank. We would like to understand the regularity of the Lyapunov exponents for random linear cocycles of dimension $d$ and rank $k < d$.

In dimension 2, for random, locally constant linear cocycles with a finitely supported measure we now have an almost complete understanding of the regularity of the Lyapunov exponent. When the rank $k = 2$, it is at least continuous due to [11], with several results in the generic setting for intermediate modulus of continuity. When the $k = 1$, the present work shows that it satisfies an almost everywhere dichotomy between analyticity and discontinuity (at least in $\mathrm{Mat}_2^+(\mathbb{R})$).

All the other cases for $d \geq 3$ and $k < d$ are still not known and we believe that they should present another type of behavior. Although these cocycles do not have full rank, our best guess now would be that there are cases where the Lyapunov exponent still has some good intermediate regularity, differently from the dichotomic behavior observed in dimension 2. Therefore, we complete this work with one last question.

**Question 5.** Study the continuity properties of the Lyapunov exponents of random cocycles in dimension $d$ and rank $k$.

# Bibliography


[1] Artur Amorim, Marcelo Durães, and Aline Melo. Analiticity of the lyapunov exponents of random products of matrices, 2025.

[2] Artur Avila, Jairo Bochi, and Jean-Christophe Yoccoz. Uniformly hyperbolic finite-valued $\mathrm{SL}(2,\mathbb{R})$-cocycles. *Comment. Math. Helv.*, 85(4):813–884, 2010.

[3] Artur Avila, Alex Eskin, and Marcelo Viana. Continuity of the lyapunov exponents of random matrix products, 2023.

[4] Alexandre Baraviera and Pedro Duarte. Approximating Lyapunov exponents and stationary measures. *J. Dynam. Differential Equations*, 31(1):25–48, 2019.

[5] Pablo G. Barrientos and Dominique Malicet. Mostly contracting random maps. Preprint, arXiv:2412.03729 [math.DS] (2024), 2024.

[6] Yves Benoist and Jean-François Quint. Central limit theorem for linear groups. *Ann. Probab.*, 44(2):1308–1340, 2016.

[7] Jamerson Bezerra, Ao Cai, Pedro Duarte, Catalina Freijo, and Silvius Klein. A dynamical Thouless formula. *Adv. Math.*, 438:50, 2024. Id/No 109446.

[8] Jamerson Bezerra and Mauricio Poletti. Random product of quasi-periodic cocycles. *Proc. Amer. Math. Soc.*, 149(7):2927–2942, 2021.

[9] Jairo Bochi. Genericity of zero Lyapunov exponents. *Ergodic Theory Dynam. Systems*, 22(6):1667–1696, 2002.

[10] Jairo Bochi and Nicolas Gourmelon. Some characterizations of domination. *Math. Z.*, 263(1):221–231, 2009.

[11] Carlos Bocker-Neto and Marcelo Viana. Continuity of Lyapunov exponents for random two-dimensional matrices. *Ergodic Theory and Dynamical Systems*, pages 1–30, 2016.

[12] Philippe Bougerol. Théorèmes limite pour les systèmes linéaires à coefficients markoviens. *Probab. Theory Related Fields*, 78(2):193–221, 1988.

[13] Ao Cai, Pedro Duarte, and Silvius Klein. Statistical properties for mixing Markov chains with applications to dynamical systems. Preprint, arXiv:2210.16908 [math.DS] (2022), 2022.

[14] Ao Cai, Pedro Duarte, and Silvius Klein. Furstenberg theory of mixed random-quasiperiodic cocycles. *Commun. Math. Phys.*, 402(1):447–487, 2023.

[15] Ao Cai, Marcelo Durães, Silvius Klein, and Aline Melo. Hölder continuity of the lyapunov exponent for markov cocycles via furstenberg's formula, 2022.

[16] Soo Bong Chae. *Holomorphy and calculus in normed spaces*, volume 92 of *Pure Appl. Math., Marcel Dekker*. Marcel Dekker, Inc., New York, NY, 1985.

[17] Argyrios Christodoulou. Parameter spaces of locally constant cocycles. *Int. Math. Res. Not.*, 2022(17):13590–13628, 2022.

[18] Pedro Duarte, Marcelo Durães, Tomé Graxinha, and Silvius Klein. Random 2d linear cocycles I: dichotomic behavior. Preprint, arXiv:2503.21050 [math.DS] (2025), 2025.

[19] Pedro Duarte, Marcelo Durães, Tomé Graxinha, and Silvius Klein. Random 2d linear cocycles II: statistical properties. Preprint, arXiv:2505.00146 [math.DS] (2025), 2025.

[20] Pedro Duarte and Catalina Freijo. Continuity of the Lyapunov exponents of non-invertible random cocycles with constant rank. *Nonlinearity*, 37(10):Paper No. 105008, 23, 2024.

[21] Pedro Duarte and Silvius Klein. *Lyapunov exponents of linear cocycles; continuity via large deviations*, volume 3 of *Atlantis Studies in Dynamical Systems*. Atlantis Press, 2016.

[22] Pedro Duarte and Silvius Klein. *Continuity of the Lyapunov exponents of linear cocycles*. Publicações Matemáticas do IMPA. [IMPA Mathematical Publications]. Instituto Nacional de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, 2017. 31o Colóquio Brasileiro de Matemática.

[23] Pedro Duarte and Silvius Klein. Continuity, positivity and simplicity of the Lyapunov exponents for quasi-periodic cocycles. *J. Eur. Math. Soc. (JEMS)*, 21(7):2051–2106, 2019.

[24] Pedro Duarte and Silvius Klein. Large deviations for products of random two dimensional matrices. *Comm. Math. Phys.*, 375(3):2191–2257, 2020.

[25] Pedro Duarte, Silvius Klein, and Manuel Santos. A random cocycle with non Hölder Lyapunov exponent. *Discrete Contin. Dyn. Syst.*, 39(8):4841–4861, 2019.

[26] R. Durret. *Probability: theory and examples*. Wadsworth & Brooks Publ., 1996. 2nd edition.

[27] Marcelo Durães. Hölder continuity for Lyapunov exponents of random linear cocycles. Master's thesis, Pontifícia Universidade Católica do Rio de Janeiro, 2021.

[28] H. Furstenberg and H. Kesten. Products of random matrices. *Ann. Math. Statist.*, 31:457–469, 1960.

[29] H. Furstenberg and Yu. Kifer. Random matrix products and measures in projective spaces. *Israel J. Math*, 10:12–32, 1983.

[30] Michael Goldstein and Wilhelm Schlag. Fine properties of the integrated density of states and a quantitative separation property of the Dirichlet eigenvalues. *Geom. Funct. Anal.*, 18(3):755–869, 2008.

[31] M. I. Gordin and B. A. Lifshits. The central limit theorem for stationary Markov processes. *Sov. Math., Dokl.*, 19:392–394, 1978.

[32] Matthew Jacques and Ian Short. Semigroups of isometries of the hyperbolic plane. *Int. Math. Res. Not.*, 2022(9):6403–6463, 2022.

[33] Yuri Kifer. Perturbations of random matrix products. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 61(1):83–95, mar 1982.

[34] Yuri Kifer. *Ergodic theory of random transformations*. Progress in probability and statistics. Birkhäuser, 1986.

[35] Elaís C. Malheiro and Marcelo Viana. Lyapunov exponents of linear cocycles over Markov shifts. *Stoch. Dyn.*, 15(3):1550020, 27, 2015.

[36] Sean Meyn, Richard L. Tweedie, and Peter W. Glynn. *Markov Chains and Stochastic Stability*. Cambridge Mathematical Library. Cambridge University Press, 2 edition, 2009.

[37] É. Le Page. Théorèmes limites pour les produits de matrices aléatoires. In *Probability measures on groups (Oberwolfach, 1981)*, volume 928 of *Lecture Notes in Math.*, pages 258–303. Springer-Verlag, 1982.

[38] É. Le Page. Régularité du plus grand exposant caractéristique des produits de matrices aléatoires indépendantes et applications. *Ann. Inst. H. Poincaré Probab. Statist.*, 25:109–142, 1989.

[39] Y. Peres. Analytic dependence of Lyapunov exponents on transition probabilities. In *Lyapunov exponents (Oberwolfach, 1990)*, volume 1486 of *Lecture Notes in Math.*, pages 64–80. Springer-Verlag, 1991.

[40] D. Ruelle. Analyticity properties of the characteristic exponents of random matrix products. *Adv. in Math.*, 32:68–80, 1979.

[41] El Hadji Yaya Tall and Marcelo Viana. Moduli of continuity for the Lyapunov exponents of random $\mathrm{GL}(2)$-cocycles. *Trans. Amer. Math. Soc.*, 373(2):1343–1383, 2020.

[42] Marcelo Viana. *Lectures on Lyapunov exponents*, volume 145 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2014.

[43] Marcelo Viana. (Dis)continuity of Lyapunov exponents. *Ergodic Theory Dyn. Syst.*, 40(3):577–611, 2020.

[44] Jean-Christophe Yoccoz. Some questions and remarks about $\mathrm{SL}(2, \mathbf{R})$, cocycles. In *Modern dynamical systems and applications*, pages 447–458. Cambridge Univ. Press, Cambridge, 2004.