Pontifícia Universidade Católica
do Rio de Janeiro

Aline de Melo Machado

Markovian, quasiperiodic and mixed dynamical systems

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
January 2023

Pontifícia Universidade Católica
do Rio de Janeiro

Aline de Melo Machado

Markovian, quasiperiodic and mixed dynamical systems

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. Alexander Eduardo Arbieto Mendoza
Universidade Federal do Rio de Janeiro - UFRJ

Prof. Bruno Rodrigues Santiago
Universidade Federal Fluminense - UFF

Prof. Carlos Humberto Vásquez Ehrenfeld
Pontifica Universidad Católica de Valparaíso - PUCV

Profa. Luciana Silva Salgado
Universidade Federal do Rio de Janeiro - UFRJ

Prof. Manuel Stadlbauer
Universidade Federal do Rio de Janeiro - UFRJ

Prof. Mauricio José Poletti Merlo
Universidade Federal do Ceará - UFC

Rio de Janeiro, January the 13th, 2023

Aline de Melo Machado

Has a Bachelor degree in Mathematics from Universidade Federal Fluminense, Volta Redonda-RJ, Brasil and Master's degree in Mathematics from the Pontifical Catholic University, Rio de Janeiro, Brazil.

Bibliographic data

de Melo Machado, Aline

Markovian, quasiperiodic and mixed dynamical systems / Aline de Melo Machado; advisor: Silvius Klein. – 2023.

101 f: il. color. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Matemática, 2023.

Inclui bibliografia

1. Matemática – Teses. 2. Cociclo linear. 3. Expoente de Lyapunov. 4. Média de Birkhoff. 5. Estimativas de grandes desvios. 6. Sistema 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.

Acknowledgments

I would like to thank my parents José Cláudio and Adriana and my sister Julia for the great support and encouragement, which were fundamental for the conclusion of my PhD's degree.

I would like to sincerely thank my advisor Prof. Silvius Klein for the continuous support, for his patience, motivation, academic advises and immense knowledge during these six years of post-graduation.

Besides my advisor, I would like to thank the members of the thesis committee, Alexander Arbieto, Bruno Santiago, Carlos Vásquez, Luciana Salgado, Manuel Stadlbauer and Maurício Poletti, for accepting the invitation and for the several useful comments that helped significantly in the text of this thesis.

I would also like to thank Ao Cai for the friendship, support and for consistently improving my work with many suggestions and discussions.

My sincere thanks also goes to Carlos Henrique, Creuza Nascimento, Kátia Aguiar and Mariana Soares for helping me to fulfill all the bureaucracy requirements.

I thank all the professors of PUC-Rio and UFRJ for the support and knowledge. Specially, Jaqueline Siqueira, Katrin Gelfert, Lorenzo Díaz, Marcos Craizer, Renata Rosa and Sinésio Pesco.

I would also like to thank my professors from my undergraduate period at UFF-Volta Redonda. In particular, I would like to thank Marina Sequeiros and Honorio Fernando and a special thanks to Ivan Aguilar for all his support since I was a young woman until now, ten years later.

I thank all the friends of PUC-Rio for the relaxed moments inside and outside of the university. In particular, I thank Gabriel Dias, Jéssica Bittencourt, João Marcos, Makson Santos and Raul Chavez. Besides, I would like to thank Camila Sobrinho and Odylo Costa for the friendship, conversations and the great moments together (even during the pandemic time).

I would also like to thank all the mathematician women that helped me a lot during this process. Specially, I thank Catalina Freijó, Luize D'Urso, Marisa Cantarino, Pêdra Andrade, Simone Freitas and Tamires Pinto. Their friendship and support were fundamental during all this years.

A special thanks to Marcelo Durães. His help and his optimism were fundamental for me to keep going through my academic career. I am fortunate to have such a great and special person in my life.

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

Abstract

de Melo Machado, Aline; Klein, Silvius (Advisor). **Markovian, quasiperiodic and mixed dynamical systems**. Rio de Janeiro, 2023. 101p. Tese de Doutorado – Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

We study several models of base dynamics and linear cocycles over such systems. We establish effective rates of convergence of the Birkhoff averages of toral translations. We derive large deviations estimates for mixed Markov-quasiperiodic dynamics. We prove continuity properties of the Lyapunov exponents of linear cocycles over Markov shifts. Besides their intrinsic interest, these results prepare the ground for a larger project concerned with the study of linear cocycles over mixed Markov-quasiperiodic base dynamics. As crucial steps in this study, we obtain a version of Kifer's non random filtration and an upper large deviations estimate for such systems.

Keywords

Linear Cocycle; Lyapunov exponent; Birkhoff averages; Large deviations estimates; Markov system.

Resumo

de Melo Machado, Aline; Klein, Silvius. **Sistemas dinâmicos markovianos, quase periódicos e mistos**. Rio de Janeiro, 2023. 101p. Tese de Doutorado – Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

Estudamos vários modelos de dinâmica de base e cociclos lineares sobre tais sistemas. Estabelecemos taxas efetivas de convergência das médias de Birkhoff de translações do toro. Obtemos estimativas de grandes desvios para dinâmicas mistas Markov-quase periódicas. Provamos propriedades de continuidade dos expoentes de Lyapunov de cociclos lineares sobre deslocamentos de Markov. Além do interesse intrínseco, estes resultados preparam o terreno para um projeto maior voltado para o estudo de cociclos lineares sobre dinâmica de base mista Markov-quase periódica. Como passos essenciais neste estudo, obtemos uma versão da filtração não aleatória de Kifer e uma estimativa de grandes desvios por cima para tais sistemas.

Palavras-chave

Cociclo linear; Expoente de Lyapunov; Média de Birkhoff; Estimativas de grandes desvios; Sistema de Markov.

Table of contents

1
Introduction

We study limiting properties of additive and multiplicative processes associated to the iterates of Markov shifts, toral translations and skew products thereof.

Let M be a compact metric space, let $f\colon M \to M$ be a continuous transformation and let ν be an f-invariant, ergodic probability measure on M.

Given a continuous observable $\varphi\colon M \to \mathbb{R}$, let

$$\varphi^{(n)}(x) := \varphi(x) + \varphi(f(x)) + \cdots + \varphi(f^{n-1}(x))$$

bet its n-th Birkhoff sum relative to the ergodic system (M, f, ν).

By the additive ergodic theorem, the Birkhoff averages $\frac{1}{n}\varphi^{(n)}(x)$ converge to the space average $\int_M \varphi \, d\nu$ for ν-a.e. $x \in M$. In particular, the convergence also holds in measure, that is, for all $\varepsilon > 0$,

$$\nu \left\{ x \in M : \left| \frac{1}{n}\varphi^{(n)}(x) - \int_M \varphi \, d\nu \right| > \varepsilon \right\} \to 0 \tag{1.1}$$

as $n \to \infty$.

When, for an appropriate class of observables, the above convergence has a certain rate, ideally exponential, we say that our ergodic system satisfies large deviations estimates (LDE). Such estimates are already available for various systems with some (non uniform) hyperbolicity, see [9].

At the other end of ergodic behavior, in the case of the torus translation, because of its unique ergodicity, the convergence in the ergodic theorem is uniform, so for n large enough, the measure of the set in (1.1) is zero. The question is then if, under appropriate assumptions, there is an *effective* uniform rate of convergence of the Birkhoff averages for this model.

Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ be the one dimensional torus endowed with the Lebesgue measure m and let $T\colon \mathbb{T} \to \mathbb{T}$, $Tx = x + \omega$ be the translation on $\mathbb{T}$ by an irrational frequency ω. Assume that ω satisfies the following generic *Diophantine condition.*

$$\operatorname{dist}(k\omega, \mathbb{Z}) \geq \frac{\gamma}{|k| \, \log^2(|k| + 1)}$$

for some $\gamma > 0$ and for all $k \in \mathbb{Z} \setminus \{0\}$.

The first result of this work is the following sharp convergence rate of the Birkhoff sums for Hölder continuous observables.

Theorem 1.0.1 *Assume that the observable φ is an α-Hölder continuous function on $\mathbb{T}$ and that the frequency $\omega \in \mathbb{T}$ satisfies the Diophantine condition above. Then for all integers N we have*

$$\left\| \frac{1}{N} \varphi^{(N)} - \int_{\mathbb{T}} \varphi \right\|_{\infty} \leq \text{const} \left(\frac{1}{\gamma} \log \frac{1}{\gamma} \right) \|\varphi\|_{\alpha} \frac{\log^{3\alpha} N}{N^{\alpha}} \tag{1.2}$$

where const is a universal constant and $\|\varphi\|_{\alpha}$ is the α-Hölder norm of φ.

A similar but less sharp result also holds for the higher dimensional torus translation.

We say that $\boldsymbol{\omega} \in \mathbb{T}^d (d \geq 1)$ satisfies *a Diophantine condition* if there exist $\gamma > 0$ and $A > d$ such that

$$\|\boldsymbol{k} \cdot \boldsymbol{\omega}\| = \text{dist}\,(\boldsymbol{k} \cdot \boldsymbol{\omega},\, \mathbb{Z}) \geq \frac{\gamma}{|\boldsymbol{k}|^A}$$

for all $\boldsymbol{k} \in \mathbb{Z}^d$ with $|\boldsymbol{k}| \neq 0$.

Theorem 1.0.2 *Let φ be an α-Hölder continuous function on $\mathbb{T}^d$, let $\boldsymbol{\omega} \in \mathbb{T}^d$ be a frequency that satisfies the Diophantine condition above and let $T_{\boldsymbol{\omega}} \colon \mathbb{T}^d \to \mathbb{T}^d$ be the corresponding torus translation. Then for all $N \geq 1$ we have*

$$\left\| \frac{1}{N} \phi^{(N)} - \int_{\mathbb{T}^d} \phi \right\|_{\infty} \leq \frac{\text{const}}{\gamma} \|\phi\|_{\alpha} \frac{1}{N^{\beta}}$$

where $\beta = \frac{\alpha}{A+d}$.

With the behavior of the Birkhoff sums of both predominantly hyperbolic and deterministic (i.e. toral translations) systems well understood, it is natural to consider *mixed* models, i.e., certain skew products combining hyperbolic and deterministic systems, such as the direct product between a (strongly mixing) Markov shift and a torus translation.

Let Σ be a compact metric space and let $\text{Prob}(\Sigma)$ be the space of probability measures on Σ, endowed with the weak* topology. Consider a Markov transition kernel on M, that is, a continuous map $M \ni x \mapsto K_x \in \text{Prob}(\Sigma)$. We assume that K is uniformly ergodic, in the sense that its powers converge uniformly in the total variation norm. In particular K admits a unique stationary measure $\mu \in \text{Prob}(\Sigma)$. We refer to (Σ, K, μ) as a Markov system.

Let $X = \Sigma^{\mathbb{Z}}$, let $\mathbb{P}$ be the Markov measure on X with initial distribution μ and transition kernel K and let $\sigma \colon X \to X$ be the forward shift. Then

$(X, \sigma, \mathbb{P})$ is an ergodic system that we call a Markov shift. This generalizes the concept of subshift of finite type. It is well known that such systems satisfy large deviations estimates, see for instance [30] and [7].

Given $\alpha \in \mathbb{T}^d$ rationally independent, consider the product map

$$f(\omega, \theta) = (\sigma\omega, \theta + \alpha).$$

It turns out that $(X \times \mathbb{T}^d, f, \mathbb{P} \times m)$ is an ergodic system, which we call a mixed Markov quasiperiodic system.

We establish a large deviations estimate for such systems with observables that depend on a finite number of coordinates.

Theorem 1.0.3 *Let K be a uniformly ergodic Markov kernel on Σ, let μ be its unique stationary measure and let $\mathbb{P}$ be the Markov measure on $X = \Sigma^{\mathbb{Z}}$. Let $\varphi\colon X \times \mathbb{T}^d \to \mathbb{R}$ be a continuous observable that depends on a finite number of coordinates of $\omega \in X$. Given any $\varepsilon > 0$, there exist $\bar{n} = \bar{n}(\varepsilon, \phi) \in \mathbb{N}$ and $c = c(\varepsilon, \phi) > 0$ such that for all $\theta \in \mathbb{T}^d$ and for all $n \geq \bar{n}$, we have*

$$\mathbb{P}\left\{\omega \in X : \left|\frac{1}{n}\sum_{j=0}^{n-1} \phi(f^j(\omega, \theta)) - \int_{X \times \mathbb{T}^d} \phi \, d(\mathbb{P} \times m)\right| \geq \varepsilon\right\} < e^{-cn}.$$

It would, of course, be interesting to consider the more general case of observables that depend on all coordinates.

An even more complex type of dynamical system is given by linear cocycles over a base ergodic system (M, f, ν). A continuous map $A\colon M \to \mathrm{GL}_m(\mathbb{R})$ determines the linear skew product $F\colon M \times \mathbb{R}^m \to M \times \mathbb{R}^m$,

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

This new dynamical system is called a linear cocycle and its iterates are given by

$$F^n(x, v) = (f^n(x), A^{(n)}(x)v),$$

where

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

are refered to as fiber iterates.

Some important examples of linear cocycles are: quasiperiodic cocycles (i.e. linear cocycles over a torus translation); random cocycles (i.e. linear cocycles over a Bernoulli shift); Markov cocycles (i.e. linear cocycles over a Markov shift); mixed random-quasiperiodic cocycles (i.e. linear cocycles over a skew product between a Bernoulli shift and toral translations) and mixed Markov-quasiperiodic cocycles (i.e. linear cocycles over a mixed Markov-quasiperiodic map).

By the subadditive ergodic theorem (or by Furstenberg-Kesten's theorem), the geometric averages $\frac{1}{n} \log \|A^{(n)}(x)\|$ converge for ν-a.e. $x \in M$ to a constant $L_1(F)$ called the maximal Lyapunov exponent of F.

The other Lyapunov exponents $L_2(F), \ldots, L_m(F)$ are defined similarly, changing the norm (or largest singular value) of the fiber iterates $A^{(n)}(x)$ by the other singular values.

Definition 1.0.1 We say that the linear cocycle A satisfies a *fiber large deviations estimate* if for every $\varepsilon > 0$ there exist $c(\varepsilon) > 0$ and $n(\varepsilon) \in \mathbb{N}$ such that

$$\mu\{x \in M : \left|\frac{1}{n} \log \|A^{(n)}(x)\| - L_1(F)\right| > \varepsilon\} < e^{-c(\varepsilon)n} \tag{1.3}$$

for all $n \geq n(\varepsilon)$.

An important question in dynamical systems concerns the continuity properties at these limiting quantities, the Lyapunov exponents, as functions of the input data.

Results on the continuity of the Lyapunov exponents are available for several models. See for instance: [10, Chapter 6] for Hölder continuity results for quasiperiodic cocyles; [10, Chapter 5], [11] and [3] for continuity results for random cocycles; [10, Chapter 5] for continuity with respect to the the fiber map A of Lyapunov exponents of Markov cocycles and [6] and [2] for continuity of Lyapunov exponent for mixed random-quasiperiodic.

We establish the *joint* Hölder continuity of the maximal Lyapunov exponent of Markov cocycle (as a function of the fiber map and the transition kernel), under a generic (irreducibility) assumptions.

More precisely, let (Σ, K, μ) be a Markov system, let $A \colon \Sigma \times \Sigma \to \mathrm{GL}_m(\mathbb{R})$ be a fiber map and consider the corresponding Markov cocycle $F = F_{(A,K)} \colon X \times \mathbb{R}^m \to X \times \mathbb{R}^m$,

$$F(\omega, v) = (\sigma\omega, A(\omega_1, \omega_0)v),$$

where $\omega = \{\omega_n\}_{n\in\mathbb{Z}} \in X = \Sigma^{\mathbb{Z}}$.

Definition 1.0.2 The Markov cocycle $F_{(A,K)}$ is called *irreducible* if there is no proper invariant section under the fiber dynamics.

Theorem 1.0.4 *Let* $A \colon \Sigma \times \Sigma \to \mathrm{GL}_m(\mathbb{R})$ *be a Lipschitz continuous cocycle and let* $K \colon \Sigma \to \mathrm{Prob}(\Sigma)$ *be a uniformly ergodic Markov kernel. Assume that:*

(i) the cocycle $F_{(A,K)}$ is irreducible,

(ii) $L_1(A,K) > L_2(A,K)$.

Then the map $(A,K) \mapsto L_1(A,K)$ is locally Hölder continuous.

An important motivation for part of this work, and a longer term project, concerns the study of linear cocycles over mixed Markov-quasiperiodic base dynamics. This projects was inspired by the ongoing work of Cai, Duarte and Klein on the stability under random noise of quasiperiodic systems (see [5], [6] and [7] for partial results in this direction). Other interesting results were recently obtained for related models by Bezerra and Poletti [2], Goldsheid [14] and by Gorodetski and Kleptsyn [15].

We are interested in deriving large deviations estimates for mixed Markov-quasiperiodic cocycles as well as continuity properties of their Lyapunov exponent. It turns out that these two problems are related. By an abstract continuity theorem of the Lyapunov exponents in [10, Chapter 3], the availability of LDE that are uniform in the data imply the Hölder continuity of the Lyapunov exponent for any space of cocycles over any base dynamics.

In this work, we derive an upper fiber LDE for such mixed cocycles.

Let (Σ, K, μ) be a Markov system, let $(X, \sigma, \mathbb{P})$ be the corresponding Markov shift and let $\alpha \in \mathbb{T}^d$ be rationally independent. Given a continuous fiber map $A\colon \Sigma \times \Sigma \times \mathbb{T}^d \to \mathrm{GL}_m(\mathbb{R})$, consider the corresponding mixed Markov-quasiperiodic cocycle $F\colon X \times \mathbb{T}^d \times \mathbb{R}^m \to X \times \mathbb{T}^d \times \mathbb{R}^m$,

$$F(\omega, \theta, v) = (\sigma\omega, \theta + \alpha, A(\omega_1, \omega_0, \theta)v),$$

where $\omega = \{\omega_n\}_{n\in\mathbb{Z}} \in X = \Sigma^{\mathbb{Z}}$.

Theorem 1.0.5 *Given a mixed Markov-quasiperiodic cocycle F as above, for any $\varepsilon > 0$ there are $\bar{n} = \bar{n}(\varepsilon) \in \mathbb{N}$ and $c = c(\varepsilon) > 0$ such that for all $\theta \in \mathbb{T}^d$*

$$\mathbb{P}\left\{\omega : \frac{1}{n}\log \|A^{(n)}(\omega, \theta)\| > L_1(F) + \varepsilon\right\} < e^{-c(\varepsilon)n}$$

for all $n \geq \bar{n}(\varepsilon)$.

This result, together with the large deviations estimates on the base (1.1), the uniform convergence (1.2) and other more technical results, like a version of Kifer's non-random filtration, prepare the ground for the completion of the aforementioned larger project.

The rest of this work is organized in six chapters as follows.

In Chapter 2 we review basic notions in ergodic theory, such as the concepts and examples of ergodic systems and the Markov operator associated to a Markov transition kernel. We also recall some important theorems in ergodic theory: Oseledets's theorem and Kifer's non-random filtration theorem.

In Chapter 3 we present the result regarding the rate of convergence of Birkhoff averages for a Diophantine torus translation with Hölder continuous observable.

In Chapter 4 we establish and prove a large deviations estimate for mixed Markov-quasiperiodic dynamical systems with observables depending on a finite number of coordinates.

In Chapter 5 we obtain the Hölder continuity of the maximal Lyapunov exponent via a Furstenberg's formula for linear cocycles over Markov shifts.

In Chapter 6 we obtain an upper large deviations estimate for linear cocycles over mixed Markov-quasiperiodic base dynamics, and as a consequence of that, the upper semi continuity of the maximal Lyapunov exponent.

In Chapter 7 we describe some related future projects.

2
General concepts

In this chapter we review the main concepts needed throughout the work. We begin with the notion of ergodic dynamical system (Section 2.1). In Section 2.2 we define the general Markov shift and introduce the Markov operator associated to a Markov transition kernel. In Section 2.3 we define the concept of linear cocycle and its Lyapunov exponents via the Furstenberg-Kesten's theorem. In Section 2.4 we recall the multiplicative ergodic theorem of Oseledets and state the Kifer non-random filtration theorem, a stronger version of Oseledets in the setting of linear cocycles over a Bernoulli shift. In Section 2.5 we introduce the concept of large deviations estimates in dynamical systems. We conclude this chapter with the statement and proof of a more general version of Kifer's non random filtration theorem (Section 2.6).

2.1
Measure preserving dynamical systems

Let $(X, \mathcal{B}, \mu)$ be a measure space and let $T\colon X \to X$ be a measurable function.

Definition 2.1.1 We say that μ is a *T-invariant measure on X*, or that T *preserves* μ, if $\mu(T^{-1}E) = \mu(E)$, for all measurable sets $E \in \mathcal{B}$.

Definition 2.1.2 Let $T\colon X \to X$ be a measurable function. We say that $(X, \mathcal{B}, \mu, T)$ is a *measure preserving dynamical system* if μ is a T-invariant probability measure.

Definition 2.1.3 Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving dynamical system and let $\phi\colon X \to \mathbb{R}$ be an absolutely integrable function (which we refer to as an observable).

i) A measurable subset $E \subset X$ is said to be *T-invariant* if $T^{-1}E = E$.

ii) $\phi\colon X \to \mathbb{R}$ is a *T-invariant function* if $\phi \circ T = \phi$, μ-a.e.

Definition 2.1.4 A measure preserving dynamical system $(X, \mathcal{B}, \mu, T)$ is said to be an *ergodic system* if $\mu(E) = 0$ or $\mu(E) = 1$ for any T-invariant subsets $E \subset X$.

Definition 2.1.5 We say that a measure preserving dynamical system (X, μ, T) is *uniquely ergodic* if $T\colon X \to X$ is a homeomorphism and μ is the *unique* T-invariant probability measure on a metric space X.

Remark 2.1 If (X, μ, T) is a uniquely ergodic system, then necessarily μ is ergodic. Indeed, suppose that there exists an invariant subset A in M with $0 < \mu(A) < 1$. Thus,

$$\mu_A(E) := \frac{\mu(E \cap A)}{\mu(A)}$$

is a different T-invariant probability measure on X, contradicting the hypothesis that μ is the unique T-invariant measure on X.

We present below examples of ergodic dynamical systems.

Example 2.1 (The Bernoulli shift) *Let Σ be a compact metric space and consider the space of sequences $X^+ = \Sigma^{\mathbb{N}}$. The Bernoulli shift is the map $\sigma : X^+ \to X^+$ defined by $\sigma(x) = \{x_{n+1}\}_{n\in\mathbb{N}}$ for $x = \{x_n\}_{n\in\mathbb{N}}$. We use the same notation for its extension to the space $X = \Sigma^{\mathbb{Z}}$ of double sided sequences. Denote by* $\mathrm{Prob}(\Sigma)$ *the space of Borel probability measures on Σ.*

Given $\mu \in \mathrm{Prob}(\Sigma)$, the shift maps $\sigma\colon X^+ \to X^+$ and $\sigma\colon X \to X$ preserves the product measure $\mu^{\mathbb{N}}$ and $\mu^{\mathbb{Z}}$ respectively. The measure preserving dynamical system $(X, \mu^{\mathbb{Z}}, \sigma)$ is called a Bernoulli shift*. It turns out that it is an ergodic system (see Proposition 4.2.7 in [33]).*

Example 2.2 (Translation on the d-dimensional torus) *Let $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$ with $d \geq 1$ be the d-dimensional torus endowed with the Haar measure m and also regarded as an additive compact group. Points on $\mathbb{T}^d$ are written as $\theta = (\theta_1, \ldots, \theta_d)$.*

Given a frequency $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{T}^d$, define the map

$$T_\alpha\colon \mathbb{T} \to \mathbb{R}, \quad T_\alpha(\theta) = \theta + \alpha$$

which is called the torus translation map.

We say that a vector $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d$ is rationally independent *if, for any integer numbers $n_0, n_1, \ldots, n_d$, we have that*

$$n_0 + n_1\alpha_1 + \cdots + n_d\alpha_d = 0 \Rightarrow n_0 = n_1 = \cdots = n_d = 0.$$

Otherwise, we say that α is rationally dependent*.*

The triple $(\mathbb{T}^d, m, T_\alpha)$ is a uniquely ergodic system if, and only if, the components of α are rationally independent (see Proposition 4.2.2 in [33]).

Example 2.3 (The affine skew product on the d-torus) *For any dimension $d \geq 2$ and an irrational number α, let $S_\alpha \colon \mathbb{T}^d \to \mathbb{T}^d$,*

$$S_\alpha(x_1, x_2, \dots, x_d) = (x_1 + x_2, x_2 + x_3, \dots, x_d + \alpha),$$

be the skew-translation map.

The map S_α is also an example of uniquely ergodic transformation for every irrational number α. (see [12]).

Birkhoff's additive ergodic theorem, one of the most important results in ergodic theory, states that given an ergodic system, the time average of an observable along the trajectories converges almost everywhere to the space average. This result can be formulated as follows (see [33], [18] for the proof).

Theorem 2.1.1 (Birkhof's Ergodic Theorem for ergodic systems) *Let $(X, \mathcal{B}, \mu, T)$ be an ergodic system and let $\phi \colon X \to X$ be an observable. Then,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \phi(T^j x) = \int_X \phi(x)\, d\mu(x)$$

for μ-a.e. $x \in X$.

2.2 Markov kernels

Let M be a compact metric space and let $\mathcal{B}$ be its Borel σ-algebra. Let $\mathrm{Prob}(M)$ denote the space of Borel probability measures on M, which we endowed with the weak* topology.

Definition 2.2.1 A *Markov kernel* on Σ is a continuous map $\mathcal{K} \colon M \to \mathrm{Prob}(M), x \mapsto \mathcal{K}_x$.

The iterates of a Markov Kernel $\mathcal{K}$ are defined recursively setting $\mathcal{K}^1 := \mathcal{K}$ and for $n \geq 2, E \in \mathcal{B}$,

$$\mathcal{K}_x^n(E) := \int_X \mathcal{K}_y^{n-1}(E)\, d\mathcal{K}_x(y).$$

Each power $\mathcal{K}^n$ is itself a Markov kernel on $(M, \mathcal{B})$.

Definition 2.2.2 A probability measure ν on $(M, \mathcal{B})$ is called $\mathcal{K}$-*stationary* if

$$\nu(E) = \int \mathcal{K}_x(E)\, d\nu(x)$$

for all $E \in \mathcal{B}$.

The above definition means that ν is stationary if ν is $\mathcal{K}_x$-invariant on average.

Definition 2.2.3 A *Markov system* is a triple $(M, \mathcal{K}, \nu)$, where $\mathcal{K}$ is a Markov Kernel on $(M, \mathcal{B})$ and ν is a $\mathcal{K}$-stationary probability measure.

Consider a Markov system $(\mathcal{K}, \nu)$ on a compact metric space M.

Definition 2.2.4 The linear operator $\mathcal{Q} = \mathcal{Q}_{\mathcal{K}} \colon L^\infty(M) \to L^\infty(M)$

$$(\mathcal{Q}f)(x) = (\mathcal{Q}_{\mathcal{K}}f)(x) := \int f(y)\, d\mathcal{K}_x(y)$$

is called a *Markov operator*. It is easy to verify that

$$(\mathcal{Q}_{\mathcal{K}}^n f)(x) := \int f(y)\, d\mathcal{K}_x^n(y)$$

for all $n \geq 1$ and $f \in L^\infty(M)$.

Let $(\mathcal{K}, \nu)$ be a Markov system. Consider the space $X^+ = M^{\mathbb{N}}$ of sequences $x = \{x_n\}_{n\in\mathbb{N}}$ with $x_n \in M$ for all $n \in \mathbb{N}$ and let $\mathcal{B}^+$ be the product σ-field $\mathcal{B}^+ = \mathcal{B}^{\mathbb{N}}$ generated by the $\mathcal{B}$-cylinders. In other words, $\mathcal{B}^+$ is generated by sets of the form

$$C(E_0, \ldots, E_m) := \{x \in X^+ : x_j \in E_j, \quad \text{for } 0 \leq j \leq m\},$$

where $E_0, \ldots, E_m \in \mathcal{B}$ are measurable sets.

Definition 2.2.5 Given any probability measure θ on $(M, \mathcal{B})$, the following expression determines a pre-measure

$$\mathbb{P}_\theta^+[C(E_0, \ldots, E_m)] := \int_{E_0}\int_{E_1} \cdots \int_{E_m} dK_{x_{m-1}}(x_m) \cdots dK_{x_1}(x_0)\, d\theta(x_0)$$

on the semi-algebra of $\mathcal{B}$-cylinders. By Carathéodory's extension theorem this pre-measures extends to a unique probability measure $\mathbb{P}_\theta^+$ on $(X^+, \mathcal{B}^+)$.

Markov system are probabilistic evolutionary models, which can also be studied in dynamical terms. Let $(\mathcal{K}, \nu)$ be a Markov system and $X = M^{\mathbb{Z}}$ be the set of double sided sequences. The *one-sided shift* is the map $\sigma : X^+ \to X^+$ such that $\sigma(\{x_n\}_{n\in\mathbb{N}}) = \{x_{n+1}\}_{n\in\mathbb{N}}$ and the *two-sided shift* is the map $\sigma \colon X \to X$ such that $\sigma(\{x_n\}_{n\in\mathbb{Z}}) = \{x_{n+1}\}_{n\in\mathbb{Z}}$. The two-sided shift is the natural extension of the one-sided shift. Then, there is a unique probability measure $\mathbb{P}_{(\mathcal{K},\nu)}$ on $(X, \mathcal{B})$ and we will refer to the measure $\mathbb{P}_{(\mathcal{K},\nu)}$ as the *Kolmogorov extension* of the Markov system $(\mathcal{K}, \nu)$.

Definition 2.2.6 Given a Markov system $(\mathcal{K}, \nu)$ let $\mathbb{P}_{(\mathcal{K},\nu)}$ be the Kolmogorov measure on $X = M^{\mathbb{Z}}$. The dynamical system $(X, \mathbb{P}_\nu, \sigma)$ is called a *Markov system.*

We denote by $\mathbb{P} = \mathbb{P}_{(\mathcal{K},\nu)}$ the Markov measure on $M^{\mathbb{N}}$ with initial distribution ν and transition kernel $\mathcal{K}$.

2.3 Linear cocycles and Lyapunov exponents

Definition 2.3.1 A *linear cocycle* over a base ergodic system (X, μ, T) is a skew-product map

$$\begin{aligned} F_A \colon X \times \mathbb{R}^m &\to X \times \mathbb{R}^m \\ (x, v) &\mapsto (Tx, A(x)v) \end{aligned}$$

where $A \colon X \to \mathrm{GL}_m(\mathbb{R})$ is a measurable map.

We identify F_A with the pair (T, A) or simply with A. Define

$$A^{(n)}(x) = A(T^{n-1}x) \cdots A(Tx)A(x).$$

Then the iterates of F_A are given by:

$$F_A^n(x, v) = (T^n x, A^{(n)}(x)v).$$

Example 2.4 (Quasiperiodic cocycles) *We call* quasiperiodic cocycle *a linear cocycle over a torus translation. That is, given $\alpha \in \mathbb{T}^d$ and $A \colon \mathbb{T}^d \to \mathrm{GL}_m(\mathbb{R})$, let $F_{(\alpha,A)} \colon \mathbb{T}^d \times \mathbb{R}^m \to \mathbb{T}^d \times \mathbb{R}^m$ such that*

$$F_{(\alpha,A)}(\theta, v) = (\theta + \alpha, A(\theta)v) .$$

For simplicity we identify the quasiperiodic cocycle $F_{(\alpha,A)}$ with the pair (α, A).

Example 2.5 (Random cocycles) *We call* random cocycle *a linear cocycle over a Bernoulli shift. That is, given $A \colon M \to \mathrm{GL}_m(\mathbb{R})$ and $X = M^{\mathbb{Z}}$, let $F = F_{(A,K)} \colon X \times \mathbb{R}^m \to X \times \mathbb{R}^m$ such that*

$$F(\omega, v) = (\sigma\omega, A(\omega_0)v) .$$

Example 2.6 (Markov cocycles) *We call* Markov cocycle *a linear cocycle over a Markov system. That is, given a Markov system (Σ, K, μ), the corresponding $(X, \mathbb{P}, \sigma)$ and given $A \colon M \times M \to \mathrm{GL}_m(\mathbb{R})$, let $F = F_{(A,K)} \colon X \times \mathbb{R}^m \to$*

$X \times \mathbb{R}^m$ *such that*

$$F(\omega, v) = (\sigma\omega, A(\omega_1, \omega_0)v)\,.$$

Its iterates are given by

$$F^n(\omega, v) = (\sigma^n\omega, A^n(\omega)v)\,,$$

where for $\omega = \{\omega_n\}_{n\in\mathbb{Z}} \in X$*,*

$$A^{(n)}(\omega) = A(\omega_n, \omega_{n-1}) \cdots A(\omega_2, \omega_1)\, A(\omega_1, \omega_0)\,.$$

Theorem 2.3.1 (Furstenberg-Kesten) *Given a* μ*-integrable cocycle* $F = F_{(T,A)}$ *of an ergodic system* (X, μ, T)*, that is,*

$$\int_X \log^+ \|A(x)\|\, d\mu(x) < +\infty$$

for μ *almost every* $x \in X$*,*

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

The number $L_1(A)$ *is called the* first Lyapunov exponent *of* A*.*

Furthermore, the Lyapunov exponents of A, denoted by $L_k(F), 1 \le k \le m$, can be characterized as the almost everywhere limits

$$L_k(F) := \lim_{n\to\infty} \frac{1}{n} \log s_k(A^{(n)}(x)),$$

where $s_k(A^n(x))$ are the singular values of the matrices $A^{(n)}(x)$.

The Lyapunov exponents depends on the data. In particular, if A is a quasiperiodic cocycle, we denote the maximal Lyapunov exponent by $L_i(\alpha, A)$. If (A, K) is a Markov cocycle, then we denote by $L_1(A, K)$ the first Lyapunov exponent of A.

2.4 The multiplicative ergodic theorem

Given a measure $\mu \in \mathrm{Prob}(\mathrm{GL}_m(\mathbb{R}))$, denote by G_μ the closed subgroup of $\mathrm{GL}_m(\mathbb{R})$ generated by the support of μ.

Let $\mathrm{Gr}(\mathbb{R}^m)$ denote the set of all linear subspaces of $\mathbb{R}^m$. In the conext of $\mathrm{GL}_m(\mathbb{R})$-valued cocycles, the Oseledets Multiplicate Ergodic Theorem for ergodic transformations can be formulated as follows (see [33] for the proof).

Theorem 2.4.1 (Oseledets) *Let* (X, μ, T) *be an ergodic system, let* $A\colon X \to \mathrm{GL}_m(\mathbb{R})$ μ*-integrable and let* $F_A\colon X \times \mathbb{R}^m \to X \times \mathbb{R}^m$ *be the corresponding* *linear cocycle. Then, there exist $\lambda_1 > \lambda_2 > \cdots > \lambda_k \geq -\infty$ and a family of measurable functions $F_j \colon X \to \mathrm{Gr}(\mathbb{R}^m), 1 \leq j \leq k$, such that for μ-almost every $x \in X$,*

(a) $A(x)F_j(x) = F_j(Tx)$ *for* $j = 1, \ldots, k$

(b) $\{0\} = F_{k+1}(x) \subsetneq F_k(x) \subsetneq \ldots \subsetneq F_2(x) \subsetneq F_1(x) = X \times \mathbb{R}^m$

(c) for every $v \in F_j(x) \setminus F_{j+1}(x)$, $\quad \lim_{n\to+\infty} \dfrac{1}{n} \log \|A^n(x)v\| = \lambda_j$.

The numbers λ_i are called the distinct Lyapunov exponents *of the linear cocycle. They coincide with the (possibly) repeated Lyapunov exponents given by the Furstenberg-Kesten's theorem. In particular, $L_1 = \lambda_1$.*

This result improves the Furstenberg-Kesten theorem in that it provides exponential rates of the convergence for the iterates $\|A^{(n)}(x)v\|$ of all vectors, rather than just for the norm of the matrices.

For random linear cocycles, Kifer [20] obtained a more precise version of the multiplicative ergodic theorem, where the filtration does not depend on the base point.

Theorem 2.4.2 (Kifer non-random filtration) *Given $\mu \in \mathrm{Prob}(\mathrm{GL}_m(\mathbb{R}))$ there exists a filtration $\mathcal{L} = (L_0, L_1, \cdots, L_r)$ with $0 \leq r \leq m$,*

$$\mathbb{R}^m = L_0 \supsetneq L_1 \supsetneq \cdots \supsetneq L_{r-1} \supsetneq L_r \supsetneq \{0\}$$

and there are numbers

$$\beta(\mu) = \beta_0(\mu) > \beta_1(\mu) > \cdots > \beta_{r-1}(\mu) > \beta_r(\mu)$$

such that for every $0 \leq j \ \leq r$

(1) each linear subspace L_j is G_μ-invariant, that is, $gL_j = L_j$ for all $g \in G_\mu$

(2) for every $v \in L_j \setminus L_{j+1}$ and $\mu^{\mathbb{N}}$-a.e. $\omega \in \Sigma^{\mathbb{N}}$,

$$\lim_{n\to\infty} \log \|A^{(n)}(\omega)v\| = \beta_j(\mu)$$

(3) the numbers $\beta_j(\mu)$ are the values

$$\int_\Sigma \int_{\mathbb{P}(\mathbb{R}^m)} \log \|gp\| \, d\nu(\hat{p}) \, d\mu(g)$$

where ν is an extremal point of $\mathrm{Prob}_\mu(\mathbb{P}(\mathbb{R}^m))$ and $\mathrm{Prob}_\mu(\mathbb{P}(\mathbb{R}^m))$ is the space of μ-stationary probability measures.

(4) for any extremal point ν *of* $\mathrm{Prob}_\mu(\mathbb{P}(\mathbb{R}^m)$ *such that*

$$\int_\Sigma \int_{\mathbb{P}(\mathbb{R}^m)} \log \|gp\| \, d\nu(\hat{p}) \, d\mu(g) = \beta_j(\mu)$$

we have $\nu(\hat{L}_j) = 1$ *and* $\nu(\hat{L}_{j+1}) = 0$.

In Section 2.6 we will formulate and prove, following [20], a more general version of this result, which will be later used to derive Kifer non random filtration-type theorems for other types of linear cocycles.

2.5 Large deviations

Let $(X, \mathcal{B}, \mu, T)$ be an ergodic dynamical system.

As mentioned in Section 2.1, Birkhoff proved that the time average of an observable along the trajectories exists almost everywhere and converges to the space average. In particular, this implies the convergence in probability. When there is a rate of convergence in probability for certain types of observables, we say that the system satisfies large deviations estimates (LDE). More precisely,

Definition 2.5.1 We say that an observable $\varphi\colon X \to \mathbb{R}$ satisfies a *base large deviations estimate* if for every $\varepsilon > 0$, there exist $c(\varepsilon) > 0$ and $n(\varepsilon) \in \mathbb{N}$ such that

$$\mu\{x \in X : \left| \frac{1}{n} \sum_{j=0}^{n-1} \varphi(T^j x) - \int \varphi \right| > \varepsilon\} < e^{-c(\varepsilon)n}$$

for all $n \geq n(\varepsilon)$.

Let $A\colon X \to \mathrm{GL}_m(\mathbb{R})$ be a linear cocycle over the dynamical system $(X, \mathcal{B}, \mu, T)$.

Large deviations estimates are available for many types of base dynamical systems, especially for systems with some hyperbolicity (see for instance Chazottes and Gouëzel [9]).

In the context of uniquely ergodic system, it was proved that the time average of a continuous observable along the trajectories converges uniformly to the space average (see p. 160 in [33]).

Since the translation on the torus $T_\alpha\colon \mathbb{T}^d \to \mathbb{T}^d$ is uniquely ergodic, the Birkhoff ergodic theorem implies that

$$\frac{1}{n} \sum_{j=0}^{n-1} \varphi(T_\alpha^j x) \to \int \varphi$$

uniformly in $x \in \mathbb{T}^d$ where $\varphi\colon \mathbb{T}^d \to \mathbb{R}$ is an observable.

In particular,

$$\left\| \frac{1}{n} \sum_{j=0}^{n-1} \varphi(T_\alpha^j(\cdot)) - \int \varphi \right\|_\infty \to 0$$

so for every $\varepsilon > 0$, there exists $n(\varepsilon) \in \mathbb{N}$ such that

$$\left| \frac{1}{n} \sum_{j=0}^{n-1} \varphi(T_\alpha^j x) - \int \varphi \right| < \varepsilon$$

for every $x \in \mathbb{T}^d$ and for every $n \geq n(\varepsilon)$.

One of the main goals of this work, is to estimate the convergence rate of the Birkhoff averages for α-Hölder continuous observable φ in the context of diophantine torus translation. We present this result in the next chapter.

Definition 2.5.2 We say that the linear cocycle A satisfies a *fiber large deviations estimate* if for every $\varepsilon > 0$, there exist $c(\varepsilon) > 0$ and $n(\varepsilon) \in \mathbb{N}$ such that

$$\mu\{x \in X : \left| \frac{1}{n} \log \|A^{(n)}(x)\| - L_1(A) \right| > \varepsilon\} < e^{-c(\varepsilon)n}$$

for every $n \geq n(\varepsilon)$.

That is, if a linear cocycle $A\colon X \to \mathrm{GL}_m(\mathbb{R})$ satisfies large deviations estimates on the fiber, then

$$L_1(A) - \varepsilon < \frac{1}{n} \log \|A^{(n)}(x)\| < L_1(A) + \varepsilon$$

except for a set of phases with exponentially small probability. When

$$\frac{1}{n} \log \|A^{(n)}(x)\| < L_1(A) + \varepsilon$$

holds with high probability, we say that the cocycle satisfies *upper large deviations estimates.*

Large deviation estimates were obtained in many contexts, in particular, in the context of random Bernoulli cocycles [3], quasiperiodic cocycles (see [10, Chapter 6] therein) and Markov cocycles (see [7] and references therein).

Large deviations estimates for linear cocycles are intimatelly related to the continuity properties of the Lyapunov exponents. It was shown (see Chapter 3 in [10]) that if a space of cocycles satisfies uniform fiber LDE (uniform in the sense that the relevant parameters are stable under perturbation of the data) then the Lyapunov exponents are Hölder continuous functions.

In Chapter 5 we will obtain the Hölder continuity of the Lyapunov exponeny for Markov operators directly, without large deviations. However, for other models of linear cocycles, establish large deviations (which is important in itself) is also crucial for deriving continuity properties of the Lyapunov exponents.

2.6 Kifer non-random filtration

Before we formulate the Kifer non-random filtration theorem, we present the following theorem that will be useful later. We will follow the argument of Y. Kifer [20].

Theorem 2.6.1 *Let $(M, \mathcal{K}, \mu)$ be a Markov system, where M is a compact metric space, and $\mathcal{F}$ be the space of Borel maps of M into itself. For every $x \in M$ and for every measurable subset $E \subset M$, there exists a probability measure ν in the space of $\mathcal{F}$ such that*

$$\mathcal{K}_x(E) = \nu\{g \in \mathcal{F} : g(x) \in E\}\,.$$

Proof. Since the space M is a compact metric space, M is Borel measurably isomorphic to a Borel subset of the unit interval $[0,1]$ (see [22]). That is, there exists a one-to-one map $\varphi : M \to [0,1]$ such that $\Gamma = \varphi(M)$ is a Borel subset of $[0,1]$ and $\varphi^{-1} : \Gamma \to M$ is also Borel.

For any $x \in M$ and a Borel subset E of $[0,1]$, define a probability measure $\tilde{\mathcal{K}}$ on $[0,1]$ such that

$$\tilde{\mathcal{K}}_x(E) = \mathcal{K}_x(\varphi^{-1}(E \cap \Gamma))$$

and define the map $z : M \times [0,1] \to [0,1]$ such that

$$z(x,\omega) = \inf\{\gamma \in [0,1] : \tilde{\mathcal{K}}_x([0,\gamma]) \geq \omega\}.$$

Fixed ω and since $\mathcal{K}_x(G)$ is a Borel function of x for any Borel subset $G \subset M$, the subset $\{x \in M : \mathcal{K}_x(\varphi^{-1}([0,a] \cap \Gamma)) < \omega\}$ is Borel. Then $z(\cdot,\omega)$ is a Borel map from M into $[0,1]$. Indeed,

$$\begin{aligned}\{x \in M : z(x,\omega) > a\} &= \{x \in M : \tilde{\mathcal{K}}_x([0,a]) < \omega\}\\ &= \{x \in M : \mathcal{K}_x(\varphi^{-1}([0,a] \cap \Gamma)) < \omega\}\,.\end{aligned}$$

Let $x_0 \in M$ and consider the map $\psi \colon [0,1] \to M$ such that

$$\psi(x) := \begin{cases} \varphi^{-1}(x), & x \in \Gamma \\ x_0, & x \in [0,1] \setminus \Gamma \end{cases}$$

For each $\omega \in [0,1]$, we may define the Borel map $f_\omega \colon M \to M, f_\omega = \psi \circ z(\cdot, \omega)$. Then, we obtain a map $\mathcal{J} \colon [0,1] \to \mathcal{F}$ such that $\mathcal{J}(\omega) = f_\omega$. Moreover, there is a measurable structure on $\mathcal{F}$ induced by the map $\mathcal{J}$ by the following: a subset $A \subset \mathcal{J}$ is measurable if $\mathcal{J}^{-1}A$ is a Borel subset of $[0,1]$.

Let m be the Lebesgue measure on $[0,1]$ and define the probability measure

$$\nu(A) = m(\mathcal{J}^1 A)$$

for any $A \subset \mathcal{F}$ such that $\mathcal{J}^{-1}A$ is a Borel subset of $[0,1]$. Fixed $x \in M$, we have

$$\begin{aligned} m\{\omega : z(x,\omega) > a\} &= m\{\omega : \tilde{\mathcal{K}}_x([0,a]) < \omega\} \\ &= 1 - \tilde{\mathcal{K}}_x([0,a]) \,. \end{aligned}$$

Hence,

$$\begin{aligned} 1 &= m(\{\omega : z(x,\omega) \in [0,a]\} \cup \{\omega : z(x,\omega) > a\}) \\ &= m\{\omega : z(x,\omega) \in [0,a]\} + m\{\omega : z(x,\omega) > a\} \\ &= m\{\omega : z(x,\omega) \in [0,a]\} + 1 - \tilde{\mathcal{K}}_x([0,a])\} \end{aligned}$$

and then $m\{\omega : z(x,\omega) \in [0,a]\} = \tilde{\mathcal{K}}_x([0,a])$. Moreover, for any Borel subset Δ of $[0,1]$, we have

$$m\{\omega : z(x,\omega) \in \Delta\} = \tilde{\mathcal{K}}_x(\Delta).$$

Therefore, for every Borel subset G of M

$$\begin{aligned} \nu\{g \in \mathcal{F} : g(x) \in G\} &= m(\mathcal{J}^{-1}\{g \in \mathcal{F} : g(x) \in G\}) \\ &= m(\{\omega : \mathcal{J}(\omega) \in \{g \in \mathcal{F} : g(x) \in G\}\}) \\ &= m\{\omega : f(\omega)x \in G\} \\ &= m\{\omega : z(x,\omega) \in \psi^{-1}G\} \\ &= \tilde{\mathcal{K}}_x(\psi^{-1}G) \\ &= \tilde{\mathcal{K}}_x(\Gamma \cap \psi^{-1}G) \\ &= \tilde{\mathcal{K}}_x(\varphi G) = \mathcal{K}_x(G) \end{aligned}$$

and this completes the proof. ■

Definition 2.6.1 Let $(\Omega, \mathcal{B}, \mathbb{P})$ be a probability space. A *filtration* on $\mathcal{B}$ is a sequence of σ-algebras $\{\mathcal{B}_n\}_{n\geq 0}$ such that $\mathcal{B}_n \subset \mathcal{B}_{n+1}$. A *martingale* is a pair $(\{\xi_n\}_{n\geq 0}, \{\mathcal{B}_n\}_{n\geq 0})$ where $\{\xi_n\}_{n\geq 0}$ is a random process on $(\Omega, \mathcal{B}, \mathbb{P})$ and $\{\mathcal{B}_n\}_{n\geq 0}$ is a filtration on $\mathcal{B}$ such that for all $n \geq 0$:

(1) ξ_n is $\mathcal{B}_n$-measurable,

(2) $\mathbb{E}[\xi_{n+1} \mid \mathcal{B}_n] = \xi_n$, $\mathbb{P}$-almost everywhere.

Theorem 2.6.2 (Martingale Convergence Theorem [26]) *Let* $(\Omega, \mathcal{B}, \mathbb{P})$ *be a probability space and let*$(\{\xi_n\}_{n\geq 0}, \{\mathcal{B}_n\}_{n\geq 0})$ *be a martingale. There exists* $\xi_\infty \in L^1(X, \mathcal{B}, \mathbb{P})$ *such that:*

(1) $\xi_n \to \xi_\infty$ $\mathbb{P}$*-almost everywhere,*

(2) $\mathbb{E}[\xi_\infty \mid \mathcal{B}_n] = \xi_n$ $\mathbb{P}$*-almost surely for every* $n \geq 0$*,*

(3) ξ *is* $\mathcal{B}_\infty$*-measurable, where* $\mathcal{B}_\infty = \sigma(\mathcal{B}_n : n \geq 0)$*.*

Let $\mathcal{F}$ be the space of all projective linear cocycles. The space $\mathcal{F}$ is endowed with a measurable structure such that the map $\mathcal{F} \times M \times \mathbb{PR}^d$, $(F, u) \mapsto Fu$ is measurable with respect to the product measurable structure in $\mathcal{F} \times M \times \mathbb{PR}^d$.

Let $\mathbb{P}^{d-1}$ be the $(d-1)$-dimensional projective space and let $m(x)$ be a positive integer-valued Borel function on M. Define $\mathcal{U}_k = \{x : m(x) = k\}$.

Definition 2.6.2 We say that $\mathcal{L}$ is a *Borel measurable subbundle* of $M \times \mathbb{R}^m$ corresponding to the function $m(x)$ if $\mathcal{L} = \cup_{x\in M}(x, \mathcal{L}_x)$ and the map $x \mapsto \mathcal{L}_x$ restricted to each $\mathcal{U}_k$ is a Borel map of $\mathcal{U}_k$ into the Grassman manifold $\mathrm{Gr}_k(\mathbb{R}^m)$.

In other words, if $\mathcal{L}$ is a Borel measurable subbundle of $M \times \mathbb{R}^m$ then, for every $x \in \mathcal{U}_k$, the k-dimensional subspaces $\mathcal{L}_x$ depend measurably on x.

Now, we are going to state and present the proof of the Kifer non-random filtration theorem following the argument in [20].

Theorem 2.6.3 *Let* $F_1, F_2, \ldots$ *be a sequence of random linear cocycle with the common distribution* n *acting on* $M \times \mathbb{PR}^{d-1}$*. Assume that* n *and a* P^**-invariant ergodic probability measure* ρ *on* M *satisfy the condition*

$$\int\int (\log^+ \|\mathcal{J}_F(x)\| + \log^+ \|\mathcal{J}_{F^{-1}}(x)\|)\, d\rho(x)dn(F) < \infty\,. \tag{2.1}$$

Then one can choose a Borel set $M_\rho \subset M$ *with* $\rho(M_\rho) = 1$ *so that for any* $x \in M_\rho$ *there exists a sequence of (non-random) linear subspaces*

$$0 \subset \mathcal{L}_x^{r(\rho)} \subset \cdots \subset \mathcal{L}_x^1 \subset \mathcal{L}_x^0 = \mathbb{R}^m \tag{2.2}$$

and a sequence of (non-random) values

$$-\infty < \beta_{r(\rho)}(\rho) < \cdots < \beta_1(\rho) < \beta_0(\rho) < \infty$$

such that for p-almost all $\omega \in \Omega$,

$$\lim_{n\to\infty} \frac{1}{n} \log \|{}^n\mathcal{J}(x,\omega)\| = \beta_0(\rho)$$

and if $\xi \in \mathcal{L}_x^i \setminus \mathcal{L}_x^{i+1}$, where $\mathcal{L}_x^i = 0$ for all $i > r(\rho)$, then

$$\lim_{n\to\infty} \frac{1}{n} \log \|{}^n\mathcal{J}(x,\omega)\xi\| = \beta_i(\rho) \quad p\text{-a.s.}$$

Moreover, the numbers $\beta_i(\rho)$ are the values which the integrals

$$\gamma(\nu) = \int\int \log \frac{\|\mathcal{J}_F(x)u\|}{\|u\|}\, d\nu(x,u)\, dn(F) \tag{2.3}$$

take on for different ergodic measures $\nu \in \mathcal{N}_\nu$, where

$$\mathcal{N}_\rho = \{\nu \in \text{Prob}(M \times \mathbb{P}^{m-1}) : \nu \text{ is } n\text{-stationary measure and } \pi\nu = \rho\}$$

and $\pi\colon M \times \mathbb{P}^{m-1} \to M$ is the natural projection. Furthermore, the dimensions of $\mathcal{L}_x^i, i = 1, \cdots, r(\rho)$ do not depend on x for every $x \in M_\rho$ and $\mathcal{L}^i = \{\mathcal{L}_x^i\}$ form Borel measurable subbundles of $M_\rho \times \mathbb{R}^m$. These subbundles are F-invariant in the sense that

$$\mathcal{J}_F \mathcal{L}_x^i = \mathcal{L}_{fx}^i \quad \rho \times n\text{-a.s.}$$

where $f = \pi F \pi^{-1}$.

Proof. First, we establish the following useful lemmas:

Lemma 2.6.1 (Kronecker's lemma) *Given a convergent series $\sum_{n=1}^{\infty} a_n$,*

$$\limsup_{n\to\infty} \sum_{j=0}^{n} \frac{j}{n} a_j = 0\,.$$

Lemma 2.6.2 *Let Z_n be a Markov chain on a topological space M and K be the Markov kernel. If Q is the correspondent Markov operator and $g \in L^\infty(\Sigma)$, then with probability one*

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

Proof. Define the sequence of random variables

$$W_n := \sum_{j=0}^{n-1} \frac{1}{j+1} [Qg(Z_j) - g(Z_{j+1})]\,.$$

Then

$$W_{n+1} = \sum_{j=0}^{n} \frac{1}{j+1}[Qg(Z_j) - g(Z_{j+1})] = W_n + \frac{Qg(Z_n) - g(Z_{n+1})}{n+1}.$$

We claim that $\{(W_n, \mathcal{B}_n)\}$ is a martingale, where $\mathcal{B}_n = \sigma\{Z_0, \ldots, Z_n\}$. In fact, since W_n is determined by Z_j with $0 \leq j \leq n$, we have

$$\begin{aligned}\mathbb{E}[W_{n+1} \mid \mathcal{B}_n] &= W_n + \frac{1}{n+1}(\mathbb{E}[Qg(Z_n) \mid \mathcal{B}_n] - \mathbb{E}[g(Z_{n+1}) \mid \mathcal{B}_n]) \\ &= W_n + \frac{1}{n+1}(Qg(Z_n) - \mathbb{E}[g(Z_{j+1}) \mid Z_n])\,.\end{aligned}$$

On the other hand,

$$\mathbb{E}[g(Z_{j+1}) \mid Z_n = x] = \int g\, dK_x = Qg(x) = Qg(Z_n)\,.$$

Hence,

$$\mathbb{E}[W_{n+1} \mid \mathcal{B}_n] = W_n$$

and consequently, the process $\{(W_n\}$ is a martingale.

By Theorem 2.6.2 (Martingale Convergence Theorem), there exists a random variable $W\colon X \to \mathbb{R}$ such that W_n converges to W almost surely. By Lemma 2.6.1 (Kronecker's lemma) we have

$$\begin{aligned}0 &= \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} (Qg)(Z_j) - g(Z_{j+1}) \\ &= \lim_{n\to\infty} \frac{1}{n} \left\{ \sum_{j=0}^{n-1} [(Qg)(Z_j) - g(Z_j) + (g(Z_j) - g(Z_{j+1}))] \right\} \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} [(Qg)(Z_j) - g(Z_j)] + \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} [g(Z_j) - g(Z_{j+1})] \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} [(Qg)(Z_j) - g(Z_j)] + \lim_{n\to\infty} \frac{g(Z_0) - g(Z_n)}{n}\end{aligned}$$

Hence,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} [(Qg)(Z_j) - g(Z_j)] = 0$$

and this completes the proof. ■

We recall the following result from H. Furstenberg and Y. Kifer [13].

Theorem 2.6.4 (Furstenberg-Kifer) *Let (Σ, K) be a Markov system on a compact metric space Σ. Given a K-Markov process $\{Z_n\}_{n\geq 0}$ and $f \in C(\Sigma)$,* *then with probability one,*

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(Z_j) \le \sup\left\{\int_\Sigma f\, d\nu : \nu \in \mathrm{Prob}_K(\Sigma)\right\}.$$

Proof. Consider a dense sequence of functions $g_1, g_2, \ldots$ in $C(\Sigma)$ and the sequence of measures

$$\mu_n(\omega) := \frac{1}{n}\sum_{j=0}^{\infty} \delta_{Z_j(\omega)}.$$

For each $i \ge 1$, let $B_i \in \mathcal{B}$ be the full probability set in Lemma 2.6.2 associated with g_i. Then $B = \cap_{i=1}^{\infty} B_i \in \mathcal{B}$ is a full probability set such that for each $\omega \in \mathcal{B}$ and all $i \ge 1$,

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

This implies that if $\nu \in \mathrm{Prob}(\Sigma)$ is any accumulation point of μ_n then for all $j \ge 1$,

$$\int (Qg_j - g_j)\, d\nu = 0.$$

On the other hand, the sequence $g_1, g_2, \ldots$ is dense, then for all $g \in C(\Sigma)$,

$$\int g\, d(Q * \nu) = \int (Qg)\, d\nu = \int g\, d\nu$$

which proves that ν is a K-stationary probability measure on Σ.

Let

$$\beta := \sup\left\{\int f\, d\nu : \nu \in \mathrm{Prob}_K(\Sigma)\right\}.$$

Fix $\omega \in B$ and choose a sequence of integers n_i such that $\frac{1}{n_i}\sum_{j=0}^{n-1} f(Z_j(\omega))$ converges to the $\limsup$. Consider a subsequence of this sequence such that the corresponding sequence of measures $\mu_{n_i}(\omega)$ converges to $\nu \in \mathrm{Prob}(\Sigma)$ in the weak-* topology. Hence, $\nu \in \mathrm{Prob}_k(\Sigma)$ and

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(Z_j) = \lim_{i\to\infty} \frac{1}{n_i}\sum_{j=0}^{n_i-1} f(Z_j) = \int f\, d\nu \le \beta$$

and this completes the proof. ■

Applying Theorem 2.6.4 to f and $-f$ one obtains

Corollary 2.6.5 *Let* (Σ, K) *be a Markov system on a compact metric space* Σ*,* $\{Z_n\}_{n\ge 0}$ *be a* K*-Markov process,* $f \in C(\Sigma)$ *and assume that* $\int_\Sigma f\, d\nu = \beta$ *for every* K*-stationary probability measure* $\nu \in \mathrm{Prob}_K(\Sigma)$*. Then, with probability* *one*

$$\lim \frac{1}{n}\sum_{j=0}^{n-1} f(Z_j) = \beta.$$

Define on $M \times \mathbb{P}^{m-1} \times \mathcal{F}$ the following continuous function:

$$\psi(x, \hat{u}, F) := \log \frac{\|\mathcal{J}_F(x)u\|}{\|u\|}$$

where $u \in \mathbb{R}^m$ is a non-zero vector on the line corresponding to $\hat{u} \in \mathbb{P}^{m-1}$. Sometimes, we will write $w = (x, u) \in M \times \mathbb{P}^{m-1}$.

Furthemore, it is easy to check that

$$\frac{1}{n+1}\sum_{j=0}^{n} \psi(\mathcal{J}_F^j(w), F_{j+1}) = \frac{1}{n} \log \left\| \mathcal{J}_{F_{n+1}}(f^n(x)) \cdots \mathcal{J}_{F_1}(x)\frac{u}{\|u\|} \right\| \tag{2.4}$$

for every $w = (x, u) \in M \times \mathbb{P}^{m-1}$. Define $\tilde{\psi}\colon M \times \mathbb{P}^{m-1} \times \mathcal{F}^{\mathbb{N}} \to \mathbb{R}$ such that

$$\tilde{\psi}(w, \omega) = \psi(w, F_1(\omega))$$

the natural extension of ψ and define the skew product transformation

$$T(w, \omega) = (F_1(\omega)w, \sigma\omega)$$

acting on $M \times \mathbb{P}^{m-1} \times \mathcal{F}^{\mathbb{N}}$. Then

$$\psi(\mathcal{J}_F^j(w), F_{j+1}(\omega)) = \tilde{\psi} \circ T^k(w, \omega)\,. \tag{2.5}$$

Assume that $\nu \in \text{Prob}(M \times \mathbb{P}^{m-1})$ is an ergodic n-stationary measure. Since

$$\int \sup_{u \in \mathbb{P}^{m-1}} |\psi(x, u, F)\, d\rho(x)\, dn(F)| < \infty$$

and from (2.4) and (2.5), for ν-almost every $w = (x, u)$

$$\lim_{n\to\infty} \frac{1}{n+1} \log \left\| \mathcal{J}_{F_{n+1}}(f^n(x)) \cdots \mathcal{J}_{F_1}(x)\frac{u}{\|u\|} \right\| = \gamma(\nu) \tag{2.6}$$

where $\gamma(\nu)$ is defined by (2.3), that is,

$$\gamma(\nu) = \int\int \log \frac{\|\mathcal{J}_F(x)u\|}{\|u\|}\, d\nu(x, u)\, dn(F)\,.$$

Moreover, if $\pi\nu = \rho \in \text{Prob}(M)$, then (2.6) implies that for ρ-almost every $x \in M$

$$\liminf_{n\to\infty} \frac{1}{n+1} \log \left\| \mathcal{J}_{F_{n+1}}(f^n(x)) \cdots \mathcal{J}_{F_1}(x)\frac{u}{\|u\|} \right\| \geq \gamma(\nu), \quad p\text{-a.s.}$$

But then also

$$\liminf_{n\to\infty} \frac{1}{n+1} \log \left\| \mathcal{J}_{F_{n+1}}(f^n(x)) \cdots \mathcal{J}_{F_1}(x) \frac{u}{\|u\|} \right\| \geq \sup_{\nu\in\mathcal{N}_\rho} \gamma(\nu), \quad \rho \times p\text{-a.s.} \tag{2.7}$$

where $\mathcal{N}_\rho = \{\nu \in \text{Prob}(M \times \mathbb{P}^{m-1}) : \nu \text{ is } n\text{-stationary measure and } \pi\nu = \rho\}$.

Now, we are going to show that, in fact, the limit in (2.7) exists and it is equal to $\sup_{\nu\in\mathcal{N}_\rho} \gamma(\nu)$.

Lemma 2.6.3 *Let $\rho \in \text{Prob}(M)$ be an ergodic invariant measure satisfying*

$$\int\int (\log^+ \|\mathcal{J}_F(x)\| + \log^+ \|\mathcal{J}_{F^{-1}}(x)\|)\, d\rho(x) dn(F) < \infty\,.$$

Then, there exists a Borel set $\mathcal{U}_p \subset M$ with $\rho(\mathcal{U}_\rho) = 1$ such that p-almost surely

$$\lim_{n\to\infty} \frac{1}{n} \log \left\| \mathcal{J}_{F_{n+1}}(f^n(x)) \cdots \mathcal{J}_{F_1}(x) \right\| = \sup_{\nu\in\mathcal{N}_\rho} \gamma(\nu) = \beta_0(\rho)$$

for every $x \in \mathcal{U}_\rho$. Furthermore, if the linear functional γ is constant for all n-stationary measures $\nu \in \mathcal{N}_\rho$, that is, $\gamma(\nu) = \beta$, then

$$\lim_{n\to\infty} \frac{1}{n} \log \left\| \mathcal{J}_{F_{n+1}}(f^n(x)) \cdots \mathcal{J}_{F_1}(x)\xi \right\| = \beta \quad p\text{-a.s.}$$

for any nonzero $\xi \in \mathbb{R}^m$ and for every $x \in \mathcal{U}_\rho$.

Proof. Define $\psi_N \colon M \times \mathbb{P}^{m-1} \times \mathcal{F}^{\mathbb{N}} \to \mathbb{R}$ and $\Psi_N \colon M \times \mathbb{P}^{m-1} \to \mathbb{R}$ such that

$$\psi_N = \max\{-N, \min\{N, \psi\}\} \quad \text{and} \quad \Psi_N = \int \psi_N \, dn.$$

By a similar argument used in Lemma 2.6.2, we can prove that for $\nu \times p$-almost surely

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} (\psi_N(\mathcal{J}_F^j(w), F_{k+1}) - \Psi_N(\mathcal{J}_F^j(w))) = 0\,. \tag{2.8}$$

Consider the Markov chain $Y_n = F^n Y_0$ on the space $M \times \mathbb{P}^{m-1}$. Then, applying Theorem 2.6.4, for ρ-almost all initial points $X_0 = \pi Y_0$,

$$\limsup_{n\to\infty} \frac{1}{n+1} \sum_{k=0} \Psi_N(Y_k) \leq \sup_{\nu\in\mathcal{N}_\rho} \int \Psi_N \, d\nu \quad p\text{-a.s.}$$

In other words, for ρ-almost all $x \in M$ and all $u \in \mathbb{P}^{m-1}$,

$$\limsup_{n\to\infty} \frac{1}{n+1} \sum_{k=0} \Psi_N(F^k(x,u)) \leq \sup_{\nu\in\mathcal{N}_\rho} \int\int \psi_N(w,F)\, d\nu(w)\, dn(F) \quad p\text{-a.s.}$$

It follows that

$$|\psi(x,u,F)-\psi_N(x,u,F)| \leq (\log^+\|\mathcal{J}_F(x)\| + \log^+\|\mathcal{J}_F^{-1}(x)\| + N)\mathbb{1}_{B_N}(x,F) \tag{2.9}$$

where

$$B_N = \{(x,F) : \max\{\log^+\|\mathcal{J}_F(x)\|, \log^+\|\mathcal{J}_F^{-1}(x)\} > N\}\,.$$

Hence, for $w=(x,u)$

$$\begin{aligned}&\left|\frac{1}{n+1}\sum_{k=0}^{n}[\psi(F^k w, F_{k+1}) - \psi_N(F^k w, F_{k+1})]\right| \leq \\ &\leq \frac{1}{n+1}\sum_{k=0}^{n}[\log^+\|\mathcal{J}_{F_{k+1}}(f^k x)\| + \log^+\|\mathcal{J}_{F_{k+1}}^{-1}(f^k x)\| + N]\mathbb{1}_{B_N}(f^k x, F_{k+1})]\,.\end{aligned} \tag{2.10}$$

Applying the random ergodic theorem, we conclude that (2.10) converges $\rho\times p$-almost surely to the limit

$$\int_{B_N}(\log^+\|\mathcal{J}_F(x)\| + \log^+\|\mathcal{J}_F^{-1}(x)\|)\,d\rho(x)\,dn(F)\,. \tag{2.11}$$

By hypothesis (2.1), the expression in (2.11) tends to zero when $N\to\infty$. This together with (2.4), (2.8), (2.9), (2.10) and (2.11) implies that for ρ-almost all x and $u\in\mathbb{P}^{m-1}$,

$$\limsup_{n\to\infty}\frac{1}{n}\log\left\|\mathcal{J}_{F_{n+1}}(f^n(x))\cdots\mathcal{J}_{F_1}(x)\hat{u}\right\| \leq \sup_{\nu\in\mathcal{N}_\rho}\gamma(\nu) \quad p\text{-a.s.} \tag{2.12}$$

Let $\{\boldsymbol{x}_i\}$ be a orthonormal basis of $\mathbb{R}^m$. Since the inequality (2.12) holds p-almost surely for any ξ_j in place of $\hat{u}$ then it follows that $\rho\times p$-almost surely

$$\limsup_{n\to\infty}\frac{1}{n}\log\left\|\mathcal{J}_{F_{n+1}}(f^n(x))\cdots\mathcal{J}_{F_1}(x)\right\| \leq \sup_{\nu\in\mathcal{N}_\rho}\gamma(\nu)\,. \tag{2.13}$$

Combining (2.7) and (2.13) together with Corollary 2.6.5, we may conclude the proof. ■

For any set of non-zero vectors Γ in $\mathbb{R}^m$, denote by $\hat{\Gamma}$ the corresponding set of points in $\mathbb{P}^{m-1}$. For any measure $\nu\in\mathrm{Prob}(\mathbb{P}^{m-1})$ denote by $\mathcal{L}(\nu)$ the minimal linear subspace $\mathcal{L}$ of $\mathbb{R}^m$ satisfying $\nu(\hat{\mathcal{L}})=1$.

It is straightforward to show that $\mathcal{L}(\nu)=\cup\{(x,\mathcal{L}_x(\nu))\}$ forms a Borel measurable subbundles of $M\times\mathbb{R}^m$.

Lemma 2.6.4 *Let $\rho\in\mathrm{Prob}(M)$ be an ergodic n-stationary probability measure and $\nu\in\mathcal{N}_\rho$. Then there exists a measurable set $V(\nu)\subset M$ such that $\nu(V(\nu))=1$ and*

$$\mathcal{L}_{f(x)}(\nu) = \mathcal{J}_F(x)\,\mathcal{L}_x(\nu)$$

for n-almost all F and for every $x \in V(\nu)$, where $f = \pi F \pi^{-1}$. Moreover, the dimension $n_x(\nu)$ of $\mathcal{L}_x(\nu)$ is an invariant function and it is equal to a constant for ρ-almost every x.

Proof. Define $\hat{\mathcal{L}}(\nu) = \cup_x (x, \hat{\mathcal{L}}_x(\nu))$. Since ρ is an n-stationary and by definition of $\mathcal{L}$,

$$1 = \nu(\hat{\mathcal{L}}(\nu)) = \int \int \nu_x(\mathcal{J}_F^{-1}(x)\hat{\mathcal{L}}_{fx}(\nu))\, d\rho(x)\, dn(F)\,.$$

Hence,

$$\nu_x(\mathcal{J}_F^{-1}(x)\hat{\mathcal{L}}_{fx}(\nu)) = 1 \quad \rho \times n\text{-a.e.}\ (x, F).$$

Then, we can choose a Borel set $V(\nu)$ with $\rho(V(\nu)) = 1$ such that the previsous equality holds for any $x \in V(\nu)$ and n-almost every F. By the minimality of $\mathcal{L}_x(\nu)$ we conclude that for every $x \in V(\nu)$,

$$\mathcal{L}_x(\nu) \subset \mathcal{J}_F^{-1}\mathcal{L}_{fx}(\nu) \quad n\text{-a.s.}$$

Hence, the dimension $n_x(\nu)$ of $\mathcal{L}_x(\nu)$ satisfies

$$n_x(\nu) \le n_{fx}(\nu) \quad n\text{-a.s.}$$

Since

$$Pn_x(\nu) = \int n_{fx}(\nu)\, dn(F)$$

and ρ is n-stationary then from the previous inequality

$$0 \le \int (n_{fx}(\nu) - n_x(\nu))\, dn(F)\, d\rho(x) = \int (P_x(\nu) - n_x(\nu))\, d\rho(x) = 0$$

and then $n_{fx}(\nu) = n_x(\nu)$ for $\rho \times n$-almost surely.

Then

$$Pn_{fx}(\nu) = n_x(\nu) \quad \rho\text{-a.s.}$$

and since ρ is ergodic, we conclude that $n_x(\nu)$ is a constant for ρ-almost every x. And this conclude the proof. ■

Recall that a measurable subbundle $\mathcal{L} = \cup_x (x, \mathcal{L}_x)$ is F-invariant $\rho \times n$-almost surely if

$$\mathcal{J}_F \mathcal{L}_x = \mathcal{L}_{fx} \quad \rho \times n\text{-a.s.} \tag{2.14}$$

where $f = \pi F \pi^{-1}$. If ρ is an ergodic n-stationary probability measure and $\mathcal{L} = \cup_x (x, \mathcal{L}_x)$ is F-invariant $\rho \times n$-almost surely then the dimension

$$d_x(\mathcal{L}) = \dim \mathcal{L}_x = d(\rho, \mathcal{L}) = const \quad \rho\text{-a.s.}$$

Then, the subbundle $\mathcal{L}$ restricted to some Borel set $\mathcal{U}(\mathcal{L}) \subset M$ with $\rho(\mathcal{U}(\mathcal{L})) =$

1 is measurably isomorphic to the direct product $\mathcal{U}(\mathcal{L}) \times \mathbb{R}^{d(\rho,\mathcal{L})}$. This isomorphism can be represented by some family of linear maps $\mathrm{J}_{\mathcal{L}}(x) \colon \mathcal{L}_x \to \mathbb{R}^{d(\rho,\mathcal{L})}$ defined for $x \in \mathcal{U}(\mathcal{L})$ and such that $(x,\xi) \in \mathcal{L}$ corresponds to $(x, \mathrm{J}_{\mathcal{L}}(x)\xi) \in \mathcal{U}(\mathcal{L}) \times \mathbb{R}^{d(\rho,\mathcal{L})}$.

For each random linear cocycle F, define the map $F^{\mathcal{L}} \colon M \times \mathbb{R}^{d(\rho,\mathcal{L})} \to M \times \mathbb{R}^{d(\rho,\mathcal{L})}$ such that $F^{\mathcal{L}}(x,\eta) = (fx, \mathcal{J}_F^{\mathcal{L}}(x)\eta)$ where $x \in \mathcal{U}(\mathcal{L})$, $f = \pi F \pi^{-1}$ and

$$\begin{cases} \mathrm{J}_{\mathcal{L}}(x)\mathcal{J}_F(x) = \mathcal{J}_F^{\mathcal{L}}(x)\mathrm{J}_{\mathcal{L}}(x), & \text{if } (x,F) \text{ satisfies (2.14)} \\ \mathcal{J}_F^{\mathcal{L}}(x) = Id, & \text{if } (x,F) \text{ does not satisfies (2.14)} \end{cases}$$

A similar argument used in the proof of the Lemma 2.6.3, we may proof that there exists $\beta(\rho,\mathcal{L})$ such that with probability one

$$\lim_{n\to\infty} \frac{1}{n} \log \left\| \mathcal{J}_{F_n}^{\mathcal{L}}(f^{n-1}x) \cdots \mathcal{J}_{F_1}^{\mathcal{L}}(x) \right\| = \beta(\rho,\mathcal{L})$$

for every $x \in \mathcal{U}(\mathcal{L})$.

Let $\nu \in \mathcal{N}_\rho$ and $\rho \in \mathrm{Prob}(M)$ are both ergodic measures. By Lemma 2.6.4, the subbundle $\mathcal{L}(\nu)$ is measurable and F-invariant $\rho \times n$-a.s. Since $\nu(\mathcal{L}(\nu)) = 1$, it follows that

$$\beta(\rho, \mathcal{L}(\nu)) = \alpha(\nu).$$

Denote by $\mathcal{C}$ the collection of all F-invariant $\rho \times p$-a.s. measurable subbundles $\mathcal{L}$ satisfying $\beta(\rho,\mathcal{L}) < \beta_0(\rho)$, where

$$\beta_0(\rho) = \sup_{\nu\in\mathcal{N}_\rho} \gamma(\nu) = \sup_{\nu\in\mathcal{N}_\rho} \int\int \log \frac{\|\mathcal{J}_F(x)u\|}{\|u\|}\, d\nu(x,u)\, dn(F)\,.$$

If $\mathcal{C} = \emptyset$ then the filtration is trivial. Suppose now that $\mathcal{C}$ is not empty then there exists $\nu \in \mathcal{N}_\rho$ with $\mathcal{L}(\nu) \in \mathcal{C}$. Since $d(\rho,\mathcal{L}) \leq m$ then $\mathcal{C}$ has a maximal element $\mathcal{L}^{\max}$ which is uniquely determined. Moreover,

$$\beta_1(\rho) = \beta(\rho, \mathcal{L}^{\max}) = \sup_{\mathcal{L}\in\mathcal{C}} \beta(\rho,\mathcal{L}) < \beta_0(\rho). \tag{2.15}$$

We claim that $\mathcal{L}^{\max}$ can be taken as $\mathcal{L}^1$ in (2.2). In fact, suppose that $\mathcal{L}$ is an F-invariant $\rho \times n$-a.s. measurable subbundle. Consider the factor $(M \times \mathbb{P}^{m-1})/\mathcal{L}$ where each two points (x,ξ) and (x,χ) in $M \times \mathbb{P}^{m-1}$ are identified if $\xi - \chi \in \mathcal{L}_x$. Moreover, since $\mathcal{L}$ is F-invarian $n \times \rho$-a.s., we have

$$\mathcal{J}_F(x)\mathbb{R}^m/\mathcal{L}_x = \mathbb{R}^m/\mathcal{L}_{fx} \quad \rho \times n\text{-a.s.}$$

Given $x \in \mathcal{U}(\mathcal{L})$, we can apply similar arguments of the previous lemmas to obtain that $\rho \times p$-a.s. the limit

$$\lim_{n\to\infty} \frac{1}{n} \log \left\| \mathcal{J}^{\mathcal{L}}_{F_n}(f^{n-1}x) \cdots \mathcal{J}^{\mathcal{L}}_{F_1}(x) \right\| = \beta(\rho, (M \times \mathbb{P}^{m-1})/\mathcal{L})$$

exists and it is non-random.

Lemma 2.6.5 *Let $\mathcal{L}$ be an F-invariant $\mu \times \rho$-a.s. measurable subbundle then*

$$\beta_0(\rho) = \max\{\beta(\rho, \mathcal{L}), \beta(\rho, (M \times \mathbb{R}^m)/\mathcal{L})\}\,.$$

Proof. Let $\mathcal{U}(\mathcal{L})$ be a measurable subset of M such that $\rho(\mathcal{U}(\mathcal{L})) = 1$, for any $x \in \mathcal{U}(\mathcal{L})$ such that

$$\mathcal{J}_F \mathcal{L}_x = \mathcal{L}_{fx} \quad \rho \times n\text{-a.s.}$$

and $d_x(\mathcal{L}) = d(\rho, \mathcal{L})$. For those x and n-almost all F, the matrices $\mathcal{J}_F(x)$ have the following form

$$\mathcal{J}_F(x) = \begin{bmatrix} \mathcal{J}^{11}_F(x) & \mathcal{J}^{12}_F(x) \\ 0 & \mathcal{J}^{22}_F(x) \end{bmatrix}$$

where $\mathcal{J}^{ij}_F(x)$ are submatrices, $\mathcal{J}^{11}_F(x)$ corresponds to the restriction of $\mathcal{J}_F(x)$ to $\mathcal{L}_x$ and $\mathcal{J}^{22}_F(x)$ corresponds to the action of $\mathcal{J}_F(x)$ on $\mathbb{R}^m/\mathcal{L}_x$.

A simple calculation gives

$$\mathcal{J}_F(x)^n(x) = \begin{bmatrix} (\mathcal{J}^{11}_F)^n(x) & C_n(x) \\ 0 & (\mathcal{J}^{22}_F)^n(x) \end{bmatrix}$$

where

$$C_n(x) := \sum_{j=0}^{n-1} (\mathcal{J}^{11}_F)^{n-i-1}(T^i x)(\mathcal{J}^{12}_F)(T^i x)(\mathcal{J}^{22}_F)^i(x).$$

Since $\max\{\|(\mathcal{J}^{11}_F)^n\|, \|(\mathcal{J}^{22}_F)^n\|\} \le \|\mathcal{J}_F(x)^n\|$ we have

$$\max\left\{\limsup \frac{1}{n} \log \|(\mathcal{J}^{11}_F)^n\|, \right.$$
$$\left. \limsup \frac{1}{n} \log \|(\mathcal{J}^{22}_F)^n\| \right\} \le \limsup \frac{1}{n} \log \|(\mathcal{J}_F(x))^n\|$$

which implies that

$$\max\{\beta(\rho, \mathcal{L}), \beta(\rho, (M \times \mathbb{R}^m)/\mathcal{L})\} \le \beta_0(\rho).$$

On the other hand, since $\mathcal{J}^{12}_F$ is bounded, the above formula shows that $\|(\mathcal{J}_F(x))^n\|$ can never grow exponentially faster than both $\|(\mathcal{J}^{11}_F)^n\|$ and

$\|(\mathcal{J}_F^{22})^n\|$. Thus,

$$\limsup \frac{1}{n} \log \|(\mathcal{J}_F(x))^n\| \leq$$
$$\leq \max\left\{\limsup \frac{1}{n} \log \|(\mathcal{J}_F^{11})^n\|, \limsup \frac{1}{n} \log \|(\mathcal{J}_F^{22})^n\|\right\}$$

and then

$$\beta_0(\rho) \leq \max\{\beta(\rho, \mathcal{L}), \beta(\rho, (M \times \mathbb{R}^m)/\mathcal{L})\}\,.$$

■

By (2.15), $\beta(\rho, \mathcal{L}^{\max}) < \beta_0(\rho)$, then by Lemma 2.6.5, $\beta(\rho, (M \times \mathbb{R}^m)/\mathcal{L}) = \beta_0(\rho)$. Applying a similar argument of Lemma 2.6.3 to the vector bundle $(M \times \mathbb{R}^m)/\mathcal{L}^{\max}$, we conclude that either

$$\lim_{n\to\infty} \frac{1}{n} \log \left\|\mathcal{J}_{F_n}(f^{n-1}x) \cdots \mathcal{J}_{F_2}(fx)\mathcal{J}_{F_1}(x)\xi\right\| = \beta_1(\rho) \quad \rho \times p\text{-a.s.}$$

for every $\xi \in \mathbb{R}^m/\mathcal{L}_x^{\max}$ or by Lemma 2.6.4, there exists an F-invariant $\rho \times n$-almost surelly non-trivial measurable subbundle $\mathcal{A}$ of $(M \times \mathbb{R}^m)/\mathcal{L}^{\max}$ with $\beta(\rho, \mathcal{A}) < \beta_0(\rho)$. Hence,there exists an F-invariant $\rho \times n$-almost surelly measurable subbundle $\tilde{\mathcal{L}}$ of $M \times \mathbb{R}^m$ such that $\tilde{\mathcal{L}} > \mathcal{L}^{\max}$ and $d(\rho, \tilde{\mathcal{L}}) > d(\rho, \mathcal{L}^{\max})$. This contradicts the maximality of $\mathcal{L}^{\max}$ and proves that

$$\lim_{n\to\infty} \frac{1}{n} \log \left\|\mathcal{J}_{F_n}(f^{n-1}x) \cdots \mathcal{J}_{F_2}(fx)\mathcal{J}_{F_1}(x)\xi\right\| = \beta_1(\rho) \quad \rho \times p\text{-a.s.}$$

for every $\xi \in \mathbb{R}^m/\mathcal{L}_x^{\max}$.

Then, take $\mathcal{L}_1 = \mathcal{L}^{\max}$. To get the next term in the filtration, we repeat above arguments for $\mathcal{L}_1$ instead of $M \times \mathbb{R}^m$. And the proof finish by induction.

■

3
Uniform convergence rate for Birkhoff sums of torus translations

In this chapter we establish effective convergence rates for the Birkhoff average of toral translations. These and other related results were published in Klein, Liu and Melo [21].

In Section 3.1 we provide the framework for our results and state them formally. In Section 3.2 we review some basic concepts about continued fractions while in Section 3.3 we review some Fourier analysis notions. Finally, we obtain estimates on the uniform convergence rate of the Birkhoff averages of a continuous observable over one-dimensional torus translations (Section 3.4) and higher dimensional torus translations (Section 3.5).

3.1
Introduction and statements

Let us recall the Birkhoff's ergodic theorem for uniquely ergodic systems.

Theorem 3.1.1 *Let $(X, \mathcal{B}, \mu, T)$ be a uniquely ergodic system and let $f: X \to \mathbb{R}$ be a continuous function. Then, the convergence of the Birkhoff averages*

$$\lim \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) = \int_X f d\mu$$

is uniform on X.

In other words, for a uniquely ergodic system and a continuous observable, the corresponding Birkhoff averages converge everywhere and uniformly.

A natural question is then: can we estimate the *convergence rate* of the Birkhoff averages for certain types of uniquely ergodic systems and observables?

We obtain a positive answer to this question in the case of a Diophantine d-dimensional torus translation with a Hölder continuous observable. We obtained similar results also for affine skew product toral transformations (see [21]). We do not include them here since they are not very relevant to the rest of this work.

Let X be a metric space. We say that a map $\phi\colon X \to \mathbb{R}$ is *α-Hölder continuous* if there exist non-negative real constants C and $\alpha \in (0,1]$ such that

$$|\phi(x) - \phi(y)| \leq C d(x,y)^\alpha \quad \text{for all} \;\; x, y \in X.$$

We denote by $C^\alpha(X)$ the Banach space of α-Hölder continuous functions on X, endowed with the usual Hölder norm

$$\|\phi\|_\alpha := \|\phi\|_\infty + \sup_{x \neq y} \frac{|\phi(x) - \phi(y)|}{d(x,y)^\alpha}\,.$$

Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ be the one dimensional torus endowed with the Lebesgue measure and let $T\colon \mathbb{T} \to \mathbb{T}$, $Tx = x + \omega$ be the translation on $\mathbb{T}$ by an irrational frequency ω, which satisfies a Diophantine condition.

Definition 3.1.1 We say that a frequency $\omega \in \mathbb{T}$ satisfies a *(strong) Diophantine condition* if there exists $\gamma > 0$ such that for all $k \in \mathbb{Z} \setminus \{0\}$ we have

$$\|k\omega\| \geq \frac{\gamma}{|k| \, \log^2(|k|+1)}\,. \tag{3.1}$$

Denote by $\mathrm{DC}(\mathbb{T})_\gamma$ the set of all frequencies $\omega \in \mathbb{T}$ satisfying (3.1). Since the series $\sum_{k\geq 1} \frac{1}{k \log^2(k+1)}$ converges, by Khinchin's theorem (see Theorem 4 in [23, Chapter II]), the set of frequencies satisfying (3.1) for some $\gamma > 0$ has full measure.

Let $\phi\colon X \to \mathbb{R}$ be a continuous observable and for every integer N let

$$\phi^{(N)}(x) := \phi(x) + \phi(x+\omega) + \ldots + \phi(x + (N-1)\omega)$$

denote the corresponding N-th Birkhoff sum.

We estimate the *convergence rate* of the Birkhoff averages for a Diophantine torus translation with a Hölder continuous observable. More precisely, we obtain the follow result.

Theorem 3.1.2 *Assume that the observable ϕ is an α-Hölder continuous function on $\mathbb{T}$ and that the frequency $\omega \in \mathbb{T}$ satisfies the Diophantine condition* (3.1)*. Then for all integers N we have*

$$\left\| \frac{1}{N}\phi^{(N)} - \int_{\mathbb{T}} \phi \right\|_\infty \leq \text{const} \left(\frac{1}{\gamma} \log \frac{1}{\gamma}\right) \|\phi\|_\alpha \, \frac{\log^{3\alpha} N}{N^\alpha}\,.$$

Here and through the chapter const will refer to a universal constant (that could be made explicit) that may change from a line to another.

It turns out that this estimate is essentially optimal, in the sense that for almost every Diophantine frequency there exists an observable $\phi \in C^\alpha(\mathbb{T})$

and a subsequence $\{n_k\}_{k\geq 1}$ such that the corresponding Birkhoff averages of ϕ over the transformation T_ω satisfy

$$\left|\frac{1}{n_k}\phi^{(n_k)}(0) - \int_{\mathbb{T}} \phi\right| \geq \text{const}\frac{1}{n_k^\alpha} \quad \text{for all} \quad k \geq 1.$$

We are not providing the proof of this optimality here, as it is not directly relevant to the rest of this work. The interested reader may consult our paper [21, Theorem 2].

The above Theorem is an extension of Denjoy-Koksma's inequality proven by M. Herman in [17, Chapter VI.3]. In Section 3.4 we present a new proof of this result. Our approach uses Fourier analysis tools (effective approximation by trigonometric polynomials along with an effective rate of decay of the Fourier coefficients of continuous functions) and estimates of (in this case, some simple) exponential sums. The interested reader may compare our approach with the more direct, elementary argument employed by M. Herman in [17, Chapter VI.3].

While indeed more technical, our argument is versatile and modular, thus we also obtain a result for higher dimensional torus translations, although not as optimal one as in the one dimensional setting.

For a multi-index $\boldsymbol{k} = (k_1, \ldots, k_d) \in \mathbb{Z}^d$, let $|\boldsymbol{k}| := \max_{1\leq j\leq d} |k_j|$, and if $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{T}^d$, let $\boldsymbol{k}\cdot\boldsymbol{x} := k_1x_1 + \ldots + k_dx_d$.

Definition 3.1.2 We say that $\boldsymbol{\omega} \in \mathbb{T}^d$ satisfies *a Diophantine condition* if there exist $\gamma > 0$ and $A > d$ such that

$$\|\boldsymbol{k}\cdot\boldsymbol{\omega}\| = \text{dist}\,(\boldsymbol{k}\cdot\boldsymbol{\omega},\, \mathbb{Z}) \geq \frac{\gamma}{|\boldsymbol{k}|^A} \tag{3.2}$$

for all $\boldsymbol{k} \in \mathbb{Z}^d$ with $|\boldsymbol{k}| \neq 0$. Denote by $\text{DC}(\mathbb{T}^d)_{\gamma,A}$ the set of all frequency vectors $\boldsymbol{\omega}$ satisfying the Diophantine condition (3.2). Given any $A > d$, the set $\bigcup_{\gamma>0}\text{DC}(\mathbb{T}^d)_{\gamma,A}$ has full measure.

Next we formulate our result on the rate of convergence of Birkhoff means of Hölder observables over a higher dimensional toral translation.

Theorem 3.1.3 *Let $\phi \in C^\alpha(\mathbb{T}^d)$, let $\boldsymbol{\omega} \in \text{DC}(\mathbb{T}^d)_{\gamma,A}$ and let $T_{\boldsymbol{\omega}}\colon \mathbb{T}^d \to \mathbb{T}^d$ be the corresponding torus translation. Then for all $N \geq 1$ we have*

$$\left\|\frac{1}{N}\,\phi^{(N)} - \int_{\mathbb{T}^d}\phi\right\|_\infty \leq \frac{\text{const}}{\gamma}\,\|\phi\|_\alpha\,\frac{1}{N^\beta}$$

where $\beta = \frac{\alpha}{A+d}$.

3.2 Some arithmetic considerations

We begin with a review of some basic arithmetic properties that will be used latter. For more detais, good references are [30, Chapter 3] and [23].

Let $\omega \in \mathbb{T} \simeq [0,1)$, consider its continued fraction expansion

$$\omega = [a_0, \ldots, a_n, \ldots] = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \cdots}}$$

and for each $n \geq 1$, the corresponding n-th (principal) convergent

$$\frac{p_n}{q_n} = [a_0, \ldots, a_n],$$

where the integers $p_n = p_n(\omega), q_n = q_n(\omega)$ are relatively prime.

It is easy to verify the following recursion formulas.

Theorem 3.2.1 *For $n \geq 2$, we have that*

$$p_{n+2} = a_{n+2} p_{n+1} + p_n,$$
$$q_{n+2} = a_{n+2} q_{n+1} + q_n.$$

The following proposition ensures that the sequence $\{q_n\}_{n\geq 1}$ of denominators of the convergents of ω is strictly increasing (for its proof, see [23, Corollary 2]).

Proposition 3.2.2 *If $a_1, a_2, \ldots$ are positive integers, then p_n and q_n are relatively primes, and the denominators $0 < q_1 < q_2 < \ldots$ form a strictly increasing sequence of integers.*

By Theorem 3.2.1 and Proposition 3.2.2, $q_{n+2} = a_{n+1}q_{n+1} + q_n \geq 2q_n$, so $q_{n+2k} \geq 2^k q_n$ for all integers n, k.

Theorem 3.2.3 *For even n, the n-th principal convergents of ω form a strictly increasing sequence converging to ω. On the other hand, for odd n, the n-th principal convergents of ω form a strictly decreasing sequence converging to ω. Furthermore, the following inequalities hold:*

$$\frac{1}{2q_{n+1}} < \frac{1}{q_{n+1} + q_n} < |q_n\omega - p_n| < \frac{1}{q_{n+1}}. \tag{3.3}$$

We denote by $\|\omega\|$ the distance between ω and the nearest integer. Then $\|q_n\omega\| = |q_n\omega - p_n|$.

Definition 3.2.4 We say that a fraction $\frac{p}{q}$, $(q > 0)$, is a *best approximation* to ω if

$$\|q\omega\| = |q\omega - p| \quad \text{and} \quad \|q'\omega\| > \|q\omega\|$$

for every $1 \leq q' < q$.

It turns out that the best approximations to ω are precisely its principal convergents. In fact, q_{n+1} is the smallest integer $j > q_n$ such that $\|j\omega\| < \|q_n\omega\|$ (see [23, Chapter 1, Theorem 6]). Noting also that $\|-t\| = \|t\|$, we conclude the following.

Proposition 3.2.5 *Given $\omega \in (0,1) \setminus \mathbb{Q}$, if $\frac{p}{q}$ is a best approximation to ω then*

$$\|j\omega\| > \frac{1}{2q} \quad \textit{for all} \quad 1 \leq |j| < q\,. \tag{3.4}$$

Proof. Suppose that $q = q_n$. By inequality (3.3),

$$\frac{1}{2q_n} < |q_{n-1}\alpha - p_{n-1}| = \|q_{n-1}\omega\|.$$

Since $|j| < q_n$ and the best approximations to ω are its principal convergents, we have that $\|q_{n-1}\omega\| < \|j\omega\|$. ■

If $\omega \in \mathrm{DC}(\mathbb{T})_\gamma$, since by (3.3) we have $\|q_n\omega\| = |q_n\omega - p_n| < \frac{1}{q_{n+1}}$, it follows that

$$q_{n+1} \leq \frac{1}{\gamma}\, q_n \log^2(q_n + 1) \quad \text{for all } n \geq 1. \tag{3.5}$$

3.3 Effective approximation by trigonometric polynomials

In this section we review some Fourier analysis notions on the additive group $\mathbb{T}^d, d \geq 1$. We start our review in the case $d = 1$ (see [32] or [19] for more details).

Given a continuous observable $\phi \colon \mathbb{T} \to \mathbb{R}$, consider the Fourier series associated with ϕ

$$\phi(x) \sim \sum_{k=-\infty}^{\infty} \widehat{\phi}(k)\, e(kx) = \int_{\mathbb{T}} \phi + \sum_{k \neq 0} \widehat{\phi}(k)\, e(kx)\,,$$

where $e(x) := e^{2\pi i x}$ and

$$\widehat{\phi}(k) = \int_0^1 \phi(x) e(-kx) dx$$

is the *k-th Fourier coefficient* of ϕ.

For every $n \geq 0$, denote by

$$S_n\phi(x) := \sum_{|k|\leq n} \widehat{\phi}(k)\, e(kx)$$

the n-th partial sum of the Fourier series of ϕ

Proposition 3.3.1 *If $\phi \in \mathbb{C}^\alpha(\mathbb{T})$, then the Fourier coefficients of ϕ have the decay*

$$\left|\widehat{\phi}(k)\right| \leq \text{const } \|\phi\|_\alpha \frac{1}{|k|^\alpha} \quad \textit{for all} \;\; k \neq 0.$$

Proof. Note that

$$\hat{\phi}(k) = \frac{1}{2\pi}\int_0^{2\pi} \phi(x)e^{-ikx}e^{-i\pi}dx = -\frac{1}{2\pi}\int_0^{2\pi} \phi(x)e^{-ik\left(x+\frac{\pi}{k}\right)}dx\,.$$

Then, by a change of variables,

$$\begin{aligned}
2\hat{\phi}(k) &= -\frac{1}{2\pi}\int_0^{2\pi} \phi(x)e^{-ik\left(x+\frac{\pi}{k}\right)}dx + \frac{1}{2\pi}\int_0^{2\pi} \phi(x)e^{-ikx}dx \\
&= \frac{1}{2\pi}\int_0^{2\pi} \left[\phi\left(x+\frac{\pi}{k}\right) - \phi(x)\right] e^{-ik\left(x+\frac{\pi}{k}\right)}dx \\
&= \frac{1}{2\pi}\int_0^{2\pi} \left[\phi\left(x+\frac{\pi}{k}\right) - \phi(x)\right] e^{-ikx}e^{-i\pi}dx \\
&= -\frac{1}{2\pi}\int_0^{2\pi} \left[\phi\left(x+\frac{\pi}{k}\right) - \phi(x)\right] e^{-ikx}dx
\end{aligned}$$

Since ϕ is a α-Hölder continuous map,

$$|\hat{\phi}(k)| \leq C\frac{\pi^\alpha}{|k|^\alpha} = \mathcal{O}\left(\frac{1}{|k|^\alpha}\right).$$

■

Definition 3.3.1 (Convolution) A typical construction of good trigonometric approximations uses integral operators of *convolution* type:

$$\varphi * K_n(x) := \int_{\mathbb{T}} \varphi(x-t)K_n(t)dt = \int_{\mathbb{T}^2} \varphi(t)K_n(x-t)dt,$$

where K_n is a trigonometric polynomial of degree n.

Since K_n is a trigonometric polynomial of degree n, the convolution $\varphi * K_n(x)$ is a trigonometric polynomial of degree $\leq n$.

Lemma 3.3.1 *Let $K_n : \mathbb{T} \to \mathbb{R}$ be a trigonometric polynomial of degree n that satisfies the following conditions for some constant $L < \infty$:*

1) $\int_{\mathbb{T}} K_n(t)dt = 1$,
2) $K_n(t) = K_n(-t)$,
3) $\int_0^{\pi} |K_n(t)|\, dt_1 \leq L$,
4) $\int_0^{\pi} nt|K_n(t)|\, dt_1 \leq L$.

Let $\varphi\colon \mathbb{T} \to \mathbb{R}$ *be an* α*-Hölder continuous function. Then there is* $M < \infty$ *such that*

$$\|\varphi * K_n - \varphi\|_\infty \leq M \cdot \frac{1}{|n|^\alpha}$$

.

Proof. By (1)-(3)

$$\begin{aligned}
|\varphi{*}K_n(x) - \varphi(x)| &= \left|\int_{\mathbb{T}} [\varphi(x-t) - \varphi(x)]\, K_n(t)\, dt\right| \\
&= 2\left|\int_0^{\pi} [\varphi(x-t) - \varphi(x)]\, K_n(t)\, dt\right| \\
&= 2\left|\int_0^{\pi} [\varphi(x+t) - \varphi(x)]\, K_n(-t)\, dt + \int_0^{\pi} [\varphi(x-t) - \varphi(x)]\, K_n(t)\, dt\right| \\
&= 2\left|\int_0^{\pi} [\varphi(x-t) - 2\varphi(x) + \varphi(x+t)]\, K_n(t)\, dt\right| \\
&\leq \int_0^{\pi} 2C|t|^\alpha |K_n(t)|\, dt \\
&= \int_0^{\pi} 2C\left|\frac{nt}{n}\right|^\alpha |K_n(t)|\, dt \\
&\leq 2C\left|\frac{1}{n}\right|^\alpha \int_0^{\pi} (nt+1)|K_n(t)|\, dt \\
&\leq M|n|^{-\alpha}
\end{aligned}$$

where $M = 2C(L+1)$.

Therefore,

$$\|\varphi * K_n - \varphi\|_\infty \leq M\, \frac{1}{|n|^\alpha}.$$

■

A family of functions $\{K_n\colon \mathbb{T} \to \mathbb{R}\}$ that satisfies the conditions (1)-(4) in Lemma 3.3.1 is sometimes called a *good kernel.*

Let

$$F_n(x) = \sum_{|k|\leq n} \left(1 - \frac{|k|}{n}\right) e(kx) = \frac{1}{n}\, \frac{\sin^2(nx/2)}{\sin^2(x/2)}$$

be the n-th Fejér kernel.

For all $n \in \mathbb{N}$, let $J_n\colon \mathbb{T} \to \mathbb{R}$ be n-th Jackson kernel, that is,

$$J_n(x) := c_n\, F_m^2(x)$$

where $m = \lfloor \frac{n}{2} \rfloor$, $F_m(x)$ is the Fejér kernel and $c_n \asymp \frac{1}{n}$ is a normalizing factor chosen so that $\int_{\mathbb{T}} J_n = 1$. Therefore, $J_n(x)$ is a trigonometric polynomial of degree $\leq n$ and uniformly in n and k, its (Fourier) coefficients have the bound $\left|\widehat{J_n}(k)\right| \lesssim 1$[1].

It turns out that the Jackson kernel $J_n(x)$ satisfies the hypotheses of Lemma 3.3.1. Hence, an important consequence that will be crucial in our proof is the following.

Theorem 3.3.2 (Jackson) *Given $\phi \in C^\alpha(\mathbb{T})$ with $\alpha \in (0,1]$, if we denote*

$$\phi_n(x) := (\phi * J_n)(x)\,,$$

then ϕ_n is a trigonometric polynomial of degree $\leq n$ and

$$\left\|\phi - \phi_n\right\|_\infty \leq \text{const } \left\|\phi\right\|_\alpha \frac{1}{n^\alpha}\,. \tag{3.6}$$

This is a quantitative version of the Weierstrass approximation theorem. Thus α-Hölder continuous functions can be approximated by trigonometric polynomials of degree $\leq n$ with error bound of order $\frac{1}{n^\alpha}$, which is *optimal* (this result is called Jackson's theorem, see [19] or [31]).

As in dimension one, now we will review some Fourier analysis notions on the additive group $\mathbb{T}^d, d \geq 2$ (see [19, Chapter 1.9] and [16, Chapter 3] for more details).

For every multi-index $\boldsymbol{k} \in \mathbb{Z}^d$, define the multiplicative characters $e_{\boldsymbol{k}} \colon \mathbb{T}^d \to \mathbb{C}$ by $e_{\boldsymbol{k}}(\boldsymbol{x}) := e(\boldsymbol{k} \cdot \boldsymbol{x})$, where $\boldsymbol{k} \cdot \boldsymbol{x} = k_1 x_1 + \ldots + k_d x_d$.

Let $\phi \in L^2(\mathbb{T}^d)$ and let $\boldsymbol{k} \in \mathbb{Z}^d$ be a multi-index. The corresponding Fourier coefficient of ϕ is then

$$\widehat{\phi}(\boldsymbol{k}) = \int_{\mathbb{T}^d} \phi(\boldsymbol{x}) \overline{e_{\boldsymbol{k}}(\boldsymbol{x})}\, d\boldsymbol{x} = \int_{\mathbb{T}^d} \phi(\boldsymbol{x}) e(-\boldsymbol{k} \cdot \boldsymbol{x})\, d\boldsymbol{x}\,.$$

Moreover, the Fourier coefficients of a function $\phi \in C^\alpha(\mathbb{T}^d)$ have the decay

$$\left|\widehat{\phi}(\boldsymbol{k})\right| \leq \text{const } \left\|\phi\right\|_\alpha \frac{1}{|\boldsymbol{k}|^\alpha} \quad \text{for all } \ \boldsymbol{k} \in \mathbb{Z}^d,\ |\boldsymbol{k}| \neq 0\,. \tag{3.7}$$

This follows from Fubini's theorem and the corresponding one variable estimate.

Consider the Fourier series expansion of $\phi \in L^2(\mathbb{T}^d)$:

$$\sum_{\boldsymbol{k} \in \mathbb{Z}^d} \widehat{\phi}(\boldsymbol{k})\, e_{\boldsymbol{k}}(\boldsymbol{x}) = \int_{\mathbb{T}^d} \phi + \sum_{|\boldsymbol{k}| \neq 0} \widehat{\phi}(\boldsymbol{k})\, e_{\boldsymbol{k}}(\boldsymbol{x})\,.$$

[1]For two *varying* quantities a and b, $a \lesssim b$ will mean that $a \leq \text{const}\, b$, where const is a universal constant.

Given $n \geq 0$, the *n-th (square) partial sum* of the Fourier series of ϕ is

$$S_n\phi(x) := \sum_{|\boldsymbol{k}|\leq n} \hat{\phi}(\boldsymbol{k})\, e_{\boldsymbol{k}}(\boldsymbol{x})\,,$$

which, as in the one variable case, may in general fail to converge (even pointwise) to ϕ.

We define the d-dimensional (square) Jackson kernel $\mathbf{J}_n \colon \mathbb{T}^d \to \mathbb{R}$ as

$$\mathbf{J}_n(x_1, \ldots, x_d) := J_n(x_1) \cdot \ldots \cdot J_n(x_d)\,. \tag{3.8}$$

Then (essentially by Fubini's theorem) $\mathbf{J}_n$ has similar properties to those of its one dimensional counterpart. More precisely, $\left|\widehat{\mathbf{J}_n}(\boldsymbol{k})\right| \lesssim 1$ uniformly in n and $\boldsymbol{k}$. Moreover, $\phi_n := \phi * \mathbf{J}_n$ is a trigonometric polynomial in d variables of degree $\leq n$ and if $\phi \in C^\alpha(\mathbb{T}^d)$, then for all $n \geq 1$,

$$\|\phi_n - \phi\|_\infty \leq \text{const } \|\phi\|_\alpha \frac{1}{n^\alpha}\,. \tag{3.9}$$

Furthermore, we have the following estimate on the (Fourier) coefficients of ϕ_n. If $\boldsymbol{k} \in \mathbb{Z}^d$ with $0 < |\boldsymbol{k}| \leq n$ then

$$\left|\widehat{\phi_n}(\boldsymbol{k})\right| = \left|\widehat{\phi * \mathbf{J}_n}(\boldsymbol{k})\right| = \left|\hat{\phi}(\boldsymbol{k})\right| \left|\widehat{\mathbf{J}_n}(\boldsymbol{k})\right| \leq \text{const } \|\phi\|_\alpha \frac{1}{|\boldsymbol{k}|^\alpha}\,, \tag{3.10}$$

where the last inequality follows from (3.7) and the fact that $\left|\widehat{\mathbf{J}_n}(\boldsymbol{k})\right| \lesssim 1$.

3.4 The one dimensional torus translation case

In this section, we establish Theorem 3.1.2 on the convergence rate for the Birkhoff averages of a Diophantine one-dimensional torus translations for Hölder observables.

Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ be the one dimensional torus endowed with the Lebesgue measure and let $T \colon \mathbb{T} \to \mathbb{T}$, $Tx = x + \omega$ be the translation on $\mathbb{T}$ by an irrational frequency ω, which satisfies a generic Diophantine condition (3.1).

Let $\phi \colon X \to \mathbb{R}$ be a continuous observable and for every integer N let

$$\phi^{(N)}(x) := \phi(x) + \phi(x + \omega) + \ldots + \phi(x + (N-1)\omega)$$

denote the corresponding Birkhoff sum.

Consider the Fourier series associated with the observable ϕ:

$$\phi(x) \sim \sum_{k=-\infty}^{\infty} \hat{\phi}(k)\, e(kx) = \int_{\mathbb{T}} \phi + \sum_{k\neq 0} \hat{\phi}(k)\, e(kx)\,.$$

Then for all $j \in \mathbb{Z}$, the Fourier series of $\phi(x + j\omega) - \int_{\mathbb{T}} \phi$ is

$$\sum_{k \neq 0} \widehat{\phi}(k)\, e(k(x + j\omega)) = \sum_{k \neq 0} \widehat{\phi}(k)\, e(jk\omega)\, e(kx)\,.$$

Hence,

$$\begin{aligned}
\frac{1}{N}\phi^{(N)}(x) - \int_{\mathbb{T}} \phi &= \frac{1}{N}\left[\phi(x) + \phi(x+\omega) + \ldots + \phi(x + (N-1)\omega) - N \int_{\mathbb{T}} \phi\right] \\
&= \frac{1}{N} \sum_{j=0}^{N-1} \sum_{k \neq 0} \widehat{\phi}(k)\, e(jk\omega)\, e(kx) \\
&= \sum_{k \neq 0} \widehat{\phi}(k) \left(\frac{1}{N} \sum_{j=0}^{N-1} e(jk\omega)\right) e(kx)\,.
\end{aligned}$$

It follows that the Fourier series of $\frac{1}{N}\phi^{(N)}(x) - \int_{\mathbb{T}} \phi$ is

$$\sum_{k \neq 0} \widehat{\phi}(k) \left(\frac{1}{N} \sum_{j=0}^{N-1} e(jk\omega)\right) e(kx) = \sum_{k \neq 0} \widehat{\phi}(k)\, \mathcal{E}_N(k\omega)\, e(kx)\,, \tag{3.11}$$

where $\mathcal{E}_N$ refers to the (averaged) exponential sum

$$\mathcal{E}_N(t) := \frac{1}{N} \sum_{j=0}^{N-1} e(jt).$$

Clearly $|\mathcal{E}_N(t)| \leq 1$. Moreover, since $\mathcal{E}_N(t)$ is the sum of a finite geometric sequence, it follows that

$$|\mathcal{E}_N(t)| = \frac{1}{N}\left|\frac{e(t)^N - 1}{e(t) - 1}\right| = \frac{1}{N}\left|\frac{1 - e(Nt)}{1 - e(t)}\right| \lesssim \frac{1}{N\,\|t\|}\,. \tag{3.12}$$

Lemma 3.4.1 *Let $\frac{p}{q}$ be a best approximation to the irrational number ω. Then*

$$\sum_{1 \leq |k| < q} |\mathcal{E}_N(k\omega)| \leq \text{const}\, q\, \frac{\log q}{N}\,.$$

Proof. Since $\|t\| = \|-t\|$ for all $t \in \mathbb{R}$, it is enough to bound $\sum_{1 \leq k < q} |\mathcal{E}_N(k\omega)|$.

Recall that by (3.4), for all $1 \leq |j| < q$,

$$\|j\omega\| > \frac{1}{2q}\,.$$

Thus for every $k, k' \in \{1, \ldots, q-1\}$ with $k \neq k'$ we have

$$\|k\omega - k'\omega\| = \|(k - k')\omega\| > \frac{1}{2q}\,. \tag{3.13}$$

Divide $\mathbb{T}$ into the $2q$ arcs

$$C_j = \left[\frac{j}{2q}, \frac{j+1}{2q}\right), \quad 0 \leq j \leq 2q-1$$

with equal length $|C_j| = \frac{1}{2q}$.

By the observation in (3.13), each arc C_j contains at most one point $k\omega \mod 1$ with $k \in \{1, \dots, q-1\}$, and by (3.4), the arcs $C_0 = [0, \frac{1}{2q})$ and $C_{2q-1} = [\frac{2q-1}{2q}, 1)$ do *not* contain any such point.

Moreover, if $x \in C_j$ with $1 \leq j \leq q-1$, then $x \in \left[0, \frac{1}{2}\right)$, and consequently, $\|x\| \geq \frac{j}{2q}$. On the other hand, if $x \in C_k$, with $q \leq k \leq 2q-1$, thus $x \in \left[\frac{1}{2}, 1\right)$. In this case, $\|x\| \geq 1 - \frac{k+1}{2q} = \frac{2q-k+1}{2q}$. Take $j = 2q-k-1$, then $k = 2q-j-1$ and

$$q \leq k \leq 2q-1 \Rightarrow 0 \leq 2q-k-1 \leq q-1 \Rightarrow 0 \leq j \leq q-1.$$

Hence, if $x \in C_k$ with $q \leq k \leq 2q-1$, it follows that $x \in C_{2q-j-1}$ with $0 \leq j \leq q-1$ and $\|x\| \geq \frac{j}{2q}$. Using (3.12), it follows that

$$\sum_{1 \leq k < q} |\mathcal{E}_N(k\omega)| \leq \sum_{1 \leq k < q} \frac{1}{N\,\|k\omega\|} \leq \frac{1}{N} \sum_{j=1}^{q-1} \frac{4q}{j} \leq \text{const}\, \frac{q}{N} \log q\,,$$

which proves the lemma. ■

Theorem 3.4.1 *Assume that the observable ϕ is an α-Hölder continuous function on $\mathbb{T}$ and that the frequency $\omega \in \mathbb{T}$ satisfies the Diophantine condition* (3.1)*. Then for all integers N we have*

$$\left\| \frac{1}{N}\, \phi^{(N)} - \int_{\mathbb{T}} \phi \right\|_{\infty} \leq \text{const} \left(\frac{1}{\gamma} \log \frac{1}{\gamma}\right) \|\phi\|_{\alpha}\, \frac{\log^{3\alpha} N}{N^{\alpha}}\,. \tag{3.14}$$

Proof. Let $\phi \in C^{\alpha}(\mathbb{T})$, fix $N \geq 1$ (the length of the Birkhoff sum) and let $1 \leq n \leq N$ (the degree of polynomial approximation) to be chosen later. Write

$$\phi = \phi_n + (\phi - \phi_n) =: \phi_n + \psi_n\,,$$

which implies that

$$\frac{1}{N}\, \phi^{(N)} - \int_{\mathbb{T}} \phi = \left(\frac{1}{N}\, \phi_n^{(N)} - \int_{\mathbb{T}} \phi_n\right) + \left(\frac{1}{N}\, \psi_n^{(N)} - \int_{\mathbb{T}} \psi_n\right). \tag{3.15}$$

From (3.6) we clearly have

$$\left\| \frac{1}{N}\, \psi_n^{(N)} - \int_{\mathbb{T}} \psi_n \right\|_{\infty} \leq \text{const}\, \|\phi\|_{\alpha}\, \frac{1}{n^{\alpha}}\,. \tag{3.16}$$

The proof is then reduced to estimating the N-th Birkhoff average of the trigonometric polynomial (of degree $\leq n$) $\phi_n = \phi * J_n$. By (3.11) applied to

ϕ_n, for all $x \in \mathbb{T}$ we have:

$$\frac{1}{N}\,\phi_n^{(N)}\,(x) - \int_{\mathbb{T}} \phi_n = \sum_{1\le |k| \le n} \widehat{\phi_n}(k)\,\mathcal{E}_N(k\omega)\,e(kx)\,.$$

For all $k \neq 0$, the corresponding Fourier coefficient of $\phi_n = J_n * \phi$ satisfies

$$\left|\widehat{\phi_n}(k)\right| = \left|\widehat{J_n}(k)\cdot\hat{\phi}(k)\right| \lesssim \left|\hat{\phi}(k)\right| \lesssim \|\phi\|_\alpha\,\frac{1}{|k|^\alpha}\,.$$

Since the sequence $\{q_j\}_{j\ge 1}$ of the denominators of the principal convergents of ω is strictly increasing (Theorem 3.2.3), there is an integer s such that $q_s \le n < q_{s+1}$.

Thus combining with the preceding, for all $x \in \mathbb{T}$ we have:

$$\begin{aligned}
\left|\frac{1}{N}\,\phi_n^{(N)}\,(x) - \int_{\mathbb{T}} \phi_n\right| &\le \sum_{1\le |k|\le n} \left|\widehat{\phi_n}(k)\right|\,|\mathcal{E}_N(k\omega)| \\
&\le \sum_{1\le |k|< q_{s+1}} \left|\widehat{\phi_n}(k)\right|\,|\mathcal{E}_N(k\omega)| \\
&\lesssim \|\phi\|_\alpha \sum_{1\le |k|<q_{s+1}} \frac{1}{|k|^\alpha}\,|\mathcal{E}_N(k\omega)| \\
&= \|\phi\|_\alpha \sum_{j=1}^{s} \sum_{q_j\le |k|<q_{j+1}} \frac{1}{|k|^\alpha}\,|\mathcal{E}_N(k\omega)| \\
&\le \|\phi\|_\alpha \sum_{j=1}^{s} \frac{1}{q_j^\alpha} \sum_{1\le |k|<q_{j+1}} |\mathcal{E}_N(k\omega)| \\
&\lesssim \|\phi\|_\alpha\,\frac{1}{N}\left(\sum_{j=1}^{s} \frac{q_{j+1}\log q_{j+1}}{q_j^\alpha}\right)
\end{aligned} \tag{3.17}$$

where the last estimate follows from Lemma 3.4.1.

If $\omega \in \mathrm{DC}(\mathbb{T})_\gamma$, then using (3.5), for all $1 \le j \le s$ we have:

$$\begin{aligned}
\frac{q_{j+1}\log(q_{j+1})}{q_j^\alpha} &\lesssim \frac{\left(\frac{1}{\gamma}\,q_j\log^2(q_j+1)\right)\,\left(\log\frac{1}{\gamma} + \log(q_j+1)\right)}{q_j^\alpha} \\
&\lesssim \left(\frac{1}{\gamma}\,\log\frac{1}{\gamma}\right)\,q_j^{1-\alpha}\,\log^3(q_j+1) \\
&\lesssim \left(\frac{1}{\gamma}\,\log\frac{1}{\gamma}\right)\,q_j^{1-\alpha}\,\log^3 n\,.
\end{aligned}$$

As mentioned earlier, $q_n \le 2^{-k} q_{n+2k}$ for all $n, k \in \mathbb{N}$, hence

$$\sum_{j=1}^{s} q_j^{1-\alpha} = \sum_{\substack{1\le j\le s\\ j \text{ odd}}} q_j^{1-\alpha} + \sum_{\substack{1\le j\le s\\ j \text{ even}}} q_j^{1-\alpha} \lesssim q_s^{1-\alpha} \le n^{1-\alpha}\,.$$

We obtained the following: if $\omega \in \mathrm{DC}(\mathbb{T})_\gamma$ and $q_s \le n < q_{s+1}$, then

$$\sum_{j=1}^{s} \frac{q_{j+1} \log q_{j+1}}{q_j^\alpha} \lesssim \left(\frac{1}{\gamma} \log \frac{1}{\gamma}\right) \log^3 n \sum_{j=1}^{s} q_j^{1-\alpha}$$
$$\lesssim \left(\frac{1}{\gamma} \log \frac{1}{\gamma}\right) n^{1-\alpha} \log^3 n \,. \tag{3.18}$$

Combining (3.15), (3.16), (3.17), and (3.18) we obtain

$$\left\| \frac{1}{N} \phi^{(N)} - \int_{\mathbb{T}} \phi \right\|_\infty \le \left\| \left(\frac{1}{N} \phi_n^{(N)} - \int_{\mathbb{T}} \phi_n \right) \right\|_\infty + \left\| \left(\frac{1}{N} \psi_n^{(N)} - \int_{\mathbb{T}} \psi_n \right) \right\|_\infty$$
$$\lesssim \|\phi\|_\alpha \frac{1}{N} \left(\sum_{j=1}^{s} \frac{q_{j+1} \log q_{j+1}}{q_j^\alpha} \right) + \text{const } \|\phi\|_\alpha \frac{1}{n^\alpha}$$
$$\lesssim \left(\frac{1}{\gamma} \log \frac{1}{\gamma}\right) \|\phi\|_\alpha \left(\frac{n^{1-\alpha} \log^3 n}{N} + \frac{1}{n^\alpha} \right) .$$

The conclusion (3.14) then follows by choosing the degree of polynomial approximation $n = \left\lfloor \frac{N}{\log^3 N} \right\rfloor$. ■

3.5 The higher dimensional torus translation case

In this section we establish our result on the rate of convergence of the Birkhoff means of Hölder observables over a higher dimensional toral translation. As in dimension one, we use approximations by trigonometric polynomials.

Theorem 3.5.1 *Let $\phi \in C^\alpha(\mathbb{T}^d)$, let $\boldsymbol{\omega} \in \mathrm{DC}(\mathbb{T}^d)_{\gamma,A}$ and let $T_{\boldsymbol{\omega}} \colon \mathbb{T}^d \to \mathbb{T}^d$ be the corresponding torus translation. Then for all $N \ge 1$ we have*

$$\left\| \frac{1}{N} \phi^{(N)} - \int_{\mathbb{T}^d} \phi \right\|_\infty \le \frac{\text{const}}{\gamma} \|\phi\|_\alpha \frac{1}{N^\beta}$$

where $\beta = \frac{\alpha}{A+d}$.

*Proof.*Fix $N \ge 1$ and let $n = o(N)$ (to be chosen later).

Let $\mathbf{J}_n \colon \mathbb{T}^d \to \mathbb{R}$ be the d-dimensional Jackson kernel defined in (3.8). Recall that $\phi_n = \phi * \mathbf{J}_n$ is a trigonometric polynomial of degree $\le n$ on $\mathbb{T}^d$, that is

$$\phi_n(\boldsymbol{x}) = \sum_{|\boldsymbol{k}| \le n} \widehat{\phi_n}(\boldsymbol{k})\, e_{\boldsymbol{k}}(\boldsymbol{x}) = \int_{\mathbb{T}^d} \phi_n + \sum_{1 \le |\boldsymbol{k}| \le n} \widehat{\phi_n}(\boldsymbol{k})\, e_{\boldsymbol{k}}(\boldsymbol{x}) \,.$$

Hence,

$$\begin{aligned}\frac{1}{N}\phi_n^{(N)}(\boldsymbol{x}) - \int_{\mathbb{T}^d}\phi_n &= \frac{1}{N}\left[\phi(\boldsymbol{x}) + \ldots + \phi(\boldsymbol{x} + (N-1)\boldsymbol{\omega}) - N\int_{\mathbb{T}}\phi\right] \\ &= \frac{1}{N}\sum_{j=0}^{N-1}\sum_{\boldsymbol{k}\neq 0}\widehat{\phi_n}(\boldsymbol{k})\, e(j\boldsymbol{k\omega} + \boldsymbol{kx}) \\ &= \sum_{k\neq 0}\widehat{\phi_n}(\boldsymbol{k})\left(\frac{1}{N}\sum_{j=0}^{N-1} e(j\boldsymbol{k\omega} + \boldsymbol{kx})\right)\end{aligned}$$

Then

$$\frac{1}{N}\phi_n^{(N)}(\boldsymbol{x}) = \int_{\mathbb{T}^d}\phi_n + \sum_{1\leq|\boldsymbol{k}|\leq n}\widehat{\phi_n}(\boldsymbol{k})\,\frac{1}{N}\,e_{\boldsymbol{k}}^{(N)}(\boldsymbol{x})\,,$$

so for all $\boldsymbol{x} \in \mathbb{T}^d$,

$$\left|\frac{1}{N}\,\phi_n^{(N)}(\boldsymbol{x}) - \int_{\mathbb{T}^d}\phi_n\right| \leq \sum_{1\leq|\boldsymbol{k}|\leq n}\left|\widehat{\phi_n}(\boldsymbol{k})\right|\,\frac{1}{N}\,\left|e_{\boldsymbol{k}}^{(N)}(\boldsymbol{x})\right|\,.$$

Let us estimate the Birkhoff sums of the multiplicative characters $e_{\boldsymbol{k}}$.

$$\begin{aligned}e_{\boldsymbol{k}}^{(N)}(\boldsymbol{x}) &= \sum_{j=0}^{N-1} e_{\boldsymbol{k}}(\boldsymbol{x} + j\boldsymbol{\omega}) = \sum_{j=0}^{N-1} e_{\boldsymbol{k}}(\boldsymbol{x})\, e_{\boldsymbol{k}}(j\boldsymbol{\omega}) \\ &= e_{\boldsymbol{k}}(\boldsymbol{x})\sum_{j=0}^{N-1} e(\boldsymbol{k}\cdot j\boldsymbol{\omega}) = e_{\boldsymbol{k}}(\boldsymbol{x})\sum_{j=0}^{N-1} e(j\,\boldsymbol{k}\cdot\boldsymbol{\omega}) \\ &= e_{\boldsymbol{k}}(\boldsymbol{x})\,\frac{1 - e(N\,\boldsymbol{k}\cdot\boldsymbol{\omega})}{1 - e(\boldsymbol{k}\cdot\boldsymbol{\omega})}\,.\end{aligned}$$

Hence for all $\boldsymbol{x} \in \mathbb{T}^d$, also using the Diophantine condition (3.2),

$$\left|e_{\boldsymbol{k}}^{(N)}(\boldsymbol{x})\right| \leq \left|\frac{1 - e(N\,\boldsymbol{k}\cdot\boldsymbol{\omega})}{1 - e(\boldsymbol{k}\cdot\boldsymbol{\omega})}\right| \leq \frac{1}{\|\boldsymbol{k}\cdot\boldsymbol{\omega}\|} \leq \frac{1}{\gamma}\,|\boldsymbol{k}|^A\,.$$

Combining this estimate on $e_{\boldsymbol{k}}^{(N)}$ with (3.10), it follows that

$$\begin{aligned}\left|\frac{1}{N}\,\phi_n^{(N)}(\boldsymbol{x}) - \int_{\mathbb{T}^d}\phi_n\right| &\leq \sum_{1\leq|\boldsymbol{k}|\leq n}\left|\widehat{\phi_n}(\boldsymbol{k})\right|\,\frac{1}{N}\,\left|e_{\boldsymbol{k}}^{(N)}(\boldsymbol{x})\right| \\ &\leq \frac{\text{const}}{\gamma}\,\|\phi\|_\alpha\,\frac{1}{N}\sum_{1\leq|\boldsymbol{k}|\leq n}\frac{1}{|\boldsymbol{k}|^\alpha}\,|\boldsymbol{k}|^A \\ &\leq \frac{\text{const}}{\gamma}\,\|\phi\|_\alpha\,\frac{n^{A+d-\alpha}}{N}\,.\end{aligned}$$

Write

$$\phi = \phi_n + (\phi - \phi_n) =: \phi_n + \psi_n\,.$$

Combining the last estimate with (3.9) we have:

$$\begin{aligned}\left\| \frac{1}{N}\,\phi^{(N)} - \int_{\mathbb{T}^d} \phi \right\|_\infty &\le \left\| \left(\frac{1}{N}\,\phi_n^{(N)} - \int_{\mathbb{T}} \phi_n \right) \right\|_\infty + \left\| \left(\frac{1}{N}\,\psi_n^{(N)} - \int_{\mathbb{T}} \psi_n \right) \right\|_\infty \\ &\le \frac{\text{const}}{\gamma}\, \|\phi\|_\alpha\, \frac{n^{A+d-\alpha}}{N} + \text{const}\, \|\phi\|_\alpha\, \frac{1}{n^\alpha} \\ &\le \frac{\text{const}}{\gamma}\, \|\phi\|_\alpha\, N^{-\frac{\alpha}{A+d}}\,,\end{aligned}$$

provided we choose $n := N^{\frac{1}{A+d}}$. ■

As noted earlier, this higher dimensional result is likely not optimal. The sharpness of the one dimensional result is due in large part to the use of continued fractions, which allowed us a finer analysis of the Fouries series of the Birkhoff sums.

4
Mixed Markov-quasiperiodic base dynamics

We call mixed Markov-quasiperiodic dynamical system the product between a Markov shift and a torus translation. The main goal of this chapter is to establish a large deviations estimates for such systems with observables depending on a finite number of coordinates.

4.1
Description of the model

Let Σ be a compact metric space, $\mathcal{F}$ be its Borel σ-algebra and denote by $\text{Prob}(\Sigma)$ the space of Borel probability measures on Σ, endowed with the weak* topology.

Let us recall some concepts that were introduced in Chapter 2. A *Markov transition kernel* on Σ is a continuous map $K\colon \Sigma \to \text{Prob}(\Sigma)$. Furthermore, we say that $\mu \in \text{Prob}(\Sigma)$ is a K-stationary measure if

$$\mu(E) = \int K_x(E)\, d\mu(x)$$

for all $E \in \mathcal{F}$.

Definition 4.1.1 We say that the kernel $K\colon \Sigma \to \text{Prob}(\Sigma)$ is *uniformly ergodic* if there exist $n \in \mathbb{N}$ and $c \in (0,1)$ such that

$$\left\| K^n_{\omega_0} - \mu \right\|_{\text{TV}} \leq c$$

for every $\omega_0 \in \Sigma$, where $\|\cdot\|_{\text{TV}}$ is the total variation norm on $\text{Prob}(\Sigma)$.

Let (Σ, K, μ) be a Markov system, that is, $K\colon \Sigma \to \text{Prob}(\Sigma)$ is a Markov kernel on Σ and μ is a K-stationary measure. Let $\mathbb{P} = \mathbb{P}_\mu = \mathbb{P}_K = \mathbb{P}_{(K,\mu)}$ denote the Markov measure on $X = \Sigma^{\mathbb{Z}}$ with initial distribution μ and transition kernel K. The *two-sided shift* is the map $\sigma : X \to X$ such that

$$\sigma(\{x_n\}_{n\in\mathbb{Z}}) = \{x_{n+1}\}_{n\in\mathbb{Z}}.$$

Then $(X, \mathbb{P}, \sigma)$ is a measure preserving dynamical system, which we call a Markov shift.

Definition 4.1.2 Let $\alpha \in \mathbb{T}^d$. We call the map $f\colon \Sigma^{\mathbb{Z}} \times \mathbb{T}^d \to \Sigma^{\mathbb{Z}} \times \mathbb{T}^d$

$$f(\omega, \theta) := (\sigma\omega, \theta + \alpha)$$

a mixed Markov-quasiperiodic dynamical system.

In other words, f is simply the product between a Markov shift and a translation on the d-dimensional torus. Hence $(X \times \mathbb{T}^d, f, \mathbb{P} \times m)$ is a measure preserving dynamical system, where m is the Lebesgue measure.

Moreover, if $\alpha \in \mathbb{T}^d$ is a rationally independent frequency, then f is ergodic (see [29, Theorem 6.1]).

4.2 Large deviations estimates

In this section we establish large deviations estimates for this base dynamics with observables that depend on a finite number of coordinates. This result hold in fact for all $\theta \in \mathbb{T}^d$, and it is uniform in θ.

We begin with the following lemma, which shows that for a full measure set of points $\omega \in X$, given any continuous observable $\phi : X \times \mathbb{T}^d \to \mathbb{R}$, the corresponding Birkhoff averages converge uniformly in $\theta \in \mathbb{T}^d$ to the space average.

Lemma 4.2.1 *Let $\mathbb{P}$ be the Kolmogorov extension of (K, μ) on $X = \Sigma^{\mathbb{Z}}$. There exists a full measure set $X' \subset X$ such that given any observable $\phi \in C^0(X \times \mathbb{T}^d)$, for all $\omega \in X'$ we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \phi(f^j(\omega, \theta)) = \int_{X \times \mathbb{T}^d} \phi \, d(\mathbb{P} \times m)$$

with uniform convergence in $\theta \in T^d$.

Proof. Let $\mathcal{X} := \operatorname{supp}(\mathbb{P})$. As $C^0(\mathcal{X} \times \mathbb{T}^d)$ is a separable space, it admits a countable and dense subset $\{\phi_j : j \geq 1\}$.

Denote by

$$B_j := \left\{ (\omega, \theta) \in \mathcal{X} \times \mathbb{T}^d : \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \phi_j(f^i(\omega, \theta)) = \int_{X \times \mathbb{T}^d} \phi_j \, d(\mathbb{P} \times m) \right\}.$$

By Theorem 2.1.1, $(\mathbb{P} \times m)(\mathcal{B}) = 1$, where $\mathcal{B} = \cap_{j\geq 1} B_j$. If we denote $\mathcal{B}_\theta = \{\omega \in \mathcal{X} : (\omega, \theta) \in \mathcal{B}\}$ then for m-almost every $\theta \in \mathbb{T}^d, \mathbb{P}(\mathcal{B}_\theta) = 1$.

Fix $\theta_0 \in \mathbb{T}^d$ such that $\mathbb{P}(\mathcal{B}_{\theta_0}) = 1$. By the density of $\{\phi_j : j \geq 1\}$ in $C^0(\mathcal{X} \times \mathbb{T}^d)$ and the definition of $\mathcal{B}$, given $\omega \in \mathcal{B}_{\theta_0}$, we have that for all

$\phi \in C^0(X \times \mathbb{T}^d)$ and $\varepsilon > 0$, there exists n_0 such that for all $n \geq n_0$ and for j large enough

$$\begin{aligned}\left|\frac{1}{n}\sum_{i=0}^{n-1}\phi(f^i(\omega,\theta_0)) - \int \phi \, d(\mathbb{P}\times m)\right| \leq &\left|\frac{1}{n}\sum_{i=0}^{n-1}\phi(f^i(\omega,\theta_0)) - \sum_{i=0}^{n-1}\phi_j(f^i(\omega,\theta_0))\right| + \\ &+\left|\frac{1}{n}\sum_{i=0}^{n-1}\phi_j(f^i(\omega,\theta_0)) - \int_{X\times\mathbb{T}^d}\phi_j \, d(\mathbb{P}\times m)\right| + \\ &+\left|\int_{X\times\mathbb{T}^d}\phi_j \, d(\mathbb{P}\times m) - \int_{X\times\mathbb{T}^d}\phi \, d(\mathbb{P}\times m)\right| \\ &\leq \frac{\varepsilon}{3}+\frac{\varepsilon}{3}+\frac{\varepsilon}{3} \\ &\leq \varepsilon.\end{aligned}$$

On the other hand,

$$\frac{1}{n}\sum_{i=0}^{n-1}\phi(f^i(\omega,\theta_0)) = \int \phi \, d\left(\frac{1}{n}\sum_{i=0}^{n-1}\delta_{f^i(\omega,\theta_0)}\right)$$

hence, for $\mathbb{P}$-almost every ω, the following weak* convergence holds:

$$\frac{1}{n}\sum_{i=0}^{n-1}\delta_{f^i(\omega,\theta_0)} \to \mathbb{P}\times m\,. \tag{4.1}$$

Consider the action of $\mathbb{T}^d$ on $X \times T^d$ given by $\theta \cdot (\omega,\theta') = (\omega, \theta+\theta')$. This action induces a convolution of measure and for all $\phi \in C(X\times\mathbb{T}^d)$,

$$\begin{aligned}\int \phi \, d(\delta_{\theta-\theta_0} * \delta_{f^j(\omega,\theta_0)}) &= \int \phi(\theta' \cdot (\omega',\theta''))\, d\delta_{\theta-\theta_0}(\theta')\, d\delta_{f^j(\omega,\theta_0)}(\omega',\theta'') \\ &= \phi((\theta-\theta_0)\cdot f^j(\omega,\theta_0)) \\ &= \phi(f^j(\omega,\theta)) \\ &= \int \phi \, \delta_{f^j(\omega,\theta)}\end{aligned}$$

that is, $\delta_{\theta-\theta_0} * \delta_{f^j(\omega,\theta_0)} = \delta_{f^j(\omega,\theta)}$ for all $j \geq 1$. Simillarly, $\delta_{\theta-\theta_0} * (\mathbb{P}\times m) = \mathbb{P}\times m$. Hence, by (4.1) and the weak* continuity of the convolution operation, for $\mathbb{P}$-almost every ω and every $\theta \in \mathbb{T}^d$,

$$\frac{1}{n}\sum_{i=0}^{n-1}\delta_{f^i(\omega,\theta)} = \frac{1}{n}\sum_{i=0}^{n-1}(\delta_{\theta-\theta_0} * \delta_{f^i(\omega,\theta_0)}) \to \delta_{\theta-\theta_0} * (\mathbb{P}\times m) = \mathbb{P}\times m.$$

Consequently, for $\mathbb{P}$-almost every ω, for all $\phi \in C^0(X\times\mathbb{T}^d)$ and all $\theta \in \mathbb{T}^d$,

$$\frac{1}{n}\sum_{i=0}^{n-1}\phi(f^i(\omega,\theta)) \to \int_{X\times\mathbb{T}^d}\phi \, d(\mathbb{P}\times m)\,.$$

Now, suppose that this convergence is not uniform in $\theta \in \mathbb{T}^d$. That is, assume that there exist $\omega \in \mathcal{B}_{\theta_0}, \varepsilon > 0, n_k \to \infty$ and $\theta_k \in \mathbb{T}^d$ for all $k \geq 1$ such that

$$\left| \frac{1}{n_k} \sum_{j=0}^{n_k-1} \phi(f^j(\omega, \theta_k)) - \int_{X \times \mathbb{T}^d} \phi \, d(\mathbb{P} \times m) \right| \geq \varepsilon.$$

Since $\mathbb{T}^d$ is compact, by passing to a subsequence we may assume that $\theta_k \to \theta$. On the other hand, by the uniform continuity of ϕ on the compact set $\mathcal{X} \times \mathbb{T}^d$, we have that for k large enough

$$|\phi(\omega', \theta_k) - \phi(\omega', \theta)| < \frac{\varepsilon}{2}, \quad \forall \omega' \in \mathcal{X}.$$

Then, for k sufficiently large we have

$$\begin{aligned} \left| \frac{1}{n_k} \sum_{j=0}^{n_k-1} \phi(f^j(\omega, \theta)) - \int_{X \times \mathbb{T}^d} \phi \, d(\mathbb{P} \times m) \right| \leq \\ &\geq \left| \frac{1}{n_k} \sum_{j=0}^{n_k-1} \phi(f^j(\omega, \theta_k)) - \int_{X \times \mathbb{T}^d} \phi \, d(\mathbb{P} \times m) \right| \\ &- \left| \frac{1}{n_k} \sum_{j=0}^{n_k-1} \phi(f^j(\omega, \theta)) - \frac{1}{n_k} \sum_{j=0}^{n_k-1} \phi(f^j(\omega, \theta_k)) \right| \\ &\geq \varepsilon - \frac{\varepsilon}{2} \\ &\geq \frac{\varepsilon}{2} \end{aligned}$$

and this contradicts the pointwise convergence for θ. ■

We are now ready to state and prove the main result of this chapter, a large deviations estimates for mixed Markov-quasiperiodic systems for observables that depend on finitely many coordinates. A related result in the case of mixed random-quasiperiodic cocycle it may be found in [5].

Theorem 4.2.1 *Let K be a uniformly ergodic Markov kernel on Σ, let μ be its unique stationary measure and let $\mathbb{P}$ be the Kolmogorov extension of (K, μ) on $X = \Sigma^{\mathbb{Z}}$. Let $\varphi \colon X \times \mathbb{T}^d \to \mathbb{R}$ be a continuous observable that depends on a finite number of coordinates of $\omega \in X$. Given any $\varepsilon > 0$, there exist $\bar{n} = \bar{n}(\varepsilon, \phi) \in \mathbb{N}$ and $c = c(\varepsilon, \phi) > 0$ such that for all $\theta \in \mathbb{T}^d$ and for all $n \geq \bar{n}$, we have*

$$\mathbb{P}\left\{ \omega \in X : \left| \frac{1}{n} \sum_{j=0}^{n-1} \phi(f^j(\omega, \theta)) - \int_{X \times \mathbb{T}^d} \phi \, d(\mathbb{P} \times m) \right| \geq \varepsilon \right\} < e^{-cn}. \tag{4.2}$$

Proof. Fix $\varepsilon > 0$. Replacing φ by $-\varphi$, it is enough to prove just the upper bound in (4.2), that is, for ω outside an exponentially small set with respect to the $\mathbb{P}$ measure and for all $\theta \in \mathbb{T}^d$, we have

$$\frac{1}{n}\sum_{j=0}^{n-1}\phi(f^j(\omega,\theta)) < \int_{X\times\mathbb{T}^d}\phi\, d(\mathbb{P}\times m) + \varepsilon\,. \tag{4.3}$$

Since φ is bounded, we may assume that $\phi \geq 0$. By Lemma 4.2.1, for $\mathbb{P}$-almost every $\omega \in X$, define $n(\omega) = n(\omega, \varepsilon)$ to be the first integer such that for every $\theta \in \mathbb{T}^d$,

$$\frac{1}{n(\omega)}\sum_{j=0}^{n(\omega)-1}\phi(f^j(\omega,\theta)) < \int_{X\times\mathbb{T}^d}\phi\, d(\mathbb{P}\times m) + \varepsilon\,.$$

For each $m \in \mathbb{N}$, define

$$\begin{aligned} U_m &:= \{\omega \in X : n(\omega) \leq m\} \\ &= \bigcup_{k=1}^{m}\left\{\omega \in X : \frac{1}{k}\sum_{j=0}^{k-1}\phi(f^j(\omega,\theta)) < \int_{X\times\mathbb{T}^d}\phi\, d(\mathbb{P}\times m) + \varepsilon \quad \forall\theta \in \mathbb{T}^d\right\}. \end{aligned}$$

The set U_m is open since f and ϕ are continuous maps and $\mathbb{T}^d$ is a compact set. On the other hand, for every $\omega \in X$, there exists $m \in \mathbb{N}$ such that $\omega \in U_m$ and clearly, for all $n \geq 1, U_n \subset U_{n+1}$, hence $U_m \nearrow X$. Consequently, $\mathbb{P}(U_m) \to \mathbb{P}(X) = 1$. Thus, there exists $N = N(\varepsilon, \phi)$ such that $\mathbb{P}(X \setminus U_N) < \varepsilon$.

For all $\omega \in U_N$, we have $1 \leq n(\omega) \leq N$ and for every $\theta \in \mathbb{T}^d$

$$\sum_{j=0}^{n(\omega)-1}\phi(f^j(\omega,\theta)) < n(\omega)\int_{X\times\mathbb{T}^d}\phi\, d(\mathbb{P}\times m) + n(\omega)\varepsilon\,. \tag{4.4}$$

Fix $\omega \in X$. We define a sequence of indices $\{n_k = n_k(\omega)\}$ and points $\{\omega_k\}$ by the following rule:

$$\omega_1 = \omega, \quad \text{and} \quad n_1 = \begin{cases} n(\omega_1), & \text{if } \omega_1 \in U_N \\ 1, & \text{if } \omega_1 \notin U_N \end{cases}$$

And, for $k \geq 1$, define

$$\omega_{k+1} = \sigma^{n_k}\omega_k, \quad \text{and} \quad n_{k+1} = \begin{cases} n(\omega_{k+1}), & \text{if } \omega_{k+1} \in U_N \\ 1, & \text{if } \omega_{k+1} \notin U_N \end{cases}$$

that is, $\omega_{k+1} = \sigma^{n_1+\cdots+n_k}\omega$.

Let $\bar{n} := \bar{n}(\varepsilon, \phi) := N\max\{\frac{\|\phi\|_0}{\varepsilon}, 1\}$, so $\bar{n} \geq N \geq n_1$. Fix any $n \geq \bar{n}$. Note that $1 \leq n_k \leq N$ for all $k \geq 1$. Hence, the sequence $a_k = \sum_{j=1}^{k} n_j$ is such that $a_k \nearrow \infty$ and, consequently, there exists $p \in \mathbb{N}$ such that

$$n_1 + \cdots + n_p \leq n \leq n_1 + \cdots + n_{p+1}$$

that is, there exists m such that $n = n_1 + \cdots + n_p + m$, where $0 \leq m < n_{p+1} \leq N$.

Observe that

$$\sum_{j=0}^{n-1}\phi(f^j(\omega,\theta)) = \sum_{j=0}^{n_1-1}\phi(f^j(\omega,\theta)) + \sum_{j=n_1}^{n_1+n_2-1}\phi(f^j(\omega,\theta)) + \cdots + \sum_{j=n_1+\cdots+n_{p-1}}^{n_1+\cdots+n_p-1}\phi(f^j(\omega,\theta)) + \sum_{j=n_1+\cdots+n_p}^{n-1}\phi(f^j(\omega,\theta)).$$

Denote, respectively, by $S_1(\omega,\theta), S_2(\omega,\theta), \cdots, S_p(\omega,\theta)$ and $S_{p+1}(\omega,\theta)$ the $p+1$ sums in the previous inequality. From (4.4), we get

$$\sum_{j=0}^{n_1-1}\phi(f^j(\omega,\theta)) < n_1\int_{X\times\mathbb{T}^d}\phi\, d(\mathbb{P}\times m) + n_1\varepsilon, \quad \text{if } \omega_1=\omega\in U_N$$

but, if $\omega_1=\omega\notin U_N$ then $n_1=1$ and $S_1(\omega,\theta)=\phi(\omega,\theta)\le\|\phi\|_0$. On the other hand,

$$\left(\sum_{j=0}^{n_1-1}\phi(f^j(\omega,\theta)) + n_1\varepsilon\right)\mathbb{1}_{U_N}(\omega) \le \sum_{j=0}^{n_1-1}\phi(f^j(\omega,\theta)) + n_1\varepsilon$$

since we assume that $\phi\ge 0$. Hence,

$$S_1(\omega,\theta)\le n_1\int_{X\times\mathbb{T}^d}\phi\, d(\mathbb{P}\times m) + n_1\varepsilon + \|\phi\|_0\cdot\mathbb{1}_{X\setminus U_N}(\omega)\,.$$

For the second sum, take $j=n_1+l$ with $l=0,\ldots,n_2-1$. Thus, we can rewrite the sum as:

$$S_2(\omega,\theta) = \sum_{j=n_1}^{n_1+n_2-1}\phi(f^j(\omega,\theta)) = \sum_{l=0}^{n_2-1}\phi(f^{l+n_1}(\omega,\theta))$$

and we can use a simular argument as the first sum $S_1(\omega,\theta)$. Hence,

$$S_2(\omega,\theta)\le n_2\int_{X\times\mathbb{T}^d}\phi\, d(\mathbb{P}\times m) + n_2\varepsilon + \|\phi\|_0\cdot\mathbb{1}_{X\setminus U_N}(\sigma^{n_1}\omega)\,.$$

Indictively, for $p\ge 1$,

$$\begin{aligned} S_p(\omega,\theta) &= \sum_{j=0}^{n_p-1}\phi(f^{j+n_1+\cdots+n_{p-1}}(\omega,\theta)) \\ &\le n_p\int_{X\times\mathbb{T}^d}\phi\, d(\mathbb{P}\times m) + n_p\varepsilon + \|\phi\|_0\cdot\mathbb{1}_{X\setminus U_N}(\sigma^{n_1+\cdots+n_{p-1}}\omega) \end{aligned}$$

and since $\phi(\omega,\theta)\le\|\phi\|_0$, we have that

$$\begin{aligned} S_{p+1}(\omega,\theta) &= \sum_{j=n_1+\cdots+n_p}^{n-1}\phi(f^j(\omega,\theta)) = \sum_{j=n_1+\cdots+n_p}^{n_1+\cdots+n_p+m-1}\phi(f^j(\omega,\theta)) \le m\cdot\|\phi\|_0 \\ &\le N\cdot\|\phi\|_0\,. \end{aligned}$$

On the other hand,

$$\begin{aligned}\mathbb{1}_{X\setminus U_N}(\omega_1)+\cdots+\mathbb{1}_{X\setminus U_N}(\omega_p) &= \mathbb{1}_{X\setminus U_N}(\omega_1)+\cdots+\mathbb{1}_{X\setminus U_N}(\sigma^{n_1+\cdots+n_{p-1}}\omega)\\ &\leq \sum_{k=0}^{p-1}\mathbb{1}_{X\setminus U_N}(\sigma^{n_1+\cdots+n_k}\omega).\end{aligned}$$

Hence,

$$\begin{aligned}\sum_{j=0}^{n-1}\phi(f^j(\omega,\theta)) &= S_1(\omega,\theta)+\cdots+S_p(\omega,\theta)+S_{p+1}(\omega,\theta)\\ &\leq (n_1+\cdots+n_p)\int_{X\times\mathbb{T}^d}\phi\, d(\mathbb{P}\times m)+(n_1+\cdots+n_p)\varepsilon\\ &+\|\phi\|_0\sum_{k=0}^{p-1}\mathbb{1}_{X\setminus U_N}(\sigma^{n_1+\cdots+n_k}\omega)+N\|\phi\|_0\\ &\leq n\int_{X\times\mathbb{T}^d}\phi\, d(\mathbb{P}\times m)+n\varepsilon+\|\phi\|_0\sum_{j=0}^{n-1}\mathbb{1}_{X\setminus U_N}(\sigma^j\omega)+N\|\phi\|_0\\ &< n\int_{X\times\mathbb{T}^d}\phi\, d(\mathbb{P}\times m)+2n\varepsilon+\|\phi\|_0\sum_{j=0}^{n-1}\mathbb{1}_{X\setminus U_N}(\sigma^j\omega).\end{aligned}$$

Then, for all $\omega\in X,\theta\in\mathbb{T}^d$ and $n\geq\bar{n}$,

$$\frac{1}{n}\sum_{j=0}^{n-1}\phi(f^j(\omega,\theta))\leq\int_{X\times\mathbb{T}^d}\phi\, d(\mathbb{P}\times m)+2\varepsilon+\|\phi\|_0\frac{1}{n}\sum_{j=0}^{n-1}\mathbb{1}_{X\setminus U_N}(\sigma^j\omega). \tag{4.5}$$

It remains to estimate the Birkhoff average over the Markov shift of the indicator function $\mathbb{1}_{X\setminus U_N}$.

Note that since the observable ϕ depends on a finite number (say k_0) of coordinates, the set U_N is determined by $k:=k_0+N$ coordinates. The same holds for its complement $X\setminus U_N$, which is a closed set. Hence, $\mathbb{1}_{X\setminus U_N}$ depends on k coordinates and its n-th Birkhoff average depends on $n+k-1$ coordinates. Then, the function

$$h:\Sigma^{n+k-1}\to\mathbb{R},\quad h(x_0,\ldots,x_{n+k-2}):=\frac{1}{n}\sum_{j=0}^{n-1}\mathbb{1}_{X\setminus U_N}(\sigma^j\omega)$$

is well defined where $\omega=\{\omega_j\}_{j\in\mathbb{N}}$ with $\omega_0=x_0,\ldots,\omega_{n+k-2}=x_{n+k-2}$.

Observe that the function h satisfies

$$|h(x_0,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_{n+k-2})-h(x_0,\ldots,x_{i-1},x_i',x_{i+1},\ldots,x_{n+k-2})|\leq\frac{2k}{n}$$

that is, h has the bounded differences property.

A result in classical probabilities, McDiarmid's inequality, states that if $X_1,\ldots,X_n$ are independent and identically distributed (i.i.d.) random variables and if h satisfies the bounded differences property, then for all $\varepsilon > 0$

$$\mathbb{P}\{|h(X_1, \dots, X_n) - \mathbb{E}h(X_1, \dots, X_n)| > \varepsilon\} \leq 2e^{-c\varepsilon^2 n}$$

where $c > 0$ depends explicitely on h. This is a generalization of the large deviations principle, since the funtion $h(x_1, \dots, x_n) = \frac{1}{n}(x_1 + \cdots + x_n)$ obviously satisfies the bounded differences property.

It turns out that McDiarmid's inequality also holds for dynamical with some hyperbolicity (see [9]). In particular it holds for Markov shifts with uniformly ergodic transition kernels. Therefore, there exists a set $\mathcal{B}_n \subset \Sigma^{n+k-1}$ with $\mathbb{P}(\mathcal{B}_n) < e^{-c(\varepsilon)n}$, where $c(\varepsilon) > 0$, so that if $(\omega_0, \dots, \omega_{n+k-2}) \notin \mathcal{B}_n$ then we have

$$h(\omega_0, \dots, \omega_{n+k-2}) - \int h(\omega_0, \dots, \omega_{n+k-2})\, d\mathbb{P} < \varepsilon\,.$$

On the other hand,

$$\begin{aligned} \int h(\omega_0, \dots, \omega_{n+k-2})\, d\mathbb{P} &= \int \frac{1}{n} \sum_{j=0}^{n-1} \mathbb{1}_{X \setminus U_N}(\sigma^j \omega)\, d\mathbb{P}(\omega) \\ &= \int \mathbb{1}_{X \setminus U_N}(\omega)\, d\mathbb{P}(\omega) \\ &= \mathbb{P}(X \setminus U_N) < 3\varepsilon \end{aligned}$$

which when combined with (4.5) implies (4.3). ■

5
Linear cocycles over Markov shifts

The main goal of this chapter is to study the continuity of the maximal Lyapunov exponent of linear cocycles over Markov shifts (which we refer to as Markov cocycles).

To this end, we formally introduce the concept of Markov cocycle (Section 5.1), then define the Markov operator and the stationary measure and study their basic properties (Section 5.2). In Section 5.3 we establish the Kifer non-random filtration for Markov cocycles, a more precise version of Oseledets theorem in this context. As a corollary, under a generic assumption (quasi-irreducibility), we obtain the uniform convergence of the expected value of the finite scale directional Lyapunov exponent. This is then used in Section 5.4 to establish the strong mixing of the Markov operator and the convergence (in an appropriate sense) of its powers to the unique stationary measure. Finally, in Section 5.5 we obtain the Hölder continuity of the Lyapunov exponent via Furstenberg's Formula.

5.1
Description of the model

Let Σ be a compact metric space and let $\mathcal{F}$ be its Borel σ-algebra. Let $\mathrm{Prob}(\Sigma)$ denote the space of Borel probability measures on Σ and consider the Wasserstein distance W_1 in the space $\mathrm{Prob}(\Sigma)$. The Kantorovich-Rubinstein theorem characterizes the Wasserstein distance as follows:

$$W_1(\mu, \nu) = \sup_{\varphi \in \mathrm{Lip}_1(\Sigma)} \left(\int \varphi \, d\mu - \int \varphi \, d\nu \right).$$

Here $\mathrm{Lip}_1(\Sigma)$ is the set of Lipschitz continuous functions on Σ with Lipschitz constant ≤ 1. It is well known that this distance metrizes the weak* topology (See [34]).

Definition 5.1.1 A *Markov kernel* is a transition map $K \colon \Sigma \to \mathrm{Prob}(\Sigma), x \mapsto K_x$, such that for any Borel set $E \in \mathcal{F}$, the function $x \mapsto K_x(E)$ is $\mathcal{F}$-measurable.

The iterates of a Markov kernel K are defined recursively setting $K^1 := K$ and for $n \geq 2, E \in \mathcal{F}$,

$$K_x^n(E) := \int_X K_y^{n-1}(E)\, dK_x(y).$$

Each power K^n is itself a Markov kernel on $(\Sigma, \mathcal{F})$.

Definition 5.1.2 A probability measure μ on $(\Sigma, \mathcal{F})$ is called *K-stationary* if

$$\mu(E) = \int K_x(E)\, d\mu(x).$$

for all $E \in \mathcal{F}$.

The above definition means that μ is stationary if μ is K_x-invariant on average.

Definition 5.1.3 A *Markov system* is a pair (K, μ), where K is a Markov Kernel on $(\Sigma, \mathcal{F})$ and μ is a K-stationary probability measure.

Let (K, μ) be a Markov system. Consider the space $X^+ = \Sigma^{\mathbb{N}}$ of sequences $x = \{x_n\}_{n \in \mathbb{N}}$ with $x_n \in \Sigma$ for all $n \in \mathbb{N}$ and let $\mathcal{F}^+$ be the product σ-field $\mathcal{F}^+ = \mathcal{F}^{\mathbb{N}}$ generated by the $\mathcal{F}$-cylinders. In other words, $\mathcal{F}^+$ is generated by sets of the form

$$C(E_0, \ldots, E_m) := \{x \in X^+ : x_j \in E_j, \quad \text{for } 0 \leq j \leq m\},$$

where $E_0, \ldots, E_m \in \mathcal{F}$ are measurable sets.

Definition 5.1.4 Given any probability measure θ on $(\Sigma, \mathcal{F})$, the following expression determines a pre-measure

$$\mathbb{P}_\theta^+[C(E_0, \ldots, E_m)] := \int_{E_0} \int_{E_1} \cdots \int_{E_m} dK_{x_{m-1}}(x_m) \cdots dK_{x_1}(x_0)\, d\theta(x_0)$$

on the semi-algebra of $\mathcal{F}$-cylinders. By Carathéodory's extension theorem this pre-measures extends to a unique probability measure $\mathbb{P}_\theta^+$ on $(X^+, \mathcal{F}^+)$.

Markov systems are probabilistic evolutionary models, which can also be studied in dynamical terms. Let (K, μ) be a Markov system and $X = \Sigma^{\mathbb{Z}}$ be the set of double sided sequences. The *one-sided shift* is the map $\sigma\colon X^+ \to X^+$ such that $\sigma(\{x_n\}_{n \in \mathbb{N}}) = \{x_{n+1}\}_{n \in \mathbb{N}}$ and the *two-sided shift* is the map $\sigma\colon X \to X$ such that $\sigma(\{x_n\}_{n \in \mathbb{Z}}) = \{x_{n+1}\}_{n \in \mathbb{Z}}$. The two-sided shift is the natural extension of the one-sided shift. Then there is a unique probability measure $\mathbb{P}_\mu$ on $(X, \mathcal{F})$ that projects to the corresponding Kolmogorov measures on $(X^+, \mathcal{F}^+)$ and we will refer to the measure $\mathbb{P}_\mu$ as the *Kolmogorov extension* of the Markov system (K, μ).

Definition 5.1.5 Given a Markov system (K, μ) let $\mathbb{P}_\mu$ be the Kolmogorov extension of (K, μ) on $X = \Sigma^{\mathbb{N}}$. The dynamical system $(X, \mathbb{P}_\mu, \sigma)$ is called a *Markov system.*

Given $x \in \Sigma$, we denote by $\mathbb{P}_x = \mathbb{P}_{(K,\delta_x)}$ the Markov measure on $\Sigma^{\mathbb{N}}$ with initial distribution δ_x and transition kernel K.

Consider a Markov system (K, μ) on a compact metric space Σ.

Definition 5.1.6 The following linear operator is called a *Markov operator*

$$(Qf)(x) = (Q_K f)(x) := \int f(y)\, dK_x(y)\,.$$

It is easy to verify that the powers Q^n of the Markov operator Q satisfy

$$(Q_K^n f)(x) := \int f(y)\, dK_x^n(y)$$

for all $n \geq 1$ and $f \in L^\infty(\Sigma)$.

Definition 5.1.7 We say that the kernel $K\colon \Sigma \to \mathrm{Prob}(\Sigma)$ is *uniformly ergodic* if there exist $n \in \mathbb{N}$ and $c \in (0, 1)$ such that

$$\left\| K_{\omega_0}^n - \mu \right\|_{\mathrm{TV}} \leq c$$

for every $\omega_0 \in \Sigma$, where $\|\cdot\|_{\mathrm{TV}}$ is the total variation norm on $\mathrm{Prob}(\Sigma)$.

This in particular implies the uniqueness (the existence is guaranteed by general principles) of a K-stationary measure μ. In this case, $K_{\omega_0}^n \to \mu$ uniformly in $\omega_0 \in \Sigma$ relative to the total variation distance. Moreover, this is equivalent to the following:

$$\|Q^n\varphi - \int \varphi\, d\mu\|_\infty \leq C\sigma^n \|\varphi\|_\infty, \quad \forall \varphi \in L^\infty(\mu)$$

where Q is the Markov operator associated with K.

A measurable function $A\colon \Sigma \times \Sigma \to \mathrm{GL}_m(\mathbb{R})$ induces the skew-product dynamical system $F = F_{(A,K)}\colon X \times \mathbb{R}^m \to X \times \mathbb{R}^m$,

$$F(\omega, v) = (\sigma\omega, A(\omega_1, \omega_0)v)\,.$$

That is, F is a linear cocycle over the base dynamics $(X, \mathbb{P}_{(K,\mu)}, \sigma)$, where the fiber dynamics is induced by the map A. We refer to such a dynamical system as a *Markov cocycle.*

Its iterates are given by

$$F^n(\omega, v) = (\sigma^n \omega, A^n(\omega) v)\,,$$

where for $\omega = \{\omega_n\}_{n\in\mathbb{Z}} \in X$,

$$A^n(\omega) = A(\omega_n, \omega_{n-1}) \cdots A(\omega_2, \omega_1)\, A(\omega_1, \omega_0)\,.$$

We identify the cocycle $F_{(A,K)}$ with the pair (A, K) and denote the corresponding Lyapunov exponents by $L_1(A, K), \ldots, L_m(A, K)$. A natural question regards their dependence on the input data. In other words, what is the regularity of the map $(A, K) \mapsto L_i(A, K)$?

In order to addres this question, let us introduce an appropriate space of cocycles and its topology.

Define the projective cocycle $\hat{F}\colon \Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^m) \to \Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^m)$ by

$$\hat{F}(\omega, \hat{v}) = (\sigma\omega, \hat{A}(\omega_1, \omega_0)\hat{v}),$$

where $\hat{v} \in \mathbb{P}(\mathbb{R}^m)$ is the projective point corresponding to a vector $v \in \mathbb{R}^n \backslash \{0\}$.

Consider the set of Markov cocycles

$$\begin{aligned}\mathcal{C} := \{(A, K)\colon\ & A\colon \Sigma \times \Sigma \to \mathrm{GL}_m(\mathbb{R}) \text{ is Lipschitz continuous and}\\ & K\colon \Sigma \to \mathrm{Prob}(\Sigma) \text{ is uniformly ergodic and}\\ & \text{continuous in the weak* topology}\}.\end{aligned}$$

This set is naturally endowed with a metric as follows:

$$d((A, K), (B, L)) := \max\{d_\infty(A, B), d_{W_1}(K, L)\},$$

where if $A, B \in \mathrm{Lip}(\Sigma \times \Sigma, \mathrm{GL}_m(\mathbb{R}))$ are two Lipschitz continuous fiber maps,

$$d_\infty(A, B) := \sup_{\omega_0, \omega_1 \in \Sigma} \|A(\omega_0, \omega_1) - B(\omega_0, \omega_1)\|$$

and the distance between two Markov kernels $K, L\colon \Sigma \to \mathrm{Prob}(\Sigma)$

$$d_{W_1}(K, L) := \sup_{w_0 \in \Sigma} W_1(K_{w_0}, L_{w_0}),$$

where W_1 is the Wasserstein distance in the space of probability measures $\mathrm{Prob}(\Sigma)$.

Let $\mathrm{Gr}(\mathbb{R}^m)$ denote the set of all linear subspaces of $\mathbb{R}^m$.

Definition 5.1.8 A measurable section $V\colon \Sigma \to \mathrm{Gr}(\mathbb{R}^m)$ is called A-invariant when

$$A(\omega_{n-1}, \omega_n)V(\omega_{n-1}) = V(\omega_n), \quad \text{for } \mathbb{P}\text{-a.e. } \omega = \{\omega\}_{n\in\mathbb{Z}} \in X$$

We say that a section $V\colon \Sigma \to \mathrm{Gr}(\mathbb{R}^m)$ is *proper* if $V(\omega_0) \subsetneq \mathbb{R}^m$ for every $\omega_0 \in \Sigma$.

Definition 5.1.9 A Markov cocycle A is called *quasi-irreducible* w.r.t. (K, μ) if there exists no measurable proper A-invariant section $V\colon \Sigma \to \mathrm{Gr}(\mathbb{R}^m)$ such that $L|_{V(\omega_0)} < L_1(A, K), \forall\, \omega_0 \in \Sigma$.

We are ready to formulate the main result of this chapter.

Theorem 5.1.1 *Let $(A, K) \in (\mathcal{C}, d)$. Assume that:*

(i) A is quasi irreducible with respect to (K, μ),

(ii) $L_1(A, K) > L_2(A, K)$.

Then there exists a neighborhood V of (A, K) in $(\mathcal{C}, d)$ where the map $(A, K) \mapsto L_1(A, K)$ is Hölder continuous.

This result extends [10, Theorem 5.1], where it was established the Hölder continuity of the Lyapunov exponents with respect to the fiber map A. In the present work we also allow the transition kernel K to vary, and prove the joint Hölder continuity in (A, K) of the exponents. Moreover, the approach used in this thesis is different from the one in [10] (which first establishes uniform large deviations type estimates for the cocycle and then deduces the Hölder continuity of the exponents from an abstract continuity theorem). The advantage of the method employed here, besides being more straightforward, is that it provides a more explicit, computable, value of the Hölder exponent.

In the case when the space Σ of symbols is finite, that is, when the base dynamics is a sub-shift of finite type, there are other results available. Fixing any fiber map, the maximal Lyapunov exponent depends analytically on the transition probabilities, see [28]. This suggests that in our more general setting, the regularity with respect to the transition kernel K might be much higher. Moreover, the continuity of the Lyapunov exponent (but without a modulus of continuity) was established in [25] without any irreducibility assumption, assuming that the fiber dynamics is two dimensional and depends on only one coordinate.

All of these results, including the one in this work, are in part inspired by the seminal works of Furstenberg and Kifer [13], Le Page [24] and Peres [27].

5.2 Stationary Measures

Let Σ be a compact metric space and let (K, μ) be the Markov system. Let $A\colon \Sigma\times\Sigma \to \mathrm{GL}_m(\mathbb{R})$ be a fiber map, which together with kernel K defines the Markov cocycle $F_{(A,K)}$. We associate to this linear cocycle the Markov kernel $\bar{K}\colon \Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m) \to \mathrm{Prob}(\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m))$ given by

$$\bar{K}(\omega_1, \omega_0, \hat{p}) = K_{\omega_1} \times \delta_{(\omega_1, A(\omega_1,\omega_0)\hat{p})}. \tag{5.1}$$

The Markov operator corresponding to this transition kernel is defined by

$$(\bar{Q}\psi)(\omega_1, \omega_0, \hat{p}) = \int \psi(\omega_2, \omega_1, \hat{A}(\omega_1, \omega_0)\hat{p})\, dK_{\omega_1}(\omega_2)$$

for every $\psi \in C^0(\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m))$.

Similarly, define the Markov kernel $K_A\colon \Sigma \times \mathbb{P}(\mathbb{R}^m) \to \mathrm{Prob}(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ by

$$K_A(\omega_0, \hat{p})(\cdot) = K_{\omega_0}(\cdot) \times \int_{\hat{A}(\omega_1,\omega_0)\hat{p}} \delta\, dK_{\omega_0}(\omega_1)$$

and consider the corresponding Markov operator $Q_{(A,K)}$ defined, for every $\phi \in C^0(\Sigma \times \mathbb{P}(\mathbb{R}^m))$, by

$$(Q_{(A,K)}\phi)(\omega_0, \hat{p}) = \int \phi(\omega_1, \hat{A}(\omega_1, \omega_0)\hat{p})\, dK_{\omega_0}(\omega_1). \tag{5.2}$$

Moreover, we introduce a projection $\Pi : C^0(\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m)) \to C^0(\Sigma \times \mathbb{P}(\mathbb{R}^m))$:

$$\Pi\psi(x_0, \hat{p}) = \int \psi(\omega_1, \omega_0, \hat{p})\, dK_{\omega_0}(\omega_1).$$

The following lemma relates these two Markov operators.

Lemma 5.2.1 *With the notations above, we have* $\Pi \circ \bar{Q} = Q_{(A,K)} \circ \Pi$.

Proof. Given any $\psi \in C^0(\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m))$, a direct computation shows that

$$\begin{aligned} Q_{(A,K)} \circ \Pi\psi(\omega_0, \hat{p}) &= \int \Pi\psi(\omega_1, \hat{A}(\omega_1, \omega_0)\hat{p})\, dK_{\omega_0}(\omega_1) \\ &= \int\int \psi(\omega_2, \omega_1, \hat{A}(\omega_1, \omega_0)\hat{p})\, dK_{\omega_1}(\omega_2)\, dK_{\omega_0}(\omega_1). \end{aligned}$$

On the other hand,

$$\begin{aligned} \Pi \circ \bar{Q}\psi(\omega_0, \hat{p}) &= \int \bar{Q}\varphi(\omega_1, \omega_0, \hat{p})\, dK_{\omega_0}(\omega_1) \\ &= \int\int \psi(\omega_2, \omega_1, \hat{A}(\omega_1, \omega_0)\hat{p})\, dK_{\omega_1}(\omega_2)\, dK_{\omega_0}(\omega_1), \end{aligned}$$

proving that $\Pi \circ \bar{Q} = Q_{(A,K)} \circ \Pi$. ■

We denote by $\mathrm{Prob}_{\bar{Q}}(\Sigma\times\Sigma\times\mathbb{P}(\mathbb{R}^m))$ and $\mathrm{Prob}_{Q}(\Sigma\times\mathbb{P}(\mathbb{R}^m))$, respectively, the convex and compact (since Σ is compact) subspace of all $\bar{K}$-stationary probability measures on $\Sigma\times\Sigma\times\mathbb{P}(\mathbb{R}^m)$ and K_A-stationary probability measure on $\Sigma\times\mathbb{P}(\mathbb{R}^m)$.

The following proposition ensures the existence of a Q-stationary probability measure in $\mathrm{Prob}(\Sigma\times\mathbb{P}(\mathbb{R}^m))$, given a $\bar{Q}$-stationary probability measure in $\mathrm{Prob}(\Sigma\times\Sigma\times\mathbb{P}(\mathbb{R}^m))$.

Proposition 5.2.1 *Given $m\in\mathrm{Prob}_{\bar{Q}}(\Sigma\times\Sigma\times\mathbb{P}(\mathbb{R}^m))$, there exists a unique probability measure $\eta\in\mathrm{Prob}(\Sigma\times\mathbb{P}(\mathbb{R}^m))$ such that $m=K\ltimes\eta$, that is, for every $\psi\in C(\Sigma\times\Sigma\times\mathbb{P}(\mathbb{R}^m))$,*

$$\int\psi(\omega_1,\omega_0,\hat{v})\,dm(\omega_1,\omega_0,\hat{v})=\int\psi(\omega_1,\omega_0,\hat{v})\,dK_{\omega_0}(\omega_1)\,d\eta(\omega_0,\hat{v}).$$

Moreover, η is a Q-stationary probability measure.

Proof. Let $m\in\mathrm{Prob}_{\bar{Q}}(\Sigma\times\Sigma\times\mathbb{P}(\mathbb{R}^m))$. Note that if $\psi_1,\psi_2\in C(\Sigma\times\Sigma\times\mathbb{P}(\mathbb{R}^m))$ are such that $\Pi\psi_1=\Pi\psi_2$ then $\int\psi_1\,dm=\int\psi_2\,dm$. In fact, suppose that $\Pi\psi_1(x,\hat{q})=\Pi\psi_2(x,\hat{q})$ for every $(x,\hat{q})\in\Sigma\times\mathbb{P}(\mathbb{R}^m)$ then

$$\begin{aligned}
&\int\psi_1(x_2,x_1,\hat{q})\;dK_{x_1}(x_2)=\int\psi_2(x_2,x_1,\hat{q})\;dK_{x_1}(x_2),\quad\forall(x_1,\hat{q})\in\Sigma\times\mathbb{P}(\mathbb{R}^m)\\
\Leftrightarrow&\int\psi_1(x_2,x_1,\hat{A}(x_1,x_0)\hat{p})\;dK_{x_1}(x_2)=\int\psi_2(x_2,x_1,\hat{A}(x_1,x_0)\hat{p})\;dK_{x_1}(x_2)\\
\Leftrightarrow&\int\bar{Q}\psi_1(x_1,x_0,\hat{p})\;dm(x_1,x_0,\hat{p})=\int\bar{Q}\psi_2(x_1,x_0,\hat{p})\;dm(x_1,x_0,\hat{p})\\
\Leftrightarrow&\int\psi_1(x_1,x_0,\hat{p})\;dm(x_1,x_0,\hat{p})=\int\psi_2(x_1,x_0,\hat{p})\;dm(x_1,x_0,\hat{p}),\forall x_0,x_1,\hat{p}.
\end{aligned}$$

By Riesz-Markov-Katutani's theorem, there exists a unique probability measure η in $\mathrm{Prob}(\Sigma\times\mathbb{P}(\mathbb{R}^m))$ such that for every $\psi\in C(\Sigma\times\Sigma\times\mathbb{P}(\mathbb{R}^m))$,

$$\int\psi\;dm=\int\phi\;d\eta,\qquad\text{where }\phi=\Pi\circ\psi.$$

In other words,

$$\int\psi(\omega_1,\omega_0,\hat{v})\,dm=\int\psi(\omega_1,\omega_0,\hat{v})\,dK_{\omega_0}(\omega_1)d\eta(\omega_0,\hat{v}).$$

Moreover, let $\psi\in C(\Sigma\times\Sigma\times\mathbb{P}(\mathbb{R}^m))$ and $\phi\in C(\Sigma\times\mathbb{P}(\mathbb{R}^m))$ such that $\phi=\Pi\circ\psi$. By the definition of η and since m is a $\bar{Q}$-stationary probability measure on $\Sigma\times\Sigma\times\mathbb{P}(\mathbb{R}^m)$, it follows that

$$\langle\phi,\eta\rangle=\langle\Pi\circ\psi,\eta\rangle=\langle\psi,m\rangle=\langle\psi,\bar{Q}^*m\rangle=\langle\bar{Q}\psi,m\rangle.$$

On the other hand,

$$\langle\phi,Q^*\eta\rangle=\langle Q\phi,\eta\rangle=\langle Q(\Pi\circ\psi),\eta\rangle=\langle\Pi\circ\bar{Q}\psi,\eta\rangle=\langle\bar{Q}\psi,m\rangle.$$

Then,

$$\langle \phi, \eta \rangle = \langle \phi, Q^* \eta \rangle$$

and consequently, η is a Q-stationary probability measure on $\Sigma \times \mathbb{P}(\mathbb{R}^m)$. ■

The converse of the previous proposition is also true, so there is a one-to-one correspondence between $\bar{K}$-stationary measures and K_A-stationary measures. Thus certain properties of the Markov operator Q can easily be transferred to $\bar{Q}$.

Consider the projection $\pi_{01} \colon \Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^m) \to \Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m)$,

$$\pi_{01}(\omega, \hat{v}) = (\omega_1, \omega_0, \hat{v})$$

for $\omega = \{\omega_n\}_{n \in \mathbb{N}} \in \Sigma^{\mathbb{N}}$.

Proposition 5.2.2 *Given $\eta \in \mathrm{Prob}_Q(\Sigma \times \mathbb{P}(\mathbb{R}^m))$, there exists an $\hat{F}$-invariant probability measure $\tilde{\eta}$ in $\mathrm{Prob}(\Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^m))$ such that $(\pi_{01})_* \tilde{\eta} = \eta$.*

Proof. Let $\psi \in C(\Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^m))$ and define $\tilde{\eta} \in \mathrm{Prob}(\Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^m))$ such that

$$\int \psi(\omega, \hat{v})\, d\tilde{\eta}(\omega, \hat{v}) := \int \psi(\omega, \hat{v})\, d\mathbb{P}_{\omega_0}(\omega)\, d\eta(\omega_0, \hat{v})\,. \tag{5.3}$$

Observe that $(\pi_{01})_* \tilde{\eta} = \eta$. In fact, let $A_0 \times \hat{B} \subset \Sigma \times \mathbb{P}(\mathbb{R}^m)$ be a cilinder. Then,

$$\begin{aligned} \tilde{\eta}(\pi_{01}^{-1}(A_0 \times \hat{B})) &= \int \mathbb{1}_{\pi_{01}^{-1}(A_0 \times \hat{B})}(\omega, \hat{v})\, d\tilde{\eta}(\omega, \hat{v}) \\ &= \int \mathbb{1}_{\Sigma \times A_0 \times \Sigma \cdots \times \hat{B}}(\omega, \hat{v})\, d\mathbb{P}_{\omega_0}(\omega)\, d\eta(\omega_0, \hat{v}) \\ &= \int \mathbb{1}_{A_0 \times \hat{B}}(\omega, \hat{v})\, d\eta(\omega_0, \hat{v}) \\ &= \eta(A_0 \times \hat{B}). \end{aligned}$$

Moreover, $\tilde{\eta}$ is an $\hat{F}$-invariant probability measure on $\Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^m)$. In fact, since $\eta \in \mathrm{Prob}_Q(\Sigma \times \mathbb{P}(\mathbb{R}^m))$, for every $\phi \in C(\Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^m))$,

$$\begin{aligned} \int_{\Sigma \times \mathbb{P}(\mathbb{R}^m)} \phi(\omega_0, \hat{v})\, d\eta(\omega_0, \hat{v}) &= \int_{\Sigma \times \mathbb{P}(\mathbb{R}^m)} (Q\phi)(\omega_0, \hat{v})\, d\eta(\omega_0, \hat{v}) \\ &= \int_{\Sigma \times \mathbb{P}(\mathbb{R}^m)} \int_{\Sigma} \phi(\omega_1, \hat{A}(\omega_1, \omega_0)\hat{v})\, dK_{\omega_0}(\omega_1)\, d\eta(\omega_0, \hat{v}). \end{aligned}$$

Hence, given $\psi \in C(\Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^m))$

$$\begin{aligned} \int \psi(\omega, \hat{v})\, d\tilde{\eta}(\omega, \hat{v}) &= \int \psi(\omega, \hat{v})\, d\mathbb{P}_{\omega_0}(\omega)\, d\eta(\omega_0, \hat{v}) \\ &= \int Q\left(\int \psi(\omega, \hat{v})\, d\mathbb{P}_{\omega_0}(\omega) \right)\, d\eta(\omega_0, \hat{v}) \\ &= \int (\psi \circ \hat{F})(\omega, \hat{v})\, d\tilde{\eta}(\omega, \hat{v}) \end{aligned}$$

proving that $\tilde{\eta}$ is an $\hat{F}$-invariant probability measure on $\Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^m)$. ■

Definition 5.2.1 An observable $\phi \in L^\infty(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ is called *η-stationary* if $Q\phi(x, \hat{v}) = \phi(x, \hat{v})$ for η-almost every $(x, \hat{v})$.

Definition 5.2.2 A Borel set $F \subset \Sigma \times \mathbb{P}(\mathbb{R}^m)$ is *η-stationary* if the indicator function $\mathbb{1}_F$ is η-stationary. That is, F is an η-stationary set if and only if for η-almost every $(\omega_0, \hat{v})$ we have $(\omega_1, \hat{A}(\omega_1, \omega_0)\hat{v}) \in F$, for K_{ω_0}-almost every $\omega_1 \in \Sigma$.

The following proposition proves that the probability η is an extremal point of $\mathrm{Prob}_Q(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ if and only if $\tilde{\eta}$ (defined by formula (5.3) is an $\hat{F}$-ergodic probability measure.

Proposition 5.2.3 *Given* $\eta \in \mathrm{Prob}_Q(\Sigma \times \mathbb{P}(\mathbb{R}^m))$, *the following are equivalent*

(i) η *is an extremal point of* $\mathrm{Prob}_Q(\Sigma \times \mathbb{P}(\mathbb{R}^m))$

(ii) If $F \subset \Sigma \times \mathbb{P}(\mathbb{R}^m)$ *is an η-stationary set then* $\eta(F) = 0$ *or* $\eta(F) = 1$

(iii) If $\phi \in L^\infty(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ *is an η-stationary function then ϕ is a constant function η-almost everywhere*

(iv) The system $(\Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^m), \hat{F}, \tilde{\eta})$ *is ergodic.*

Proof. We first prove that (i) implies (ii).

Let η be an extremal point of $\mathrm{Prob}_Q(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ and assume, by contradiction, that there exists an η-stationary subset $F \subset \Sigma \times \mathbb{P}(\mathbb{R}^m)$ such that $t := \eta(F) \in (0,1)$. Then, the same holds for $F^{\mathrm{C}} := (\Sigma \times \mathbb{P}(\mathbb{R}^m)) \setminus F$, that is, F^{C} is an η-stationary subset and $\eta(F^{\mathrm{C}}) = 1 - t \in (0,1)$.

Let η_F and $\eta_{F^{\mathrm{C}}}$ be probability measures on $\Sigma \times \mathbb{P}(\mathbb{R}^m))$ such that

$$\eta_F(E) = \frac{\eta(E \cap F)}{\eta(F)} \quad \text{and} \quad \eta_{F^{\mathrm{C}}}(E) = \frac{\eta(E \cap F^{\mathrm{C}})}{\eta(F^{\mathrm{C}})}.$$

Observe that $\eta_F \neq \eta_{F^{\mathrm{C}}}$ since $\eta_F(F) = 1$ and $\eta_{F^{\mathrm{C}}}(F) = 0$. Moreover, $\eta = t\eta_F + (1-t)\eta_{F^{\mathrm{C}}}$. So if we show that η_F and $\eta_{F^{\mathrm{C}}}$ are Q-stationary probability measures, we get a contradiction with η being an extremal point of $\mathrm{Prob}_Q(\Sigma \times \mathbb{P}(\mathbb{R}^m))$.

Since the indicator function $\mathbb{1}_F$ is an η-stationary function and η is a Q-stationary probability measure, we have that for every $\phi \in C(\Sigma \times \mathbb{P}(\mathbb{R}^m))$,

$$\begin{aligned}
&\int_{\Sigma\times\mathbb{P}(\mathbb{R}^m)} Q\phi \, d\eta_F = \frac{1}{\eta(F)}\int_F Q\phi \, d\eta \\
&= \frac{1}{\eta(F)}\int_F\int_\Sigma \phi(\omega_1, \hat{A}(\omega_1,\omega_0)\hat{v}) \, dK_{\omega_0}(\omega_1) \, d\eta(\omega_0,\hat{v}) \\
&= \frac{1}{\eta(F)}\int_{\Sigma\times\mathbb{P}(\mathbb{R}^m)}\int_\Sigma \phi(\omega_1, \hat{A}(\omega_1,\omega_0)\hat{v})\mathbb{1}_F(\omega_0,\hat{v}) \, dK_{\omega_0}(\omega_1) \, d\eta(\omega_0,\hat{v}) \\
&= \frac{1}{\eta(F)}\int_{\Sigma\times\mathbb{P}(\mathbb{R}^m)}\int_\Sigma \phi(\omega_1, \hat{A}(\omega_1,\omega_0)\hat{v})\mathbb{1}_F(\omega_1,\hat{A}(\omega_0,\omega_1)\hat{v}) \, dK_{\omega_0}(\omega_1) \, d\eta(\omega_0,\hat{v}) \\
&= \frac{1}{\eta(F)}\int_{\Sigma\times\mathbb{P}(\mathbb{R}^m)}\int_\Sigma (\phi|_F)(\omega_1, \hat{A}(\omega_1,\omega_0)\hat{v}) \, dK_{\omega_0}(\omega_1) \, d\eta(\omega_0,\hat{v}) \\
&= \frac{1}{\eta(F)}\int_{\Sigma\times\mathbb{P}(\mathbb{R}^m)} Q(\phi|_F)(\omega_0,\hat{v}) \, d\eta(\omega_0,\hat{v}) \\
&= \frac{1}{\eta(F)}\int_{\Sigma\times\mathbb{P}(\mathbb{R}^m)} (\phi|_F)(\omega_0,\hat{v}) \, d\eta(\omega_0,\hat{v}) \\
&= \frac{1}{\eta(F)}\int_F \phi(\omega_0,\hat{v}) \, d\eta(\omega_0,\hat{v}) \\
&= \int_{\Sigma\times\mathbb{P}(\mathbb{R}^m)} \phi \, d\eta_F
\end{aligned}$$

and then η_F is an Q-stationary probability measure. Analogously, η_{F^C} is also Q-stationary and this contradicts the hypothesis that η is an extremal point of $\text{Prob}_Q(\Sigma \times \mathbb{P}(\mathbb{R}^m))$.

Now, we will prove that (ii) implies (iii).

Consider the linear subspace

$$\mathcal{V} := \{\phi \in L^\infty(\Sigma \times \mathbb{P}(\mathbb{R}^m)) : \phi \text{ is an } \eta\text{-stationary function}\}.$$

Note that $\mathcal{V}$ is a lattice. In fact, take $\phi \in \mathcal{V}$. Since η is Q-stationary,

$$\int (Q|\phi| - |\phi|) \, d\eta = \int Q|\phi| \, d\eta - \int |\phi| \, d\eta = 0. \tag{5.4}$$

On the other hand, since $\phi \in \mathcal{V}$,

$$\begin{aligned}
|\phi(\omega_0,\hat{v})| &= |Q\phi(\omega_0,\hat{v})| \\
&= \left|\int_\Sigma \phi(\omega_1, \hat{A}(\omega_1,\omega_0)\hat{v}) \, dK_{\omega_0}(\omega_1)\right| \\
&\leq \int_\Sigma |\phi(\omega_1, \hat{A}(\omega_1,\omega_0)\hat{v})| \, dK_{\omega_0}(\omega_1) \\
&= Q|\phi|(\omega_0,\hat{v})
\end{aligned}$$

for η-almost every $(\omega_0,\hat{v})$. That is,

$$Q|\phi| - |\phi| \geq 0, \quad \eta\text{-almost everywhere.} \tag{5.5}$$

Combining (5.4) and (5.5), we conclude that $Q|\phi| = |\phi|$ for η-almost surely. Then $|\phi| \in \mathcal{V}$ and this proves that $\mathcal{V}$ is a lattice. Moreover, if $\phi, \psi \in \mathcal{V}$ then $\max\{\phi, \psi\} \in \mathcal{V}$.

Fix $c \in \mathbb{R}$ and consider the sub-level set

$$E = \{(x, \hat{v}) : \phi(x, \hat{v}) < c\}.$$

To prove that (ii) implies (iii) it is enough to show that E is an η-stationary set. Then, by hypothesis, $\eta(E) = 0$ or $\eta(E) = 1$ for all $c \in \mathbb{R}$, proving that ϕ is constant η-almost everywhere.

Let $\phi_n(x, \hat{v}) = \min\{1, n \cdot \max\{c - \phi(x, \hat{v}), 0\}\}$. Since $\phi \in \mathcal{V}$ and $1 \in \mathcal{V}$, we have that $\phi_n \in \mathcal{V}$. Clearly, $\phi_n \to \mathbb{1}_E$ as $n \to \infty$ then $Q\phi_n \to Q\mathbb{1}_E$ as $n \to \infty$. On the other hand, $\phi_n = Q\phi_n$ for η-almost everywhere and since $\phi_n \to \mathbb{1}_E$, it follows that $Q\mathbb{1}_E = \mathbb{1}_E, \eta$-almost surely and then the indicator function $\mathbb{1}_E$ is η-stationary.

Suppose now that (iii) is true. We will prove that $\tilde{\eta}$ is an $\hat{F}$-ergodic probability measure. Let $\psi \in L^\infty(\Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^m))$ such that $\psi \circ \hat{F} = \psi, \tilde{\eta}$-almost everywhere. It is enough to prove that ψ is a constant function $\tilde{\eta}$-almost surely.

Consider $\phi : \Sigma \times \mathbb{P}(\mathbb{R}^m) \to \mathbb{R}$ such that

$$\phi(\omega_0, \hat{v}) = \int \psi(\omega, \hat{v})\, d\mathbb{P}(\omega).$$

We claim that ϕ is constant η-almost everywhere. For this, we show that ϕ is an η-stationary function. Since $\psi \circ \hat{F} = \psi, \tilde{\eta}$-almost surely,

$$\begin{aligned}
Q\phi(\omega_0, \hat{v}) &= \int_\Sigma \phi(\omega_1, \hat{A}(\omega_1, \omega_0)\hat{v})\, dK_{\omega_0}(\omega_1) \\
&= \int_\Sigma \int_{\Sigma^{\mathbb{N}}} \psi(\ldots, \omega_2, \omega_1, \hat{A}(\omega_1, \omega_0)\hat{v})\, d\mathbb{P}(\omega)\, dK_{\omega_0}(\omega_1) \\
&= \int_{\Sigma^{\mathbb{N}}} \psi(\sigma\omega, \hat{A}(\omega_1, \omega_0)\hat{v})\, d\mathbb{P}(\omega) \\
&= \int_{\Sigma^{\mathbb{N}}} \psi \circ \hat{F}(\omega, \hat{v})\, d\mathbb{P}(\omega) \\
&= \int_{\Sigma^{\mathbb{N}}} \psi(\omega, \hat{v})\, d\mathbb{P}(\omega) = \phi(\omega_0, \hat{v}), \quad \eta\text{-almost every } (\omega_0, \hat{p}).
\end{aligned}$$

Then, there exists $c \in \mathbb{R}$ such that $\phi(\omega_0, \hat{v}) = c, \eta$-almost every $(\omega_0, \hat{v})$. Left to show that ψ does not depend on $\omega = (\omega_0, \omega_1, \ldots, \omega_k, \ldots)$. Fix $k \geq 1$, it is enough to show that ψ is constant in $(\omega_0, \omega_1, \ldots, \omega_{k-1})$.

Since ψ is $\hat{F}$-invariant, we have that $\psi = \psi \circ \hat{F}^k$ for $\tilde{\eta}$-almost everywhere. Hence,

$$\begin{aligned}
\int \psi(\{\omega_n\}_{n\in\mathbb{N}}, \hat{v})\, dK^\infty_{\omega_k}(\omega) &= \int \psi(\sigma^k(\omega), \hat{A}^k(\omega)\hat{v})\, d\mathbb{P}(\omega) \\
&= \phi(\omega_k, \hat{A}^k(\omega)\hat{v}) = c, \quad \eta\text{-almost surely.}
\end{aligned}$$

Then, ψ is constant in $(\omega_0, \ldots, \omega_{k-1}, \ldots)$ and then is constant in $(\omega, \hat{v})$, $\tilde{\eta}$-almost every $(\omega, \hat{v})$.

It remains to proof that (iv) implies (i). Assume by contradiction that η is not an extremal point of $\mathrm{Prob}_Q(\Sigma \times \mathbb{P}(\mathbb{R}^m))$. Then, there exist $t \in (0, 1)$ and $\eta_1, \eta_2 \in \mathrm{Prob}_Q(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ such that

$$\eta = t\eta_1 + (1 - t)\eta_2.$$

Then, $\tilde{\eta} = t\tilde{\eta}_1 + (1 - t)\tilde{\eta}_2$. In other words, $\tilde{\eta}$ is not an extremal point and, consequently, $\tilde{\eta}$ is not an ergodic probability measure. ■

5.3 Kifer non-random filtration

For measure preserving dynamical systems, the Oseledets multiplicative ergodic theorem improves the Furstenberg-Kesten theorem in that it provides exponential rates of convergence of the iterates $A^{(n)}(\omega)$ of the cocycle A along all directions. Kifer [20] improved the Oseledets theorem in the setting of random cocycles (i.e. linear cocycles over a Bernoulli shift) by proving the existence of an invariant filtration that does not depend on the base point, thus it is non-random. The main goal of this section is to derive a version of this result in the context of Markov cocycles. For this, we follow the argument presented in [4]. For another arguments providing a more complete version of this result, see [8]. Moreover, assuming the quasi-irreducibility of the cocycle we derive a Furstenberg-type formula and eventually the uniform convergence of the expected value of the finite scale Lyapunov exponent.

Given (K, μ) a Markov system, consider the continuous observable $\psi \colon \Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m) \to \mathbb{R}$ defined by

$$\psi(y, x, \hat{v}) = \log \|A(y, x)v\| \tag{5.6}$$

where v is any unit vector representing the projective point $\hat{v}$. The observable ψ extends naturally to a function $\tilde{\psi} \colon \Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^m) \to \mathbb{R}$ such that $\tilde{\psi} = \psi \circ \pi_{01}$.

Consider the continuous linear functional $\alpha \colon \mathrm{Prob}_Q(\Sigma \times \mathbb{P}(\mathbb{R}^m)) \to \mathbb{R}$ defined by

$$\alpha(\eta) := \int_{\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m)} \psi(y, x, \hat{v})\, dK_x(y) d\eta(x, \hat{v})$$

and define

$$\beta := \max\{\alpha(\eta) : \eta \in \mathrm{Prob}_Q(\Sigma \times \mathbb{P}(\mathbb{R}^m))\}.$$

Theorem 5.3.1 *Let* (Σ, K, μ) *be a Markov system and let* $A \colon \Sigma \times \Sigma \to \mathrm{GL}_m(\mathbb{R})$ *be a fiber map. Then*

(i) *Given any* $(\omega_0, v) \in \Sigma \times \mathbb{P}(\mathbb{R}^m)$,

$$\limsup \frac{1}{n} \log \|A^n(\omega)v\| \leq \beta$$

for $\mathbb{P}_{\omega_0}$*-almost every* $\omega \in \Sigma^{\mathbb{N}}$.

(ii) *If the functional* α *is constant, then*

$$\lim \frac{1}{n} \log \|A^n(\omega)v\| = \beta$$

for $\mathbb{P}_{\mu}$*-almost every* $\omega \in \Sigma^{\mathbb{N}}$.

(iii) *For* $\mathbb{P}_{\omega_0}$*-almost every* $\omega \in \Sigma^{\mathbb{N}}$,

$$\lim \frac{1}{n} \log \|A^n(\omega)\| = \beta\,.$$

In particular,

$$L_1(A, K) = \beta = \int_{\Sigma\times\Sigma\times\mathbb{P}(\mathbb{R}^m)} \psi(y, x, \hat{v})\; dK_x(y) d\eta(x, \hat{v})$$

which is a version of Furstenberg's formula in this setting.

Proof. Let $M = \Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m)$ and let $\bar{K}\colon M \to \text{Prob}(M)$ be the kernel defined in (5.1) and let $\psi \in C^0(M)$ be the continuous observable defined in (5.6).

For each $v \in \mathbb{R}^d \setminus \{0\}$ consider a Markov chain $Z_n^{\hat{v}}\colon \Sigma^{\mathbb{N}} \to M$ with transition kernel $\bar{K}$, defined by

$$Z_n^{\hat{v}}(\omega) := (\omega_{n+1}, \omega_n, \hat{A}^n(\omega)\hat{v}).$$

Items (i) and (ii) follow applying Theorem 2.6.4 and Theorem 2.6.5 respectively.

Let us now prove (iii). By Furstenberg-Kesten's Theorem, the following limit exists for $\mathbb{P}$-almost every $\omega \in \Sigma^{\mathbb{N}}$

$$\lim \frac{1}{n} \log \|A^n(\omega)\| = L_1(\mu).$$

Fixing a basis $\{e_1, \ldots, e_m\}$ of $\mathbb{R}^m$, define the matrix norm

$$\|g\|' = \max_{1\leq i\leq d} \|ge_i\|.$$

The set of maximizing measures

$$\mathcal{M} := \{\eta \in \mathrm{Prob}_Q(\Sigma \times \mathbb{P}(\mathbb{R}^m)) : \alpha(\eta) = \beta\}$$

is a non-empty compact convex set. By Krein-Milman Theorem there exists an extremal point η of $\mathcal{M}$ and then this measure is also an extremal point of $\mathrm{Prob}_Q(\Sigma \times \mathbb{P}(\mathbb{R}^m))$. By proposition 5.2.3, $\tilde{\eta}$ is an $\hat{F}_A$-ergodic probability measure. Thus, by Birkhoff Ergodic Theorem (Theorem 2.1.1), for $\tilde{\eta}$-almost every $(\omega, \hat{v}) \in \Sigma^{\mathbb{N}} \times \mathbb{P}(\mathbb{R}^m)$,

$$\begin{aligned}
\beta = \alpha(\eta) &= \int_{\Sigma\times\Sigma\times\mathbb{P}(\mathbb{R}^m)} \psi(x_1, x_0, \hat{v})\, dK_{x_0}(x_1) d\eta(x, \hat{v}) \\
&= \int_{\Sigma^{\mathbb{N}}\times\mathbb{P}(\mathbb{R}^m)} \tilde{\psi}(x, \hat{v})\, d\tilde{\eta}(x, \hat{v}) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \tilde{\psi}(\hat{F}^j(\omega, \hat{v})) \\
&= \lim_{n\to\infty} \frac{1}{n} \log \|A^n(\omega) v\| \\
&\leq \lim_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)\| = \lim_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)\|' \\
&= \max_{1\leq i\leq d} \lim_{n\to\infty} \frac{1}{n} \log \|A^n(\omega) e_i\| \leq \beta\,.
\end{aligned}$$

This proves (iii). ■

We are ready to state and prove a version of Kifer's non random filtration theorem for Markov cocycles.

Theorem 5.3.2 *Given a Markov cocycle (A, K), there are numbers $\beta_0, \beta_1 \in \mathbb{R}$ and an A-invariant section $\mathcal{L}_1 \colon \Sigma \to \mathrm{Gr}(\mathbb{R}^m)$ such that $\mathcal{L}_1(\omega_0) \subsetneq \mathbb{R}^m$ and for every $v \in \mathbb{R}^m \setminus \mathcal{L}_1(\omega_0)$*

$$\lim \frac{1}{n} \log \|A^n(\omega) v\| = \beta_0$$

while if $v \in \mathcal{L}_1(\omega_0)$

$$\lim \frac{1}{n} \log \|A^n(\omega) v\| \leq \beta_1\,.$$

Moreover, the numbers β_0 and β_1 are exactly the first and the second Lyapunov exponents.

Proof. Consider the Markov chain $\{(\omega_n, A^n(\omega)v)\}_{n\in\mathbb{N}}$ on $\Sigma^{\mathbb{N}} \times \mathbb{R}^m$. Recall that

$$A^n(\omega) = A(\omega_n, \omega_{n-1}) \circ \cdots A(\omega_2, \omega_1) \circ A(\omega_1, \omega_0).$$

Let $\delta_{(\omega_0, v)}$ be the initial distribution and denote by R the kernel of this Markov chain.

By Theorem 2.6.1, there exists a measure m on

$$\mathcal{F} = \{g\colon \Sigma \times \mathrm{GL}_d(R) \to \Sigma \times \mathrm{GL}_d(R) : g \text{ is a Borel map}\}$$

such that

$$R_{(x,M)}(E) = m\{g \in \mathcal{F} : g(x,M) \in E\}\,.$$

Let $g_1, g_2, \ldots, g_n, \ldots \in \mathcal{F}$ be independent and identically distributed maps relative to m. For every $\omega \in \Sigma$ and $n \geq 1$ consider $f_n\colon \Sigma \to \Sigma$ and $J_n\colon \Sigma \to \mathrm{GL}_d(\mathbb{R})$ such that $g_n(\omega_0, \mathrm{Id}) := (f_n(\omega_0), J_n(\omega_0))$. Define $f^0(\omega_0) = \omega_0$, $J^0(\omega_0) = \mathrm{Id}$ and for every $n \geq 1$,

$$f^n = f_n \circ \cdots \circ f_1,$$

$$J^n(\omega) = J_n(f^{n-1}(\omega_0))J_{n-1}(f^{n-2}(\omega_0)) \cdots J_2(f^1(\omega_0))J_1(\omega_0)\,.$$

The sequence $\{(f^n(\omega_0), J^n(\omega_0)), n \in \mathbb{N}\}$ is a version of the Markov chain $(\omega_n, A^n(\omega))$. In fact, let $\mathcal{F}(n)$ be the σ-algebra generated by $\{(f^p(\omega_0), J^p(\omega_0)), p \leq n\}$. Take $y = f^n(\omega_0)$ and $M = J^n(\omega_0)$. If A is a borel subset of Σ and B is a borelian subset of $\mathcal{C}$, we have

$$\begin{aligned}
\mathbb{E}((f^{n+1}(x), J^{n+1}(x)) \in A \times B|\mathcal{F}(n)) &= \mathbb{P}(f_{n+1}(y) \in A, J_{n+1}(y)M \in B) \\
&= \mathbb{P}(g_{n+1}(y, \mathrm{Id}) \in A \times BM^{-1}) \\
&= m\{g : g(y, \mathrm{Id}) \in A \times BM^{-1}\} \\
&= R_{(y,\mathrm{Id})}(A \times BM^{-1}) \\
&= R_{(y,M)}(A \times B)\,.
\end{aligned}$$

Define $F_n\colon \Sigma \times \mathbb{R}^d \to \Sigma \times \mathbb{R}^d$ such that

$$F_n(x, v) = (f_n(x), J_n(x)v).$$

Note that F_n are independent and identically distributed maps and given any initial $(\omega_0, v) \in \Sigma \times \mathbb{R}^d$, the Markov chain $F_n \circ \cdots \circ F_1(\omega_0, v)$ in $\Sigma \times \mathbb{R}^d$ is a version of the Markov chain $(\omega_n, A^n(\omega)v)$ with the same initial distribution $\delta_{(\omega_0,v)}$.

To conclude the proof apply Theorem 2.6.3 to the Markov chain F_n. ■

An immediate consequence of the quasi-irreducibility condition (Definition 5.1.9) is the following:

Corollary 5.3.3 *Assume that A is quasi-irreducible. Then the non-random filtration is trivial, that is, $\mathcal{L}_1(\omega_0) = \{0\}$ for every $\omega_0 \in \Sigma$. In particular,*

$\alpha(\eta) \equiv \beta = L_1(A, K)$ *for all* $\eta \in \mathrm{Prob}_Q(\Sigma \times \mathbb{P}(\mathbb{R}^m))$, *so the functional* α *is constant.*

Proof. Assume by contradiction that the non-random filtration is not trivial. Then $\mathcal{L}_1(\omega_0) \neq \{0\}$ for all $\omega_0 \in \Sigma$. Moreover, $\mathcal{L}_1$ is A-invariant and if $v \in \mathcal{L}_1(\omega_0)$

$$\lim \frac{1}{n} \log \|A^n(\omega)v\| \leq \beta_1 < \beta_0$$

for $\mathbb{P}_{\omega_0}$-almost every $\omega \in \Sigma^{\mathbb{N}}$. Contradicting the quasi-irreducibility of A.

This proves the non-random filtration is trivial. By item (iii) of Theorem 2.6.3, we have

$$\alpha(\eta) \equiv \beta_0 = \beta = L_1(A) = \lim_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)\|, \mathbb{P}_{\omega_0}\text{-a.e.}\, \omega \in \Sigma^{\mathbb{N}}$$

for all η that is an extremal point of $\mathrm{Prob}_Q(\Sigma \times \mathbb{P}(\mathbb{R}^m))$. By Krein-Milman and the linearity of $\alpha(\cdot)$, the result follows. ■

The following theorem ensures the uniform convergence of the expected value in ω_0.

Theorem 5.3.4 *Let* A *be a Markov cocycle over a Markov system* (K, μ) *such that* A *and* A^{-1} *are both measurable. If* A *is quasi-irreducible and* $L_1(A, K) > L_2(A, K)$ *then*

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E}_{\omega_0}(\log \|A^{(n)}(\omega)v\|) = L_1(A, K),$$

with uniform convergence in $(\omega_0, \hat{v}) \in \Sigma \times \mathbb{S}^{m-1}$.

Proof. Since A is quasi-irreducible, by Corollary 5.3.3 and the Lebesgue dominated convergence theorem, we have the pointwise convergence:

$$\lim_{n\to\infty} \mathbb{E}_{\omega_0}\left(\frac{1}{n} \log \|A^{(n)}(\omega)v\|\right) = L_1(A) \tag{5.7}$$

for every $(\omega_0, \hat{v}) \in \Sigma \times \mathbb{S}^{m-1}$.

Assume first by contradiction that fixing $\omega_0 \in \Sigma$, the convergence in $\hat{v}$ is not uniform. Then there exist $\delta > 0$ and a sequence $\{v_n\}_{n\in\mathbb{N}} \subset \mathbb{S}^{m-1}$ such that

$$\left|\mathbb{E}_{\omega_0}\left(\frac{1}{n} \log \|A^{(n)}(\omega)v_n\|\right) - L_1(A, K)\right| \geq \delta, \quad \forall k \geq 1.$$

Since

$$\mathbb{E}_{\omega_0}\left(\frac{1}{n} \log \|A^{(n)}(\omega)v_n\|\right) \leq$$
$$\leq \mathbb{E}_{\omega_0}\left(\frac{1}{n} \log \|A^{(n)}(\omega)\|\right) \to L_1(A, K) < L_1(A, K) + \frac{\delta}{2},$$

we have that for $n \leq N$ with N large enough, it can not happen that

$$\mathbb{E}_{\omega_0}\left(\frac{1}{n}\log\|A^{(n)}(\omega)v_n\|\right) \geq L_1(A,K)+\delta.$$

Thus we only need to consider the case with

$$\mathbb{E}_{\omega_0}\left(\frac{1}{n}\log\|A^{(n)}(\omega)v_n\|\right) \leq L_1(A,K)-\delta\,.$$

We are going to prove that this cannot happen either. First, we claim that

$$\liminf_{n\to\infty}\frac{\|A^n(\omega)v_n\|}{\|A^n(\omega)\|} = c(\omega) > 0$$

for $\mathbb{P}_\mu$-almost every $\omega \in \Sigma^{\mathbb{N}}$.

Note that

$$\frac{\|A^n(\omega)v_n\|}{\|A^n(\omega)\|} \geq |v_n\cdot \bar{v}^{(n)}(A)| \to |v\cdot \bar{v}^{(\infty)}(A)|$$

where $\bar{v}^{(n)}(A)(\omega) \in \mathbb{S}^{m-1}$ is the most expanding direction of the n-th iterate of $A^n(\omega)$ and $\bar{v}^{(\infty)}(A)(\omega)$ is such that $\bar{v}^{(n)}(A)(\omega) \to \bar{v}^{(\infty)}(A)(\omega)$.

On the other hand, $\bar{v}^{(\infty)}(A)^{\perp}$ is the sum of all invariant subspaces in the Osedelets decomposition associated with Lyapunov exponents $< L_1(A,K)$. Then, the quasi-irreducibility implies that

$$\liminf\frac{\|A^n(\omega)v_n\|}{\|A^n(\omega)\|} > 0$$

$\mathbb{P}_\mu$-almost surely. Therefore,

$$\frac{1}{n}\log\frac{\|A^n(\omega)v_n\|}{\|A^n(\omega)\|} \to 0$$

almost surely as $n \to \infty$. Using the Dominated Convergence Theorem

$$\begin{aligned}\lim\frac{1}{n}\mathbb{E}_x\left[\log\|A^n(\omega)v_n\|\right] &= \lim\frac{1}{n}\mathbb{E}_x\left[\log\|A^n(\omega)\|\right] + \lim\frac{1}{n}\mathbb{E}_x\left[\log\frac{\|A^n(\omega)v_n\|}{\|A^n(\omega)\|}\right]\\ &= L_1(A,K)+0 = L_1(A,K)\end{aligned}$$

which establishes the claim and proves that the limit

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}_{\omega_0}(\log\|A^{(n)}(\omega)v\|) = L_1(A,K),$$

is uniform in $\hat{v} \in \mathbb{S}^{m-1}$ when $\omega_0 \in \Sigma$ is fixed.

Now it remains to prove that the previous limit is uniform in both $(\omega_0, \hat{v})$. For every $M \in \mathrm{GL}_m(\mathbb{R})$, consider

$$\chi(M) = \sup\{\log \|M\|, \log \|M\|^{-1}\}$$

and define $g_n \colon \Sigma^{\mathbb{N}} \to \mathbb{R}$ such that

$$\begin{aligned} g_n(\omega) &= g_n(\omega_0) \\ &= \sup\left\{\left|\frac{1}{n}\mathbb{E}_{\omega_0}(\log \|A^{(n)}(\omega)v\|) - L_1(A,K)\right| : v \in \mathbb{S}^{m-1}, \omega = \{\omega_n\} \in \Sigma^{\mathbb{N}}\right\}. \end{aligned}$$

Since $-\chi(M) \leq \log \|Mu\| \leq \chi(M)$ for every $u \in \mathbb{S}^{m-1}$, the sequence g_n is uniformly bounded and

$$|g_n(\omega)| \leq \frac{1}{n}\mathbb{E}_{\omega_0}(\chi(A^{(n)}(\omega)) + |L_1(A,K)|\,.$$

Moreover, by (5.7), $g_n(\omega) \to 0$ when $n \to \infty$ for $\mathbb{P}_\mu$-a.e. $\omega \in \Sigma^{\mathbb{N}}$. Hence,

$$\lim_{n\to\infty} \int g_n(\omega)\; d\mathbb{P}_\mu(\omega) = 0.$$

Let $p \in \mathbb{N}$ such that $p < n$ and consider

$$v_p = \frac{A^p(\omega)v}{\|A^p(\omega)v\|}\,.$$

Then

$$\begin{aligned} &\left|\frac{1}{n}\mathbb{E}_{\omega_0}(\log \|A^{(n)}(\omega)v\|) - L_1(A,K)\right| \leq \\ &\leq \left|\frac{1}{n}\mathbb{E}_{\omega_0}(\log \|A^{(n-p)}(f^p\omega)v_p\|) - L_1(A,K)\right| + \left|\frac{1}{n}\mathbb{E}_{\omega_0}(\log \|A^{(p)}(\omega)v\|)\right| \\ &\leq \left|\frac{1}{n}\mathbb{E}_{\omega_0}(\mathbb{E}_{\omega_p}(\log \|A^{(n-p)}(f^p\omega)v_p\|)) - L_1(A,K)\right| + \frac{1}{n}\mathbb{E}_{\omega_0}(\chi(A^{(p)}(\omega))) \\ &\leq \mathbb{E}_{\omega_0}\left(\frac{1}{n-p}\mathbb{E}_{\omega_p}(\log \|A^{(n-p)}(f^p\omega)v_p\|) - L_1(A,K)\right) + \frac{p}{n}(|L_1(A,K)| + a)\,. \end{aligned}$$

Hence,

$$g_n(\omega) \leq \mathbb{E}_{\omega_0}(g_{n-p}(f^p\omega)) + \frac{p}{n}(|L_1(A,K)| + a)\,.$$

Since K is uniformly ergodic, there exists a sequence $\epsilon(p)$, where $\epsilon(p) \to 0$ when $p \to \infty$, such that

$$\sup_{\omega_0 \in \Sigma}\left|\int g_{n-p}(y)\; dK^p_{\omega_0}(y) - \int g_{n-p}(y)\; d\mathbb{P}_{\omega_0}(y)\right| \leq \epsilon(p)$$

for every n.

Then, for every p,

$$\lim_{n\to\infty} \sup_{\omega_0\in\Sigma} g_n(\omega_0) \le \lim_{n\to\infty} \int g_{n-p}(y)\, dK^p_{\omega_0}(y) + \epsilon(p) \le \epsilon(p),$$

and this concludes the proof. ■

5.4 Strong Mixing of the Markov Operator

In this section we show that the Markov operator $Q_{(A,K)}$ acts as a contraction on an appropriate space of observables.

Consider on the projective space $\mathbb{P}(\mathbb{R}^m)$ the distance

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

where p and q are representatives of $\hat{p}$ and $\hat{q}$ respectively.

Given $0<\alpha\le 1$ and $\phi\in L^\infty(\Sigma\times\mathbb{P}(\mathbb{R}^m))$, we define the Hölder seminorm v_α by:

$$v_\alpha(\phi) = \sup_{\substack{\omega_0\in\Sigma\\ \hat{p}\neq\hat{q}}} \frac{|\phi(\omega_0,\hat{p})-\phi(\omega_0,\hat{q})|}{\delta(\hat{p},\hat{q})^\alpha},$$

Definition 5.4.1 The space of α-Hölder continuous functions on $\mathbb{P}(\mathbb{R}^m)$ is given by

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

The Hölder norm $\|\cdot\|_\alpha$ on this space is defined by

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

Moreover, consider the average Hölder constant:

$$k_\alpha(A,K) = \sup_{\substack{\omega_0\in\Sigma\\ \hat{p}\neq\hat{q}}} \int \frac{\delta(\hat{A}(\omega_1,\omega_0)\hat{p},\hat{A}(\omega_1,\omega_0)\hat{q})^\alpha}{\delta(\hat{p},\hat{q})^\alpha}\, dK_{\omega_0}(\omega_1).$$

Let $Q_{A,K}\colon C^0(\Sigma\times\mathbb{P}(\mathbb{R}^m)) \to C^0(\Sigma\times\mathbb{P}(\mathbb{R}^m))$ be the Markov operator defined in (5.2). Namely,

$$(Q_{(A,K)}\phi)(\omega_0,\hat{p}) = \int \phi(\omega_1,\hat{A}(\omega_1,\omega_0)\hat{p})\, dK_{\omega_0}(\omega_1)$$

for every $\phi\in C^0(\Sigma\times\mathbb{P}(\mathbb{R}^m))$.

Proposition 5.4.1 *For all* $\phi \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$,

$$v_\alpha(Q_{A,K}(\phi)) \leq v_\alpha(\phi) k_\alpha(A, K).$$

Proof. Given $\phi \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ and $\hat{p}, \hat{q} \in \Sigma \times \mathbb{P}(\mathbb{R}^m)$,

$$\begin{aligned}
&\frac{|Q_{A,K}(\phi)(\hat{p}) - Q_{A,K}(\phi)(\hat{q})|}{\delta(\hat{p},\hat{q})^\alpha} = \\
&= \frac{\left|\int \phi(\omega_1, \hat{A}(\omega_1,\omega_0)\hat{p}) - \phi(\omega_1, \hat{A}(\omega_1,\omega_0)\hat{q})\, dK_{\omega_0}(\omega_1)\right|}{\delta(\hat{p},\hat{q})^\alpha} \\
&\leq \int \left| \frac{\phi(\omega_1, \hat{A}(\omega_1,\omega_0)\hat{p}) - \phi(\omega_1, \hat{A}(\omega_1,\omega_0)\hat{q})}{\delta(\hat{p},\hat{q})^\alpha} \right| dK_{\omega_0}(\omega_1) \\
&\leq v_\alpha(\phi) \int \frac{\delta(\hat{A}(\omega_1,\omega_0)\hat{p}, \hat{A}(\omega_1,\omega_0)\hat{q})^\alpha}{\delta(\hat{p},\hat{q})^\alpha}\, dK_{\omega_0}(\omega_1)\,.
\end{aligned}$$

Now, taking the supremum in $\omega_0 \in \Sigma$ and $\hat{p} \neq \hat{q} \in \mathbb{P}(\mathbb{R}^m)$ on both sides we conclude the proof. ■

Proposition 5.4.2 *The sequence* $\{k_\alpha(A^n, K^n)\}_n$ *is sub-multiplicative:*

$$k_\alpha(A^{m+n}, K^{m+n}) \leq k_\alpha(A^n, K^n) \cdot k_\alpha(A^m, K^m).$$

Proof. By definition,

$$\begin{aligned}
k_\alpha(A^{m+n}, K^{m+n}) &= \sup_{\substack{\omega_0 \in \Sigma \\ \hat{p} \neq \hat{q}}} \int \frac{\delta(\hat{A}^{m+n}(\omega)\hat{p}, \hat{A}^{m+n}(\omega)\hat{q})^\alpha}{\delta(\hat{p},\hat{q})^\alpha}\, dK^{m+n}_{\omega_0}(\omega_1, \omega_2, \ldots, \omega_{m+n}) \\
&\leq k_\alpha(A^n, K^n) \cdot k_\alpha(A^m, K^m)
\end{aligned}$$

and then, the sequence $\{k_\alpha(A^n, K^n)\}_n$ is sub-multiplicative. ■

Recall that we consider the metric space $(\mathcal{C}, d)$, where

$$\begin{aligned}
\mathcal{C} := \{(A,K)\colon\ & A\colon \Sigma \times \Sigma \to \mathrm{GL}_m(\mathbb{R}) \text{ is Lipschitz continuous and} \\
& K\colon \Sigma \to \mathrm{Prob}(\Sigma) \text{ is uniformly ergodic and} \\
& \text{continuous in the weak* topology.}\}
\end{aligned}$$

is the set of Markov cocycles and d is the metric defined as follows:

$$d((A,K),(B,L)) := \max\{d_\infty(A,B), d_{W_1}(K,L)\},$$

Proposition 5.4.3 *Fix* $n \in \mathbb{N}$. *The map* $\mathcal{C} \ni (A,K) \mapsto (A^n, K^n)$ *is Lipschitz with respect to the metric* d.

Proof. Note that the linear map $A \mapsto A^n$ is Lipschitz with constant $C(n)$ that depends on n but not on the kernel. Moreover, we claim that the map $K \mapsto K^n$ is also Lipschitz with constant n: for every Markov kernel $L\colon \Sigma \to \mathrm{Prob}(\Sigma)$,

$$\begin{aligned}
d(K^2, L^2) = & \sup_{\substack{\omega_0\in\Sigma \\ \varphi\in \mathrm{Lip}_1(\Sigma)}} \left| \int \varphi(\omega_2)\, d(K^2_{\omega_0}(\omega_2) - L^2_{\omega_0}(\omega_2)) \right| \\
= & \sup_{\substack{\omega_0\in\Sigma \\ \varphi\in \mathrm{Lip}_1(\Sigma)}} \left| \int \varphi(\omega_2)\, dK_{\omega_1}(\omega_2) dK_{\omega_0}(\omega_1) - \int \varphi(\omega_2)\, dL_{\omega_1}(\omega_2) dL_{\omega_0}(\omega_1) \right| \\
\leq & \sup_{\substack{\omega_0\in\Sigma \\ \varphi\in \mathrm{Lip}_1(\Sigma)}} \left| \int \varphi(\omega_2)\, dK_{\omega_1}(\omega_2) dK_{\omega_0}(\omega_1) - \int \varphi(\omega_2)\, dK_{\omega_1}(\omega_2) dL_{\omega_0}(\omega_1) \right| + \\
& + \left| \int \varphi(\omega_2)\, dK_{\omega_1}(\omega_2) dL_{\omega_0}(\omega_1) - \int \varphi(\omega_2)\, dL_{\omega_1}(\omega_2) dL_{\omega_0}(\omega_1) \right| \\
\leq & \, 2d(K, L).
\end{aligned}$$

The proof of the claim follows by induction. Hence, the joint map $(A, K) \mapsto (A^n, K^n)$ is Lipschitz with constant the maximum between $C(n)$ and n. ■

Lemma 5.4.1 *Given a pair $(A, K) \in (\mathcal{C}, d)$, then for all $\alpha > 0$,*

$$k_\alpha(A, K) \leq \sup_{\substack{\omega_0\in\Sigma \\ \hat{p}\in\mathbb{P}(\mathbb{R}^m)}} \int_\Sigma \left(\frac{s_1(\hat{A}(\omega_1, \omega_0)) s_2(\hat{A}(\omega_1, \omega_0))}{\|\hat{A}(\omega_1, \omega_0)\hat{p}\|^2} \right)^\alpha dK_{\omega_0}(\omega_1),$$

where $s_1(\cdot)$ and $s_2(\cdot)$ are the first and second singular values.

Proof. It is enough to prove that given $\alpha > 0$ and two points $\hat{p}, \hat{q} \in \mathbb{P}(\mathbb{R}^m)$, we have

$$\left[\frac{\delta(\hat{A}(\omega)\hat{p}, \hat{A}(\omega)\hat{q})}{\delta(\hat{p}, \hat{q})} \right]^\alpha \leq \frac{|s_1(A(\omega)) s_2(A(\omega))|^\alpha}{2} \left[\frac{1}{\|A(\omega)p\|^{2\alpha}} + \frac{1}{\|A(\omega)q\|^{2\alpha}} \right]$$

for every $\omega_0 \in \Sigma$.

In fact, if we integrate with respect to the measure K_{ω_0} and take the supremum in $\hat{p} \neq \hat{q}$ on both sides of this inequality, we conclude the lemma.

By the definition of the projective distance,

$$\left[\frac{\delta(\hat{A}(\omega)\hat{p}, \hat{A}(\omega)\hat{q})}{\delta(\hat{p}, \hat{q})} \right]^\alpha = \left[\frac{\|A(\omega_1, \omega_0)p \wedge A(\omega_1, \omega_0)q\|}{\|A(\omega)p\| \|A(\omega)q\|} \right]^\alpha \left[\frac{\|p\| \|q\|}{\|p \wedge q\|} \right]^\alpha. \tag{5.8}$$

On the other hand, by the exterior product property,

$$\|A(\omega_1, \omega_0)p \wedge A(\omega_1, \omega_0)q\| = |s_1(A(\omega_1, \omega_0)) s_2(A(\omega_1, \omega_0))| \|p \wedge q\|.$$

which when combined with (5.8), we have

$$\left[\frac{\delta(\hat{A}(\omega)\hat{p}, \hat{A}(\omega)\hat{q})}{\delta(\hat{p}, \hat{q})}\right]^\alpha = \left[\frac{|s_1(A(\omega))s_2(A(\omega))|}{\|A(\omega)p\|\|A(\omega)q\|}\right]^\alpha$$
$$\leq \frac{|s_1(A(\omega))s_2(A(\omega))|^\alpha}{2}\left[\frac{1}{\|A(\omega)p\|^{2\alpha}} + \frac{1}{\|A(\omega)q\|^{2\alpha}}\right]$$

since the geometric mean is less or equal than the arithmetic mean. ■

Proposition 5.4.4 *Let $(A, K) \in (\mathcal{C}, d)$. Assume that*

(i) A is quasi irreducible with respect to (K, μ),

(ii) $L_1(A, K) > L_2(A, K)$.

Then, there are numbers $\delta > 0$, $0 < \alpha < 1$, $0 < \sigma < 1$ and $n \in \mathbb{N}$ such that for all $(B, L) \in (\mathcal{C}, d)$ with $d((B, L), (A, K)) < \delta$ one has $k_\alpha(B^n, L^n) < \sigma$.

Proof. By theorem 5.3.4, given $(A, K) \in (\mathcal{C}, d)$ satisfying assumptions (i) and (ii), we have

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}_{\omega_0}(\log \|A^{(n)}(\omega)v\|^{-2}) = -2L_1(A, K),$$

with uniform convergence in $(\omega_0, \hat{v}) \in \Sigma \times \mathbb{S}^{d-1}$.

Hence, by choosing ϵ small enough e.g $\frac{1}{4}(L_1(A, K) - L_2(A, K))$ and n sufficiently large, we conclude that for all $(\omega_0, v) \in \Sigma \times \mathbb{S}^{m-1}$

$$\mathbb{E}_{\omega_0}(\log \|A^{(n)}(\omega)v\|^{-2}) \leq n(-2L_1(A, K) + \epsilon).$$

Let $\Lambda_2 A$ denote its corresponding second exterior power. Note that

$$\|\Lambda_2 M\| = s_1(\Lambda_2 M) = s_1(M) \cdot s_2(M)$$

for every $M \in \mathrm{GL}_m(\mathbb{R})$. So in particular, $L_1(\Lambda_2 A) = L_1(A) + L_2(A)$.

Applying Theorem 5.3.1 to the cocycle $\Lambda_2 A$, we get that for every $\omega_0 \in \Sigma$ and for $\mathbb{P}_{\omega_0}$-almost every $\omega \in \Sigma^{\mathbb{N}}$,

$$\frac{1}{n}\log \|\Lambda_2 A^n(\omega)\| \to L_1(\Lambda_2 A)\,.$$

By dominated convergence,

$$\lim_{n\to\infty} \mathbb{E}_{\omega_0}\left(\frac{1}{n}\log \|\Lambda_2 A^{(n)}(\omega)\|\right) \to L_1(\Lambda_2 A),$$

this shows that

$$\lim_{n\to\infty} \mathbb{E}_{\omega_0}\left(\frac{1}{n}\log s_1(A^n(\omega)) \cdot s_2(A^n(\omega))\right) \to L_1(A, K) + L_2(A, K)$$

So for n large enough

$$\mathbb{E}_{\omega_0}\left(\frac{1}{n}\log s_1 s_2(A^n(\omega))\right) \leq L_1(A,K) + L_2(A,K) + \varepsilon$$

Since $L_1(A,K) > L_2(A,K)$, for n sufficiently large,

$$\mathbb{E}_{\omega_0}\log\left[\frac{s_1(A^{(n)}(\omega))s_2(A^{(n)}(\omega))}{\|A^{(n)}(\omega)v\|^2}\right] \leq -1, \tag{5.9}$$

for all $(\omega_0, \hat{v}) \in \Sigma \times \mathbb{S}^{m-1}$.

By the inequality $e^x \leq 1 + x + \frac{x^2}{2}e^{|x|}$, we conclude that for every $v \in \mathbb{S}^{m-1}$ and every $\omega_0 \in \Sigma$,

$$\begin{aligned}
\mathbb{E}_{\omega_0}\left[\frac{|s_1(A^{(n)}(w))||s_2(A^{(n)}(\omega))|}{\|A^{(n)}(\omega)v\|^2}\right]^\alpha &= \mathbb{E}_{\omega_0}\exp\left(\log\left[\frac{|s_1(A^{(n)}(w))||s_2(A^{(n)}(\omega))|}{\|A^{(n)}(\omega)v\|^2}\right]^\alpha\right)\\
&\leq 1 + \mathbb{E}_{\omega_0}\left[\alpha\log\frac{|s_1(A^{(n)}(\omega))||s_2(A^{(n)}(\omega))|}{\|A^{(n)}(\omega)v\|^2}\right] + \\
&+ \mathbb{E}_{\omega_0}\left[\frac{\alpha^2}{2}\log^2\frac{|s_1(A^{(n)}(\omega))||s_2(A^{(n)}(\omega))|}{\|A^{(n)}(\omega)v\|^2}e^{\frac{|\alpha\log|s_1(A^{(n)}(\omega))s_2(A^{(n)}(\omega))||}{\|A^{(n)}(\omega)v\|^2}}\right]\\
&\leq 1 - \alpha + C\frac{\alpha^2}{2}.
\end{aligned}$$

Note that C is a constant that depends only on (A,K) and n. Thus, we can choose α small enough such that

$$\mathbb{E}_{\omega_0}\left[\frac{|s_1(A^{(n)}(w))||s_2(A^{(n)}(\omega))|}{\|A^{(n)}(\omega)v\|^2}\right]^\alpha < 1.$$

Hence, by Lemma 5.4.1,

$$k_\alpha(A,K) \leq \sup_{\substack{\omega_0\in\Sigma\\ \hat{p}\in\mathbb{P}(\mathbb{R}^m)}} \mathbb{E}_{\omega_0}\exp(\log\left[\frac{|s_1(A^{(n)}(w))||s_2(A^{(n)}(\omega))|}{\|A^{(n)}(\omega)v\|^2}\right]^\alpha) \leq 1.$$

Moreover, $k_\alpha(A,K)$ depends continuously on (A,K) and, by proposition 5.4.3, the map $(A,K) \mapsto (A^{(n)}, K^n)$ is Lipschitz, therefore we can extend the result to a neighborhood of (A,K). ■

Let $(\mathcal{E}, \|\cdot\|_{\mathcal{E}})$ be a Banach space where $\mathcal{E} \subset C^0(M)$ is Q-invariant in the sense that $\varphi \in \mathcal{E}$ if, and only if, $Q\varphi \in \mathcal{E}$. Moreover, we assume that the constant function $\mathbb{1} \in \mathcal{E}$ and that the inclusion of $\mathcal{E} \subset C^0(M)$ is continuous, namely $\|\varphi\|_\infty \leq C_1\|\varphi\|_{\mathcal{E}}$ for some constant $C_1 < \infty$. We also assume that Q is bounded (or continuous) on $(\mathcal{E}, \|\cdot\|_{\mathcal{E}})$, i.e. $\|Q\varphi\|_{\mathcal{E}} \leq C_2\|\varphi\|_{\mathcal{E}}$ with $C_2 < \infty$. In practice we will have $C_1 = C_2 = 1$.

Definition 5.4.2 (Strong mixing) Let $(M, \mathcal{K}, \nu)$ be an abstract Markov system and let $\mathcal{Q}\colon C^0(M) \to C^0(M)$,

$$(\mathcal{Q}\varphi)(x) := \int_M \varphi(y)\, dK_x(y)$$

be the associated Markov operator. Let $(\mathcal{E}, \|\cdot\|_{\mathcal{E}})$ *be a Banach space as above. We say that* $\mathcal{Q}$ *is* strongly mixing *on* $\mathcal{E}$ *if there are* $C < \infty$ *and* $\sigma \in (0,1)$ *such that for all* $n \in \mathbb{N}$ *and* $\varphi \in \mathcal{E}$,

$$\left\| Q^n\varphi - \int_M \varphi d\mu \right\|_\infty \leq C\sigma^n \|\varphi\|_{\mathcal{E}} .$$

Note that for the Markov system (Σ, K, μ) that generates the base dynamics, by assumption, we have $K^n_{\omega_0} \to \mu$ uniformly in $\omega_0 \in \Sigma$, which is equivalent to the following:

$$\left\| Q^n\varphi - \int_M \varphi d\mu \right\|_\infty \leq C\sigma^n \|\varphi\|_\infty$$

for some $C < \infty$ and $0 < \sigma < 1$.

Thus Q is strongly mixing on $L^\infty(\mu)$ (this is strongest form of strong mixing).

We will prove that the Markov operator $Q_{(A,K)}$, corresponding to the Markov cocycle (A, K), is strongly mixing on the space of Hölder functions $\mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$. Indeed, this holds in a neighborhood of A as shown by the following theorem.

Theorem 5.4.5 *Given* $(A, K) \in (\mathcal{C}, d)$ *such that the assumptions of proposition 5.4.4 are satisfied, there exist constants* $C < \infty$, $0 < \sigma < 1$ *and a neighborhood* U *of* (A, K) *in* $(\mathcal{C}, d)$ *such that for all* $(B, L) \in U$, $Q_{B,L}$ *is strongly mixing on* $\mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$*:*

$$\left\| Q^n_{B,L}\varphi - \int_{\Sigma\times\mathbb{P}(\mathbb{R}^m)} \varphi d\eta_{B,L} \right\|_\infty \leq C\sigma^n \|\varphi\|_\alpha , \forall\, \varphi \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m)).$$

Moreover, since the v_α *seminorm is also exponentially contracting on* $\mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$*, we further get*

$$\left\| Q^n_{B,L}\varphi - \int_{\Sigma\times\mathbb{P}(\mathbb{R}^m)} \varphi d\eta_{B,L} \right\|_\alpha \leq C\sigma^n \|\varphi\|_\alpha , \forall\, \varphi \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m)).$$

Remark 5.1 *The contraction constant* σ *can be choosen as the square root of the minimum between the contracting rate of the* v_α *seminorm and the* *convergence rate of the kernel* $L : \Sigma \to \mathrm{Prob}(\Sigma)$*, and we may always choose them to be the same by setting properly the size of the neighbourhood.*

Proof. Take $(B, L) \in U$ where U is given by Proposition 5.4.4, $n \geq m$ with $n, m \in \mathbb{N}$, $\varphi \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ and $\eta_{B,L}$ any L_B-stationary measure, define 4 families of transformations in the following way: for any $(\omega_0, p) \in \Sigma \times \mathbb{P}(\mathbb{R}^m)$,

(1) $(T^{(0)}_{B,L,n}\varphi)(\omega_0, p) := (Q^n_{B,L}\varphi)(\omega_0, p) = \mathbb{E}_{\omega_0}[\varphi(\omega_n, B^{(n)}p)]$.
(2) $(T^{(1)}_{B,L,n,m}\varphi)(\omega_0, p) := \mathbb{E}_{\omega_0}[\varphi(\omega_n, (B^{(m)} \circ T^{n-m})p)]$.
(3) $(T^{(2)}_{B,L,m}\varphi)(\omega_0, p) := \mathbb{E}_\mu[\varphi(\omega_n, B^{(m)}p)]$, constant in ω_0, thus we denote it by $(T^{(2)}_{B,L,m}\varphi)(p)$ which is a compact transformation.
(4) $(T^{(3)}_{B,L}\varphi)(\omega_0, p) := \int \varphi d\eta_{B,L}$, constant.

Then it is straightforward to obtain the following inequalities:

(1) $\left|(T^{(0)}_{B,L,n}\varphi)(\omega_0, p) - T^{(1)}_{B,L,n,m}\varphi)(\omega_0, p)\right| \leq C\sigma^m \|\varphi\|_\alpha$ for the same σ using the contracting property of the v_α seminorm.
(2) $\left|T^{(1)}_{B,L,n,m}\varphi)(\omega_0, p) - (T^{(2)}_{B,L,m}\varphi)(p)\right| \leq C\sigma^{n-m} \|\varphi\|_\alpha$ using the uniform convergence rate of $L^{n-m}_{\omega_0} \to \mu$.
(3) $\left|(T^{(2)}_{B,L,m}\varphi)(p) - (T^{(2)}_{B,L,n}\varphi)(p)\right| \leq C\sigma^m \|\varphi\|_\alpha$ using again the contracting property of the v_α seminorm.

For simplicity, we may set $n = 2m$ in (1) and (2), and set $n = l$ in (3), then by (1)-(3), we have for all $B \in U$ and $\varphi \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$,

$$\left\|Q^{2m}_{B,L}\varphi - T^{(2)}_{B,L,l}\varphi\right\|_\infty \leq 3C\sigma^m \|\varphi\|_\alpha .$$

Note that the sequence $\{T^{(2)}_{B,L,l}\varphi\}_{l\geq 0}$ is relatively compact in $C(\mathbb{P}(\mathbb{R}^m))$. Then the set S_φ of its limit points in $(C(\mathbb{P}(\mathbb{R}^m)), \|\cdot\|_\infty)$ is non-empty. Considering any $g \in S_\varphi$, we claim that

$$g = \int \varphi d\eta_{B,L} = T^{(3)}_{B,L}\varphi.$$

Let us prove the claim. Take a subsequence $l_j \to \infty$ such that $\{T^{(2)}_{B,L,l_j}\varphi\}_{j\geq 0}$ converges to g in the previous inequality, we get

$$\left\|Q^{2m}_{B,L}\varphi - g\right\|_\infty \leq 3C\sigma^m \|\varphi\|_\alpha .$$

On the other hand, we have $v_\alpha(Q^{2m}_{B,L}\varphi) \leq C\sigma^{2m} \|\varphi\|_\alpha$. This implies $v_\alpha(g) = 0$ which further implies that g is constant. Finally, using the condition that $\eta_{B,L}$ is L_B-stationary, we have

$$\int Q^{2m}_{B,L}\varphi d\eta_{B,L} = \int \varphi d\eta_{B,L}$$

which equals g. Take $\sigma' = \sigma^{\frac{1}{2}}$ as a new parameter, and this completes the proof. ∎

Corollary 5.4.6 (Uniqueness of the stationary measure) *Given $(A, K) \in (\mathcal{C}, d)$ such that the assumptions of proposition 5.4.4 are satisfied, the Markov operator $Q_{B,L}$ has a unique stationary measure $\eta_{(B,L)}$ for every $(B, L) \in U$, which further gives that $\bar{L}_B$ has a unique stationary measure $L \ltimes \eta_{B,L}$.*

Proof. Assume there are two different stationary measures $\eta_{B,L}$ and $\eta'_{B,L}$, using Theorem 5.4.5 and the fact that $\mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ is dense in $L^\infty(\Sigma \times \mathbb{P}(\mathbb{R}^m))$, we get $\eta_{B,L} = \eta'_{B,L}$. ∎

As a result, we can finally upgrade the Furstenberg formula as follows.

Theorem 5.4.7 *Given (A, K) in $(\mathcal{C}, d)$ such that the assumptions of proposition 5.4.4 are satisfied, there exists a neighborhood U of $(A, K) \in (\mathcal{C}, d)$ such that for every $(B, L) \in U$*

$$L_1(B, L) = \int_{\Sigma\times\Sigma\times\mathbb{P}(\mathbb{R}^m)} \psi(y, x, \hat{v})\, dL_x(y) d\eta_{B,L}(x, \hat{v})$$

where $\psi : \Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m) \to \mathbb{R}$ is such that

$$\psi(y, x, \hat{v}) = \log \frac{\|A(y, x)v\|}{\|v\|}$$

and $\eta_{B,L}$ is the unique $Q_{B,L}$-stationary measure.

5.5 Joint Continuity of the Lyapunov Exponents

In this section we establish the Hölder continuity of the maximal Lyapunov exponents as a function of the input data, namely the fiber map A and the transition kernel K. We use the technique introduced in [1]. This approach, via the Furstenberg Formula, enables the study of this type of continuity without the need of going through the theory of large deviations. Moreover, it also has the advantage of providing a precise computation of the Hölder exponent.

Lemma 5.5.1 *Given two differents $p, q \in \mathbb{P}\mathbb{R}^d \setminus \{0\}$,*

$$\left\| \frac{p}{\|p\|} - \frac{q}{\|q\|} \right\| \leq \|p - q\| \max\left\{ \frac{1}{\|p\|}, \frac{1}{\|q\|} \right\}.$$

Corollary 5.5.1 *Let $g_1, g_2 \in \mathrm{SL}_d(\mathbb{R})$ and consider the projective actions, $\hat{g}_i \colon \mathbb{P}(\mathbb{R}^m) \to \mathbb{P}(\mathbb{R}^m)$ taking $\hat{p}$ to $\frac{g_i(p)}{\|g_i(p)\|}$ for $i = 1, 2$.*

Then

$$\delta(\hat{g}_1\hat{p}, \hat{g}_2\hat{p}) \leq \|g_1 - g_2\| \max\left\{\frac{1}{\|g_1(p)\|}, \frac{1}{\|g_2(p)\|}\right\}.$$

Proof. Observe that

$$\delta(\hat{g}_1\hat{p}, \hat{g}_2\hat{p}) \leq \left\|\frac{g_1(p)}{\|g_1(p)\|} - \frac{g_2(q)}{\|g_2(q)\|}\right\|$$

Then, by Lemma 5.5.1,

$$\begin{aligned}\left\|\frac{g_1(p)}{\|g_1(p)\|} - \frac{g_2(q)}{\|g_2(q)\|}\right\| &\leq \|g_1(p) - g_2(p)\| \max\left\{\frac{1}{\|g_1(p)\|}, \frac{1}{\|g_2(p)\|}\right\} \\ &\leq \|g_1 - g_2\| \max\left\{\frac{1}{\|g_1(p)\|}, \frac{1}{\|g_2(p)\|}\right\}\end{aligned}$$

■

In the next proposition, we show that for every quasi-irreducible Markov cocycle and every $n \in \mathbb{N}$, the map $A \mapsto Q_A^n$ is locally Hölder. We extend it to the mixing Markov case in (A, K). More precisely, we prove the following

Proposition 5.5.2 *Let $(A, K) \in (\mathcal{C}, d)$. Assume that:*

(i) A is quasi irreducible with respect to (K, μ),

(ii) $L_1(A, K) > L_2(A, K)$.

Then, there exists $\delta > 0$, such that for all (B, L) and (D, T) in $(\mathcal{C}, d)$ satisfying $d((A, K), (B, L)) < \delta$ and $d((A, K), (D, T)) < \delta$, for all $f \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ and every $n \in \mathbb{N}$,

$$\|Q_{B,L}^n f - Q_{D,T}^n f\|_\infty \leq C d((B, L), (D, T))^\alpha v_\alpha(f).$$

Proof. First consider the case $n = 1$. Since

$$\begin{aligned}&\|(Q_{B,L} - Q_{D,T})(f)\|_\infty \\ &= \sup_{\substack{\hat{v} \in \mathbb{P}(\mathbb{R}^m) \\ \omega_0 \in \Sigma}} \left|\int_\Sigma f(\omega_1, B(\omega_1, \omega_0)v)\, dL_{\omega_0}(\omega_1) - \int_\Sigma f(\omega_1, D(\omega_1, \omega_0)v)\, dT_{\omega_0}(\omega_1)\right|\end{aligned}$$

For every $\omega_0 \in \Sigma$ and $v \in \mathbb{P}(\mathbb{R}^m)$,

$$\begin{aligned}
&\|(Q_{B,L} - Q_{D,T})(f)\|_\infty \\
&\leq \sup_{\substack{\hat{v}\in\mathbb{P}(\mathbb{R}^m)\\ \omega_0\in\Sigma}} \left|\int_\Sigma f(\omega_1, B(\omega_1,\omega_0)v) - f(\omega_1, D(\omega_1,\omega_0)v)\; dL_{\omega_0}(\omega_1)\right| + \\
&+ \sup_{\substack{\hat{v}\in\mathbb{P}(\mathbb{R}^m)\\ \omega_0\in\Sigma}} \left|\int_\Sigma f(\omega_1, D(\omega_1,\omega_0)v)\; dL_{\omega_0}(\omega_1) - \int_\Sigma f(\omega_1, D(\omega_1,\omega_0)v)\; dT_{\omega_0}(\omega_1)\right|.
\end{aligned}$$

Since f is Hölder, we can bound the first term by

$$\begin{aligned}
&\sup_{\substack{\hat{v}\in\mathbb{P}(\mathbb{R}^m)\\ \omega_0\in\Sigma}} \left|\int_\Sigma f(\omega_1, B(\omega_1,\omega_0)v) - f(\omega_1, D(\omega_1,\omega_0)v)\; dL_{\omega_0}(\omega_1)\right| \leq \\
&\leq \sup_{\substack{\hat{v}\in\mathbb{P}(\mathbb{R}^m)\\ \omega_0\in\Sigma}} \int_\Sigma |f(\omega_1, B(\omega_1,\omega_0)v) - f(\omega_1, D(\omega_1,\omega_0)v)|\; dL_{\omega_0}(\omega_1) \\
&\leq \int_\Sigma v_\alpha(f)\; \delta(B(\omega_1,\omega_0)v, D(\omega_1,\omega_0)v)^\alpha\; dL_{\omega_0}(\omega_1) \\
&\leq v_\alpha(f)\; d_\infty(B,D)^\alpha \\
&\leq v_\alpha(f)\; d((B,L),(D,T))^\alpha
\end{aligned}$$

since $\int_\Sigma \delta(B(\omega_1,\omega_0)v, D(\omega_1,\omega_0)v)^\alpha\; dL_{\omega_0}(\omega_1) \leq d_\infty(B,D)^\alpha$

Now we proceed to estimate the second term. By Corollary 5.5.1, for every $\pi \in \Pi(L_{\omega_0}, T_{\omega_0})$:

$$\begin{aligned}
&\sup_{\substack{\hat{v}\in\mathbb{P}(\mathbb{R}^m)\\ \omega_0\in\Sigma}} \left|\int_\Sigma f(\omega_1, D(\omega_1,\omega_0)v)\; dL_{\omega_0}(\omega_1) - \int_\Sigma f(z_1, D(z_1,\omega_0)v)\; dT_{\omega_0}(z_1)\right| \\
&= \sup_{\substack{\hat{v}\in\mathbb{P}(\mathbb{R}^m)\\ \omega_0\in\Sigma}} \left|\int_{\Sigma\times\Sigma} f(\omega_1, D(\omega_1,\omega_0)v) - f(z_1, D(z_1,\omega_0)v)\; d\pi(\omega_1,z_1)\right| \\
&\leq v_\alpha(f) \sup_{\substack{\hat{v}\in\mathbb{P}(\mathbb{R}^m)\\ \omega_0\in\Sigma}} \int_{\Sigma\times\Sigma} \delta\left(D(\omega_1,\omega_0)v, D(z_1,\omega_0)v\right)^\alpha\; d\pi(\omega_1,z_1)
\end{aligned}$$

Hence,

$$\begin{aligned}
&\sup_{\substack{\hat{v}\in\mathbb{P}(\mathbb{R}^m)\\ \omega_0\in\Sigma}} \left|\int_\Sigma f(\omega_1, D(\omega_1,\omega_0)v)\; dL_{\omega_0}(\omega_1) - \int_\Sigma f(z_1, D(z_1,\omega_0)v)\; dT_{\omega_0}(z_1)\right| \\
&\leq v_\alpha(f) \sup_{\substack{\hat{v}\in\mathbb{P}(\mathbb{R}^m)\\ \omega_0\in\Sigma}} \int_{\Sigma\times\Sigma} \|D(\omega_1,\omega_0) - D(z_1,\omega_0)\|^\alpha \times \\
&\times \max\left\{\frac{1}{\|D(\omega_1,\omega_0)(v)\|}, \frac{1}{\|D(z_1,\omega_0)(v)\|}\right\}^\alpha\; d\pi(\omega_1,z_1).
\end{aligned}$$

Since Σ is compact, there exists a constant $C_1 > 0$, such that

$$\max\left\{\frac{1}{\|D(\omega_1,\omega_0)(v)\|}, \frac{1}{\|D(z_1,\omega_0)(v)\|}\right\}^{\alpha} \leq C_1.$$

Then, for every $\pi \in \Pi(L_{\omega_0}, T_{\omega_0})$, we can bound the second term by

$$\begin{aligned}
&C_1 v_\alpha(f) \sup_{\omega_0\in\Sigma} \int_{\Sigma\times\Sigma} \|D(\omega_1,\omega_0) - D(z_1,\omega_0)\|^{\alpha} \, d\pi(\omega_1, z_1) \\
&\leq C_2 v_\alpha(f) \sup_{\omega_0\in\Sigma} \left(\int_{\Sigma\times\Sigma} \|(\omega_1,\omega_0) - (z_1,\omega_0)\| \, d\pi(\omega_1,z_1)\right)^{\alpha} \\
&\leq C_2 v_\alpha(f) \; d(L,T)^{\alpha} \\
&\leq C_2 v_\alpha(f) \; d((B,L),(D,T))^{\alpha},
\end{aligned}$$

where on the second line we used the Lipschitz continuity of the map $x \mapsto D(x,y)$ and Jensen's inequality together with the concavity of the function $t \mapsto t^{\alpha}$, which holds when $t \in [0,\infty)$ and $\alpha \in (0,1]$.

Therefore, we conclude the case $n = 1$:

$$\|(Q_{B,L} - Q_{D,T})(f)\|_{\infty} \leq C_2 v_\alpha(f) \; d((B,L),(D,T))^{\alpha}.$$

Now observe that the difference $Q^n_{B,L} - Q^n_{D,T}$ can be written as a telescopic sum as follows:

$$\begin{aligned}
Q^n_{B,L} - Q^n_{D,T} &= Q^n_{B,L} - Q_{D,T} \circ Q^{n-1}_{B,L} + Q_{D,T} \circ Q^{n-1}_{B,L} - \cdots + Q^{n-1}_{D,T} \circ Q_{B,L} - Q^n_{D,T} \\
&= \sum_{i=0}^{n-1} Q^i_{D,T} \circ (Q_{B,L} - Q_{D,T}) \circ Q^{n-i-1}_{B,L}.
\end{aligned}$$

Then, using the triangle inequality, the fact that the norm of the Markov operator is 1 and the case $n = 1$, we obtain:

$$\begin{aligned}
\|Q^n_{B,L}(f) - Q^n_{D,T}(f)\| &= \left\|\sum_{i=0}^{n-1} Q^i_{D,T} \circ (Q_{B,L} - Q_{D,T}) \circ Q^{n-i-1}_{B,L}(f)\right\|_{\infty} \\
&\leq \sum_{i=0}^{n-1} \left\|Q^i_{D,T} \circ (Q_{B,L} - Q_{D,T}) \circ Q^{n-i-1}_{B,L}(f)\right\|_{\infty} \\
&\leq \sum_{i=0}^{n-1} \left\|(Q_{B,L} - Q_{D,T}) \circ Q^{n-i-1}_{B,L}(f)\right\|_{\infty} \\
&\leq C_2 d((B,L),(D,T))^{\alpha} \sum_{i=0}^{n-1} v_\alpha\left(Q^{n-i-1}_{B,L}(f)\right).
\end{aligned}$$

Since the operator contracts its seminorm v_α (see propositions 5.4.1 and 5.4.4), we conclude that there exists $\delta > 0$ and $\sigma < 1$ such that, if $d((A,K),(B,L)) < \delta$, then $k_\alpha(B^n, L^n) < \sigma$. Moreover, since k_α is sub multiplicative, we conclude:

$$\begin{aligned}\|Q^n_{B,L}(f) - Q^n_{D,T}(f)\| &\leq C_2 d((B,L),(D,T))^\alpha v_\alpha(f) \sum_{i=0}^{\infty} k_\alpha(B^i, L^i) \\ &\leq C d((B,L),(D,T))^\alpha v_\alpha(f).\end{aligned}$$

■

Since the map $(A,K) \mapsto Q^n_{A,K}$ is locally Hölder and $Q^n_{A,K}$ converges to the stationary measure $\eta_{A,K}$ (in the sense of 5.4.5), we can finally prove that the map $(A,K) \mapsto \eta_{A,K}$ is also locally Hölder.

Corollary 5.5.3 *Let $(A,K) \in (\mathcal{C}, d)$ such that the assumptions of proposition 5.4.4 are satisfied. Then there exists $\delta > 0$ such that for all (B,L) and (D,T) that are δ-close to (A,K) and for all $f \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$, we have*

$$\left| \int f \, d\eta_{B,L} - \int f \, d\eta_{D,T} \right| \leq C d((B,L),(D,T))^\alpha v_\alpha(f).$$

Proof. By lemma 5.4.6, there are unique stationary measures $\eta_{B,L}$ and $\eta_{D,T}$ associated with the Markov kernels L_B and T_D respectively. Moreover, by theorem 5.4.5,

$$\lim_{n\to\infty} Q^n_{B,L}(f) = \left(\int f \, d\eta_{B,L} \right) \mathbf{1} \quad \text{and} \quad \lim_{n\to\infty} Q^n_{D,T}(f) = \left(\int f \, d\eta_{D,T} \right) \mathbf{1},$$

where $\mathbf{1}$ is the constant function such that $\mathbf{1}(\omega_0, v) = 1$.

Therefore, we conclude that:

$$\begin{aligned}\left| \int f \, d\eta_{B,L} - \int f \, d\eta_{D,T} \right| &\leq \sup_{n\to\infty} \|Q^n_{B,L}(f) - Q^n_{D,T}(f)\| \\ &\leq C d((B,L),(D,T))^\alpha v_\alpha(f).\end{aligned}$$

■

Given a kernel $K\colon \Sigma \to \text{Prob}(\Sigma)$ and a probability measure η on $\Sigma \times \mathbb{P}(\mathbb{R}^m)$, let $K \ltimes \eta$ be a probability measure on $\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m)$ such that for every $\psi \in C(\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m))$,

$$\int \psi(\omega_1, \omega_0, \hat{v}) \, d(K \ltimes \eta)(\omega_1, \omega_0, \hat{v}) = \int \psi(\omega_1, \omega_0, \hat{v}) \, dK_{\omega_0}(\omega_1) \, d\eta(\omega_0, \hat{v}).$$

An immediate consequence of the previous corollary is that the map $(A,K) \mapsto m_{A,K} := K \ltimes \eta_{A,K}$ is also locally Hölder, where $m_{A,K}$ is the unique stationary measure on $\text{Prob}(\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m))$ associated to the kernel $\bar{K}$.

We are now ready to prove the local Hölder continuity of the Lyapunov exponents.

Theorem 5.5.4 *Let* $(A, K) \in (\mathcal{C}, d)$*. Assume that:*

(i) *A is quasi irreducible with respect to* (K, μ)*,*

(ii) $L_1(A, K) > L_2(A, K)$*.*

Then, there exists a neighborhood V *of* (A, K) *in* $(\mathcal{C}, d)$ *where the map* $(A, K) \mapsto L_1(A, K)$ *is Hölder continuous.*

Proof. By hypotheses (i) and (ii), we are in the setting of Theorem 5.4.7, thus we can express the top Lyapunov exponent $L_1(A, K)$ as

$$L_1(A, K) = \int_{\Sigma\times\Sigma\times\mathbb{P}(\mathbb{R}^m)} \psi_A(\omega_1, \omega_0, \hat{v})\, dK_{\omega_0}(\omega_1) d\eta_{A,K}(\omega_0, \hat{v}) = \int \psi_A \, dm_{A,K} \ ,$$

where $\psi_A(\omega_1, \omega_0, \hat{v}) = \log \frac{\|A(\omega_1,\omega_0)v\|}{\|v\|}$ and $m_{A,K}$ is the unique stationary measure associated to the Markov kernel $\bar{K}_A$.

Moreover, there exists a neighborhood V of (A, K) in $(\mathcal{C}, d)$ such that for every (B, L) and (D, T) in V, we can express their top Lyapunov exponent using Furstenberg's Formula.

Therefore, by corollary 5.5.3 and the fact that $A \mapsto \psi_A$ is locally Lipschitz, we estimate:

$$\begin{aligned}
&|L_1(B, L) - L_1(D, T)| = \left| \int \psi_B \, dm_{B,L} - \int \psi_D \, dm_{D,T} \right| \\
&\leq \left| \int \psi_B \, dm_{B,L} - \int \psi_B \, dm_{D,T} \right| + \left| \int \psi_B \, dm_{D,T} - \int \psi_D \, dm_{D,T} \right| \\
&\leq C d((B, L), (D, T))^\alpha
\end{aligned}$$

and this concludes the proof. ■

The Hölder coefficient α above is computable based on the input data. More precisely, we iterate the cocycle (A, K) a sufficient number n of times, until the inequality (5.9) holds (the existence of such a number of iterates is guaranteed by the our assumptions). Then α is chosen such that $1-\alpha+C\frac{\alpha^2}{2} < 1$, where the constant C depends explicitly on the data.

6
Linear cocycles over mixed Markov-quasiperiodic dynamics

In this chapter we consider linear cocycles over mixed Markov-quasiperiodic base dynamics. We begin whith the definition of the model (Section 6.1) and after which we derive a version of Kifer's non-random filtration in this setting (Section 6.2). In the last section we obtain an upper large deviations estimate for the fiber dynamics, and as a consequence of that, the upper semi continuity of the Lyapunov exponent. This forms the starting point of the study of mixed Markov-quasiperiodic cocycles, whose ultimate goal (left for a future project) is to establish full large deviations and Hölder continuity of the Lyapunov exponent for such cocycles.

6.1
The model

Let us recall the base dynamics that was introduced in the Chapter 4.

Let (Σ, K, μ) be a Markov system, that is, Σ is a compact metric space, $K\colon \Sigma \to \text{Prob}(\Sigma)$ is a Markov kernel and μ is a K-stationary measure. Let $\mathbb{P} = \mathbb{P}_K = \mathbb{P}_{(K,\mu)}$ denote the Markov measure on $X = \Sigma^{\mathbb{Z}}$ with initial distribution μ and transition kernel K. The two-sided shift is the map $\sigma : X \to X$ such that

$$\sigma(\{x_n\}_{n\in\mathbb{Z}}) = \{x_{n+1}\}_{n\in\mathbb{Z}}.$$

Then $(X, \mathbb{P}, \sigma)$ is a measure preserving dynamical system, which we call a Markov shift.

Definition 6.1.1 Let $\alpha \in \mathbb{T}^d$. We call the map $f\colon \Sigma^{\mathbb{Z}} \times \mathbb{T}^d \to \Sigma^{\mathbb{Z}} \times \mathbb{T}^d$ by

$$f(\omega, \theta) := (\sigma\omega, \theta + \alpha)$$

a *mixed Markov-quasiperiodic dynamical system.*

Note that f is the product between a Markov shift and a translation on the d-dimensional torus. Hence $(f, \mathbb{P} \times m)$ is a measure preserving dynamical system, where m is the Lebesgue measure on T^d. Moreover, if $\alpha \in \mathbb{T}^d$ is a rationally independent frequency, then f is ergodic (see [29, Theorem 6.1]).

A measurable function $A\colon \Sigma \times \Sigma \times \mathbb{T}^d \to \mathrm{GL}_m(\mathbb{R})$ induces the skew-product dynamical system $F = F_{(A,K)}\colon \Sigma^{\mathbb{Z}} \times \mathbb{T}^d \times \mathbb{R}^m \to \Sigma^{\mathbb{Z}} \times \mathbb{T}^d \times \mathbb{R}^m$,

$$F(\omega, \theta, v) = (\sigma\omega, \theta + \alpha, A(\omega_1, \omega_0, \theta)v)\,.$$

That is, F is a linear cocycle over the base dynamics $(\Sigma^{\mathbb{Z}} \times \mathbb{T}^d, \mathbb{P}_{(K,\mu)}, f)$, where the fiber dynamics is induced by the map A. We refer to such a dynamical system as a *mixed Markov-quasiperiodic cocycle.*

Its iterates are given by

$$F^n(\omega, \theta, v) = (\sigma^n\omega, \theta + n\alpha, A^n(\omega, \theta)v)$$

where for $\omega = \{\omega_n\}_{n\in\mathbb{Z}} \in X$,

$$A^n(\omega, \theta) = A(\omega_n, \omega_{n-1}, \theta + (n-1)\alpha) \cdots A(\omega_2, \omega_1, \theta + \alpha)\, A(\omega_1, \omega_0, \theta)\,.$$

By the Furstenberg-Kesten theorem (Theorem 2.3.1), the limit

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n(\omega, \theta)\|$$

exists for $\mathbb{P}_\mu \times m$ a.e. $(\omega, \theta) \in X \times \mathbb{T}^d$, where m is the Lebesgue measure. Since the base dynamics is ergodic with respect to $\mathbb{P}_\mu \times m$, the limit is a constant that depends on A and the kernel K and it is called the maximal Lyapunov exponent of the cocycle F, which we denote by $L_1(A, K)$.

We identify the cocycle $F_{A,K}$ with the pair (A, K) and denote the corresponding Lyapunov exponent by $L_1(A, K), \ldots, L_m(A, K)$.

6.2
Kifer non-random filtration

In this section, we present the statement and the proof of the Kifer Non-Random filtration in the case of Markov quasi-periodic cocycle.

First, we introduce some terminology.

Definition 6.2.1 We say that a Markov chain $\{X_n\}_{n\geq 0}$ is a *version* of another Markov chain $\{Y_n\}_{n\geq 0}$ over $\mathbb{P}_{x_0}$ when they have the same kernel K and the initial distribution is δ_{x_0}.

Now, we are now ready to statement and prove the to Its proof was based the Kifer non-random filtration in the context of Markov quasi-periodic cocycle. Its proof is based on [4, Lemma 2.6].

Theorem 6.2.1 *For $(\mu \times m)$-almost every $(\omega_0, \theta) \in \Sigma \times \mathbb{T}^d$, there exists A-invariant section $\mathcal{L}_1\colon \Sigma \times \mathbb{T}^d \to \mathrm{Gr}(\mathbb{R}^d)$ and two numbers β_0, β_1 such that*

$\mathcal{L}_1(\omega_0, \theta) \subsetneq \mathbb{R}^m$ *and for every* $v \in \mathbb{R}^m \setminus \mathcal{L}_1(\omega_0, \theta)$

$$\lim \frac{1}{n} \log \|A^n(\omega, \theta)v\| = \beta_0$$

while if $v \in \mathcal{L}_1(\omega_0)$

$$\lim \frac{1}{n} \log \|A^n(\omega, \theta)v\| \leq \beta_1 \,.$$

Moreover, the numbers β_0 *and* β_1 *are exactly the first and the second Lyapunov exponents.*

Proof. Let $\mathcal{C}$ be the set of quasi-periodic cocycles, that is,

$$\mathcal{C} := \{\tilde{C} \colon \mathbb{T}^d \times \mathbb{R}^d \to \mathbb{T}^d \times \mathbb{R}^d : \tilde{C}(\theta, v) = (\theta + \alpha, C(\theta)v), \text{ where } C(\theta) \in \mathrm{GL}_d(\mathbb{R})\} \,.$$

Let K be a kernel on the space of symbols Σ and a measurable map $A \colon \Sigma \times \Sigma \to \mathcal{C}$. Consider the map $J \colon \mathbb{T}^d \times \mathbb{R}^m \to \mathbb{T}^d \times \mathbb{R}^m$ such that $J(\theta, v) = (\theta, v)$ and define the map $A^n \colon \Sigma^Z \to \mathcal{C}$ such that $A^0(\omega) = J$ and for every $n \geq 1$

$$A^n(\omega) = A(\omega_n, \omega_{n-1}) \circ A(\omega_{n-1}, \omega_{n-2}) \circ \cdots \circ A(\omega_1, \omega_0).$$

Consider the Markov chain $(\omega_n, A^n(\omega))$ with initial distribution $\delta_{(\omega_0, J)}$ and let R be the kernel of this chain.

By Theorem 2.6.1, there exists a measure m on

$$\mathcal{F} = \{g \colon \Sigma \times \mathcal{C} \to \Sigma \times \mathcal{C} : g \text{ is a Borel map}\}$$

such that

$$R_{(x, \tilde{C})}(E) = m\{g \in \mathcal{F} : g(x, \tilde{C}) \in E\} \,.$$

Let $g_1, g_2, \ldots, g_n, \ldots \in \mathcal{F}$ be independent and identically distributed maps relative to m. For every $\omega \in \Sigma$ and $n \geq 1$ consider $f_n \colon \Sigma \to \Sigma$ and $J_n \colon \Sigma \to \mathcal{C}$ such that $g_n(\omega_0, J) := (f_n(\omega_0), J_n(\omega_0))$. Define $f^0(\omega_0) = \omega_0$, $J^0(\omega_0) = J$ and for every $n \geq 1$,

$$f^n = f_n \circ \cdots \circ f_1,$$

$$J^n(\omega) = J_n(f^{n-1}(\omega_0)) J_{n-1}(f^{n-2}(\omega_0)) \cdots J_2(f^1(\omega_0)) J_1(\omega_0) \,.$$

The sequence $\{(f^n(\omega_0), J^n(\omega_0)), n \in \mathbb{N}\}$ is a version of the Markov chain $(\omega_n, A^n(\omega))$. In fact, let $\mathcal{F}(n)$ be the σ-algebra generated by $\{(f^p(\omega_0), J^p(\omega_0)), p \leq n\}$. Take $y = f^n(\omega_0)$ and $M = J^n(\omega_0)$. If A is a borel subset of Σ and B is a borelian subset of $\mathcal{C}$, we have

$$\begin{aligned}\mathbb{E}((f^{n+1}(x), J^{n+1}(x)) \in A \times B | \mathcal{F}(n)) &= \mathbb{P}(f_{n+1}(y) \in A, J_{n+1}(y)M \in B) \\ &= \mathbb{P}(g_{n+1}(y, J) \in A \times BM^{-1}) \\ &= m\{g : g(y, J) \in A \times BM^{-1}\} \\ &= R_{(y,J)}(A \times BM^{-1}) \\ &= R_{(y,M)}(A \times B) \,.\end{aligned}$$

Define $F_n \colon \Sigma \times \mathbb{T}^d \times \mathbb{R}^d \to \Sigma \times \mathbb{T}^d \times \mathbb{R}^d$ such that

$$F_n(x, \theta, u) = (f_n(x), J_n(x)(\theta, u)).$$

Note that F_n are independent and identically distributed maps and given any initial $(\omega_0, \theta, v) \in \Sigma \times \mathbb{T}^d \times \mathbb{R}^d$, the Markov chain $F_n \circ \cdots \circ F_1(\omega_0, \theta, v)$ in $\Sigma \times \mathbb{T}^d \times \mathbb{R}^d$ is a version of the Markov chain $(\omega_n, A^n(\omega)(\theta, v))$ with the same initial distribution $\delta_{(\omega_0, \theta, v)}$.

We conclude the rest of the proof applying Theorem 2.6.3 to the Markov chain F_n. ■

6.3 Upper large deviations on the fiber

In this section we obtain an upper large deviations estimate for the fiber dynamics, and as a consequence of that, the upper semi continuity of the maximal Lyapunov exponent.

For this, we start with a simple topological lemma to be used later.

Lemma 6.3.1 *Let (M, d) be a metric space and let ν be a Borel probability measure on M. Given a closed set $L \subset M$ and $\varepsilon > 0$ there are an open sets $D \supset L$ such that $\nu(D) < \nu(L) + \varepsilon$ and a Lipschitz continuous function $g \colon M \to [0, 1]$ such that $\mathbb{1}_L \leq g \leq \mathbb{1}_D$.*

Proof. For every $\delta > 0$, let L_δ be the open δ-neighborhood of L, that is,

$$L_\delta := \{x \in M : d(x, L) < \delta\} \,.$$

Since L is a closed set we have that $L = \cap_{\delta > 0} L_\delta$. Then, $\nu(L_\delta) \to \nu(L)$ as $\delta \to 0$ and, consequently, there exists $\delta_0 = \delta_0(L, \varepsilon, \nu) > 0$ such that $\nu(L_{\delta_0}) < \nu(L) + \varepsilon$.

Consider $D := L_{\delta_0}$ and note that $d(L, D^C) = d(L, L_{\delta_0}^C) \geq \delta_0 > 0$. Define the function $g\colon M \to \mathbb{R}$ such that

$$g(x) := \frac{d(x, D^C)}{d(x, D^C) + d(x, L)} .$$

It is easy to check that g is a Lipschitz continuous map with $\|g\|_{\text{Lip}} \leq \frac{1}{\delta_0}$, while clearly $\mathbb{1}_L \leq g\mathbb{1}_D$. ■

Fix a number $M < \infty$. Let $\mathcal{C}_M = \{(\alpha, A) \in \mathcal{C} : \|A\|_0 \leq M\}$, and consider the mixed Markov-quasiperiodic dynamics f on $\Sigma^{\mathbb{Z}} \times \mathbb{T}^d$.

Theorem 6.3.1 *Let (K, μ) be a Markov system. Given any $\varepsilon > 0$, there are $\delta = \delta(\varepsilon, K, M) > 0, \bar{n} = \bar{n}(\varepsilon, K, M) \in \mathbb{N}$ and $c = c(\varepsilon, K, M) > 0$ such that for all kernel $K'\colon \Sigma \to \text{Prob}(\Sigma)$ with $d_{W_1}(K, K') < \delta$, for all $\theta \in \mathbb{T}^d$ and for all $n \geq \bar{n}$ we have*

$$\mathbb{P}_{K',\mu'} \left\{ \omega \in X : \frac{1}{n} \log \|A^n(\omega)(\theta)\| \geq L_1(K) + \varepsilon \right\} < e^{-cn} \tag{6.1}$$

where μ' is the stationary measure of K'. Moreover, the map $K' \to L_1(K')$ is upper semicontinuous with respect to the Wasserstein metric d_{W_1}.

A related result in the case of mixed random-quasiperiodic cocycle it may be found in [5].

Proof. Let

$$a_n(\omega, \theta) := \log \|A^n(\omega)(\theta)\|$$

and note that $\{a_n\}_{n\in\mathbb{N}}$ is an f-subadditive sequence, that is, for all $n, m \in \mathbb{N}$ and $(\omega, \theta) \in X \times \mathbb{T}^d$ we have

$$a_{n+m}(\omega, \theta) \leq a_n(\omega, \theta) + a_m(F^n(\omega, \theta)) .$$

For $(\omega, \theta) \in X \times \mathbb{T}^d$, let $n(\omega, \theta)$ be the first positive integer n such that

$$\frac{1}{n} a_n(\omega, \theta) < L_1(K) + \varepsilon . \tag{6.2}$$

For each $m \in \mathbb{N}$, define

$$U_m := \{(\omega, \theta) \in X \times \mathbb{T}^d : n(\omega, \theta) \leq m\}.$$

By Theorem , $n(\omega, \theta)$ is defined for $(\mathbb{P}_{(K,\mu)} \times m)$-almost every (ω, θ). Hence, U_m increases to a full $(\mathbb{P}_{(K,\mu)} \times m)$-measure set as $m \to \infty$. Then, there exists $N = N(\varepsilon, K)$ such that $(\mathbb{P}_{(K,\mu)} \times m)(X \setminus U_N) < \varepsilon$.

Let $C = C(M, K) := \sup\{\log \|A(\omega)(\theta) : \omega \in X, \theta \in \mathbb{T}^d\|\} < \infty$. Fix $(\omega, \theta) \in X \times \mathbb{T}^d$ and define the sequence of indices $\{n_k = n_k(\omega, \theta)\}_{k\geq 1}$ and points $\{(\omega_k, \theta_k)\}$ as follows:

$$(\omega_1, \theta_1) = (\omega, \theta), \quad \text{and} \quad n_1 = \begin{cases} n(\omega_1, \theta_1), & \text{if } (\omega_1, \theta_1) \in U_N \\ 1, & \text{if } (\omega_1, \theta_1) \notin U_N \end{cases}$$

And, for $k \geq 1$, define

$$(\omega_{k+1}, \theta_{k+1}) = f^{n_k}(\omega_k, \theta_k)$$

and

$$n_{k+1} = \begin{cases} n(\omega_{k+1}, \theta_{k+1}), & \text{if } (\omega_{k+1}, \theta_{k+1}) \in U_N \\ 1, & \text{if } (\omega_{k+1}, \theta_{k+1}) \notin U_N \end{cases}$$

that is, $(\omega_{k+1}, \theta_{k+1}) = f^{n_1+\cdots+n_k}(\omega, \theta)$.

Let $\bar{n} := \bar{n}(\varepsilon, L, M) := N \max\{\frac{C}{\varepsilon}, 1\}$, so $\bar{n} \geq N \geq n_1$. Fix any $n \geq \bar{n}$. Note that $1 \leq n_k \leq N$ for all $k \geq 1$. Hence, the sequence $a_k = \sum_{j=1}^{k} n_j$ is such that $a_k \nearrow \infty$ and, consequently, there exists $p \in \mathbb{N}$ such that

$$n_1 + \cdots + n_p \leq n \leq n_1 + \cdots + n_{p+1}$$

that is, there exists m such that $n = n_1 + \cdots + n_p + m$, where $0 \leq m < n_{p+1} \leq N$.

Using the subadditivity of the sequence $\{a_n\}_{n\geq 1}$ it follows that

$$\begin{aligned} a_n(\omega, \theta) \leq a_{n_1}(\omega, \theta) + a_{n_2}(f^{n_1}(\omega, \theta)) + \cdots + a_{n_p}(f^{n_1+\cdots+n_{p-1}}(\omega, \theta)) \\ + a_m(f^{n_1+\cdots+n_p}(\omega, \theta)) \,. \end{aligned}$$

From (6.2), we get

$$a_{n_1}(\omega, \theta) \leq n_1(L_1(L) + \varepsilon), \quad \text{if } (\omega, \theta) \in U_N$$

but, if $(\omega, \theta) \notin U_N$ then $n_1 = 1$ and $a_{n_1}(\omega, \theta) \leq C$. Hence,

$$a_{n_1}(\omega, \theta) \leq n_1(L_1(L) + \varepsilon) + C \cdot \mathbb{1}_{X \setminus U_N}(\omega, \theta) \,.$$

For the second term,

$$a_{n_2}(f^{n_1}(\omega, \theta)) \leq n_2(L_1(L) + \varepsilon), \quad \text{if } f^{n_1}(\omega, \theta) \in U_N$$

otherwise, $n_2 = 1$ and $a_{n_2}(\omega, \theta) \leq C$. Hence,

$$a_{n_2}(f^{n_1}(\omega, \theta)) \leq n_2(L_1(L) + \varepsilon) + C \cdot \mathbb{1}_{X \setminus U_N}(f^{n_1}(\omega, \theta)) \,.$$

Indictively, for $k \geq 1$,

$$a_{n_k}(f^{n_1+\cdots+n_{k-1}}(\omega,\theta)) \leq n_k(L_1(L)+\varepsilon) + C \cdot \mathbb{1}_{X\setminus U_N}(f^{n_1+\cdots+n_{k-1}}(\omega,\theta)) \,.$$

Using the subadditivity of the sequence $\{a_n\}$, it follows that

$$a_n(\omega,\theta) \leq (n_1+\cdots+n_p)(L_1(L)+\varepsilon) + C\sum_{j=0}^{n-1} \mathbb{1}_{X\setminus U_N}(f^j(\omega,\theta)) + CN \,.$$

Hence for all $(\omega,\theta) \in X \times \mathbb{T}^d$ and for all $n \geq \bar{n}$ we have

$$\frac{1}{n}\log \|A^n(\omega)(\theta)\| \leq L_1(L) + 2\varepsilon + C \cdot \frac{1}{n}\sum_{j=0}^{n-1} \mathbb{1}_{X\setminus U_N}(f^j(\omega,\theta)) \,.$$

Note that the closed set $U_m^{\mathrm{C}} \subset X \times \mathbb{T}^d$ is determined by the coordinates $\omega_0, \ldots, \omega_{N-1}$ and θ. By Lemma 6.3.1, there are an open set $D \supset U_m$ with $(\mathbb{P}_{K,\mu} \times m)(D) < 2\varepsilon$ and a Lipschitz continuous function $g\colon X \times \mathbb{T}^d \to [0,1]$ which depend only on the coordinates $\omega_0, \ldots, \omega_{N-1}, \theta$ such that $\mathbb{1}_{U_m^{\mathrm{C}}} \leq g \leq \mathbb{1}_D$.

Then, for all $(\omega,\theta) \in X \times \mathbb{T}^d$ and $n \geq \bar{n}$ we have

$$\frac{1}{n}\sum_{j=0}^{n-1} \mathbb{1}_{U_m^{\mathrm{C}}}(f^j(\omega,\theta)) < \frac{1}{n}\sum_{j=0}^{n-1} g(f^j(\omega,\theta)) \,.$$

Applying Theorem 4.2.1 to g, for any kernel $K'\colon \Sigma \to \mathrm{Prob}(\Sigma)$ that is sufficiently close to $K\colon \Sigma \to \mathrm{Prob}(\Sigma)$ in the Wasserstein distance d_{W_1} and for all $\theta \in \mathbb{T}^d$ we have:

$$\frac{1}{n}\sum_{j=0}^{n-1} g(f^j(\omega,\theta)) < \int g \, d(\mathbb{P}_{K,\mu} \times m) + \varepsilon$$

for ω outside a set of $\mathbb{P}_{K',\mu}$-measure $< e^{-cn}$, where $c = c(\varepsilon, K) > 0$.

Moreover,

$$\int g \, d(\mathbb{P}_{K,\mu} \times m) \leq \int \mathbb{1}_D \, d(\mathbb{P}_{K,\mu} \times m) = (\mathbb{P}_{K,\mu} \times m)(D) < 2\varepsilon$$

which combined with previous estimates proves (6.1).

Let K' be a kernel close enough to K in the Wasserstein metric d_{W_1}. Using the estimate (6.1) and integrating with respect to the measure $\mathbb{P}_{K',\mu} \times m$ for all large enough

$$\int \frac{1}{n}\log \|A^n(\omega,\theta)\| \, d(\mathbb{P}_{K',\mu} \times m) < L_1(K) + 2\varepsilon \,.$$

Letting $n \to \infty$, we conclude that $L_1(K') < L_1(K) + 2\varepsilon$. ■

7
Future works

A first future project concerns the study of fiber large deviations estimates (See definition 2.5.2) and of Hölder continuity properties of the Lyapunov exponents of mixed Markov-quasiperiodic cocycle as functions of the input data.

Using Kifer's non-random filtration (Theorem 2.6.3), we will establish an analogue in this context of the uniform convergence (Theorem 5.3.4) for Markov cocycles.

The Markov operator corresponding to mixed Markov-quasiperiodic cocycles will not be strongly mixing in the sense of Definition 5.4.2. By analogy with the mixed random-quasiperiodic cocycles studied by Cai, Duarte and Klein, the spectrum of this operator, when restricted to the space of Hölder continuous observables, will likely consist of a set $K \subset \bar{D}_\sigma(0), o \leq \sigma < 1$ plus the *entire* unit circle, which is a type of spectral gap, but not in the sense usually understood in the literature. The uniform convergence rate in Chapter 3 will play a vital role in dealing with the peripheral spectrum (which, again, is the entire unit circle).

A second future project concerns large deviations estimates for mixed Markov-quasiperiodic dynamical system with more general observables. The approach used to prove Theorem 4.2.1 will clearly not be applicable in this more general setting.

A third future project is the study of the stability of the Lyapunov exponents of quasiperiodic cocycles under Markovian noise.

Bibliography

[1] BARAVIERA, A.; DUARTE, P.. **Approximating Lyapunov exponents and stationary measures**. J. Dynam. Differential Equations, 31(1):25–48, 2019.

[2] BEZERRA, J.; POLETTI, M.. **Random product of quasi-periodic cocycles**. Proc. Amer. Math. Soc., 149(7):2927–2942, 2021.

[3] BOCKER-NETO, C.; VIANA, M.. **Continuity of Lyapunov exponents for random two-dimensional matrices**. Ergodic Theory Dynam. Systems, 37(5):1413–1442, 2017.

[4] BOUGEROL, P.. **Théorèmes limite pour les systèmes linéaires à coefficients markoviens**. Probab. Theory Related Fields, 78(2):193–221, 1988.

[5] CAI, A.; DUARTE, P. ; KLEIN, S.. **Mixed random-quasiperiodic cocycles**. preprint arXiv:2109.09544 [math.DS], 2021.

[6] CAI, A.; DUARTE, P. ; KLEIN, S.. **Furstenberg theory of mixed random-quasiperiodic cocycles**. preprint arXiv:2201.04745 [math.DS], 2021.

[7] CAI, A.; DUARTE, P. ; KLEIN, S.. **Statistical properties for mixing markov chains with applications to dynamical systems**. preprint arXiv:2210.16908 [math.DS], 2022.

[8] CAI, A.; DURAES, M.; KLEIN, S. ; MELO, A.. **Hölder continuity of the lyapunov exponent for markov cocycles via furstenberg's formula**. preprint arXiv:2212.00174 [math.DS], 2022.

[9] CHAZOTTES, J.-R.; GOUËZEL, S.. **Optimal concentration inequalities for dynamical systems**. Comm. Math. Phys., 316(3):843–889, 2012.

[10] DUARTE, P.; KLEIN, S.. **Lyapunov exponents of linear cocycles; continuity via large deviations**, volumen 3 de **Atlantis Studies in Dynamical Systems**. Atlantis Press, 2016.

[11] DUARTE, P.; KLEIN, S.. **Large deviations for products of random two dimensional matrices**. Comm. Math. Phys., 375(3):2191–2257, 2020.

[12] FURSTENBERG, H.. **Strict ergodicity and transformation of the torus**. Amer. J. Math., 83:573–601, 1961.

[13] FURSTENBERG, H.; KIFER, Y.. **Random matrix products and measures on projective spaces**. Israel J. Math., 46(1-2):12–32, 1983.

[14] GOLDSHEID, I.. **Exponential growth of products of non-stationary Markov-dependent matrices**. Int. Math. Res. Not. IMRN, (8):6310–6346, 2022.

[15] GORODETSKI, A.; KLEPTSYN, V.. **Non-stationary version of furstenberg theorem on random matrix products**, 2022.

[16] GRAFAKOS, L.. **Classical Fourier analysis**, volumen 249 de **Graduate Texts in Mathematics**. Springer, New York, third edition, 2014.

[17] HERMAN, M.-R.. **Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations**. Inst. Hautes Études Sci. Publ. Math., (49):5–233, 1979.

[18] KATZNELSON, Y.; WEISS, B.. **A simple proof of some ergodic theorems**. Israel J. Math., 42(4):291–296, 1982.

[19] KATZNELSON, Y.. **An introduction to harmonic analysis**. Cambridge Mathematical Library. Cambridge University Press, Cambridge, third edition, 2004.

[20] KIFER, Y.. **Ergodic theory of random transformations**, volumen 10 de **Progress in Probability and Statistics**. Birkhäuser Boston, Inc., Boston, MA, 1986.

[21] KLEIN, S.; LIU, X.-C. ; MELO, A.. **Uniform convergence rate for Birkhoff means of certain uniquely ergodic toral maps**. Ergodic Theory Dynam. Systems, 41(11):3363–3388, 2021.

[22] KURATOWSKI, K.. **Topology: Volume I**. Número v. 1. Elsevier Science, 2014.

[23] LANG, S.. **Introduction to Diophantine approximations**. Springer-Verlag, New York, second edition, 1995.

[24] LE PAGE, E.. **Régularité du plus grand exposant caractéristique des produits de matrices aléatoires indépendantes et applications**. Ann. Inst. H. Poincaré Probab. Statist., 25(2):109–142, 1989.

[25] MALHEIRO, E. C.; VIANA, M.. **Lyapunov exponents of linear cocycles over Markov shifts**. Stoch. Dyn., 15(3):1550020, 27, 2015.

[26] MARTIN, N.; ENGLAND, J.. **Mathematical Theory of Entropy**. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2011.

[27] PERES, Y.. **Analytic dependence of Lyapunov exponents on transition probabilities**. In: LYAPUNOV EXPONENTS (OBERWOLFACH, 1990), volumen 1486 de **Lecture Notes in Math.**, p. 64–80. Springer, Berlin, 1991.

[28] PERES, Y.. **Domains of analytic continuation for the top Lyapunov exponent**. Ann. Inst. H. Poincaré Probab. Statist., 28(1):131–148, 1992.

[29] PETERSEN, K.. **Ergodic theory**, volumen 2 de **Cambridge Studies in Advanced Mathematics**. Cambridge University Press, Cambridge, 1989. Corrected reprint of the 1983 original.

[30] QUEFFÉLEC, H.; QUEFFÉLEC, M.. **Diophantine approximation and Dirichlet series**, volumen 2 de **Harish-Chandra Research Institute Lecture Notes**. Hindustan Book Agency, New Delhi, 2013.

[31] SHADRIN, A.. **Lecture notes for the course on approximation theory at the University of Cambridge**. http://www.damtp.cam.ac.uk/user/na/PartIIIat/at.html.

[32] STEIN, E. M.; SHAKARCHI, R.. **Fourier analysis**, volumen 1 de **Princeton Lectures in Analysis**. Princeton University Press, Princeton, NJ, 2003.

[33] VIANA, M.; OLIVEIRA, K.. **Foundations of Ergodic Theory**. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2016.

[34] VILLANI, C.. **Optimal transport**, volumen 338 de **Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]**. Springer-Verlag, Berlin, 2009. Old and new.