CONTINUITY OF LYAPUNOV EXPONENTS
FOR RANDOM 2D MATRICES

CARLOS BOCKER-NETO AND MARCELO VIANA

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 2-by-2 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, Kesten [10], that there exist numbers $\lambda_+$ and $\lambda_-$ such that

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

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 $\lambda_+$ and $\lambda_-$ vary continuously with the coefficients of $(A_1, \ldots, A_m, p_1, \ldots, p_m)$ at all points.

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

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

The situation in Theorem A corresponds to the special case when the measures have finite supports: $\nu = p_1 \delta_{A_1} + \cdots + p_m \delta_{A_m}$ and $\nu' = p'_1 \delta_{A'_1} + \cdots + p'_m \delta_{A'_m}$. Clearly, the support of $\nu'$ is Hausdorff close to the support of $\nu$ if $A'_i$ is close to $A_i$, $p_i$ for all $i$. In this regard, recall that we assume that all $p_i > 0$: the conclusion of Theorem A may fail if this condition is removed (see Remark 8.5).

Date: September 24, 2013.

C.B.-N. was supported by a CNPq and FAPERJ doctoral scholarship. M.V. is partially supported by CNPq, FAPERJ, and PRONEX-Dynamical Systems.
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 [22], 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 Section 8 indicates that similar phenomena may occur also for more regular cocycles. A detailed discussion of these and related issues will appear in Section 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 \( F : \mathcal{V} \to \mathcal{V}_{f(x)} \) 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 \( \mu \) be an \( f \)-invariant probability measure on \( M \) with \( \log \|(F_x)^{\pm 1}\| \in L^1(\mu) \). It follows from the sub-additive ergodic theorem (Kingman [19]) that the extremal Lyapunov exponents

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

are well-defined \( \mu \)-almost everywhere.

The theorem of Oseledets [23] provides a more detailed statement. Namely, at \( \mu \)-almost every point \( x \in M \), there exist numbers \( \hat{\lambda}_1(F, x) > \cdots > \hat{\lambda}_{k(x)}(F, x) \) and linear subspaces \( \mathcal{V}_x = V_x^1 > V_x^2 > \cdots > V_x^{k(x)} > \{0\} = V_x^{k(x)+1} \) such that

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

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

\[
F_x(E_x^j) = E_x^{j+1} \quad \text{and} \quad \lim_{n \to \infty} \frac{1}{n} \log \|F^n_x(v)\| = \hat{\lambda}_j(F, x) \quad \text{for all } v \in E_x^j \setminus \{0\}.
\]

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

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

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

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, \( d = \dim \mathcal{V} \). The average Lyapunov exponents of \( F \) are defined by

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

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

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

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

where $\Lambda^k$ denotes the $k$th exterior power. Similarly, the sum of the $k$ smallest Lyapunov exponents is always lower semi-continuous.

However, Lyapunov exponents are, usually, discontinuous functions of the data. A number of results, both positive and negative, will be recalled in a while. 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 $A : X \to \text{GL}(2, \mathbb{C})$ 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

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

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

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

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

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

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

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

We prove Theorem C in Sections 3 through 6, and we deduce Theorem B from it in Section 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 $\nu'$ is contained in a small neighborhood of the support of $\nu$; moreover, $\lambda_{\pm}(A, p) = \lambda_{\pm}(\nu)$ and $\lambda_{\pm}(B, q) = \lambda_{\pm}(\nu')$. In Section 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
$B : X \to \text{GL}(2, \mathbb{C})$, let $E^{s}_{B,x}$ and $E^{u}_{B,x}$ be the Oseledets subspaces of the corresponding cocycle at a point $x \in M$ (when they exist). Assume that $\lambda_{-}(A, p) < \lambda_{+}(A, p)$. Then, for any $\varepsilon > 0$,

$$\mu\left( \{ x \in M : \angle(E^{u}_{A,x}, E^{u}_{B,x}) < \varepsilon \text{ and } \angle(E^{s}_{A,x}, E^{s}_{B,x}) < \varepsilon \} \right)$$

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 [27] proved real-analytic dependence of the largest exponent on the cocycle, for linear cocycles admitting an invariant convex cone field. Short afterwards, Furstenberg, Kifer [11, 17] and Hennion [14] proved continuity of the largest exponent of i.i.d. random matrices, under a condition of almost irreducibility. Some reducible cases were treated by Kifer and Slud [17, 18], who also observed that discontinuities may occur when the probability vector degenerates ([17], see Remark 8.5 below). Stability of Lyapunov exponents under certain random perturbations was obtained by Young [31].

For i.i.d. random matrices satisfying strong irreducibility and the contraction property, Le Page [24, 25] proved local Hölder continuity, and even smoothness, of the largest exponent on the cocycle; the assumptions ensure that the largest exponent is simple (multiplicity 1), by work of Guivarc’h, Raugi [13] and Gol’dsheid, Margulis [12]. For i.i.d. random matrices over Bernoulli and Markov shifts, Peres [26] showed that simple exponents are locally real-analytic functions of the transition data.

A construction of Halperin quoted by Simon, Taylor [28] shows that for every $\alpha > 0$ one can find random Schrödinger cocycles

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

(the $V_n$ are i.i.d. random variables) near which the exponents fail to be $\alpha$-Hölder continuous. Thus, the previously mentioned results of Le Page can not be improved. Johnson [16] found examples of discontinuous dependence of the exponent on the energy $E$, for Schrödinger cocycles over quasi-periodic flows. Recently, Bourgain, Jitomirskaya [8, 9] proved continuous dependence of the exponents on the energy $E$, for one-dimensional quasi-periodic Schrödinger cocycles: $V_n = V(f^n(\theta))$ where $V : S^1 \to \mathbb{R}$ 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 [20, 21] considered $L^\infty$ cocycles with values in $\text{SL}(2, \mathbb{R})$ and proved that, as long as the base dynamics is aperiodic, discontinuities always exist: the set of cocycles with non-zero exponents is never open. This was refined to the continuous case by Bochi [4, 5]: an $\text{SL}(2, \mathbb{R})$-cocycle is a continuity point in the $C^0$ topology if and only if it is uniformly hyperbolic or else the exponents vanish. This statement was inspired by Mañé’s surprising announcement in [22]. Indeed, and most strikingly, the theorem of Mañé-Bochi [5, 22] remains true restricted to the subset of $C^0$ derivative cocycles, that is, of the form $F = Df$ for some $C^1$ area preserving diffeomorphism $f$. Moreover, this has been extended to cocycles and diffeomorphisms in arbitrary dimension, by Bochi, Viana [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, Cong [2], Arbieto, Bochi [1]. Consequently, they are precisely the continuity points for the Lyapunov exponents relative to the \(L^p\) topology.

These results show that discontinuity of Lyapunov exponents is quite common among cocycles with low regularity. Locally constant cocycles, as we deal with here, sit at the opposite end of the regularity spectrum, and the results in the present paper show that in this context continuity does hold at every point. For cocycles with intermediate regularities the continuity problem is very much open. However, our construction in Section 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, Viana [3] studied the continuity of the Lyapunov exponents in the very broad context of smooth cocycles. The continuity criterium in [3, Section 6] was the starting point for the proof of our Theorem C.

**Acknowledgements.** We are grateful to Artur Avila, Jairo Bochi, and Jiagang Yang for several useful conversations. Lemma 7.1 is due to Artur Avila.

## 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 Section 6.

Let \(\mathcal{P}(X)\) be the space of probability measures on \(X\) and let \(\mathcal{G}(X)\) and \(\mathcal{S}(X)\) denote the spaces of bounded measurable functions from \(X\) to \(\text{GL}(2, \mathbb{C})\) and \(\text{SL}(2, \mathbb{C})\), 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 Sections 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 \(P \in \text{SL}(2, \mathbb{C})\) and \(\theta : X \to \mathbb{C} \setminus \{0\}\) such that

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

for every \(x\) in some full \(p\)-measure set \(Z \subset X\).
Proposition 3.1 is contained in the main results of Furstenberg, Kifer [11] and Hennion [14], as well as in Proposition 6.3 of Avila, Viana [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 sequel. For the details, see the aforementioned papers or Chapter 5 of [29].

Given \((A,p)\) in \(\mathcal{S}(X) \times \mathcal{P}(X)\), a probability measure \(\eta\) on \(\mathbb{P}(\mathbb{C}^2)\) is called \((A,p)\)-stationary if
\[
\int \psi(\xi) \, d\eta(\xi) = \int \int \psi(A_x \xi) \, d\eta(\xi) \, dp(x)
\]
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 \(\text{Stat}(A,p)\) of \((A,p)\)-stationary measures is always nonempty: that is because \(\eta \mapsto \int (A_x)_* \eta \, dp(x)\) is a continuous operator in the space of probability measures of on \(\mathbb{P}(\mathbb{C}^2)\) and so, by Tychonoff - Schauder, it has some fixed point. Another useful property is that \(\text{Stat}(A,p)\) varies in a semi-continuous fashion with the data \((A,p)\):

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

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
\[
\phi : M \times \mathbb{P}(\mathbb{C}^2) \to \mathbb{R}, \quad \phi(x, [v]) = \log \frac{\|A_{x_0} v\|}{\|v\|}.
\]
Since \(\phi\) 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 \((A,p) \in \mathcal{S}(X) \times \mathcal{P}(X)\),
\[
\lambda_+(A,p) = \max \left\{ \int \phi(x, \xi) \, d\eta(\xi) \, dp(x) : \eta \in \text{Stat}(A,p) \right\}.
\]

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

**Lemma 3.4.** For any \((A,p) \in \mathcal{S}(X) \times \mathcal{P}(X)\), if \(\eta \in \text{Stat}(A,p)\) 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
\[
\lim_{n \to \pm \infty} \frac{1}{n} \log \|A^n_x v\| = \lambda_-(A,p) \quad \text{for} \ v \in L \quad \text{and} \ p\text{-almost every} \ x.
\]

We call \((A,p)\) irreducible if there exists no \((A,p)\)-invariant subspace, that is, no one-dimensional subspace \(L \subset \mathbb{C}^2\) 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
\[
\lambda_+(A,p) = \int \phi(x, \xi) \, d\eta(\xi) \, dp(x) \quad \text{for every} \ \eta \in \text{Stat}(A,p).
\]
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 \(S(X) \times P(X)\). By Lemma 3.3, for each \(k\) there exists some \(\eta_k \in \text{Stat}(A_k, p_k)\) that realizes the largest Lyapunov exponent:

\[
\lambda_+(A_k, p_k) = \int \phi_k(x, \xi) d\eta_k(\xi) dp_k(x), \quad \phi_k(x, [v]) = \log \left\| A_k x v \right\| / \left\| v \right\|.
\]

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

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

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

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

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

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

Notice that (3) \& (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 \((A, p) \in S(X) \times P(X)\) 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 \(P = \text{id}\) and

\[
A_x = \left( \begin{array}{cc} \theta_x & 0 \\ 0 & \theta_x^{-1} \end{array} \right) \quad \text{for all } x \in Z.
\]

We will always consider pairs \((B, q) \in V(A, p, \gamma, Z)\), that give full weight to \(Z\). Thus, it is no restriction either 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, \(\delta_h\) and \(\delta_v\), are both \((A, p)\)-stationary measures. In particular, \(\text{Stat}(A, p)\) 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 \(\text{Stat}(A, p)\) that realizes the Lyapunov exponent \(\lambda_+(A, p)\), namely the Dirac mass \(\delta_h:\)

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

Let us check that Proposition 3.6 is a consequence. Since \(\lambda_+\) is always upper semi-continuous, 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

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

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

1. \(\phi_B(x, \xi) \geq -m - \tau\) for every \((x, \xi)\) 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 \(d(p, q) < \gamma\);
3. there exists \(\varepsilon > 0\) such that \(\phi_B(x, \xi) \geq \log |\theta_x| - \tau\) for every \((x, \xi)\) with \(\xi \in H^\varepsilon\) and every \(B\) with \(d(A, B) < \gamma\).

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

\[
\int \phi_B(x, \xi) \, d\eta(\xi) = \int_{H^\varepsilon} \phi_B(x, \xi) \, d\eta(\xi) + \int_{H^\varepsilon} \phi_B(x, \xi) \, d\eta(\xi)
\]

\[
\geq \eta(H^\varepsilon)(\log |\theta_x| - \tau) - (m + \tau)\delta
\]

for every \(x\). The choice of \(\delta\) 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 $\varepsilon$-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 $r$-close to one of the numbers $\pm \int \log |\theta_x| \, dp(x)$ and, hence, is $2r$-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 $2r$-close to either $\lambda_+(A,p)$ or $\lambda_-(A,p)$. Assuming that $r$ is small enough, this implies that $|\lambda_x(A,p) - \lambda_x(B,q)| < 2r$ for both $x \in \{+,-\}$. In particular, we get the claim also in this case.

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 $\infty$ and the vertical direction $v$ is identified with $0$. 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 $\eta$ 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 $\zeta$ 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 $\eta(D(0,\varepsilon^{-1})) \leq \delta$ for any $(B,q)$-stationary probability measure $\eta$ 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, $D(z_0,r) = \{ z \in \mathbb{C} : |z - z_0| \leq r \}$. The proof of this proposition will appear in Section 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 $z \in \overline{\mathbb{C}} \setminus \{0\}$ converges to $\infty$ when $n \to +\infty$ and it converges to 0 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 $\delta_\infty$ is an attractor and $\delta_\infty$ is a repeller for $f_A$:

$$\lim_{n \to +\infty} f^{(\lambda)}_A \xi \to \delta_\infty \quad \text{if} \quad \xi(\{0\}) = 0 \quad \text{and} \quad \lim_{n \to -\infty} f^{(\lambda)}_A \xi \to \delta_0 \quad \text{if} \quad \xi(\{\infty\}) = 0.$$
Lemma 4.1. Given any $\varepsilon > 0$ and $\delta > 0$, there exists $\gamma > 0$ such that
$$\eta(\mathbb{D}(0, \varepsilon^{-1}) \setminus \mathbb{D}(0, \varepsilon)) \leq \delta$$
for every $(B, q)$-stationary measure $\eta$ and every $(B, q) \in V(A, p, \gamma)$.

Proof. Let $Q_\varepsilon = \mathbb{D}(0, \varepsilon^{-1}) \setminus \hat{B}(0, \varepsilon)$ and suppose that there exists a sequence $(B_k, q_k)$ converging to $(A, p)$ and $(B_k, q_k)$-stationary measures $\eta_k$ such that $\eta_k(Q_\varepsilon) \geq \delta$. By compactness and Lemma 3.2, we may suppose that $\eta_k$ converges to some $(A, p)$-stationary measure $\eta$. Since $Q_\varepsilon$ is closed, $\eta(Q_\varepsilon) \geq \limsup \eta_k(Q_\varepsilon) \geq \delta$. This contradicts the fact that all $(A, p)$-stationary measures are supported on $\{0, \infty\}$. This contradiction proves that $\eta(\mathbb{D}(0, \varepsilon^{-1}) \setminus \mathbb{D}(0, \varepsilon)) \leq \eta(Q_\varepsilon) \leq \delta$. \qed

Thus, for proving Proposition 3.8 we must show that the stationary measures of irreducible cocycles near $(A, p)$ have small mass in the neighborhood of $0$. The key property that distinguishes $\delta_0$ among the elements of $\text{Stat}(A, p)$ 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 cocycle $(B, q)$ 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, $\eta(\mathbb{D}(0, \varepsilon) \setminus \mathbb{D}(0, \rho))$ and $\eta(\mathbb{D}(0, \rho))$, where $0 \leq \rho < \varepsilon$ is very small if $(B, q)$ is close to having $0$ 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 Sections 5 and 6.

The following example illustrates these issues and could serve as a guideline for what follows. Take $p$ to be supported on exactly two points, with equal masses, corresponding to Möbius transformations
$$B_1(z) = 9z^2 \quad \text{and} \quad B_2(z) = \frac{2^{-1}z + b}{cz + 2}$$
with $b$ and $c$ close to zero. In this case, $\rho$ 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 $|z| \geq \rho$ then $B_1^n(z)$ leaves $\mathbb{D}(0, \varepsilon)$ rapidly, because $0$ is a strongly repelling fixed point for $B_1$. If $|z| < \rho$ then $|B_2(z)| \geq \rho$ and so the sequence $B_2^n B_1(z)$ also leaves $\mathbb{D}(0, \varepsilon)$ in a small number of iterates. One deduces that, in either case $\mathbb{D}(0, \varepsilon) \setminus \mathbb{D}(0, \rho)$ 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 $\rho$ 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 Section 5 and depends on the stationary measure as well as on the cocycle.

4.2. Discretization. Let us begin by introducing a convenient discretization procedure. Fix $h > 0$ such that $\int \log |\theta_x| \, dp(x) > 6h$. For each $x \in X$, let $s_x$ be the unique integer number such that
$$\log |\theta_x| - 2h < hs_x \leq \log |\theta_x| - h.$$  \hspace{1cm} (6)
As immediate consequences, we get:
$$e^{-2h|\theta_x|} < e^{hs_x} \leq e^{-h|\theta_x|} < \|A\| \quad \text{for all } x \in X$$  \hspace{1cm} (7)
Lemma 4.3. Let \( \Delta : \mathbb{C} \to \mathbb{C} \) be \( \Delta(z) = e^{2h_s} z \). The relations (7) e (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

\[
s \geq |s_x| \quad \text{for every } x \in X \quad \text{and} \quad hs \geq \log(2\|A\|).
\]

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

Given any measurable set \( K \subset X \), define \( D^K_x : \mathbb{C} \to \mathbb{C} \) by \( D^K_x(z) = e^{2h_x^K} z \), where

\[
s^K_x = s_x \quad \text{if } x \in K \quad \text{and} \quad s^K_x = -s \quad \text{if } x \in X \setminus K.
\]

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

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

Define \( \alpha = 1/s \). Then

\[
\int hs^K_x dp(x) \geq 2h \quad \text{for every } K \subset X \text{ with } p(X \setminus K) \leq \alpha.
\]

4.3. Contractions. We need a few elementary facts about the behavior of contraction on a disk \( D(0,a) \). Let \( \lambda < 1 \) and \( \Phi : D(0,a) \to D(0,a) \) be a \( \lambda \)-contraction.

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

(a) \( D(0,r) \cap \Phi(D(0,r)) = \emptyset \) for all \( 0 \leq r < \frac{|w_0|}{2} \)

(b) \( \Phi(D(0,R)) \subset D(0,R) \) for all \( R \geq \frac{|w_0|}{1 - \lambda} \)

(c) \( D(0,\frac{1-\lambda}{2} s) \cap \Phi(D(0,\frac{1-\lambda}{2} s)) = \emptyset \) whenever \( \Phi(D(0,s)) \not\subset D(0,s) \).

**Proof.** It is clear that \( \Phi(D(0,r)) \) is contained in \( D(w_0, r) \) and \( D(0,r) \cap D(w_0, r) = \emptyset \) 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 \( R \geq |w_0|/(1 - \lambda) \) and \( |z| \leq R \). This proves part (b). Then, \( \Phi(D(0,s)) \not\subset D(0,s) \) implies \( s < |w_0|/(1 - \lambda) \), that is, \( (1 - \lambda)s/2 < |w_0|/2 \). By (a), this implies (c). \( \square \)

**Lemma 4.3.** Let \( \tau > 0 \) and \( \Lambda > \lambda \). If the fixed point of \( \Phi \) is contained in \( D(0,\tau) \) then:

\[
\Phi(D(0,r)) \subset D(0,\Lambda r) \quad \text{for all } a \geq r \geq C\tau, \quad \text{where} \quad C = \frac{1 + \lambda}{\Lambda - \lambda}.
\]
Proof. Let $z_0$ be the fixed point of $\Phi$. Then

$$\Phi(\mathbb{D}(0, r)) \subset \Phi(\mathbb{D}(z_0, r + \tau)) \subset \mathbb{D}(z_0, \lambda(r + \tau)) \subset \mathbb{D}(0, \lambda(r + \tau) + \tau)$$

The assumption $r \geq (1 + \lambda)\tau/(\Lambda - \lambda)$ implies that $\lambda(r + \tau) + \tau \leq \Lambda r$ and, therefore,

$$\Phi(\mathbb{D}(0, r)) \subset \mathbb{D}(0, \Lambda r).$$

\hfill $\square$

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

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 $\Phi$ 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 $\Phi$ and compactness of $\mathbb{D}(0, r_1)$, 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 $\Phi$ is a $\lambda$-contraction. \hfill $\square$

4.4. Applications to cocycles. Here are a few applications of the lemmas in Section 4.3 to the context we are interested in:

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 if $B_x^{-1}(\mathbb{D}(0, r)) \cap \mathbb{D}(0, \|A\|^2r) \neq \emptyset$ then

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

Proof. Clearly, the diameter of $A_x^{-1}(\mathbb{D}(0, r))$ is bounded by $2|\theta_x|^{-2}r \leq 2\|A\|^2r$, for every $r$ and every $x$. Then we may find $\gamma > 0$ such that $d(A, B) < \gamma$ then the diameter of $B_x^{-1}(\mathbb{D}(0, r))$ is less than $3\|A\|^2r$, for every $r$ and every $x$. Then

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

To conclude, use the second part of (9). \hfill $\square$

Lemma 4.6. Given $r_0 > 0$ there is $\gamma > 0$ such that if $d(A, B) < \gamma$ and $r \in [r_0, 1]$ then

$$B_x^{-1}(\mathbb{D}(0, r)) \subset \mathbb{D}(0, e^{-2hr}r) = D_x^{-1}(\mathbb{D}(0, r)).$$

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

Lemma 4.7. There exist $c \in (0, 1)$ and $\gamma > 0$ such that if $d(A, B) < \gamma$ then

$$B_x^{-1}(\mathbb{D}(0, r)) \not\subset D_x^{-1}(\mathbb{D}(0, r)) \quad \text{implies} \quad \mathbb{D}(0, cr) \cap B_x^{-1}(\mathbb{D}(0, cr)) = \emptyset.$$

Proof. As observed before, every $\Phi_x = D_xB_x^{-1}$ is an $e^{-h}$-contraction on $\mathbb{D}(0, 1)$ if $B$ is close enough to $A$. Then, Lemma 4.2(c) means that $B_x^{-1}(\mathbb{D}(0, r)) \not\subset D_x^{-1}(\mathbb{D}(0, r))$ implies

$$\mathbb{D}(0, c_1e^{-2hr}r) \cap B_x^{-1}(\mathbb{D}(0, c_1r)) = \emptyset,$$

where $c_1 = \frac{1 - e^{-h}}{2}$. 

\hfill $\square$
Since \( s \geq 0 \) and \( s \geq s_x \) for every \( x \), this relation implies
\[
\mathbb{D}(0, c_1 e^{-2hs_x}) \cap B_x^{-1}(\mathbb{D}(0, c_1 e^{-2hs})) = \emptyset.
\]
So, we may take \( c = c_1 e^{-2hs} \).

Recall that \( X_+ \) denotes the set of points \( x \in X \) for which \( s_x > 0 \). As a particular case of (11), we have that \( p(X_+) > 2\alpha \).

**Lemma 4.8.** There exist \( C > 1 \) and \( \gamma > 0 \) such that if \( d(A, B) < \gamma \) then
\[
B_x^{-1}(\mathbb{D}(0, r)) \subset \mathbb{D}(0, e^{-2hs_x}r) \subset \mathbb{D}(0, e^{-2h}r)
\]
for every \( r \geq C\tau \) and any \( x \in X_+ \) such that \( B_x \) has a fixed point in \( \mathbb{D}(0, \tau) \).

**Proof.** For each \( x \in X_+ \), we have \( \log |\theta_x| \geq h(s_x + 1) \) and so \( A^{-1}_x(z) = \theta_x^2 z \) is an \( e^{-2h(s_x+1)} \)-contraction. This contraction rate is bounded from zero, since \( s_x \leq s \) for every \( x \). Thus, we can find \( \gamma > 0 \) such that \( d(A, B) < \gamma \) implies that \( B_x^{-1} \) is an \( e^{-2h(s_x+2)} \)-contraction on \( \mathbb{D}(0, 1) \), for every \( x \in X_+ \). By Lemma 4.3 (with \( \Phi = B_x^{-1} \) and \( \lambda = e^{-2h(s_x+2)} \) and \( \Lambda = e^{-2hs_x} \)),
\[
B_x^{-1}(\mathbb{D}(0, r)) \subset \mathbb{D}(0, e^{-2hs_x}r)
\]
for every \( r \geq C_x \tau \), where
\[
C_x = \frac{1 + e^{-2h(s_x+2)}}{e^{-2hs_x} - e^{-2h(s_x+2)}} = \frac{e^{2h(s_x+2)} + 1}{e^{2h} - 1}.
\]
Take \( C = (e^{2h(s+2)} + 1)/(e^{2h} - 1) \) and observe that \( C_x \leq C \). Thus, (12) holds for every \( x \in X_+ \) and every \( r \geq C\tau \).

5. Adapted radii

The following definition plays a center part in our arguments. We say that a radius \( r \geq 0 \) is \((B, q, \eta)\)-adapted on a measurable set \( K \subset M \) if
\[
\int \eta(B_x^{-1}(\mathbb{D}(0, r))) \, dq(x) \leq \int \eta((D^K_x)^{-1}(\mathbb{D}(0, r))) \, dq(x).
\]
(13)

For \( x \) and \( r \) fixed, \((D^K_x)^{-1}(\mathbb{D}(0, r)) = \mathbb{D}(0, e^{-2hs_x}r)\) can only increase when the set \( K \) increases (because \( s_x \geq -s \) for every \( x \in X \)). So, the condition (13) becomes stronger as the set \( K \) becomes larger.

For each measurable set \( K \subset X \) with \( p(X \setminus K) \leq \alpha \), define
\[
\rho(K) = \rho(B, q, \eta, K) = \inf\{ r \geq 0 : \text{every } s \in [r, 1] \text{ is } (B, q, \eta)\text{-adapted on } K \} \quad (14)
\]
For any fixed \( r_0 > 0 \), it follows from Lemma 4.6 that every \( r \in [r_0, 1] \) is \((B, q, \eta)\)-adapted on \( X \) if \( (B, q) \) is close enough to \((A, p)\). Taking \( r_0 = \varepsilon/(2\|A\|) \), we find that
\[
\rho(K) \leq \frac{\varepsilon}{2\|A\|} \quad \text{for any measurable } K \subset X \text{ with } p(X \setminus K) \leq \alpha. \quad (15)
\]

**Proposition 5.1.** Given \( \varepsilon > 0 \) and \( \delta > 0 \) there exists \( \gamma > 0 \) such that
\[
\eta(\mathbb{D}(0, \varepsilon) \setminus \mathbb{D}(0, \rho(K))) \leq \delta
\]
for any measurable set \( K \) with \( p(X \setminus K) \leq \alpha \), any \((B, q)\)-stationary measure \( \eta \) and any \((B, q) \in V(A, p, \gamma) \).
Proposition 5.1 will be proved in Section 5.2. The following direct consequence is the main conclusion in this section. Define
\[ \rho = \rho(B, q, \eta) = \inf \{ \rho(K) : \rho(X \setminus K) \leq \alpha \} \] (16)

**Corollary 5.2.** Given \( \varepsilon > 0 \) and \( \delta > 0 \) there exists \( \gamma > 0 \) such that
\[ \eta(D(0, \varepsilon) \setminus D(0, \rho)) \leq \delta \]
for any \((B, q)\)-stationary measure \( \eta \) and any \((B, q) \in V(A, p, \gamma)\).

**Proof.** Take \( K_j \) with \( \rho(K_j) \setminus \rho \) and notice that \( D(0, \rho) = \cap_j D(0, \rho(K_j)) \). □

### 5.1. Two auxiliary lemmas.

For proving Proposition 5.1, it is convenient to discretize phase space as well. Define \( I_j(r) = D(0, e^{-(2j-2)h}r) \setminus D(0, e^{-2jhr}) \) for each \( j \in \mathbb{Z} \) and \( r > 0 \). Clearly, for any fixed \( r \), the sequence \( (I_j(r))_j \) is invariant under \( \Delta \) and every \( D_x \). So, it is also invariant under every \( D^K_x \), for any \( K \subset X \).

**Lemma 5.3.** If \( r > 0 \) is \((B, q, \eta)\)-adapted on \( K \) then
\[ \int_{K+} \sum_{j=1}^{s^K} \eta(I_j(r)) \, dq(x) \leq \int_{K-} \sum_{j=s^K+1}^{0} \eta(I_j(r)) \, dq(x). \]

If \( e^{2ht}r \) is \((B, q, \eta)\)-adapted on \( K \) for every \( t = 0, 1, \ldots, n \) then
\[ \int_{K+} \sum_{j=1}^{s^K} \eta(I_{t+j}(r)) \, dq(x) \leq \int_{K-} \sum_{j=s^K+1}^{0} \eta(I_{t+j}(r)) \, dq(x), \text{ for } t = 0, 1, \ldots, n. \]

**Proof.** Define
\[ L_x(r) = \begin{cases} D(0, r) \setminus (D^K_x)^{-1}(D(0, r)) = D(0, r) \setminus D(0, e^{-2hs^K_x}r) & \text{for } x \in K_+ \\ \emptyset & \text{otherwise} \\ (D^K_x)^{-1}(D(0, r)) \setminus D(0, r) = D(0, e^{-2hs^K_x}r) \setminus D(0, r) & \text{for } x \in K_- \end{cases} \]

Using that \( r \) is \((B, q, \eta)\)-adapted and \( \eta \) is \((B, q)\)-stationary, we find that
\[ \int_{K+} (\eta(D(0, r)) - \eta(D(0, e^{-2hs^K_x}r))) \, dq(x) \leq \int (\eta(D(0, r)) - \eta(B^{-1}_x(D(0, r))) \, dq(x) = 0. \]

The left hand side coincides with \( \int_{K+} \eta(L_x(r)) \, dq(x) - \int_{K-} \eta(L_x(r)) \, dq(x) \). So,
\[ \int_{K+} \eta(L_x(r)) \, dq(x) \leq \int_{K-} \eta(L_x(r)) \, dq(x) \]

Now, to get the first claim, just notice that \( L_x(r) = \bigcup_{j=1}^{s^K} I_j(r) \) if \( x \in K_+ \) and \( L_x(r) = \bigcup_{j=s^K+1}^{0} I_j(r) \) if \( x \in K_- \) (where \( \cup \) 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 5.4.** 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 \( n_\ast = \sup \{ n_x : x \in Y_\ast \} \) for \( \ast \in \{+, -\} \). Suppose that there exist \( \tau > 0, n \geq 0 \) and a probability measurable \( q \) on \( X \) such that
Then
\[
\sum_{j=1}^{n} a_j \leq \frac{n_+ + n_-}{\tau} \sum_{j=-n_{-}+1}^{0} a_j
\]

Proof. Begin by noticing that
\[
\sum_{t=0}^{n} \sum_{j=1}^{n_x} a_{j+t} = \sum_{t=1}^{n} \sum_{j=1}^{n_x} a_j \geq n_x \left( \sum_{j=1}^{n} a_j - \sum_{j=1}^{0} a_j \right)
\]
and, similarly,
\[
\sum_{t=0}^{n} \sum_{j=-n_{-}+1}^{0} a_{j+t} = \sum_{t=-n_{-}+1}^{0} \sum_{j=1}^{n_x} a_j \leq n_x \left( \sum_{j=1}^{n} a_j + \sum_{j=-n_{-}+1}^{0} a_j \right)
\]

Adding the inequalities (b) over all \( t = 0, \ldots, n \) and using (17)-(18),
\[
\int_{Y_+} n_x \left[ \sum_{j=1}^{n} a_j - \sum_{j=1}^{n_x} a_j \right] dq(x) \leq \int_{Y_-} n_x \left[ \sum_{j=1}^{n} a_j + \sum_{j=-n_{-}+1}^{0} a_j \right] dq(x).
\]

Then, using the inequality (a),
\[
\tau \sum_{j=1}^{n} a_j \leq \int_{Y_+} n_x \sum_{j=1}^{n_x} a_j dq(x) + \int_{Y_-} n_x \sum_{j=-n_{-}+1}^{0} a_j dq(x)
\]
\[
\leq n_+ \int_{Y_+} n_x \sum_{j=1}^{n_x} a_j dq(x) + n_- \int_{Y_-} \sum_{j=-n_{-}+1}^{0} a_j dq(x)
\]

Using the inequality (b) with \( t = 0 \), it follows that
\[
\tau \sum_{j=1}^{n} a_j \leq (n_+ + n_-) \int_{Y_-} \sum_{j=-n_{-}+1}^{0} a_j dq(x) \leq \sum_{j=-n_{-}+1}^{0} a_j q(Y_-)
\]

This implies the conclusion of the lemma.

5.2. **Proof of Proposition 5.1.** Take \( n_x = |s^K_\| \) for each \( x \in X \). Observe that \( n_x \leq s \) for every \( x \in X \), as a consequence of (9). For any probability \( q \),
\[
\int_{K_+} n_x dq(x) - \int_{K_-} n_x dq(x) = \int \log h s^K_\] dq(x)
\]
By (10), the right hand side is bounded below by \( 2h \) when \( q = p \). In particular,
\[
\int_{K_+} n_x dq(x) - \int_{K_-} n_x dq(x) \geq h
\]
whenever \( d(p, q) < \gamma \), with \( \gamma \) sufficiently small. By Lemma 4.1, up to reducing \( \gamma \) we may suppose that
\[
\eta(\mathbb{D}(0, \varepsilon^{-1}) \setminus \mathbb{D}(0, \varepsilon)) \leq \frac{h \delta}{2s}
\]
for any \((B, q)\)-stationary measure and any \((B, q) \in V(A, p, \gamma)\).

Consider any \( r_0 > \rho(K) \) and then take \( r_1 \in [\varepsilon, 1] \) such that \( r_0 = r_1 e^{-2hn} \) for some \( n \geq 0 \). By the definition of \( \rho(K) \) in (14), every \( r \in [r_0, 1] \) is \((B, q, \eta)\)-adapted on \( K \).
In particular, this holds for \( r = r_1 e^{-2ht} \) for every \( t = 0, 1, \ldots, n \). Let \( a_j = \eta(I_j(r_1)) \) for \( j \in \mathbb{Z} \). Then the conclusion of Lemma 5.3 may be written as:

\[
\int_{K_j} \sum_{j=1}^{n_x} a_{j+t} \, dq(x) \leq \int_{K_j} \sum_{j=-n_x+1}^{0} a_{j+t} \, dq(x) \quad \text{for all } t = 0, 1, \ldots, n. \tag{20}
\]

Properties (19) and (20) correspond to hypotheses (a) and (b) in Lemma 5.4. From that lemma we get that

\[
\sum_{j=1}^{n} a_j \leq \frac{2s}{h} \sum_{j=-x+1}^{0} a_j. \tag{21}
\]

The left hand side of (21) coincides with

\[
\eta(\mathbb{D}(0, r_1) \setminus \mathbb{D}(0, r_1 e^{-2ht})) = \eta(\mathbb{D}(0, r_1) \setminus \mathbb{D}(0, r_0)) \geq \eta(\mathbb{D}(0, \varepsilon) \setminus \mathbb{D}(0, r_0))
\]

(because \( r_1 \geq \varepsilon \)). The right hand side of (21) coincides with

\[
\frac{2s}{h} \eta(\mathbb{D}(0, r_1 e^{-2hs}) \setminus \mathbb{D}(0, r_1)) \leq \frac{2s}{h} \eta(\mathbb{D}(0, \varepsilon^{-1}) \setminus \mathbb{D}(0, \varepsilon)).
\]

(because \( \varepsilon \leq r_1 \leq 1 \) and \( e^{2hs} < \varepsilon^{-1} \), as long as \( \varepsilon \) is sufficiently small). Hence, the inequality (21) implies

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

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, at this point it suffices to show that \( \eta(\mathbb{D}(0, \rho)) \leq \text{const} \delta \) for every \((B, q)\)-stationary measure \( \eta \), if \((B, q)\) is close enough to \((A, p)\) and satisfies \( q(z \in X : B_x(z) = z) < 1 \).

The case when \( \rho = 0 \) is easy, because the next lemma implies that \( \mathbb{D}(0, 0) = \{0\} \) always has measure zero:

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

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

\[
l_a = \eta(F) = \int \eta(B_x^{-1}(F)) \, dq(x) \leq \int l_a \, dq(x) = l_a.
\]

This implies \( \eta(B_x^{-1}(F)) = al \) 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 \( \theta_{z_1} > 1 \) 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 M"obius transformation \( B_x \) is hyperbolic, with fixed points close to 0 and \( \infty \), 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, \# \( F \leq 2 \). 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 \( x \), contradicting the hypothesis. Otherwise,
Case 1: \[ p(\{x \in X : B_x^{-1}(\mathbb{D}(0,r)) \subset \mathbb{D}(0,e^{-2hs_x}r)\}) \leq \alpha \quad \text{for some} \ 0 \leq r < \rho. \] This is handled by the following lemma:

**Lemma 6.2.** If \( p(\{x \in X : B_x^{-1}(\mathbb{D}(0,r)) \subset \mathbb{D}(0,e^{-2hs_x}r)\}) \leq \alpha \) \( \text{for some} \ 0 \leq r < \rho \) then \( \eta(\mathbb{D}(0,\rho)) \leq 2\rho. \)

**Proof.** Applying Lemma 4.4 to the \( \beta \)-contractions \( \Phi_x = D_x B_x^{-1} \), we get that the set \( K_t = \{x \in X : B_x^{-1}(\mathbb{D}(0,t)) \subset \mathbb{D}(0,e^{-2hs_x}t)\} \) increases as \( t \) increases. So, assuming there exists some \( r \) as in the statement, it can chosen arbitrarily close to \( \rho \). Fix \( r \in (\|A\|^{-2}\rho,\rho) \) and let \( K = K_r \). We have \( p(X \setminus K) \leq \alpha \), by assumption, and so the definition of \( \rho \) gives that \( r < \rho(K) \). Then, by the definition of \( \rho(K) \) in (16), there exists \( t \in (r,\rho) \) that is not \( (B,q,\eta) \)-adapted on \( K \). In other words,

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

This implies that there exists \( y \in X \) such that

\[
\eta(B_y^{-1}(\mathbb{D}(0,t))) > \eta(\mathbb{D}(0,e^{-2hs_y}t)) \geq \eta(\mathbb{D}(0,e^{-2hs_x}t)).
\]

(recall that \( s_y^K \leq s_x \) 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:

\[
\eta(B_y^{-1}(\mathbb{D}(0,t))) > \eta(\mathbb{D}(0,e^{-2hs_y}t)) = \eta(\mathbb{D}(0,e^{2hs_x}t))
\]

The choice of \( t \) together with (9) give that \( e^{2hs_x}t > \|A\|^2 t > \|A\|^2 r > \rho \). It follows that

\[
\eta(B_y^{-1}(\mathbb{D}(0,t))) > \eta(\mathbb{D}(0,\rho)). \tag{22}
\]

Another consequence is that \( B_y^{-1}(\mathbb{D}(0,t)) \subset \mathbb{D}(0,e^{2hs_x}t) \). By Lemma 4.5, this implies

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

By (15), we have that \( t < \rho \leq \varepsilon/(2\|A\|) \). It follows that \( B_y^{-1}(\mathbb{D}(0,t)) \subset \mathbb{D}(0,\varepsilon) \), because \( \|B_y^{-1}\| \leq 2\|A\| \) if \( B \) is close enough to \( A \). Summarizing,

\[
B_y^{-1}(\mathbb{D}(0,t)) \subset \mathbb{D}(0,\varepsilon) \setminus \mathbb{D}(0,\|A\|^2 t) \subset \mathbb{D}(0,\varepsilon) \setminus \mathbb{D}(0,\rho).
\]

Consequently, using Corollary 5.2,

\[
\eta(B_y^{-1}(\mathbb{D}(0,t))) \leq \eta(\mathbb{D}(0,\varepsilon) \setminus \mathbb{D}(0,\rho)) \leq \delta. \tag{23}
\]

From (22) and (23) we get that \( \eta(\mathbb{D}(0,\rho)) < \delta. \)
Case 2: $p\{x \in X : B_x^{-1}(\mathbb{D}(0, r)) \not\subset \mathbb{D}(0, e^{-2hsx_r})\} > \alpha$ for every $0 \leq r < \rho$. For each $z \in \mathbb{D}(0, 1)$ and $r \in [0, 1]$, define

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

Let $c$ and $C$ be as the Lemmas 4.7 and 4.8, respectively, and then let $\ell \geq 0$ be the smallest integer such that $e^{-2\ell} \leq c$. Keep in mind that $c$, $C$ and $\ell$ depend only on $A$. So, the same is true about

$$\omega = 8C^2e^{4h\ell}\alpha^{-1}. \quad (24)$$

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 $\rho_0 \in [0, 1]$ such that

1. $p(\Gamma(z_0, \rho_0)) \geq 2\omega^{-1}$;
2. $p(X_+ \setminus \Gamma(z_0, Ce^{2h\ell}\rho_0)) \geq \alpha$ if $\rho_0 > 0$.

**Proof.** Clearly, $\Gamma(0, 1) = X_+$ if $B$ is close enough to $A$. Then, (11) implies that $p(\Gamma(0, 1)) > 2\alpha > 2\omega^{-1}$. Let $\rho_0$ 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 $\rho_0$ 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 $\rho_0$ 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 $\rho_1$. 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,

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

The definition of $\omega$ in (24) is such that this last expression is equal to $\alpha$. \hfill $\Box$

**Remark 6.4.** If $B$ is close to $A$ then the point $z_0$ is close to zero and the radius $\rho_0$ is small. Indeed, given any $r_0 > 0$, we have $\Gamma(0, r_0) = X_+$ 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, $z_0 \in \mathbb{D}(0, 2r_0)$.

In particular, it is no restriction to assume that

$$Ce^{2h\ell}\rho_0 \leq \frac{\varepsilon}{2\|A\|} \quad (25)$$

and we will do so in what follows. 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.** For $0 \leq r \leq Ce^{2h\ell}\rho_0$,

$$p\{x \in X : B_x^{-1}(\mathbb{D}(0, r)) \not\subset \mathbb{D}(0, e^{-2hsx_r})\} \geq \alpha. \quad (26)$$

**Proof.** If $x \in X_+$ is such that $B_x^{-1}(\mathbb{D}(0, r)) \subset \mathbb{D}(0, e^{-2hsx_r})$ then $B_x$ 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

$$X_+ \setminus \Gamma(0, r) \subset \{ x \in X : B_x^{-1}(\mathbb{D}(0, r)) \not\subset \mathbb{D}(0, e^{-2hsx_r})\}.$$ 

Then the claim follows from Lemma 6.3(b), with $r = Ce^{2h\ell}\rho_0$. \hfill $\Box$
Lemma 6.6. (a) $\eta(\mathbb{D}(0, c\rho)) \leq 2s\eta(\mathbb{D}(0, \varepsilon) \setminus \mathbb{D}(0, c\rho))$;
(b) $\eta(\mathbb{D}(0, C\rho_0)) \leq 2s\eta(\mathbb{D}(0, \varepsilon) \setminus \mathbb{D}(0, C\rho_0))$.

Proof. Let $K = \{ x \in X : B_x^{-1}(\mathbb{D}(0, \rho)) \subset \mathbb{D}(0, e^{-2\varepsilon x}\rho) \}$. Since we assume that $p(X \setminus K) > \alpha$, we have $q(X \setminus K) > \alpha/2 = 1/(2s)$ for every $q$ in a neighborhood of $p$. Since $\eta$ is stationary,

$$
\int_{X \setminus K} (\eta(\mathbb{D}(0, c\rho)) - \eta(B_x^{-1}(\mathbb{D}(0, c\rho)))) \, dp(x) = \int_K (\eta(B_x^{-1}(\mathbb{D}(0, c\rho))) - \eta(\mathbb{D}(0, c\rho))) \, dp(x).$$

(26)

From Lemma 4.7 we get that $\mathbb{D}(0, c\rho)$ and its pre-image under $B_x$ are disjoint for every $x \in X \setminus K$. We also know that the pre-image is contained in $\mathbb{D}(0, \varepsilon)$: that follows from (15) and the fact that the norm of $B_x^{-1}$ is bounded by $2\|A\|$ if $B$ is close enough to $A$. Consequently,

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

For every $x \in X$,

$$
\eta(B_x^{-1}(\mathbb{D}(0, c\rho))) - \eta(\mathbb{D}(0, \varepsilon)) \leq \eta(\mathbb{D}(0, \varepsilon)) - \eta(\mathbb{D}(0, c\rho)) = \eta(\mathbb{D}(0, \varepsilon) \setminus \mathbb{D}(0, c\rho)).
$$

Replacing the last two estimates on the left hand side and the right hand side of (26), respectively, we obtain that

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

This yields,

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

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

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

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

instead. By Corollary 6.5, $p(X \setminus K) \geq \alpha$ and so $q(X \setminus K) \geq \alpha/2 = 1/(2s)$ for every $q$ in a neighborhood of $p$. Using (25), we get that the pre-image of $\mathbb{D}(0, Ce^{2\varepsilon \ell \rho_0})$ under any $B_x$ is contained in $\mathbb{D}(0, \varepsilon)$. So, the same arguments as in the previous paragraph yield

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

Since $Ce^{2\varepsilon \ell} \geq 1$, this implies the conclusion in part (b) of the lemma.

Lemma 6.7. For any $C\rho_0 \leq r \leq \varepsilon/(2\|A\|)$,

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

Proof. Lemma 4.8 implies that

$$
q(\Gamma(0, \rho_0))\eta(\mathbb{D}(0, \varepsilon) \setminus \mathbb{D}(0, e^{-2\varepsilon r})) = \int_{\Gamma(0, \rho_0)} (\eta(\mathbb{D}(0, r)) - \eta(\mathbb{D}(0, e^{-2\varepsilon r}))) \, dq(x)
$$

$$
\leq \int_{\Gamma(0, \rho_0)} (\eta(\mathbb{D}(0, r)) - \eta(B_x^{-1}(\mathbb{D}(0, r)))) \, dq(x)
$$

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

$$
\int_{X \setminus \Gamma(0, \rho_0)} (\eta(B_x^{-1}(\mathbb{D}(0, r))) - \eta(\mathbb{D}(0, r))) \, dq(x) \leq \eta(\mathbb{D}(0, \varepsilon) \setminus \mathbb{D}(0, r)).
$$
Putting these two inequalities together,
\[ \omega^{-1} \eta\left( \mathbb{D}(0, r) \setminus \mathbb{D}(0, e^{-2h}r) \right) \leq \eta\left( \mathbb{D}(0, \varepsilon) \setminus \mathbb{D}(0, r) \right). \]
This implies the claim in the lemma. \( \square \)

**Corollary 6.8.** \( \eta\left( \mathbb{D}(0, \varepsilon) \right) \leq \kappa \delta, \) where \( \kappa = (1 + 2s)(1 + \omega)^\ell \) depends only on \( A. \)

**Proof.** First, suppose that \( Ce^{2h} \rho_0 \leq \rho. \) Then we may apply Lemma 6.7 to every \( r = e^{-2h}j, j = 0, \ldots, \ell - 1. \) So, using also Corollary 5.2,
\[ \eta\left( \mathbb{D}(0, \varepsilon) \setminus \mathbb{D}(0, e^{-2h}j) \right) \leq (1 + \omega)^j \eta\left( \mathbb{D}(0, \varepsilon) \setminus \mathbb{D}(0, \rho) \right) \leq \delta(1 + \omega)^j. \]
In view of the definition of \( \ell, \) this implies
\[ \eta\left( \mathbb{D}(0, \varepsilon) \setminus \mathbb{D}(0, c\rho) \right) \leq \eta\left( \mathbb{D}(0, \varepsilon) \setminus \mathbb{D}(0, e^{-2h}j) \right) \leq \delta(1 + \omega)^j. \]
Combined with Lemma 6.6(a), this gives \( \eta\left( \mathbb{D}(0, c\rho) \right) \leq 2s\delta(1 + \omega)^j. \) Adding the last two inequalities, we obtain
\[ \eta\left( \mathbb{D}(0, \varepsilon) \right) \leq (1 + 2s)\delta(1 + \omega)^j. \] (27)

Now, suppose that \( Ce^{2h} \rho_0 > \rho \) (in particular, \( \rho_0 > 0. \)) Then, just as before,
\[ \eta\left( \mathbb{D}(0, \varepsilon) \setminus \mathbb{D}(0, C\rho_0) \right) \leq (1 + \omega)^\ell \eta\left( \mathbb{D}(0, \varepsilon) \setminus \mathbb{D}(0, Ce^{2h}j) \right) \]
\[ \leq (1 + \omega)^\ell \eta\left( \mathbb{D}(0, \varepsilon) \setminus \mathbb{D}(0, \rho) \right) \leq \delta(1 + \omega)^\ell. \]
Lemma 6.6(b) gives \( \eta\left( \mathbb{D}(0, C\rho_0) \right) \leq 2s\delta(1 + \omega)^\ell. \) Adding these two inequalities,
\[ \eta\left( \mathbb{D}(0, \varepsilon) \right) \leq (1 + 2s)\delta(1 + \omega)^\ell. \] (28)
The inequalities (27) and (28) imply the conclusion of the corollary. \( \square \)

By Lemma 4.1 and Corollary 6.8:
\[ \eta\left( \mathbb{D}(0, \varepsilon^{-1}) \right) \leq \eta\left( \mathbb{D}(0, \varepsilon^{-1}) \setminus \mathbb{D}(0, \varepsilon) \right) + \eta\left( \mathbb{D}(0, \varepsilon) \right) \leq (1 + \kappa)\delta. \]
This gives Proposition 3.8 when \( z_0 = 0. \) To finish, let us explain how the latter assumption 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 \( \tilde{B}_z = H \cdot B_z \cdot H^{-1} \) is uniformly close to \( B. \) A measure \( \eta \) is \((B, q)\)-stationary if and only if \( \tilde{H}_x \eta \) is \((\tilde{B}, q)\)-stationary. It is clear that \( q\{x \in X : B_z(z) = z\} \leq 1 \) for all \( z \) and only if \( q\{x \in X : \tilde{B}_z(z) = z\} \leq 1 \) for all \( z. \) Analogously, the set \( \tilde{\Gamma}(z, r) \) of points \( x \) such that \( \tilde{B}_z \) has a fixed point on \( D(z, r) \) coincides with \( \Gamma(z + z_0, r) \) for every \( z \) and \( r. \) In particular, by Lemma 6.3,
\[ p\left( \tilde{\Gamma}(0, \rho_0) \right) \geq 2\omega^{-1} \quad \text{and} \quad p\left( X_+ \setminus \tilde{\Gamma}(0, Ce^{2h}j\rho_0) \right) \geq \alpha \text{ if } \rho_0 > 0. \]
So, we may apply the previous arguments to \( \tilde{B}, q, \) and \( \tilde{H}, \) to get that
\[ \eta\left( \mathbb{D}(0, \varepsilon^{-1}) - z_0 \right) = \tilde{\eta}\left( \mathbb{D}(0, \varepsilon^{-1}) \right) \leq (1 + \kappa)\delta \] (29)
for any \((B, q)\)-stationary measure \( \eta \) and any \((B, q)\) that satisfies the assumptions in the present section. Since \( z_0 \) is small,
\[ \left( \mathbb{D}(0, \varepsilon^{-1}) - z_0 \right) \cup \left( \mathbb{D}(0, \varepsilon^{-1}) \setminus \mathbb{D}(0, \varepsilon) \right) \supset \mathbb{D}(0, \varepsilon^{-1}). \]
Thus, combining (29) with Lemma 4.1, we find that \( \eta\left( \mathbb{D}(0, \varepsilon^{-1}) \right) \leq (2 + \kappa)\delta. \)

The proof of Proposition 3.8 is now complete.
7. Proof of Theorem B

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

Lemma 7.1 (Avila). Let $X$ be a metric space such that every bounded closed subset is compact, and let $\nu$ be any Borel probability measure in $X$ whose support $Z = \text{supp}\nu$ is bounded.

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

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

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

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

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

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

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

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

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

$$\sum_{i=1}^n |\mu(Q_i) - \nu(Q_i)| < \varepsilon$$

for every $\mu \in V$.

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

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

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

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

where $p_{i,j} = \nu(a_{i,j})$. Denote $I_i = I \setminus \bigcup_{j \in J(i)} I_{i,j}$. 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 \( X \) is a polish space, that is, a complete separable metric space. Since all Borel non-atomic probabilities on polish spaces are isomorphic (see Ito [15, § 2.4]), 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 \( \mu \). By setting \( \phi \equiv a_{i,j} \) on each \( Z_i \times I_i,j \) we extend \( \phi_i \) to a measurable map \( Z_i \times I \to Q_i \) that still sends the restriction of \( q \) to the restriction of \( \mu \). Gluing all these extensions we obtain a measurable map \( \phi : Z \times I \to X \) such that \( \phi \cdot q = \mu \). By construction, \( \phi(x,t) \in Q_i \) for every \( x \in Z_i \) and \( t \in I \). This implies that \( d(\phi(x,t), x) \leq \text{diam} Q_i < \varepsilon \) for all \( (x,t) \in Z \times I \). Finally,

\[
\|q - (\nu \times \lambda)\| = \sum_{i=1}^{n} \|\left( \frac{\mu(Q_i)}{\nu(Q_i)} - 1 \right)(\nu \times \lambda) (Z_i \times I)\|
\]

\[
= \sum_{i=1}^{n} |\mu(Q_i) - \nu(Q_i)| < \varepsilon.
\]

The proof of the lemma is complete. \( \Box \)

Now, given \( \rho > 0 \), let \( \nu \) be a probability measure in \( \text{GL}(2, \mathbb{C}) \) with compact support. Consider \( X = \text{supp} \nu \times I, \ p = \nu \times \lambda \) and \( A : X \to \text{GL}(2, \mathbb{C}) \) given by \( A(x,t) = x \). From Theorem C, there is \( \varepsilon > 0 \) such that \( |\lambda_+ (A,p) - \lambda_+ (B,q)| < \rho \) 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 \( \delta \) such that if \( \nu' \in V \) and \( \text{supp} \nu' \subset B_\delta(\text{supp} \nu) \) then there exist \( B : X \to \text{GL}(2, \mathbb{C}) \) and a probability measure \( q \) on \( X \) such that \( d(p,q) < \varepsilon \), \( d(A,B) < \varepsilon \) and \( \nu' = B \cdot q \). 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, 22], where it is shown that Lyapunov exponents are often discontinuous relative to the \( C^0 \) topology.

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

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

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

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

Consider on \( M \) the Bernoulli measure \( \mu \) associated to an arbitrary probability vector \( p = (p_1, p_2) \) with positive entries.

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

\[
A(x) = \begin{pmatrix} \sigma & 0 \\ 0 & \sigma^{-1} \end{pmatrix} \quad \text{if } x_0 = 1 \quad \text{and} \quad A(x) = \begin{pmatrix} \sigma^{-1} & 0 \\ 0 & \sigma \end{pmatrix} \quad \text{if } x_0 = 2.
\]
Observe that the Lyapunov exponents are given by \( \lambda_k(A, p) = \pm |p_1 - p_2| \log \sigma \). 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 \textit{discontinuity} for the Lyapunov exponents relative to the \( C^r \) topology:

**Theorem 8.1.** For any \( r > 0 \) such that \( 2^{2r} < \sigma \) there exist \( B : M \to \text{SL}(2, \mathbb{R}) \) 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 [20] and Bochi [4, 5]. Here is an outline. The unperturbed cocycle \( A \) preserves both the horizontal line bundle \( H_x = \{x\} \times \mathbb{R}(1,0) \) and the vertical line bundle \( V_x = \{x\} \times \mathbb{R}(0,1) \). Then, the Oseledets subspaces must coincide with \( H_x \) and \( V_x \) almost everywhere. We choose cylinders \( Z_n \subset M \) whose first \( n \) iterates \( f^i(Z_n) \), \( 0 \leq i \leq n-1 \) are pairwise disjoint. Then we construct cocycles \( B_n \) by modifying \( A \) on some of these iterates so that

\[
B_n^m(x)H_x = V_{f^m(x)} \quad \text{and} \quad B_n^m(x)V_x = H_{f^m(x)} \quad \text{for all} \ x \in Z_n.
\]

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

\[
\| B_n - A \|_r \leq \text{const} \left( \frac{2^{2r}}{\sigma} \right)^{n/2}
\]

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

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

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

\[
R(x) = \begin{cases} 1 & \varepsilon_n \\ \varepsilon_n & 1 \end{cases} \quad \text{if} \ x \in Z_n \cup f^{2k}(Z_n)
\]

\[
R(x) = \text{id} \quad \text{in all other cases.}
\]

and then take \( B_n = AR_n \).

**Lemma 8.2.** \( B_n^m(x)H_x = V_{f^m(x)} \) and \( B_n^m(x)V_x = H_{f^m(x)} \) for all \( x \in Z_n \).

**Proof.** Notice that for any \( x \in Z_n \),

\[
B_n^k(x)H_x = \mathbb{R}(\varepsilon_n, 1) \quad \text{and} \quad B_n^k(x)V_x = V_{f^k(x)}
\]

\[
B_n^{k+1}(x)H_x = V_{f^{k+1}(x)} \quad \text{and} \quad B_n^{k+1}(x)V_x = \mathbb{R}(-\varepsilon_n, 1)
\]

\[
B_n^{2k}(x)H_x = V_{f^{2k}(x)} \quad \text{and} \quad B_n^{2k}(x)V_x = \mathbb{R}(-1, \varepsilon_n).
\]

The claim follows by iterating one more time. \( \square \)

**Lemma 8.3.** There exists \( C > 0 \) such that \( \| B_n - A \|_r \leq C \left( \frac{2^{2r}}{\sigma} \right)^k \) for every \( n \).
Proof. Let \( L_n = A - B_n \). Clearly, \( \sup \| L \| \leq \sup \| A \| \| \text{id} - R_n \| \) and this is bounded by \( \sigma \epsilon_n \). Now let us estimate the second term in the definition (30). If \( x \) and \( y \) are not in the same cylinder \([0; a]\) then \( d(x, y) = 1 \), and so
\[
\frac{\| L_n(x) - L_n(y) \|}{d(x, y)^r} \leq 2 \sup \| L_n \| \leq 2 \sigma \epsilon_n. \tag{31}
\]
From now on we suppose \( x \) and \( y \) belong to the same cylinder. Then, since \( A \) is constant on cylinders,
\[
\frac{\| L_n(x) - L_n(y) \|}{d(x, y)^r} = \frac{\| A(x)(R_n(x) - R_n(y)) \|}{d(x, y)^r} \leq \sigma \frac{\| R_n(x) - R_n(y) \|}{d(x, y)^r}.
\]
If neither \( x \) nor \( y \) belong to \( Z_n \cup f^k(Z_n) \cup f^{2k}(Z_n) \) then \( R_n(x) \) and \( R_n(y) \) are both equal to \( \text{id} \), and so the expression on the right vanishes. If \( x \) and \( y \) belong to the same \( f^i(Z_n) \) then \( R_n(x) = R_n(y) \) 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 \( d(x, y) \geq 2^{-2k} \) and so, using once more that \( \| \text{id} - R_n \| \leq \epsilon_n \) at every point,
\[
\frac{\| L_n(x) - L_n(y) \|}{d(x, y)^r} \leq \sigma \frac{\| R_n(x) - R_n(y) \|}{d(x, y)^r} \leq 2 \sigma \epsilon_n 2^{2kr}.
\]
Noting that this bound is worse than (31), we conclude that
\[
\| L_n \| \leq \sigma \epsilon_n + 2 \sigma \epsilon_n 2^{2kr} \leq 3 \sigma \left( \frac{2^{2r}}{\sigma} \right)^k.
\]
Now it suffices to take \( C = 3 \sigma \). \( \square \)

Now we want to prove that \( \lambda_{\pm}(B_n) = 0 \) for every \( n \). Let \( \mu_n \) be the normalized restriction of \( \mu \) to \( Z_n \) and \( f_n : Z_n \to Z_n \) be the first return map (defined on a full measure subset). Indeed,
\[
Z_n = \bigcup_{b \in B} [0; w, b, w] \quad \text{(up to a zero measure subset)}
\]
where \( w = (1, \ldots, 1, 2, \ldots, 2, 2) \) and the union is over the set \( B \) of all finite words
\( b = (b_1, \ldots, b_n) \) not having \( w \) as a sub-word. Moreover,
\[
f_n \mid [0; w, b, w] = f^{n+s} \mid [0; w, b, w] \quad \text{for each } b \in B.
\]
Thus, \((f_n, \mu_n)\) is a Bernoulli shift with an infinite alphabet \( B \) and probability vector given by \( p_b = \mu_n([0; w, b, w]) \). Let \( B_n : Z_n \to \text{SL}(2, \mathbb{R}) \) be the cocycle induced by \( B \) over \( f_n \), that is,
\[
B_n \mid [0; w, b, w] = B_{n+s} \mid [0; w, b, w] \quad \text{for each } b \in B.
\]
It is a well known basic fact (see [30, Proposition 2.9], for instance) that the Lyapunov spectrum of the induced cocycle is obtained multiplying the Lyapunov spectrum of the original cocycle by the average return time. In our setting this means
\[
\lambda_{\pm}(B_n) = \frac{1}{\mu(Z_n)} \lambda_{\pm}(B_n).
\]
Therefore, it suffices to prove that \( \lambda_{\pm}(B_n) = 0 \) for every \( n \).
Indeed, suppose the Lyapunov exponents of \( B_n \) are non-zero and let \( E^u_x \oplus E^s_x \) be the Oseledets splitting (defined almost everywhere in \( Z_n \)). Consider the probability measure \( m^u \) and \( m^s \) for the cocycle \( B_n \) defined by
\[
m^*(B) = \mu(\{ (x, (x, E^s_x)) \in B \}) = \int \delta_{(x, E^s_x)}(B) \, d\mu(x)
\]
for \( * \in \{ s, u \} \) and any measurable subset \( B \). The key observation is that, as a consequence of Lemma 8.2, the cocycle \( B_n \) permutes the vertical and horizontal subbundles:
\[
B_n(x)H_x = V_{f_n(x)} \quad \text{and} \quad B_n(x)V_x = H_{f_n(x)} \quad \text{for all} \ x \in Z_n.
\]
Let \( m \) be the measure defined on \( M \times \mathbb{P}(\mathbb{R}^2) \) by
\[
m_n(x) = \frac{1}{2}(\mu_n(\{ x \in Z_n : V_x \in X \}) + \mu_n(\{ x \in Z_n : H_x \in X \})).
\]
That is, \( m_n \) projects down to \( \mu_n \) and its disintegration is given by \( x \mapsto (\delta_{H_x} + \delta_{V_x})/2 \). It is clear from (32) that \( m_n \) is \( B_n \)-invariant.

**Lemma 8.4.** The probability measure \( m_n \) is ergodic.

**Proof.** Suppose there is an invariant set \( X \subset M \times \mathbb{P}(\mathbb{R}^2) \) with \( m_n(X) \in (0, 1) \). Let \( X_0 \) be the set of \( x \in Z_n \) whose fiber \( X \cap (\{ x \} \times \mathbb{P}(\mathbb{R}^2)) \) contains neither \( H_x \) nor \( V_x \). In view of (32), \( X_0 \) is an \( f_n \)-invariant set and so its \( \mu_n \)-measure is either 0 or 1. 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 \( x \in Z_n \) whose fiber contains both \( H_x \) and \( V_x \). Now let \( X_H \) be the set of \( x \in Z_n \) whose fiber contains \( H_x \) but not \( V_x \), and let \( X_V \) be the set of \( x \in Z_n \) whose fiber contains \( V_x \) but not \( H_x \). The previous observations show that \( X_H \cup X_V \) has full \( \mu_n \)-measure and it follows from (32) that
\[
f_n(X_H) = X_V \quad \text{and} \quad f_n(X_V) = X_H.
\]
Thus, \( \mu_n(X_H) = 1/2 = \mu_n(X_V) \) and \( f_n^2(X_H) = X_H \) and \( f_n^2(X_V) = X_V \). This is a contradiction because \( f_n \) is Bernoulli and, in particular, the second iterate is ergodic. \( \square \)

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

Then, by Lemma 8.4, \( m_n \) must coincide with either \( m^s \) and \( m^u \). This is a contradiction, because the conditional probabilities of \( m_n \) are supported on exactly two points on each fiber, whereas the conditional probabilities of either \( m^u \) and \( m^s \) are Dirac masses on a single point. This contradiction proves that the Lyapunov exponents of \( B_n \) do vanish for every \( n \), and that concludes the proof of Theorem 8.1.

The same kind of argument shows that, in general, one can not expect continuity to hold when some of the probabilities \( p_i \) vanishes:
Remark 8.5. (Kifer [17]) 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

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

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 (32). 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

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

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 $\lambda_+(A) = \log 2$.

References


