Continuity of the Lyapunov Exponents of Linear Cocycles

Publicações Matemáticas

Continuity of the Lyapunov Exponents of Linear Cocycles

Pedro Duarte
Universidade de Lisboa

Silvius Klein
PUC-Rio

impa

31º Colóquio Brasileiro de Matemática

Estrada Dona Castorina, 110
22460-320 Rio de Janeiro, RJ

Impresso no Brasil / Printed in Brazil

Capa: Noni Geiger / Sérgio R. Vaz

31º Colóquio Brasileiro de Matemática

- Álgebra e Geometria no Cálculo de Estrutura Molecular - C. Lavor, N. Maculan, M. Souza e R. Alves
- **Continuity of the Lyapunov Exponents of Linear Cocycles - Pedro Duarte e Silvius Klein**
- Estimativas de Área, Raio e Curvatura para *H*-superfícies em Variedades Riemannianas de Dimensão Três - William H. Meeks III e Álvaro K. Ramos
- Introdução aos Escoamentos Compressíveis - José da Rocha Miranda Pontes, Norberto Mangiavacchi e Gustavo Rabello dos Anjos
- Introdução Matemática à Dinâmica de Fluídos Geofísicos - Breno Raphaldini, Carlos F.M. Raupp e Pedro Leite da Silva Dias
- Limit Cycles, Abelian Integral and Hilbert's Sixteenth Problem - Marco Uribe e Hossein Movasati
- Regularization by Noise in Ordinary and Partial Differential Equations - Christian Olivera
- Topological Methods in the Quest for Periodic Orbits - Joa Weber
- Uma Breve Introdução à Matemática da Mecânica Quântica - Artur O. Lopes

Distribuição:
IMPA
Estrada Dona Castorina, 110
22460-320 Rio de Janeiro, RJ
e-mail: ddic@impa.br
http://www.impa.br

ISBN: 978-85-244-0433-7

Contents

Preface

The purpose of this book is to present a *gentler introduction* to our method of studying continuity properties of Lyapunov exponents of linear cocycles. To that extent, we chose to illustrate this method in the *simplest setting* that is still both relevant in dynamical systems and applicable to mathematical physics problems.

In ergodic theory, a linear cocycle is a skew-product map acting on a vector bundle, which preserves the linear bundle structure and induces a measure preserving dynamical system on the base. The vector bundle is usually assumed to be trivial; the base dynamics is an ergodic measure preserving transformation on some probability space, while the fiber action is induced by a matrix-valued measurable function on the base. Lyapunov exponents quantify the average exponential growth of the iterates of the cocycle along invariant subspaces of the fibers, which are called Oseledets subspaces.

An important class of examples of linear cocycles are the ones associated to a discrete, one-dimensional, ergodic Schrödinger operator. Such an operator is the discretized version of a quantum Hamiltonian. Its potential is given by a time-series, that is, it is obtained by sampling an observable (called the potential function) along the orbit of an ergodic transformation.

The study of the continuity properties of the Lyapunov exponents as the input data (e.g. the fiber dynamics) is perturbed constitutes an active research topic in dynamical systems, both in Brazil and elsewhere.

A general research area in dynamical systems is the study of statistical properties like large deviations, for an observable sampled along the iterates of the system. This theory is well developed for rather general classes of base dynamical systems. However, when it comes to the dynamics induced by a linear cocycle on the projective space, this topic is much less understood. In fact, even when the base dynamics is a Bernoulli shift, this problem is not completely solved.

Both the continuity properties of the Lyapunov exponents and the statistical properties of the iterates of a linear cocycle are important tools in the study of the spectra of discrete Schrödinger operators in mathematical physics.

In our recent research monograph [16], we established a connection between these two research topics in dynamical systems. To wit, we proved that if a linear cocycle satisfies certain large deviation type (LDT) estimates, which are uniform in the data, then necessarily the corresponding Lyapunov exponents (LE) vary continuously with the data. Furthermore, this result is quantitative, in the sense that it provides a modulus of continuity which depends on the strength of the large deviations. We referred to this general result as the abstract continuity theorem (ACT). We then showed that such LDT estimates hold for certain types of linear cocycles over Markov shifts and over toral translations, thus ensuring the applicability of the general continuity result to these models.

The setting of the abstract continuity theorem (ACT) chosen for this book consists of $\mathrm{SL}_2(\mathbb{R})$-valued linear cocycles (i.e. linear cocycles with values in the group of two by two real matrices of determinant one).

The proof of the ACT consists of an inductive procedure that establishes continuity of relevant quantities for finite, larger and larger number of iterates of the system. This leads to continuity of the limit quantities, the Lyapunov exponents. The inductive procedure is based upon a deterministic result on the composition of a long chain of linear maps, called the *Avalanche Principle* (AP).

Furthermore, we establish uniform LDT estimates for $\mathrm{SL}_2(\mathbb{R})$-valued linear cocycles over a Bernoulli shift and over a one dimensional torus translation. The ACT is then applicable to these models.

In this setting, the formulation of the statements is significantly simplified and many arguments become less technical, while retaining most features present in the general setup.

While all results described in this book are consequences of their more general counterparts obtained elsewhere, in several instances, the formulation or proof presented here are new. For example: the formulation of the ACT in Section 3.2, the proof of the Hölder continuity of the Oseledets splitting in Section 3.4, the formulation of the LDT for quasi-periodic cocycles in Section 5.1, the proof of Sorets-Spencer theorem in Section 5.6, appear in print for the first time in this form.

One of the objectives of this book is to popularize these types of problems with the hope that the theory grows to become applicable to other types of systems, besides random and quasi-periodic cocycles.

The target audience we had in mind while writing this book was postgraduate students, as well as researchers with interests in this subject, but not necessarily experts in it. As such, we tried to make the presentation self contained modulo graduate textbooks on various topics.

The reader should be familiar with basic notions in ergodic theory, probabilities, Fourier analysis and functional analysis, usually provided by standard postgraduate courses on these subjects.

Two reference textbooks to keep handy are M. Viana and K. Oliveira [61] on ergodic theory and M. Viana [60] on Lyapunov exponents. They cover most of what one needs to know for the first three chapters of this book. Familiarity with Markov chains is helpful in understanding the approach used in the fourth chapter, and for that, D. Levin and Y. Peres [42, Chapter 1] suffices. Finally, the last chapter requires a nontrivial amount of complex and harmonic analysis tools, for which we recommend T. Gamelin [23] and C. Muscalu and W. Schlag [45, Chapters 1-3]. More precise references are provided within each chapter.

The book is organized as follows.

In Chapter 1 we review basic notions in ergodic theory and we introduce linear cocycles and Lyapunov exponents. We end the chapter with a discussion of some parallels between ergodic theorems and limit theorems in probabilities. These types of analogies will prove important all throughout this book.

In Chapter 2 we formulate the avalanche principle, describe the needed geometrical considerations and present its proof.

In Chapter 3 we describe our version of large deviations estimates then formulate and prove the abstract continuity theorem for the Lyapunov exponent and the Oseledets splitting.

In Chapter 4 we derive a uniform large deviation estimate for linear cocycles over the Bernoulli shift. The ACT is then applicable and it implies the continuity of the Lyapunov exponent for this model. We also present an adaptation of the original argument of Le Page for the continuity of the LE, without large deviations.

In Chapter 5 we derive a uniform large deviation estimate for linear cocycles over the one dimensional torus translation, assuming that the translation frequency satisfies some generic arithmetic assumptions and that the cocycles depend analytically on the base point. The ACT is then also applicable to this model.

In both Chapter 4 and Chapter 5 we describe the applicability of these results to Schrödinger cocycles.

All chapters end with bibliographical notes summarizing relevant related results.

Furthermore, all chapters contain *exercises*, which have two functions. The statements formulated in each exercise are needed in the arguments. Moreover, they are meant to help the reader practise her growing familiarity with the subject matter.

These notes, as well as the the idea of offering an advanced course in the 31st Colóquio Brasileiro de Matemática, grew out of our respective seminar presentations in Lisbon and Rio de Janeiro, during the last few months.

The first author would like to thank his colleagues in Lisbon, João Lopes Dias, José Pedro Gaivão and Telmo Peixe, for attending talks on this subject and for their suggestions.

The second author would like to thank students and postdocs at IMPA and PUC-Rio, including Jamerson Douglas Bezerra, Catalina Freijo, Xiaochuan Liu, Karina Marin, Mauricio Poletti, Adriana Sánchez and Elhadji Yaya Tall. His interaction with them lead to a simpler formulation of the ACT and provided the motivation for taking on the task of finding a less technical approach to our method.

The first author was partially supported by National Funding from FCT-Fundação para a Ciência e a Tecnologia, under the project: UID/ MAT/04561/2013.

Part of this work was performed while the second author was a CNPq senior postdoctoral fellow at IMPA, under the CNPq grant 110960/2016-5. He is grateful to the host institution and to the grant awarding institution for their support.

Special thanks are due to Teresa and Jaqueline for their help and patience during the writing of this book.

Chapter 1

Linear Cocycles

1.1 The definition and examples of ergodic systems

Given a probability space $(X, \mathcal{F}, \mu)$, a *measure preserving transformation* is an $\mathcal{F}$-measurable map $T : X \to X$ such that

$$\mu(T^{-1}(A)) = \mu(A), \quad \text{for all } A \in \mathcal{F}.$$

A *measure preserving dynamical system* (MPDS) is any triple (X, μ, T) where (X, μ) is a probability space (the σ-field $\mathcal{F}$ is implicit to X) and $T : X \to X$ is a measure preserving transformation.

We refer to elements of X as *phases*. The sequence of iterates $\{T^n x\}_{n \geq 0}$ is called the *orbit* of the phase x.

Definition 1.1. We say that the MPDS (X, μ, T) is *ergodic* if there is no T-invariant measurable set $A = T^{-1}(A)$ such that $0 < \mu(A) < 1$.

Definition 1.2. We say that the MPDS (X, μ, T) is *mixing* when for all $A, B \in \mathcal{F}$,

$$\lim_{n \to +\infty} \mu(A \cap T^{-n}(B)) = \mu(A)\, \mu(B)\ .$$

Mixing MPDS are always ergodic, but the converse is not true in general.

Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ be the one dimensional torus. When convenient, we identify the torus $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ (an additive group) with the unit circle $\mathbb{S}^1 \subset \mathbb{C}$ (a multiplicative group) via the map $x + \mathbb{Z} \mapsto e(x) := e^{2\pi i x}$, but we maintain the additive notation, e.g. we write $x + y \pmod 1$ instead of $e(x)e(y)$.

For $d \geq 1$, let $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$ be the *d-dimensional torus*. The normalized Haar measure denoted by $|\cdot|$ on the σ-field $\mathcal{F}$ of Borel sets determines a probability space $(\mathbb{T}^d, \mathcal{F}, |\cdot|)$.

We mention below a few classes of MPDS on the torus.

Example 1.1 (toral translations)**.** Given $\omega \in \mathbb{R}^d$, the translation map $T : \mathbb{T}^d \to \mathbb{T}^d$, $Tx := x + \omega \pmod 1$, preserves the Haar measure. This MPDS is ergodic if and only if the components of ω are rationally independent. Toral translations are never mixing.

Example 1.2 (toral endomorphisms)**.** Given a matrix $M \in \mathrm{GL}(d, \mathbb{Z})$, the *endomorphism* $T : \mathbb{T}^d \to \mathbb{T}^d$, $Tx := M\,x \pmod 1$, preserves the Haar measure. The endomorphism T is ergodic if and only if the spectrum of M does not contain any root of unity. Ergodic toral automorphisms are always mixing.

The composition of a toral endomorphisms with a translation is called an *affine endomorphism*. This provides another class of MPDS on the torus. See [62] for the characterization of the ergodic properties of affine endomorphisms.

Let Σ be a compact metric space and consider the space of sequences $X = \Sigma^{\mathbb{Z}}$. The (two-sided) *shift* is the homeomorphism $T \colon X \to X$ defined by $Tx := \{x_{n+1}\}_{n \in \mathbb{Z}}$ for $x = \{x_n\}_{n \in \mathbb{Z}}$. Denote by $\mathrm{Prob}(\Sigma)$ the space of Borel probability measures on Σ.

Example 1.3 (Bernoulli shifts)**.** Given $\mu \in \mathrm{Prob}(\Sigma)$, the shift map $T \colon X \to X$ preserves the product probability measure $\mu^{\mathbb{Z}}$. The MPDS $(X, \mu^{\mathbb{Z}}, T)$ is called a *Bernoulli shift*. Bernoulli shifts are ergodic and mixing.

A *stochastic matrix* is any square matrix $P = (p_{ij}) \in \mathrm{Mat}_m(\mathbb{R})$ such that $p_{ij} \geq 0$ for all $i, j = 1, \ldots, m$ and $\sum_{i=1}^m p_{ij} = 1$ for all $j = 1, \ldots, m$. A stochastic matrix P is called *primitive* if there exists $n \geq 1$ for which the power matrix $P^n = (p_{ij}^{(n)})$ has all entries strictly positive, i.e. $p_{ij}^{(n)} > 0$ for $i, j = 1, \ldots, m$.

A vector $q = (q_1, \ldots, q_m)$ with non-negative entries $q_j \geq 0$ such that $\sum_{j=1}^m q_j = 1$ is called a *probability vector*.

A probability vector q is said to be *P-stationary* if $P\,q = q$, i.e. $q_i = \sum_{j=1}^m p_{ij}\, q_j$ for all $i = 1, \ldots, m$.

Example 1.4 (Markov shifts of finite type)**.** Given a pair (P, q) consisting of a stochastic matrix $P \in \mathrm{Mat}_m(\mathbb{R})$ and a P-stationary probability vector q, consider the space of sequences $X = \Sigma^{\mathbb{Z}}$ over the finite alphabet $\Sigma = \{1, \ldots, m\}$ and the Σ-valued random process

$$\xi_n\{x_j\}_{j\in\mathbb{Z}} := x_n$$

defined over X.

Then there is a unique probability measure $\mathbb{P}_{P,q}$ over the Borel σ-algebra of X such that

(a) $\mathbb{P}_{P,q}[\,\xi_0 = i\,] = q_i$ for $i = 1, \ldots, m$,

(b) $\mathbb{P}_{P,q}[\,\xi_n = i \,|\, \xi_{n-1} = j\,] = p_{ij}$ for all $i, j = 1, \ldots, m$,.

The (two-sided) shift $T\colon X \to X$ preserves the measure $\mathbb{P}_{P,q}$ and the MPDS $(X, \mathbb{P}_{P,q}, T)$ is called a *Markov shift of finite type*.

The support of the measure $\mathbb{P}_{P,q}$ is the following subspace of admissible sequences

$$X(P) := \{\, \{x_j\}_{j\in\mathbb{Z}} \in X : p_{x_j x_{j-1}} > 0 \quad \text{for all } j \in \mathbb{Z}\}$$

known as a *subshift of finite type*.

The system $(X, \mathbb{P}_{P,q}, T)$ is mixing if and only if P is primitive.

1.2 The additive and subadditive ergodic theorems

Given a probability space (X, μ), we denote by $L^1(X, \mu)$ the space of measurable functions $\varphi : X \to \mathbb{R}$ that are absolutely integrable:

$$\mathbb{E}_\mu(|\varphi|) := \int_X |\varphi|\, d\mu < +\infty\ .$$

These functions will be called *observables*.

A simplified version of the Birkhoff (additive) ergodic theorem (BET) reads as follows.

Theorem 1.1. *Given an ergodic MPDS* (X, μ, T)*, for any observable* φ *and for* μ *almost every point* $x \in X$*,*

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

In other words, the additive ergodic theorem says that given an observable φ, if we denote by

$$S_n \varphi(x) := \sum_{j=0}^{n-1} \varphi(T^j x)$$

the corresponding Birkhoff sums, then a typical Birkhoff average $\frac{1}{n} S_n \varphi(x)$ converges to the space average of φ.

The subadditive ergodic theorem of Kingman generalizes Birkhoff's ergodic theorem. We formulate it below in a slightly simplified way.

Theorem 1.2. *Let* (X, μ, T) *be an ergodic MPDS. Given a sequence of measurable functions* $f_n \colon X \to \mathbb{R}$ *such that* $f_1 \in L^1(X, \mu)$ *and*

$$f_{n+m} \le f_n + f_m \circ T^n \quad \text{for all } \; n, m \ge 0 \; ,$$

the sequence $\{\int f_n \, d\mu\}_{n \ge 0}$ *is subadditive, i.e.,*

$$\int_X f_{n+m} \, d\mu \le \int_X f_n \, d\mu + \int_X f_m \, d\mu \quad \text{for all } \; n, m \ge 0 \; ,$$

and for μ*-a.e.* $x \in X$*, we have*

$$\lim_{n \to \infty} \frac{1}{n} f_n(x) = \lim_{n \to \infty} \frac{1}{n} \int_X f_n \, d\mu = \inf_{n \ge 1} \frac{1}{n} \int_X f_n \, d\mu < \infty \; .$$

The proofs of these fundamental theorems in ergodic theory can be found in most monographs on the subject (see for instance [61]). We would also like to mention the simple proofs of Y. Katznelson and B. Weiss [33] that use a stopping time argument which was later employed in other settings (e.g. [19, 32] and [16, Section 3.2]) as well.

1.3 Linear cocycles and the Lyapunov exponent

Let (X, μ, T) be an MPDS which throughout this book is assumed to be ergodic. A linear cocycle over (X, μ, T) is a skew-product map

$$F_A \colon X \times \mathbb{R}^2 \to X \times \mathbb{R}^2$$

given by

$$X \times \mathbb{R}^2 \ni (x, v) \mapsto (Tx, A(x)v) \in X \times \mathbb{R}^2 ,$$

where

$$A \colon X \to \mathrm{SL}_2(\mathbb{R})$$

is a measurable function.

Hence T is the base dynamics while A defines the fiber action. Since the base dynamics will be fixed, we may identify the cocycle with its fiber action A.

The forward iterates F_A^n of a linear cocycle F_A are given by $F_A^n(x, v) = (T^n x, A^{(n)}(x)v)$, where

$$A^{(n)}(x) := A(T^{n-1}x) \ldots A(Tx)\, A(x) \quad (n \in \mathbb{N}) .$$

Exercise 1.5. Show that if $g \in \mathrm{SL}_2(\mathbb{R})$, then $\|g\| \geq 1$ and $\|g^{-1}\| = \|g\|$. Recall that $\|\cdot\|$ refers to the operator norm of a matrix.

A cocycle A is said to be *μ-integrable* if

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

Note that since the matrix $A(x) \in \mathrm{SL}_2(\mathbb{R})$, its norm is ≥ 1.

Because norms behave sub-multiplicatively with matrix products, the sequence of functions

$$f_n(x) := \log\|A^{(n)}(x)\|$$

is subadditive.

Thus Kingman's ergodic theorem is applicable and we have the following.

Definition 1.3. Given a μ-integrable cocycle A, the μ-a.e. limit

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

exists and it is called the (maximal) *Lyapunov exponent* (LE) of A. Moreover,

$$L(A) = \lim_{n\to\infty} \int_X \frac{1}{n} \log\|A^{(n)}(x)\| d\mu(x) = \inf_{n\geq 1} \int_X \frac{1}{n} \log\|A^{(n)}(x)\| d\mu(x).$$

From the point of view of the base dynamics, two important classes of linear cocycles are the quasi-periodic and the random cocycles, which we define below.

Example 1.6. A *quasi-periodic cocycle* is any cocycle $A\colon \mathbb{T}^d \to \mathrm{SL}_2(\mathbb{R})$ over an ergodic torus translation $T : \mathbb{T}^d \to \mathbb{T}^d$.

If $Tx := x+\omega \pmod 1$ then $\omega \in \mathbb{R}^d$ is called the *frequency vector* of the cocycle.

Example 1.7. Let Σ be a compact metric space and let μ be a probability measure on Σ. Let $(X, \mu^{\mathbb{Z}}, T)$ be the Bernoulli shift, where $X = \Sigma^{\mathbb{Z}}$ is the space of sequences in Σ.

A function $\tilde{A} : X \to \mathrm{SL}_2(\mathbb{R})$ is called a *random Bernoulli cocycle* if $\tilde{A}$ depends only on the first coordinate x_0, that is, if

$$\tilde{A}\{x_n\}_{n\in\mathbb{Z}} = A(x_0)$$

for some measurable function $A\colon \Sigma \to \mathbb{R}$.

From the point of view of the fiber action, an important example of a linear cocycle is the Schrödinger cocycle, which appears in the study of the discrete, ergodic operators in mathematical physics. We briefly introduce these concepts (see [13] for more on this subject).

Example 1.8. Consider an invertible MPDS (X, μ, T) and a bounded observable $\varphi\colon X \to \mathbb{R}$. Let $x \in X$ be any phase. At every site n on the integer lattice $\mathbb{Z}$ we define the potential

$$v_n(x) := \varphi(T^n x)\,.$$

The discrete Schrödinger operator with potential $n \mapsto v_n(x)$ is the operator $H(x)$ defined on $l^2(\mathbb{Z})$ as follows.

If $\psi = \{\psi_n\}_{n\in\mathbb{Z}} \in l^2(\mathbb{Z})$, then

$$\big[H(x)\psi\big]_n := -(\psi_{n+1} + \psi_{n-1}) + v_n(x)\psi_n \quad \text{for all } n \in \mathbb{Z}\,.$$

Consider the Schrödinger (i.e. eigenvalue) equation

$$H(x)\,\psi = E\,\psi\,,$$

for some energy (i.e eigenvalue) $E \in \mathbb{R}$ and state (i.e. eigenvector) $\psi = \{\psi_n\}_{n\in\mathbb{Z}}$.

Define the associated *Schrödinger cocycle* as the cocycle (T, A_E), where

$$A_E(x) := \begin{bmatrix} \varphi(x) - E & -1 \\ 1 & 0 \end{bmatrix} \in \mathrm{SL}_2(\mathbb{R})\,.$$

Note that the Schrödinger equation above is a second order finite difference equation. An easy calculation shows that its formal solutions are given by

$$\begin{bmatrix} \psi_{n+1} \\ \psi_n \end{bmatrix} = A_E^{(n+1)}(x) \cdot \begin{bmatrix} \psi_0 \\ \psi_{-1} \end{bmatrix}\,,$$

where for all $n \in \mathbb{N}$, $A_E^{(n)}(x)$ is the n-th iterate of $A_E(x)$.

We will return to this example in each of the next chapters, showing how the results obtained are applicable to Schrödinger cocycles.

1.4 Some probabilistic considerations

Consider a scalar random process, i.e. a sequence $\xi_0, \xi_1, \ldots, \xi_{n-1}, \ldots$ of random variables with values in $\mathbb{R}$, and denote by

$$S_n := \sum_{j=0}^{n-1} \xi_j$$

the corresponding *additive* (sum) process.

The strong law of large numbers says that if the random variables defining the process are independent, identically distributed (i.i.d.)

and absolutely integrable, then the average process converges almost surely:

$$\frac{1}{n}\, S_n \to \mathbb{E}(\xi_0) = \int x \, d\mu(x) \quad \text{as } n \to \infty\,,$$

where $\mu \in \text{Prob}(\mathbb{R})$ is their common probability distribution.

Given an MPDS (X, μ, T), any observable $\varphi \colon X \to \mathbb{R}$ determines a sequence of real valued random variables

$$\xi_n := \varphi \circ T^n\,. \tag{1.1}$$

These random variables are identically distributed and absolutely integrable, but in general they are not independent.

Let us note that any i.i.d. sequence $\{\xi_n\}_n$ of random variables can be realized as the type of process given in (1.1), with (X, μ, T) being a Bernoulli shift and φ being an observable on the space of sequences X that depends only on the zeroth coordinate of the sequence.

Birkhoff's ergodic theorem says that even in the absence of independence, a very weak form thereof, the ergodicity of the system, ensures the convergence of the time averages in (1.1) to the space average. Thus Birkhoff's ergodic theorem can be seen as the generalization and the analogue in dynamical systems of the strong law of large numbers from probabilities.

Let us now consider a sequence $M_0, M_1, \ldots, M_{n-1}, \ldots$ of i.i.d. random variables with values in $\text{SL}_2(\mathbb{R})$. Denote by

$$\Pi^{(n)} := M_{n-1} \cdot \ldots \cdot M_1 \cdot M_0$$

the corresponding *multiplicative* (product) process.

The Furstenberg-Kesten theorem,[1] the analogue of the strong law of large numbers for multiplicative processes, says that the following geometric average of the process converges almost surely:

$$\frac{1}{n}\, \log \|\Pi^{(n)}\| \to L(\mu) \quad \text{as } n \to \infty,$$

where $\mu \in \text{Prob}(\text{SL}_2(\mathbb{R}))$ is the common probability distribution of the random variables. The a.s. limit $L(\mu)$ is called the (maximal) Lyapunov exponent of the process.

[1]The setting of Furstenberg-Kesten's theorem is actually a bit more general: instead of i.i.d., the sequence of random matrices is assumed metrically transitive and stationary (see [20]).

Furthermore, any absolutely integrable linear cocycle over an ergodic MPDS (X, μ, T), i.e. any matrix-valued observable $A \colon X \to \mathrm{SL}_2(\mathbb{R})$, determines the sequence of random matrices $M_n := A \circ T^n$. Note that the corresponding multiplicative process is exactly $A^{(n)}(x)$, the n-th iterate of the cocycle A. Moreover, the sequence $\{M_n\}$ is metrically transitive and stationary, but in general not independent. An independent multiplicative process can be realized as a random Bernoulli cocycle.

The Furstenberg-Kesten theorem, or the more general Kingman's ergodic theorem, are applicable and ensure the existence of the maximal Lyapunov exponent of this multiplicative process (or equivalently, of the linear cocycle).

These analogies with limit theorems in probabilities will be expanded and will prove important in the next chapters of this book.

1.5 The multiplicative ergodic theorem

Let $\mathrm{Gr}_1(\mathbb{R}^2)$ denote the Grassmannian of 1-dimensional linear subspaces (lines) $\ell \subset \mathbb{R}^2$. In the context of $\mathrm{SL}_2(\mathbb{R})$-valued cocycles, the Oseledets Multiplicative Ergodic Theorem (MET) for invertible ergodic transformations can be formulated as follows (see [60, Theorem 3.20] for the proof).

Theorem 1.3. *Let (X, μ, T) be an invertible, ergodic MPDS.*

Let $F_A \colon X \times \mathbb{R}^2 \to X \times \mathbb{R}^2$, $F_A(x, v) = (Tx, A(x)\, v)$, where $A \colon X \to \mathrm{SL}_2(\mathbb{R})$ is a μ-integrable linear cocycle with $L(A) > 0$.

There exists a measurable decomposition $\mathbb{R}^2 = \mathcal{E}^+(x) \oplus \mathcal{E}^-(x)$, with $\mathcal{E}^\pm \colon X \to \mathrm{Gr}_1(\mathbb{R}^2)$ measurable, such that for μ-almost every $x \in X$,

(a) $A(x)\, \mathcal{E}^\pm(x) = \mathcal{E}^\pm(Tx)$,

(b) $\displaystyle\lim_{n \to \pm\infty} \frac{1}{n} \log \| A^{(n)}(x)\, v \| = L(A)$, *for all* $v \neq 0$ *in* $\mathcal{E}^+(x)$,

(c) $\displaystyle\lim_{n \to \pm\infty} \frac{1}{n} \log \| A^{(n)}(x)\, v \| = -L(A)$, *for all* $v \neq 0$ *in* $\mathcal{E}^-(x)$,

(d) $\displaystyle\lim_{n \to \pm\infty} \frac{1}{n} \log \left| \sin \measuredangle (\mathcal{E}^+(T^n x), \mathcal{E}^-(T^n x)) \right| = 0.$

Definition 1.4. By the MET there exists a full measure T-invariant set of points such that statements (a)-(d) in the MET hold. The elements of this set are called Oseledets regular points.

1.6 Bibliographical notes

All the background in Ergodic Theory reviewed here (Birkhoff, Kigman and Oseledets theorems) can be found in the book of K. Oliveira and M. Viana [61]. The book of M. Viana [60] gives the reader a broad perspective on the on the specific topic of Lyapunov exponents.

Chapter 2

The Avalanche Principle

2.1 Introduction and statement

Given two sequences of positive real numbers M_n and N_n with geometric growth and a positive real number $\varepsilon > 0$, we will say that M_n and N_n are ε-*asymptotic*, and write $M_n \overset{\varepsilon}{\asymp} N_n$, if for all $n \geq 0$,

$$e^{-n\varepsilon} \leq \frac{M_n}{N_n} \leq e^{n\varepsilon}.$$

Let $\mathrm{GL}_d(\mathbb{R})$ denote the general linear group of real $d \times d$ matrices. Given $g_0, g_1, \ldots, g_n \in \mathrm{GL}_d(\mathbb{R})$, the relation

$$\|g_{n-1} \, \cdots \, g_1 \, g_0\| \overset{\varepsilon}{\asymp} \|g_{n-1}\| \, \cdots \, \|g_1\| \, \|g_0\|$$

can only hold if some highly non typical alignment between the matrices g_i occurs. In fact, typically one has

$$\|g_{n-1} \, \cdots \, g_1 \, g_0\| \ll e^{-n\,a} \, \|g_{n-1}\| \, \cdots \, \|g_1\| \, \|g_0\|$$

for some not so small $a > 0$. This motivates the following definition.

Definition 2.1. Given matrices $g_0, g_1, \ldots, g_{n-1} \in \mathrm{GL}_d(\mathbb{R})$, their expansion rift is the ratio

$$\rho(g_0, g_1, \ldots, g_{n-1}) := \frac{\|g_{n-1} \cdots g_1\, g_0\|}{\|g_{n-1}\| \cdots \|g_1\|\, \|g_0\|} \in (0,1].$$

The Avalanche Principle roughly says that under some general assumptions the expansion rift of a product of matrices behaves multiplicatively, in the sense that

$$\rho(g_0, g_1, \ldots, g_{n-1}) \overset{\delta}{\asymp} \rho(g_0, g_1) \cdots \rho(g_{n-2}, g_{n-1})$$

for some small positive number δ.

Before formulating it we need to recall some basic concepts and fix their notations.

Given $g \in \mathrm{GL}_d(\mathbb{R})$ let

$$s_1(g) \geq s_2(g) \geq \ldots \geq s_d(g) > 0$$

denote the sorted *singular values* of g. By definition these are the eigenvalues of the positive definite matrix $(g^* g)^{1/2}$. The first singular value $s_1(g)$ is the usual operator norm

$$s_1(g) = \max_{x \in \mathbb{R}^d \setminus \{0\}} \frac{\|gx\|}{\|x\|} =: \|g\|.$$

The last singular value of g is the least expansion factor of g, regarded as a linear transformation, and it can be characterized by

$$s_d(g) = \min_{x \in \mathbb{R}^d \setminus \{0\}} \frac{\|gx\|}{\|x\|} = \|g^{-1}\|^{-1}.$$

From the definition of the singular values it follows that

$$\left|\det g\right| = \prod_{j=1}^{d} s_j(g).$$

Definition 2.2. The *gap* (or the *singular gap*) of $g \in \mathrm{GL}_d(\mathbb{R})$ is the ratio between its first and second singular values,

$$\mathrm{gr}(g) := \frac{s_1(g)}{s_2(g)}.$$

Remark 2.1. If g is a matrix in $\mathrm{SL}_2(\mathbb{R})$, i.e., if $\det(g) = 1$, then $\mathrm{gr}(g) = \|g\|^2$.

Let $\mathbb{P}(\mathbb{R}^d)$ denote the *projective space*, consisting of all lines through the origin in the Euclidean space $\mathbb{R}^d$. Points in $\mathbb{P}(\mathbb{R}^d)$ are equivalence classes $\hat{x}$ of non-zero vectors $x \in \mathbb{R}^d$. We consider the projective distance $\delta\colon \mathbb{P}(\mathbb{R}^d) \times \mathbb{P}(\mathbb{R}^d) \to [0,1]$

$$\delta(\hat{x}, \hat{y}) := \frac{\|x \wedge y\|}{\|x\|\,\|y\|} = \sin\left(\angle(x,y)\right).$$

For readers not familiar with exterior products, we note that

$$\|x \wedge y\| = \|x\|\,\|y\|\,\sin\left(\angle(x,y)\right)$$

is nothing but the area of the parallelogram spanned by the vectors x and y.

We will denote by g^* the transpose of a matrix g. The eigenvectors of $g^* g$ are called *singular vectors* of g. Each singular vector of g is hence associated with a singular value of g (eigenvalue of $(g^* g)^{1/2}$).

Definition 2.3. Given $g \in \mathrm{GL}_d(\mathbb{R})$ such that $\mathrm{gr}(g) > 1$, the *most expanding direction* of g is the singular direction $\hat{\mathfrak{v}}(g) \in \mathbb{P}(\mathbb{R}^d)$ associated with the first singular value $s_1(g)$ of g. Let $\mathfrak{v}(g)$ be any of the two unit vector representatives of the projective point $\hat{\mathfrak{v}}(g)$. Finally, we set $\hat{\mathfrak{v}}^*(g) := \hat{\mathfrak{v}}(g^*)$ and $\mathfrak{v}^*(g) := \mathfrak{v}(g^*)$.

Any matrix $g \in \mathrm{GL}_d(\mathbb{R})$ maps the most expanding direction of g to the most expanding direction of g^*, multiplying vectors by the factor $s_1(g) = \|g\|$. In other words

$$g\,\mathfrak{v}(g) = \pm s_1(g)\,\mathfrak{v}^*(g).$$

The matrix g also induces a projective map $\hat{g}\colon \mathbb{P}(\mathbb{R}^d) \to \mathbb{P}(\mathbb{R}^d)$, $\hat{g}(\hat{x}) := \widehat{gx}$, for which one has

$$\hat{g}\,\hat{\mathfrak{v}}(g) = \hat{\mathfrak{v}}^*(g) \quad \text{and} \quad \widehat{g^*}\,\hat{\mathfrak{v}}^*(g) = \hat{\mathfrak{v}}(g). \tag{2.1}$$

We can now state the Avalanche Principle. See [16, Theorem 2.1]

Theorem 2.1. *There are universal constants* $c_i > 0$, $i = 0, 1, 2, 3$, *such that given* $0 < \kappa \leq c_0 \varepsilon^2$ *and* $g_0, g_1, \ldots, g_n \in \mathrm{GL}_d(\mathbb{R})$, *if*

(G) $\mathrm{gr}(g_j) \geq \kappa^{-1}$ *for* $j = 0, 1, \ldots, n-1$,

(A) $\frac{\|g_j\, g_{j-1}\|}{\|g_j\|\,\|g_{j-1}\|} \geq \varepsilon$ *for* $j = 1, \ldots, n-1$,

then, writing $g^n := g_{n-1}\,\cdots\,g_1\,g_0$,

(1) $\max\{\,\delta(\hat{\mathfrak{v}}(g^n), \hat{\mathfrak{v}}(g_0)),\ \delta(\hat{\mathfrak{v}}^*(g^n), \hat{\mathfrak{v}}^*(g_{n-1}))\,\} \leq c_2\,\kappa\,\varepsilon^{-1}$

(2) $\rho(g_0, g_1, \ldots, g_{n-1}) \overset{c_3\kappa/\varepsilon^2}{\asymp} \rho(g_0, g_1)\,\cdots\,\rho(g_{n-2}, g_{n-1})$.

Condition (G) will be referred to as the *gap assumption* because it imposes a lower bound on the gaps of the matrices g_j. Hypothesis (A) will be referred to as the *angle assumption*, a terminology to be explained later (see Remark 2.2).

Conclusion (1) of the AP says that the most expanding direction of the product matrix g^n is nearly aligned with the corresponding most expanding direction of the first matrix g_0. In other words

$$\delta(\,\hat{\mathfrak{v}}(g^n),\,\hat{\mathfrak{v}}(g_0)\,) \lesssim \kappa\,\varepsilon^{-1} \tag{2.2}$$

It also states a similar alignment between the images of most expanding directions of g^n and g_{n-1}.

Conclusion (2) of the AP is equivalent to

$$\frac{\|g_{n-1}\,\cdots\,g_1\,g_0\|\,\|g_{n-2}\|\,\cdots\,\|g_1\|}{\|g_1\,g_0\|\,\cdots\,\|g_{n-2}\,g_{n-1}\|} \overset{c_3\kappa/\varepsilon^2}{\asymp} 1$$

which taking logarithms reads as

$$\Big|\log\|g_{n-1}\,\cdots\,g_1\,g_0\| + \sum_{j=1}^{n-2}\log\|g_j\| - \sum_{j=1}^{n-1}\log\|g_j\,g_{j-1}\|\,\Big| \leq c_3\,\frac{\kappa}{\varepsilon^2}\,n.$$

Finally, dividing by n one gets

$$\frac{1}{n}\log\|g^n\| = -\frac{1}{n}\sum_{j=1}^{n-2}\log\|g_j\| + \frac{1}{n}\sum_{j=1}^{n-1}\log\|g_j\, g_{j-1}\| + \mathcal{O}(\frac{\kappa}{\varepsilon^2}) \quad (2.3)$$

In Chapter 3, formula (2.3) plays a key role in the inductive proof of the continuity of the LE and the Oseledets decomposition.

2.2 Staging the proof

The projective distance $\delta\colon \mathbb{P}(\mathbb{R}^d)\times\mathbb{P}(\mathbb{R}^d)\to[0,1]$ determines a complementary angle function $\alpha\colon \mathbb{P}(\mathbb{R}^d)\times\mathbb{P}(\mathbb{R}^d)\to[0,1]$ defined by

$$\alpha(\hat{x},\hat{y}) := \frac{|x\cdot y|}{\|x\|\,\|y\|} = \cos\left(\angle(x,y)\right).$$

The complementarity of the functions δ and α is expressed by (see 2.1)

$$\alpha(\hat{x},\hat{y})^2 + \delta(\hat{x},\hat{y})^2 = 1.$$

The following exotic operation will be used to express an upper bound on the expansion rift of two matrices. Consider the algebraic operation

$$a\oplus b := a + b - a\,b$$

on the set $[0,1]$. The transformation $\Phi : ([0,1],\oplus) \to ([0,1],\cdot)$, $\Phi(x) := 1-x$, is a semigroup isomorphism.

Proposition 2.1. *For any $a,b,c\in[0,1]$,*

(1) $0\oplus a = a$,

(2) $1\oplus a = 1$,

(3) $a\oplus b = (1-b)\,a + b = (1-a)\,b + a$,

(4) $a\oplus b < 1 \;\Leftrightarrow\; a<1$ *and* $b<1$,

(5) $a\le b \;\Rightarrow\; a\oplus c \le b\oplus c$,

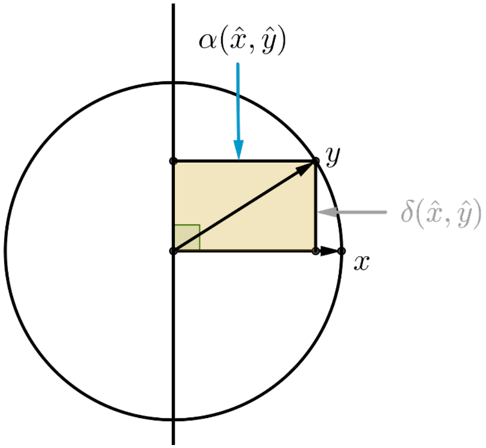

Figure 2.1: Angles α and δ

(6) $b > 0 \;\Rightarrow\; (ab^{-1} \oplus c)\, b \leq a \oplus c$,

(7) $a\, c + b\sqrt{1-a^2}\,\sqrt{1-c^2} \leq \sqrt{a^2 \oplus b^2}$.

Proof. Items (1)-(5) are left as exercises. Item (6) holds because

$$(ab^{-1} \oplus c)\, b = (ab^{-1} + c - cab^{-1})\, b = a + c\, b - ca \leq a + c - ca = a \oplus c.$$

For the last item consider the linear function $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) := a\,x + b\sqrt{1-a^2}\, y$ and the circle quarter $\Gamma = \{(c, \sqrt{1-c^2}) \colon c \in [0,1]\}$. The Lagrange multiplier method shows that $\max_{(x,y)\in\Gamma} f(x,y) = \sqrt{a^2 \oplus b^2}$, the extreme being attained at the point $(c, \sqrt{1-c^2})$ with $c = a/\sqrt{a \oplus b}$. This proves (7). □

Lemma 2.2. *Given $g \in \mathrm{GL}_d(\mathbb{R})$ with $\mathrm{gr}(g) > 1$, $\hat{x} \in \mathbb{P}(\mathbb{R}^d)$ and a unit vector $x \in \hat{x}$, writing $\alpha = \alpha(\hat{x}, \hat{\mathfrak{v}}(g))$ we have*

(a) $\alpha \leq \frac{\|gx\|}{\|g\|} \leq \sqrt{\alpha^2 \oplus \mathrm{gr}(g)^{-2}}$,

(b) $\delta(\hat{g}\,\hat{x}, \hat{\mathfrak{v}}^*(g)) \leq \alpha^{-1}\mathrm{gr}(g)^{-1}\,\delta(\hat{x}, \hat{\mathfrak{v}}(g))$,

(c) *The map* $\hat g\colon \mathbb{P}(\mathbb{R}^d) \to \mathbb{P}(\mathbb{R}^d)$ *has Lipschitz constant* $\lesssim \frac{r+\sqrt{1-r^2}}{\mathrm{gr}(g)\,(1-r^2)}$ *over the disk* $\{\hat x \in \mathbb{P}(\mathbb{R}^d)\colon \delta(\hat x, \hat{\mathfrak{v}}(g)) \le r\}$.

Proof. Let us denote $\sigma = \mathrm{gr}(g)$. Choose the unit vector $v = \mathfrak{v}(g)$ so that $\angle(v,x)$ is non obtuse. Then $x = \alpha\, v + u$ with $u \perp v$ and $\|u\| = \sqrt{1-\alpha^2}$. Letting $v^* = \mathfrak{v}^*(g)$, we have $gx = \alpha\,\|g\|\,v^* + gu$ with $gu \perp v^*$ and $\|gu\| \le \sqrt{1-\alpha^2}\, s_2(g) = \sqrt{1-\alpha^2}\,\|g\|/\sigma$.

We define the number $0 \le \kappa \le \sigma^{-1}$ so that $\|gu\| = \sqrt{1-\alpha^2}\,\kappa\,\|g\|$. Hence

$$\alpha^2\,\|g\|^2 \le \alpha^2\,\|g\|^2 + \|gu\|^2 = \|gx\|^2\ ,$$

and also

$$\begin{aligned}\|gx\|^2 &= \alpha^2\,\|g\|^2 + \|gu\|^2 = \|g\|^2\,\left(\alpha^2 + (1-\alpha^2)\kappa^2\right)\\ &= \|g\|^2\,\left(\alpha^2 \oplus \kappa^2\right) \le \|g\|^2\,\left(\alpha^2 \oplus \sigma^{-2}\right)\ ,\end{aligned}$$

which proves (a).

Using (a), item (b) follows from

$$\begin{aligned}\delta(\hat g\,\hat x, \hat{\mathfrak{v}}^*(g)) &= \frac{\|g\,v \wedge gx\|}{\|gv\|\,\|gx\|} = \frac{\|g\,v \wedge gx\|}{\|g\|\,\|gx\|} = \frac{\|v^* \wedge gx\|}{\|gx\|}\\ &= \frac{\|gu\|}{\|gx\|} \le \frac{\sqrt{1-\alpha^2}\,\|g\|}{\sigma\,\|gx\|} \le \frac{\delta(\hat x, \hat{\mathfrak{v}}(g))}{\alpha\,\sigma}\ .\end{aligned}$$

With the notation introduced in Exercise 2.1, we have the following formula for the derivative of the projective map $\hat g\colon \mathbb{P}(\mathbb{R}^d) \to \mathbb{P}(\mathbb{R}^d)$ (see Exercicise 2.2),

$$(D\hat g)_{\hat x}\, v = \frac{g\,v - \left(\frac{g\,x}{\|g\,x\|}\cdot g\,v\right)\frac{g\,x}{\|g\,x\|}}{\|g\,x\|} = \frac{1}{\|g\,x\|}\,\pi^{\perp}_{gx/\|gx\|}(g\,v).$$

To prove (c), take unit vectors $v = \mathfrak{v}(g)$ and $v^* = \mathfrak{v}^*(g)$ such that $g\,v = \|g\|\,v^*$. Because v is the most expanding direction of g we have

$$\|\pi^{\perp}_{v^*} \circ g\| = \|g \circ \pi^{\perp}_{v}\| \le s_2(g) = \sigma^{-1}\|g\|.$$

Given $\hat x$ such that $\delta(\hat x, \hat{\mathfrak{v}}(g)) \le r$, and a unit vector $x \in \hat x$, by (a)

$$\frac{\|g\|}{\|gx\|} \le \frac{1}{\alpha(\hat x, \hat{\mathfrak{v}}(g))} \le \frac{1}{\sqrt{1-r^2}}\ . \tag{2.4}$$

Using (b) we get

$$\delta(\hat{g}\hat{x}, \hat{\mathfrak{v}}^*(g)) \leq \frac{\delta(\hat{x}, \hat{\mathfrak{v}}(g))}{\alpha(\hat{x}, \hat{\mathfrak{v}}(g))\,\mathrm{gr}(g)} \leq \frac{r}{\sigma\sqrt{1-r^2}}.$$

Hence

$$(D\hat{g})_x\, v = \frac{1}{\|gx\|}\,\pi^{\perp}_{v^*}(g\,v) + \frac{1}{\|gx\|}\left(\pi^{\perp}_{gx/\|gx\|} - \pi^{\perp}_{v^*}\right)(g\,v)\,.$$

Thus, by (2.4) and Exercise 2.1 we have

$$\begin{aligned}\|(D\hat{g})_x\| &\leq \frac{\|g\|}{\sigma\,\|gx\|} + \frac{\delta(\hat{g}\hat{x}, \hat{\mathfrak{v}}^*(g))\,\|g\|}{\|gx\|}\\ &\leq \frac{1}{\sigma\sqrt{1-r^2}} + \frac{r}{\sigma\,(1-r^2)} = \frac{r+\sqrt{1-r^2}}{\sigma\,(1-r^2)}\,.\end{aligned}$$

Let $d(\hat{u},\hat{v})$ denote the Riemannian distance (arclength) on $\mathbb{P}(\mathbb{R}^d)$. Since $d(\hat{u},\hat{v}) = \arcsin(\delta(\hat{u},\hat{v}))$, the δ-ball $B(\hat{\mathfrak{v}}, r) := \{\hat{x}\colon \delta(\hat{x},\hat{\mathfrak{v}}(g)) \leq r\} = \{\hat{x}\colon d(\hat{x},\hat{\mathfrak{v}}(g)) \leq \arcsin r\}$ is a convex Riemannian disk. By the Mean Value Theorem, the map $\hat{g}|_{B(\hat{\mathfrak{v}},r)}$ has Lipschitz constant $\leq \frac{r+\sqrt{1-r^2}}{\sigma\,(1-r^2)}$ with respect to the Riemannian distance d. Since $\delta \leq d \leq \frac{\pi}{2}\,\delta$, the map $\hat{g}|_{B(\hat{\mathfrak{v}},r)}$ has also Lipschitz constant $\leq \frac{\pi}{2}\,\frac{r+\sqrt{1-r^2}}{\sigma\,(1-r^2)}$ with respect to δ. $\square$

Exercise 2.1. Given a unit vector $v \in \mathbb{R}^d$, $\|v\| = 1$, denote by $\pi_v, \pi_v^{\perp}\colon \mathbb{R}^d \to \mathbb{R}^d$ the orthogonal projections $\pi_v(x) := (v\cdot x)\,v$, respectively $\pi_v^{\perp}(x) := x - (v\cdot x)\,v$. Prove that for all unit vectors $u, v \in \mathbb{R}^d$,

$$\|\pi_v^{\perp} - \pi_u^{\perp}\| = \|\pi_v - \pi_u\| = \delta(\hat{u},\hat{v}).$$

Hint: Let $\rho(x) := \|\pi_u(x) - \pi_v(x)\| = \|(x\cdot u)\,u - (x\cdot v)\,v\|$. Prove that $\max_{\|x\|=1}\rho(x) = \delta(\hat{u},\hat{v})$ and this maximum is attained along the plane spanned by u and v.

Exercise 2.2. Given $g \in \mathrm{GL}_d(\mathbb{R})$ and $\hat{x} \in \mathbb{P}(\mathbb{R}^d)$, $x \in \hat{x}$ a non-zero representative and $v \in x^{\perp} = T_{\hat{x}}\mathbb{P}(\mathbb{R}^d)$, prove that

$$(D\hat{g})_{\hat{x}}\, v = \frac{1}{\|g\,x\|}\,\pi^{\perp}_{gx/\|gx\|}(g\,v).$$

Next corollary is a reformulation of items (b) and (c) of Lemma 2.2, which is more suitable for application.

Corollary 2.3. *Given* $g \in \mathrm{GL}_d(\mathbb{R})$ *such that* $\mathrm{gr}(g) \geq \kappa^{-1}$, *define*

$$\Sigma_\varepsilon := \{\hat{x} \in \mathbb{P}(\mathbb{R}^d)\colon\ \alpha(\hat{x}, \hat{\mathfrak{v}}(g)) \geq \varepsilon\,\} = B\left(\hat{\mathfrak{v}}(g), \sqrt{1-\varepsilon^2}\right).$$

Given a point $\hat{x} \in \Sigma_\varepsilon$,

(a) $\delta(\hat{g}\,\hat{x}, \hat{g}\,\hat{\mathfrak{v}}(g)) \leq \frac{\kappa}{\varepsilon}\,\delta(\hat{x}, \hat{\mathfrak{v}}(g))$,

(b) The map $\hat{g}|_{\Sigma_\varepsilon}\colon \Sigma_\varepsilon \to \mathbb{P}(\mathbb{R}^d)$ *has Lipschitz constant* $\lesssim \frac{\kappa}{\varepsilon^2}$.

Definition 2.4. Given $g, g' \in \mathrm{GL}_d(\mathbb{R})$ with $\mathrm{gr}(g) > 1$ and $\mathrm{gr}(g') > 1$ we define their lower angle as

$$\alpha(g, g') := \alpha(\hat{\mathfrak{v}}^*(g), \hat{\mathfrak{v}}(g')).$$

The upper angle between g and g' is

$$\beta(g, g') := \sqrt{\mathrm{gr}(g)^{-2} \oplus \alpha(g, g')^2 \oplus \mathrm{gr}(g')^{-2}}.$$

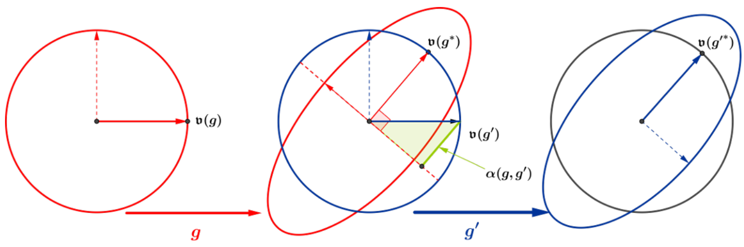

Figure 2.2: The (lower) angle between two matrices

Lemma 2.4. *Given* $g, g' \in \mathrm{GL}_d(\mathbb{R})$ *if* $\mathrm{gr}(g) > 1$ *and* $\mathrm{gr}(g') > 1$ *then*

$$\alpha(g, g') \leq \frac{\|g'\,g\|}{\|g'\|\,\|g\|} \leq \beta(g, g').$$

Proof. Let $\alpha := \alpha(g,g') = \alpha(\hat{\mathfrak{v}}^*(g), \hat{\mathfrak{v}}(g'))$ and take unit vectors $v = \mathfrak{v}(g)$, $v^* = \mathfrak{v}^*(g)$ and $v' = \mathfrak{v}(g')$ such that $v^* \cdot v' = \alpha > 0$ and $g\, v = \|g\|\, v^*$.

Since $\hat{g}\, \hat{\mathfrak{v}}(g) = \hat{\mathfrak{v}}^*(g)$, $w = \frac{g\, v}{\|g\, v\|}$ is a unit representative of $\hat{w} = \hat{\mathfrak{v}}^*(g)$. Hence, applying Lemma 2.2 (a) to g' and $\hat{w}$, we get

$$\alpha(g,g')\, \|g'\| = \alpha(\hat{w}, \hat{\mathfrak{v}}(g'))\, \|g'\| \leq \|\frac{g' g v}{\|g v\|}\| \leq \frac{\|g' g\|}{\|g\|}\ ,$$

which proves the first inequality.

For the second inequality, consider a unit vector $w \in \mathbb{R}^d$, representative of a projective point $\hat{w} \in \mathbb{P}(\mathbb{R}^d)$, such that $a := w \cdot v = \alpha(\hat{w}, \hat{\mathfrak{v}}(g)) \geq 0$. Then $w = a\, v + \sqrt{1-a^2}\, u$, where u is a unit vector orthogonal to v. It follows that $g\, w = a\, \|g\|\, v^* + \sqrt{1-a^2}\, g\, u$ with $g\, u \perp v^*$, and $\|g\, u\| = \kappa\, \|g\|$ for some $0 \leq \kappa \leq \mathrm{gr}(g)^{-1}$. Therefore

$$\frac{\|g\, w\|^2}{\|g\|^2} = a^2 + (1-a^2)\, \kappa^2 = a^2 \oplus \kappa^2\ .$$

and

$$\frac{g\, w}{\|g\, w\|} = \frac{a}{\sqrt{a^2 \oplus \kappa^2}}\, v^* + \frac{\sqrt{1-a^2}}{\sqrt{a^2 \oplus \kappa^2}}\, \frac{g\, u}{\|g\|}\ .$$

The vector v' can be written as $v' = \alpha\, v^* + w'$ with $w' \perp v^*$ and $\|w'\| = \sqrt{1-\alpha^2}$. Set now $b := \alpha(\hat{g}\hat{w}, \hat{\mathfrak{v}}(g'))$. Then

$$\begin{aligned} b = \Big|\frac{g\, w}{\|g\, w\|} \cdot v'\Big| &\leq \frac{\alpha\, a}{\sqrt{a^2 \oplus \kappa^2}} + \frac{\sqrt{1-a^2}}{\sqrt{a^2 \oplus \kappa^2}}\, \frac{\big|g\, u \cdot v'\big|}{\|g\|} \\ &\leq \frac{\alpha\, a}{\sqrt{a^2 \oplus \kappa^2}} + \frac{\kappa\, \sqrt{1-a^2}}{\sqrt{a^2 \oplus \kappa^2}}\, \Big|\frac{g\, u}{\|g\, u\|} \cdot w'\Big| \\ &\leq \frac{\alpha\, a}{\sqrt{a^2 \oplus \kappa^2}} + \frac{\kappa\, \sqrt{1-a^2}}{\sqrt{a^2 \oplus \kappa^2}}\, \|w'\| \\ &= \frac{\alpha\, a}{\sqrt{a^2 \oplus \kappa^2}} + \frac{\kappa\, \sqrt{1-a^2}\, \sqrt{1-\alpha^2}}{\sqrt{a^2 \oplus \kappa^2}} \leq \frac{\sqrt{\alpha^2 \oplus \kappa^2}}{\sqrt{a^2 \oplus \kappa^2}}\ . \end{aligned}$$

We use (7) of Proposition 2.1 on the last inequality. Finally, by

Lemma 2.2 (a) applied to g' and the unit vector $gw/\|gw\|$,

$$\begin{aligned}\|g'\,g\,w\| &\leq \|g'\|\,\sqrt{b^2 \oplus \mathrm{gr}(g')^{-2}}\,\|g\,w\| \\ &\leq \|g'\|\,\|g\|\,\sqrt{b^2 \oplus \mathrm{gr}(g')^{-2}}\,\sqrt{a^2 \oplus \kappa^2} \\ &\leq \|g'\|\,\|g\|\,\sqrt{\kappa^2 \oplus \alpha^2 \oplus \mathrm{gr}(g')^{-2}} \leq \beta(g,g')\,\|g'\|\,\|g\|\ ,\end{aligned}$$

where on the two last inequalities use items (6) and (5) of Proposition 2.1. □

Remark 2.2. Assumption (A) of the AP is essentially equivalent to

$$\alpha(g_{j-1}, g_j) \geq \varepsilon, \qquad \text{for all } j = 1, \ldots, n-1,$$

and it will be referred to as the *angle* assumption of the AP. In fact, the above condition is slightly stronger than (A), which in turn implies that

$$\alpha(g_{j-1}, g_j) \geq \frac{\varepsilon}{\sqrt{1 + 2\,\frac{\kappa^2}{\varepsilon^2}}}, \qquad \text{for all } j = 1, \ldots, n-1,$$

Given matrices $g_0, g_1, \ldots, g_{n-1} \in \mathrm{GL}_d(\mathbb{R})$, for $1 \leq j \leq n$ we write

$$g^j := g_{j-1}\,\cdots\,g_1\,g_0.$$

Lemma 2.5. *If* $\mathrm{gr}(g_j) > 1$ *and* $\mathrm{gr}(g^j) > 1$, *for* $1 \leq j \leq n$ *, then*

$$\prod_{j=1}^{n-1} \alpha(g^j, g_j) \leq \frac{\|g_{n-1}\,\cdots\,g_1\,g_0\|}{\|g_{n-1}\|\,\cdots\,\|g_1\|\,\|g_0\|} \leq \prod_{j=1}^{n-1} \beta(g^j, g_j).$$

Proof. By definition $g^n = g_{n-1} \ldots g_1 g_0$, and by convention $g^0 = I$. Hence $\|g_{n-1} \ldots g_1 g_0\| = \prod_{i=0}^{n-1} \frac{\|g^{i+1}\|}{\|g^i\|}$. This implies that

$$\begin{aligned}\frac{\|g_{n-1} \ldots g_1 g_0\|}{\|g_{n-1}\| \ldots \|g_1\|} &= \left(\prod_{i=0}^{n-1} \frac{1}{\|g_i\|}\right) \left(\prod_{i=0}^{n-1} \frac{\|g^{i+1}\|}{\|g^i\|}\right) \\ &= \prod_{i=0}^{n-1} \frac{\|g_i\,g^i\|}{\|g_i\|\,\|g^i\|}\ .\end{aligned}$$

It is now enough to apply Lemma 2.4 to each factor. □

2.3 The proof of the avalanche principle

Let us now prove the AP. By the previous lemma

$$\prod_{j=1}^{n-1} \frac{\alpha(g^j, g_j)}{\beta(g_{j-1}, g_j)} \leq \frac{\rho(g_0, \ldots, g_{n-1})}{\prod_{j=1}^{n-1} \rho(g_{j-1}, g_j)} \leq \prod_{j=1}^{n-1} \frac{\beta(g^j, g_j)}{\alpha(g_{j-1}, g_j)} \tag{2.5}$$

The strategy for conclusion (2) is to prove that the factors

$$\frac{\alpha(g^j, g_j)}{\beta(g_{j-1}, g_j)} \quad \text{and} \quad \frac{\beta(g_{j-1}, g_j)}{\alpha(g^j, g_j)}$$

are all very close to 1, with logarithms of order $\kappa\,\varepsilon^{-2}$. From conclusion (1) of the AP, apllied to the sequence of matrices $g_0, g_1, \ldots, g_j$,

$$\max\left\{ \delta\left(\mathfrak{v}^*(g^j), \mathfrak{v}^*(g_{j-1})\right), \delta\left(\mathfrak{v}(g^j), \mathfrak{v}(g_0)\right) \right\} \leq \kappa\,\varepsilon^{-1}, \tag{2.6}$$

for all $j = 1, \ldots, n$. Before proving (1) let us finish the proof of (2). From (2.6) we get

$$\begin{aligned} \left|\log \frac{\alpha(g^j, g_j)}{\alpha(g_{j-1}, g_j)}\right| &\leq \frac{\left|\alpha(g^j, g_j) - \alpha(g_{j-1}, g_j)\right|}{\min\{\alpha(g^j, g_j), \alpha(g_{j-1}, g_j)\}} \\ &\lesssim \frac{\delta(\mathfrak{v}^*(g^j), \mathfrak{v}^*(g_{j-1}))}{\varepsilon} \lesssim \frac{\kappa}{\varepsilon^2}. \end{aligned}$$

From the definition of the upper angle β we also have

$$\left|\log \frac{\beta(g^j, g_j)}{\alpha(g^j, g_j)}\right| \leq \log \sqrt{1 + 2\,\frac{\kappa^2}{\varepsilon^2}} \leq \frac{\kappa^2}{\varepsilon^2} \ll \frac{\kappa}{\varepsilon^2}.$$

These relations imply the existence of a universal positive constant c_3 such that

$$\begin{aligned} \left|\log \frac{\alpha(g^j, g_j)}{\beta(g_{j-1}, g_j)}\right| &\leq \left|\log \frac{\alpha(g^j, g_j)}{\alpha(g_{j-1}, g_j)}\right| + \left|\log \frac{\alpha(g_{j-1}, g_j)}{\beta(g_{j-1}, g_j)}\right| \\ &\lesssim \frac{\kappa}{\varepsilon^2} + \frac{\kappa^2}{\varepsilon^2} \leq c_3\,\frac{\kappa}{\varepsilon^2}. \end{aligned}$$

and

$$\left|\log \frac{\beta(g_{j-1},g_j)}{\alpha(g^j,g_j)}\right| \le \left|\log \frac{\beta(g_{j-1},g_j)}{\alpha(g_{j-1},g_j)}\right| + \left|\log \frac{\alpha(g_{j-1},g_j)}{\alpha(g^j,g_j)}\right|$$
$$\lesssim \frac{\kappa^2}{\varepsilon^2} + \frac{\kappa}{\varepsilon^2} \le c_3\,\frac{\kappa}{\varepsilon^2}.$$

Hence, from (2.5) we infer that

$$e^{-c_3\,\kappa\,\varepsilon^{-2}\,n} \le \frac{\rho(g_0,g_1,\ldots,g_{n-1})}{\prod_{j=1}^{n-1}\rho(g_{j-1},g_j)} \le e^{c_3\,\kappa\,\varepsilon^{-2}\,n}$$

which proves conclusion (2) of the AP.

To finish, we prove (1), addressing first the inequality

$$\delta\left(\mathfrak{v}(g^n),\mathfrak{v}(g_0)\right) \le \kappa\,\varepsilon^{-1}. \tag{2.7}$$

Consider the circular sequence of projective maps defined by the matrices

$$g_0, g_1, \ldots, g_{n-1}, g_{n-1}^*, \ldots, g_1^*, g_0^*.$$

Writing $\hat{\mathfrak{v}}_i := \hat{\mathfrak{v}}(g_i)$ and $\hat{\mathfrak{v}}_i^* := \hat{\mathfrak{v}}^*(g_i)$, we look at the sequence

$$\hat{\mathfrak{v}}_0 \overset{\hat g_0}{\mapsto} \hat{\mathfrak{v}}_0^*,\ \hat{\mathfrak{v}}_1 \overset{\hat g_1}{\mapsto} \hat{\mathfrak{v}}_1^*,\ \cdots,\ \hat{\mathfrak{v}}_{n-1} \overset{\hat g_{n-1}}{\mapsto} \hat{\mathfrak{v}}_{n-1}^*,$$
$$\hat{\mathfrak{v}}_{n-1}^* \overset{\hat g_{n-1}^*}{\mapsto} \hat{\mathfrak{v}}_{n-1},\ \cdots,\ \hat{\mathfrak{v}}_1^* \overset{\hat g_1^*}{\mapsto} \hat{\mathfrak{v}}_1,\ \hat{\mathfrak{v}}_0^* \overset{\hat g_0^*}{\mapsto} \hat{\mathfrak{v}}_0$$

as a closed pseudo-orbit for the given circular sequence of maps. To simplify the notation we will write

$$g_n = g_{n-1}^*,\ \ldots,\ g_{2n-2} = g_1^*,\ g_{2n-1} = g_0^*,$$
$$\hat{\mathfrak{v}}_n = \hat{\mathfrak{v}}_{n-1}^*,\ \hat{\mathfrak{v}}_n^* = \hat{\mathfrak{v}}_{n-1},\ \ldots,\ \hat{\mathfrak{v}}_{2n-1} = \hat{\mathfrak{v}}_0^*,\ \hat{\mathfrak{v}}_{2n-1}^* = \hat{\mathfrak{v}}_0.$$

We use a shadowing argument (see Figure 2.3) to prove the existence of a contracting fixed point $\tilde{\mathfrak{v}} \in \mathbb{P}(\mathbb{R}^d)$, which is $\kappa\,\varepsilon^{-1}$-near $\mathfrak{v}_0$, of the projective map

$$\widehat{(g^n)^* g^n} = \hat g_{2n-1}\cdots\hat g_n \hat g_{n-1}\cdots\hat g_1 \hat g_0.$$

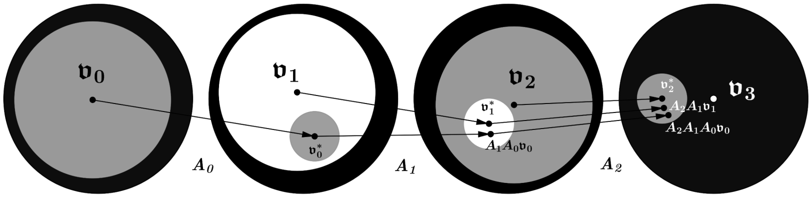

Figure 2.3: Shadowing property for contracting projective maps

Since $\hat{\mathfrak{v}}(g^n)$ is the unique contracting fixed point of this map, we must have $\tilde{\mathfrak{v}} = \hat{\mathfrak{v}}(g^n)$ and (2.7) follows.

For each $i = 0, 1, \ldots, 2n-1$ and $j = 0, 1, \ldots, 2n-i$, set

$$\hat{\mathfrak{v}}_i^j := \hat{g}_{i+j-1} \,\ldots\, \hat{g}_{i+1}\, \hat{g}_i\, \hat{\mathfrak{v}}_i, \tag{2.8}$$

so that, for each $0 \le i \le 2n-1$, the sequence of points

$$\mathfrak{v}_i = \hat{\mathfrak{v}}_i^0 \mapsto \hat{\mathfrak{v}}_i^* = \hat{\mathfrak{v}}_i^1 \mapsto \hat{\mathfrak{v}}_i^2 \mapsto \hat{\mathfrak{v}}_i^3 \mapsto \cdots \mapsto \hat{\mathfrak{v}}_i^{2n-i}$$

is a true orbit of the given chain of projective maps. By remark 2.2, instead of (A) we can assume that our sequence of matrices satisfies $\alpha(g_{j-1}, g_j) \ge \varepsilon$ for all $j = 0, 1, \ldots, n-1$. This implies that $\alpha(\hat{\mathfrak{v}}_{i-1}^*, \hat{\mathfrak{v}}_i) \ge \varepsilon$, or equivalently $\delta(\hat{\mathfrak{v}}_i, \hat{\mathfrak{v}}_{i-1}^*) \le \sqrt{1-\varepsilon^2}$, for all $i = 0, 1, \ldots, 2n-1$. By (a) of Corollary 2.3 we have

$$\delta(\hat{\mathfrak{v}}_i^1, \hat{\mathfrak{v}}_{i-1}^2) = \delta(\hat{g}_i\hat{\mathfrak{v}}_i, \hat{g}_i\hat{\mathfrak{v}}_{i-1}^*) \le \kappa\,\varepsilon^{-1} \qquad \text{for all } \ i = 0, 1, \ldots, 2n-1.$$

Applying item (b) of the same corollary inductively we get (see Figure 2.4)

$$\delta(\hat{\mathfrak{v}}_i^{j+1}, \hat{\mathfrak{v}}_{i-1}^{j+2}) = \delta((\hat{g}_{i+j}\ldots\hat{g}_i)\,\hat{\mathfrak{v}}_i, (\hat{g}_{i+j}\ldots\hat{g}_{i-1})\,\hat{\mathfrak{v}}_{i-1}^*) \le (\kappa\,\varepsilon^{-1})\,(\kappa\,\varepsilon^{-2})^j$$

for all $j = 0, 1, \ldots, 2n-i-1$. The details of the inductive verification of applicability of Corollary 2.3 are left to the reader (see

Exercise 2.3). Hence

$$\delta(\hat{g}^{2n}\hat{\mathfrak{v}}_0, \hat{\mathfrak{v}}_0) = \delta(\hat{\mathfrak{v}}_0^{2n}, \hat{\mathfrak{v}}_{2n-1}^1) \leq \sum_{i=1}^{2n-1} \delta(\hat{\mathfrak{v}}_i^{2n-i}, \hat{\hat{\mathfrak{v}}}_{i-1}^{2n-i+1})$$
$$\leq \kappa\,\varepsilon^{-1} \sum_{i=1}^{2n-1} (\kappa\,\varepsilon^{-2})^{2n-i-1} \leq \frac{\kappa\,\varepsilon^{-1}}{1-\kappa\,\varepsilon^{-2}} \lesssim \kappa\,\varepsilon^{-1}.$$

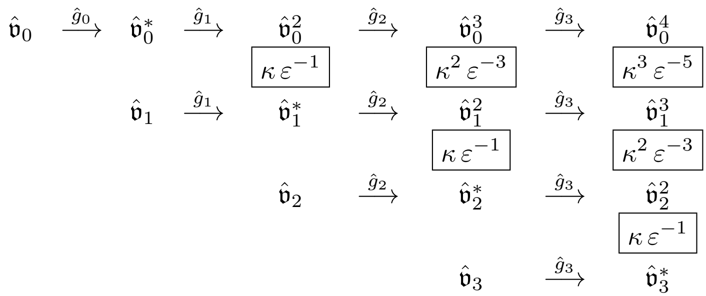

Figure 2.4: Orbits of the chain of projective maps $\hat{g}_0, \ldots, \hat{g}_{n-1}$

This proves that $\hat{g}^{2n}$ maps the ball B of radius $\sqrt{1-\varepsilon^2}$ around $\hat{\mathfrak{v}}_0$ into itself with contracting Lipschitz factor $\mathrm{Lip}(\hat{g}^{2n}|_B) \leq (\kappa\,\varepsilon^{-2})^{2n} \ll 1$. Thus, the (unique) fixed point $\tilde{\mathfrak{v}}$ of the map $\hat{g}^{2n}$ in the ball B is $\kappa\,\varepsilon^{-1}$ near to $\hat{\mathfrak{v}}_0$. As explained above, this proves that

$$\delta(\hat{\mathfrak{v}}(g^n), \hat{\mathfrak{v}}(g_0)) \lesssim \kappa\,\varepsilon^{-1}.$$

The second inequality in (1) reduces to (2.7) if the argument is applied to the sequence of transpose matrices $g_{n-1}^*, \ldots, g_1^*, g_0^*$.

Exercise 2.3. Consider the projective points $\hat{\mathfrak{v}}_i^j$ defined in (2.8) and prove that for all $i = 0, 1, \ldots, 2n-1$ and $j = 0, 1, \ldots, 2n-i-1$,

$$\delta(\hat{\mathfrak{v}}_i^{j+1}, \hat{\mathfrak{v}}_{i-1}^{j+2}) \leq (\kappa\,\varepsilon^{-1})\,(\kappa\,\varepsilon^{-2})^j.$$

Exercise 2.4. Given $g \in \mathrm{GL}_d(\mathbb{R})$ and $\hat{u}, \hat{v} \in \mathbb{P}(\mathbb{R}^d)$, prove that if $\hat{g}\,\hat{u} = \hat{v}$ then $\widehat{g^*}\,(\hat{v}^\perp) = \hat{u}^\perp$.

The following chain of exercises leads to an inequality (Exercise 2.9) which is needed in Chapter 3. The *relative distance* between $g, g' \in \mathrm{GL}_d(\mathbb{R})$ is defined by

$$d_{\mathrm{rel}}(g, g') := \frac{\|g - g'\|}{\max\{\|g\|, \|g'\|\}} \,.$$

Notice that this relative distance is not a metric.

Exercise 2.5. For all $p, q \in \mathbb{R}^d \setminus \{0\}$,

$$\| \frac{p}{\|p\|} - \frac{q}{\|q\|} \| \leq \max\left\{\|p\|^{-1}, \|q\|^{-1}\right\} \, \|p - q\| \,.$$

Exercise 2.6. For all $g_1, g_2 \in \mathrm{GL}_d(\mathbb{R})$ and any unit vector $p \in \mathbb{R}^d$,

$$\delta(\,\hat{g}_1\hat{p},\, \hat{g}_2\hat{p}\,) \leq \max\{\|g_1\, p\|^{-1}, \|g_2\, p\|^{-1}\} \, \|g_1 - g_2\| \,.$$

Exercise 2.7. Let (X, d) be a complete metric space, $T_1 \colon X \to X$ a Lipschitz contraction with $\mathrm{Lip}(T_1) < \kappa < 1$, $x_1^* = T_1(x_1^*)$ a fixed point, and $T_2 \colon X \to X$ any other map with a fixed point $x_2^* = T_2(x_2^*)$. Prove that

$$d(x_1^*, x_2^*) \leq \frac{1}{1 - \kappa}\, d(T_1, T_2) \,,$$

where $d(T_1, T_2) := \sup_{x \in X} d(T_1(x), T_2(x))$.

Consider now the set of normalized positive definite matrices

$$\mathcal{P} := \{g \in \mathrm{GL}_d(\mathbb{R}) \colon \|g\| = 1,\ g^* = g > 0\}$$

and the projection $P \colon \mathrm{GL}_d(\mathbb{R}) \to \mathcal{P}$, $P(g) := g^*\, g/\|g\|^2$.

Exercise 2.8. Show that for all $g, h \in \mathrm{GL}_d(\mathbb{R})$,

1. $\hat{\mathfrak{v}}(g) = \hat{\mathfrak{v}}(P(g))$,

2. $d_{\mathrm{rel}}(P(g), P(h)) \leq 4\, d_{\mathrm{rel}}(g, h)$.

Exercise 2.9. Given $g_1, g_2 \in \mathrm{GL}_d(\mathbb{R})$, if $\mathrm{gr}(g_1) \geq 10$, $\mathrm{gr}(g_2) \geq 10$ and $d_{\mathrm{rel}}(g_1, g_2) \leq \frac{1}{80}$ then

$$\delta\,(\,\hat{\mathfrak{v}}(g_1),\, \hat{\mathfrak{v}}(g_2)\,) \leq 12\, d_{\mathrm{rel}}(g_1, g_2)\ .$$

Hint: By Exercise 2.8 it is enough to show that given $h_1, h_2 \in \mathcal{P}$, if $\mathrm{gr}(h_1) \geq 100$, $\mathrm{gr}(h_2) \geq 100$ and $d_{\mathrm{rel}}(h_1, h_2) \leq \frac{1}{20}$ then

$$\delta\,(\,\hat{\mathfrak{v}}(h_1),\, \hat{\mathfrak{v}}(h_2)\,) \leq 3\, d_{\mathrm{rel}}(h_1, h_2)\ .$$

Let $\hat{p}_0 = \hat{\mathfrak{v}}(h_1)$, take $\delta = \frac{1}{5}$, consider the ball $B = B_\delta(\hat{p}_0)$ w.r.t. the metric δ, and establish the following facts:

1. $\hat{h}_1(B) \subseteq B$ (use item (b) of Lemma 2.2 (b))
2. $\|h_1 p\| \geq \frac{1}{2}$ for any unit vector p with $\hat{p} \in B$,
3. $\|h_2 p\| \geq \frac{1}{2}$ for any unit vector p with $\hat{p} \in B$,
4. $\delta(\hat{h}_1\, \hat{p}, \hat{h}_2\, \hat{p}) \leq 2\,\|h_1 - h_2\|$ for all $\hat{p} \in B$ (use Exercise 2.6),
5. The projective map $\hat{h}_1$ has Lipschitz constant $\leq \frac{1}{50}$ on B (use item (c) of Lemma 2.2),
6. $\hat{h}_2(B) \subseteq B$ (use the two previous items),
7. $\delta(\hat{\mathfrak{v}}(h_1), \hat{\mathfrak{v}}(h_2)) \leq 3\,\|h_1 - h_2\| = 3\, d_{\mathrm{rel}}(h_1, h_2)$ (use Exercise 2.7).

2.4 Bibliographical notes

The AP was introduced by M. Goldstein and W. Schlag [25, Proposition 2.2] as a technique to obtain Hölder continuity of the LE for quasi-periodic Schrödinger cocycles. In its original version, the AP applies to chains of unimodular matrices in $\mathrm{SL}_2(\mathbb{C})$, and the length of the chain is assumed to be less than some lower bound on the norms of the matrices. Note that for unimodular matrices, the gap ratio and the norm are two equivalent measurements. Still in this unimodular setting, for matrices in $\mathrm{SL}_2(\mathbb{R})$, J. Bourgain and S. Jitomirskaya [11, Lemma 5] relaxed the constraint on the length of the chain of matrices, and later J. Bourgain [10, Lemma 2.6] removed it, at the cost of slightly weakening the conclusion of the AP.

Later, W. Schlag [53, lemma 1] generalized the AP to invertible matrices in $\mathrm{GL}_d(\mathbb{C})$. Moreover, an earlier draft of [2] that C. Sadel has shared with the authors contained his version of the AP for $\mathrm{GL}_d(\mathbb{C})$ matrices. Both of these higher dimensional APs assume some bound on the length of the chains of matrices.

The version of the AP in these notes does not require this assumption and was established by the authors in [14, Theorem 3.1]. As a by-product of its more geometric approach conclusion (1) of Theorem 2.3 was added to the AP. This provides a quantitative control on the most expanding directions of the matrix product. In [16] a more general AP is described, one which holds for (possibly non-invertible) matrices in $\mathrm{Mat}_d(\mathbb{R})$.

Chapter 3

The Abstract Continuity Theorem

3.1 Large deviations type estimates

The ergodic theorems formulated in the previous chapter imply convergence in measure of the corresponding quantities. The main assumption of the continuity results in this chapter is that the averages corresponding to the fiber dynamics satisfy a precise, *quantitative* convergence in measure estimate. In order to describe these large deviations type (LDT) estimates, let us return to the analogy with limit theorems from classical probabilities.

Consider a sequence $\xi_0, \xi_1, \dots, \xi_{n-1}, \dots$ of random variables with values in $\mathbb{R}$, and let $S_n := \xi_0 + \xi_1 + \ldots + \xi_{n-1}$. If the process is independent, identically distributed and if its first moment is finite, then the average $\frac{1}{n} S_n$ converges almost surely to the mean $\mathbb{E}(\xi_0)$. In particular it also converges in measure:

$$\mathbb{P}\left[\left| \frac{1}{n} S_n - \mathbb{E}(\xi_0) \right| > \epsilon \right] \to 0 \quad \text{as } n \to \infty \,.$$

The event $\left| \frac{1}{n} S_n - \mathbb{E}(\xi_0) \right| > \epsilon$ is called a *tail event*. The asymptotic behavior of tail events forms the subject of the theory of large deviations (see [49]). A classical result in this theory is the following theorem due to Cramér.

Theorem 3.1. *Let* $\{\xi_n\}$ *be an i.i.d. random process with mean* $\mu = \mathbb{E}(\xi_0)$*. If the process has finite exponential moments, i.e. if the moment generating function* $M(t) := \mathbb{E}[e^{t\,X_0}] < \infty$ *for all* $t > 0$*, then*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[\, \big|\frac{1}{n}\, S_n - \mu\big| > \varepsilon \right] = -I(\varepsilon)$$

where $I(\varepsilon) := \sup_{t>0}(t\,\varepsilon - \log M(t) + t\,\mu)$ *is called the rate function.*

In other words, if n is large enough, the probability of the tail event is exponentially small:

$$\mathbb{P}\left[\, \big|\frac{1}{n}\, S_n - \mathbb{E}(\xi_0)\big| > \epsilon \right] \asymp e^{-I(\epsilon)\, n}\,,$$

for some rate function $I(\epsilon)$.

An analogue of Cramér's large deviations principle for multiplicative processes holds as well, and it was obtained by E. Le Page [38]. The result in [38] holds assuming certain conditions (strong irreducibility and contraction) on the support of the probability distribution of the process. We note that while these assumptions are generic, they do exclude interesting examples. Removing these assumptions has lately become the subject of intense work by several authors.

In classical probabilities, the theory of large deviations is part of a larger subject, that of *concentration inequalities*, which provide bounds on the deviation of a random variable from a constant, generally its expected value. Hoeffding's inequality, which we formulate below, is a standard example of a concentration inequality (see [58]).

Theorem 3.2. *Let* $\xi_0, \xi_1, \ldots, \xi_{n-1}$ *be an independent*[1] *random process with values in* $\mathbb{R}$ *and let* $S_n := \xi_0 + \xi_1 + \ldots + \xi_{n-1}$ *be its sum. If the process is almost surely bounded, i.e. if for some finite constant* C*,* $\big|\xi_i\big| \le C$ *a.s. for all* $i = 0, \ldots n-1$*, then*

$$\mathbb{P}\left[\, \Big|\frac{1}{n}\, S_n - \mathbb{E}\big(\frac{1}{n}\, S_n\big)\Big| > \epsilon \right] < 2\, e^{-\frac{1}{2C^2}\,\epsilon^2 n}\,. \tag{3.1}$$

[1]The process need not be identically distributed and it need not be infinite.

Note that compared to Cramér's large deviations principle, Hoeffding's inequality only provides an upper bound for the tail event (when the given random process is infinite). However, it has the advantage of being a *finite scale* rather than an asymptotic result. Moreover, it is a *quantitative* estimate that depends explicitly and *uniformly* on the data. More precisely, note that if we perturb the process slightly in the L^∞-norm, the a.s. bound C will not change much, so the bound (3.1) on the probability of deviation from the mean will not change much either.

There are analogues of such concentration inequalities for certain classes of *base* dynamical systems (see [12]). In this book we are concerned with such estimates for the *fiber* dynamics of linear cocycles.

Consider an MPDS (X, μ, T) and let $A\colon X \to \mathrm{SL}_2(\mathbb{R})$ be a μ-integrable linear cocycle over it. For every $n \in \mathbb{N}$, denote by

$$A^{(n)}(x) := A(T^{n-1}x) \, \dots \, A(Tx) \, A(x) \, ,$$

its n-th iterate and consider the geometric average

$$u_A^{(n)}(x) := \frac{1}{n} \, \log\|A^{(n)}(x)\| \, .$$

We denote the mean of this average by

$$L^{(n)}(A) := \int_X u_A^{(n)}(x) d\mu(x) = \int_X \frac{1}{n} \, \log\|A^{(n)}(x)\| d\mu(x) \, ,$$

and refer to it as a *finite scale Lyapunov exponent* of A. That is because as $n \to \infty$, the finite scale Lyapunov exponent $L^{(n)}(A)$ converges to $L(A)$, the (infinite scale) Lyapunov exponent of A.

We are now ready to introduce our concept of concentration inequality or large deviation type (LDT) estimate for a linear cocycle.

Definition 3.1. A cocycle $A\colon X \to \mathrm{SL}_2(\mathbb{R})$ satisfies an LDT estimate if there is a constant $c > 0$ and for every small enough $\epsilon > 0$ there is $\bar{n} = \bar{n}(\epsilon) \in \mathbb{N}$ such that for all $n \geq \bar{n}$,

$$\mu\Big\{ x \in X \colon \left| \frac{1}{n} \log\|A^{(n)}(x)\| - L^{(n)}(A) \right| > \epsilon \Big\} < e^{-c\epsilon^2 n} \, . \qquad (3.2)$$

Note that since $L^{(n)}(A) \to L(A)$, we may substitute in (3.2) the Lyapunov exponent $L(A)$ for the finite scale quantity $L^{(n)}(A)$.

3.2 The formulation of the abstract continuity theorem

In this chapter we establish a criterion for the continuity of the Lyapunov exponent and of the Oseledets splitting seen as functions of the cocycle (i.e. of the fiber dynamics). We refer to this continuity criterion as the abstract continuity theorem (ACT). This result is quantitative, in the sense that it provides a modulus of continuity.

Given an MPDS (X, μ, T), let $(\mathcal{C}, d)$ be a metric space of linear cocycles $A \colon X \to \mathrm{SL}_2(\mathbb{R})$ over this base dynamics.

The main assumption required by the method employed here is the availability of a uniform LDT estimate for each cocycle in this metric space. We say that a cocycle $A \in \mathcal{C}$ satisfies a *uniform* LDT if the constants[2] c and $\bar{n}$ in Definition 3.1 above are stable under small perturbations of A. We formulate this more precisely below.

Definition 3.2. A cocycle $A \in \mathcal{C}$ satisfies a uniform LDT if there are constants $\delta > 0$, $c > 0$ and for every small enough $\epsilon > 0$ there is $\bar{n} = \bar{n}(\epsilon) \in \mathbb{N}$ such that

$$\mu \left\{ x \in X \colon \left| \frac{1}{n} \log \| B^{(n)}(x) \| - L^{(n)}(B) \right| > \epsilon \right\} < e^{-c\epsilon^2 n} \tag{3.3}$$

for all cocycles $B \in \mathcal{C}$ with $d(B, A) < \delta$ and for all $n \geq \bar{n}$.

Remark 3.1. Note that at this point, it is not clear that we get an equivalent definition of the uniform LDT by substituting in (3.3) the limiting quantity $L(B)$ for the finite scale quantity $L^{(n)}(B)$. That is because while $L^{(n)}(B) \to L(B)$, the convergence is not a-priori known to be uniform in B. However, in the course of proving the abstract continuity theorem, we will also derive this uniform convergence. Thus a-posteriori, (3.3) will be equivalent with

$$\mu \left\{ x \in X \colon \left| \frac{1}{n} \log \| B^{(n)}(x) \| - L(B) \right| > \epsilon \right\} < e^{-c\epsilon^2 n}$$

for all B in the vicinity of A and all scales $n \geq \bar{n}$ (see Remark 3.2).

[2] We will refer to the constants c and $\bar{n}$ as the LDT parameters of A. They depend on A, and in general they may blow up as A is perturbed.

Let us denote by $\mathcal{C}^\star$ the set of cocycles $A \in \mathcal{C}$ with $L(A) > 0$. For any cocycle $A \in \mathcal{C}^\star$, we denote the subspaces (lines) of its Oseledets splitting in the MET 1.3 by $\mathcal{E}_A^\pm(x)$. Thus for almost every $x \in X$ we have the (T, A)-invariant splitting $\mathbb{R}^2 = \mathcal{E}_A^+(x) \oplus \mathcal{E}_A^-(x)$.

Because of our identification between a line in $\mathbb{R}^2$ and a point in the projective space $\mathbb{P}(\mathbb{R}^2)$, the components of the Oseledets decomposition are functions $\mathcal{E}_A^\pm \colon X \to \mathbb{P}(\mathbb{R}^2)$.

Let $L^1(X, \mathbb{P}(\mathbb{R}^2))$ be the space of all Borel measurable functions $\mathcal{E} \colon X \to \mathbb{P}(\mathbb{R}^2)$. On this space we consider the distance

$$d(\mathcal{E}_1, \mathcal{E}_2) := \int_X \delta(\mathcal{E}_1(x), \mathcal{E}_2(x))\, d\mu(x)\,,$$

where the quantity under the integral sign refers to the distance between points in the projective space $\mathbb{P}(\mathbb{R}^2)$.

We may now formulate the ACT.

Theorem 3.3. *Let (X, μ, T) be an MPDS and let $(\mathcal{C}, d)$ be a metric space of $\mathrm{SL}_2(\mathbb{R})$-valued cocycles over it. We assume the following:*

(i) $\|A\| \in L^\infty(X, \mu)$ *for all* $A \in \mathcal{C}$.

(ii) $d(A, B) \geq \|A - B\|_{L^\infty}$ *for all* $A, B \in \mathcal{C}$.

(iii) Every cocycle $A \in \mathcal{C}^\star$ *satisfies the uniform LDT* (3.3).

Then the following statements hold.

1a. The Lyapunov exponent $L \colon \mathcal{C} \to \mathbb{R}$ *is a continuous function. In particular,* $\mathcal{C}^\star$ *is an open set in* $(\mathcal{C}, d)$.

1b. On $\mathcal{C}^\star$*, the Lyapunov exponent is a locally Hölder continuous function.*

2a. The Oseledets splitting components $\mathcal{E}^\pm \colon \mathcal{C}^\star \to L^1(X, \mathbb{P}(\mathbb{R}^2))$, $A \mapsto \mathcal{E}_A^\pm$*, are locally Hölder continuous functions.*

2b. In particular, for any $A \in \mathcal{C}^\star$*, there are constants* $K < \infty$ *and* $\alpha > 0$ *such that if* B_1, B_2 *are in a small neighborhood of* A*, then*

$$\mu\left\{x \in X \colon \delta(\mathcal{E}_{B_1}^\pm(x),\, \mathcal{E}_{B_2}^\pm(x)) > d(B_1, B_2)^\alpha\right\} < K\, d(B_1, B_2)^\alpha\,.$$

In the next two chapters, under appropriate assumptions, we will establish the uniform LDT for random Bernoulli cocycles and respectively for quasi-periodic cocycles. Thus the ACT will be applicable to linear cocycles over these types of base dynamics, proving the continuity of the corresponding Lyapunov exponent and Oseledets splitting components.

Regarding the structure of the fiber dynamics, the ACT is applicable to the Schrödinger cocycles defined in Example 1.8. Indeed, let (X, μ, T) be an MPDS and let $\varphi \colon X \to \mathbb{R}$ be a bounded observable.

For every $E \in \mathbb{R}$, consider the Schrödinger cocycle

$$A_E(x) := \begin{bmatrix} \varphi(x) - E & -1 \\ 1 & 0 \end{bmatrix},$$

and let the one parameter family $\mathcal{C} := \{A_E \colon E \in \mathbb{R}\}$ be the corresponding space of cocycles, endowed with the distance:

$$d(A_{E_1}, A_{E_2}) := \left| E_1 - E_2 \right| = \|A_{E_1} - A_{E_2}\|_\infty .$$

Since the only quantity that varies is the parameter E, denote the Lyapunov exponent $L(A_E) =: L(E)$ and the Oseledets splitting components $\mathcal{E}^{\pm}_{A_E} =: \mathcal{E}^{\pm}_{E}$. With this setup we have the following.

Corollary 3.1. *Assume that for all parameters E we have $L(E) > 0$ and the cocycle A_E satisfies the uniform LDT* (3.3) *with parameters given by some absolute constants. Then the Lyapunov exponent $L(E)$ and the Oseledets splitting components $\mathcal{E}^{\pm}_{E}$ are Hölder continuous functions of E.*

In the next two chapters we will apply this result to random and respectively to quasi-periodic cocycles. Furthermore, for each model we will describe a criterion for the positivity of the Lyapunov exponent.

The proof of the ACT for the Lyapunov exponent uses an inductive procedure in the number of iterates of the cocycle.[3]

[3]The continuity of the components of the Oseledets splitting will be established in a more direct manner. However, the argument requires as an input the continuity and other related properties of the LE.

1. In Proposition 3.2 we show that given any $n \in \mathbb{N}$, the finite scale Lyapunov exponent $L^{(n)}(A)$ is a Lipschitz continuous function of the cocycle.

   This is easy to see, as $L^{(n)}(A)$ is obtained by performing a finite number of operations and then an integration. However, the Lipschitz constant depends on the number n of iterations, hence this argument cannot be taken to the limit.

2. In Proposition 3.3 we establish the main technical ingredient of the proof, the inductive step procedure, which can be described as follows.

   If the finite scale LE $L^{(n)}(A)$, at a scale $n = n_0$, does not vary much as the cocycle A is slightly perturbed, then the same will hold, save for a small, explicit error, at a larger scale $n = n_1$.

   The argument is based on the avalanche principle, whose applicability is ensured by the LDT estimates. Moreover, because of the exponential decay in the LDT estimates, the scale n_1 can be taken exponentially large in n_0.

3. The inductive step procedure will imply Proposition 3.4, which establishes a *uniform* (in cocycle) rate of convergence of the finite scale Lyapunov exponent $L^{(n)}$ to the (infinite scale) Lyapunov exponent L.

4. This uniform rate of convergence will ensure that some of the regularity of the finite scale LE at an initial scale will be carried over to the limit, thus establishing the theorem.

Let us comment further on this last step, in order to help the reader anticipate the direction of the argument.

If a sequence of continuous functions on a metric space converges uniformly, then the limit is itself a continuous function. The content of the following exercise is a quantitative statement in the same sprit, establishing a modulus of continuity for the limit function.

Exercise 3.1. Let (M, d) and (N, d) be two metric spaces, let $V \subset M$ be a subset (say a ball) and let $f_n \colon M \to N$, $n \geq 1$, be a sequence of functions. Assume the following:

(i) The sequence $\{f_n\}_n$ convergence uniformly on V to a function f at an exponential rate, i.e. for some $c > 0$ we have

$$d(f_n(a), f(a)) \le e^{-c\,n} \quad \text{for all } a \in V \quad \text{and for all } n \ge 1\,.$$

(ii) There is $C > 0$ such that for all $a, b \in V$ and for all $n \ge 1$,

$$\text{if} \quad d(a,b) < e^{-C\,n} \quad \text{then} \quad d(f_n(a), f_n(b)) \le e^{-c\,n}\,.$$

Then for all $x, y \in V$ we have

$$d(f(x), f(y)) \le 3\,e^c\,d(x,y)^{\frac{c}{C}}\,,$$

that is, f is Hölder continuous on V with Hölder exponent $\alpha = \frac{c}{C}$.

It is clear that the statement of this exercise can be tweaked (or it will be clear, after solving the exercise) to derive some modulus of continuity for the limit function if the rate of convergence was slower.

3.3 Continuity of the Lyapunov exponent

We are in the setting and under the assumptions of the abstract continuity theorem 3.3. Various context-universal constants (i.e. constants depending only on the given data) will appear throughout this section. In order to ease the presentation and not have to keep track of all such constants, given $a, b \in \mathbb{R}$ we will write $a \lesssim b$ if $a \le C\,b$ for some context-universal constant $0 < C < \infty$. Moreover, for $n \in \mathbb{N}$ and $x \in \mathbb{R}$, the notation $n \asymp x$ means $|n - x| \le 1$.

Exercise 3.2. For any $A \in \mathcal{C}$ show that the following bounds hold for a.e. $x \in X$ and for all $n \in \mathbb{N}$:

$$0 \le \log\|A^{(n)}(x)\| \le n\,\log\|A\|_{L^\infty}\,.$$

Conclude that if $B \in \mathcal{C}$ with $d(B, A) \le 1$, then for a.e. $x \in X$ and for all $n \in \mathbb{N}$:

$$0 \le \frac{1}{n}\,\log\|B^{(n)}(x)\| \le C\,,$$

where $C = C(A) := \log\,(1 + \|A\|_{L^\infty})$.

Proposition 3.2 (finite scale continuity). *Let $A \in \mathcal{C}$. There is a constant $C = C(A) < \infty$ such that for all cocycles $B_1, B_2 \in \mathcal{C}$ with $d(B_i, A) \le 1$, $i = 1, 2$, for all iterates $n \ge 1$ and for a.e. phase $x \in X$,*

$$\left| \frac{1}{n} \log\|B_1^{(n)}(x)\| - \frac{1}{n} \log\|B_2^{(n)}(x)\| \right| \le e^{C\,n}\, d(B_1, B_2)\,. \tag{3.4}$$

In particular,

$$\left| L^{(n)}(B_1) - L^{(n)}(B_2) \right| \le e^{C\,n}\, d(B_1, B_2)\,. \tag{3.5}$$

Proof. Let $B \in \mathcal{C}$ be any cocycle with $d(B, A) \le 1$. By Exercise 3.2, $1 \le \|B^{(n)}(x)\| \le e^{C\,n}$ for all $n \in \mathbb{N}$ and for a.e. $x \in X$.

Applying the mean value theorem to the function log and using the above inequalities, for a.e. $x \in X$ we have:

$$\begin{aligned}
&\left| \frac{1}{n} \log\|B_1^{(n)}(x)\| - \frac{1}{n} \log\|B_2^{(n)}(x)\| \right| \\
&\quad\le \frac{1}{n} \frac{1}{\min\{\|B_1^{(n)}(x)\|, \|B_2^{(n)}(x)\|\}} \left| \|B_1^{(n)}(x)\| - \|B_2^{(n)}(x)\| \right| \\
&\quad\le \frac{1}{n} \|B_1^{(n)}(x) - B_2^{(n)}(x)\| \\
&\quad\le \frac{1}{n} \sum_{i=0}^{n-1} e^{C(n-1)} \|B_1(T^i(x) - B_2(T^i(x))\| \\
&\quad\le \frac{1}{n} e^{C\,n} \sum_{i=0}^{n-1} \|B_1 - B_2\|_{L^\infty} \le e^{C\,n}\, d(B_1, B_2)\,.
\end{aligned}$$

This proves (3.4). Integrating in x proves (3.5). □

The above proposition shows that the finite scale LE functions $L^{(n)} \colon \mathcal{C} \to \mathbb{R}$ are continuous (in fact, Lipschitz continuous, but with Lipschitz constant depending exponentially on n). Since, as a consequence of Kingman's subadditive ergodic theorem (see Definition 1.3), for all $A \in \mathcal{C}$, $L(A) = \inf_{n \ge 1} L^{(n)}(A)$, we may conclude, according to the exercise below, that L is an *upper semi-continuous* function.

This is a general fact about the Lyapunov exponent. However, its lower semi-continuity (and hence continuity) requires further assumptions on the space of cocycles.

Exercise 3.3. Let (M, d) be a metric space and let $f_n \colon M \to \mathbb{R}$, $n \geq 1$ be a sequence of upper semi-continuous functions. Consider $f(a) := \inf_{n \geq 1} f_n(a)$ the pointwise infimum of these functions.

Prove that f is upper semi-continuous. Find examples showing that f need not be continuous.

Exercise 3.4. Let (M, d) be a metric space, let f be an upper semi-continuous function on it and let $a \in M$. Prove that if $f(a) = 0$ and $f \geq 0$ in a neighborhood of a, then f is continuous at a.

Exercise 3.5. Let $A \in \mathcal{C}$ be a cocycle and let $n_0 < n_1$ be two integers. If $n_1 = n \cdot n_0 + q$, where $0 \leq q < n_0$, then for a.e. $x \in X$ we have

$$\left| \frac{1}{n_1} \log \| A^{(n_1)}(x) \| - \frac{1}{n \, n_0} \log \| A^{(n \, n_0)}(x) \| \right| \leq C \frac{q}{n_1} \leq C \frac{n_0}{n_1} \,,$$

where $C = C(A)$ is the constant in Exercise 3.2.

Proposition 3.3 (inductive step procedure)**.** *Let $A \in \mathcal{C}^\star$ and let $c, \bar{n}$ denote its (uniform) LDT parameters.*

Fix $\epsilon := \frac{L(A)}{100} > 0$ and denote $c_1 := \frac{c}{2} \, \epsilon^2$.

There are constants $C = C(A) < \infty$, $\delta = \delta(A) > 0$, $\overline{n_0} = \overline{n_0}(A) \in \mathbb{N}$, such that for any $n_0 \geq \overline{n_0}$, if the inequalities

$$\text{(a)} \quad L^{(n_0)}(B) - L^{(2n_0)}(B) < \eta_0 \tag{3.6}$$

$$\text{(b)} \quad \left| L^{(n_0)}(B) - L^{(n_0)}(A) \right| < \theta_0 \tag{3.7}$$

hold for a cocycle $B \in \mathcal{C}$ with $d(B, A) < \delta$, and if the positive numbers η_0, θ_0 satisfy

$$\theta_0 + 2\eta_0 < L(A) - 6\epsilon \,, \tag{3.8}$$

then for an integer n_1 such that

$$n_1 \asymp e^{c_1 \, n_0} \,, \tag{3.9}$$

we have:

$$\left| \; L^{(n_1)}(B) + L^{(n_0)}(B) - 2L^{(2n_0)}(B) \; \right| < C \frac{n_0}{n_1} \,. \tag{3.10}$$

Furthermore,

$$\text{(a++)} \quad L^{(n_1)}(B) - L^{(2n_1)}(B) < \eta_1 \tag{3.11}$$

$$\text{(b++)} \quad \left| L^{(n_1)}(B) - L^{(n_1)}(A) \right| < \theta_1 \,, \tag{3.12}$$

where

$$\theta_1 = \theta_0 + 4\eta_0 + C\,\frac{n_0}{n_1}\,, \tag{3.13}$$

$$\eta_1 = C\,\frac{n_0}{n_1}\,. \tag{3.14}$$

Proof. By Exercise 3.2, there is a finite constant $C = C(A)$ such that if $B \in \mathcal{C}$ with $d(B, A) \leq 1$, then

$$\left\| \frac{1}{n} \log \|B^{(n)}\| \right\|_{L^\infty} \leq C\,. \tag{3.15}$$

In particular, $L^{(n)}(B) \leq C$ and $L(B) \leq C$.

Let δ be the size of the ball around $A \in \mathcal{C}$ where the uniform LDT with parameters $c, \bar{n}$ holds. Fix $B \in \mathcal{C}$ with $d(B, A) < \delta$.

The integer $\overline{n_0}$ is chosen large enough to ensure that various estimates are applicable at scales $n \geq \overline{n_0}$. That is:

- $\overline{n_0} \geq \bar{n}$, so that the uniform LDT for A applies if $n \geq \overline{n_0}$;
- $\left| L^{(n)}(A) - L(A) \right| < \epsilon$ for $n \geq \overline{n_0}$, which is ensured by the fact that $L^{(n)}(A) \to L(A)$ as $n \to \infty$;
- Various concrete asymptotic inequalities, like $n^2 \ll e^{c/2\,\epsilon^2 n}$, hold for $n \geq \overline{n_0}$.

Note tat $\overline{n_0}$ depends only on A (since ϵ was fixed).

Fix the scales n_0 and n_1 such that $n_0 \geq \overline{n_0}$ and $n_1 \asymp e^{c_1\, n_0}$.

We may assume that $n_1 = n\, n_0$ for some $n \in \mathbb{N}$. Otherwise, by Exercise 3.5, our estimates will accrue an extra error of order $\frac{n_0}{n_1}$, which is compatible with the conclusions of this proposition.

The goal is to use the Avalanche Principle, more precisely (2.3), to relate the block of length n_1 (i.e. the product of n_1 matrices) $B^{(n_1)}(x)$ to blocks of length n_0 for sufficiently many phases x; averaging in x will then establish (3.10), from which everything else follows.

Let us then define, for every $0 \leq i \leq n-1$,

$$g_i = g_i(x) := B^{(n_0)}(T^{i\, n_0}\, x)\,.$$

Then clearly

$$g^{(n)} = B^{(n_1)}(x)$$

and for all $1 \leq i \leq n-1$,

$$\begin{aligned} g_i\, g_{i-1} &= B^{(n_0)}(T^{n_0}\ T^{(i-1)\, n_0}\, x)\, B^{(n_0)}(T^{(i-1)\, n_0}\, x) \\ &= B^{(2n_0)}(T^{(i-1)n_0} x). \end{aligned}$$

The fiber LDT applied to the cocycle B at scales n_0 and $2n_0$ will ensure that the geometric conditions in the AP are satisfied except for a small set of phases. Indeed, for all scales $m \geq \overline{n_0}$, if x is outside a set $\mathscr{B}_m$ of measure $< e^{-c\,\epsilon^2 m}$, then

$$-\,\epsilon < \frac{1}{m} \log\|B^{(m)}(x)\| - L^{(m)}(B) < \epsilon\,. \tag{3.16}$$

The gap condition will follow by using the left hand side of (3.16) at scale n_0, the assumption (3.7), and the positivity of $L(A)$.

If $x \notin \mathscr{B}_{n_0}$, then

$$\begin{aligned} \frac{1}{n_0} \log\|B^{(n_0)}(x)\| &> L^{(n_0)}(B) - \epsilon \\ &> L^{(n_0)}(A) - \theta_0 - \epsilon \\ &\geq L(A) - \theta_0 - \epsilon. \end{aligned}$$

We conclude that for $x \notin \mathscr{B}_{n_0}$, where $\mu(\mathscr{B}_{n_0}) < e^{-c\,\epsilon^2 n_0}$,

$$\mathrm{gr}(B^{(n_0)}(x)) = \|B^{(n_0)}(x)\|^2 > e^{2n_0\,(L(A)-\theta_0-\epsilon)} =: \frac{1}{\varkappa_{ap}}\,. \tag{3.17}$$

Next we address the validity of the angles condition.

Applying the left hand side of (3.16) at scale $2n_0$, we have that for $x \notin \mathscr{B}_{2n_0}$,

$$\frac{1}{2n_0} \log\|B^{(2n_0)}(x)\| > L^{(2n_0)}(B) - \epsilon\,.$$

Applying the right hand side of of (3.16) at scale n_0, we have that for $x \notin \mathcal{B}_{n_0} \cup T^{-n_0}\mathcal{B}_{n_0}$,

$$\frac{1}{n_0} \log\|B^{(n_0)}(x)\| < L^{(n_0)}(B) + \epsilon$$
$$\frac{1}{n_0} \log\|B^{(n_0)}(T^{n_0}x)\| < L^{(n_0)}(B) + \epsilon.$$

Combining the last three estimates, for $x \notin \mathcal{B}_{2n_0} \cup \mathcal{B}_{n_0} \cup T^{-n_0}\mathcal{B}_{n_0} =: \tilde{\mathcal{B}}_{n_0}$, where $\mu(\tilde{\mathcal{B}}_{n_0}) < 3e^{-c\,\epsilon^2 n_0}$, we have that:

$$\frac{\|B^{(2n_0)}(x)\|}{\|B^{(n_0)}(T^{n_0}x)\| \, \|B^{(n_0)}(x)\|} > \frac{e^{2n_0(L^{(2n_0)}-\epsilon)}}{e^{2n_0(L^{(n_0)}+\epsilon)}} = e^{2n_0(L^{(2n_0)}-L^{(n_0)}-2\epsilon)}.$$

Using the inductive assumption (3.6) we conclude:

$$\frac{\|B^{(2n_0)}(x)\|}{\|B^{(n_0)}(T^{n_0}x)\| \, \|B^{(n_0)}(x)\|} > e^{-2n_0(\eta_0+2\epsilon)} =: \epsilon_{ap} \,. \tag{3.18}$$

Note that (3.8) implies $\frac{\varkappa_{ap}}{\epsilon_{ap}^2} = e^{-n_0\ (L(A)-\theta_0-2\eta_0-5\epsilon)} < e^{-\epsilon\, n_0}$, hence $\varkappa_{ap} \ll \epsilon_{ap}^2$.

Let $\bar{\mathcal{B}}_{n_0} := \bigcup_{i=0}^{n-1} T^{-i\,n_0}\, \tilde{\mathcal{B}}_{n_0}$. Note that

$$\mu(\bar{\mathcal{B}}_{n_0}) < 3\,n\,e^{-c\,\epsilon^2 n_0} = 3\,\frac{n_1}{n_0}\,e^{-c\,\epsilon^2 n_0} < e^{c/2\,\epsilon^2 n_0}\,e^{-c\,\epsilon^2 n_0}$$
$$= e^{-c/2\,\epsilon^2 n_0} = e^{-c_1\,n_0}\,.$$

Moreover, note that when $x \notin \bar{\mathcal{B}}_{n_0}$, the geometric conditions (3.17) and (3.18) hold for the phases $x, T^{n_0}x, \ldots, T^{(n-1)n_0}x$. That is, the blocks of length n_0 defined earlier satisfy:

$$\operatorname{gr}(g_i) > \frac{1}{\varkappa_{ap}} \quad \text{for all} \quad 0 \le i \le n-1\,,$$
$$\frac{\|g_i\, g_{i-1}\|}{\|g_i\|\,\|g_{i-1}\|} > \epsilon_{ap} \quad \text{for all} \quad 1 \le i \le n-1\,.$$

Therefore, we can apply the estimate (2.3) in the avalanche principle and obtain:

$$\Big|\, \log\|g^{(n)}\| + \sum_{i=1}^{n-2} \log\|g_i\| - \sum_{i=1}^{n-1} \log\|g_i\, g_{i-1}\| \,\Big| \lesssim n \cdot \frac{\varkappa_{ap}}{\epsilon_{ap}^2}\,.$$

Rewriting this in terms of matrix blocks we have that for $x \notin \bar{\mathscr{B}}_{n_0}$, where $\mu(\bar{\mathscr{B}}_{n_0}) < e^{-c_1 n_0}$,

$$\Big| \log\|B^{(n_1)}(x)\| + \sum_{i=1}^{n-2} \log\|B^{(n_0)}(T^{i n_0}\, x)\| - \sum_{i=1}^{n-1} \log\|B^{(2n_0)}(T^{(i-1)n_0}\, x)\| \Big| \lesssim n\, e^{-\epsilon\, n_0}. \tag{3.19}$$

Divide both sides of (3.19) by $n_1 = n\, n_0$ to get that for all $x \notin \bar{\mathscr{B}}_{n_0}$ we have

$$\Big| \frac{1}{n_1} \log\|B^{(n_1)}(x)\| + \frac{1}{n} \sum_{i=1}^{n-2} \frac{1}{n_0} \log\|B^{(n_0)}(T^{i n_0}\, x)\| - \frac{2}{n} \sum_{i=1}^{n-1} \frac{1}{2n_0} \log\|B^{(2n_0)}(T^{(i-1)n_0}\, x)\| \Big| \lesssim e^{-\epsilon\, n_0}\,.$$

Denote by $f(x)$ the function on the left hand side of the estimate above; then $\left|f(x)\right| \lesssim e^{-\epsilon\, n_0}$ for $x \notin \bar{\mathscr{B}}_{n_0}$ and using (3.15), for a.e. $x \in X$ we have $\left|f(x)\right| \lesssim C$. Moreover,

$$\int_X f(x)\, \mu(dx) = L^{(n_1)}(B) + \frac{n-2}{n}\, L^{(n_0)}(B) - \frac{2(n-1)}{n} L^{(2n_0)}(B)\,,$$

hence

$$\begin{aligned} \Big| L^{(n_1)}(B) + \frac{n-2}{n}\, L^{(n_0)}(B) - \frac{2(n-1)}{n} L^{(2n_0)}(B) \Big| &\le \int_X \left|f(x)\right| \mu(dx) \\ = \int_{\bar{\mathscr{B}}_{n_0}^{\complement}} \left|f(x)\right| \mu(dx) + \int_{\bar{\mathscr{B}}_{n_0}} \left|f(x)\right| \mu(dx) &\lesssim e^{-\epsilon\, n_0} + C\, \mu(\bar{\mathscr{B}}_{n_0}) \\ \lesssim e^{-\epsilon\, n_0} + C\, e^{-c_1\, n_0} \lesssim C\, e^{-c_1\, n_0} &< C\, \frac{n_0}{n_1}\,. \end{aligned}$$

Therefore,

$$\Big| L^{(n_1)}(B) + \frac{n-2}{n}\, L^{(n_0)}(B) - \frac{2(n-1)}{n} L^{(2n_0)}(B) \Big| < C\, \frac{n_0}{n_1}\,.$$

The term on the left hand side of the above inequality can be written in the form

$$\mid L^{(n_1)}(B) + L^{(n_0)}(B) - 2L^{(2n_0)}(B) - \frac{2}{n}\,[L^{(n_0)}(B) - L^{(2n_0)}(B)] \mid,$$

hence we conclude:

$$\begin{aligned} &\mid L^{(n_1)}(B) + L^{(n_0)}(B) - 2L^{(2n_0)}(B) \mid \\ &\qquad < C\,\frac{n_0}{n_1} + \frac{2}{n}[L^{(n_0)}(B) - L^{(2n_0)}(B)] \lesssim C\frac{n_0}{n_1}\,. \end{aligned} \tag{3.20}$$

Clearly the same argument leading to (3.20) will hold for $2n_1$ instead of n_1, which via the triangle inequality proves (3.11), that is, the conclusion (a++).

We can rewrite (3.20) in the form

$$\mid L^{(n_1)}(B) - L^{(n_0)}(B) + 2[L^{(n_0)}(B) - L^{(2n_0)}(B)] \mid < C\frac{n_0}{n_1}\,. \tag{3.21}$$

Using (3.21) for B and A we get:

$$\begin{aligned} &\left|L^{(n_1)}(B) - L^{(n_1)}(A)\right| \\ &\quad < \mid L^{(n_1)}(B) - L^{(n_0)}(B) + 2[L^{(n_0)}(B) - L^{(2n_0)}(B)] \mid \\ &\quad + \mid L^{(n_1)}(A) - L^{(n_0)}(A) + 2[L^{(n_0)}(A) - L^{(2n_0)}(A)] \mid \\ &\quad + \left|L^{(n_0)}(B) - L^{(n_0)}(A)\right| \\ &\quad + 2\left|L^{(n_0)}(B) - L^{(2n_0)}(B)\right| + 2\left|L^{(n_0)}(A) - L^{(2n_0)}(A)\right| \\ &\quad < \theta_0 + 4\eta_0 + C\frac{n_0}{n_1}\,, \end{aligned}$$

which establishes (3.12), that is, the conclusion (b++) of the proposition. □

Proposition 3.4 (rate of convergence)**.** *Let $A \in \mathcal{C}^\star$. There are constants $\delta_1 > 0$, $\overline{n_1} \in \mathbb{N}$, $c_2 > 0$, $K < \infty$, all depending only on A, such that the following hold.*

$$\left|L(B) - L^{(n)}(B)\right| < K\,\frac{\log n}{n}\,, \tag{3.22}$$

$$\left|L(B) + L^{(n)}(B) - 2L^{(2n)}(B)\right| < e^{-c_2\,n}\,, \tag{3.23}$$

for all $n \geq \overline{n_1}$ and for all $B \in \mathcal{C}$ with $d(B, A) < \delta_1$.

Proof. We will apply repeatedly the inductive step procedure in Proposition 3.3. The constants $\epsilon, c_1, C, \delta, \overline{n_0}$ appearing in this proof are the once introduced there.

First we use the finite scale continuity in Proposition 3.2 to create a neighborhood of A and a large enough interval of scales $\mathcal{N}_0$, such that if $n_0 \in \mathcal{N}_0$ and if B is in that neighborhood, then the assumptions (3.6), (3.7) in Proposition 3.3 hold.

Indeed, let $n_0^- \asymp \overline{n_0}$, $n_0^+ \asymp e^{\overline{n_0}}$ and define

$$\mathcal{N}_0 := [n_0^-, n_0^+] .$$

Let $\delta_1 := \min\{\delta, e^{-3\, n_0^+}\}$.

By Proposition 3.2, if $B \in \mathcal{C}$ with $d(B, A) \leq 1$, then for all $n \geq 1$,

$$\left| L^{(n)}(B) - L^{(n)}(A) \right| \leq e^{C\, n}\, d(B, A) .$$

Let $B \in \mathcal{C}$ with $d(B, A) < \delta_1$ and let $n_0 \in \mathcal{N}_0$. Then $n_0 \leq n_0^+$ and we have

$$\left| L^{(2n_0)}(B) - L^{(2n_0)}(A) \right| < e^{C\, 2n_0}\, \delta_1 \leq e^{C\, 2n_0^+}\, e^{-3C\, n_0^+} = e^{-C\, n_0^+} < \epsilon ,$$

since $\overline{n_0}$ is assumed large enough.

Similarly we have

$$\left| L^{(n_0)}(B) - L^{(n_0)}(A) \right| < e^{-2C\, n_0^+} < \epsilon =: \theta_0 .$$

Since also

$$\left| L^{(2n_0)}(A) - L^{(n_0)}(A) \right| \leq \left| L^{(2n_0)}(A) - L(A) \right| + \left| L(A) - L^{(n_0)}(A) \right| < 2\epsilon ,$$

from the last three inequalities we conclude that

$$\left| L^{(n_0)}(B) - L^{(2n_0)}(B) \right| \leq \epsilon + 2\epsilon + \epsilon = 4\epsilon =: \eta_0 .$$

Moreover,

$$\theta_0 + 2\eta_0 = \epsilon + 8\epsilon = 9\epsilon < L(A) - 6\epsilon .$$

We conclude that the assumptions (3.6), (3.7), (3.8) in Proposition 3.3 hold at any scale $n_0 \in \mathcal{N}_0$ and for any cocycle B with $d(B, A) < \delta_1$.

Let $n_1^- \asymp e^{c_1 \, n_0^-}$, $n_1^+ \asymp e^{c_1 \, n_0^+}$ and define

$$\mathcal{N}_1 := [n_1^-, \, n_1^+] \, .$$

If $n_1 \in \mathcal{N}_1$, then clearly there is $n_0 \in \mathcal{N}_0$ such that $n_1 \asymp e^{c_1 \, n_0}$ (hence $n_0 \lesssim \log n_1$).

We may then apply Proposition 3.3 with the pair of scales n_0, n_1 and obtain the following:

$$\left| \; L^{(n_1)}(B) + L^{(n_0)}(B) - 2L^{(2n_0)}(B) \; \right| < C \frac{n_0}{n_1} < K \, \frac{\log n_1}{n_1} \, ,$$

for some constant $K = K(A) < \infty$.

Furthermore,

$$\begin{aligned} L^{(n_1)}(B) - L^{(2n_1)}(B) &< \eta_1 \\ \left| L^{(n_1)}(B) - L^{(n_1)}(A) \right| &< \theta_1 \, , \end{aligned}$$

where

$$\begin{aligned} \theta_1 &= \theta_0 + 4\eta_0 + C \, \frac{n_0}{n_1} < 17\epsilon + K \, \frac{\log n_1}{n_1} \, , \\ \eta_1 &= C \, \frac{n_0}{n_1} < K \, \frac{\log n_1}{n_1} \, . \end{aligned}$$

Moreover,

$$\begin{aligned} \theta_1 + 2\eta_1 &\le 17\epsilon + K \, \frac{\log n_1}{n_1} + 2K \, \frac{\log n_1}{n_1} = 17\epsilon + 3K \, \frac{\log n_1}{n_1} < 20\epsilon \\ &< L(A) - 6\epsilon \, . \end{aligned}$$

This shows that the assumptions of Proposition 3.3 are again satisfied for all scales $n_1 \in \mathcal{N}_1$ and for all cocycles B with $d(B, A) < \delta_1$, so we can continue the process.

Let $n_2^- \asymp e^{c_1 \, n_1^-}$, $n_2^+ \asymp e^{c_1 \, n_1^+}$ and define

$$\mathcal{N}_2 := [n_2^-, \, n_2^+] \, .$$

Note that the intervals of scales $\mathcal{N}_1$ and $\mathcal{N}_2$ overlap.

If $n_2 \in \mathcal{N}_2$, then clearly there is $n_1 \in \mathcal{N}_1$ such that $n_2 \asymp e^{c_1 n_1}$ (hence $n_1 \lesssim \log n_2$).

We may then apply Proposition 3.3 with the pair of scales n_1, n_2 and obtain the following:

$$\mid L^{(n_2)}(B) + L^{(n_1)}(B) - 2L^{(2n_1)}(B) \mid < C\frac{n_1}{n_2} < K\,\frac{\log n_2}{n_2}\,.$$

Furthermore,

$$L^{(n_2)}(B) - L^{(2n_2)}(B) < \eta_2$$
$$\left|L^{(n_2)}(B) - L^{(n_2)}(A)\right| < \theta_2\,,$$

where

$$\theta_2 = \theta_1 + 4\eta_1 + C\,\frac{n_1}{n_2} < 17\epsilon + 5K\,\frac{\log n_1}{n_1} + K\,\frac{\log n_2}{n_2}\,,$$
$$\eta_2 = C\,\frac{n_1}{n_2} < K\,\frac{\log n_2}{n_2}\,.$$

It is now becoming clear that we can continue this argument inductively. That is because at each step k, the error η_k in the estimate (3.6) is very small, $\eta_k \le K\,\frac{\log n_k}{n_k}$ when $k \ge 1$, while the error θ_k in the estimate (3.7), starting with $k \ge 2$, only increases by a term of order $\frac{\log n_k}{n_k}$. However, the series $\sum_{k\ge1}\frac{\log n_k}{n_k}$ is summable, and its sum is of order $\frac{\log n_1}{n_1} < \epsilon$. The error θ_k is then of order ϵ for all k, thus the assumption (3.8) is always satisfied.

Therefore we obtain a sequence of overlapping intervals of scales $\{\mathcal{N}_k\}_{k\ge1}$. Their union covers all natural numbers $n \ge n_1^-$.

Define the threshold $\overline{n_1} := n_1^-$ and let $n \ge \overline{n_1}$. Then there is $k \ge 0$ such that $n \in \mathcal{N}_{k+1}$, so there is also $n_k \in \mathcal{N}_k$ such that

$$n = n_{k+1} \asymp e^{c_1\, n_k}\,.$$

The conclusions of Proposition 3.3 hold with the pair of scales n_k, n_{k+1}. Let us first use (3.11) and conclude that

$$L^{(n_{k+1})}(B) - L^{(2n_{k+1})}(B) < \eta_{k+1} < K\,\frac{\log n_{k+1}}{n_{k+1}}\,.$$

This shows that for all $n \geq \overline{n_1}$, we have

$$L^{(n)}(B) - L^{(2n)}(B) < K \, \frac{\log n}{n} \, ,$$

from which we can easily conclude (see the Exercise 3.6 below) that

$$\left| L^{(n)}(B) - L(B) \right| \lesssim K \, \frac{\log n}{n} \, ,$$

establishing (3.22).

We now use the conclusion (3.10) of Proposition 3.3 and conclude that

$$\begin{aligned} \mid L^{(n_{k+1})}(B) + L^{(n_k)}(B) - 2L^{(2n_k)}(B) \mid &< K \, \frac{\log n_{k+1}}{n_{k+1}} \\ &\leq K \, c_1 \, n_k \, e^{-c_1 \, n_k} < e^{-\frac{c_1}{2} \, n_k} \, . \end{aligned}$$

On the other hand, applying (3.22) with $n = n_{k+1}$ we have

$$\left| L^{(n_{k+1})}(B) - L(B) \right| \lesssim K \, \frac{\log n_{k+1}}{n_{k+1}} < e^{-\frac{c_1}{2} \, n_k} \, .$$

The last two inequalities then imply

$$\mid L(B) + L^{(n_k)}(B) - 2L^{(2n_k)}(B) \mid < 2e^{-\frac{c_1}{2} \, n_k} < e^{-\frac{c_1}{3} \, n_k} \, .$$

This establishes (3.23) for every $n \geq \overline{n_1}$ as well. □

Exercise 3.6. Let $\{x_n\}_{n \geq 1}$ be a sequence of real numbers that converges to x and assume that for all n,

$$|x_n - x_{2n}| \leq K \, \frac{\log n}{n} \, .$$

Prove that

$$|x_n - x| \lesssim K \, \frac{\log n}{n} \, .$$

Proof of Theorem 3.3 parts 1a. and 1b. Let $A \in \mathcal{C}$ with $L(A) > 0$. We wish to prove that in a small neighborhood of A, the function L is Hölder continuous. For that, recall Exercise 3.1 and the discussion preceding it.

The finite scale LE functions $L^{(n)} \to L$ as $n \to \infty$. However, the rate of convergence given by (3.22) is too slow. Instead, we use the sequence of functions

$$f_n := -L^{(n)} + 2L^{(2n)} \colon \mathcal{C} \to \mathbb{R}\,.$$

Clearly $f_n(B) \to -L(B)+2L(B) = L(B)$ as $n \to \infty$, for all $B \in \mathcal{C}$. Moreover, in a small neighborhood of A, by (3.23) in Proposition 3.4, this rate of convergence is exponential: for all $n \geq \overline{n_1}$ we have

$$\left|L(B) - f_n(B)\right| = \left|L(B) + L^{(n)}(B) - 2L^{(2n)}(B)\right| \leq e^{-c_2\, n}\,.$$

By the finite scale continuity in Proposition 3.2, if $B_1, B_2 \in \mathcal{C}$ are such that $d(B_1, B_2) < e^{-2(C+c_2)\, n}$, then for $m = n$ and $m = 2n$,

$$\left|L^{(m)}(B_1) - L^{(m)}(B_2)\right| \leq e^{C\, m} d(B_1, B_2) < e^{C\, 2n}\, e^{-2(C+c_2)\, n} = e^{-2c_2\, n}.$$

Thus $d(f_n(B_1), f_n(B_2)) < 3e^{-2c_2\, n} < e^{-c_2\, n}$.

From Exercise 3.1 we conclude that L is Hölder continuous (with exponent $\alpha = \frac{c_2}{2(C+c_2)}$) in a neighborhood of A, thus establishing part 1.b of Theorem 3.3.

Continuity at cocycles with zero Lyapunov exponents is immediate, due to upper semicontinuity (see Exercise 3.4), and this proves part 1a. □

Remark 3.2. Let $A \in \mathcal{C}^\star$. The estimate (3.22) in Proposition 3.4 shows that there is a neighborhood $\mathcal{V}$ of A in $\mathcal{C}$ and a threshold $\overline{n_1} \in \mathbb{N}$, such that the finite scale Lyapunov exponents $L^{(n)}$, $n \geq \overline{n_1}$ converge uniformly to L on $\mathcal{V}$. Combined with the continuity of the LE, this implies the following.

For every small $\epsilon > 0$, there is $n(\epsilon)$ such that for all $n \geq n(\epsilon)$ and for all $B \in \mathcal{V}$ we have

$$\left|L(B) - L(A)\right| \leq \epsilon \quad \text{and} \quad \left|L^{(n)}(B) - L(A)\right| \leq \epsilon\,.$$

Therefore, a-posteriori the uniform LDT can be formulated in the following stronger way. There is a neighborhood $\mathcal{V}$ of A and a constant $c > 0$ such that for all small $\epsilon > 0$, there is $\bar{n}(\epsilon)$ such that

$$\mu\Big\{x \in X \colon \left|\frac{1}{n}\log\|B^{(n)}(x)\| - L(A)\right| > \epsilon\Big\} < e^{-c\epsilon^2 n}$$

for all $B \in \mathcal{V}$ and $n \geq \bar{n}(\epsilon)$.

3.4 Continuity of the Oseledets splitting

We begin by introducing some concepts needed in the argument.

Given a cocycle $F_A\colon X \times \mathbb{R}^2 \to X \times \mathbb{R}^2$, $F_A(x, v) = (Tx, A(x)v)$, determined by a function $A\colon X \to \mathrm{SL}_2(\mathbb{R})$, its *inverse* is the map $F_A^{-1}\colon X \times \mathbb{R}^2 \to X \times \mathbb{R}^2$, $F_A^{-1}(x, v) = (T^{-1}x, A(T^{-1}x)^{-1}\, v)$. The iterates of the inverse cocycle $A^{-1}\colon X \to \mathrm{SL}_2(\mathbb{R})$ satisfy for all $n \in \mathbb{N}$ and $x \in X$,

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

Similarly the *adjoint* of F_A is the map $F_{A^*}\colon X \times \mathbb{R}^2 \to X \times \mathbb{R}^2$, defined by $F_{A^*}(x, v) = (T^{-1}x, A(T^{-1}x)^*\, v)$. The iterates of the adjoint cocycle $A^*\colon X \to \mathrm{SL}_2(\mathbb{R})$ satisfy for all $n \in \mathbb{N}$ and $x \in X$,

$$(A^*)^{(n)}(x) = A^{(n)}(T^{-n}x)^*\ .$$

Given $g \in \mathrm{GL}_2(\mathbb{R})$, let $\mathfrak{v}_+(g) = \mathfrak{v}(g)$ be a most expanding unit vector of g and denote by $\mathfrak{v}_-(g)$ a least expanding unit vector of g. Then $\{\mathfrak{v}_+(g), \mathfrak{v}_-(g)\}$ is a singular vector basis of g. As before let $\hat{\mathfrak{v}}_\pm(g)$ be the projective point determined by $\mathfrak{v}_\pm(g)$.

Any projective point $\hat{p} \in \mathbb{P}$ determines a unique line $\ell \in \mathrm{Gr}_1(\mathbb{R}^2)$, and this correspondence is one-to-one and onto. From now on, we make the identification $\mathrm{Gr}_1(\mathbb{R}^2) \equiv \mathbb{P}(\mathbb{R}^2)$.

We will write $\mathcal{E}_A^\pm(x)$ instead of $\mathcal{E}^\pm(x)$ to emphasize the dependence on A of the Oseledets decomposition of the cocycle A.

Remark 3.3. It follows from the proof of [60, Theorem 3.20] that if $L(A) > 0$ then for μ-almost every $x \in X$,

$$\mathcal{E}_A^\pm(x) = \lim_{n\to+\infty} \hat{\mathfrak{v}}_-(A^{(\mp n)}(x)).$$

Exercise 3.7. Given $g \in \mathrm{GL}_2(\mathbb{R})$ prove that $\hat{\mathfrak{v}}_\pm(g^{-1}) = \hat{\mathfrak{v}}_\mp(g^*)$.

Exercise 3.8. Prove that if $L(A) > 0$ then for μ-a.e. $x \in X$,

$$\mathcal{E}_{A^*}^+(x) = \lim_{n\to+\infty} \hat{\mathfrak{v}}_+(A^{(n)}(x)).$$

Hint: Use Remark 3.3 and Exercise 3.7.

In [16], assuming $L(A) > 0$, we define the sequence of partial functions $\overline{\mathfrak{v}}^{(n)}(A)\colon X \to \mathbb{P}(\mathbb{R}^2)$,

$$\overline{\mathfrak{v}}^{(n)}(A)(x) := \begin{cases} \hat{\mathfrak{v}}(A^{(n)}(x)) & \text{if } \mathrm{gr}(A^{(n)}(x)) > 1 \\ \text{undefined} & \text{otherwise.} \end{cases}$$

By Proposition 4.4 in [16], this sequence converges μ-almost everywhere to a (total) measurable function $\overline{\mathfrak{v}}^{(\infty)}(A)\colon X \to \mathbb{P}(\mathbb{R}^2)$,

$$\overline{\mathfrak{v}}^{(\infty)}(A)(x) := \lim_{n\to+\infty} \overline{\mathfrak{v}}^{(n)}(A)(x).$$

This limit also exists by Exercise 3.8.

Exercise 3.9. Prove that if a cocycle A is μ-integrable then its adjoint A^* is also μ-integrable. Moreover, show that A and its adjoint A^* have the same Lyapunov exponent $L(A) = L(A^*)$.

Exercise 3.10. Consider a cocycle A such that $L(A) > 0$. Prove that $\mathcal{E}_A^{\pm} = (\mathcal{E}_{A^*}^{\mp})^{\perp}$. In particular, $\mathcal{E}_A^{+} = \overline{\mathfrak{v}}^{(\infty)}(A^*)$ and $\mathcal{E}_A^{-} = (\overline{\mathfrak{v}}^{(\infty)}(A))^{\perp}$.

Let $A \in \mathcal{C}^{\star}$, so that A satisfies the uniform LDT estimates (3.3). Given $\epsilon > 0$ write $L = L(A)$ and define the set

$$\Omega_{n,\epsilon}(A) := \left\{ x \in X\colon \left|\frac{1}{m} \log\|A^{(m)}(x)\| - L\right| \leq \epsilon,\ \forall m \geq n \right\}.$$

Exercise 3.11. Show that $\lim_{n\to+\infty} \mu(\Omega_{n,\epsilon}(A)) = 1$ for all $\epsilon > 0$.

Exercise 3.12. Given $A \in \mathcal{C}^{\star}$, show that

$$\mu(X \setminus \Omega_{n,\epsilon}(B)) \lesssim e^{-n\,c\,\epsilon^2}$$

for all $\epsilon > 0$, $n \in \mathbb{N}$ and any cocycle $B \in \mathcal{C}$ close enough to A.

The proof of the next proposition uses the argument in [60, Lemma 3.16]). Due to the assumed availability of the LDT estimate, the argument becomes quantitative.

Proposition 3.5 (rate of convergence)**.** *Let $A \in \mathcal{C}^{\star}$. There are constants $\delta > 0$, $\epsilon > 0$, $\overline{n_0} \in \mathbb{N}$ and $C < \infty$, all depending only on A, such that*

$$d\left(\overline{\mathfrak{v}}^{(n)}(B), \overline{\mathfrak{v}}^{(\infty)}(B)\right) < C\, e^{-n\,c\,\epsilon^2}$$

for all $n \geq \overline{n_0}$ and for all $B \in \mathcal{C}$ with $d(B, A) < \delta$.

Proof. Take $0 < \gamma < L(A)$. By the continuity of the LE we can assume that δ is small enough so that

$$\inf\{L(B)\colon B \in \mathcal{C} \text{ and } d(B,A) < \delta\} > \gamma.$$

Next choose $\epsilon > 0$ so that for all $B \in \mathcal{C}$ with $d(B,A) < \delta$

$$L(B) - \epsilon \gg \gamma \gg c\,\epsilon^2$$

where $c > 0$ is the LDT parameter in (3.3)

Let $v_m = v_m(x)$ be a unit vector in $\overline{\mathfrak{v}}^{(m)}(A)(x) = \hat{\mathfrak{v}}_+(A^{(m)}(x))$ and, similarly, let $u_m = u_m(x)$ be a unit vector in $\hat{\mathfrak{v}}_-(A^{(m)}(x))$. Then $\{v_m(x), u_m(x)\}$ is a singular vector basis of $A^{(m)}(x)$.

Consider now the following cocycle over the same base transformation T:

$$\widetilde{A}(x) := A(x)^{-*} = (A(x)^{-1})^* = (A(x)^*)^{-1}.$$

By Exercise 3.7 we have

$$\hat{u}_m(x) = \hat{\mathfrak{v}}_+(\widetilde{A}^{(m)}(x)) \text{ and } \hat{v}_m(x) = \hat{\mathfrak{v}}_-(\widetilde{A}^{(m)}(x)).$$

Let $\alpha_m(x) := \measuredangle(\hat{v}_m(x), \hat{v}_{m+1}(x))$, so that $\sin\alpha_m = \delta(\hat{v}_m, \hat{v}_{m+1})$. Then

$$v_m(x) = (\sin\alpha_m)\, u_{m+1}(x) + (\cos\alpha_m)\, v_{m+1}(x)$$

which implies that

$$\begin{aligned}\|\widetilde{A}^{(m+1)}(x) v_m(x)\| &\geq \left|\sin\alpha_m\right| \|\widetilde{A}^{(m+1)}(x) u_{m+1}(x)\| \\ &= \left|\sin\alpha_m\right| \|\widetilde{A}^{(m+1)}(x)\|.\end{aligned}$$

Since $\widetilde{A}^{(m)}(x) \in \mathrm{SL}_2(\mathbb{R})$, $\|\widetilde{A}^{(m)}(x)\, v_m(x)\| = \|\widetilde{A}^{(m)}(x)\|^{-1}$ and

$$\begin{aligned}\delta(\hat{v}_m(x), \hat{v}_{m+1}(x)) = \left|\sin\alpha_m\right| &\leq \frac{\|\widetilde{A}^{(m+1)}(x)\, v_m(x)\|}{\|\widetilde{A}^{(m+1)}(x)\|} \\ &\leq \frac{\|\widetilde{A}(T^m x)\|\, \|\widetilde{A}^{(m)}(x)\, v_m(x)\|}{\|\widetilde{A}^{(m+1)}(x)\|} = \frac{\|\widetilde{A}(T^m x)\|}{\|\widetilde{A}^{(m+1)}(x)\|\, \|\widetilde{A}^{(m)}(x)\|}.\end{aligned}$$

Because A is an SL_2-cocycle, $\|A\|_\infty = \|\widetilde{A}\|_\infty$ and $\Omega_{n,\epsilon}(A) = \Omega_{n,\epsilon}(\widetilde{A})$. Hence, for all $m \geq n$ and $x \in \Omega_{n,\epsilon}(A)$

$$\begin{aligned}\frac{1}{m}\,\log\delta(\hat{v}_m(x),\hat{v}_{m+1}(x)) &\leq \frac{\log\|\widetilde{A}\|_\infty}{m} - \frac{m+1}{m}\,(L-\epsilon) - (L-\epsilon)\\ &\leq \frac{\log\|A\|_\infty}{m} - 2\,\gamma < -\gamma\end{aligned}$$

the last inequality holds provided $n > \log\|A\|_\infty/\gamma$.

Thus for all $m \geq n$,

$$\delta(\hat{v}_m(x),\hat{v}_{m+1}(x)) \leq e^{-m\,\gamma},$$

which implies that for all $x \in \Omega_{m,\epsilon}(A)$,

$$\delta\left(\overline{\mathfrak{v}}^{(m)}(A)(x),\overline{\mathfrak{v}}^{(\infty)}(A)(x)\right) < C\,e^{-m\,\gamma} \ll e^{-n\,c\,\epsilon^2}$$

with $C = (1-e^{-\gamma})^{-1}$.

Since $\mu(X \setminus \Omega_{n,\epsilon}(A)) \lesssim e^{-n\,c\,\epsilon^2}$ the conclusion follows by taking the average in x.

Since all bounds are uniform in a δ-neighborhood of A, the same convergence rate holds for all $B \in \mathcal{C}^\star$ with $d(B,A) < \delta$. $\square$

Proposition 3.6 (finite scale continuity)**.** *Given $\epsilon > 0$, there is a constant $C_1 = C_1(A,\epsilon) < \infty$, such that for any $B_1, B_2 \in \mathcal{C}$ with $d(B_i, A) < \delta$, $i = 1, 2$, if $n \geq \overline{n}(\epsilon)$ and $d(B_1, B_2) < e^{-C_1\,n}$, then for x outside a set of measure $\lesssim e^{-n\,c\,\epsilon^2}$*

$$\delta\left(\overline{\mathfrak{v}}^{(n)}(B_1)(x),\,\overline{\mathfrak{v}}^{(n)}(B_2)(x)\right) < e^{-n\,c\,\epsilon^2}\,. \tag{3.24}$$

Moreover,

$$d\left(\overline{\mathfrak{v}}^{(n)}(B_1),\overline{\mathfrak{v}}^{(n)}(B_2)\right) \lesssim e^{-n\,c\,\epsilon^2}\,. \tag{3.25}$$

Proof. Take $0 < \gamma < L(A)$. By the continuity of the LE we can assume that δ is small enough so that

$$\inf\{L(B)\colon B \in \mathcal{C} \text{ and } d(B,A) < \delta\} > \gamma.$$

Next choose $\epsilon > 0$ sufficiently small so that

$$L(B) - \epsilon > \gamma.$$

For each $n \geq \bar{n}(\epsilon)$ define the deviation set

$$\mathcal{B}_n(B) := \{x \in X : \frac{1}{n} \log\|B^{(n)}(x)\| < L(B) - \epsilon\}$$

which has exponentially small measure $\mu(\mathcal{B}_n(B)) < e^{-nc\epsilon^2}$.

Given two cocycles $B_1, B_2 \in \mathcal{C}$ with $d(B_i, A) < \delta$ $(i = 1, 2)$ and an integer $n \geq \bar{n}(\epsilon)$ take $x \notin \mathcal{B}_n(B_1) \cup \mathcal{B}_n(B_2)$ and set $g_i := B_i^{(n)}(x)$.

Firstly note that

$$\mathrm{gr}(g_i) = \|B_i^{(n)}(x)\|^2 \geq e^{2n\,(L(B_i)-\epsilon)} > e^{2n\gamma} \gg 1,$$

so in particular $\overline{\mathfrak{v}}^{(n)}(B_i)(x) = \overline{\mathfrak{v}}(B_i^{(n)}(x))$ are well defined.

Since for every x, $\|B_i(x)\| < C_0 = C(A) < \infty$, we have

$$\|g_i\| = \|B_i^{(n)}(x)\| < e^{C_0\,n}.$$

Moreover, assuming $d(B_1, B_2) < e^{-C_1\,n}$, with C_1 to be chosen later,

$$\begin{aligned}\|g_1 - g_2\| = \|B_1^{(n)}(x) - B_2^{(n)}(x)\| &\leq n\, e^{C_0\,(n-1)}\, d(B_1, B_2)\\ &< e^{-(C_1 - 2C_0)\,n}.\end{aligned}$$

If we choose $C_1 > 2C_0 - \gamma + c\,\epsilon^2$, then

$$d_{\mathrm{rel}}(g_1, g_2) = \frac{\|g_1 - g_2\|}{\max\{\|g_1\|, \|g_2\|\}} \leq e^{-(C_1 - 2C_0 + \gamma)\,n} < e^{-n\,c\,\epsilon^2} \ll 1\,.$$

Then Exercise 2.9 applies, and we conclude:

$$\delta(\overline{\mathfrak{v}}(g_1), \overline{\mathfrak{v}}(g_2)) \leq 12\, d_{\mathrm{rel}}(g_1, g_2) \lesssim e^{-n\,c\,\epsilon^2}.$$

This proves (3.24), while (3.25) follows by integration in x. □

Proof of Theorem 3.3 parts 2a. and 2b. The following functions are defined on $\mathcal{C}^\star$ and take values in $L^1(X, \mathbb{P}(\mathbb{R}^2))$.

$$f_n(A) := \overline{\mathfrak{v}}^{(n)}(A) \text{ and } f(A) := \overline{\mathfrak{v}}^{(\infty)}(A)$$
$$g_n(A) := \overline{\mathfrak{v}}^{(n)}(A^*) \text{ and } g(A) := \overline{\mathfrak{v}}^{(\infty)}(A^*)$$
$$h_n(A) := \overline{\mathfrak{v}}^{(n)}(A)^\perp \text{ and } f(A) := \overline{\mathfrak{v}}^{(\infty)}(A)^\perp\,,$$

where if $\mathcal{E} \in L^1(X, \mathbb{P}(\mathbb{R}^2))$, then the notation $\mathcal{E}^\perp$ refers to

$$X \ni x \mapsto \mathcal{E}^\perp(x) := (\mathcal{E}(x))^\perp \in \mathbb{P}(\mathbb{R}^2)\,.$$

Propositions 3.6 and 3.5 ensure that assumptions (i) and (ii) of Exercise 3.1 are satisfied for the sequence f_n, hence the limit f is Hölder continuous.

The same argument applied instead to the cocycle A^* shows that g is Hölder continuous.

Since $\perp\colon \mathbb{P}(\mathbb{R}^2) \to \mathbb{P}(\mathbb{R}^2)$ is an isometry, the Hölder continuity of h follows from that of f.

Thus we proved item 2a.

Item 2b follows from item 2a by applying Chebyshev's inequality. Indeed, if $\mathcal{E}^\pm\colon \mathcal{C}^\star \to L^1(X, \mathbb{P}(\mathbb{R}^2))$, $A \mapsto \mathcal{E}^\pm_A$, are locally Hölder continuous functions with Hölder exponent α, then

$$d(\mathcal{E}^\pm_{B_1}, \mathcal{E}^\pm_{B_2}) \lesssim d(B_1, B_2)^\alpha\,.$$

Let $f^\pm\colon X \to \mathbb{R}$ be the measurable functions

$$f^\pm(x) := \delta(\mathcal{E}^\pm_{B_1}(x), \mathcal{E}^\pm_{B_2}(x))\,.$$

Thus

$$\begin{aligned}\|f^\pm\|_{L^1} = \int_X f^\pm(x)\, d\mu(x) &= \int_X \delta(\mathcal{E}^\pm_{B_1}(x), \mathcal{E}^\pm_{B_2}(x))\, d\mu(x)\\ &= d(\mathcal{E}^\pm_{B_1}, \mathcal{E}^\pm_{B_2}) \lesssim d(B_1, B_2)^\alpha\,.\end{aligned}$$

Applying Chebyshev's inequality to $f^\pm$ we have

$$\mu\left\{x \in X\colon f^\pm(x) \geq d(B_1, B_2)^{\frac{\alpha}{2}}\right\} \leq \frac{\|f^\pm\|_{L^1}}{d(B_1, B_2)^{\frac{\alpha}{2}}} \lesssim d(B_1, B_2)^{\frac{\alpha}{2}}\,.$$

Thus

$$\mu\left\{x \in X\colon \delta(\mathcal{E}^\pm_{B_1}(x),\, \mathcal{E}^\pm_{B_2}(x)) \geq d(B_1, B_2)^{\frac{\alpha}{2}}\right\} \lesssim d(B_1, B_2)^{\frac{\alpha}{2}}\,,$$

which completes the proof. □

3.5 Bibliographical notes

This type of abstract continuity result for the Lyapunov exponents was not available prior to our monograph [16]. However, our method has its origin in M. Goldstein and W. Schlag [25], where the first version of the avalanche principle appeared and the use of large deviations was employed in establishing Hölder continuity for quasi-periodic Schrödinger cocycles. Furthermore, W. Schlag [53] hints at the modularity of this type of argument, an observation that motivated us to pursue this type of approach further.

Continuity of the Oseledets decomposition for $GL_2(\mathbb{C})$-valued random i.i.d. cocycles was obtained by C. Bocker-Neto and M. Viana in [6]. Their result is not quantitative but it requires no generic assumptions (such as irreducibility) on the space of cocycles. Other related results were recently obtained in [3, 4]. A different type of continuity property, namely stability of the Lyapunov exponents and of the Oseledets decomposition under random perturbations of a fixed cocycle, was studied in [40, 47].

Chapter 4

Random Cocycles

4.1 Introduction and statement

Given a compact metric space (Σ, d) consider the space of sequences $\Omega_\Sigma = \Sigma^{\mathbb{Z}}$ endowed with the product topology. The homeomorphism $T\colon \Omega_\Sigma \to \Omega_\Sigma$, $T\{\omega_i\}_{i\in\mathbb{Z}} := \{\omega_{i+1}\}_{i\in\mathbb{Z}}$, is called the *full shift map*.

Let $\mathrm{Prob}(\Sigma)$ be the space of Borel probability measures on Σ. For a given measure $\mu \in \mathrm{Prob}(\Sigma)$ consider the product probability measure $\mathbb{P}_\mu = \mu^{\mathbb{Z}}$ on Ω_Σ. Then $(\Omega_\Sigma, \mathbb{P}_\mu, T)$ is an ergodic transformation, referred to as a *full Bernoulli shift*.

Let $L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$ be the space of bounded Borel-measurable functions $A\colon \Sigma \to \mathrm{SL}_2(\mathbb{R})$. Given $A \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$ and $\mu \in \mathrm{Prob}(\Sigma)$ they determine a measurable function $\tilde{A}\colon \Omega_\Sigma \to \mathrm{SL}_2(\mathbb{R})$, $\tilde{A}\{\omega_n\}_{n\in\mathbb{Z}} := A(\omega_0)$, and hence a linear cocycle $F_{(A,\mu)}\colon \Omega_\Sigma \times \mathbb{R}^2 \to \Omega_\Sigma \times \mathbb{R}^2$ over the Bernoulli shift $(\Omega_\Sigma, \mathbb{P}_\mu, T)$. We refer to the cocycle $F_{(A,\mu)}$ as a *random cocycle*, and identify the pair (A, μ) with the map $F_{(A,\mu)}$. The n-th iterate $F^n_{(A,\mu)} = F_{(A^n,\mu^n)}$ is the random cocycle determined by the pair (A^n, μ^n) where $\mu^n := \mu \times \cdots \times \mu \in \mathrm{Prob}(\Sigma^n)$ and where $A^n\colon \Sigma^n \to \mathrm{GL}_2(\mathbb{R})$ is the function

$$A^n(x_0, x_1, \ldots, x_{n-1}) := A(x_{n-1}) \cdots A(x_1)\, A(x_0).$$

The Lyapunov exponent of the random cocycle (A, μ) is simply denoted by $L(A)$, assuming the underlying measure μ is fixed.

A measure $\nu \in \mathrm{Prob}(\mathbb{P}(\mathbb{R}^2))$ is called *stationary* w.r.t. (A, μ) if

$$\nu = \int_\Sigma \hat{A}(x)_* \nu \, d\mu(x).$$

Furstenberg's formula [22] states that for any random cocycle (A, μ) with $A \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$ there exists at least one stationary measure $\nu \in \mathrm{Prob}(\mathbb{P}(\mathbb{R}^2))$ such that

$$L(A) = \int_\Sigma \int_{\mathbb{P}(\mathbb{R}^2)} \log\|A(x)\, p\| \, d\nu(\hat{p}) \, d\mu(x). \tag{4.1}$$

In this formula p stands for a unit representative of $\hat{p} \in \mathbb{P}(\mathbb{R}^2)$.

Given a cocycle (A, μ) and a line $\ell \subset \mathbb{R}^2$ invariant under all matrices $A(x)$ with $x \in \Sigma$, the pair $(A|_\ell, \mu)$ represents the linear cocycle obtained restricting $F_{(A,\mu)} \colon \Omega_\Sigma \times \mathbb{R}^2 \to \Omega_\Sigma \times \mathbb{R}^2$ to the 1-dimensional sub-bundle $\Omega_\Sigma \times \ell$. Because the process $L_n := \log\|\tilde{A}^{(n)}|_\ell\|$ is additive, by Birkhoff's ergodic theorem the Lyapunov exponent of $(A|_\ell, \mu)$ is

$$L(A|_\ell) = \int_\Sigma \log\|A(x)|_\ell\| \, d\mu(x).$$

Definition 4.1 (Définition 2.7 in [7])**.** A cocycle (A, μ) is called *quasi-irreducible* if there is no invariant line $\ell \subset \mathbb{R}^2$ which is invariant under all matrices of the cocycle, i.e., such that $A(x)\ell = \ell$ for μ-a.e. $x \in \Sigma$, and where $L(A|_\ell) < L(A)$.

As we will see (Propositions 4.2 and 4.6) , if a cocycle (A, μ) is quasi-irreducible then it admits a unique stationary measure $\nu \in \mathrm{Prob}(\mathbb{P}(\mathbb{R}^2))$. In this case $L(A)$ is uniquely determined by ν through Furstenberg's formula (4.1).

The next theorem states that random quasi-irreducible cocycles with positive Lyapunov exponent satisfy a uniform LDT.

Theorem 4.1. *Given* $\mu \in \mathrm{Prob}(\Sigma)$ *and* $A \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$ *assume*

(1) (A, μ) *is quasi-irreducible,*

(2) $L(A) > 0$.

There exist constants $\delta = \delta(A, \mu) > 0$, $C = C(A, \mu) < \infty$, $\kappa = \kappa(A, \mu) > 0$ *and* $\varepsilon_0 = \varepsilon_0(A) > 0$ *such that for all* $\|A - B\|_\infty < \delta$ *with* $B \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$, *for all* $0 < \varepsilon < \varepsilon_0$ *and* $n \in \mathbb{N}$ *we have*

$$\mathbb{P}_\mu \left[\left| \frac{1}{n} \log \|B^{(n)}\| - L(B) \right| > \varepsilon \right] \leq C\, e^{-\kappa\, \varepsilon^2\, n}.$$

Denote by $L^1(\Omega, \mathbb{P}(\mathbb{R}^2))$ the space of all measurable functions $\mathcal{E} \colon \Omega \to \mathbb{P}(\mathbb{R}^2)$ endowed with the following L^1 metric

$$d_\mu(\mathcal{E}, \mathcal{E}') := \mathbb{E}_\mu[\, \delta(\mathcal{E}, \mathcal{E}')\,] = \int_\Omega \delta(\mathcal{E}(x), \mathcal{E}'(x))\, d\mathbb{P}_\mu(x).$$

Given a cocycle $A \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$ with $L(A) > 0$, its Oseledets decomposition determines the two sections $\mathcal{E}^\pm_A \in L^1(\Omega, \mathbb{P}(\mathbb{R}^2))$ introduced in Chapter 1. The first part of the following theorem is due to E. Le Page [39][1].

Theorem 4.2. *Given* $\mu \in \mathrm{Prob}(\Sigma)$ *and* $A \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$ *assume*

(1) (A, μ) *is quasi-irreducible,*

(2) $L(A) > 0$.

Then there exists a neighborhood $\mathcal{V}$ *of* A *in* $L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$ *such that the function* $L \colon \mathcal{V} \to \mathbb{R}$, $B \mapsto L(B)$ *is Hölder continuous.*

Moreover, the Oseledets sections $\mathcal{V} \ni B \mapsto \mathcal{E}^\pm_B \in L^1(\Omega, \mathbb{P}(\mathbb{R}^2))$ *are also Hölder continuous w.r.t. the metric* d_μ.

4.2 Continuity of the Lyapunov exponent

In this section we provide a direct proof E. Le Page's theorem, the first half of Theorem 4.2. Like the second part, this is also a consequence of the ACT (Theorem 3.3) and Theorem 4.1. The proof presented here was adapted from [5].

[1] The theorem in [39] is formulated in a slightly more particular setting, for one parameter families of cocycles.

Throughout this chapter we set $\mathbb{P} := \mathbb{P}(\mathbb{R}^2)$. Recall the projective distance $\delta\colon \mathbb{P}\times\mathbb{P} \to [0,+\infty)$ introduced in Chapter 2 and given by

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

where p and q are representative vectors of the projective points $\hat{p}$ and $\hat{q}$ respectively.

Let $L^\infty(\mathbb{P})$ be the space of bounded Borel measurable functions $\phi\colon \mathbb{P}\to\mathbb{C}$. Given $\phi\in L^\infty(\mathbb{P})$ and $0<\alpha\le 1$, define

$$\|\phi\|_\infty := \sup_{\hat{p}\in\mathbb{P}} \left|\phi(\hat{p})\right|,$$

$$v_\alpha(\phi) := \sup_{\hat{p}\neq\hat{q}} \frac{\left|\phi(\hat{p})-\phi(\hat{q})\right|}{\delta(\hat{p},\hat{q})^\alpha},$$

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

Then

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

is the space of α-Hölder continuous functions on $\mathbb{P}$. The value $v_\alpha(\phi)$ will be referred to as the *Hölder constant* of ϕ. By convention we set $\mathcal{H}_0(\mathbb{P})$ to be the space of continuous functions on $\mathbb{P}$.

Exercise 4.1. Show that for all $\varphi\in\mathcal{H}_0(\mathbb{P})$ and $\hat{p}\in\mathbb{P}$,

$$\|\varphi-\varphi(\hat{p})\|_\infty \le v_0(\varphi) \le \|\varphi\|_\infty .$$

Exercise 4.2. Set $\mathbf{1}$ to be the constant function 1 and prove that $(\mathcal{H}_\alpha(\mathbb{P}), \|\cdot\|_\alpha)$ is a Banach algebra with unity $\mathbf{1}$.

Exercise 4.3. Show that $\{(\mathcal{H}_\alpha(\mathbb{P}), v_\alpha)\}_{\alpha\in[0,1]}$ is a family of semi-normed spaces such that for all $0\le\alpha<\beta\le 1$, $0\le t\le 1$ and $\varphi\in\mathcal{H}_\alpha(\mathbb{P})$,

1. $\mathcal{H}_\alpha(\mathbb{P})\subset\mathcal{H}_\beta(\mathbb{P})$ (monotonicity),
2. $v_\alpha(\varphi)\le v_\beta(\varphi)$ (monotonicity),
3. $v_{(1-t)\alpha+t\beta}(\varphi) \le v_\alpha(\varphi)^{1-t} v_\beta(\varphi)^t$ (convexity).

Given a cocycle $(A, \mu) \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R})) \times \mathrm{Prob}(\Sigma)$ we define its *Markov operator* $\mathcal{Q}_A = \mathcal{Q}_{A,\mu} \colon L^\infty(\mathbb{P}) \to L^\infty(\mathbb{P})$ by

$$\mathcal{Q}_{A,\mu}(\phi)(\hat{p}) := \int_\Sigma \phi(\hat{A}(x)\, \hat{p})\, d\mu(x),$$

where $\hat{A}(x) \colon \mathbb{P} \to \mathbb{P}$ stands for the projective action of $A(x)$.

Define also the quantity

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

measuring the average Hölder constant of $\hat{p} \mapsto \hat{A}(x)(\hat{p})$. Next proposition clarifies the importance of this measurement.

Proposition 4.1. *For all $\phi \in \mathcal{H}_\alpha(\mathbb{P})$,*

$$v_\alpha(\mathcal{Q}_{A,\mu}(\phi)) \leq \kappa_\alpha(A, \mu)\, v_\alpha(\phi).$$

Proof. Given $\phi \in \mathcal{H}_\alpha(\mathbb{P})$, and $\hat{p}, \hat{q} \in \mathbb{P}$,

$$\begin{aligned} \left| \mathcal{Q}_A(\phi)(\hat{p}) - \mathcal{Q}_A(\phi)(\hat{q}) \right| &\leq \int_{\mathbb{P}} \left| \phi(\hat{A}(x)\, \hat{p}) - \phi(\hat{A}(x)\, \hat{q}) \right| d\mu(x) \\ &\leq v_\alpha(\varphi) \int_{\mathbb{P}} \delta(\hat{A}(x)\, \hat{p}, \hat{A}(x)\, \hat{q})^\alpha\, d\mu(x) \\ &\leq v_\alpha(\varphi)\, \kappa_\alpha(A, \mu)\, \delta(\hat{p}, \hat{q})^\alpha \end{aligned}$$

which proves the proposition. □

Exercise 4.4. Prove that for all $n \in \mathbb{N}$,

$$(\mathcal{Q}_{A,\mu})^n = \mathcal{Q}_{A^n, \mu^n}.$$

Exercise 4.5. Show the sequence $\kappa_\alpha(A^n, \mu^n)$ is sub-multiplicative, i.e., for all $n, m \in \mathbb{N}$,

$$\kappa_\alpha(A^{n+m}, \mu^{n+m}) \leq \kappa_\alpha(A^n, \mu^n)\, \kappa_\alpha(A^m, \mu^m).$$

Definition 4.2 (See Definition II.1 in [29])**.** A bounded linear operator $\mathcal{Q}\colon \mathcal{B} \to \mathcal{B}$ on a Banach space $\mathcal{B}$ is called quasi-compact and simple if its spectrum admits a decomposition in disjoint closed sets $\mathrm{spec}(\mathcal{Q}) = \Sigma_0 \cup \{\lambda_0\}$ such that $\lambda_0 \in \mathbb{C}$ is a simple eigenvalue of $\mathcal{Q}$ and $|\lambda| < \lambda_0$ for all $\lambda \in \Sigma_0$.

Proposition 4.2. *Let $(A, \mu) \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R})) \times \mathrm{Prob}(\Sigma)$ be a cocycle such that for some $0 < \alpha \leq 1$ and $n \geq 1$*

$$\kappa_\alpha(A^n, \mu^n)^{\frac{1}{n}} \leq \sigma < 1.$$

Then the operator $\mathcal{Q} = \mathcal{Q}_{A,\mu}\colon \mathcal{H}_\alpha(\mathbb{P}) \to \mathcal{H}_\alpha(\mathbb{P})$ is quasi-compact and simple. More precisely there exists a (unique) stationary measure $\nu \in \mathrm{Prob}(\mathbb{P})$ w.r.t. (A, μ) such that defining the subspace

$$\mathcal{N}_\alpha(\nu) := \left\{\varphi \in \mathcal{H}_\alpha(\mathbb{P})\colon \int_{\mathbb{P}} \varphi \, d\nu = 0\right\}$$

the operator $\mathcal{Q}$ has the following properties

1. $\mathrm{spec}(\mathcal{Q}\colon \mathcal{H}_\alpha(\mathbb{P}) \to \mathcal{H}_\alpha(\mathbb{P})) \subset \{1\} \cup \mathbb{D}_\sigma(0)$,
2. $\mathcal{H}_\alpha(\mathbb{P}) = \mathbb{C}\,\mathbf{1} \oplus \mathcal{N}_\alpha(\nu)$ *is a $\mathcal{Q}$-invariant decomposition,*
3. *$\mathcal{Q}$ fixes every function in $\mathbb{C}\,\mathbf{1}$ and acts as a contraction with spectral radius $\leq \sigma$ on $\mathcal{N}_\alpha(\nu)$.*

Proof. The semi-norm v_α induces a norm on the quotient $\mathcal{H}_\alpha(\mathbb{P})/\mathbb{C}\mathbf{1}$. By Proposition 4.1, $\mathcal{Q}^n$ acts on $\mathcal{H}_\alpha(\mathbb{P})/\mathbb{C}\mathbf{1}$ as a σ^n-contraction. Hence $\mathcal{Q}$ also acts on $\mathcal{H}_\alpha(\mathbb{P})/\mathbb{C}\mathbf{1}$ as a contraction with spectral radius $\leq \sigma$. Since $\mathcal{Q}$ also fixes the constant functions in $\mathbb{C}\mathbf{1}$, it is a quasi-compact operator with simple eigenvalue 1 (associated to eigen-space $\mathbb{C}\,\mathbf{1}$) and inner spectral radius $\leq \sigma$. Thus $\mathrm{spec}(\mathcal{Q}) \subset \{1\} \cup \mathbb{D}_\sigma(0)$.

By spectral theory [51, Chap. XI] there exists a $\mathcal{Q}$-invariant decomposition $\mathcal{H}_\alpha(\mathbb{P}) = \mathbb{R}\,\mathbf{1} \oplus \mathcal{N}_\alpha$ such that $\mathcal{Q}$ acts as a contraction with spectral radius $\leq \sigma$ on $\mathcal{N}_\alpha$. Thus we can define a linear functional $\Lambda\colon \mathcal{H}_\alpha(\mathbb{P}) \to \mathbb{C}$ setting $\Lambda(c\,\mathbf{1} + \psi) := c$ for $\psi \in \mathcal{N}_\alpha$. This functional has several properties:

- $\Lambda(\mathbf{1}) = 1$.
- Λ is positive, i.e., $\varphi \geq 0$ implies $\Lambda(\varphi) \geq 0$. Given $\varphi = c\,\mathbf{1} + \psi \geq 0$

with $\psi \in \mathcal{N}_\alpha$, we have $0 \leq \mathcal{Q}^n(c\,\mathbf{1}+\psi) = c+\mathcal{Q}^n(\psi)$ for all $n \geq 0$. Since $\psi \in \mathcal{N}_\alpha$ we have $\lim_{n\to+\infty} \mathcal{Q}^n(\psi) = 0$ which implies $\Lambda(\varphi) = c \geq 0$.

▪ Λ is continuous w.r.t. the norm $\|\cdot\|_\infty$. Indeed given any real function $\varphi \in \mathcal{H}_\alpha(\mathbb{P})$ using

$$-\|\varphi\|_\infty \mathbf{1} \leq \varphi \leq \|\varphi\|_\infty \mathbf{1}$$

by positivity of $\mathcal{Q}$ it follows that

$$\big|\Lambda(\varphi)\big| \leq \Lambda(\mathbf{1})\, \|\varphi\|_\infty = \|\varphi\|_\infty .$$

This implies that Λ is also continuous over complex functions.

▪ Λ extends to positive linear functional $\tilde{\Lambda}\colon \mathcal{C}(\mathbb{P}) \to \mathbb{C}$ because by Stone-Weierstrass theorem the sub-algebra $\mathcal{H}_\alpha(\mathbb{P})$ is dense in $\mathcal{C}(\mathbb{P})$.

▪ Finally by Riesz Theorem there exists a Borel probability $\nu \in \mathrm{Prob}(\mathbb{P})$ such that $\tilde{\Lambda}(\varphi) = \int_{\mathbb{P}} \varphi\, d\nu$ for all $\varphi \in \mathcal{C}(\mathbb{P})$.

Since by definition $\mathcal{N}_\alpha$ is the kernel of Λ, one has $\mathcal{N}_\alpha = \mathcal{N}_\alpha(\nu)$. □

Next we verify that the hypothesis of Proposition 4.2 is satisfied under the assumptions of Theorem 4.1.

Lemma 4.3. *Let (A,μ) be a quasi-irreducible $\mathrm{SL}_2(\mathbb{R})$-cocycle such that $L(A) > 0$. Then*

$$\lim_{n\to+\infty} \frac{1}{n}\, \mathbb{E}_\mu \left[\, \log\|\tilde{A}^{(n)}\, p\|\, \right] = L(A)$$

with uniform convergence in $\hat{p} \in \mathbb{P}$, where $p \in \hat{p}$ is a unit vector.

Proof. Let $F \subset \Omega$ be a T-invariant Borel set with full probability, $\mathbb{P}_\mu(F) = 1$, consisting of Oseledets regular points. For any $\omega \in F$ we have the Oseledets decomposition $\mathbb{R}^2 = E^+(\omega) \oplus E^-(\omega)$ which is invariant under the cocycle action, i.e., $\tilde{A}(\omega)E^\pm(\omega) = E^\pm(T\omega)$, for all $\omega \in F$. Moreover, given $\omega \in F$ and a unit vector $p \in \mathbb{R}^2$, either $p \in E^-(\omega)$, or else

$$\lim_{n\to+\infty} \frac{1}{n}\, \log\|\tilde{A}^{(n)}(\omega)\, p\| = L(A). \tag{4.2}$$

Consider now the linear subspace

$$S := \left\{ p \in \mathbb{R}^2 \colon\ p \in E^-(\omega),\ \mathbb{P}_\mu\text{-almost surely} \right\}.$$

Since $\hat{A}(\omega_0)\, E^-(\omega) = E^-(T\omega)$ for all $\omega \in F$, it follows easily that $A(x)\, S = S$ for all $x \in \Sigma$. On the other hand, since $L(A|_S) \leq -L(A) < 0$, we must have $\dim S \leq 1$. If $\dim S = 1$ then the cocycle (A, μ) is not quasi-irreducible. Therefore $S = \{0\}$, because (A, μ) is quasi-irreducible, which in turn implies that (4.2) holds for all $p \in \mathbb{R}^2$ $\mathbb{P}_\mu$-almost surely. Because the functions $\frac{1}{n} \log\|\tilde{A}^{(n)} p\|$ are uniformly bounded by $\log\|A\|_\infty < \infty$, by the Dominated Convergence Theorem, $\frac{1}{n}\, \mathbb{E}_\mu \left[\log\|\tilde{A}^{(n)}\, p\| \right]$ converges pointwise to $L(A)$.

Assume now that this convergence is not uniform, in order to get a contradiction. This assumption implies the existence of a sequence of unit vectors $p_n \in \mathbb{R}^2$ and a positive number $\delta > 0$ such that for all large n,

$$\frac{1}{n}\, \mathbb{E}_\mu \left[\log\|\tilde{A}^{(n)}\, p_n\| \right] \leq L(A) - \delta.$$

By compactness of the unit circle we can assume that p_n converges to a unit vector $p \in \mathbb{R}^2$. We claim that $\frac{1}{n}\, \mathbb{E}_\mu \left[\log\|\tilde{A}^{(n)}\, p_n\| \right]$ converges to $L(A)$, which contradicts the previous bound. Notice that

$$\frac{\|\tilde{A}^{(n)}\, p_n\|}{\|\tilde{A}^{(n)}\|} \geq \left| p_n \cdot \overline{\mathfrak{v}}^{(n)}(\tilde{A}) \right| \to \left| p \cdot \overline{\mathfrak{v}}^{(\infty)}(\tilde{A}) \right|.$$

On the other hand, since $p \cdot \overline{\mathfrak{v}}^{(\infty)}(\tilde{A}) = 0$ is equivalent to $p \in \overline{\mathfrak{v}}^{(\infty)}(\tilde{A})^\perp = E^-(\tilde{A})$, the fact $S = \{0\}$ implies that $\mathbb{P}_\mu$-almost surely

$$\liminf_{n \to +\infty} \frac{\|\tilde{A}^{(n)}\, p_n\|}{\|\tilde{A}^{(n)}\|} > 0.$$

Therefore $\frac{1}{n} \log \frac{\|\tilde{A}^{(n)}\, p_n\|}{\|\tilde{A}^{(n)}\|}$ converges to zero $\mathbb{P}_\mu$-almost surely, and using again the Dominated Convergence Theorem

$$\begin{aligned} \lim_{n \to \infty} \frac{1}{n}\, \mathbb{E}_\mu \left[\log\|\tilde{A}^{(n)}\, p_n\| \right] &= \lim_{n \to \infty} \frac{1}{n}\, \mathbb{E}_\mu \left[\log\|\tilde{A}^{(n)}\| \right] \\ &\quad + \frac{1}{n}\, \mathbb{E}_\mu \left[\log \frac{\|\tilde{A}^{(n)}\, p_n\|}{\|\tilde{A}^{(n)}\|} \right] \\ &= L(A) + 0 = L(A) \end{aligned}$$

which establishes the claim and finishes the proof. □

Exercise 4.6. Given $g \in \mathrm{SL}_2(\mathbb{R})$ and a unit vector $p \in \mathbb{R}^2$, prove that the projective map $\hat{g} \colon \mathbb{P} \to \mathbb{P}$ has derivative

$$(\hat{g})'(\hat{p}) = \|gp\|^{-2}.$$

Hint: Use Exercise 2.2. Given unit vectors $p, v \in \mathbb{R}^2$ with $v \perp p$ notice that, because $g \in \mathrm{SL}_2(\mathbb{R})$, one has

$$1 = \|v \wedge p\| = \|(gv) \wedge (gp)\| = \|gp\| \, \|D\pi^{\perp}_{gp/\|gp\|}(gv)\|.$$

Proposition 4.4. *Given $\alpha > 0$ and unit vectors $x, y \in \mathbb{R}^2$,*

$$\left[\frac{\delta(\hat{A}(\hat{x}), \hat{A}(\hat{y}))}{\delta(\hat{x}, \hat{y})}\right]^{\alpha} \leq \frac{1}{2}\left\{\frac{1}{\|Ax\|^{2\alpha}} + \frac{1}{\|Ay\|^{2\alpha}}\right\}.$$

Proof. Given unit vectors $x, y \in \mathbb{R}^2$, we have $\|Ax \wedge Ay\| = \|x \wedge y\|$ because $A \in \mathrm{SL}_2(\mathbb{R})$. Thus

$$\begin{aligned}\left[\frac{\delta(\hat{A}(\hat{x}), \hat{A}(\hat{y}))}{\delta(\hat{x}, \hat{y})}\right]^{\alpha} &= \left[\frac{\|Ax \wedge Ay\|}{\|Ax\|\|Ay\|} \, \frac{1}{\|x \wedge y\|}\right]^{\alpha} \\ &= \frac{1}{\|Ax\|^{\alpha}} \, \frac{1}{\|Ay\|^{\alpha}} \leq \frac{1}{2}\left\{\frac{1}{\|Ax\|^{2\alpha}} + \frac{1}{\|Ay\|^{2\alpha}}\right\}\end{aligned}$$

where we have used that $\sqrt{a\,b} \leq \frac{1}{2}\,\{a + b\}$ with $a = \|Ax\|^{-2\alpha}$ and $b = \|Ay\|^{-2\alpha}$. □

Proposition 4.5. *Given a cocycle $(A, \mu) \in L^{\infty}(\Sigma, \mathrm{SL}_2(\mathbb{R})) \times \mathrm{Prob}(\Sigma)$,*

$$\kappa_{\alpha}(A, \mu) = \sup_{\hat{x} \in \mathbb{P}(\mathbb{R}^d)} \mathbb{E}_{\mu}\left[\,\|Ax\|^{-2\alpha}\,\right] \quad \textit{for all} \;\; \alpha > 0$$

where x is a unit representative of $\hat{x} \in \mathbb{P}$.

Proof. For the first inequality ($\leq$) just average the one in Proposition 4.4 and then take sup. The converse inequality ($\geq$) follows from Exercise 4.6 and the Mean Value Theorem. □

Next proposition says that the Markov operator $\mathcal{Q}_{A,\mu}$ of a quasi-irreducible cocycle with positive Lyapunov exponent acts contractively on the semi-normed space $(\mathcal{H}_{\alpha}(\mathbb{P}), v_{\alpha})$, for some small α and some large enough iterate. Moreover, this behavior is uniform in a neighborhood of A. It follows by Proposition 4.2 that the Markov operator $\mathcal{Q}_{A,\mu} \colon \mathcal{H}_{\alpha}(\mathbb{P}) \to \mathcal{H}_{\alpha}(\mathbb{P})$ is quasi-compact and simple.

Proposition 4.6. *Let* $(A, \mu) \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R})) \times \mathrm{Prob}(\Sigma)$ *be a quasi-irreducible cocycle with* $L(A) > 0$*. There are numbers* $\delta > 0$, $0 < \alpha < 1$, $0 < \kappa < 1$ *and* $n \in \mathbb{N}$ *such that for all* $B \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$ *with* $\|B - A\|_\infty < \delta$ *one has* $\kappa_\alpha(B^n, \mu^n) \le \kappa$*.*

Proof. We have

$$\begin{aligned}\lim_{n\to+\infty} \frac{1}{n}\, \mathbb{E}_{\mu^n}[\,\log\|A^n\, p\|^{-2}\,] &= \lim_{n\to+\infty} \frac{1}{n} \int_\Omega \log\|\tilde{A}^{(n)}\, p\|^{-2}\, d\mathbb{P}_\mu \\ &= -2 \lim_{n\to+\infty} \frac{1}{n} \int_\Omega \log\|\tilde{A}^{(n)} p\|\, d\mathbb{P}_\mu \\ &= -2\, L(A) < 0.\end{aligned}$$

Since $\frac{1}{n}\int_\Omega \log\|\tilde{A}^{(n)} p\|\, d\mathbb{P}_\mu$ converges uniformly in $\hat{p}$ to $L(A)$, for some n large enough we have for all unit vectors $p \in \mathbb{R}^2$

$$\mathbb{E}_{\mu^n}[\,\log\|A^n\, p\|^{-2}\,] \le -1.$$

To finish the proof, using the following inequality

$$e^x \le 1 + x + \frac{x^2}{2}\, e^{|x|}$$

we get for all unit vectors $p \in \mathbb{R}^2$,

$$\begin{aligned}\mathbb{E}_{\mu^n}\left[\,\|A^n\, p\|^{-2\alpha}\,\right] &= \mathbb{E}_{\mu^n}\left[e^{\alpha\, \log\|A^n\, p\|^{-2}}\right] \\ &\le \mathbb{E}_{\mu^n}\left[1 + \alpha\, \log(\|A^n\, p\|^{-2}) + \frac{\alpha^2}{2}\, \|A^n\|^2\, \log^2(\|A^n\, p\|^{-2})\right] \\ &\le 1 - \alpha + K\, \frac{\alpha^2}{2}\end{aligned}$$

for some positive constant $K = K(A, n)$. Thus, taking α small

$$\kappa_\alpha(A^n, \mu^n) \le \kappa := 1 - \alpha + K\, \frac{\alpha^2}{2} < 1.$$

By Proposition 4.5, the measurement $\kappa_\alpha(A, \mu)$ depends continuously on $A \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$. Therefore the above bound κ can be made uniform, i.e., valid for all cocycles in a neighborhood of A. $\square$

Definition 4.3. Given $A, B \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$ and $\mu \in \mathrm{Prob}(\Sigma)$, define

$$\Delta_\alpha(A,B) := \sup_{\hat{p}\in\mathbb{P}} \mathbb{E}_\mu \left[d(\hat{A}(\hat{p}), \hat{B}(\hat{p}))^\alpha \right].$$

Exercise 4.7. Show that Δ_α is a metric on the space $L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$ such that

$$\Delta_\alpha(A,B) \leq (\|A-B\|_\infty)^\alpha.$$

By propositions 4.2 and 4.6, any quasi-irreducible cocycle (A,μ) with $L(A) > 0$ has a unique stationary measure $\nu_A \in \mathrm{Prob}(\mathbb{P})$. Next proposition says that ν_A depends continuously on A.

Proposition 4.7. *Let $A, B \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$ and $\mu \in \mathrm{Prob}(\Sigma)$. Assume that $\kappa := \kappa_\alpha(A,\mu) < 1$ for some $0 < \alpha \leq 1$. Then for all $n \in \mathbb{N}$ and $\varphi \in \mathcal{H}_\alpha(\mathbb{P})$,*

$$\|\mathcal{Q}^n_{A,\mu}(\varphi) - \mathcal{Q}^n_{B,\mu}(\varphi)\|_\infty \leq \frac{\Delta_\alpha(A,B)}{1-\kappa}\, v_\alpha(\varphi).$$

Moreover, if also $\kappa_\alpha(B,\mu) < 1$ then for all $\varphi \in \mathcal{H}_\alpha(\mathbb{P})$,

$$\left|\int_{\mathbb{P}} \varphi\, d\nu_A - \int_{\mathbb{P}} \varphi\, d\nu_B\right| \leq \frac{\Delta_\alpha(A,B)}{1-\kappa}\, v_\alpha(\varphi).$$

Proof. First notice that

$$\begin{aligned} \|\mathcal{Q}_{A,\mu}(\varphi) - \mathcal{Q}_{B,\mu}(\varphi)\|_\infty &\leq \sup_{\hat{p}\in\mathbb{P}} \int_\Sigma \left|\varphi(\hat{A}(x)\,\hat{p}) - \varphi(\hat{B}(x)\,\hat{p}))\right| d\mu(x) \\ &\leq v_\alpha(\varphi) \sup_{\hat{p}\in\mathbb{P}} \int_\Sigma \delta(\hat{A}(x)\,\hat{p}, \hat{B}(x)\,\hat{p})^\alpha\, d\mu(x) \\ &= \Delta_\alpha(A,B)\, v_\alpha(\varphi). \end{aligned}$$

From this inequality and the relation

$$\mathcal{Q}^n_{A,\mu} - \mathcal{Q}^n_{B,\mu} = \sum_{i=0}^{n-1} \mathcal{Q}^i_{B,\mu} \circ (\mathcal{Q}_{A,\mu} - \mathcal{Q}_{B,\mu}) \circ \mathcal{Q}^{n-i-1}_{A,\mu}$$

we get

$$
\begin{aligned}
\|\mathcal{Q}^n_{A,\mu}(\varphi) - \mathcal{Q}^n_{B,\mu}(\varphi)\|_\infty &\leq \sum_{i=0}^{n-1} \|\mathcal{Q}^i_{B,\mu}(\mathcal{Q}_{A,\mu} - \mathcal{Q}_{B,\mu})(\mathcal{Q}^{n-i-1}_{A,\mu}(\varphi))\|_\infty \\
&\leq \sum_{i=0}^{n-1} \|(\mathcal{Q}_{A,\mu} - \mathcal{Q}_{B,\mu})(\mathcal{Q}^{n-i-1}_{A,\mu}(\varphi))\|_\infty \\
&\leq \sum_{i=0}^{n-1} \Delta_\alpha(A,B)\, v_\alpha(\mathcal{Q}^{n-i-1}_{A,\mu}(\varphi))) \\
&\leq \Delta_\alpha(A,B)\, v_\alpha(\varphi) \sum_{i=0}^{n-1} \kappa^{n-i-1} \\
&\leq \frac{\Delta_\alpha(A,B)}{1-\kappa}\, v_\alpha(\varphi).
\end{aligned}
$$

This proves the first inequality of the proposition. Finally, since $\lim_{n\to+\infty} \mathcal{Q}^n_{A,\mu}(\varphi) = \left(\int_{\mathbb{P}} \varphi\, d\nu_A\right) \mathbf{1}$ and $\lim_{n\to+\infty} \mathcal{Q}^n_{B,\mu}(\varphi) = \left(\int_{\mathbb{P}} \varphi\, d\nu_B\right) \mathbf{1}$, one has

$$
\Big|\int_{\mathbb{P}} \varphi\, d\nu_A - \int_{\mathbb{P}} \varphi\, d\nu_B\Big| \leq \sup_n \|\mathcal{Q}^n_{A,\mu}(\varphi) - \mathcal{Q}^n_{B,\mu}(\varphi)\|_\infty \leq \frac{\Delta_\alpha(A,B)}{1-\kappa}\, v_\alpha(\varphi)
$$

which proves the second inequality. □

We end this section with a proof of E. Le Page's theorem. This theorem also follows from Theorems 3.3 and 4.1.

Theorem 4.3 (E. Le Page)**.** *Let $(A,\mu) \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R})) \times \mathrm{Prob}(\Sigma)$ be quasi-irreducible with $L(A,\mu) > 0$. Then there are positive constants $\alpha > 0$, $C < \infty$ and $\delta > 0$ such that for all $B_1, B_2 \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$ if $\|B_j - A\|_\infty < \delta$, $j = 1, 2$, then*

$$
\big|L(B_1) - L(B_2)\big| \leq C\,(\|B_1 - B_2\|_\infty)^\alpha.
$$

Proof. Given $A \in \mathrm{SL}_2(\mathbb{R})$ let $\varphi_A \colon \mathbb{P} \to \mathbb{R}$ be the function

$$
\varphi_A(\hat{p}) := \log\|A\, p\|
$$

where $p \in \mathbb{R}^2$ stands for a unit representative of $\hat{p}$. The function $\mathrm{SL}_2(\mathbb{R}) \ni A \mapsto \varphi_A \in \mathcal{H}_1(\mathbb{P})$ is locally Lipschitz. Given $R > 0$ there is a constant $C = C_R < \infty$ such that

$$\|\varphi_A - \varphi_B\|_\infty \leq C_R \, \|A - B\|$$

for all $A, B \in \mathrm{SL}_2(\mathbb{R})$ such that $\max\{\|A\|, \|B\|\} \leq R$.

Let (A, μ) be a quasi-irreducible cocycle with $L(A) > 0$. By Proposition 4.6 there exist $n \in \mathbb{N}$, $0 < \alpha < 1$ and $0 < \kappa < 1$ such that $\kappa_\alpha(B^n, \mu^n) \leq \kappa$ for all cocycles B near A. Since the map $A \mapsto A^n$ is locally Lipschitz we can without loss of generality suppose $n = 1$.

Then, by Furstenberg's formula (4.1)

$$\begin{aligned}
\left|L(B_1) - L(B_2)\right| &\leq \mathbb{E}_\mu \left[\left| \int \varphi_{B_1} \, d\nu_{B_1} - \int \varphi_{B_2} \, d\nu_{B_2} \right| \right] \\
&\leq \mathbb{E}_\mu \left[\left| \int \varphi_{B_1} \, d\nu_{B_1} - \int \varphi_{B_1} \, d\nu_{B_2} \right| \right] \\
&\qquad + \mathbb{E}_\mu \left[\left| \int \varphi_{B_1} \, d\nu_{B_2} - \int \varphi_{B_2} \, d\nu_{B_2} \right| \right] \\
&\leq \frac{\Delta_\alpha(B_1, B_2)}{1-\kappa} \, v_\alpha(\varphi_{B_1}) + \mathbb{E}_\mu \left[\int \left| \varphi_{B_1} - \varphi_{B_2} \right| d\nu_{B_2} \right] \\
&\leq \frac{v_1(\varphi_{B_1})}{1-\kappa} \, \|B_1 - B_2\|_\infty^\alpha + C_R \, \|B_1 - B_2\|_\infty
\end{aligned}$$

where R is a uniform bound on the norms of the matrices $B_j(x)$ with $j = 1, 2$. This proves that L is locally Hölder continuous in a neighborhood of the cocycle A. □

4.3 Large deviations for sum processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $\{\xi_n \colon \Omega \to \mathbb{R}\}_{n \geq 0}$ a random stationary process, with $\mu = \mathbb{E}(\xi_n)$ for all $n \in \mathbb{N}$.

Definition 4.4. The sum process $S_n = \xi_0 + \xi_1 + \cdots + \xi_{n-1}$ is said to satisfy an LDT estimate if there exist constants $c > 0$ and $C < \infty$ such that for all small enough $\varepsilon > 0$ and $n \geq 1$,

$$\mathbb{P}\left[\left| \frac{1}{n} \, S_n - \mu \right| > \varepsilon \right] \leq C \, e^{-c\,\varepsilon^2 n}.$$

The following result is a particular version of the more general and more precise large deviation principle of H. Cramér formulated in Theorem 3.1.

Proposition 4.8. *Let* $\{\xi_n\}_{n\geq 0}$ *be a random i.i.d. process consisting of bounded random variables. Then its sum process* $S_n = \xi_0 + \xi_1 + \cdots + \xi_{n-1}$ *satisfies an LDT estimate.*

Its proof makes use of some basic ingredients, namely Chebyshev's inequality, cumulant generating functions and Legendre's transform.

Exercise 4.8 (Chebyshev's inequality)**.** Show that for any random variable $\xi\colon \Omega \to \mathbb{R}$ and any positive real numbers λ and t

$$\mathbb{P}\left[\, \left|\xi - \mathbb{E}[\xi]\right| \geq \lambda \,\right] \leq e^{-\lambda\, t}\, \mathbb{E}[\, e^{t\, \left|\xi - \mathbb{E}[\xi]\right|}\,].$$

Let $\xi : \Omega \to \mathbb{R}$ be a bounded random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The function $c_\xi\colon \mathbb{R} \to \mathbb{R}$,

$$c_\xi(t) := \log \mathbb{E}[\, e^{t\,\xi}\,]$$

is called the *second characteristic function* of ξ, also known as the *cumulant generating function* of ξ (see [43]).

Proposition 4.9. *Let* $\xi\colon \Omega \to \mathbb{R}$ *be a bounded random variable. Then*

(1) c_ξ *is an analytic convex function,*

(2) $c_\xi(0) = 0$,

(3) $(c_\xi)'(0) = \mathbb{E}(\xi)$,

(4) $c_\xi(t) \geq t\,\mathbb{E}(\xi)$, *for all* $t \in \mathbb{R}$,

Proof. For the first part of (1) notice that the boundedness of ξ implies that the parametric integral $\mathbb{E}(e^{z\,\xi})$ and its formal derivative $\mathbb{E}(e^{z\,\xi}\,\xi)$ are well-defined continuous functions on complex plane. Hence $\mathbb{E}(e^{z\,\xi})$ is an entire analytic function. Since $c_\xi(0) = \log \mathbb{E}(\mathbf{1}) = \log 1 = 0$, (2) follows. Property (3) holds because

$$(c_\xi)'(0) = \mathbb{E}(\xi\, \mathbf{1})/\mathbb{E}(\mathbf{1}) = \mathbb{E}(\xi).$$

The convexity of c_ξ follows by Hölder inequality, with conjugate exponents $p = 1/s$ and $q = 1/(1-s)$, where $0 < s < 1$. In fact, for all $t_1, t_2 \in \mathbb{R}$,

$$\begin{aligned} c_\xi(s\,t_1 + (1-s)\,t_2) &= \log \mathbb{E}[\left(e^{t_1\,\xi}\right)^s \,\left(e^{t_2\,\xi}\right)^{1-s}] \\ &\leq \log\left(\mathbb{E}[e^{t_1\,\xi}]\right)^s \left(\mathbb{E}[e^{t_2\,\xi}]\right)^{1-s} \\ &= s\,c_\xi(t_1) + (1-s)\,c_\xi(t_2)\;. \end{aligned}$$

Finally the convexity, together with (2) and (3), implies (4). □

Exercise 4.9. Let $\xi\colon \Omega \to \mathbb{R}$ be a bounded non-constant random variable. Prove that

$$(c_\xi)''(t) = \mathbb{E}_{\mathbb{P}_t}[\,\xi^2\,] - \mathbb{E}_{\mathbb{P}_t}[\,\xi\,]^2 =: \mathrm{Var}_{\mathbb{P}_t}(\xi)$$

where $\mathbb{P}_t := e^{t\,\xi}\,\mathbb{P}/\mathbb{E}[e^{t\,\xi}]$. Using Jensen's inequality, conclude that c_ξ is strictly convex.

The *Legendre transform* is an involutive non-linear operator acting on smooth strictly convex functions. Let $\mathcal{C}$ denote the space of smooth strictly convex functions $c\colon I \to \mathbb{R}$, defined on some open interval $I \subset \mathbb{R}$ and such that $c''(t) > 0$ for all $t \in I$.

Definition 4.5. Given $c \in \mathcal{C}$, its Legendre transform is the function $\hat{c} = \mathcal{L}(c)$ defined by

$$\hat{c}(\varepsilon) := \sup_{t\in I} \varepsilon\,t - c(t)\;,$$

over the interval $\hat{I} = (\hat{\varepsilon}_1, \hat{\varepsilon}_2)$ with $\hat{\varepsilon}_1 := \inf_{t\in I} c'(t)$, $\hat{\varepsilon}_2 := \sup_{t\in I} c'(t)$.

The Legendre transform is involutive in the sense that if $c \in \mathcal{C}$ then $\mathcal{L}(c) \in \mathcal{C}$ and $\mathcal{L}^2(c) = c$. (see [1, Section 14C, Chapter 3]).

Exercise 4.10. Let $c \in \mathcal{C}$ be a strictly convex function with domain I such that $0 \in \mathrm{int}(I)$ and $c(0) = c'(0) = 0$. Prove that its Legendre transform $\hat{c}$ has domain $\hat{I}$ such that $0 \in \mathrm{int}(\hat{I})$ and $\hat{c}(0) = (\hat{c})'(0) = 0$.

Exercise 4.11. Prove that the Legendre transform of the function $c(t) := \frac{h\,t^2}{2}$, defined for $t \in]-t_0, t_0[$ with $t_0, h > 0$, is the function $\hat{c}(\varepsilon) := \frac{\varepsilon^2}{2\,h}$, defined for $\varepsilon \in]-\varepsilon_0, \varepsilon_0[$ where $\varepsilon_0 = h\,t_0$.

Proof of Proposition 4.8. Let $\mu = \frac{1}{n}\,\mathbb{E}(S_n) = \mathbb{E}(\xi_m)$ for all $n, m \in \mathbb{N}$. By Chebyshev's inequality (Exercise 4.8) we have that for all $t \in \mathbb{R}$

$$\mathbb{P}[\,\frac{1}{n}\,S_n - \mu > \varepsilon\,] \leq e^{-t\,n\,\varepsilon}\mathbb{E}[\,e^{t(S_n - n\,\mu)}\,].$$

Let $c(t)$ be the cumulant generating function of $\xi_0 - \mu$. By Proposition 4.9 and Exercise 4.9, $c(t)$ is a strictly convex function such that $c(0) = c'(0) = 0$. On the other hand, because the process $\{\xi_n\}_{n\geq 0}$ is i.i.d. the sum S_n has cumulant generating function $c_{S_n}(t) = n\,c(t)$. Hence for all $\varepsilon > 0$ and $t > 0$

$$\mathbb{P}[\,\frac{1}{n}\,S_n - \mu > \varepsilon\,] \leq e^{-t\,n\,\varepsilon}\,e^{n\,c(t)} = e^{-n\,(\varepsilon\,t - c(t))}.$$

For each $\varepsilon > 0$, in order to minimize the right-hand-side upper-bound we choose $\tau(\varepsilon) := \operatorname{argmax}_t \varepsilon\,t - c(t)$. The upper-bound associated to this choice is then expressed in therms of the Legendre transform $\hat{c}(\varepsilon)$ of $c(t)$, i.e., for all $n \in \mathbb{N}$

$$\mathbb{P}[\,\frac{1}{n}\,S_n - \mu > \varepsilon\,] \leq e^{-n\,\hat{c}(\varepsilon)}.$$

By Exercise 4.10, $\hat{c}(\varepsilon)$ is a strictly convex function such that $\hat{c}(0) = (\hat{c})'(0) = 0$. In particular $0 < \hat{c}(\varepsilon) < \kappa\,\varepsilon^2$ for some $\kappa > 0$ and all $\varepsilon > 0$. A similar bound on lower deviations (below average) is driven applying the same method to the symmetric process. This proves that S_n satisfies a LDT estimate. □

Next exercise isolates the assumption on the cumulant generating function of a sum process that allows for the same conclusion as in Proposition 4.8. To understand the assumption think of $\{\xi_n\}_{n\geq 0}$ as a normalized process with zero average, so that $c_{S_n}(t)$ is a convex function with $c_{S_n}(0) = c'_{S_n}(0) = 0$.

Exercise 4.12. Let $S_n = \xi_0 + \xi_1 + \cdots + \xi_{n-1}$ be the sum of a bounded process $\{\xi_n\}_{n\geq 0}$. Assume there exists a strictly convex function $c \in \mathcal{C}$ with domain $I = (-t_0, t_0)$, $c(0) = c'(0) = 0$ and $d > 0$ such that for all $0 < t < t_0$ and $n \in \mathbb{N}$

$$c_{S_n}(t) = \log \mathbb{E}[\,e^{t\,S_n}\,] \leq d + n\,c(t). \qquad (4.3)$$

Prove that for all $\varepsilon \in \hat{I}$ (the domain of $\hat{c}$) and all $n \in \mathbb{N}$,

$$\mathbb{P}\left[\frac{1}{n} S_n > \varepsilon\right] \leq e^d \, e^{-n \, \hat{c}(\varepsilon)},$$

where $\hat{c}(\varepsilon)$ denotes the Legendre transform of $c(t)$. Moreover, if $c''(t) \leq h$ for $t \in (-t_0, t_0)$ then for all $0 \leq \varepsilon < h \, t_0$ and all $n \in \mathbb{N}$

$$\mathbb{P}\left[\frac{1}{n} S_n > \varepsilon\right] \leq e^d \, e^{-n \, \frac{\varepsilon^2}{2h}}.$$

We explain now a spectral approach due to S. V. Nagaev to derive an upper-bound on cumulant generating functions of sum processes associated to certain Markov processes. This will lead to a LDT estimate.

Let X be a compact metric space and $\mathcal{F}$ be its Borel σ-field. As before, $\mathrm{Prob}(X)$ will denote the space of Borel probability measures on X. We denote by $L^\infty(X)$ the Banach space of bounded measurable functions $\xi \colon X \to \mathbb{C}$, endowed with the usual sup-norm $\|\cdot\|_\infty$.

Definition 4.6. A *Markov kernel* is a function $K \colon X \to \mathrm{Prob}(X)$, $x \mapsto K_x$, such that for any Borel set $E \in \mathcal{F}$, the function $x \mapsto K_x(E)$ is $\mathcal{F}$-measurable. A Markov kernel K determines the following linear operator $\mathcal{Q}_K \colon L^\infty(X) \to L^\infty(X)$,

$$(\mathcal{Q}_K \phi)(x) := \int_X \phi(y) \, dK_x(y).$$

We refer to K as the kernel of $\mathcal{Q}_K$ and to $\mathcal{Q}_K$ as the Markov operator of K.

Exercise 4.13. Given a random cocycle (A, μ) with $A \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$ and $\mu \in \mathrm{Prob}(\Sigma)$, prove that $\mathcal{Q}_{A,\mu}$ is the Markov operator with kernel

$$K_{\hat{p}} := \int_\Sigma \delta_{\hat{A}(x)\, \hat{p}} \, d\mu(x).$$

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

$$K^n_x(E) := \int_X K^{n-1}_y(E) \, dK_x(y).$$

Exercise 4.14. Given a Markov kernel K, prove that for all $n \in \mathbb{N}$, K^n is the kernel of the power operator $(\mathcal{Q}_K)^n$, i.e., $(Q_K)^n = Q_{K^n}$.

Definition 4.7. Given a Markov kernel K on $(X, \mathfrak{F})$, a measure $\mu \in \mathrm{Prob}(X)$ is said to be K-*stationary* when for all $E \in \mathfrak{F}$,

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

We call *Markov system* to any pair (K, μ) where K is a Markov kernel K on $(X, \mathfrak{F})$ and $\mu \in \mathrm{Prob}(X)$ is a K-stationary probability measure.

The topological product space $X^{\mathbb{N}}$ is compact and metrizable. Its Borel σ-field $\mathcal{F}$ is generated by the *cylinders*, i.e., sets of the form

$$C(E_0, \ldots, E_m) := \{(x_j)_{j\geq 0} \in X^{\mathbb{N}} \colon x_j \in E_j \ \text{ for } \ 0 \leq j \leq m\}$$

with $E_0, \ldots, E_m \in \mathfrak{F}$. Given $\theta \in \mathrm{Prob}(X)$, the following expression determines a pre-measure over the cylinder semi-algebra on $X^{\mathbb{N}}$

$$\mathbb{P}_\theta[C(E_0, \ldots, E_m)] := {}_{E_m} \cdots \int_{E_0} d\theta(x_0) \prod_{j=1}^{m} dK_{x_{j-1}}(x_j).$$

By Carathéodory's extension theorem this pre-measure extends to a unique probability measure $\mathbb{P}_\theta$ on $(X^{\mathbb{N}}, \mathcal{F})$. This construction, due to A. Kolmogorov, is such that the process $e_n \colon X^{\mathbb{N}} \to X$ defined by $e_n\{x_j\}_{j\geq 0} := x_n$, satisfies for all $E \in \mathfrak{F}$,

1. $\mathbb{P}_\theta[\, e_0 \in E\,] = \theta(E)$,

2. $\mathbb{P}_\theta[\, e_n \in E \,|\, e_{n-1} = x\,] = K_x(E)$ for all $x \in X$ and $n \geq 1$.

By construction $\{e_n\}_{n\geq 0}$ is a Markov process with initial distribution θ and transition kernel K on the probability space $(X^{\mathbb{N}}, \mathcal{F}, \mathbb{P}_\theta)$. Any Markov process can in fact be realized in this way.

Exercise 4.15. If $\mu \in \mathrm{Prob}(X)$ is K-stationary prove that $\{e_n\}_{n\geq 0}$ is a stationary process on $(X^{\mathbb{N}}, \mathcal{F}, \mathbb{P}_\mu)$.

Given a Markov system (K, μ) we will refer to the probability measure $\mathbb{P}_\mu$ on $(X^{\mathbb{N}}, \mathcal{F})$ as the *Kolmogorov extension* of (K, μ). We call *sum process* of an observable $\xi \in L^\infty(X)$ the sum of the stationary process $\{\xi_n := \xi \circ e_n\}_{n \geq 0}$

$$\begin{aligned} S_n(\xi)(x) &:= \xi_0(x) + \xi_1(x) + \cdots + \xi_{n-1}(x) \\ &= \xi(x_0) + \xi(x_1) + \cdots + \xi(x_{n-1}), \end{aligned}$$

where $x = \{x_n\}_{n \geq 0} \in X^{\mathbb{N}}$, on $(X^{\mathbb{N}}, \mathcal{F}, \mathbb{P}_\mu)$. We aim to establish an LDT estimate for $S_n(\xi)$, at least for some subclass of observables $\xi \in L^\infty(X)$.

The spectral method we are about to explain analyzes a one parameter family of operators $\mathcal{Q}_t$ that matches the Markov operator $\mathcal{Q}_K$ for $t = 0$.

Definition 4.8. A Markov kernel K on $(X, \mathfrak{F})$ and a measurable observable $\xi \in L^\infty(X)$ determine the following family of so called *Laplace-Markov* operators $\mathcal{Q}_{K,\xi,t} \colon L^\infty(X) \to L^\infty(X)$,

$$(\mathcal{Q}_{K,\xi,t}\phi)(x) := \int_X \phi(y)\, e^{t\,\xi(y)}\, dK_x(y),$$

where t is a real or complex parameter.

Since $L^\infty(X)$ is a Banach algebra, the multiplication operator $\mathcal{M}_{e^{t\xi}} \colon L^\infty(X) \to L^\infty(X)$, $\phi \mapsto \phi\, e^{t\xi}$, is bounded. Moreover, the dependence of these operators on t is analytic. Hence, because $\mathcal{Q}_{K,\xi,t} = \mathcal{Q}_K \circ \mathcal{M}_{e^{t\xi}}$, the Laplace-Markov family is also analytic on t.

In the sequel, the probability $\mathbb{P}_\theta$ and the associated expected value $\mathbb{E}_\theta$ for a Dirac mass $\theta = \delta_x$ with $x \in X$, will be denoted by $\mathbb{P}_x$ and $\mathbb{E}_x$, respectively. Notice that $(\mathcal{Q}_{K,\xi,t}\phi)(x) = \mathbb{E}_x[\,\phi\, e^{t\xi}\,]$ and in particular $(\mathcal{Q}_{K,\xi,t}\mathbf{1})(x) = \mathbb{E}_x[\,e^{t\xi}\,]$. This relation extends inductively.

Exercise 4.16. Prove that for all $n \geq 1$,

$$(\mathcal{Q}_t^n \mathbf{1})(x) = \mathbb{E}_x\left[\, e^{t\, S_n(\xi)}\,\right]$$

where $\mathcal{Q}_t = \mathcal{Q}_{K,\xi,t}$.

To get large deviations through Nagaev's method we make a strong assumption on the Markov system (K,μ), namely that for some Banach sub-algebra $\mathcal{B} \subset L^\infty(X)$ with $\mathbf{1} \in \mathcal{B}$, $\mathcal{Q}_K \colon \mathcal{B} \to \mathcal{B}$ is a quasi-compact and simple operator.

Let $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ be a complex Banach algebra and also a *lattice*, in the sense that $\overline{f}, |f| \in \mathcal{B}$ for all $f \in \mathcal{B}$. Assume also $\mathbf{1} \in \mathcal{B}$, $\mathcal{B} \subset L^\infty(X)$ and the inclusion $\mathcal{B} \hookrightarrow L^\infty(X)$ is continuous: $\|f\|_\infty \leq \|f\|_{\mathcal{B}}$ for all $f \in \mathcal{B}$.

Definition 4.9. We say that a Markov system (K,μ) acts simply and quasi-compactly on $\mathcal{B}$ if there are constants $C < \infty$ and $0 < \sigma < 1$ such that for all $f \in \mathcal{B}$ and all $n \geq 0$,

$$\|\mathcal{Q}_K f - (\textstyle\int f \, d\mu)\, \mathbf{1}\|_{\mathcal{B}} \leq C\, \sigma^n \, \|f\|_{\mathcal{B}}.$$

Let $\mathcal{L}(\mathcal{B})$ be the Banach algebra of bounded linear operators on $\mathcal{B}$ and denote by $\|T\|_{\mathcal{B}}$ the operator norm of $T \in \mathcal{L}(\mathcal{B})$.

Theorem 4.4. *Let (K,μ) be a Markov system which acts simply and quasi-compactly on a Banach sub-algebra $\mathcal{B} \subset L^\infty(X)$ satisfying the above assumptions. Then given $\xi \in \mathcal{B}$ there are constants $\kappa, \varepsilon_0 > 0$ and $C' < \infty$ such that for all $x \in X$, $0 < \varepsilon < \varepsilon_0$ and $n \in \mathbb{N}$*

$$\mathbb{P}_x \left[\, \left| \frac{1}{n}\, S_n(\xi) - \int_X \xi \, d\mu \right| > \varepsilon \,\right] \leq C' \, e^{-n\, \kappa\, \varepsilon^2}.$$

The constants C', κ and ε_0 depend on $\|\mathcal{Q}_0\|_{\mathcal{B}}$, $\|\xi\|_{\mathcal{B}}$ and on the constants C and σ in Definition 4.9 controlling the action of $\mathcal{Q}_0$ on $\mathcal{B}$.

Proof. Given $\xi \in \mathcal{B}$, the Laplace-Markov family $\mathcal{Q}_t := \mathcal{Q}_{K,\xi,t}$ extends to an entire function $\mathcal{Q} \colon \mathbb{C} \to \mathcal{L}(\mathcal{B})$. The reason for this is the factorization $\mathcal{Q}_t = \mathcal{Q}_0 \circ \mathcal{M}_{e^{t\xi}}$. Because $\mathcal{B}$ is a Banach algebra,

$$\mathbb{C} \ni t \mapsto e^{t\xi} := \sum_{n=0}^{\infty} \frac{t^n\, \xi^n}{n!} \in \mathcal{B}$$

is an analytic function, while $f \mapsto \mathcal{M}_f$ is an isometric embedding of $\mathcal{B}$ into $\mathcal{L}(\mathcal{B})$ as a Banach sub-algebra.

By hypothesis there are constants $C_0 < \infty$ and $0 < \sigma_0 < 1$ such that for all $f \in \mathcal{B}$ and $n \in \mathbb{N}$,

$$\|\mathcal{Q}_0^n f - (\textstyle\int f \, d\mu)\, \mathbf{1}\|_{\mathcal{B}} \leq C_0\, \sigma_0^n \, \|f\|_{\mathcal{B}}.$$

It follows from this inequality that $\mathrm{spec}(\mathcal{Q}_0) \subset \{1\} \cup \mathbb{D}_{\sigma_0}(0)$ and $\mathbf{1}$ is a simple eigenvalue of $\mathcal{Q}_0$. There is a $\mathcal{Q}_0$-invariant decomposition $\mathcal{B} = E_0 \oplus H_0$ with $E_0 = \mathbb{C}\,\mathbf{1}$ and $H_0 := \{f \in \mathcal{B} \colon \int f\, d\mu = 0\}$. Moreover $P_0 \colon \mathcal{B} \to \mathcal{B}$, $P_0 f := (\int f\, d\mu)\, \mathbf{1}$, is the spectral projection associated with the eigenvalue 1.

Exercise 4.17. From the two measurements $\|\mathcal{Q}_0\|_{\mathcal{B}}$ and $\|\xi\|_{\mathcal{B}}$ find constants $M, C_1 < \infty$ such that for all $f \in \mathcal{B}$ and $|t| \leq 1$,

$$\|\mathcal{Q}_t f\|_{\mathcal{B}} \leq M\, \|f\|_{\mathcal{B}} \quad \text{and} \quad \|\mathcal{Q}_t f - \mathcal{Q}_0 f\|_{\mathcal{B}} \leq C_1\, |t|\, \|f\|_{\mathcal{B}}.$$

Using the bounds from Exercise 4.17, the spectral decomposition $\mathcal{B} = E_0 \oplus H_0$ persists for small t. More precisely, fixing $\sigma \in (\sigma_0, 1)$, say $\sigma := \frac{1+\sigma_0}{2}$, there are constants $t_0 > 0$ and $C_2 < \infty$, there are subspaces $E_t, H_t \subset \mathcal{B}$ and there are analytic functions $\mathbb{D}_{t_0}(0) \ni t \mapsto \lambda(t) \in \mathbb{C}$ and $\mathbb{D}_{t_0}(0) \ni t \mapsto P_t \in \mathcal{L}(\mathcal{B})$ such that for all $|t| < t_0$

1. $\mathcal{B} = E_t \oplus H_t$ is a $\mathcal{Q}_t$-invariant decomposition,
2. $\dim(E_t) = 1$,
3. P_t is the projection onto E_t parallel to H_t,
4. $P_t \circ \mathcal{Q}_t = \mathcal{Q}_t \circ P_t = \lambda(t)\, P_t$,
5. $\mathcal{Q}_t f = \lambda(t)\, f$ for all $f \in E_t$,
6. $\lambda(0) = 1$ and $\left|\lambda(t)\right| > (1+\sigma)/2$,
7. $\|\mathcal{Q}_t^n f - \lambda(t)^n\, P_t f\|_{\mathcal{B}} \leq C_2\, \sigma^n\, \|f\|_{\mathcal{B}}$ for all $f \in \mathcal{B}$ and $n \in \mathbb{N}$,
8. $\mathrm{spec}(\mathcal{Q}_t|_{H_t}) \subset \mathbb{D}_\sigma(0)$,
9. $\|P_t f\|_{\mathcal{B}} \leq C_2\, \|f\|_{\mathcal{B}}$ for all $f \in \mathcal{B}$,
10. $\|P_t f - P_0 f\|_{\mathcal{B}} \leq C_2\, |t|\, \|f\|_{\mathcal{B}}$ for all $f \in \mathcal{B}$.

These properties describe the continuous dependence of the spectral decomposition of the Laplace-Markov operator $\mathcal{Q}_t$ on the parameter t. Their proof uses Spectral Theory (see for instance [51, Chapter XI]). More precisely, they can be proven with a Taylor development of the Cauchy integral formula for the spectral projection P_t. See [16, Proposition 5.12].

Exercise 4.18. Track the dependence of t_0 and C_2 on the constants C_0, σ_0, $\|\mathcal{Q}_0\|_{\mathcal{B}}$ and $\|\xi\|_{\mathcal{B}}$.

Exercise 4.19. Prove that $\mathbb{R} \ni t \mapsto \lambda(t)$ is a real analytic function with strictly positive values. **Hint:** $\mathcal{Q}_t$ is a positive operator.

Exercise 4.20. Prove that $c(t) := \log \lambda(t)$ is a real analytic function such that $c(0) = 0$ and $c'(0) = \mathbb{E}_\mu(\xi)$. **Hint:** By the Implicit Function Theorem there exists an analytic function $\mathbb{D}_{t_0}(0) \ni t \mapsto f_t \in \mathcal{B}$ such that $\mathcal{Q}_t\, f_t = \lambda(t)\, f_t$ and $\int f_t\, d\mu = 1$ for all t. Differentiating this relation prove that $\lambda'(0) = \int \mathcal{Q}_0' \mathbf{1}\, d\mu = \mathbb{E}_\mu(\xi)$.

Exercise 4.21. Derive an absolute upper bound h for the second derivative of $c(t) := \log \lambda(t)$ on some compact interval contained in $(-t_0, t_0)$, e.g., $c''(t) \leq h$ for all $|t| \leq \frac{t_0}{2}$. Show explicitly the dependence of h on the constants M and t_0. **Hint:** The function $\lambda(t)$ is analytic on $\mathbb{D}_{t_0}(0)$. Use Cauchy's integral formula for $c''(t)$.

We can normalize the process ξ_n, adding up some constant term, so that it has zero average. There is no loss of generality in assuming that $\mathbb{E}_\mu(\xi) = 0$. By exercises 4.20 and 4.21, $c(0) = c'(0) = 0$, and $c(t) \leq \frac{h\,t^2}{2}$ for all $|t| < \frac{t_0}{2}$.

Using properties 7. and 10. above we have for all $|t| < t_0$

$$\begin{aligned}
\left|\mathbb{E}_x[\, e^{t\, S_n(\xi)}\,] - \lambda(t)^n\right| &= \left|(\mathcal{Q}_t^n \mathbf{1})(x) - \lambda(t)^n\right| \leq \|\mathcal{Q}_t^n \mathbf{1} - \lambda(t)^n\, \mathbf{1}\|_{\mathcal{B}} \\
&\leq \|\mathcal{Q}_t^n \mathbf{1} - \lambda(t)^n\, P_t \mathbf{1}\|_{\mathcal{B}} + \lambda(t)^n \|P_t \mathbf{1} - \mathbf{1}\|_{\mathcal{B}} \\
&\leq C_2\, \sigma^n + \lambda(t)^n \|P_t \mathbf{1} - P_0 \mathbf{1}\|_{\mathcal{B}} \\
&\leq C_2\, \sigma^n + \lambda(t)^n\, C_2\, \|\mathbf{1}\|_{\mathcal{B}}\, |t|.
\end{aligned}$$

Hence there exists $d > 0$ such that for all $|t| < \frac{t_0}{2}$ and $n \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{E}_x[\, e^{t\, S_n(\xi)}\,] &\leq e^{n\, c(t)}\, (1 + C_2\, \|\mathbf{1}\|_{\mathcal{B}}\, |t|) + C_2\, \sigma^n \\
&\leq e^{d + n\, c(t)} \leq e^{d + n\, \frac{h\, t^2}{2}}.
\end{aligned}$$

Applying Exercise 4.12, it follows that for all $0 \leq \varepsilon < h\, \frac{t_0}{2}$ and $n \in \mathbb{N}$,

$$\mathbb{P}_x[\, \frac{1}{n}\, S_n(\xi) > \varepsilon\,] \leq e^d\, e^{-n\, \frac{\varepsilon^2}{2\, h}}.$$

Combining this with the analogous inequality for $-\xi$ we get

$$\mathbb{P}_x[\, \frac{1}{n}\, |S_n(\xi)| > \varepsilon\,] \leq 2\, e^d\, e^{-n\, \frac{\varepsilon^2}{2\,h}}.$$

This shows that $S_n(\xi)$ satisfies an LDT estimate. Moreover, all constants h, t_0, $\varepsilon_0 = h\, \frac{t_0}{2}$ and d depend only on the specified measurements C_0, σ_0, $\|\mathcal{Q}_0\|_{\mathcal{B}}$ and $\|\xi\|_{\mathcal{B}}$. □

Remark 4.1. The constant $C' = 2\, e^d$ in the previous proof can be kept small provided n is large enough.

4.4 Large deviations for random cocycles

Given a measure preserving dynamical system $T\colon \Omega \to \Omega$ and a measurable cocycle $A\colon \Omega \to \mathrm{SL}_2(\mathbb{R})$, the process $L_n(A)(\omega) := \log\|A^{(n)}(\omega)\|$ is *sub-additive* in the sense that for all $n, m \in \mathbb{N}$,

$$L_{n+m}(A) \leq L_n(A) \circ T^m + L_m(A).$$

A related *additive* process over the skew-product map $\hat{F}\colon \Omega \times \mathbb{P} \to \Omega \times \mathbb{P}$, $\hat{F}(\omega, \hat{p}) := (T\omega, \hat{A}(\omega)\, \hat{p})$, can be defined as follows

$$S_n(A)(\omega, \hat{p}) := \log\|A^{(n)}(\omega)\, p\|,$$

where p is a unit representative of $\hat{p}$. The observable $\xi_A\colon \Omega \times \mathbb{P} \to \mathbb{R}$

$$\xi_A(\omega, \hat{p}) := \log\|A(\omega) p\|,$$

'watches' the one-step fiber expansion along the direction $\hat{p}$, while $S_n(A)$ is precisely the sum process associated with ξ_A.

Exercise 4.22. Prove that $S_n(A) = \sum_{j=0}^{n-1} \xi_A \circ \hat{F}^j$ for all $n \in \mathbb{N}$.

The additive process $S_n(A)$ is essential to reduce Theorem 4.1 (on LDT estimates for random cocycles) to Theorem 4.4 (with LDT estimates for additive processes).

Consider the space $L^\infty(\Sigma\times\Sigma\times\mathbb{P})$ of bounded measurable functions $\phi\colon \Sigma\times\mathbb{P}\to\mathbb{C}$. Given a function $\phi\in L^\infty(\Sigma\times\Sigma\times\mathbb{P})$ and $0<\alpha\leq 1$, define

$$\begin{aligned}
\|\phi\|_\infty &:= \sup_{x,y\in\Sigma,\,\hat{p}\in\mathbb{P}} \left|\phi(x,y,\hat{p})\right|,\\
v_\alpha(\phi) &:= \sup_{\substack{x,y\in\Sigma\\ \hat{p}\neq\hat{q}}} \frac{\left|\phi(x,y,\hat{p})-\phi(x,y,\hat{q})\right|}{\delta(\hat{p},\hat{q})^\alpha},\\
\|\phi\|_\alpha &:= v_\alpha(\phi)+\|\phi\|_\infty
\end{aligned}$$

and set

$$\mathcal{H}_\alpha(\Sigma\times\Sigma\times\mathbb{P}) := \{\,\phi\in L^\infty(\Sigma\times\Sigma\times\mathbb{P}) :\ \|\phi\|_\alpha<\infty\,\}.$$

Exercise 4.23. Prove that $(\mathcal{H}_\alpha(\Sigma\times\Sigma\times\mathbb{P}), \|\cdot\|_\alpha)$ is a Banach algebra with unity.

Given a cocycle $(A,\mu)\in L^\infty(\Sigma,\mathrm{SL}_2(\mathbb{R}))\times\mathrm{Prob}(\Sigma)$ we define the Markov operator $\mathcal{Q}_A=\mathcal{Q}_{A,\mu}\colon L^\infty(\Sigma\times\Sigma\times\mathbb{P})\to L^\infty(\Sigma\times\Sigma\times\mathbb{P})$ by

$$\mathcal{Q}_A(\phi)(x,y,\hat{p}) := \int_\Sigma \phi(y,z,\hat{A}(y)\,\hat{p})\,d\mu(z).$$

Exercise 4.24. Show that the Markov operator $\mathcal{Q}_A$ is determined by the following kernel on the compact metric space $\Sigma\times\Sigma\times\mathbb{P}$

$$K^A_{(x,y,\hat{p})} = \int_\Sigma \delta_{(y,z,\hat{A}(y)\,\hat{p})}\,d\mu(z).$$

Let $L^\infty(\Sigma\times\mathbb{P})$, resp. $\mathcal{H}_\alpha(\Sigma\times\mathbb{P})$, be the subspace of functions $\phi\in L^\infty(\Sigma\times\Sigma\times\mathbb{P})$, resp. $\phi\in\mathcal{H}_\alpha(\Sigma\times\Sigma\times\mathbb{P})$, which do not depend on the first variable.

Exercise 4.25. Prove that $\mathcal{Q}_A$ maps $L^\infty(\Sigma\times\Sigma\times\mathbb{P})$ into $L^\infty(\Sigma\times\mathbb{P})$ and hence induces an operator $\mathcal{Q}_A\colon L^\infty(\Sigma\times\mathbb{P})\to L^\infty(\Sigma\times\mathbb{P})$.

Exercise 4.26. Prove by induction in $n\in\mathbb{N}$ that for any function $\phi\in L^\infty(\Sigma\times\mathbb{P})$

$$\begin{aligned}
(\mathcal{Q}_A^n\phi)(x_0,\hat{p}) &= \int_{\Sigma^n} \phi\left(x_n, \hat{A}(x_{n-1})\cdots\hat{A}(x_1)\,\hat{A}(x_0)\,\hat{p}\right)\,d\mu^n(x_1,\cdots,x_n)\\
&= \int_{\Sigma^n} \phi\left(x_n, \widehat{A^{n-1}}\,\hat{A}(x_0)\,\hat{p}\right)\,d\mu^n(x_1,\cdots,x_n)
\end{aligned}$$

where $A^{n-1} = A^{n-1}(x_1, \ldots, x_{n-1})$

Exercise 4.27. Prove that for all $A \in \mathrm{SL}_2(\mathbb{R})$ and $\hat{p}, \hat{q} \in \mathbb{P}$,

$$\frac{1}{\|A\|^2} \leq \frac{\delta(\hat{A}\,\hat{p}, \hat{A}\,\hat{q})}{\delta(\hat{p}, \hat{q})} \leq \|A\|^2.$$

Exercise 4.28. Prove that for all $\phi \in L^\infty(\Sigma \times \mathbb{P})$ and $n \in \mathbb{N}$

$$v_\alpha(\mathcal{Q}_A^n \phi) \leq \|A\|_\infty^{2\alpha}\, \kappa_\alpha(A^{n-1}, \mu^{n-1})\, v_\alpha(\phi).$$

Conclude that $\mathcal{Q}_A$ acts simply and quasi-compactly on $\mathcal{H}_\alpha(\Sigma \times \mathbb{P})$, with stationary measure $\mu \times \nu_A$. Prove also that $\mu \times \mu \times \nu_A$ is the unique stationary measure of K^A on $\Sigma \times \Sigma \times \mathbb{P}$. **Hint:** Use the exercises 4.26 and 4.27 to prove the above inequality.

Next consider the observable $\xi_A \in L^\infty(\Sigma \times \Sigma \times \mathbb{P})$

$$\xi_A(x, y, \hat{p}) := \log\|A(x)p\|,$$

Exercise 4.29. Consider the set $\Omega \subset (\Sigma\times\Sigma\times\mathbb{P})^{\mathbb{N}}$ of (K^A-admissible) sequences $\omega = \{\omega_n\}_{n\in\mathbb{N}}$ such that for some pair of sequences $\{x_n\} \subset \Sigma$ and $\{\hat{p}_n\} \subset \mathbb{P}$ one has $\omega_n = (x_n, x_{n+1}, \hat{p}_n)$ and $\hat{p}_{n+1} = \hat{A}(x_n)\,\hat{p}_n$ for all $n \in \mathbb{N}$.

Show that Ω has full measure w.r.t. the Kolmogorov extension $\mathbb{P}_{\mu\times\mu\times\nu_A}$ of $(K^A, \mu \times \mu \times \nu_A)$.

For a K^A-admissible sequence ω

$$\begin{aligned} S_n(\xi_A)(\omega) &= \sum_{j=0}^{n-1} \xi_A(x_j, x_{j-1}, \hat{p}_j) \\ &= \sum_{j=0}^{n-1} \log\|A(x_j)\, \frac{A(x_{j-1}) \cdots A(x_1)\, A(x_0)p_0}{\|A(x_{j-1}) \cdots A(x_1)\, A(x_0)p_0\|}\| \\ &= \log\|A(x_{n-1}) \cdots A(x_1)\, A(x_0)p_0\|. \end{aligned}$$

Proof of Theorem 4.1. Let $(A,\mu) \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R})) \times \mathrm{Prob}(\Sigma)$ be a quasi-irreducible random cocyle such that $L(A) > 0$ and let $\nu_A \in \mathrm{Prob}(\mathbb{P})$ be its stationary probability measure. By Exercise 4.28, the Markov operator $\mathcal{Q}_A$ acts simply and quasi-compactly on the Banach sub-algebra $\mathcal{H}_\alpha(\Sigma\times\mathbb{P})$. Hence by Theorem 4.4 applied to the Markov system $(K^A, \mu\times\mu\times\nu_A)$ there are constants $\varepsilon_0, C, h > 0$ such that for all $0 < \varepsilon < \varepsilon_0$, $(x,\hat{p}) \in \Sigma\times\mathbb{P}$ and $n \in \mathbb{N}$,

$$\mathbb{P}_x \left[\, \big| \frac{1}{n} \, \log\|A^{(n)}\, p\| - L(A) \big| \geq \varepsilon \, \right] \leq C \, e^{-\frac{\varepsilon^2}{2\,h}\, n} \, .$$

Averaging in $x \in \Sigma$ w.r.t. μ, for all $\hat{p} \in \mathbb{P}$ we get that

$$\mathbb{P}_\mu \left[\, \big| \frac{1}{n} \, \log\|A^{(n)}\, p\| - L(A) \big| \geq \varepsilon \, \right] \leq C \, e^{-\frac{\varepsilon^2}{2\,h}\, n} \, .$$

Choose the canonical basis $\{e_1, e_2\}$ of $\mathbb{R}^2$ and consider the following norm $\|\cdot\|'$ on the space of matrices $\mathrm{Mat}_2(\mathbb{R})$, $\|M\|' := \max_{j=1,2}\|M\, e_j\|$. Since this norm is equivalent to the operator norm, we have for all $\hat{p} \in \mathbb{P}$ and $n \in \mathbb{N}$,

$$\|A^{(n)}\, p\| \leq \|A^{(n)}\| \lesssim \|A^{(n)}\|' = \max_{j=1,2}\|A^{(n)}\, e_j\| \, .$$

Thus a simple comparison of the deviation sets gives

$$\mathbb{P}_\mu \left[\, \big| \frac{1}{n} \, \log\|A^{(n)}\| - L(A) \big| \geq \varepsilon \, \right] \lesssim e^{-\frac{\varepsilon^2}{2\,h}\, n}$$

for all $0 < \varepsilon < \varepsilon_0$ and $n \in \mathbb{N}$.

To finish the proof, let us explain why this LDT estimate is uniform in a neighborhood of A. By Proposition 4.6 there are constants $0 < \alpha < 1$, $0 < \kappa < 1$ and $n \in \mathbb{N}$ such that $\kappa_\alpha(B^n, \mu^n) \leq \kappa^n$ for every cocycle B in some neighborhood $\mathcal{V}$ of A. Hence, for every $B \in \mathcal{V}$, the Markov operator $\mathcal{Q}_B$ acts simply and quasi-compactly on $\mathcal{H}_\alpha(\Sigma\times\Sigma\times\mathbb{P})$. This behavior is described in Definition 4.9. The respective constants are $\sigma = \kappa$ and $C = \kappa_\alpha(B,\mu)^n$, where n is the constant fixed above. Simple calculations show that $\|\xi_B\|_\alpha \leq \|B\|_\infty^2$ and $\|\mathcal{Q}_B\|_{\mathcal{H}_\alpha} \leq \max\{1, \kappa_\alpha(B,\mu)\}$. Therefore, since these upper-bound measurements depend continuously on B, the LDT parameters from Theorem 4.4 can be kept constant in the neighborhood $\mathcal{V}$. $\square$

Proof of Theorem 4.2. This theorem follows as an application of the ACT (Theorem 3.3). Assumptions (i) and (ii) of the ACT are automatically satisfied while hypothesis (iii) of the ACT holds by Theorem 4.1. The conclusions of Theorem 4.2 follow then by Theorem 3.3. □

4.5 Consequence for Schrödinger cocycles

The goal of this section is to discuss the applicability of Theorem 4.1 on LDT estimates and of Theorem 4.2 on the continuity of the LE and of the Oseledets splitting to random Bernoulli Schrödinger cocycles.

Let $\mu \in \mathrm{Prob}(\mathbb{R})$ be a probability measure with compact support, set $\Sigma := \mathrm{supp}(\mu)$ and consider the function $A\colon \mathbb{R} \times \Sigma \to \mathrm{SL}_2(\mathbb{R})$

$$A_E(x) = A(E, x) := \begin{bmatrix} x - E & -1 \\ 1 & 0 \end{bmatrix}.$$

For each $E \in \mathbb{R}$, the pair (μ, A_E) determines a random cocycle which is also a Schrödinger cocycle as defined in Example 1.8.

Proposition 4.10. *If* $\mathrm{supp}(\mu)$ *contains more than one point then the Schrödinger cocycle* (A, μ) *satisfies* $L(A) > 0$.

Proof. See [13, Theorem 4.3], or else Exercise 4.30 below. □

The proof of this fact uses on the following classical theorem of H. Furstenberg [22].

Theorem 4.5. *Given* $\mu \in \mathrm{Prob}(\Sigma)$ *and* $A \in L^\infty(\Sigma, \mathrm{SL}_2(\mathbb{R}))$ *assume:*

(a) The subgroup generated by the set of matrices $\{A(x)\colon x \in \Sigma\}$ *is not compact.*

(b) There is no finite subset $L \subset \mathbb{P}(\mathbb{R}^2)$, $L \neq \emptyset$, *such that for all* $x \in \Sigma$, $A(x)L = L$.

Then $L(A) > 0$.

Proof. We refer the reader to the book [60, Theorem 6.11] for the proof of this statement. □

Exercise 4.30. Given the family of $\mathrm{SL}_2(\mathbb{R})$ matrices $M_x := \begin{bmatrix} x & -1 \\ 1 & 0 \end{bmatrix}$ show that:

(a) $M_a\, M_b^{-1} = \begin{bmatrix} 1 & a-b \\ 0 & 1 \end{bmatrix}$ for any $a, b \in \mathbb{R}$.

(b) $M_a^{-1}\, M_b = \begin{bmatrix} 1 & 0 \\ a-b & 1 \end{bmatrix}$ for any $a, b \in \mathbb{R}$.

(c) If $\#\mathrm{supp}(\mu) > 1$ then the subgroup generated by $\{A_E(x) \colon x \in \mathrm{supp}(\mu)\}$ is not compact.

(d) There is no finite subset $\emptyset \neq L \subset \mathbb{P}(\mathbb{R}^2)$ such that $\hat{M}_a\, \hat{M}_b^{-1}\, L = L$ and $\hat{M}_a^{-1}\, \hat{M}_b\, L = L$ for some pair of real numbers $a \neq b$.

(e) $L(A_E, \mu) > 0$, for all $E \in \mathbb{R}$.

Proposition 4.11. *If* $\mathrm{supp}(\mu)$ *contains more than one point then the Lyapunov exponent* $L(E) := L(A_E)$ *and the Oseledets splitting components* $\mathcal{E}^\pm_E := \mathcal{E}^\pm_{A_E}$ *are Hölder continuous functions of* E*.*

Proof. Follows from Proposition 4.10, Theorem 4.1 and Corollary 3.1, or, alternatively, from Proposition 4.10 and Theorem 4.2. □

4.6 Bibliographical notes

The study of random cocycles goes back to the seminal work of H. Furstenberg [22] where the positivity criterion in Theorem 4.5 was established for $\mathrm{GL}_d(\mathbb{R})$-cocycles. Since then, the scope of Furstenberg's theory has been greatly extended, namely with similar criteria for the simplicity of the LE [50, 27].

Regarding the continuity of the LE of random linear cocycles, the first result was established by H. Furstenberg and Y. Kifer [21] for generic (irreducible and contracting) $\mathrm{GL}_d(\mathbb{R})$-cocycles. For random cocycles generated by finitely many matrices, Y. Peres [48] proved the analyticity of the top LE as a function of the probability vector. The regularity of the top LE with respect to the matrices is much more subtle. On one hand a theorem of Ruelle [52] shows that this dependence on the matrices is analytic for uniformly hyperbolic cocycles, but on the other hand an example of a random Schrödinger cocycle due to B. Halperin (see Simon-Taylor [54]) shows that this modulus of continuity is in general not better than Hölder. In [39] E. Le Page proved the Hölder continuity of the top LE (the analogue of Theorem 4.3) for irreducible $\mathrm{GL}_d(\mathbb{R})$-cocycles with a gap between the first two LE. The full continuity, without any generic assumption, for general $\mathrm{GL}_2(\mathbb{R})$-cocycles was established by C. Bocker-Neto and M. Viana in [6]. The analogue of this result for random $\mathrm{GL}_d(\mathbb{R})$-cocycles has been announced by A. Avila, A. Eskin and M. Viana (see [60, Note 10.7]. An extension of [6] to a particular type of cocycles over Markov systems (particular in the sense that the cocycle still depends on one coordinate, as in the Bernoulli case) was obtained by E. Malheiro and M. Viana in [44]. For the interested reader, a general one-stop reference for continuity results for random cocycles is M. Viana's monograph [60].

We provide now a few notes on large deviation and other limit theorems for Markov processes in general and for random cocycles in particular. In [46] S. V. Nagaev proved a central limit theorem for stationary Markov chains. In his approach Nagaev uses the spectral properties of a quasi-compact Markov operator acting on some space of bounded measurable functions. Nagaev's approach was used by V. Tutubalin [59] to establish the first central limit theorem in the context of random cocycles. This method was used by E. Le Page to obtain more general central limit theorems, as well as a large deviation principle [38]. Later P. Bougerol extended Le Page's approach, proving similar results for Markov type random cocycles [8]. The book of P. Bougerol and J. Lacroix [9], on random i.i.d. products of matrices, is an excellent introduction on the subject in [38, 8]. More recentely, the book of H. Hennion and L. Hervé [29] describes a powerful abstract setting where the method of Nagaev can be applied to derive limit theorems.

These notes are based in our manuscript [16] where we established *uniform* LDT estimates and continuity of the LE for strongly mixing Markov cocycles. We remark that the known large deviation principles, in the context of random cocycles (see [7, 9, 29]), did not provide the required uniformity in the LDT estimates. The presentation here is significantly shortened because of our simpler setting: we consider bounded measurable random (Bernoulli) $\mathrm{SL}_2(\mathbb{R})$-cocycles. The proof of E Le Page's theorem (Theorem 4.3) is taken from [5].

Chapter 5

Quasi-periodic Cocycles

5.1 Introduction and statement

Let $\mathbb{T}$ be the one dimensional torus, regarded as the additive group $\mathbb{R}/\mathbb{Z}$, with the interval $[0,1)$ chosen as a model for this quotient space. Moreover, any function $f\colon \mathbb{T} \to \mathbb{C}$ may be identified with its 1-periodic lifting $f^\sharp\colon \mathbb{R} \to \mathbb{C}$.

Through these identifications we endow $\mathbb{T}$ with a probability Borel measure denoted by $|\cdot|$, namely the Lebesgue measure from $\mathbb{R}$ restricted to $[0,1)$. Furthermore, given a function f on $\mathbb{T}$, concepts such as continuity or differentiability correspond to the continuity or differentiability of the periodic lifting $f^\sharp$ on $\mathbb{R}$.

We will stop distinguishing between $x + \mathbb{Z} \in \mathbb{R}/\mathbb{Z}$ and $x \in \mathbb{R}$, or between f on $\mathbb{T}$ and $f^\sharp$ on $\mathbb{R}$, and instead let x or f respectively refer to either, depending on the context.

Let $L^1(\mathbb{T})$ be the space of all measurable functions $f\colon \mathbb{T} \to \mathbb{C}$ that are absolutely integrable with respect to the measure $|\cdot|$, that is,

$$\int_{\mathbb{T}} |f(t)|\, dt < \infty\,.$$

We will call any $f \in L^1(\mathbb{T})$ an *observable*.

Occasionally we use the notation $\langle f \rangle := \int_{\mathbb{T}} f$ for the integral of an observable f, and refer to this number also as the mean of f.

Let $\omega \in \mathbb{R} \setminus \mathbb{Q}$ be a number which we call *frequency*, and define the transformation $T = T_\omega \colon \mathbb{T} \to \mathbb{T}$,

$$Tx := x + \omega \mod \mathbb{Z}$$

to be the translation by ω.

Then $(\mathbb{T}, |\cdot|, T)$ is an ergodic MPDS, called the torus translation.

The results in this chapter require an arithmetic assumption[1] on the frequency. That is, we will assume that ω is not merely irrational, but it is not well approximated by rationals with small denominators.

Definition 5.1. We say that a frequency $\omega \in \mathbb{T}$ satisfies a Diophantine condition if

$$\|k\omega\| := \operatorname{dist}(k\omega, \mathbb{Z}) \geq \frac{\gamma}{|k| \, (\log |k|)^2} \tag{5.1}$$

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

It can be shown that the set of frequencies satisfying this condition has measure $1 - \mathcal{O}(\gamma)$, hence almost every frequency $\omega \in \mathbb{T}$ will satisfy (5.1) for some $\gamma > 0$.

The estimates derived in this chapter (e.g. the LDT) will not depend on the frequency ω per se, but on the constant γ. Since ω will be fixed throughout, we will not emphasize that dependence on γ and in fact for simplicity drop it from notations.

All throughout this chapter, if $x \in \mathbb{R}$, we will use the notation

$$e(x) := e^{2\pi i x} \in \mathbb{S},$$

where $\mathbb{S}$ is the unit circle regarded as a subset (and multiplicative subgroup) of $\mathbb{C}$, the complex plane. This defines a 1-periodic surjective function $e \colon \mathbb{R} \to \mathbb{S} \subset \mathbb{C}$, with the property that $e(x) = e(y)$ if and only if $x - y \in \mathbb{Z}$.

Hence e induces an isomorphism $\mathbb{T} \ni x + \mathbb{Z} \mapsto e(x) \in \mathbb{S}$. This map is also continuous, measure preserving and it conjugates the torus translation T_ω with the circle rotation (by angle $2\pi\omega$).

[1]This type of assumption is sufficient for our purposes. It is known at this point, but just as a folklore result, that some type of arithmetic assumption is necessary for the kind of results obtained in this chapter.

Therefore, when convenient, via $x \mapsto e(x)$, we will identify $\mathbb{T}$ with $\mathbb{S}$ as metric spaces, probability spaces and MPDS (relative to the translation and respectively rotation).

Furthermore, any function $f\colon \mathbb{T} \to \mathbb{C}$ is identified, when need be, with the function $f^\flat\colon \mathbb{S} \to \mathbb{C}$, $f^\flat(e(x)) = f(x)$, and again, we will stop distinguishing in notations f from $f^\flat$.

Let $L^2(\mathbb{T})$ be the space of all square integrable functions on $\mathbb{T}$. Endowed with the inner product

$$\langle f, g \rangle = \int_{\mathbb{T}} f \, \overline{g} \,,$$

$L^2(\mathbb{T})$ is a Hilbert space. The (multiplicative) characters $e_n\colon \mathbb{T} \to \mathbb{C}$, $e_n(x) := e(nx) = e^{2\pi i n x}$, form an orthonormal basis. The coefficients $\widehat{f}(n)$ of a function $f \in L^2(\mathbb{T})$ relative to this basis are called the *Fourier coefficients* of f, so

$$\widehat{f}(n) := \int_{\mathbb{T}} f(t) \, \overline{e_n(t)} \, dt = \int_{\mathbb{T}} f(t) \, e(-nt) \, dt \,.$$

Thus any $f \in L^2(\mathbb{T})$ may be expanded into the Fourier series $f(x) = \sum_{n\in\mathbb{Z}} \widehat{f}(n) \, e_n(x)$, where the convergence of the infinite sum and the equality are understood in the L^2 sense.

Through the identification $\mathbb{T} \equiv \mathbb{S} \subset \mathbb{C}$, an observable on $\mathbb{T}$ may be afforded additional analytic properties.

For example, we say that a function $f = f(x)$ on $\mathbb{T}$ has a *holomorphic extension* to a domain Ω, where $\mathbb{T} \subset \Omega \subset \mathbb{C}$, if there is holomorphic function from Ω to $\mathbb{C}$, whose values on $\mathbb{T}$ are those of $f(z)$. By the interior uniqueness theorem for holomorphic functions, if such a function exists, it is unique, and we denote it by $f = f(z)$.

Any *real analytic* function on $\mathbb{T}$ (i.e. a function $f\colon \mathbb{T} \to \mathbb{R}$ that expands as a power series locally near any point) has a holomorphic extension to a complex domain $\Omega \supset \mathbb{T}$. Any such domain contains an annulus of a certain width around $\mathbb{T}$.

Let $\rho > 0$ and let

$$\mathcal{A} = \mathcal{A}_\rho := \{z \in \mathbb{C}\colon 1 - \rho < |z| < 1 + \rho\}$$

be the annulus of width ρ around the torus $\mathbb{T}$. This domain (hence its width) will be fixed once and for all.

Definition 5.2. We define $C^\omega_\rho(\mathbb{T}, \mathbb{R})$ to be the set of all real analytic functions $f\colon \mathbb{T} \to \mathbb{R}$ that have a holomorphic extension to the annulus $\mathscr{A}_\rho$, extension which is continuous up to the boundary of the annulus.

Endowed with the norm

$$\|f\|_\rho := \sup_{z \in \mathscr{A}_\rho} |f(z)| \, ,$$

the set $C^\omega_\rho(\mathbb{T}, \mathbb{R})$ is a Banach space.

Exercise 5.1. Show that $(C^\omega_\rho(\mathbb{T}, \mathbb{R}), \|\cdot\|_\rho)$ is indeed a Banach space.

We are finally ready to introduce the space of *analytic quasi-periodic cocycles.*

Definition 5.3. We define $C^\omega_\rho(\mathbb{T}, \mathrm{SL}(2, \mathbb{R}))$ to be the set of all functions $A\colon \mathbb{T} \to \mathrm{SL}_2(\mathbb{R})$ that have a holomorphic extension to $\mathscr{A}_\rho$, which is continuous up to the boundary.

By holomorphicity of a matrix valued function, we simply understand the holomorphicity of each of its entries.

That is, $A = (a_{ij}) \in C^\omega_\rho(\mathbb{T}, \mathrm{SL}(2, \mathbb{R}))$ means $a_{ij} \in C^\omega_\rho(\mathbb{T}, \mathbb{R})$ for all $1 \le i, j \le 2$.

We define on $C^\omega_\rho(\mathbb{T}, \mathrm{SL}(2, \mathbb{R}))$ the distance:

$$d(A, B) = \|A - B\|_r := \sup_{z \in \mathscr{A}_\rho} \|A(z) - B(z)\| \, .$$

With this distance, $C^\omega_\rho(\mathbb{T}, \mathrm{SL}(2, \mathbb{R}))$ is a complete metric space.

It is in fact a closed subspace of the Banach space of functions $A\colon \mathbb{T} \to \mathrm{Mat}_2(\mathbb{R})$ having a holomorphic extension to $\mathscr{A}_\rho$.

Exercise 5.2. Verify the assertions formulated above about the metric on $C^\omega_\rho(\mathbb{T}, \mathrm{SL}(2, \mathbb{R}))$.

We fix a frequency ω, consider the translation by ω on the torus $\mathbb{T}$ and regard the matrix valued functions $A \in C^\omega_\rho(\mathbb{T}, \mathrm{SL}(2, \mathbb{R}))$ as linear cocycles over this translation. Thus the space of analytic quasi-periodic cocycles is identified with $C^\omega_\rho(\mathbb{T}, \mathrm{SL}(2, \mathbb{R}))$.

We may now formulate the main results of this chapter.

Theorem 5.1. *Let $A \in C^\omega_\rho(\mathbb{T}, \mathrm{SL}(2,\mathbb{R}))$ be a quasi-periodic cocycle, let $\omega \in \mathbb{T}$ be a frequency satisfying the Diophantine condition* (5.1) *and let $C < \infty$ be a constant such that $\log\|A\|_r < C$.*

For every small $\epsilon > 0$ there is $\bar{n} = \bar{n}(\epsilon, C) \in \mathbb{N}$ such that for all $n \geq \bar{n}$,

$$\left| \left\{ x \in \mathbb{T} \colon \left| \frac{1}{n} \log\|A^{(n)}(x)\| - L^{(n)}(A) \right| > \epsilon \right\} \right| < e^{-c\,\epsilon^2\, n}, \tag{5.2}$$

where $c = \mathcal{O}\big(\frac{1}{C}\big)$.

Remark 5.1. We note that since the LDT parameters $\bar{n}$ and c only depend on the uniform constant C, (5.2) is a *uniform* LDT estimate.

We also note the fact that unlike in the random case, the positivity of the Lyapunov exponent is not needed in order to obtain an LDT estimate for such quasi-periodic cocycles.

Theorem 5.2. *Assume that $\omega \in \mathbb{T}$ satisfies the Diophantine condition* (5.1)*. Then the Lyapunov exponent*

$$L \colon C^\omega_\rho(\mathbb{T}, \mathrm{SL}(2,\mathbb{R})) \to \mathbb{R}$$

is a continuous function.

Furthermore, if $A \in C^\omega_\rho(\mathbb{T}, \mathrm{SL}(2,\mathbb{R}))$ with $L(A) > 0$, then there is a neighborhood $\mathcal{V}$ of A in $C^\omega_\rho(\mathbb{T}, \mathrm{SL}(2,\mathbb{R}))$ such that:

1. *The Lyapunov exponent L is Hölder continuous on $\mathcal{V}$.*
2. *The components $\mathcal{E}^\pm \colon \mathcal{V} \to L^1(\mathbb{T}, \mathbb{P}(\mathbb{R}^2))$, $A \mapsto \mathcal{E}^\pm_A$ of the Oseledets splitting are Hölder continuous.*

5.2 Staging the proof

For any observable $f \colon \mathbb{T} \to \mathbb{R}$ and integer $n \in \mathbb{N}$, let

$$S_n f(x) := f(x) + f(Tx) + \ldots + f(T^{n-1}x)$$

be its n-th Birkhoff sum.

For any cocycle $A \in C^\omega_\rho(\mathbb{T}, \mathrm{SL}(2,\mathbb{R}))$ and integer $n \in \mathbb{N}$ consider the function on the torus $\mathbb{T}$,

$$u^{(n)}_A(x) := \frac{1}{n} \log\|A^{(n)}(x)\|\,.$$

Since $A(x)$ has the holomorphic extension $A(z)$ to $\mathscr{A}$, the iterates $A^{(n)}(x)$ also extend holomorphically to $\mathscr{A}$. Therefore, each function $u_A^{(n)}(x)$ has the *subharmonic* extension to $\mathscr{A}$:

$$u_A^{(n)}(z) = \frac{1}{n} \log\|A^{(n)}(z)\| .$$

It is easy to see that the functions $u_A^{(n)}(z)$ are uniformly bounded in z, n and even A. The subharmonicity and the uniform boundedness of these functions play a crucial rôle in the derivation of the LDT estimates. Let us explain this point.

The functions $u_A^{(n)}(x)$ are *almost invariant* under the base transformation T in the sense that

$$\left|u_A^{(n)}(x) - u_A^{(n)}(Tx)\right| \leq \frac{C}{n} ,$$

for some constant $C = C(A) < \infty$ and for all $n \in \mathbb{N}$.

This almost invariance property implies, via the triangle inequality, that for all $j \in \mathbb{N}$,

$$\left|u_A^{(n)}(x) - u_A^{(n)}(T^j x)\right| \leq \frac{Cj}{n} .$$

Thus if R is an integer with $R = o(n)$, for all $x \in \mathbb{T}$ we have

$$\left|u_A^{(n)}(x) - \frac{1}{R} S_R u_A^{(n)}(x)\right| \leq \frac{CR}{n} = o(1) . \tag{5.3}$$

By Birkhoff's ergodic theorem, as $R \to \infty$, the averages $\frac{1}{R} S_R f(x)$ of an observable $f \in L^1(\mathbb{T})$ converge pointwise almost everywhere to the mean $\int_{\mathbb{T}} f$, so

$$\frac{1}{R} S_R u_A^{(n)}(x) \to \int_{\mathbb{T}} u_A^{(n)} = L^{(n)}(A) \quad \text{ as } R \to \infty \;\text{ for a.e. } x \in \mathbb{T} .$$

Our goal is to establish an LDT estimate for the cocycle A, that is, an estimate of the form

$$\left|\{x \in \mathbb{T} \colon \left|u_A^{(n)}(x) - \int_{\mathbb{T}} u_A^{(n)}\right| > \epsilon\}\right| < \iota(\epsilon, n) ,$$

where $\iota(\epsilon, n)$ decays rapidly (e.g. exponentially) as $n \to \infty$.

Because of the almost invariance (5.3), this task then reduces to proving a *quantitative* version of the Birkhoff ergodic theorem, one that only depends on some measurement on the observable and not on the observable per se.

More precisely, we need a statement of the form

$$\left|\{x \in \mathbb{T} \colon \left|\frac{1}{R} S_R u(x) - \int_{\mathbb{T}} u\right| > \epsilon\}\right| < \iota(\epsilon, n)\,, \tag{5.4}$$

which should apply with the *same* rate function ι to *all* observables $u = u_A^{(n)}$, corresponding to the iterates $A^{(n)}$, $n \in \mathbb{N}$ of the cocycle A. In fact, in order to derive a uniform LDT, (5.4) should apply, with the same rate ι to observables $u = u_B^{(n)}$, corresponding to the iterates $B^{(n)}$ of all cocycles B in a small neighborhood of A.

The functions $u_B^{(n)}(x) := \frac{1}{n} \log\|B^{(n)}(x)\|$ have bounded subharmonic extensions to $\mathscr{A}$, and it is not hard to see that the bound is uniform in B and n.

Thus it is sufficient to establish a quantitative Birkhoff ergodic theorem (qBET) like (5.4) for any observable u with a subharmonic extension to $\mathscr{A}$ that is bounded by a constant C, where the rate function ι depends only on this bound C.[2]

The derivation of this qBET, which we obtain in Theorem 5.6,[3] is the core of this chapter, and it is achieved by using various concepts and results in harmonic analysis, potential theory and analytic number theory. We give a hint of what is to come.

Expand the observable u into a Fourier series. Since the mean of a function is its zeroth Fourier coefficient, subtracting the mean we have that

$$u(x) - \int_{\mathbb{T}} u = \sum_{k \neq 0} \widehat{u}(k)\, e(kx)\,,$$

so taking the Birkhoff averages, we have

$$\frac{1}{R} S_R u(x) - \int_{\mathbb{T}} u = \sum_{k \neq 0} \widehat{u}(k)\, \frac{1}{R} S_R\, e(kx)\,.$$

[2]The rate ι will also depend on the frequency ω (more precisely, on its arithmetic properties). However, ω is fixed.

[3]In fact, instead of the usual Birkhoff averages, we will consider higher order averages.

Therefore, in order to estimate the sum of the Fourier modes above, we need the following two ingredients.

1. A rate of decay of the Fourier coefficients of u. This will be established in Theorem 5.5 using harmonic analysis and potential theory tools, chief amongst them, the Riesz representation theorem. We will review most of the concepts needed. However, at the first perusal of this chapter, the reader may just accept the validity of the decay (5.10) and move on with the rest of the argument.

2. An estimate on the Birkhoff averages $\frac{1}{R} S_R\, e_k(x)$ of the characters $e_k(x) = e(kx)$. This is precisely where the arithmetic condition on the frequency ω is used.

5.3 Uniform measurements on subharmonic functions

As already mentioned, subharmonic functions play a crucial rôle in the derivation of the LDT estimates for quasi-periodic cocycles. We will derive some measurements on a subharmonic function that depend only on its uniform bound and on its domain.

We begin with a review of some notions in potential theory, more specifically: the Green's function and the harmonic measure of a domain, subharmonic functions and their basic properties. Standard reference books on this topic are [28, 41] and [24]. For an easier dive into this subject consider [23].

All throughout this section, $\Omega \subset \mathbb{C}$ will be a bounded domain with analytic boundary.[4]

We denote by $g(z, \zeta) = g(z, \zeta; \Omega)$ the *Green's function* for Ω with pole at ζ. Let us recall the basic properties of the Green's function that are needed here (see [23, Chapter XV.6]).

For every $\zeta \in \Omega$, the function $z \mapsto g(z, \zeta)$ is harmonic on $\Omega \setminus \{\zeta\}$, and

$$H(z, \zeta) := g(z, \zeta) + \log |z - \zeta|$$

[4] We say that a domain has analytic boundary if its boundary consists of a finite number of disjoint, simple, closed analytic curves. For instance, any disk, and also any annulus, has analytic boundary.

is harmonic at ζ, hence everywhere on Ω, as a function of z.

Moreover, $H(z,\zeta)$ is smooth on $\Omega \times \Omega$.

Furthermore, the Green's function $g(z,\zeta)$ is smooth on the set $\{(z,\zeta) \in \Omega \times \Omega\colon z \neq \zeta\}$, and $g(z,\zeta) = g(\zeta, z)$.

Finally, $g(z,\zeta) > 0$ on $\Omega \times \Omega$ and $g(z,\zeta) \to 0$ as $z \to \partial\Omega$.

We note that for every $\zeta \in \Omega$, $z \mapsto H(z,\zeta)$ is in fact the (Perron) solution to the Dirichlet problem on Ω with the continuous function $\partial\Omega \ni z \mapsto \log|z-\zeta|$ as boundary value.

Given a domain Ω as above and $z \in \Omega$, the *harmonic measure* at z with respect to Ω is a Borel probability measure $\nu_z = \nu_{z,\Omega}$ on $\partial\Omega$ such that for every Borel set $E \subset \partial\Omega$, the function $z \mapsto \nu_z(E)$ is harmonic on Ω and if f is continuous on $\partial\Omega$, then

$$H_f(z) := \int_{\partial\Omega} f(\zeta) d\nu_z(\zeta) \tag{5.5}$$

is the solution to the Dirichlet problem on Ω with boundary value f.

In probabilistic terms, $\nu_z(E)$ represents the probability that a Brownian motion which started inside the domain Ω at z, exits Ω through the subset $E \subset \partial\Omega$.

We also note that the harmonic measure and the Green's function of a domain Ω are related by the formula

$$\nu_z = -\frac{1}{2\pi}\frac{\partial g_z}{\partial n}\, ds\,,$$

where $\frac{\partial g_z}{\partial n}$ is the exterior normal derivative of $g_z(\zeta) = g(z,\zeta)$.

This formula is a direct consequence of (5.5) and of Green's third identity (which can be found in [23]).

We now review the notion of *subharmonic function.*

Definition 5.4. A function $u\colon \Omega \to [-\infty, \infty)$ is called subharmonic in the domain $\Omega \subset \mathbb{C}$ if for every $z \in \Omega$, u is upper semicontinuous [5] at z and it satisfies the sub-mean value property:

$$u(z) \leq \int_0^1 u(z + re(\theta))d\theta\,,$$

for some $r_0(z) > 0$ and for all $r \leq r_0(z)$.

[5] All of our subharmonic functions will be finite, i.e. $u\colon \Omega \to \mathbb{R}$, nonnegative and continuous rather than just upper semicontinuos.

Basic examples of subharmonic functions are $\log |z - z_0|$ or more generally, $\log |f(z)|$ for some analytic function $f(z)$ or $\int \log |z - \zeta| \, d\mu(\zeta)$ for some positive measure with compact support in $\mathbb{C}$.

The maximum of a finite collection of subharmonic functions is subharmonic, while the supremum of a collection (not necessarily finite) of subharmonic functions is subharmonic provided it is upper semicontinuous. In particular this implies that if $A \colon \Omega \to \mathrm{SL}_2(\mathbb{R})$ is a matrix valued analytic function, then

$$u(z) := \log\|A(z)\| = \sup_{\|v\|, \|w\| \leq 1} \log\left|\langle A(z)\, v, w \rangle\right|$$

is subharmonic in Ω.

Note that on every compact set $K \subset \Omega$, the function $u(z)$ defined above has the bounds

$$0 \leq u(z) \leq \log \sup_K \|A(z)\| \,.$$

A fundamental result in the theory of subharmonic functions is the Riesz representation theorem, which we formulate below.

Theorem 5.3. *Let $u \colon \Omega \to \mathbb{R}$ be a subharmonic function. There is a unique Borel measure μ on Ω called the Riesz measure of u, such that for every compactly contained subdomain $\Omega' \Subset \Omega$,*

$$u(z) = \int_{\Omega'} \log |z - \zeta| \, d\mu(\zeta) + h(z) \,,$$

where $h(z)$ is a harmonic function on Ω'.

We conclude this summary with a general version of the Poisson-Jensen formula, see [28, Theorem 3.14].

Theorem 5.4. *Let $u \colon \Omega \to \mathbb{R}$ be a subharmonic function and let $\Omega_0 \Subset \Omega$. For every $z \in \Omega_0$ we have*

$$u(z) = \int_{\partial\Omega_0} u(\zeta) d\nu_z(\zeta) - \int_{\Omega_0} g(z, \zeta) d\mu(\zeta) \,,$$

where $d\nu_z$ is the harmonic measure at z w.r.t. Ω_0, $g(z, \zeta)$ is the Green's function of Ω_0 and μ is the Riesz measure of u in Ω.

The goal of this section is to prove that if $u(z)$ is bounded, then the total mass of its Riesz measure (which we call the Riesz mass of u) and the L^∞ norm of the harmonic function h are bounded by a constant that depends only on the bound on u and on the domain Ω.

Proposition 5.1. *Let $u\colon \Omega \to \mathbb{R}$ be a subharmonic function. Assume that*

$$|u(z)| \le C \quad \text{for all } z \in \Omega\,.$$

Consider the Riesz representation of u on the subdomain $\Omega' \Subset \Omega$

$$u(z) = \int_{\Omega'} \log|z-\zeta|\, d\mu(\zeta) + h(z)\,.$$

Let $\Omega'' \Subset \Omega'$ be another subdomain.

There is a constant $C(\Omega, \Omega', \Omega'') < \infty$ such that

$$\mu(\Omega') + \|h\|_{L^\infty(\Omega'')} \le C(\Omega, \Omega', \Omega'')\, C\,.$$

Proof. We will adapt the proof of [26, Lemma 2.2].

Consider another domain Ω_0 such that $\Omega' \Subset \Omega_0 \Subset \Omega$. Let $g(z,\zeta)$ be the Green's function for Ω_0 and let ν_z be the harmonic measure at z w.r..t. Ω_0. By the Poisson-Jensen formula in Theorem 5.4 above, for all $z \in \Omega_0$ we have

$$u(z) = \int_{\partial\Omega_0} u(\zeta) d\nu_z(\zeta) - \int_{\Omega_0} g(z,\zeta) d\mu(\zeta)\,. \tag{5.6}$$

Since $g(z,\zeta) > 0$ on $\Omega_0 \times \Omega_0$, and Ω' is compactly contained in Ω_0,

$$\inf_{(z,\zeta)\in\Omega'\times\Omega'} g(z,\zeta) =: c_1 > 0\,.$$

Note that c_1 is a constant that only depends on the domains Ω_0 and Ω', hence on Ω and Ω'.

From (5.6), for all $z \in \Omega'$, we then get

$$\begin{aligned} c_1\, \mu(\Omega') &\le \int_{\Omega'} g(z,\zeta) d\mu(\zeta) \le \int_{\Omega_0} g(z,\zeta) d\mu(\zeta) \\ &= \int_{\partial\Omega_0} u(\zeta) d\nu_z(\zeta) - u(z) \le C + C = 2C. \end{aligned}$$

Hence

$$\mu(\Omega') \leq \frac{2}{c_1} C =: C_1 C ,$$

which establishes the desired bound on the Riesz mass of u.

We note that in fact, for some constant $C_2(\Omega, \Omega') < \infty$, we also have (and this will be needed below) that

$$\mu(\Omega_0) \leq C_2 C .$$

To see this, consider a domain Ω_1 such that $\Omega_0 \Subset \Omega_1 \Subset \Omega$ and repeat the same argument shown above by considering the Green's function and the harmonic measure corresponding to the (larger) domain Ω_1 instead (in other words, Ω_1 will play the rôle of Ω_0 and Ω_0 that of Ω').

Deriving the bound on the harmonic part of the Riesz representation theorem requires more work, as we first have to identify more precisely this function.

Recall that $H(z, \zeta) := g(z, \zeta) + \log |z - \zeta|$ is harmonic in z on $\Omega_0 \supset \Omega'$, for all $\zeta \in \Omega_0 \supset \Omega'$. In particular, the function

$$\Omega' \ni z \mapsto \int_{\Omega'} H(z, \zeta) d\mu(\zeta)$$

is also harmonic.

Recall also that $g(z, \zeta)$ is harmonic in z on $\Omega_0 \setminus \{\zeta\}$, hence if $\zeta \in \Omega_0 \setminus \Omega'$, then $g(z, \zeta)$ is harmonic in z on the domain $\Omega' \subset \Omega_0 \setminus \{\zeta\}$. In particular, the function

$$\Omega' \ni z \mapsto \int_{\Omega_0 \setminus \Omega'} g(z, \zeta) d\mu(\zeta)$$

is also harmonic.

Let us write

$$\int_{\Omega_0} g(z, \zeta) d\mu(\zeta) = \int_{\Omega'} g(z, \zeta) d\mu(\zeta) + \int_{\Omega_0 \setminus \Omega'} g(z, \zeta) d\mu(\zeta)$$
$$\int_{\Omega'} g(z, \zeta) d\mu(\zeta) = \int_{\Omega'} H(z, \zeta) d\mu(\zeta) - \int_{\Omega'} \log |z - \zeta| \, d\mu(\zeta) .$$

Using the last two formulas, we can now rewrite (5.6) as

$$u(z) = \int_{\Omega'} \log|z-\zeta|\, d\mu(\zeta) \\ + \int_{\partial\Omega_0} u(\zeta) d\nu_z(\zeta) - \int_{\Omega'} H(z,\zeta) d\mu(\zeta) - \int_{\Omega_0\setminus\Omega'} g(z,\zeta) d\mu(\zeta)\,.$$

Then the harmonic part $h(z)$ of the Riesz representation of $u(z)$ can be described as the sum of the following harmonic functions:

$$h(z) = \int_{\partial\Omega_0} u(\zeta) d\nu_z(\zeta) - \int_{\Omega'} H(z,\zeta) d\mu(\zeta) - \int_{\Omega_0\setminus\Omega'} g(z,\zeta) d\mu(\zeta)\,.$$

It is now easy to bound h on a slightly smaller subdomain $\Omega'' \Subset \Omega'$.

For the first integral we use the fact that $u(z)$ is bounded and the harmonic measure is a probability measure:

$$\left|\int_{\partial\Omega_0} u(\zeta) d\nu_z(\zeta)\right| \le \int_{\partial\Omega_0} |u(\zeta)|\, d\nu_z(\zeta) \le C\,.$$

For the second integral we use the fact $H(z,\zeta)$ is smooth on $\Omega_0 \times \Omega_0$ and $\Omega'' \times \Omega'$ is compactly contained in $\Omega_0 \times \Omega_0$, so

$$\sup_{(z,\zeta)\in\Omega''\times\Omega'} |H(z,\zeta)| =: C_3 < \infty\,,$$

where C_3 depends on $\Omega_0, \Omega', \Omega''$, hence on $\Omega, \Omega', \Omega''$.

Then

$$\left|\int_{\Omega'} H(z,\zeta) d\mu(\zeta)\right| \le \int_{\Omega'} \big|H(z,\zeta)\big| d\mu(\zeta) \le C_3\, \mu(\Omega') \le C_3\, C_1\, C\,.$$

Finally, the third integral is the one that requires the restriction $z \in \Omega''$. Indeed, since $g(z,\zeta)$ is smooth on $\{(z,\zeta) \in \Omega_0 \times \Omega_0 \colon z \ne \zeta\}$ and it extends continuously to the set $\{(z,\zeta) \in \overline{\Omega_0} \times \overline{\Omega_0} \colon z \ne \zeta\}$ which contains $\Omega'' \times (\Omega_0 \setminus \Omega')$, it follows that

$$\sup_{(z,\zeta)\in\Omega''\times(\Omega_0\setminus\Omega')} |g(z,\zeta)| =: C_4 < \infty\,,$$

where C_4 depends on $\Omega, \Omega', \Omega''$.

Then

$$\left| \int_{\Omega_0 \setminus \Omega'} g(z,\zeta) d\mu(\zeta) \right| \le \int_{\Omega_0 \setminus \Omega'} |g(z,\zeta)| \, d\mu(\zeta) \le C_4 \, \mu(\Omega_0) \le C_4 \, C_2 \, C \, .$$

Putting it all together, we have that for all $z \in \Omega''$.

$$|h(z)| \le (1 + C_3 \, C_1 + C_4 \, C_2) \, C \, ,$$

which completes the proof. □

Our subharmonic functions are defined on an annulus $\mathcal{A} = \mathcal{A}_\rho$, whose width ρ was fixed once and for all. We may apply the above with the subdomains $\mathcal{A}'' \Subset \mathcal{A}' \Subset \mathcal{A}$ chosen as the annuli of width $\frac{\rho}{3}$ and $\frac{\rho}{2}$ respectively. Thus in particular we have.

Corollary 5.2. *Let $u \colon \mathcal{A} \to \mathbb{R}$ be a bounded subharmonic function, and let*

$$u(z) = \int_{\mathcal{A}'} \log |z - \zeta| \, d\mu(\zeta) + h(z)$$

be its Riesz representation on the smaller annulus $\mathcal{A}'$. Assume that

$$\sup_{z \in \mathcal{A}} |u(z)| \le C \, .$$

Then

$$\mu(\mathcal{A}') + \|h\|_{L^\infty(\mathcal{A}'')} \lesssim C \, . \tag{5.7}$$

We call the estimate in Corollary 5.2 above a *uniform measurement* on the bounded subharmonic function $u(z)$, since it does not depend on the function u per se, but only on its bound C. It is precisely this measurement that will determine the parameters in the qBET for the observable $u(x)$.

5.4 Decay of the Fourier coefficients

If $f \in C^1(\mathbb{T})$, using integration by parts, for every integer $k \neq 0$ we have that

$$\begin{aligned}\widehat{f'}(k) &= \int_0^1 f'(x) e(-kx) dx \\ &= f(x) e(-kx)\Big|_0^1 - \int_0^1 f(x)(-2\pi i k) e(-kx) dx \\ &= 2\pi i k \int_0^1 f(x)\, e(-kx) dx = 2\pi i k\, \widehat{f}(k)\,,\end{aligned}$$

hence

$$\big|\widehat{f}(k)\big| = \frac{\big|\widehat{f'}(k)\big|}{2\pi}\,\frac{1}{|k|} \leq \frac{\|f'\|_\infty}{2\pi}\,\frac{1}{|k|}\,.$$

Thus the Fourier coefficients of a continously differentiable function decay like $\frac{1}{|k|}$.

Weaker types of regularity still imply some rate of decay of the Fourier coefficients. For instance (see [45, Section 1.4.4]), if f is α-Hölder, then

$$\big|\widehat{f}(k)\big| = \mathcal{O}\Big(\frac{1}{|k|^\alpha}\Big) \quad \text{as} \quad |k| \to \infty\,.$$

For a function f with no such regularity properties besides mere continuity, all we can say about the decay of its Fourier coefficients is that $\widehat{f}(k) \to 0$ as $|k| \to \infty$ (this is the Riemann-Lebesgue lemma).

The remarkable fact about functions $u \colon \mathbb{T} \to \mathbb{R}$ that have a subharmonic extension to $\mathcal{A}$, is that while they generally have no regularity properties besides (semi-)continuity, their Fourier coefficients still decay like those of a continuously differentiable function, and this is what we set out to prove in this section.

We begin with a summary of the harmonic analysis concepts needed later in this section (see [45, Chapters 2 and 3]) for more details.

Let u be a harmonic function in the neighborhood of the closed unit disk $\overline{\mathbb{D}}$.

By the Poisson integral formula we have the representation:

$$u(z) = \int_{\mathbb{T}} u(e(t)) \, \frac{1 - \left|z\right|^2}{\left|z - e(t)\right|^2} \, dt \qquad \text{for all } \; z \in \mathbb{D} \, .$$

Writing $z = r\, e(x)$ with $0 \leq r < 1$, this formula becomes

$$u(re(x)) = \int_{\mathbb{T}} u(t) \, \frac{1 - r^2}{1 - 2r \cos(2\pi(x - t)) + r^2} \, dt \, .$$

We define the *Poisson kernel* as the family of functions $P_r \colon \mathbb{T} \to \mathbb{R}$, with $0 \leq r < 1$ and

$$P_r(x) := \frac{1 - r^2}{1 - 2r \cos(2\pi x) + r^2} \, .$$

Thus the values of u in $\mathbb{D}$ can be obtained from the values of u on the boundary $\mathbb{T}$ of $\mathbb{D}$ via the convolution[6] with the Poisson kernel:

$$u(re(x)) = (f * P_r)(x) \, ,$$

where $f(x) = u(e(x))$.

Conversely, given a continuous function $f \colon \mathbb{T} \to \mathbb{R}$, the function $u_f \colon \mathbb{D} \to \mathbb{R}$,

$$u_f(re(x)) := (f * P_r)(x)$$

is harmonic in $\mathbb{D}$ and its boundary value is f, since $u(z) \to f$ uniformly as $z \to \mathbb{T}$ radially.[7]

Hence the convolution with the Poisson kernel solves the Dirichlet problem for the Laplace equation, i.e. the problem of finding a harmonic function on the unit disk $\mathbb{D}$ with prescribed boundary condition.

The results of the following exercise are called the Cauchy estimates for harmonic functions. They are the analogues of the Cauchy estimates for holomorphic functions, which are derived via Cauchy's integral formula.

---

[6] Recall that the convolution of two functions $f, g \colon \mathbb{T} \to \mathbb{R}$ is the function $f * g$ on $\mathbb{T}$ defined by $f * g\,(x) := \int_{\mathbb{T}} f(x - t)\, g(t)\, dt$.

[7] More precisely, if for $0 \leq r < 1$ we define $F_r(x) := u(re(x))$, then $F_r \to f$ uniformly as $r \to 1$.

Exercise 5.3. Use the Poisson integral formula to show that if u is a harmonic function in a neighborhood of $\overline{\mathbb{D}}$, then

$$\big|\nabla u(0)\big| \le C_1\, M\,,$$

where C_1 is an absolute constant and $M := \sup_{z\in\mathbb{T}} |u(z)|$.

Derive from this that if u is harmonic in the neighborhood of some closed disk $\overline{D}(a,\rho)$, and if $|u(z)| \le M$ on its boundary, then

$$\big|\nabla u(a)\big| \le C_1\, \frac{1}{\rho}\, M\,.$$

We consider again the Dirichlet problem, but this time assuming less (than continuous) regularity for the boundary function.

Let $f \in L^2(\mathbb{T})$. Since the Poisson kernel $\{P_r\}_{0<r<1}$ is an approximate identity, the convolution of f with the kernel converges in the L^2-norm to f:

$$P_r * f \to f \quad \text{in } L^2(\mathbb{T}) \quad \text{as } r \to 1\,. \tag{5.8}$$

Let us also note that $P_r > 0$, and that, again because the Poisson kernel is an approximate identity,

$$\int_T P_r(x)dx = 1\,.$$

Moreover, for $z = r\,e(x) \in \mathbb{D}$, we put

$$P(z) = P(re(x)) := P_r(x)$$

Then P is a harmonic function in $\mathbb{D}$ since it is easy to see that

$$P(z) = \Re\left(\frac{1+z}{1-z}\right)\,.$$

Then if we define, as before, for $z = r\,e(x) \in \mathbb{D}$,

$$u_f(z) = u_f(re(x)) := (P_r * f)(x)\,,$$

then u_f is a harmonic function in $\mathbb{D}$ and by (5.8), its L^2 boundary value is f, in the sense that $u(z) \to f$ in L^2 as $z \to \mathbb{T}$ radially.

Thus the convolution with the Poisson kernel also solves the Dirichlet problem with L^2 boundary condition.

Let $u\colon \mathbb{D} \to \mathbb{R}$ be a harmonic function. The *harmonic conjugate* of u is the unique harmonic function $\widetilde{u}\colon \mathbb{D} \to \mathbb{R}$ such that $u + i\widetilde{u}$ is holomorphic in $\mathbb{D}$ and $\widetilde{u}(0) = 0$.

The harmonic conjugate of P is of course the function

$$Q(z) := \Im\left(\frac{1+z}{1-z}\right),$$

which defines the conjugate Poisson kernel

$$Q_r(x) = Q(re(x)) := \frac{2r\sin(2\pi x)}{1 - 2r\cos(2\pi x) + r^2}.$$

Hence if $f \in L^2(\mathbb{T})$, then the harmonic conjugate of its harmonic extension[8] u_f is

$$\widetilde{u_f}(re(x)) = (Q_r * f)(x).$$

It turns out (see [45, Corollary 3.14]) that $\widetilde{u_f}(re(x))$ converges as $r \to 1$ for a.e. $x \in \mathbb{T}$, and that it also converges in the L^2 norm. The limit, denoted by $\mathcal{H}(f)$, is called the *Hilbert transform* (or the conjugate function) of f.

Therefore, we obtain an operator $\mathcal{H}\colon L^2(\mathbb{T}) \to L^2(\mathbb{T})$,

$$\mathcal{H}(f)(x) = \lim_{r\to 1}\,(Q_r * f)(x).$$

$\mathcal{H}(f)$ represents the boundary value of the harmonic conjugate of the harmonic extension of f to the unit disk.

Exercise 5.4. Prove that if $0 \le r < 1$ then $\mathcal{H}(P_r) = Q_r$.

The Hilbert transform is a bounded linear operator.[9] It is related to the Fourier transform via the following identity on Fourier

[8] understood in the L^2 sense described above.

[9] This is a result of Marcel Riesz, the brother of Frigyes Riesz (who is responsible for the representation theorem for subharmonic functions described earlier). M. Riesz was the doctoral student of L. Fejér, whose kernel we will use in the next section; one of his doctoral students in Stockholm was H. Cramér, who is responssible for the large deviations principle formulated before—one big, happy, mathematical family.

coefficients:

$$\widehat{\mathcal{H}(f)}(k) = -i\,\mathrm{sgn}(k)\widehat{f}(k) \quad \text{for all} \quad k \in \mathbb{Z},\, k \neq 0\,.$$

Thus we conclude that if $f \in L^2(\mathbb{T})$, then for all $k \in \mathbb{Z}$ with $k \neq 0$,

$$\left|\widehat{\mathcal{H}(f)}(k)\right| = \left|\widehat{f}(k)\right|. \tag{5.9}$$

The following exercises contain the remaining properties of the Hilbert transform needed in the proof of Theorem 5.5 below.

Exercise 5.5. Show that for every $f \in L^2(\mathbb{T})$,

$$\mathcal{H}(\mathcal{H}(f)) = -f + u_f(0) = -f + \int_{\mathbb{T}} f\,.$$

Exercise 5.6. Show that the function $\log|e(x) - 1| \in L^2(\mathbb{T})$.

You may also show that for every analytic function $f\colon \mathbb{T} \to \mathbb{R}$, if $f \not\equiv 0$, then $\log|f| \in L^p(\mathbb{T})$, for all $1 \leq p < \infty$.

Consider the *saw-tooth function* $s\colon \mathbb{R} \to \mathbb{R}$,

$$s(x) := \begin{cases} \{x\} - \frac{1}{2} & \text{if } x \notin \mathbb{Z} \\ 0 & \text{if } x \in \mathbb{Z} \end{cases}$$

where $\{x\}$ is the fractional part of x. Note that $s(x)$ is 1-periodic.

Exercise 5.7. Prove that

$$\log|e(x) - 1| = -\pi\,\mathcal{H}(s)(x)\,.$$

To solve this exercise, you may follow the following steps.

1. Set out to actually compute $\mathcal{H}(\log|e(x) - 1|)$ instead, and then use Exercise 5.5.

2. Note that $\log|1 - z|$ is the harmonic extension to $\mathbb{D}$ of $\log|e(x) - 1|$, and that $\mathrm{Log}(1 - z)$ is holomorphic on $\mathbb{D}$, where Log refers to the principal branch of the logarithm.

3. Show that if $x \in [0, 1]$, then $\mathrm{Arg}(1 - e(x)) = \pi\, s(x)$, where Arg refers to the principal argument.

Exercise 5.8. Compute the Fourier coefficients of the saw-tooth function and conclude that for all $k \in \mathbb{Z}$, $k \neq 0$,

$$\widehat{s}(k) = \mathcal{O}\big(\frac{1}{k}\big) \,.$$

Lemma 5.3. *Let* $u_1(x) := \log |e(x) - 1|$. *Then*

$$\left|\widehat{u_1}(k)\right| \lesssim \frac{1}{|k|} \quad \textit{for all } k \neq 0 \,.$$

Proof. We simply combine the exercises above to get: $u_1 = -\pi \, \mathcal{H}(s)$, so

$$\left|\widehat{u_1}(k)\right| = \pi \left|\widehat{\mathcal{H}(s)}(k)\right| = \pi \left|\widehat{s}(k)\right| \lesssim \frac{1}{|k|} \,.$$

□

Theorem 5.5. *Let* $u(x)$ *be a function on* $\mathbb{T}$ *with a bounded subharmonic extension to* $\mathscr{A}$. *Assume that*

$$\sup_{z \in \mathscr{A}} |u(z)| \leq C \,.$$

Then the Fourier coefficients of u *have the decay*

$$\left|\widehat{u}(k)\right| \lesssim C \, \frac{1}{|k|} \quad \textit{for all } k \neq 0 \,. \tag{5.10}$$

Proof. We use the Riesz representation theorem for subharmonic functions and Corollary 5.2 regarding the uniform measurements on its components. Let $\mathscr{A}'' \Subset \mathscr{A}' \Subset \mathscr{A}$ and consider

$$u(z) = \int_{\mathscr{A}'} \log |z - \zeta| \, d\mu(\zeta) + h(z)$$

the Riesz representation of u on the annulus $\mathscr{A}'$.

Since $|u(z)| \leq C$ for all $z \in \mathscr{A}$,

$$\mu(\mathscr{A}') + \|h\|_{L^\infty(\mathscr{A}'')} \lesssim C \,. \tag{5.11}$$

We first bound the Fourier coefficients of the harmonic function h. By the observation at the beginning of this section, and since harmonic functions are smooth,

$$\left|\widehat{h}(k)\right| \le \frac{\|h'\|_{L^\infty(\mathbb{T})}}{2\pi} \, \frac{1}{|k|} \, .$$

However, h is harmonic on the annulus $\mathcal{A}'$ of width $\frac{\rho}{2}$ for some $\rho > 0$, and for every $x \in \mathbb{T}$, the closed disk $\overline{D}(x, \frac{\rho}{4}) \subset \mathcal{A}''$. Then by Exercise 5.3, for every $x \in \mathbb{T}$ we have that

$$|h'(x)| \le C_1 \, \frac{1}{\rho/4} \, \sup_{z \in \overline{D}(x, \frac{\rho}{4})} |h(z)| \lesssim \|h\|_{L^\infty(\mathcal{A}'')} \lesssim C \, .$$

We can then conclude that

$$\left|\widehat{h}(k)\right| \lesssim C \, \frac{1}{|k|} \, .$$

We now consider the logarithmic potential

$$v(z) := \int_{\mathcal{A}'} \log|z - \zeta| \, d\mu(\zeta) \, .$$

The goal is to show that for every $\zeta \in \mathbb{C}$, the Fourier coefficients of the function $u_\zeta(e(x)) := \log|e(x) - \zeta|$ are of order $\frac{1}{k}$, and then the result will follow by integration in ζ. We achieve this in several steps, depending on where ζ lies in the complex plane.

- When $\zeta = 1$, Lemma 5.3 proves the estimate.
- When $\zeta = e(x_0) \in \mathbb{T}$,

$$u_\zeta(e(x)) = \log|e(x) - e(x_0)| = \log|e(x - x_0) - 1| = u_1(x - x_0) \, .$$

Using the change of variables $x - x_0 = t$, we conclude that

$$\widehat{u_\zeta}(k) = e(-x_0) \, \widehat{u_1}(k) \, ,$$

so we conclude that

$$\left|\widehat{u_\zeta}(k)\right| \lesssim \frac{1}{|k|} \, .$$

▪ Now let $\zeta = r$, where $0 \le r < 1$.

Using integration by parts, we have

$$\begin{aligned}\left|\widehat{u_\zeta}(k)\right| &= \Big|\int_0^1 \log|e(x) - r|\, e(-kx)\, dx\Big| \\ &= \frac{1}{|k|}\Big|\int_0^1 \frac{d}{dx}\big[\log|e(x)-r|\big]\, e(-kx)\, dx\Big| = \frac{1}{|k|}\left|\widehat{f}(k)\right|,\end{aligned}$$

where $f(x) := \frac{d}{dx}\big[\log|e(x)-r|\big]$.

An easy calculation shows that

$$\begin{aligned}f(x) &= \frac{d}{dx}\big[\log|e(x)-r|\big] = \frac{1}{2}\frac{d}{dx}\log|e(x)-r|^2 \\ &= \frac{1}{2}\,\frac{2\pi\, 2r\sin(2\pi x)}{1 - 2r\cos(2\pi x) + r^2} = \pi\, Q_r(x) = \pi\, \mathcal{H}(P_r)(x)\,,\end{aligned}$$

where the last equality is due to Exercise 5.4.

Then

$$\begin{aligned}\left|\widehat{u_\zeta}(k)\right| &= \pi\,\frac{1}{|k|}\left|\widehat{\mathcal{H}(P_r)}(k)\right| = \pi\,\frac{1}{|k|}\left|\widehat{P_r}(k)\right| \\ &\le \pi\,\frac{1}{|k|}\,\|P_r\|_{L^1(\mathbb{T})} = \pi\,\frac{1}{|k|}\,.\end{aligned}$$

▪ When $|\zeta| < 1$, so $\zeta = re(x_0)$ with $0 < r < 1$, we simply perform a rotation by x_0 and reduce the problem to the previous case. Indeed,

$$u_\zeta(e(x)) = \log|e(x) - re(x_0)| = \log|e(x - x_0) - r|\,,$$

so we conclude as before by making the change of variables $x - x_0 = t$.

▪ Finally, consider the case $|\zeta| > 1$. This will be reduced to the case $|\zeta| < 1$ using the following exercise.

Exercise 5.9. Let $\zeta \in \mathbb{C}$ with $|\zeta| > 1$. Define $\zeta^* := \overline{\zeta^{-1}}$, the complex conjugate of its inverse, so that $|\zeta^*| = |\zeta|^{-1} < 1$.

Show that for all $e(x) \in \mathbb{T}$,

$$\log|e(x) - \zeta| = \log|e(x) - \zeta^*| + \log|\zeta|\,.$$

The Fourier coefficients of a constant function are all zero (except for the zeroth coefficient). Then from the exercise above,

$$\widehat{u_\zeta}(k) = \widehat{u_{\zeta^*}}(k) \quad \text{for all } \; k \neq 0\,,$$

which completes the proof in this case as well, since $|\zeta^*| < 1$.

We conclude that for all $\zeta \in \mathbb{C}$, $\zeta \neq 0$,

$$\left|\widehat{u_\zeta}(k)\right| \lesssim \frac{1}{|k|}\,,$$

where the underlying factor in $\lesssim$ is an absolute constant.

We may now estimate the Fourier coefficients of

$$v(x) = \int_{\mathscr{A}'} \log|e(x) - \zeta|\, d\mu(\zeta)\,.$$

We have:

$$\begin{aligned}\widehat{v}(k) &= \int_{\mathbb{T}} \left(\int_{\mathscr{A}'} \log|e(x) - \zeta|\, d\mu(\zeta) \right) e(-kx)\, dx \\ &= \int_{\mathscr{A}'} \left(\int_{\mathbb{T}} \log|e(x) - \zeta|\; e(-kx)\, dx \right) d\mu(\zeta) = \int_{\mathscr{A}'} \widehat{u_\zeta}(k)\, d\mu(\zeta)\,.\end{aligned}$$

Then for all $k \neq 0$,

$$|\widehat{v}(k)| \leq \int_{\mathscr{A}'} |\widehat{u_\zeta}(k)|\; d\mu(\zeta) \lesssim \frac{1}{|k|}\, \mu(\mathscr{A}') \lesssim C\, \frac{1}{|k|}\,,$$

which completes the proof. □

5.5 The proof of the large deviation type estimate

Let $A \in C^\omega_\rho(\mathbb{T}, \mathrm{SL}(2,\mathbb{R}))$ be a quasi-periodic cocycle.

Recall that each function $u_A^{(n)}(x)$, $n \geq 1$ extends to a subharmonic function $u_A^{(n)}(z)$ on $\mathscr{A}$, where

$$u_A^{(n)}(z) := \frac{1}{n}\, \log\|A^{(n)}(z)\|\,.$$

The next exercise shows that these subharmonic functions are bounded. The bound is uniform in n and also uniform in the cocycle.

Exercise 5.10. Show that for all $n \geq 1$ and for all $z \in \mathscr{A}$ we have:

$$0 \leq u_A^{(n)}(z) \leq \log\|A^{(n)}\|_r \,.$$

Furthermore, conclude that this bound is also uniform in the cocycle: there is $C = C(A) < \infty$ such that for all $B \in C^\omega_\rho(\mathbb{T}, \mathrm{SL}(2,\mathbb{R}))$with $\|B - A\|_r \leq 1$,

$$0 \leq u_B^{(n)}(z) \leq C \,.$$

Next we show that the functions $u_A^{(n)}(x)$ are *almost invariant* under the base transformation T in the following sense.

Proposition 5.4 (almost invariance)**.** *For all $x \in \mathbb{T}$ and for all $n \geq 1$ we have*

$$\left|u_A^{(n)}(x) - u_A^{(n)}(Tx)\right| \leq \frac{2\,\log\|A\|_r}{n} \,.$$

Proof. Fix $x \in \mathbb{T}$ and $n \geq 1$. Then

$$\begin{aligned}
&\frac{1}{n}\log\|A^{(n)}(x)\| - \frac{1}{n}\log\|A^{(n)}(Tx)\| \\
&\qquad = \frac{1}{n}\log\frac{\|A(T^n x)^{-1}\left[A(T^n x)\,A(T^{n-1}x)\,\ldots\,A(Tx)\right]\,A(x)\|}{\|A(T^n x)\,A(T^{n-1}x)\,\ldots\,A(Tx)\|} \\
&\qquad \leq \frac{1}{n}\log\left[\|A(T^n x)^{-1}\|\,\|A(x)\|\right] \leq \frac{2\,\log\|A\|_r}{n} \,.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\frac{1}{n}\log\|A^{(n)}(Tx)\| - \frac{1}{n}\log\|A^{(n)}(x)\| \\
&\qquad = \frac{1}{n}\log\frac{\|A(T^n x)\left[A(T^{n-1}x)\,\ldots\,A(Tx)\,A(x)\right]\,A(x)^{-1}\|}{\|A(T^{n-1}x)\,\ldots\,A(x)\|} \\
&\qquad \leq \frac{1}{n}\log\left[\|A(T^n x)\|\,\|A(x)^{-1}\|\right] \leq \frac{2\,\log\|A\|_r}{n} \,.
\end{aligned}$$

The conclusion then follows. □

As mentioned before, the main ingredient in the proof of the LDT estimate for quasi-periodic cocycles is a sharp quantitative Birkhoff ergodic theorem (qBET).

It is this result that requires an arithmetic assumption on the frequency. Recall that the frequency $\omega \in \mathbb{T}$ satisfies a Diophantine condition[10] if

$$\|k\omega\| \geq \frac{\gamma}{|k|\,(\log|k|)^2} \tag{5.12}$$

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

We briefly present the arithmetic considerations needed in the sequel. For full details on this topic, see S. Lang [37, Chapter 1].

Let $\omega \in \mathbb{T}$ be an irrational frequency and consider its continued fraction representation $\omega = [a_0, a_1, \dots, a_n, \dots]$. For every $n \geq 1$, let

$$\frac{p_n}{q_n} = [a_0, a_1, \dots, a_n]$$

be its n-th principal convergent.

The denominators of the principal convergents form an increasing sequence

$$0 < q_1 < \ldots < q_n < q_{n+1} < \ldots .$$

Moreover, by [37, Chapter 1, Theorem 5],

$$\frac{1}{2q_{n+1}} < \|q_n\omega\| = |q_n\omega - p_n| < \frac{1}{q_{n+1}} . \tag{5.13}$$

We call a *best approximation* to ω any (reduced) fraction $\frac{p}{q}$ such that $\|q\omega\| = |q\omega - p|$ and

$$\|j\omega\| > \|q\omega\| \quad \text{for all} \quad 1 \leq j < q .$$

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 [37, Chapter 1, Theorem 6]).

Therefore, if $\frac{p}{q}$ is a best approximation to ω, then $q = q_{n+1}$ for some $n \geq 0$, so for all $1 \leq j < q = q_{n+1}$ we must have that $\|j\omega\| \geq \|q_n\omega\|$. Using (5.13), $\|q_n\omega\| > \frac{1}{2q_{n+1}} = \frac{1}{2q}$.

Therefore, if $\frac{p}{q}$ is a best approximation to ω, then

$$\|j\omega\| > \frac{1}{2q} \quad \text{for all} \quad 1 \leq j < q .$$

[10] A weaker arithmetic condition will suffice, see [64].

As mentioned above, the denominators of the principal convergents, hence those of the best approximations to ω, form an increasing sequence. It turns out that if ω satisfies the Diophantine condition (5.12), then the frequency of their occurrence can be better specified.

Indeed, for any large enough integer R, let $n+1$ be the first integer j such that $q_j > R$. Then $q_n \leq R < q_{n+1}$, and using (5.13) and (5.12),

$$q_{n+1} < \frac{1}{\|q_n \omega\|} \lesssim q_n \, (\log q_n)^2 < R \, (\log R)^2 \, .$$

Thus $R < q_{n+1} \lesssim R \, (\log R)^2$.

We can summarize these considerations into the following lemma, which will be needed later.

Lemma 5.5. *Assume that the frequency ω satisfies the Diophantine condition* (5.12)*. Then for every large enough integer R there is a best approximation $\frac{p}{q}$ to ω such that*

$$R < q \lesssim R \, (\log R)^2 \, .$$

Moreover, if $1 \leq j < q$ then $\|j\omega\| > \frac{1}{2q}$.

Let $u \colon \mathbb{T} \to \mathbb{R}$ be an observable and for every $R \in \mathbb{N}$, let

$$(S_R u)(x) = \sum_{j=0}^{R-1} u(x + j\omega)$$

be the corresponding Birkhoff sums.

It is expected that if the averages $\frac{1}{R} \, S_R u \, (x)$ have nice convergence properties, then averaging again might improve these convergence properties.[11] Let us then consider the second order averages

$$\begin{aligned}\frac{1}{R}\, S_R \left(\frac{1}{R} S_R u\right)(x) &= \frac{1}{R^2}\sum_{j_1=0}^{R-1}\sum_{j_2=0}^{R-1} u(x+(j_1+j_2)\omega)\\ &= \frac{1}{R^2}\sum_{j=0}^{2(R-1)} c_R(j)\, u(x+j\omega)\,,\end{aligned}$$

where

$$c_R(j) := \#\{(j_1,j_2)\colon j_1+j_2=j,\ 0\le j_1,j_2\le R-1\}\,.$$

Note that

$$\frac{1}{R^2}\sum_{j=0}^{2(R-1)} c_R(j) = 1.$$

We are now ready to formulate a qBET with second order averages. As we will see at the end of this section, the fact that we consider second order rather than first order averages will be of no importance for establishing the LDT estimate.

Theorem 5.6. *Let $u\colon \mathbb{T}\to\mathbb{R}$ and let $\omega\in\mathbb{T}$. Assume that ω satisfies the Diophantine assumption* (5.12) *and that $u(x)$ has a subharmonic extension to the annulus $\mathscr{A}$ so that*

$$|u(z)|\le C \quad \textit{for all}\ \ z\in\mathscr{A}\,.$$

There is $c=\mathcal{O}\left(\frac{1}{C}\right)$ and for every $\epsilon>0$ there exists an integer $R_0=R_0(\epsilon,C)$ such that for all $R\ge R_0$ we have:

$$\Big|\{x\in\mathbb{T}\colon \Big|\frac{1}{R^2}\sum_{j=0}^{2(R-1)} c_R(j)\, u(x+j\omega)-\langle u\rangle\Big|>\epsilon\}\Big| < e^{-c\,\epsilon\, R}\,. \tag{5.14}$$

[11] This approach is inspired by the pointwise convergence of the partial sums of a Fourier series: while the partial sums of the Fourier series of a continuous function may fail to converge uniformly, or even pointwise everywhere, their Cesàro averages do converge uniformly by Fejér's theorem (see [36, Theorem 2.3]).

Remark 5.2. Second order averages are not strictly necessary for deriving such a qBET (see M. Goldstein and W. Schlag [25]). However, the argument we present here via second order averages is technically simpler. Furthermore, applying the same approach presented in the proof below to the usual Birkhoff averages instead of the second order averages, will lead to an estimate of the measure of the exceptional set just shy of the exponential decay needed (see Exercise 5.11).

Proof. Expand $u = u(x)$ into a Fourier series:

$$u(x) = \sum_{k\in\mathbb{Z}} \widehat{u}(k)\, e(kx) = \langle u\rangle + \sum_{k\neq 0} \widehat{u}(k)\, e(kx)\,.$$

The convergence of such a Fourier series (all throughout) should be understood in the L^2-norm.[12]

Then for $j \in \mathbb{Z}$

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

It follows that

$$\frac{1}{R^2}\sum_{j=0}^{2(R-1)} c_R(j)\, u(x+j\omega) - \langle u\rangle$$
$$= \sum_{k\neq 0} \widehat{u}(k) \left(\frac{1}{R^2} \sum_{j=0}^{2(R-1)} c_R(j)\, e(jk\omega) \right) e(kx)\,.$$

Consider the Fejér-type kernel of order 2:

$$F_R(t) := \left(\frac{1}{R} \sum_{j=0}^{R-1} e(jt) \right)^2 \,.$$

[12]The function u is bounded, so it belongs to $L^2(\mathbb{T})$. Thus the convergence of its Fourier series in $L^2(\mathbb{T})$ is of course an elementary fact, and this is all we need in the sequel. However, we note that in our setting, since $u(x)$ is continuous on $\mathbb{T}$ and its Fourier coefficients have the decay $\widehat{u}(k) = \mathcal{O}(\frac{1}{|k|})$, the convergence is in fact *pointwise* and uniform (see [36, Theorem 15.3]).

It is easy to see that

$$F_R(t) = \frac{1}{R^2} \sum_{j=0}^{2(R-1)} c_R(j)\, e(jt)\,.$$

Therefore,

$$\frac{1}{R^2} \sum_{j=0}^{2(R-1)} c_R(j)\, u(x + j\omega) - \langle u \rangle = \sum_{k \neq 0} \widehat{u}(k)\, F_R(k\omega)\, e(kx)\,. \quad (5.15)$$

We will estimate the right hand side of (5.15).

For that, let us recall that since $u(z)$ is subharmonic on $\mathscr{A}$ and it is uniformly bounded by C, from Theorem 5.5 we get :

$$|\widehat{u}(k)| \lesssim C \frac{1}{|k|}\,. \quad (5.16)$$

We will also have to estimate $F_R(k\omega)$. To this end, note that

$$|F_R(t)| = \frac{1}{R^2} \left| \frac{1 - e(Rt)}{1 - e(t)} \right|^2 \le \frac{1}{R^2\, \|t\|^2}.$$

Since obviously $|F_R(t)| \le 1$, the kernel satisfies the bound

$$|F_R(t)| \le \min\{1, \frac{1}{R^2\, \|t\|^2}\} \le \frac{2}{1 + R^2\, \|t\|^2}\,.$$

Since ω satisfies the Diophantine condition (5.12), by Lemma 5.5, there is a best approximation $\frac{p}{q}$ to ω so that

$$R < q \lesssim R\, (\log R)^2\,. \quad (5.17)$$

Moreover,

$$\|j\omega\| > \frac{1}{2q} \quad \text{if } 1 \le j < q\,. \quad (5.18)$$

Split the right hand side of the sum in (5.15) as:

$$\sum_{k \neq 0} \widehat{u}(k)\, e(kx)\, F_R(k\omega) = \sum_{0<|k|<R^{1/2}} + \sum_{R^{1/2} \le |k| < q} + \sum_{|q| \le |k| < K} + \sum_{|k| \ge K} \quad (5.19)$$

where $\log K \sim \epsilon\, R$.

In fact, it will be clear later that we should choose $K := e^{\frac{1}{C}\epsilon\, R}$.

The first three sums in (5.19), denoted by $S_1(x)$, $S_2(x)$ and $S_3(x)$ respectively, will be uniformly bounded in x by ϵ, while the forth sum, denoted by $S_4(x)$ will be estimated in the L^2-norm.

$$|S_1(x)| \leq \sum_{0<|k|<R^{1/2}} |\widehat{u}(k)|\, |F_R(k\omega)| \lesssim C \sum_{0<|k|<R^{1/2}} \frac{1}{|k|} \frac{1}{R^2 \|k\omega\|^2}\,.$$

Using the Diophantine condition (5.12), we have that

$$\|k\omega\| \gtrsim \frac{1}{|k|\, (\log|k|)^2} \quad \text{so} \quad \frac{1}{|k|\, \|k\omega\|^2} \lesssim |k|\, (\log|k|)^4\,.$$

Then for all $x \in \mathbb{T}$,

$$\begin{aligned} |S_1(x)| &\lesssim C\, \frac{1}{R^2} \sum_{0<|k|<R^{1/2}} |k|\, (\log|k|)^4 \\ &\lesssim C\, \frac{1}{R^2}\, R^{1/2}\, (\log R)^4\, R^{1/2} = C\, \frac{(\log R)^4}{R} < \epsilon \end{aligned}$$

provided $R \geq R_0(\epsilon, C)$.

To bound $S_2(x)$ and $S_3(x)$ we need the following estimate.

Let $I \subset \mathbb{Z}$ be an interval of size $|I| < q$. Then for all $k, k' \in I$ with $k \neq k'$, since $|k - k'| \leq |I| < q$, the inequality (5.18) implies

$$\|k\omega - k'\omega\| > \frac{1}{2q}\,.$$

Divide $\mathbb{T}$ into the $2q$ arcs $C_j = [\frac{j}{2q}, \frac{j+1}{2q})$, $0 \leq j \leq 2q-1$, with equal length $\left|C_j\right| = \frac{1}{2q}$. By the previous observation each arc C_j contains at most one point $k\omega$ with $k \in I$. Moreover, if $x \in C_j$ with $0 \leq j \leq q-1$ then $\|x\| \geq \frac{j}{2q}$ and similarly, if $x \in C_{2q-j-1}$ with $0 \leq j \leq q-1$ then $\|x\| \geq \frac{j}{2q}$. From this we derive:

$$\begin{aligned} \sum_{k\in I} |F_R(k\omega)| \leq \sum_{k\in I} \frac{2}{1 + R^2 \left\|k\omega\right\|^2} &\leq \sum_{j=0}^{q-1} \frac{4}{1 + R^2 (\frac{j}{2q})^2} \\ &\leq \int_0^q \frac{4}{1 + \frac{R^2}{4q^2}\, x^2}\, dx \lesssim \frac{q}{R}\,. \end{aligned}$$

Thus for any interval $I \subset \mathbb{Z}$ of size $< q$,

$$\sum_{k \in I} |F_R(k\omega)| \lesssim \frac{q}{R}\,. \tag{5.20}$$

Using (5.20) and then (5.17), it follows that for all $x \in \mathbb{T}$,

$$\begin{aligned}
|S_2(x)| &\le \sum_{R^{1/2} \le |k| < q} |\widehat{u}(k)|\, |F_R(k\omega)| \lesssim C \sum_{R^{1/2} \le |k| < q} \frac{1}{|k|}\, |F_R(k\omega)| \\
&\le C \frac{1}{R^{1/2}} \sum_{1 \le |k| < q} |F_R(k\omega)| \lesssim C\, \frac{1}{R^{1/2}} \frac{q}{R} \\
&< C\, \frac{1}{R^{1/2}}\, \frac{R(\log R)^2}{R} = C\, \frac{(\log R)^2}{R^{1/2}} < \epsilon\,,
\end{aligned}$$

provided $R \ge R_0(\epsilon, C)$.

Similarly,

$$\begin{aligned}
|S_3(x)| \le &\sum_{|q| \le |k| < K} |\widehat{u}(k)|\, |F_R(k\omega)| \lesssim C \sum_{|q| \le |k| < K} \frac{1}{|k|}\, |F_R(k\omega)| \\
&= C \sum_{1 \le s < K/q} \left(\sum_{sq \le |k| < (s+1)q} \frac{1}{|k|}\, |F_R(k\omega)| \right) \lesssim C \sum_{1 \le s < K/q} \frac{1}{sq} \frac{q}{R} \\
&= C \frac{1}{R} \sum_{1 \le s < K/q} \frac{1}{s} \le C \frac{1}{R} \log \frac{K}{q} < C \frac{1}{R} \log K = C \frac{1}{R}\, \frac{1}{C}\, \epsilon R = \epsilon\,.
\end{aligned}$$

We conclude that

$$|S_1(x)| + |S_2(x)| + |S_3(x)| \lesssim \epsilon. \tag{5.21}$$

for all $x \in \mathbb{T}$ and for $R \ge R_0(\epsilon, C)$.

We now estimate $S_4(x)$ in the L^2-norm. By Parseval's identity,

$$\begin{aligned}
\int_{\mathbb{T}} |S_4(x)|^2 &= \int_{\mathbb{T}} \Big| \sum_{|k| \ge K} \widehat{u}(k)\, F_R(k\omega)\, e(kx) \Big|^2 dx = \sum_{|k| \ge K} |\widehat{u}(k)|^2\, \big|F_R(k\omega)\big|^2 \\
&\le \sum_{|k| \ge K} |\widehat{u}(k)|^2 \lesssim C^2 \sum_{|k| \ge K} \frac{1}{|k|^2} \le C^2\, \frac{1}{K} = C^2\, e^{-\frac{1}{C}\,\epsilon\, R}.
\end{aligned}$$

Using Chebyshev' s inequality we get:

$$\Big|\{x \in \mathbb{T} \colon\, |S_4(x)| \gtrsim \epsilon\}\Big| \lesssim \frac{C^2\, e^{-\frac{1}{C}\,\epsilon\, R}}{\epsilon^2} = \left(\frac{C}{\epsilon}\right)^2\, e^{-\frac{\epsilon}{C}\, R} < e^{-\frac{\epsilon}{2C}\, R},$$

provided $R \geq R_0(\epsilon, C)$.

From (5.15), (5.19) and (5.21) we have

$$\{x \in \mathbb{T} \colon \Big|\frac{1}{R^2} \sum_{j=0}^{2(R-1)} c_R(j)\, u(x+j\omega) - \langle u \rangle\,\Big| \gtrsim \epsilon\} \subset \{x \in \mathbb{T} \colon\, |S_4(x)| \gtrsim \epsilon\}.$$

Combined with the previous estimate, this completes the proof. □

Exercise 5.11. Under the assumptions of Theorem 5.6, prove that for every $\epsilon > 0$

$$\Big|\{x \in \mathbb{T} \colon \Big|\frac{1}{R} \sum_{j=0}^{R-1} u(x+j\omega) - \langle u \rangle\,\Big| > \epsilon\}\Big| < e^{-c\,\epsilon\, R/\log R},$$

for all $R \geq R_0(\epsilon, C)$ and for some constant c of order $\frac{1}{C}$.

We now have all the required ingredients to establish the LDT for quasi-periodic cocycles.

Proof of Theorem 5.1. Let $A \in C^\omega_\rho(\mathbb{T}, \mathrm{SL}(2,\mathbb{R}))$ be a quasi-periodic cocycle, and let $C < \infty$ be such that $\log\|A\|_r < C$.

We will combine the qBET in Theorem 5.6 above with the almost invariance in Proposition 5.4.

Fix $\epsilon > 0$ and let $\bar{n}$ be a large enough integer, $\bar{n} = \epsilon\, R_0$, where $R_0 = R_0(\epsilon, C)$ is the threshold in Theorem 5.6.

Let $n \geq \bar{n}$ and choose an integer R with $R \asymp \frac{\epsilon}{C}\, n$, so $R \geq R_0$.

Applying Theorem 5.6 to the subharmonic function $u = u_A^{(n)}$ we have:

$$\Big|\frac{1}{R^2} \sum_{j=0}^{2(R-1)} c_R(j)\, u_A^{(n)}(x+j\omega) - \left\langle u_A^{(n)} \right\rangle\Big| \leq \epsilon\,, \tag{5.22}$$

for all phases x outside of a set of measure $< e^{-c\,\epsilon\,R}$, where c is of order $\frac{1}{C}$.

On the other hand, by almost invariance, we have that

$$\left|u_A^{(n)}(x) - u_A^{(n)}(Tx)\right| \le \frac{2\,\log\|A\|_r}{n} < \frac{2C}{n}\,,$$

hence for $0 \le j \le 2(R-1)$,

$$\left|u_A^{(n)}(x + j\omega) - u_A^{(n)}(x)\right| < \frac{2\,C\,j}{n} < \frac{4\,C\,R}{n} \le 4\epsilon\,.$$

We then have, for all $x \in \mathbb{T}$,

$$\left|u_A^{(n)}(x) - \frac{1}{R^2}\sum_{j=0}^{2(R-1)} c_R(j)\, u_A^{(n)}(x + j\omega)\right| < 4\epsilon\,. \tag{5.23}$$

From (5.22) and (5.23) we conclude that

$$\left|u_A^{(n)}(x) - \left\langle u_A^{(n)}\right\rangle\right| \le 5\epsilon$$

for x outside of a set of measure

$$< e^{-c\,\epsilon\,R} = e^{-c\,\epsilon\,\frac{\epsilon}{C}\,n} \le e^{-c^2\,\epsilon^2\,n}\,.$$

This concluses the proof of the theorem. □

Proof of Theorem 5.2. The ACT in Theorem 3 is clearly applicable.

That is because if $A \in C^\omega_\rho(\mathbb{T}, \mathrm{SL}(2,\mathbb{R}))$, then in particular A is continuous on $\mathbb{T}$ and hence bounded.

Moreover, for any $A, B \in C^\omega_\rho(\mathbb{T}, \mathrm{SL}(2,\mathbb{R}))$,

$$\begin{aligned} d(A,B) = \|A - B\|_r &= \sup_{z\in\mathcal{A}}\|A(z) - B(z)\| \\ &\ge \sup_{z\in\mathbb{T}}\|A(z) - B(z)\| = \|A - B\|_\infty\,. \end{aligned}$$

Finally, every $A \in C^\omega_\rho(\mathbb{T}, \mathrm{SL}(2,\mathbb{R}))$ satisfies a uniform LDT, as shown above. □

5.6 Consequence for Schrödinger cocycles

The goal of this section is to discuss the applicability of Theorem 5.1 on LDT estimates and of Theorem 5.2 on the continuity of the LE and of the Oseledets splitting to analytic quasi-periodic Schrödinger cocycles.

An analytic quasi-periodic Schrödinger cocycle is a linear cocycle over a torus translation, having the following structure:

$$A_{\lambda,E}(x) := \begin{bmatrix} \lambda\, v(x) - E & -1 \\ 1 & 0 \end{bmatrix},$$

where $\lambda \in \mathbb{R}$ is a coupling constant and $v \in C^{\omega}_{\rho}(\mathbb{T}, \mathbb{R})$ is the potential function. In this setting, λ and v are fixed, and the energy parameter $E \in \mathbb{R}$ varies.

Clearly Theorem 5.1 applies, so A_E satisfies an LDT estimate, with the same parameters for all E in a given compact interval. Moreover, the Lyapunov exponent $L(E) = L(A_{\lambda,E})$ depends continuously on E.

The remaining question is whether this dependence is Hölder continuous. According to Corollary 3.1 of the abstract continuity theorem, this happens when $L(E) > 0$.

Unlike in the random case, the Lyapunov exponent of a quasi-periodic Schrödinger cocycle is not always positive.

Having a large enough coupling constant λ is a sufficient condition for the positivity of $L(E)$ for all $E \in \mathbb{R}$. This result, which we formulate below, is due to E. Sorets and T. Spencer, see [55].

Theorem 5.7. *Let $\omega \in \mathbb{R} \setminus \mathbb{Q}$ and let $v \colon \mathbb{T} \to \mathbb{R}$ be a non constant real analytic function. There is a constant $\lambda_0 = \lambda_0(v) < \infty$ such that for all $\lambda \in \mathbb{R}$ with $|\lambda| \geq \lambda_0$, and for all $E \in \mathbb{R}$, we have*

$$L(A_{\lambda,E}) \geq \log |\lambda| - \mathcal{O}(1)\,. \tag{5.24}$$

In particular, by possibly increasing λ_0, $L(A_{\lambda,E})$ is positive and of order $\log|\lambda|$.

We present a different proof of this theorem, one which also ensures the better, optimal bound (5.24), compared with the original result in [55]. The argument we use here is an adaptation/simplification of the argument we used in [15] to establish the positivity of the Lyapunov exponents for more general types of quasi-periodic cocycles.

The following result, which we reproduce from [15], is the main analytic tool used to establish lower bounds on Lyapunov exponents. It is based on a convexity argument for means of subharmonic functions.

Proposition 5.6. *Let $u(z)$ be a subharmonic function on a neighborhood of the annulus $\mathcal{A}_\rho = \{z \colon 1-\rho \leq |z| \leq 1+\rho\}$. Assume that:*

$$u(z) \leq S \quad \text{for all } z : |z| = 1+\rho\,, \tag{5.25}$$

$$u(z) \geq \gamma \quad \text{for all } z : |z| = 1+y_0\,, \tag{5.26}$$

where $0 \leq y_0 < \rho$.

Then

$$\int_{\mathbb{T}} u(x)\, dx \geq \frac{1}{1-\alpha}\,(\gamma - \alpha S)\,, \tag{5.27}$$

where

$$\alpha = \frac{\log(1+y_0)}{\log(1+\rho)} \sim \frac{y_0}{\rho}\,. \tag{5.28}$$

Proof. The proof is a simple consequence of a general result on subharmonic functions, used to derive Hardy's convexity theorem (see Theorem 1.6 and the Remark following it in [18]). This result says that given a subharmonic function $u(z)$ on an annulus, its mean along concentric circles is log - convex. That is, if we define

$$m(r) := \int_{|z|=r} u(z)\, \frac{dz}{2\pi}$$

and if

$$\log r = (1-\alpha) \log r_1 + \alpha \log r_2 \tag{5.29}$$

for some $0 < \alpha < 1$, then

$$m(r) \leq (1-\alpha)\, m(r_1) + \alpha\, m(r_2)\,. \tag{5.30}$$

It can be shown, using say Green's theorem, that if $u(z)$ were *harmonic*, then $m(r)$ would be log - affine. Then the above result for subharmonic functions would follow using the principle of harmonic majorant (see [18] for details).

We apply (5.30) with $r = 1 + y_0$, $r_1 = 1$, $r_2 = 1 + \rho$, so for (5.29) to hold, α will be chosen as in (5.28). Then the convexity property (5.30) implies:

$$m(1 + y_0) \le (1 - \alpha)\, m(1) + \alpha\, m(1 + \rho)\,, \tag{5.31}$$

where

$$m(1) = \int_{|z|=1} u(z)\, \frac{dz}{2\pi} = \int_{\mathbb{T}} u(x) dx\,, \tag{5.32}$$

$$m(1 + \rho) = \int_{|z|=1+\rho} u(z)\, \frac{dz}{2\pi} \le S\,, \tag{5.33}$$

$$m(1 + y_0) = \int_{|z|=1+y_0} u(z)\, \frac{dz}{2\pi} \ge \gamma\,, \tag{5.34}$$

and (5.33) and (5.34) are due to (5.25) and (5.26) respectively.

The estimate (5.27) then follows from (5.31) - (5.34). □

We will need the following exercise on products of hyperbolic matrices.

Exercise 5.12. Let $n \in \mathbb{N}$ and for every index $0 \le j \le n-1$ consider the $\mathrm{SL}_2(\mathbb{R})$ matrix

$$g_j := \begin{bmatrix} a_j & -1 \\ 1 & 0 \end{bmatrix}.$$

Assume that for all $0 \le j \le n - 1$,

$$|a_j| \ge \lambda > 2\,.$$

Consider the product of these matrices

$$g^{(n)} := g_{n-1} \, \ldots \, g_1\, g_0 = \begin{bmatrix} s_n & * \\ t_n & * \end{bmatrix},$$

where $*$ stands for unspecified matrix entries.

Prove that

$$|t_n| < s_n \quad \text{and} \quad |s_n| \ge (\lambda - 1)^n\,.$$

Conclude that $\|g^{(n)}\| \ge (\lambda - 1)^n$.

From the previous exercise, we can easily derive the conclusion of Theorem 5.7 in the case when E is large enough relative to λ.[13]

Exercise 5.13. Assume that $|\lambda| > 2$ and that $|E| \geq |\lambda| \, (\|v\|_\infty + 1)$. Prove that

$$L(A_{\lambda,E}) \geq \log\left(|\lambda| - 1\right).$$

A crucial ingredient in this approach of proving Theorem 5.7 is the uniqueness theorem for holomorphic functions. That is, a holomorphic function has only a finite number of zeros in a compact set, unless it is identically zero.

Use this fact to solve the following exercise.

Exercise 5.14. Let $v \in C^\omega_\rho(\mathbb{T}, \mathbb{R})$, let $0 < \delta < \rho$ and let $\mathcal{I} \subset \mathbb{R}$ be a compact interval. Define

$$\epsilon_0(v) := \inf_{t \in \mathcal{I}} \sup_{r \in [1, 1+\delta]} \inf_{x \in [0,1]} |v(re(x)) - t| \,.$$

Prove that if v is non-constant then $\epsilon_0(v) > 0$.

Proof of Theorem 5.7. Since v is real analytic on $\mathbb{T}$, it has a holomorphic extension to the annulus $\mathscr{A}_\rho$, for some $\rho > 0$.

Let $\mathcal{I} := [-C, C]$, where $C = \|v\|_\infty + 1$.

Fix $0 < \delta < \rho$ and let $\epsilon_0(v)$ be the constant from Exercise 5.14. Since v is assumed non-constant, $\epsilon_0(v) > 0$.

Let $\lambda_0 > \frac{4}{\epsilon_0(v)}$ and fix any coupling constant λ with $|\lambda| \geq \lambda_0$.

By Exercise 5.13, it is enough to consider E such that $|E| \leq C \, |\lambda|$.

Fix such an energy E and let $t := \frac{E}{|\lambda|}$. Then $t \in \mathcal{I}$, so

$$\sup_{r \in [1, 1+\delta]} \inf_{x \in [0,1]} |v(re(x)) - t| \geq \epsilon_0(v) \,.$$

It follows that there is $r_0 \in [1, 1+\delta]$ such that for all $x \in [0, 1]$,

$$|v(r_0 \, e(x)) - t| \geq \frac{\epsilon_0(v)}{2} \,.$$

[13] However, this is not the interesting case of the theorem. The relevant energy parameters E are in the vicinity of the spectrum of the corresponding Schrödinger operator, which is contained in the interval $[-C\,|\lambda|\,, C\,|\lambda|]$, where $C = \mathcal{O}(\|v\|_\infty)$.

Put $r_0 = 1 + y_0$, where $0 \leq y_0 \leq \delta < \rho$. Since $t := \frac{E}{|\lambda|}$, from the previous estimate we get: for all z with $|z| = 1 + y_0$,

$$|\lambda\, v(z) - E| \geq |\lambda| \, \frac{\epsilon_0(v)}{2} \geq |\lambda_0| \, \frac{\epsilon_0(v)}{2} > 2\,. \tag{5.35}$$

The torus translation T, which we now regard as the rotation on $\mathbb{S}$, extends to the annulus $\mathscr{A}_\rho$, and it leaves all circles centered at 0 invariant. Thus (5.35) implies that for all $j \in \mathbb{N}$,

$$\left|\lambda\, v(T^j z) - E\right| \geq |\lambda| \, \frac{\epsilon_0(v)}{2} > 2\,. \tag{5.36}$$

Fix any $n \in \mathbb{N}$ and consider the matrices

$$g_j = g_j(z) := \begin{bmatrix} \lambda\, v(T^j z) - E & -1 \\ 1 & 0 \end{bmatrix}, \qquad \text{where } 0 \leq j \leq n-1\,,$$

whose product is clearly $A^{(n)}_{\lambda,E}(z)$.

Estimate (5.36) shows that if $|z| = 1 + y_0$, then these matrices satisfy the assumptions in Exercise 5.12, and we conclude that

$$\|A^{(n)}_{\lambda,E}(z)\| \geq \left(|\lambda| \, \frac{\epsilon_0(v)}{2} - 1\right)^n \,.$$

Taking logarithms on both sides, we get that for all $z\colon\ |z| = 1+y_0$,

$$\frac{1}{n} \log \|A^{(n)}_{\lambda,E}(z)\| \geq \log \left(|\lambda| \, \frac{\epsilon_0(v)}{2} - 1\right) = \log|\lambda| - \mathcal{O}(1) =: \gamma(\lambda)\,. \tag{5.37}$$

Moreover, since for all $z \in \mathscr{A}_\rho$, we clearly have the upper bound

$$\|g_j(z)\| \leq |\lambda| \, \|v\|_r + |E| \leq |\lambda| \, (\|v\|_r + \|v\|_\infty + 1) =: C_1 \, |\lambda| \,,$$

we derive that

$$\|A^{(n)}_{\lambda,E}(z)\| \leq (C_1 \, |\lambda|)^n \,.$$

Taking logarithms on both sides we conclude that for all $z \in \mathscr{A}_\rho$,

$$\frac{1}{n} \log \|A^{(n)}_{\lambda,E}(z)\| \leq \log\, (C_1 \, |\lambda|) = \log|\lambda| + \mathcal{O}(1) =: S(\lambda)\,. \tag{5.38}$$

The estimates (5.37) and (5.38) ensure that the hypothesis of the Proposition 5.6 hold for the subharmonic function

$$u(z) = \frac{1}{n} \log \|A^{(n)}_{\lambda,E}(z)\| \,.$$

Thus we conclude that its mean on $\mathbb{T}$ has the lower bound

$$\int_{\mathbb{T}} \frac{1}{n} \log \|A^{(n)}_{\lambda,E}(x)\| = \int_{\mathbb{T}} u(x) dx \geq \log|\lambda| - \mathcal{O}(1) \,.$$

As this holds for all $n \in \mathbb{N}$, taking the limit as $n \to \infty$ we conclude that

$$L(A_{\lambda,E}) \geq \log|\lambda| - \mathcal{O}(1) \,,$$

which completes the proof of the theorem. □

5.7 Bibliographical notes

Let us begin by noting that the results in this chapter have a counterpart, albeit a slightly weaker one[14], regarding *higher dimensional* torus translations. In fact, it is in this higher dimensional setting that this method of proving continuity of the LE via large deviations proves its versatility, as all other available approaches are essentially one dimensional.

In some sense, the strongest result on continuity of the Lyapunov exponents for quasi-periodic cocycles in the *one-dimensional* torus translation case is due to A. Ávila, S. Jitomirskaya and C. Sadel (see [2]). The authors prove joint continuity in cocycle and frequency at all points (A, ω) with ω irrational. The cocycles considered are analytic and $\mathrm{Mat}_m(\mathbb{C})$-valued. A previous work of S. Jitomirskaya and C. Marx [30] established a similar result for $\mathrm{Mat}_2(\mathbb{C})$-valued cocycles, using a different approach. We note here that both approaches rely crucially on the convexity of the Lyapunov exponent of the complexified cocycle, as a function of the imaginary variable, by establishing first continuity away from the torus. This method immediately breaks down in the higher dimensional torus translation case.

[14] The available results provide a weak-Hölder modulus of continuity in the higher dimensional setting.

The approaches of [2] and [30] are independent of any arithmetic constraints on the translation frequency ω and they do not use large deviations. However, the results are not quantitative, in the sense that the they do not provide any modulus of continuity of the Lyapunov exponents. All available quantitative results, from the classic result of M. Goldstein and W.Schlag (see [25]) to more recent results such as [56, 57, 64, 14, 17], use some type of large deviations, whose derivation depends upon imposing appropriate arithmetic conditions on ω.

We note that in the (more particular) context of Schrödinger cocycles, joint continuity in the energy parameter and the frequency translation was proven for the one dimensional torus translation case by J. Bourgain and S. Jitomirskaya (see [11]) and for the higher dimensional torus translation case by J. Bourgain (see [10]). Both papers used weaker versions of large deviation estimates, proven under weak arithmetic (i.e. restricted Diophantine) conditions on ω, although eventually the results were made independent of any such restrictions.

Continuity properties of the Lyapunov exponents were also established for certain *non-analytic* quasi-periodic models (see [34, 35, 63]). Moreover, the reader may also consult the recent surveys [13], [31].

Bibliography

[1] V. I. Arnol'd. *Mathematical methods of classical mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 199? Translated from the 1974 Russian original by K. Vogtmann and A. Weinstein, Corrected reprint of the second (1989) edition.

[2] Artur Ávila, Svetlana Jitomirskaya, and Christian Sadel. Complex one-frequency cocycles. *J. Eur. Math. Soc. (JEMS)*, 16(9):1915–1935, 2014.

[3] L. Backes. A note on the continuity of Oseledets subspaces for fiber-bunched cocycles. *preprint*, pages 1–6, 2015.

[4] Lucas Backes and Mauricio Poletti. Continuity of Lyapunov exponents is equivalent to continuity of Oseledets subspaces. *preprint*, pages 1–16, 2015.

[5] Alexandre Baraviera and Pedro Duarte. Approximating Lyapunov exponents and stationary measures. *preprint*, pages 1–27, 2017.

[6] Carlos Bocker-Neto and Marcelo Viana. Continuity of Lyapunov exponents for random two-dimensional matrices. *Ergodic Theory and Dynamical Systems*, pages 1–30, 2016.

[7] Philippe Bougerol. Théorèmes limite pour les systèmes linéaires à coefficients markoviens. *Probab. Theory Related Fields*, 78(2):193–221, 1988.

[8] Philippe Bougerol. Théorèmes limite pour les systèmes linéaires à coefficients markoviens. *Probab. Theory Related Fields*, 78(2):193–221, 1988.

[9] Philippe Bougerol and Jean Lacroix. *Products of random matrices with applications to Schrödinger operators*, volume 8 of *Progress in Probability and Statistics.* Birkhäuser Boston, Inc., Boston, MA, 1985.

[10] J. Bourgain. Positivity and continuity of the Lyapounov exponent for shifts on $\mathbb{T}^d$ with arbitrary frequency vector and real analytic potential. *J. Anal. Math.*, 96:313–355, 2005.

[11] J. Bourgain and S. Jitomirskaya. Continuity of the Lyapunov exponent for quasiperiodic operators with analytic potential. *J. Statist. Phys.*, 108(5-6):1203–1218, 2002. Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays.

[12] Jean-René Chazottes and Sébastien Gouëzel. Optimal concentration inequalities for dynamical systems. *Communications in Mathematical Physics*, 316(3):843–889, 2012.

[13] D Damanik. Schrödinger operators with dynamically defined potentials: a survey. *preprint*, pages 1–80, 2015. to appear in Ergodic Theory and Dynamical Systems.

[14] Pedro Duarte and Silvius Klein. Continuity of the Lyapunov Exponents for Quasiperiodic Cocycles. *Comm. Math. Phys.*, 332(3):1113–1166, 2014.

[15] Pedro Duarte and Silvius Klein. Positive Lyapunov exponents for higher dimensional quasiperiodic cocycles. *Comm. Math. Phys.*, 332(1):189–219, 2014.

[16] Pedro Duarte and Silvius Klein. *Lyapunov exponents of linear cocycles; continuity via large deviations*, volume 3 of *Atlantis Studies in Dynamical Systems.* Atlantis Press, 2016.

[17] Pedro Duarte and Silvius Klein. Continuity, positivity and simplicity of the Lyapunov exponents for quasi-periodic cocycles. *to appear in J. Eur. Math. Soc. (JEMS)*, pages 1–64, 2017.

[18] Peter L. Duren. *Theory of H^p spaces.* Pure and Applied Mathematics, Vol. 38. Academic Press, New York, 1970.

[19] Alex Furman. On the multiplicative ergodic theorem for uniquely ergodic systems. *Ann. Inst. H. Poincaré Probab. Statist.*, 33(6):797–815, 1997.

[20] H. Furstenberg and H. Kesten. Products of random matrices. *Ann. Math. Statist.*, 31:457–469, 1960.

[21] H. Furstenberg and Y. Kifer. Random matrix products and measures on projective spaces. *Israel J. Math.*, 46(1-2):12–32, 1983.

[22] Harry Furstenberg. Noncommuting random products. *Trans. Amer. Math. Soc.*, 108:377–428, 1963.

[23] Theodore W. Gamelin. *Complex analysis.* Undergraduate Texts in Mathematics. Springer-Verlag, New York, 2001.

[24] John B. Garnett and Donald E. Marshall. *Harmonic measure*, volume 2 of *New Mathematical Monographs.* Cambridge University Press, Cambridge, 2008. Reprint of the 2005 original.

[25] Michael Goldstein and Wilhelm Schlag. Hölder continuity of the integrated density of states for quasi-periodic Schrödinger equations and averages of shifts of subharmonic functions. *Ann. of Math. (2)*, 154(1):155–203, 2001.

[26] Michael Goldstein and Wilhelm Schlag. Fine properties of the integrated density of states and a quantitative separation property of the Dirichlet eigenvalues. *Geom. Funct. Anal.*, 18(3):755–869, 2008.

[27] Y. Guivarc'h and A. Raugi. Frontière de Furstenberg, propriétés de contraction et théorèmes de convergence. *Z. Wahrsch. Verw. Gebiete*, 69(2):187–242, 1985.

[28] W. K. Hayman and P. B. Kennedy. *Subharmonic functions. Vol. I.* Academic Press [Harcourt Brace Jovanovich, Publishers], London-New York, 1976. London Mathematical Society Monographs, No. 9.

[29] Hubert Hennion and Loïc Hervé. *Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness*, volume 1766 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2001.

[30] S. Jitomirskaya and C. A. Marx. Analytic quasi-perodic cocycles with singularities and the Lyapunov exponent of extended Harper's model. *Comm. Math. Phys.*, 316(1):237–267, 2012.

[31] S. Jitomirskaya and C. A. Marx. Dynamics and spectral theory of quasi-periodic Schrödinger-type operators. *preprint*, pages 1–44, 2015. to appear in Ergodic Theory and Dynamical Systems.

[32] Svetlana Jitomirskaya and Rajinder Mavi. Continuity of the measure of the spectrum for quasiperiodic Schrödinger operators with rough potentials. *Comm. Math. Phys.*, 325(2):585–601, 2014.

[33] Yitzhak Katznelson and Benjamin Weiss. A simple proof of some ergodic theorems. *Israel J. Math.*, 42(4):291–296, 1982.

[34] Silvius Klein. Anderson localization for the discrete one-dimensional quasi-periodic Schrödinger operator with potential defined by a Gevrey-class function. *J. Funct. Anal.*, 218(2):255–292, 2005.

[35] Silvius Klein. Localization for quasiperiodic Schrödinger operators with multivariable Gevrey potential functions. *J. Spectr. Theory*, 4:1–53, 2014.

[36] T. W. Körner. *Fourier analysis*. Cambridge University Press, Cambridge, second edition, 1989.

[37] Serge Lang. *Introduction to Diophantine approximations*. Springer-Verlag, New York, second edition, 1995.

[38] Émile Le Page. Théorèmes limites pour les produits de matrices aléatoires. In *Probability measures on groups (Oberwolfach, 1981)*, volume 928 of *Lecture Notes in Math.*, pages 258–303. Springer, Berlin-New York, 1982.

[39] Émile Le Page. 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.

[40] F. Ledrappier and L.-S. Young. Stability of Lyapunov exponents. *Ergodic Theory Dynam. Systems*, 11(3):469–484, 1991.

[41] B. Ya. Levin. *Lectures on entire functions*, volume 150 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1996. In collaboration with and with a preface by Yu. Lyubarskii, M. Sodin and V. Tkachenko, Translated from the Russian manuscript by Tkachenko.

[42] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times. With a chapter on "Coupling from the past" by James G. Propp and David B. Wilson.* Providence, RI: American Mathematical Society (AMS), 2009.

[43] Eugene Lukacs. *Characteristic functions*. Hafner Publishing Co., New York, 1970. Second edition, revised and enlarged.

[44] Elaís C. Malheiro and Marcelo Viana. Lyapunov exponents of linear cocycles over Markov shifts. *Stoch. Dyn.*, 15(3):1550020, 27, 2015.

[45] Camil Muscalu and Wilhelm Schlag. *Classical and multilinear harmonic analysis. Vol. I*, volume 137 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2013.

[46] S. V. Nagaev. Some limit theorems for stationary Markov chains. *Teor. Veroyatnost. i Primenen.*, 2:389–416, 1957.

[47] Gunter Ochs. Stability of Oseledets spaces is equivalent to stability of Lyapunov exponents. *Dynam. Stability Systems*, 14(2):183–201, 1999.

[48] Yuval Peres. Analytic dependence of Lyapunov exponents on transition probabilities. In *Lyapunov exponents (Oberwolfach, 1990)*, volume 1486 of *Lecture Notes in Math.*, pages 64–80. Springer, Berlin, 1991.

[49] Firas Rassoul-Agha and Timo Seppäläinen. *A course on large deviations with an introduction to Gibbs measures*, volume 162 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2015.

[50] A. Raugi and J. Rosenberg. *Fonctions harmoniques et théorèmes limites pour les marches aléatoires sur les groupes*. Number no. 54 in Bulletin de la Société mathématique de France. Mémoire. Société mathématique de France, 1977.

[51] Frigyes Riesz and Béla Sz.-Nagy. *Functional analysis*. Frederick Ungar Publishing Co., New York, 1955. Translated by Leo F. Boron.

[52] D. Ruelle. Analycity properties of the characteristic exponents of random matrix products. *Adv. in Math.*, 32(1):68–80, 1979.

[53] Wilhelm Schlag. Regularity and convergence rates for the Lyapunov exponents of linear cocycles. *J. Mod. Dyn.*, 7(4):619–637, 2013.

[54] Barry Simon and Michael Taylor. Harmonic analysis on $\mathrm{SL}(2,\mathbf{R})$ and smoothness of the density of states in the one-dimensional Anderson model. *Comm. Math. Phys.*, 101(1):1–19, 1985.

[55] Eugene Sorets and Thomas Spencer. Positive Lyapunov exponents for Schrödinger operators with quasi-periodic potentials. *Comm. Math. Phys.*, 142(3):543–566, 1991.

[56] Kai Tao. Continuity of Lyapunov exponent for analytic quasi-periodic cocycles on higher-dimensional torus. *Front. Math. China*, 7(3):521–542, 2012.

[57] Kai Tao. Hölder continuity of Lyapunov exponent for quasi-periodic Jacobi operators. *Bull. Soc. Math. France*, 142(4):635–671, 2014.

[58] Terence Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.

[59] V. N. Tutubalin. Limit theorems for a product of random matrices. *Teor. Verojatnost. i Primenen.*, 10:19–32, 1965.

[60] M. Viana. *Lectures on Lyapunov Exponents.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2014.

[61] Marcelo Viana and Krerley Oliveira. *Foundations of ergodic theory*, volume 151 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 2016.

[62] P. Walters. *An Introduction to Ergodic Theory.* Graduate Texts in Mathematics. Springer New York, 2000.

[63] Yiqian Wang and Zhenghe Zhang. Uniform positivity and continuity of Lyapunov exponents for a class of C^2 quasiperiodic Schrödinger cocycles. *J. Funct. Anal.*, 268(9):2525–2585, 2015.

[64] Jiangong You and Shiwen Zhang. Hölder continuity of the Lyapunov exponent for analytic quasiperiodic Schrödinger cocycle with weak Liouville frequency. *Ergodic Theory Dynam. Systems*, 34(4):1395–1408, 2014.

Index