

# CONTINUITY OF LYAPUNOV EXPONENTS FOR RANDOM 2D MATRICES

CARLOS BOCKER-NETO AND MARCELO VIANA

ABSTRACT. The Lyapunov exponents of locally constant  $GL(2, \mathbb{C})$ -cocycles over Bernoulli shifts depend continuously on the cocycle and on the invariant probability. The Oseledets decomposition also depends continuously on the cocycle, in measure.

## CONTENTS

1. Introduction	2
2. Continuity of Lyapunov exponents	2
2.1. Linear cocycles	2
2.2. Continuity problem	3
2.3. Continuity theorem	4
2.4. Previous results	5
3. Proof of Theorem C	6
3.1. Reducing to the diagonal case	7
3.2. Handling the diagonal case	13
4. Proof of the Key Proposition	13
4.1. Preliminary observations	14
4.2. Auxiliary statements	15
4.3. Proof of Proposition 4.2	18
5. Main estimates	22
5.1. Mass away from the vertical	22
5.2. Proof of Proposition 4.8	25
5.3. Mass close to the vertical	27
5.4. Proof of Proposition 4.10	29
6. Consequences of Theorem C	29
6.1. Proof of Theorem B	29
6.2. Proof of Theorem D	31
7. Concluding remarks	33
7.1. An example of discontinuity	33
7.2. Open problems	36
References	37

---

C.B.-N. was supported by a CNPq and FAPERJ doctoral scholarship. M.V. is partially supported by CNPq, FAPERJ, and PRONEX-Dynamical Systems.

## 1. INTRODUCTION

Let  $A_1, \dots, A_m$  be invertible 2-by-2 matrices and  $p_1, \dots, p_m$  be (strictly) positive numbers with  $p_1 + \dots + p_m = 1$ . Consider

$$L^n = L_{n-1} \cdots L_1 L_0, \quad n \geq 1,$$

where the  $L_j$  are independent random variables with identical probability distributions, given by

$$\text{probability}(\{L_j = A_i\}) = p_i \quad \text{for all } j \geq 0 \text{ and } i = 1, \dots, m.$$

It is a classical fact, going back to Furstenberg, Kesten [14], that there exist numbers  $\lambda_+$  and  $\lambda_-$  such that

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

almost surely. The results in this paper imply that these *extremal Lyapunov exponents* always vary continuously with the choice of the matrices and the probability weights:

**Theorem A.** *The extremal Lyapunov exponents  $\lambda_+$  and  $\lambda_-$  depend continuously on  $(A_1, \dots, A_m, p_1, \dots, p_m)$  at all points.*

This conclusion holds in much more generality. Indeed, we may take the probability distribution of the random variables  $L_j$  to be any probability measure  $\nu$  on  $\text{GL}(2, \mathbb{C})$  with compact support. Let  $\lambda_+(\nu)$  and  $\lambda_-(\nu)$ , respectively, denote the values of the (almost certain) limits in (1). Then we have:

**Theorem B.** *For every  $\varepsilon > 0$  there exists  $\delta > 0$  and a weak\* neighborhood  $V$  of  $\nu$  in the space of probability measures on  $\text{GL}(2, \mathbb{C})$  such that  $|\lambda_{\pm}(\nu) - \lambda_{\pm}(\nu')| < \varepsilon$  for every probability measure  $\nu' \in V$  whose support is contained in the  $\delta$ -neighborhood of the support of  $\nu$ .*

The situation in Theorem A corresponds to the special case when the measures have finite supports:

$$\nu = p_1 \delta_{A_1} + \dots + p_m \delta_{A_m} \quad \text{and} \quad \nu' = p'_1 \delta_{A'_1} + \dots + p'_m \delta_{A'_m}.$$

Clearly, the support of  $\nu'$  is Hausdorff close to the support of  $\nu$  if  $A'_i$  is close to  $A_i$ ,  $p_i$  for all  $i$ . In this regard, recall that we assume that all  $p_i > 0$ : the conclusion of Theorem A may fail if this condition is removed, as we will recall in Remark 7.5.

## 2. CONTINUITY OF LYAPUNOV EXPONENTS

In this section we put the previous results in a broader context and give a convenient translation of Theorem B to the theory of linear cocycles.

**2.1. Linear cocycles.** Let  $\pi : \mathcal{V} \rightarrow M$  be a finite-dimensional (real or complex) vector bundle and  $F : \mathcal{V} \rightarrow \mathcal{V}$  be a *linear cocycle* over some measurable transformation  $f : M \rightarrow M$ . By this we mean that  $\pi \circ F = f \circ \pi$  and the actions  $F_x : \mathcal{V}_x \rightarrow \mathcal{V}_{f(x)}$  on the fibers are linear isomorphisms. Take  $\mathcal{V}$  to be endowed with some measurable Riemannian metric, that is, an Hermitian product on each fiber depending measurably on the base point. Let  $\mu$  be an  $f$ -invariant probability measure on  $M$  such that

$$\log \|(F_x)^{\pm 1}\| \in L^1(\mu).$$

It follows from the sub-additive ergodic theorem (Kingman [24]) that the numbers

$$\lambda_+(F, x) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \|F_x^n\| \quad \text{and} \quad \lambda_-(F, x) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \|(F_x^n)^{-1}\|^{-1}$$

are well-defined  $\mu$ -almost everywhere.

The theorem of Oseledets [30] provides a more detailed statement. Namely, at  $\mu$ -almost every point  $x \in M$ , there exist numbers

$$\hat{\lambda}_1(F, x) > \cdots > \hat{\lambda}_{k(x)}(F, x)$$

and a filtration

$$\mathcal{V}_x = V_x^1 > V_x^2 > \cdots > V_x^{k(x)} > \{0\} = V_x^{k(x)+1} \quad (2)$$

such that  $F_x(V_x^j) = V_{f(x)}^j$  and

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \|F_x^n(v)\| = \hat{\lambda}_j(F, x) \quad \text{for all } v \in V_x^j \setminus V_x^{j+1}.$$

When  $f$  is invertible one can say more: there exists a splitting

$$\mathcal{V}_x = E_x^1 \oplus E_x^2 \oplus \cdots \oplus E_x^{k(x)} \quad (3)$$

such that  $F_x(E_x^j) = E_{f(x)}^j$  and

$$\lim_{n \rightarrow \pm\infty} \frac{1}{n} \log \|F_x^n(v)\| = \hat{\lambda}_j(F, x) \quad \text{for all } v \in E_x^j \setminus \{0\}.$$

The number  $k(x) \geq 1$  and the *Lyapunov exponents*  $\hat{\lambda}_j(F, \cdot)$  are measurable functions of the point  $x$ , with

$$\hat{\lambda}_1(F, x) = \lambda_+(F, x) \quad \text{and} \quad \hat{\lambda}_{k(x)}(F, x) = \lambda_-(F, x),$$

and they are constant on the orbits of  $f$ . In particular, they are constant  $\mu$ -almost everywhere if  $\mu$  is ergodic.

**2.2. Continuity problem.** Next, let  $\lambda_1(F, x) \geq \cdots \geq \lambda_d(F, x)$  be the list of all Lyapunov exponents, where each is counted according to its multiplicity  $m_j(x) = \dim V_x^j - \dim V_x^{j+1}$  ( $= \dim E_x^j$  in the invertible case). Of course,  $d = \dim \mathcal{V}$ . The *average Lyapunov exponents* of  $F$  are defined by

$$\lambda_i(F, \mu) = \int \lambda_i(F, \cdot) d\mu, \quad \text{for } i = 1, \dots, d.$$

The results in this paper are motivated by the following basic question:

**Problem 2.1.** What are the continuity points of

$$(F, \mu) \mapsto (\lambda_1(F, \mu), \dots, \lambda_d(F, \mu)) ?$$

It is well known that the sum of the  $k$  largest Lyapunov exponents

$$(F, \mu) \mapsto \lambda_1(F, \mu) + \cdots + \lambda_k(F, \mu) \quad (4)$$

(any  $1 \leq k < d$ ) is upper semi-continuous, relative to the  $L^\infty$ -norm in the space of cocycles and the pointwise topology in the space of probabilities (the smallest topology that makes  $\mu \mapsto \int \psi d\mu$  continuous for every bounded measurable function  $\psi$ ). Indeed, this is an easy consequence of the identity

$$\lambda_1(F, \mu) + \cdots + \lambda_k(F, \mu) = \inf_{n \geq 1} \frac{1}{n} \int \log \|\Lambda^k(F_x^n)\| d\mu(x) \quad (5)$$

where  $\Lambda^k$  denotes the  $k$ th exterior power. Similarly, the sum of the  $k$  smallest Lyapunov exponents is always lower semi-continuous. However, Lyapunov exponents are, usually, *discontinuous* functions of the data. A number of results, both positive and negative, will be recalled in a while.

**2.3. Continuity theorem.** Let  $\mathcal{X}$  be a polish space, that is, a separable completely metrizable topological space. Let  $p$  be a probability measure on  $\mathcal{X}$  and  $A : \mathcal{X} \rightarrow \text{GL}(2, \mathbb{C})$  be a measurable function such that

$$\log \|A^{\pm 1}\| \quad \text{are bounded.} \quad (6)$$

Let  $f : M \rightarrow M$  be the shift map on  $M$  and  $\mu = p^{\mathbb{Z}}$ . Consider the linear cocycle

$$F : M \times \mathbb{C}^2 \rightarrow M \times \mathbb{C}^2, \quad F(\mathbf{x}, v) = (f(\mathbf{x}), A_{x_0}(v)),$$

where  $x_0 \in \mathcal{X}$  denotes the zeroth coordinate of  $\mathbf{x} \in M$ . In the spaces of cocycles and probability measures on  $\mathcal{X}$  we consider the distances defined by, respectively,

$$d(A, B) = \sup_{x \in \mathcal{X}} \|A_x - B_x\| \quad d(p, q) = \sup_{|\phi| \leq 1} \left| \int \phi d(p - q) \right| \quad (7)$$

where the second sup is over all measurable functions  $\phi : \mathcal{X} \rightarrow \mathbb{R}$  with  $\sup |\phi| \leq 1$ . In the space of pairs  $(A, p)$  we consider the topology determined by the bases of neighborhoods

$$V(A, p, \gamma, \mathcal{Z}) = \{(B, q) : d(A, B) < \gamma, q(\mathcal{Z}) = 1, d(p, q) < \gamma\} \quad (8)$$

where  $\gamma > 0$  and  $\mathcal{Z}$  is any measurable subset of  $\mathcal{X}$  with  $p(\mathcal{Z}) = 1$ .

**Theorem C.** *The extremal Lyapunov exponents  $\lambda_{\pm}(A, p) = \lambda_{\pm}(F, \mu)$  depend continuously on  $(A, p)$  at all points.*

We prove Theorem C in Sections 3 and 4, and we deduce Theorem B from it in Section 6. Theorem C can also be deduced from Theorem B: if  $d(A, B)$  and  $d(p, q)$  are small then  $\nu' = B_*q$  is close to  $\nu = A_*p$  in the weak\* topology, and the support of  $\nu'$  is contained in a small neighborhood of the support of  $\nu$ ; moreover,  $\lambda_{\pm}(A, p) = \lambda_{\pm}(\nu)$  and  $\lambda_{\pm}(B, q) = \lambda_{\pm}(\nu')$ . In this way one even gets a more general version of Theorem C, where  $\mathcal{X}$  can be any measurable space.

Our arguments also show that the Oseledets decomposition depends continuously on the cocycle in measure. Given  $B : \mathcal{X} \rightarrow \text{GL}(2, \mathbb{C})$ , let  $E_{B, \mathbf{x}}^s$  and  $E_{B, \mathbf{x}}^u$  be the Oseledets subspaces of the corresponding cocycle at a point  $\mathbf{x} \in M$  (when they exist).

**Theorem D.** *Suppose  $\lambda_-(A, p) < \lambda_+(A, p)$ . For any sequence  $A^k : \mathcal{X} \rightarrow \text{GL}(2, \mathbb{C})$  such that  $d(A^k, A) \rightarrow 0$ , and for any  $\varepsilon > 0$ , we have*

$$\mu(\{x \in M : \angle(E_{A, x}^u, E_{A^k, x}^u) < \varepsilon \text{ and } \angle(E_{A, x}^s, E_{A^k, x}^s) < \varepsilon\}) \rightarrow 1.$$

A few words are in order on our choice of the topology (8). As we are going to see, the proof of Theorem C splits into two cases, depending on whether the cocycle is almost irreducible (Section 3.1) or diagonal (Section 3.2). In the irreducible case, continuity of the Lyapunov exponents was known before ([15, 18], see also [3]) and only requires the weak\* topology. In a nutshell, this is because in the irreducible case

$$\lambda_+(A, p) = \int \log \frac{\|A(\mathbf{x})(v)\|}{\|v\|} d\mu(\mathbf{x}) d\eta(v) \quad (9)$$

for *every* stationary measure  $\eta$  (Furstenberg's formula); then one only has to note that the set of stationary measures varies semi-continuously with the data. The main point in the proof Theorem C is to handle the diagonal case, where (9) breaks down. That is where we need the full strength of (8).

Restricted to the space of pairs  $(A, p)$  where  $A$  is continuous (and bounded), it suffices to consider the neater bases of neighborhoods

$$V(A, p, \varepsilon) = \{(B, q) : d(A, B) < \varepsilon, \text{supp } q \subset \text{supp } p, d(p, q) < \varepsilon\}. \quad (10)$$

However, this will not be used in the present paper.

**2.4. Previous results.** The problem of dependence of Lyapunov exponents on the linear cocycle or the base dynamics has been addressed by several authors. In a pioneer work, Ruelle [36] proved real-analytic dependence of the largest exponent on the cocycle, for linear cocycles admitting an invariant convex cone field. Short afterwards, Furstenberg, Kifer [15, 22] and Hennion [18] proved continuity of the largest exponent of i.i.d. random matrices, under a condition of almost irreducibility. Some reducible cases were treated by Kifer and Slud [22, 23], who also observed that discontinuities may occur when the probability vector degenerates ([22], cf. Remark 7.5 below).

For i.i.d. random matrices satisfying strong irreducibility and the contraction property, Le Page [31, 32] proved local Hölder continuous, and even smooth, dependence of the largest exponent on the cocycle; the assumptions ensure that the largest exponent is simple (multiplicity 1), by work of Guivarc'h, Raugi [17] and Gol'dsheid, Margulis [16]. For i.i.d. random matrices over Bernoulli and Markov shifts, Peres [33] showed that simple exponents are locally real-analytic functions of the transition data.

A construction of Halperin quoted by Simon, Taylor [37] shows that for every  $\alpha > 0$  one can find *random Schrödinger cocycles*

$$\begin{pmatrix} E - V_n & -1 \\ 1 & 0 \end{pmatrix}$$

(the  $V_n$  are i.i.d. random variables) near which the exponents fail to be  $\alpha$ -Hölder continuous. Thus, the previously mentioned results of Le Page can not be improved. Johnson [20] found examples of discontinuous dependence of the exponent on the energy  $E$ , for Schrödinger cocycles over quasi-periodic flows. Recently, Bourgain, Jitomirskaya [11, 12] proved continuous dependence of the exponents on the energy  $E$ , for one-dimensional *quasi-periodic* Schrödinger cocycles:  $V_n = V(f^n(\theta))$  where  $V : S^1 \rightarrow \mathbb{R}$  is real-analytic and  $f$  is an irrational circle rotation.

Going back to linear cocycles, the answer to the continuity problem is bound to depend on the class of cocycles under consideration, including its topology. Knill [25, 26] considered  $L^\infty$  cocycles with values in  $\text{SL}(2, \mathbb{R})$  and proved that, as long as the base dynamics is aperiodic, discontinuities always exist: the set of cocycles with non-zero exponents is never open. This was refined to the continuous case by Bochi [5, 6]: an  $\text{SL}(2, \mathbb{R})$ -cocycle is a continuity point in the  $C^0$  topology if and only if it is uniformly hyperbolic or else the exponents vanish. This statement was inspired by Mañé's surprising announcement in [29]. Indeed, and most strikingly, the theorem of Mañé-Bochi [6, 29] remains true restricted to the subset of  $C^0$  derivative cocycles, that is, of the form  $F = Df$  for some  $C^1$  area preserving diffeomorphism  $f$ . Moreover, this has been extended to cocycles and diffeomorphisms

in arbitrary dimension, by Bochi, Viana [7, 8]. Let us also note that linear cocycles whose exponents are all equal form an  $L^p$ -residual subset, for any  $p \in [1, \infty)$ , by Arnold, Cong [2], Arbieto, Bochi [1]. Consequently, they are precisely the continuity points for the Lyapunov exponents relative to the  $L^p$  topology.

These results show that discontinuity of Lyapunov exponents is quite common among cocycles with low regularity. Locally constant cocycles, as we deal with here, sit at the opposite end of the regularity spectrum, and the results in the present paper show that in this context continuity does hold at every point. For cocycles with intermediate regularities the continuity problem is very much open. However, our construction in Section 7.1 shows that for any  $r \in (0, \infty)$  there exist locally constant cocycles over Bernoulli shifts that are points of discontinuity for the Lyapunov exponents in the space of all  $r$ -Hölder cocycles. We will return to this topic in the final section.

Recently, Avila, Viana [3] studied the continuity of the Lyapunov exponents in the very broad context of *smooth* cocycles. The continuity criterium in [3, Section 5] was the starting point for the proof of our Theorem C.

This paper is organized as follows. In Section 3 we reduce Theorem C to a key result on stationary measures of nearby cocycles. The latter is proved in Sections 4 and 5. In Section 6 we deduce Theorems B and D. Finally, in Section 7 we describe an example of discontinuity of Lyapunov exponents for Hölder cocycles, and we close with a short list of open problems and conjectures.

*Acknowledgements.* We are grateful to Artur Avila, Jairo Bochi, and Jiagang Yang for several useful conversations. Lemma 6.1 is due to Artur Avila.

### 3. PROOF OF THEOREM C

We start with a simple observation. Let  $\mathcal{P}(\mathcal{X})$  be the space of probability measures on  $\mathcal{X}$  and let  $\mathcal{G}(\mathcal{X})$  and  $\mathcal{S}(\mathcal{X})$  denote the spaces of bounded measurable functions from  $\mathcal{X}$  to  $\text{GL}(2, \mathbb{C})$  and  $\text{SL}(2, \mathbb{C})$ , respectively. Given any  $A \in \mathcal{G}(\mathcal{X})$  let  $B \in \mathcal{S}(\mathcal{X})$  and  $c : \mathcal{X} \rightarrow \mathbb{C}$  be such that  $A_x = c_x B_x$  for every  $x \in \mathcal{X}$ . Although  $c_x = (\det A_x)^{1/2}$  and  $B_x$  are determined up to sign only, choices can be made consistently in a neighborhood, so that  $B$  and  $c$  depend continuously on  $A$ . It is also easy to see that the Lyapunov exponents are related by

$$\lambda_{\pm}(A, p) = \lambda_{\pm}(B, p) + \int \log |c_x| dp(x)$$

Thus, since the last term depends continuously on  $(A, p)$  relative to the topology defined by (8), continuity of the Lyapunov exponents on  $\mathcal{S}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$  yields continuity on the whole  $\mathcal{G}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$ . So, we may suppose from the start that  $A \in \mathcal{S}(\mathcal{X})$ . Observe also that in this case one has

$$\lambda_+(A, p) + \lambda_-(A, p) = 0.$$

From here on the proof has two main steps. First, we reduce the problem to the case when the matrices are simultaneously diagonalizable:

**Proposition 3.1.** *If  $(A, p) \in \mathcal{S}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$  is a point of discontinuity for  $\lambda_+$  then there is  $P \in \text{SL}(2, \mathbb{C})$  and  $\theta : \mathcal{X} \rightarrow \mathbb{C} \setminus \{0\}$  such that*

$$PA_x P^{-1} = \begin{pmatrix} \theta_x & 0 \\ 0 & \theta_x^{-1} \end{pmatrix}$$

for all  $x \in \mathcal{Z}$ , where  $\mathcal{Z} \subset \mathcal{X}$  is a full  $p$ -measure set. In particular,  $A_x A_y = A_y A_x$  for all  $x, y \in \mathcal{Z}$ .

Then we rule out the diagonal case as well:

**Proposition 3.2.** *Let  $(A, p) \in \mathcal{S}(X) \times \mathcal{P}(\mathcal{X})$  be such that  $A$  is as in the conclusion of Proposition 3.1. Then  $(A, p)$  is a point of continuity for  $\lambda_+$ .*

The proofs of these two propositions are given in the next couple of sections. In view of the previous observations, they contain the proof of Theorem C.

**3.1. Reducing to the diagonal case.** The proof of Proposition 3.1 is a simplified version of ideas of Avila, Viana [3], partly inspired by Bonatti, Gomez-Mont, Viana [10]. For the sake of completeness, and also because our setting is not strictly contained in [3], we give the full arguments. The definitions and preliminary results apply to functions  $A$  with values in  $\mathrm{GL}(d, \mathbb{C})$ , for any  $d \geq 2$ .

The *local stable set*  $W_{loc}^s(\mathbf{x})$  of  $\mathbf{x} \in M$  is the set of all  $\mathbf{y} = (y_n)_{n \in \mathbb{Z}}$  such that  $x_n = y_n$  for all  $n \geq 0$ . The *local unstable set*  $W_{loc}^u(\mathbf{x})$  is defined similarly, considering  $n < 0$  instead. The *projective cocycle* associated to  $A : \mathcal{X} \rightarrow \mathrm{GL}(d, \mathbb{C})$  is defined by

$$F_A : M \times \mathbb{P}(\mathbb{C}^d) \rightarrow M \times \mathbb{P}(\mathbb{C}^d), \quad (\mathbf{x}, [v]) \mapsto (f(\mathbf{x}), [A(\mathbf{x})v])$$

where  $A(\mathbf{x}) = A_{x_0}$  for every  $\mathbf{x} \in M$ .

**3.1.1. Invariant  $u$ -states.** Let  $\mathcal{M}(p)$  denote the set of probability measures in  $M \times \mathbb{P}(\mathbb{C}^d)$  that project down to  $\mu$ . A *disintegration* of  $m \in \mathcal{M}(p)$  is a measurable function assigning to each point  $\mathbf{x} \in M$  a probability  $m_{\mathbf{x}}$  with  $m_{\mathbf{x}}(\{\mathbf{x}\} \times \mathbb{P}(\mathbb{C}^d)) = 1$  and such that

$$m(E) = \int m_{\mathbf{x}}(E) d\mu(\mathbf{x}), \quad \text{for every measurable } E \subset M \times \mathbb{P}(\mathbb{C}^d).$$

A disintegration always exists in this setting; moreover, it is essentially unique. See Rokhlin [34] and [9, Appendix C.6].

A probability  $m \in \mathcal{M}(p)$  is a  *$u$ -state* if some disintegration  $\mathbf{x} \mapsto m_{\mathbf{x}}$  is constant on every local unstable set, restricted to a full  $\mu$ -measure subset of  $M$ . Then the same is true for every disintegration, by essential uniqueness; moreover, one can choose the disintegration so that it is constant on local unstable sets on the whole  $M$ . If  $m$  is an invariant probability then we say that  $m$  is an *invariant  $u$ -state*. The definition of *invariant  $s$ -states* is analogous, considering local stable sets instead, and the same observations apply.

An  *$su$ -state* is a probability which is both a  $u$ -state and an  $s$ -state.

**Lemma 3.3.** *A probability  $m \in \mathcal{M}(p)$  is an invariant  $su$ -state if and only if  $m = \mu \times \eta$  for some probability measure  $\eta$  on  $\mathbb{P}(\mathbb{C}^d)$  invariant under the action of  $A_x$  for  $p$ -almost every  $x \in \mathcal{X}$ .*

*Proof.* The “if” part is not used in this paper, so we leave the proof to the reader. To prove the “only if” part notice that, by assumption,  $m$  admits disintegrations  $\mathbf{x} \mapsto m_{\mathbf{x}}^u$ , constant on local unstable sets, and  $\mathbf{x} \mapsto m_{\mathbf{x}}^s$ , constant on local stable sets. By essential uniqueness, there exists a full  $\mu$ -measure set  $X \subset M$  such that  $m_{\mathbf{x}}^u = m_{\mathbf{x}}^s$  for all  $\mathbf{x} \in X$ . The assumption on  $\mu$  implies that  $\mu = \mu^u \times \mu^s$  where  $\mu^u$  is a probability on the set positive one-sided sequences  $(x_n)_{n \geq 0}$  and  $\mu^s$  is a probability on the set negative one-sided sequences  $(x_n)_{n < 0}$ . Fix  $\bar{\mathbf{x}} \in M$  such that  $W_{loc}^u(\bar{\mathbf{x}})$  intersects  $X$  on a full  $\mu^u$ -measure set. Then let  $\eta = m_{\bar{\mathbf{x}}}^u$ . The local stable

sets through the points of  $X \cap W_{loc}^u(\mathbf{x})$  fill-in a full  $\mu$ -measure subset of  $M$ . Thus,  $\eta = m_{\mathbf{x}}^s$  at  $\mu$ -almost every point and so the constant family  $\mathbf{x} \mapsto m_{\mathbf{x}} = \eta$  is a disintegration of  $m$ . This means that  $m = \mu \times \eta$ . Finally, the fact that  $\mu$  and  $m$  are invariant gives  $A(\mathbf{x})_* m_{\mathbf{x}} = m_{f(\mathbf{x})}$  at  $\mu$ -almost every point and that implies  $(A_x)_* \eta = \eta$  for  $p$ -almost every  $x \in \mathcal{X}$ , as claimed.  $\square$

**Lemma 3.4.** *If  $\lambda_{\pm}(A, p) = 0$  then every  $F_A$ -invariant measure  $m$  in  $\mathcal{M}(p)$  is an  $su$ -state.*

*Proof.* This is a direct consequence of Ledrappier [27, Theorem 1]. Indeed, let  $\mathcal{B}^s$  be the  $\sigma$ -algebra of measurable subsets of  $M$  which are unions of entire local stable sets. Clearly,  $f$  and  $F_A$  are  $\mathcal{B}^s$ -measurable. Hence, Ledrappier's theorem gives that the disintegration of any  $F_A$ -invariant probability  $m \in \mathcal{M}(p)$  is  $\mathcal{B}^s$ -measurable modulo zero  $\mu$ -measure sets. This is the same as saying that  $m$  is an  $s$ -state. Analogously, one proves that  $m$  is a  $u$ -state.  $\square$

Let us consider the function  $\phi_A : M \times \mathbb{P}(\mathbb{C}^d) \rightarrow \mathbb{R}$  defined by

$$\phi_A(\mathbf{x}, [v]) = \log \frac{\|A(\mathbf{x})v\|}{\|v\|}. \quad (11)$$

**Lemma 3.5.** *For every  $A : \mathcal{X} \rightarrow \text{GL}(d, \mathbb{C})$  and every  $F_A$ -invariant probability measure  $m \in \mathcal{M}(p)$ ,*

$$\lambda_-(A, p) \leq \int \phi_A dm \leq \lambda_+(A, p).$$

*Proof.* For every  $(\mathbf{x}, [v]) \in M \times \mathbb{P}(\mathbb{C}^d)$  and  $n \geq 1$ ,

$$\sum_{j=0}^{n-1} \phi_A(F_A^j(\mathbf{x}, [v])) \leq \log \|A^n(\mathbf{x})\|.$$

Integrating with respect to any probability  $m \in \mathcal{M}(p)$ ,

$$\frac{1}{n} \int \sum_{j=0}^{n-1} \phi_A \circ F_A^j dm \leq \frac{1}{n} \int \log \|A^n(\mathbf{x})\| d\mu(\mathbf{x}).$$

The right hand side converges to  $\lambda_+(A, p)$  and, assuming  $m$  is invariant, the left hand side coincides with  $\int \phi_A dm$ . This gives the upper bound in the statement. The lower bound is analogous.  $\square$

Now let  $A$  take values in  $\text{SL}(2, \mathbb{C})$ . We want to show that the upper bound in Lemma 3.5 is attained at some  $u$ -state and the lower bound is attained at some  $s$ -state. When  $\lambda_{\pm}(A, p) = 0$  this is a trivial consequence of Lemma 3.4. So, it is no restriction to suppose that  $\lambda_+(A, p) > 0 > \lambda_-(A, p)$ .

Let  $E_{\mathbf{x}}^u \oplus E_{\mathbf{x}}^s$  be the Oseledets splitting of  $F_A$ , defined at  $\mu$ -almost every  $\mathbf{x}$ . Consider the probabilities  $m^u$  and  $m^s$  defined on  $M \times \mathbb{P}(\mathbb{C}^2)$  by

$$m^*(B) = \mu(\{\mathbf{x} : (\mathbf{x}, E_{\mathbf{x}}^*) \in B\}) = \int \delta_{(\mathbf{x}, E_{\mathbf{x}}^*)}(B) d\mu(\mathbf{x}) \quad (12)$$

for  $*$   $\in \{s, u\}$  and any measurable subset  $B$ . It is clear that  $m^u$  and  $m^s$  are invariant under  $F_A$  and project down to  $\mu$ . Moreover, their disintegrations are given by

$$\mathbf{x} \mapsto \delta_{(\mathbf{x}, E_{\mathbf{x}}^*)} \quad \text{for } * \in \{s, u\}.$$

Since  $E_{\mathbf{x}}^u$  depends only on  $\{A_{x_n} : n < 0\}$  and  $E_{\mathbf{x}}^s$  depends only on  $\{A_{x_n} : n \geq 0\}$ , we get that  $m^u$  is a  $u$ -state and  $m^s$  is an  $s$ -state.



**Lemma 3.6.** *Every  $F_A$ -invariant probability measure  $m \in \mathcal{M}(p)$  is a convex combination  $m = \alpha m^u + \beta m^s$ , for some  $\alpha, \beta \geq 0$  with  $\alpha + \beta = 1$ .*

*Proof.* Given  $\kappa > 0$ , define  $X_\kappa$  to be the set of all  $(\mathbf{x}, [v]) \in M \times \mathbb{P}(\mathbb{C}^2)$  such that the Oseledets splitting  $E_{\mathbf{x}}^u \oplus E_{\mathbf{x}}^s$  is defined at  $\mathbf{x}$  and  $[v]$  splits  $v = v^u + v^s$  with  $\kappa^{-1}\|v^s\| \leq \|v^u\| \leq \kappa\|v^s\|$ . Since the two Lyapunov exponents are distinct, any point of  $X_\kappa$  returns at most finitely many times to  $X_\kappa$ . So, by Poincaré recurrence,  $m(X_\kappa) = 0$  for every  $\kappa$ . This means that  $m$  gives full weight to  $\{(\mathbf{x}, E_{\mathbf{x}}^u), (\mathbf{x}, E_{\mathbf{x}}^s) : \mathbf{x} \in M\}$  and so it is a convex combination of  $m^u$  and  $m^s$ .  $\square$

**Lemma 3.7.**  $\lambda_+(A, p) = \int \phi_A dm^u$  and  $\lambda_-(A, p) = \int \phi_A dm^s$ .

*Proof.* Let  $v_{\mathbf{x}}^u$  be a unit vector in the Oseledets subspace  $E_{\mathbf{x}}^u$ . Then

$$\begin{aligned} \lambda_+(A, \mathbf{x}) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \|A^n(\mathbf{x})v_{\mathbf{x}}^u\| = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \log \|A(f^j(\mathbf{x}))v_{f^j(\mathbf{x})}^u\| \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \phi_A(f^j(\mathbf{x}), E_{f^j(\mathbf{x})}^u) = \tilde{\phi}_A(\mathbf{x}, E_{\mathbf{x}}^u) \end{aligned}$$

for  $\mu$ -almost every  $\mathbf{x}$ , where  $\tilde{\phi}_A$  is the Birkhoff average of  $\phi_A$  for  $F_A$ . Hence,

$$\lambda_+(A, p) = \int \tilde{\phi}_A(\mathbf{x}, E_{\mathbf{x}}^u) d\mu(\mathbf{x}) = \int \tilde{\phi}_A dm^u = \int \phi_A dm^u.$$

Analogously,  $\lambda_-(A, p) = \int \phi_A dm^s$ . This completes the proof.  $\square$

**Remark 3.8.** It follows from Lemma 3.6 that  $m^u$  is the unique invariant measure  $m$  such that  $\lambda_+(A, p) = \int \phi_A dm$ .

3.1.2. *Stationary measures.* Given  $(B, q)$  in  $\mathcal{S}(X) \times \mathcal{P}(\mathcal{X})$ , a probability  $\eta$  on  $\mathbb{P}(\mathbb{C}^2)$  is called  $(B, q)$ -stationary if

$$\eta = \int (B_x)_* \eta dq(x). \quad (13)$$

The next lemma asserts that the stationary measures are the projections to  $\mathbb{P}(\mathbb{C}^2)$  of the  $u$ -states of the corresponding cocycle. We are going to denote  $M^u = \mathcal{X}^{\mathbb{Z}^+}$  and  $M^s = \mathcal{X}^{\mathbb{Z}^-}$ . Notice that  $q^{\mathbb{Z}} = \mu^s \times \mu^u$  where  $\mu^*$  is a measure on  $M^*$ , for  $* \in \{s, u\}$ .

**Lemma 3.9.** *If  $m$  is an invariant  $u$ -state for  $(B, q)$  then its projection  $\eta$  to  $\mathbb{P}(\mathbb{C}^2)$  is a  $(B, q)$ -stationary measure. Conversely, given any  $(B, q)$ -stationary  $\eta$  there exists an invariant  $u$ -state that projects to  $\eta$ .*

*Proof.* Let  $\mathbf{x} \mapsto m_{\mathbf{x}}$  be a disintegration of  $m$  constant along unstable leaves. For any measurable set  $I \subset \mathbb{P}(\mathbb{C}^2)$ ,

$$\eta(I) = m(M \times I) = \int m_{\mathbf{x}}(M \times I) d\mu(\mathbf{x}) = \int m_{f(\mathbf{x})}(M \times I) d\mu(\mathbf{x})$$

because  $\mu$  is  $f$ -invariant. Since  $m$  is  $F_B$ -invariant, the expression on the right hand side may be rewritten as

$$\begin{aligned} &\int B(\mathbf{x})_* m_{\mathbf{x}}(M \times I) d\mu(\mathbf{x}) \\ &= \int_{M^s} \left( \int_{M^u} B(\mathbf{x})_* m_{\mathbf{x}}(M \times I) d\mu^u(\mathbf{x}^u) \right) d\mu^s(\mathbf{x}^s). \end{aligned}$$

Since the disintegration is constant on local unstable sets and  $B(\mathbf{x}^s, \mathbf{x}^u)$  depends only on  $\mathbf{x}^s$  (we write  $B(\mathbf{x}^s)$  instead), this last expression coincides with

$$\begin{aligned} & \int_{\Sigma^s} B(\mathbf{x}^s)_* \left( \int_{\Sigma^u} m_{\mathbf{x}^u}(M \times I) d\mu^u(\mathbf{x}^u) \right) d\mu^s(\mathbf{x}^s) \\ &= \int_{\Sigma^s} B(\mathbf{x}^s)_* \eta(I) d\mu^s(\mathbf{x}^s) = \int B(\mathbf{x})_* \eta(I) d\mu(\mathbf{x}) = \int (B_x)_* \eta(I) dq(x). \end{aligned}$$

Thus,  $\eta = \int (B_x)_* \eta dq(x)$  as claimed.

Conversely, given any  $(B, q)$ -stationary measure  $\eta$ , consider the sequence of functions

$$m^n : \mathbf{x} \mapsto m_{\mathbf{x}}^n = B^n(f^{-n}(\mathbf{x}))_* \eta$$

with values in the space of probabilities on  $\mathbb{P}(\mathbb{C}^2)$ . It is clear from the definition that each  $m^n$  is measurable with respect to the  $\sigma$ -algebra  $\mathcal{F}^n$  of subsets of  $M$  generated by the cylinders

$$[-n : \Delta_{-n}, \dots, \Delta_{-1}] = \{\mathbf{x} \in M : x_i \in \Delta_i \text{ for } i = -n, \dots, -1\},$$

where the  $\Delta_i$  are measurable subsets of  $\mathcal{X}$ . These  $\sigma$ -algebras  $\mathcal{F}^n$  form a non-decreasing sequence. We claim that  $(m^n, \mathcal{F}^n)$  is a martingale, that is,

$$\int_C m^{n+1} d\mu = \int_C m^n d\mu \quad \text{for every } C \in \mathcal{F}^n \text{ and every } n \geq 1. \quad (14)$$

To prove this, it suffices to treat the case when  $C$  is a cylinder  $[-n : \Delta_{-n}, \dots, \Delta_{-1}]$ . Then, for any  $n \geq 1$ ,

$$\begin{aligned} \int_C A^{n+1}(f^{-n-1}(\mathbf{x}))_* \eta d\mu(\mathbf{x}) &= \int_C A^n(f^{-n}(\mathbf{x}))_* A(f^{-n-1}(\mathbf{x}))_* \eta d\mu(\mathbf{x}) \\ &= \int_C A^n(f^{-n}(\mathbf{x}))_* \left[ \int_{\mathcal{X}} (A_y)_* \eta dp(y) \right] d\mu(\mathbf{x}) \\ &= \int_C A^n(f^{-n}(\mathbf{x}))_* \eta d\mu \end{aligned}$$

because  $\eta$  is stationary. This proves the claim (14). Then, by the martingale convergence theorem (see [13, Chapter 5]), there exists a function  $\mathbf{x} \mapsto m_{\mathbf{x}}$  such that  $m_{\mathbf{x}}^n$  converges  $\mu$ -almost everywhere to  $m_{\mathbf{x}}$  in the weak\* topology. Let  $m$  be the probability measure defined on  $M \times \mathbb{P}(\mathbb{C}^2)$  by

$$m(E) = \int m_{\mathbf{x}}(E \cap (\{\mathbf{x}\} \times \mathbb{P}(\mathbb{C}^2))) d\mu(\mathbf{x})$$

for any measurable set  $E$ . By construction, the disintegration  $\mathbf{x} \mapsto m_{\mathbf{x}}$  is constant on every  $\{\mathbf{x}^s\} \times M^u$ . This means that  $m$  is a  $u$ -state. Also by construction,  $m_{f(\mathbf{x})} = A(\mathbf{x})_* m_{\mathbf{x}}$  for  $\mu$ -almost every  $\mathbf{x} \in M$ . This proves that the  $u$ -state  $m$  is invariant. Moreover, by (14) and the assumption that  $\eta$  is stationary,

$$m^n(M \times I) = m^1(M \times I) = \int_M (A_x)_* \eta(I) dp(x) = \eta(I)$$

for every  $n \geq 1$  and any measurable set  $I \subset \mathbb{P}(\mathbb{C}^2)$ . This means that  $m^n$  projects to  $\eta$  for every  $n \geq 1$ . Then so does the limit  $m$ . This completes the proof of the lemma.  $\square$

We are also going to show that the projection of  $m^u$  to the projective space  $\mathbb{P}(\mathbb{C}^2)$  completely determines the Lyapunov exponents:

**Lemma 3.10.** *Let  $m$  be a  $u$ -state realizing  $\lambda_+(A, p)$  and let  $\eta$  be its projection to  $\mathbb{P}(\mathbb{C}^2)$ . Then*

$$\lambda_+(A, p) = \int \int \log \phi_A(x, v) d\eta([v]) dp(x)$$

*Proof.* Suppose first that  $\lambda_+(A, p) = 0$ . By Lemmas 3.4 and 3.3, every  $F_A$ -invariant probability  $m$  that projects down to  $\mu$  realizes the largest exponent and is a product measure  $m = \mu \times \eta$ . Thus, in this case, the lemma follows immediately from Fubini's Theorem. If  $\lambda_+(A, p) > 0$ , then  $m^u$  is the unique  $u$ -state that realizes  $\lambda_+$ . Then a straightforward calculation,

$$\begin{aligned} \lambda_+(A, p) &= \int_M \log \|A(\mathbf{x})E_{\mathbf{x}}^u\| d\mu = \int_{M^s} \int_{M^u} \log \|A(\mathbf{x}^s)E_{\mathbf{x}^u}^u\| d\mu(\mathbf{x}^u) d\mu^s \\ &= \int_{\mathcal{X}} \int_{M^u} \log \|A_y E_{\mathbf{x}^u}^u\| d\mu(\mathbf{x}^u) dp(y) \\ &= \int_{\mathcal{X}} \int_{M^u} \int_{\mathbb{P}(\mathbb{C}^2)} \log \frac{\|A_y v\|}{\|v\|} d\delta_{E_{\mathbf{x}^u}^u} d\mu(\mathbf{x}^u) dp(y) \\ &= \int_{\mathcal{X}} \int_{\mathbb{P}(\mathbb{C}^2)} \log \frac{\|A_y v\|}{\|v\|} d\eta([v]) dp(y), \end{aligned}$$

concludes the proof of the lemma.  $\square$

**Lemma 3.11.** *If  $(A^k, p^k)_k$  converges to  $(A, p)$  and  $\eta^k$  is a sequence of  $(A^k, p^k)$ -stationary measure converging to  $\eta$  then  $\eta$  is an  $(A, p)$ -stationary measure.*

*Proof.* We have to show that

$$\lim_k \int (A_x^k)_* \eta^k dp^k = \int (A_x)_* \eta dp$$

in the weak\* sense. Let  $\phi : \mathbb{P}(\mathbb{C}^2) \rightarrow \mathbb{R}$  be a continuous function. Then

$$\left| \int \int \phi(A_x^k v) d\eta^k dp^k - \int \int \phi(A_x v) d\eta dp \right| \leq a_k + b_k + c_k$$

where

$$\begin{aligned} a_k &= \left| \int \int \phi(A_x^k v) d\eta^k dp^k - \int \int \phi(A_x v) d\eta^k dp^k \right| \\ b_k &= \left| \int \int \phi(A_x v) d\eta^k dp^k - \int \int \phi(A_x v) d\eta dp^k \right| \\ c_k &= \left| \int \int \phi(A_x v) d\eta dp^k - \int \int \phi(A_x v) d\eta dp \right| \end{aligned}$$

It is clear that  $(a_k)_k$  converges to zero, because  $\|A_x^k - A_x\|$  converges uniformly to zero and  $\phi$  is uniformly continuous. To prove that  $b_k$  converges to zero we argue as follows. Given  $\varepsilon > 0$ , fix  $\delta > 0$  such that  $|\phi(v) - \phi(w)| < \varepsilon/3$  for all  $v, w \in \mathbb{P}(\mathbb{C}^2)$  such that  $d(v, w) < \delta$ . Since the image of  $A$  is contained in a compact subset of  $\text{SL}(2, \mathbb{C})$ , there are  $B_1, \dots, B_n \in \text{SL}(2, \mathbb{C})$  such that their  $\delta$ -neighborhoods cover  $A(\mathcal{X})$ . The assumption that  $(\eta^k)_k$  converges to  $\eta$  in the weak\* topology implies that there exists  $k_0 \in \mathbb{N}$  such that

$$\left| \int \phi(B_i v) d\eta^k - \int \phi(B_i v) d\eta \right| < \varepsilon/3$$

for all  $k > k_0$  and for all  $i = 1, \dots, n$ . Then we can use the triangle inequality to conclude that

$$\left| \int \phi(A_x v) d\eta^k - \int \phi(A_x v) d\eta \right| \leq \varepsilon$$

for all  $k > k_0$ . Integrating with respect to  $p^k$  we conclude that  $b_k \leq \varepsilon$  for all  $k > k_0$ . This proves that  $b_k$  converges to 0. Finally, it is clear that  $a_k$  converges to zero, because our assumptions imply that  $(p^k)_k$  converges strongly to  $p$ . The proof of the lemma is complete.  $\square$

3.1.3. *Proof of Proposition 3.1.* Notice that  $\lambda_+$  is non-negative and, as observed in (4)–(5),

$$(A, p) \mapsto \lambda_+(A, p) = \inf \frac{1}{n} \int \log \|A^n(\mathbf{x})\| d\mu(\mathbf{x}) \quad (15)$$

is upper-semicontinuous for the topology defined by (8). So, if  $(A, p) \in \mathcal{S}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$  is a discontinuity point for the largest Lyapunov exponent then  $\lambda_+(A, p) > 0$  and there is a sequence  $(A^k, p^k)_k$  converging to  $(A, p)$  as  $k \rightarrow \infty$  such that

$$\lim_k \lambda_+(A^k, p^k) < \lambda_+(A, p).$$

As we have seen, for each  $k$  there exists some  $(A^k, p^k)$ -stationary measure  $\eta^k$  satisfying

$$\int_{\mathcal{X}} \int_{\mathbb{P}(\mathbb{C}^2)} \log \|A_x^k v\| d\eta^k(v) dp^k(x) = \lambda_+(A^k, p^k).$$

Up to restricting to a subsequence, we may assume that  $(\eta^k)_k$  converges in the weak\* topology to some probability measure  $\eta$  on  $\mathbb{P}(\mathbb{C}^2)$ . Then  $\eta$  is an  $(A, p)$ -stationary measure, by Lemma 3.11. Using Lemma 3.10 we see that

$$\begin{aligned} \int_{\mathcal{X}} \int_{\mathbb{P}(\mathbb{C}^2)} \log \|A_x v\| d\eta(v) dp(x) &= \lim_k \int_{\mathcal{X}} \int_{\mathbb{P}(\mathbb{C}^2)} \log \|A_x^k v\| d\eta^k(v) dp^k(x) \\ &< \lambda_+(A, p) = \int_{\mathcal{X}} \int_{\mathbb{P}(\mathbb{C}^2)} \log \|A_x v\| d\eta^u(v) dp(x) \end{aligned}$$

where  $\eta^u$  is the projection of  $m^u$ . In particular, by Lemma 3.9, there exists an invariant  $u$ -state  $m \neq m^u$ . It follows, using Lemma 3.6, that

$$m = \alpha m^u + \beta m^s \quad \text{with } \alpha + \beta = 1 \text{ and } \beta \neq 0.$$

This implies that  $m^s$  is a  $u$ -state, because it is a linear combination of  $m$  and  $m^u$ . Hence  $m^s$  is an  $su$ -state. In view of Lemma 3.3 this means that the Oseledets subspace  $E_{\mathbf{x}}^s$  is constant on a full  $\mu$ -measure set. Let  $F^s \in \mathbb{P}(\mathbb{C}^2)$  denote this constant. Analogously, using that  $(A, p)$  is a discontinuity point for the smallest Lyapunov exponent, we find  $F^u \in \mathbb{P}(\mathbb{C}^2)$  such that  $E_{\mathbf{x}}^u = F^u$  for  $\mu$ -almost every  $\mathbf{x}$ . It is clear that  $F^u$  and  $F^s$  are both invariant under  $A_x$ , for  $p$ -almost every  $x \in \mathcal{X}$ , because  $\mu = p^{\mathcal{Z}}$ . This means that there exists  $\mathcal{Z} \subset \mathcal{X}$  with  $p(\mathcal{Z}) = 1$  such that the linear operators defined by the  $A_y$ ,  $y \in \mathcal{Z}$  have a common eigenbasis, which is precisely the first claim in the proposition. The last claim (commutativity) is a trivial consequence. This completes the proof of Proposition 3.1.

**3.2. Handling the diagonal case.** Here we prove Proposition 3.2. Let  $(A, p) \in \mathcal{S}(X) \times \mathcal{P}(\mathcal{X})$  and  $\mathcal{Z}$  be as in the conclusion of Proposition 3.1 and consider any  $p \in \mathcal{P}(\mathcal{X})$ . Since conjugacies preserve the Lyapunov exponents, we may suppose  $P = \text{id}$  and

$$A_x = \begin{pmatrix} \theta_x & 0 \\ 0 & \theta_x^{-1} \end{pmatrix} \quad \text{for all } x \in \mathcal{Z}. \quad (16)$$

Notice that the Lyapunov exponents of  $(A, p)$  are

$$\pm \int_{\mathcal{Z}} \log |\theta_x| dp(x). \quad (17)$$

If they vanish then  $(A, p)$  is automatically a continuity point, and so there is nothing to prove. Otherwise, it is no restriction to suppose

$$\int_{\mathcal{Z}} \log |\theta_x| > 0. \quad (18)$$

Let  $V_\varepsilon$  be the  $\varepsilon$ -neighborhood of the horizontal direction in  $\mathbb{P}(\mathbb{C}^2)$  and  $\mathcal{Z}$  be as given in Proposition 3.1. The key step in the proof of Theorem C is the following

**Proposition 3.12.** *Given  $\varepsilon > 0$  and  $\delta > 0$  there exists  $\gamma > 0$  such that if  $(B, q) \in V(A, p, \gamma, \mathcal{Z})$  and there is no one-dimensional subspace invariant under all  $B_x$  for  $x$  in a full  $q$ -measure then  $\eta(V_\varepsilon^c) \leq \delta$  for any  $(B, q)$ -stationary measure  $\eta$ .*

The proof of Proposition 3.12 will be given in Section 4. Right now, let us conclude the proof of Proposition 3.2.

Let  $(B, q) \in \mathcal{S}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$  be close to  $(A, p)$  in the sense of (8). First, suppose there exists some one-dimensional subspace  $r \subset \mathbb{C}^2$  invariant under all the  $B_x$ ,  $x$  in a  $q$ -full measure. Then  $r$  must be close to either the vertical axis or the horizontal axis: that is because (18) implies  $|\theta_x| \neq 1$  for some  $q$ -positive measure subset. Then the Lyapunov exponent of  $(B, q)$  along  $r$  is close to one of the exponents (17). Since the other exponent is symmetric, this proves that the Lyapunov exponents of  $(B, q)$  are close to the Lyapunov exponents of  $(A, p)$ . Now assume  $B$  does not admit any invariant one-dimensional subspace. Let  $M > 0$  such that  $M^{-1}\|v\| < \|B_x v\| < M\|v\|$  for  $p$ -almost every  $x \in \mathcal{X}$ , all  $v \in \mathbb{C}^2$  and  $d(A, B) < 1$ . Let  $0 \ll \varepsilon \ll \delta \ll \rho \ll 1$ . Let  $m$  be any  $u$ -state realizing the largest Lyapunov exponent of  $(B, q)$ , and  $\eta$  its projection on  $P(\mathbb{C}^2)$ . By Proposition 3.12,

$$\begin{aligned} \int_{\mathbb{P}(\mathbb{C}^2)} \log \frac{\|B_x v\|}{\|v\|} d\eta([v]) &= \int_{V_\varepsilon} \log \frac{\|B_x v\|}{\|v\|} d\eta([v]) + \int_{V_\varepsilon^c} \log \frac{\|B_x v\|}{\|v\|} d\eta([v]) \\ &\geq -\delta \log M + \eta(V_\varepsilon)(\log |\theta_x| - \delta) \end{aligned}$$

for  $q$ -almost every  $x \in \mathcal{X}$ . Together with Lemma 3.10, this implies

$$\lambda_+(B, q) > \eta(V_\varepsilon)\lambda_+(A, p) - \delta(\log M + \eta(V_\varepsilon)) > \lambda_+(A, p) - \rho.$$

Upper semi-continuity gives  $\lambda_+(B, q) \leq \lambda_+(A, p) + \rho$ . Thus, we have shown that  $(A, p)$  is indeed a continuity point for the Lyapunov exponents.

This reduces the proof of Proposition 3.2 and Theorem C to proving Proposition 3.12.

#### 4. PROOF OF THE KEY PROPOSITION

Here we give a suitable reformulation of Proposition 3.12 and reduce its proof to two technical estimates, Propositions 4.8 and 4.10, whose proof will be presented in the next section.

**4.1. Preliminary observations.** As a first step we note that under the assumptions of the proposition all stationary measures are non-atomic.

**Lemma 4.1.** *There exists  $\gamma > 0$  such that if  $(B, q) \in V(A, p, \gamma, \mathcal{Z})$  and there is no one-dimensional subspace of  $\mathbb{R}^2$  invariant under  $B_x$  for every  $x$  in a full  $q$ -measure, then every  $(B, q)$ -stationary measure is non-atomic.*

*Proof.* By assumption,  $A$  is diagonal and the Lyapunov exponents do not vanish. So, we may take  $\gamma > 0$  so that if  $(B, q) \in V(A, p, \gamma, \mathcal{Z})$  then  $B_x$  is hyperbolic and its eigenspaces are close to the horizontal and vertical directions, for every  $x$  in some set  $\mathcal{L} \subset \mathcal{X}$  with  $q(\mathcal{L}) > 0$ . Then any finite set of one-dimensional subspaces invariant under any  $B_x$ ,  $x \in \mathcal{L}$  has at most two elements. Moreover, they must coincide with the eigenspaces of  $B_x$  and, consequently, are actually fixed under  $B_x$ . Since we assume there is no one-dimensional subspace fixed by  $B_x$  for  $\mu$ -almost every  $x$ , it follows that there is no finite set of one-dimensional subspaces invariant under  $B_x$  for  $\mu$ -almost every  $x$ .

Now let us suppose  $\eta$  has some atom. Let  $z_1, \dots, z_N$  be the atoms with the largest mass, say,  $\eta(\{z_i\}) = a$  for  $i = 1, \dots, N$ . Since  $\eta$  is a stationary measure,

$$\eta(\{B_x^{-1}(z_1), \dots, B_x^{-1}(z_N)\}) = \eta(\{z_1, \dots, z_N\}) = Na$$

for  $q$ -almost every  $x \in \mathcal{X}$ . Moreover, in view of the previous paragraph, we have  $\{B_x^{-1}(z_1), \dots, B_x^{-1}(z_N)\} \neq \{z_1, \dots, z_N\}$  for a positive  $q$ -measure subset of points  $x$ . This implies that there exists  $z \neq z_i$  for  $i = 1, \dots, N$  such that  $\eta(\{z\}) = a$ . That contradicts the choice of the  $z_i$  and so the lemma is proved.  $\square$

Let  $\phi : \mathbb{P}(\mathbb{C}^2) \rightarrow \mathbb{C} \cup \{\infty\}$ ,  $\phi([z_1, z_2]) = z_1/z_2$  be the standard identification between the complex projective space and the Riemann sphere. Then the projective action of a linear map

$$B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

corresponds to a Möbius transformation on the sphere

$$\hat{B} : \mathbb{C} \cup \{\infty\} \rightarrow \mathbb{C} \cup \{\infty\} \quad \hat{B}(z) = \frac{az + b}{cz + d},$$

in the sense that  $\phi \circ B = \hat{B} \circ \phi$ . It follows that a measure  $\xi$  in projective space is  $(B, q)$ -stationary if and only if the measure  $\eta = \phi_* \xi$  on the sphere satisfies  $\eta = \int (\hat{B}_x)_* \eta dq(x)$ . Then the measure  $\eta$  is also said to be  $(B, q)$ -stationary. Clearly,  $\eta$  is non-atomic if and only if  $\xi$  is.

This means that the key Proposition 3.12 may be restated as

**Proposition 4.2.** *Given  $\varepsilon > 0$  and  $\delta > 0$  there exist  $\gamma > 0$  such that if  $(B, q) \in V(A, p, \gamma, \mathcal{Z})$  and  $q(\{x \in \mathcal{X} : \hat{B}_x(z) = z\}) < 1$  for all  $z \in \mathbb{C} \cup \{\infty\}$  then*

$$\eta(B(0, \varepsilon^{-1})) \leq \delta$$

for any  $(B, q)$ -stationary probability measure  $\eta$  on  $\mathbb{C} \cup \{\infty\}$ .

The proof of this proposition will appear in the next section. Let us briefly comment on the statement and the overall strategy of the proof. As mentioned before, the set  $\text{Stat}(A, p)$  of stationary measures varies in a semi-continuous fashion with the data: if  $(B, q)$  is close to  $(A, p)$  then every  $(B, q)$ -stationary measure is close to  $\text{Stat}(A, p)$ . This is not sufficient for our purposes because in the diagonal case there

are several stationary measures, not all of which realize the largest Lyapunov exponent. Indeed, the assumption that both the vertical direction and the horizontal direction are invariant under almost every  $A_x$  implies that both associated Dirac masses on the Riemann sphere,  $\delta_0$  and  $\delta_\infty$ , are  $(A, p)$ -stationary measures, and so  $\text{Stat}(A, p)$  the whole line segment between these two Dirac masses.

To establish continuity of the Lyapunov exponents we need to prove the much finer fact that stationary measures of nearby (irreducible) cocycles are close to the one element of  $\text{Stat}(A, p)$ , namely  $\delta_\infty$ , that realizes the Lyapunov exponent  $\lambda_+(A, p)$ . That is the meaning of the key proposition. The reason we may restrict ourselves to irreducible cocycles is because in the reducible case continuity follows from a different, and much easier argument, as we have seen.

The crucial property that singles out  $\delta_\infty$  among all  $(A, p)$ -stationary measures is the fact that it is an attractor for the random walk defined by  $(A, p)$  on  $\mathbb{P}(\mathbb{C}^2)$ . Indeed, the random trajectory  $A^n(x)\xi$  of any  $\xi \in \mathbb{P}(\mathbb{C}^2) \setminus \{0\}$  converges to  $\infty$  almost surely. Consequently, the forward iterates of any probability  $\eta$  with  $\eta(\{0\}) = 0$  under the dynamics

$$f_A : \eta \mapsto \int (A_x)_* \eta dp(x) \quad (19)$$

induced by  $A$  in the space of the probability measures of  $\mathbb{P}(\mathbb{C}^2)$  converge to  $\delta_\infty$ .

The heart of the proof is, thus, a robustness theorem for certain random walks. We prove that the attractor persists for all nearby irreducible cocycles: if  $(B, q)$  is close enough to  $(A, p)$  and there is no one-dimensional subspace invariant under  $q$ -almost every  $B_x$ , then  $f_B$  possesses an attractor that is strongly concentrated near  $\infty$ , and draws the forward iterates of every Dirac mass. In particular, every fixed point  $\eta$  of the operator  $f_B$  must be strongly concentrated near  $\infty$ , as claimed.

While the details are fairly lengthy, the main ideas in the proof are very natural, so that applications of this approach to much more general situations can be expected. In particular, there is some promising progress in the setting of Hölder continuous (not locally constant) two-dimensional cocycles over hyperbolic systems.

**4.2. Auxiliary statements.** Recall, from (16) and (18), that

$$A_x = \begin{pmatrix} \theta_x & 0 \\ 0 & \theta_x^{-1} \end{pmatrix} \quad \text{with} \quad \int \log |\theta_x| dp(x) > 0 \quad (20)$$

for every  $x \in \mathcal{Z}$ . By definition,  $q(\mathcal{Z}) = 1$  for all  $(B, q) \in V(A, p, \gamma, \mathcal{Z})$ . Thus, up to restricting all cocycles to a full measure subset, which does not affect the Lyapunov exponents, we may assume that  $\mathcal{Z} = \mathcal{X}$ . We do so in all that follows. Let  $B$ ,  $q$ , and  $\eta$  be as in the statement.

**Lemma 4.3.** *There are  $\beta, \sigma \in (0, 1)$ ,  $k \in \mathbb{N}$ , positive numbers  $(\sigma_x)_{x \in \mathcal{X}}$ , and integers  $(s_x)_{x \in \mathcal{X}}$  such that*

- (a)  $0 < \|A\|^{-1}/4 \leq \sigma_x \leq \beta|\theta_x|$  for all  $x \in \mathcal{X}$
- (b)  $\sigma_x = \sigma^{s_x}$  for all  $x \in \mathcal{X}$
- (c)  $\int \log \sigma_x dp(x) > 4/k$ .

*Proof.* Fix  $k \in \mathbb{N}$  large enough so that  $\int \log |\theta_x| dp(x) > 7/k$ . Define  $\log \beta = \log \sigma = -1/k$ . For each  $x \in \mathcal{X}$ , define

$$r_x = [k \log |\theta_x|], \quad s_x = \begin{cases} r_x - 1 & \text{if } r_x \neq 1 \\ r_x - 2 & \text{if } r_x = 1 \end{cases} \quad \log \sigma_x = -\frac{s_x}{k}.$$

Properties (a) and (b) follow immediately. Moreover,

$$\int \log \sigma_x dp(x) \geq \int (\log |\theta_x| - 3/k) dp(x) > 4/k$$

as claimed in (c). The proof is complete.  $\square$

Let  $\sigma$ ,  $\beta$ ,  $\sigma_x$ , and  $s_x$ , be as in Lemma 4.3. We partition  $\mathcal{X} = \mathcal{X}_- \cup \mathcal{X}_+$ , where  $\mathcal{X}_-$  is the subset of  $x \in \mathcal{X}$  with  $s_x < 0$  (i.e.  $\sigma_x > 1$ ) and  $\mathcal{X}_+$  is the subset of  $x \in \mathcal{X}$  with  $s_x > 0$  (i.e.  $\sigma_x < 1$ ). For each  $x \in \mathcal{X}$ , let

$$D_x = \begin{pmatrix} \sigma_x & 0 \\ 0 & \sigma_x^{-1} \end{pmatrix} \quad \text{and} \quad \hat{D}_x(z) = \sigma_x^2 z. \quad (21)$$

Consider also

$$D_{sp} = \begin{pmatrix} \sigma^\tau & 0 \\ 0 & \sigma^{-\tau} \end{pmatrix} \quad \text{and} \quad \hat{D}_{sp}(z) = \sigma^{2\tau} z, \quad (22)$$

where  $\tau$  is the smallest integer such that  $\sigma^\tau \leq \|A\|^{-1}/4$ . Given any  $\mathcal{K} \subset \mathcal{X}$ , let  $K$  be the cocycle defined by

$$K_x = \begin{pmatrix} k_x & 0 \\ 0 & k_x^{-1} \end{pmatrix} \quad \text{where} \quad k_x = \begin{cases} \sigma_x & \text{if } x \in \mathcal{K} \\ \sigma^\tau & \text{if } x \in \mathcal{X} \setminus \mathcal{K}. \end{cases} \quad (23)$$

**Lemma 4.4.** *There exist  $\alpha > 0$  and  $\tilde{\alpha} > 0$  such that, given any measurable set  $\mathcal{K} \subset \mathcal{X}$  with  $p(\mathcal{K}) \geq 1 - \alpha$ ,*

$$\int \log k_x dp(x) \geq 2/k \quad \text{and} \quad p(\{x : k_x > 1\}) \geq \tilde{\alpha}.$$

*Proof.* Taking  $\alpha = (-k \log \sigma^\tau)^{-1}$ , we have

$$\begin{aligned} \int \log k_x dp &\geq \int \log \sigma_x dp + \int_{\mathcal{X} \setminus \mathcal{K}} (\log \sigma^\tau - \log \sigma_x) dp \\ &\geq 4/k + 2 \log \sigma^\tau p(\mathcal{X} \setminus \mathcal{K}) \geq 2/k. \end{aligned}$$

This proves the first claim. The second one is a direct consequence, with  $\tilde{\alpha} = 2/(k \sup k_x)$ .  $\square$

For  $z_0 \in \mathbb{C}$  and  $r \geq 0$ , we denote  $B(z_0, r) = \{z \in \mathbb{C} : |z - z_0| \leq r\}$ . Given  $B$ ,  $C \in \mathcal{S}(X)$  and  $\mathcal{Y} \subset \mathcal{X}$  we say that  $r \geq 0$  is  $(B, \mathcal{Y})$ -centered with respect to  $C$  if

$$\hat{B}_x^{-1}(B(0, r)) \subset \hat{C}_x^{-1}(B(0, r)) \quad \text{for every } x \in \mathcal{Y}. \quad (24)$$

When  $\mathcal{Y} = \mathcal{X}$  we just say that  $r$  is  $B$ -centered with respect to  $C$ . Given  $B$ ,  $C \in \mathcal{S}(\mathcal{X})$ ,  $q \in \mathcal{P}(\mathcal{X})$ , and a  $(B, q)$ -stationary measure  $\eta$ , we say that  $r \geq 0$  is  $(B, q, \eta)$ -targeted with respect to  $C$  if

$$\int \eta(\hat{B}_x^{-1}(B(0, r))) dq(x) \leq \int \eta(\hat{C}_x^{-1}(B(0, r))) dq(x) \quad (25)$$

**Remark 4.5.** If  $r \geq 0$  is  $B$ -centered (respectively,  $(B, q, \eta)$ -targeted) with respect to  $D$  then it is also  $B$ -centered (respectively,  $(B, q, \eta)$ -targeted) with respect to the cocycle  $K$  defined in (23). That is because  $\hat{D}_x^{-1}(B(0, r)) \subset \hat{K}_x^{-1}(B(0, r))$  for any  $x \in \mathcal{X}$ .

The following simple facts will be useful in what follows:

**Lemma 4.6.** *Given  $\rho > 0$  there is  $\gamma > 0$  such every  $r \in [\rho, \rho^{-1}]$  is  $B$ -centered with respect to  $D$  for every  $B \in \mathcal{S}(X)$  with  $d(A, B) < \gamma$ .*



*Proof.* By assumption,  $\pm \log |\theta_x|$ ,  $x \in \mathcal{X}$  is bounded. Write

$$B_x^{-1} = \begin{pmatrix} a_x & b_x \\ c_x & d_x \end{pmatrix}.$$

The condition  $d(A, B) < \gamma$  implies that  $|a_x - \theta_x^{-1}|$ ,  $|b_x|$ ,  $|c_x|$ ,  $|d_x - \theta_x|$  are all less than  $c_1\gamma$  for some constant  $c_1$  independent of  $x$  and  $\gamma$ . Given  $\rho > 0$ , assume first that  $\gamma \leq \rho^2$ . Then, for any  $|z| \in [\rho, \rho^{-1}]$ ,

$$|\hat{B}_x^{-1}(z)| \leq \frac{|a_x z| + |b_x|}{|d_x| - |c_x z|} \leq \frac{|a_x| + c_1\sqrt{\gamma}}{|d_x| - c_1\sqrt{\gamma}}|z| \leq \frac{|\theta_x^{-1}|}{|\theta_x|} \frac{1 + c_2\sqrt{\gamma}}{1 - c_2\sqrt{\gamma}}|z|$$

where  $c_2$  is also independent of  $x$  and  $\gamma$ . Thus, there exists  $\gamma_0 > 0$ , independent of  $x \in \mathcal{X}$  such that, if  $d(A, B) < \gamma \leq \gamma_0$  then

$$|\hat{B}_x^{-1}(z)| \leq (\beta|\theta_x|)^{-2}|z| \leq \sigma_x^{-2}|z| = |\hat{D}_x^{-1}(z)|$$

for every  $x \in \mathcal{X}$  and  $|z| \in [\rho, \rho^{-1}]$ . This gives that every  $r \in [\rho, \rho^{-1}]$  is  $B$ -centered with respect to  $D$ , as claimed.  $\square$

**Lemma 4.7.** *There are  $\gamma > 0$  and  $c > 0$  such that if  $d(A, B) < \gamma$  and  $x \in X_0$  is such that  $\hat{B}_x$  has a fixed point in  $B(0, \rho)$ , for some  $\rho < c^{-1}$ , then every  $r \in [c\rho, 1]$  is  $(B, \{x\})$ -centered with respect to  $D$ .*

*Proof.* First, take  $\gamma > 0$  such that  $d(A, B) < \gamma$  implies that  $\hat{B}_x^{-1}$  is a  $\lambda_x$ -contraction with

$$\frac{\|A\|^{-1}}{2} = \lambda \leq \lambda_x \leq (1 + \rho_0)^{-2}$$

and  $\hat{D}_x^{-1}(z) = \Lambda_x z$  with  $\lambda_x \leq \beta\Lambda_x$ . Then choose  $c > 0$  large enough so that  $\lambda^{-1} < c(\beta^{-1} - 1 - c^{-1})$ . It follows that

$$|\hat{B}_x^{-1}(z)| \leq [c^{-1} + \lambda_x(1 + c^{-1})]r \leq \Lambda_x r$$

whenever  $|z| \leq r$  and  $r \in [c\rho, 1]$ . In other words,

$$\hat{B}_x^{-1}(B(0, r)) \subset \hat{D}_x^{-1}(B(0, r)) \quad \text{for all } r \in [c\rho, 1],$$

as claimed. This proves the lemma.  $\square$

The proof of Proposition 4.2 relies on a couple of technical results, Propositions 4.8 and 4.10, that we state in the sequel and whose proofs will appear in Section 5. The first proposition gives a bound on the mass of the stationary measure away from the vertical (and the horizontal) direction. Fix  $\alpha > 0$  as in Lemma 4.4, once and for all.

**Proposition 4.8.** *Given  $\varepsilon > 0$  and  $\delta > 0$  there exists  $\gamma > 0$  such that if  $d(A, B) < \gamma$  and  $d(p, q) < \gamma$  then*

$$\eta(B(0, \varepsilon^{-1}) \setminus B(0, r_0)) \leq \delta$$

for any  $(B, q)$ -stationary measure  $\eta$  and any  $r_0 \in (0, 1)$  such that every  $r \in [r_0, \varepsilon^{-1}]$  is  $(B, \mathcal{K})$ -centered with respect to  $D$  for some measurable set  $\mathcal{K}$  with  $p(\mathcal{K}) \geq 1 - \alpha$ .

What we actually use is the following consequence:

**Corollary 4.9.** *Given  $\varepsilon > 0$  and  $\delta > 0$  there exist  $\gamma > 0$  such that if  $d(A, B) < \gamma$  and  $d(p, q) < \gamma$  then either  $\eta(B(0, \varepsilon^{-1})) \leq \delta$  or there exists  $r_0 \in (0, 1)$  such that*

$$\eta(B(0, \varepsilon^{-1}) \setminus B(0, r_0)) \leq \delta$$

and  $p(\{x \in \mathcal{X} : \hat{B}_x^{-1}(B(0, r_0)) \not\subset \hat{D}_x^{-1}(B(0, r_0))\}) \geq \alpha$ .

*Proof.* Let  $r_1 \geq 0$  be the infimum of all  $r \in (0, 1)$  such that

$$\eta(B(0, \varepsilon^{-1}) \setminus B(0, r)) < \delta.$$

If  $r_1 = 0$  then  $\eta(B(0, \varepsilon^{-1}) \setminus \{0\}) \leq \delta$ . Since  $\eta$  has no atoms, by Lemma 4.1, it follows that  $\eta(B(0, \varepsilon^{-1})) \leq \delta$ . This proves the corollary in this case. Now, suppose  $r_1 > 0$ . Then,  $\eta(B(0, \varepsilon^{-1}) \setminus B(0, r_1)) \geq \delta$ , and so, by Proposition 4.8,

$$A_1 = \{x \in \mathcal{X} : \hat{B}_x^{-1}(B(0, r_1)) \not\subset \hat{D}_x^{-1}(B(0, r_1))\}$$

has  $p(A_1) > \alpha$ . Let  $1 > r_2 > r_3 > \dots$  be a decreasing sequence converging to  $r_1$ , and

$$A_k = \{x \in \mathcal{X} : \hat{B}_x^{-1}(B(0, r_k)) \not\subset \hat{D}_x^{-1}(B(0, r_k))\},$$

for  $k = 2, 3, \dots$ . Notice that  $\liminf_k A_k \supset A_1$  and so, by the Lemma of Fatou,  $\liminf_k p(A_k) \geq p(A_1) > \alpha$ . In particular, there is  $N \geq 2$  such that  $p(A_N) \geq \alpha$ . The proof is complete, taking  $r_0 = r_N$ .  $\square$

Our second technical proposition will allow us to bound the mass of the stationary measure close to the vertical direction:

**Proposition 4.10.** *There are  $\gamma > 0$  and  $N \geq 1$  such that if  $d(A, B) < \gamma$  and  $r_0 \in [0, 1]$  and  $x \in \mathcal{X}$  are such that  $\hat{B}_x^{-1}(B(0, r_0)) \not\subset \hat{D}_x^{-1}(B(0, r_0))$ , then*

$$\mathcal{D} \cap \hat{B}_x^{-1}(\mathcal{D}) = \emptyset, \quad \text{where } \mathcal{D} = \begin{cases} \hat{D}_x^{-N}(B(0, r_0)) & \text{if } x \in \mathcal{X}_- \\ \hat{D}_x^N(B(0, r_0)) & \text{if } x \in \mathcal{X}_+. \end{cases}$$

*In particular,  $B(0, \sigma^{2N\tau} r_0) \cap \hat{B}_x^{-1}(B(0, \sigma^{2N\tau} r_0)) = \emptyset$ .*

**4.3. Proof of Proposition 4.2.** The assumption  $\lambda_+(A, p) > 0$  implies that there exist  $\alpha_0 > 0$  and  $\rho_0 > 0$  such that

$$X_0 = \{x \in \mathcal{X} : |\theta_x| > 1 + \rho_0\}$$

has  $p(X_0) \geq \alpha_0$ . Let  $c > 0$  and  $N \geq 1$  be fixed as in Lemma 4.7 and Proposition 4.10, respectively. Denote  $\beta_0 = 2\alpha_0/(1 + 8c^2\sigma^{-4\tau N})$ . For each  $z \in \mathbb{C}$  and  $\rho \in [0, 1)$ , define

$$\Gamma(z, \rho) = \{x \in X_0 : \hat{B}_x \text{ has some fixed point in } B(z, \rho)\}.$$

In particular,  $\Gamma(z, 0)$  is the set of  $x \in X_0$  such that  $z$  is fixed under  $\hat{B}_x$ . Observe also that  $\Gamma(z, 0) = \bigcap_{\rho > 0} \Gamma(z, \rho)$ .

**Lemma 4.11.** *Given  $\rho > 0$  there exist  $\gamma > 0$  and  $\lambda_0 \in (0, 1)$  such that if  $d(A, B) < \gamma$  then  $\hat{B}_x^{-1}(B(0, r)) \subset B(0, \lambda_0 r)$  for every  $1 > r \geq c\rho$  and every  $x \in \Gamma(0, \rho)$ . In particular, there is  $\kappa \geq 1$  such that  $\hat{B}_x^{-\kappa}(B(0, r)) \subset B(0, \sigma^{2\tau} r)$  for  $1 > r \geq c\rho\sigma^{-2\tau}$  and  $x \in \Gamma(0, \rho)$ .*

*Proof.* Take  $\gamma > 0$  small enough to ensure that every  $\hat{B}_x^{-1}$ ,  $x \in X_0$  is a contraction on the ball  $B(0, 1)$ , with uniform contraction rate  $\lambda \in (0, 1)$ . Then, consider  $\lambda_0 = \lambda(1 + c^{-1}) + c^{-1}$ . Fix  $x \in X_0$  and let  $z_0 \in B(0, 1)$  be the unique fixed point of  $\hat{B}_x^{-1}$ . For any  $z$  with  $|z| = r \geq c\rho$ ,

$$|\hat{B}_x^{-1}(z)| \leq \rho + \lambda|z - z_0| \leq \rho + \lambda(r + \rho) \leq [c^{-1} + \lambda(1 + c^{-1})]r \leq \lambda_0 r.$$

This proves the first claim in the statement. To get the second statement, just take  $\kappa \geq 1$  to be the smallest positive integer such that  $\lambda_0^\kappa \leq \sigma^{2\tau}$ .  $\square$

We distinguish two cases in the proof of the proposition. First, we take the cocycle to be “reducible”, in the sense that the  $B_x$  have a common invariant line, for a subset of values of  $x \in X_0$  with sizable mass. More precisely, we suppose that

$$p(\Gamma(z_0, 0)) \geq \beta_0 \quad \text{for some } z_0 \in B(0, 1). \quad (26)$$

It is no restriction to suppose that  $z_0 = 0$ , as we will see in a while, so let us do that for the time being. Then, (26) implies that  $q(\Gamma(0, 0)) \geq \beta_0/2$  for every  $q$  in a neighborhood of  $p$ . Suppose, by contradiction, that  $\eta(B(0, \varepsilon^{-1})) > \delta$ . Then, by Corollary 4.9, there exists  $r_0 \in (0, 1)$  such that

$$\eta(B(0, \varepsilon^{-1}) \setminus B(0, r_0)) \leq \delta \quad \text{and} \quad p(Y) \geq \alpha, \quad (27)$$

where  $Y = \{x \in \mathcal{X} : \hat{B}_x^{-1}(B(0, r_0)) \not\subset \hat{D}_x^{-1}(B(0, r_0))\}$ . The latter implies that  $q(Y) \geq \alpha/2$  for every  $q$  sufficiently close to  $p$ . Lemma 4.11 implies that

$$\begin{aligned} & q(\Gamma(0, 0))\eta(B(0, r_0) \setminus B(0, \lambda_0 r_0)) \\ &= \int_{\Gamma(0, 0)} \eta(B(0, r_0)) - \eta(B(0, \lambda_0 r_0)) \, dq(x) \\ &\leq \int_{\Gamma(0, 0)} \left( \eta(B(0, r_0)) - \eta(\hat{B}_x^{-1}(B(0, r_0))) \right) \, dq(x) \end{aligned}$$

Since  $\eta$  is stationary, the last expression coincides with

$$\int_{\mathcal{X} \setminus \Gamma(0, 0)} \left( \eta(\hat{B}_x^{-1}(B(0, r_0))) - \eta(B(0, r_0)) \right) \, dq(x),$$

which is, clearly, bounded above by  $\eta(B(0, \varepsilon^{-1}) \setminus B(0, r_0))$ . In this way, using (27), we find that

$$q(\Gamma(0, 0))\eta(B(0, r_0) \setminus B(0, \lambda_0 r_0)) \leq \eta(B(0, \varepsilon^{-1}) \setminus B(0, r_0)) \leq \delta.$$

Recall that  $q(\Gamma(0, 0)) \geq \beta_0/2$ . Then, using (27) once more,

$$\eta(B(0, \varepsilon^{-1}) \setminus B(0, \lambda_0 r_0)) \leq \delta + 2\delta\beta_0^{-1}.$$

Arguing by induction we get that

$$\eta(B(0, \varepsilon^{-1}) \setminus B(0, \lambda_0^j r_0)) \leq \delta(1 + 2\beta_0^{-1})^j \quad \text{for every } j \geq 0.$$

In particular, this holds for  $j = \kappa N$ . Hence, cf. Lemma 4.11,

$$\eta(B(0, \varepsilon^{-1}) \setminus B(0, \sigma^{2\tau N} r_0)) \leq \delta(1 + 2\beta_0^{-1})^{\kappa N}. \quad (28)$$

Denote  $\mathcal{B}_0 = B(0, \sigma^{2\tau N} r_0)$ . From Proposition 4.10 we get that  $\mathcal{B}_0$  and its pre-image under  $\hat{B}_x$  are disjoint for every  $x \in Y$ . So, (28) implies

$$\eta(\hat{B}_x^{-1}(\mathcal{B}_0)) \leq \delta(1 + 2\beta_0^{-1})^{\kappa N} \quad \text{for every } x \in Y. \quad (29)$$

Since  $\eta$  is stationary,

$$\begin{aligned} q(Y)\eta(\mathcal{B}_0) &= \int_{\mathcal{X} \setminus Y} \eta(\hat{B}_x^{-1}(\mathcal{B}_0) \setminus \mathcal{B}_0) \, dq(x) + \int_Y \eta(\hat{B}_x^{-1}(\mathcal{B}_0)) \, dq(x) \\ &\leq \int_{\mathcal{X} \setminus Y} \eta(B(0, \varepsilon^{-1}) \setminus \mathcal{B}_0) \, dq(x) + \int_Y \eta(\hat{B}_x^{-1}(\mathcal{B}_0)) \, dq(x) \end{aligned}$$

Recall that  $q(Y) \geq \alpha/2$ . Hence, using (28) and (29),

$$\eta(\mathcal{B}_0) \leq 4\delta\alpha^{-1}(1 + 2\beta_0^{-1})^{\kappa N}. \quad (30)$$

Adding (28) and (30) we conclude that

$$\eta(B(0, \varepsilon^{-1})) \leq \tilde{c}\delta, \quad \tilde{c} = (1 + 4\alpha^{-1})(1 + 2\beta_0^{-1})^{\kappa N}. \quad (31)$$

So far we have been assuming that the fixed point sits at  $z_0 = 0$ . Let us now explain how this assumption can be removed. Notice that for every  $x \in X_0$  the matrix  $A_x$  is diagonal, its larger eigenvalue is far from the unit circle, and the corresponding eigenvector is horizontal. Thus, an attracting fixed point  $z_0 \in B(0, 1)$  as in (26) must be close to zero (in other words, the direction it represents is close to horizontal) if the cocycle  $B$  is close to  $A$ . Define

$$H = \begin{pmatrix} a_0 & -b_0 \\ b_0 & a_0 \end{pmatrix},$$

where  $(a_0, b_0) \approx (1, 0)$  be a unit vector in the direction represented by  $z_0$ , and then consider the cocycle  $C$  defined by  $C_x = H B_x H^{-1}$ . Clearly,  $\hat{B}_x^{-1}(z_0) = z_0$  translates to  $\hat{C}_x^{-1}(0) = z_0$ . Moreover, if  $\eta$  is  $(B, q)$ -stationary then  $H_*\eta$  is  $(C, q)$ -stationary. Thus, we can use the arguments in the previous paragraph to conclude that

$$H_*\eta(B(0, 2\varepsilon^{-1})) \leq \tilde{c}\delta.$$

Finally,  $H(B(0, \varepsilon^{-1})) \subset B(0, 2\varepsilon^{-1})$  because  $H$  is close to the identity, and so it follows that

$$\eta(B(0, \varepsilon^{-1})) \leq \tilde{c}\delta \quad (32)$$

also in this case. One can easily dispose of the factor  $\tilde{c}$ . So, the proof of Proposition 4.2 in the reducible case is complete.

Now, we assume that the cocycle is “irreducible”, in the sense that  $p(\Gamma(z, 0)) < \beta_0$  for all  $z \in B(0, 1)$ . We need the following lemma:

**Lemma 4.12.** *There exists  $\gamma > 0$  such that if  $d(A, B) < \gamma$  and  $p(\Gamma(z, 0)) < \beta_0$  for all  $z \in B(0, 1)$  then for each (small)  $\varsigma > 0$  there exist  $z_0 \in B(0, 1)$  and  $\rho_0 > 0$  such that*

- (a)  $p(\Gamma(z, \rho_0)) \leq p(\Gamma(z_0, \rho_0)) + \varsigma$  for all  $z \in B(0, 1)$ ;
- (b)  $\beta_0/4 \leq p(\Gamma(z_0, \rho_0)) \leq \beta_0$ ;
- (c)  $p(X_0 \setminus \Gamma(z_0, c\sigma^{-2\tau N}\rho_0)) \geq \beta_0/2$ .

*Proof.* Let  $\varrho = \inf\{r > 0 : p(\Gamma(z, r)) > \beta_0 \text{ for some } z \in B(0, 1)\}$ . We claim that  $\varrho > 0$ . Indeed, suppose that for each  $n \in \mathbb{N}$  there exists  $z_n \in B(0, 1)$  such that  $p(\Gamma(z_n, 1/n)) > \beta_0$ . We may suppose that  $(z_n)_n$  converges to some  $\tilde{z} \in B(0, 1)$ . Then  $p(\Gamma(\tilde{z}, r)) > \beta_0$  for any  $r > 0$ , and so  $p(\Gamma(\tilde{z}, 0)) \geq \beta_0$ . The latter contradicts the hypothesis, and so our claim is proved. Now, define  $\rho_0 = 9\varrho/10$  and let

$$S = \sup\{p(\Gamma(z, \rho_0)) : z \in B(0, 1)\}.$$

Notice that  $S \leq \beta_0$ , because  $\rho_0 < \varrho$ . We claim that  $S > \beta_0/4$ . Indeed, by the definition of  $\varrho$ , one may find  $z \in B(0, 1)$  such that  $p(\Gamma(z, 11\varrho/10)) > \beta_0$ . It is easy to check that  $\Gamma(z, 11\varrho/10)$  may be covered with not more than four sets  $p(\Gamma(z', \rho_0))$ ,  $z' \in B(0, 1)$ . Then,  $p(\Gamma(z', \rho_0)) > \beta_0/4$  for some choice of  $z'$ , and that proves the claim. Now, given any small  $\varsigma > 0$ , take  $z_0 \in B(0, 1)$  such that  $p(\Gamma(z_0, \rho_0)) + \varsigma > S$ . Properties (a) and (b) follow immediately from the previous considerations. We are left to prove (c). Clearly, one can find  $G \subset \mathbb{C}$  with  $\#G \leq 4c^2\sigma^{-4\tau N}$  such that

$\{\Gamma(z, \rho_0) : z \in G\}$  covers  $\Gamma(z_0, c\sigma^{-2\tau N}\rho_0)$ . Consequently, since the supremum  $S \leq \beta_0$ .

$$\begin{aligned} \mu(X_0 \setminus \Gamma(z_0, c\sigma^{-2\tau N}\rho_0)) &\geq p(X_0) - \sum_{z \in G} \mu(\Gamma(z, \rho_0)) \\ &\geq \alpha_0 - 4c^2\sigma^{-4\tau N}\beta_0, \end{aligned}$$

Now notice that  $\beta_0$  was defined in such a way that this last expression is equal to  $\beta_0/2$ . This completes the proof of the lemma.  $\square$

Let  $z_0$  and  $\rho_0 > 0$  be as given by Lemma 4.12, for some sufficiently small  $\varsigma > 0$ . For the same reasons as in the reducible case, it is no restriction to suppose that  $z_0 = 0$ . Define  $X_1 = X_0 \setminus \Gamma(0, c\sigma^{-2\tau N}\rho_0)$ . By parts (c) and (d) of Lemma 4.12,

$$p(\Gamma(0, \rho_0)) \geq \beta_0/4 \quad \text{and} \quad p(X_1) \geq \beta_0/2. \quad (33)$$

Suppose, by contradiction, that  $\eta(B(0, \varepsilon^{-1})) > \delta$ . Then take  $r_0 \in (0, 1)$  as in Corollary 4.9:

$$\eta(B(0, \varepsilon^{-1}) \setminus B(0, r_0)) \leq \delta \quad \text{and} \quad p(Y_0) \geq \alpha, \quad (34)$$

where  $Y_0 = \{x \in \mathcal{X} : \hat{B}_x^{-1}(B(0, r_0)) \not\subset \hat{D}_x^{-1}(B(0, r_0))\}$ . Let  $r_1 = \max\{r_0, c\sigma^{-2\tau N}\rho_0\}$  and

$$Y_1 = \{x \in \mathcal{X} : \hat{B}_x^{-1}(B(0, r_1)) \not\subset \hat{D}_x^{-1}(B(0, r_1))\}.$$

Lemma 4.12 and (34) imply

$$\eta(B(0, \varepsilon^{-1}) \setminus B(0, r_1)) \leq \delta \quad \text{and} \quad p(Y_1) \geq \beta_1 \quad (35)$$

where  $\beta_1 = \min\{\alpha, \beta_0/2\}$ . Indeed, the first claim in (35) is a direct consequence of (34). For the second claim there are two cases. If  $r_1 = r_0$  then  $Y_1 = Y_0$  and so (34) yields  $p(Y_1) \geq \alpha$ . If  $r_1 = c\sigma^{-2\tau N}\rho_0$  we may use Lemma 4.12(c), together with the observation that

$$Y_1^c = \{x \in \mathcal{X} : \hat{B}_x^{-1}(B(0, r_1)) \subset \hat{D}_x^{-1}(B(0, r_1))\} \subset \Gamma(0, r_1),$$

to conclude that  $p(Y_1) \geq \beta_0/2$ . This establishes (35). It follows that

$$q(\Gamma(0, \rho)) > \beta_0/8 \quad \text{and} \quad q(Y_1) > \beta_1/2,$$

as long as  $d(p, q)$  is sufficiently small. Lemma 4.11 implies that

$$\begin{aligned} q(\Gamma(0, \rho_0))\eta\left(B(0, r_1) \setminus B(0, \lambda_0 r_1)\right) \\ &= \int_{\Gamma(0, \rho_0)} \left(\eta(B(0, r_1)) - \eta(B(0, \lambda_0 r_1))\right) dq(x) \\ &= \int_{\Gamma(0, \rho_0)} \left(\eta(B(0, r_1)) - \eta(\hat{B}_x^{-1}(B(0, r_1)))\right) dq(x) \end{aligned}$$

Since  $\eta$  is stationary, the last expression coincides with

$$\int_{\mathcal{X} \setminus \Gamma(0, \rho_0)} \left(\eta(\hat{B}_x^{-1}(B(0, r_1))) - \eta(B(0, r_1))\right) dq(x),$$

which is, clearly, bounded above by  $\eta(B(0, \varepsilon^{-1}) \setminus B(0, r_1))$ . Using (35), it follows that

$$q(\Gamma(0, \rho))\eta\left(B(0, r_1) \setminus B(0, \lambda_0 r_1)\right) \leq \delta.$$

Then, using (35) once more,

$$\eta(B(0, \varepsilon^{-1}) \setminus B(0, \lambda_0 r_1)) \leq \delta(1 + 8\beta_0^{-1}) \quad (36)$$

Arguing by induction we get that

$$\eta(B(0, \varepsilon^{-1}) \setminus B(0, \lambda_0^j r_1)) \leq \delta(1 + 8\beta_0^{-1})^j \quad \text{for every } j \geq 0$$

(the cases  $j = 0$  and  $j = 1$  are given by (35) and (36), respectively). In particular, this holds for  $j = \kappa N$ . Hence, cf. Lemma 4.11,

$$\eta(B(0, \varepsilon^{-1}) \setminus B(0, \sigma^{2\tau N} r_1)) \leq \delta(1 + 8\beta_0^{-1})^{\kappa N}. \quad (37)$$

Denote  $\mathcal{B}_1 = B(0, \sigma^{2\tau N} r_1)$ . From Proposition 4.10 we get that  $\mathcal{B}_1$  and its pre-image under  $\hat{B}_x$  are disjoint for every  $x \in Y_1$ . So, (37) implies

$$\eta(\hat{B}_x^{-1}(\mathcal{B}_1)) \leq \delta(1 + 8\beta_0^{-1})^{\kappa N} \quad \text{for every } x \in Y_1. \quad (38)$$

Since  $\eta$  is stationary,

$$\begin{aligned} q(Y_1)\eta(\mathcal{B}_1) &= \int_{\mathcal{X} \setminus Y_1} \eta(\hat{B}_x^{-1}(\mathcal{B}_1) \setminus \mathcal{B}_1) dq(x) + \int_{Y_1} \eta(\hat{B}_x^{-1}(\mathcal{B}_1)) dq(x) \\ &\leq \int_{\mathcal{X} \setminus Y_1} \eta(B(0, \varepsilon^{-1}) \setminus \mathcal{B}_1) dq(x) + \int_{Y_1} \eta(\hat{B}_x^{-1}(\mathcal{B}_1)) dq(x) \end{aligned}$$

Combining  $q(Y_1) \geq \beta_1/2$  with (37) and (38), we find that

$$\eta(\mathcal{B}_1) \leq 4\delta\beta_1^{-1}(1 + 8\beta_0^{-1})^{\kappa N}. \quad (39)$$

Adding (37) and (39) we conclude that

$$\eta(B(0, \varepsilon^{-1})) \leq \tilde{c}\delta, \quad \tilde{c} = (1 + 4\beta_1^{-1})(1 + 8\beta_0^{-1})^{\kappa N}. \quad (40)$$

That completes the proof in the irreducible case, under the assumption that  $z_0 = 0$ . This assumption can be removed in just the same way as before in the reducible case, and so our argument is complete.

## 5. MAIN ESTIMATES

All we have to do to finish the proof of Proposition 4.2 is to prove Propositions 4.8 and 4.10.

**5.1. Mass away from the vertical.** In this section we prepare the proof of Proposition 4.8. Let  $\sigma < 1$  be as in Lemma 4.3. For each  $\mathcal{K} \subset \mathcal{X}$  consider the associated cocycle  $K$ , as defined in (23). Clearly,

$$\hat{K}_x(B(0, r\sigma^{2j})) = B(0, r\sigma^{2j+2s_x}) \quad (41)$$

for every  $r > 0$ ,  $x \in \mathcal{X}$ , and  $j \in \mathbb{Z}$ . Define

$$I_j(r) = B(0, r\sigma^{2j-2}) \setminus B(0, r\sigma^{2j}) \quad (42)$$

for  $j \in \mathbb{Z}$  and

$$L_x(r) = \begin{cases} B(0, r) \setminus \hat{K}_x^{-1}(B(0, r)) & \text{for } x \in \mathcal{X}_- \\ \hat{K}_x^{-1}(B(0, r)) \setminus B(0, r) & \text{for } x \in \mathcal{X}_+ \end{cases} \quad (43)$$

where  $\mathcal{X} = \mathcal{X}_- \cup \mathcal{X}_+$  denotes the partition associated to the cocycle  $K$ , that is, such that  $k_x > 1$  for  $x \in \mathcal{X}_-$  and  $k_x < 1$  for  $x \in \mathcal{X}_+$ . Notice that  $\mathcal{X} \setminus \mathcal{K} \subset \mathcal{X}_+$  because  $k_x = \sigma^\tau$  for all  $x \in \mathcal{X} \setminus \mathcal{K}$ .

**Lemma 5.1.** *If  $r > 0$  is  $(B, q, \eta)$ -targeted with respect to  $K$  then*

- (1)  $\int_{\mathcal{X}_-} \eta(L_x(r)) dq(x) \leq \int_{\mathcal{X}_+} \eta(L_x(r)) dq(x)$
- (2)  $\int_{\mathcal{X}_-} \sum_{j=1}^{-s_x} \eta(I_j(r)) dq(x) \leq \int_{\mathcal{X}_+} \sum_{j=-s_x+1}^0 \eta(I_j(r)) dq(x)$ .

More generally, given  $n \geq 0$ , if  $r\sigma^{2t}$  is  $(B, q, \eta)$ -targeted with respect to  $K$  for  $t = 0, \dots, n$ , then

$$\int_{\mathcal{X}_-} \sum_{j=t+1}^{t-s_x} \eta(I_j(r)) dq(x) \leq \int_{\mathcal{X}_+} \sum_{j=t-s_x+1}^t \eta(I_j(r)) dq(x) \text{ for } t = 0, \dots, n.$$

*Proof.* Denote  $J = B(0, r)$ . Using that  $r$  is  $(B, q, \eta)$ -centered and  $\eta$  is  $(B, q)$ -stationary

$$\int \left( \eta(J) - \eta(\hat{K}_x^{-1}(J)) \right) dq(x) \leq \int \left( \eta(J) - \eta(\hat{B}_x^{-1}(J)) \right) dq(x) = 0.$$

By the definition (43), the left hand side coincides with

$$\int_{\mathcal{X}_-} \eta(L_x(r)) dq(x) - \int_{\mathcal{X}_+} \eta(L_x(r)) dq(x).$$

This proves the first claim. The second one is a direct consequence: just note that, by (41),

$$L_x(r) = \begin{cases} B(0, r) \setminus B(0, r\sigma^{-2s_x}) = \bigsqcup_{j=1}^{-s_x} I_j(r) & \text{for } x \in \mathcal{X}_- \\ B(0, r\sigma^{-2s_x}) \setminus B(0, r) = \bigsqcup_{j=-s_x+1}^0 I_j(r) & \text{for } x \in \mathcal{X}_+. \end{cases}$$

The last claim follows, noticing  $I_j(r\sigma^{2t}) = I_{j+t}(r)$  for all  $j$  and  $r$ .  $\square$

Let  $\alpha$  and  $\gamma$  be the constants in Lemma 4.4.

**Corollary 5.2.** *If  $p(\mathcal{K}) \geq 1 - \alpha$  and  $r > 0$  is  $(B, q, \eta)$ -targeted with respect to  $K$  then*

$$\eta(I_1(r)) \leq \tilde{\alpha}^{-1} \sum_{j=-n_++1}^0 \eta(I_j(r)), \quad \text{where } n_+ = \sup_{x \in \mathcal{X}_+} |s_x|.$$

*Proof.* Lemma 4.4 gives that  $p(\mathcal{X}_-) \geq \tilde{\alpha}$ . Part (2) of Lemma 5.1 implies

$$p(\mathcal{X}_-) \eta(I_1(r)) \leq p(\mathcal{X}_+) \sum_{j=-n_x+1}^0 \eta(I_j(r))$$

The conclusion of the corollary follows, immediately.  $\square$

**Remark 5.3.** If  $r$  is  $B$ -centered then the conclusions of Lemma 5.1 and Corollary 5.2 hold for every  $(B, q)$ -stationary measure  $\eta$ .

**Lemma 5.4.** *There exists  $\gamma > 0$  such that if  $d(A, B) < \gamma$  and  $r \in [0, 1]$  and  $x \in \mathcal{X}$  are such that  $\hat{B}_x^{-1}(B(0, r)) \cap B(0, r) \neq \emptyset$ , then*

$$\hat{B}_x^{-1}(B(0, r)) \cup B(0, r) \subset \hat{D}_{sp}^{-1}(B(0, r)).$$

*Proof.* Take  $\gamma > 0$  such that if  $d(A, B) < \gamma$  then the diameter of  $\hat{B}_x^{-1}(B(0, r))$  is less than  $3\|A\|^2 r$ , for every  $r \in [0, 1]$  and  $x \in \mathcal{X}$ . Then,  $\hat{B}_x^{-1}(B(0, r)) \cap B(0, r) \neq \emptyset$  implies

$$\hat{B}_x^{-1}(B(0, r)) \cup B(0, r) \subset B(0, 4\|A\|^2 r) \subset \hat{D}_{sp}^{-1}(B(0, r)).$$

This proves the claim.  $\square$

We also need the following calculus result. In the application, for proving Proposition 4.8, we will take  $n_x = |s_x|$  and  $a_j = \eta(I_j(r))$ .

**Lemma 5.5.** *Let  $(n_x)_{x \in \mathcal{X}}$  be a bounded family of positive integers and  $(a_j)_{j \in \mathbb{Z}}$  be a sequence of non-negative real numbers. Assume that*

$$(a) \quad 0 < S \leq \int_{\mathcal{X}_-} n_x dq(x) - \int_{\mathcal{X}_+} n_x dq(x) \text{ and}$$

$$(b) \quad \int_{\mathcal{X}_-} \sum_{j=t+1}^{t+n_x} a_j dq(x) \leq \int_{\mathcal{X}_+} \sum_{j=t-n_x+1}^t a_j dq(x) \text{ for } t = 0, \dots, n.$$

Denote  $n_- = \sup\{n_x : x \in \mathcal{X}_-\}$  and  $n_+ = \sup\{n_x : x \in \mathcal{X}_+\}$ . Then

$$\sum_{j=1}^n a_j \leq \left( \frac{n_- + n_+}{S} \right) \sum_{j=-n_+}^0 a_j.$$

*Proof.* Begin by noting that

$$\sum_{t=0}^n \sum_{j=t+1}^{t+n_x} a_j = \sum_{l=1}^{n_x} \sum_{j=l}^{n+l} a_j \geq n_x \left( \sum_{j=1}^n a_j - \sum_{j=1}^{n_x} a_j \right) \quad (44)$$

and that

$$\sum_{t=0}^n \sum_{j=t-n_x+1}^t a_j = \sum_{l=-n_x+1}^0 \sum_{j=l}^{n+l} a_j \leq n_x \left( \sum_{j=1}^n a_j + \sum_{j=-n_x+1}^0 a_j \right) \quad (45)$$

So, adding the inequalities (b) over all  $t = 0, \dots, n$  and using (44)-(45),

$$\int_{\mathcal{X}_-} n_x \left[ \sum_{j=1}^n a_j - \sum_{j=1}^{n_x} a_j \right] dq(x) \leq \int_{\mathcal{X}_+} n_x \left[ \sum_{j=1}^n a_j + \sum_{j=-n_x+1}^0 a_j \right] dq(x)$$

or, equivalently,

$$S \sum_{j=1}^n a_j \leq \int_{\mathcal{X}_-} n_x \sum_{j=1}^{n_x} a_j dq(x) + \int_{\mathcal{X}_+} n_x \sum_{j=-n_x+1}^0 a_j dq(x)$$

This implies, using the inequality (b) once more,

$$\begin{aligned} S \sum_{j=1}^n a_j &\leq n_- \int_{\mathcal{X}_-} \sum_{j=1}^{n_x} a_j dq(x) + n_+ \int_{\mathcal{X}_+} \sum_{j=-n_x+1}^0 a_j dq(x) \\ &\leq (n_- + n_+) \int_{\mathcal{X}_+} \sum_{j=-n_x+1}^0 a_j dq(x). \end{aligned}$$

This last expression is bounded above by  $(n_- + n_+) \sum_{j=-n_+}^0 a_j$ . In this way we get the conclusion of the lemma.  $\square$

Define  $\alpha_s = \sum_{j=(s-1)n_+}^{sn_+} a_j$  for each  $s \geq 0$ . In the same setting as Lemma 5.5, we obtain

**Corollary 5.6.** *Let  $n = s_0 n_+$  for some integer  $s_0 \geq 1$ . There is  $s \in \{1, \dots, s_0\}$  such that*

$$\alpha_s \leq \left( \frac{n_- + n_+}{s_0 S} \right) \alpha_0.$$

*Proof.* The conclusion of Lemma 5.5 may be rewritten

$$\sum_{s=1}^{s_0} \alpha_s = \sum_{j=1}^n a_j \leq \left( \frac{n_- + n_+}{S} \right) \sum_{j=-n_+}^0 a_j = \left( \frac{n_- + n_+}{S} \right) \alpha_0.$$



This implies that  $\min_{1 \leq s \leq s_0} \alpha_j \leq (n_- + n_+) \alpha_0 / (s_0 S)$ , as claimed.  $\square$

**5.2. Proof of Proposition 4.8.** The claim will follow from applying Lemma 5.5 and Corollary 5.6 to appropriate data. As before, let  $K$  be the cocycle and  $\mathcal{X} = \mathcal{X}_+ \cup \mathcal{X}_-$  be the partition associated to a given set  $\mathcal{K} \subset \mathcal{X}$ . We break the presentation of the proof into three steps:

**Step 1:** Define  $S(p) = \int_{\mathcal{X}_-} n_x dp(x) - \int_{\mathcal{X}_+} n_x dp(x)$ , where

$$n_x = \frac{|\log k_x|}{|\log \sigma|} = \begin{cases} |s_x| & \text{if } x \in \mathcal{K} \\ \tau & \text{if } x \in \mathcal{X} \setminus \mathcal{K}. \end{cases}$$

Let  $r_0 \geq 0$  be such that every  $r \in [r_0, \varepsilon^{-1}]$  is  $(B, \mathcal{K})$ -centered with respect to  $D$ . Take  $\mathcal{K}$  to have been chose such that  $p(\mathcal{K}) \geq 1 - \alpha$ . Then, by Lemma 4.4,

$$S(p) = \int_{\mathcal{X}} -\frac{\log k_x}{\log \sigma} dp(x) = \int_{\mathcal{X}} \frac{\log k_x}{|\log \sigma|} dp(x) > \frac{2}{k |\log \sigma|} > 0.$$

Consequently, there exist  $\gamma > 0$  and  $S > 0$  such that  $S(q) > S$  for every  $q$  with  $d(p, q) < \gamma$ . This corresponds to condition (a) in Lemma 5.5. Given  $\varepsilon > 0$  and  $\delta > 0$ , let  $n = s_0 n_+ = s_0 \tau$  for some integer

$$s_0 \geq \left( \frac{n_- + n_+}{S} \right)^2 \delta^{-1}.$$

Fix also  $R > \sigma^{-2n} \varepsilon^{-1}$ . By Lemma 4.6, there exists  $\gamma > 0$  such that if  $d(A, B) < \gamma$  then every  $r \in [(R\sigma^{-2})^{-1}, R\sigma^{-2}]$  is  $B$ -centered with respect to  $D$ . This applies to  $y\sigma^{2j}$  for every  $j = 0, 1, \dots, n$  and any  $y \in [R, R\sigma^{-2}]$ , because  $y\sigma^{2j} > \varepsilon^{-1} > (R\sigma^{-2})^{-1}$ . Fix  $y \in [R, R\sigma^{-2}]$  and define  $a_j(y) = \eta(I_j(y))$  for  $j \in \mathbb{Z}$ . Then Lemma 5.1 gives

$$\int_{\mathcal{X}_-} \sum_{j=t+1}^{t+n_x} a_j(y) dq(x) \leq \int_{\mathcal{X}_+} \sum_{j=t-n_x+1}^t a_j(y) dq(x)$$

for all  $t = 0, \dots, n$ . This corresponds to condition (b) in Lemma 5.5. Thus, we are in a position to apply Corollary 5.6: we conclude that there exists  $s \in \{1, \dots, s_0\}$  such that

$$\alpha_s(y) \leq \left( \frac{n_- + n_+}{s_0 S} \right) \alpha_0(y) \leq \left( \frac{S}{n_- + n_+} \right) \delta \alpha_0(y) \leq \left( \frac{S}{n_- + n_+} \right) \delta \quad (46)$$

Notice that, by definition,

$$\alpha_s(y) = \eta(B(0, y\sigma^{2(s-1)n_+}) \setminus B(0, y\sigma^{2sn_+})).$$

**Step 2:** Fix  $r_1 \geq 1$  such that every  $r \in [r_1 \sigma^{-2}, 1]$  is  $(B, q, \eta)$ -targeted with respect to  $K$  and either  $r_1 \leq r_0$  or  $r_1$  is not  $(B, q, \eta)$ -targeted. We are going to estimate  $\eta(B(0, \varepsilon^{-1}) \setminus B(0, r_1))$ , with the aid of Lemma 5.5. Condition (a) in the lemma is just the same as before. Concerning condition (b), notice that  $y \in [R, R\sigma^{-2}]$  above may always be chosen so that  $r_1 = z\sigma^{2\bar{n}}$  for some  $\bar{n} \in \mathbb{N}$ , where  $z = y\sigma^{2sn_+}$ . Then  $z\sigma^{2t}$  is  $(B, q, \eta)$ -targeted for every  $t = 0, 1, \dots, \bar{n} - 1$ , due to our choice of  $r_1$ , and so Lemma 5.1 gives

$$\int_{\mathcal{X}_-} \sum_{j=t+1}^{t+n_x} a_j(z) dq(x) \leq \int_{\mathcal{X}_+} \sum_{j=t-n_x+1}^t a_j(z) dq(x)$$

for all  $t = 0, \dots, \bar{n} - 1$ . Thus, applying Lemma 5.5,

$$\sum_{j=1}^{\bar{n}} a_j(z) \leq \left( \frac{n_- + n_+}{S} \right) \sum_{j=-n_++1}^0 a_j(z).$$

The left hand side coincides with (recall that  $z \geq R\sigma^{2n} > \varepsilon^{-1}$ )

$$\eta(B(0, z) \setminus B(0, z\sigma^{2\bar{n}-2})) \geq \eta(B(0, \varepsilon^{-1}) \setminus B(0, r_1\sigma^{-2})).$$

The sum on the right hand side coincides with

$$\eta(B(0, z\sigma^{-2n_+}) \setminus B(0, z)) = \alpha_0(z) = \alpha_s(y).$$

Consequently,

$$\eta(B(0, \varepsilon^{-1}) \setminus B(0, r_1\sigma^{-2})) \leq \left( \frac{n_- + n_+}{S} \right) \alpha_s(y). \quad (47)$$

The relations (46) and (47) yield

$$\eta(B(0, \varepsilon^{-1}) \setminus B(0, r_1\sigma^{-2})) \leq \delta. \quad (48)$$

Moreover, Corollary 5.2 gives that

$$\begin{aligned} \eta(I_1(r_1\sigma^{-2})) &\leq \tilde{\alpha}^{-1} \sum_{j=-n_++1}^0 a_j(r_1\sigma^{-2}) \\ &\leq \tilde{\alpha}^{-1} \eta(B(0, \varepsilon^{-1}) \setminus B(0, r_1\sigma^{-2})) \leq \tilde{\alpha}^{-1} \delta \end{aligned} \quad (49)$$

Combining (48) and (49) we obtain

$$\eta(B(0, \varepsilon^{-1}) \setminus B(0, r_1)) \leq (1 + \tilde{\alpha}^{-1})\delta. \quad (50)$$

**Step 3:** If  $r_1 \leq r_0$  then (50) implies the conclusion of the proposition (the factor on the right hand side can be avoided replacing  $\delta$  by a convenient multiple throughout the argument). Otherwise,  $r_1$  is not  $(B, q, \eta)$ -targeted with respect to  $K$ , and so we must have

$$\eta(\hat{B}_x^{-1}B(0, r_1)) > \eta(\hat{K}_x^{-1}B(0, r_1)) \quad (51)$$

for some  $x \in \mathcal{X}$ . Notice that  $x$  must belong to  $\mathcal{X} \setminus \mathcal{K}$ , since

$$\eta(\hat{B}_y^{-1}B(0, r_1)) \leq \eta(\hat{D}_y^{-1}B(0, r_1)) = \eta(\hat{K}_y^{-1}B(0, r_1))$$

for every  $y \in \mathcal{K}$ , because  $r_1$  is  $(B, \mathcal{K})$ -centered. Then (51) becomes

$$\eta(\hat{B}_x^{-1}B(0, r_1)) > \eta(\hat{D}_{sp}^{-1}B(0, r_1)).$$

It follows, using Lemma 5.4, that  $B(0, r_1)$  and  $\hat{B}_x^{-1}(B(0, r_1))$  are disjoint. By (50), this implies that

$$\eta(\hat{B}_x^{-1}(B(0, r_1))) \leq (1 + \tilde{\alpha}^{-1})\delta.$$

Moreover,  $\eta(B(0, r_1)) \leq \eta(\hat{D}_{sp}^{-1}B(0, r_1)) \leq \eta(\hat{B}_x^{-1}B(0, r_1))$  and so the previous relation implies that

$$\eta(B(0, r_1)) \leq (1 + \tilde{\alpha}^{-1})\delta.$$

Using (50) once more, we conclude that  $\eta(B(0, \varepsilon^{-1})) \leq 2(1 + \tilde{\alpha}^{-1})\delta$ . This implies the conclusion of the proposition (as before, the factor on the right hand side can be avoided replacing  $\delta$  by a convenient multiple throughout the argument), and so the proof of the proposition is complete.

**5.3. Mass close to the vertical.** Next, we are going to prepare the proof of Proposition 4.10. A Möbius transformation  $h$  is a  $\gamma_0$ -deformation of  $f(z) = \lambda z$  if  $h(z) = (az + b)/(cz + d)$  for some choice of the coefficients satisfying

$$\max \left\{ \left| \frac{a}{\lambda} \right| - 1, |b|, |c|, |d| - 1 \right\} < \gamma_0.$$

**Lemma 5.7.** *Given  $\beta_0, \sigma_0 \in (0, 1)$  there are  $\gamma_0 > 0$  and  $N_0 \in \mathbb{N}$  such that for any  $f(z) = \lambda z$  and  $g(z) = \Lambda z$  with  $|\lambda| \leq \beta_0 |\Lambda|$  and  $|\Lambda| \leq \sigma_0$ , and for any  $\gamma_0$ -deformation  $\tilde{f}$  of  $f$ , we have*

$$\tilde{f}(g^{N_0}(B(0, r))) \cap g^{N_0}(B(0, r)) = \emptyset.$$

for any  $r \in [0, 1]$  such that  $\tilde{f}(B(0, r)) \not\subset g(B(0, r))$ .

*Proof.* Fix  $N_0 \in \mathbb{N}$  such that

$$|\Lambda|^{N_0-1} \leq \sigma_0^{N_0-1} \leq \frac{1 - \beta_0}{100} \quad (52)$$

and  $\gamma_0 > 0$  given by

$$\gamma_0 = \frac{1 - \beta_0}{100} < \frac{1}{100} \quad (53)$$

Write  $\tilde{f}(z) = (az + b)/(cz + d)$ . If  $\tilde{f}(0) = 0$  then  $b = 0$  and (53) gives

$$|\tilde{f}(z)| \leq \frac{|az|}{|d| - |c|} \leq \frac{1 + \gamma_0}{1 - 2\gamma_0} |\lambda z| \leq \beta_0^{-1} |\lambda z| \leq |\Lambda z|$$

whenever  $|z| \leq 1$ . This means that  $\tilde{f}(B(0, r)) \subset g(B(0, r))$  for all  $r \leq 1$ , in which case we have nothing to do. So, let us suppose that  $b \neq 0$ . Take

$$r_0 = \frac{10|b|}{|\Lambda|(1 - \beta_0)}.$$

Then  $|\tilde{f}(z)| \leq |\Lambda z|$  for every  $|z| \in [r_0, 1]$ . Indeed,

$$|\tilde{f}(z)| \leq \frac{|az| + |b|}{|d| - |c|} \leq \frac{|\lambda|(1 + \gamma_0) + |\Lambda|(1 - |\beta_0|)/10}{1 - 2\gamma_0} |z|$$

and, in view of (53), the right hand side is bounded by

$$\frac{\beta_0(1 + \gamma_0) + 10\gamma_0}{1 - 2\gamma_0} |\Lambda z| \leq \frac{\beta_0 + 20\gamma_0}{1 - 2\gamma_0} |\Lambda z| \leq |\Lambda z|.$$

This gives that  $\tilde{f}(B(0, r)) \subset g(B(0, r))$  for every  $r \in [r_0, 1]$ . Now consider  $r \in [0, r_0]$ . By (52),

$$|\Lambda|^{N_0} r \leq |\Lambda|^{N_0} r_0 \leq \frac{|\Lambda|(1 - \beta_0)}{100} \frac{10|b|}{|\Lambda|(1 - \beta_0)} \leq \frac{|b|}{10} \leq \frac{|b|}{5|d|}$$

and that means that

$$g^{N_0}(B(0, r)) \subset B\left(0, \frac{|b|}{5|d|}\right). \quad (54)$$

The relation (53) also leads to

$$|\tilde{f}'(z)| \leq \frac{|ad| + |bc|}{(|d| - |cz|)^2} \leq \frac{|\lambda|(1 + \gamma_0)^2 + (\gamma_0|\lambda|)^2}{(1 - 2\gamma_0)^2}$$

for all  $|z| \leq 1$ . Hence, using (53) once more,

$$|\tilde{f}'(z)| \leq \frac{1 + 4\gamma_0}{1 - 4\gamma_0} |\lambda| \leq \beta_0^{-1} |\lambda| \leq |\Lambda| \leq 1$$

That implies

$$\tilde{f}(B(0, \frac{|b|}{5|d|})) \subset B(\frac{b}{d}, \frac{|b|}{5|d|}) \quad (55)$$

From (54) and (55) we get that  $g^{N_0}(B(0, r)) \cap \tilde{f}(g^{N_0}(B(0, r))) = \emptyset$  for all  $r \in [0, r_0]$ . This completes the proof of the lemma.  $\square$

**Lemma 5.8.** *Given  $\beta_0, \sigma_0 \in (0, 1)$  there exist  $\gamma_0 > 0$  and  $N_0 \in \mathbb{N}$  such that for any  $f(z) = \lambda z$  and  $g(z) = \Lambda z$  with  $|\lambda| \leq \beta_0 |\Lambda|$  and  $|\lambda| \leq \sigma_0$ , and for any  $\tilde{g}$  whose inverse is a  $\gamma_0$ -deformation of  $g$ , we have*

$$f^{N_0}(B(0, r)) \cap \tilde{g}^{-1}(f^{N_0}(B(0, r))) = \emptyset$$

for any  $r \in [0, 1]$  such that  $\tilde{g}^{-1}(B(0, r)) \not\subset f^{-1}(B(0, r))$ .

*Proof.* Fix  $N_0 \geq 1$  such that

$$|\lambda|^{N_0} \leq \sigma_0^{N_0} \leq \frac{1 - \beta_0}{100}. \quad (56)$$

Fix  $\gamma_0 > 0$  such that

$$\gamma_0 = \frac{1 - \beta_0}{100} < \frac{1}{100}. \quad (57)$$

Write  $\tilde{g}(z) = (az + b)/(cz + d)$ . Suppose that  $b = 0$ . The assumption that  $\tilde{g}^{-1}$  is a  $\gamma_0$ -deformation of  $g^{-1}$ , together with (57), gives

$$|\tilde{g}^{-1}(z)| \leq \frac{|dz|}{|a| - |c|} \leq \frac{1 + \gamma_0}{1 - 2\gamma_0} |\Lambda^{-1}z| \leq \beta_0 |\Lambda^{-1}z| \leq |\lambda^{-1}z|$$

whenever  $|z| \leq 1$ . This means that  $\tilde{g}^{-1}(B(0, r)) \subset f^{-1}(B(0, r))$  for every  $r \leq 1$ , in which case there is nothing to do. Now, let us suppose that  $b \neq 0$ . Take

$$r_0 = \frac{10|b||\lambda|}{(1 - \beta_0)}.$$

Then  $|\tilde{g}^{-1}(z)| \leq |\lambda^{-1}z|$  for every  $|z| \in [r_0, 1]$ . Indeed,

$$|\tilde{g}^{-1}(z)| \leq \frac{|dz| + |b|}{|a| - |c|} \leq \frac{|\Lambda^{-1}|(1 + \gamma_0) + |(1 - |\beta_0|)/(10|\lambda|)}{1 - 2\gamma_0} |z|$$

and, in view of (57), the right hand side is bounded by

$$\frac{\beta_0(1 + \gamma_0) + 10\gamma_0}{1 - 2\gamma_0} |\lambda^{-1}z| \leq \frac{\beta_0 + 20\gamma_0}{1 - 2\gamma_0} |\lambda^{-1}z| \leq |\lambda^{-1}z|.$$

This means that  $\tilde{g}^{-1}(B(0, r)) \subset f^{-1}(B(0, r))$  for every  $r \in [r_0, 1]$ . Now consider  $r \in [0, r_0]$ . By (56)

$$|\lambda|^{N_0} r \leq |\lambda|^{N_0} r_0 \leq \frac{(1 - \beta_0)}{100} \frac{10|b||\lambda|}{(1 - \beta_0)} \leq \frac{|b\lambda|}{10} \leq \frac{|b\lambda|}{5|a|}$$

and that means that

$$f^{N_0}(B(0, r)) \subset B(0, \frac{|b\lambda|}{5|a|}) \subset B(0, \frac{|b|}{5|a|}). \quad (58)$$

Recalling that  $|\lambda| \leq \min\{1, |\Lambda|\}$ , the relation (57) also gives

$$|(\tilde{g}^{-1})'(z)| \leq \frac{|ad| + |bc|}{(|a| - |c|)^2} \leq \frac{|\Lambda^{-1}|(1 + \gamma_0)^2 + \gamma_0^2}{(1 - 2\gamma_0)^2} \leq 2|\lambda^{-1}|$$

for all  $|z| \leq 1$ . This implies

$$\tilde{g}^{-1}\left(B\left(0, \frac{|b\lambda|}{5|a|}\right)\right) \subset B\left(\frac{b}{a}, \frac{2|b|}{5|a|}\right). \quad (59)$$

From (58) and (59) we get that  $f^{N_0}(B(0, r)) \cap \tilde{g}^{-1}(f^{N_0}(B(0, r))) = \emptyset$  for every  $r \in [0, r_0]$ . This completes the proof of the lemma.  $\square$

**5.4. Proof of Proposition 4.10.** If  $d(A, B) < \gamma$  then every  $\hat{B}_x^{-1}$  is a  $(C\gamma)$ -deformation of  $f = \hat{A}_x^{-1}$ , where the constant  $C = \sup_{x \in \mathcal{X}} |\theta_x|$  depends only on  $A$ . Indeed,

$$B_x = \begin{pmatrix} a_x & b_x \\ c_x & d_x \end{pmatrix} \quad \text{with} \quad |a_x - \theta_x|, |b_x|, |c_x|, |d_x - \theta_x^{-1}| < \gamma$$

yields

$$\hat{B}_x^{-1} = \frac{d_x z - b_x}{-c_x z + a_x} = \frac{d_x \theta_x^{-1} z - b_x \theta_x^{-1}}{-c_x \theta_x^{-1} z + a_x \theta_x^{-1}}$$

with  $|d_x \theta_x^{-1} - \theta_x^{-2}|, |b_x \theta_x^{-1}|, |c_x \theta_x^{-1}|, |a_x \theta_x^{-1} - 1| < \gamma |\theta_x|^{-1} \leq C\gamma |\theta_x|^{-2}$ . Take

$$f = \hat{A}_x^{-1} \quad \text{and} \quad g = \hat{D}_x^{-1} \quad \text{for each } x \in \mathcal{X}_-.$$

Observe that  $f(z) = |\theta_x|^{-2}|z|$  and  $g(z) = \sigma_x^{-2}|z|$ . Since  $\sigma_x \leq \beta|\theta_x|$  and  $\sigma_x \geq \sigma^{-1}$  (see Lemma 4.3), we may apply Lemma 5.7 with  $\beta_0 = \beta^2$  and  $\sigma_0 = \sigma^2$ . Using also the observation in the previous paragraph, we get that there exist  $\gamma_- > 0$  and  $N_- \in \mathbb{N}$  such that

$$\hat{B}_x^{-1}(\hat{D}_x^{-N_-}(B(0, r))) \cap \hat{D}_x^{-N_-}(B(0, r)) = \emptyset \quad \text{for } x \in \mathcal{X}_-$$

if  $d(A, B) < \gamma_-$  and  $r \in [0, 1]$  is such that  $\hat{B}_x^{-1}(B(0, r)) \not\subset \hat{D}_x^{-1}(B(0, r))$ . Now take

$$f = \hat{D}_x \quad \text{and} \quad g = \hat{A}_x \quad \text{for each } x \in \mathcal{X}_+.$$

Then  $f(z) = \sigma_x^2|z|$  and  $g(z) = |\theta_x|^2|z|$  and so we are in the setting of Lemma 5.8, with  $\beta_0 = \beta^2$  and  $\sigma_0 = \sigma^2$ . In this way we find  $\gamma_+ > 0$  and  $N_+ \in \mathbb{N}$  such that

$$\hat{B}_x^{-1}(\hat{D}_x^{N_+}(B(0, r))) \cap \hat{D}_x^{N_+}(B(0, r)) = \emptyset \quad \text{for } x \in \mathcal{X}_+$$

if  $d(A, B) < \gamma_+$  and  $r \in [0, 1]$  is such that  $\hat{B}_x^{-1}(B(0, r)) \not\subset \hat{D}_x^{-1}(B(0, r))$ . To complete the proof of the proposition, just take  $\gamma = \min\{\gamma_-, \gamma_+\}$  and  $N = \max\{N_-, N_+\}$ .

## 6. CONSEQUENCES OF THEOREM C

In this section we deduce Theorem B and Theorem D.

**6.1. Proof of Theorem B.** The main step is the following lemma. Let  $\lambda$  be the Lebesgue measure on the unit interval  $I$ , and let  $\|\eta\|$  denote the total variation of a signed measure  $\eta$ .

**Lemma 6.1** (Avila). *Let  $X$  be a metric space such that every bounded closed subset is compact, and let  $\nu$  be any Borel probability measure in  $X$  whose support  $Z = \text{supp } \nu$  is bounded.*

*For every  $\varepsilon > 0$  there is  $\delta > 0$  and a weak\* neighborhood  $V$  of  $\nu$  such that every probability  $\mu \in V$  whose support is contained in  $B_\delta(Z)$  may be written as  $\phi_* q = \mu$  for some probability  $q$  on  $Z \times I$  satisfying  $\|q - (\nu \times \lambda)\| < \varepsilon$  and some measurable map  $\phi : Z \times I \rightarrow X$  such that  $d(\phi(x, t), x) < \varepsilon$  for all  $x \in Z$ .*

*Proof.* We claim that for any  $\delta > 0$  there exists a cover  $\mathcal{Q}$  of  $B_\delta(Z)$  by disjoint measurable sets  $Q_i$ ,  $i = 1, \dots, n$  with  $\nu(Q_i) > 0$  and  $\nu(\partial Q_i) = 0$  and  $\text{diam } Q_i < 12\delta$ . This can be seen as follows. For each  $x \in Z$  take  $r_x \in (\delta, 2\delta)$  such that  $\nu(\partial B(x, r_x)) = 0$ . Then  $\{B(x, r_x) : x \in Z\}$  is a cover of the closure of  $B_\delta(Z)$ , a bounded closed set. Let  $\{V_1, V_2, \dots, V_k\}$  be a finite subcover. By construction,  $\text{diam } V_i < 4\delta$  and  $\nu(V_i) > 0$  and  $\nu(\partial V_i) = 0$  for every  $i$ . Consider the partition  $\mathcal{P}$  of  $\cup_{i=1}^k V_i$  into the sets  $V_1^* \cap \dots \cap V_k^*$ , where each  $V_i^*$  is either  $V_i$  or its complement. Define

$$Q_1 = V_1 \cup \{P \in \mathcal{P} : \nu(P) = 0 \text{ and } P \subset V_i \text{ with } V_i \cap V_1 \neq \emptyset\}.$$

Then define  $Q_2 \subset X$  as follows. If  $V_2 \subset Q_1$  then  $Q_2 = \emptyset$ ; otherwise, notice that  $\nu(V_2 \setminus Q_1) > 0$ , and then take

$$Q_2 = V_2 \cup \{P \in \mathcal{P} : \nu(P) = 0 \text{ and } P \subset V_i \text{ with } V_i \cap V_2 \neq \emptyset\} \setminus Q_1$$

More generally, for every  $2 \leq l \leq k$ , assume that  $Q_1, \dots, Q_{l-1}$  have been defined and then let  $Q_l = \emptyset$  if  $V_l \subset \cup_{i=1}^{l-1} Q_i$  and

$$Q_l = V_l \cup \{P \in \mathcal{P} : \nu(P) = 0 \text{ and } P \subset V_i \text{ with } V_i \cap V_l \neq \emptyset\} \setminus \cup_{i=1}^{l-1} Q_i$$

if  $\nu(V_l \setminus \cup_{i=1}^{l-1} Q_i) > 0$ . Those of these sets  $Q_i$  that are non-empty form a cover  $\mathcal{Q}$  as in our claim.

Proceeding with the proof of the lemma, take  $\delta = \varepsilon/12$  and assume that the neighborhood  $V$  is small enough that

$$\sum_{i=1}^n |\mu(Q_i) - \nu(Q_i)| < \varepsilon \quad \text{for every } \mu \in V.$$

Let  $Z_i = \text{supp } \nu \cap Q_i$  for each  $i = 1, \dots, n$ . Clearly,  $\nu(Z_i) = \nu(Q_i)$ . Let  $q$  be the measure on  $Z \times I$  that coincides with

$$\frac{\mu(Q_i)}{\nu(Q_i)} (\nu \times \lambda)$$

restricted to each  $Z_i \times I$ . For each  $i$ , let  $a_{i,j}$ ,  $j \in J(i)$  be the atoms of  $\mu$  contained in  $Q_i$  (the set  $J(i)$  may be empty). Moreover, let  $I_{i,j}$ ,  $j \in J(i)$  be disjoint subsets of  $I$  such that

$$\lambda(I_{i,j}) = \frac{p_{i,j}}{\mu(Q_i)} \quad \text{for all } j \in J(i),$$

where  $p_{i,j} = \nu(a_{i,j})$ . Denote  $I_i = I \setminus \cup_{j \in J(i)} I_{i,j}$ . Then

$$q(Z_i \times I_i) = \mu(Q_i) - \sum_{j \in J(i)} p_{i,j} = \mu(Q_i \setminus \{a_{i,j} : j \in J(i)\}).$$

The assumption implies that  $X$  is a polish space, that is, a complete separable metric space. Since all Borel non-atomic probabilities on polish spaces are isomorphic (see Ito [19, § 2.4]), the previous equality ensures that there exists an invertible measurable map

$$\phi_i : Z_i \times I_i \rightarrow Q_i \setminus \{a_{i,j} : j \in J(i)\}$$

mapping the restriction of  $q$  to the restriction of  $\mu$ . By setting  $\phi \equiv a_{i,j}$  on each  $Z_i \times I_{i,j}$  we extend  $\phi_i$  to a measurable map  $Z_i \times I \rightarrow Q_i$  that still sends the restriction of  $q$  to the restriction of  $\mu$ . Gluing all these extensions we obtain a measurable map  $\phi : Z \times I \rightarrow X$  such that  $\phi_* q = \mu$ . By construction,  $\phi(x, t) \in Q_i$

for every  $x \in Z_i$  and  $t \in I$ . This implies that  $d(\phi(x, t), x) \leq \text{diam } Q_i < \varepsilon$  for all  $(x, t) \in Z \times I$ . Finally,

$$\begin{aligned} \|q - (\nu \times \lambda)\| &= \sum_{i=1}^n \left\| \left( \frac{\mu(Q_i)}{\nu(Q_i)} - 1 \right) (\nu \times \lambda) \Big|_{(Z_i \times I)} \right\| \\ &= \sum_{i=1}^n |\mu(Q_i) - \nu(Q_i)| < \varepsilon. \end{aligned}$$

The proof of the lemma is complete.  $\square$

Now, given  $\rho > 0$ , let  $\nu$  be a probability measure in  $\text{GL}(2, \mathbb{C})$  with compact support. Consider  $\mathcal{X} = \text{supp } \nu \times I$ ,  $p = \nu \times \lambda$  and  $A : \mathcal{X} \rightarrow \text{GL}(2, \mathbb{C})$  given by  $A(x, t) = x$ . From Theorem C, there is  $\varepsilon > 0$  such that  $|\lambda_{\pm}(A, p) - \lambda_{\pm}(B, q)| < \rho$  for all  $(B, q)$  such that  $d(p, q) < \varepsilon$  and  $d(A, B) < \varepsilon$ . On the other hand, Lemma 6.1 implies that there exist a weak\* neighborhood  $V$  and  $\delta$  such that if  $\nu' \in V$  and  $\text{supp } \nu' \subset B_{\delta}(\text{supp } \nu)$  then there exist  $B : \mathcal{X} \rightarrow \text{GL}(2, \mathbb{C})$  and a probability measure  $q$  on  $\mathcal{X}$  such that  $d(p, q) < \varepsilon$ ,  $d(A, B) < \varepsilon$  and  $\nu' = B_*q$ . Noting that  $\lambda_{\pm}(\nu) = \lambda_{\pm}(A, p)$  and  $\lambda_{\pm}(\nu') = \lambda_{\pm}(B, q)$ , we obtain Theorem B.

**6.2. Proof of Theorem D.** Recall that we denote  $M = \mathcal{X}^{\mathbb{Z}}$ .

**Lemma 6.2.** *Let  $(\mu^k)_k$  be a sequence of probabilities on  $M$  converging to  $\mu$  in the weak\* topology. Let  $(m^k)_k$  be a sequence of probabilities on  $M \times \mathbb{P}(\mathbb{C}^2)$  projecting down to  $(\mu^k)_k$ . Then there exists a subsequence of  $(m^k)_k$  converging, in the weak\* topology, to some probability  $m$  that projects down to  $\mu$ .*

*In particular, the space  $\mathcal{M}(p)$  of probability measures on  $M \times \mathbb{P}(\mathbb{C}^2)$  that project down to  $\mu$  is compact for the weak\* topology.*

*Proof.* Since  $M$  and  $M \times \mathbb{P}(\mathbb{C}^2)$  are polish spaces, we may apply Prohorov's theorem (see Billingsley [4, § 6]) in either of these spaces: a sequence of probabilities  $(\xi^k)_k$  has weak\*-converging subsequences if and only if for each  $\varepsilon > 0$  there is a compact set  $K_{\varepsilon}$  such that  $\xi^k(K_{\varepsilon}) > 1 - \varepsilon$  for any  $k \geq 1$ . By assumption,  $(\mu^k)_k$  converges to  $\mu$  in the weak\* topology. Thus, given any  $\varepsilon > 0$ , there is some compact set  $K_{\varepsilon} \subset M$  such that  $\mu^k(K_{\varepsilon}) > 1 - \varepsilon$ , for any  $k \geq 1$ . Then  $\hat{K}_{\varepsilon} = K_{\varepsilon} \times \mathbb{P}(\mathbb{C}^2)$  is compact and  $m^k(\hat{K}_{\varepsilon}) = \mu^k(K_{\varepsilon}) > 1 - \varepsilon$  for any  $k \geq 1$ . This ensures that  $(m^k)_k$  has weak\*-converging subsequences, as claimed. Considering the special case when the sequence  $(\mu^k)_k$  is constant equal to  $\mu$ , one gets the last part of the lemma.  $\square$

**Lemma 6.3.** *Let  $A^k : \mathcal{X} \rightarrow \text{GL}(2, \mathbb{C})$ ,  $k \geq 1$  be such that  $d(A^k, A) \rightarrow 0$ , and let  $F_{A^k} : M \times \mathbb{P}(\mathbb{C}^2) \rightarrow M \times \mathbb{P}(\mathbb{C}^2)$ ,  $k \geq 1$  be the associated projective cocycles. Let  $(m^k)_k$  be a sequence of probability measures on  $M \times \mathbb{P}(\mathbb{C}^2)$  such that  $m^k$  projects down to  $\mu$  for all  $k$ . If  $(m^k)_k$  converges to  $m$ , then  $((F_{A^k})_* m^k)_k$  converges to  $(F_A)_* m$ , in the weak\* topology.*

*Proof.* Let  $\varphi : M \times \mathbb{P}(\mathbb{C}^2) \rightarrow \mathbb{R}$  be any uniformly continuous bounded function. By the theorem of Lusin, given any  $\varepsilon > 0$  there is some compact set  $K \subset M$  such that  $\mu(K) > 1 - \varepsilon$  and  $A : M \rightarrow \text{SL}(2, \mathbb{C})$  is continuous restricted to  $K$ . Then  $\varphi \circ F_A : K \times \mathbb{P}(\mathbb{C}^2) \rightarrow \mathbb{R}$  is continuous and so, by the extension theorem of Tietze, it admits some continuous extension  $\tilde{\varphi} : M \times \mathbb{P}(\mathbb{C}^2) \rightarrow \mathbb{R}$  to the whole space, with

$\|\tilde{\varphi}\| \leq \|\varphi\|$ . We have

$$\begin{aligned} & \left| \int \varphi d(F_{A^k})_* m^k - \int \varphi d(F_A)_* m \right| = \left| \int \varphi \circ F_{A^k} dm^k - \int \varphi \circ F_A dm \right| \\ & \leq \left| \int \varphi \circ F_{A^k} dm^k - \int \varphi \circ F_A dm^k \right| + \left| \int \varphi \circ F_A dm^k - \int \varphi \circ F_A dm \right| \end{aligned}$$

The first term on the right hand side converges to zero when  $k \rightarrow \infty$ , because  $\varphi \circ F_{A^k}$  converges uniformly to  $\varphi \circ F_A$ . The last term admits the following bound:

$$\begin{aligned} & \left| \int \varphi \circ F_A dm^k - \int \varphi \circ F_A dm \right| \\ & \leq \left| \int \tilde{\varphi} dm^k - \int \tilde{\varphi} dm \right| + 2\|\varphi\|(m^k + m)(K^c \times \mathbb{P}(\mathbb{C}^2)) \\ & \leq \left| \int \tilde{\varphi} dm^k - \int \tilde{\varphi} dm \right| + 4\|\varphi\|\mu(K^c) \end{aligned}$$

The first term on the right hand side converges to zero when  $k \rightarrow \infty$ , because  $\tilde{\varphi}$  is continuous, and the second term is bounded by  $4\|\varphi\|\varepsilon$ . Since  $\varepsilon > 0$  is arbitrary, this proves that

$$\int \varphi d(F_{A^k})_* m^k \rightarrow \int \varphi d(F_A)_* m \quad \text{as } k \rightarrow \infty,$$

for any uniformly continuous bounded function  $\varphi$ . So (see Theorem 2.1 in Billingsley [4]), the sequence  $(F_{A^k})_* m^k$  converges weakly\* to  $(F_A)_* m$ , as claimed.  $\square$

**Corollary 6.4.** *Suppose that  $\lambda_+(A, p) > 0$  and let  $m^u$  be the  $u$ -state defined by (12). Let  $(A^k)_k$  be such that  $d(A^k, A) \rightarrow 0$  as  $k \rightarrow \infty$ . For each  $k \geq 1$ , let  $m_k^u$  be an invariant  $u$ -state for  $(A^k, p)$  realizing  $\lambda_+(A^k, p)$ . Then  $(m_k^u)_k$  converges to  $m^u$  in the weak\* topology.*

*Proof.* In view of the compactness Lemma 6.2, we only have to show that every accumulation point  $m$  of the sequence  $(m_k^u)_k$  coincides with  $m^u$ . Restricting to a subsequence if necessary, we may suppose that  $m$  is the limit of  $(m_k^u)_k$ , not just an accumulation point. We claim that  $m$  is an  $F$ -invariant probability. By definition, every  $m_k^u$  projects down to  $\mu$ . Then, we may use Lemma 6.3 to conclude that  $(F_{A^k})_* m_k^u$  converges to  $(F_A)_* m$  as  $k \rightarrow \infty$ . Since each  $m_k^u$  is assumed to be  $F_{A^k}$ -invariant, this proves that  $(F_A)_* m = m$ , as claimed. The assumption implies that  $(\phi_{A^k})_k$  converges to  $\phi_A$ , uniformly on  $M \times \mathbb{P}(\mathbb{C}^2)$ . Consequently,

$$\int \phi_A dm = \lim \int \phi_{A^k} dm_k^u.$$

In addition, using Theorem C:

$$\lim \int \phi_{A^k} dm_k^u = \lim \lambda_+(A^k, p) = \lambda_+(A, p) = \int \phi_A dm^u$$

This proves that  $m$  realizes  $\lambda_+(A, p)$ . From Remark 3.8 we conclude that  $m = m^u$ . This completes the proof of the corollary.  $\square$

Let us deduce Theorem D. We only have to show that

$$\mu(\{x \in M : \angle(E_{A,x}^u, E_{A^k,x}^u) < \varepsilon\}) \rightarrow 1 \quad \text{when } k \rightarrow \infty,$$

as the statement about stable spaces  $E_A^s$  is analogous. By Theorem C, the assumption  $\lambda_+(A, p) > 0$  implies  $\lambda_+(A^k, p) > 0$  for every large  $k$ . Let  $m^u$  and  $m_k^u$ ,  $k \geq 1$  be



$u$ -states for  $A$  and  $A^k$ ,  $k \geq 1$  defined as in (12). By Corollary 6.4,  $(m_k^u)_k$  converges to  $m^u$  in the weak\* topology. The map  $\psi : M \rightarrow \mathbb{P}(\mathbb{C}^2)$ ,  $\psi(x) = E_{A,\mathbf{x}}^u$  is measurable map and its graph has full  $m^u$ -measure. By the theorem of Lusin, given any  $\varepsilon > 0$  we may find a compact set  $K \subset M$  such that the restriction  $\psi_K$  to  $K$  is continuous and the  $m^u$ -measure of its graph is bigger than  $1 - \varepsilon$ . Given  $\delta > 0$ , let  $V$  be an open neighborhood of the graph of  $\psi_K$  such that

$$V \cap (K \times \mathbb{P}(\mathbb{C}^2)) \subset V_\delta := \{(\mathbf{x}, \xi) \in K \times \mathbb{P}(\mathbb{C}^2) : \angle(\xi, \psi(\mathbf{x})) < \delta\}.$$

By the definition of  $m_k^u$ ,

$$m_k^u(V_\delta) = \mu(\{\mathbf{x} \in K : d(E_{A^k,\mathbf{x}}^u, E_{A,\mathbf{x}}^u) < \delta\}).$$

By weak\* convergence,  $\liminf m_k^u(V) \geq m^u(V) \geq 1 - \varepsilon$ . On the other hand,  $m_k^u(K \times \mathbb{P}(\mathbb{C}^2)) = \mu(K) \geq 1 - \varepsilon$  for every  $k$ . Thus,

$$m_k^u(V_\delta) \geq m_k^u(V \cap (K \times \mathbb{P}(\mathbb{C}^2))) \geq 1 - 3\varepsilon$$

for every large  $k$ . Hence,  $\mu(\{\mathbf{x} \in M : d(E_{A^k,\mathbf{x}}^u, E_{A,\mathbf{x}}^u) < \delta\}) \geq 1 - 3\varepsilon$  for every large  $k$ . Since  $\delta$  and  $\varepsilon$  are arbitrary, this proves Theorem D.

## 7. CONCLUDING REMARKS

We are going to describe a construction of points of discontinuity of the Lyapunov exponents as functions of the cocycle, relative to some Hölder topology. This builds on and refines [5, 6, 8, 29], where it is shown that Lyapunov exponents are often discontinuous relative to the  $C^0$  topology. In the final section we list a few related open problems.

**7.1. An example of discontinuity.** Let  $M = \Sigma_2$  be the shift with 2 symbols, endowed with the metric  $d(\mathbf{x}, \mathbf{y}) = 2^{-N(\mathbf{x}, \mathbf{y})}$ , where

$$N(\mathbf{x}, \mathbf{y}) = \sup\{n \geq 0 : x_n = y_n \text{ whenever } |n| < N\}.$$

For any  $r \in (0, \infty)$ , the  $C^r$  norm in the space of  $r$ -Hölder continuous functions  $L : M \rightarrow \mathcal{L}(\mathbb{C}^d, \mathbb{C}^d)$  is defined by

$$\|L\|_r = \sup_{\mathbf{x} \in M} \|L(\mathbf{x})\| + \sup_{\mathbf{x} \neq \mathbf{y}} \frac{\|L(\mathbf{x}) - L(\mathbf{y})\|}{d(\mathbf{x}, \mathbf{y})^r}.$$

Consider on  $M$  the Bernoulli measure associated to any probability vector  $(p_1, p_2)$  with positive entries and  $p_1 \neq p_2$ . Given any  $\sigma > 1$ , consider the (locally constant) cocycle  $A : M \rightarrow \text{SL}(2, \mathbb{R})$  defined by

$$A(\mathbf{x}) = \begin{pmatrix} \sigma & 0 \\ 0 & \sigma^{-1} \end{pmatrix} \quad \text{if } x_0 = 1$$

and

$$A(\mathbf{x}) = \begin{pmatrix} \sigma^{-1} & 0 \\ 0 & \sigma \end{pmatrix} \quad \text{if } x_0 = 2.$$

**Theorem 7.1.** *For any  $r > 0$  such that  $2^{2r} < \sigma$  there exist  $B : M \rightarrow \text{SL}(2, \mathbb{R})$  with vanishing Lyapunov exponents and such that  $\|A - B\|_r$  is arbitrarily close to zero.*

Since the Lyapunov exponents  $\lambda_{\pm}(A) = \pm|p_1 - p_2| \log \sigma$  of  $A : M \rightarrow \text{SL}(2, \mathbb{R})$  are non-zero, it follows that  $A$  is a point of discontinuity for the Lyapunov exponents relative to the  $C^r$  topology.

The proof of Theorem 7.1 is an adaptation of ideas of Knill [25] and Bochi [5, 6]. Here is an outline. Notice that the unperturbed cocycle  $A$  preserves both the horizontal line bundle  $H_{\mathbf{x}} = \{\mathbf{x}\} \times \mathbb{R}(1, 0)$  and the vertical line bundle  $V_{\mathbf{x}} = \{\mathbf{x}\} \times \mathbb{R}(0, 1)$ . Then, the Oseledets subspaces must coincide with  $H_{\mathbf{x}}$  and  $V_{\mathbf{x}}$  almost everywhere. We choose cylinders  $Z_n \subset M$  whose first  $n$  iterates  $f^i(Z_n)$ ,  $0 \leq i \leq n-1$  are pairwise disjoint. Then we construct cocycles  $B_n$  by modifying  $A$  on some of these iterates so that

$$B_n^n(x)H_{\mathbf{x}} = V_{f^n(\mathbf{x})} \quad \text{and} \quad B_n^n(x)V_{\mathbf{x}} = H_{f^n(\mathbf{x})} \quad \text{for all } \mathbf{x} \in Z_n.$$

We deduce that the Lyapunov exponents of  $B_n$  vanish. Moreover, by construction, each  $B_n$  is constant on every atom of some finite partition of  $M$  into cylinders. In particular,  $B_n$  is Hölder continuous for every  $r > 0$ . From the construction we also get that

$$\|B_n - A\|_r \leq \text{const} (2^{2r}/\sigma)^{n/2} \quad (60)$$

decays to zero as  $n \rightarrow \infty$ . This is how we get the claims in the theorem. Now let us fill-in the details of the proof.

Let  $n = 2k + 1$  for some  $k \geq 1$  and  $Z_n = [0; 2, \dots, 2, 1, \dots, 1, 1]$  where the symbol 2 appears  $k$  times and the symbol 1 appears  $k + 1$  times. Notice that the  $f^i(Z_n)$ ,  $0 \leq i \leq 2k$  are pairwise disjoint. Let

$$\varepsilon_n = \sigma^{-k} \quad \text{and} \quad \delta_n = \arctan \varepsilon_n. \quad (61)$$

Define  $R : M \rightarrow \text{SL}(2, \mathbb{R})$  by

$$\begin{aligned} R(\mathbf{x}) &= \text{rotation of angle } \delta_n \quad \text{if } \mathbf{x} \in f^k(Z_n) \\ R(\mathbf{x}) &= \begin{pmatrix} 1 & 0 \\ \varepsilon_n & 1 \end{pmatrix} \quad \text{if } \mathbf{x} \in Z_n \cup f^{2k}(Z_n) \\ R(\mathbf{x}) &= \text{id} \quad \text{in all other cases.} \end{aligned}$$

and then take  $B_n = AR_n$ .

**Lemma 7.2.**  $B_n^n(\mathbf{x})H_{\mathbf{x}} = V_{f^n(\mathbf{x})}$  and  $B_n^n(\mathbf{x})V_{\mathbf{x}} = H_{f^n(\mathbf{x})}$  for all  $\mathbf{x} \in Z_n$ .

*Proof.* Notice that for any  $\mathbf{x} \in Z_n$ ,

$$\begin{aligned} B_n^k(\mathbf{x})H_{\mathbf{x}} &= \mathbb{R}(\varepsilon_n, 1) \quad \text{and} \quad B_n^k(\mathbf{x})V_{\mathbf{x}} = V_{f^k(\mathbf{x})} \\ B_n^{k+1}(\mathbf{x})H_{\mathbf{x}} &= V_{f^{k+1}(\mathbf{x})} \quad \text{and} \quad B_n^{k+1}(\mathbf{x})V_{\mathbf{x}} = \mathbb{R}(-\varepsilon_n, 1) \\ B_n^{2k}(\mathbf{x})H_{\mathbf{x}} &= V_{f^{2k}(\mathbf{x})} \quad \text{and} \quad B_n^{2k}(\mathbf{x})V_{\mathbf{x}} = \mathbb{R}(-1, \varepsilon_n). \end{aligned}$$

The claim follows by iterating one more time.  $\square$

**Lemma 7.3.** *There exists  $C > 0$  such that  $\|B_n - A\|_r \leq C (2^{2r}/\sigma)^k$  for every  $n$ .*

*Proof.* Let  $L_n = A - B_n$ . Clearly,  $\sup \|L\| \leq \sup \|A\| \|\text{id} - R_n\|$  and this is bounded by  $\sigma\varepsilon_n$ . Now let us estimate the second term in the definition (60). If  $\mathbf{x}$  and  $\mathbf{y}$  are not in the same cylinder  $[0; a]$  then  $d(\mathbf{x}, \mathbf{y}) = 1$ , and so

$$\frac{\|L_n(\mathbf{x}) - L_n(\mathbf{y})\|}{d(\mathbf{x}, \mathbf{y})^r} \leq 2 \sup \|L_n\| \leq 2\sigma\varepsilon_n. \quad (62)$$

From now on we suppose  $\mathbf{x}$  and  $\mathbf{y}$  belong to the same cylinder. Then, since  $A$  is constant on cylinders,

$$\frac{\|L_n(\mathbf{x}) - L_n(\mathbf{y})\|}{d(\mathbf{x}, \mathbf{y})^r} = \frac{\|A(\mathbf{x})(R_n(\mathbf{x}) - R_n(\mathbf{y}))\|}{d(\mathbf{x}, \mathbf{y})^r} \leq \sigma \frac{\|R_n(\mathbf{x}) - R_n(\mathbf{y})\|}{d(\mathbf{x}, \mathbf{y})^r}.$$

If neither  $\mathbf{x}$  nor  $\mathbf{y}$  belong to  $Z_n \cup f^k(Z_n) \cup f^{2k}(Z_n)$  then  $R_n(\mathbf{x})$  and  $R_n(\mathbf{y})$  are both equal to  $\text{id}$ , and so the expression on the right vanishes. If  $\mathbf{x}$  and  $\mathbf{y}$  belong to the same  $f^i(Z_n)$  then  $R_n(\mathbf{x}) = R_n(\mathbf{y})$  and so, once more, the expression on the right vanishes. We are left to consider the case when one of the points belongs to some  $f^i(Z_n)$  and the other one does not. Then  $d(\mathbf{x}, \mathbf{y}) \geq 2^{-2k}$  and so, using once more that  $\|\text{id} - R_n\| \leq \varepsilon_n$  at every point,

$$\frac{\|L_n(\mathbf{x}) - L_n(\mathbf{y})\|}{d(\mathbf{x}, \mathbf{y})^r} \leq \sigma \frac{\|R_n(\mathbf{x}) - R_n(\mathbf{y})\|}{d(\mathbf{x}, \mathbf{y})^r} \leq 2\sigma\varepsilon_n 2^{2kr}.$$

Noting that this bound is worst than (62), we conclude that

$$\|L_n\|_r \leq \sigma\varepsilon_n + 2\sigma\varepsilon_n 2^{2kr} \leq 3\sigma (2^{2r}/\sigma)^k$$

Now it suffices to take  $C = 3\sigma$ .  $\square$

Now we want to prove that  $\lambda_{\pm}(B_n) = 0$  for every  $n$ . Let  $\mu_n$  be the normalized restriction of  $\mu$  to  $Z_n$  and  $f_n : Z_n \rightarrow Z_n$  be the first return map (defined on a full measure subset). Indeed,

$$Z_n = \bigsqcup_{b \in \mathcal{B}} [0; w, b, w] \quad (\text{up to a zero measure subset})$$

where  $w = (1, \dots, 1, 2, \dots, 2, 2)$  and the union is over the set  $\mathcal{B}$  of all finite words  $b = (b_1, \dots, b_s)$  not having  $w$  as a sub-word. Moreover,

$$f_n | [0; w, b, w] = f^{n+s} | [0; w, b, w] \quad \text{for each } b \in \mathcal{B}.$$

Thus,  $(f_n, \mu_n)$  is a Bernoulli shift with an infinite alphabet  $\mathcal{B}$  and probability vector given by  $p_b = \mu_n([0; w, b, w])$ . Let  $\hat{B}_n : Z_n \rightarrow \text{SL}(2, \mathbb{R})$  be the cocycle induced by  $B$  over  $f_n$ , that is,

$$\hat{B}_n | [0; w, b, w] = B_n^{n+s} | [0; w, b, w] \quad \text{for each } b \in \mathcal{B}.$$

It is a well known basic fact (see [38, Proposition 2.9], for instance) that the Lyapunov spectrum of the induced cocycle is obtained multiplying the Lyapunov spectrum of the original cocycle by the average return time. In our setting this means

$$\lambda_{\pm}(\hat{B}_n) = \frac{1}{\mu(Z_n)} \lambda_{\pm}(B_n).$$

Therefore, it suffices to prove that  $\lambda_{\pm}(\hat{B}_n) = 0$  for every  $n$ .

Indeed, suppose the Lyapunov exponents of  $\hat{B}_n$  are non-zero and let  $E_{\mathbf{x}}^u \oplus E_{\mathbf{x}}^s$  be the Oseledets splitting (defined almost everywhere in  $Z_n$ ). Consider the probability measures  $m^u$  and  $m^s$  for the cocycle  $\hat{B}_n$  defined as in (12). The key observation is that, as a consequence of Lemma 7.2, the cocycle  $\hat{B}_n$  permutes the vertical and horizontal subbundles:

$$\hat{B}_n(\mathbf{x})H_{\mathbf{x}} = V_{f_n(\mathbf{x})} \quad \text{and} \quad \hat{B}_n(\mathbf{x})V_{\mathbf{x}} = H_{f_n(\mathbf{x})} \quad \text{for all } \mathbf{x} \in Z_n. \quad (63)$$

Let  $m$  be the measure defined on  $M \times \mathbb{P}(\mathbb{R}^2)$  by

$$m_n(X) = \frac{1}{2}(\mu_n(\{\mathbf{x} \in Z_n : V_{\mathbf{x}} \in X\}) + \mu_n(\{\mathbf{x} \in Z_n : H_{\mathbf{x}} \in X\})).$$

That is,  $m_n$  projects down to  $\mu_n$  and its disintegration is given by  $\mathbf{x} \mapsto (\delta_{H_{\mathbf{x}}} + \delta_{V_{\mathbf{x}}})/2$ . It is clear from (63) that  $m_n$  is  $\hat{B}_n$ -invariant.

**Lemma 7.4.** *The probability measure  $m_n$  is ergodic.*

*Proof.* Suppose there is an invariant set  $\mathcal{X} \subset M \times \mathbb{P}(\mathbb{R}^2)$  with  $m_n(\mathcal{X}) \in (0, 1)$ . Let  $X_0$  be the set of  $\mathbf{x} \in Z_n$  whose fiber  $\mathcal{X} \cap (\{\mathbf{x}\} \times \mathbb{P}(\mathbb{R}^2))$  contains neither  $H_{\mathbf{x}}$  nor  $V_{\mathbf{x}}$ . In view of (63),  $X_0$  is an  $f_n$ -invariant set and so its  $\mu_n$ -measure is either 0 or 1. Since  $m_n(\mathcal{X}) > 0$ , we must have  $\mu_n(X_0) = 0$ . The same kind of argument shows that  $\mu_n(X_2) = 0$ , where  $X_2$  is the set of  $\mathbf{x} \in Z_n$  whose fiber contains both  $H_{\mathbf{x}}$  and  $V_{\mathbf{x}}$ . Now let  $X_H$  be the set of  $\mathbf{x} \in Z_n$  whose fiber contains  $H_{\mathbf{x}}$  but not  $V_{\mathbf{x}}$ , and let  $X_V$  be the set of  $\mathbf{x} \in Z_n$  whose fiber contains  $V_{\mathbf{x}}$  but not  $H_{\mathbf{x}}$ . The previous observations show that  $X_H \cup X_V$  has full  $\mu_n$ -measure and it follows from (63) that

$$f_n(X_H) = X_V \quad \text{and} \quad f_n(X_V) = X_H.$$

Thus,  $\mu_n(X_H) = 1/2 = \mu_n(X_V)$  and  $f_n^2(X_H) = X_H$  and  $f_n^2(X_V) = X_V$ . This is a contradiction because  $f_n$  is Bernoulli and, in particular, the second iterate is ergodic.  $\square$

By Lemma 3.6, the invariant measure  $m_n$  is a linear combination of  $m^u$  and  $m^s$ . Then, in view of Lemma 7.4,  $m_n$  must coincide with either  $m^s$  and  $m^u$ . This is a contradiction, because the conditional probabilities of  $m_n$  are supported on exactly two points on each fiber, whereas the conditional probabilities of either  $m^u$  and  $m^s$  are Dirac masses on a single point. This contradiction proves that the Lyapunov exponents of  $\hat{B}_n$  do vanish for every  $n$ , and that concludes the proof of Theorem 7.1.

The same kind of argument shows that, in general, one can expect continuity to hold when some of the probabilities  $p_i$  vanishes:

**Remark 7.5.** (Kifer [22]) Take  $d = 2$ , a probability vector  $p = (p_1, p_2)$  with non-negative coefficients, and a cocycle  $A = (A_1, A_2)$  defined by

$$A_1 = \begin{pmatrix} \sigma & 0 \\ 0 & \sigma^{-1} \end{pmatrix} \quad \text{and} \quad A_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

where  $\sigma > 1$ . By the same arguments as we used before,  $\lambda_{\pm}(A, p) = 0$  for every  $p \in \Lambda_2$ . In this regard, observe that the cocycle induced by  $A$  over the cylinder  $[0; 2]$  exchanges the vertical and horizontal directions, just as in (63). Now, it is clear that  $\lambda_{\pm}(A, (1, 0)) = \pm \log \sigma$ . Thus, the Lyapunov exponents are discontinuous at  $(A, (1, 0))$ .

## 7.2. Open problems.

**Problem 7.6.** Does continuity extend to

- (a) unbounded cocycles satisfying an integrability condition ?
- (b) locally constant cocycles over Bernoulli shifts ?
- (c) locally constant cocycles over Markov shifts ?
- (d) locally constant cocycles in any dimension  $d$  ?
- (e) Hölder continuous cocycles satisfying the fiber bunched condition ?

In (a) we have in mind the condition  $\log \|A^{\pm 1}\| \in L^1(\mu)$ . Since it involves both the cocycle and the probability measure, in this case the topology should be defined in the space of pairs  $(A, p)$ . By *locally constant* in (b) and (c) we mean that  $A(\mathbf{x})$  depends on a bounded number of coordinates of  $\mathbf{x}$ . We have treated the case when

$A(\mathbf{x})$  depends only on the zeroth coordinate of  $\mathbf{x}$ . In (d) it suffices to consider the largest Lyapunov exponent: then, using exterior powers in a well-known way (see Peres [33], for instance), one would get continuity for all Lyapunov exponents. An interesting special case to look at are symplectic cocycles, that is, such that every  $A(\mathbf{x})$  preserves some given symplectic form. See [3] for the definition of the fiber bunching condition in (e). By Theorem 7.1 we can not expect continuity to hold for general Hölder cocycles. On the other hand, the hypotheses of the theorem is incompatible with fiber bunching.

**Problem 7.7.** Can we say more about the regularity of the Lyapunov exponents as functions of the cocycle: Hölder continuity? Lipschitz continuity? Differentiability?

Partial answers and related results were obtained by Le Page [32] and Peres [33].

#### REFERENCES

- [1] A. Arbieto and J. Bochi.  $L^p$ -generic cocycles have one-point Lyapunov spectrum. *Stoch. Dyn.*, 3:73–81, 2003.
- [2] L. Arnold and N. D. Cong. On the simplicity of the Lyapunov spectrum of products of random matrices. *Ergod. Th. & Dynam. Sys.*, 17:1005–1025, 1997.
- [3] A. Avila and M. Viana. Extremal Lyapunov exponents: an invariance principle and applications. *Inventiones Math.*, 181:115–178, 2010.
- [4] P. Billingsley. *Convergence of probability measures*. John Wiley & Sons Inc., New York, 1968.
- [5] J. Bochi. Discontinuity of the Lyapunov exponents for non-hyperbolic cocycles. Preprint [www.mat.puc-rio.br/~jairo/](http://www.mat.puc-rio.br/~jairo/).
- [6] J. Bochi. Genericity of zero Lyapunov exponents. *Ergod. Th. & Dynam. Sys.*, 22:1667–1696, 2002.
- [7] J. Bochi.  $C^1$ -generic symplectic diffeomorphisms: partial hyperbolicity and zero centre Lyapunov exponents. *J. Inst. Math. Jussieu*, 8:49–93, 2009.
- [8] J. Bochi and M. Viana. The Lyapunov exponents of generic volume-preserving and symplectic maps. *Ann. of Math.*, 161:1423–1485, 2005.
- [9] C. Bonatti, L. J. Díaz, and M. Viana. *Dynamics beyond uniform hyperbolicity*, volume 102 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, 2005.
- [10] C. Bonatti, X. Gómez-Mont, and M. Viana. Généricité d'exposants de Lyapunov non-nuls pour des produits déterministes de matrices. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 20:579–624, 2003.
- [11] J. Bourgain. Positivity and continuity of the Lyapounov exponent for shifts on  $\mathbb{T}^d$  with arbitrary frequency vector and real analytic potential. *J. Anal. Math.*, 96:313–355, 2005.
- [12] J. Bourgain and S. Jitomirskaya. Continuity of the Lyapunov exponent for quasiperiodic operators with analytic potential. *J. Statist. Phys.*, 108:1203–1218, 2002.
- [13] L. Breiman. *Probability*. Addison-Wesley, 1968.
- [14] H. Furstenberg and H. Kesten. Products of random matrices. *Ann. Math. Statist.*, 31:457–469, 1960.
- [15] H. Furstenberg and Yu. Kifer. Random matrix products and measures in projective spaces. *Israel J. Math*, 10:12–32, 1983.
- [16] I. Ya. Gol'dsheid and G. A. Margulis. Lyapunov indices of a product of random matrices. *Uspekhi Mat. Nauk.*, 44:13–60, 1989.
- [17] Y. Guivarc'h and A. Raugi. Products of random matrices : convergence theorems. *Contemp. Math.*, 50:31–54, 1986.
- [18] H. Hennion. Loi des grands nombres et perturbations pour des produits réductibles de matrices aléatoires indépendantes. *Z. Wahrsch. Verw. Gebiete*, 67:265–278, 1984.
- [19] K. Itô. *Introduction to probability theory*. Cambridge University Press, 1984.
- [20] R. Johnson. Lyapounov numbers for the almost periodic Schrödinger equation. *Illinois J. Math.*, 28:397–419, 1984.
- [21] J. L. Kelley. *General Topology*. Van Norstrand, 1955.

- [22] Yu. Kifer. Perturbations of random matrix products. *Z. Wahrsch. Verw. Gebiete*, 61:83–95, 1982.
- [23] Yu. Kifer and E. Slud. Perturbations of random matrix products in a reducible case. *Ergodic Theory Dynam. Systems*, 2:367–382 (1983), 1982.
- [24] J. Kingman. The ergodic theorem of subadditive stochastic processes. *J. Royal Statist. Soc.*, 30:499–510, 1968.
- [25] O. Knill. The upper Lyapunov exponent of  $SL(2, \mathbf{R})$  cocycles: discontinuity and the problem of positivity. In *Lyapunov exponents (Oberwolfach, 1990)*, volume 1486 of *Lecture Notes in Math.*, pages 86–97. Springer, 1991.
- [26] O. Knill. Positive Lyapunov exponents for a dense set of bounded measurable  $SL(2, \mathbf{R})$ -cocycles. *Ergod. Th. & Dynam. Sys.*, 12:319–331, 1992.
- [27] F. Ledrappier. Positivity of the exponent for stationary sequences of matrices. In *Lyapunov exponents (Bremen, 1984)*, volume 1186 of *Lect. Notes Math.*, pages 56–73. Springer, 1986.
- [28] P. A. Loeb and E. Talvila. Lusin’s theorem and Bochner integration. *Scientia Mathematica Japonicae Online*, 10:55–62, 2004.
- [29] R. Mañé. Oseledec’s theorem from the generic viewpoint. In *Procs. International Congress of Mathematicians, Vol. 1, 2 (Warsaw, 1983)*, pages 1269–1276, Warsaw, 1984. PWN Publ.
- [30] V. I. Oseledets. A multiplicative ergodic theorem: Lyapunov characteristic numbers for dynamical systems. *Trans. Moscow Math. Soc.*, 19:197–231, 1968.
- [31] É. Le Page. Théorèmes limites pour les produits de matrices aléatoires. In *Probability measures on groups (Oberwolfach, 1981)*, volume 928 of *Lecture Notes in Math.*, pages 258–303. Springer, 1982.
- [32] É. Le Page. Régularité du plus grand exposant caractéristique des produits de matrices aléatoires indépendantes et applications. *Ann. Inst. H. Poincaré Probab. Statist.*, 25:109–142, 1989.
- [33] Y. Peres. Analytic dependence of Lyapunov exponents on transition probabilities. In *Lyapunov exponents (Oberwolfach, 1990)*, volume 1486 of *Lecture Notes in Math.*, pages 64–80. Springer, 1991.
- [34] V. A. Rokhlin. On the fundamental ideas of measure theory. *A. M. S. Transl.*, 10:1–52, 1952. Transl. from *Mat. Sbornik* 25 (1949), 107–150.
- [35] D. Rudolph. *Fundamentals of measurable dynamics*. Oxford Science Publications. The Clarendon Press Oxford University Press, 1990. Ergodic theory on Lebesgue spaces.
- [36] D. Ruelle. Analyticity properties of the characteristic exponents of random matrix products. *Adv. in Math.*, 32:68–80, 1979.
- [37] B. Simon and M. Taylor. Harmonic analysis on  $SL(2, \mathbf{R})$  and smoothness of the density of states in the one-dimensional Anderson model. *Comm. Math. Phys.*, 101:1–19, 1985.
- [38] M. Viana. Lyapunov exponents of Teichmüller flows. In *Partially hyperbolic dynamics, laminations, and Teichmüller flow*, volume 51 of *Fields Inst. Commun.*, pages 139–201. Amer. Math. Soc., 2007.

IMPA, EST. D. CASTORINA 110, 22460-320 RIO DE JANEIRO, RJ, BRAZIL  
*E-mail address:* bocker@impa.br, viana@impa.br