Stochastics and Dynamics
Vol. 16, No. 3 (2016) 1660010 (54 pages)

DOI: 10.1142/S0219493716600108

Large deviation type estimates for iterates of linear cocycles

Pedro Duarte

Departamento de Matemática and CMAF-CIO,
Faculdade de Ciências da Universidade de Lisboa,
P-1749-016, Lisbon, Portugal
pmduarte@fc.ul.pt

Silvius Klein

Department of Mathematical Sciences,
Norwegian University of Science and Technology (NTNU),
NO-7491, Trondheim, Norway
silvius.klein@math.ntnu.no

Received 17 September 2014
Accepted 29 July 2015
Published 22 January 2016

We describe some methods used to derive large deviation type (LDT) estimates for quantities associated to random and quasi-periodic linear cocycles. We then explain how such LDT estimates can be used in an inductive scheme to prove continuity properties of the Lyapunov exponents as functions of the cocycle. This is a survey of recent work to appear in a research monograph.

Keywords: Linear cocycles; Lyapunov exponents; large deviations; Schrödinger operators.

AMS Subject Classification: 37H15, 34D08, 60F10, 82B44

1. Introduction

This paper surveys the use of certain large deviation type (LDT) estimates for dynamical systems in the study of continuity properties of the Lyapunov exponents of linear cocycles. A fully detailed version of this work will appear in the research monograph [17], see also [14, 13, 15, 16].

A linear cocycle is the dynamical system underlying a skew-product map acting on a vector bundle. The base dynamics is given by an ergodic transformation, while the action on the fiber is given by a matrix valued function with a certain regularity. Lyapunov exponents are quantities that measure the average exponential growth of the iterates of the cocycle along the fibers (see [1]).

Two classes of general linear cocycles have been extensively studied so far: the class of random cocycles, where the base dynamics is a Bernoulli shift, and the class of quasi-periodic cocycles, where the base dynamics is a torus translation.

We study here the continuity properties of the Lyapunov exponents for a general class of linear cocycles over a fixed base dynamics. We identify the cocycle with the matrix valued function that determines its fiber action. Continuity is meant as a function of the cocycle. Our method applies to both random and quasi-periodic cocycles and it gives quantitative results, i.e. a modulus of continuity under appropriate further conditions.

The main assumptions required by the method described in this paper to prove continuity of the Lyapunov exponents, are the availability of LDT estimates on the base dynamics for a rich enough class of observables, and the existence of uniform LDT estimates on the fiber dynamics. Large deviations in classical probabilities or for multiplicative systems associated to a dynamical system describe the asymptotic behavior of tail events in terms of a rate function. We require a somewhat different type of large deviations. Instead of a precise rate, only a good estimate on the decay of the tail event is needed. However, we require that such estimates hold for all iterates of the system, after a certain threshold, and that in the case of the fiber dynamics, this threshold of applicability as well as various other parameters describing the LDT estimates are stable under small perturbations of the cocycle, a property we refer to as *uniform* fiber LDT estimates.

For quasi-periodic models, base LDT estimates are a consequence of the unique ergodicity of the system, while for the random i.i.d. model, base LDT estimates are a consequence of the classical Cramér's theorem.

Uniform fiber LDT estimates for quasi-periodic models were first obtained by J. Bourgain and M. Goldstein in [8] for *Schrödinger cocycles*, and used in their study of spectral properties of discrete quasi-periodic Schrödinger operators. For random models, fiber LDT estimates follow from the work of E. Le Page [36] (the Bernoulli case) and P. Bougerol [4] (the Markov case). However, these estimates lack the *uniformity* we require in the proof of continuity of the Lyapunov exponents. We obtain uniform base and fiber LDT estimates for Bernoulli and Markov cocycles by following a more general and abstract method described in [28].

Kingman's subadditive ergodic theorem allows us to describe Lyapunov exponents as limits, when the number n of iterates grows, of finite scale Lyapunov exponents, which are defined as the phase space average of quantities related to the singular values of the nth iterate of the cocycle.

The mechanism for obtaining quantitative continuity properties of the limit objects (i.e. the Lyapunov exponents) is a deterministic result called the avalanche principle which was first established for $\mathrm{SL}(2,\mathbb{R})$ matrices by M. Goldstein and W. Schlag in [23].

Roughly speaking, the avalanche principle (AP) allows us to relate singular values of a long block (i.e. product) of matrices to certain averages of singular values of individual components of the block. This holds provided certain geometric conditions (which we call "gaps" and "angles" conditions) on the individual components are satisfied. The gap condition means that a pattern on the relative sizes of consecutive singular values holds uniformly for all elements of the block, while the angle condition ensures that most expanding singular directions of consecutive elements of the block are not almost orthogonal, hence they are not canceling each other out.

In order to effectively apply the AP to long blocks made up of iterates of a cocycle, the geometric conditions need to be satisfied for a large enough set of phases. This is where the LDT estimates on the fiber action are used, as they turn estimates on finite scale Lyapunov exponents, which are phase space *averages*, into *pointwise* estimates which hold for a large number of phases, and correspondingly they imply the geometric conditions of the AP for that large set of phases.

The sharpness of the LDT determines how long a block of matrices in the AP can be, before running out of phases satisfying the geometric conditions. This argument is then used repeatedly, in an inductive procedure, where the previous long block becomes a typical component of the next much larger block, and the LDT estimate is used again to guarantee the geometric conditions for sufficiently many phases, and hence the applicability of the AP in the next stage of the induction.

This method of proving continuity of Lyapunov exponents was first introduced by M. Goldstein and W. Schlag in [23] in the context of *quasi-periodic*, analytic Schrödinger cocycles, where the base dynamics is a torus translation by a *Diophantine* frequency.

Continuity results for Lyapunov exponents of random cocycles satisfying an irreducibility condition go back to H. Furstenberg, Y. Kifer [21] and E. Le Page [37]. More recently, continuity results for general random cocycles were obtained in [3, 39], see also M. Viana's monograph [51] for a more detailed account of these results.

To summarize, this paper presents an introduction to some of the methods used to derive LDT estimates for quasi-periodic and random cocycles, and an abstract scheme we have developed to prove continuity of Lyapunov exponents of cocycles satisfying such estimates. This method is versatile enough to apply to both quasi-periodic models (one or multivariable Diophantine torus translations) and random models (Bernoulli and Markov systems), and possibly to other base dynamics; it provides a modulus of continuity (whose strength depends on the sharpness of the LDT) in the neighborhood of a simple Lyapunov exponent; it is flexible enough to apply to higher dimensional cocycles; it is general enough to imply and to extend most already known *quantitative* continuity results.

This survey is organized as follows. In Sec. 2 we describe and compare different types of large deviations for random processes and dynamical systems; given a linear cocycle, we then introduce our concepts of base and fiber LDT estimates, to be used in this paper. In Sec. 3, we explain the use of tools from harmonic analysis (e.g., BMO estimates) and analytic number theory (e.g., Erdös–Turán inequalities) in deriving such LDT estimates for quasi-periodic cocycles. In Sec. 4, we explain the use of functional analysis tools (e.g., perturbation theory of quasi-compact operators) in deriving such LDT estimates for random cocycles. Section 5 describes the abstract continuity theorem of the Lyapunov exponents assuming the availability of the LDT estimates introduced in Sec. 2 and its applicability to quasi-periodic and random cocycles; furthermore, we describe the inductive procedure, based on the avalanche principle, leading to its proof. We conclude the paper by indicating other possible uses of the LDT estimates.

2. Definitions of LDT Estimates

In probability theory and harmonic analysis there are several inequalities describing the deviation of a function from its mean. The most basic result of this kind is Chebyshev's inequality. We formulate it in its exponential form. For any $t, \lambda > 0$ and any random variable X

$$\mathbb{P}[|X - \mathbb{E}(X)| \geq \lambda] \leq e^{-\lambda t}\mathbb{E}[e^{t|X-\mathbb{E}(X)|}]. \tag{2.1}$$

A fundamental result in harmonic analysis, concerning functions of bounded mean oscillation (BMO), is John–Nirenberg's inequality. Given $f \in L^1(\mathbb{T})$ let

$$\|f\|_{\mathrm{BMO}} := \sup_I \langle |f - \langle f \rangle_I| \rangle_I,$$

where the sup is taken over all intervals $I \subset \mathbb{T}$ and $\langle f \rangle_I = \frac{1}{|I|}\int_I f$.

Then if $\|f\|_{\mathrm{BMO}} < +\infty$, John–Nirenberg's inequality states that

$$|\{x \in \mathbb{T} : |f - \langle f \rangle_{\mathbb{T}}| \geq \lambda\}| \leq e^{-c\lambda/\|f\|_{\mathrm{BMO}}}, \tag{2.2}$$

where c is a universal constant.

Let $X_0, X_1, X_2, \ldots$ be a real valued random process and denote by $S_n = \sum_{j=0}^{n-1} X_j$ the corresponding sum process. *Tail* events of this process correspond to the deviation of its averages $\frac{1}{n}S_n$ from their means $\mathbb{E}(\frac{1}{n}S_n)$.

There are several types of large deviation inequalities describing tail events, such as Chernoff bounds (see [48]), which we formulate for a random i.i.d. process $\{X_n\}$:

$$\mathbb{P}\left[\left|\frac{1}{n}S_n - \mu\right| \geq \lambda\right] < C\max\{e^{-(c\lambda^2/\sigma^2)n}, e^{-(\lambda/K)n}\} \tag{2.3}$$

for some universal constants $C, c > 0$ and where $\mu = \mathbb{E}(X_0)$, $\sigma^2 = \mathrm{var}(X_0)$ and $K = \|X_0\|_\infty$.

The asymptotic behavior of tail events forms the subject of the theory of large deviations (see [43]). A classical result in this theory is the following theorem due to Cramér.

Theorem 2.1. *If the random process $\{X_n\}$ is i.i.d. with mean $\mu = \mathbb{E}(X_0)$ and finite moment generating function $M(t) := \mathbb{E}[e^{tX_0}] < +\infty$ for all $t > 0$, then*

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

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

We now give a general formulation of the large deviation principle (see [43]). Given an increasing sequence of integers $\{r_n\}$ and a lower semi-continuous function $I : \mathbb{R} \to [0, +\infty)$, we say that the random process $\{X_n\}$ satisfies a *large deviation principle* with normalizing sequence $\{r_n\}$ and rate function I, if for any closed set $F \subset \mathbb{R}$,

$$\limsup_{n\to+\infty} \frac{1}{r_n} \log \mathbb{P}\left[\frac{1}{n} S_n \in F\right] \leq - \inf_{x\in F} I(x),$$

and for any open set $G \subset \mathbb{R}$,

$$\liminf_{n\to+\infty} \frac{1}{r_n} \log \mathbb{P}\left[\frac{1}{n} S_n \in G\right] \geq - \inf_{x\in G} I(x).$$

We note that the large deviation principle holds under the assumptions of Theorem 2.1 with $r_n = n$ and the rate function specified in that theorem. For other large deviation principles, including Markov processes, see for instance [50].

Given a dynamical system (X, μ, T), any observable $\xi : X \to \mathbb{R}$ determines the random process $X_n = \xi \circ T^n$. Let $\langle \xi \rangle = \int_X \xi d\mu$ be the mean of this random process, i.e. the space average of the observable, and let $S_n\xi := \sum_{j=0}^{n-1} \xi \circ T^j$ be the corresponding sum process, i.e. the usual Birkhoff sums.

There are many results available regarding large deviations for dynamical systems (see for instance [32, 38, 45, 53]).

Starting with work of H. Furstenberg there has been interest in finding analogues of the classical limit theorems in probabilities for non-commuting random products. Let ν be a probability measure on the group $\mathrm{GL}(m, \mathbb{R})$ of invertible matrices, and let $g_0, g_1, g_2, \ldots$ be a matrix valued i.i.d. process with common distribution ν. We use the notation $g^{(n)}$ for the product process $g^{(n)} := g_{n-1} \ldots g_1 g_0$, and refer to this context as *Furstenberg's setting.*

Furstenberg and Kesten (see [20]) proved that the sequence $\frac{1}{n} \log\|g^{(n)}\|$ converges ν-a.s. to the maximal Lyapunov exponent

$$L_1(\nu) := \lim_{n\to\infty} \int_{\mathrm{GL}(m,\mathbb{R})} \frac{1}{n} \log\|g\| \nu^n(dg),$$

where ν^n stands for the nth convolution power of ν.

This convergence statement is the analogue of the law of large numbers for non-commuting products. The following theorem, due to E. Le Page (see [5, 36]), is the corresponding large deviation principle.

Theorem 2.2. *Let ν be a measure in $\mathrm{GL}(m, \mathbb{R})$ with finite exponential moment and such that the semigroup T_ν generated by the support of ν is strongly irreducible and contracting. Then there exist constants $c, \varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$ and $v \in \mathbb{R}^m$,*

$$\lim_{n\to+\infty} \frac{1}{n} \log \mathbb{P}\left[\left|\frac{1}{n} \log\|g^{(n)} v\| - L_1(\nu)\right| > \varepsilon\right] = -I(\varepsilon),$$

with rate function $I(\varepsilon) := \sup_{0<t<c}(\varepsilon t - \log \lambda(t) + tL_1(\nu))$, *and where* $\lambda(t)$ *denotes the maximum modulus eigenvalue of a Laplace–Markov family of operators* Q_t *associated with the distribution* ν.

A similar result for Markov processes was obtained by P. Bougerol (see [4]).

A more general setting for studying products of matrices is provided by linear cocycles. Given a base dynamical system (X, μ, T) and a measurable function $A : X \to \mathrm{Mat}(m, \mathbb{R})$ we call *linear cocycle* the skew-product map $F : X \times \mathbb{R}^m \to X \times \mathbb{R}^m$ defined by

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

The iterates of this map are given by $F^n(x, v) = (T^n x, A^{(n)}(x)v)$, where $A^{(n)}(x) := A(T^{n-1}x) \dots A(Tx)A(x)$. We will fix the base dynamics T and identify the cocycle F with the function A defining its fiber action.

A cocycle A is said to be *integrable* if $\log^+\|A\| \in L^1(\mu)$.

Given an ergodic system (X, μ, T) and an integrable cocycle A, by Kingman's ergodic theorem the following limits exist for all $1 \le j \le m$, and μ-a.e. $x \in X$,

$$\Lambda_j(A) = \lim_{n\to+\infty} \frac{1}{n} \log\|\wedge_j A^{(n)}(x)\| = \lim_{n\to+\infty} \frac{1}{n} \log \sum_{k=1}^{j} s_k(A^{(n)}(x)),$$

where given $g \in \mathrm{Mat}(m, \mathbb{R})$, $\wedge_j g$ denotes the jth exterior power of g, and the numbers $s_1(g) \ge s_2(g) \ge \cdots \ge s_m(g) \ge 0$ stand for its sorted singular values. The Lyapunov exponents of the cocycle A can then be characterized by

$$L_j(A) = \Lambda_j(A) - \Lambda_{j-1}(A) = \lim_{n\to+\infty} \frac{1}{n} \log s_j(A^{(n)}(x)),$$

with the convention that $\Lambda_0(A) = 0$. The Lyapunov spectrum of a cocycle A is the sequence of its Lyapunov exponents

$$L_1(A) \ge L_2(A) \ge \cdots \ge L_m(A) \ge -\infty.$$

We denote by $L_1^{(n)}(A) := \int_X \frac{1}{n} \log\|A^{(n)}\| d\mu$ the *finite scale* Lyapunov exponent of A, so that by Kingman's ergodic theorem

$$L_1(A) = \lim_{n\to\infty} L_1^{(n)}(A).$$

We note that Furstenberg's setting is obtained by choosing the base dynamics (X, μ, T) to be a Bernoulli shift, with $X = \mathrm{GL}(m, \mathbb{R})^{\mathbb{N}}$, $\mu = \nu^{\mathbb{N}}$ the product Bernoulli measure, $T(g_n)_{n\ge0} = (g_{n+1})_{n\ge0}$ the shift map, and the fiber action to be $A(g_n)_{n\ge0} = g_0$.

Moreover, the quasi-periodic setting refers to a base dynamics consisting of an ergodic finite dimensional torus translation and a fiber action which depends analytically on the base point.

The large deviation principle for sums of scalar, and respectively for products of matrix valued random processes, are asymptotic results. Our study of continuity properties of Lyapunov exponents of linear cocycles does not require asymptotic statements, but only good upper bounds on the measure of tail events, for the random processes given by the base and fiber dynamics. We call these bounds *large deviation type* (LDT) estimates.

To describe these LDT estimates we introduce the following formalism. From now on, $\underline{\epsilon}, \underline{\iota} : (0, \infty) \to (0, \infty)$ will represent functions that describe respectively, the size of the deviation from the mean and the measure of the deviation set. We assume that the deviation size functions $\underline{\epsilon}(t)$ are non-increasing. We assume that the deviation set measure functions $\underline{\iota}(t)$ are continuous and strictly decreasing to 0, as $t \to \infty$. We use the notation $\epsilon_n := \underline{\epsilon}(n)$ and $\iota_n := \underline{\iota}(n)$ for integers n.

Let $\mathcal{P}$ be a set of triplets $\underline{p} = (\underline{n_0}, \underline{\epsilon}, \underline{\iota})$, where $\underline{n_0}$ is an integer and $\underline{\epsilon}$ and $\underline{\iota}$ are deviation functions. An element $\underline{p} \in \mathcal{P}$ is referred to as an LDT parameter.

We now define the base and fiber LDT estimates.

Definition 2.1. An observable $\xi : X \to \mathbb{R}$ satisfies a base-LDT estimate w.r.t. a space of parameters $\mathcal{P}$ if for every $\epsilon > 0$ there is $\underline{p} = \underline{p}(\xi, \epsilon) \in \mathcal{P}$, $\underline{p} = (\underline{n_0}, \underline{\epsilon}, \underline{\iota})$, such that for all $n \geq \underline{n_0}$ we have $\epsilon_n \leq \epsilon$ and

$$\mu \left\{ x \in X : \left| \frac{1}{n} S_n \xi(x) - \langle \xi \rangle \right| > \epsilon_n \right\} < \iota_n. \tag{2.4}$$

Note that if an observable ξ satisfies a large deviation principle with rate function $I(\epsilon)$, then it also satisfies a base-LDT estimate with parameters $\underline{\epsilon}(t) \equiv \epsilon$ and $\underline{\iota}(t) \equiv e^{-tI(\epsilon)}$.

Definition 2.2. A measurable cocycle $A : X \to \mathrm{Mat}(m, \mathbb{R})$ satisfies a fiber-LDT estimate w.r.t. a space of parameters $\mathcal{P}$ if for every $\epsilon > 0$ there is $\underline{p} = \underline{p}(A, \epsilon) \in \mathcal{P}$, $\underline{p} = (\underline{n_0}, \underline{\epsilon}, \underline{\iota})$, such that for all $n \geq \underline{n_0}$ we have $\epsilon_n \leq \epsilon$ and

$$\mu \left\{ x \in X : \left| \frac{1}{n} \log\| A^{(n)}(x) \| - L_1^{(n)}(A) \right| > \epsilon_n \right\} < \iota_n. \tag{2.5}$$

In Furstenberg's setting, Theorem 2.2 implies the fiber-LDT estimate with parameters $\underline{\epsilon}(t) \equiv \epsilon$ and $\underline{\iota}(t) \equiv e^{-tI(\epsilon)}$.

We use LDT estimates to prove *continuity* of the Lyapunov exponents as functions of the cocycle, where the space of cocycles is endowed with a distance. For this we need a stronger form of the fiber-LDT, one that is *uniform* in a neighborhood of the cocycle, in the sense that estimate (2.5) holds with the same LDT parameter for all nearby cocycles.

Definition 2.3. A measurable cocycle A satisfies a uniform fiber-LDT if for all $\epsilon > 0$ there are $\delta = \delta(A, \epsilon) > 0$ and $\underline{p} = \underline{p}(A, \epsilon) \in \mathcal{P}$, $\underline{p} = (\underline{n_0}, \underline{\epsilon}, \underline{\iota})$, such that if B is a measurable cocycle with $\mathrm{dist}(B, A) < \delta$ and if $n \geq \underline{n_0}$ then $\epsilon_n \leq \epsilon$ and

$$\mu \left\{ x \in X : \left| \frac{1}{n} \log\| B^{(n)}(x) \| - L_1^{(n)}(B) \right| > \epsilon_n \right\} < \iota_n.$$

We remark that Theorem 2.2 does not provide a uniform fiber LDT estimate, and hence it cannot be employed directly in our scheme for proving continuity of

Lyapunov exponents. The same remark applies to the Markov case studied in [4]. However, the spectral theory approach developed in these works can be adapted to derive uniform fiber LDT estimates.

Proving base and fiber (uniform) LDT estimates for quasi-periodic cocycles uses harmonic analysis and potential theory tools, along with the arithmetic properties of the torus translation.

3. Deriving LDT for Quasi-Periodic Cocycles

The goal of this section is to describe some of the methods used for deriving LDT estimates for quasi-periodic models. Central to these methods are the use of the subharmoncity of quantities related to the iterates of the cocycle, and the use of the arithmetic properties of the frequency defining the base dynamics. We formulate the main assumptions on the model. We review relevant previous results. We indicate the main steps in the invertible case by reducing the proof of fiber LDT estimates to base LDT estimates on subharmonic functions, with parameters depending only on certain uniform measurements on the observable. We give some hints of the machinery behind the proof of such base LDT estimates by considering a toy model of subharmonic functions. We indicate some of the difficulties and the ways to overcome them in the non-invertible (but not identically singular) case.

3.1. *The model*

Let $Tx = x + \omega$ be the translation on the torus $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$, $d \geq 1$, by a rationally independent vector ω. This ergodic map defines the base dynamics, and it is assumed fixed.

Let $A : \mathbb{T}^d \to \mathrm{Mat}(m, \mathbb{R})$ be a matrix valued real *analytic* function, so $A(x)$ has an extension $A(z)$ to $\mathcal{A}_r^d = \mathcal{A}_r \times \cdots \times \mathcal{A}_r$, where $\mathcal{A}_r$ is the annulus $\{z \in \mathbb{C} : 1 - r < |z| < 1 + r\}$ of width $2r$.

The iterates $A^{(n)}(x) := A(x + (n-1)\omega) \cdots A(x + \omega)A(x)$ of the cocycle are also analytic on $\mathcal{A}_r^d$.

In order to treat occurrences of small denominators, the translation vector will be assumed to satisfy a generic *Diophantine condition*:

$$\|k \cdot \omega\| \geq \frac{t}{|k|^{d+\delta_0}} \tag{3.1}$$

for some $t > 0$, $\delta_0 > 0$ and for all $k \in \mathbb{Z}^d \backslash \{0\}$, where for any real number x, $\|x\| := \min_{k \in \mathbb{Z}} |x - k|$.

For every integer $m \geq 1$, let $C_r^\omega(\mathbb{T}^d, \mathrm{Mat}(m, \mathbb{R}))$ be the vector space of matrix valued analytic functions on $\mathcal{A}_r^d$, with a continuous extension up to the boundary. Endowed with the norm $\|A\|_r := \sup_{z \in \mathcal{A}_r^d} \|A(z)\|$, $C_r^\omega(\mathbb{T}^d, \mathrm{Mat}(m, \mathbb{R}))$ is a Banach space.

In a previous work (see [12]) we studied $\mathrm{GL}(m, \mathbb{R})$-valued analytic cocycles. Here we will allow our cocycles to have *singularities* (i.e. points of non-invertibility), as long as they are not identically singular (which in particular ensures that all Lyapunov exponents are finite).

3.2. *Literature review*

Fiber LDT estimates for quasi-periodic cocycles were first obtained in the context of studying spectral properties of discrete, one-dimensional, quasi-periodic Schrödinger operators. These operators, denoted by $H(x)$, act on $l^2(\mathbb{Z}, \mathbb{R}) \ni \psi = \{\psi_n\}_n$ by

$$[H(x)\psi]_n := -(\psi_{n+1} + \psi_{n-1} - 2\psi_n) + v(T^n x)\psi_n, \tag{3.2}$$

and describe the Hamiltonian of a quantum particle on the lattice $\mathbb{Z}$.

The term $[\Delta\psi]_n := (\psi_{n+1} + \psi_{n-1} - 2\psi_n)$ defines the discrete Laplacian, while $v(T^n x) = v(x + n\omega)$ is the potential at site n on the integer lattice. The potential is defined by a function $v : \mathbb{T}^d \to \mathbb{R}$ which is assumed analytic.

The associated discrete Schrödinger (i.e. eigenvalue) equation

$$[H(x)\psi]_n = E\psi_n$$

for the state $\psi = \{\psi_n\}_n$ and the energy E is a second order finite differences equation, which is solved formally by the iterates of the cocycle:

$$A_E(x) := \begin{bmatrix} v(x) + 2 - E & -1 \\ 1 & 0 \end{bmatrix} \in \mathrm{SL}(2, \mathbb{R}).$$

The cocycle (or rather the one-parameter family of cocycles indexed by the energy parameter $E \in \mathbb{R}$) $A_E(x)$ is called a quasi-periodic *Schrödinger* cocycle. Properties such as uniformity of the LDT estimates, or continuity of the Lyapunov exponent are understood with respect to this energy parameter. Note that $A_E(x) \in \mathrm{SL}(2, \mathbb{R})$ so $A_E^{(n)}(x) \in \mathrm{SL}(2, \mathbb{R})$, hence $\|A_E^{(n)}(x)\| \geq 1$ for all n, x, E.

LDT estimates for quasi-periodic Schrödinger cocycles play an important role in the study of the spectral properties of the corresponding operator (see J. Bourgain's monograph [6]) as well as in the study of quantitative positivity and continuity properties of the Lyapunov exponent (regarded as a function of the energy E) and of a related physical quantity called the integrated density of states (see for instance [6, 24]).

The first such estimates were obtained by J. Bourgain and M. Goldstein in [8] and used to establish pure point spectrum with exponentially decaying eigenfunctions for the operator (3.2) with potential $v_\lambda(x) = \lambda v(x)$ and $\lambda \gg 1$.

Phrased in the language we have introduced in Sec. 2, the results obtained in [8] (for both $d = 1$ ad $d > 1$) provide fiber LDT estimates with deviation size function $\underline{\epsilon}(t) \equiv t^{-a}$ and deviation measure function $\underline{\iota}(t) \equiv e^{-t^b}$ for some absolute constants $a, b > 0$ and threshold of applicability $n \geq \underline{n_0}$ depending on the Diophantine condition (3.1), the sup-norm of $v(x)$ and the size of E (which can be taken in a fixed compact). Therefore, this fiber LDT is *uniform* in E.

Later, M. Goldstein and W. Schlag (see [23]) proved sharper fiber LDT estimates in the one variable case $d = 1$, assuming a stronger Diophantine condition. Their result was subsequently improved in [52].

LDT estimates for other related models, but with stronger limitations, were proven afterwards, see for instance [10, 9, 7, 33, 34].

Uniform fiber LDT estimates for more general, $\mathrm{Mat}(2, \mathbb{R})$-valued cocycles admitting *singularities* (e.g., points where they are not invertible) were obtained in the *one-frequency* torus translation case in [29] (see also references therein). In our work (see [15, 17]), we obtain uniform fiber LDT estimates for higher dimensional, $\mathrm{Mat}(m, \mathbb{R})$-valued cocycles with singularities, for both one and *several* variables torus translations.

3.3. *Main ingredients for proving fiber LDT*

We first note that for a *continuous* observable ξ, the convergence in Birkhoff's ergodic theorem is *uniform*, hence the base LDT estimate (2.4) holds automatically with $\underline{\epsilon}(t) \equiv \epsilon$, $\underline{\iota}(t) \equiv 0$ and $\underline{n_0}$ depending on ϵ and ξ, but in a very *non-explicit* way.

We will show how the proof of fiber LDT estimates can be reduced to having base LDT estimates for a certain class of observables, with the LDT parameters depending *very explicitly* on the observable (thus ensuring strong uniformity).

Given $A \in C^\omega_r(\mathbb{T}^d, \mathrm{Mat}(m, \mathbb{R}))$, if we denote by

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

then the fiber LDT estimate (2.5) can be written as

$$|\{x \in \mathbb{T}^d : |u_A^{(n)}(x) - \langle u_A^{(n)}\rangle| > \epsilon_n\}| < \iota_n. \tag{3.3}$$

We will establish such an estimate with $\epsilon_n = n^{-a}$ and $\iota_n = e^{-cn^b}$, for some absolute constants $a, b, c > 0$.

Let $d = 1$. Since the maps $A^{(n)}(z)$ are holomorphic on $\mathcal{A}_r$, the maps $u_A^{(n)}(z)$ are *subharmonic* on $\mathcal{A}_r$ (a good reference on subharmonic functions is [27]).

When $d > 1$, the maps $u_A^{(n)}(z)$ are *pluri subharmonic* on $\mathcal{A}_r^d$, i.e. they are subharmonic along any complex line, and in particular, they are subharmonic in each coordinate.

(Pluri) subharmonicity of the maps $u_A^{(n)}(z)$ associated with the iterates of the cocycles will play a crucial role in deriving fiber LDT estimates.

Note that these maps always have the trivial upper bound

$$u_A^{(n)}(z) \le \log\|A\|_r \quad \text{for all } z \in \mathcal{A}_r^d,\ \ n \in \mathbb{N}.$$

Moreover, if $A(x) \in \mathrm{SL}(2, \mathbb{R})$, which is the case of the Schrödinger cocycles, then we also have the trivial lower bound

$$u_A^{(n)}(z) \ge 0 \quad \text{for all } z \in \mathcal{A}_r^d,\ \ n \in \mathbb{N}.$$

More generally, using Cramer's formula, $\mathrm{GL}(m,\mathbb{R})$-valued cocycles $A(x)$ have the uniform bounds

$$-\log\|A^{-1}\|_r \le u_A^{(n)}(z) \le \log\|A\|_r \quad \text{for all } z \in \mathcal{A}_r^d,\ n \in \mathbb{N}.$$

Therefore, if $A \in C_r^\omega(\mathbb{T}^d, \mathrm{GL}(m,\mathbb{R}))$, then the maps $u_A^{(n)}(z)$ are pluri subharmonic and uniformly bounded on $\mathcal{A}_r^d$ (the latter will not be the case for singular cocycles, which we discuss separately in Sec. 3.5).

Of course, not all function sequences $u^{(n)}(x)$ with a (pluri) subharmonic, uniformly bounded extension satisfy an estimate like (3.3).

However, the function sequences $u_A^{(n)}(x)$ have another crucial feature: they are *almost invariant* under the base transformation, in the sense that

$$|u_A^{(n)}(x) - u_A^{(n)}(Tx)| \lesssim \frac{1}{n} \quad \text{for all } x \in \mathbb{T}^d,\ n \ge 1. \tag{3.4}$$

This holds provided $A(x) \in \mathrm{GL}(m,\mathbb{R})$, and can be seen through a simple computation.

If we apply this almost invariance $m \asymp n^{1-\epsilon}$ times, we get

$$\left| u_A^{(n)}(x) - \frac{1}{m}\sum_{j=0}^{m-1} u_A^{(n)}(T^j x) \right| \lesssim \frac{m}{n} = n^{-\epsilon} \quad \text{for all } x \in \mathbb{T}^d,\ n \ge 1$$

or

$$\left| u_A^{(n)}(x) - \frac{1}{m} S_m u_A^{(n)}(x) \right| \lesssim n^{-\epsilon} \quad \text{for all } x \in \mathbb{T}^d,\ n \ge 1.$$

Therefore, to derive the fiber LDT (3.3), it is enough to prove that

$$\left| \left\{ x \in \mathbb{T}^d : \left| \frac{1}{m} S_m u_A^{(n)}(x) - \langle u_A^{(n)} \rangle \right| > m^{-a} \right\} \right| < e^{-cm^b}, \tag{3.5}$$

for some absolute constants $a, b, c > 0$.

In other words, it would be enough to prove *base* LDT estimates for the maps $u_A^{(n)}$, but in a way that they apply *uniformly* for all $n \ge 1$ and for any cocycle near A. Therefore, uniform fiber LDTs follow from base LDT for (pluri) subharmonic observables, provided the LDT parameters depend only on some *measurements* on the observable (such as the width r of its domain and its sup-norm) and on the base dynamics (which is assumed fixed).

3.4. *Estimates on subharmonic functions*

The goal here is to show that given a (pluri) subharmonic function $u(z)$ or $\mathcal{A}_r^d$,

$$|\{x \in \mathbb{T}^d : |S_n u(x) - n\langle u \rangle| > n^{1-a}\}| < e^{-cn^b} \tag{3.6}$$

for some universal constants $a, b, c > 0$ and for all $n \ge \underline{n_0}$, where $\underline{n_0}$ may only depend on some uniform measurements on the observable u (and on the Diophantine condition on ω).

We describe some ideas used to derive (3.6). To warm up, let $d = 1$ and consider a very simple (yet relevant) example of a subharmonic function, $u(z) = \log|z - 1|$, whose restriction to $\mathbb{T}$ has the form $u(x) = \log|e(x) - 1|$, where we use the notation $e(x) := e^{2\pi i x}$. We will prove the base-LDT (3.6) for this subharmonic function, following W. Schlag's address at the 2003 ICMP in Lisbon (see [46]).

Applying John–Nirerberg's inequality (2.2) to the function $S_n u(x)$, we have

$$|\{x \in \mathbb{T} : |S_n u(x) - n\langle u \rangle\}| > n^{1-a}| < e^{-cn^{1-a}/\|S_n u\|_{\mathrm{BMO}}}. \tag{3.7}$$

Therefore, in order to establish (3.6), it would be enough to derive a good upper bound on the BMO norm of $S_n u$, namely to prove at least that $\|S_n u\|_{\mathrm{BMO}} = o(n)$ as $n \to \infty$.

One should note that $u(x) = \log|e(x) - 1| \notin L^\infty(\mathbb{T})$, and in fact this is a standard example of a BMO function which is not in L^∞. Even if our subharmonic function $u(x)$ were bounded, so that $\|S_n u\|_{L^\infty} \lesssim n$, estimating the BMO-norm by the L^∞-norm would only give $\|S_n u\|_{\mathrm{BMO}} \lesssim n$ which clearly is not enough for our purposes.

Lemma 3.1. *Let* $u(x) = \log|e(x) - 1|$. *Then*

$$\|S_n u\|_{\mathrm{BMO}} \lesssim n^\delta \tag{3.8}$$

for some $0 < \delta < 1$ *which depends only on the Diophantine condition on* ω.

Proof. As noted earlier, L^∞ is a proper subset of BMO, with $u(x)$ being a typical BMO but not L^∞ function. However, u is the image of an L^∞ function via a singular integral operator, the *Hilbert transform*.

Informally, we could think of the Hilbert transform $\mathcal{H}$ of a function on $\mathbb{T}$ as being given by the boundary values of the harmonic conjugate of its harmonic extension to the unit disk (see Chap. 3 in [41]).

Let

$$s(x) := \begin{cases} \{x\} - \dfrac{1}{2} & \text{if } x \notin \mathbb{Z} \\[2ex] 0 & \text{if } x \in \mathbb{Z} \end{cases}$$

be the saw-tooth function, where $\{x\}$ is the fractional part of x.

An elementary calculation that involves computing $\arg(e(x) - 1)$ shows that

$$u(x) = \log|e(x) - 1| = \mathcal{H}(s)(x),$$

hence, since the Hilbert transform commutes with translations,

$$S_n u(x) = \sum_{j=0}^{n-1} \log|e(x + j\omega) - 1| = \mathcal{H}\left(\sum_{j=0}^{n-1} s(\cdot + j\omega)\right)(x).$$

A deep result in harmonic analysis, due to C. Fefferman, implies that the Hilbert transform is a bounded operator from L^∞ to BMO (see Chap. 6 in [18]).

Therefore,

$$\|S_n u\|_{\mathrm{BMO}} = \left\| \mathcal{H}\left(\sum_{j=0}^{n-1} s(\cdot + j\omega) \right) \right\|_{\mathrm{BMO}} \lesssim \left\| \sum_{j=0}^{n-1} s(\cdot + j\omega) \right\|_{L^\infty}. \tag{3.9}$$

It turns out that the L^∞ norm of $\sum_{j=0}^{n-1} s(\cdot + j\omega)$ has a number theoretical meaning: it is comparable to the *discrepancy* of the sequence $\{j\omega\}$. A good reference for the discussion following below is [40].

The discrepancy of a sequence $\{x_j\}_{j\in\mathbb{N}}$ of points on $\mathbb{T}$ measures how well distributed they are on the torus.

We may define it as

$$D_n\{x_j\} := \sup_{\alpha,\beta} |\#\{0 \le j \le n-1 : x_j \in [\alpha, \beta)\} - n(\beta - \alpha)|.$$

We say that the sequence $\{x_j\}$ is uniformly distributed on the torus if $D_n\{x_j\} = o(n)$ as $n \to \infty$. Clearly

$$\begin{aligned} D_n\{x_j\} &= \sup_{\alpha,\beta} \left| \sum_{j=0}^{n-1} \mathbb{1}_{[\alpha,\beta)}(x_j) - (\beta - \alpha) \right| \\ &= \sup_{\alpha,\beta} \left| \sum_{j=0}^{n-1} \mathbb{1}_{[0,\beta-\alpha)}(x_j - \alpha) - (\beta - \alpha) \right|, \end{aligned}$$

where $\mathbb{1}_I(x)$ is the indicator function of the interval I.

One can easily verify the following relationship between the indicator function of an interval and the saw-tooth function $s(x)$:

$$\mathbb{1}_{[0,\alpha)}(x) = s(x - \alpha) - s(x) + \alpha, \tag{3.10}$$

valid if $x \neq 0$, $x \neq \alpha \pmod 1$.

Define

$$\Delta_n\{x_j\} := \sup_{\alpha} \left| \sum_{j=0}^{n-1} s(x_j - \alpha) \right|.$$

Using (3.10), one can show that

$$\Delta_n\{x_j\} \le D_n\{x_j\} \le 2\Delta_n\{x_j\},$$

hence $\Delta_n\{x_j\}$ can be regarded as another kind of discrepancy.

Recall from (3.9) that

$$\begin{aligned} \|S_n u\|_{\mathrm{BMO}} &\lesssim \left\| \sum_{j=0}^{n-1} s(\cdot + j\omega) \right\|_{L^\infty} = \sup_{\alpha} \left\| \sum_{j=0}^{n-1} s(j\omega - \alpha) \right\| \\ &= \Delta_n\{j\omega\} \le D_n\{j\omega\}. \end{aligned}$$

This shows that in order to get an estimate on the BMO norm of $S_n u$, we need an estimate on the discrepancy of the sequence $\{j\omega\}_{j\in\mathbb{N}}$. It is an elementary fact that this sequence is uniformly distributed for any irrational ω, hence $D_n\{j\omega\} = o(n)$ as $n \to \infty$.

However, we need a more *quantitative* estimate, and this is where the Diophantine condition (which describes the irrationality of ω quantitatively) comes into play. We recall the following classical Erdös–Turán inequality (see [40]): for any sequence $\{x_j\} \subset \mathbb{T}$ and for any integers n, K we have:

$$D_n\{x_j\} \leq \frac{n}{K+1} + 3\sum_{k=1}^{K} \frac{1}{k} \left| \sum_{j=0}^{n-1} e(kx_j) \right|. \tag{3.11}$$

Apply this inequality to the sequence $x_j = j\omega$, so the exponential sum in (3.11) becomes

$$\sum_{j=0}^{n-1} e(kx_j) = \sum_{j=0}^{n-1} e(kj\omega) = \frac{e(kn\omega) - 1}{e(k\omega) - 1},$$

hence, using the Diophantine condition on ω

$$\left| \sum_{j=0}^{n-1} e(kj\omega) \right| \leq \frac{1}{|e(k\omega) - 1|} = \frac{1}{\|k\omega\|} \lesssim |k|^{1+\delta_0}.$$

Erdös–Turán's inequality (3.11) then says

$$D_n\{j\omega\} \leq \frac{n}{K+1} + 3\sum_{k=1}^{K} k^{\delta_0} \lesssim \frac{n}{K} + K^{\delta_0+1} \lesssim n^{\delta}$$

provided we choose $K \asymp n^{1/(\delta_0+2)}$, so $\delta = \frac{\delta_0+1}{\delta_0+2} \in (0,1)$. □

This proof contains many of the ingredients necessary to derive the base LDT estimate (3.6) for subharmonic functions: BMO estimates, John–Nirerberg's inequality, a quantitative description of the uniform distribution of the sequence $\{j\omega\}$.

While general subharmonic functions are more complex, and they require a finer analysis, the difficult part comes from handling certain sums of functions not unlike $\log|z-1|$. Indeed, standard examples of subharmonic functions are $u(z) = \log|f(z)|$, for some holomorphic function $f(z)$. If $\zeta_1, \ldots, \zeta_N$ are the zeros of the $f(z)$ in a compact set Ω, then on that set $f(z) = g(z) \cdot \prod_{j=1}^{N}(z - \zeta_j)$, where $g(z)$ is analytic and free of zeros, hence $u(z) = \log|g(z)| + \sum_{j=1}^{N} \log|z - \zeta_j|$.

Since $g(z)$ has no zeros on Ω, $h(z) := \log|g(z)|$ is harmonic. Putting $d\mu(\zeta) := \sum_{j=1}^{N} \delta z_j(\zeta)$ to be the sum of the Dirac measures corresponding to the zeros of $f(z)$, we obtain the following representation of $u(z)$ on Ω:

$$u(z) = h(z) + \int \log|z - \zeta| d\mu(\zeta).$$

In fact, by the Riesz representation theorem, *any* subharmonic function $u(z)$ has such a representation, for some harmonic function $h(z)$ and some compactly supported measure μ called its Riesz measure.

Most of the time, due to its smoothness, the harmonic part is harmless, and the difficult part in obtaining the desired estimates is reduced to the study of the logarithmic potential $\int \log|z-\zeta| d\mu(\zeta)$.

We keep $d = 1$ (i.e. the one variable translation case) for now, and describe some of the steps in the proof of (3.6) for general subharmonic functions.

We first introduce additional assumptions concerning certain uniform measurements on the function $u(z)$. One number that will stay fixed throughout is the width r of its domain. The other is the "Riesz mass" of u, i.e. the total mass $\|\mu\|$ of its Riesz measure.

As noted earlier, in the case of Schrödinger, thus $\mathrm{SL}(2,\mathbb{R})$-valued cocycles, the subharmonic functions $u_A^{(n)}(z)$ corresponding to its iterates have the trivial bounds

$$0 \le u_A^{(n)}(z) \le \log\|A\|_r \quad \text{for all } z \in \mathcal{A}_r,\ \ n \in \mathbb{N},$$

and in the case of $\mathrm{GL}(m,\mathbb{R})$-valued cocycles we have the bounds

$$-\log\|A^{-1}\|_r \le u_A^{(n)}(z) \le \log\|A\|_r.$$

In any case, there is a constant $C(A) < \infty$, which is stable under perturbations of the cocycle, such that

$$\sup_{z\in\mathcal{A}_r} |u_A^{(n)}(z)| \le C(A) \quad \text{for all } n \in \mathbb{N},$$

hence the maps $u_A^{(n)}$ are uniformly bounded on $\mathcal{A}_r$ in n and A.

It turns out that the width r of the domain of a subharmonic function $u(z)$ and its sup-norm C over the domain completely determine its Riesz mass.

Subharmonic functions, however, may not be bounded from below, and in fact they may attain the value $-\infty$. That will be the case with the subharmonic functions associated to iterates of a cocycle with *singularities*. A quantitative version of the Riesz representation theorem due to M. Goldstein and W. Schlag (see [25]) implies an estimate on the Riesz measure under the weaker conditions:

$$\sup_{z\in\mathcal{A}_r} u(z) \le C \quad \text{and} \quad \sup_{x\in\mathbb{T}} u(x) \ge -C. \tag{3.12}$$

Then $\|\mu\| \lesssim \mathcal{S}$, where $\mathcal{S}$ is a constant depending only on r and C.

A crucial ingredient in proving (3.6) for an observable $u(x)$ is having an estimate on the decay of its Fourier coefficients, one that depends only on some measurements on $u(x)$, thus applying uniformly to all maps $u_A^{(n)}(x)$ corresponding to iterates of the cocycle (or to iterates of nearby cocycles). One should note that general subharmonic functions lack smoothness, hence such estimates are nontrivial.

Lemma 3.2. *Let $u(x)$ be a function on $\mathbb{T}$ with a subharmonic extension to $\mathcal{A}_r$. Assume that its Riesz mass is bounded by $\mathcal{S}$. Then its Fourier coefficients have the decay*

$$|\hat{u}(k)| \lesssim \mathcal{S} \cdot \frac{1}{|k|} \quad \textit{for all } k \neq 0. \tag{3.13}$$

Proof. We only sketch the proof of this result for the toy model $u(x) = \log|e(x)-1|$. The general case expands upon this simple example and uses the Riesz representation theorem (see [6, 15] for details). We use again the fact that $u = \mathcal{H}(s)$, where $s(x)$ is the saw-tooth function.

It is easy to see from its definition that the Hilbert transform is related to the Fourier transform via the identity $\widehat{\mathcal{H}(f)}(k) = -i\,\mathrm{sign}(k)\hat{f}(k)$, for all $k \in \mathbb{Z}$, $k \neq 0$. Then

$$|\hat{u}(k)| = |\widehat{\mathcal{H}(s)}(k)| = |\hat{s}(k)| \lesssim \frac{1}{|k|},$$

where the last inequality follows from a direct elementary calculation of the Fourier coefficients of the saw-tooth function. $\square$

Proposition 3.1. *Let $u(x)$ be a function on $\mathbb{T}$ with a subharmonic extension to $\mathcal{A}_r$. Assume that its Riesz mass is bounded by $\mathcal{S}$. Then for some explicit constants $a, b, c > 0$ and for all $n \geq \underline{n_0}$, where $\underline{n_0}$ depends only on ω, we have:*

$$\left|\left\{x \in \mathbb{T} : \left|\frac{1}{n}S_n u(x) - \langle u \rangle\right| > \mathcal{S} n^{-a}\right\}\right| < e^{-cn^b}. \tag{3.14}$$

Proof. We sketch the argument. Expand $u(x)$ into its Fourier series

$$u(x) = \langle u \rangle + \sum_{k\neq 0} \hat{u}(k) e(kx).$$

Then

$$\begin{aligned}\frac{1}{n}S_n u(x) &= \frac{1}{n}\sum_{j=0}^{n-1} u(x + j\omega)\\ &= \langle u \rangle + \sum_{k\neq 0} \hat{u}(k)e(kx)\left(\frac{1}{n}\sum_{j=0}^{n-1} e(jk\omega)\right)\\ &= \langle u \rangle + \sum_{k\neq 0} \hat{u}(k)e(kx)K_n(k\omega),\end{aligned}$$

where we denoted by $K_n(t)$ the Fejér kernel

$$K_n(t) = \frac{1}{n}\sum_{j=0}^{n-1} e(jt) = \frac{1}{n}\frac{1-e(nt)}{1-e(t)},$$

which clearly has the bound

$$|K_n(t)| \leq \min\left\{1, \frac{1}{n\|t\|}\right\}. \tag{3.15}$$

We write

$$\frac{1}{n}S_n u(x) - \langle u \rangle = \sum_{k\neq 0} \hat{u}(k)K_n(k\omega)e(kx),$$

and the goal is to estimate the above in the L^2-norm and then to apply Chebyshev's inequality. By Parseval,

$$\left\| \frac{1}{n} S_n u - \langle u \rangle \right\|_{L^2}^2 = \sum_{k \neq 0} |\hat{u}(k)|^2 |K_n(k\omega)|^2$$

$$= \sum_{0<|k|<K} |\hat{u}(k)|^2 |K_n(k\omega)|^2 + \sum_{|k| \geq K} |\hat{u}(k)|^2 |K_n(k\omega)|^2.$$

The sum above was split into two parts, with the splitting point chosen to optimize the sum of the estimates. In the second sum we rely only on the decay (3.13) of the Fourier coefficients (provided K is large enough), and simply bound the Fejér kernel by 1. In the first sum, the weakness of the decay of the Fourier coefficients is compensated by the decay of the Fejér kernel. This is ensured by the Diophantine condition (3.1) on ω, since from (3.15) we have:

$$|K_n(k\omega)| \leq \frac{1}{n\|k\omega\|} \lesssim \frac{|k|^{1+\delta_0}}{n}.$$

In the end, for some constant $a > 0$, we get the following:

$$\left\| \frac{1}{n} S_n u - \langle u \rangle \right\|_{L^2}^2 \lesssim \mathcal{S} n^{-a}$$

and by Chebyshev,

$$\left| \left\{ x \in \mathbb{T} : \left| \frac{1}{n} S_n u(x) - \langle u \rangle \right| > \mathcal{S} n^{-a/3} \right\} \right| < n^{-4a/3}. \tag{3.16}$$

This is an LDT estimate, but of a much weaker form than needed, since the measure of the exceptional set decays only polynomially in the scale. We need a way to *boost* such a weak LDT estimate to a much stronger one, and this is done again through the use of BMO estimates and John–Nirenberg's inequality.

The following is a crucial result called the splitting lemma (see [6, 25, 46, 15] for full details on its proof).

Lemma 3.3. *Let $u(x)$ be a function as in Lemma* 3.2. *Assume that the following a priori estimate holds*:

$$|\{x \in \mathbb{T} : |u(x) - \langle u \rangle| > \epsilon_0\}| < \epsilon_1, \tag{3.17}$$

where $\epsilon_1 \ll \epsilon_0$. Then

$$\|u\|_{\mathrm{BMO}} \lesssim \epsilon_0 + (\mathcal{S}\epsilon_1)^{1/2}.$$

This lemma is then applied with $\frac{1}{n} S_n u(x)$ playing the role of $u(x)$, $\epsilon_0 := \frac{C}{r} n^{-a/3}$ and $\epsilon_1 := n^{-4a/3}$, hence the assumption (3.17) follows from the weak LDT (3.16) and it leads to a BMO estimate on $\frac{1}{n} S_n u$. Applying John–Nirenberg's inequality (2.2), we obtain the desired stronger LDT (3.14). □

Let us briefly discuss the *multivariable* case. For simplicity, let $d = 2$. The maps $u_A^{(n)}(z_1, z_2)$ are now *pluri* subharmonic, i.e. subharmonic in each variable (in fact, along any complex line).

The proof of Proposition 3.1 in the multivariable case follows the same procedure as in the one variable case. However, pluri subharmonic functions do not have a representation like the one given by Riesz' theorem for subharmonic functions. Because of this, proving (3.14) requires a more delicate analysis that involves applying the technical results (i.e. Lemma 3.2 on the decay of the Fourier coefficients and Lemma 3.17, the splitting lemma) in *each* variable.

In order to do that, we need uniform estimates on the Riesz mass of the restrictions of $u(z_1, z_2)$ along each horizontal and vertical line. In the Schrödinger, $\mathrm{SL}(2, \mathbb{R})$, or more generally, $\mathrm{GL}(m, \mathbb{R})$-valued cocycles cases, this is automatic, as the maps $u_A^{(n)}(z_1, z_2)$ to which we need to apply there estimates are uniformly bounded (in n and A).

We will discuss in the next subsection the singular case, where these bounds do not hold.

3.5. *Singular but non-identically singular case*

In this section we discuss proving uniform fiber LDT estimates for analytic cocycles $A : \mathbb{T}^d \to \mathrm{Mat}(m, \mathbb{R})$, with $\det[A(x)] \not\equiv 0$ (see our preprint [15] and the upcoming monograph [17] for full details). We refer to these as cocycles with *singularities*, as they may have points of non-invertibility, but we assume that they are not identically singular. In particular, by analyticity, this ensures that the set of singularities of such a cocycle and of its iterates has zero Lebesgue measure.

As mentioned earlier, uniform fiber LDT estimates for cocycles with singularities were obtained in [29] in the *one variable* case ($d = 1$) and for $m = 2$.

The issue of singularity is especially delicate in the *several* variables case. One obstacle, for instance, is the fact that an analytic function of several variables may vanish identically along some hyperplanes, while not being globally identically zero. Related to this, a pluri subharmonic function may be identically $-\infty$ along some hyperplanes, while not being globally $-\infty$.

A crucial tool in our analysis is a result that shows that the obstacle described above for an analytic function can be removed with an appropriate change of coordinates. Another crucial tool in our analysis is the observation that while the pluri subharmonic functions corresponding to iterates of a cocycle may have singularities as described above, these singularities can be captured by certain *analytic* functions.

Here are more details regarding this approach.

Let $A(x)$ be a cocycle, $A(z)$ be its complex extension to $\mathcal{A}_r^d$ and denote by $f_A(z) := \det[A(z)]$ its determinant, which is an analytic function on $\mathcal{A}_r^d$, assumed non-identically zero.

We are going to obtain some bounds on the pluri subharmonic functions $u_A^{(n)}(z)$ associated with the iterates of the cocycle. The upper bound $u_A^{(n)}(z) \le \log\|A\|_r < \infty$

is trivial. To obtain an estimate from below, we apply Cramer's rule: $\det[M] \cdot I = M \cdot \mathrm{adj}(M)$ to the matrices $M = A^{(n)}(x)$ and get:

$$\begin{aligned}\prod_{i=0}^{n-1} f_A(T^i z) \cdot I &= \prod_{i=0}^{n-1} \det[A(T^i z)] \cdot I \\ &= A(T^{n-1}z) \cdots A(z) \cdot \mathrm{adj}(A(z)) \cdots \mathrm{adj}(A(T^{n-1}z)) \\ &= A^{(n)}(z) \cdot \mathrm{adj}(A(z)) \cdots \mathrm{adj}(A(T^{n-1}z)).\end{aligned}$$

Clearly

$$\|\mathrm{adj}(A(z))\| \lesssim \|A(z)\|^{m-1} \leq \|A\|_r^{m-1} \quad \text{for all } z.$$

This, together with the trivial upper bound, imply the following:

$$-C_1 + \frac{1}{n}\sum_{i=0}^{n-1} \log|f_A(T^i z)| \leq u_A^{(n)}(z) \leq C_1, \tag{3.18}$$

for some $C_1 = C_1(A) \sim |\log\|A\|_r|$ and for all $z \in \mathcal{A}_r^d$.

In other words, while $u_A^{(n)}(z)$ may fail to be bounded from below, and in fact it may be identically $-\infty$ along some hyperplanes, these singularities are captured by averages of a simpler pluri subharmonic function, $\log|f_A(z)|$, where $f_A(z)$ is *analytic* and $f_A(z) \not\equiv 0$.

The bounds (3.18) are *stable* under perturbations, in the sense that if $B \approx A$, then $C_1(B) \approx C_1(A)$ and $f_B \approx f_A$.

Let $d = 2$, for simplicity. Our goal is to ensure that (3.12) hold for the functions $u_A^{(n)}(z)$ (uniformly in n and A) along each horizontal and vertical line, i.e. that:

$$\sup_{z_2 \in \mathcal{A}_r} u_A^{(n)}(z_1, z_2) \leq C_1 \quad \text{and} \quad \sup_{x_2 \in \mathbb{T}} u_A^{(n)}(x_1, x_2) \geq -C_2 \tag{3.19}$$

hold for all $z_1 \in \mathcal{A}_r$, all $x_1 \in \mathbb{T}$, as well as with the roles of the variables interchanged.

The second bound in (3.19) is obviously wrong along lines where $u_A^{(n)}$ is identically $-\infty$.

To circumvent this obstacle, we show that given an analytic function f on $\mathcal{A}_r^d$ with $f \not\equiv 0$, there is a global change of coordinates $x' = Mx$ on $\mathbb{T}^d$, given by some matrix $M \in \mathrm{SL}(d, \mathbb{Z})$, such that in the *new coordinates*, f does not vanish identically along any horizontal or vertical lines. Moreover, this applies uniformly in a neighborhood of f.

By replacing $u_A^{(n)}$ by $u_A^{(n)} \circ M$, f_A by $f_A \circ M$ and ω by $M^{-1}\omega$, we may assume that in (3.18) the analytic function f_A does not vanish identically along any horizontal or vertical lines, and hence $\log|f_A(z)|$ is not identically $-\infty$ along any such lines. Using this, but not without more additional effort, we can ensure that (3.19) hold, and with that, we can proceed as in the previous subsection to obtain an appropriate base-LDT for pluri subharmonic functions, applicable uniformly to the maps $u_A^{(n)}(z)$.

The only ingredient needed to derive the fiber LDTs from such strongly uniform base LDT estimates, is an almost invariance principle like (3.4). This principle does not necessarily hold for cocycles with singularities, because when we estimate $|u_A^{(n)}(x) - u_A^{(n)}(Tx)|$, we end up having to bound a term of the form $\log\|A^{-1}(x)\|$.

An upper bound on $\|A^{-1}(x)\|$ is correlated with a lower bound on $f_A(x) = \det[A(x)]$. Therefore, we need to understand quantitatively the set where $f_A(x) \approx 0$, which means deriving a Łojaziewicz-type inequality.

More precisely, it can be shown (see [23, 33, 34]), that given an analytic function f on $\mathcal{A}_r^d$ with $f \not\equiv 0$, there are constants $C < \infty$ and $b > 0$, depending only on f, such that for all $\epsilon > 0$ we have:

$$|\{x \in \mathbb{T}^d : |f(x)| < \epsilon\}| < C\epsilon^b. \tag{3.20}$$

In fact, we need a uniform statement, in the sense that as we perturb f, the constants C and b do not change significantly. This was shown to hold for $d = 1$ in [29] and it follows easily from the more general and quantitative approach to Łojasiewicz inequalities in [33] for $d = 1$ and [34] for $d > 1$.

Applying (3.20) to f_A, we ensure that $\det[A(x)]$ has a good lower bound for sufficiently many phases x, hence $\|A^{-1}(x)\|$ has a good upper bound for sufficiently many phases as well. In the end, we obtain an almost invariance principle with a weaker bound than in (3.4) and valid only outside an exponentially small set of phases, but that is enough for our needs and it leads to the following theorem.

Theorem 3.1. *Given* $A \in C_r^\omega(\mathbb{T}^d, \mathrm{Mat}(m, \mathbb{R}))$ *with* $\det[A(x)] \not\equiv 0$ *and* $\omega \in \mathbb{T}^d$ *Diophantine, there are constants* $\delta = \delta(A) > 0$, $n_0 = n_0(A, \omega) \in \mathbb{N}$, $c = c(A) > 0$, $a = a(\omega) > 0$ *and* $b = b(\omega) > 0$ *such that if* $\|B - A\|_r \le \delta$ *and* $n \ge n_0$ *then*

$$\left|\left\{x \in \mathbb{T}^d : \left|\frac{1}{n}\log\|B^{(n)}(x)\| - L_1^{(n)}(B)\right| > n^{-a}\right\}\right| < e^{-cn^b}. \tag{3.21}$$

In fact, this method leads to more than just a fiber LDT for the top Lyapunov exponent. The norm of a matrix $g \in \mathrm{Mat}(m, \mathbb{R})$ is its top singular value $s_1(g)$. It can be shown that (3.21) holds with the norm replaced by *any* singular value of $B^{(n)}(x)$, thus leading to LDT estimates corresponding to each individual Lyapunov exponent.

4. Deriving LDT for Random Cocycles

The goal of this section is to provide LDT estimates for random cocycles over strongly mixing Markov shifts.

4.1. *Literature review*

We mention briefly some of the origins of this subject.

One is the aforementioned Furstenberg's work, started with the proof by H. Furstenberg and H. Kesten of a law of large numbers for random i.i.d. products of matrices [20], and later abstracted by Furstenberg to a seminal theory on random products in semisimple Lie groups [22]. In this context, a first central limit theorem was proved by V. N. Tutubalin in [49]. Since its origin, the scope of Furstenberg's theory has been greatly extended by many contributions (see for instance [44, 26]).

Another source is a central limit theorem of S. V. Nagaev for stationary Markov chains (see [42]). In his approach Nagaev uses the spectral properties of a quasi-compact Markov operator acting on some space of bounded measurable functions. This method was used by E. Le Page to obtain more general central limit theorems, as well as a large deviation principle, for random i.i.d. products of matrices [36]. Later P. Bougerol extended Le Page's approach, proving similar results for Markov type random products of matrices (see [4]).

The book of P. Bougerol and J. Lacroix [5], on random i.i.d. products of matrices, is an excellent introduction to this subject. More recentely, the book of H. Hennion and L. Hervé [28] describes a powerful abstract setting where the method of Nagaev can be applied to derive limit theorems. It contains several applications, including dynamical systems and linear cocycles, that illustrate the method. In Sec. 4.4 we specialize the setting in [28] to prove an abstract LDT theorem which is still enough for our purposes.

4.2. *The model*

Before stating the base and fiber LDT theorems we need to describe the random cocycle models to which they apply.

Let Σ be a compact metric space and $\mathcal{F}$ its Borel σ-field.

Definition 4.1. A *Markov kernel* is a function $K : \Sigma \times \mathcal{F} \to [0, 1]$ such that

(1) for every $x \in \Sigma$, $A \mapsto K(x, A)$ is a probability measure in Σ, also denoted by K_x,
(2) for every $A \in \mathcal{F}$, the function $x \mapsto K(x, A)$ is $\mathcal{F}$-measurable.

The iterated Markov kernels are defined recursively, setting

(a) $K^1 = K$,
(b) $K^{n+1}(x, A) = \int_\Sigma K^n(y, A)K(x, dy)$, for all $n \geq 1$.

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

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

$$\mu(A) = \int K(x, A)\mu(dx).$$

A set $A \in \mathcal{F}$ is said to be *K-invariant* when $K(x, A) = 1$ for all $x \in A$ and $K(x, A) = 0$ for all $x \in X\backslash A$. A K-stationary measure μ is called *ergodic* when there is no K-invariant set $A \in \mathcal{F}$ such that $0 < \mu(A) < 1$. As usual, ergodic measures are the extremal points in the convex set of K-stationary measures.

Definition 4.2. A *Markov system* is a pair (K, μ), where K is a Markov kernel on $(\Sigma, \mathfrak{F})$ and μ is a K-stationary probability measure.

Given some initial probability measure μ on Σ there is a canonical construction, due to Kolmogorov, of a probability space $(X, \mathcal{F}, \mathbb{P}_\mu)$ and a Markov stochastic process $\{e_n : X \to \Sigma\}_{n\geq 0}$ with initial distribution μ and transition kernel K, i.e. for all $x \in \Sigma$ and $A \in \mathfrak{F}$,

(1) $\mathbb{P}_\mu[e_0 \in A] = \mu(A)$,
(2) $\mathbb{P}_\mu[e_n \in A \,|\, e_{n-1} = x] = K(x, A)$.

We briefly outline this construction. Elements in Σ are called *states*. Set $X^+ = \Sigma^{\mathbb{N}}$ as the space of *state sequences* $x = (x_n)_{n\in\mathbb{N}}$, with $x_n \in \Sigma$ for all $n \in \mathbb{N}$, and let $\mathcal{F}$ be the product σ-field $\mathcal{F} = \mathfrak{F}^{\mathbb{N}}$ generated by the $\mathfrak{F}$-cylinders, i.e. by the sets of the form

$$C(A_0, \ldots, A_m) := \{x \in X^+ : x_j \in A_j, \text{ for } 0 \leq j \leq m\},$$

where $A_0, \ldots, A_m \in \mathfrak{F}$ are measurable sets. The (topological) product space X^+ can be made a metric space. The σ-field $\mathcal{F}$ coincides with the Borel σ-field of the compact space X^+. The following expression determines a pre-measure

$$\mathbb{P}_\mu[C(A_0, \ldots, A_m)] := \int_{A_m} \cdots \int_{A_0} \mu(dx_0) \prod_{j=1}^{m} K(x_{j-1}, dx_j)$$

over the semi-algebra of $\mathfrak{F}$-cylinders. By Carathéodory's extension theorem this pre-measure extends to a unique probability measure $\mathbb{P}_\mu$ on $\mathcal{F}$. It follows from this definition that (1) and (2) hold for the sequence of random variables $e_n : X^+ \to \Sigma$, defined by $e_n(x) := x_n$ for $x = (x_n)_{n\in\mathbb{N}}$. It also follows that the process $\{e_n\}_{n\geq 0}$ is stationary w.r.t. $(X, \mathcal{F}, \mathbb{P}_\mu)$ if and only if μ is a K-stationary measure.

Markov systems are probabilistic evolutionary models, but they can also be studied in dynamical terms. For that we introduce the *shift mappings*. The *one-sided shift* is the map $T : X^+ \to X^+$ defined by $T(x_n)_{n\geq 0} = (x_{n+1})_{n\geq 0}$. The map T is continuous, and hence $\mathcal{F}$-measurable. The measure $\mathbb{P}_\mu$ is preserved by T, i.e. $T_*\mathbb{P}_\mu = \mathbb{P}_\mu$, if and only if μ is K-stationary. A similar characterization holds on ergodicity. If μ is a K-stationary measure then $\mathbb{P}_\mu$ is ergodic w.r.t. T if and only if μ is ergodic. The process e_n is dynamically generated by the observable e_0 in the sense that $e_n = e_0 \circ T^n$, for all $n \geq 0$.

The one-sided shift $T : X^+ \to X^+$ is not invertible but it admits the *two-sided shift* $T : X \to X$ defined on $X = \Sigma^{\mathbb{Z}}$ by $T(x_n)_{n\in\mathbb{Z}} = (x_{n+1})_{n\in\mathbb{Z}}$, as its *natural extension*. This means $T : X \to X$ is a homeomorphism which makes the following diagram commutative, and factors any other homeomorphism with the same property

$$\begin{array}{ccc} X & \xrightarrow{\;T\;} & X \\ {\scriptstyle\pi}\big\downarrow & & \big\downarrow{\scriptstyle\pi} \\ X^+ & \xrightarrow{\;T\;} & X^+ \end{array} \qquad (4.1)$$

The vertical arrows stand for the natural projection $\pi : X \to X^+$, $\pi(x_n)_{n\in\mathbb{Z}} = (x_n)_{n\in\mathbb{N}}$. The measure $\mathbb{P}_\mu$ on X^+ can be extended to a probability measure, still denoted $\mathbb{P}_\mu$, on X, which is preserved by both T and π, i.e. $T_*\mathbb{P}_\mu = \mathbb{P}_\mu$ and $\pi_*\mathbb{P}_\mu = \mathbb{P}_\mu$. We shall refer to these two measures, respectively on the spaces X and X^+, as the *Kolmogorov extensions* of the Markov system (K, μ). For the sake of notational simplicity we use the same letter $\mathcal{F}$ to denote the σ-fields generated by $\mathfrak{F}$-cylinders on both spaces X^+ and X.

Definition 4.3. Given a Markov system (K, μ) the measure preserving dynamical system $(T, X, \mathcal{F}, \mathbb{P}_\mu)$ is called a Markov shift.

Let $(L^\infty(\Sigma), \|\cdot\|_\infty)$ denote the Banach algebra of complex bounded $\mathfrak{F}$-measurable functions with the sup-norm $\|f\|_\infty = \sup_{x\in\Sigma} |f(x)|$.

Definition 4.4. (Condition (A1) in [4]) We say that (K, μ) is strongly mixing if there are constants $C > 0$ and $0 < \rho < 1$ such that for every $f \in L^\infty(\Sigma)$, all $x \in \Sigma$ and $n \in \mathbb{N}$,

$$\left| \int_\Sigma f(y) K^n(x, dy) - \int_\Sigma f(y) \mu(dy) \right| \leq C\rho^n \|f\|_\infty.$$

Remark 4.1. If (K, μ) is strongly mixing then μ is the unique K-stationary measure and the Markov shift $(T, X, \mathcal{F}, \mathbb{P}_\mu)$ is mixing.

Examples of strongly mixing systems arise naturally from Markov kernels satisfying the *Doeblin condition* (see [11]). We say that K satisfies the Doeblin condition if there is a positive finite measure ρ on $(\Sigma, \mathfrak{F})$ and some $\varepsilon > 0$ such that for all $x \in \Sigma$ and $A \in \mathfrak{F}$,

$$K(x, A) \geq 1 - \varepsilon \Rightarrow \rho(A) \geq \varepsilon.$$

Given $A \in \mathfrak{F}$, define

$$L^\infty(A) := \{f \in L^\infty(\Sigma) : f|_{\Sigma\setminus A} \equiv 0\},$$

which is a closed Banach sub-algebra of $(L^\infty(\Sigma), \|\cdot\|_\infty)$.

Proposition 4.1. *If a Markov kernel K satisfies the Doeblin condition then there are sets $\Sigma_1, \ldots, \Sigma_m$ in $\mathfrak{F}$ and probability measures $\nu_1, \ldots, \nu_m$ on Σ such that for all $i, j = 1, \ldots, m$,*

1. $\Sigma_i \cap \Sigma_j = \emptyset$ *when* $i \neq j$,
2. Σ_i *is* K*-forward invariant, i.e.* $K(x, \Sigma_i) = 1$ *for* $x \in \Sigma_i$,
3. ν_i *is* K*-stationary and ergodic with* $\nu_i(\Sigma_j) = \delta_{ij}$,
4. $lim_{n\to+\infty} K^n(x, \Sigma_1 \cup \cdots \cup \Sigma_m) = 1$, *with geometric uniform speed of convergence, for all* $x \in \Sigma$,
5. $\nu(\Sigma_1 \cup \cdots \cup \Sigma_m) = 1$, *for every* K*-stationary probability* ν.

Moreover, for every $1 \leq i \leq m$ *there is an integer* $p_i \in \mathbb{N}$ *and measurable sets* $\Sigma_{i,1}, \ldots, \Sigma_{i,p_i} \in \mathcal{F}$ *such that*

(1) $\{\Sigma_{i,1}, \ldots, \Sigma_{i,p_i}\}$ *is a partition of* Σ_i,
(2) $K(x, \Sigma_{i,j+1}) = 1$ *for* $x \in \Sigma_{i,j}$ *and* $1 \leq j \leq p_i$, *with* $\Sigma_{i,p_i+1} = \Sigma_{i,1}$,
(3) $(\Sigma_{i,j}, K^{p_i})$ *is strongly mixing for all* $1 \leq j \leq p_i$.

Proof. See Sec. V-5 in [11]. □

We now state the two main theorems on LDT estimates.

Let us begin with the base LDT theorem. Recall that $X = \Sigma^{\mathbb{Z}}$. Consider the metric $\widetilde{d} : X \times X \to [0, 1]$

$$\widetilde{d}(x, x') := 2^{-\inf\{|k| : k \in \mathbb{Z}, x_k \neq x'_k\}},$$

for all $x = (x_k)_{k \in \mathbb{Z}}$ and $x' = (x'_k)_{k \in \mathbb{Z}}$ in X. Note that X is not compact for the topology induced by $\widetilde{d}$, unless Σ is finite. Given $k \in \mathbb{N}$, $\alpha > 0$ and $f \in L^\infty(X)$ define

$$v_k(f) := \sup\{|f(x) - f(y)| : \widetilde{d}(x, y) \leq 2^{-k}\},$$

$$v_\alpha(f) := \sup\{2^{\alpha k} v_k(f) : k \in \mathbb{N}\},$$

$$\|f\|_\alpha := \|f\|_\infty + v_\alpha(f),$$

$$\mathcal{H}_\alpha(X) := \{f \in L^\infty(X) : v_\alpha(f) < +\infty\}.$$

The last set, $\mathcal{H}_\alpha(X)$, is the space of Hölder continuous functions with exponent α w.r.t. the distance $\widetilde{d}$ on X. In fact it follows easily from the definition that

$$v_\alpha(f) = \sup_{x \neq x'} \frac{|f(x) - f(x')|}{\widetilde{d}(x, x')^\alpha}.$$

Proposition 4.2. *For all* $0 \leq \alpha \leq 1$, $(\mathcal{H}_\alpha(X), \|\cdot\|_\alpha)$ *is a unital Banach algebra, and also a lattice.*

Proof. See Proposition 5.4 in [17] or Proposition 1.4 in [16]. □

We say that a function $f : X \to \mathbb{C}$ is *future independent* if $f(x) = f(y)$ for any $x, y \in X$ such that $x_k = y_k$ for all $k \leq 0$. Define

$$\mathcal{H}_\alpha(X^-) := \{f \in \mathcal{H}_\alpha(X) : f \text{ is future independent}\}. \tag{4.2}$$

Clearly $\mathcal{H}_\alpha(X^-)$ is a closed sub-algebra of $\mathcal{H}_\alpha(X)$, and hence a unital Banach algebra itself.

We can now state the base LDT theorem.

Theorem 4.1. *Let* (K, μ) *be a strongly mixing Markov system. Then for any* $0 < \alpha \leq 1$ *and any observable* $\xi \in \mathcal{H}_\alpha(X^-)$ *there are constants* $C = C(\xi) > 0$,

$k = k(\xi) > 0$ and $\varepsilon_0 = \varepsilon_0(\xi) > 0$ such that for all $0 < \varepsilon < \varepsilon_0$, $x \in \Sigma$ and $n \in \mathbb{N}$,

$$\mathbb{P}_\mu \left[\left| \frac{1}{n} \sum_{j=0}^{n-1} \xi \circ T^j - \mathbb{E}_\mu(\xi) \right| \geq \varepsilon \right] \leq C e^{-k\varepsilon^2 n}.$$

Moreover, the constants C, k and ε_0 depend only on K and $\|\xi\|_\alpha$, and hence can be kept constant when K is fixed and ξ ranges over any bounded set in $\mathcal{H}_\alpha(X^-)$.

For the fiber LDT estimates we introduce spaces of measurable random cocycles over a strongly mixing Markov system (K, μ).

Definition 4.5. We define $\mathcal{B}^\infty_m = \mathcal{B}^\infty_m(K)$ to be the space of bounded measurable functions $A : \Sigma \times \Sigma \to \mathrm{GL}(m, \mathbb{R})$ with bounded inverse A^{-1}. This is a metric space with the distance

$$d_\infty(A, B) := \|A - B\|_\infty.$$

Each $A \in \mathcal{B}^\infty_m$ determines the linear cocycle $F_A : X \times \mathbb{R}^m \to X \times \mathbb{R}^m$,

$$F_A(x, v) := (Tx, A(x)v),$$

where we identify A with the function $A : X \to \mathrm{GL}(m, \mathbb{R})$, $A(x) := A(x_0, x_1)$, for all $x = (x_n)_{n\in\mathbb{Z}} \in X$. The iterates of F_A are the maps $F^n_A : X \times \mathbb{R}^m \to X \times \mathbb{R}^m$

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

with $A^{(n)} : X \to \mathrm{GL}(m, \mathbb{R})$ defined by

$$A^{(n)}(x) := A(x_{n-1}, x_n) \cdots A(x_1, x_2)A(x_0, x_1).$$

Let $\mathrm{Gr}(\mathbb{R}^m)$ denote the Grassmann manifold of the Euclidean space $\mathbb{R}^m$. An $\mathcal{F}$-measurable function $V : \Sigma \to \mathrm{Gr}(\mathbb{R}^m)$ is said to be an *A-invariant* family of linear subspaces when

$$A(x_{n-1}, x_n)V(x_{n-1}) = V(x_n) \quad \text{for } \mathbb{P}_\mu \text{ -a.e. } x \in X.$$

The ergodicity of $(T, X, \mathbb{P}_\mu)$, or that of (K, μ), implies that the subspaces $V(x)$ have constant dimension $\mathbb{P}_\mu$-a.e., which we denoted by $\dim(V)$. We say that this family is *proper* if $0 < \dim(V) < m$. Next we introduce the concept of irreducible cocycle (see Definition 2.7 in [4]).

Definition 4.6. A cocycle $A \in \mathcal{B}^\infty_m(K)$ is said to be irreducible w.r.t. (K, μ) if it admits no measurable proper A-invariant family of linear subspaces.

We prove the following uniform fiber LDT theorem.

Theorem 4.2. *Given a Markov system (K, μ) and $A \in \mathcal{B}^\infty_m(K)$ if*

(1) *(K, μ) is strongly mixing,*
(2) *A is irreducible,*
(3) *$L_1(A) > L_2(A)$,*

then there exists a neighborhood $\mathcal{V}$ *of* A *in* $\mathcal{B}_m^\infty(K)$ *and there are constants* $C > 0$, $k > 0$ *and* $\varepsilon_0 > 0$ *such that for all* $0 < \varepsilon < \varepsilon_0$, $B \in \mathcal{V}$ *and* $n \in \mathbb{N}$,

$$\mathbb{P}_\mu \left[\left| \frac{1}{n} \log\|B^{(n)}\| - L_1(B) \right| \geq \varepsilon \right] \leq C e^{-k\varepsilon^2 n}.$$

4.3. *Spectral method*

The proofs of Theorems 4.1 and 4.2 follow from the spectral method in [28], which was originally due to S. V. Nagaev. We sketch the strategy, introducing some needed concepts.

Let $\mathcal{B}$ be a Banach space, and $\mathcal{L}(\mathcal{B})$ denote the Banach algebra of bounded linear operators $T : \mathcal{B} \to \mathcal{B}$. Given $T \in \mathcal{L}(\mathcal{B})$, we denote its spectrum by $\sigma(T)$, and its spectral radius by

$$\rho(T) = \lim_{n\to+\infty} \|T^n\|^{1/n} = \inf_{n\geq 0} \|T^n\|^{1/n}.$$

Definition 4.7. The operator T is called quasi-compact if there is a T-invariant decomposition $\mathcal{B} = F \oplus \mathcal{H}$ such that $\dim F < +\infty$ and the spectral radius of $T|_{\mathcal{H}}$ is (strictly) less than the absolute value $|\lambda|$ of any eigenvalue λ of $T|_F$. T is called quasi-compact and simple when furthermore $\dim F = 1$. In this case $\sigma(T|_F)$ consists of a single simple eigenvalue referred to as the maximal eigenvalue of T.

Definition 4.8. We call observed Markov system any triple (K, μ, ξ), where (K, μ) is a Markov system and $\xi : \Sigma \to \mathbb{R}$ is a measurable observable.

Define the sum process $S_n(\xi) := \sum_{j=0}^{n-1} \xi \circ e_j$ on $(X^+, \mathcal{F})$.

Let $\mathbb{P}_x$ denote the probability measure $\mathbb{P}_x := \mathbb{P}_{\delta_x}$ on X^+, where δ_x is the Dirac measure at a point $x \in \Sigma$. With this notation we have $\mathbb{P}_\mu = \int_\Sigma \mathbb{P}_x \mu(dx)$ for any probability measure μ on Σ. The expectations $\mathbb{E}_x$ and $\mathbb{E}_\mu$ refer to the probability measures $\mathbb{P}_x$ and $\mathbb{P}_\mu$, respectively.

Consider an observed Markov system (K, μ, ξ).

The linear operator

$$(Qf)(x) = (Q_K f)(x) := \int_X f(y) K(x, dy),$$

is called a *Markov operator*. It operates on $\mathcal{F}$-measurable functions on Σ, mapping L^p functions to L^p functions, for any $1 \leq p \leq \infty$. We shall write Q instead of Q_K when the kernel K is fixed.

The linear operator

$$(Q_\xi f)(x) = (Q_{K,\xi} f)(x) := \int_X f(y) e^{\xi(y)} K(x, dy),$$

is called a *Laplace–Markov operator*. It also operates on $\mathcal{F}$-measurable functions on Σ, but the domain of Q_ξ depends on the observable ξ.

Because (K, μ) is strongly mixing (see Definition 4.4), the Markov operator Q_K is quasi-compact and simple on $L^\infty(\Sigma)$. In fact these two statements are easily seen to be equivalent. Let $\mathcal{B} \subset L^\infty(\Sigma)$ be a Banach space where the Markov operator $Q = Q_K : \mathcal{B} \to \mathcal{B}$ is still quasi-compact, and the Laplace–Markov operator $Q_\xi : \mathcal{B} \to \mathcal{B}$ is bounded. By spectral continuity, if t is small then $Q_{t\xi} : \mathcal{B} \to \mathcal{B}$ is also quasi-compact and simple. Let $\mathbf{1}$ denote the constant function, $\mathbf{1}(x) = 1$. The operator Q has eigenvalue 1 associated to the eigenfunction $\mathbf{1}$. Thus, for t small the operator $Q_{t\xi}$ has a simple eigenvalue $\lambda(t)$ associated to some eigenfunction $v(t) \in \mathcal{B}$. Under general assumptions, the functions $\lambda(t)$ and $v(t)$ are analytic in t, with $\lambda(0) = 1$ and $v(0) = \mathbf{1}$. We can normalize $v(t)$ so that $\mathbb{E}_\mu[v(t)] = 1$ for all t.

Assume now that $\mathbb{E}_\mu[\xi] = 0$, or otherwise take $\overline{\xi} = \xi - \mathbb{E}_\mu[\xi]$ instead of ξ. A simple calculation shows that $\mathbb{E}_\mu[e^{tS_n(\xi)}] = \mathbb{E}_\mu[Q_{t\xi}^n \mathbf{1}]$ (see Lemma 4.1). Thus $\mathbb{E}_\mu[e^{tS_n(\xi)}] = \mathbb{E}_\mu[Q_{t\xi}^n \mathbf{1}] \approx \mathbb{E}_\mu[Q_{t\xi}^n v(t)] = \lambda(t)^n$ (see Proposition 4.7). Another simple computation shows that $c(t) := \log \lambda(t)$ is a non-negative convex function such that $c(0) = 0$ and $c'(0) = \mathbb{E}_\mu[\xi] = 0$. Therefore, choosing $h > c''(0)$, by Chebyshev's inequality (see (2.1)) we have

$$\begin{aligned} \mathbb{P}_\mu[S_n(\xi) > n\varepsilon] &\leq e^{-tn\varepsilon}\mathbb{E}_\mu[e^{tS_n(\xi)}] \approx e^{-n(\varepsilon t - c(t))} \\ &\leq e^{-n(\varepsilon t - \frac{ht^2}{2})}, \end{aligned}$$

where this inequality holds for every $t \approx 0$. Finally, optimizing t (see the proof of Theorem 4.3) we get the LDT estimate

$$\mathbb{P}_\mu[S_n(\xi) > n\varepsilon] \lesssim e^{-n\frac{\varepsilon^2}{2h}}.$$

The irreducibility assumption is essential to prove Theorem 4.2. The proof exploits the fact that for irreducible cocycles there is a Banach algebra of measurable functions, independent of the cocycle, where the associated Laplace–Markov operators act as quasi-compact and simple operators (see Sec. 4.6). For reducible cocycles this fact could be true, and lead to fiber LDT estimates, but only with a Banach algebra tailored to the cocycle. Hence the same scheme of proof would not provide the required uniformity.

4.4. *Abstract setting*

We discuss now a setting, consisting of the assumptions (B1)–(B7) and (A1)–(A4) below, where an abstract LDT theorem is proved, and from which Theorems 4.1 and 4.2 will be deduced. The context here specializes a more general setting in [28].

Let Σ be a compact metric space, and $\mathfrak{X}$ be a set of observed Markov systems (K, μ, ξ) over Σ, endowed with some distance d.

Besides $\mathfrak{X}$, this setting consists of a scale of complex Banach algebras $(\mathcal{B}_\alpha, \|\cdot\|_\alpha)$ indexed in $\alpha \in [0,1]$, where each $\mathcal{B}_\alpha$ is a space of bounded measurable functions on Σ. We assume that there exist seminorms $v_\alpha : \mathcal{B}_\alpha \to [0, +\infty)$ such that for all $0 \leq \alpha \leq 1$,

(B1) $\|f\|_\alpha = v_\alpha(f) + \|f\|_\infty$, for all $f \in \mathcal{B}_\alpha$,
(B2) $\mathcal{B}_0 = L^\infty(\Sigma)$, and $\|\cdot\|_0$ is equivalent to $\|\cdot\|_\infty$,

(B3) $\mathcal{B}_\alpha$ is a lattice, i.e. if $f \in \mathcal{B}_\alpha$ then $\overline{f}, |f| \in \mathcal{B}_\alpha$,
(B4) $\mathcal{B}_\alpha$ is a Banach algebra with unity $\mathbf{1} \in \mathcal{B}_\alpha$ and $v_\alpha(\mathbf{1}) = 0$.

Assume also that this family is a *scale of normed spaces* in the sense that for all $0 \le \alpha_0 < \alpha_1 < \alpha_2 \le 1$,

(B5) $\mathcal{B}_{\alpha_2} \subset \mathcal{B}_{\alpha_1} \subset \mathcal{B}_{\alpha_0}$,
(B6) $v_{\alpha_0}(f) \le v_{\alpha_1}(f) \le v_{\alpha_2}(f)$, for all $f \in \mathcal{B}_{\alpha_2}$,
(B7) $v_{\alpha_1}(f) \le v_{\alpha_0}(f)^{\frac{\alpha_2-\alpha_1}{\alpha_2-\alpha_0}} v_{\alpha_2}(f)^{\frac{\alpha_1-\alpha_0}{\alpha_2-\alpha_0}}$, for all $f \in \mathcal{B}_{\alpha_2}$.

Finally assume there exists an interval $[\alpha_1, \alpha_0] \subset (0,1]$ with $\alpha_1 < \frac{\alpha_0}{2}$ such that for all $\alpha \in [\alpha_1, \alpha_0]$ the space $\mathfrak{X}$ satisfies:

(A1) $(K, \mu, -\xi) \in \mathfrak{X}$ whenever $(K, \mu, \xi) \in \mathfrak{X}$.
(A2) The Markov operators $Q_K : \mathcal{B}_\alpha \to \mathcal{B}_\alpha$ are uniformly quasi-compact and simple, i.e. there exist constants $C > 0$ and $0 < \sigma < 1$ such that for all $(K, \mu, \xi) \in \mathfrak{X}$ and $f \in \mathcal{B}_\alpha$,

$$\|Q_K^n f - \langle f, \mu \rangle \mathbf{1}\|_\alpha \le C \sigma^n \|f\|_\alpha.$$

(A3) The operators $Q_{K,z\xi}$ act continuously on the Banach algebras $\mathcal{B}_\alpha$, uniformly in $(K, \mu, \xi) \in \mathfrak{X}$ and z small. More precisely, we assume there are constants $b > 0$ and $M > 0$ such that for all $i = 0, 1, 2$, $|z| < b$ and $f \in \mathcal{B}_\alpha$,

$$Q_{K,z\xi}(f\xi^i) \in \mathcal{B}_\alpha \quad \text{and} \quad \|Q_{K,z\xi}(f\xi^i)\|_\alpha \le M\|f\|_\alpha.$$

(A4) The family of functions $\mathfrak{X} \ni (K, \mu, \xi) \mapsto Q_{K,z\xi}$, indexed in the disk $|z| \le b$, is Hölder equi-continuous in the sense that there exists $0 < \theta \le 1$ such that for all $|z| \le b$, $f \in \mathcal{B}_\alpha$ and $(K_1, \mu_1, \xi_1), (K_2, \mu_2, \xi_2) \in \mathfrak{X}$,

$$\|Q_{K_1,z\xi_1} f - Q_{K_2,z\xi_2} f\|_\infty \le M\|f\|_\alpha d((K_1, \mu_1, \xi_1), (K_2, \mu_2, \xi_2))^\theta.$$

The interval $[\alpha_1, \alpha_0]$ will be referred to as the *range* of the scale of Banach algebras. The reason to work with this range instead of $[0, 1]$ is twofold: in the fiber LDT theorem we will need to take α_0 small enough to have contraction in (A2), but at the same time α_1 bounded away from 0 to have uniformity in this contraction.

The necessity of the condition $\alpha_1 < \frac{\alpha_0}{2}$ is explained in Remark 4.2.

The positive constants C, σ, M, b and θ above will be referred to as the *setting constants*.

An example of a scale of Banach algebras satisfying (B1)–(B7) is formed by the spaces of α-Hölder continuous functions on Σ w.r.t. some normalized distance $d : \Sigma \times \Sigma \to [0, 1]$. The norms on these spaces are defined as follows: for all $\alpha \in (0, 1]$ and $f \in L^\infty(\Sigma)$, let

$$\|f\|_\alpha := v_\alpha(f) + \|f\|_\infty, \quad \text{with } v_\alpha(f) := \sup_{\substack{x,y\in\Sigma \\ x\neq y}} \frac{|f(x) - f(y)|}{d(x, y)^\alpha}.$$

Proposition 4.3. *The family of normed spaces*

$$\mathcal{H}_\alpha(\Sigma) := \{f \in L^\infty(\Sigma) : v_\alpha(f) < +\infty\}, \quad \alpha \in [0,1]$$

satisfies (B1)–(B7).

Proof. See for instance [35]. □

Examples of contexts satisfying all assumptions (B1)–(B7) and (A1)–(A4) are provided by the applications in Secs. 4.5 and 4.6.

Assumption (A1) allows us to reduce deviations below average to deviations above average, thus shortening proofs.

(A2) is the main assumption: for $(K,\mu,\xi) \in \mathfrak{X}$ all Markov operators $Q_K : \mathcal{B}_\alpha \to \mathcal{B}_\alpha$ are quasi-compact and simple, uniformly in (K,μ,ξ). This will imply that, possibly decreasing b, all Laplace–Markov operators $Q_{K,z\xi} : \mathcal{B}_\alpha \to \mathcal{B}_\alpha$ are also quasi-compact and simple, uniformly in (K,μ,ξ) and $|z| < b$.

(A3) is a regularity assumption. The operators $Q_{K,z\xi}$ act continuously on $\mathcal{B}_\alpha$, uniformly in (K,μ,ξ) and $|z| < b$. Moreover, it implies that $\mathbb{D}_b \ni z \mapsto Q_{K,z\xi} \in \mathcal{L}(\mathcal{B}_\alpha)$, is an analytic function.

Finally (A4) implies that the function $(K,\mu,\xi) \mapsto \lambda_{K,\xi}(z)$ is uniformly Hölder continuous. Here $\lambda_{K,\xi}(z)$ denotes the maximal eigenvalue of $Q_{K,z\xi}$.

These facts follow from the propositions stated below.

Assume the scale $(\mathcal{B}_\alpha, \|\cdot\|_\alpha)$ is fixed satisfying (B1)–(B7). Given a Markov system (K,μ), consider the space

$$\mathfrak{X}^\alpha_{K,L} := \{(K,\mu,\xi) : \xi \in \mathcal{B}_\alpha, \|\xi\|_\alpha \le L\}$$

of observed Markov systems over (K,μ). We will identify $\mathfrak{X}^\alpha_{K,L}$ as a subspace of $\mathcal{B}_\alpha$ and endow it with the corresponding norm distance.

Proposition 4.4. *Given a Markov system* (K,μ) *and* $L > 0$, *if* $\mathfrak{X}^\alpha_{K,L}$ *satisfies* (A2) *with setting constants* (C,σ) *then for any* $b > 0$ *there exists* $M > 0$ *such that* $\mathfrak{X}^\alpha_{K,L}$ *satisfies* (A1)–(A4) *with setting constants* C, σ, M, b *and* $\theta = 1$. *Moreover, the map* $\mathfrak{X}^\alpha_{K,L} \to \mathcal{L}(\mathcal{B}_\alpha)$, $(K,\mu,\xi) \mapsto Q_{K,\xi}$, *is analytic.*

Proof. Since $\mathcal{B}_\alpha$ is a Banach algebra, given $\xi \in \mathcal{B}_\alpha$, the multiplication operator $D_\xi : \mathcal{B}_\alpha \to \mathcal{B}_\alpha$, $D_\xi f := \xi f$, is uniformly bounded for $\xi \in \mathfrak{X}^\alpha_{K,L}$. Thus, because $Q_{K,\xi} = Q_K \circ D_{e^\xi}$, the map $Q_{K,*} : \mathcal{B}_\alpha \to \mathcal{L}(\mathcal{B}_\alpha)$, $\xi \mapsto Q_{K,\xi}$, is analytic. Condition (A1) holds trivially, and (A3) follows from the previous considerations. (A3) implies there exists $M > 0$ such that for all $\alpha_1 \le \alpha \le \alpha_0$ and all $f \in \mathcal{B}_\alpha$,

$$\|Q_K f\|_\alpha \le M\|f\|_\alpha.$$

A simple computation, using that $(\mathcal{B}_\alpha, \|\cdot\|_\alpha)$ is a Banach algebra, shows that for all $\alpha_1 \le \alpha \le \alpha_0$, $f \in \mathcal{B}_\alpha$ and $\xi_1, \xi_2 \in \mathfrak{X}^\alpha_{K,L}$,

$$\|Q_{K,\xi_1} f - Q_{K,\xi_2} f\|_\alpha \le M e^L \|\xi_1 - \xi_2\|_\alpha \|f\|_\alpha.$$

This implies (A4). □

It follows from (A3) that $Q_{K,z\xi} \in \mathcal{L}(\mathcal{B}_\alpha)$, for all $z \in \mathbb{D}_b$. In particular, the function $Q_{K,*\xi} : \mathbb{D}_b \to \mathcal{L}(\mathcal{B}_\alpha)$, $z \mapsto Q_{K,z\xi}$, is well-defined, for every $(K, \mu, \xi) \in \mathfrak{X}$.

Proposition 4.5. *The function $Q_{K,*\xi} : \mathbb{D}_b \to \mathcal{L}(\mathcal{B}_\alpha)$ is analytic with*

$$\frac{d}{dz} Q_{K,z\xi}(f) = Q_{K,z\xi}(f\xi) \quad \textit{for } f \in \mathcal{B}_\alpha,$$

for all $(K, \mu, \xi) \in \mathfrak{X}$, and $\alpha_1 \leq \alpha \leq \alpha_0$.

Proof. See Proposition 5.10 in [17] or Proposition 2.3 in [16]. □

Next proposition focuses on the quasi-compactness and simplicity of $Q_z = Q_{K,z\xi}$, and can be proved using arguments in [36, 4]. See also Proposition 5.11 in [17] or Proposition 2.4 in [16].

Proposition 4.6. *Consider a metric space $\mathfrak{X}$ of observed Markov systems satisfying* (A1)–(A4) *in the range $[\alpha_1, \alpha_0] \subset (0, 1]$ with setting constants C, σ, M, b and θ.*

Given $\varepsilon > 0$ there exist $C', M' > 0$ and $0 < b_0 < b$ such that the following statement holds: for all $(K, \mu, \xi) \in \mathfrak{X}$, $z \in \mathbb{D}_{b_0}$ and $\alpha_1 \leq \alpha \leq \alpha_0$ there exist: a one-dimensional subspace $E_z = E_{K,z\xi} \subset \mathcal{B}_\alpha$, a hyperplane $H_z = H_{K,z\xi} \subset \mathcal{B}_\alpha$, a number $\lambda(z) = \lambda_{K,\xi}(z) \in \mathbb{C}$, and a linear map $P_z = P_{K,z\xi} \in \mathcal{L}(\mathcal{B}_\alpha)$ such that

(1) *$\mathcal{B}_\alpha = E_z \oplus H_z$ is a Q_z-invariant decomposition,*
(2) *P_z is a projection onto E_z, parallel to H_z,*
(3) *$Q_z \circ P_z = P_z \circ Q_z = \lambda(z)\, P_z$,*
(4) *$Q_z f = \lambda(z) f$ for all $f \in E_z$,*
(5) *$z \mapsto \lambda(z)$ is analytic in a neighborhood of $\overline{\mathbb{D}}_{b_0}$,*
(6) *$|\lambda(z)| \geq 1 - \varepsilon$.*

Furthermore, for all $f \in \mathcal{B}_\alpha$,

(7) *$\|Q_z^n f - \lambda(z)^n P_z f\|_\alpha \leq C'(\sigma + \varepsilon)^n \|f\|_\alpha$,*
(8) *$\|P_z f\|_\alpha \leq C' \|f\|_\alpha$,*
(9) *$\|P_z f - P_0 f\|_\alpha \leq C' |z| \|f\|_\alpha$*

and for all $z \in \mathbb{D}_{b_0}$ and $(K_1, \mu_1, \xi_1), (K_2, \mu_2, \xi_2) \in \mathfrak{X}$,

(10) *$|\lambda_{K_1,\xi_1}(z) - \lambda_{K_2,\xi_2}(z)| \leq M' d((K_1, \mu_1, \xi_1), (K_2, \mu_2, \xi_2))^{\frac{\theta}{2}}$.*

Remark 4.2. The condition $\alpha_1 < \frac{\alpha_0}{2}$ and the assumption (A4) are only needed to prove (10).

If the hypothesis of Proposition 4.4 are satisfied, with (A2) holding for one α, then, except for (10), all conclusions of Proposition 4.6 hold for that particular α. Since by Proposition 4.4, the mapping $\xi \mapsto Q_{K,\xi}$ is analytic, it follows that

$\xi \mapsto \lambda(\xi) =$ maximal eigenvalue of $Q_{K,\xi}$, is also an analytic function. Hence in this case there exists $M' > 0$ such that for all $z \in \mathbb{D}_{b_0}$, and $\xi_1, \xi_2 \in \mathfrak{X}^\alpha_{K,L}$,

$$|\lambda_{K,\xi_1}(z) - \lambda_{K,\xi_2}(z)| \leq M' \|\xi_1 - \xi_2\|_\alpha.$$

This statement implies (10). Hence

Remark 4.3. If for some $0 < \alpha \leq 1$ the space $\mathfrak{X}^\alpha_{K,L}$ satisfies (A2) with setting constants (C, σ) then all conclusions of Proposition 4.6 hold, with a Lipschitz modulus of continuity in (10).

Let $c_{K,\xi}(t) := \log \lambda_{K,\xi}(t)$, where $\lambda_{K,\xi}(t)$ denotes the maximal eigenvalue of $Q_{K,t\xi}$. Note that $\lambda_{K,\xi}(t) > 0$ because $Q_{K,t\xi}$ is a positive operator for all $t \in \mathbb{R}$.

We can now state the abstract LDT theorem.

Theorem 4.3. *Let* $(\mathcal{B}_\alpha, \|\cdot\|_\alpha)$ *be a scale of Banach algebras satisfying* (B1)–(B7), *and* $\mathfrak{X}$ *be a space of observed Markov systems for which assumptions* (A1)–(A4) *hold.*

Consider $(K_0, \mu_0, \xi_0) \in \mathfrak{X}$ *with* $h > (c_{K_0,\xi_0})''(0)$. *Then there exist a neighborhood* $\mathcal{V}$ *of* $(K_0, \mu_0, \xi_0) \in \mathfrak{X}$, $C > 0$ *and* $\varepsilon_0 > 0$ *such that for all* $(K, \mu, \xi) \in \mathcal{V}$, $0 < \varepsilon < \varepsilon_0$, $x \in \Sigma$ *and* $n \in \mathbb{N}$,

$$\mathbb{P}_x \left[\left| \frac{1}{n} S_n(\xi) - \mathbb{E}_\mu(\xi) \right| \geq \varepsilon \right] \leq C e^{-\frac{\varepsilon^2}{2h} n}. \tag{4.3}$$

Remark 4.4. The proof of this theorem shows that conclusion (4.3) holds for any $(K, \mu, \xi) \in \mathfrak{X}$ such that $h > (c_{K,\xi})''(0)$.

Corollary 4.1. *Assume* $\mathfrak{X}^\alpha_{K,L}$ *satisfies* (A2).

Given $\xi_0 \in \mathfrak{X}^\alpha_{K,L}$ *with* $h > (c_{K,\xi_0})''(0)$, *there exist a neighborhood* $\mathcal{V}$ *of* ξ_0 *in* $\mathfrak{X}^\alpha_{K,L}$, $C > 0$ *and* $\varepsilon_0 > 0$ *such that for all* $\xi \in \mathcal{V}$, $0 < \varepsilon < \varepsilon_0$, $x \in \Sigma$ *and* $n \in \mathbb{N}$, *the inequality* (4.3) *holds.*

Proof. Combine Proposition 4.4 with Theorem 4.3. □

Remark 4.5. Averaging in x w.r.t. μ the probabilities in Theorem 4.3 we get for all $0 < \varepsilon < \varepsilon_0$, $(K, \mu, \xi) \in \mathcal{V}$ and $n \in \mathbb{N}$,

$$\mathbb{P}_\mu \left[\left| \frac{1}{n} S_n(\xi) - \mathbb{E}_\mu(\xi) \right| \geq \varepsilon \right] \leq C e^{-\frac{\varepsilon^2}{2h} n}.$$

In the rest of this section assume that (A1)–(A4) hold for the space $\mathfrak{X}$ of observed Markov, and take $0 < b_0 < b$ according to Proposition 4.6.

Lemma 4.1. *For all* $(K, \mu, \xi) \in \mathfrak{X}$, $n \in \mathbb{N}$, $z \in \mathbb{D}_{b_0}$ *and* $x \in \Sigma$,

$$((Q_{K,z\xi})^n \mathbf{1})(x) = \mathbb{E}_x[e^{z S_n(\xi)}] = \int_\Omega e^{z S_n(\xi)} d\mathbb{P}_x.$$

In particular, for all $z \in \mathbb{D}_{b_0}$,

$$\mathbb{E}_\mu((Q_{K,z\xi})^n \mathbf{1}) = \mathbb{E}_\mu[e^{z S_n(\xi)}].$$

Proof. In fact,

$$((Q_{K,z\xi})^n \mathbf{1})(x_0) = \int_{\Sigma^n} e^{z \sum_{j=1}^n \xi(x_j)} \prod_{j=0}^{n-1} K(x_j, dx_{j+1})$$
$$= \mathbb{E}_{x_0}[e^{zS_n(\xi)}]$$

and averaging this relation in the variable x_0 w.r.t. μ we derive the second identity. □

Proposition 4.7. *There exist $C_1 > 0$ and a sequence δ_n converging geometrically to 0 such that for all $(K, \mu, \xi) \in \mathfrak{X}$, $t \in (-b_0, b_0)$, $x \in \Sigma$ and $n \in \mathbb{N}$*

$$|\log \mathbb{E}_x[e^{tS_n(\xi)}] - n \log \lambda_{K,\xi}(t)| \le C_1|t| + \delta_n.$$

Proof. See Proposition 5.12 in [17] or Proposition 2.8 in [16]. □

Proof. (of Theorem 4.3) Denote by $\mathcal{H}(\overline{\mathbb{D}}_{b_0})$ the Banach space of analytic functions on $\mathbb{D}_{b_0}$ with a continuous extension to $\overline{\mathbb{D}}_{b_0}$. By Proposition 4.6(5), $\lambda = \lambda_{K,\xi} \in \mathcal{H}(\overline{\mathbb{D}}_{b_0})$ for all $(K, \mu, \xi) \in \mathfrak{X}$. Since $h > (c_{K,\xi})''(0)$, by Proposition 4.6(10) there is a neighborhood $\mathcal{V}$ of the observed Markov system $(K_0, \mu_0, \xi_0) \in \mathfrak{X}$ such that $h > (c_{K,\xi})''(0)$ for all $(K, \mu, \xi) \in \mathcal{V}$. Assume $\mathbb{E}_\mu(\xi) = 0$. Otherwise work with $\xi' = \xi - \mathbb{E}_\mu(\xi)\,\mathbf{1}$, for which $\mathbb{E}_\mu(\xi') = 0$. By Proposition 4.7,

$$\frac{1}{n} \log \mathbb{E}_x[e^{tS_n(\xi)}] \le c_{K,\xi}(t) + \frac{C_1|t| + \delta_n}{n},$$

with δ_n decreasing to 0 geometrically. Because $h > (c_{K,\xi})''(0)$, using again the equi-continuity in Proposition 4.6(10) there exists a small neighborhood $(-t_0, t_0)$, of $t = 0$ such that $c_{K,\xi}(t) \le \frac{ht^2}{2}$, for all $|t| < t_0$ and $(K, \mu, \xi) \in \mathcal{V}$. Applying Chebyshev's inequality (2.1), for $|t| < t_0$

$$\begin{aligned} \mathbb{P}_x[S_n(\xi) > n\varepsilon] &\le e^{-tn\varepsilon} \mathbb{E}_x[e^{tS_n(\xi)}] \\ &\le e^{-(t\varepsilon - c(t))n + C_1|t| + \delta_n} \\ &\le e^{-(t\varepsilon - \frac{ht^2}{2})n + C_1|t| + \delta_n}. \end{aligned}$$

Define now

$$C := 2e^{C_1 t_0 + \sup_{n \ge 0} \delta_n}.$$

Given $0 < \varepsilon < \varepsilon_0 := ht_0$, pick $t = \frac{\varepsilon}{h} \in (0, t_0)$. This choice of t minimizes the function $g(t) = e^{-(t\varepsilon - \frac{ht^2}{2})}$. Then

$$\mathbb{P}_x[S_n(\xi) > n\varepsilon] \le e^{-\frac{\varepsilon^2}{2h}n + C_1 \frac{\varepsilon}{h} + \delta_n} \le \frac{1}{2} C e^{-\frac{\varepsilon^2}{2h}n}.$$

By (A1), we derive the same conclusion for $-\xi$,

$$\mathbb{P}_x[S_n(\xi) < -n\varepsilon] = \mathbb{P}_\mu[S_n(-\xi) > n\varepsilon] \le \frac{1}{2} C e^{-\frac{\varepsilon^2}{2h}n}.$$

Thus, for all $(K, \mu, \xi) \in \mathcal{V}$, $0 < \varepsilon < \varepsilon_0$ and $n \in \mathbb{N}$,

$$\mathbb{P}_x[|S_n(\xi)| > n\varepsilon] \leq C e^{-\frac{\varepsilon^2}{2h} n}. \qquad \square$$

4.5. *Base LDT estimates*

To derive Theorem 4.1 from Theorem 4.3 we specify the data $(\mathcal{B}_\alpha, \|\cdot\|_\alpha)$ and $\mathfrak{X}$, and check the validity of the assumptions (B1)–(B7) and (A1)–(A4).

Consider a strongly mixing Markov system (K, μ) on the compact metric space Σ. Let $X^- = \Sigma^{\mathbb{Z}_0^-}$ be the space of sequences in Σ indexed in $\mathbb{Z}_0^- := \{\dots, -2, -1, 0\}$. As before, X^- is a compact metric space, and we denote by $\mathcal{F}$ its Borel σ-field. The kernel K on Σ induces another Markov kernel $\widetilde{K}$ on X^- defined by

$$\widetilde{K}_{(\dots,x_{-1},x_0)} := \int_\Sigma \delta_{(\dots,x_{-1},x_0,x_1)} K(x_0, dx_1).$$

Let $\mathbb{P}_\mu^-$ denote the Kolmogorov extension of (K, μ), which is also the unique $\widetilde{K}$-stationary measure. Theorem 4.3 will be applied to the Markov system $(X^-, \widetilde{K})$.

The spaces $\mathcal{H}_\alpha(X^-)$, defined in (4.2), can be regarded as consisting of $\mathcal{F}$-measurable functions on X^-. They form the scale of Banach algebras satisfying (B1)–(B7), where the Markov operators $Q_{\widetilde{K}}$ act.

As noted before the spaces $\mathcal{H}_\alpha(X)$ are Banach algebras, as well as lattices, for all $\alpha \in [0, 1]$ (see Proposition 4.2). For $\alpha = 0$, the seminorm v_0 measures the variation of f because

$$v_0(f) = \sup\{|f(x) - f(x')| : x, x' \in X\}.$$

Hence $\mathcal{H}_0(X) = L^\infty(X)$, while the norm $\|\cdot\|_0$ is equivalent to $\|\cdot\|_\infty$. These considerations show that $\{\mathcal{H}_\alpha(X)\}_{\alpha\in[0,1]}$ satisfies the assumptions (B1)–(B4). For the remaining ones, (B5)–(B7), see [35]. Because each $\mathcal{H}_\alpha(X^-)$ is a sub-algebra (and a sub-lattice) of $\mathcal{H}_\alpha(X)$ properties (B1)–(B7) hold for $\{\mathcal{H}_\alpha(X^-)\}_{\alpha\in[0,1]}$ as well.

Let $\mathfrak{X}^\alpha_{\widetilde{K},L}$ be the space of observed Markov systems $(X^-, \widetilde{K}, \xi)$ with $\xi \in \mathcal{H}_\alpha(X^-)$ and $\|\xi\|_\alpha \leq L$. This space is identified as a subspace of $\mathcal{H}_\alpha(X^-)$, and endowed with the corresponding norm distance.

The Markov operator of the kernel $\widetilde{K}$ is $Q_{\widetilde{K}} : L^\infty(X^-) \to L^\infty(X^-)$,

$$(Q_{\widetilde{K}} f)(\dots, x_{-1}, x_0) := \int_\Sigma f(\dots, x_{-1}, x_0, x_1) K(x_0, dx_1).$$

This operator acts continuously on $\mathcal{H}_a(X^-)$.

Proposition 4.8. *For all* $f \in \mathcal{H}_\alpha(X^-)$ *and* $n \in \mathbb{N}$,

(1) $\|(Q_{\widetilde{K}})^n f\|_\infty \leq \|f\|_\infty$,
(2) $v_\alpha((Q_{\widetilde{K}})^n f) \leq \max\{2\|(Q_{\widetilde{K}})^n f\|_\infty, 2^{-n\alpha} v_\alpha(f)\}$.

Proof. See Proposition 5.13 in [17] or Proposition 3.1 in [16]. $\square$

Next proposition shows that $\mathfrak{X}^{\alpha}_{\widetilde{K},L}$ satisfies (A2) with range $[\alpha_1, 1]$ for any given $\alpha_1 > 0$. The setting constants $C > 0$ and $0 < \sigma < 1$ depend on the number α_1.

Proposition 4.9. *If (K, μ) is strongly mixing, then given $0 < \alpha_1 < 1$ there are constants $C > 0$ and $0 < \sigma < 1$ such that for all $\alpha_1 \leq \alpha \leq 1$, $Q_{\widetilde{K}} : \mathcal{H}_\alpha(X^-) \to \mathcal{H}_\alpha(X^-)$ is quasi-compact and simple with spectral constants C and σ, i.e. for all $f \in \mathcal{H}_\alpha(X^-)$,*

$$\|(Q_{\widetilde{K}})^n f - \langle f, \mathbb{P}^-_\mu \rangle \mathbf{1}\|_\alpha \leq C\sigma^n \|f\|_\alpha.$$

Proof. See Proposition 5.14 in [17] or Proposition 3.2 in [16]. □

Proposition 4.4, and the preceding one, imply that $\mathfrak{X}^{\alpha}_{\widetilde{K},L}$ satisfies all the assumptions (A1)–(A4) in the range $[\alpha_1, 1]$, which makes Theorem 4.3 applicable.

Proof. (of Theorem 4.1) Consider $0 < \alpha \leq 1$ and $\xi \in \mathcal{H}_\alpha(X^-)$. Take $\alpha_1 < \alpha$ and let $L > \|\xi\|_\alpha$. Then $\xi \in \mathfrak{X}^{\alpha}_{\widetilde{K},L}$. By Proposition 4.9 the space $\mathfrak{X}^{\alpha}_{\widetilde{K},L}$ satisfies (A2) in the range $[\alpha_1, 1]$, and hence, by Proposition 4.4, it satisfies all (A1)–(A4) in the same range.

For $\delta > 0$ small, the function $\hat{\lambda}_K : \mathfrak{X}^{\alpha}_{\widetilde{K},\delta} \to \mathbb{C}$, $\xi \mapsto \hat{\lambda}_K(\xi) =$ maximal eigenvalue of $Q_{K,\xi}$, is well-defined and analytic. The analyticity follows from Proposition 4.4. Decreasing δ we can assume that $\hat{\lambda} = \hat{\lambda}_K$ is bounded. Choose $b > 0$ small, such that $b\,L < \delta$, and note that $\lambda_{K,\xi}(z) = \hat{\lambda}(z\,\xi)$ for all $z \in \mathbb{D}_b$. Hence the family of analytic functions $\{\lambda_{K,\xi}(z)\}_{\xi \in \mathfrak{X}^{\alpha}_{\widetilde{K},L}}$ is uniformly bounded over $\overline{\mathbb{D}_b}$. Shrinking b, the derivatives $\lambda'_{K,\xi}(z)$ and $\lambda''_{K,\xi}(z)$ are also bounded. Thus, there exists $h > 0$ such that $(c_{K,\xi})''(t) < h$ for all $t \in [-b, b]$ and $\xi \in \mathfrak{X}^{\alpha}_{\widetilde{K},L}$. By Theorem 4.3 and Remark 4.4 there are constants ε_0 and $C > 0$ such that for all $\xi \in \mathfrak{X}^{\alpha}_{\widetilde{K},L}$, $0 < \varepsilon < \varepsilon_0$ and all $n \in \mathbb{N}$,

$$\mathbb{P}^-_\mu \left[\left| \frac{1}{n} S_n(\xi) - \mathbb{E}^-_\mu(\xi) \right| \geq \varepsilon \right] \leq C e^{-\frac{\varepsilon^2}{2h} n}.$$

Consider the natural (measure preserving) projection $\pi : X \to X^-$. Since

$$\pi^{-1} \left[\left| \frac{1}{n} S_n(\xi) - \mathbb{E}^-_\mu(\xi) \right| \geq \varepsilon \right] = \left\{ x \in X : \left| \frac{1}{n} \sum_{j=0}^{n-1} \xi(T^j(x)) - \int_X \xi \, d\mathbb{P}_\mu \right| \geq \varepsilon \right\}$$

all observables $\xi \in \mathfrak{X}^{\alpha}_{\widetilde{K},L}$ satisfy a uniform base-LDT estimate.

The constants δ, b, h, ε_0 and C depend all on L and $\hat{\lambda}_K$, i.e. on L and K. □

4.6. *Fiber LDT estimates*

Finally, we use Theorem 4.3 to establish the fiber LDT Theorem 4.2. For that we specify the data $(\mathcal{B}_\alpha, \|\cdot\|_\alpha)$ and $\mathfrak{X}$, and check that (B1)–(B7) and (A1)–(A4) hold.

Consider the space $\mathcal{B}^\infty_m(K)$ of random cocycles over a Markov system (K, μ) introduced in Definition 4.5. For each cocycle $A \in \mathcal{B}^\infty_m(K)$ we define a Markov kernel on $\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m)$ by

$$K_A(x, y, p) := \int_\Sigma \delta_{(y,z,A(y,z)p)} K(y, dz). \tag{4.4}$$

Theorem 4.3 will be applied to the Markov systems $(\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m), K_A)$ with $A \in \mathcal{B}^\infty_m(K)$. The observables for which we derive fiber LDT estimates are the functions $\xi_A : \Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m) \to \mathbb{R}$ defined by

$$\xi_A(x, y, p) := \log\|A(x, y)p\|. \tag{4.5}$$

Next we introduce the scale of Banach algebras satisfying (B1)–(B7). First we specify a distance on the real projective space $\mathbb{P}(\mathbb{R}^m)$. Let

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

where each point $p \in \mathbb{P}(\mathbb{R}^m)$ is identified with any of its representative vectors $p \in \mathbb{R}^m \setminus \{0\}$. Given $0 \le \alpha \le 1$ and a bounded measurable function $f : \Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m) \to \mathbb{C}$, define

$$\|f\|_\alpha := v_\alpha(f) + \|f\|_\infty, \tag{4.6}$$

$$v_\alpha(f) := \sup_{\substack{x,y,\in\Sigma \\ p\neq q}} \frac{|f(x, y, p) - f(x, y, q)|}{\delta(p, q)^\alpha}. \tag{4.7}$$

Let $\mathcal{H}_\alpha(\Sigma\times\Sigma\times\mathbb{P}(\mathbb{R}^m))$ be the space of functions $f \in L^\infty(\Sigma\times\Sigma\times\mathbb{P}(\mathbb{R}^m))$ such that $v_\alpha(f) < +\infty$, i.e. the space of measurable functions which are α-Hölder continuous in the last variable. This is an algebra because:

Proposition 4.10. *For $0 \le \alpha \le 1$, the function v_α is a seminorm on $\mathcal{H}_\alpha(\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m))$ such that for $f, g \in \mathcal{H}_\alpha(\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m))$,*

$$v_\alpha(fg) \le \|f\|_\infty v_\alpha(g) + \|g\|_\infty v_\alpha(f).$$

That $(\mathcal{H}_\alpha(\Sigma\times\Sigma\times\mathbb{P}(\mathbb{R}^m)), \|\cdot\|_\alpha)$ is a unital Banach algebra, as well as a lattice, can be easily checked. Thus, conditions (B1), (B3) and (B4) follow. For $\alpha = 0$, the seminorm v_0 measures the variation of f on the projective coordinate

$$v_0(f) = \sup\{|f(x, y, p) - f(x, y, p')| : x, y \in \Sigma, p, p' \in \mathbb{P}(\mathbb{R}^m)\}.$$

Hence $\mathcal{H}_0(\Sigma\times\Sigma\times\mathbb{P}(\mathbb{R}^m)) = L^\infty(\Sigma\times\Sigma\times\mathbb{P}(\mathbb{R}^m))$, while the norm $\|\cdot\|_0$ is equivalent to $\|\cdot\|_\infty$. This proves (B2). Conditions (B5) and (B6) hold because the projective metric δ takes values in $[0, 1]$. Finally (B7) follows from the equality

$$\frac{\Delta}{d^{\alpha_1}} = \left(\frac{\Delta}{d^{\alpha_0}}\right)^{\frac{\alpha_2-\alpha_1}{\alpha_2-\alpha_0}} \left(\frac{\Delta}{d^{\alpha_2}}\right)^{\frac{\alpha_1-\alpha_0}{\alpha_2-\alpha_0}},$$

which holds for all $\Delta \ge 0$ and $d > 0$.

Let now $\mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ be the subspace of functions $f(x, y, p)$ in $\mathcal{H}_\alpha(\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m))$ that do not depend on the first coordinate x. This subspace is clearly a closed sub-algebra of $\mathcal{H}_\alpha(\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m))$. Therefore, the family $\{(\mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m)), \|\cdot\|_\alpha)\}_{\alpha \in [0,1]}$ is another scale of Banach sub-algebras satisfying the assumptions (B1)–(B7).

Given a cocycle $A \in \mathcal{B}^\infty_m(K)$, consider the linear transformation $Q_A : L^\infty(\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m)) \to L^\infty(\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m))$ defined by

$$(Q_A f)(x, y, p) := \int_\Sigma f(y, z, A(y, z)p) K(y, dz). \tag{4.8}$$

This is the Markov operator associated with the kernel (4.4).

We can now introduce the metric space of observed Markov systems

$$\mathfrak{X} := \{(K_A, \mu_A, \pm\xi_A) : A \in \mathcal{B}^\infty_m, A \text{ irreducible}, L_1(A) > L_2(A)\}.$$

This space is identified with a subspace of $\mathcal{B}^\infty_m$, and endowed with the distance

$$\operatorname{dist}((K_A, \mu_A, \xi_A), (K_B, \mu_B, \xi_B)) := d_\infty(A, B).$$

Assumption (A1) is clear from the definition of $\mathfrak{X}$.

Since $(Q_A f)(x, y, p)$ does not depend on the coordinate x, the Markov operator Q_A leaves invariant the subspace of functions $f(x, y, p)$ that are constant in x. Next, we are going to see that Q_A acts invariantly on the subspace $\mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$.

Given $A \in \mathcal{B}^\infty_m(K)$ and $0 < \alpha \le 1$, define for all $n \in \mathbb{N}$,

$$\kappa^n_\alpha(A) := \sup_{x \in \Sigma, p \ne q} \mathbb{E}_x \left[\left(\frac{\delta(A^{(n)} p, A^{(n)} q)}{\delta(p, q)} \right)^\alpha \right] \in [0, +\infty]. \tag{4.9}$$

The following lemma highlights the importance of this quantity.

Lemma 4.2. *Given $A \in \mathcal{B}^\infty_m(K)$, $f \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ and $n \in \mathbb{N}$,*

$$v_\alpha(Q^n_A f) \le \kappa^n_\alpha(A) v_\alpha(f).$$

Proof. See Lemma 5.5 in [17] or Lemma 3.6 in [16]. □

Lemma 4.3. *The sequence $\{\kappa^n_\alpha(A)\}_{n \ge 0}$ is sub-multiplicative, i.e.*

$$\kappa^{n+\ell}_\alpha(A) \le \kappa^n_\alpha(A) \kappa^\ell_\alpha(A) \quad \text{for } n, \ell \in \mathbb{N}.$$

In particular,

$$\lim_{n \to +\infty} \kappa^n_\alpha(A)^{1/n} = \inf\{\kappa^n_\alpha(A)^{1/n} : n \in \mathbb{N}\}.$$

Proof. See Lemma 5.6 in [17] or Lemma 3.7 in [16]. □

These constants become finite provided α is small enough.

Lemma 4.4. *Given $A \in \mathcal{B}^\infty_m(K)$ and $n \in \mathbb{N}$, for all $0 < \alpha \le \frac{1}{4n}$*

$$\kappa^n_\alpha(A) \le \max\{\|A\|_\infty, \|A^{-1}\|_\infty\}.$$

Proof. See Lemma 5.7 in [17] or Lemma 3.8 in [16]. □

From the previous lemma, we see that the operator Q_A leaves the subspace $\mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ invariant, for all small enough $\alpha > 0$. To prove that Q_A is quasi-compact and simple, the hypothesis (2)–(3) in Theorem 4.2 are essential. They are used in the following lemma.

Lemma 4.5. *Given $A \in \mathcal{B}_m^\infty(K)$ such that $(K_A, \mu_A, \xi_A) \in \mathfrak{X}$, for some $n \in \mathbb{N}$, all $x \in \Sigma$ and $p \neq q$ in $\mathbb{P}(\mathbb{R}^m)$,*

$$\mathbb{E}_x \left[\log \frac{\delta(A^{(n)}p, A^{(n)}q)}{\delta(p,q)} \right] \leq -1.$$

Proof. See Lemma 5.9 in [17] or Lemma 3.10 in [16]. □

Proposition 4.11. *Given $A \in \mathcal{B}_m^\infty(K)$ such that $(K_A, \mu_A, \xi_A) \in \mathfrak{X}$, there exist $\mathcal{V}$, a neighborhood of A in $\mathcal{B}_m^\infty(K)$, and there are positive constants $0 < \alpha_1 < \frac{\alpha_0}{2} < \alpha_0$, $C > 0$ and $0 < \sigma < 1$ such that*

$$v_\alpha(Q_B^n f) \leq C\sigma^n v_\alpha(f),$$

for all $B \in \mathcal{V}$, $\alpha \in [\alpha_1, \alpha_0]$, $f \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ and $n \in \mathbb{N}$.

Proof. For a fixed cocycle $A \in \mathcal{B}_m^\infty(K)$ such that $(K_A, \mu_A, \xi_A) \in \mathfrak{X}$, the contraction statement, w.r.t. the seminorm v_α, follows from Proposition 4.5. The generalization of this property to a neighborhood $\mathcal{V}$ of A follows from the modulus of continuity

$$|\kappa_\alpha^n(A) - \kappa_\alpha^n(B)| \leq C_n d_\infty(A, B),$$

where the constant C_n depends on n.

See also Lemma 5.17 in [17] or Proposition 3.11 in [16] for a complete proof. □

Next proposition implies (A2).

Proposition 4.12. *Given $A \in \mathcal{B}_m^\infty(K)$ such that $(K_A, \mu_A, \xi_A) \in \mathfrak{X}$, there exist a neighborhood $\mathcal{V}$ of $A \in \mathcal{B}_m^\infty(K)$, a range $0 < \alpha_1 < \frac{\alpha_0}{2} < \alpha_0 \leq 1$ and there are constants $C > 0$ and $0 < \sigma < 1$ such that for all $B \in \mathcal{V}$, $\alpha \in [\alpha_1, \alpha_0]$ and $f \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$,*

$$\|Q_B^n f - \langle f, \mu_B \rangle \mathbf{1}\|_\alpha \leq C\sigma^n \|f\|_\alpha,$$

where μ_B denotes the (unique) K_B-stationary measure on $\Sigma \times \mathbb{P}(\mathbb{R}^m)$.

Proof. This proposition follows from the proof of Theorem 3.7 in [4], using the conclusion of Proposition 4.11. See also Proposition 5.18 in [17] or Proposition 3.12 in [16] for a complete proof. □

Like the Markov operator Q_A, the Laplace–Markov operator $Q_{A,z}$ of the observed Markov system (K_A, μ_A, ξ_A),

$$(Q_{A,z} f)(x,p) = \int_\Sigma f(y, A(x,y)p) \|A(x,y)\|^z K(x, dy),$$

acts invariantly on the subspaces $\mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$, with small enough $\alpha > 0$. Choose $0 < \alpha_1 < \alpha_0 \leq 1$ according to Proposition 4.11.

Assumption (A3) is automatically satisfied because $\|A\|_\infty < \infty$ and $\|A^{-1}\|_\infty < \infty$ which imply that $\xi_A \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ for all $\alpha > 0$. Note that $Q_{A,z} = Q_A \circ D_{e^{z\xi_a}}$, where $D_{e^{z\xi_a}}$ denotes the multiplication operator by $e^{z\xi_a}$. This is a bounded operator because $\mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ is a Banach algebra containing the function $e^{z\xi_a}$.

Finally the next lemma proves (A4).

Lemma 4.6. *Given* $A, B \in \mathcal{B}^\infty_m(K)$ *and* $b > 0$, *there is a constant* $C_2 > 0$ *such that for all* $f \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$, *and all* $z \in \mathbb{C}$ *such that* $\operatorname{Re} z \leq b$,

$$\|Q_{A,z} f - Q_{B,z} f\|_\infty \leq C_2 d_\infty(A,B)^\alpha \|f\|_\alpha.$$

Proof. See Lemma 5.10 in [17] or Lemma 3.14 in [16]. □

Proof. (of Theorem 4.2) The space of observed Markov systems $\mathfrak{X}$ satisfies all assumptions (A1)–(A4). Hence, by Theorem 4.3, there exist a neighborhood $\mathcal{V}$ of $A \in \mathcal{B}^\infty_m(K)$ and constants $\varepsilon_0, C, h > 0$ such that for all $B \in \mathcal{V}$, $0 < \varepsilon < \varepsilon_0$, $(x,p) \in \Sigma \times \mathbb{P}(\mathbb{R}^m)$ and $n \in \mathbb{N}$,

$$\mathbb{P}_x \left[\left| \frac{1}{n} \log \|B^{(n)} p\| - L_1(B,\mu) \right| \geq \varepsilon \right] \leq C e^{-\frac{\varepsilon^2}{2h} n},$$

and integrating w.r.t. μ we get for all $p \in \mathbb{P}(\mathbb{R}^m)$,

$$\mathbb{P}_\mu \left[\left| \frac{1}{n} \log \|B^{(n)} p\| - L_1(B,\mu) \right| \geq \varepsilon \right] \leq C e^{-\frac{\varepsilon^2}{2h} n}.$$

Choose the canonical basis $\{e_1, \ldots, e_d\}$ of $\mathbb{R}^m$ and consider the following norm $\|\cdot\|'$ on $\mathrm{Mat}_d(\mathbb{R})$, $\|M\|' := \max_{1 \leq j \leq d} \|M\, e_j\|$. Since this norm is equivalent to the operator norm, for all $B \in \mathcal{V}$, $p \in \mathbb{P}(\mathbb{R}^m)$ and $n \in \mathbb{N}$,

$$\|B^{(n)} p\| \leq \|B^{(n)}\| \lesssim \|B^{(n)}\|' = \max_{1 \leq j \leq d} \|B^{(n)} e_j\|.$$

Thus a simple comparison of the deviation sets gives

$$\mathbb{P}_\mu \left[\left| \frac{1}{n} \log \|B^{(n)}\| - L_1(B,\mu) \right| \geq \varepsilon \right] \lesssim e^{-\frac{\varepsilon^2}{2h} n}$$

for all $B \in \mathcal{V}$, $0 < \varepsilon < \varepsilon_0$ and $n \in \mathbb{N}$. □

5. Continuity of Lyapunov Exponents

In this section we describe an *abstract*, modular approach to proving continuity of the Lyapunov exponents, which uses an inductive procedure based on the deterministic, general avalanche principle. The main advantage of this approach, besides the fact that it provides *quantitative* estimates, is its *versatility.* This approach applies to quasi-periodic cocycles (one and multivariable torus translations), to random cocycles (Bernoulli and Markov systems) and to any other types of base dynamics as long as appropriate LDT estimates are satisfied. Moreover, compared to other available *quantitative* continuity results, this approach allows for weaker assumptions, e.g., for quasi-periodic models, unlike in [23, 47], we do not have to assume positivity/simplicity of the Lyapunov exponents, while for the random i.i.d. model, unlike in [37], we do not have to assume a contraction property.

5.1. *Literature review*

The problem of continuity of the Lyapunov exponents for analytic, quasi-periodic cocycles has been widely studied. The works of M. Goldstein and W. Schlag in [23] and J. Bourgain and S. Jitomirskaya in [9] are classic papers on the subject. They refer to *Schrödinger* cocycles, as defined in Sec. 3, and continuity is understood relative to the energy parameter and/or the frequency.

In [9], the authors prove joint continuity in energy E and frequency ω, at all points (E, ω) with ω irrational.

In [23], for the one frequency case, assuming a strong Diophantine condition on the frequency, the authors prove a sharp fiber LDT estimate and establish the avalanche principle (AP) for $\mathrm{SL}(2, \mathbb{R})$ matrices. Based on these ingredients, they develop an inductive procedure that leads to Hölder continuity of the (top) Lyapunov exponent as a function of the energy E, under the assumption of a positive lower bound on the Lyapunov exponent. A similar approach is applied to the multifrequency Diophantine torus translation case, leading to weak-Hölder continuity of the Lyapunov exponent, the weaker modulus of continuity being due to a weaker version of the fiber LDT estimate available in this case.

Extensions of the ideas and results in [23] to other related models were obtained in [10, 33, 34].

J. Bourgain proved in [7] joint continuity in energy and frequency for the multifrequency torus translation model.

A higher dimensional version of the AP, along with a higher dimensional version of the result in [23], were obtained in [47] for Schrödinger-like cocycles, under the restrictive assumption that all Lyapunov exponents are simple. It was also indicated in [47] that this method is in some sense *modular*, a statement that motivated in part our recent work being surveyed here.

With motivations that are both intrinsic and related to mathematical physics problems (e.g., spectral properties of Jacobi-type operators), the study of continuity properties of the Lyapunov exponents has been extended from Schrödinger cocycles to more general ones, including higher dimensional cocycles and/or cocycles with singularities. Each extension comes with significant technical challenges, requiring new methods.

C. Marx and S. Jitomirskaya proved joint continuity in energy and frequency (one frequency case) for $\mathrm{Mat}(2, \mathbb{C})$-valued analytic cocycles (see [30] and references therein). Using a different approach, A. Ávila, S. Jitomirskaya, C. Sadel extended this result to multidimensional (i.e. $\mathrm{Mat}(m, \mathbb{C})$-valued) analytic cocycles (see [2]).

We note that both results mentioned above ([30, 2]) treating *one* frequency torus translations, rely crucially on the *convexity* of the top Lyapunov exponent of the complexified cocycle as a function of the imaginary variable, by firstly establishing continuity *away* from the torus. This approach immediately breaks down in the *multifrequency* case.

Our work in [12] presents a geometric, conceptual approach to the avalanche principle, which allows us to generalize it to higher dimensions, namely to blocks of $\mathrm{GL}(m, \mathbb{R})$ matrices, and further (see [14, 17]), to any blocks of non-zero matrices in $\mathrm{Mat}(m, \mathbb{R})$. We use this general AP in [12] to prove Hölder (or weak-Hölder for multifrequency translations) continuity of the Lyapunov exponents of $\mathrm{GL}(m, \mathbb{R})$-valued analytic cocycles in a neighborhood of a cocycle with *simple* Lyapunov exponents. Moreover, continuity of all Lyapunov exponents (but without a modulus of continuity) holds everywhere, regardless of the multiplicity of the Lyapunov exponents.

Our next goal was to handle cocycles with singularities (i.e. not necessarily $\mathrm{GL}(m, \mathbb{R})$-valued), which, as explained in Sec. 3, is especially delicate in the multifrequency case. We will explain later in this section how the uniform fiber LDT in Theorem 3.1 leads to continuity of all Lyapunov exponents and to weak-Höllder continuity of simple Lyapunov exponents for such cocycles.

While unlike in [30, 2], we do require Diophantine translations, and the frequency is fixed, our method applies equally to translations on the one or the *higher* dimensional torus.

At the other end of the type of ergodic behavior of the base dynamics — the random case, continuity results for linear cocycles over Bernoulli shifts in the *generic* case go back to H. Furstenberg and Kifer, see [21].

E. Le Page proved in [37] Hölder continuity of the top Lyapunov exponent for a one-parameter family of cocycles over the Bernoulli shift, under irreducibility and contraction assumptions, which are assumed to hold uniformly throughout this family.

Compared with this theorem, our recent results being surveyed here, do not require any contraction assumption and provide continuity of all exponents (regardless of the gaps in the Lyapunov spectrum) and Hölder continuity (in the presence of gaps). The statement is about continuity in the space of irreducible cocycles and not just for one-parameter families. It is also more general since we address cocycles over mixing Markov shifts, and not just over the Bernoulli shifts. We are not aware of any generalization of Le Page's theorem, on the continuity of the top Lyapunov exponent, for irreducible cocycles over strongly mixing Markov shifts.

C. Bocker-Neto and M. Viana [3] proved continuity of the Lyapunov exponents for two-dimensional cocycles over Bernoulli shifts without any irreducibility assumptions (the result does not provide a modulus of continuity though). A higher dimensional version of this result was announced by A. Ávila, A. Eskin and M. Viana (see the monograph [51]). An extension of results from [3] 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 in [39]. We note, for the interested reader, that a general one-stop reference for continuity results for random cocycles is M. Viana's monograph [51].

5.2. *Abstract continuity theorem*

Let (X, μ, T) be an ergodic system. Before we formulate the abstract continuity theorem (ACT) of the Lyapunov exponents of a linear cocycle, we need a few more definitions.

Definition 5.1. A space of measurable cocycles $\mathcal{C}$ is any class of matrix valued functions $A : X \to \mathrm{Mat}(m, \mathbb{R})$, where $m \in \mathbb{N}$ is not fixed, such that every $A : X \to \mathrm{Mat}(m, \mathbb{R})$ in $\mathcal{C}$ has the following properties:

(1) A is measurable.
(2) $\|A\| \in L^\infty(\mu)$.
(3) The exterior powers $\wedge_k A : X \to \mathrm{Mat}_{\binom{m}{k}}(\mathbb{R})$ are in $\mathcal{C}$, for $k \leq m$.

Each subspace $\mathcal{C}_m := \{A \in \mathcal{C} \mid A : X \to \mathrm{Mat}_m(\mathbb{R})\}$ is *a priori* endowed with a distance $\mathrm{dist}\colon \mathcal{C}_m \times \mathcal{C}_m \to [0, +\infty)$ which is at least as fine as the L^∞ distance, i.e. for all $A, B \in \mathcal{C}_m$ we have

$$\mathrm{dist}(B, A) \geq \|B - A\|_{L^\infty}.$$

We assume a correlation between the distances on each of these subspaces, namely the map $\mathcal{C}_m \ni A \mapsto \wedge_k A \in \mathcal{C}_{\binom{m}{k}}$ is locally Lipschitz.

The functions $\frac{1}{n}\log\|A^{(n)}(x)\|$ are integrable, and their integrals are what we called in Sec. 2 finite scale (top) Lyapunov exponents. We need stronger integrability assumptions on these functions.

Definition 5.2. A cocycle $A \in \mathcal{C}$ is called L^2-bounded if there is a constant $C < \infty$, which we call its L^2-bound, such that for all $n \geq 1$ we have:

$$\left\| \frac{1}{n}\log\|A^{(n)}(\cdot)\| \right\|_{L^2} < C. \tag{5.1}$$

A cocycle $A \in \mathcal{C}_m$ is called uniformly L^2-bounded, if the above bound holds uniformly near A.

Given a cocycle $A \in \mathcal{C}$ and an integer $N \in \mathbb{N}$, denote by $\mathcal{F}_N(A)$ the algebra generated by the sets $\{x \in X : \|A^{(n)}(x)\| \leq c\}$ or $\{x \in X : \|A^{(n)}(x)\| \geq c\}$ where $c \geq 0$ and $0 \leq n \leq N$.

Let Ξ be a set of measurable functions $\xi : X \to \mathbb{R}$, which we call observables. Let $A \in \mathcal{C}$.

Definition 5.3. We say that Ξ and A are compatible if for every integer $N \in \mathbb{N}$, for every set $F \in \mathcal{F}_N(A)$ and for every $\epsilon > 0$, there is an observable $\xi \in \Xi$ such that:

$$\mathbb{1}_F \leq \xi \quad \text{and} \quad \int_X \xi d\mu \leq \mu(F) + \epsilon. \tag{5.2}$$

Theorem 5.1. *Consider an ergodic system (X, μ, T), a space of measurable cocycles $\mathcal{C}$, a set of observables Ξ, a set of LDT parameters $\mathcal{P}$ and assume the following*:

(1) Ξ *is compatible with every cocycle* $A \in \mathcal{C}$.
(2) *Every observable* $\xi \in \Xi$ *satisfies a base LDT w.r.t.* $\mathcal{P}$.
(3) *Every* $A \in \mathcal{C}$ *with* $L_1(A) > -\infty$ *is uniformly* L^2*-bounded.*
(4) *Every cocycle* $A \in \mathcal{C}$ *for which* $L_1(A) > L_2(A)$ *satisfies a uniform fiber LDT w.r.t.* $\mathcal{P}$.

Then all Lyapunov exponents $L_k \colon \mathcal{C}_m \to [-\infty, \infty)$, $1 \leq k \leq m$, $m \in \mathbb{N}$ *are continuous functions of the cocycle.*

Moreover, given $A \in \mathcal{C}$ *and* $1 \leq k \leq m$, *if the Lyapunov exponent* $L_k(A)$ *is simple, then locally near* A *the map* L_k *has a modulus of continuity* $\omega(h) := [\underline{\iota}(c \log \frac{1}{h})]^{1/2}$ *for some* $c = c(A) > 0$ *and for some deviation measure function* $\underline{\iota} = \underline{\iota}(A)$ *corresponding to an LDT parameter in the set* $\mathcal{P}$.

5.3. *Deriving continuity*

We describe the applicability of the ACT to the quasi-periodic and random models, for which we have already derived LDT estimates in Secs. 3 and 4.

Quasi-periodic models

For every $m \geq 1$, let $\mathcal{C}_m$ be the set of cocycles $A \in C^\omega_r(\mathbb{T}^d, \mathrm{Mat}(m, \mathbb{R}))$ with $\det[A(x)] \not\equiv 0$. Then $\mathcal{C}_m$ is an open set in $C^\omega_r(\mathbb{T}^d, \mathrm{Mat}(m, \mathbb{R}))$, which we equip with the induced distance.

Theorem 5.2. *For all dimensions* m, *assuming* ω *Diophantine, the maps* $L_k : \mathcal{C}_m \to \mathbb{R}$, $1 \leq k \leq m$ *are continuous. Moreover, if* $A \in \mathcal{C}_m$ *is such that* $L_k(A)$ *is simple, then locally near* A *the map* L_k *is weak-Hölder continuous.*

Proof. We explain briefly how the assumptions of the ACT are satisfied in this setting.

Given $A \in \mathcal{C}_m$ and $N \in \mathbb{N}$, note that the sets $\{x \in \mathbb{T}^d : \|A^{(n)}(x)\| \leq c\}$ or $\{x \in \mathbb{T}^d : \|A^{(n)}(x)\| \geq c\}$ for some $1 \leq n \leq N$ and $c > 0$ are closed Jordan measurable, so the algebra $\mathcal{F}_N(A)$ generated by them consists only of Jordan measurable sets.

Let $\Xi := \mathcal{C}_0(\mathbb{T}^d)$ be the set of all *continuous* observables $\xi : \mathbb{T}^d \to \mathbb{R}$. By the regularity of the Borel measure, there is an open set $U \supseteq \overline{F}$ such that $|U| \le |F| + \epsilon$. By Urysohn's lemma, there is a continuous function $\xi \in \Xi$ such that $0 \le \xi \le 1$, $\xi \equiv 1$ on $\overline{F}$ and $\xi \equiv 0$ on $U^{\complement}$. Then

$$\mathbb{1}_F \le \xi \quad \text{and} \quad \int_{\mathbb{T}^d} \xi dx \le |U| \le |F| + \epsilon,$$

which shows that Ξ is compatible with every cocycle in $\mathcal{C}$.

The (uniform) L^2-boundedness follows from (3.18) and the (uniform) Łojasiewicz inequality (3.20). The latter is used to prove that if f is analytic on $\mathcal{A}_r^d$ and $f \not\equiv 0$, then $\|\log|f|\|_{L^2} \le C(f) < \infty$, and the bound $C(f)$ is stable under perturbations.

Finally, base LDTs hold trivially for this model, while uniform fiber LDTs were established in Theorem 3.1 with deviation functions $\underline{\epsilon}(t) \equiv t^{-a}$ and $\underline{\iota}(t) \equiv e^{-ct^b}$. Therefore, all Lyapunov exponents are continuous.

Moreover, defining $\mathcal{P}$ to be the set of triplets $(\underline{n_0}, \underline{\epsilon}, \underline{\iota})$ where $\underline{n_0} \in \mathbb{N}$, $\underline{\epsilon}(t) \equiv t^{-a}$ and $\underline{\iota}(t) \equiv e^{-ct^b}$ with $a, b, c > 0$, a simple calculation shows that the modulus of continuity locally near simple Lyapunov exponents is $\omega(h) = e^{-c[\log(1/h)]^b}$ for some $c, b > 0$. We call such a modulus of continuity weak-Hölder. $\square$

Random models

Let (K, μ) be a Markov system. Consider the space $\mathcal{B}_m^\infty(K)$ of random cocycles $A : \Sigma \times \Sigma \to \mathrm{GL}(m, \mathbb{R})$ introduced in Definition 4.5. A cocycle $A \in \mathcal{B}_m^\infty(K)$ is called totally irreducible if all its exterior powers are irreducible. Denote by $\mathcal{I}_m^\infty(K)$ the space of totally irreducible cocycles in $\mathcal{B}_m^\infty(K)$.

Theorem 5.3. *If (K, μ) is strongly mixing then all Lyapunov exponents $L_k : \mathcal{I}_m^\infty(K) \to [-\infty, \infty)$, $1 \le k \le m$, are continuous functions of the cocycle. Moreover, given $A \in \mathcal{I}_m^\infty(K)$, if the Lyapunov spectrum of A is simple, then locally near A, all Lyapunov exponents are Hölder continuous functions of the cocycle.*

Proof. We are going to apply Theorem 5.1 to the space $\mathcal{C}_m = \mathcal{I}_m^\infty(K)$ of totally irreducible measurable cocycles over the Markov dynamical system $(T, X, \mathcal{F}, \mathbb{P}_\mu)$. Consider the space $\mathcal{P}$ of LDT parameters $\underline{p} = (\underline{n_0}, \underline{\epsilon}, \underline{\iota})$ with $\underline{n_0} \in \mathbb{N}$, $\underline{\epsilon}(t) \equiv \varepsilon$ and $\underline{\iota}(t) \equiv e^{-ct}$, for some constants $\varepsilon, c > 0$.

The chosen set of observables is the Banach algebra $\Xi = \mathcal{H}_\alpha(X^-)$, for some $\alpha > 0$ small enough.

The compatibility condition (1) in Theorem 5.1 is automatic because Ξ contains all functions $\mathbb{1}_F$ with $F \in \mathcal{F}_N(A)$, $N \in \mathbb{N}$ and $A \in \mathcal{B}_m^\infty(K)$.

The base-LDT assumption (2) for every observable $\xi \in \Xi$ is a consequence of Theorem 4.1.

Condition (3) holds because, by the definition of the metric on $\mathcal{B}_m^\infty(K)$, every cocycle in $\mathcal{B}_m^\infty(K)$ is uniformly L^2-bounded.

Finally, the fiber-LDT assumption (4) follows from Theorem 4.2.

Thus, applying Theorem 5.1, we see that all Lyapunov exponents are continuous functions on $\mathcal{I}_m^\infty(K)$.

A simple computation shows that the modulus of continuity associated to the choice $\mathcal{P}$ (of the space of LDT parameters) is the modulus of Hölder continuity.

□

The concept of irreducibility extends to linear cocycles depending on finitely many symbols, over Markov systems of finite order. There is a standard procedure to reduce them to cocycles in the previous setting. This reduction corresponds to redefining the space of symbols to be some product Σ^k, with large k.

Remark 5.1. Theorem 5.3 (on the continuity of all Lyapunov exponents on the appropriate space of irreducible cocycles) extends to cocycles depending on finitely many symbols, and over Markov systems of finite order (see Theorem 5.5 in [17] or Theorem 4.1 in [16]).

5.4. *On the proof of the ACT*

We describe the inductive procedure, based on the avalanche principle, which we use to prove the ACT.

The avalanche principle

Consider a long chain of matrices $g_0, g_1, \ldots, g_{n-1}$ in $\mathrm{Mat}(m, \mathbb{R})$. The aim of the AP is to relate the expansion $\|g_{n-1} \cdots g_1 g_0\|$ of the product $g_{n-1} \cdots g_1 g_0$ to the product $\|g_{n-1}\| \cdots \|g_1\| \|g_0\|$ of the individual expansions $\|g_j\|$.

Given quantities M_n and N_n, with exponential growth $M_n, N_n \gtrsim e^{na}$ where $a > 0$, we say (in rough terms) that they are *almost asymptotic*, and write $M_n \asymp N_n$, when $e^{-n\epsilon} \leq M_n/N_n \leq e^{n\epsilon}$ for some $0 < \epsilon \ll a$.

In general it is not true that $\|g_{n-1} \cdots g_1 g_0\| \asymp \|g_{n-1}\| \cdots \|g_1\| \|g_0\|$, unless some atypically sharp alignment of the singular directions of the matrices g_j occurs.

Given $g_0, g_1 \in \mathrm{Mat}(m, \mathbb{R})$ non-zero matrices, let us call *expansion rift* of g_0, g_1 the number $\rho(g_0, g_1) := \frac{\|g_1 g_0\|}{\|g_1\| \|g_0\|} \in [0, 1]$. This number measures the break of expansion in the matrix product $g_1 g_0$. More generally, we define the *expansion rift* of a chain of matrices $g_0, g_1, \ldots, g_{n-1}$ in $\mathrm{Mat}(m, \mathbb{R})$ to be the number

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

We call *gap ratio* of a matrix $g \in \mathrm{Mat}(m, \mathbb{R})$ the quotient between its largest and second largest singular values, which can also be expressed as

$$\mathrm{gr}(g) := \frac{\|g\|^2}{\|\wedge_2 g\|},$$

where $\wedge_2 g$ denotes the second exterior power of g. Note that $\mathrm{gr}(g) \geq 1$. We will say that g has a *first singular gap* when $\mathrm{gr}(g) > 1$.

With this terminology, the AP says that given any long chain of matrices $g_0, g_1, \ldots, g_{n-1} \in \mathrm{Mat}(m, \mathbb{R})$, where the gap ratio of each matrix is large and the expansion rift of any pair of consecutive matrices is never too small, then the expansion rift of the product behaves multiplicatively, in the (almost asymptotic) sense that

$$\rho(g_0, g_1, \ldots, g_{n-1}) \asymp \rho(g_0, g_1)\rho(g_1, g_2) \cdots \rho(g_{n-2}, g_{n-1}), \tag{5.3}$$

or, equivalently,

$$\frac{\|g_{n-1} \cdots g_1 g_0\| \|g_1\| \cdots \|g_{n-2}\|}{\|g_1 g_0\| \cdots \|g_{n-1} g_{n-2}\|} \asymp 1.$$

More precisely, the hypotheses of the AP are, for some $\varepsilon, \kappa > 0$

$$\kappa \ll \varepsilon^2, \quad \mathrm{gr}(g_j) \geq \kappa^{-1} \quad \text{and} \quad \rho(g_{j-1}, g_j) \geq \varepsilon \quad \text{for all } j. \tag{5.4}$$

To explain this multiplicative behavior we introduce a geometric quantity, referred to as the *angle* between g and g', that will be compared with the expansion rift $\rho(g, g')$.

Consider on $\mathbb{P}(\mathbb{R}^m)$ the projective distance defined by

$$\delta(\hat{u}, \hat{v}) := \sqrt{1 - \frac{\langle u, v\rangle^2}{\|u\|^2 \|v\|^2}} = \frac{\|u \wedge v\|}{\|u\| \|v\|},$$

where u and v are non-zero vectors representing $\hat{u}, \hat{v} \in \mathbb{P}(\mathbb{R}^m)$. The complementary quantity

$$\alpha(\hat{u}, \hat{v}) := \frac{|\langle u, v\rangle|}{\|u\| \|v\|}$$

will be called (with some abuse of terminology) the *angle* between the projective points $\hat{u}, \hat{v} \in \mathbb{P}(\mathbb{R}^m)$.

Define $v(g) \in \mathbb{R}^m$ to be the *most expanding* unit vector of a matrix $g \in \mathrm{Mat}(m, \mathbb{R})$, in the sense that $\|g v(g)\| = \|g\|$. When g has a first singular gap this vector is unique up to a sign, thus determining a well-defined projective point $\overline{\mathfrak{v}}(g) \in \mathbb{P}(\mathbb{R}^m)$, which represents a singular direction of g. If $\mathrm{gr}(g) > 1$, then $g v(g) = \pm v(g^*)$, where g^* stands for the transpose of g.

Given $g, g' \in \mathrm{Mat}(m, \mathbb{R})$ with first singular gaps, the *angle* between g and g' is the angle between the most expanding directions $\overline{\mathfrak{v}}(g^*)$ and $\overline{\mathfrak{v}}(g')$, i.e.

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

We also set

$$\beta(g, g') := \sqrt{\mathrm{gr}(g)^{-2} \oplus \alpha(g, g')^2 \oplus \mathrm{gr}(g')^{-2}}, \tag{5.5}$$

where the symbol $\oplus$ stands for the operation

$$a \oplus b := a + b - ab.$$

With this operation, $[0, 1]$ is a commutative semigroup.

Proposition 5.1. *Given* $g, g' \in \mathrm{Mat}(m, \mathbb{R})$ *with* $\mathrm{gr}(g), \mathrm{gr}(g') > 1$,

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

Proof. See Proposition 2.18 in [17] or Proposition 2.14 in [14]. □

The following inequality

$$1 \leq \frac{\beta(g, g')}{\alpha(g, g')} \leq \sqrt{1 + \frac{1}{\mathrm{gr}(g)^2 \alpha(g, g')^2} + \frac{1}{\mathrm{gr}(g')^2 \alpha(g, g')^2}} \tag{5.6}$$

shows that if $\mathrm{gr}(g), \mathrm{gr}(g') \geq \kappa^{-1}$ and $\alpha(g, g') \geq \varepsilon$, with $\kappa \ll \varepsilon^2$, then the expansion rift $\rho(g, g')$ is well approximated by the angle $\alpha(g, g')$. In fact we have in this case

$$\alpha(g, g') \leq \rho(g, g') \leq e^{\kappa} \alpha(g, g'). \tag{5.7}$$

Given a chain of matrices $g_0, g_1, \ldots, g_{n-1} \in \mathrm{Mat}(m, \mathbb{R})$, we use the notation $g^{(i)} := g_{i-1} \ldots g_0$, for any $1 < i < n$. From Proposition 5.1, by induction we get

Proposition 5.2. *Given a chain* $g_0, g_1, \ldots, g_{n-1} \in \mathrm{Mat}(m, \mathbb{R})$, *if all matrices* g_i *and* $g^{(i)}$ *have first singular gaps then*

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

From the assumption (5.4) of the AP and (5.7),

$$\alpha(g_{i-1}, g_i) \approx \rho(g_{i-1}, g_i) \approx \beta(g_{i-1}, g_i).$$

Thus, to infer (5.3), because of Proposition 5.2, it is enough to check that $\alpha(g^{(i)}, g_i) \approx \alpha(g_{i-1}, g_i)$, for all i. For this, we use a shadowing argument based on the contracting behavior of the action of the matrices g_j on the projective space $\mathbb{P}(\mathbb{R}^m)$.

Each matrix $g \in \mathrm{Mat}(m, \mathbb{R})$ induces the following partial mapping $\varphi_g : \mathbb{P}(\mathbb{R}^m) \backslash \mathrm{K}_g \to \mathbb{P}(\mathbb{R}^m)$, $\varphi_g \hat{v} := \widehat{g v}$, for all $\hat{v} \notin \mathrm{K}_g$, where K_g denotes the projective subspace determined by the kernel of g.

Proposition 5.3. *Given* $g \in \mathrm{Mat}(m, \mathbb{R})$ *and* $\hat{v} \neq \hat{u}$ *in* $\mathbb{P}(\mathbb{R}^m) \backslash \mathrm{K}_g$,

$$\frac{\delta(\varphi_g(\hat{v}), \varphi_g(\hat{u}))}{\delta(\hat{v}, \hat{u})} \leq \frac{1}{\mathrm{gr}(g) \alpha(\hat{v}, \overline{\mathfrak{v}}(g)) \alpha(\hat{u}, \overline{\mathfrak{v}}(g))}.$$

Proof. Given $\hat{v} \in \mathbb{P}(\mathbb{R}^m)$, let v be one of its unit vector representatives. Since we have $v = \alpha(\hat{v}, \overline{\mathfrak{v}}(g)) v(g) + w$, with w orthogonal to $v(g)$, it follows that $gv = \|g\| \alpha(\hat{v}, \overline{\mathfrak{v}}(g)) v(g) + gw$, with gw orthogonal to $v(g^*)$. Thus $\|gv\| \geq \|g\| \alpha(\hat{v}, \overline{\mathfrak{v}}(g))$.

Similarly, given $\hat{u} \neq \hat{v}$, if u is one of its unit vector representatives, $\|gu\| \geq \|g\|\alpha(\hat{u}, \overline{\mathfrak{v}}(g))$. Therefore,

$$\begin{aligned}
\frac{\delta(\varphi_g(\hat{u}), \varphi_g(\hat{v}))}{\delta(\hat{u}, \hat{v})} &= \frac{\|gu \wedge gv\|}{\|gu\|\|gv\|} \frac{1}{\|u \wedge v\|} \\
&\leq \frac{\|(\wedge_2 g)(u \wedge v)\|}{\|u \wedge v\|} \frac{1}{\|g\|^2 \alpha(\hat{u}, \overline{\mathfrak{v}}(g))\alpha(\hat{v}, \overline{\mathfrak{v}}(g))} \\
&\leq \frac{\|\wedge_2 g\|}{\|g\|^2} \frac{1}{\alpha(\hat{u}, \overline{\mathfrak{v}}(g))\alpha(\hat{v}, \overline{\mathfrak{v}}(g))} \\
&= \frac{1}{\mathrm{gr}(g)\alpha(\hat{u}, \overline{\mathfrak{v}}(g))\alpha(\hat{v}, \overline{\mathfrak{v}}(g))}.
\end{aligned}$$

□

The previous proposition and the assumption (5.4) imply that each map φ_{g_j} is a strong contraction in some neighborhood of $\overline{\mathfrak{v}}(g_j)$ of the following form

$$\Sigma_\varepsilon(g_j) := \{\hat{v} \in \mathbb{P}(\mathbb{R}^m) : \alpha(\hat{v}, \overline{\mathfrak{v}}(g_j)) \geq \varepsilon\}.$$

We have $\mathrm{gr}(g) = \mathrm{gr}(g^*)$ and $\alpha(g, g') = \alpha(g'^*, g^*)$. Hence the assumption (5.4) on the chain $g_0, \ldots, g_{n-1}$ is also satisfied by the transpose chain $g_{n-1}^*, \ldots, g_0^*$. Since for any matrix g, $\varphi_g(\overline{\mathfrak{v}}(g)) = \overline{\mathfrak{v}}(g^*)$ and $\varphi_{g^*}(\overline{\mathfrak{v}}(g^*)) = \overline{\mathfrak{v}}(g)$, we infer from (5.4) that $\overline{\mathfrak{v}}(g_{j-1}^*) \in \Sigma_{\frac{\varepsilon}{2}}(g_j)$ and $\overline{\mathfrak{v}}(g_j) \in \Sigma_{\frac{\varepsilon}{2}}(g_{j-1}^*)$, for all $j = 1, \ldots, n-1$. The cyclic sequence of projective points

$$\overline{\mathfrak{v}}(g_{i-1}^*) \mapsto \cdots \mapsto \overline{\mathfrak{v}}(g_0^*) \mapsto \overline{\mathfrak{v}}(g_0) \mapsto \cdots \mapsto \overline{\mathfrak{v}}(g_{i-1}) \mapsto \overline{\mathfrak{v}}(g_{i-1}^*)$$

is a *pseudo-orbit* for the sequence of projective mappings associated to the chain $g_{i-1}^*, \ldots, g_0^*, g_0, \ldots, g_{i-1}$. Because the projective action of the matrix $g^{(i)}(g^{(i)})^* = (g_{i-1} \cdots g_0)(g_0^* \cdots g_{i-1}^*)$ fixes the point $\overline{\mathfrak{v}}(g^{(i)*})$, the strong contracting behavior of these maps implies that the pseudo-orbit above is *shadowed* by the true orbit of $\overline{\mathfrak{v}}(g^{(i)*})$ under the sequence of maps $\varphi_{g_{i-1}^*}, \ldots, \varphi_{g_0^*}, \varphi_{g_0}, \ldots, \varphi_{g_{i-1}}$. Therefore $\overline{\mathfrak{v}}(g^{(i)*}) \approx \overline{\mathfrak{v}}(g_{i-1}^*)$, which in turn proves $\alpha(g^{(i)}, g_i) \approx \alpha(g_{i-1}, g_i)$.

The sketched argument has other collateral consequences such as the exponential growth of $\mathrm{gr}(g^{(n)})$ and the proximity relations $\overline{\mathfrak{v}}(g^{(n)*}) \approx \overline{\mathfrak{v}}(g_{n-1}^*)$ and $\overline{\mathfrak{v}}(g^{(n)}) \approx \overline{\mathfrak{v}}(g_0)$ (see [14, 17]).

We now state the avalanche principle.

Theorem 5.4. *There exists a constant $c > 0$ such that given $0 < \varepsilon < 1$, $0 < \kappa \leq c\varepsilon^2$ and $g_0, g_1, \ldots, g_{n-1} \in \mathrm{Mat}(m, \mathbb{R})$, if*

(a) $\mathrm{gr}(g_i) \geq \kappa^{-1}$, *for $0 \leq i \leq n-1$, and*
(b) $\alpha(g_{i-1}, g_i) \geq \varepsilon$, *for $1 \leq i \leq n-1$,*

then

$$\left\| \log\|g^{(n)}\| + \sum_{i=1}^{n-2} \log\|g_i\| - \sum_{i=1}^{n-1} \log\|g_i g_{i-1}\| \right\| \lesssim n\frac{\kappa}{\varepsilon^2}.$$

Next corollary is a practical reformulation of the AP's assumptions.

Corollary 5.1. *There exists* $c > 0$ *such that given* $0 < \varepsilon < 1$, $0 < \kappa \leq c\varepsilon^2$ *and* $g_0, g_1, \ldots, g_{n-1} \in \mathrm{Mat}(m, \mathbb{R})$, *if*

$$\text{(gaps)} \quad \mathrm{gr}(g_i) \geq \frac{1}{\kappa} \qquad \textit{for all } 0 \leq i \leq n-1$$

$$\text{(angles)} \quad \frac{\|g_i g_{i-1}\|}{\|g_i\| \|g_{i-1}\|} \geq \varepsilon \quad \textit{for all } 1 \leq i \leq n-1$$

then the same conclusion holds.

The inductive procedure

The proof of the ACT consists of two main steps: the finite scale continuity — which shows that if the number of iterates of the cocycle is fixed, then the corresponding finite scale top Lyapunov exponent depends continuously on the cocycle, and the inductive step — which shows that this behavior does not change significantly as we increase the scale.

Let us assume that the finite scale continuity is already available (this step is not too difficult). This means that for any large enough but fixed scale n_0, if $B \approx A$ then

$$L_1^{(n_0)}(B) = L_1^{(n_0)}(A) + o(1). \tag{5.8}$$

Let $n_1 \gg n_0$ be another much larger scale. The goal is to prove something of the form

$$L_1^{(n_1)}(B) = L_1^{(n_0)}(B) + \mathcal{O}\left(\frac{n_0}{n_1}\right), \tag{5.9}$$

where the implicit constant in $\mathcal{O}(\frac{n_0}{n_1})$ is independent of B, i.e. the estimate is uniform in a neighborhood of the cocycle A.

Applying (5.9) to both B and A and using (5.8), we would get

$$L_1^{(n_1)}(B) = L_1^{(n_1)}(A) + o(1) + \mathcal{O}\left(\frac{n_0}{n_1}\right).$$

Continuing this with another scale $n_2 \gg n_1$, we would get

$$L_1^{(n_2)}(B) = L_1^{(n_2)}(A) + o(1) + \mathcal{O}\left(\frac{n_0}{n_1}\right) + \mathcal{O}\left(\frac{n_1}{n_2}\right),$$

and so on. As the sequence $\{n_k\}$ of scales increases fast, the sum $\mathcal{O}(\frac{n_0}{n_1}) + \mathcal{O}(\frac{n_1}{n_2}) + \cdots$ will be negligible, hence in the limit we would get that

$$L_1(B) = L_1(A) + o(1),$$

thus proving continuity of the top Lyapunov exponent. A modulus of continuity in the limit would follow, amid some loss, from having one at an initial finite scale.

This procedure will not work exactly as described, because (5.9) is not necessarily true as stated. However, a more complex relation, but having a similar flavor will hold true.

An estimate in the spirit of (5.9) relates the space average of the function $\frac{1}{n_1}\log\|B^{(n_1)}(x)\|$ to that of the function $\frac{1}{n_0}\log\|B^{(n_0)}(x)\|$. For these space averages to be comparable, it would be enough if the corresponding functions were pointwise comparable for all but a small set of phases, i.e. that

$$\frac{1}{n_1}\log\|B^{(n_1)}(x)\| = \frac{1}{n_0}\log\|B^{(n_0)}(x)\| + \mathcal{O}\left(\frac{n_0}{n_1}\right), \tag{5.10}$$

for all x outside a set of small measures.

Assume n_1 is a multiple of n_0, so $n_1 = nn_0$, $n \gg 1$. Fix a phase $x \in X$. Then $B^{(n_1)}(x)$, which is a block (i.e. a product) of length n_1, can be divided into n blocks, each of length n_0, which we denote by $g_0, g_1, \ldots, g_{n-1}$. Hence

$$g_i = g_i(x) = B^{(n_0)}(T^{in_0}x) \quad \text{and} \quad g^{(n)} = g_{n-1}\cdots g_1 g_0 = B^{(n_1)}(x).$$

Let us assume for a moment that the AP is applicable to these matrices (i.e. for the given phase x). This will imply that

$$\frac{1}{n}\log\|g^{(n)}\| = -\frac{1}{n}\sum_{i=1}^{n-2}\log\|g_i\| + \frac{1}{n}\sum_{i=1}^{n-1}\log\|g_i g_{i-1}\| + \text{“error”},$$

hence

$$\begin{aligned}\frac{1}{n}\log\|B^{(n_1)}(x)\| &= -\frac{1}{n}\sum_{i=1}^{n-2}\log\|B^{(n_0)}(T^{in_0}x)\| \\ &\quad + \frac{1}{n}\sum_{i=1}^{n-1}\log\|B^{(n_0)}(T^{in_0}x)B^{(n_0)}(T^{(i-1)n_0}x)\| + \text{“error”}.\end{aligned}$$

Since $B^{(n_0)}(T^{n_0}y)B^{(n_0)}(y) = B^{(2n_0)}(y)$, after diving both sides by n_0 and remembering that $n_1 = nn_0$, we get from the above

$$\begin{aligned}\frac{1}{n_1}\log\|B^{(n_1)}(x)\| &= -\frac{1}{n}\sum_{i=1}^{n-2}\frac{1}{n_0}\log\|B^{(n_0)}(T^{in_0}x)\| \\ &\quad + 2\frac{1}{n}\sum_{i=1}^{n-1}\frac{1}{2n_0}\log\|B^{(2n_0)}(T^{(i-1)n_0}x)\| + \text{“error”}.\end{aligned} \tag{5.11}$$

Now let us assume that the AP was in fact applicable for all but a small set of phases. Hence (5.11) will hold for all x outside a set of small measures. Averaging in x, using the uniform L^2-boundedness assumption (and skipping some technicalities), we conclude that

$$L_1^{(n_1)}(B) = -L_1^{(n_0)}(B) + 2L_1^{(2n_0)}(B) + \text{“error”}. \tag{5.12}$$

Assuming moreover that the "error" coming from the AP is in this setting something like $\mathcal{O}(\frac{n_0}{n_1})$, estimate (5.12) is in fact not very unlike (5.9). It relates the finite scale Lyapunov exponent at scale n_1, to its counterparts at scales $n_0, 2n_0$. The scales n_0 and $2n_0$ have the same order of magnitude, hence we may assume that the initial finite scale continuity property applies to both of them. The inductive argument sketched above and based upon (5.9) would work in a similar manner with (5.12) instead, leading to the continuity of the Lyapunov exponents.

For all of this to work, the assumptions "gaps" and "angles" of the AP (say in its formulation from Corollary 5.1) should hold for all but a small set of phases x. This is where the fiber LDT estimates and the information carried at the current scale (i.e. the inductive hypothesis) come into play.

Indeed, by the uniform fiber LDT, if $B \approx A$, then for all x outside a set of small measure, the following hold

$$\frac{1}{n_0}\log\|B^{(n_0)}(x)\| = L_1^{(n_0)}(B) + o(1),$$
$$\frac{1}{n_0}\log\|B^{(n_0)}(T^{n_0}x)\| = L_1^{(n_0)}(B) + o(1),$$
$$\frac{1}{2n_0}\log\|B^{(2n_0)}(x)\| = L_1^{(2n_0)}(B) + o(1).$$

If, moreover,

$$L_1^{(n_0)}(B) - L_1^{(2n_0)}(B) < \eta \ll 1,$$

which follows from the finite scale continuity and the fact that the initial scale n_0 is large enough, then an easy algebraic calculation shows that for all such phases x we have:

$$\frac{\|B^{(2n_0)}(x)\|}{\|B^{(n_0)}(T^{n_0}x)\|\|B^{(n_0)}(x)\|} > e^{-2n_0(\eta+o(1))}.$$

This will imply the "angle" condition in the AP.

The "gap" condition is a more delicate issue. This point is made especially difficult by the fact that we are working with higher dimensional ($m \geq 2$), and possibly not (everywhere) invertible cocycles.

The treatment of this issue is what makes our inductive procedure differ most from the one in [23]. Moreover, unlike quasi-periodic models, fiber LDTs for random models are only available in the presence of a gap in the Lyapunov spectrum. This approach is also the main reason that random models can be treated in the same inductive scheme and with fewer restrictions than in [37].

The complete details of this argument can be found in [13, 17]. We only mention here that a key ingredient of the argument is a type of *uniform* upper semicontinuity property of the top Lyapunov exponent. Such a property was previously established in [19, 31] for *uniquely ergodic systems*. Bernoulli or Markov systems are *not* uniquely ergodic. We obtain a weaker statement, sufficient for our needs, that applies to these models as well. This is where the compatibility condition and the base large deviation are used, as substitutes for unique ergodicity. Here is the statement of the "nearly" upper semicontinuity property.

Proposition 5.4. *Let $A \in \mathcal{C}_m$ be a measurable cocycle such that Ξ and A are compatible and every observable $\xi \in \Xi$ satisfies a base LDT. Assume that $L_1(A) > -\infty$ and that A is L^2-bounded. Then for every $\epsilon > 0$, there are $\delta = \delta(A, \epsilon) > 0$, $n_0 = n_0(A, \epsilon) \in \mathbb{N}$ and $\underline{\iota} = \underline{\iota}(A, \epsilon)$ corresponding to an LDT parameter in $\mathcal{P}$, such that if $B \in \mathcal{C}_m$ with $d(B, A) < \delta$, and if $n \geq n_0$, then the upper bound*

$$\frac{1}{n} \log\|B^{(n)}(x)\| \leq L_1(A) + \epsilon, \tag{5.13}$$

holds for all x outside of a set of measure $< \iota_n$.

6. Further Work

As mentioned before, LDT estimates for Schrödinger cocycles have also been used to study integer lattice, one-dimensional, quasi-periodic Schrödinger operators, e.g., to derive lower bounds on Lyapunov exponents, spectral properties of the operator, continuity properties of the integrated density of states, estimates on the measure of the spectrum etc. The availability of such estimates for more general cocycles, discussed in this survey, makes it likely that similar kinds of problems can be approached for more general types of discrete quasi-periodic operators, such as band-lattice Schrödinger operators (which approximate higher dimensional lattice Schrödinger operators) or Jacobi type operators.

We have work in progress dealing with *identically singular* cocycles $A(x)$ (i.e. $\det[A(x)] \equiv 0$) in the multivariable case (the one variable case has been treated in [2]). The completion of this project will have other immediate consequences, such as criteria for positivity of Lyapunov exponents and simplicity of the Lyapunov spectrum for higher dimensional Schrödinger and Jacobi type cocycles. No such results are currently available in the multivariable case.

Regarding random cocycles, an interesting open problem is proving (uniform) fiber LDT estimates for *reducible* cocycles (our current work requires irreducibility). Moreover, another difficult and interesting open problem is establishing *quantitative* continuity properties of the Lyapunov exponents in the vicinity of a reducible cocycle, either through our scheme involving LDTs or through other means.

Finally, it would be interesting to see if our scheme for proving continuity of Lyapunov exponents is applicable to cocycles over different kinds of base dynamics. As indicated earlier, base LDT estimates are already available, so the challenge is to prove uniform fiber LDTs for such models.

Acknowledgments

The first author was supported by Fundação para a Ciência e a Tecnologia, UID/MAT/04561/2013.

The second author was supported by the Norwegian Research Council project no. 213638, "Discrete Models in Mathematical Analysis".

The second author is grateful to Jaqueline Siqueira for the help with the preparation of the manuscript.

References

1. L. Arnold, *Random Dynamical Systems*, Springer Monographs in Mathematics (Springer-Verlag, 1998).
2. A. Ávila, S. Jitomirskaya and C. Sadel, Complex one-frequency cocycles, *J. Eur. Math. Soc.* **16** (2014) 1915–1935.
3. C. Bocker-Neto and M. Viana, Continuity of Lyapunov exponents for random 2d matrices, preprint (2010), 1–38.
4. P. Bougerol, Théorèmes limite pour les systèmes linéaires à coefficients markoviens, *Probab. Theory Relat. Fields* **78** (1988) 193–221.
5. P. Bougerol and J. Lacroix, *Products of Random Matrices with Applications to Schrödinger Operators*, Progress in Probability and Statistics, Vol. 8 (Birkhäuser, 1985).
6. J. Bourgain, *Green's Function Estimates for Lattice Schrödinger Operators and Applications*, Annals of Mathematics Studies, Vol. 158 (Princeton Univ. Press, 2005).
7. ——, 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** (2005) 313–355.
8. J. Bourgain and M. Goldstein, On nonperturbative localization with quasi-periodic potential, *Ann. Math. (2)* **152** (2000) 835–879.
9. J. Bourgain and S. Jitomirskaya, Continuity of the Lyapunov exponent for quasiperiodic operators with analytic potential, *J. Statist. Phys.* **108** (2002) 1203–1218, Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays.
10. J. Bourgain, M. Goldstein and W. Schlag, Anderson localization for Schrödinger operators on $\mathbb{Z}$ with potentials given by the skew-shift, *Commun. Math. Phys.* **220** (2001) 583–621.
11. J. L. Doob, *Stochastic Processes*, Wiley Classics Library (John Wiley & Sons, 1990), Reprint of the 1953 original (Wiley-Interscience).
12. P. Duarte and S. Klein, Continuity of the Lyapunov exponents for quasiperiodic cocycles, *Commun. Math. Phys.* **332** (2014) 1113–1166.
13. ——, An abstract continuity theorem of the Lyapunov exponents, preprint (2015).
14. ——, The avalanche principle and other estimates on Grassmann manifolds, preprint (2015).
15. ——, Large deviations for quasi-periodic cocycles with singularities, preprint (2015).
16. ——, Large deviations type estimates for random cocycles, preprint (2015).
17. ——, *Lyapunov Exponents of Linear Cocycles*; *Continuity via Large Deviations*, Atlantis Studies in Dynamical Systems, Vol. 3 (Atlantis Press, 2016).
18. J. Duoandikoetxea, *Fourier Analysis*, Graduate Studies in Mathematics, Vol. 29 (Amer. Math. Soc., 2001), translated and revised from the 1995 Spanish original by David Cruz-Uribe.
19. A. Furman, On the multiplicative ergodic theorem for uniquely ergodic systems, *Ann. Inst. H. Poincaré Probab. Statist.* **33** (1997) 797–815.
20. H. Furstenberg and H. Kesten, Products of random matrices, *Ann. Math. Statist.* **31** (1960) 457–469.

21. H. Furstenberg and Y. Kifer, Random matrix products and measures on projective spaces, *Israel J. Math.* **46** (1983) 12–32.
22. H. Furstenberg, Noncommuting random products, *Trans. Amer. Math. Soc.* **108** (1963) 377–428.
23. M. Goldstein and W. Schlag, Hölder continuity of the integrated density of states for quasi-periodic Schrödinger equations and averages of shifts of subharmonic functions, *Ann. Math. (2)* **154** (2001) 155–203.
24. ———, On Schrödinger operators with dynamically defined potentials, *Mosc. Math. J.* **5** (2005) 577–612; 742–743.
25. ———, Fine properties of the integrated density of states and a quantitative separation property of the Dirichlet eigenvalues, *Geom. Funct. Anal.* **18** (2008) 755–869.
26. Y. Guivarc'h and A. Raugi, Frontière de Furstenberg, propriétés de contraction et théorèmes de convergence, *Z. Wahrsch. Verw. Gebiete* **69** (1985) 187–242.
27. W. K. Hayman and P. B. Kennedy, *Subharmonic Functions*, Vol. I (Academic Press, 1976).
28. H. Hennion and L. Hervé, *Limit Theorems for Markov Chains and Stochastic Properties of Dynamical Systems by Quasi-Compactness*, Lecture Notes in Mathematics, Vol. 1766 (Springer-Verlag, 2001).
29. S. Jitomirskaya and C. A. Marx, Continuity of the Lyapunov exponent for analytic quasi-periodic cocycles with singularities, *J. Fixed Point Theory Appl.* **10** (2011) 129–146.
30. ———, Analytic quasi-perodic cocycles with singularities and the Lyapunov exponent of extended Harper's model, *Commun. Math. Phys.* **316** (2012) 237–267.
31. S. Jitomirskaya and R. Mavi, Continuity of the measure of the spectrum for quasiperiodic Schrödinger operators with rough potentials, *Commun. Math. Phys.* **325** (2014) 585–601.
32. Y. Kifer, Large deviations in dynamical systems and stochastic processes, *Trans. Amer. Math. Soc.* **321** (1990) 505–524.
33. S. Klein, Anderson localization for the discrete one-dimensional quasi-periodic Schrödinger operator with potential defined by a Gevrey-class function, *J. Funct. Anal.* **218** (2005) 255–292.
34. ———, Localization for quasiperiodic Schrödinger operators with multivariable Gevrey potential functions, *J. Spectr. Theory* **4** (2014) 1–53.
35. S. G. Krein and Yu. I. Petunin, Scales of banach spaces, *Russ. Math. Surv.* **21** (1966) 85.
36. É. Le Page, Théorèmes limites pour les produits de matrices aléatoires, in *Probability Measures on Groups* (Oberwolfach, 1981), Lecture Notes in Math., Vol. 928 (Springer, 1982), pp. 258–303.
37. ———, Régularité du plus grand exposant caractéristique des produits de matrices aléatoires indépendantes et applications, *Ann. l'inst. H. Poincaré (B) Probab. Statist.* **25** (1989) 109–142.
38. A. O. Lopes, Entropy and large deviation, *Nonlinearity* **3** (1990) 527–546.
39. E. Malheiro and M. Viana, Lyapunov exponents of linear cocycles over Markov shifts, preprint (2014).
40. H. L. Montgomery, *Ten Lectures on the Interface Between Analytic Number Theory and Harmonic Analysis*, CBMS Regional Conference Series in Mathematics, Vol. 84 (Amer. Math. Soc., 1994).
41. C. Muscalu and W. Schlag, *Classical and Multilinear Harmonic Analysis*, Vol. I, Cambridge Studies in Advanced Mathematics, Vol. 137 (Cambridge Univ. Press, 2013).

42. S. V. Nagaev, Some limit theorems for stationary Markov chains, *Teor. Veroyatnost. i Primenen.* **2** (1957) 389–416.
43. F. Rassoul-Agha and T. Seppäläinen, A course on large deviations with an introduction to Gibbs measures, Graduate Studies in Mathematics, Vol. 162 (Amer. Math. Soc., 2015).
44. A. Raugi and J. Rosenberg, *Fonctions Harmoniques et Théorèmes Limites pour les Marches Aléatoires sur les Groupes*, Bulletin de la Société mathématique de France, Mémoire, no. 54 (Société mathématique de France, 1977).
45. L. Rey-Bellet and L.-S. Young, Large deviations in non-uniformly hyperbolic dynamical systems, *Ergodic Theory Dynam. Systems* **28** (2008) 587–612.
46. W. Schlag, *On Discrete Schrödinger Operators with Stochastic Potentials*, XIV Int. Congress on Mathematical Physics (World Scientific, 2005), pp. 206–215.
47. ———, Regularity and convergence rates for the Lyapunov exponents of linear cocycles, *J. Mod. Dyn.* **7** (2013) 619–637.
48. T. Tao, *Topics in Random Matrix Theory*, Graduate Studies in Mathematics, Vol. 132 (Amer. Math. Soc., 2012).
49. V. N. Tutubalin, Limit theorems for a product of random matrices, *Teor. Verojatnost. i Primenen.* **10** (1965) 19–32.
50. S. R. S. Varadhan, *Large Deviations and Applications*, École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87, Lecture Notes in Math., Vol. 1362 (Springer, 1988), pp. 1–49.
51. M. Viana, *Lectures on Lyapunov Exponents* (Cambridge Univ. Press, 2014).
52. J. You and S. Zhang, Hölder continuity of the Lyapunov exponent for analytic quasiperiodic Schrödinger cocycle with weak Liouville frequency, *Ergodic Theory Dynam. Systems* **34** (2014) 1395–1408.
53. L.-S. Young, Large deviations in dynamical systems, *Trans. Amer. Math. Soc.* **318** (1990) 525–543.