Ergod. Th. & Dynam. Sys. (2015), **0**, 1–30
doi:10.1017/etds.2015.116

Continuity of Lyapunov exponents for random 2D matrices

Q1

CARLOS BOCKER-NETO† and MARCELO VIANA‡

† *Departamento de Matemática, Universidade Federal da Paraíba, Cidade Universitária, 58051-900 João Pessoa, PB, Brazil*
(e-mail: carlos@mat.ufpb.br)

‡ *IMPA, Est. D. Castorina 110, Jardim Botânico, 22460-320 Rio de Janeiro, RJ, Brazil*
(e-mail: viana@impa.br)

(*Received* 17 *July* 2014 *and accepted in revised form* 23 *September* 2015)

Abstract. The Lyapunov exponents of locally constant GL(2, ℂ)-cocycles over Bernoulli shifts vary continuously with the cocycle and the invariant probability measure.

1. *Introduction*

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

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

where the L_j, $j \geq 0$ are independent random variables such that the probability of $\{L_j = A_i\}$ is equal to p_i for all $j \geq 0$ and $i = 1, \ldots, m$.

It is a classical fact, going back to Furstenberg and Kesten [**11**], that there exist numbers λ_+ and λ_- such that

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

almost surely. The purpose of this paper is to prove that these *extremal Lyapunov exponents* always vary continuously with the choice of the matrices and the probability weights.

THEOREM A. *The extremal Lyapunov exponents λ_+ and λ_- vary continuously with the coefficients of $(A_1, \ldots, A_m, p_1, \ldots, p_m)$ at all points.*

Actually, continuity holds much more generally: we may take the probability distribution of the random variables L_j to be any compactly supported probability measure ν on GL(2, ℂ). Let $\lambda_+(\nu)$ and $\lambda_-(\nu)$, respectively, denote the values of the (almost certain) limits in (1). Then we get the following theorem.

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

The situation in Theorem A corresponds to the special case when the measures have finite supports: $\nu = p_1\delta_{A_1} + \cdots + p_m\delta_{A_m}$ and $\nu' = p'_1\delta_{A'_1} + \cdots + p'_m\delta_{A'_m}$. Clearly, the support of ν' is Hausdorff close to the support of ν 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 (see Remark 8.5).

Although the behavior of Lyapunov exponents as functions of the defining data has been investigated by several authors, it is still far from being well understood. This is partly because this behavior is very subtle and depends in a delicate way on the precise set-up. Positive results have been obtained in some specific situations. However, Mañé [**23**] and Bochi [**5**] showed that continuity of the Lyapunov exponents is actually *rare* among continuous 2D cocycles: often, it holds only when the Lyapunov exponents vanish identically. In fact, our construction in §8 indicates that similar phenomena may occur also for more regular cocycles. A detailed discussion of these and related issues will appear in §2.3.

2. *Continuity of Lyapunov exponents*

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

2.1. *Linear cocycles.*

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

Let μ be an f-invariant probability measure on M with $\log \|(F_x)^{\pm 1}\| \in L^1(\mu)$. It follows, from the subadditive ergodic theorem [**20**], that the *extremal Lyapunov exponents*

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

are well defined μ-almost everywhere.

The theorem of Oseledets [**24**] provides a more detailed statement. Namely, at μ-almost every point $x \in M$, there exist numbers $\hat\lambda_1(F, x) > \cdots > \hat\lambda_{k(x)}(F, x)$ and linear subspaces $\mathcal{V}_x = V^1_x > V^2_x > \cdots > V^{k(x)}_x > \{0\} = V^{k(x)+1}_x$ such that

$$F_x(V^j_x) = V^j_{f(x)} \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{n} \log \|F^n_x(v)\| = \hat\lambda_j(F, x) \quad \text{for all } v \in V^j_x \setminus V^{j+1}_x.$$

When f is invertible one can say more: at μ-almost every $x \in M$ there exists a splitting $\mathcal{V}_x = E^1_x \oplus E^2_x \oplus \cdots \oplus E^{k(x)}_x$ such that

$$F_x(E^j_x) = E^j_{f(x)} \quad \text{and} \quad \lim_{n\to\pm\infty} \frac{1}{n} \log \|F^n_x(v)\| = \hat\lambda_j(F, x) \quad \text{for all } v \in E^j_x \setminus \{0\}.$$

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 μ-almost everywhere if μ is ergodic.

Now, 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, $d =$ dimension of $\mathcal{V}$. The *average Lyapunov exponents* of F are defined by

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

The results in this paper are motivated by the following basic question: *what are the continuity points of* $(F, \mu) \mapsto (\lambda_1(F, \mu), \ldots, \lambda_d(F, \mu))$?

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

$$(F, \mu) \mapsto \lambda_1(F, \mu) + \cdots + \lambda_k(F, \mu) \quad (\text{any } 1 \leq k < d)$$

is upper semicontinuous, relative to the L^∞-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 ψ). Indeed, this is an easy consequence of the identity

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

Q2 where Λ^k denotes the kth exterior power. Similarly, the sum of the k smallest Lyapunov exponents is always lower semicontinuous.

However, Lyapunov exponents are, usually, *discontinuous* functions of the data. A number of results, both positive and negative, will be recalled shortly. Right now, let us reformulate our main statement in this language.

2.2. *Continuity theorem.* Let X be a polish space, that is, a separable completely metrizable topological space. Let p be a probability measure on X and $A : X \to \mathrm{GL}(2, \mathbb{C})$ be a measurable bounded function: that is, such that $\log \|A^{\pm 1}\|$ are bounded. Let $f : M \to M$ be the shift map on $M = X^{\mathbb{Z}}$ (also a polish space) and let $\mu = p^{\mathbb{Z}}$. Consider the linear cocycle

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

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

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

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

$$V(A, p, \gamma, Z) = \{(B, q) : d(A, B) < \gamma, q(Z) = 1, d(p, q) < \gamma\}, \tag{2}$$

where $\gamma > 0$ and $Z \subset X$ is any measurable set with $p(Z) = 1$. We will write $V(A, p, \gamma) = V(A, p, \gamma, X)$.

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 §§3–6, and we deduce Theorem B from it in §7. 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 ν' is contained in a small neighborhood of the support of ν; moreover, $\lambda_{\pm}(A, p) = \lambda_{\pm}(\nu)$ and $\lambda_{\pm}(B, q) = \lambda_{\pm}(\nu')$. In §8 we show that locally constant cocycles may be discontinuity points for the Lyapunov exponents in the space of Hölder continuous cocycles.

It is not difficult to deduce from our arguments that the Oseledets decomposition also depends continuously on the cocycle, in the following sense. Given $B : X \to \mathrm{GL}(2, \mathbb{C})$, let $E^s_{B,\mathbf{x}}$ and $E^u_{B,\mathbf{x}}$ be the Oseledets subspaces of the corresponding cocycle at a point $\mathbf{x} \in M$ (when they exist). Assume that $\lambda_-(A, p) < \lambda_+(A, p)$. Then, for any $\varepsilon > 0$,

$$\mu(\{\mathbf{x} \in M : \angle(E^u_{A,\mathbf{x}}, E^u_{B,\mathbf{x}}) < \varepsilon \text{ and } \angle(E^s_{A,\mathbf{x}}, E^s_{B,\mathbf{x}}) < \varepsilon\}) \quad \text{is close to 1}$$

if $d(A, B)$ is close to zero. The details will not be included here.

2.3. *Related 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 [**28**] proved real-analytic dependence of the largest exponent on the cocycle, for linear cocycles admitting an invariant convex cone field. Shortly afterwards, Furstenberg and Kifer [**12**, **18**] and Hennion [**15**] proved continuity of the largest exponent of independent and identically distributed random matrices, under a condition of almost irreducibility. Some reducible cases were treated by Kifer and Slud [**18**, **19**], who also observed that discontinuities may occur when the probability vector degenerates ([**18**], see Remark 8.5 below). Stability of Lyapunov exponents under certain random perturbations was obtained by Young [**33**].

For independent and identically distributed random matrices satisfying strong irreducibility and the contraction property, Le Page [**25**, **26**] proved local Hölder continuity, and even smoothness, of the largest exponent on the cocycle; the assumptions ensure that the largest exponent is simple (multiplicity one), by work of Guivarc'h and Raugi [**14**] and Gol'dsheid and Margulis [**13**]. For independent and identically distributed random matrices over Bernoulli and Markov shifts, Peres [**27**] showed that simple exponents are locally real-analytic functions of the transition data.

A construction of Halperin quoted by Simon and Taylor [**29**] 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 independent and identically distributed random variables) near which the exponents fail to be α-Hölder continuous. Thus the previously mentioned results of Le Page cannot be improved. Johnson [**17**] found examples of discontinuous dependence of the exponent on the energy E, for Schrödinger cocycles over quasiperiodic flows. Recently, Bourgain and Jitomirskaya [**8**, **9**] proved continuous dependence of the exponents on the energy E, for one-dimensional *quasiperiodic* Schrödinger cocycles: $V_n = V(f^n(\theta))$, where $V : S^1 \to \mathbb{R}$ is real-analytic and f is an irrational circle rotation.

Going back to general linear cocycles, the answer to the continuity problem is bound to depend on the class of cocycles under consideration, including its topology. Knill [**21**, **22**] considered L^∞ cocycles with values in SL(2, $\mathbb{R}$) and proved that, as long as the base dynamics are aperiodic, discontinuities always exist: the set of cocycles with non-zero exponents is never open. This was refined to the continuous case by Bochi [**4**, **5**]: an 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 [**23**]. Indeed, and most strikingly, the theorem of Mañé and Bochi [**5**, **23**] remains true, when 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 and Viana [**6**, **7**]. 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 and Cong [**2**] and Arbieto and 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 §8 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 continuous cocycles. Altogether, our results suggest the following.

CONJECTURE. *For any $r > 0$, Lyapunov exponents always vary continuously on the realm of fiber-bunched (see* [**3**] *for the definition) r-Hölder continuous cocycles.*

Recently, Avila and Viana [**3**] studied the continuity of the Lyapunov exponents in the very broad context of *smooth* cocycles. The continuity criterion in [**3**, §6] was the starting point for the proof of our Theorem C.

3. *Proof of Theorem C*

In this section, we reduce Theorem C to a statement about the random walks induced by pairs (B, q) close to (A, p). The proof of this statement (Propositions 3.7–3.8) will be given in §6.

Let $\mathcal{P}(X)$ be the space of Borel probability measures on the polish space X and let $\mathcal{G}(X)$ and $\mathcal{S}(X)$ denote the spaces of bounded measurable functions from X to GL(2, $\mathbb{C}$) and SL(2, $\mathbb{C}$), respectively. Given any $A \in \mathcal{G}(X)$, let $B \in \mathcal{S}(X)$ and $c : X \to \mathbb{C}$ be such that $A_x = c_x B_x$ for every $x \in X$. 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 (2), continuity of the Lyapunov exponents on $\mathcal{S}(X) \times \mathcal{P}(X)$ yields continuity on the whole $\mathcal{G}(X) \times \mathcal{P}(X)$. So we may suppose from the start that $A \in \mathcal{S}(X)$. Observe also that, in this case, one has $\lambda_+(A, p) + \lambda_-(A, p) = 0$.

From here on, the proof of Theorem C has two main parts, that we present in §§3.1 and 3.2, respectively. By *point of (dis)continuity* we will mean a point of (dis)continuity for either (and, hence, both) extremal Lyapunov exponents $\lambda_\pm$.

3.1. *Non-diagonal case.* First, we reduce the problem to the case when the matrices are simultaneously diagonalizable.

PROPOSITION 3.1. *If a pair $(A, p) \in \mathcal{S}(X) \times \mathcal{P}(X)$ is a point of discontinuity, then $\lambda_+(A, p) > 0 > \lambda_-(A, p)$ and there are $P \in \mathrm{SL}(2, \mathbb{C})$ and $\theta : X \to \mathbb{C}\setminus\{0\}$ such that*

$$P A_x P^{-1} = \begin{pmatrix} \theta_x & 0 \\ 0 & \theta_x^{-1} \end{pmatrix} \quad \text{for every } x \text{ in some full } p\text{-measure set } Z \subset X.$$

Proposition 3.1 is contained in the main results of Furstenberg and Kifer [**12**] and Hennion [**15**], as well as in Avila and Viana [**3**, Proposition 6.3]. We are going to give an outline of the proof, for the reader's convenience and also because it allows us to introduce some of the ideas that will be used in the subsequent work. For the details, see the aforementioned papers or [**31**, Ch. 5].

Given (A, p) in $\mathcal{S}(X) \times \mathcal{P}(X)$, a probability measure η on $\mathbb{P}(\mathbb{C}^2)$ is called (A, p)-*stationary* if

$$\int \psi(\xi)\, d\eta(\xi) = \iint \psi(A_x \xi)\, d\eta(\xi)\, dp(x)$$

for every bounded measurable function $\psi : \mathbb{P}(\mathbb{C}^2) \to \mathbb{C}$ (note that A_x denotes both a matrix and its action on the projective space).

The set $\mathrm{Stat}(A, p)$ of (A, p)-stationary probability measures is always non-empty: that is because $\eta \mapsto \int (A_x)_* \eta\, dp(x)$ is a continuous operator in the space $\mathcal{M}$ of Borel probability measures on $\mathbb{P}(\mathbb{C}^2)$ and so, by Tychonoff–Schauder, it has some fixed point. In this regard, note that $\mathbb{P}(\mathbb{C}^2)$ is endowed with the weak* topology, which makes it compact, convex and metrizable. Another useful property is that $\mathrm{Stat}(A, p)$ varies in a semicontinuous fashion with the data (A, p).

LEMMA 3.2. *If $(A_k, p_k)_k$ converges to (A, p) in $\mathcal{S}(X) \times \mathcal{P}(X)$ and $(\eta_k)_k$ are probability measures with $\eta_k \in \mathrm{Stat}(A_k, p_k)$ for every k, then $\eta \in \mathrm{Stat}(A, p)$.*

The reason why stationary measures are useful in our context is because one can express the Lyapunov exponents in terms of these measures. For this, let us consider the function

$$\phi : M \times \mathbb{P}(\mathbb{C}^2) \to \mathbb{R}, \quad \phi(\mathbf{x}, [v]) = \log \frac{\|A_{x_0} v\|}{\|v\|}.$$

Since ϕ depends only on x_0 and $[v]$, we may also view it as a function on $X \times \mathbb{P}(\mathbb{C}^2)$.

LEMMA 3.3. *For any $(A, p) \in \mathcal{S}(X) \times \mathcal{P}(X)$,*

$$\lambda_+(A, p) = \max\left\{ \int \phi(x, \xi)\, d\eta(\xi)\, dp(x) : \eta \in \mathrm{Stat}(A, p) \right\}.$$

From Lemmas 3.2 and 3.3 one immediately gets that $(A, p) \mapsto \lambda_+(A, p)$ is upper semicontinuous, as was mentioned previously. In particular, every (A, p) such that $\lambda_\pm(A, p) = 0$ is a point of continuity.

LEMMA 3.4. *For any* $(A, p) \in \mathcal{S}(X) \times \mathcal{P}(X)$*, if* $\eta \in \mathrm{Stat}(A, p)$ *is such that*

$$\int \phi(x, \xi)\, d\eta(\xi)\, dp(x) < \lambda_+(A, p),$$

then there is $L \in \mathbb{P}(\mathbb{C}^2)$ *with* $\eta(\{L\}) > 0$ *and* $A_x L = L$ *for* p*-almost every* x *and*

$$\lim_{n\to\pm\infty} \frac{1}{n} \log \|A_x^n v\| = \lambda_-(A, p) \quad \text{for } v \in L \text{ and } p\text{-almost every } x.$$

We call a pair (A, p) *irreducible* if there exists no (A, p)*-invariant subspace*: that is, no one-dimensional subspace $L < \mathbb{C}^2$ such that $A_x L = L$ for p-almost every x. Lemmas 3.3 and 3.4 have the following immediate consequence.

COROLLARY 3.5. *If* $(A, p) \in \mathcal{S}(X) \times \mathcal{P}(X)$ *is irreducible, then*

$$\lambda_+(A, p) = \int \phi(x, \xi)\, d\eta(\xi)\, dp(x) \quad \text{for every } \eta \in \mathrm{Stat}(A, p).$$

It is easy to deduce that if (A, p) is irreducible then it is a point of continuity. Recall that we only need to consider the case when $\lambda_+(A, p) > 0 > \lambda_-(A, p)$. Let $(A_k, p_k)_k$ be any sequence converging to (A, p) in $\mathcal{S}(X) \times \mathcal{P}(X)$. By Lemma 3.3, for each k there exists some $\eta_k \in \mathrm{Stat}(A_k, p_k)$ that realizes the largest Lyapunov exponent: that is

$$\lambda_+(A_k, p_k) = \int \phi_k(x, \xi)\, d\eta_k(\xi)\, dp_k(x), \qquad \phi_k(x, [v]) = \log \frac{\|A_{k,x} v\|}{\|v\|}.$$

Up to restricting to a subsequence, we may suppose that $(\eta_k)_k$ converges to some probability η, relative to the weak* topology. Combining Lemma 3.2 and Corollary 3.5, we get that $\eta \in \mathrm{Stat}(A, p)$ and

$$\lambda_+(A, p) = \int \phi(x, \xi)\, d\eta(\xi)\, dp(x).$$

Our assumptions imply that there exists a compact set $K \subset \mathrm{GL}(2)$ that contains the supports of p and every p_k. The sequence $(\phi_k)_k$ converges to ϕ uniformly on $K \times \mathbb{P}(\mathbb{C}^2)$ and then it follows that

$$\int \phi_k(x, \xi)\, d\eta_k(\xi)\, dp_k(x) \to \int \phi(x, \xi)\, d\eta(\xi)\, dp(x).$$

This proves that $\lambda_+(A, p) = \lim_k \lambda_+(A_k, p_k)$.

Next, suppose that (A, p) admits exactly one invariant subspace L. The previous arguments remain valid, and so (A, p) is still a point of continuity, unless

$$\lim_{n\to\pm\infty} \frac{1}{n} \log \|A_x^n v\| = \lambda_-(A, p) \quad \text{for } v \in L \text{ and } p\text{-almost every } x. \tag{3}$$

Let us also consider the cocycle defined by A over the inverse f^{-1}. It is clear that the Lyapunov exponents of the two cocycles, over f and over f^{-1}, coincide. For the same reason, (A, p) is a point of continuity over f if and only if it is a point of continuity over f^{-1}. By the previous arguments applied to the cocycle over f^{-1}, this does happen unless

$$\lim_{n\to\pm\infty} \frac{1}{n} \log \|A_x^{-n} v\| = \lambda_-(A, p) \quad \text{for } v \in L \text{ and } p\text{-almost every } x. \tag{4}$$

Notice that (3) and (4) are incompatible, because $\lambda_-(A, p) \neq 0$. Thus (A, p) is still a point of continuity if it admits a unique invariant subspace.

Thus for $A(A, p)$ to be a point of discontinuity it must admit two or more invariant subspaces, precisely as stated in Proposition 3.1.

3.2. *Diagonal case.* The key point in this paper is that we are able to prove continuity in the diagonal case as well.

PROPOSITION 3.6. *If $(A, p) \in \mathcal{S}(X) \times \mathcal{P}(X)$ is as in the conclusion of Proposition 3.1, then it is a point of continuity.*

In preparation for the proof of Proposition 3.6, let us make a few observations. Since conjugacies preserve the Lyapunov exponents, it is no restriction to suppose that $P = \mathrm{id}$ and

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

We will always consider pairs $(B, q) \in V(A, p, \gamma, Z)$, that give full weight to Z. Thus it is no restriction to suppose that $Z = X$. Notice that the Lyapunov exponents of (A, p) coincide with the values of $\pm \int \log |\theta_x| \, dp(x)$ and, by assumption, they are non-zero. Up to a further conjugacy, reversing the roles of the two axes, we may suppose that

$$\lambda_+(A, p) = \int \log |\theta_x| \, dp(x) > 0. \tag{5}$$

The arguments in the previous section break down in the present context, because now there are several stationary measures, not all of which realize the largest Lyapunov exponent. Indeed, the fact that both the horizontal direction and the vertical direction are invariant under almost every A_x means that the corresponding Dirac masses, δ_h and δ_v, are both (A, p)-stationary measures. In particular, $\mathrm{Stat}(A, p)$ contains the whole line segment between these two Dirac masses (in fact, the two sets coincide).

To get continuity of the Lyapunov exponents we will have to prove the much finer fact that the stationary measures of (irreducible) nearby cocycles are close to the one element of $\mathrm{Stat}(A, p)$ that realizes the Lyapunov exponent $\lambda_+(A, p)$, namely, the Dirac mass δ_h. That is the content of the next proposition. The notion of an irreducible pair was introduced right before Corollary 3.5.

PROPOSITION 3.7. *Given $\varepsilon > 0$ and $\delta > 0$, there exists $\gamma > 0$ such that $\eta(H_\varepsilon^c) \leq \delta$ for any (B, q)-stationary measure η and any irreducible pair $(B, q) \in V(A, p, \gamma)$, where H_ε denotes the ε-neighborhood of the horizontal direction $h \in \mathbb{P}(\mathbb{C}^2)$.*

Let us check that Proposition 3.6 is a consequence. Since λ_+ is always upper semicontinuous, it suffices to show that, given $\tau > 0$, there is $\gamma > 0$ such that $\lambda_+(B, q) > \lambda_+(A, p) - 4\tau$ for every $(B, q) \in V(A, p, \gamma)$.

First, suppose that (B, q) is irreducible. Let $m = \sup_x |\log|\theta_x||$. For each $B \in \mathcal{S}(X)$, denote

$$\phi_B : X \times \mathbb{P}(\mathbb{C}^2) \to \mathbb{R}, \quad \phi_B(x, [v]) = \log \frac{\|B_x v\|}{\|v\|}.$$

Note that $\phi_A(x, h) = \log|\theta_x| \geq -m$ for every x. Then, if γ is small enough:

(1) $\phi_B(x, \xi) \geq -m - \tau$ for every (x, ξ) and every B with $d(A, B) < \gamma$;

(2) $\int \log|\theta_x|\, dq(x) \geq \int \log|\theta_x|\, dp(x) - \tau$ for every q with $d(p, q) < \gamma$; and

(3) there exists $\varepsilon > 0$ such that $\phi_B(x, \xi) \geq \log|\theta_x| - \tau$ for every (x, ξ) with $\xi \in H_\varepsilon$ and every B with $d(A, B) < \gamma$.

Fix $\delta > 0$ such that $(m + \tau)\delta < \tau$. Let η be any (B, q)-stationary measure that realizes the largest Lyapunov exponent. Proposition 3.7 gives that $\eta(H_\varepsilon^c) \leq \delta$, as long as γ is small enough. So

$$\begin{aligned}\int \phi_B(x, \xi)\, d\eta(\xi) &= \int_{H_\varepsilon} \phi_B(x, \xi)\, d\eta(\xi) + \int_{H_\varepsilon^c} \phi_B(x, \xi)\, d\eta(\xi)\\ &\geq \eta(H_\varepsilon)(\log|\theta_x| - \tau) - (m + \tau)\delta\end{aligned}$$

for every x. The choice of δ ensures that the expression on the right-hand side is bounded below by $\log|\theta_x| - 3\tau$. Integrating with respect to q, we obtain that

$$\lambda_+(B, q) \geq \int \log|\theta_x|\, dq(x) - 3\tau \geq \int \log|\theta_x|\, dp(x) - 4\tau = \lambda_+(A, p) - 4\tau.$$

This proves our claim in the irreducible case.

Now suppose that (B, q) admits some invariant one-dimensional subspace L. Observe that L must be close to either the horizontal direction or the vertical direction. Indeed, consider any $\varepsilon > 0$. The condition (5) implies that $|\theta_x| \neq 1$ for every x in some $Z \subset X$ with $p(Z) > 0$. On the one hand, $q(Z) > 0$ for any probability q such that $d(p, q)$ is small. On the other hand, if $x \in Z$ and $d(A, B)$ is small, the matrix B_x can have no invariant subspace outside the ε-neighborhoods of the horizontal and vertical axes. This justifies our observation. Then, assuming that $\varepsilon > 0$ is small enough, the Lyapunov exponent of (B, q) along the subspace L is τ-close to one of the numbers $\pm \int \log|\theta_x|\, dq(x)$ and, hence, is 2τ-close to one of the numbers $\pm \int \log|\theta_x|\, dp(x)$. This means, in other words, that either $\lambda_+(B, q)$ or $\lambda_-(B, q)$ is 2τ-close to either $\lambda_+(A, p)$ or $\lambda_-(A, p)$. Assuming that τ is small enough, this implies that $|\lambda_*(A, p) - \lambda_*(B, q)| < 2\tau$ for both $* \in \{+, -\}$. In particular, we get the claim in this case also.

This reduces Proposition 3.6 and Theorem C to Proposition 3.7. Before proceeding to prove this proposition, it is convenient to reformulate it as follows.

Let $\phi : \mathbb{P}(\mathbb{C}^2) \to \overline{\mathbb{C}}$, $\phi([z_1, z_2]) = z_1/z_2$ be the standard identification between the complex projective space and the Riemann sphere. The horizontal direction h is identified with ∞ and the vertical direction v is identified with zero. The projective action of a linear map

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

corresponds to the Möbius transformation on the sphere defined by

$$B : \overline{\mathbb{C}} \to \overline{\mathbb{C}}, \quad z \mapsto \frac{az + b}{cz + d}$$

(we will use the same notation for a linear map and the corresponding Möbius transformation). It follows that a measure η in projective space is (B, q)-stationary if and only if its image $\zeta = \phi_* \eta$ on the sphere satisfies $\zeta = \int (B_x)_* \zeta \, dq(x)$. We will say that ζ is a (B, q)-stationary measure on the sphere.

Thus Proposition 3.7 may be restated as follows.

PROPOSITION 3.8. *Given $\varepsilon > 0$ and $\delta > 0$ there is $\gamma > 0$ so that $\eta(\mathbb{D}(0, \varepsilon^{-1})) \le \delta$ for any (B, q)-stationary probability measure η on the Riemann sphere and any $(B, q) \in V(A, p, \gamma)$ such that $q(\{x \in X : B_x(z) = z\}) < 1$ for all $z \in \overline{\mathbb{C}}$.*

Here, and in what follows, $\mathbb{D}(z_0, r) = \{z \in \mathbb{C} : |z - z_0| \le r\}$. The proof of this proposition will appear in §6.

4. *Preliminaries*

In this section we collect a few simple facts that will be used in the proof of Proposition 3.8.

4.1. *Transient regime.* Since $A_x(z) = \theta_x^2 z$ for every z, the relation (5) implies that, almost surely, the orbit $A_x^n(z)$ of any $z \in \mathbb{C} \setminus \{0\}$ converges to ∞ when $n \to +\infty$ and it converges to zero when $n \to -\infty$. Consider the dynamics

$$f_A : \xi \mapsto \int (A_x)_* \xi \, dp(x)$$

induced by (A, p) in the space of the probability measures of the sphere. It follows that δ_∞ is an attractor and δ_0 is a repeller for f_A: that is

$$\lim_{n \to +\infty} f_A^n \xi \to \delta_\infty \quad \text{if } \xi(\{0\}) = 0 \quad \text{and} \quad \lim_{n \to -\infty} f_A^n \xi \to \delta_0 \quad \text{if } \xi(\{\infty\}) = 0$$

with respect to the weak* topology. In particular, every (A, p)-stationary measure must be supported on $\{0, \infty\}$.

LEMMA 4.1. *Given any $\varepsilon > 0$ and $\delta > 0$, there exists $\gamma > 0$ such that*

$$\eta(\mathbb{D}(0, \varepsilon^{-1}) \setminus \mathbb{D}(0, \varepsilon)) \le \delta$$

for every (B, q)-stationary measure η and every $(B, q) \in V(A, p, \gamma)$.

Proof. Let $Q_\varepsilon = \{z \in \mathbb{C} : \varepsilon \le |z| \le \varepsilon^{-1}\}$ and suppose that there exists a sequence (B_k, q_k) converging to (A, p) and (B_k, q_k)-stationary measures η_k such that $\eta_k(Q_\varepsilon) \ge \delta$. By compactness and Lemma 3.2, we may suppose that η_k converges to some (A, p)-stationary measure η. Since Q_ε is closed, $\eta(Q_\varepsilon) \ge \limsup \eta_k(Q_\varepsilon) \ge \delta$. This contradicts the fact that all (A, p)-stationary measures are supported on $\{0, \infty\}$. This contradiction proves that $\eta(\mathbb{D}(0, \varepsilon^{-1}) \setminus \mathbb{D}(0, \varepsilon)) \le \eta(Q_\varepsilon) \le \delta$. □

Thus, to prove Proposition 3.8, we must show that the stationary measures of irreducible cocycles near (A, p) have small mass in the neighborhood of zero. The key property that distinguishes δ_0 among the elements of $\mathrm{Stat}(A, p)$ is that, as observed previously, it is a *repeller* for the dynamics f_A. That basic observation underlies all our arguments.

The main difficulty for bounding $\eta(\mathbb{D}(0, \varepsilon))$ is that the problem is inherently non-compact: the conclusion of Proposition 3.8 is generally false when the pair (B, q) is reducible; thus, estimates must take into account how close an irreducible cocycle is to being reducible. The way we handle this is, roughly speaking, by splitting the mass $\eta(\mathbb{D}(0, \varepsilon))$ into two parts, $\eta(\mathbb{D}(0, \varepsilon)\backslash\mathbb{D}(0, \rho))$ and $\eta(\mathbb{D}(0, \rho))$, where $0 \le \rho < \varepsilon$ is very small if (B, q) is close to having zero as a fixed point (that is, having the vertical direction v as an eigenspace). Then we estimate the two parts using two different approaches, in §§5 and 6.

The following example illustrates these issues and can be used as a guideline for what follows. Take p to be supported on exactly two points, with equal masses, corresponding to Möbius transformations

$$B_1(z) = 9z \quad \text{and} \quad B_2(z) = \frac{2^{-1}z + b}{cz + 2}$$

with b and c close to zero. In this case, ρ may be defined in terms of the distance between the fixed point 0 of B_1 and its image under B_2: that is, in terms of $|b|$. If $|z| \ge \rho$ then $B_1^n(z)$ leaves $\mathbb{D}(0, \varepsilon)$ rapidly, because zero is a strongly repelling fixed point for B_1. If $|z| < \rho$ then $|B_2(z)| \ge \rho$ and so the sequence $B_1^n B_2(z)$ also leaves $\mathbb{D}(0, \varepsilon)$ in a small number of iterates. One deduces that, in either case $\mathbb{D}(0, \varepsilon)\backslash\mathbb{D}(0, \rho)$ or $\mathbb{D}(0, \rho)$, the average time to exit $\mathbb{D}(0, \varepsilon)$ is small. Building on this, one obtains that both sets have small mass, relative to any stationary measure.

The reader should be warned, however, that the choice of the threshold radius ρ is a lot more delicate in our general situation than in such a simple example. The way we implement it is through the notion of *adapted radius* that will appear in §5 and it depends on the stationary measure as well as on the cocycle.

4.2. *Discretization.* We begin by introducing a convenient discretization procedure. We emphasize that this procedure depends only on the pair (A, p): the numbers $h > 0$, $s \in \mathbb{Z}$, $s_x \in \mathbb{Z}$ and $\alpha > 0$ that we introduce in the subsequent work depend only on (A, p) and they are fixed here, once and for all.

Fix $h > 0$ such that $\int \log|\theta_x|\,dp(x) > 6h$. For each $x \in X$, let s_x be the unique integer such that

$$\log|\theta_x| - 2h < hs_x \le \log|\theta_x| - h. \tag{6}$$

As immediate consequences, we get (denote $\|A\| = \sup_{x\in X} \|A_x\|$)

$$e^{-2h}|\theta_x| < e^{hs_x} \le e^{-h}|\theta_x| < \|A\| \quad \text{for all } x \in X \tag{7}$$

and

$$\int hs_x\,dp(x) > 4h. \tag{8}$$

Define $D_x : \overline{\mathbb{C}} \to \overline{\mathbb{C}}$ by $D_x(z) = e^{2hs_x}z$. The relations (7) and (8) mean that D_x is definitely (slightly) more contracting than $A_x(z) = \theta_x^2 z$ but, nevertheless, is still dilating on average. Fix an integer $s > 0$, large enough so that

$$s \ge |s_x| \quad \text{for every } x \in X \text{ and } hs \ge \log(2\|A\|). \tag{9}$$

Then define $\Delta : \overline{\mathbb{C}} \to \overline{\mathbb{C}}$ by $\Delta(z) = e^{-2hs}z$.

Given any measurable set $K \subset X$, define $D_x^K : \overline{\mathbb{C}} \to \overline{\mathbb{C}}$ by $D_x^K(z) = e^{2hs_x^K} z$, where

$$s_x^K = \begin{cases} s_x & \text{if } x \in K \\ -s & \text{if } x \in X \backslash K. \end{cases} \tag{10}$$

In other words, D_x^K coincides with D_x on the set K and is constant and equal to the strong contraction Δ on the complement of K. By (8),

$$\int hs_x^K \, dp(x) \geq 4h - \int_{X \backslash K} h(s + s_x) \, dp(x) \geq 4h - 2p(X \backslash K)hs.$$

Define $\alpha = 1/s$. Then

$$\int hs_x^K \, dp(x) \geq 2h \quad \text{for every } K \subset X \text{ with } p(X \backslash K) \leq \alpha. \tag{11}$$

Let $K_+ = \{x \in \mathcal{M} : s_x^K > 0\}$ be the region where D_x^K is an expansion and $K_- = \{x \in \mathcal{M} : s_x^K < 0\}$ be the region where D_x^K is a contraction. Notice that $X \backslash K \subset X_-$ because $s_x^K = -s$ for all $x \in X \backslash K$. Moreover, by (11)

$$p(K_+)hs \geq \int_{K_+} hs_x^K \, dp(x) \geq \int hs_x^K \, dp(x) \geq 2h \tag{12}$$

and so $p(K_+) \geq 2\alpha$ for every $K \subset X$ with $p(X \backslash K) \leq \alpha$.

4.3. *Contractions.* We need a few elementary facts about the behavior of contractions on a closed disk $\mathbb{D}(0, a) = \{z \in \mathbb{C} : |z| \leq a\}$, where $a > 0$ is fixed. Let $\lambda < 1$ and $\Phi : \mathbb{D}(0, a) \to \mathbb{D}(0, a)$ be a λ-contraction.

LEMMA 4.2. *Suppose that* $w_0 = \Phi(0)$ *is different from* 0. *Then:*

(a) $\mathbb{D}(0, r) \cap \Phi(\mathbb{D}(0, r)) = \emptyset$ *for all* $0 \leq r < |w_0|/2$;

(b) *if* $a \geq |w_0|/1 - \lambda$, *then* $\Phi(\mathbb{D}(0, R)) \subset \mathbb{D}(0, R)$ *for all* $a \geq R \geq |w_0|/1 - \lambda$; *and*

(c) *if* $0 \leq \hat{r} \leq a$ *and* $\Phi(\mathbb{D}(0, \hat{r})) \not\subset \mathbb{D}(0, \hat{r})$, *then*

$$\mathbb{D}\left(0, \frac{1-\lambda}{2}\hat{r}\right) \cap \Phi\left(\mathbb{D}\left(0, \frac{1-\lambda}{2}\hat{r}\right)\right) = \emptyset.$$

Proof. It is clear that $\Phi(\mathbb{D}(0, r))$ is contained in $\mathbb{D}(w_0, r)$ and $\mathbb{D}(0, r) \cap \mathbb{D}(w_0, r) = \emptyset$ when $r < |w_0|/2$. This proves part (a). Next, observe that

$$|\Phi(z)| \leq |\Phi(z) - \Phi(0)| + |\Phi(0)| \leq \lambda|z| + |w_0| \leq \lambda R + (1 - \lambda)R = R$$

if $a \geq R \geq |w_0|/(1 - \lambda)$ and $|z| \leq R$. This proves part (b). Then, $\Phi(\mathbb{D}(0, \hat{r})) \not\subset \mathbb{D}(0, \hat{r})$ implies that $\hat{r} < |w_0|/(1 - \lambda)$: that is, $(1 - \lambda)\hat{r}/2 < |w_0|/2$. By (a), this implies (c). □

LEMMA 4.3. *Let* $\tau > 0$ *and* $1 \geq \Lambda > \lambda > 0$ *with* $((1 + \lambda)/(\Lambda - \lambda))\tau \leq a$. *If the fixed point of* Φ *is in* $\mathbb{D}(0, \tau)$, *then*

$$\Phi(\mathbb{D}(0, r)) \subset \mathbb{D}(0, \Lambda r) \quad \textit{for all } r \in [C\tau, a] \textit{ where } C = (1 + \lambda)/(\Lambda - \lambda).$$

Proof. Let $z_0 \in \mathbb{D}(0, \tau)$ be the fixed point of Φ and be $z \in \mathbb{D}(0, r)$ with $a \geq r \geq C\tau$. Then

$$|\Phi(z)| \leq |\Phi(z) - z_0| + |z_0| \leq \lambda|z - z_0| + |z_0| \leq \lambda(r + \tau) + \tau.$$

The assumption $r \geq (1 + \lambda)\tau/(\Lambda - \lambda)$ implies that $\lambda(r + \tau) + \tau \leq \Lambda r$ and, therefore, $|\Phi(z)| \leq \Lambda r$: that is, $\Phi(\mathbb{D}(0, r)) \subset \mathbb{D}(0, \Lambda r)$. □

LEMMA 4.4. *There is* $0 \leq r_1 \leq a$ *such that* $\{r \in [0, a] : \Phi(\mathbb{D}(0, r)) \subset \mathbb{D}(0, r)\} = [r_1, a]$.

Proof. Let r_1 be the infimum of $r \geq 0$ such that $\Phi(\mathbb{D}(0, s)) \subset \mathbb{D}(0, s)$ for all $s \geq r$. Clearly, $\Phi(\mathbb{D}(0, r_1)) \subset \mathbb{D}(0, r_1)$. We claim that $\Phi(\mathbb{D}(0, r)) \not\subset \mathbb{D}(0, r)$ for all $r < r_1$. Indeed, suppose that there is $r_2 < r_1$ such that $\Phi(\mathbb{D}(0, r_2)) \subset \mathbb{D}(0, r_2)$. By the choice of r_1 and the fact that Φ is continuous, there is $\xi_0 \in \mathbb{D}(0, r_1)$ with $|\xi_0| = r_1$ such that $|\Phi(\xi_0)| = r_1$: if $|\Phi(z)| < r_1$ for all $z \in \mathbb{D}(0, r_1)$ then, by continuity of Φ and compactness of Q3 $\mathbb{D}(0, r_1)$, there would be $\delta > 0$ such that $|\Phi(z)| < r_1 - \delta$ for $z \in \mathbb{D}(0, r_1)$; the latter would contradict the choice of r_1. Let $\eta_0 = r_2\xi_0/|\xi_0| \in \mathbb{D}(0, r_2)$. Then, we would have $|\Phi(\xi_0) - \Phi(\eta_0)| \geq r_1 - r_2 \geq |\xi_0 - \eta_0|$, which would also contradict the assumption that Φ is a λ-contraction. □

4.4. *Applications to cocycles.* Here are a few applications of the lemmas in §4.3 to the context that we are interested in. Let $A \in \mathcal{S}(X)$ be given. The parameter $\gamma > 0$ in the statements is the radius of a neighborhood of A on which certain properties hold. Reducing γ just reduces this neighborhood and, thus, can only weaken the claim. So all the statements in this section extend automatically to every $\gamma > 0$ that is sufficiently small.

LEMMA 4.5. *There exists* $\gamma > 0$ *such that if* $d(A, B) < \gamma$ *and* $r \in [0, 1]$ *and* $x \in X$ *are such that* $B_x^{-1}(\mathbb{D}(0, r)) \cap \mathbb{D}(0, \|A\|^2 r) \neq \emptyset$, *then*

$$B_x^{-1}(\mathbb{D}(0, r)) \cup \mathbb{D}(0, \|A\|^2 r) \subset \mathbb{D}(0, e^{2hs} r) = \Delta^{-1}(\mathbb{D}(0, r)).$$

Proof. Clearly, the diameter of $A_x^{-1}(\mathbb{D}(0, r))$ is bounded by $2|\theta_x|^{-2} r \leq 2\|A\|^2 r$, for every r and every x. Take $\gamma > 0$ to be sufficiently small so that $d(A, B) < \gamma$ implies that the diameter of $B_x^{-1}(\mathbb{D}(0, r))$ is less than $3\|A\|^2 r$ for every r and every x. Then

$$B_x^{-1}(\mathbb{D}(0, r)) \cap \mathbb{D}(0, \|A\|^2 r) \neq \emptyset \Rightarrow B_x^{-1}(\mathbb{D}(0, r)) \cup \mathbb{D}(0, \|A\|^2 r) \subset \mathbb{D}(0, 4\|A\|^2 r).$$

To conclude, use the second part of (9). □

LEMMA 4.6. *Given* $0 < r_0 \leq 1$ *there exists* $\gamma > 0$ *such that, if* $d(A, B) < \gamma$ *and* $r \in [r_0, 1]$,

$$B_x^{-1}(\mathbb{D}(0, r)) \subset \mathbb{D}(0, e^{-2hs_x} r) = D_x^{-1}(\mathbb{D}(0, r)) \quad \textit{for every } x \in X.$$

Proof. Let $r_0 \in (0, 1]$ be fixed. By (6), every $D_x A_x^{-1}$, $x \in X$ is an e^{-2h}-contraction fixing the origin. Let $C = (1 + e^{-h})/(1 - e^{-h})$. Then, assuming that γ is sufficiently small, every $\Phi_x = D_x B_x^{-1}$, $x \in X$ is an e^{-h}-contraction on $\mathbb{D}(0, 1)$ and its fixed point is in $\mathbb{D}(0, C^{-1} r_0)$. By Lemma 4.3 (with $a = 1$ and $\lambda = e^{-h}$ and $\Lambda = 1$ and $\tau = C^{-1} r_0$), it follows that $\Phi_x(\mathbb{D}(0, r)) \subset \mathbb{D}(0, r)$ for all $x \in X$ and $1 \geq r \geq r_0$. In other words, $B_x^{-1}(\mathbb{D}(0, r)) \subset D_x^{-1}(\mathbb{D}(0, r))$ for all $x \in X$ and $1 \geq r \geq r_0$. □

Remark 4.7. The fact that $\Phi_x = D_x B_x^{-1}$ is an e^{-h}-contraction on $\mathbb{D}(0, 1)$ for every $x \in X$, if B is close enough to A, will be used a few times in the subsequent work.

COROLLARY 4.8. *There exists $\gamma > 0$ such that, if $d(A, B) < \gamma$ and $\varepsilon < e^{-2hs}$,*

$$B_x^{-1}(\mathbb{D}(0, 1)) \subset \mathbb{D}(0, \varepsilon^{-1}) \quad \textit{for every } x \in X.$$

Proof. Recall that $s \geq -s_x$ for every x and apply Lemma 4.6 with $r = r_0 = 1$. □

Next, define

$$c_1 = \frac{1 - e^{-h}}{2} \quad \text{and} \quad c = c_1 e^{-2hs}. \tag{13}$$

These numbers depend only on A, because h and s have been fixed depending only on A.

LEMMA 4.9. *There exists $\gamma > 0$ such that, if $d(A, B) < \gamma$,*

$$\mathbb{D}(0, \mathrm{cr}) \cap B_x^{-1}(\mathbb{D}(0, \mathrm{cr})) = \emptyset$$

for every $x \in X$ and $0 < r < 1$ such that $B_x^{-1}(\mathbb{D}(0, r)) \not\subset D_x^{-1}(\mathbb{D}(0, r))$.

Proof. As observed before (Remark 4.7), every $\Phi_x = D_x B_x^{-1}$ is an e^{-h}-contraction on $\mathbb{D}(0, 1)$ if B is close enough to A. Let $x \in X$ and $0 < r < 1$ be as in the statement. The hypothesis $B_x^{-1}(\mathbb{D}(0, r)) \not\subset D_x^{-1}(\mathbb{D}(0, r))$ may be rewritten as $\Phi_x(\mathbb{D}(0, r)) \not\subset \mathbb{D}(0, r)$. Applying Lemma 4.2(c), with $a = 1$ and $\lambda = e^{-h}$ and $\hat{r} = r$, we conclude that

$$\mathbb{D}(0, c_1 r) \cap \Phi_x(\mathbb{D}(0, c_1 r)) = \emptyset.$$

Using the definitions of D_x and Φ_x, this may be rewritten as

$$\mathbb{D}(0, c_1 e^{-2hs_x} r) \cap B_x^{-1}(\mathbb{D}(0, c_1 r)) = \emptyset$$

and, since $s \geq 0$ and $s \geq s_x$ for every x, this relation implies that

$$\mathbb{D}(0, c_1 e^{-2hs} r) \cap B_x^{-1}(\mathbb{D}(0, c_1 e^{-2hs} r)) = \emptyset,$$

just as claimed. □

Recall that X_+ denotes the set of points $x \in X$ for which $s_x > 0$. As a particular case of (12), taking $K = X$, we have that $p(X_+) > 2\alpha$. Define

$$C = \frac{2e^{2hs}}{1 - e^{-h}}. \tag{14}$$

Keep in mind that C depends only on A, because h and s have been fixed, depending only on A.

LEMMA 4.10. *There exists $\gamma > 0$ such that, if $d(A, B) < \gamma$ and $0 < C\tau \leq 1$,*

$$B_x^{-1}(\mathbb{D}(0, r)) \subset \mathbb{D}(0, e^{-2hs_x} r) \subset \mathbb{D}(0, e^{-2h} r)$$

for every $r \in [C\tau, 1]$ and any $x \in X_+$ such that the fixed point of B_x is in $\mathbb{D}(0, \tau)$.

Proof. For each $x \in X_+$, we have that $\log|\theta_x| \geq h(s_x + 1)$ and so, in particular, $A_x^{-1}(z) = \theta_x^2 z$ is an $e^{-2h(s_x+1)}$-contraction on $\mathbb{D}(0, 1)$. Thus, assuming that $\gamma > 0$ is small enough, $d(A, B) < \gamma$ implies that B_x^{-1} is an $e^{-2h(s_x+1/2)}$-contraction on $\mathbb{D}(0, 1)$ for every $x \in X_+$. Let $a = 1$ and $\Lambda_x = e^{-2hs_x}$ and $\lambda_x = e^{-h}e^{-2hs_x}$. Then, applying Lemma 4.3 to $\Phi = B_x^{-1}$, we obtain that if the fixed point of B_x^{-1} is in $\mathbb{D}(0, \tau)$, then

$$B_x^{-1}(\mathbb{D}(0, r)) \subset \mathbb{D}(0, \Lambda_x r) = \mathbb{D}(0, e^{-2hs_x} r) \tag{15}$$

for every $r \in [C_x\tau, 1]$, where

$$C_x = \frac{1 + \lambda_x}{\Lambda_x - \lambda_x}$$

and it is assumed that $0 < C_x\tau \leq 1$. Note that $C_x \leq C$ for every x, because $h > 0$ and $s_x \leq s$ and $s \geq 0$. Thus (15) holds for $1 \geq r \geq C\tau > 0$ and every $x \in X_+$ such that the fixed point of B_x^{-1} is in $\mathbb{D}(0, \tau)$. □

5. *Adapted radii*

The following definition plays a central part in our arguments. Given a pair $(B, q) \in \mathcal{S}(X) \times \mathcal{P}(X)$ and a (B, q)-stationary measure η, we say that $r \geq 0$ is a (B, q, η)-*adapted radius* on a measurable set $K \subset X$ if

$$\int \eta(B_x^{-1}(\mathbb{D}(0, r)))\, dq(x) \leq \int \eta((D_x^K)^{-1}(\mathbb{D}(0, r)))\, dq(x). \tag{16}$$

For x and r fixed, $(D_x^K)^{-1}(\mathbb{D}(0, r)) = \mathbb{D}(0, e^{-2hs_x^K} r)$ can only decrease when the set K increases (because $s_x \geq -s$ for every $x \in X$). So the condition (16) becomes stronger as the set K becomes larger.

For each measurable set $K \subset X$ with $p(X \backslash K) \leq \alpha$, define

$$\rho(B, q, \eta, K) = \inf\{r \in [0, 1] : \text{every } s \in [r, 1] \text{ is } (B, q, \eta)\text{-adapted on } K\}. \tag{17}$$

Sometimes we write $\rho(K)$ to mean $\rho(B, q, \eta, K)$, if B, q and η are fixed and no confusion can arise from this simplification.

Applying Lemma 4.6 with $r_0 = 1$ we get that if γ is sufficiently small, depending only on A, then $B_x^{-1}(\mathbb{D}(0, 1)) \subset D_x^{-1}(\mathbb{D}(0, 1))$ for every $x \in X$ and any B such that $d(A, B) < \gamma$. In particular, if $(B, q) \in V(A, p, \gamma)$ and η is a (B, q)-stationary measure, then $r_0 = 1$ is (B, q, η)-adapted. This ensures that $\rho(B, q, \eta, K)$ is well defined for any such (B, q, η) and any measurable $K \subset X$ with $p(X \backslash K) \leq \alpha$.

PROPOSITION 5.1. *Given $\varepsilon > 0$ and $\delta > 0$, there exists $\gamma > 0$ such that for any $(B, q) \in V(A, p, \gamma)$, any (B, q)-stationary measure η and any measurable set K with $p(X \backslash K) \leq \alpha$,*

$$\eta(\mathbb{D}(0, \varepsilon) \backslash \mathbb{D}(0, \rho(K))) \leq \delta \quad \textit{where } \rho(K) = \rho(B, q, \eta, K).$$

Proposition 5.1 will be proved in §5.2. The following direct consequence is the main conclusion in this section. Define

$$\rho = \rho(B, q, \eta) = \inf\{\rho(B, q, \eta, K) : p(X \backslash K) \leq \alpha\}. \tag{18}$$

Sometimes we write ρ to mean $\rho(B, q, \eta)$, if B, q and η are fixed and no confusion can arise from doing so.

COROLLARY 5.2. *Given $\varepsilon > 0$ and $\delta > 0$, there exists $\gamma > 0$ such that, for any $(B, q) \in V(A, p, \gamma)$ and any (B, q)-stationary measure η,*

$$\eta(\mathbb{D}(0, \varepsilon) \setminus \mathbb{D}(0, \rho)) \le \delta \quad \textit{where } \rho = \rho(B, q, \eta).$$

Proof. Take K_j with $\rho(K_j) \searrow \rho$ and notice that $\mathbb{D}(0, \rho) = \bigcap_j \mathbb{D}(0, \rho(K_j))$. □

Remark 5.3. Reducing γ just reduces the neighborhood $V(A, p, \gamma)$, which can only weaken the statements of Proposition 5.1 and Corollary 5.2. Thus both statements hold true for every sufficiently small γ.

5.1. *Two auxiliary lemmas.* To prove Proposition 5.1, it is convenient to discretize the phase space as well. Define $I_j(r) = \mathbb{D}(0, e^{-(2j-2)h} r) \setminus \mathbb{D}(0, e^{-2jh} r)$ for each $j \in \mathbb{Z}$ and $r > 0$. Clearly, for any fixed r, the sequence $(I_j(r))_j$ is invariant under Δ and every D_x. So it is also invariant under every D_x^K, for any $K \subset X$.

LEMMA 5.4. *If $r > 0$ is (B, q, η)-adapted on K, then*

$$\int_{K_+} \sum_{j=1}^{s_x^K} \eta(I_j(r))\, dq(x) \le \int_{K_-} \sum_{j=s_x^K+1}^{0} \eta(I_j(r))\, dq(x).$$

If $e^{-2ht} r$ is (B, q, η)-adapted on K for every $t = 0, 1, \ldots, n$, then

$$\int_{K_+} \sum_{j=1}^{s_x^K} \eta(I_{t+j}(r))\, dq(x) \le \int_{K_-} \sum_{j=s_x^K+1}^{0} \eta(I_{t+j}(r))\, dq(x) \quad \textit{for } t = 0, 1, \ldots, n.$$

Proof. Define

$$L_x(r) = \begin{cases} \mathbb{D}(0, r) \setminus (D_x^K)^{-1}(\mathbb{D}(0, r)) = \mathbb{D}(0, r) \setminus \mathbb{D}(0, e^{-2hs_x^K} r) & \text{for } x \in K_+, \\ \emptyset & \text{otherwise}, \\ (D_x^K)^{-1}(\mathbb{D}(0, r)) \setminus \mathbb{D}(0, r) = \mathbb{D}(0, e^{-2hs_x^K} r) \setminus \mathbb{D}(0, r) & \text{for } x \in K_-. \end{cases}$$

Using that r is (B, q, η)-adapted and η is (B, q)-stationary, we find that

$$\begin{aligned} &\int (\eta(\mathbb{D}(0, r)) - \eta(\mathbb{D}(0, e^{-2hs_x^K} r)))\, dq(x) \\ &\quad \le \int (\eta(\mathbb{D}(0, r)) - \eta(B_x^{-1}(\mathbb{D}(0, r))))\, dq(x) = 0. \end{aligned}$$

The left-hand side coincides with $\int_{K_+} \eta(L_x(r))\, dq(x) - \int_{K_-} \eta(L_x(r))\, dq(x)$. So

$$\int_{K_+} \eta(L_x(r))\, dq(x) \le \int_{K_-} \eta(L_x(r))\, dq(x).$$

Now, to get the first claim, just notice that $L_x(r) = \bigsqcup_{j=1}^{s_x^K} I_j(r)$ if $x \in K_+$ and $L_x(r) = \bigsqcup_{j=s_x^K+1}^{0} I_j(r)$ if $x \in K_-$ (where $\bigsqcup$ denotes disjoint union). The last claim is an immediate consequence, because $I_{j+t}(r) = I_j(e^{-2ht} r)$ for every j, t and r. □

We also need the following abstract fact.

LEMMA 5.5. *Let $X \to \mathbb{N}$, $x \mapsto n_x$ be a bounded measurable function and let $(a_j)_{j\in\mathbb{Z}}$ be a sequence of non-negative real numbers. Given measurable subsets Y_+ and Y_- of X, denote $n_* = \sup\{n_x : x \in Y_*\}$ for $* \in \{+, -\}$. Suppose that there exist $\tau > 0$, $n \geq 0$ and a probability measurable q on X such that:*

(a) $0 < \tau \leq \int_{Y_+} n_x\, dq(x) - \int_{Y_-} n_x\, dq(x)$; *and*

(b) $\int_{Y_+} \sum_{j=1}^{n_x} a_{j+t}\, dq(x) \leq \int_{Y_-} \sum_{j=-n_x+1}^{0} a_{j+t}\, dq(x)$ *for* $t = 0, \ldots, n$.

Then

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

Proof. Begin by noticing that

$$\sum_{t=0}^{n} \sum_{j=1}^{n_x} a_{j+t} = \sum_{l=1}^{n_x} \sum_{j=l}^{n+l} a_j \geq \sum_{l=1}^{n_x} \sum_{j=n_x+1}^{n+1} a_j \geq n_x \left(\sum_{j=1}^{n} a_j - \sum_{j=1}^{n_x} a_j \right) \tag{19}$$

and, similarly,

$$\begin{aligned} \sum_{t=0}^{n} \sum_{j=-n_x+1}^{0} a_{j+t} &= \sum_{l=-n_x+1}^{0} \sum_{j=l}^{n+l} a_j \\ &\leq \sum_{l=-n_x+1}^{0} \sum_{j=-n_x+1}^{n} a_j \leq n_x \left(\sum_{j=1}^{n} a_j + \sum_{j=-n_x+1}^{0} a_j \right). \end{aligned} \tag{20}$$

Adding the inequalities (b) over all $t = 0, \ldots, n$ and using (19)–(20),

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

Then, using the inequality (a),

$$\begin{aligned} \tau \sum_{j=1}^{n} a_j &\leq \int_{Y_+} n_x \sum_{j=1}^{n_x} a_j\, dq(x) + \int_{Y_-} n_x \sum_{j=-n_x+1}^{0} a_j\, dq(x) \\ &\leq n_+ \int_{Y_+} \sum_{j=1}^{n_x} a_j\, dq(x) + n_- \int_{Y_-} \sum_{j=-n_x+1}^{0} a_j\, dq(x). \end{aligned}$$

Using the inequality (b) with $t = 0$, it follows that

$$\tau \sum_{j=1}^{n} a_j \leq (n_+ + n_-) \int_{Y_-} \sum_{j=-n_x+1}^{0} a_j\, dq(x) \leq (n_+ + n_-) \sum_{j=-n_-+1}^{0} a_j q(Y_-).$$

This implies the conclusion of the lemma. □

5.2. *Proof of Proposition 5.1.* The family of functions $x \in X \mapsto s_x^K$ defined in (10) is uniformly bounded: by definition, $|s_x^K| \leq s$ for any measurable set $K \subset X$ and every $x \in X$. Thus we may choose $\gamma > 0$ such that

$$\left| \int s_x^K\, dp(x) - \int s_x^K\, dq(x) \right| < 1 \tag{21}$$

for every $q \in \mathcal{P}(X)$ such that $d(p,q) < \gamma$ and every measurable set $K \subset X$.

Fix any $\varepsilon < e^{-2hs}$. By Lemma 4.1, reducing γ if necessary, we may suppose that

$$\eta(\mathbb{D}(0,\varepsilon^{-1}) \setminus \mathbb{D}(0,\varepsilon)) \le \frac{h\delta}{2s}$$

for every (B,q)-stationary measure η and any pair $(B,q) \in V(A,p,\gamma)$.

Let (B,q,η) be fixed and $K \subset X$ be any measurable set with $p(X \setminus K) \le \alpha$. Define $n_x = |s_x^K|$ for each $x \in X$. Then

$$\int_{K_+} n_x \, dq(x) - \int_{K_-} n_x \, dq(x) = \int s_x^K \, dq(x).$$

Combining (21) with (11) through the triangle inequality, we deduce that

$$\int_{K_+} n_x \, dq(x) - \int_{K_-} n_x \, dq(x) = \int s_x^K \, dq(x) \ge 1 \tag{22}$$

whenever $d(p,q) < \gamma$.

Consider any $1 \ge r_0 > \rho(K)$ and then take $r_1 \in [\varepsilon, 1]$ such that $r_0 = r_1 e^{-2hn}$ for some $n \ge 0$. By the definition of $\rho(K)$ in (17), every $r \in [r_0, 1]$ is (B,q,η)-adapted on K. In particular, this holds for $r = r_1 e^{-2ht}$ for every $t = 0, 1, \dots, n$. Let $a_j = \eta(I_j(r_1))$ for $j \in \mathbb{Z}$. Then the conclusion of Lemma 5.4 may be written as

$$\int_{K_+} \sum_{j=1}^{n_x} a_{j+t} \, dq(x) \le \int_{K_-} \sum_{j=-n_x+1}^{0} a_{j+t} \, dq(x) \quad \text{for all } t = 0, 1, \dots, n. \tag{23}$$

Properties (22) and (23) correspond to hypotheses (a) and (b) in Lemma 5.5. From this lemma we get that

$$\sum_{j=1}^{n} a_j \le \frac{2s}{h} \sum_{j=-s+1}^{0} a_j. \tag{24}$$

The left-hand side of (24) coincides with

$$\eta(\mathbb{D}(0,r_1) \setminus \mathbb{D}(0, r_1 e^{-2hn})) = \eta(\mathbb{D}(0,r_1) \setminus \mathbb{D}(0,r_0)) \ge \eta(\mathbb{D}(0,\varepsilon) \setminus \mathbb{D}(0,r_0))$$

(because $r_1 \ge \varepsilon$). The right-hand side of (24) coincides with

$$\frac{2s}{h} \eta(\mathbb{D}(0, r_1 e^{2hs}) \setminus \mathbb{D}(0,r_1)) \le \frac{2s}{h} \eta(\mathbb{D}(0,\varepsilon^{-1}) \setminus \mathbb{D}(0,\varepsilon)),$$

(because $\varepsilon \le r_1 \le 1$ and $e^{2hs} < \varepsilon^{-1}$, as long as ε is sufficiently small). Hence, the inequality (24) implies that

$$\eta(\mathbb{D}(0,\varepsilon) \setminus \mathbb{D}(0,r_0)) \le \frac{2s}{h} \eta(\mathbb{D}(0,\varepsilon^{-1}) \setminus \mathbb{D}(0,\varepsilon)) \le \delta.$$

Making $r_0 \to \rho(K)$ one gets the conclusion of the proposition.

6. *Proof of Proposition 3.8*

In view of Lemma 4.1 and Corollary 5.2–Remark 5.3, at this point it suffices to show that

$$\eta(\mathbb{D}(0,\rho)) \le \text{const}\,\delta \quad (\text{the number } \rho = \rho(B, q, \eta) \text{ was defined in (18)})$$

for every (B, q)-stationary measure η and every pair (B, q) close enough to (A, p) and such that $q(\{x \in X : B_x(z) = z\}) < 1$ for every $z \in \overline{\mathbb{C}}$.

The case when $\rho = 0$ is easy, because the next lemma implies that $\mathbb{D}(0, 0) = \{0\}$ always has measure zero. For the same reason as in Remark 5.3, the statement extends automatically to every $\gamma > 0$ sufficiently small.

LEMMA 6.1. *There exists $\gamma > 0$ such that, if the pair $(B, q) \in V(A, p, \gamma)$ satisfies $q(\{x \in X : B_x(z) = z\}) < 1$ for all $z \in \overline{\mathbb{C}}$, every (B, q)-stationary measure η is non-atomic.*

Proof. Suppose that η has some atom. Let $a_0 > 0$ be the largest mass of any atom and let $F = \{z_1, \dots, z_l\}$ be the set of atoms with $\eta(\{z_i\}) = a_0$. Then $\eta(E) \le a_0 \# E$ for any finite set $E \subset \overline{\mathbb{C}}$, and the equality holds if and only if $E \subset F$. Since η is a stationary measure,

$$la_0 = \eta(F) = \int \eta(B_x^{-1}(F))\, dq(x) \le \int la_0\, dq(x) = la_0.$$

This implies that $\eta(B_x^{-1}(F)) = a_0 l$ for q-almost every x which, in view of the previous observations, implies that $B_x^{-1}(F) = F$ for q-almost every x. Clearly, (5) implies that $|\theta_x| > 1$ for every x in some $Y \subset X$ with $p(Y) > 0$. If (B, q) is close to (A, p) then $q(Y) > 0$ and the Möbius transformation B_x is hyperbolic, with fixed points close to zero and ∞, for every $x \in Y$. Then, F must be contained in the set of fixed points of B_x for any $x \in Y$. In particular, $\# F \le 2$. If F consists of a single point z_1 then the invariance property $B_x^{-1}(F) = F$ for q-almost every x means that $B_x(z_1) = z_1$ for q-almost every x, which contradicts the hypothesis. Otherwise, $F = \{z_1, z_2\}$ with z_1 close to zero and z_2 close to ∞. Since A_x fixes both zero and ∞ and we take B to be close to A, $B_x(z_1) \ne z_2$ and $B_x(z_2) \ne z_1$ for every x. Thus the invariance property of F translates to $B_x(z_i) = z_i$ for $i = 1, 2$ and q-almost every x. Arguing just as in the previous case, we reach a contradiction. These contradictions prove that η cannot have atoms. □

For the remainder of the proof, suppose that $\rho > 0$. Consider $\varepsilon < e^{-2hs}$, where h and α are the constants introduced in the §4.2. Throughout, it is understood that η is a (B, q)-stationary measure and $(B, q) \in V(A, p, \gamma)$ for some $\gamma > 0$ sufficiently small (conditions are imposed along the way) depending only on A and ε and δ.

For each $t \in [0, 1]$, define

$$K_t = \{x \in X : B_x^{-1}(\mathbb{D}(0, t)) \subset \mathbb{D}(0, e^{-2hs_x} t)\}.$$

Applying Lemma 4.4 to $\Phi_x = D_x B_x^{-1}$ and $a = 1$ (we have seen in Remark 4.7 that Φ_x is an e^{-h}-contraction on $\mathbb{D}(0, 1)$ for every $x \in X$), we find that the function

$$[0, 1] \ni t \mapsto K_t \quad \text{is non-decreasing.} \tag{25}$$

Let us distinguish two cases.

Case 1. $p(X \backslash K_r) \le \alpha$ for some $r \in [0, \rho)$. This is handled by the following lemma.

LEMMA 6.2. *If $p(X\backslash K_r) \le \alpha$ for some $r \in [0, \rho)$, then $\eta(\mathbb{D}(0, \rho)) \le 2\delta$.*

Proof. The observation (25) implies that $t \mapsto p(X\backslash K_t)$ is non-increasing. Thus r may be chosen arbitrarily close to ρ. Fix $r \in (\|A\|^{-2}\rho, \rho)$ and let $K = K_r$. The hypothesis implies that $p(X\backslash K) \le \alpha$ and then the definition of ρ in (18) gives that $r < \rho(K)$. Then, by the definition of $\rho(K)$ in (17), there exists $t \in (r, \rho)$ that is not (B, q, η)-adapted on K. In other words,

$$\int \eta(B_x^{-1}(\mathbb{D}(0, t)))\, dq(x) > \int \eta(\mathbb{D}(0, e^{-2hs_x^K} t))\, dq(x).$$

This implies that there exists $y \in X$ such that

$$\eta(B_y^{-1}(\mathbb{D}(0, t))) > \eta(\mathbb{D}(0, e^{-2hs_y^K} t)) \ge \eta(\mathbb{D}(0, e^{-2hs_y} t)),$$

(recall that $s_x^K \le s_x$ for every x). In particular, $y \notin K_t$ and so, by the observation at the beginning of this proof, $y \notin K$. Consequently, the previous relation can be strengthened: that is

$$\eta(B_y^{-1}(\mathbb{D}(0, t))) > \eta(\mathbb{D}(0, e^{-2hs_y^K} t)) = \eta(\mathbb{D}(0, e^{2hs} t)). \tag{26}$$

The choice of t together with (9) give that $e^{2hs}t > \|A\|^2 t > \|A\|^2 r > \rho$. Thus

$$\eta(B_y^{-1}(\mathbb{D}(0, t))) > \eta(\mathbb{D}(0, \rho)). \tag{27}$$

Another consequence of (26) is that

$$B_y^{-1}(\mathbb{D}(0, t)) \not\subset \mathbb{D}(0, e^{2hs} t). \tag{28}$$

Take $\gamma > 0$ to be small enough (depending only on A) for the assertion of Lemma 4.5 to be valid in this setting. Applying the lemma with $r = t$, (28) implies that

$$B_y^{-1}(\mathbb{D}(0, t)) \cap \mathbb{D}(0, \|A\|^2 t) = \emptyset \quad \text{and so} \quad B_y^{-1}(\mathbb{D}(0, t)) \cap \mathbb{D}(0, \rho) = \emptyset.$$

On the other hand, Corollary 4.8 gives that $B_y^{-1}(\mathbb{D}(0, t)) \subset \mathbb{D}(0, \varepsilon^{-1})$. So

$$B_y^{-1}(\mathbb{D}(0, t)) \subset \mathbb{D}(0, \varepsilon^{-1})\backslash\mathbb{D}(0, \|A\|^2 t) \subset \mathbb{D}(0, \varepsilon^{-1})\backslash\mathbb{D}(0, \rho). \tag{29}$$

Take $\gamma > 0$ to be small enough (depending only on A and ε and δ) for the assertions of Lemma 4.1 and Corollary 5.2 to hold in this setting. Then

$$\eta(\mathbb{D}(0, \varepsilon^{-1})\backslash\mathbb{D}(0, \varepsilon)) \le \delta \quad \text{and} \quad \eta(\mathbb{D}(0, \varepsilon)\backslash\mathbb{D}(0, \rho)) \le \delta.$$

By (29), this implies that

$$\eta(B_y^{-1}(\mathbb{D}(0, t))) \le \eta(\mathbb{D}(0, \varepsilon^{-1})\backslash\mathbb{D}(0, \rho)) \le 2\delta. \tag{30}$$

From (27) and (30) we get that $\eta(\mathbb{D}(0, \rho)) \le 2\delta$, as claimed. □

Case 2. $p(X\backslash K_r) > \alpha$ for every $r \in [0, \rho)$. It is clear that, reducing γ if necessary, B_x has a unique fixed point in $\mathbb{D}(0, 2)$ for all $x \in X_+$. So, for each $z \in \mathbb{D}(0, 1)$ and $r \in [0, 1]$, define

$$\Gamma(z, r) = \{x \in X_+ : \text{the fixed point of } B_x \text{ is in } \mathbb{D}(z, r)\}.$$

Let $c \in (0, 1)$ and $C > 1$ be as defined in (13) and (14), respectively. Then let $\ell \ge 0$ be the smallest integer such that $e^{-2h\ell} < c$. Keep in mind that c, C and ℓ depend only on A. So the same is true about

$$\omega = 8C^2 e^{4h\ell} \alpha^{-1}. \tag{31}$$

The reason for this definition will become apparent in the proof of the next lemma.

LEMMA 6.3. *There exist $z_0 \in \mathbb{D}(0, 1)$ and $\rho_0 \in [0, C^{-1}e^{-2h\ell}]$ such that:*
(a) $p(\Gamma(z_0, \rho_0)) \geq 2\omega^{-1}$*; and*
(b) $p(X_+ \backslash \Gamma(z_0, Ce^{2h\ell}\rho_0)) \geq \alpha$ *if* $\rho_0 > 0$.

Proof. Clearly, $\Gamma(0, C^{-1}e^{-2h\ell}) = X_+$ if B is close enough to A. Then (12) implies that $p(\Gamma(0, C^{-1}e^{-2h\ell})) > 2\alpha > 2\omega^{-1}$. Let ρ_0 be the infimum of the values of $r > 0$ such that $p(\Gamma(z, r)) \geq 2\omega^{-1}$ for some $z \in \mathbb{D}(0, 1)$. Consider $(r_k)_k$ decreasing to ρ_0 and $(z_k)_k$ in $\mathbb{D}(0, 1)$ such that $p(\gamma(z_k, r_k)) \geq 2\omega^{-1}$ for every k. Let z_0 be any accumulation point of $(z_k)_k$. Given any $r > \rho_0$, we have $\mathbb{D}(z_k, r_k) \subset \mathbb{D}(z_0, r)$, and so $\Gamma(z_k, r_k) \subset \Gamma(z_0, r)$, for arbitrarily large values of k. This implies that $p(\Gamma(z_0, r)) \geq 2\omega^{-1}$ for every $r > \rho_0$ and, consequently, $p(\Gamma(z_0, \rho_0)) \geq 2\omega^{-1}$. This gives part (a).

To prove part (b), suppose that $\rho_0 > 0$ and let $\rho_1 = 99\rho_0/100$. The definition of ρ_0 entails $p(\Gamma(z, \rho_1)) < 2\omega^{-1}$ for every $z \in \mathbb{D}(0, 1)$. Clearly, any ball of radius $Ce^{2h\ell}\rho_0$ can be covered with $4C^2e^{4h\ell}$ balls of radius ρ_1. Thus we can find $G \subset \mathbb{D}(0, 1)$ with $\#G \leq 4C^2e^{4h\ell}$ such that $\{\Gamma(z, \rho_1) : z \in G\}$ covers $\Gamma(z_0, Ce^{2h\ell}\rho_0)$. Then

$$p(X_+ \backslash \Gamma(z_0, Ce^{2h\ell}\rho_0)) \geq p(X_+) - \sum_{z \in G} p(\Gamma(z, \rho_1)) > 2\alpha - 4C^2e^{4h\ell}2\omega^{-1}.$$

The definition of ω in (31) is such that this last expression is equal to α. □

Remark 6.4. If B is close to A then the point z_0 is close to zero and the radius ρ_0 is small. More precisely, given any $r_0 > 0$, we have $\Gamma(0, r) = X_+$ for every $r \in [r_0, 1]$, as long as B is close enough to A. Then the previous construction yields $\rho_0 \leq r_0$. Moreover, $\Gamma(z, r)$ is empty for any $r \in [0, r_0]$ and any z with $|z| > 2r_0$. So $z_0 \in \mathbb{D}(0, 2r_0)$.

Also, observe that $Ce^{2h\ell}\rho_0 \leq 1$ for all B close sufficiently to A. For the time being, let us suppose that $z_0 = 0$. This assumption will be removed at the end of the section.

COROLLARY 6.5. $p(X \backslash K_r) \geq \alpha$ *for* $0 \leq r \leq Ce^{2h\ell}\rho_0$.

Proof. The observation (25) implies that $r \mapsto p(X \backslash K_r)$ is non-increasing. Thus it suffices to consider $r = Ce^{2h\ell}\rho_0$. If $x \in X_+$ is such that $B_x^{-1}(\mathbb{D}(0, r)) \subset \mathbb{D}(0, e^{-2hs_x}r)$, then B_x^{-1} is a contraction that maps $\mathbb{D}(0, r)$ inside itself. Consequently, B_x has a fixed point in $\mathbb{D}(0, r)$; in other words, $x \in \Gamma(0, r)$. This proves that

$$X_+ \backslash \Gamma(0, r) \subset X \backslash K_r.$$

Then the claim follows from Lemma 6.3(b). □

LEMMA 6.6.
(a) $\eta(\mathbb{D}(0, c\hat{\rho})) \leq 2s\eta(\mathbb{D}(0, \varepsilon^{-1}) \backslash \mathbb{D}(0, c\hat{\rho}))$ *for all* $\hat{\rho} \in [0, \rho)$.
(b) $\eta(\mathbb{D}(0, C\rho_0)) \leq 2s\eta(\mathbb{D}(0, \varepsilon^{-1}) \backslash \mathbb{D}(0, C\rho_0))$.

Proof. Let $K = K_{\hat{\rho}}$ for some $\hat{\rho} \in [0, \rho)$. The assumption of Case 2 together with (25) imply that $p(X \backslash K) > \alpha$. So $q(X \backslash K) > \alpha/2 = 1/(2s)$ for every q in a neighborhood of p. Since η is stationary,

$$\begin{aligned}\int_{X \backslash K} (\eta(\mathbb{D}(0, c\hat{\rho})) - \eta(B_x^{-1}(\mathbb{D}(0, c\hat{\rho})))) \, dp(x) \\ = \int_K (\eta(B_x^{-1}(\mathbb{D}(0, c\hat{\rho}))) - \eta(\mathbb{D}(0, c\hat{\rho}))) \, dp(x). \quad (32)\end{aligned}$$

Reducing $\gamma > 0$, if necessary (depending only on A) we may assume that the assertions of Corollary 4.8 and Lemma 4.9 hold in this setting: in particular (taking $r = \hat{\rho}$ in Lemma 4.9),

$$B_x^{-1}(\mathbb{D}(0, c\hat{\rho})) \subset B_x^{-1}(\mathbb{D}(0, 1)) \subset \mathbb{D}(0, \varepsilon^{-1}) \quad \text{and} \quad B_x^{-1}(\mathbb{D}(0, c\hat{\rho})) \cap \mathbb{D}(0, c\hat{\rho}) = \emptyset$$

for every $x \in X \setminus K$. Consequently,

$$\eta(B_x^{-1}(\mathbb{D}(0, c\hat{\rho}))) \le \eta(\mathbb{D}(0, \varepsilon^{-1}) \setminus \mathbb{D}(0, c\hat{\rho})) \quad \text{for every } x \in X \setminus K.$$

For every $x \in X$, we have the general inequality

$$\begin{aligned} \eta(B_x^{-1}(\mathbb{D}(0, c\hat{\rho}))) - \eta(\mathbb{D}(0, c\hat{\rho})) &\le \eta(\mathbb{D}(0, \varepsilon^{-1})) - \eta(\mathbb{D}(0, c\hat{\rho})) \\ &= \eta(\mathbb{D}(0, \varepsilon^{-1}) \setminus \mathbb{D}(0, c\hat{\rho})). \end{aligned}$$

Replacing the last two estimates on the left-hand side and the right-hand side of (32), respectively, we obtain that

$$q(X \setminus K)(\eta(\mathbb{D}(0, c\hat{\rho})) - \eta(\mathbb{D}(0, \varepsilon^{-1}) \setminus \mathbb{D}(0, c\hat{\rho}))) \le q(K)\eta(\mathbb{D}(0, \varepsilon^{-1}) \setminus \mathbb{D}(0, c\hat{\rho})).$$

This yields

$$\eta(\mathbb{D}(0, c\hat{\rho})) \le q(X \setminus K)^{-1}\eta(\mathbb{D}(0, \varepsilon^{-1}) \setminus \mathbb{D}(0, c\hat{\rho})) \le 2s\eta(\mathbb{D}(0, \varepsilon^{-1}) \setminus \mathbb{D}(0, c\hat{\rho})),$$

as we wanted to prove. This gives part (a).

Part (b) follows from the same arguments, with $\hat{\rho}$ replaced by $Ce^{2h\ell}\rho_0$ and

$$K = \{x \in X : B_x^{-1}(\mathbb{D}(0, Ce^{2h\ell}\rho_0)) \subset \mathbb{D}(0, Ce^{-2hs_x + 2h\ell}\rho_0)\}$$

instead. By Corollary 6.5, $p(X \setminus K) \ge \alpha$ and so $q(X \setminus K) \ge \alpha/2 = 1/(2s)$ for every q in a neighborhood of p. Since $\mathbb{D}(0, Ce^{2h\ell}\rho_0) \subset \mathbb{D}(0, 1)$, Corollary 4.8 implies that the preimage of $\mathbb{D}(0, Ce^{2h\ell}\rho_0)$ under any B_x is contained in $\mathbb{D}(0, \varepsilon^{-1})$. So the same arguments as in the previous paragraph yield

$$\eta(\mathbb{D}(0, cCe^{2h\ell}\rho_0)) \le s\eta(\mathbb{D}(0, \varepsilon^{-1}) \setminus \mathbb{D}(0, cCe^{2h\ell}\rho_0)).$$

Since $ce^{2h\ell} \ge 1$, this implies the conclusion in part (b) of the lemma. □

LEMMA 6.7. *For any $C\rho_0 \le r \le 1$,*

$$\eta(\mathbb{D}(0, \varepsilon^{-1}) \setminus \mathbb{D}(0, e^{-2h}r)) \le (1 + \omega)\eta(\mathbb{D}(0, \varepsilon^{-1}) \setminus \mathbb{D}(0, r)).$$

Proof. Lemma 4.10 implies that

$$\begin{aligned} q(\Gamma(0, \rho_0))\eta(\mathbb{D}(0, r) \setminus \mathbb{D}(0, e^{-2h}r)) &= \int_{\Gamma(0,\rho_0)} (\eta(\mathbb{D}(0, r)) - \eta(\mathbb{D}(0, e^{-2h}r)))\, dq(x) \\ &\le \int_{\Gamma(0,\rho_0)} (\eta(\mathbb{D}(0, r)) - \eta(B_x^{-1}(\mathbb{D}(0, r))))\, dq(x). \end{aligned}$$

Since η is stationary, the last expression coincides with

$$\int_{X \setminus \Gamma(0,\rho_0)} (\eta(B_x^{-1}(\mathbb{D}(0, r))) - \eta(\mathbb{D}(0, r)))\, dq(x) \le \eta(\mathbb{D}(0, \varepsilon^{-1}) \setminus \mathbb{D}(0, r)).$$

Putting these two inequalities together,

$$\omega^{-1}\eta(\mathbb{D}(0, r) \setminus \mathbb{D}(0, e^{-2h}r)) \le \eta(\mathbb{D}(0, \varepsilon^{-1}) \setminus \mathbb{D}(0, r)).$$

This implies the claim in the lemma. □

The next corollary completes the proof of Proposition 3.8 when $z_0 = 0$. Observe that the constant $\kappa > 0$ in the statement depends only on A.

COROLLARY 6.8. *$\eta(\mathbb{D}(0, \varepsilon^{-1})) \le \kappa\delta$, where $\kappa = 2(1 + 2s)(1 + \omega)^{\ell} > 0$.*

Proof. First, suppose that $Ce^{2h\ell}\rho_0 < \rho$. Then we may apply Lemma 6.7 to every $r = e^{-2hj}\rho$, $j = 0, \ldots, \ell - 1$. So, using also Corollary 5.2 and Lemma 4.1,

$$\eta(\mathbb{D}(0, \varepsilon^{-1})\backslash\mathbb{D}(0, e^{-2h\ell}\rho)) \le (1 + \omega)^{\ell}\eta(\mathbb{D}(0, \varepsilon^{-1})\backslash\mathbb{D}(0, \rho)) \le 2\delta(1 + \omega)^{\ell}.$$

Choose $\hat{\rho} \in [Ce^{2h\ell}\rho_0, \rho)$ close enough to ρ that $c\hat{\rho} \ge e^{-2h\ell}\rho$ (keep in mind that $c > e^{-2h\ell}$, by the definition of ℓ). Then

$$\eta(\mathbb{D}(0, \varepsilon^{-1})\backslash\mathbb{D}(0, c\hat{\rho})) \le \eta(\mathbb{D}(0, \varepsilon^{-1})\backslash\mathbb{D}(0, e^{-2h\ell}\rho)) \le 2\delta(1 + \omega)^{\ell}.$$

Combining this with Lemma 6.6(a), we find that $\eta(\mathbb{D}(0, c\hat{\rho})) \le 4s(1 + \omega)^{\ell}\delta$. Adding these last two inequalities, we obtain that

$$\eta(\mathbb{D}(0, \varepsilon^{-1})) \le 2(1 + 2s)(1 + \omega)^{\ell}\delta. \tag{33}$$

This proves the claim in this case.

Now suppose that $Ce^{2h\ell}\rho_0 \ge \rho$ (in particular, $\rho_0 > 0$). Then, just as before,

$$\begin{aligned}\eta(\mathbb{D}(0, \varepsilon^{-1})\backslash\mathbb{D}(0, C\rho_0)) &\le (1 + \omega)^{\ell}\eta(\mathbb{D}(0, \varepsilon^{-1})\backslash\mathbb{D}(0, Ce^{2h\ell}\rho_0))\\ &\le (1 + \omega)^{\ell}\eta(\mathbb{D}(0, \varepsilon^{-1})\backslash\mathbb{D}(0, \rho)) \le 2(1 + \omega)^{\ell}\delta.\end{aligned}$$

Lemma 6.6(b) gives $\eta(\mathbb{D}(0, C\rho_0)) \le 4s(1 + \omega)^{\ell}\delta$. Adding these two inequalities,

$$\eta(\mathbb{D}(0, \varepsilon^{-1})) \le 2(1 + 2s)(1 + \omega)^{\ell}\delta. \tag{34}$$

The inequalities (33) and (34) imply the conclusion of the corollary. □

To finish, let us explain how the assumption $z_0 = 0$ can be removed.

As observed in Remark 6.4, the point z_0 is necessarily close to zero if B is close to A. Then $H : \overline{\mathbb{C}} \to \overline{\mathbb{C}}$, $H(z) = z - z_0$ is uniformly close to the identity, and so the cocycle $\tilde{B}$ defined by $\tilde{B}_x = H \cdot B_x \cdot H^{-1}$ is uniformly close to B. A measure η is (B, q)-stationary if and only if $\tilde{\eta} = H_*\eta$ is $(\tilde{B}, q)$-stationary. It is clear that $q(\{x \in X : B_x(z) = z\}) < 1$ for all z if and only if $q(\{x \in X : \tilde{B}_x(z) = z\}) < 1$ for all z. Analogously, the set $\tilde{\Gamma}(z, r)$ of points x such that the fixed point of $\tilde{B}_x$ is in $\mathbb{D}(z, r)$ coincides with $\Gamma(z + z_0, r)$ for every z and r. In particular, by Lemma 6.3,

$$p(\tilde{\Gamma}(0, \rho_0)) \ge 2\omega^{-1} \quad \text{and} \quad p(X_+\backslash\tilde{\Gamma}(0, Ce^{2h\ell}\rho_0)) \ge \alpha \text{ if } \rho_0 > 0.$$

So we may apply the previous arguments to $\tilde{B}$, q, and $\tilde{\eta}$, to get that

$$\eta(\mathbb{D}(0, \varepsilon^{-1}) - z_0) = \tilde{\eta}(\mathbb{D}(0, \varepsilon^{-1})) \le (1 + \kappa)\delta \tag{35}$$

for any (B, q)-stationary measure η and any (B, q) that satisfies the assumptions in the present section. Since z_0 is small,

$$(\mathbb{D}(0, \varepsilon^{-1}) - z_0) \cup (\mathbb{D}(0, \varepsilon^{-1})\backslash\mathbb{D}(0, \varepsilon)) \supset \mathbb{D}(0, \varepsilon^{-1}).$$

Thus, combining (35) with Lemma 4.1, we find that $\eta(\mathbb{D}(0, \varepsilon^{-1})) \le (2 + \kappa)\delta$.

The proof of Proposition 3.8 is now complete.

7. *Proof of Theorem B*

Let λ be the Lebesgue measure on the unit interval I, and let $\|\eta\|$ denote the total variation of a signed measure η.

LEMMA 7.1. (Avila) *Let Y be a metric space such that every bounded closed subset is compact, and let ν be any Borel probability measure on Y such that the support $Z = \operatorname{supp} \nu$ is bounded.*

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

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 $\operatorname{diam} Q_i < 12\delta$. This can be seen as follows. For each $x \in Z$ take $r_x \in (\delta, 2\delta)$ such that $\nu(\partial \mathbb{D}(x, r_x)) = 0$. Then $\{\mathbb{D}(x, r_x) : x \in Z\}$ 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, $\operatorname{diam} V_i < 4\delta$ and $\nu(V_i) > 0$ and $\nu(\partial V_i) = 0$ for every i. Consider the partition $\mathcal{P}$ of $\bigcup_{i=1}^k V_i$ 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 $Q_2 \subset Y$ 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 $2 \le l \le k$, assume that $Q_1, \dots, Q_{l-1}$ have been defined, and then let $Q_l = \emptyset$ if $V_l \subset \bigcup_{i=1}^{l-1} Q_i$ and

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

if $\nu(V_l \setminus \bigcup_{i=1}^{l-1} Q_i) > 0$. Those of the 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 so that

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

Let $Z_i = \operatorname{supp} \nu \cap Q_i$ 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 μ 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 $I_i = I \setminus \bigcup_{j\in J(i)} I_{i,j}$. 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 Y is a polish space, that is, a complete separable metric space. Since all Borel non-atomic probabilities on polish spaces are isomorphic (see Ito [**16**, § 2.4] or [**32**, Theorem 8.5.4]), the previous equality ensures that there exists an invertible measurable map

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

mapping the restriction of q to the restriction of μ. By setting $\phi \equiv a_{i,j}$ on each $Z_i \times I_{i,j}$ we extend ϕ_i to a measurable map $Z_i \times I \to Q_i$ that still sends the restriction of q to the restriction of μ. Gluing all these extensions, we obtain a measurable map $\phi : Z \times I \to X$ 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 $d(\phi(x, t), x) \le \operatorname{diam} Q_i < \varepsilon$ for all $(x, t) \in Z \times I$. Finally,

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

The proof of the lemma is complete. □

Now, given $\rho > 0$, let ν be a probability measure in $Y = \mathrm{GL}(2, \mathbb{C})$ with compact support. Consider $X = \operatorname{supp} \nu \times I$, $p = \nu \times \lambda$ and $A : X \to \mathrm{GL}(2, \mathbb{C})$ 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 7.1 implies that there exist a weak* neighborhood V and δ such that, if $\nu' \in V$ and $\operatorname{supp} \nu' \subset B_\delta(\operatorname{supp} \nu)$, there exist $B : X \to \mathrm{GL}(2, \mathbb{C})$ and a probability measure q on 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.

8. *An example of discontinuity*

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 [**4**, **5**, **7**, **23**], where it is shown that Lyapunov exponents are often discontinuous relative to the C^0 topology.

Let $M = \Sigma_2$ be the shift with two symbols, endowed with the metric $d(\mathbf{x}, \mathbf{y}) = 2^{-N(\mathbf{x},\mathbf{y})}$, where

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

For any $r \in (0, \infty)$, the H^r norm in the space of r-Hölder continuous functions $L : M \to \mathcal{L}(\mathbb{C}^d, \mathbb{C}^d)$ 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 μ associated with an arbitrary probability vector $p = (p_1, p_2)$ with positive entries.

Given any $\sigma > 1$, consider the (locally constant) cocycle $A : M \to \mathrm{SL}(2, \mathbb{R})$ defined by

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

Observe that the Lyapunov exponents are given by $\lambda_\pm(A, p) = \pm|p_1 - p_2| \log \sigma$. In particular, they are non-zero if $p_1 \neq p_2$. Then it follows from the next theorem that (A, p) is a point of *discontinuity* for the Lyapunov exponents relative to the H^r topology.

THEOREM 8.1. *For any $r > 0$ such that $2^{2r} < \sigma$ there exist $B : M \to \mathrm{SL}(2, \mathbb{R})$ with vanishing Lyapunov exponents and such that $\|A - B\|_r$ is arbitrarily close to zero.*

The proof of Theorem 8.1 is an adaptation of ideas of Knill [**21**] and Bochi [**4**, **5**]. Here is an outline. 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 $f^i(Z_n)$, $0 \le i \le n-1$ 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

$$\|B_n - A\|_r \le \mathrm{const}(2^{2r}/\sigma)^{n/2} \tag{36}$$

decays to zero as $n \to \infty$. 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 $k \ge 1$ 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 $f^i(Z_n)$, $0 \le i \le 2k$ are pairwise disjoint. Let

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

Define $R : M \to \mathrm{SL}(2, \mathbb{R})$ by

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

and then take $B_n = A R_n$.

LEMMA 8.2. $B_n^n(\mathbf{x}) H_\mathbf{x} = V_{f^n(\mathbf{x})}$ *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 $\mathbf{x} \in Z_n$,

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

The claim follows by iterating one more time. □

LEMMA 8.3. *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, $\sup \|L\| \leq \sup \|A\| \, \|\mathrm{id} - R_n\|$ and this is bounded by $\sigma\varepsilon_n$. Now let us estimate the second term in the definition (36). If $\mathbf{x}$ and $\mathbf{y}$ are not in the same cylinder $[0; a]$, then $d(\mathbf{x}, \mathbf{y}) = 1$, and so

$$\frac{\|L_n(\mathbf{x}) - L_n(\mathbf{y})\|}{d(\mathbf{x}, \mathbf{y})^r} \leq 2 \sup \|L_n\| \leq 2\sigma\varepsilon_n. \tag{37}$$

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 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 $\|\mathrm{id} - R_n\| \leq \varepsilon_n$ 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 worse than (37), 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 $C = 3\sigma$. □

We want to prove that $\lambda_{\pm}(B_n) = 0$ for every n. Let μ_n be the normalized restriction of μ to Z_n and $f_n : Z_n \to Z_n$ 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 $w = (1, \ldots, 1, 2, \ldots, 2, 2)$ and the union is over the set $\mathcal{B}$ of all finite words $b = (b_1, \ldots, b_s)$ not having w as a subword. Moreover,

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

Thus (f_n, μ_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 $\hat{B}_n : Z_n \to \mathrm{SL}(2, \mathbb{R})$ be the function induced by B_n over f_n: that is,

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

It is a well-known basic fact (see [**30**, Proposition 2.9], for instance) that the Lyapunov spectrum of the induced function is obtained by multiplying the Lyapunov spectrum of the original function 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 $E^u_{\mathbf{x}} \oplus E^s_{\mathbf{x}}$ be the Oseledets splitting (defined almost everywhere in Z_n). Consider the probability measures m^u and m^s defined on $Z_n \times \mathbb{P}(\mathbb{R}^2)$ by

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

for $* \in \{s, u\}$ and any measurable subset B of $Z_n \times \mathbb{P}(\mathbb{R}^2)$. The key observation is that, as a consequence of Lemma 8.2, the cocycle

$$F_{\hat{B}_n} : Z_n \times \mathbb{P}(\mathbb{R}^2) \to Z_n \times \mathbb{P}(\mathbb{R}^2), \quad F_{\hat{B}_n}(x, v) = (f_n(x), \hat{B}_n(x)v)$$

permutes the vertical and horizontal subbundles: that is

$$\hat{B}_n(\mathbf{x})H_{\mathbf{x}} = V_{f_n(\mathbf{x})} \quad \text{and} \quad \hat{B}_n(\mathbf{x})V_{\mathbf{x}} = H_{f_n(\mathbf{x})} \quad \text{for all } \mathbf{x} \in Z_n. \tag{38}$$

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

$$m_n(B) = \tfrac{1}{2}\mu_n(\{\mathbf{x} \in Z_n : (\mathbf{x}, V_{\mathbf{x}}) \in B\}) + \tfrac{1}{2}\mu_n(\{\mathbf{x} \in Z_n : (\mathbf{x}, H_{\mathbf{x}}) \in B\})$$

for any measurable subset B of $Z_n \times \mathbb{P}(\mathbb{R}^2)$. That is, m_n projects down to μ_n and its disintegration is given by $\mathbf{x} \mapsto (\delta_{H_{\mathbf{x}}} + \delta_{V_{\mathbf{x}}})/2$. It is clear from (38) that m_n is $F_{\hat{B}_n}$-invariant.

LEMMA 8.4. *The probability measure m_n is ergodic.*

Proof. Suppose there is an invariant set $X \subset Z_n \times \mathbb{P}(\mathbb{R}^2)$ with $m_n(X) \in (0, 1)$. Let X_0 be the set of $\mathbf{x} \in Z_n$ whose fiber $X \cap (\{\mathbf{x}\} \times \mathbb{P}(\mathbb{R}^2))$ contains neither $(\mathbf{x}, H_{\mathbf{x}})$ nor $(\mathbf{x}, V_{\mathbf{x}})$. In other words, the complement X_0^c is the image of the intersection

$$X \cap \{(\mathbf{x}, [v]) \in Z_n \times \mathbb{P}(\mathbb{R}^2) : [v] = H_{\mathbf{x}} \text{ or } [v] = V_{\mathbf{x}}\}$$

under the canonical projection $\pi : Z_n \times \mathbb{P}(\mathbb{R}^2) \to Z_n$. Since this intersection is a measurable subset of $Z_n \times \mathbb{P}(\mathbb{R}^2)$ and $\mathbb{P}(\mathbb{R}^2)$ is a polish space, we may use [**10**, Theorem III.23] (see [**31**, Proposition 4.5]) to conclude that X_0^c is a measurable subset of Z_n, up to zero μ_n-measure. Thus the same is true about X_0.

In view of (38), X_0 is an f_n-invariant set and so its μ_n-measure is either zero or one. Since $m_n(X) > 0$, 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 $(\mathbf{x}, H_{\mathbf{x}})$ and $(\mathbf{x}, V_{\mathbf{x}})$. Now let X_H be the set of $\mathbf{x} \in Z_n$ whose fiber contains $(\mathbf{x}, H_{\mathbf{x}})$ but not $(\mathbf{x}, V_{\mathbf{x}})$, and let X_V be the set of $\mathbf{x} \in Z_n$ whose fiber contains $(\mathbf{x}, V_{\mathbf{x}})$ but not $(\mathbf{x}, H_{\mathbf{x}})$. The previous observations show that $X_H \cup X_V$ has full μ_n-measure and it follows from (38) 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. □

It is easy to see that m_n is a convex combination of the probabilities m^u and m^s. Indeed, given $\kappa > 0$, define X_κ to be the set of all $(\mathbf{x}, [v]) \in Z_n \times \mathbb{P}(\mathbb{R}^2)$ such that the Oseledets splitting $E^u_{\mathbf{x}} \oplus E^s_{\mathbf{x}}$ 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_κ returns at most a finite number of times to X_κ. So, by Poincaré recurrence, $m_n(X_\kappa) = 0$ for every κ. This means that m_n gives full weight to $\{(\mathbf{x}, E^u_{\mathbf{x}}), (\mathbf{x}, E^s_{\mathbf{x}}) : \mathbf{x} \in Z_n\}$ and so it is a convex combination of m^u and m^s.

Then, by Lemma 8.4, m_n must coincide with either m^s or 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 or m^s are Dirac masses on a single point. This contradiction proves that the Lyapunov exponents of B_n do vanish for every n, and that concludes the proof of Theorem 8.1.

The same kind of argument shows that, in general, one cannot expect continuity to hold when some of the probabilities p_i vanishes.

Remark 8.5. [**18**] 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 (38). Now it is clear that $\lambda_\pm(A, (1, 0)) = \pm\log\sigma$. Thus the Lyapunov exponents are discontinuous at $(A, (1, 0))$.

Remark 8.6. A variation of the previous idea yields another example of discontinuity, relative to the L^q-topology, any $q \in [1, \infty)$. Let $X = \mathbb{N}$ and p be supported on the whole X. Define

$$A_x \equiv \begin{pmatrix} 2 & 0 \\ 0 & 2^{-1} \end{pmatrix} \quad \text{and} \quad A_k(x) = \begin{cases} A_x & \text{if } x \neq k, \\ R_{\pi/2} & \text{otherwise}, \end{cases}$$

where $R_{\pi/2}$ is the rotation by $\pi/2$. Note that $(A_k)_k \to A$ in the L^p sense. However, $\lambda_+(A_k) = 0$ for every k, whereas $\lambda_+(A) = \log 2$.

Acknowledgements. We are grateful to Artur Avila, Jairo Bochi and Jiagang Yang for several useful conversations. Lemma 7.1 is due to Artur Avila. We are also grateful to the anonymous referee for a thorough revision of the paper that helped to improve the presentation. C.B.-N. was supported by a CNPq and FAPERJ doctoral scholarship. M.V. is partially supported by CNPq, FAPERJ and PRONEX-Dynamical Systems.

REFERENCES

[**1**] A. Arbieto and J. Bochi. L^p-generic cocycles have one-point Lyapunov spectrum. *Stoch. Dyn.* **3** (2003), 73–81.

[**2**] L. Arnold and N. D. Cong. On the simplicity of the Lyapunov spectrum of products of random matrices. *Ergod. Th. & Dynam. Sys.* **17** (1997), 1005–1025.

[**3**] A. Avila and M. Viana. Extremal Lyapunov exponents: an invariance principle and applications. *Invent. Math.* **181**(1) (2010), 115–189.

[**4**] J. Bochi. Discontinuity of the Lyapunov exponents for non-hyperbolic cocycles. *Preprint*, www.mat.puc-rio.br/∼jairo/.

[**5**] J. Bochi. Genericity of zero Lyapunov exponents. *Ergod. Th. & Dynam. Sys.* **22** (2002), 1667–1696.

[**6**] J. Bochi. C^1-generic symplectic diffeomorphisms: partial hyperbolicity and zero centre Lyapunov exponents. *J. Inst. Math. Jussieu* **8** (2009), 49–93.

[**7**] J. Bochi and M. Viana. The Lyapunov exponents of generic volume-preserving and symplectic maps. *Ann. of Math. (2)* **161** (2005), 1423–1485.

[**8**] 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** (2005), 313–355.

[**9**] J. Bourgain and S. Jitomirskaya. Continuity of the Lyapunov exponent for quasiperiodic operators with analytic potential. *J. Stat. Phys.* **108** (2002), 1203–1218.

[**10**] C. Castaing and M. Valadier. *Convex Analysis and Measurable Multifunctions (Lecture Notes in Mathematics, 580)*. Springer, 1977.

[**11**] H. Furstenberg and H. Kesten. Products of random matrices. *Ann. Math. Statist.* **31** (1960), 457–469.

[**12**] H. Furstenberg and Yu. Kifer. Random matrix products and measures in projective spaces. *Israel J. Math.* **10** (1983), 12–32.

[**13**] I. Ya. Gol'dsheid and G. A. Margulis. Lyapunov indices of a product of random matrices. *Uspekhi Mat. Nauk* **44** (1989), 13–60.

[**14**] Y. Guivarc'h and A. Raugi. Products of random matrices: convergence theorems. *Contemp. Math.* **50** (1986), 31–54.

[**15**] H. Hennion. Loi des grands nombres et perturbations pour des produits réductibles de matrices aléatoires indépendantes. *Z. Wahrsch. Verw. Gebiete* **67** (1984), 265–278.

[**16**] K. Itō. *Introduction to Probability Theory*. Cambridge University Press, 1984.

[**17**] R. Johnson. Lyapounov numbers for the almost periodic Schrödinger equation. *Illinois J. Math.* **28** (1984), 397–419.

[**18**] Yu. Kifer. Perturbations of random matrix products. *Z. Wahrsch. Verw. Gebiete* **61** (1982), 83–95.

[**19**] Yu. Kifer and E. Slud. Perturbations of random matrix products in a reducible case. *Ergod. Th. & Dynam. Sys.* **2** (1983), 367–382, 1982.

[**20**] J. Kingman. The ergodic theory of subadditive stochastic processes. *J. Roy. Statist. Soc.* **30** (1968), 499–510.

[**21**] O. Knill. The upper Lyapunov exponent of SL(2, **R**) cocycles: discontinuity and the problem of positivity. *Lyapunov Exponents (Oberwolfach, 1990) (Lecture Notes in Mathematics, 1486)*. Springer, 1991, pp. 86–97.

[**22**] O. Knill. Positive Lyapunov exponents for a dense set of bounded measurable $SL(2, \mathbf{R})$-cocycles. *Ergod. Th. & Dynam. Sys.* **12** (1992), 319–331.

[**23**] R. Mañé. Oseledec's theorem from the generic viewpoint. *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Warsaw, 1983)*. PWN Publishers, Warsaw, 1984, pp. 1269–1276.

[**24**] V. I. Oseledets. A multiplicative ergodic theorem: Lyapunov characteristic numbers for dynamical systems. *Trans. Moscow Math. Soc.* **19** (1968), 197–231.

[**25**] É. Le Page. Théorèmes limites pour les produits de matrices aléatoires. *Probability Measures on Groups (Oberwolfach, 1981) (Lecture Notes in Mathematics, 928)*. Springer, 1982, pp. 258–303.

[**26**] É. Le Page. Régularité du plus grand exposant caractéristique des produits de matrices aléatoires indépendantes et applications. *Ann. Inst. H. Poincaré Probab. Stat.* **25** (1989), 109–142.

[**27**] Y. Peres. Analytic dependence of Lyapunov exponents on transition probabilities. *Lyapunov Exponents (Oberwolfach, 1990) (Lecture Notes in Mathematics, 1486)*. Springer, 1991, pp. 64–80.

[**28**] D. Ruelle. Analyticity properties of the characteristic exponents of random matrix products. *Adv. Math.* **32** (1979), 68–80.

[**29**] B. Simon and M. Taylor. Harmonic analysis on SL(2, **R**) and smoothness of the density of states in the one-dimensional Anderson model. *Comm. Math. Phys.* **101** (1985), 1–19.

[**30**] M. Viana. Lyapunov exponents of Teichmüller flows. *Partially Hyperbolic Dynamics, Laminations, and Teichmüller Flow (Fields Institute Communications, 51)*. American Mathematical Society, 2007, pp. 139–201.

[**31**] M. Viana. *Lectures on Lyapunov Exponents*. Cambridge University Press, 2014.

[**32**] M. Viana and K. Oliveira. *Foundations of Ergodic Theory*. Cambridge University Press, 2015.

[**33**] L.-S. Young. Random perturbations of matrix cocycles. *Ergod. Th. & Dynam. Sys.* **6** (1986), 627–637.

AUTHOR QUERIES

Q1 (page 1)
The distinction between surnames can be ambiguous, therefore to ensure accurate tagging for indexing purposes online (eg for PubMed entries), please check that the highlighted surnames have been correctly identified, that all names are in the correct order and spelt correctly.

Ok! All names are correct.

Q2 (page 3)
Please confirm whether $\wedge^k$ or Λ^k is OK here.

The correct way is marked in yellow (\wedge).

Q3 (page 13)
Extra closing bracket deleted. Please check.

Ok, thanks!

Q4 (page 30)
Please update preprint information for Ref. [4], if possible.

Replace that link by http://www.mat.uc.cl/~jairo.bochi/docs/discont.pdf

Q5 (page 30)
Are there any particular pages for refs [10], [16], [31] and [32] please?

more precise refs were inserted directly in the text (e.g. page 6, line 20).

Q6 (page 30)
Please provide the place of publisher for the Refs [10, 16, 21, 25, 27, 31, 32].

10, 21, 25, 27 : Heidelberg
16, 31, 32: Cambridge, UK