LARGE DEVIATIONS FOR PRODUCTS OF RANDOM TWO DIMENSIONAL MATRICES

PEDRO DUARTE AND SILVIUS KLEIN

Abstract. We establish large deviation type estimates for i.i.d. products of two dimensional random matrices with finitely supported probability distribution. The estimates are stable under perturbations and require no irreducibility assumptions. In consequence, we obtain a uniform local modulus of continuity for the corresponding Lyapunov exponent regarded as a function of the support of the distribution. This in turn has consequences on the modulus of continuity of the integrated density of states and on the localization properties of random Jacobi operators.

1. Introduction and statements

Consider a multiplicative random process; that is, let \( \{ g_n \}_{n \in \mathbb{Z}} \) be a sequence of i.i.d. random matrices relative to a compactly supported probability measure \( \mu \) on the general linear group \( \text{GL}_d(\mathbb{R}) \), and let \( g^{(n)} := g_{n-1} \ldots g_1 g_0 \) denote the partial products process. The Furstenberg-Kesten theorem, the multiplicative analogue of the law of large numbers, implies that the average process \( \frac{1}{n} \log \| g^{(n)} \| \) converges almost surely to a number \( L^+ (\mu) \), called the (maximal) Lyapunov exponent of the process.

A natural problem in this context is to obtain a more quantitative version of the convergence \( \frac{1}{n} \log \| g^{(n)} \| \to L^+ (\mu) \), or in other words, to establish large deviation type estimates or other statistical properties for such multiplicative processes.

A second problem concerns the continuity of the limit quantity, the Lyapunov exponent \( L^+ (\mu) \), with respect to the input data (e.g. the measure \( \mu \), or just its support), as the data varies in an appropriate sense.

Both of these problems are highly relevant in the spectral theory of the discrete random (e.g. Schrödinger or Jacobi) operators in mathematical physics.
These topics were studied in the 80s by H. Furstenberg and Y. Kifer and by E. Le Page in a generic setting (i.e. under irreducibility\footnote{Irreducibility refers to the non existence of proper subspaces invariant under the closed semigroup $T^\mu$ generated by the support of the measure $\mu$. There are different versions of this notion.} and contractivity\footnote{Contractivity refers to the existence in $T^\mu$ of matrices with arbitrarily large gaps between consecutive singular values.} assumptions on the support of the measure). Their results have served well the mathematical physics community, as they apply to the Anderson tight binding model (i.e. the discrete random Schrödinger operator on the integer lattice or on a strip lattice). They do not, however, apply to all Jacobi operators, as the corresponding eigenvalue equation may lead to a multiplicative process whose underlying probability measure has reducible (non generic) support.

More recently, the continuity of the Lyapunov exponent(s) has been considered in the general (not necessarily generic) setting by C. Bocker and M. Viana and by A. Avila, A. Eskin and M. Viana. Their results do not provide a modulus of continuity, and in consequence are not applicable to spectral theory problems\footnote{A more detailed review of such results—including another new quantitative statement—follows.}

We have previously established a link between the two problems formulated above, in the general setting of linear cocycles over an ergodic base dynamical system. More precisely, we have shown that if a given cocycle satisfies certain large deviation type (LDT) estimates, which are uniform in the data, then necessarily the corresponding Lyapunov exponent (LE) varies continuously with the data, and with a modulus of continuity that depends explicitly on the strength of the LDT estimates.

In this paper we establish such uniform LDT estimates for locally constant $GL_2(\mathbb{R})$ cocycles over a full Bernoulli shift in a finite number of symbols. This corresponds to a $GL_2(\mathbb{R})$-valued random multiplicative process whose probability distribution is a finite sum of point masses\footnote{We note that the case of an absolutely continuous probability distribution was already studied by E. Le Page.}.

As a consequence of our general theory, we derive a local modulus of continuity (namely weak-Hölder) for the Lyapunov exponent regarded as a function of the support of the measure. Furthermore, this implies the same local modulus of continuity for the integrated density of states (IDS) of a random Jacobi operator, near every energy level with positive Lyapunov exponent.
This project was in part motivated by a question posed to the second author by G. Stolz, related to his work with J. Chapman on (dynamical) localization for certain disordered quantum spin systems (see \cite{4}). The results in this paper partly address that question, and the approach we use has the potential to completely solve it in the future.

We are grateful to G. Stolz for his question, as it helped us channel our attention to a simpler yet still interesting model, and thus to a more attainable goal. We would also like to thank S. Griffiths for his help simplifying a probabilistic argument in the prison break section.

A case study. Let \( \{ \omega_n \}_{n \in \mathbb{Z}} \) be an i.i.d. sequence of random variables with common distribution \( \mu \), such that \( \text{supp } \mu = [a, b] \subset (0, \infty) \) and \( \int \log t \, d\mu(t) \neq 0 \). A toy model for the system considered in \cite{4} is the Jacobi operator \( H_\omega \) on \( l^2(\mathbb{Z}) \) defined as follows. If \( \psi = \{ \psi_n \}_{n \in \mathbb{Z}} \in l^2(\mathbb{Z}) \),

\[
[H_\omega \psi]_n := \begin{cases} 
\omega_n \psi_{n-1} + \psi_{n+1} & \text{if } n \text{ is even} \\
\psi_{n-1} + \omega_n \psi_{n+1} & \text{if } n \text{ is odd}.
\end{cases}
\] (1.1)

The operator \( H_\omega \) is thus an infinite, tridiagonal, selfadjoint random matrix, with zero diagonal and symmetric off-diagonals of the form \( 1, \omega, 1, \omega, 1, \omega, 1, \ldots \)

It is convenient to consider the two-step transfer matrices of the corresponding eigenvalue equation \( H_\omega \psi = E \psi \), i.e. the matrices \( g_n^E \) satisfying

\[
\begin{bmatrix} \psi_{2n+2} \\
\psi_{2n+1} \end{bmatrix} = g_n^E \begin{bmatrix} \psi_{2n} \\
\psi_{2n-1} \end{bmatrix}.
\]

Then

\[
g_n^E = \begin{bmatrix} (E^2 - 1) \frac{1}{\omega_n} & -E \omega_n \\
E \frac{1}{\omega_n} & -\omega_n \end{bmatrix},
\] (1.2)

so for every energy parameter \( E \), \( \{ g_n^E \}_{n \in \mathbb{Z}} \) is an \( \text{SL}_2(\mathbb{R}) \)-valued i.i.d. multiplicative random process.

Also, the maximal LE of the operator (1.1) at energy \( E \) is half the Lyapunov exponent \( L^+(E) \) of this process.

If \( E \neq 0 \) it can be shown that the support of the underlying probability measure satisfies the irreducibility and contraction properties. This implies the positivity (by Furstenberg’s theorem) and the Hölder continuity (by Le Page’s theorem) of the maximal Lyapunov exponent.

However, when \( E = 0 \), the process (1.2) is diagonal:

\[
g_n^0 = \begin{bmatrix} -\frac{1}{\omega_n} & 0 \\
0 & -\omega_n \end{bmatrix},
\]

hence reducible, while \( L^+(0) = |E(\log \omega_0)| = |\int \log t \, d\mu(t)| > 0. \)
Therefore, in the vicinity of the energy level $E = 0$, the study of the operator (1.1) falls outside the scope of Le Page’s theorem. However, when the probability distribution $\mu$ has finite support, the results in this paper are immediately applicable and they provide a local modulus of continuity of the LE, and thus of the IDS. This in turn implies the Wegner estimates used in the multiscale analysis that leads to the localization of the operator (1.1).

The setting. Let $\Sigma = \{1, \ldots, k\}$ be a finite space of symbols and let $p = (p_1, \ldots, p_k)$ with $p_i > 0$ for all $i = 1, \ldots, k$ and $p_1 + \ldots + p_k = 1$ be a probability vector. Consider the space $X = \Sigma^\mathbb{Z}$ of all sequences $x = \{x_n\}_{n \in \mathbb{Z}}$ of symbols from $\Sigma$, endowed with the product probability measure $\mathbb{P} = p^\mathbb{Z}$. Then the map $T: X \to X$, $T\{x_n\}_{n \in \mathbb{Z}} := \{x_{n+1}\}_{n \in \mathbb{Z}}$ is an ergodic transformation called the (full) Bernoulli shift.

Any function from $\Sigma$ to the general linear group $\text{GL}_2(\mathbb{R})$, that is, any $k$-tuple $(A_1, \ldots, A_k)$ of two by two invertible matrices determines the locally constant function $A: X \to \text{GL}_2(\mathbb{R})$, $A\{x_n\}_{n \in \mathbb{Z}} := A_{x_0}$.

The corresponding random cocycle, or linear cocycle over the Bernoulli shift $T$ is the transformation $F_A: X \times \mathbb{R}^2 \to X \times \mathbb{R}^2$, $F_A(x, v) := (Tx, A(x)v)$.

We identify the transformation $F_A$ with the function $A$ and furthermore with the $k$-tuple $(A_1, \ldots, A_k)$, and refer to either of them as a cocycle. We denote by $\text{GL}_2(\mathbb{R})^k$ the space all such cocycles/$k$-tuples.

We endow the space $\text{GL}_2(\mathbb{R})^k$ with the uniform distance $d(A, B) := \sup_{x \in X} \|A(x) - B(x)\| = \max_{1 \leq j \leq k} \|A_j - B_j\|$, where $A = (A_1, \ldots, A_k)$ and $B = (B_1, \ldots, B_k)$.

The $n$-th iterate of a cocycle $A \in \text{GL}_2(\mathbb{R})^k$ is given by

$$A^{(n)}(x) = A(T^{n-1}x) \ldots A(Tx)A(x) = A_{x_{n-1}} \ldots A_{x_1}A_{x_0},$$

and it encodes a product of $n$ i.i.d. random matrices.

Given any cocycle $A$, by the Furstenberg-Kesten theorem there exist two numbers, $L^+(A)$ and $L^-(A)$, called the Lyapunov exponents of $A$, such that $\mathbb{P}$-almost surely,

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

This is the multiplicative analogue of the law of large numbers.
It is clear that $L^+(A) \geq L^-(A)$. When $L^+(A) > L^-(A)$, by the Oseledets multiplicative ergodic theorem, there exists a (proper) measurable decomposition $\mathbb{R}^2 = \mathcal{E}^+_A(x) \oplus \mathcal{E}^-_A(x)$ such that $\mathbb{P}$-almost surely $A(x) \mathcal{E}^+_A(x) = \mathcal{E}^+_A(Tx)$ and

$$\lim_{n \to \pm \infty} \frac{1}{n} \log \|A^n(x)v\| = L^+(A) \quad \text{for all } v \in \mathcal{E}^+_A(x) \setminus \{0\}$$

$$\lim_{n \to \pm \infty} \frac{1}{n} \log \|A^n(x)v\| = L^-(A) \quad \text{for all } v \in \mathcal{E}^-_A(x) \setminus \{0\}.$$ 

The projective line, denoted by $\mathbb{P} = \mathbb{P}(\mathbb{R}^2)$, is the space of all lines (1-dimensional linear subspaces) in $\mathbb{R}^2$. Henceforth whenever $\hat{p} \in \mathbb{P}$ the letter ‘$p$’ will stand for any non-zero vector $p \in \hat{p}$. A natural metric is defined in $\mathbb{P}$ by

$$d(\hat{p}, \hat{q}) := \frac{\|p \wedge q\|}{\|p\| \|q\|} = |\sin \angle(p, q)|.$$ 

We identify the lines $\mathcal{E}^ \pm_A(x)$ with points in the projective space $\mathbb{P}(\mathbb{R}^2)$, so the components of the Oseledets decomposition are regarded as functions $\mathcal{E}^\pm_A : X \to \mathbb{P}(\mathbb{R}^2)$. Let $L^1(X, \mathbb{P}(\mathbb{R}^2))$ be the space of all Borel measurable functions $\mathcal{E} : X \to \mathbb{P}(\mathbb{R}^2)$. On this space we consider the distance

$$d(\mathcal{E}_1, \mathcal{E}_2) := \mathbb{E}_x d(\mathcal{E}_1(x), \mathcal{E}_2(x)).$$

Note that if $A \in \text{GL}_2(\mathbb{R})^k$ is a cocycle with $L^+(A) > L^-(A)$, then $\mathcal{E}^\pm_A \in L^1(X, \mathbb{P}(\mathbb{R}^2))$.

**Statements.** Let us formally introduce the concept of large deviations for iterates of a linear cocycle.

**Definition 1.1.** A cocycle $A : X \to \text{GL}_2(\mathbb{R})$ satisfies an exponential LDT estimate if there is a constant $c > 0$ and for every small enough $\epsilon > 0$ there is $\bar{n} = \bar{n}(A, \epsilon) \in \mathbb{N}$ such that for all $n \geq \bar{n}$,

$$\mathbb{P} \left[ \left| \frac{1}{n} \log \|A^n\| - L^+(A) \right| > \epsilon \right] < e^{-cn}.$$ 

We say that a cocycle $A \in \text{GL}_2(\mathbb{R})^k$ satisfies a uniform exponential LDT estimate if the constants $c$ and $\bar{n}$ above are stable under small perturbations of $A$. We formulate this more precisely below.

---

5We refer to the constants $c$ and $\bar{n}$ as the LDT parameters of $A$. They depend on $A$, and in general they may blow up as $A$ is perturbed.
Definition 1.2. A cocycle $A \in \text{GL}_2(\mathbb{R})^k$ satisfies a uniform exponential LDT estimate if there are constants $\delta > 0$, $c > 0$ and for every small enough $\epsilon > 0$ there is $\bar{n} = \bar{n}(A, \epsilon) \in \mathbb{N}$ such that

$$P \left[ \frac{1}{n} \log \| B^{(n)} \| - L^+(B) > \epsilon \right] < e^{-cn^2}$$

for all cocycles $B \in \text{GL}_2(\mathbb{R})^k$ with $d(B, A) < \delta$ and for all $n \geq \bar{n}$.

Below we define a weaker version of the LDT estimate, where instead of the exponential, we have a sub-exponential decay of the probability of the tail event.

Definition 1.3. A cocycle $A : X \to \text{GL}_2(\mathbb{R})$ satisfies a sub-exponential LDT estimate if there are constants $\bar{n} \in \mathbb{N}$ and $a, b > 0$ such that for all $n \geq \bar{n}$,

$$P \left[ \frac{1}{n} \log \| A^{(n)} \| - L^+(A) > n^{-a} \right] < e^{-n^b}.$$

Such an estimate is called uniform if the LDT parameters $\bar{n}, a, b$ are stable under perturbations of the cocycle $A$ in $\text{GL}_2(\mathbb{R})^k$.

Definition 1.4. A cocycle $A = (A_1, \ldots, A_k) \in \text{GL}_2(\mathbb{R})^k$ is called diagonalizable if the matrices $A_1, \ldots, A_k$ are simultaneously diagonalizable over $\mathbb{R}$.

We are now ready to formulate the main result of this paper, from which everything else follows.

Theorem 1.1. Consider the space $\text{GL}_2(\mathbb{R})^k$ of locally constant cocycles over a full Bernoulli shift in $k$ symbols. Let $A \in \text{GL}_2(\mathbb{R})^k$ be a cocycle with $L^+(A) > L^-(A)$.

If $A$ is diagonalizable, then it satisfies a uniform sub-exponential LDT estimate.

If $A$ is not diagonalizable, then it satisfies a uniform exponential LDT estimate.

In order to formulate our result on the continuity of the LE and that of the Oseledets decomposition, we define some specific moduli of continuity.

Definition 1.5. Let $(M, d)$ and $(N, d)$ be two metric spaces and let $f : M \to N$ be a function.

We say that $f$ is locally Hölder continuous near $a \in M$ if there are a neighborhood $\mathcal{N}_a$ of $a$ in $M$ and constants $C < \infty$ and $\alpha > 0$ such that for all $b_1, b_2 \in \mathcal{N}_a$ we have

$$d(f(b_1), f(b_2)) \leq C \cdot d(b_1, b_2)^\alpha.$$
Moreover, \( f \) is locally weak-Hölder continuous near \( a \in M \) if there are a neighborhood \( \mathcal{N}_a \) of \( a \) in \( M \) and constants \( C < \infty \), \( \alpha > 0 \) and \( \sigma \in (0, 1] \) such that for all \( b_1, b_2 \in \mathcal{N}_a \) we have
\[
d(f(b_1), f(b_2)) \leq C \exp \left( -\alpha \left( \log \frac{1}{d(b_1, b_2)} \right)^\sigma \right).
\]

Note that the case \( \sigma = 1 \) corresponds to Hölder continuity.

The following continuity statements are direct consequences of Theorem 1.1 and the abstract continuity theorem in [6] or [7].

**Theorem 1.2.** Consider the space \( \text{GL}_2(\mathbb{R})^k \) of locally constant cocycles over a full Bernoulli shift in \( k \) symbols. Then the Lyapunov exponents functions \( L^\pm : \text{GL}_2(\mathbb{R})^k \to \mathbb{R} \) are continuous at all points.

Furthermore, let \( A \in \text{GL}_2(\mathbb{R})^k \) be a cocycle with \( L^+(A) > L^-(A) \).

If \( A \) is diagonalizable, then locally near \( A \), the Lyapunov exponents \( L^+ \) and \( L^- \) are weak-Hölder continuous.

Moreover, locally near \( A \), the maps \( B \mapsto E^\pm_B \in L^1(X, \mathbb{P}(\mathbb{R}^2)) \) are well defined and weak-Hölder continuous.

If \( A \) is not diagonalizable, then the same results hold but with the stronger, Hölder modulus of continuity.

We note that if \( A \) is a reducible cocycle with \( L^+(A) = L^-(A) \), then near \( A \) the modulus of continuity of the LE may drastically deteriorate (below weak-Hölder), as shown in our recent work [8] with M. Santos.

**Applications to mathematical physics.** We now describe some consequences of the continuity result above to the spectral theory of random Jacobi operators.

Let \( \{v_n\}_{n \in \mathbb{Z}} \) and \( \{w_n\}_{n \in \mathbb{Z}} \) be two i.i.d. sequences of real-valued, random variables (with possibly different distributions) such that moreover \( \{v_n\} \) and \( \{w_n\} \) are independent from each other. Assume also that \( v_1, w_1 \) are almost surely bounded and \( \mathbb{E}(\log |w_1|) > -\infty \).

The corresponding random Jacobi operator is the operator \( H_\omega \) on \( l^2(\mathbb{Z}) \) defined as follows: if \( \psi = \{\psi_n\}_{n \in \mathbb{Z}} \in l^2(\mathbb{Z}) \), then for all \( n \in \mathbb{Z} \),
\[
[H_\omega \psi]_n := -(w_{n+1}(\omega)\psi_{n+1} + w_n(\omega)\psi_{n-1}) + v_n(\omega)\psi_n.
\]

(1.6)

Denote by \( P_n : l^2(\mathbb{Z}) \to \mathbb{R}^n \) the coordinate projection to the range \( \{0, 1, 2, \ldots, n - 1\} \subset \mathbb{Z} \), and let
\[
H^{(n)}_\omega := P_n H_\omega P_n^*,
\]

(1.7)

Thus we obtain another proof of the result of C. Bocker and M. Viana on the continuity of the LE on the whole space of cocycles, in the finite support setting.
where $P_n^*$ stands for the adjoint of $P_n$. $H_\omega^{(n)}$ is called the finite volume truncation of $H$. By ergodicity, the following almost sure limit exists

$$N(E) := \lim_{n \to \infty} \frac{1}{n} \#((-\infty, E] \cap \text{Spectrum of } H_\omega^{(n)}),$$

and the function $E \mapsto N(E)$ is called the integrated density of states (IDS) of the ergodic operator $H_\omega$ (see [5]).

The LE and the IDS are related via the Thouless formula:

$$L(E) = \int_{\mathbb{R}} \log|E - E'| dN(E'),$$

which essentially describes one function as the Hilbert transform of the other. Thus while the IDS is known to always be log-Hölder continuous, a stronger modulus of continuity can be derived from that of the LE.

We recall some notions of localization for a random operator. More details can be found in the self contained introduction [17] to the topic of Anderson localization.

**Definition 1.6.** Let $I \subset \mathbb{R}$ be an open interval of energies such that $I$ intersects the almost sure spectrum of $H_\omega$.

An operator $H_\omega$ satisfies spectral localization in $I$ if almost surely $H_\omega$ has pure point spectrum in $I$ (i.e. $I$ contains no continuous spectrum).

A stronger (and also more physically relevant) property is that of dynamical localization, which ensures that solutions of the time dependent Schrödinger-type equation $H_\omega \psi(t) = i \partial_t \psi(t)$ stay localized in space, uniformly for all times.

An operator $H_\omega$ satisfies (a strong form of) dynamical localization in $I$ if for every compact subinterval $J \subset I$, there are constants $C < \infty$ and $\gamma, \sigma > 0$ such that for all $j, k \in \mathbb{Z}$,

$$\mathbb{E} \left( \sup_{t \in \mathbb{R}} \left| \langle e_j, e^{-itH_\omega} 1_J(H_\omega) e_k \rangle \right| \right) \leq C e^{-\gamma |j-k|^\sigma},$$

where $\{e_j\}_{j \in \mathbb{Z}}$ is the canonical orthonormal basis in $l^2(\mathbb{Z})$ and $1_J(H_\omega)$ denotes the spectral projection of $H_\omega$ onto the interval $J$. Both $e^{-itH_\omega}$ and $1_J(H_\omega)$ are defined via functional calculus for selfadjoint operators.

**Theorem 1.3.** Consider the random Jacobi operator $H_\omega$ defined in (1.6), and assume moreover that the distributions of the random variables $\{v_n\}, \{w_n\}$ have finite support.

Let $E_0$ be any energy level such that $L^+(E_0) > 0$. There exists an open interval $I$ containing $E_0$ such that on $I$, the Lyapunov exponent $L^+$ is positive, the integrated density of states $N$ is weak-Hölder continuous and almost surely, the operator $H_\omega$ satisfies dynamical localization (provided $I$ intersects the almost sure spectrum of $H_\omega$).
We remark that the case study (1.1) does not exactly fit the above theorem, since the sequence of ‘weights’ \( \{w_n\}_{n \in \mathbb{Z}} \) is given by

\[ \ldots, \omega_{-1}, 1, \omega_0, 1, \omega_1, \ldots, \]

thus it is not independent. It is however Markov, hence the Thouless formula relating the LE to the IDS still holds (it holds for every ergodic system); moreover, as the LE of \( H_\omega \) can be obtained from the i.i.d. two-step transfer matrices \( g_n^E \) in (1.2), its weak-Hölder continuity is ensured, and with it, all the other properties (modulus of continuity of the IDS and localization near the energy \( E = 0 \)).

**Related results.** We first review previous work on limit theorems for multiplicative random processes. E. Le Page proved central limit theorems, as well as a large deviation principle [13] for generic (strongly irreducible and contracting) \( \text{GL}_d(\mathbb{R}) \)-cocycles. Later P. Bougerol extended Le Page’s approach, proving similar results for Markov type random cocycles [3]. Their results are asymptotic and do not provide uniformity of the large deviation parameters in the data. In [6] (see also [7]) we established, under similar (but weaker) assumptions, finitary (i.e. non asymptotic) and uniform LDT estimates for strongly mixing Markov (so in particular also Bernoulli) cocycles.

Regarding the continuity of the LE of random linear cocycles, the first result was obtained by H. Furstenberg and Y. Kifer [9] for generic (irreducible and contracting) \( \text{GL}_d(\mathbb{R}) \)-cocycles. In [14] E. Le Page proved the Hölder continuity of the maximal LE for a one-parameter family of random \( \text{GL}_d(\mathbb{R}) \)-cocycles satisfying similar generic assumptions. A new proof of Le Page’s result (with a formulation for the whole space of quasi-irreducible cocycles) was recently obtained by A. Baraviera and P. Duarte [1].

The continuity in the whole space of random \( \text{GL}_2(\mathbb{R}) \)-cocycles (without any generic assumption), was established by C. Bocher-Neto and M. Viana in [2]. The analogue of this result for random \( \text{GL}_d(\mathbb{R}) \)-cocycles \( (d \geq 2) \) has been announced by A. Avila, A. Eskin and M. Viana (see [20], Note 10.7). An extension of [2] to a particular type of cocycles over Markov systems (particular in the sense that the cocycle still depends on one coordinate, as in the Bernoulli case) was obtained by E. Malheiro and M. Viana in [16].

Very recently, and using completely different methods from ours, E.Y. Tall and M. Viana [18] considered the problem of establishing a pointwise modulus of continuity for the LE in the whole space of random \( \text{GL}_2(\mathbb{R}) \)-cocycles. They established Hölder continuity at cocycles \( A \) with \( L^+(A) > L^-(A) \) and a weak (log-Hölder) minimum modulus of
continuity at all points. Moreover, compared to Theorem 1.2 in this paper, the results in [18] apply to any compactly supported probability distribution (and not just to finitely supported ones). However, they lack uniformity (providing a pointwise rather than a local modulus of continuity), and in particular are not suitable for mathematical physics applications.

An interesting question arising from this comparison is whether a uniform (rather than pointwise) Hölder modulus of continuity holds in the vicinity of a cocycle with simple Lyapunov exponents, or else whether our weak-Hölder result is optimal.

Other natural problems related to the results in this paper are the finite state Markov setting (which we treat in a forthcoming paper), the general compact (rather than finite) support distribution case, and the much more challenging higher dimensional cocycles setting.

The proof structure. In [6], we introduced a general method for establishing a modulus of continuity for the LE (and for the Oseledets decomposition) of linear cocycles, that employs as a black box the availability of uniform LDT estimates. Such estimates are to be obtained separately, using methods specific to the model of linear cocycles considered (we treat the irreducible random and the quasi-periodic models in the book). Our more recent monograph [7] provides a self-contained and less technical introduction to this theory, as it is concerned with the $\text{SL}_2(\mathbb{R})$ setting, which suffices for this paper.

The continuity Theorem 1.2 is thus a direct consequence of the large deviation estimates in Theorem 1.1, while the modulus of continuity of the IDS in Theorem 1.3 follows via the Thouless formula from that of the LE. Furthermore, a good enough modulus of continuity of the IDS implies a sharp enough Wegner estimate which, via multiscale analysis (a standard technique in the theory of random Schrödinger type operators) leads to the localization statement formulated above.

We now describe the structure of the proof of the uniform LDT estimates.

In [7] we derived uniform LDT estimates for quasi-irreducible cocycles, that is, for cocycles satisfying a rather weak form of irreducibility. It turns out that any random SL$_2(\mathbb{R})$ cocycle is diagonalizable or else, itself or its inverse is quasi-irreducible. This will allow us to reduce the problem to SL$_2(\mathbb{R})$-valued diagonalizable cocycles (see Section 2).

---

7This method is based on ideas introduced by M. Goldstein and W. Schlag [10] in their study of quasi-periodic Schrödinger operators.
A diagonalizable cocycle trivially satisfies LDT estimates (or any other limit theorems), and it does so uniformly within the set \( \mathcal{D}_{\text{Diag}} \) of all diagonalizable cocycles. The problem is that this set is not open. Nearby quasi-irreducible cocycles will satisfy LDT estimates, but there is no a priori reason why their LDT parameters do not blow up while approaching \( \mathcal{D}_{\text{Diag}} \), and herein lies the challenge of the proof.

In fact, referring to Definition 1.3 we will show that \( a, b \) are some universal constants, and the problem of stability under perturbations of the LDT parameters will reduce to the stability of \( \bar{n} \), the minimum number of iterates required for the LDT to start applying. The idea of the proof is to introduce a quantitative measurement of irreducibility, and relate it to that minimum number of iterates \( \bar{n} \). There are different ways of measuring the irreducibility of a cocycle (see Section 3, where these measurements are introduced and related).

For the purpose of this introduction, given a cocycle \( B \), consider its irreducibility measurement to be the distance \( d(B, \mathcal{D}_{\text{Diag}}) \) to the set of diagonalizable cocycles. If the cocycle \( B \) is close enough to a diagonalizable cocycle \( B_\flat \), then up to a certain finite number of iterates \( m(B) \) that depends on \( d(B, B_\flat) \), the LDT estimate of \( B_\flat \) transfers over to \( B \) by proximity. On the other hand, after a large enough number of iterates \( \tilde{m}(B) \), the irreducibility kicks in and it provides an LDT for \( B \) (see Section 5). There are then two issues left to address.

The first is to derive an explicit relationship between the irreducibility measurement \( d(B, \mathcal{D}_{\text{Diag}}) \) of a cocycle and the minimum number of iterates \( \tilde{m}(B) \) required so that the (quantitative) LDT for quasi-irreducible cocycles applies. A key ingredient in this derivation is a probabilistic (random walk) argument which we refer to as the prison break (see Section 4). We note that this is the only part in this paper requiring the finite support condition on the probability distribution.

The second is to bridge the range of values between the iterates \( m(B) \) and \( \tilde{m}(B) \). This is achieved in Section 6 by performing an ‘almost linearization’ of the multiplicative process by means of the Avalanche Principle—a deterministic result on the product of a long chain of matrices satisfying some geometric conditions (see [6, 7]).

2. Reduction to diagonalizable cocycles

In this section we show that establishing uniform LDT estimates for \( \text{GL}_2(\mathbb{R}) \) random cocycles and a modulus of continuity for their LE can be reduced to the setting of \( \text{SL}_2(\mathbb{R}) \) valued diagonalizable cocycles.
Special linear group reduction. If \( g \in \text{GL}_2(\mathbb{R}) \), let
\[
g_* := \frac{1}{\sqrt{|\det g|}} \in \text{SL}_2^\pm(\mathbb{R}),
\]
where \( \text{SL}_2^\pm(\mathbb{R}) \) is the group of two dimensional matrices with determinant \( \pm 1 \). As in this context it makes no difference whether the determinant of a matrix is 1 or \(-1\), we will disregard the sign of the determinant and simply write \( \text{SL}_2(\mathbb{R}) \).

Moreover, for a cocycle \( A \in \text{GL}_2(\mathbb{R})^k \), denote by \( A_* \) the induced \( \text{SL}_2(\mathbb{R})^k \) cocycle given by
\[
A_*(x) := \frac{1}{\sqrt{|\det A(x)|}} A(x).
\]

Note that since for any matrix \( g_* \in \text{SL}_2(\mathbb{R}) \) the operator norm satisfies
\[
\|g_*^{-1}\| = \|g_*\| \geq 1,
\]
if \( A_* \in \text{SL}_2(\mathbb{R})^k \), then
\[
L^+(A_*) \geq 0 \geq L^-(A_*) \quad \text{and} \quad L^-(A_*) = -L^+(A_*).
\]

Because of this, from now on, if \( A_* \in \text{SL}_2(\mathbb{R})^k \), we will denote its maximal Lyapunov exponent \( L^+(A_*) \) by \( L(A_*) \) and refer to it as the Lyapunov exponent of \( A_* \).

**Proposition 2.1.** Let \( A \in \text{GL}_2(\mathbb{R})^k \) be a cocycle. If \( A_* \) satisfies a uniform LDT estimate in \( \text{SL}_2(\mathbb{R})^k \), then \( A \) satisfies a similar LDT estimate in \( \text{GL}_2(\mathbb{R})^k \). Moreover, if \( L : \text{SL}_2(\mathbb{R})^k \to \mathbb{R} \) has a certain (at most Lipschitz) modulus of continuity near \( A_* \), then \( L^+ : \text{GL}_2(\mathbb{R})^k \to \mathbb{R} \) has the same modulus of continuity near \( A \).

**Proof.** It is easy to verify that the maps
\[
\text{GL}_2(\mathbb{R}) \ni g \mapsto \log |\det g| \in \mathbb{R} \quad \text{and} \quad \text{GL}_2(\mathbb{R}) \ni g \mapsto g_* \in \text{SL}_2(\mathbb{R})
\]
are locally Lipschitz continuous.

This then implies that the maps
\[
\text{GL}_2(\mathbb{R})^k \ni A \mapsto \xi_A := \log |\det A| \in L^\infty(X, \mathbb{R}) \quad \text{and}
\]
\[
\text{GL}_2(\mathbb{R})^k \ni A \mapsto A_* \in \text{SL}_2(\mathbb{R})^k
\]
are locally Lipschitz as well.

The continuity statement in the proposition is then evident. To derive the LDT statement, note that
\[
\frac{1}{n} \log \|A^{(n)}\| = \frac{1}{n} \log \|A_*^{(n)}\| + \frac{1}{2n} \sum_{i=0}^{n-1} \xi_A \circ T^i. \tag{2.1}
\]

It is then enough to establish a uniform LDT estimate for the Birkhoff sum on the right hand side.
Let us recall the Hoeffding inequality from classical probabilities (see for instance [19]), which will be used again later.

**Lemma 2.2.** Let $\xi_0, \xi_1, \ldots, \xi_{n-1}$ be a (finite) random process and denote by $S_n := \xi_0 + \xi_1 + \ldots + \xi_{n-1}$ the corresponding sum process.

Assume that for $0 \leq j \leq n-1$ the random variables $\xi_j$ are independent and that almost surely $|\xi_j| \leq K$. Then for all $\epsilon > 0$ we have
\[
P \left( \left| \frac{1}{n} S_n - \mathbb{E} \left( \frac{1}{n} S_n \right) \right| > \epsilon \right) \leq e^{-\frac{\epsilon^2}{2K^2 n}}.
\]

What makes this estimate important for our purposes (more so than say, Cramér’s large deviation principle) is its uniformity in the input data: the parameters in the estimate depend only on the almost sure bound on the process.

If in (2.1) we put $\xi_i := \xi_A \circ T^i$, then Hoeffding’s inequality applies, and it does so uniformly for all nearby cocycles, since $B \mapsto \xi_B$ is Lipschitz, so the almost sure bound on $\xi_B$ is uniform in $B$. □

From now on we will only consider SL$_2(\mathbb{R})$ cocycles.

**Reduction to diagonalizable cocycles.** We first define a weak form of irreducibility.

Let $A \in$ SL$_2(\mathbb{R})^k$ and assume that the line $\ell \subset \mathbb{R}^2$ is invariant under all matrices (regarded as linear transformations) $A_j$ with $j \in \Sigma$. In other words, $A_j \ell = \ell$ for all $j \in \Sigma$, and we can consider the restriction $A\mid_\ell$ of the cocycle $A$ to the one dimensional subspace $\ell$. This restriction may be described as $A\mid_\ell v = \lambda_A v$ for some function $\lambda_A : \Sigma \rightarrow \mathbb{R}$ and unit vector $v \in \ell$. The process $\log \|A\mid_\ell\|$ is thus additive, and by Birkhoff’s ergodic theorem, the Lyapunov exponent of $A\mid_\ell$ is $L(A\mid_\ell) = \int_{\Sigma} \log |\lambda_A| \, dp = \mathbb{E} (\log \|A\mid_\ell\|)$.

Note also that from the Oseledets theorem, $L(A\mid_\ell) = L^+(A) = L(A)$ or $L(A\mid_\ell) = L^-(A) = -L(A)$.

**Definition 2.1.** A cocycle $A \in$ SL$_2(\mathbb{R})^k$ is called quasi-irreducible if there is no line $\ell \subset \mathbb{R}^2$ such that $\ell$ is invariant under all matrices $A_j$ with $j \in \Sigma$ and $L(A\mid_\ell) < L(A)$. \(^8\)

**Definition 2.2.** A cocycle $A \in$ SL$_2(\mathbb{R})^k$ is called diagonalizable if there exist two transversal invariant lines $\ell$ and $\ell'$, that is, $A_j \ell = \ell$ and $A_j \ell' = \ell'$ for all $j \in \Sigma$.

\(^8\)We later prove (see Lemma 3.1) that when $L(A) > 0$, this is equivalent to the concept of quasi-irreducibility introduced in [3] (see Définition 2.7).
We denote by \( \mathcal{D} \) the set of all diagonalizable cocycles in \( \text{SL}_2(\mathbb{R})^k \) and by \( \mathcal{D}^* \) the set of cocycles \( A \in \mathcal{D} \) with \( L(A) > 0 \).

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

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

It is then clear that \( L(A^{-1}) = L(A) \).

Moreover, if the line \( \ell \subset \mathbb{R}^2 \) is \( A \)-invariant then it is also invariant for \( A^{-1} \) and \( L(A^{-1}|_\ell) = \int_T \log |\lambda^{-1}_A| \, dp = -L(A|_\ell) \).

**Lemma 2.3.** For a cocycle \( A \in \text{SL}_2(\mathbb{R})^k \), the following dichotomy holds: \( A \) is diagonalizable or else, \( A \) or \( A^{-1} \) is quasi-irreducible.

**Proof.** Let \( A \in \text{SL}_2(\mathbb{R})^k \) be a non diagonalizable cocycle. Then either \( A \) has no invariant lines, so in particular it is quasi-irreducible, or it has exactly one invariant line \( \ell \). In this case, either \( L(A|_\ell) = L(A) \), so \( A \) is quasi-irreducible, or \( L(A|_\ell) = -L(A) \). But then \( \ell \) is also invariant for \( A^{-1} \) and \( L(A^{-1}) = L(A) = -L(A|_\ell) = L(A^{-1}|_\ell) \), so \( A^{-1} \) is quasi-irreducible. \( \square \)

**Lemma 2.4.** If a cocycle \( A \in \text{SL}_2(\mathbb{R})^k \) satisfies a uniform LDT estimate, then \( A^{-1} \) satisfies the same uniform LDT estimate.

**Proof.** It is easy to see that the map \( \text{SL}_2(\mathbb{R})^k \ni B \mapsto B^{-1} \in \text{SL}_2(\mathbb{R})^k \) is locally bi-Lipschitz. Then for any cocycle \( B^{-1} \) near \( A^{-1} \) (which ensures the proximity of \( B \) to \( A \)), since

\[
(B^{-1})^{(n)}(x) = B^{(n)}(T^{-n}x)^{-1} \quad \text{and} \quad L(B^{-1}) = L(B),
\]

it follows that the event

\[
\left[ \frac{1}{n} \log \| (B^{-1})^{(n)} \| - L(B^{-1}) \right] > n^{-a} \]

\[
T^{-n} \left[ \frac{1}{n} \log \| B^{(n)} \| - L(B) \right] > n^{-a} \cdot
\]

We then conclude that

\[
\mathbb{P} \left[ \left| \frac{1}{n} \log \| (B^{-1})^{(n)} \| - L(B^{-1}) \right| > n^{-a} \right] = \mathbb{P} \left[ \left| \frac{1}{n} \log \| B^{(n)} \| - L(B) \right| > n^{-a} \right],
\]

which shows that \( B^{-1} \) satisfies the same LDT as \( B \). \( \square \)
Remark 2.1. We have already established exponential uniform LDT estimates for quasi-irreducible cocycles (see Theorem 4.1 in [7]). Thus the two lemmas above reduce the proof of Theorem 1.1 to establishing (sub-exponential) uniform (relative to the whole space $\text{SL}_2(\mathbb{R})^k$) LDT estimates for diagonalizable $\text{SL}_2(\mathbb{R})$ cocycles.

**LDT in the set of diagonalizable cocycles.** We derive a uniform LDT estimate *within* the set $\text{Diag}^*$ of diagonalizable cocycles with positive Lyapunov exponent.

**Theorem 2.1.** Let $A \in \text{Diag}$ with $L(A) > 0$. There are constants $\delta = \delta(A) > 0$, $c = c(A) > 0$ and $C = C(A) < \infty$ such that if $D \in \text{Diag}$ is any diagonalizable cocycle with $d(D, A) < \delta$, then

$$
\mathbb{P} \left[ \frac{1}{n} \log \|D^{(n)}\| - L(D) > \varepsilon \right] \leq e^{-c(A)\varepsilon^2 n}
$$

for all $\varepsilon > 0$ and $n \geq \frac{C(A)}{\varepsilon}$.

*Proof.* Let $A = (A_1, \ldots, A_k) \in \text{Diag}^*$. This is equivalent to the existence of a matrix $p_A \in \text{GL}_2(\mathbb{R})$ and a diagonal cocycle $D_A : X \to \text{SL}_2(\mathbb{R})$, such that

$$A(x) = p_A D_A(x) p_A^{-1} \quad \text{for all } x \in X. \quad (2.2)$$

We write

$$D_A(x) = \begin{bmatrix} \theta_A(x) & 0 \\ 0 & \theta_A(x)^{-1} \end{bmatrix},$$

where $\theta_A : X \to \mathbb{R} \setminus \{0\}$ is a locally constant function (in other words, it can be identified with a vector $(\theta_{A,1}, \ldots, \theta_{A,k}) \in \mathbb{R}^k$).

The numbers $\theta_A(x)$ and $\theta_A(x)^{-1}$ are the eigenvalues of the matrix $A(x)$, while the columns of the matrix $p_A$ (which simultaneously diagonalizes $A(x), x \in X$) are corresponding eigenvectors. We show that we can choose $p_A$ and $D_A$ so that $A \mapsto p_A$ and for all $x \in X, A \mapsto D_A(x)$ are locally continuous (in fact, locally analytic).

First note that

$$L(A) = \| \mathbb{E} \log |\theta_A| \| = \left| \sum_{j=1}^k p_j \log |\theta_{A,j}| \right|. \quad (2.3)$$

Indeed, for all $x \in X$ we have $A^{(n)}(x) = p_A D_A^{(n)}(x) p_A^{-1}$, and

$$D_A^{(n)}(x) = \begin{bmatrix} \theta_A^{(n)}(x) & 0 \\ 0 & \theta_A^{(n)}(x)^{-1} \end{bmatrix} \quad \text{where } \theta_A^{(n)}(x) := \prod_{j=n-1}^0 \theta_A(T^j x).$$

Thus

$$\|D_A^{(n)}(x)\| = \max \left\{ |\theta_A^{(n)}(x)|, |\theta_A^{(n)}(x)^{-1}| \right\}$$
and it follows that for all $x \in X$,

$$
\left| \frac{1}{n} \log \| A^{(n)}(x) \| - \frac{1}{n} \sum_{j=0}^{n-1} \log \| \theta_A(T^j x) \| \right| \leq \frac{1}{n} \left( \log \| p_A \| + \log \| p_A^{-1} \| \right).
$$

(2.4)

Applying the Furstenberg-Kesten theorem to the cocycle $A$ and the Birkhoff ergodic theorem to the observable $\log |\theta_A|$, we derive (2.3).

Then since $L(A) > 0$ and $L(A) = \sum_{j=1}^k p_j \log |\theta_{A,j}|$, there is an index $j$ such that $|\theta_{A,j}| \neq 1$. Therefore, at least one of the matrices $A_j$, say $A_1$ is hyperbolic.

In particular, its eigenvalues $\theta_{A,1}$ and $\theta_{A,1}^{-1}$ are simple. As a consequence of the general perturbation theory of linear operators (see [11], or the more modern approach [12]) we have the following. As we perturb the cocycle $A$ in a small enough neighborhood in $\text{Diag}^*$, its first component $A_1$ is perturbed in a small neighborhood in $\text{GL}_2(\mathbb{R})$, hence the corresponding eigenvalues $\theta_{A,1}$ and $\theta_{A,1}^{-1}$ remain simple (and they can be chosen to depend analytically on $A_1$). Furthermore, we can choose the corresponding eigenvectors $u_{A,1}^+$ and $u_{A,1}^-$ to (locally) depend analytically on $A_1$, hence in particular they depend Lipschitz continuously on $A$. The matrix $p_A$ whose column vectors are $u_{A,1}^+$ and $u_{A,1}^-$ diagonalizes $A_1$, hence it diagonalizes $A_1, \ldots, A_k$.

We conclude that in (2.2) we can choose the matrix $p_A$ in such a way that $\text{Diag}^* \ni A \mapsto p_A$ is locally Lipschitz continuous, and in particular it is locally bounded. Combined with the fact that the function $\text{GL}_2(\mathbb{R}) \ni g \mapsto g^{-1} \in \text{GL}_2(\mathbb{R})$ is locally Lipschitz continuous, we derive the same property for $\text{Diag}^* \ni A \mapsto p_A^{-1}$.

Finally, since $D_A = p_A^{-1} A p_A$, the map $\text{Diag}^* \ni A \mapsto D_A$ is also locally Lipschitz continuous, hence $\text{Diag}^* \ni A \mapsto \log |\theta_A|$ are locally Lipschitz continuous and bounded as well.

From the above it follows in particular that there is a locally bounded function $C: \text{Diag}^* \to (0, \infty)$ such that $(\log \| p_A \| + \log \| p_A^{-1} \|) \leq C(A)$ and $\| \log |\theta_A|\| \leq C(A)$.

Therefore, (2.4) implies that for all $x \in X$,

$$
\left| \frac{1}{n} \log \| A^{(n)}(x) \| - \frac{1}{n} \sum_{j=0}^{n-1} \log \| \theta_A(T^j x) \| \right| \leq \frac{C(A)}{n} \leq \frac{\epsilon}{2} \quad (2.5)
$$

if $n \geq \frac{2C(A)}{\epsilon}$.
Furthermore, the i.i.d. random variables \( \xi_{A,j} := \log |\theta_A \circ T^j| \) are bounded by \( C(A) \), hence Hoeffding’s inequality in Lemma 2.2 implies

\[
\mathbb{P} \left[ \left| \frac{1}{n} \sum_{j=0}^{n-1} \log |\theta_A(T^j x)| - \mathbb{E} \log |\theta_A| \right| \geq \frac{\epsilon}{2} \right] \leq e^{-\frac{\epsilon^2}{8C(A)^2} n}. \tag{2.6}
\]

The conclusion follows by combining (2.3), (2.5) and (2.6). □

Remark 2.2. From the above considerations we also conclude that the restriction \( L : \text{Diag}^* \to \mathbb{R} \) of the Lyapunov exponent is a locally Lipschitz continuous function. That is because \( L(A) = \left| \mathbb{E} \log |\theta_A| \right| \) and \( A \mapsto \log |\theta_A| \) is locally Lipschitz. Furthermore, \( L \) is continuous on \( \text{Diag} \), as continuity at \( \text{SL}_2(\mathbb{R}) \) cocycles with zero LE is automatic.

3. Irreducibility Measurements

In this section we introduce two irreducibility measurements \( \rho(B) \) and \( N(B) \) for a cocycle \( B \), which we then relate to the distance \( d(B, \text{Diag}) \) of the cocycle \( B \) to the space of diagonalizable cocycles.

Lemma 3.1. Let \( A \) be an \( \text{SL}_2(\mathbb{R}) \) valued cocycle with \( L(A) > 0 \). The following statements are equivalent:

(i) \( A \) is quasi-irreducible.

(ii) \( \frac{1}{n} \log \| A^n v \| \to L(A) \) \( \mathbb{P} \)-almost surely, for all \( v \in \mathbb{R}^2 \), \( v \neq 0 \).

Proof. The implication (ii) \( \implies \) (i) is obvious. If there is an \( A \)-invariant line \( \ell \), picking \( v \in \ell \), \( v \neq 0 \), the following holds \( \mathbb{P} \)-almost surely:

\[
L(A|_{\ell}) = \lim_{n \to \infty} \frac{1}{n} \log \| A^n (x) v \| = L(A), \quad \text{so} \quad L(A|_{\ell}) = L(A).
\]

Now we prove the implication (i) \( \implies \) (ii).

Let \( F \subset X \) be a \( T \)-invariant Borel set with full probability, \( \mathbb{P}(F) = 1 \), consisting of Oseledets regular points. Thus for all \( x \in F \) we have the Oseledets decomposition \( \mathbb{R}^2 = \mathcal{E}^+(x) \oplus \mathcal{E}^-(x) \) into proper subspaces, which are invariant under the cocycle action, i.e., \( A(x) \mathcal{E}^\pm(x) = \mathcal{E}^\pm(Tx) \). Moreover, given any \( x \in F \) and \( v \in \mathbb{R}^2 \), \( v \neq 0 \),

\[
\lim_{n \to \infty} \frac{1}{n} \log \| A^n (x) v \| = \begin{cases} -L(A) & \text{if} \ v \in \mathcal{E}^-(x), \\
L(A) & \text{if} \ v \notin \mathcal{E}^-(x). \end{cases}
\]

On the other hand, by Theorem A in \([9]\), for every \( v \in \mathbb{R}^2 \), \( v \neq 0 \), there is a constant \( \lambda_v \) and a set \( F_v \subset X \) with \( \mathbb{P}(F_v) = 1 \) such that for all \( x \in F_v \),

\[
\lim_{n \to \infty} \frac{1}{n} \log \| A^n (x) v \| = \lambda_v.
\]
Hence for all nonzero vectors \( v \), \( \lambda_v = L(A) \) or \( \lambda_v = -L(A) \) and
\[
\text{either } \lim_{n \to \infty} \frac{1}{n} \log \| A^n(x)v \| = L(A) \quad \text{for all } x \in F_v, \\
or \lim_{n \to \infty} \frac{1}{n} \log \| A^n(x)v \| = -L(A) \quad \text{for all } x \in F_v.
\]
Consider the set
\[
S := \{ v \in \mathbb{R}^2 : \text{or } \text{for all } x \in F_v \}
\]
It is easy to see that \( S \) is a linear subspace. We prove that \( S \) is invariant under the cocycle. Because locally constant random cocycles over full Bernoulli shifts factor through the one-sided shift, without loss of generality we may assume that \( T : X \to X \) is the lateral shift with \( X = \Sigma^\mathbb{N} \).

Given \( i \in \Sigma \) consider the conditional probability measure, \( \mathbb{P}_i(F) := \mathbb{P}(F|x_0 = i) \), defined for Borel sets \( F \subset X \). By definition the family of probability measures \( \{ \mathbb{P}_i \}_{i \in \Sigma} \) is the desintegration of \( \mathbb{P} \) through the canonical projection \( \pi : X \to \Sigma \), \( \pi(x) := x_0 \). Because \( \mathbb{P} \) is the product measure \( \mathbb{P}^\Sigma \), it follows that \( \mathbb{P}_i(F) = \mathbb{P}(F_i) \), where for any Borel set \( F \subset X \) we define \( F_i := \{ x \in X : ix \in F \} \).

Now, given \( v \in S \) with \( v \neq 0 \), consider the full probability event \( F^v := \{ x \in X : v \in \mathcal{E}^-(x) \} \). If \( i \in \Sigma \), since
\[
1 = \mathbb{P}(F^v) = \int_{\Sigma} \mathbb{P}_j(F^v) \, dp(j) = \int_{\Sigma} \mathbb{P}(F_j^v) \, dp(j)
\]
the event \( F_i^v = \{ x \in X : v \in \mathcal{E}^-(ix) \} \) has also probability one. By invariance of the Oseledets splitting, if \( v \in \mathcal{E}^-(ix) \) then \( A_i v \in \mathcal{E}^-(x) \), which implies that \( A_i v \in \mathcal{E}^-(x) \) \( \mathbb{P} \)-almost surely, thus proving that \( A_i v \in S \). Therefore, \( S \) is an \( A \)-invariant linear subspace of \( \mathbb{R}^2 \).

If \( \dim S = 2 \), then \( \mathbb{P} \)-almost surely, \( \mathcal{E}^-(x) = \mathbb{R}^2 \), which contradicts the fact that \( L(A) > 0 \). If \( \dim(S) = 1 \) then \( S \) is an \( A \)-invariant line and picking \( v \in S \), \( v \neq 0 \), \( \mathbb{P} \)-almost surely we have
\[
L(A|_S) = \lim_{n \to \infty} \frac{1}{n} \log \| A^n(x)v \| = -L(A).
\]
Thus \( L(A|_S) = -L(A) \), which contradicts the quasi-irreducibility of \( A \). It follows that \( S = \{ 0 \} \), so for every \( v \neq 0 \), \( \frac{1}{n} \log \| A^n(x)v \| \to -L(A) \) does not happen \( \mathbb{P} \)-almost surely. But then, from the preceding dichotomy, \( \frac{1}{n} \log \| A^n(x)v \| \to L(A) \) must happen \( \mathbb{P} \)-almost surely. \( \square \)
Lemma 3.2. Let $A \in \text{SL}_2(\mathbb{R})^k$ be a cocycle. If $A$ is quasi-irreducible and $L(A) > 0$, then

$$\lim_{n \to \infty} \min_{v \in S^1} \mathbb{E} \left[ \frac{1}{n} \log \| A^n v \| \right] = L(A).$$

Proof. Fix any vector $v \in S^1$. Using the previous lemma, we have that

$$\frac{1}{n} \log \| A^n v \| \to L(A) \quad \mathbb{P}\text{-almost surely.}$$

Since the functions $\frac{1}{n} \log \| A^n v \|$ are uniformly bounded by $\log \| A \|_{\infty} < \infty$, by the dominated convergence theorem, $\frac{1}{n} \mathbb{E} \left[ \log \| A^n v \| \right]$ converges to $L(A)$.

Assume now that this convergence is not uniform (in $v \in S^1$), in order to get a contradiction. This assumption implies the existence of a sequence of unit vectors $v_n \in \mathbb{R}^2$ and a positive number $\delta > 0$ such that for all large $n$,

$$\frac{1}{n} \mathbb{E} \left[ \log \| A^n v_n \| \right] \leq L(A) - \delta. \quad (3.1)$$

We claim now (and postpone the motivation for the end of the proof) that $\mathbb{P}$-almost surely,

$$\liminf_{n \to \infty} \frac{\| A^n v_n \|}{\| A^n \|} > 0. \quad (3.2)$$

Then $\mathbb{P}$-almost surely, $\frac{1}{n} \log \frac{\| A^n v_n \|}{\| A^n \|} \to 0$. Therefore, using again the dominated convergence theorem, we get

$$\lim_{n \to \infty} \frac{1}{n} \mathbb{E} \left[ \log \| A^n v_n \| \right] = \lim_{n \to \infty} \frac{1}{n} \mathbb{E} \left[ \log \| A^n \| \right] + \frac{1}{n} \mathbb{E} \left[ \log \frac{\| A^n v_n \|}{\| A^n \|} \right] = L(A) + 0 = L(A),$$

which contradicts (3.1).

To complete the proof we verify the claim (3.2). The argument depends upon certain geometrical considerations related to our proof of the Oseledets multiplicative ergodic theorem in [6, Chapter 4].

We review below these considerations (see also Section 2.2 in [6]).

Given a matrix $g \in \text{SL}_2(\mathbb{R})$, if $\| g \| > 1$ then its singular values are distinct, so its singular directions (most and least expanding) are

---

9This claim, in a more general setting, is also proven in [3, Proposition 2.8], using ingredients in the proof of the Oseledets multiplicative ergodic theorem given by Ledrappier [15].
well defined. The same holds also for the transpose matrix $g^*$. Then there are two orthonormal singular bases of $\mathbb{R}^2$, \{${v}_+(g), {v}_-(g)$\} and \{${v}_+(g^*), {v}_-(g^*)$\} such that

$$g {v}_+(g) = \|g\| {v}(g^*) \quad \text{and} \quad g {v}_-(g) = \|g\|^{-1} {v}(g^*).$$

If $w \in \mathbb{R}^2$ is any other vector, then using the Pythagorean’s theorem,

$$\|gw\| \|g\| \geq |w \cdot {v}_+(g) + (g)|.$$

Denote also by $\hat{v}_+(g) \in \mathbb{P}(\mathbb{R}^2)$ the projective point corresponding to the unit vector $v_+(g)$. With these notations, given the cocycle $A$, we define the sequence of partial functions $\{v_n(A)\}$ by\[v_n(A)(x) := \begin{cases} 
\hat{v}_+(A(n)(x)) & \text{if } \|A(n)(x)\| > 1 \\
\text{undefined} & \text{otherwise.}
\end{cases}\]

Since $L(A) > 0$, by Proposition 4.4 in [6], this sequence converges $\mathbb{P}$-almost surely to a (total) measurable function $v_∞(A) : X \to \mathbb{P}(\mathbb{R}^2)$,

$$v_∞(A)(x) := \lim_{n \to +\infty} v_n(A)(x).$$

As it turns out from our proof of the Oseledets theorem (see the beginning of the proof of Theorem 4.4 in [6]), the Oseledets subspace corresponding to the second Lyapunov exponent $-L(A)$ is the orthogonal complement of the most expanding direction of the cocycle $A$, that is, $\mathbb{P}$-almost surely, $E^-(x) = v_∞(A)(x)⊥$.

Claim 3.2 now follows immediately. By the compactness of the unit circle we can assume that the sequence $\{v_n\}$ converges to a unit vector $v \in \mathbb{R}^2$. Then for $\mathbb{P}$-almost every $x$, and for $n$ large enough,

$$\|A(n)(x) v_n\| \|A(n)(x)\| \geq |v_n \cdot {v}_+(A(n)(x))| \to |v \cdot v_∞(A)(x)|.$$

But if $v \in \mathbb{S}^1$, then $v \cdot v_∞(A)(x) > 0$, since otherwise we would have $v \in v_∞(A)(x)⊥ = E^-(x)$, which happens with probability zero. □

**Definition 3.1.** Let $A$ be a quasi-irreducible cocycle with $L(A) > 0$. We define $N(A)$ to be the least $n \in \mathbb{N}$ such that for all $v \in \mathbb{S}^1$ one has

$$\mathbb{E} \left[ \frac{1}{n} \log\|A(n)v\| \right] > \frac{1}{2} L(A).$$

Next we show that if $A$ is instead a diagonalizable cocycle, then Lemma 3.2 does not hold, and as $B \to A$, where $B$ is quasi-irreducible, the measurement $N(B) \to \infty$.

**Proposition 3.3.** Let $A$ be a diagonalizable cocycle with $L(A) > 0$. 
(1) There is \( v_A \in \mathbb{S}^1 \) such that for all \( n \geq 1 \),
\[
\mathbb{E} \left( \frac{1}{n} \log \| A^{(n)} v_A \| \right) = -L(A).
\]

(2) There are a constant \( c = c(A) > 0 \) and a neighborhood \( \mathcal{N}_A \) of \( A \) such that if \( B \in \mathcal{N}_A \) is a quasi-irreducible cocycle with \( L(B) > 0 \), then
\[
N(B) \geq c \log \left( \frac{d(B, \text{Diag})}{L(B)} \right).
\]
In particular, \( N(B) \geq c \log (d(B, A)^{-1}) \to 0 \) as the neighborhood \( \mathcal{N}_A \) shrinks.

Proof. Following the same approach used in the proof of Theorem 2.1, there are unit vectors \( u^+, u^- \) and there is a locally constant function \( \theta: X \to \mathbb{R} \setminus \{0\} \) such that for all \( x \in X \),
\[
A(x) u^+ = \theta(x) u^+ \quad \text{and} \quad A(x) u^- = \theta(x)^{-1} u^-.
\]

Then for all \( n \geq 1 \),
\[
\frac{1}{n} \log \| A^{(n)}(x) u^+ \| = \frac{1}{n} \sum_{j=0}^{n-1} \log |\theta(T^j x)| \quad \text{and}
\]
\[
\frac{1}{n} \log \| A^{(n)}(x) u^- \| = -\frac{1}{n} \sum_{j=0}^{n-1} \log |\theta(T^j x)|.
\]

By the Oseledets theorem and the law of large numbers, as \( n \to \infty \), one of the two quantities above converges to \( L(A) = |\mathbb{E} (\log |\theta|)| \) and the other to \(-L(A)\). With no loss of generality, we may assume that it is the former quantity that converges to \(-L(A)\), and so put \( v_A := u^+ \).

By the invariance of the measure, for all \( n \geq 1 \),
\[
\mathbb{E} \left( \frac{1}{n} \sum_{j=0}^{n-1} \log |\theta \circ T^j| \right) = \mathbb{E} (\log |\theta|).
\]

We conclude that for all \( n \geq 1 \) we have
\[
\mathbb{E} \left( \frac{1}{n} \log \| A^{(n)} v_A \| \right) = \mathbb{E} (\log |\theta|) = -L(A),
\]
which completes the proof of the first item.

To prove the second item, we fix a neighborhood \( \mathcal{N}_A \) of \( A \) such that, by Remark 2.2, we have that \( L(D) > \frac{1}{2} L(A) \) for all diagonalizable cocycles \( D \in \mathcal{N}_A \). Let \( 1 < L < \infty \) be such that for all \( B \in \mathcal{N}_A \),
\[
\max_{x \in X} \max \left\{ \|B(x)\|, \|B(x)^{-1}\| \right\} < L
\]
and put \( K := \log L \).
We decrease, if necessary, the neighborhood $N_A$, to ensure that, if $\delta$ denotes its diameter, then $\delta < (L(A)/2)^2$.

Fix a quasi-irreducible cocycle $B \in N_A$, with $L(B) > 0$. Let $D$ be any diagonalizable cocycle in $N_A$.

Then $\mathbb{P}$-almost surely and for every unit vector $v$ and for all $n \geq 1$, by the mean value theorem we have

$$
\left| \frac{1}{n} \log \| B^{(n)}(x) v \| - \frac{1}{n} \log \| D^{(n)}(x) v \| \right| \leq \frac{1}{n} \min \{ \| B^{(n)}(x) v \|, \| D^{(n)}(x) v \| \}.
$$

Moreover,

$$
\| B^{(n)}(x) v - D^{(n)}(x) v \| \\
\leq \sum_{j=0}^{n-1} \| B(T^{n-1}x) \| \ldots \| B(T^jx) - D(T^jx) \| \ldots \| D(x) \| \\
< n L^n d(B, D),
$$

while

$$
\| B^{(n)}(x) v \| \geq \| B^{(n)}(x)^{-1} \|^{-1} \geq L^{-n} \quad \text{and} \quad \| D^{(n)}(x) v \| \geq \| D^{(n)}(x)^{-1} \|^{-1} \geq L^{-n}.
$$

Therefore,

$$
\left| \frac{1}{n} \log \| B^{(n)}(x) v \| - \frac{1}{n} \log \| D^{(n)}(x) v \| \right| \leq L^{2n} d(B, D) = e^{2Kn} d(B, D).
$$

Choosing $v = v_D$, taking expectations and using item 1, we have that for all $n \geq 1$,

$$
\mathbb{E} \left( \frac{1}{n} \| B^{(n)} v_D \| \right) \leq \mathbb{E} \left( \frac{1}{n} \| D^{(n)} v_D \| \right) + e^{2Kn} d(B, D) \\
= -L(D) + e^{2Kn} d(B, D) \leq 0,
$$

provided

$$
e^{2Kn} < d(B, D)^{-1/2} < \frac{L(A)}{2} d(B, D)^{-1}.
$$

Therefore, we conclude that for all integers $n \leq \frac{1}{4K} \log d(B, D)^{-1},$

$$
\mathbb{E} \left( \frac{1}{n} \| B^{(n)} v_D \| \right) \leq 0 < \frac{L(B)}{2}.
$$

This shows that $N(B) > \frac{1}{4K} \log d(B, D)^{-1}$. But since $D$ was an arbitrary diagonalizable cocycle in a neighborhood of $A$, we conclude:

$$
N(B) \geq \frac{1}{4K} \log \left( d(B, \text{Diag})^{-1} \right),
$$

which completes the proof. $\square$
Given $A = (A_1, \ldots, A_k) \in \mathcal{D}iag^*$, we denote by $\hat{e}_\pm(A) \in \mathbb{P}$ the points corresponding to the invariant lines where the Lyapunov exponent of $A$ is positive, respectively negative. Moreover, $e_\pm(A)$ will stand for unit vectors representing these projective points $\hat{e}_\pm(A)$. Because $A \in \mathcal{D}iag$ there is a family of non-zero numbers $\{a_j: j = 1, \ldots, k\}$ such that $A_j e_+(A) = a_j e_+(A)$ and $A_j e_-(A) = a_j^{-1} e_-(A)$, for all $j = 1, \ldots, k$.

Then we define

$$\Sigma_H(A) := \{ i \in \{1, \ldots, k\} : |a_i| \geq e^{L(A)} \}.$$  

This set is non-empty because $L(A) = \sum_{j=1}^k p_j \log |a_j| > 0$.

Likewise, given a hyperbolic matrix $g \in \text{SL}_2(\mathbb{R})$, we denote by $\hat{e}_\pm(g) \in \mathbb{P}$ the eigen-directions of $g$ associated to eigenvalues of $g$ with absolute value larger than 1, respectively less than 1.

Given $A \in \mathcal{D}iag^*$, we define the projective cones$^{10}$

$$\mathcal{D}_\pm(\delta) = \mathcal{D}_\pm(A, \delta) := \{ \hat{x} \in \mathbb{P}: d(\hat{x}, \hat{e}_\pm(A)) < \delta d(\hat{x}, \hat{e}_\mp(A)) \}$$  \hspace{1cm} (3.3)

where the radius $\delta = \delta(A) > 0$ is chosen sufficiently small so that

$$\max_{1 \leq i \leq k} \sup_{v, v' \in \mathcal{D}_\pm(2\delta)} \left| \log \|A_i v\| - \log \|A_i v'\| \right| < \frac{1}{20} L(A).$$  \hspace{1cm} (3.4)

Take $L < \infty$ such that $\max\{ \|A\|_\infty, \|A^{-1}\|_\infty \} < L$. Next we fix a neighborhood $\mathcal{N}_A := \{B \in C: \|A - B\|_\infty < \varepsilon\}$ of $A$ in the space of cocycles $C = \text{SL}_2(\mathbb{R})^k$ such that for all $B \in \mathcal{N}_A$:

N0:  \hspace{0.5cm} (a) $d(A, B) < \frac{\delta}{L} e^{-\frac{L(A)}{4} \left( e^{-\frac{L(A)}{12}} - 1 \right)}$,

(b) $d(A, B) < \frac{L^{-1} \sqrt{e^{-\frac{L(A)}{12}} - 1}}{1 + \sqrt{e^{-\frac{L(A)}{12}} - 1}}$,

(c) $d(A, B) < \delta L^{-1} e^{-\frac{L(A)}{8} \left( 1 - e^{-\frac{L(A)}{24}} \right)}$.

N1: For all $i \in \Sigma_H(A)$, the matrix $B_i$ is hyperbolic. This holds because $B_i \approx A_i$ for all $i$ and $A_i$ is hyperbolic for $i \in \Sigma_H(A)$. In particular the directions $\hat{e}_\pm(B_i)$ are well-defined for all $B \in \mathcal{N}_A$ and $i \in \Sigma_H(A)$.

N2: $\frac{4}{3} L(A) > L(B)$ for all $B \in \mathcal{N}_A$. This holds because the LE is an upper-semicontinuous function.

N3: $\|B\|_\infty < L$ and $\|B^{-1}\|_\infty < L$.

N4: For all $i \in \Sigma_H(A)$, $d(\hat{e}_+(B_i), \hat{e}_-(B_i)) > \frac{1}{2} d(\hat{e}_+(A), \hat{e}_-(A)) =: c$.

$^{10}$ With this definition, $\mathcal{D}_\pm(\delta) = \mathcal{D}_\pm(\delta^{-1})^C$, for all $\delta > 0$. 
N5: \( \max \sup_{1 \leq i, j \leq k; \psi, \psi' \in \Delta_{+}(A, 2\delta)} |\log \|A_{i}v\| - \log \|B_{i}v'\| | < \frac{1}{12} L(A) \). This holds provided \( \|A - B\|_{\infty} \) is smaller than a certain fraction of \( L(A) \), because of (3.4) and the following inequality

\[
|\log \|A_{i}v'\| - \log \|B_{i}v'\| | \leq L \|A_{i} - B_{i}\| \||v'||,
\]

which follows from N3.

Other constraints on \( N_{A} \) will be imposed below. The statements of this section and the next refer to the neighborhood \( N_{A} \).

Given \( g \in \text{SL}_{2}(\mathbb{R}) \) we denote by \( \hat{g} : \mathbb{P} \to \mathbb{P} \) the projective action of this matrix, \( \hat{g}x := \hat{g}x \).

Then for all \( B \in N_{A} \) we define

\[
\rho^{+}(B) := \max_{i \in \Sigma_{H}(A)} \max_{1 \leq j \leq k} d(\hat{B}_{j}\hat{e}_{+}(B_{i}), \hat{e}_{+}(B_{i}))
\]

\[
\rho^{-}(B) := \max_{i \in \Sigma_{H}(A)} \max_{1 \leq j \leq k} d(\hat{B}_{j}\hat{e}_{-}(B_{i}), \hat{e}_{-}(B_{i}))
\]

\[
\rho(B) := \max\{\rho^{-}(B), \rho^{+}(B)\}.
\]

**Lemma 3.4.** For all \( B \in N_{A} \), \( \rho(B) = 0 \) if and only if \( B \) is diagonalizable. Moreover, \( \rho^{+}(B^{-1}) \gg \rho^{-}(B) \) and \( \rho^{-}(B^{-1}) \gg \rho^{+}(B) \).

**Proof.** If \( \rho(B) = 0 \) then for some \( i_{0} \in \Sigma_{H}(A) \) and all \( j = 1, \ldots, k \), \( \hat{B}_{j}\hat{e}_{+}(B_{i_{0}}) = \hat{e}_{+}(B_{i_{0}}) \) and \( \hat{B}_{j}\hat{e}_{-}(B_{i_{0}}) = \hat{e}_{-}(B_{i_{0}}) \), which implies that \( B \) is diagonalizable. Conversely, if the matrices \( B_{j} \) are simultaneously diagonalizable there exists a pair of distinct projective points \( \hat{e}_{1} \) and \( \hat{e}_{2} \) which are fixed under the action of all the matrices \( B_{j} \). Taking \( i_{0} \in \Sigma_{H}(A) \), since the matrix \( B_{i_{0}} \) is hyperbolic these two points must coincide with \( \hat{e}_{+}(B_{i_{0}}) \) and \( \hat{e}_{-}(B_{i_{0}}) \). Hence \( \rho(B) = \rho^{-}(B) = \rho^{+}(B) = 0 \).

By N3, \( \|B_{j}\|^{2} \|B_{j}^{-1}\|^{2} \leq L^{4} \) for all \( B \in N_{A} \), and hence by Remark 4.1 (see also Lemma 2.11 in [6]) for every projective point \( \hat{x} \)

\[
\frac{1}{L^{4}} \leq \frac{d(\hat{B}_{j}\hat{x}, \hat{x})}{d(\hat{x}, B_{j}^{-1}\hat{x})} \leq L^{4}.
\]

Applying these inequalities to the points \( \hat{x} = \hat{e}_{\pm}(B_{i}) \) with \( i \in \Sigma_{H}(A) \) we get \( \rho^{+}(B^{-1}) \leq L^{4} \rho_{-}(B) \), \( \rho^{-}(B) \leq L^{4} \rho_{+}(B^{-1}) \), \( \rho^{+}(B) \leq L^{4} \rho_{-}(B^{-1}) \) and \( \rho^{-}(B^{-1}) \leq L^{4} \rho^{+}(B) \), which completes the proof. \( \square \)

**Proposition 3.5.** Given \( A \in \text{Diag} \) with \( L(A) > 0 \), there exists a constant \( C = C(A) < \infty \) such that for all \( B \in N_{A} \),

\[
d(B, \text{Diag}) \leq C \rho(B) \quad \text{and} \quad \rho(B) \leq C d(B, A).
\]

In particular, for every \( B \in N_{A} \),

either \( d(B, \text{Diag}) \leq C \rho^{-}(B) \) or \( d(B, \text{Diag}) \leq C \rho_{-}(B^{-1}) \).
Proof. Let $L$ and $c$ be the constants defined in assumptions N3-N4 and take the constant $C = C(L, c)$ provided by Proposition 3.6 below. For all $B \in \mathcal{N}_A$ and $i \in \Sigma_H(A)$, $d(\hat{e}_+(B_i), \hat{e}_-(B_i)) > c$ and $\|B\|_\infty < L$. Choose the sign $\epsilon = \pm$ so that $\rho(B) = \rho_\epsilon(B)$, and take the maximizing index $i_0 \in \Sigma_H(A)$ such that for some $j = 1, \ldots, k$,

$$\rho(B) = d(\hat{B}_j \hat{e}_\epsilon(B_{i_0}), \hat{e}_\epsilon(B_{i_0})).$$

Let $\hat{p}^+ := \hat{e}_+(B_{i_0})$ and $\hat{p}^- := \hat{e}_-(B_{i_0})$. By Proposition 3.6 for every $i = 1, \ldots, k$ there exists $B^*_i \in \text{SL}_2(\mathbb{R})$ such that $\hat{B}^*_i \hat{p}^+ = \hat{p}^+$, $\hat{B}^*_i \hat{p}^- = \hat{p}^-$ and $\|B^*_i - B_i\| \leq C \rho(B)$. Hence $B^* = (B^*_1, \ldots, B^*_k) \in \mathcal{C} = \text{SL}_2(\mathbb{R})^k$ is a diagonalizable cocycle such that

$$d(B, \text{Diag}) \leq \|B - B^*\|_\infty \leq C \rho(B).$$

Finally we need to prove that $\rho(B) \lesssim d(A, B)$. For $i \in \Sigma_H$, since $A_i$ is hyperbolic its eigenvalues are simple and the function $B_i \mapsto \hat{e}_\pm(B_i)$ is analytic on a neighborhood of $A_i$. Hence there exists $C = C(A) < \infty$ such that $d(\hat{e}_\pm(B_i), \hat{e}_\pm(A_i)) \leq C d(A, B)$, for all $i \in \Sigma_H$ and $B \in \mathcal{N}_A$. Using that the projective action of any matrix $A_j$ has Lipschitz constant $\text{Lip}(\hat{A}_j) \leq \|A_j\|^2$, we obtain for all $i \in \Sigma_H$ and $1 \leq j \leq k$,

$$d(\hat{B}_j \hat{e}_\pm(B_i), \hat{e}_\pm(B_i)) \leq d(\hat{B}_j \hat{e}_\pm(B_i), \hat{A}_j \hat{e}_\pm(B_i)) + d(\hat{A}_j \hat{e}_\pm(B_i), \hat{e}_\pm(A_i))$$

$$+ d(\hat{e}_\pm(A_i), \hat{e}_\pm(B_i))$$

$$\leq \|B_j e_\pm(B_i) - A_j e_\pm(B_i)\| + d(\hat{A}_j \hat{e}_\pm(B_i), \hat{A}_j \hat{e}_\pm(A_i))$$

$$+ d(\hat{e}_\pm(A_i), \hat{e}_\pm(B_i))$$

$$\leq \|B_j - A_j\| + (\|A_j\|^2 + 1) d(\hat{e}_\pm(A_i), \hat{e}_\pm(B_i))$$

$$\leq (1 + (L^2 + 1)C) d(A, B).$$

Taking the maximum in $i \in \Sigma_H$ and $1 \leq j \leq k$, this inequality implies that $\rho(B) \lesssim d(A, B)$. \hfill \Box

**Proposition 3.6.** Given $c > 0$ and $1 \leq L < \infty$ there exists $C = C(L, c) < \infty$ with the following property:

For any $g \in \text{SL}_2(\mathbb{R})$ and any two points $\hat{p}^+ \in \mathbb{P}$ such that $\|g\| \leq L$ and $d(\hat{p}^+, \hat{p}^-) \geq c$ there exists $g^* \in \text{SL}_2(\mathbb{R})$ such that $\hat{g}^* \hat{p}^+ = \hat{p}^+$, $\hat{g}^* \hat{p}^- = \hat{p}^-$ and

$$\|g^* - g\| \leq C \max\{d(\hat{g} \hat{p}^+, \hat{p}^+), d(\hat{g} \hat{p}^-, \hat{p}^-)\}.$$ 

**Proof.** Take unit vectors $p^\pm \in \hat{p}^\pm$ such that $\angle(p^+, p^-)$ is not obtuse. Let $m \in \text{GL}_2(\mathbb{R})$ be the matrix with columns $gp^+ / \|gp^+\|$ and $gp^- / \|gp^-\|$. Similarly, let $m' \in \text{GL}_2(\mathbb{R})$ be the matrix with columns $p^+ \hat{p}^-$ and $\hat{p}^- \hat{p}^-$. Consider the matrix $g' = m'm^{-1}g \in \text{GL}_2(\mathbb{R})$. By construction this matrix satisfies $\hat{g}' \hat{p}^+ = \hat{p}^+$ and $\hat{g}' \hat{p}^- = \hat{p}^-$. \hfill \Box
Next notice that
\[ \| g' - g \| = \| (m' - m) m^{-1} g \| = \| m' - m \| \| m^{-1} \| \| g \|. \]
Since \( \| g \| \leq L \), we now need to derive bounds for \( \| m' - m \| \) and \( \| m^{-1} \| \).
For any matrix \( a \in \text{Mat}_2(\mathbb{R}) \) and in particular for \( a = m' - m \),
\[ \| a \| \leq 2 \max \{ \| a e_1 \|, \| a e_2 \| \} \]
where \( \{ e_1, e_2 \} \) stands for the canonical basis of \( \mathbb{R}^2 \). On the other hand, given \( \hat{x}, \hat{y} \in \mathbb{P} \), if we choose the unit vectors \( x \in \hat{x} \) and \( y \in \hat{y} \) to form a non obtuse angle then
\[ \| x - y \| \leq \sqrt{2} d(\hat{x}, \hat{y}). \]
Combining these two facts we get
\[ \| m' - m \| \leq 2 \sqrt{2} \max \{ d(\hat{g} \hat{p}^+, \hat{p}^+), d(\hat{g} \hat{p}^-, \hat{p}^-) \}. \]
For any matrix \( g \in \text{SL}_2(\mathbb{R}) \), the projective map \( \hat{g} : \mathbb{P} \to \mathbb{P} \) has Lipschitz constant \( \text{Lip}(\hat{g}) \leq \| g \|^2 \) and the same is true about \( g^{-1} \) because \( \| g^{-1} \| = \| g \| \). Hence, since in our case \( \| g \| \leq L \),
\[ |\det m| = \frac{\| gp^+ \wedge gp^- \|}{\| gp^+ \| \| gp^- \|} = d(\hat{g} \hat{p}^+, \hat{g} \hat{p}^-) \geq \frac{1}{L^2} d(\hat{p}^+, \hat{p}^-) \geq \frac{c}{L^2}. \]
Therefore
\[ \| m^{-1} \| \leq \frac{\| m \|}{|\det m|} \leq \frac{2}{|\det m|} \leq \frac{2 L^2}{c}. \]
Together these bounds imply that
\[ \| g' - g \| \leq 4 \sqrt{2} L^3 c^{-1} \max \{ d(\hat{g} \hat{p}^+, \hat{p}^+), d(\hat{g} \hat{p}^-, \hat{p}^-) \}. \]
Finally, by Lemma 3.7 below we can replace \( g' \in \text{GL}_2(\mathbb{R}) \) by \( g^* := g'_e \in \text{SL}_2(\mathbb{R}) \) and still have
\[ \| g^* - g \| \leq \max \{ d(\hat{g} \hat{p}^+, \hat{p}^+), d(\hat{g} \hat{p}^-, \hat{p}^-) \}. \]
We finish explaining why this lemma can be applied. First notice that the previous estimate on \( |\det m| \) gives
\[ \frac{c}{L^2} \leq |\det m| = d(\hat{g} \hat{p}^-, \hat{g} \hat{p}^-) \leq 1. \]
Constructing a smooth deformation \( m_t \) from \( m_0 = m' \) to \( m_1 = m \) by matrices with unitary columns we have
\[ \| m_t \| \leq 2 \quad \text{and} \quad \frac{c}{L^2} \leq |\det m_t| \leq 1. \]
Hence \( g_t := m'_m^{-1} m g \) is a smooth deformation from \( g \) to \( g' \) such that \( \| g_t \| \leq 4 L^3 c^{-1} \) and
\[ |\det g_t| = \frac{|\det m'|}{|\det m_t|} \geq d(\hat{p}^+, \hat{p}^-) \geq c. \]
This shows that Lemma 3.7 can be applied with an appropriate new constant $\tilde{L} = \tilde{L}(L, c) < \infty$. 

\textbf{Lemma 3.7.} Given $1 \leq L < \infty$ there exist $C = C(L) < \infty$ such that for all $g \in \text{SL}_2(\mathbb{R})$ with $\|g\| < L$ and all $g' \in \text{GL}_2(\mathbb{R})$ with $\|g'\| < L$, $|\det g'| \geq L^{-1}$ and such that these bounds for $g'$ hold along a smooth deformation from $g$ to $g'$, then $g_* := |\det g'|^{-\frac{1}{2}} g' \in \text{SL}_2(\mathbb{R})$ satisfies

$$\|g - g_*\| \leq C \|g - g'\|.$$ 

\textit{Proof.} By Jacobi’s formula for the derivative of the determinant,

$$(D \det)_g(h) = \text{tr}(h \, g^{-1}) \det g, \quad h \in \text{Mat}_2(\mathbb{R}).$$

Hence, for all $g \in \text{GL}_2(\mathbb{R})$, $\|(D \det)_g\| \leq 2 \|g^{-1}\| \|\det g\| = 2 \|g\|$. Using the assumptions and the Mean Value Theorem we get

$$\|\det g|^{-1/2} - |\det g'|^{-1/2}\| \leq L^{5/2} \|g - g'\|.$$ 

Therefore

$$\|g - g_*\| \leq \|g - |\det g'|^{-\frac{1}{2}} g\| + |\det g'|^{-\frac{1}{2}} \|g - g'\|$$

$$\leq \|\det g|^{-\frac{1}{2}} - |\det g'|^{-\frac{1}{2}}\| \|g\| + |\det g'|^{-\frac{1}{2}} \|g - g'\|$$

$$\leq (L^{5/2} L + L^{1/2}) \|g - g'\|$$

which proves this lemma with $C(L) := L^{7/2} + L^{1/2}$. \hfill \Box

The next proposition establishes a relation between the quasi-irreducibility measurements $\rho(B)$ and $N(B)$. We postpone its proof to the next section.

\textbf{Proposition 3.8.} Given $A \in \text{Diag}$ with $L(A) > 0$ there is a constant $c = c(A) > 0$ such that if the neighborhood $\mathcal{N}_A$ of $A$ is small enough then for all $B \in \mathcal{N}_A$,

$$\rho_-(B) \leq e^{-c N(B)} \quad \text{and} \quad \rho_+(B) \leq e^{-c N(B^{-1})}.$$ 

In other words,

$$N(B) \leq c^{-1} \log(\rho_-(B)^{-1}) \quad \text{and} \quad N(B^{-1}) \leq c^{-1} \log(\rho_+(B)^{-1}).$$

\textbf{Remark 3.1.} Combining Propositions 3.3, 3.5 and 3.8 we conclude that all three irreducibility measurements introduced in this section, namely $N(B)$, $\rho(B)$ and $d(B, \text{Diag})$ are essentially equivalent.

More precisely, let us put $N(A) := \infty$ if $A$ is diagonalizable (which, by Proposition 3.3, is a reasonable convention). Moreover, for any cocycle $B$, define $N(B) := \max\{N(B), N(B^{-1})\}$. 
Then for any diagonalizable cocycle $A$ with $L(A) > 0$, there are constants $c_1, c_2, c_3 > 0$ and a neighborhood $N_A$ of $A$ such that for all $B \in N_A$, the following hold:

$$c_1 \bar{N}(B) \leq \log \rho(B)^{-1} \leq c_2 \log d(B, \text{Diag})^{-1} \leq c_3 \bar{N}(B).$$

4. Prison break

Any cocycle $A \in \mathcal{C} = \text{SL}_2(\mathbb{R})^k$ induces a random walk on the projective space. If the cocycle $A$ is diagonalizable with $L(A) > 0$ then the random walk starting at a point $\hat{x} \neq \hat{e}_-(A)$ converges rapidly to $\hat{e}_+(A)$. This is not the case for $\hat{x} = \hat{e}_-(A)$ where the random walk gets permanently stuck. For a quasi-irreducible cocycle $B \in \mathcal{C}$ very close to $A$ the corresponding random walk may also wander for a while near the point $\hat{e}_-(A)$ after which it rapidly moves to somewhere near $\hat{e}_+(A)$. It is natural to ask for how long one needs to wait until it becomes likely to see the random walk moving in a neighborhood of $\hat{e}_+(A)$. The answer depends of course on the distance from $B$ to $A$. Proposition 4.12 makes a precise quantitative statement which answers this question.

The next subsection provides an abstract scheme for the proof of Proposition 4.12. We illustrate it with a prison break metaphor.

Imagine a prisoner whose movement in the world $\Sigma$ is modeled by a random walk $\xi_n$. The confinement constraints on the prisoner’s movements are encoded in the transition probabilities of this random walk. Let $\Sigma_0$ be the prisoner’s cell, $\Sigma_1$ be the prison area and $\Sigma_2$ be the state where the prisoner serves his sentence.

Assume the following about the prisoner.

1. The probability that he escapes from the cell $\Sigma_0$ within a time $n_0$ is small but positive.
2. He has a large probability of evading the prison $\Sigma_1$ within a time $n_1$ once he is outside the cell.
3. The probability of him fleeing the state $\Sigma_2$ within a time $n_2$ is again large once he is out of prison.
4. The chances of him ever being caught again in the prison $\Sigma_1$ are very slim after he gets abroad.

From these assumptions one concludes that after a long enough time, of the form $qn_0 + n_1 + n_2$ for some possibly large $q \in \mathbb{N}$, the prisoner will very likely stay permanently out of the jail $\Sigma_1$.

The next subsection quantifies this statement in Proposition 4.3.
Abstract scheme. Let $K$ be a stochastic kernel on a compact space of symbols $\Sigma$ and denote by $\mathcal{F}$ the Borel $\sigma$-algebra of $\Sigma$. Let $\text{Prob}(\Sigma)$ denote the space of probability measures on $(\Sigma, \mathcal{F})$. Let $\{\xi_n\}_{n \geq 0}$ be the process on $X = \Sigma^\mathbb{N}$ defined by $\xi_n\{x_j\}_{j \in \mathbb{N}} := x_n$.

Given $\nu \in \text{Prob}(\Sigma)$ there exists a unique probability measure $P_\nu$ on $X = \Sigma^\mathbb{N}$ such that $\{\xi_n\}_{n \geq 0}$ is a Markov process with transition kernel $K$ and initial distribution $\nu$, in the sense that for all $E \in \mathcal{F}$ and $x \in \Sigma$

1. $P_\nu[\xi_n \in E | \xi_{n-1} = x] = K_x(E)$ \ (n \geq 1),
2. $P_\nu[\xi_0 \in E] = \nu(E)$.

To emphasize the measure $P_\nu$ underlying the process $\xi_n$, we will write $\xi_\nu^n$, or simply $\xi_\nu^x$ when $\nu = \delta_x$, instead of $\xi^n$.

Given $x \in \Sigma$ and $E \in \mathcal{F}$ with $x \in E$ define the probability of escaping from $E$ in time $n$

$$P_n(x, E) := \mathbb{P}_x[ \text{ for some } 0 \leq j \leq n, \xi_j^x \notin E]$$

as well as the complementary probability of remaining in $E$ for time $n$

$$P_n^*(x, E) := \mathbb{P}_x[ \text{ for all } 0 \leq j \leq n, \xi_j^x \in E].$$

Note that

$$P_n(x, E) + P_n^*(x, E) = 1.$$ 

Proposition 4.1. Given $x_0 \in E \subset F \subset \Sigma$ in $\mathcal{F}$,

$$P_{n+m}(x_0, F) \geq P_m(x_0, E) \left( \inf_{x \in \Sigma \setminus E} P_n(x, F) \right).$$

Proof. Define a stopping time $\tilde{j}: X \to \mathbb{N} \cup \{\infty\},$

$$\tilde{j} := \min\{l \geq 0: \xi_l^{x_0} \notin E\}$$

together with a random variable $\tilde{x}: X \to \Sigma$ such that $\tilde{x} = \xi_l^{x_0}$ whenever $\tilde{j} < \infty$. By construction $\tilde{x}$ takes values in $\Sigma \setminus E$ if $\tilde{j} < \infty$.

$$P_{n+m}(x_0, F) \geq \mathbb{P}_{x_0}[\tilde{j} \leq m \text{ and for some } 0 \leq j \leq n, \xi_j^{\tilde{x}} \notin F]$$

$$= \sum_{l=0}^{m} \mathbb{P}_{x_0}[ \text{ for some } 0 \leq j \leq n, \xi_j^{\tilde{x}} \notin F | \tilde{j} = l] \mathbb{P}_{x_0}[\tilde{j} = l]$$

$$\geq \left( \inf_{x \in \Sigma \setminus E} P_n(x, F) \right) \sum_{l=0}^{m} \mathbb{P}_{x_0}[\tilde{j} = l]$$

$$= \left( \inf_{x \in \Sigma \setminus E} P_n(x, F) \right) P_m(x_0, E).$$
Proposition 4.2. \( \mathcal{P}_n^*(E) \leq \mathcal{P}_n^*(E) \mathcal{P}_m^*(E) \) for all \( n, m \geq 0 \).

Proof. Let \( X_n = \Sigma^{n+1} \) and denote by \( \pi_n : X \rightarrow X_n \) the canonical projection \( \pi_n \{x_j\}_{j \in \mathbb{N}} = (x_0, x_1, \ldots, x_n) \). Let \( \mathbb{P}_x^{(n)} := (\pi_n)_* \mathbb{P}_x \) and note that \( \mathbb{P}_x^{(n)}(\{x\} \times E^n) = \mathcal{P}_n^*(x, E) \).

Define the family of events

\[
E_{n,m}^x := [\xi_j^x \in E, \quad \forall n + 1 \leq j \leq n + m].
\]

Using the law of total probability and then the fact that \( \{\xi_n^x\} \) is a Markov chain, we obtain the following:

\[
\mathcal{P}_n^*(x, E) = \mathbb{P}_x \left[ \text{for all } 0 \leq j \leq n + m \text{, } \xi_j^x \in E \right]
= \int_{\{x\} \times E^n} \mathbb{P}_x \left[ E_{n,m}^x \mid \xi_j^x = x_j, \forall 0 \leq j \leq n \right] d\mathbb{P}_x^{(n)}(x_0, x_1, \ldots, x_n)
= \int_{\{x\} \times E^n} \mathbb{P}_x \left[ E_{n,m}^x \mid \xi_n^x = x_n \right] d\mathbb{P}_x^{(n)}(x_0, x_1, \ldots, x_n).
\]

But given any \( y \in E \), since \( \{\xi_n^x\} \) is a Markov chain,

\[
\mathbb{P}_x \left[ E_{n,m}^x \mid \xi_n^x = y \right] = \mathbb{P}_x \left[ \xi_{n+1}^x \in E, \ldots, \xi_{n+m}^x \in E \mid \xi_n^x = y \right] = \mathbb{P}_y \left[ \xi_1^y \in E, \ldots, \xi_m^y \in E \right] = \mathcal{P}_m^*(y, E) \leq \mathcal{P}_m^*(E).
\]

Combining this with the previous identity we have:

\[
\mathcal{P}_n^*(x, E) \leq \mathcal{P}_m^*(E) \mathcal{P}_x^{(n)}(\{x\} \times E^n) = \mathcal{P}_m^*(E) \mathcal{P}_n^*(x, E),
\]
and the conclusion follows by taking the supremum over \( x \in E \). \( \square \)

Consider now the prisoner’s context: three measurable sets \( \Sigma_0 \subset \Sigma_1 \subset \Sigma_2 \subset \mathcal{F}, \) the cell, the prison and the state, respectively.

Proposition 4.3. Given \( r > 0 \) assume that:

(A1) \( \mathcal{P}_n^*(x, \Sigma_0) \geq b_0 > 0 \) for all \( x \in \Sigma_0, \)
(A2) \( \mathcal{P}_n^*(x, \Sigma_1) \geq 1 - r/4 \) for all \( x \in \Sigma_1 \setminus \Sigma_0, \)
(A3) \( \mathcal{P}_n^*(x, \Sigma_2) \geq 1 - r/4 \) for all \( x \in \Sigma_2 \setminus \Sigma_1, \)
(A4) \( \mathcal{P}_n^*(x, \Sigma_2^c) \geq 1 - r/4 \) for all \( x \in \Sigma_2^c \) and \( n \geq 0. \)

If \( (1 - b_0)^q < r/4 \) for some \( q \in \mathbb{N} \) then setting \( N := q n_0 + n_1 + n_2, \) one has that for all \( x \in \Sigma, \)

\[
\mathbb{P}_x \left[ \exists j \geq N, \xi_j^x \in \Sigma_1 \right] < r.
\]
Proof. From assumption (A1) we get \( P^*_{n_0}(x, \Sigma_0) < 1 - b_0 \) for all \( x \in \Sigma_0 \). Taking the sup this implies that \( P^*_{n_0}(\Sigma_0) < 1 - b_0 \). By Proposition 4.2
\[
P^*_{n_0}(\Sigma_0) < (1 - b_0)^q < r/4.
\]
Thus \( P_{qn_0}(x, \Sigma_0) \geq 1 - r/4 \) for all \( x \in \Sigma_0 \).
Combining this fact with assumptions (A2)-(A3) and applying Proposition 4.1 we get that for all \( x \in \Sigma_2 \),
\[
P_{N}(x, \Sigma_2) \geq (1 - r/4)^3. \tag{4.1}
\]
Finally, the claim in the proposition reduces to showing that
\[
P \left[ \xi^x_{j_{0}} \in \Sigma^C_1, \ \forall \ j \geq N \right] \geq (1 - r/4)^4 > 1 - r
\]
for all \( x_0 \in \Sigma \).
If \( x_0 \in \Sigma^C_2 \) this follows from (A4) because \( 1 - r/4 > 1 - r \).
Otherwise, let \( x_0 \in \Sigma_2 \) and consider the stopping time \( \bar{j} : X \rightarrow \mathbb{N} \) defined by
\[
\bar{j} := \min \{l \geq 0 : \xi^x_{i} \notin \Sigma_2\}.
\]
Define also the random variable \( \bar{x} : X \rightarrow \Sigma, \ \bar{x} := \xi^x_{i} \). By construction
\( \bar{x} \) takes values in \( \Sigma^C_2 \). Then, using assumption (A4) and (4.1),
\[
P \left[ \xi^x_{j_{0}} \in \Sigma^C_1, \ \forall \ j \geq N \right] \geq P_{x_0} \left[ \bar{j} \leq N \ \text{and for all} \ j \geq 0, \ \bar{x} \in \Sigma^C_1 \right]
\]
\[
= \sum_{l=1}^{N} P_{x_0} \left[ \text{for all} \ j \geq 0, \ \bar{x} \in \Sigma^C_1 | \bar{j} = l \right] P_{x_0}[\bar{j} = l]
\]
\[
\geq \left( \inf_{l \in \mathbb{N}} \inf_{x \in \Sigma^C_2} P^*_{l}(x, \Sigma^C_1) \right) \sum_{l=1}^{N} P_{x_0}[\bar{j} = l]
\]
\[
= \left( \inf_{l \in \mathbb{N}} \inf_{x \in \Sigma^C_2} P^*_{l}(x, \Sigma^C_1) \right) P_{N}(x_0, \Sigma_2)
\]
\[
\geq (1 - r/4) (1 - r/4)^3 = (1 - r/4)^4 > 1 - r .
\]
□

The setting. A given probability vector \( p = (p_1, \ldots, p_k) \in \mathbb{R}^k_+ \) and a space of symbols \( \Sigma = \{1, \ldots, k\} \) determine the Bernoulli measure \( \mathbb{P}_p := p^\mathbb{Z} \) in the space of sequences \( X = \Sigma^\mathbb{Z} \), which is invariant under the full shift map \( T : X \rightarrow X \). Assume now that \( A = (A_1, \ldots, A_k) \in \mathcal{C} = \text{SL}_2(\mathbb{R})^k \) is a diagonalizable cocycle with \( L(A) > 0 \) and \( B = \)
$(B_1, \ldots, B_k) \in N_A$ is a nearby cocycle such that $\rho(B) = \rho_-(B) > 0$, which henceforth will be denoted by $\rho_B^{11}$.

Given $\hat{x} \in \mathbb{P}$, consider the random walk $\{ \xi_n^\hat{x} = \xi_n^\hat{x}(B) : X \to \mathbb{P} \}_{n \geq 0}$ defined by

$$\xi_n^\hat{x}\{\omega_j\}_{j \in \mathbb{Z}} := \hat{B}_{\omega_{n-1}} \ldots \hat{B}_{\omega_1} \hat{B}_{\omega_0} \hat{x}.$$

Let $C = C(A)$ be $\frac{10}{9}$ of the homonymous constant in Proposition 3.5. Take a diagonalizable cocycle $D = (D_1, \ldots, D_k) \in \text{Diag}$ which nearly minimizes the distance to $B$, say $d(B, D) < \frac{10}{9} d(B, \text{Diag})$. By Proposition 3.5, $d(B, D) \leq C \rho_B$ and $\rho_B \leq C d(A, B)$. We can assume that $d(B, D) < d(A, B)$ for otherwise we would take $D = A$. Hence

$$d(A, D) \leq d(A, B) + d(B, D) < 2 d(A, B). \quad (4.2)$$

Figure 1 illustrates the relative positions of $A$, $B$ and $D$ in $N_A$.

Up to a conjugation we may assume that $D$ is diagonal with

$$D_i = \begin{bmatrix} d_i & 0 \\ 0 & d_i^{-1} \end{bmatrix} \quad \text{for} \quad 1 \leq i \leq k$$

and so $\hat{e}_+(D) = (1 : 0)$ and $\hat{e}_-(D) = (0 : 1)$.

We will be using the following two projective coordinate systems $\psi_\pm : \mathbb{P} \to \mathbb{R} \cup \{ \infty \}$ defined respectively by $\psi_-(x : y) := x/y$ and $\psi_+(x : y) := y/x$. Since $D$ is diagonal we have the following. With respect to $\psi_+$, the projective point $\hat{e}_+(D)$ corresponds to 0, the projective point $\hat{e}_-(D)$ corresponds to $\infty$, while with respect to $\psi_-$, the roles are reversed. Moreover, $\psi_\pm(D_\pm(D, r)) = (-r, r)$, where the projective cones $D_\pm(D, r)$ were defined in (3.3).

Proposition 3.8 of the previous section, relating the measurements $N(B)$ and $\rho(B)$, will be proved through the abstract prison break argument encapsulated in Proposition 4.3. Before specifying the sets $\Sigma_0$, $\Sigma_1$ and $\Sigma_2$, and for clarity, we collect below a list of constants (depending only on $A$) and conditions on the neighborhood $N_A$ that will be needed in the argument (as well as to define the sets $\Sigma_i$). The following details may be skipped be at first reading.

Besides the LE $L(A)$ of $A$, let $\delta = \delta(A)$ be the size of the cones in assumption N5, let $L$ be the upper bound in assumption N3 and let $C = C(A)$ be the constant introduced above, which is $\frac{10}{9}$ of the homonymous constant in Proposition 3.5. Set the probability threshold

$$r = r(A) := \frac{L(A)}{42 L(A) + 60 \log L} \quad (4.3)$$

11The case where $\rho(B) = \rho_+(B) > 0$ reduces to the previous one applied to the inverse cocycle $B^{-1}$ and will not addressed here.
for the application of Proposition 4.3. Let
\[ M_0 := \frac{Ce^{-\frac{5}{3}L(A)}}{1 - e^{-\frac{5}{3}L(A)}} + \frac{C}{1 - e^{-\frac{5}{3}L(A)}} + 1, \]
(4.4)
the parameter in Lemma 4.6 below. For all 1 \( \leq j \leq k \) set
\[ \tilde{\lambda}_j := e^{2 \log |a_j| - \frac{L(A)}{3}}, \]
(4.5)
\[ \lambda^*_j := e^{2 \log |a_j| - \frac{L(A)}{2}}. \]
(4.6)
These families of numbers will be used to define two random walks for comparison with \( \xi_n^x(B) \), see lemmas 4.8 and 4.10 below. Next choose \( s = s(A) < \infty \) such that
\[ s > \max\left\{ 3 (\tilde{\lambda}_j - \lambda^*_j)^{-1}: 1 \leq j \leq k \right\}, \]
(4.7)
\[ s > \max\left\{ (1 + (\lambda^*_j)^{-1}) ((\lambda^*_j)^{-1} - \tilde{\lambda}_j^{-1})^{-1}: 1 \leq j \leq k \right\}. \]
(4.8)
A simple calculation gives
\[ \lambda_j - \lambda^*_j \leq L^2 e^{-\frac{L(A)}{6}} (1 - e^{-\frac{L(A)}{6}}), \]
\[ (\lambda^*_j)^{-1} - \tilde{\lambda}_j^{-1})^{-1} \leq \frac{L^2 e^{-\frac{L(A)}{2}}}{e^{-\frac{L(A)}{6}} - 1}. \]
and these bounds can be used to derive an explicit expression for \( s(A) \). This parameter \( s \) is used in lemmas 4.8 and 4.10. Define also
\[
\hat{c} = \hat{c}(A) := \frac{1}{18} \left( \frac{L(A)}{2 \log L + L(A)/2} \right)^2.
\] (4.9)

The parameter \( \hat{c} \) will be the rate of exponential decay in Hoeffding’s inequality for the random walk determined by the numbers \( \lambda_j^* \). Next let \( l_0 = l_0(A) \in \mathbb{N} \) be the smallest integer such that
\[
\frac{e^{-\hat{c} l_0}}{1 - e^{-\hat{c}}} < \frac{r}{8}.
\] (4.10)

and let
\[
\kappa = \kappa(A) := l_0 \max \{- \log \lambda_j^*: 1 \leq j \leq k\}.
\] (4.11)

These parameters, \( \hat{c} \) and \( \kappa \), appear in the proof of Proposition 4.7. Finally set
\[
M = M(A) := e^{\kappa(A)} s(A) M_0(A).
\] (4.12)

We now impose several assumptions that will further restrict the size of the neighborhood \( \mathcal{N}_A \). In all statements N6-N11, \( B \) is an arbitrary cocycle in \( \mathcal{N}_A \) while \( D \in \mathcal{D} \text{diag} \) is taken, as above, near \( B \). When we say that a certain assumption holds we mean that there exists a sufficiently small neighborhood \( \mathcal{N}_A \) of \( A \) such that all cocycles \( B \in \mathcal{N}_A \) satisfy that assumption.

N6: \( M (\rho_B)^{1/2} < \delta \). Holds because \( \rho_B \leq C d(A, B) \) (Proposition 3.5).

N7: \( d(B^{-1}, D^{-1}) \leq C \rho_B \) and \( L(D) > \frac{9}{10} L(A) \). The following inequality
\[
\|B^{-1} - D^{-1}\|_\infty \leq L^2 \|B - D\|_\infty
\]
shows that if we replace \( C \) by \( C L^2 \) then the first inequality holds without changing \( \mathcal{N}_A \). Because \( D \) and \( A \) are diagonalizable cocycles, in view of Remark 2.2 the second assumption on \( L(D) \) holds as well.

N8: \( |d_i| \geq e^{L(A)/2} \) for all \( i \in \Sigma_H(A) \), where \( D = (D_1, \ldots, D_k) \) and some coordinates are fixed where \( D_i = \begin{bmatrix} d_i & 0 \\ 0 & d_i^{-1} \end{bmatrix} \). This holds by (4.2) because \( |a_i| \geq e^{L(A)} \).

N9: \( \mathcal{D}_\pm(D, \delta) \subset \mathcal{D}_\pm(A, 2\delta) \). Holds because \( D \) is near \( A \), by (4.2).
Given \( i \in \Sigma_H(A) \), one has \( D_i^{\pm 1}\mathcal{D}_\pm(\delta) \subset \mathcal{D}_\pm(e^{-L(A)}\delta) \). For instance if \( \hat{x} \in \mathcal{D}_-(\delta) \), using N8,
\[
d(\hat{D}_i^{-1}\hat{x}, \hat{D}_i^{-1}\hat{e}_-(D)) = d(\hat{D}_i^{-1}\hat{x}, \hat{D}_i^{-1}\hat{e}_-(D)) = \frac{d(\hat{x}, \hat{e}_-(D))}{\|D_i^{-1}x\| |d_i|} \leq \delta \frac{d(\hat{x}, \hat{e}_+(D))}{\|D_i^{-1}x\| |d_i|} \leq \frac{\delta}{\|D_i^{-1}x\| |d_i|} \|D_i^{-1}x\| \|D_i^{-1}e_+(D)\| = \frac{\delta}{e^{L(A)}} d(\hat{D}_i^{-1}\hat{x}, \hat{e}_+(D))
\]
which implies that \( \hat{D}_i^{-1}\hat{x} \in \mathcal{D}_-(e^{-L(A)}\delta) \).

N10: \( \hat{B}_i^{\pm 1}\mathcal{D}_\pm(D, \delta) \subset \mathcal{D}_\pm(D, \delta) \) for all \( i \in \Sigma_H(A) \). This follows by continuity from the previous considerations.

N11: \( \log(\rho_B^{-1}) > \max\{2r, l_0 L(A)\} \). This holds by (4.2). Note that \( \rho_B \to 0 \) as the size of the neighborhood \( N_A \) decreases.

Next consider the (projective arc) sets w.r.t. the cocycle \( D \)
\[
\Sigma_0 := \mathcal{D}_-(M\rho_B), \quad \Sigma_1 := \mathcal{D}_-(\delta^{-1}) \quad \text{and} \quad \Sigma_2 := \mathcal{D}_-(\rho_B^{-1}).
\]
By N6, \( M\rho_B < \delta \) which implies that
\[
\Sigma_0 \subset \Sigma_1 \subset \Sigma_2 \subset \mathbb{P}.
\]
In the prisoner’s metaphor these sets are the ‘cell’, the ‘prison’ and the ‘state’. Throughout the rest of this section the projective cones \( \mathcal{D}_\pm(a) \) always refer to the diagonal cocycle \( D \).

Establishing Assumption (A1).

**Proposition 4.4.** There exist constants \( b_0 > 0 \) and \( n_0 \in \mathbb{N} \), depending on \( A \) and on the probability vector \( p \), such that for all \( \hat{x} \in \Sigma_0 \),
\[
\mathbb{P}_\hat{x} \left[ \exists j \leq n_0 \text{ such that } \xi_j^\hat{x}(B) \notin \Sigma_0 \right] \geq b_0.
\]

The proof of this proposition requires some lemmas.

**Remark 4.1.** Given \( g \in GL_2(\mathbb{R}) \) and unit vectors \( x, y \in \mathbb{R}^d \),
\[
d(\hat{g} \hat{x}, \hat{g} \hat{y}) = \frac{\|gx \land gy\|}{\|gx\| \|gy\|} = \frac{|\det g| \|x \land y\|}{\|gx\| \|gy\|} = \frac{|\det g|}{\|gx\| \|gy\|} d(\hat{x}, \hat{y}).
\]

**Lemma 4.5.** Given \( i \in \Sigma_H(A) \) for all \( \hat{x}, \hat{y} \in \mathcal{D}_\pm(\delta) \),
\[
d(B_i^{\pm 1}\hat{x}, B_i^{\pm 1}\hat{y}) \leq e^{-\frac{\delta}{2} L(A)} d(\hat{x}, \hat{y}).
\]
Proof. Given \( i \in \Sigma_H(A) \), by N5 and N9, for any \( \hat{v} \in \mathcal{D}_-(\delta) \),
\[
\|B^i_v\| \leq e^{-\log|a_i| + \frac{L(A)}{12}} < e^{-\frac{5}{3}L(A)}.
\]
Because of N10, we have for all \( \hat{v} \in \mathcal{D}_-(\delta) \),
\[
\|B_i^{-1}v\| > e^{\frac{5}{3}L(A)}\|B_iB_i^{-1}v\| \geq e^{\frac{5}{3}L(A)}
\]
which in turn implies that for all \( \hat{x}, \hat{y} \in \mathcal{D}_-(\delta) \)
\[
d(\hat{B}_i^{-1}\hat{x}, \hat{B}_i^{-1}\hat{y}) = \frac{d(\hat{x}, \hat{y})}{\|B_i^{-1}x\| \|B_i^{-1}y\|} \leq e^{-\frac{2}{3}L(A)} d(\hat{x}, \hat{y}).
\]
This proves one inequality. The other is analogous. \( \square \)

Lemma 4.6. For all \( 1 \leq j \leq k \),
\[
d(\hat{B}_j^{\pm 1} \hat{e}_\pm(D), \hat{e}_\pm(D)) \leq M_0 \rho_B.
\]

Proof. Because of N10, which applies to the cocycle \( A \) as well, we have \( \hat{e}_\pm(A) \in \mathcal{D}_\pm(\delta) \) and \( \hat{e}_\pm(B_i) \in \mathcal{D}_\pm(\delta) \) for all \( i \in \Sigma_H(A) \). We claim that for every \( i \in \Sigma_H(A) \),
\[
d(\hat{e}_\pm(B_i), \hat{e}_\pm(D)) \leq \frac{C}{1 - e^{-\frac{2}{3}L(A)}} \rho_B.
\] (4.13)
By Lemma 4.5 we have
\[
d(\hat{e}_\pm(B_i), \hat{e}_\pm(D)) \leq d(\hat{B}_i^{\pm 1} \hat{e}_\pm(B_i), \hat{B}_i^{\pm 1} \hat{e}_\pm(D)) + d(\hat{B}_i^{\pm 1} \hat{e}_\pm(D), \hat{B}_i^{\pm 1} \hat{e}_\pm(D))
\leq e^{-\frac{2}{3}L(A)} d(\hat{e}_\pm(B_i), \hat{e}_\pm(D)) + \|B_i^{\pm 1} - D_i^{\pm 1}\|_\infty,
\]
which implies by N7 that
\[
(1 - e^{-\frac{2}{3}L(A)}) d(\hat{e}_\pm(B_i), \hat{e}_\pm(D)) \leq C \rho_B
\]
and proves the claim.

Therefore, using (4.13),
\[
d(\hat{B}_j^{\pm 1} \hat{e}_\pm(D), \hat{e}_\pm(D)) \leq d(\hat{B}_j^{\pm 1} \hat{e}_\pm(D), \hat{B}_j^{\pm 1} \hat{e}_\pm(B_i))
\]
\[
+ d(\hat{B}_j^{\pm 1} \hat{e}_\pm(B_i), \hat{e}_\pm(B_i)) + d(\hat{e}_\pm(B_i), \hat{e}_\pm(D))
\leq e^{-\frac{2}{3}L(A)} d(\hat{e}_\pm(D), \hat{e}_\pm(B_i)) + \rho_B + \frac{C}{1 - e^{-\frac{2}{3}L(A)}} \rho_B,
\]
\[
\leq \frac{C e^{-\frac{2}{3}L(A)}}{1 - e^{-\frac{2}{3}L(A)}} \rho_B + \rho_B + \frac{C}{1 - e^{-\frac{2}{3}L(A)}} \rho_B = M_0 \rho_B.
\] \( \square \)
Proof of Proposition 4.4. Fix $i \in \Sigma_H(A)$ and $1 \leq j \leq k$ such that
\[
\rho_B = d(\hat{B}_j \hat{e}_-(B_i), \hat{e}_-(B_i)).
\]
Set $b_0 = p_i^* p_j > 0$, where $n_0 \in \mathbb{N}$ is large enough but yet to be chosen depending only on $A$. We claim that for any $\hat{x} \in \Sigma_0$ either $\hat{B}_i^{n_0} \hat{B}_j \hat{x} \notin \Sigma_0$, if $\hat{x}$ is close enough to $\hat{e}_-(B_i)$, or else $\hat{B}_i^{n_0} \hat{x} \notin \Sigma_0$ otherwise. In either case the probability of these events is at least $b_0$.

Next we need to explain how to fix $n_0 \in \mathbb{N}$, depending on $A$, so that one of the above alternative claims holds.

Given $\hat{x} \in \Sigma_0$, the condition $d(\hat{x}, \hat{e}_-(B_i)) < \frac{\rho_B}{L^2 + 1}$ describes the first case, in which we have
\[
d(\hat{B}_j \hat{x}, \hat{e}_-(B_i)) \geq d(\hat{B}_j \hat{e}_-(B_i), \hat{e}_-(B_i)) - d(\hat{B}_j \hat{e}_-(B_i), \hat{B}_j \hat{x}) \geq \rho_B - L^2 d(\hat{e}_-(B_i), \hat{x}) \geq \rho_B - L^2 \frac{\rho_B}{L^2 + 1} = \frac{\rho_B}{L^2 + 1}.
\]
Note that by N3 and Remark 4.1, $d(\hat{B}_j \hat{x}, \hat{B}_j \hat{y}) \leq L^2 d(\hat{x}, \hat{y})$ for all $\hat{x}, \hat{y} \in \mathbb{P}$. The previous inequality shows that for any $\hat{x} \in \Sigma_0$, either $d(\hat{x}, \hat{e}_-(B_i)) \geq \frac{\rho_B}{L^2 + 1}$ or else $d(\hat{B}_j \hat{x}, \hat{e}_-(B_i)) \geq \frac{\rho_B}{L^2 + 1}$ and hence reduces the proof to the first case. Assuming $d(\hat{x}, \hat{e}_-(B_i)) \geq \frac{\rho_B}{L^2 + 1}$ we have to find $n_0 \in \mathbb{N}$ such that
\[
d(\hat{B}_i^{n_0} \hat{x}, \hat{e}_-(B_i)) \geq M' \rho_B
\]
with $M' := M + \frac{C}{1 - e^{-\frac{L(A)}{2}}}$. Using (4.13) this implies
\[
d(\hat{B}_i^{n_0} \hat{x}, \hat{e}_-(D)) \geq \frac{C}{1 - e^{-\frac{L(A)}{2}}} \rho_B = M' \rho_B
\]
thus proving that $\hat{B}_i^{n_0} \hat{x} \notin \Sigma_0$.

To finish, we find $n_0 \in \mathbb{N}$ such that (4.14) holds. Since by N6 $\Sigma_0 \subset \mathcal{D}_-(\delta)$, using Lemma 4.5 we have that
\[
d(\hat{B}_i^{n_0} \hat{x}, \hat{e}_-(B_i)) \geq e^{-\frac{L(A)}{3} n_0 A} d(\hat{x}, \hat{e}_-(B_i)) \geq e^{-\frac{L(A)}{3} n_0 A} \frac{\rho_B}{L^2 + 1} \geq M' \rho_B
\]
provided we take $n_0 > \frac{3}{5} \frac{\log[M' (L^2 + 1)]}{L(A)}$. Thus we can pick $n_0 \in \mathbb{N}$ depending only on $A$ such that (4.14) holds whenever $d(\hat{x}, \hat{e}_-(B_i)) \geq \frac{\rho_B}{L^2 + 1}$. $\square$
Establishing Assumption (A2).

Proposition 4.7. If \( n_1(B) := \lceil \frac{2}{E(A)} \log(\rho_B^{-1}) \rceil \) then for all \( \hat{x} \in \Sigma_1 \setminus \Sigma_0 \),
\[
\mathbb{P}_x \left[ \exists j \leq n_1, \xi_j^\hat{x}(B) \notin \Sigma_1 \right] \geq 1 - r/4.
\]

Consider the projective coordinate system \( \psi = \psi \) introduced above, in which \( \hat{e}_-(D) = 0 \) and \( \hat{e}_+(D) = \infty \).

We will keep denoting by \( \hat{B}_j \) the action of \( B_j \) expressed in the previous projective coordinate, i.e., we write \( \hat{B}_j(x) := \psi(B_j(x : 1)) \). Consider the family of numbers \( \{\lambda^*_j\}_{1 \leq j \leq k} \) defined in (4.6). A simple calculation gives
\[
\sum_{j=1}^{k} p_j \log \lambda^*_j = \frac{3L(A)}{2} > 0. \tag{4.15}
\]

Lemma 4.8. For all \( x \in [s_M, \delta^{-1}] \) and \( 1 \leq j \leq k \),
\[
\hat{B}_j(x) \geq M_0 \rho_B + \lambda^*_j (x - M_0 \rho_B) =: \hat{B}_j^*(x). \tag{4.16}
\]

Proof. By Lemma 4.6, the projective point \( \hat{B}_j(\hat{e}-(D)) \) has coordinate \( |\hat{B}_j(0)| \leq 2M_0 \rho_B \).

Writing \( B_j = \begin{bmatrix} b_j & c_j \\ d_j & e_j \end{bmatrix} \) one has \( \hat{B}_j(x) = (b_j x + c_j)(d_j x + e_j)^{-1} \). Since \( B_j \in \text{SL}_2(\mathbb{R}) \), \( |d_j| \leq ||B-D||_{\infty} < d(A,B) \) and \( 0 \leq x \leq \delta^{-1} \) we have
\[
\hat{B}_j'(x) = \frac{1}{(d_j x + e_j)^2} \geq \frac{1}{(|e_j| + \delta^{-1} d(A,B))^2}.
\]

Because \( \hat{e}_\pm(A) \in \mathcal{D}_\pm(\delta) \), by N10, and also by N5 and N9,
\[
|e_j| \leq ||B \circ e_-(D)|| \leq e^{-\log|a_j| + \frac{L(A)}{12}} = \lambda_j^{-\frac{1}{2}} e^{-\frac{L(A)}{12}} < \tilde{\lambda}_j^{-\frac{1}{2}}.
\]

Thus for all \( j = 1, \ldots, k \) and \( x \in [0, \delta^{-1}] \),
\[
\hat{B}_j'(x) \geq \frac{1}{(\lambda_j^{-\frac{1}{2}} e^{-\frac{L(A)}{12}} + \delta^{-1} d(A,B))^2} \geq \frac{\tilde{\lambda}_j e^{\frac{L(A)}{6}}}{(1 + \delta^{-1} L e^{\frac{L(A)}{4}} d(A,B))^2}.
\]

By N0 (a)
\[
d(A, B) < \frac{\delta}{L} e^{\frac{L(A)}{12}} - \frac{1}{L e^{\frac{L(A)}{4}}},
\]
which is equivalent to
\[
\gamma = \frac{\frac{L(A)}{6}}{(1 + \delta^{-1} L e^{\frac{L(A)}{4}} d(A,B))^2} > 1.
\]
Therefore \( \hat{B}_j'(x) \geq \gamma \hat{\lambda}_j > \bar{\lambda}_j \) and by the mean value theorem for all \( 0 \leq x \leq \delta^{-1} \),

\[
\hat{B}_j(x) \geq \hat{B}_j(0) + \bar{\lambda}_j x \geq -2M_0 \rho_B + \bar{\lambda}_j x.
\]

Since \( \hat{B}_j'(x) \geq \bar{\lambda}_j > \lambda_j^* = (\check{B}_j^*)'(x) \), to see that \( \hat{B}_j(x) \geq \check{B}_j^*(x) \) for all \( x \in [sM_0 \rho_B, \delta^{-1}] \), it is now enough to check that

\[
-2M_0 \rho_B + \bar{\lambda}_j (sM_0 \rho_B) \geq \check{B}_j^*(sM_0 \rho_B)
\]

\[
= M_0 \rho_B + \lambda_j^* (sM_0 \rho_B - M_0 \rho_B),
\]

which is equivalent to

\[
-2 + \bar{\lambda}_j s \geq 1 + \lambda_j^* (s - 1).
\]

Hence, because (4.7) is enough for this, by the definition of \( s = s(A) \) inequality (4.16) holds for all \( x \in [sM_0 \rho_B, \delta^{-1}] \).

Consider the i.i.d. process \( \eta_j : X \to \mathbb{R}_+ \), \( \eta_j \{ \omega_j \}_{j \in \mathbb{Z}} : = \log \lambda_{\omega_j}^* \) and its associated sum process \( S_n = \sum_{j=0}^{n-1} \eta_j = \sum_{j=0}^{n-1} \eta_0 \circ T^j \). Note that the processes \( \{ \eta_n \} \) and \( \{ S_n \} \) are completely determined by the cocycle \( A \).

The probability of escaping in assumption (A2) is estimated comparing the random walks \( \xi_n^\omega \) and \( S_n \) by means of (4.16).

**Proof of Proposition 4.7.** By (4.15) one has \( \mathbb{E}[\eta_0] = \frac{3L(A)}{2} \). By Hoeffding’s inequality (Lemma 2.2) the constant \( \check{c} = \check{c}(A) \) in (4.9) is such that the following LDT estimate holds for all \( n \in \mathbb{N} \),

\[
P_p[S_n < nL(A)] < e^{-\check{c}n}. \tag{4.17}
\]

Indeed the i.i.d. process \( \eta_j \) satisfies

\[
|\eta_j| = |\log \lambda_{\omega_j}^*| = |2 \log |a_{\omega_j}| - L(A)/2| \leq 2 \log L + L(A)/2 =: K
\]

while the event \( [S_n < nL(A)] \) is contained in the deviation set

\[
[|n^{-1}S_n - \mathbb{E}(n^{-1}S_n)| > \epsilon] \quad \text{with} \quad \epsilon := \frac{L(A)}{3}.
\]

Since \( \check{c} = \frac{e^2}{2K^2} \), the previous LDT estimate follows from Hoeffding’s inequality.

Consider now \( l_0 \in \mathbb{N} \) and \( \kappa < \infty \) respectively defined by (4.10) and in (4.11). Then \( \sum_{j=l_0}^{\infty} e^{-\check{c}j} = \frac{e^{-\check{c}l_0}}{1-e^{-\check{c}}} < \frac{\epsilon}{8} \) and

\[
P_p[S_j \geq -\kappa, \forall j = 1, \ldots, l_0] = 1. \tag{4.18}
\]

This implies that the event \( \Omega = [S_j \geq -\kappa, \forall j \geq 1] \) has probability \( P_p(\Omega) \geq 1 - \frac{\epsilon}{8} \).
Given \( \hat{x} \in \Sigma_1 \setminus \Sigma_0 \), we have \( t = |\psi(\hat{x})| \in [M \rho_B, \delta^{-1}] \). Inductively, and while \( \xi_n^x \in \Sigma_1 \), one obtains by successive applications of inequality (4.16) that for all \( \omega = \{\omega_j\}_{j \in \mathbb{Z}} \in \Omega \),
\[
|\psi(\xi_n^x(\omega))| = \hat{B}_{\omega_{n-1}} \ldots \hat{B}_{\omega_1} \hat{B}_{\omega_0}(t) \geq B_{\omega_{n-1}}^{*} \ldots B_{\omega_1}^{*} \hat{B}_{\omega_0}^{*}(t)
\geq M_0 \rho_B + e^{S_n}(t - M_0 \rho_B)
\geq e^{-\kappa} M \rho_B = s M_0 \rho_B
\]
which in particular justifies that we can keep inductively applying inequality (4.16).

Consider now the event \( B = [S_{n_1} < n_1 L(A)] \) which by (4.17) has probability \( < e^{-\hat{e} n_1} \). By N11, \( l_0 < n_1 \) and by (4.10), \( e^{-\hat{e} l_0} < \frac{1}{8} \). Then the event \( \Omega' = \Omega \setminus B \) has probability
\[
\mathbb{P}_p(\Omega') \geq \mathbb{P}_p(\Omega) - \mathbb{P}_p(B) \geq 1 - \frac{r}{8} - e^{-\hat{e} n_1} > 1 - \frac{r}{4}.
\]
Assume now, by contradiction, that for some \( \omega \in \Omega' \) one has for all \( j \leq n_1, \xi_n^x = \xi_n^x(\omega) \in \Sigma_1 \). This implies that \( |\psi(\xi_n^x)| \in [M \rho_B, \delta^{-1}] \) for all \( j \leq n_1 \). We can assume that the number \( s = s(A) \) defined by (4.7) and (4.8) satisfies \( s \geq 2 \). Because \( M_0 \geq 1 \), it follows that \( M \geq M_0 + 1 \). Then the previous inductive chain of inequalities implies that
\[
|\psi(\xi_n^x)| \geq M_0 \rho_B + e^{S_{n_1}}(t - M_0 \rho_B)
\geq e^{n_1 L(A)}(t - M_0 \rho_B) \geq e^{n_1 L(A)}(M - M_0) \rho_B
\geq e^{n_1 L(A)} \rho_B = \frac{1}{\rho_B^2} \rho_B = \rho_B^{-1} > \delta^{-1},
\]
which contradicts \( \xi_n^x \in \Sigma_1 \). This concludes the proof. \( \square \)

**Establishing Assumption (A3).**

**Proposition 4.9.** If \( \eta_2(B) := \lceil \frac{1}{L(A)} \log(\rho_B^{-1}) \rceil \) then for all \( \hat{x} \in \Sigma_2 \setminus \Sigma_1 \),
\[
\mathbb{P}_{\hat{x}} \left[ \exists j \leq n_2, \xi_n^x \notin \Sigma_2 \right] \geq 1 - r/4.
\]

Consider the projective coordinate system \( \psi = \psi_\perp \) introduced above, in which \( \Sigma^6_1 = \mathcal{D}_- (\delta^{-1})^6 = [-\delta, \delta] \) and \( \Sigma^6_2 = \mathcal{D}_- ((M \rho_B)^{-1})^6 = [-M \rho_B, M \rho_B] \).

We will keep denoting by \( \hat{B}_j \) the action of \( B_j \) expressed in the previous projective coordinate, i.e., we write \( \hat{B}_j(x) := \psi(\hat{B}_j(1 : x)) \).

Consider again the parameters \( \hat{\lambda}_j, \lambda_j^* \) defined in (4.5) and (4.6).

**Lemma 4.10.** For all \( x \in [s M_0 \rho_B, (e^{\kappa} + 1) \delta] \) and \( 1 \leq j \leq k \),
\[
\hat{B}_j(x) \leq M_0 \rho_B + (\lambda_j^*)^{-1}(x - M_0 \rho_B) =: \hat{B}_j^*(x).
\] (4.19)
Proof. By Lemma 4.6, the projective point $\hat{B}_j(\hat{e}_+(D))$ has coordinate $|\hat{B}_j(0)| \leq 2 M_0 \rho_B$.

Writing $B_j = \begin{bmatrix} b_j & c_j \\ d_j & e_j \end{bmatrix}$ one has $\hat{B}_j(x) = (d_j + e_j x) (b_j + c_j x)^{-1}$. Since $B_j \in \text{SL}_2(\mathbb{R}), |c_j| \leq \|B - D\|_\infty < d(A, B)$ and $0 \leq x \leq \delta^{-1}$ we have

$$\hat{B}_j'(x) = \frac{1}{(b_j + c_j x)^2} \leq \frac{1}{(|b_j| - \delta^{-1} d(A, B))^2}.$$  

Because $|d_j| \leq d(A, B)$ and $|b_j| \geq |a_j| - d(A, B) \geq L^{-1} - d(A, B)$,

$$\frac{|b_j|}{\sqrt{b_j^2 + d_j^2}} \geq \frac{1}{\sqrt{1 + \frac{d(A, B)^2}{(L^{-1} - d(A, B))^2}}} > e^{-\frac{L(A)}{24}},$$

where the last inequality follows from N0(b). Therefore

$$|b_j| \geq e^{-\frac{L(A)}{24}} \sqrt{b_j^2 + d_j^2} = e^{-\frac{L(A)}{24}} \|B_j e_+(D)\| \geq e^{-\frac{L(A)}{24}} e^{\log |a_j| - \frac{L(A)}{12}} = \lambda_j^2 e^{\frac{L(A)}{24}} > \bar{\lambda}_j^2.$$  

Thus for all $j = 1, \ldots, k$ and $x \in [0, \delta^{-1}]$,

$$\hat{B}_j'(x) \leq \frac{1}{\left(\lambda_j^2 e^{\frac{L(A)}{24}} - \delta^{-1} d(A, B)\right)^2} \leq \frac{\bar{\lambda}_j^{-1} e^{-\frac{L(A)}{24}}}{(1 - \delta^{-1} L e^{\frac{L(A)}{8}} d(A, B))^2}.$$  

By N0 (c)

$$d(A, B) < \delta L^{-1} e^{-\frac{L(A)}{4}} \left(1 - e^{-\frac{L(A)}{24}}\right),$$

which is equivalent to

$$\gamma = \frac{e^{-\frac{L(A)}{24}}}{(1 - \delta^{-1} L e^{\frac{L(A)}{8}} d(A, B))^2} < 1.$$  

Therefore $\hat{B}_j'(x) \leq \gamma \bar{\lambda}_j^{-1} < \bar{\lambda}_j^{-1}$ and by the mean value theorem for all $0 \leq x \leq \delta^{-1}$,

$$\hat{B}_j(x) \leq \hat{B}_j(0) + \bar{\lambda}_j^{-1} x \leq 2 M_0 \rho_B + \bar{\lambda}_j^{-1} x.$$  

Since $\hat{B}_j'(x) \leq \bar{\lambda}_j^{-1} < (\lambda_j^*)^{-1} = (\hat{B}_j^*)'(x)$, inequality $\hat{B}_j(x) \leq \hat{B}_j^*(x)$, for all $x \in [s M_0 \rho_B, \delta^{-1}]$, follows from

$$2 M_0 \rho_B + \bar{\lambda}_j^{-1} (s M_0 \rho_B) \leq \hat{B}_j^*(s M_0 \rho_B) = M_0 \rho_B + (\lambda_j^*)^{-1} (s M_0 \rho_B - M_0 \rho_B),$$

which is equivalent to

$$2 + \bar{\lambda}_j^{-1} s \leq 1 + (\lambda_j^*)^{-1} (s - 1).$$
Hence, because (4.8) is enough for this, by the definition of \( s = s(A) \), inequality (4.19) holds for all \( x \in [s M_0 \rho_B, \delta^{-1}] \). Finally, because \( \delta \) can be made small enough we may assume that \( (e^\kappa + 1) \delta < \delta^{-1} \).

The probability of escaping in assumption (A3) is estimated comparing again the random walks \( \xi_n^\hat{x} \) and \( S_n \), now by means of (4.19). The argument is analogous to the one used for assumption (A2).

**Proof of Proposition 4.9.** Consider the constants \( \hat{c} = \hat{c}(A) > 0 \), \( l_0 = l_0(A) \) and \( \kappa = \kappa(A) < \infty \), respectively defined in (4.9), (4.10) and (4.11) and consider the event

\[
\Omega' = \{ S_{n_2} > n_2 L(A) \text{ and } S_j \geq -\kappa, \forall j \geq 1 \}
\]

which, as we have seen in the proof of Proposition 4.7, has probability \( \mathbb{P}_p(\Omega) \geq 1 - \frac{r}{4} \).

Given \( \hat{x} \in \Sigma_2 \setminus \Sigma_1 \), we have \( t = |\psi(\hat{x})| \in [M \rho_B, \delta] \). Inductively, and while \( \xi_n^\hat{x} \in \Sigma_2 \), one obtains by successive applications of inequality (4.19) that for all \( \omega = \{\omega_j\}_{j \in \mathbb{Z}} \in \Omega' \),

\[
|\psi(\xi_n^\hat{x}(\omega))| = \hat{B}_{\omega_{n-1}} \ldots \hat{B}_{\omega_1} \hat{B}_{\omega_0}(t) \leq B_{\omega_{n-1}}^* \ldots \hat{B}_{\omega_1}^* \hat{B}_{\omega_0}(t) \\
\leq M_0 \rho_B + e^{-S_n}(t - M_0 \rho_B) \\
\leq M_0 \rho_B + e^\kappa \delta < \delta (e^\kappa + 1)
\]

which in particular justifies that we can keep inductively applying (4.19).

Establishing Assumption (A4).

**Proposition 4.11.** For all \( \hat{x} \in \Sigma_2^\complement \),

\[
\mathbb{P}_\hat{x} \left[ \forall j \geq 0, \xi_j^\hat{x}(B) \notin \Sigma_1 \right] \geq 1 - r/4.
\]
Proof. Consider the projective coordinate $\psi = \psi_+$ introduced in the previous subsection. Through this coordinate system we make the identifications $\Sigma_2^c \equiv \mathcal{D}_-((M \rho_B)^{-1})^c = [-M \rho_B, M \rho_B]$ and $\Sigma_1^c \equiv \mathcal{D}_-(\delta)^c = [-\delta, \delta]$.

Given $\hat{x} \in \Sigma_2^c$ and $N \in \mathbb{N}$, consider the event

$$\Omega_N := [\exists 1 \leq j \leq N \text{ such that } \xi_j^\hat{x} \in \Sigma_1].$$

Our goal is to prove that $\mathbb{P}(\Omega_N) < r/4$ for all $N \in \mathbb{N}$.

Consider the following family of events indexed in $0 \leq i \leq N$:

$$\mathcal{E}_i^N := \left[ \xi_i^\hat{x} \not\in \Sigma_2 \text{ and } \forall i < l \leq N, \xi_l^\hat{x} \in \Sigma_2 \right].$$

Because $\hat{x} \not\in \Sigma_2$, these events partition the space of sequences $X$.

As before, let $S_n := \sum_{j=0}^{n-1} \eta \circ T^j$ be the sum process associated with the i.i.d. process generated by the random variable $\eta: X \rightarrow \mathbb{R}$, $\eta(\omega) := \log \lambda_{\omega}^*$. Given $\omega \in \Omega_N \cap \mathcal{E}_i^N$, let $t_i := \psi(\xi_i^\hat{x}(\omega)) < M \rho_B$ and $n \in [1, N - i]$ be the first integer such that $\xi_{i+n}^\hat{x} \in \Sigma_1$. Proceeding inductively, from inequality (4.19) we get that

$$\delta \leq |\psi(\xi_{i+n}^\hat{x}(\omega))| = B_{\omega_{i+n-1}} \cdots B_{\omega_{i+1}} B_{\omega_i}(t_i) \leq \hat{B}^*_{\omega_{i+n-1}} \cdots B^*_{\omega_{i+1}} B^*_{\omega_i}(t_i)
\leq M_0 \rho_B + e^{-S_n(T^i \omega)} (t_i - M_0 \rho_B)
< M_0 \rho_B + e^{-S_n(T^i \omega)} (M - M_0) \rho_B.
$$

By N6 this implies that $e^{-S_n(T^i \omega)} > 1$ and hence that

$$\delta < e^{-S_n(T^i \omega)} M \rho_B.$$

In other words we have established the inclusion

$$\Omega_N \cap \mathcal{E}_i^N \subset T^{-i} \left[ \exists n \geq 1 \text{ such that } e^{-S_n} M \rho_B > \delta \right],$$

which implies that

$$\mathbb{P}[\Omega_N | \mathcal{E}_i^N] \leq \mathbb{P} \left[ \exists n \geq 1 \text{ such that } e^{-S_n} M \rho_B > \delta \right].$$

Next, define $l = l(A) := \max\{\log(\lambda_j^*)^{-1} : 1 \leq j \leq k\}$ so that $S_n \geq -nl$ for all $n \geq 0$. Note also that $l = \kappa/l_0 > 0$. By assumption N6 we have $M \rho_B^{1/2} < \delta$. Since

$$e^{-S_n} M \rho_B > \delta \quad \Rightarrow \quad e^{S_n} < \delta^{-1} M \rho_B < \rho_B^{1/2}
\quad \Rightarrow \quad -nl \leq S_n < \frac{1}{2} \log(\rho_B) < 0
\quad \Rightarrow \quad n \geq \frac{1}{2l} \log(\rho_B^{-1}) \text{ and } S_n < n L(A).$$

it follows that

$$\mathbb{P}[\Omega_N | \mathcal{E}_i^N] \leq \mathbb{P} \left[ \exists n \geq \frac{1}{2l} \log(\rho_B^{-1}) \text{ such that } S_n < n L(A) \right]$$
which by (4.17), (4.10) and assumption N11 has probability
\[
e^{-\hat{c}_l \log(\rho - 1_B)} \frac{e^{-\hat{c}_l \log(\rho_B^{-1})}}{1 - e^{-\hat{c}_l}} < \frac{e^{-\hat{e}_l}}{1 - e^{-\hat{c}_l}} < \frac{r}{8} < \frac{r}{4}.
\]

Finally, applying the Law of Total Probabilities
\[
P(\Omega_N) = \sum_{i=0}^{n-1} P[\Omega_N | \mathcal{E}_i^N] P(\mathcal{E}_i^N)
\]
\[
< \frac{r}{4} \sum_{i=0}^{n-1} P(\mathcal{E}_i^N) = \frac{r}{4}.
\]

\[\square\]

**Conclusion.** We now apply the general prison break result to our setting.

**Proposition 4.12.** Let \( A \in C \) be a diagonalizable cocycle such that \( L(A) > 0 \). There exists \( c_0 = c_0(A) < \infty \) such that if \( B \in \mathcal{N}_A \) with \( \rho_B = \rho_-(B) > 0 \) and \( n_B := c_0 \log(\rho_B^{-1}) \) then for all \( \hat{x} \in \mathbb{P} \)
\[
P[\xi_j(B) \in \Sigma_1] < r.
\]

**Proof.** Take \( n_0 \in \mathbb{N} \) and \( b_0 > 0 \) as given by Proposition 4.4. Next choose \( q = q(A) \in \mathbb{N} \) such that \( q \geq \frac{1}{b_0} \log(4/r) \), which implies that \((1 - b_0)^q < \frac{r}{4} \). Take \( n_1 = n_1(B) \) and \( n_2 = n_2(B) \in \mathbb{N} \) as given by propositions 4.7 and 4.9.

By Proposition 4.11, \( P[\xi_j(B) \in \Sigma_1] \geq 1 - r/4 \), for all \( \hat{x} \in \Sigma_1 \).

Therefore all assumptions (A1)-(A4) of Proposition 4.3 are satisfied, the first three within the times \( n_0, n_1 \) and \( n_2 \) respectively. Reducing the neighborhood \( \mathcal{N}_A \) we can assume that \( \log \rho_B^{-1} > n_0 \) for all \( B \in \mathcal{N}_A \). Since the time \( N = qn_0 + n_1 + n_2 \) is bounded above by \( c_0 \log(\rho_B^{-1}) \) with \( c_0 = q + \frac{3}{L(A)} \), \( c_0 \) being a constant depending only on \( A \), the conclusion of this proposition follows from that of Proposition 4.3. \[\square\]

**Proof of Proposition 3.8.** The second inequality reduces to applying the first one to the inverse cocycle \( B^{-1} \). Thus, from now on we will focus on the first.

Since \( A \in \text{Diag}^* \) we may assume without loss of generality that \( A \) is a diagonal cocycle,
\[
A_i = \begin{bmatrix} a_i & 0 \\ 0 & a_i^{-1} \end{bmatrix} \quad \text{for } 1 \leq i \leq k
\]
and so $\hat{e}_+(A) = (1 : 0)$ and $\hat{e}_-(A) = (0 : 1)$. In the coordinate system $\psi_+$ associated with the diagonal cocycle $A$, the projective action of the matrix $A_i$ is given by $\hat{A}_i x = a_i^{-2} x$. In these coordinates, the random walk $\xi_n^\hat{v} = \xi_n^\hat{v}(A)$ becomes a random walk $\xi_n^x$ on the real line with starting point $x$, where $\hat{v} = (1 : x)$.

Next consider the i.i.d. process $\eta_n$, where $\eta_n = \log |a_i|$ with probability $p_i$, and denote by $S_n = \sum_{j=0}^{n-1} \eta_j$ the corresponding sum process. A simple verification shows that

$$\mathbb{E}\left[\frac{1}{n} S_n\right] = \mathbb{E}[\eta_0] = L(A).$$

By Hoeffding’s inequality there exist positive constants $n_0 \in \mathbb{N}$ and $c_1$ depending only on $A$ such that for all $n \geq n_0$,

$$\mathbb{P}\left[\exists j \geq n, \frac{1}{j} S_j < \frac{9}{10} L(A)\right] < e^{-c_1 n}. \quad (4.21)$$

Let $c_0 = c_0(A)$ and $n_B = c_0 \log \rho_B^{-1}$ be the constants provided by Proposition 4.12. Making $N_A$ small enough, since by Proposition 3.5 $\rho_B \lesssim d(A, B)$, we can assume that $n_B \geq n_0$. Then take the union $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ of the sets defined in (4.20) and (4.21), i.e.,

$$\mathcal{B}_1 := \left\{ \omega \in X : \exists j \geq n_B \quad \xi_j^\hat{v}(B)(\omega) \in \Sigma_1 \right\} ,$$

$$\mathcal{B}_2 := \left\{ \omega \in X : \frac{1}{n} S_n(\omega) < \frac{9}{10} L(A) \text{ for some } n \geq n_B \right\} .$$

For each matrix $B_j$ consider the projective function $\varphi_{B_j}(\hat{v}) := \log \|B_j v\|$. By N5 and N9, for all $\hat{v} \in \mathcal{D}_+(A, 2\delta)$ and $1 \leq j \leq k$

$$\left| \varphi_{B_j}(\hat{v}) - \log |a_j| \right| = \left| \log \|B_j v\| - \log \|A_j e_+(A)\| \right| \leq \frac{L(A)}{12} .$$

Given $\omega = \{\omega_j\}_{j \in \mathbb{Z}} \notin \mathcal{B}$ and $\hat{v} \in \mathbb{P}$, because $\xi_j^\hat{v}(B) \in \Sigma_1^\hat{v} = \overline{\mathcal{D}_+(\delta)} \subset \mathcal{D}_+(A, 2\delta)$ for all $n \geq n_B$,

$$\frac{1}{n} \log \|B^{(n)} v\| - \frac{1}{n} S_n \leq \frac{1}{n} \sum_{j=0}^{n-1} \left| \varphi_{B_{\omega_j}}(\xi_j^\hat{v}(B)) - \log |a_{\omega_j}| \right|$$

$$\leq \frac{n_B}{n} 2 \log L + \frac{n - n_B}{n} \frac{L(A)}{12}$$

$$\leq 2 \frac{n_B}{n} \log L + \frac{L(A)}{12} \leq \frac{L(A)}{10}$$

where the last inequality holds for $n \geq \frac{120 \log L}{L(A)} n_B$. 

Hence, since $\omega \notin \mathcal{B}_2$ and $n \geq n_0$,
\[
\frac{1}{n} \log \|B^{(n)}v\| \geq \frac{1}{n}S_n - \frac{L(A)}{10} \geq \left(\frac{9}{10} - \frac{1}{10}\right) L(A) > \frac{7}{10} L(A).
\]
Therefore, integrating we have for every $\hat{v} \in \mathcal{P}$ and $n \geq \frac{120 \log L}{L(A)} n_B$,
\[
\mathbb{E} \left[\frac{1}{n} \log \|B^{(n)}v\|\right] \geq -(\log L) \mathbb{P}(\mathcal{B}) + (1 - \mathbb{P}(\mathcal{B})) \frac{7 L(A)}{10}
\geq \frac{2 L(A)}{3} > \frac{L(B)}{2}
\]
where the last inequality holds by assumption N2, and the one next-to-last is also valid provided
\[
\mathbb{P}(\mathcal{B}) < \frac{L(A)}{30 \log L + 21 L(A)} = 2r.
\tag{4.22}
\]
Note from the above that in particular we also have
\[
L(B) \geq \frac{2 L(A)}{3}
\tag{4.23}
\]
for all $B \in N_A$. By (4.21) and Proposition 4.12, the inequality
\[
\mathbb{P}(\mathcal{B}) < r + e^{-c_1 n_B} < 2r
\]
is equivalent to
\[
n_B \geq \frac{1}{c_1} \log \left(42 + \frac{60 \log L}{L(A)}\right).
\]
Since the right-hand-side of this inequality depends only on $A$, reducing the neighborhood $N_A$, we can assume that it holds for all $B \in N_A$. Therefore, setting $K = \frac{120 \log L}{L(A)}$, for all $n \geq K n_B$ and $\hat{v} \in \mathcal{P}$, relation (4.22) holds which in turn implies that
\[
\mathbb{E} \left[\frac{1}{n} \log \|B^{(n)}v\|\right] > \frac{L(B)}{2}.
\]
By Definition 3.1 this proves that
\[
N(B) \leq K n_B \leq K c_0 \log (\rho_-(B)^{-1}).
\]
The proposition is established with $c = K^{-1} c_0^{-1}$.

**Remark 4.2.** As a by-product of the previous proof we derive the fact (to be used later) that $L(B) \geq 2 L(A)/3$.

An improvement of this argument leads to the lower semi-continuity of the LE, which when combined with the upper-semicontinuity shows that the LE function $A \mapsto L(A)$ is continuous at any diagonalizable cocycle $A \in \text{SL}_2(\mathbb{R})^k$ with $L(A) > 0$. As remarked in [1], E. Le Page's
Theorem on the Hölder continuity of the LE holds for quasi-irreducible SL$_2$-cocycles $A$ such that $L(A) > 0$. For any non diagonalizable SL$_2$-cocycle $A$ such that $L(A) > 0$ either $A$ or else the inverse cocycle $A^{-1}$ are quasi-irreducible. Finally, since the continuity of the LE at $A$ is equivalent to the continuity at $A^{-1}$, combining these facts we get an alternative proof of C. Boker-Neto and M. Viana’s Theorem [2] on the general continuity of the LE for SL$_2$($\mathbb{R}$) and GL$_2$($\mathbb{R}$) random cocycles in the finite support setting.

5. Quantitative LDT for irreducible cocycles

In this section we establish uniform LDT estimates for non-diagonalizable cocycles in a neighborhood of a diagonalizable one.

**Theorem 5.1.** Given a cocycle $A \in \text{Diag}$ with $L(A) > 0$ there exists a neighborhood $N_A$ of $A$ in $C = \text{SL}_2(\mathbb{R})^k$ such that for any cocycle $B \in N_A$ with $\rho_-(B) > 0$

$$
P \left[ \left| \frac{1}{n} \log \| B^{(n)} \| - L(B) \right| > n^{-1/6} \right] < e^{-n^{1/6}}$$

for all $n \geq \lfloor c^{-1} \log \rho_-(B)^{-1} \rfloor^{54}$.

The proof will be based on a functional analytic argument.

Let $(\mathcal{B}, \| \cdot \|_\mathcal{B})$ be a Banach space and $Q: \mathcal{B} \to \mathcal{B}$ a bounded linear operator on $\mathcal{B}$. We say that $Q$ is quasi-compact and simple if there exists a $Q$-invariant decomposition $\mathcal{B} = E_0 \oplus E_1$ into closed subspaces $E_0, E_1$ such that for some constants $0 < \sigma < \lambda$:

1. $\dim E_1 = 1$ and $Q(\phi) = \lambda \phi$ for all $\phi \in E_1$;
2. The operator $Q|_{E_0}: E_0 \to E_0$ has spectral radius $< \sigma$.

Denoting by $P: \mathcal{B} \to E_1$ the spectral projection there is a constant $C < \infty$ such that for all $\phi \in \mathcal{B}$,

1. $Q(P(\phi)) = P(Q(\phi)) = \lambda P(\phi)$;
2. $\| P(\phi) \|_\mathcal{B} \leq C \| \phi \|_\mathcal{B}$;
3. $\| Q^n(\phi) - \lambda^n P(\phi) \|_\mathcal{B} \leq C \sigma^n \| \phi \|_\mathcal{B}$, for any $n \in \mathbb{N}$.

We will refer to $\sigma, \lambda$ and $C$ as the spectral constants of the quasi-compact and simple operator $Q$.

Let $\Sigma = \{1, \ldots, k\}$ and $\mathbb{P} = \mathbb{P}(\mathbb{R}^2)$. Define the space $L^\infty(\Sigma \times \mathbb{P})$ of all complex bounded measurable functions $\phi: \Sigma \times \mathbb{P} \to \mathbb{C}$ and let $\| \phi \|_\infty$ denote the usual sup norm of an observable $\phi \in L^\infty(\Sigma \times \mathbb{P})$. 
Given $0 < \alpha \leq 1$ define also

$$v_\alpha(\phi) := \max_{i \in \Sigma} \sup_{\hat{p} \neq \hat{q}} \frac{|\phi(i, \hat{p}) - \phi(i, \hat{q})|}{d(\hat{p}, \hat{q})^\alpha}.$$  

A function $\phi \in L^\infty(\Sigma \times \mathbb{P})$ is said to be $\alpha$-Hölder continuous if $v_\alpha(\phi) < +\infty$. Denote by $\mathcal{H}_\alpha(\Sigma \times \mathbb{P})$ the space of $\alpha$-Hölder continuous observables

$$\mathcal{H}_\alpha(\Sigma \times \mathbb{P}) := \{ \phi \in L^\infty(\Sigma \times \mathbb{P}) : v_\alpha(\phi) < +\infty \}$$

and consider on this space the norm

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

These spaces form a scale of Banach algebras [6, Propositions 5.10 and 5.19]. Denote by $\mathcal{H}_\alpha(\mathbb{P})$, respectively $L^\infty(\mathbb{P})$, the Banach sub-algebras of the previous algebras formed by functions $\phi(i, \hat{x})$ which do not depend on the variable $i$.

Throughout the rest of this section let $A \in SL_2(\mathbb{R})^k$ be a diagonalizable cocycle such that $L(A) > 0$ and $p = (p_1, \ldots, p_k) \in \mathbb{R}_+^k$ be the associated probability vector underlying the Bernoulli measure $\mathbb{P}_p = P^\Sigma$. Let $N_A$ be the neighborhood of $A$ fixed in Section 3. Then Proposition 3.8 holds for all cocycles $B \in N_A$. Throughout the proofs of this section the neighborhood $N_A$ of $A$ may be shrunk in order to ensure that some relations regarding certain measurements of $A$ hold.

Given a cocycle $B \in N_A$, its projective action determines the Markov operator $Q_B : L^\infty(\Sigma \times \mathbb{P}) \to L^\infty(\Sigma \times \mathbb{P})$ defined by

$$Q_B(\phi)(i, \hat{x}) = \int_\Sigma \phi(l, \hat{B}_i \hat{x}) \, dp(l). \quad (5.1)$$

This operator is associated to the kernel $K_B : \Sigma \times \mathbb{P} \to \text{Prob}(\Sigma \times \mathbb{P})$, $K_B(i, \hat{x}) := \int_\Sigma \delta_{(l, \hat{B}_i \hat{x})} \, dp(l)$. The cocycle $B$ also determines the observable $\xi_B : \Sigma \times \mathbb{P} \to \mathbb{P}$, $\xi_B(i, \hat{x}) := \log\|B_i x\|$, and the family of Laplace-Markov operators $Q_{B,t} : L^\infty(\Sigma \times \mathbb{P}) \to L^\infty(\Sigma \times \mathbb{P})$

$$(Q_{B,t}\phi)(i, \hat{x}) := \int_\Sigma e^{t \xi_B(l, \hat{B}_i \hat{x})} \phi(l, \hat{B}_i \hat{x}) \, dp(l), \quad (5.2)$$

which includes the Markov operator at $t = 0$, i.e., $Q_{B,0} = Q_B$.

If $L(B) > 0$ and $\rho_-(B) > 0$ then the random cocycle $B$ is quasi-irreducible. By [7] Proposition 4.6 there exists $\alpha > 0$ small enough such that the Banach sub-algebra $\mathcal{H}_\alpha(\Sigma \times \mathbb{P})$ is invariant under the Markov operator $Q_B$, and moreover $Q_B$ acts on this space as a quasi-compact and simple operator. We will estimate the spectral constants.
of $Q_B$ in terms of the irreducibility measurement $N = N(B)$ (see Definition 3.1 Proposition 3.8).

By spectral continuity, for all small $t \approx 0$, the Laplace-Markov operator $Q_{B,t}$ is also a quasi-compact and simple. We will derive bounds on the spectral constants of $Q_{B,t}$ and on the size of the neighborhood of $t = 0$ where these bounds hold. Again, all these estimates are expressed in terms of the measurement $N = N(B)$.

Finally in the last part of this section we use the previous bounds on spectral constants to prove Theorem 5.1.

**Spectral constants of the Markov operator.** Take any cocycle $B \in N_A$ such that $\rho_-(B) > 0$ and write $N = N(B)$. Let $L$ be the upper bound introduced in Section 3 (not to be confused with the LE $L(B)$), before the definition of the neighborhood $N_A$, for which we have

$$\max_{1 \leq i \leq k} \max \{\|B_i\|, \|B_i^{-1}\|\} < L.$$ 

Given $0 < \alpha \leq 1$ and $n \in \mathbb{N}$ define

$$\kappa^n_\alpha(B) := \sup_{\hat{x} \neq \hat{y}} \mathbb{E} \left[ \left( \frac{d(\hat{B}^{(n)}_n \hat{x}, \hat{B}^{(n)}_n \hat{y})}{d(\hat{x}, \hat{y})} \right)^\alpha \right]$$

$$= \sup_{\hat{x} \neq \hat{y}} \sum_{j_n=1}^k \cdots \sum_{j_1=1}^{k} p_{j_n} \cdots p_{j_1} \left[ \frac{d(\hat{B}_{j_n} \cdots \hat{B}_{j_1} \hat{x}, \hat{B}_{j_n} \cdots \hat{B}_{j_1} \hat{y})}{d(\hat{x}, \hat{y})} \right]^\alpha.$$ 

This measurement plays a key role in the determination of the spectral constants of the operator $Q_B$. First, a straightforward calculation uncovers its sub-multiplicative behavior, i.e., for all $n, m \in \mathbb{N}$,

$$\kappa^{n+m}_\alpha(B) \leq \kappa^n_\alpha(B) \kappa^m_\alpha(B). \quad (5.3)$$ 

When $\alpha$ is small these quantities are bounded.

**Lemma 5.1.** If $0 < \alpha < \frac{1}{4n}$ then $\kappa^n_\alpha(B) \leq L$.

**Proof.** See [6, Lemma 5.7].

Next lemma discloses the spectral character of the numbers $\kappa^n_\alpha(B)$.

**Lemma 5.2.** For all $\phi \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P})$,

$$v_\alpha(Q^n_B \phi) \leq L^{2\alpha} \kappa^{n-1}_\alpha(B) v_\alpha(\phi).$$

**Proof.** A simple computation shows that

$$Q^n_B \phi(i, \hat{x}) = \sum_{j_n=1}^k \cdots \sum_{j_1=1}^k p_{j_n} \cdots p_{j_1} \phi(j_n, \hat{B}_{j_n-1} \cdots \hat{B}_1 \hat{B}_i \hat{x}).$$
Therefore
\[ |(Q^α_B)(i, ̂x) - (Q^α_P)(i, ̂y)| \leq \]
\[ \leq v_α(φ) \sum_{j_n=1}^k \cdots \sum_{j_1=1}^k p_{j_n} \cdots p_{j_1} \left[ \frac{d(\hat{B}_{j_n-1} \cdots \hat{B}_{j_1} \, \hat{B}_i \, \hat{B}_{j_n-1} \cdots \hat{B}_{j_1} \, \hat{B}_i \, \hat{y})}{d(\hat{x}, \hat{y})} \right]^α \]
\[ \leq v_α(φ) \mathbb{E} \left[ \left( \frac{d(\hat{B}^{(n-1)} \hat{B}_i \, \hat{B}^{(n-1)} \hat{B}_i \, \hat{y})}{d(\hat{B}_i \, \hat{y})} \right)^α \right]. \]
Hence by Remark 4.1
\[ d(\hat{B}_i \, \hat{x}, \hat{B}_i \, \hat{y}) \leq \|B_i^{-1}\|^2 d(\hat{x}, \hat{y}), \]
which implies that \( v_α(Q^α_B) \leq L^{2α} \kappa^α(B) v_α(φ). \)

The spectral bound \( \kappa^α(B) \) can be estimated as the following one variable maximum.

**Lemma 5.3.** \( \kappa^α(B) = \max_{\|x\|=1} \mathbb{E} \left[ \|B^{(n)} \, x\|^{-2α} \right] \)

**Proof.** See [7 Lemma 4.5].

**Proposition 5.4.** Let \( α = \frac{1}{4 L (\log^2 L) / N^2} \) and \( σ = \left( 1 - \frac{1}{8 L (\log^2 L) / N^2} \right)^{1/N} \).
Then the operator \( Q_B : \mathcal{H}_α(Σ × P) → \mathcal{H}_α(Σ × P) \) is quasi-compact and simple with spectral constants \( λ = 1, \ σ \) and \( C ≍ 2 L \) as \( N → +∞. \)

**Proof.** As we decrease the size of \( N_A, \ N = N(B) \) tends to \( +∞ \) and by [4.23] \( L(B)^{-1} \) is bounded away from \( +∞. \) Hence we can assume that \( N_A \) is small enough so that \( N(B) > L(B)^{-1} \) for all \( B ∈ N_A. \) By Definition 3.1, for any unit vector \( v ∈ \mathbb{R}^2, \)
\[ \frac{1}{N} \mathbb{E} \left[ \log \|B^{(N)} \, v\|^{-2} \right] = -\frac{2}{N} \mathbb{E} \left[ \log \|B^{(N)} \, v\| \right] < -L(B) \]
which implies that
\[ \mathbb{E} \left[ \log \|B^{(N)} \, v\|^{-2} \right] \leq -N L(B) < -1. \] (5.4)
A standard calculation (see the proof of [7 Proposition 4.6]) shows that for any unit vector \( v ∈ \mathbb{R}^2 \) and for any \( α > 0 \) letting \( x = α \log(\|B^{(N)} \, v\|^{-2}) \)
\[ \|B^{(N)} \, v\|^{-2α} = e^x ≤ 1 + x + \frac{x^2}{2} e^x \]
\[ ≤ 1 + α \log \|B^{(N)} \, v\|^{-2} + \frac{α^2}{2} \|B^{(N)}\|^{2α} \log^2 \|B^{(N)}\|^2 \]
\[ ≤ 1 + α \log \|B^{(N)} \, v\|^{-2} + 2 α^2 L^{2αN} N^2 \log^2 L. \]
Hence taking the expected value and using \[5.4\] we get

\[
\mathbb{E} \left[ \| B^{(N)} v \|^{-2\alpha} \right] \leq 1 - \alpha + 2\alpha^2 L^{2\alpha N} N^2 \log^2 L.
\]

Fix now \( \alpha = \frac{1}{4L(\log^2 L) N^2} \). If \( N \) is large enough then \( L^{2\alpha N} = L^{\frac{1}{2L\log^2 L} N^2} \ll L \) and so

\[
2\alpha^2 L^{2\alpha N} N^2 \log^2 L \leq \frac{1}{8L(\log^2 L) N^2} = \frac{\alpha}{2}.
\]

Therefore

\[
\mathbb{E} \left[ \log \| B^{(N)} v \|^{-2\alpha} \right] \leq 1 - \alpha + \frac{\alpha}{2} = 1 - \frac{\alpha}{2}.
\]

Thus, by Lemma \[5.3\] one has \( \kappa^N_\alpha(B) \leq 1 - \frac{\alpha}{2} \), which is equivalent to \( \kappa^N_\alpha(B) \leq \sigma^N \). Given \( n \geq 1 \) write \( n - 1 = qN + r \) with \( q \in \mathbb{N} \) and \( 0 \leq r < N \). Because \( r < N \), we have \( 0 < \alpha \ll \frac{1}{4N} < \frac{1}{N} \), which by Lemma \[5.1\] implies that \( \kappa^r_\alpha(B) \leq L \). Hence, by Lemma \[5.2\] and the sub-multiplicative relation \[5.3\] we have

\[
v_\alpha(Q^n_B \phi) \leq v_\alpha(\phi) L^{2\alpha \kappa_{\alpha}^{-1}(B)}
\]

\[
\leq v_\alpha(\phi) L^{O(\frac{1}{N})} \kappa_{\alpha}^N(B)^q \kappa_{\alpha}^{r}(B)
\]

\[
\leq v_\alpha(\phi) L^{1+O(\frac{1}{N})} \sigma^N \sigma^{r-N} \leq v_\alpha(\phi) L^{1+O(\frac{1}{N})} \sigma^{-N} 
\]

\[
\leq C_N \sigma^N v_\alpha(\phi),
\]

with \( C_N := \frac{L^{1+O(\frac{1}{N})} \sigma^{-N}}{L^{1-\frac{1}{2}}} \approx L \) as \( N \rightarrow +\infty \).

Let now \( \nu = \nu_B \in \text{Prob}(\mathbb{P}) \) be the (unique) stationary measure of the cocycle \( B \) and define \( \mu_B := (\sum_{l=1}^{k} p_l \delta_{t_l}) \times \nu_B \in \text{Prob}(\Sigma \times \mathbb{P}) \), i.e.,

\[
\int \phi d\mu_B := \sum_{l=1}^{k} p_l \int_{\mathbb{P}} \phi(l, \hat{x}) d\nu_B(\hat{x}).
\]

Consider the operator \( P : \mathcal{H}_\alpha(\Sigma \times \mathbb{P}) \rightarrow \mathcal{H}_\alpha(\Sigma \times \mathbb{P}) \) defined by \( P(\phi) := (\int \phi d\mu_B) \mathbf{1} \), where \( \mathbf{1} \) stands for the constant function 1. This operator is the spectral projection onto the eigen-space \( E_1 \subset \mathcal{H}_\alpha(\Sigma \times \mathbb{P}) \) of constant functions. It has norm 1 because

\[
\| P(\phi) \|_\alpha = \| P(\phi) \|_\infty = \| \int \phi d\mu \| \leq \| \phi \|_\infty \leq \| \phi \|_\alpha.
\]

By the spectral character of the projection \( P, Q_B \circ P = P \circ Q_B = P \). Hence, denoting the kernel of \( P \) by \( E_0 := \{ \phi \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}) : P(\phi) = 0 \} \) we get a direct sum decomposition \( \mathcal{H}_\alpha(\Sigma \times \mathbb{P}) = E_0 \oplus E_1 \) which is invariant under the Markov operator \( Q_B \).

To finish note that

\[
v_\alpha(Q^n_B \phi - P(\phi)) = v_\alpha(Q^n_B \phi) \leq C_N \sigma^n v_\alpha(\phi) \leq C_N \sigma^n \| \phi \|_\alpha.
\]
We need a similar bound on the decay of \( \| Q^n_B \phi - P(\phi) \|_{\infty} \). For that we introduce two seminorms on \( L^\infty(\Sigma \times \mathbb{P}) \).

\[
\begin{align*}
v_0(\phi) &:= \max_{1 \leq i \leq k} \sup_{\hat{x} \neq \hat{y}} |\phi(i, \hat{x}) - \phi(i, \hat{y})|, \\
\text{osc}(\phi) &:= \max_{1 \leq i, j \leq k} \sup_{\hat{x}, \hat{y} \in \mathbb{P}} |\phi(i, \hat{x}) - \phi(j, \hat{y})|.
\end{align*}
\]

It is easily seen that for all \( \phi \in L^\infty(\Sigma \times \mathbb{P}) \)

1. \( v_0(\phi) \leq v_\alpha(\phi) \) for all \( \alpha > 0 \)
2. \( \text{osc}(Q_B \phi) \leq v_0(\phi) \),
3. \( \| \phi \|_{\infty} \leq \text{osc}(\phi) \) if \( \int \phi \, d\mu = 0 \).

The first inequality holds because \( P \) has diameter 1. For the second note that

\[
\begin{align*}
|Q_B(i, \hat{x}) - Q_B(j, \hat{y})| &\leq \sum_{l=1}^{k} p_l |\phi(l, \hat{B}_i \hat{x}) - \phi(l, \hat{B}_j \hat{y})| \\
&\leq \sum_{l=1}^{k} p_l v_0(\phi) = v_0(\phi).
\end{align*}
\]

Finally the third inequality holds because

\[
|\phi(i, \hat{x})| = |\phi(i, \hat{x}) - \int \phi \, d\mu| \\
\leq \int |\phi(i, \hat{x}) - \phi(j, \hat{y})| \, d\mu(j, \hat{y}) \leq \text{osc}(\phi).
\]

Using these three inequalities we have

\[
\begin{align*}
\| Q^n_B \phi - P(\phi) \|_{\infty} &\leq \text{osc}(Q^n_B \phi - P(\phi)) = \text{osc}(Q_B(Q^{n-1}_B \phi - P(\phi))) \\
&\leq v_0(Q^{n-1}_B \phi - P(\phi)) \leq v_\alpha(Q^{n-1}_B \phi - P(\phi)) \\
&\leq C_N \sigma^{n-1} v_\alpha(\phi) = \sigma^{-1} C_N \sigma^n v_\alpha(\phi).
\end{align*}
\]

Therefore

\[
\begin{align*}
\| Q^n_B \phi - P(\phi) \|_{\alpha} &\leq C_N (1 + \sigma^{-1}) \sigma^n v_\alpha(\phi) \\
&\leq C_N (1 + \sigma^{-1}) \sigma^n \| \phi \|_{\alpha},
\end{align*}
\]

with \( C_N (1 + \sigma^{-1}) \approx 2 L \). \( \square \)
Spectral constants of the Laplace-Markov operator. Consider a cocycle $B \in \mathcal{N}_A$ such that $\rho_+(B) > 0$ and write $N = N(B)$. From the previous section, the spectral constants of the Markov operator $Q = Q_B$ acting on $\mathcal{H}_\alpha(\Sigma \times \mathbb{P})$ are $\lambda_M = 1$, $\sigma_M = (1 - \frac{\omega}{2})^{1/N}$ and $C_M < 3L$, where $\alpha = \frac{1}{4L(\log^2 L) N^2}$. The Laplace-Markov operator $Q_t = Q_{B,t}$ is a positive operator, i.e., $Q_t(\phi) \geq 0$ whenever $\phi \geq 0$. Hence for $t \approx 0$ it has a positive maximal eigenvalue $\lambda(t) = 1 + O(t)$ ($t \to 0$).

In this section we use a continuity argument to derive the spectral constants for $Q_t$ on the same space $\mathcal{H}_\alpha(\Sigma \times \mathbb{P})$. The subscript $M$ in the spectral constants $\lambda_M, \sigma_M$ and $C_M$ of the Markov operator $Q_B$ stands for ‘Markov’. Likewise we will use a subscript $L$ for ‘Laplace’ for the spectral constants $\lambda_L, \sigma_L$ and $C_L$ of the Laplace-Markov operator $Q_{B,L}$. By simply applying [6, Proposition 5.12] with $\varepsilon = \frac{1 - \sigma}{3}$, so that $\sigma < \sigma + \varepsilon < 1 - \varepsilon < 1$, one gets spectral constants for $Q_t$ of the following magnitude: $\lambda_L > 1 - \varepsilon$, $\sigma_L < \sigma + \varepsilon$ and $C_L < (c_1 N^4)^{c_0} N^2$, for some positive constants $c_0$ and $c_1$ depending only on the cocycle $A$.

Note that the bound on $C_L$ grows super-exponentially as $N \to +\infty$. Even worse, these bounds hold only for exponentially small parameters $t$, i.e., for $|t| \ll (c_1 N^4)^{-c_0} N^2$. Therefore these bounds are far from enough to prove Theorem 5.1. The strategy to overcome this difficulty is a trade-off between the bounds $\sigma_L$ and $C_L$. Allowing $\sigma_L$ to be much closer to 1 we obtain for $C_L$ a polynomial bound in $N$, which holds for polynomially small parameters $t$.

The distance from $\sigma$ to 1 will be a reference measurement, asymptotically (as $N \to +\infty$) given by

$$1 - \sigma \asymp -\log \sigma = -\frac{1}{N} \log (1 - \frac{1}{8L(\log^2 L) N^2}) \approx \frac{1}{8L(\log^2 L) N^2}. \quad (5.5)$$

The next proposition provides a tool for implementing this trade-off.

Proposition 5.5. Given $0 < C < \infty$ and $\frac{2}{3} \leq \sigma < 1$, there exist constants $C' \lesssim \frac{c^2}{(1 - \sigma)^2}$ and $\beta_0 \gtrsim \frac{(1 - \sigma)^4}{c^2}$ such that given $0 < \beta < \beta_0$ if \{c'\}_n \geq 0 is a sequence with $c'_0 = C$ and for all $n \geq 1$

$$0 < c'_n \leq C \sigma^n + \beta (c'_0 c'_{n-1} + c'_1 c'_{n-2} + \ldots + c'_{n-1} c'_0)$$

then for all $n \geq 0$,

$$c'_n \leq C \sigma^n + C' \beta \sigma^{\sqrt{n}} < C' \sigma^{\sqrt{n}}. \quad (5.5)$$

Proof. The proof is based on Lemma 5.7 below. Let us set

$$c_n := C \sigma^n,$$

$$\hat{c}_n := c_n + \beta (c_0 c_{n-1} + c_1 c_{n-2} + \ldots + c_{n-1} c_0).$$
Define also (recursively) \( c_0^* := C \) and for all \( n \geq 1 \),
\[
c_n^* := c_n + \beta \left( c_0^* c_{n-1}^* + c_1^* c_{n-2}^* + \ldots + c_{n-1}^* c_0^* \right) .
\]
By induction we have \( c_n' \leq c_n^* \) for all \( n \geq 0 \). Consider the constant

\[
C' := \frac{6C^2}{(1 - \sigma)^2} .
\]

We claim that for all \( j \in \mathbb{N} \),
\[
|c_j^* - c_j| \leq C' \beta \sigma \sqrt{j} .
\]

This implies that
\[
c_j' \leq c_j^* \leq |c_j^* - c_j| + |c_j| \leq (C' \beta + C) \sigma \sqrt{j} < C' \sigma \sqrt{j},
\]
where the last inequality holds because
\[
C' \beta + C < C' \quad \text{something obvious for } \beta > 0 \text{ small enough. Thus claim } (5.7) \text{ implies this proposition.}
\]

Let us now prove the claim. Assume, by induction hypothesis, that (5.7), and in particular (5.8), hold for all \( 0 \leq j \leq n \). Using Lemma 5.7, we have
\[
|c_{n+1}^* - c_{n+1}| \leq |c_{n+1}^* - \hat{c}_{n+1}| + |\hat{c}_{n+1} - c_{n+1}|
\]
\[
\leq \beta \sum_{j=0}^{n} |c_j^* c_{n-j}^* - c_j c_{n-j}| + \beta \sum_{j=0}^{n} c_j c_{n-j}
\]
\[
\leq \beta \sum_{j=0}^{n} \left\{ c_j^* c_{n-j}^* - c_{n-j}^* - c_{n-j} + c_{n-j} |c_j^* - c_j| \right\} + \beta \sum_{j=0}^{n} C^2 \sigma^n
\]
\[
\leq 2 \beta^2 C'^2 \sum_{j=0}^{n} \sigma \sqrt{j+\sqrt{n-j}} + \beta C^2 (n+1) \sigma^n
\]
\[
\leq \frac{40 \beta^2 C'^2}{(1 - \sigma)^2} \sigma \sqrt{n} + \beta C^2 (n+1) \sigma^{n/8} \sigma \sqrt{n+1}
\]
\[
\leq \frac{\beta^2}{\sigma (1 - \sigma)^2} \frac{40 C'^2}{(1 - \sigma)^2} \sigma \sqrt{n+1} + \beta C^2 \frac{3 C^2}{\sigma^{1/8} (1 - \sigma)} \sigma \sqrt{n+1}
\]
\[
\leq \beta \frac{40 \beta C'^2 + 3 C^2}{\sigma (1 - \sigma)^2} \sigma \sqrt{n+1} < \beta C' \sigma \sqrt{n+1}.
\]

We have used above the following inequalities
\[
(1) \frac{n}{8} + \sqrt{n+1} \leq n, \quad \forall n \geq 2
\]
\[(2) \quad \sigma^{\sqrt{n}} \leq \frac{\sigma^{\sqrt{n+1}}}{\sigma}, \quad \forall \ n \geq 0,\]
\[(3) \quad (n+1)\sigma^{n/8} \leq \frac{-8}{e\sigma^{1/8}\log\sigma} < \frac{3}{\sigma^{1/8}(1-\sigma)}, \quad \forall \ n \geq 0,\]
\[(4) \quad \frac{40\beta(C')^2 + 3C^2}{(1-\sigma)^2} < \sigma C'.\]

The first two inequalities are clear. Inequality (3) is obtained computing the maximum of the function \(f(x) = (x + 1)\sigma^{x/8}\) over the half-line \([0, +\infty[.\)

Finally, since \(\frac{2}{3} \leq \sigma < 1\) we have
\[(4) \iff 40\beta C'^2 + 3C^2 < 6\sigma C^2 \iff 40\beta C'^2 < C^2\]
\[\iff 40 \times 36\beta C^4 < C^2(1-\sigma)^4\]
\[\iff 1440\beta C^2 < (1-\sigma)^4\]
\[\iff \beta < \beta_0 := \frac{(1-\sigma)^4}{1440C^2}.\]

Hence inequality (4) holds for all \(0 < \beta < \beta_0.\) \(\square\)

**Lemma 5.6.** Given \(0 < \sigma < 1,\)
\[\sum_{n=0}^{\infty} \sigma^{\sqrt{n}} \leq \frac{2}{(1-\sigma)^2} + \frac{1}{1-\sigma}.\]

**Proof.** Given \(n \in \mathbb{N}, \lfloor \sqrt{n} \rfloor = j \iff j^2 \leq n < (j+1)^2,\) so that
\[\left| \{n \in \mathbb{N} : \lfloor \sqrt{n} \rfloor = j \} \right| = (j+1)^2 - j^2 = 2j + 1.\]
Thus
\[\sum_{n=0}^{\infty} \sigma^{\sqrt{n}} \leq \sum_{n=0}^{\infty} \sigma^{\lfloor \sqrt{n} \rfloor} = \sum_{j=0}^{\infty} \left| \{n \in \mathbb{N} : \lfloor \sqrt{n} \rfloor = j \} \right| \sigma^j\]
\[= \sum_{j=0}^{\infty} (2j + 1) \sigma^j = \frac{2\sigma}{(1-\sigma)^2} + \frac{1}{1-\sigma}.\]
\(\square\)

**Lemma 5.7.** Given \(0 < \sigma < 1\)
\[\sum_{j=0}^{n} \sigma^{\sqrt{j} + \sqrt{n-j} - \sqrt{n}} \leq \frac{20}{(1-\sigma)^2}, \quad \text{for all } n \geq 1.\]

**Proof.** For \(n = 1\) the sum equals 2 and the inequality is obvious. From now on we assume that \(n \geq 2.\) For \(0 \leq j \leq n,\) one has \(\sqrt{j} \sqrt{n-j} \leq \frac{n}{2},\)
which implies that \( \sqrt{n + 2 \sqrt{j} \sqrt{n - j}} \leq \sqrt{2n} \), and hence that

\[
\frac{1}{\sqrt{2n}} \leq \frac{1}{\sqrt{n + 2 \sqrt{j} \sqrt{n - j}}} \leq \frac{2}{\sqrt{n + 2 \sqrt{j} \sqrt{n - j} + \sqrt{n}}}
\]

Given \( 0 < \sigma < 1 \) and using the previous inequality we get

\[
\sum_{j=0}^{n} \sigma^{\sqrt{j+n-j} - \sqrt{n}} = \sum_{j=0}^{n} \sigma^{\sqrt{(\sqrt{j+n-j})^2 - \sqrt{n}}}
\]

\[
= \sum_{j=0}^{n} \sigma^{\sqrt{n+2 \sqrt{j} \sqrt{n-j} - \sqrt{n}}}
\]

\[
= \sum_{j=0}^{n} \sigma^{\sqrt{n+2 \sqrt{j} \sqrt{n-j} + \sqrt{n}}}
\]

\[
\leq \sum_{j=0}^{n} \sigma^{\sqrt{n+2 \sqrt{n}} / \sqrt{2n}} \leq 2 \sum_{j=0}^{n/2} \sigma^{\sqrt{(n-j) / 2n}}.
\]

For \( 0 \leq j \leq \frac{n}{2} \), we have \( \frac{n-j}{2n} \geq \frac{1}{4} \). Hence

\[
\sqrt{\frac{j (n-j)}{2n}} \geq \sqrt{\frac{j}{4}}.
\]

Thus by Lemma 5.6

\[
\sum_{j=0}^{n/2} \sigma^{\sqrt{(n-j) / 2n}} \leq \sum_{j=0}^{n/2} \sigma^{\sqrt{\frac{1}{4}}} \leq \sum_{j=0}^{\infty} \sigma^{\sqrt{\frac{1}{4}}} \leq \frac{2}{(1-\sigma^2)} + \frac{1}{1-\sigma^2}
\]

On the other hand

\[
\frac{1}{1-\sigma^2} = \frac{1+\sigma^2}{1-\sigma} \leq \frac{2}{1-\sigma}.
\]

Therefore

\[
\sum_{j=0}^{n} \sigma^{\sqrt{j+n-j} - \sqrt{n}} \leq 2 \sum_{j=0}^{n/2} \sigma^{\sqrt{(n-j) / 2n}}
\]

\[
\leq \frac{16}{(1-\sigma)^2} + \frac{4}{1-\sigma} \leq \frac{20}{(1-\sigma)^2}.
\]

Finally we establish the spectral constants of the Laplace-Markov family of operators.
Proposition 5.8. There are positive constants $\sigma_L$, $\lambda_L$, $C_L$ and $b_L$ with the following properties. For every $z \in D(0)$ the space $\mathcal{H}_\alpha(\Sigma \times \mathbb{P})$ admits a one dimensional subspace $E_z$ and a co-dimension 1 subspace $H_z$, there exist a number $\lambda(z) \in \mathbb{C}$ and a linear map $P_z \in L(\mathcal{H}_\alpha(\Sigma \times \mathbb{P}))$ such that for all $f \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P})$,

1. $b_L = N^{-20}$, $1 - \sigma_L \asymp N^{-8}$, $1 - \lambda_L \asymp N^{-10}$ and $C_L \asymp N^6$ as $N \to +\infty$.
2. $\mathcal{H}_\alpha(\Sigma \times \mathbb{P}) = E_z \oplus H_z$ is a $Q_z$-invariant decomposition,
3. $P_z$ is a projection onto $E_z$, parallel to $H_z$,
4. $Q_z(P_z f) = P_z(Q_z f) = \lambda(z) P_z f$,
5. $Q_z f = \lambda(z) f$ if $f \in E_z$,
6. $z \mapsto \lambda(z)$ is analytic in a neighborhood of $D(0)$,
7. $|\lambda(z)| \geq \lambda_L$,
8. $|\lambda'(z)| \leq 2C_L$,
9. $|\lambda''(z)| \leq 4b_L^{-1}C_L$,
10. $\|Q_z^n f - \lambda(z)^n P_z f\| \leq C_L \sigma^{n/2} \|f\|$,
11. $\|Q_z^n f - \lambda(z)^n P_z f\| \leq 2C_L (\sigma_L)^n \|f\|$,
12. there is a spectral gap: $\sigma_L < \lambda_L^2$,
13. $\|P_z f\| \leq C_L \|f\|$,
14. $\|P_z f - P_0 f\| \leq C_L |z| \|f\|$.

Proof. Consider the abstract setting discussed in [6, Section 5.2.1]. The scale of Banach algebras $\mathcal{H}_\alpha(\Sigma \times \mathbb{P})$, with $\alpha \in [0, 1]$, satisfies assumptions (B1)-(B7).

For each cocycle $B \in N_A$ consider the stochastic kernel $K_B$, the corresponding stationary measure $\mu_B$ on $\Sigma \times \mathbb{P}$ and the observable $\xi_B: \Sigma \times \mathbb{P} \to \mathbb{R}$,

$$\xi_B(i, \hat{x}) := \log \|B_i x\|$$

where $x$ stands for a unit representative of $\hat{x} \in \mathbb{P}$. In [6] we refer to the tuple $(K_B, \mu_B, \xi_B)$ as an observed Markov space on $\Sigma \times \mathbb{P}$. Denote by $\mathfrak{X}$ the space of all observed Markov spaces of the form $(K_B, \mu_B, \pm \xi_B)$ with $B \in N_A$ and $\rho_-(B) > 0$. This space satisfies assumptions (A1)-(A3) in [6, Section 5.2.1].

Assumption (A1) follows automatically from the definition of $\mathfrak{X}$. Assumption (A2) is a consequence of Proposition 5.4 with the constants therein. Let us now focus on assumption (A3). A simple calculation shows that $\|\xi_B\|_{\infty} \leq \log L$ and

$$\left| \log \|B_i x\| - \log \|B_i y\| \right| \leq \frac{\|B_i x\| - \|B_i y\|}{L^{-1}} \leq L \|B_i x - B_i y\| \leq L^2 \|x - y\| \leq \sqrt{2} L^2 d(\hat{x}, \hat{y})$$
which implies that \( v_\alpha(\xi_B) \leq v_1(\xi_B) \leq \sqrt{2}L^2 \). Therefore, \( \|\xi_B\|_\alpha < 2L^2 \).

The family of Laplace-Markov operators \( Q_z : \mathcal{H}_\alpha(\Sigma \times P) \rightarrow \mathcal{H}_\alpha(\Sigma \times P) \) in [5.2] can be written as \( Q_z(f) = \mathcal{Q}_0(f e^{z\xi_B}) \) and since \( \mathcal{H}_\alpha(\Sigma \times P) \) is a Banach algebra, the map \( C \ni z \mapsto Q_z \in \mathcal{L}(\mathcal{H}_\alpha(\Sigma \times P)) \) is an entire (analytic) function. By Proposition 5.4, the family of Laplace-Markov operators

\[
\mathcal{Q}_0 f \quad \text{with} \quad 1 + C' \sigma \approx 2L + 1. \quad \text{Hence, for } i = 0, 1, 2,
\]

\[
\|Q_z(f \xi_B^i)\|_\alpha = \|Q_0(f \xi_B^i e^{z\xi_B})\|_\alpha \leq (2L + 1) \|f \xi_B^i e^{z\xi_B}\|_\alpha
\]

\[
\leq (2L + 1) \|\xi_B\|_\alpha \|e^{z\xi_B}\|_\alpha \|f\|_\alpha
\]

\[
\leq (2L + 1) (2L^2)^2 e^{2|z|L^2} \|f\|_\alpha.
\]

Therefore, setting \( b = (10L^2)^{-1} \) and \( M = 10L^5 \), we have for all \( z \in D_b \)

\[
\|Q_z(f \xi_B^i)\|_\alpha \leq 8e^{1/5}L^5 \|f\|_\alpha \leq M \|f\|_\alpha
\]

which proves (A3).

Item (1) will follow from the choices made below.

Except (7)-(12), all other items of this proposition follow from [6, Proposition 5.12 and Remark 5.3]. Actually, apart from (7)-(12), everything else follows from a general spectral continuity argument and the quantitative lemma [6, Lemma 5.2]. Items (13) and (14) for instance hold with \( C_L = (3 + 3C)^2 (1 - \sigma)^{-1} M \), with \( C = 2L \) and \( M = 10L^5 \). Using the asymptotic expression (5.5) for \( 1 - \sigma \) we see that we can take \( C_L = \kappa_0 L^{10} N^6 \), where \( \kappa_0 \) stands for some absolute constant. Moreover, these bounds hold for all \( |z| \leq L^{-10} N^{-6} \).

Let us now prove (10). From now on we will focus our analysis of the family of operators \( Q_z \) on the smaller disk \( D_{4b_L}(0) \) with radius \( 4N^{-20} \), i.e., \( b_L := N^{-20} \). Consider the family of operators \( N_z := Q_z - \lambda(z) P_z \). The same lemma, [6, Lemma 5.2], invoked to justify (14) shows that \( \|\frac{d}{dz} N_z\| \leq C_L \) for all \( z \in D_{4b_L}(0) \).

Defining

\[
c^j := \sup_{z \in D_{4b_L}(0)} \|N_z^j\| \quad (j \geq 0)
\]

the previous inequality shows that

\[
\|\frac{d}{dz} (N_z^n)\| \leq \sum_{j=1}^n \|N_z^{n-j}\| \|\frac{d}{dz} N_z\| \|N_z^{j-1}\| \leq C_L \sum_{j=1}^n c^j c^j_{n-j}. \]

Next set \( c_j := C' \sigma^j \), so that \( \|N_z^j\| \leq c_j \). Then, for all \( z \in D_{4b_L}(0) \),

\[
\|N_z^n\| \leq \|N_z^n\| + 4C_L b_L \sum_{j=1}^n c^j c^j_{n-j} \leq c_n + 4C_L b_L \sum_{j=1}^n c^j c^j_{n-j}.
\]
Setting $\beta := 4 C_L b_L$, this implies that
\[ c'_n \leq c_n + \beta (c'_0 c'_{n-1} + c'_1 c'_{n-2} + \ldots + c'_{n-1} c'_0). \]
By Proposition 5.5, since for large $N$ we have that
\[ \beta = 4 C_L b_L \approx \frac{L^{10} N^6}{N^{20}} \ll \frac{1}{2^{14} L^6 (\log^8 L) N^{12}} = \frac{(1 - \sigma)^4}{C^2} =: \beta_0 \]
we have for all $n \geq 0$ and $z \in \mathcal{D}_{4b_L}(0)$,
\[ \|Q^n - \lambda(z)^n P_z\| = \|N^n\| \leq c'_n \leq C' \sigma \sqrt{n} \leq C_L \sigma \sqrt{n} \]
which proves (10). Note that increasing if necessary the value of $\kappa_0$ we ensure that
\[ C' \approx C^2 (1 - \sigma)^{-2} = 2^8 L^4 (\log^4 L) N^6 < \kappa_0 L^{10} N^6 = C_L. \]
Consider the numbers
\[ \sigma_L := \exp\left(-\frac{1}{N^8}\right) \quad \text{and} \quad \lambda_L := \exp\left(-\frac{1}{N^{10}}\right) \]
which satisfy $1 - \sigma_L \approx N^{-8}$ and $1 - \lambda_L \approx N^{-10}$. With these definitions item (12) holds because $2 N^{-10} \ll N^{-8}$.
To prove item (11) consider the integer $n_0 = N^8$. The inequality
\[ C' \sigma^{\sqrt{n_0}} = C' \exp\left(-\sqrt{n_0} \log \sigma^{-1}\right) \leq \exp\left(-\frac{n_0}{N^8}\right) = \sigma^{n_0}_L \]
holds because
\[ \sqrt{n_0} \log \sigma^{-1} \approx \frac{N^4}{N^8} \approx 1 \sim n_0 = N^8. \]
Hence $\|N^{n_0}_z\| \lesssim \sigma^{n_0}_L$, for all $z \in \mathcal{D}_{4b_L}(0)$. Given $n \geq 1$, performing an integer division we write $n = q n_0 + r$ with $0 \leq r < n_0$. Thus, by item (10) we get for all $z \in \mathcal{D}_{4b_L}(0)$
\[ \|N^n_z\| \leq \|N^n_z\| \|N^{n_0}_z\|^q \lesssim C_L (\sigma_L)^{q n_0} \lesssim C_L (\sigma_L)^{n_0} \lesssim C_L (\sigma_L)^n \]
which proves (11). Notice that $\sigma^{n_0}_L = e^{-1} \approx 1$.
We focus now on the proof of items (7)-(9). Consider the family of operators $L_z = Q_z \circ P_z$. By [6, Lemma 5.2], for all $z \in \mathcal{D}_{4b_L}(0)$
\[ \|L_z - L_0\| \leq C_L |z|. \]
Let $1$ denote the constant function 1 on $\Sigma \times \mathbb{P}$ and $\mu_B$ be the stationary probability measure on $\Sigma \times \mathbb{P}$. Then for all $z \in \mathcal{D}_{4b_L}(0)$,
\[ \lambda(z) = \frac{\langle L_z 1, \mu_B \rangle}{\langle P_z 1, \mu_B \rangle}. \]
The denominator above does not vanish because
\[ \langle P_1, \mu_B \rangle \geq \langle P_0, \mu_B \rangle - \|P_0 - P_1\|_\infty \geq 1 - C_L b_L \approx 1. \]

Hence, for all \( z \in \mathbb{D}_{4b_L}(0) \),
\[
|\lambda(z) - 1| \lesssim |\langle L_1, \mu_B \rangle - \langle P_1, \mu_B \rangle| \\
\leq |\langle L_1 - L_0, \mu_0 \rangle| + |\langle P_1 - P_0, \mu_0 \rangle| \\
\leq 2 C_L b_L \approx N^{-14} \ll N^{-10}.
\]

This proves (7).
By Cauchy’s integral formula, for all \( z \in \mathbb{D}_{2b_L}(0) \)
\[
|\lambda'(z)| \leq \frac{1}{2\pi} \int_{|\zeta - z| = 4b_L} \frac{|\lambda(\zeta) - 1|}{|\zeta - z|^2} |d\zeta| \leq \frac{8\pi b_L}{2\pi} \frac{2C_L b_L}{(2b_L)^2} = 2C_L,
\]
which proves (8). Similarly, for all \( z \in \mathbb{D}_{b_L}(0) \)
\[
|\lambda''(z)| \leq \frac{1}{2\pi} \int_{|\zeta - z| = 2b_L} \frac{|\lambda'(\zeta)|}{|\zeta - z|^2} |d\zeta| \leq \frac{4\pi b_L}{2\pi} \frac{2C_L}{b_L^2} = 4b_L^{-1} C_L,
\]
which proves (9). \( \square \)

**Proving large deviation estimates.**

**Proposition 5.9.** Given a cocycle \( A \in \text{Diag} \) with \( L(A) > 0 \) there exists a neighborhood \( N_A \) of \( A \) in \( C = \text{SL}_2(\mathbb{R})^k \) such that for any cocycle \( B \in N_A \) with \( \rho_-(B) > 0 \), if \( N = N(B) \) there are positive constants \( h \approx N^{20} \) and \( C \approx N^6 \) such that for all \( 0 < \epsilon < 1 \) and \( n \geq 1 \),
\[
\mathbb{P} \left[ \frac{1}{n} \log \|B(n)\| - L(B) \right] > \epsilon \right] \leq C e^{-\epsilon^2 n}.
\]

**Proof.** The proof of this result reduces to [7, Theorem 4.1] or [6, Theorem 5.3]. We outline it here for the reader’s convenience.

As before let \( X = \Sigma^\mathbb{Z} \) and \( T : X \to X \) denote the shift homeomorphism. Define the skew-product map \( F_B : X \times \mathbb{P} \to X \times \mathbb{P} \) by \( F_B(x, \hat{p}) := (Tx, \hat{B}_{x_0} \hat{p}) \). Recall that \( \nu_B \in \text{Prob}(\mathbb{P}) \) stands for the (unique) stationary measure of the cocycle \( B \) on \( \mathbb{P} \). Then \( \mathbb{P}_B = (\hat{p}^\infty) \times \nu_B \) is an \( F_B \)-invariant probability measure on \( X \times \mathbb{P} \). Next consider the random process \( e_n : X \times \mathbb{P} \to \Sigma \times \mathbb{P} \), \( e_n(x, \hat{p}) := (x_n, \hat{B}^{(n)}(x) \hat{p}) \), which satisfies \( e_n = e_0 \circ F_B^n \) for all \( n \geq 0 \). Given any probability \( \theta \in \text{Prob}(\Sigma \times \mathbb{P}) \) there is a unique probability measure \( \mathbb{P}_{\theta} \in \text{Prob}(X \times \mathbb{P}) \) such that for any measurable set \( E \subset \Sigma \times \mathbb{P} \),
\begin{enumerate}
  \item \( \mathbb{P}_{\theta}[e_0 \in E] = \theta(E) \),
  \item \( \mathbb{P}_{\theta}[e_{n+1} \in E | e_n = (i, \hat{p})] = K_B((i, \hat{p}), E) \).
\end{enumerate}
We will refer to $P_\theta$ as the Kolmogorov extension of $\theta$. Because $\mu_B = \left( \sum_{l \in \Sigma} p_l \delta_l \right) \times \nu_B$ is the stationary measure of the stochastic kernel $K_B$, the measure $P_B$ is the Kolmogorov extension of $\theta = \mu_B$. We denote by $P_{(i, \hat{p})}$ the Kolmogorov extension of a Dirac mass $\theta = \delta_{(i, \hat{p})}$. Averaging these Dirac masses we recover the invariant probability

$$P_B = \int_{\Sigma \times \mathcal{P}} P_{(i, \hat{p})} d\mu_B(i, \hat{p}). \quad (5.9)$$

Let $\xi_B : \Sigma \times \mathcal{P} \to \mathbb{R}$ be the observable defined by $\xi_B(i, \hat{p}) := \log \|B_i p\|$ where $p$ stands for a unit vector representative of $\hat{p}$. This observable determines the random process $\eta_n : X \times \mathcal{P} \to \mathbb{R}$, $\eta_n := \xi_B \circ e_n$, and a corresponding sum process $S_n = \sum_{j=0}^{n-1} \eta_j$. For all $(x, \hat{p}) \in X \times \mathcal{P},$

$$S_n(x, \hat{p}) = \sum_{j=0}^{n-1} \log \|B_{x_j} \ldots B_{x_0} p\| = \log \|B^{(n)}(x) p\| \quad (5.10)$$

where $p$ is a unit vector representative of $\hat{p}$. This relates the additive process $S_n$ to the sub-additive one $\log \|B^{(n)}\|$. The process $\{\eta_n\}_{n \geq 0}$ is stationary w.r.t. $P_B$ and by Furstenberg’s formula we have

$$\mathbb{E}[\eta_n] = \int_{\Sigma \times \mathcal{P}} \xi_B d\mu_B = L(B).$$

In particular $\frac{1}{n} \mathbb{E}(S_n) = L(B)$ for all $n \geq 1$. We may assume that $L(B) = 0$ (otherwise we work with the GL-cocycle $\tilde{B}$ obtained by multiplying $B$ with the factor $e^{-L(B)}$).

By Chebyshev’s inequality we get the following upper bound

$$\mathbb{P}_{(i, \hat{p})} \left[ \frac{1}{n} S_n \geq \varepsilon \right] \leq e^{-nt\varepsilon} \mathbb{E}_{(i, \hat{p})}[e^{tS_n}] \quad (5.11)$$

where $t$ and $\varepsilon$ are positive parameters and $\mathbb{E}_{(i, \hat{p})}$ stands for expectation w.r.t. the probability measure $P_{(i, \hat{p})}$. The right-hand-side expected value can be expressed in terms of the operator $Q_{B,t}$:

$$\mathbb{E}_{(i, \hat{p})}[e^{tS_n}] = (Q_{B,t}^n 1)(i, \hat{p}).$$

For $t \in \mathbb{R}$, the Laplace-Markov operators $Q_{B,t}$ are positive operators on the Banach lattice $H_\alpha(\Sigma \times \mathcal{P})$ with $\alpha = (4L(\log^2 L)N^2)^{-1}$. By items (7) and (11) of Proposition 5.8 they are quasi-compact and simple operators for all $0 < t < b_L = N^{-20}$. Hence $Q_{B,t}$ has a unique positive maximal eigenvalue $\lambda(t)$ which depends analytically on $t \in \mathbb{R}$. The
uniform spectral decomposition of this family of operators implies that

$$(Q^n_{B,t} \mathbf{1})(i, \hat{p}) = \lambda(t)^n + |(Q^n_{B,t} \mathbf{1})(i, \hat{p}) - \lambda(t)^n|$$

$$\leq \lambda(t)^n + \|Q^n_{B,t} \mathbf{1} - \lambda(t)^n \mathbf{1}\|_\alpha$$

$$\leq \lambda(t)^n + \lambda(t)^n \|P_t \mathbf{1} - \mathbf{1}\|_\alpha + \|Q^n_{B,t} \mathbf{1} - \lambda(t)^n P_t \mathbf{1}\|_\alpha$$

$$\leq \lambda(t)^n (1 + C_L |t|) + C_L \sigma_L^2 \leq 2 C_L \lambda(t)^n.$$ 

The last inequality holds because $\sigma_L < 1 \leq \lambda(t)$, a claim proven below.

Hence

$$\mathbb{E}_{(i, \hat{p})}[e^{t S_n}] = (Q^n_{B,t} \mathbf{1})(i, \hat{p}) \leq 2 C_L \lambda(t)^n = 2 C_L e^{n c(t)} \quad (5.12)$$

with $c(t) := \log \lambda(t)$ for $t \not\in \mathbb{D}_{b}\lambda(0)$. The function $c(t)$ satisfies $c(0) = 0$ because $\lambda(0) = 1$, and $c'(0) = 0$ because $L(B) = 0$. Now, $\lambda(t) \geq 1$ is equivalent to $c(t) \geq 0$ which follows by Jensen’s inequality and the concavity of the logarithmic function. In fact

$$c(t) = \lim_{n \to +\infty} \frac{1}{n} \log \|Q^n_{B,t}\| \geq \lim_{n \to +\infty} \frac{1}{n} \log \|Q^n_{B,t} \mathbf{1}\|_\infty$$

$$\geq \lim_{n \to +\infty} \frac{1}{n} \log \mathbb{E}_{(i, \hat{p})}[e^{t S_n}] \geq \lim_{n \to +\infty} \frac{1}{n} \mathbb{E}_{(i, \hat{p})}[t S_n] = 0.$$ 

Note that by [11 Lemma 4.3], $\frac{1}{n} \mathbb{E}_{(i, \hat{p})}[S_n] = \frac{1}{n} \mathbb{E} [\log \|B^{(n)} \lambda B_{i,\hat{p}}\|] \| B_{i,\hat{p}} \|$ converges to $L(B) = 0$ uniformly in $\hat{p}$.

By item (9) of Proposition 5.8 we have

$$h := \max_{t \in \mathbb{D}_{b}\lambda(0)} |c''(t)| \leq 4 b^{-1}_L \varepsilon \mathcal{N}^2.$$ 

For all $0 < t < b_L$ one has $c(t) \leq \frac{h}{2} t^2$ and combining (5.11) with (5.12)

$$\mathbb{P}_{(i, \hat{p})} \left[ \frac{1}{n} S_n \geq \varepsilon \right] \leq 2 C_L e^{-n \left( t \varepsilon - \frac{h}{2} t^2 \right)}.$$ 

Choosing the value of $t$ which minimizes the right-hand-side of the previous inequality leads to the following upper bound

$$\mathbb{P}_{(i, \hat{p})} \left[ \frac{1}{n} S_n \geq \varepsilon \right] \leq 2 C_L e^{-n \hat{\varphi}(\varepsilon)} \text{ with } \hat{\varphi}(\varepsilon) := \max_{|t| < b_L} \left( t \varepsilon - \frac{h}{2} t^2 \right)$$

where $\hat{\varphi}(\varepsilon)$ is the Legendre transform of the quadratic function $\varphi : (-b_L, b_L) \to \mathbb{R}$, $\varphi(t) := \frac{h}{2} t^2$. A simple calculation shows that $\hat{\varphi}(\varepsilon) = \frac{\varepsilon^2}{2h}$ for all $|\varepsilon| < h b_L \approx \mathcal{N}^4$. Averaging these inequalities in $(i, \hat{p})$ w.r.t. $\mu_B$, through (5.9) we get for all $n \geq 1$ and $0 < \varepsilon < \mathcal{N}^4$

$$\mathbb{P}_B \left[ \frac{1}{n} S_n \geq \varepsilon \right] \leq 2 C_L e^{-\frac{\varepsilon^2}{2h}}.$$
This provides an estimate for deviations of \( \frac{1}{n} S_n \) above the average. Applying the same method to the symmetric process, the same bound holds for deviations below the average. We have assumed so far that \( L(B) = 0 \). In the general case by (5.10) we have for all \( n \geq 1 \)
\[
\mathbb{P}_B \left[ \left| \frac{1}{n} \log \| B^{(n)} p \| - L(B) \right| \geq \varepsilon \right] \leq 4 C_L e^{-n \varepsilon^2 / 2n}.
\]
Finally, a standard argument allows us to substitute in the previous LDT estimate \( n^{-1} \log \| B^{(n)} p \| \) by \( n^{-1} \log \| B^{(n)} \| \), see the proof of [7, Theorem 4.1]. This concludes the proof.

**Proof of Theorem 5.1.** Take \( \varepsilon = n^{-1/6} \) with \( n \geq \left( c \log \rho_-(B)^{-1} \right)^{5/4} \).
By Proposition 3.8 this implies that \( n \geq N^{5/4} \). Then by Proposition 5.9, the event \( \Delta_n := \left[ n^{-1} \log \| B^{(n)} \| - L(B) \right] > n^{-1/6} \) has probability
\[
\mathbb{P}(\Delta_n) \leq N^6 e^{-\frac{n^{-1/3} N^{-26}}{2} n} \leq n^{1/9} e^{-\frac{1}{2} n^{1/3} \frac{1}{4} \frac{26}{34}}
\]
\[
\leq n^{1/9} e^{-\frac{1}{2} n^{1/3} \frac{10}{34}} \ll e^{-n^{1/8}}
\]
The last inequality holds for all sufficiently large \( n \), so decreasing the size of the neighborhood \( N_A \) we can always assume that \( N \) is large enough that this holds for all \( n \geq N^{5/4} \). This proves the theorem. \( \square \)

### 6. The Bridging Argument

In this section we prove a uniform LDT estimate in the vicinity of a diagonalizable cocycle (for which an LDT estimate already holds at all scales) up to a certain *finite scale* that depends on the proximity to the set of diagonalizable cocycles.

For every cocycle \( A \in \text{SL}_2(\mathbb{R})^k \) and scale \( n \in \mathbb{N} \), the number
\[
L^{(n)}(A) := \mathbb{E} \left[ \frac{1}{n} \log \| A^{(n)} \| \right]
\]
will be referred to as the *finite scale* LE of \( A \).

Note that as \( n \to \infty \), \( L^{(n)}(A) \to L(A) \) (the infinite scale LE).

**Proposition 6.1.** Let \( A \in \text{Diag} \) with \( L(A) > 0 \). There are constants \( \delta = \delta(A) > 0 \) and \( \bar{n} = \bar{n}(A) \in \mathbb{N} \) such that if \( B \in \text{SL}_2(\mathbb{R})^k \) is any cocycle with \( d(B, A) \leq \delta \), then for all scales \( n \) in the range
\[
\bar{n} \leq n \leq \left( \log d(B, \text{Diag})^{-1} \right)^{3/4}
\]
we have
\[
L_n(B) \geq \frac{L(A)}{3},
\]
\[
|L_n(B) - L_n(A)| \leq 2n^{-1/5},
\]
\[
P\left[\left| \frac{1}{n} \log \| B_n \| - L_n(B) \right| > n^{-1/5} \right] \leq e^{-n^{1/3}}.
\]

**Proof.** By Theorem 2.1, for every cocycle \( D \in \text{Diag}^* \), if \( \epsilon > 0 \) and \( n \geq \frac{C(D)}{\epsilon} \) we have
\[
P\left[\left| \frac{1}{n} \log \| D_n \| - L(D) \right| > \epsilon \right] \leq e^{-c(D)\epsilon^2 n},
\]
where \( C, c: \text{Diag}^* \to (0, \infty) \) are locally bounded functions.

Therefore, there are \( \delta_0 = \delta_0(A) > 0 \) and \( \bar{n} = \bar{n}(A) \in \mathbb{N} \) such that if \( D \in \text{Diag} \) with \( d(A, D) < \delta_0 \) and if \( n \geq \bar{n} \), putting \( \epsilon = n^{-1/4} \) above, we have
\[
P\left[\left| \frac{1}{n} \log \| D_n \| - L(D) \right| > n^{-1/4} \right] \leq e^{-n^{1/3}}. \tag{6.1}
\]

Since by Remark 2.2 on \( \text{Diag} \) the LE is continuous, by possibly decreasing \( \delta_0 \) we may assume that \( L(D) \geq \frac{L(A)}{2} \) for all cocycles \( D \in \text{Diag} \) with \( d(A, D) \leq \delta_0 \).

Let \( K = K(A) < \infty \) be such that \( \log \| B \|_{\infty} \leq K \) for all cocycles \( B \in \text{SL}_2(\mathbb{R})^k \) with \( d(A, B) \leq \delta_0 \).

An easy calculation then shows that for any two cocycles \( B \) and \( D \) in this \( \delta_0 \)-neighborhood of \( A \), for all scales \( m \in \mathbb{N} \) and all points \( x \in X \) we have
\[
\left| \frac{1}{m} \log \| B^{(m)}(x) \| - \frac{1}{m} \log \| D^{(m)}(x) \| \right| \leq e^{Km} d(B, D). \tag{6.2}
\]

Let \( \delta = \delta(A) > 0 \) be such that \( \delta < \min\{e^{-\bar{n}^2}, \frac{\delta_0}{3}\} \).

Fix any cocycle \( B \in \text{SL}_2(\mathbb{R})^k \) such that \( d(B, A) \leq \delta \).

Note that since \( A \in \text{Diag} \), we have \( d(B, \text{Diag}) \leq d(B, A) \leq \delta \), but \( d(B, \text{Diag}) \) could be significantly smaller than \( \delta \) (even zero). In order to get a (uniform) LDT estimate for \( B \) by proximity, at scales \( n \) in a reasonably long range, we will use the proximity of \( B \) to a ‘quasi-closest’ diagonalizable cocycle \( B_\flat \), rather than to \( A \). We will distinguish between the cases \( d(B, \text{Diag}) = 0 \) and \( d(B, \text{Diag}) > 0 \).

If \( d(B, \text{Diag}) = 0 \) then for every \( n \) there is \( D_n \in \text{Diag} \) such that \( d(B, D_n) \leq e^{-Kn} n^{-1/4} < \delta_0 \) if \( n \) is large enough.
Then applying (6.2) with $D = D_n$ we get that for all $x \in X$,

$$
\left| \frac{1}{n} \log \|B_n^{(n)}(x)\| - \frac{1}{n} \log \|D_n^{(n)}(x)\| \right| \leq e^{Kn} d(B, D_n) < n^{-1/4}.
$$

On the other hand, the LDT estimate (6.1) is applicable to the diagonalizable cocycle $D_n$ and for all $n \geq \bar{n}$ we have

$$
P \left[ \left| \frac{1}{n} \log \|D_n^{(n)}\| - L(D_n) \right| > n^{-1/4} \right] \leq e^{-n^{1/3}}.
$$

Combining the previous two inequalities, we have that

$$
\left| \frac{1}{n} \log \|B_n^{(n)}(x)\| - L(B_n) \right| < 2n^{-1/4} < n^{-1/5}
$$

with probability $> 1 - e^{-n^{1/3}}$, and provided $n$ is large enough.

Integrating in $x$ we get

$$
|L^{(n)}(B) - L(D_n)| \leq 2n^{-1/4} + 2Ke^{-n^{1/3}} < n^{-1/5},
$$

provided that $n \geq \bar{n}$ (if necessary, we slightly increase $\bar{n}$ so that asymptotic inequalities like the ones above hold for all $n \geq \bar{n}$).

Then $B$ satisfies the following LDT estimate

$$
P \left[ \left| \frac{1}{n} \log \|B^{(n)}\| - L^{(n)}(B) \right| > n^{-1/5} \right] \leq e^{-n^{1/3}}
$$

for all $n \geq \bar{n}$, and we are done with the case $d(B, \text{Diag}) = 0$.

Now assume that $d(B, \text{Diag}) > 0$, so there is $B_\flat \in \text{Diag}$ such that

$$
d(B, B_\flat) \leq 2d(B, \text{Diag}) (\leq 2\delta).
$$

Of course, $d(B, B_\flat) \geq d(B, \text{Diag})$, hence

$$
d(B, B_\flat) \approx d(B, \text{Diag}).
$$

Moreover, by the triangle inequality we also have

$$
d(A, B_\flat) \leq 3\delta < \delta_0,
$$

so (6.1) applies to $B_\flat$ and we have that for all $n \geq \bar{n}$,

$$
P \left[ \left| \frac{1}{n} \log \|B_\flat^{(n)}\| - L(B_\flat) \right| > n^{-1/4} \right] \leq e^{-n^{1/3}}.
$$

We transfer the LDT (6.3) to $B$ by proximity, in a certain range of scales. Put

$$
n_0 := \left( \log d(B, \text{Diag})^{-1} \right)^{1/4}, \quad \text{so } d(B, \text{Diag}) = e^{-n_0^{3}}
$$
and note that since $\delta < e^{-\tilde{n}^2}$, we have $\tilde{n} \leq n_0^2$. Now we derive an LDT for $B$ at any scale $n$ with $\tilde{n} \leq n \leq n_0^3$, using the same procedure as above. Applying (6.2) with $D = B_e$ we get that for all $x \in X$,

$$\left| \frac{1}{n} \log \|B^{(n)}(x)\| - \frac{1}{n} \log \|B_e^{(n)}(x)\| \right| \leq e^{Kn} d(B, B_e) < e^{Kn_0^3} e^{-n_0^4} < n^{-1/4}.$$ 

Combining the above with (6.3) we get

$$\left| \frac{1}{n} \log \|B^{(n)}\| - L(B) \right| < 2n^{-1/4} < n^{-1/5}$$

with probability $> 1 - e^{-n^{1/3}}$.

Integrating in $x$ we have

$$\left| L^{(n)}(B) - L(B_e) \right| < n^{-1/5}.$$

Then $|L^{(2n)}(B) - L^{(n)}(B)| \leq 2n^{-1/5}$ and $B$ satisfies the LDT estimate

$$\mathbb{P} \left[ \left| \frac{1}{n} \log \|B^{(n)}\| - L^{(n)}(B) \right| > n^{-1/5} \right] \leq e^{-n^{1/3}}$$

for all scales $n$ in the range $\tilde{n} \leq n \leq n_0^3 = (\log d(B, \text{Diag})^{-1})^{3/4}$.

Finally, in the same range of scales,

$$L^{(n)}(B) > L(B_e) - n^{-1/5} \geq \frac{L(A)}{2} - n^{-1/5} > \frac{L(A)}{3}$$

provided $\tilde{n}$ is larger than a negative power of $L(A)$. 

The next goal is to extend the range of the LDT (at the cost of weakening the estimate), for scales up to an arbitrary power of $\log d(B, \text{Diag})^{-1}$.

For that we use the Avalanche Principle (AP). Let us recall its statement (see for instance Theorem 2.1. in [7]).

**Lemma 6.2.** Let $0 < \kappa \ll \epsilon$. Given $g_0, g_1, \ldots, g_{n-1} \in \text{SL}_2(\mathbb{R})$ a sequence of matrices, if the geometric conditions

1. (angle) $\frac{\|g_i g_{i-1}\|}{\|g_i\| \|g_{i-1}\|} \geq \epsilon$

2. (gap) $\|g_i\| \geq \frac{1}{\kappa}$

are satisfied for all indices $i$, then the following holds:

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

We now formulate the main result of this section.
**Theorem 6.1.** Let $A \in \text{Diag}$ with $L(A) > 0$ and fix any integer $p \in \mathbb{N}$. There are constants $\delta = \delta(A) > 0$ and $\overline{\pi} = \overline{\pi}(A) \in \mathbb{N}$ such that if $B \in \text{SL}_2(\mathbb{R})^k$ is any cocycle with $d(B,A) \leq \delta$, then

$$\mathbb{P}\left[\frac{1}{n} \log \|B^{(n)}\| - L^{(n)}(B) \leq n^{-1/5}\right] \leq e^{-n^{2/p}}$$

holds for all scales $n$ with $\overline{\pi} \leq n \leq (\log d(B,\text{Diag})^{-1})^p$.

**Proof.** The constants $\delta, \overline{\pi}$, as well as the bound $K$ are the same as in Proposition 6.1. Fix any cocycle $B \in \text{SL}_2(\mathbb{R})^k$ with $d(B,A) \leq \delta$.

When $d(B,\text{Diag}) = 0$ there is nothing to prove, as the LDT in Proposition 6.1 holds for all scales $n \geq \overline{\pi}$.

Assume that $d(B,\text{Diag}) > 0$ and put $n_0 := (\log d(B,\text{Diag})^{-1})^{1/4}$. Then by Proposition 6.1 for $m$ in the range $\overline{\pi} \leq m \leq n_0^4$ and for $x \notin \mathcal{B}_m$, where $\mathbb{P}[\mathcal{B}_m] \leq e^{-m^{1/3}}$, we have that

$$L^{(m)}(B) \geq \frac{L(A)}{3} \tag{6.4}$$

and the following uniform LDT holds:

$$-m^{-1/5} \leq \frac{1}{m} \log \|B^{(m)}(x)\| - L^{(m)}(B) \leq m^{-1/5}. \tag{6.5}$$

Let $N \in \mathbb{N}$ with $n_0^3 \leq N \leq n_0^{4p} = (\log d(B,\text{Diag})^{-1})^p$. Write $N = (n-1)n_0 + m_0$, with $n_0 \leq m_0 < 2n_0$, so that $n \geq N^{\frac{1}{n_0^2}} \leq n_0^{4p-1}$.

The LDT (6.5) will allow us to apply the Avalanche Principle to matrix blocks of length $n_0$ and $m_0$, for a large enough set of phases $x$.

Indeed, fix $x \in X$ and define

$$g_i = g_i(x) := B^{(n_0)}(T^{i n_0} x) \quad \text{if } 0 \leq i \leq n - 2 \quad \text{and}$$

$$g_{n-1} = g_{n-1}(x) := B^{(m_0)}(T^{(n-1)n_0} x)$$

Then clearly

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

and for all $1 \leq i \leq n - 2$,

$$g_i g_{i-1} = B^{(n_0)}(T^{i n_0} x) B^{(n_0)}(T^{(i-1)n_0} x) \quad \text{while}$$

$$g_{n-1} g_{n-2} = B^{(n_0+m_0)}(T^{(n-2)n_0} x).$$

The validity of the gap condition in the AP in Lemma 6.2, for a large number of phases $x$, is due to the left hand side of (6.5) and the lower bound (6.4), both applied at scales $n_0,m_0$. 
Indeed, if \( x \notin \mathcal{B}_{n_0} \), then
\[
\frac{1}{n_0} \log \| B^{(n_0)}(x) \| > L^{(n_0)}(B) - n_0^{-1/5} > \frac{L(A)}{4} =: c > 0,
\]
and the same holds at scale \( m_0 \).

Then for \( x \notin \bigcup_{i=0}^{n_2-2} T^{-i n_0} \mathcal{B}_{n_0} \cup T^{-(n-1)n_0} \mathcal{B}_{n_0} \), which is a set of measure \( \lesssim n_0 e^{-n^{1/3}} \), and for all indices \( 0 \leq i \leq n - 1 \), we have
\[
\| g_i \| = \| g_i(x) \| \geq e^{c n_0} =: \frac{1}{\kappa}.
\]

The validity of the angle condition in the AP is derived in a similar way. If \( x \in \mathcal{B}_{2n_0} \cup \mathcal{B}_{n_0} \cup T^{-n_0} \mathcal{B}_{n_0} =: \mathcal{B}_{n_0} \), then
\[
\frac{\| B^{(2n_0)}(x) \|}{\| B^{(n_0)}(T^{n_0}x) \|} \| B^{(n_0)}(x) \| \geq e^{2 n_0(L^{(2n_0)}(B)-n_0^{-1/5})} e^{2 n_0(L^{(n_0)}(B)+n_0^{-1/5})}.
\]
\[
= e^{2 n_0(L^{(2n_0)}(B)-L^{(n_0)}(B))-4 n_0^{4/5}} \geq e^{-8 n_0^{4/5}} =: \epsilon.
\]

Then for all \( x \notin \bigcup_{i=0}^{n_2-2} T^{-i n_0} \mathcal{B}_{n_0} \), and for all indices \( 1 \leq i \leq n - 2 \),
\[
\frac{\| g_i g_{i-1} \|}{\| g_i \| \| g_{i-1} \|} \geq e^{-8 n_0^{4/5}} =: \epsilon,
\]
A similar estimate holds for the index \( i = n - 1 \).

Note also that
\[
\frac{\kappa}{\epsilon} = \frac{e^{-c n_0}}{e^{-8 n_0^{4/5}}} < e^{-c/2 n_0} \ll 1.
\]

We may then apply the AP to the list of matrices \( g_0, \ldots, g_{n-2}, g_{n-1} \) and get
\[
\log \| B^{(X)}(x) \| = -\sum_{i=1}^{n-2} \log \| B^{(n_0)}(T^{i n_0} x) \| + \sum_{i=1}^{n-2} \log \| B^{(2n_0)}(T^{(i-1)n_0} x) \|
\]
\[
+ \log \| B^{(n_0+m_0)}(T^{(n-2)n_0} x) \| + \mathcal{O}(n e^{-c n_0}), \quad (6.6)
\]

for all \( x \) outside a set \( \mathcal{B}_0 \) with \( \mathbb{P} [ \mathcal{B}_0 ] < e^{-n_0^{1/3}} n_0^4 p < e^{-n_0^{1/4}} \).

The right hand side of (6.6) can be broken down into three sums of i.i.d. random variables.

Indeed, since the matrix blocks \( B^{(n_0)}(T^{i n_0} x) \) use disjoint sets of \( x \)-coordinates for distinct values of the index \( i \), the sequence of random variables \( \{ \log \| B^{(n_0)}(T^{i n_0} x) \| : i = 1, \ldots, n - 2 \} \) is already independent
(and of course identically distributed). The common expected value of these random variables is \( n_0 L^{(n_0)}(B) \) and their common (and uniform)
The \(L^\infty\) bound is \(n_0 K\). Hoeffding’s inequality (Lemma 2.2) is then applicable and it implies, putting \(\epsilon = n_0 n^{-1/4}\),

\[
P\left[ \left| \sum_{i=1}^{n-2} \log \|B^{(n_0)}(T^{i n_0} x)\| - (n - 2) n_0 L^{(n_0)}(B) \right| > n_0 n^{3/4} \right] < e^{-\frac{1}{2K} n^{1/2}} < e^{-n_0}.
\]

Let us denote by \(\mathcal{B}_1\) the event above, so \(P[\mathcal{B}_1] < e^{-n_0}\).

The other sequence of random variables

\[
\{ \log \|B^{(2n_0)}(T^{i n_0} x)\| : i = 1, \ldots, n - 2 \}
\]

is not independent, as consecutive matrix blocks defining the terms of the sequence depend on interlacing sets of \(x\)-coordinates. However, we may split it into two sequences, corresponding to the even and respectively the odd values of the indices \(i\):

\[
\{ \log \|B^{(2n_0)}(T^{i n_0} x)\| : i = 1, \ldots, n - 2, \text{ i odd} \},
\]

\[
\{ \log \|B^{(2n_0)}(T^{i n_0} x)\| : i = 1, \ldots, n - 2, \text{ i even} \},
\]

each of which consisting now of independent random variables, as the corresponding matrix blocks depend of disjoint sets of \(x\)-coordinates.

The common expected value is now \(2n_0 L^{(2n_0)}(B)\) and the uniform \(L^\infty\) bound is \(2n_0 K\). Applying Hoeffding’s inequality with \(\epsilon = 2n_0 n^{-1/4}\) to each of the two i.i.d. sequences above and then summing, we conclude as before that

\[
P\left[ \sum_{i=1}^{n-2} \log \|B^{(2n_0)}(T^{i n_0} x)\| - (n - 2) 2n_0 L^{(2n_0)}(B) > 4n_0 n^{3/4} \right] < e^{-n_0}.
\]

Let us denote by \(\mathcal{B}_2\) the event above, so \(P[\mathcal{B}_2] < e^{-n_0}\).

Combining (6.6) with the previous two estimates derived via Hoeffding’s inequality, we obtain

\[
\log \|B^{(N)}(x)\| = - (n - 2) n_0 L^{(n_0)}(B) + (n - 2) 2n_0 L^{(2n_0)}(B) + O(n_0 n^{3/4})
\]

\[
+ \log \|B^{(n_0+n_0)}(T^{(n-2)n_0} x)\| + O(n e^{-c n_0}) \quad (6.7)
\]

for all \(x \notin \mathcal{B} := \mathcal{B}_0 \cup \mathcal{B}_1 \cup \mathcal{B}_2\), where \(P[\mathcal{B}] < e^{-n_0^{1/4}} + e^{-n_0} + e^{-n_0} < e^{-n_0^{1/5}}\).

Note that since \(N \asymp n n_0 \asymp n_0^{4p} < e^{c n_0}\), then

\[
\frac{(n - 2) n_0}{N} = 1 + O(n^{-1}), \quad \frac{(n - 2) 2n_0}{N} = 2 + O(n^{-1}),
\]

\[
\frac{n_0 n^{3/4}}{N} = O(n^{-1/4}), \quad \frac{n e^{-c n_0}}{N} = O(n^{-1}).
\]


Dividing both sides of (6.7) by $N$ we then have:

$$\frac{1}{N} \log \|B^{(N)}(x)\| = -L^{(n_0)}(B) + 2L^{(2n_0)}(B) + O(n^{-1/4})$$

for all $x \notin B$, where $\mathbb{P}[B] < e^{-n_0^{1/5}} < n^{-1/4}$.

Integrating in $x$,

$$L^{(N)}(B) = -L^{(n_0)}(B) + 2L^{(2n_0)}(B) + O(n^{-1/4}) + O(K \mathbb{P}[B])$$

$$= -L^{(n_0)}(B) + 2L^{(2n_0)}(B) + O(n^{-1/4}).$$

We may then conclude that if $x \notin B$ then

$$\frac{1}{N} \log \|B^{(N)}(x)\| = L^{(N)}(B) + O(n^{-1/4}).$$

Thus

$$\mathbb{P} \left[ \left| \frac{1}{N} \log \|B^{(N)}\| - L^{(N)}(B) \right| > n^{-1/5} \right] \leq \mathbb{P}[B] < e^{-n_0^{1/5}} < e^{-N^{1/4}},$$

which completes the proof.

\[ \square \]

7. Proofs of the main results

We combine the results of the previous sections to derive a uniform LDT estimate in the space of all random cocycles over a finite set of symbols, and as a consequence of that, a modulus of continuity for the Lyapunov exponent.

Proofs of Theorems 1.1 and 1.2. By Remark 2.1, it is enough to establish uniform (relative to the whole space $SL_2(\mathbb{R})^k$) sub-exponential LDT estimates for diagonalizable $SL_2(\mathbb{R})$ cocycles with positive LE.

Let $A \in \text{Diag}$ with $L(A) > 0$.

By Theorem 2.1, there are constants $\delta_1 > 0$, $c > 0$ and $C < \infty$ depending on $A$, such that if $D \in \text{Diag}$ is any diagonalizable cocycle with $d(D, A) \leq \delta_1$, then

$$\mathbb{P} \left[ \left| \frac{1}{n} \log \|D^{(n)}\| - L(D) \right| > \epsilon \right] \leq e^{-c(A) \epsilon^2 n}$$

for all $\epsilon > 0$ and $n \geq \frac{C(A)}{\epsilon}$.

In particular, putting $\epsilon := n^{-1/4}$, there is $n_1 = n_1(A) \in \mathbb{N}$ such that for all $n \geq n_1$, we have

$$\mathbb{P} \left[ \left| \frac{1}{n} \log \|D^{(n)}\| - L(D) \right| > n^{-1/4} \right] \leq e^{-n^{1/3}}.$$

Integrating in $x$, we get (for a constant $K = K(A) < \infty$),

$$\left| L^{(n)}(D) - L(D) \right| \leq n^{-1/4} + Ke^{-n^{1/3}} < 2n^{-1/4} \ll n^{-1/5},$$

70 P. DUARTE AND S. KLEIN
so we can essentially replace $L(D)$ by $L^{(n)}(D)$.

We conclude the following. There are constants $\delta_1 > 0$ and $\bar{n}_1 \in \mathbb{N}$ depending on $A$, such that if $D \in \mathcal{Diag}$ and $d(D, A) \leq \delta_1$, then for all $n \geq \bar{n}_1$ we have the following:

$$
P\left[ \left\| \frac{1}{n} \log \|D^{(n)}\| - L^{(n)}(D) \right\| > n^{-1/5} \right] \leq e^{-n^{1/3}}. \quad (7.1)
$$

If $B \notin \mathcal{Diag}$, then $\rho(B) = \max\{\rho_-(B), \rho_+(B)\} > 0$, so either $\rho_-(B) > 0$ or $\rho_+(B) > 0$. But since $\rho_-(B^{-1}) \approx \rho_+(B)$, the latter case is equivalent to $\rho_-(B^{-1}) > 0$.

By Theorem [5.1], there are constant $\delta_2 > 0$ and $c > 0$ depending on $A$, such that if $B \in \text{SL}_2(\mathbb{R})^k$ is any cocycle with $d(B, A) \leq \delta_2$ and $\rho_-(B) > 0$, then

$$
P\left[ \left\| \frac{1}{n} \log \|B^{(n)}\| - L(B) \right\| > n^{-1/6} \right] < e^{-n^{1/6}} \quad (7.2)
$$

holds for all $n \geq \lceil c^{-1} \log (1/\rho_-(B)) \rceil^{54}$.

As before, we may essentially replace in $(7.2) L(B)$ by $L^{(n)}(B)$.

Moreover, $A^{-1}$ is also diagonalizable and $L(A^{-1}) = L(A) > 0$, so $(7.2)$ holds for $B^{-1}$ too, provided it is close enough to $A^{-1}$. Since the map $\text{SL}_2(\mathbb{R})^k \ni B \mapsto B^{-1} \in \text{SL}_2(\mathbb{R})^k$ is locally bi-Lipschitz, that proximity is ensured by the proximity of $B$ to $A$ (we may have to decrease $\delta_2$ slightly). Furthermore, by Lemma [2.4] any LDT that holds for a cocycle holds for its inverse.

From these considerations we derive the following. There are constants $\delta_2 > 0$ and $c > 0$ depending on $A$, such that if $B \in \text{SL}_2(\mathbb{R})^k$ is any cocycle with $d(B, A) \leq \delta_2$ and $\rho_-(B^{\pm}) > 0$, then

$$
P\left[ \left\| \frac{1}{n} \log \|B^{(n)}\| - L^{(n)}(B) \right\| > n^{-1/7} \right] < e^{-n^{1/6}} \quad (7.3)
$$

holds for all $n \geq \lceil c^{-1} \log (1/\rho_-(B^{\pm})) \rceil^{54}$.

Recall from Proposition [3.5] that there are constants $\delta_3 > 0$ and $C < \infty$ depending on $A$, such that for all cocycles $B$ with $d(B, A) \leq \delta_3$, either $d(B, \mathcal{Diag}) \leq C \rho_-(B)$ or $d(B, \mathcal{Diag}) \leq C \rho_-(B^{-1})$.

Choose the sign $\varepsilon = \pm$ so that $d(B, \mathcal{Diag}) \leq C \rho_-(B^\varepsilon)$.

Let $\delta_4 := \min\{\delta_2, \delta_3\}$. If $B \notin \mathcal{Diag}$ (so in particular $d(B, \mathcal{Diag}) > 0$) and if $d(B, A) \leq \delta_4$, then $(7.3)$ holds for all $n \geq \lceil c^{-1} \log (1/\rho_-(B^\varepsilon)) \rceil^{54}$.

But $c^{-1} \log (1/\rho_-(B^\varepsilon)) \leq c^{-1} \log C + c^{-1} \log d(B, \mathcal{Diag})^{-1}$.

Since $d(B, \mathcal{Diag}) \leq d(B, A) \leq \delta_4 \ll 1$, by possibly decreasing $\delta_4$, we may assume that

$$c^{-1} \log C + c^{-1} \log d(B, \mathcal{Diag})^{-1} \leq (\log d(B, \mathcal{Diag})^{-1})^{1+1/9}.$$
Consequently, for any cocycle $B \in \text{SL}_2(\mathbb{R})^k$ with $d(B, A) \leq \delta_4$, if $B \notin \text{Diag}$ then

$$\mathbb{P}\left[\frac{1}{n} \log \|B^{(n)}\| - L^{(n)}(B) > n^{-1/7}\right] < e^{-n^{1/6}}$$

(7.4) holds for all $n \geq (\log d(B, \text{Diag}) - 1)^{60}$.

Finally, we apply the bridging argument in Theorem 6.1 with $p = 60$. There are constants $\delta_5 > 0$ and $\bar{n}_2 \in \mathbb{N}$ depending on $A$, such that if $B \in \text{SL}_2(\mathbb{R})^k$ is any cocycle with $d(B, A) \leq \delta_5$ and if $B \notin \text{Diag}$, then

$$\mathbb{P}\left[\frac{1}{n} \log \|B^{(n)}\| - L^{(n)}(B) > n^{-1/5}\right] \leq e^{-n^{1/5}}$$

(7.5) holds for all scales $n$ with $\bar{n}_2 \leq n \leq (\log d(B, \text{Diag}) - 1)^{60}$.

From (7.4) and (7.5) we derive a uniform LDT for non diagonalizable cocycles, while (7.1) provides the same for diagonalizable cocycles. We conclude: there are constants $\delta > 0$ and $\bar{n} \in \mathbb{N}$ depending on $A$, so that for any cocycle $B \in \text{SL}_2(\mathbb{R})^k$ with $d(B, A) \leq \delta$ and for all $n \geq \bar{n}$, we have

$$\mathbb{P}\left[\frac{1}{n} \log \|B^{(n)}\| - L^{(n)}(B) > n^{-1/5}\right] \leq e^{-n^{1/5}}.$$  

(7.6)

This is the form of the uniform LDT required by our abstract continuity theorem (ACT) (see Section 3.2 in [7]). Consequently, the continuity properties of the LE and of the Oseledets decomposition formulated in Theorem 1.2 follow from (7.6) by means of the ACT (Theorem 3.3 in [7]). The estimate (7.6) also implies (see Proposition 3.4 in [7]) a uniform rate of convergence of the finite scale LE $L^{(n)}(B)$ to $L(B)$. Therefore, we may essentially replace in (7.6) $L^{(n)}(B)$ by $L(B)$, and derive a uniform LDT as formulated in Definition 1.3, with $a = -1/6$ and $b = 1/1200$.

Proof of Theorem 1.3. Let $\{v_n\}_{n \in \mathbb{Z}}$ be an i.i.d. sequence of random variables with common probability distribution $\mu$ and let $\{w_n\}_{n \in \mathbb{Z}}$ be another i.i.d. sequence of random variables with distribution $\nu$, such that moreover $\{v_n\}$ and $\{w_n\}$ are independent from each other. Consider the random Jacobi operator $H_\omega$ acting on $l^2(\mathbb{Z})$ and given by

$$[H_\omega \psi]_n := -(w_{n+1}(\omega)\psi_{n+1} + w_n(\omega)\psi_{n-1}) + v_n(\omega)\psi_n.$$  

(7.7)

\[12^{\text{In [7] the uniform LDT is exponential, and as a result, the modulus of continuity of the LE and that of the Oseledets decomposition are Hölder. As alluded to at the end of Section 3.2 of [7] (or using our more general ACT in [6]), a sub-exponential uniform LDT leads to a weak-Hölder modulus of continuity.}}\]
We assume that the supports of the distributions \( \mu \) and \( \nu \) are finite, and that \( \mathbb{E}(\log |w_1|) > -\infty \), that is, \( 0 \not\in \text{supp} (\nu) \).

The Schrödinger (or eigenvalue) equation \( H_\omega \psi = E \psi \) for \( E \in \mathbb{R} \) and \( \psi = \{ \psi_n \}_{n \in \mathbb{Z}} \in \mathbb{R}^\mathbb{Z} \), is a second order finite difference equation. It can be written as the matrix first order finite difference equation

\[
\begin{bmatrix}
\psi_{n+1} \\
\psi_n
\end{bmatrix} = g_n^E \begin{bmatrix}
\psi_n \\
\psi_{n-1}
\end{bmatrix},
\]

where

\[
g_n^E = \begin{bmatrix}
v_n - E & w_n \\
w_{n+1} & w_n
\end{bmatrix}.
\]

The Lyapunov exponent \( L^+(E) \) of the operator \([7.7]\) at energy \( E \) is thus the Lyapunov exponent \( L^+(g^E) \) of the i.i.d. multiplicative process \( g^E := \{ g_n^E \}_{n \in \mathbb{Z}} \). This process can be conjugated with the i.i.d. SL\(_2(\mathbb{R})\) process \( \tilde{g}^E := \{ \tilde{g}_n^E \}_{n \in \mathbb{Z}} \), where

\[
\tilde{g}_n^E = \begin{bmatrix}
v_n - E & -w_n \\
w_n & 0
\end{bmatrix},
\]

in the sense that

\[
g_n^E = \begin{bmatrix}
w_{n+1} & 0 \\
0 & 1
\end{bmatrix}^{-1} \tilde{g}_n^E \begin{bmatrix}
w_n & 0 \\
0 & 1
\end{bmatrix}.
\]

Thus for all \( E \in \mathbb{R} \) we have

\[
L^+(E) = L^+(g^E) = L(\tilde{g}^E) \geq 0.
\]

Theorem 1.2 is applicable to \( \tilde{g}^E \) and it implies that the function \( E \mapsto L(\tilde{g}^E) = L^+(E) \) is continuous. In particular, if \( E_0 \in \mathbb{R} \) is such that \( L^+(E_0) > 0 \), there is an open interval \( J \ni E_0 \) such that \( L^+|_J > 0 \) and \( L^+|_J \) is weak Hölder continuous.

By means of the Thouless formula, the weak Hölder continuity of the IDS also follows.

Finally, let us discuss the dynamical localization of the operator \([7.7]\). Theorem 5.1 in [4] establishes such a result (for the more general setting of block Jacobi operators) under the assumptions of Furstenberg’s theory. However, these assumptions are only needed insofar as to guarantee the positivity of the LE and the Hölder continuity of the IDS, where the latter is used to derive a Wegner estimate. In our case, the positivity of the LE is assumed, while the weak Hölder continuity of the IDS still implies a (good enough, albeit weaker) Wegner estimate. \( \square \)
Acknowledgments. The first author was supported by Fundação para a Ciência e a Tecnologia, under the projects: UID/MAT/04561/2013 and PTDC/MAT-PUR/29126/2017. The second author has been supported in part by the CNPq research grant 306369/2017-6 (Brazil) and by a research productivity grant from his current institution (PUC-Rio). He would also like to acknowledge the support of the Isaac Newton Institute for Mathematical Sciences in Cambridge, UK, where the work on this project first started, during the PEPW04 workshop in 2015.

References

[1] Alexandre Baraviera and Pedro Duarte, *Approximating Lyapunov exponents and stationary measures*, preprint (2017), 1–27.

[2] Carlos Bocker-Neto and Marcelo Viana, *Continuity of Lyapunov exponents for random two-dimensional matrices*, Ergodic Theory and Dynamical Systems (2016), 1–30.

[3] Philippe Bougerol, *Théorèmes limite pour les systèmes linéaires à coefficients markoviens*, Probab. Theory Related Fields 78 (1988), no. 2, 193–221. MR 945109 (89i:60122)

[4] Jacob Chapman and Günter Stolz, *Localization for random block operators related to the XY spin chain*, Ann. Henri Poincaré 16 (2015), no. 2, 405–435. MR 3302603

[5] David Damanik, *Schrödinger operators with dynamically defined potentials*, Ergodic Theory Dynam. Systems 37 (2017), no. 6, 1681–1764. MR 3681983

[6] Pedro Duarte and Silvius Klein, *Lyapunov exponents of linear cocycles; continuity via large deviations*, Atlantis Studies in Dynamical Systems, vol. 3, Atlantis Press, 2016.

[7] ______, *Continuity of the Lyapunov exponents of linear cocycles*, Publicações Matemáticas, 31º Colóquio Brasileiro de Matemática, IMPA, 2017, available at [https://impa.br/wp-content/uploads/2017/08/31CBM_02.pdf](https://impa.br/wp-content/uploads/2017/08/31CBM_02.pdf).

[8] Pedro Duarte, Silvius Klein, and Manuel Santos, *A random cocycle with non Hölder Lyapunov exponent*, preprint, 2018.

[9] H. Furstenberg and Y. Kifer, *Random matrix products and measures on projective spaces*, Israel J. Math. 46 (1983), no. 1-2, 12–32. MR 727020

[10] Michael Goldstein and Wilhelm Schlag, *Hölder continuity of the integrated density of states for quasi-periodic Schrödinger equations and averages of shifts of subharmonic functions*, Ann. of Math. (2) 154 (2001), no. 1, 155–203. MR 1847592 (2002h:82055)

[11] Tosio Kato, *Perturbation theory for linear operators*, Classics in Mathematics, Springer-Verlag, Berlin, 1995, Reprint of the 1980 edition. MR 1335452

[12] Benoît Kloeckner, *Effective perturbation theory for linear operators*, preprint [arXiv:1703.09425](https://arxiv.org/abs/1703.09425) 2017.

[13] Émile Le Page, *Théorèmes limites pour les produits de matrices aléatoires*, Probability measures on groups (Oberwolfach, 1981), Lecture Notes in Math., vol. 928, Springer, Berlin-New York, 1982, pp. 258–303. MR 669072 (84d:60012)
[14] Émile Le Page, *Régularité du plus grand exposant caractéristique des produits de matrices aléatoires indépendantes et applications*, Ann. Inst. H. Poincaré Probab. Statist. **25** (1989), no. 2, 109–142. MR 1001021

[15] F. Ledrappier, *Quelques propriétés des exposants caractéristiques*, École d’été de probabilités de Saint-Flour, XII—1982, Lecture Notes in Math., vol. 1097, Springer, Berlin, 1984, pp. 305–396. MR 876081

[16] Elaís C. Malheiro and Marcelo Viana, *Lyapunov exponents of linear cocycles over Markov shifts*, Stoch. Dyn. **15** (2015), no. 3, 1550020, 27. MR 3349975

[17] Günter Stolz, *An introduction to the mathematics of Anderson localization*, Entropy and the quantum II, Contemp. Math., vol. 552, Amer. Math. Soc., Providence, RI, 2011, pp. 71–108. MR 2868042

[18] El Hadji Yaya Tall and Marcelo Viana, *Moduli of continuity for Lyapunov exponents of random GL(2) cocycles*, preprint, 2018.

[19] Terence Tao, *Topics in random matrix theory*, Graduate Studies in Mathematics, vol. 132, American Mathematical Society, Providence, RI, 2012. MR 2906465 (2012k:60023)

[20] M. Viana, *Lectures on Lyapunov exponents*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2014.

**Departamento de Matemática and CMAF, Faculdade de Ciências, Universidade de Lisboa, Portugal**

*E-mail address: pduarte@ptmat.fc.ul.pt*

**Departamento de Matemática, Pontificia Universidade Católica do Rio de Janeiro (PUC-Rio), Brazil**

*E-mail address: silviusk@impa.br*