Random product of matrices (Seminar notes at UZH, 2025)

Yuxiang Jiao
https://yuxiangjiao.github.io

§0 Revisit: Oseledets' multiplicative ergodic theorem

Let $(X, m, \mathscr{B})$ be a probability space and $f : X \to X$ be a measure preserving transformation. Recall that a **linear cocycle** over the system $(X, m, \mathscr{B}, f)$ is given by a (measurable) map $A : X \to \mathrm{GL}_d(\mathbb{R})$ for some positive integer d. The linear cocycle induces a natural skew-product over f given by

$$F : X \times \mathbb{R}^d \to X \times \mathbb{R}^d, \quad (x, v) \mapsto (fx, A(x)v).$$

Then the iteration of F is given by

$$F^n(x) = (f^n(x), A^n(x)), \quad \text{where } A^n(x) = A(f^{n-1}x) \cdots A(fx)A(x).$$

Remark 0.1 Here, the terminology "linear cocycle" also refers to the map

$$A : X \times \mathbb{Z}_{\geqslant 0} \to \mathrm{GL}_d(\mathbb{R}), \quad (x, n) \mapsto A^n(x).$$

More general, if the base system (X, f) is replaced by a (semi)group action $G \curvearrowright X$ then a "linear cocycle" is a map

$$A : X \times G \to \mathrm{GL}_d(\mathbb{R})$$

satisfying the "cocycle condition"

$$A(x, g_1 g_2) = A(g_2.x, g_1) A(x, g_2), \quad \forall g_1, g_2 \in G, x \in X.$$

This identifies the concept of cocycles in the theory of group cohomology.

Theorem 0.2 (Oseledets' Multiplicative Ergodic Theorem)
Assume that A satisfies the integrability condition $\log^+ \|A(\cdot)\| \in L^1(X, m)$. Then there exists a forward invariant set $\widetilde{X} \in \mathcal{A}$ with full measure such that for each $x \in \widetilde{X}$, there exist

- a positive integer $1 \leqslant k(x) \leqslant d$;
- $k(x)$ different numbers $\lambda_1(x) > \lambda_2(x) > \cdots > \lambda_k(x)$, called Lyapunov exponents;
- a filtration of subspaces of $\mathbb{R}^d$ as

$$\mathbb{R}^d = L_1(x) \supsetneqq L_2(x) \supsetneqq \cdots \supsetneqq L_k(x) \supsetneqq L_{k+1}(x) = \{0\},$$

such that

(i) All the objects depend measurably on $x \in \widetilde{X}$ and satisfy the cocycle invariance

$$k(fx) = k(x), \quad \lambda_i(fx) = \lambda_i(x), \quad A(x)V_i(x) = V_i(fx).$$

(ii) For every $x \in \widetilde{X}$ and $1 \leqslant i \leqslant k(x)$, we have

$$\lim_{n\to+\infty} \frac{1}{n} \log \|A^n(x)v\| = \lambda_i(x), \quad \forall v \in V_i(x) \setminus V_{i+1}(x).$$

(iii) For every $x \in \widetilde{X}$, we have

$$\lambda_1(x) = \lim_{n\to+\infty} \frac{1}{n} \log \|A^n(x)\|, \quad \lambda_k(x) = \lim_{n\to+\infty} \frac{1}{n} \log \|A^n(x)^{-1}\|^{-1}.$$

(iv) If we denote $d_i(x) := \dim V_i(x) - \dim V_{i+1}(x)$ then we have

$$\sum_{i=1}^{k} d_i(x)\lambda_i(x) = \lim_{n\to+\infty} \frac{1}{n} \log|\det A(x)|.$$

In particular, if A takes values in $\mathrm{SL}_d(\mathbb{R})$ then $\sum_{i=1}^{k} d_i(x)\lambda_i(x) = 0$.

Remark 0.3 Recall that every positive definite symmetric B can be diagonalized as $B = Q^{-1}AQ$, where A is a diagonal matrix with positive diagonal entries. Hence we can define $B^\alpha := Q^{-1}A^\alpha Q$ for every $\alpha \in \mathbb{R}$. One can show that the limit

$$\lim_{n\to+\infty} (A^n(x)^t A^n(x))^{\frac{1}{2n}}$$

exists almost everywhere, and and that its distinct eigenvalues are $e^{\lambda_1} > \cdots > e^{\lambda_i}$, each with multiplicity d_i. This derives items (iii) and (iv) in the theorem.

Assume for simplicity that the base system is ergodic. Then the integer k and the Lyapunov exponents $\lambda_1 > \cdots > \lambda_k$ are independent of $x \in X$. Some natural questions arise from the theorem:

Question 0.4

1. When does the cocycle A has a **simple Lyapunov spectrum**, i.e., when does it admit d-different Lyapunov exponents?
2. Assuming that A takes values in $\mathrm{SL}_d(\mathbb{R})$, when does A possess a **positive Lyapunov exponent**, i.e., when does A have at least two different Lyapunov exponents?

Example 0.5

Let A be a **constant cocycle**, given by $A = A_0 \in \mathrm{GL}_d(\mathbb{R})$. Then the Lyapunov exponents of A are the logarithms of the absolute value of eigenvalues of A_0. For most choices of A_0, the cocycle A has a simple Lyapunov spectrum.

For general cocycles, we also expect that most of them should have a simple Lyapunov spectrum. To this end, we will consider Question 0.4 in one of the simplest yet still interesting cases: the base system is Bernoulli, and the cocycle is locally constant — also known as the **random product of matrices**.

§1 Random product of matrices and Lyapunov exponents

As a starting point, we recall the strong law of large numbers.

Theorem 1.1 (Strong Law of Large Numbers)
Let $X_0, X_1, \cdots, X_n, \cdots$ be a sequence of i.i.d. random variables. Assume that the expectation $\mathbf{E}\,|X_0| < \infty$. Then

$$\frac{1}{n}(X_0 + X_1 + \cdots + X_{n-1}) \to \mathbf{E}\,X_0 \quad \text{almost surely.}$$

Let $Y_n = e^{X_n}$. Then $Y_0, Y_1, \cdots$ is a sequence of i.i.d. positive random variables.

Corollary 1.2
Assuming $\mathbf{E}\log Y_0 = \mathbf{E}\,X_0 > 0$, then

$$Y_0 Y_1 \cdots Y_n \to +\infty, \quad \text{exponentially fast almost surely .}$$

We now wish to generalize this result to the case of non-commuting random products.

Example 1.3
Let us start with the most famous matrix in dynamical systems, $g_1 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$, together with its "mirror" $g_2 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$. We consider a random sequence $A_0, A_1, \cdots$ where each A_i independently takes the value g_1 or g_2 with equal probability. Do we still have that

$$A^n := A_{n-1} \cdots A_1 A_0$$

goes to infinity exponentially fast almost surely? The answer is **Yes**.

To be precise, let μ be a probability measure on $\mathrm{SL}_d(\mathbb{R})$. Let $X = (\mathrm{SL}_d(\mathbb{R}))^{\mathbb{N}}$ be endowed with the product probability measure $\mathbf{P} = \mu^{\otimes\mathbb{N}}$. A point $x \in X$ represents one realization of a sequence of matrices $x = (A_0(x), A_1(x), \cdots)$. Under the product measure $\mathbf{P} = \mu^{\otimes\mathbb{N}}$, the coordinate maps $(A_i : X \to \mathrm{SL}_d(\mathbb{R}))$ form an i.i.d. family of random variables with common distribution μ. We aim to study the asymptotic behavior of

$$A^n = A^n(x) := A_{n-1} \cdots A_1 A_0$$

as $n \to +\infty$. A natural integrability condition is

$$\mathbf{E}\log\|A_0\| = \int \log\|A_0\|\,\mathrm{d}\mu < \infty.$$

Definition 1.4. We define the **extremal Lyapunov exponents** as

$$\lambda_+ := \lim_{n\to\infty}\frac{1}{n}\log\|A^n\|, \quad \lambda_- := \lim_{n\to\infty}\frac{1}{n}\log\left\|(A^n)^{-1}\right\|^{-1}.$$

They are called the **upper and lower Lyapunov exponents**, respectively.

The existence and almost sure constancy of these limits follow from Oseledects' multiplicity ergodic theorem (or Kingman's subadditive ergodic theorem) through the following dynamical formalism.

Dynamical formalism.

Let $\sigma : X \to X$ be the left shift map $(A_0(x), A_1(x), \cdots) \mapsto (A_1(x), A_2(x), \cdots)$. Then $(X, \mathbf{P}, \sigma)$ is a probability measure preserving system. The cocycle $A : X \to (\mathrm{SL}_d(\mathbb{R}))^{\mathbb{N}}$ is given by $x = (A_0(x), A_1(x), \cdots) \mapsto A_0(x)$. Then we have

$$A^n(x) = A(\sigma^{n-1}(x)) \cdots A(\sigma(x))A(x).$$

Remark 1.5 To make life easier, we assume that μ is finitely supported, given by

$$\mu = \sum_{i=0}^{N-1} p_i \delta_{B_i},$$

where $\sum_{i=0}^{N-1} p_i = 1$. In this case, the system reduces to $X = \Sigma = \{0, 1, \cdots, N-1\}^{\mathbb{N}}$, which is compact. We slightly abuse notation by also viewing μ as a probability measure on $\{0, 1, \cdots, N-1\}$, then $(\Sigma, \mathbf{P}, \sigma)$ is a **Bernoulli shift**. The cocycle A is **locally constant** since each cylinder set $\{i\} \times \Sigma \subset \Sigma$ is open for every $i \in \{0, 1, \cdots, N-1\}$. Although compactness is not essential for the statement or the proof, it makes the picture much easier to visualize.

By Oseledet's theorem (Theorem 0.2) for the cocycle A, since $(X, \mathbf{P}, \sigma)$ is ergodic, there exist k different Lyapunov exponents $\lambda_1 > \cdots > \lambda_k$ of A. Moreover,

$$\lambda_1 = \lim_{n\to+\infty} \frac{1}{n} \log \|A^n\| = \lambda_+, \quad \lambda_k = \lim_{n\to+\infty} \frac{1}{n} \log \left\|(A^n)^{-1}\right\|^{-1} = \lambda_-.$$

Recall that A takes values in $\mathrm{SL}_d(\mathbb{R})$. In view of item (iv) in Theorem 0.2, we have

$$\lambda_+ > \lambda_- \quad \text{iff} \quad k \geqslant 2 \quad \text{iff} \quad \lambda_+ > 0.$$

§2 Examples and main results

Example 2.1 (Deterministic cases)

Let $g_0 \neq \mathrm{Id} \in \mathrm{SL}_2(\mathbb{R})$ be a fixed matrix. Let μ be the corresponding Dirac measure $\mu = \delta_{g_0}$ on $\mathrm{SL}_2(\mathbb{R})$. This setting gives rise to three distinct cases:

- If g_0 is hyperbolic, then $\lambda_+ > 0$ and $\|A^n\|$ goes to infinity exponentially fast.
- If g_0 is parabolic, then $\lambda_+ = 0$ and $\|A^n\|$ goes to infinity polynomially fast.
- If g_0 is elliptic, then $\lambda_+ = 0$ and $\|A^n\|$ is uniformly bounded.

Deterministic cases are not very interesting. However, the last two types naturally extend to non-deterministic settings, providing further examples with non-positive Lyapunov exponents.

Example 2.2 (Compact case)

Assume that μ is supported on a compact subgroup of $\mathrm{SL}_2(\mathbb{R})$. That is, there exists $g_0 \in \mathrm{SL}_2(\mathbb{R})$ such that μ is supported on $g_0^{-1}\mathrm{SO}_2(\mathbb{R})g_0$. Then we have $A^n \in g_0^{-1}\mathrm{SO}_2(\mathbb{R})g_0$ for every $n \geqslant 0$. Therefore $\|A^n\| \leqslant \|g_0^{-1}\|\|g_0\|$, which is uniformly bounded, and hence $\lambda_+ = 0$.

Example 2.3 (Parabolic cases)

Assume that μ is supported on the unipotent subgroup $\left\{\begin{bmatrix} 1 & s \\ & 1 \end{bmatrix} : s \in \mathbb{R}\right\}$. Then $A_n = \begin{bmatrix} 1 & Y_n \\ & 1 \end{bmatrix}$, where Y_n is a sequence of i.i.d. random variables. We have

$$A^n = \begin{bmatrix} 1 & Y_0 + \cdots + Y_{n-1} \\ & 1 \end{bmatrix}.$$

By the law of large numbers, $\|A^n\|$ grows linearly almost surely and hence $\lambda_+ = 0$.

The reason λ_+ is not positive in this example is that the matrices in $\operatorname{supp}\mu$ share a common invariant direction in $\mathbb{R}^2$. That is, the group generated by $\operatorname{supp}\mu$ acts reducibly on $\mathbb{R}^2$.

Example 2.4 (Reducible cases)

Assume that μ is supported on the positive diagonal subgroup $\left\{\begin{bmatrix} e^t & \\ & e^{-t} \end{bmatrix} : t \in \mathbb{R}\right\}$. Then $A_n = \begin{bmatrix} e^{X_n} & \\ & e^{-X_n} \end{bmatrix}$, where X_n is a sequence of i.i.d. random variables. By the low of large number, $\frac{1}{n}\log\|A^n\| \to |\mathbf{E}\, X_n|$. If $\mathbf{E}\, X_n = 0$ then $\lambda_+ = 0$.

In fact, not only reducible actions may produce a zero Lyapunov exponent, "almost reducible" ones can do so as well. The following, slightly more delicate example illustrates this phenomenon.

Example 2.5 (Almost reducible cases)

Let $g_0 = \begin{bmatrix} e & \\ & e^{-1} \end{bmatrix}$, a hyperbolic matrix that expands along the x-axis and contracts along the y-axis. Let $g_1 = \begin{bmatrix} & 1 \\ -1 & \end{bmatrix}$, which interchanges the two coordinate axes. Let $\mu = \frac{1}{2}\delta_{g_0} + \frac{1}{2}\delta_{g_1}$. Because of the rotations introduced by g_1, the expansion and contraction effects along the axes almost cancel out. In fact, if we denote by t_k the index of the k-th occurrence of g_2 (that is, $A_{t_k} = g_2$), and set $\xi_k = t_k - t_{k-1} - 1$ then

$$\log\|A^n\| \sim \frac{\xi_1 - \xi_2 + \cdots + \xi_{k-1} - \xi_k}{n} \to 0 \text{ a.e.}\,.$$

We note that all of the preceding examples arise from certain algebraic obstructions. We summarize these obstructions in the following definitions.

Definition 2.6. A probability measure μ on $\mathrm{SL}_d(\mathbb{R})$ is said to be **irreducible** if there is no nontrivial proper subspace $V \subset \mathbb{R}^d$ such that $A(V) \subseteq V$ for μ-almost every $A \in \mathrm{SL}_d(\mathbb{R})$.

Definition 2.7. A probability measure μ on $\mathrm{SL}_d(\mathbb{R})$ is said to be **strongly irreducible** if there is no finite union $W = V_1 \cup V_2 \cup \cdots \cup V_\ell$ of nontrivial proper subspaces $V_i \subseteq \mathbb{R}^d$ such that $A(W) \subseteq W$ for μ-almost every $A \in \mathrm{SL}_d(\mathbb{R})$.

Definition 2.8. A probability measure μ on $\mathrm{SL}_d(\mathbb{R})$ is said to be **non-compact** if the semigroup generated by $\operatorname{supp}\mu$ is not contained in any compact subgroup of $\mathrm{SL}_d(\mathbb{R})$.

Theorem 2.9 (Furstenberg) If μ is strongly irreducible and non-compact then $\lambda_+ > 0$.

Exercise 2.10. Check Example 1.3 satisfies the assumptions and hence $\lambda_+ > 0$.

In order to establish the simplicity of the Lyapunov spectrum, Furstenberg introduced certain contracting conditions for the action of $\operatorname{supp}\mu$ on various projective spaces. Let $T_\mu \subset \mathrm{SL}_d(\mathbb{R})$ be the semigroup generated by $\operatorname{supp}\mu$.

Definition 2.11. A probability measure μ on $\mathrm{SL}_d(\mathbb{R})$ is said to be **proximal** if there exists a sequence $(g_n) \subset T_\mu$ such that $g_n/\|g_n\|^{-1} \to B \in \mathrm{Mat}_d(\mathbb{R})$ with rank $B = 1$. Equivalently, there exists a hyperplane $V \subset \mathbb{R}^d$ (i.e. $\dim V = d-1$) and a point $[v_0] \in \mathbb{P}(\mathbb{R}^d)$ such that $[g_n v] \to [v_0]$ for every $[v] \in \mathbb{P}(\mathbb{R}^d) \setminus \mathbb{P}(V)$.

Definition 2.12. For each $1 \leqslant p \leqslant d-1$, a probability measure μ on $\mathrm{SL}_d(\mathbb{R})$ is said to be **p-proximal** (resp. **p-strongly irreducible**) if it is proximal (resp. strongly irreducible) when regarded as a probability measure on $\mathrm{SL}(\wedge^p \mathbb{R}^d)$.

Theorem 2.13 (Furstenberg)

If μ is p-strongly irreducible and p-proximal for every $1 \leqslant p \leqslant d-1$ then μ has a simple Lyapunov spectrum.

However, these conditions are often difficult to verify in practice. Later, Gol'dsheid and Margulis showed that all obstructions preventing the Lyapunov spectrum from being simple are of algebraic nature.

Theorem 2.14 (Gol'dsheid-Margulis)

Assume that T_μ is Zariski dense in $\mathrm{SL}_d(\mathbb{R})$ under the real Zariski topology. Then μ has a simple Lyapunov spectrum.

In fact, if T_μ is Zariski dense, then it is straightforward to verify that μ is:

- non-compact: otherwise, T_μ would preserve a non-degenerate bilinear form on $\mathbb{R}^d$, that is, there would exist some $Q \in \mathrm{GL}_d(\mathbb{R})$ such that $g^t Q g = Q$ for every $g \in T_\mu$.
- irreducible: otherwise, suppose that T_μ preserves a nontrivial proper subspace $V \subset \mathbb{R}^d$. Let $e_1, \cdots, e_\ell$ be a basis of V and $e_{\ell+1}, \cdots, e_d$ be a basis of $V^\perp$ then
$$\langle g v_i, v_j \rangle = 0, \quad \forall g \in T_\mu,\ \forall 1 \leqslant i \leqslant \ell < j \leqslant d,$$
which shows that T_μ satisfies a nontrivial algebraic relation, contradicting Zariski density.

Exercise 2.15. Show that if T_μ is Zariski dense then μ is p-strongly irreducible for all $1 \leqslant p \leqslant d-1$.

The most nontrivial part of the Gol'dsheid-Margulis proof is to verify that Zariski density in fact implies the proximality condition on each projective space $\mathbb{P}(\wedge^p \mathbb{R}^d)$. We also refer the reader to Chapter 6 of Benoist and Quint's textbook *Random walks on reductive groups*, which shows that the subset of T_μ consisting of loxodromic elements[1] is still Zariski dense.

[1] In the case of $\mathrm{SL}_d(\mathbb{R})$, an element g is called **loxodromic** if it has exactly d eigenvalues with distinct absolute values. Hence the sequence (g^n) is p-proximal for all p.

§3 Stationary measures

We now introduce an important tool for studying random dynamics. For a deterministic (continuous) dynamical system $f : M \to M$ on a compact space M, one typically studies the ergodic properties of f through an f-invariant probability measure on M. In the random setting, the probability measure μ defines a random walk (random dynamics) on $\mathbb{R}^d$: at each step, a point v moves to gv with probability $\mu(g)$. To obtain a compact phase space, we consider the induced projective action of $\mathrm{SL}_d(\mathbb{R})$ on $M = \mathbb{RP}^{d-1} = \mathbb{P}(\mathbb{R}^d)$ given by

$$g.[v] = [gv], \quad \forall [v] \in \mathbb{P}(\mathbb{R}^d), g \in \mathrm{SL}_d(\mathbb{R}).$$

Given a probability measure η on $\mathbb{P}(\mathbb{R}^d)$, we define the **convolution** $\mu * \eta$ to be the probability measure on $\mathbb{P}(\mathbb{R}^d)$ satisfying

$$\int_{\mathbb{P}(\mathbb{R}^d)} \phi([v])\, \mathrm{d}(\mu * \eta)([v]) = \int_{\mathbb{P}(\mathbb{R}^d)} \int_{\mathrm{SL}_d(\mathbb{R})} \phi([gv])\, \mathrm{d}\mu(g) \mathrm{d}\eta([v]),$$

for every continuous function ϕ on $\mathbb{P}(\mathbb{R}^d)$.

Definition 3.1. A probability measure ν on $\mathbb{P}(\mathbb{R}^d)$ is called **μ-stationary** if $\mu * \nu = \nu$. A μ-stationary measure ν is said to be **ergodic** if it cannot be written as a nontrivial convex combination of two distinct μ-stationary measures.

Remark 3.2 A stationary measure is invariant on average under the random walk.

Remark 3.3 Every $\mathrm{supp}\,\mu$-invariant measure is μ-stationary. But there may not exist a common invariant measure on $\mathbb{P}(\mathbb{R}^d)$ that is preserved by every $g \in \mathrm{supp}\,\mu$. We will see later that the assumptions in Theorem 2.9 in fact exclude this possibility.

Recall that the cocycle A induces a skew-product

$$F : X \times \mathbb{R}^d \to X \times \mathbb{R}^d, \quad (x, v) \mapsto (\sigma(x), A(x)v).$$

We may also consider its projectivization, given by

$$\mathbb{P}F : X \times \mathbb{P}(\mathbb{R}^d) \to X \times \mathbb{P}(\mathbb{R}^d), \quad (x, [v]) \mapsto (\sigma(x), [A(x)v]).$$

Recall that $X = (\mathrm{SL}_d(\mathbb{R}))^{\mathbb{N}}$ is endowed with the σ-invariant measure $\mathbf{P} = \mu^{\otimes \mathbb{N}}$. The following elementary observation reveals the connection between μ-stationary measures on $\mathbb{P}(\mathbb{R}^d)$ and $\mathbb{P}F$-invariant measures on the corresponding skew-product.

Proposition 3.4
1. A probability measure ν on $\mathbb{P}(\mathbb{R}^d)$ is μ-stationary iff $\mathbf{P} \times \nu$ is $\mathbb{P}F$-invariant.
2. A μ-stationary measure on $\mathbb{P}(\mathbb{R}^d)$ is ergodic iff $\mathbf{P} \times \nu$ is $\mathbb{P}F$-ergodic.

Exercise 3.5. Prove this proposition.

The existence of a stationary measure on $\mathbb{P}(\mathbb{R}^d)$ is guaranteed by the Krylov–Bogolyubov argument, as follows. For each n, let μ^{*n} (the **multiplicative convolution**) be the probability measure on $\mathrm{SL}_d(\mathbb{R})$ defined by $\mu^{*n} := (\mathrm{Mult}_n)_*(\mu^{\otimes n})$, where

$$\mathrm{Mult}_n : (\mathrm{SL}_d(\mathbb{R}))^n \to \mathrm{SL}_d(\mathbb{R}), \quad (g_1, \cdots, g_n) \mapsto g_1 \cdots g_n.$$

Let η be an arbitrary probability measure on $\mathbb{P}(\mathbb{R}^d)$, and let

$$\eta_n = \frac{1}{n}\sum_{k=0}^{n-1} \mu^{*n} * \eta.$$

Then η_n is a sequence of probability measures on $\mathbb{P}(\mathbb{R}^d)$. Because the space of probability measures on a compact metric space is compact with respect to the weak-$*$ topology. Let ν be a limit point of $(\eta_n)_{n=0}^{\infty}$, then ν is a μ-stationary measure on $\mathbb{P}(\mathbb{R}^d)$.

Properties of stationary measures

Proposition 3.6

Let ν be a μ-stationary measure on $\mathbb{P}(\mathbb{R}^d)$. If μ is strongly irreducible then $\eta(\mathbb{P}(V)) = 0$ for any proper projective subspace $\mathbb{P}(V)$.

Proof. Otherwise, let

$$d_0 := \min\{\dim V : \nu(\mathbb{P}(V)) > 0\} < d - 1.$$

Note that for any distinct subspaces $V_1, V_2 \subset \mathbb{R}^d$ with $\dim V_1 = \dim V_2 = d_0$, we have $\nu(\mathbb{P}(V_1 \cap V_2)) = 0$, since $\dim(V_1 \cap V_2) < d_0$. Hence, for every $c > 0$, the set

$$\mathcal{M}_c := \{\mathbb{P}(V) \subset \mathbb{P}(\mathbb{R}^d) : \dim V = d_0,\ \nu(\mathbb{P}(V)) \geqslant c\}$$

is finite. Let $c_0 > 0$ be the largest constant such that $\mathcal{M}_{c_0}$ is nonempty, and let

$$\mathcal{M} = \mathcal{M}_{c_0} = \{\mathbb{P}(V) \subset \mathbb{P}(\mathbb{R}^d) : \dim V = d_0,\ \nu(\mathbb{P}(V)) = c_0\}.$$

Since ν is μ-stationary, for every $\mathbb{P}(V) \in \mathcal{M}$, we have

$$c_0 = \nu(\mathbb{P}(V)) = \mu * \nu(\mathbb{P}(V)) = \int_{\mathrm{SL}_d(\mathbb{R})} \nu(\mathbb{P}(g^{-1}V))\, \mathrm{d}\mu(g) \leqslant c_0.$$

Equality holds only if $\mathbb{P}(g^{-1}V) \in \mathcal{M}$ for μ-almost every $g \in \mathrm{SL}_d(\mathbb{R})$. Therefore, the union $\bigcup_{\mathbb{P}(V)\in\mathcal{M}} V$ is invariant under g for μ-almost every $g \in \mathrm{SL}_d(\mathbb{R})$, which contradicts the strong irreducibility assumption. $\square$

Definition 3.7. A probability measure η on $\mathbb{P}(\mathbb{R}^d)$ is said to be **proper** if every proper projective subspace $\mathbb{P}(V) \subset \mathbb{P}(\mathbb{R}^d)$ has measure zero.

Now we consider an arbitrary matrix $B \neq 0 \in \mathrm{Mat}_d(\mathbb{R})$. Note that B need not have full rank. The induced projective map $B : \mathbb{P}(\mathbb{R}^d) \to \mathbb{P}(\mathbb{R}^d)$ s not defined on the projective subspace $\mathbb{P}(B^{-1}(0))$. However, if η is a proper probability measure on $\mathbb{P}(\mathbb{R}^d)$ then $\eta(\mathbb{P}(B^{-1}(0))) = 0$, so B is defined η-almost everywhere on $\mathbb{P}(\mathbb{R}^d)$. In particular, $B_*\eta$ is a well-defined probability measure on $\mathbb{P}(\mathbb{R}^d)$.

Exercise 3.8. Let η be a proper probability measure on $\mathbb{P}(\mathbb{R}^d)$. Let $(B_n)_{n=1}^{\infty} \subset \mathrm{Mat}_d(\mathbb{R})$ be a sequence of nonzero matrices such that $B_n \to B \neq 0$. Show that $(B_n)_*\eta \to B_*\eta$ in the weak* topology.

Proposition 3.9

Let η be a proper probability measure on $\mathbb{RP}^{d-1}$. Then the stabilizer $H(\eta) := \{B \in \mathrm{SL}_d(\mathbb{R}) : B_*\eta = \eta\}$ is a compact subgroup in $\mathrm{SL}_d(\mathbb{R})$.

Proof. The fact that $H(\eta)$ is a subgroup of $\mathrm{SL}(d, \mathbb{R})$ follows from the definition directly. It is also closed by the exercise. It suffices to show that the norm $\|B\|$ is bounded for $B \in H(\eta)$. Otherwise, let $(B_n)_{n=1}^{\infty} \subseteq H(\eta)$ be a sequence of matrices such that $\|B_n\| \to \infty$. Let $\widetilde{B}_n = B_n / \|B_n\|$. Passing to a subsequence if necessary, we can assume that $\widetilde{B}_n \to \widetilde{B} \in \mathrm{Mat}_d(\mathbb{R})$. Note that $\det \widetilde{B}_n = \|B_n\|^{-d} \to 0$. The limit matrix $\widetilde{B}$ does not have full rank. Therefore, $\widetilde{B}_*\eta$ is supported on a proper subspace and hence $\widetilde{B}_*\eta \neq \eta$. However, $(\widetilde{B}_n)_*\eta = B_{n*}\eta = \eta$ for every n. Hence $\widetilde{B}_*\eta = \lim (\widetilde{B}_n)_*\eta = \eta$. We get a contradiction. □

Corollary 3.10

Under the assumptions of Theorem 2.9, there is no common invariant measure on $\mathbb{P}(\mathbb{R}^d)$ that is preserved by every $g \in \operatorname{supp} \mu$.

§4 A Probabilistic proof for $d = 2$

Let $X = (\mathrm{SL}_2(\mathbb{R}))^{\mathbb{N}}$ and σ be the left shift map. Let $A : X \to \mathrm{SL}_2(\mathbb{R})$ be the cocycle defined as before. Let ν be a μ-stationary measure on $\mathbb{P}(\mathbb{R}^2)$, which is proper by the previous discussion. Then $\mathbf{P} \times \nu$ is invariant under the skew product $\mathbb{P}F : X \times \mathbb{P}(\mathbb{R}^d) \to X \times \mathbb{P}(\mathbb{R}^d)$.

We now interpret the Lyapunov exponent as a Birkhoff average. Note that the logarithm of the norm of matrix products is subadditive but not additive; hence it is more convenient to consider the induced action on vectors in $\mathbb{R}^2$. Accordingly, we define the measurable function

$$\Phi : X \times \mathbb{P}(\mathbb{R}^2) \to \mathbb{R}, \quad (x, [v]) \mapsto \log \frac{\|A(x)v\|}{\|v\|},$$

which is well-defined.

By Oseledets' multiplicative ergodic theorem (Theorem 0.2), one of the following holds for almost every $x \in X$,

- either for every $v \neq 0 \in \mathbb{R}^2$, we have $\frac{1}{n} \log \|A^n(x)v\| \to \lambda_+$;
- or there exists a one-dimensional subspace $L(x) \subset \mathbb{R}^2$ such that for every $v \in \mathbb{R}^2 \setminus L(x)$, we have $\frac{1}{n} \log \|A^n(x)v\| \to \lambda_+$.

Since ν is proper, we have $\nu(\mathbb{P}(L(x))) = 0$ even in the latter case. Consequently, for $(\mathbf{P} \times \nu)$-almost every $(x, [v]) \in X \times \mathbb{P}(\mathbb{R}^2)$,

$$\frac{1}{n} \log \frac{\|A^n(x)v\|}{\|v\|} \to \lambda_+.$$

On the other hand, the left-hand side is a Birkhoff average for the skew product:

$$\frac{1}{n} \sum_{k=0}^{n-1} \Phi \circ (\mathbb{P}F)^k(x, [v]) = \frac{1}{n} \left(\log \frac{\|A(x)v\|}{\|v\|} + \cdots + \log \frac{\|A^n(x)v\|}{\|A^{n-1}v\|} \right) = \frac{1}{n} \log \frac{\|A^n(x)v\|}{\|v\|}.$$

Recall Example 2.3, which shows that the divergence $\|A^n(x)\| \to \infty$ for $\mathbf{P}$-almost every x does not necessarily imply that $\lambda_+ > 0$. Nevertheless, the interpretation above allows us to apply the following proposition from ergodic theory, which reduces the theorem to showing that $\|A^n(x)v\| \to \infty$ for $(\mathbf{P} \times \nu)$-almost every $(x, [v])$.

Proposition 4.1

Let $T : (Y, m) \to (Y, m)$ be a measure preserving transformation on a probability space. If $\phi \in L^1(m)$ satisfying

$$\sum_{k=0}^{n-1} \phi \circ T^k \to +\infty$$

m-almost everywhere, then $\int \phi \, \mathrm{d}m > 0$.

Proof. Let

$$\overline{\phi} := \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \phi \circ T^k \geqslant 0.$$

Then $\int \phi = \int \overline{\phi} \geqslant 0$. Suppose $\int \phi = 0$ then $\overline{\phi} = 0$ almost everywhere.

For every integer $\ell \geqslant 0$, define

$$P_\ell := \left\{ y \in Y : \sum_{k=0}^{n-1} \phi \circ T^k(y) \geqslant 2^{-\ell}, \ \forall n \geqslant 1 \right\}, \quad Q_\ell := \bigcup_{k=0}^{\infty} T^{-k} P_\ell.$$

We first show that $\bigcup_{\ell=0}^{\infty} Q_\ell$ is a full-measure set.

Assume that $y \notin \bigcup_{\ell=0}^{\infty} Q_\ell$. We will construct a sequence of integers $0 = n_0 < n_1 < n_2 < \cdots$ inductively as follows. Suppose n_ℓ has been defined for some $\ell \geqslant 0$. Since $y \notin Q_\ell$, we have $T^{n_\ell}(y) \notin P_\ell$. Hence there exists an integer $n_{\ell+1} > n_\ell$ such that

$$\sum_{k=0}^{n_{\ell+1}-n_\ell-1} \phi \circ T^k(T^{n_\ell} y) < 2^{-\ell}.$$

Consequently,

$$\sum_{k=0}^{n_\ell - 1} \phi \circ T^k(y) = \sum_{\ell=0}^{\ell-1} \sum_{k=0}^{n_{\ell+1}-n_\ell-1} \phi \circ T^k(T^{n_\ell} y) < \sum_{\ell=0}^{\ell-1} 2^{-\ell} < 2.$$

Therefore, $\sum_{k=0}^{n-1} \phi \circ T^k(y)$ does not diverge to $+\infty$. This shows that $\bigcup_{\ell=0}^{\infty} Q_\ell$ must be a full-measure set.

Now we compare $\overline{\phi}$ with the Birkhoff average of $\mathbb{1}_{P_\ell}$. For every $y \in Q_\ell$, let n_1 be the smallest positive integer such that $T^{n_1}(x) \in P_\ell$. Then we have

$$\sum_{k=0}^{n-1} \phi \circ T^k(y) \geqslant \sum_{k=0}^{n_1-1} \phi \circ T^k(y) + 2^{-\ell} \sum_{k=n_1}^{n-1} \mathbb{1}_{P_\ell} \circ T^k(y).$$

Let $\overline{\mathbb{1}}_{P_\ell}$ be the Birkhoff average of $\mathbb{1}_{P_\ell}$. We obtain $\overline{\phi} \geqslant 2^{-\ell} \overline{\mathbb{1}}_{P_\ell}$ almost everywhere on Q_ℓ. By our assumption that $\int \overline{\phi} = 0$, we have

$$0 \geqslant \int_{Q_\ell} \overline{\mathbb{1}}_{P_\ell} \, \mathrm{d}m = \int \mathbb{1}_{P_\ell} \, \mathrm{d}m = m(P_\ell).$$

Hence $m(Q_\ell) = 0$ for every $\ell \geqslant 0$. This contradicts that $\bigcup_{\ell=0}^{\infty} Q_\ell$ is a full-measure set. □

Convergence of measures

For $g \in \mathrm{SL}_2(\mathbb{R})$, let g^t be the transpose of g. Let μ^t be the probability measure on $\mathrm{SL}_2(\mathbb{R})$ which is the push forward of μ under the transpose map. Let $\widetilde{\nu}$ be a μ^t-stationary measure on $\mathbb{P}(\mathbb{R}^2)$, which is also proper.

Lemma 4.2

For $\mathbf{P}$ almost every $x \in X$, there exists a probability measure $\widetilde{\nu}_x$ on $\mathbb{P}(\mathbb{R}^2)$ such that

$$(A^n(x)^t)_*\widetilde{\nu} \to \eta_x$$

as $n \to +\infty$ in the weak* topology. Moreover, for μ-almost every $g \in \mathrm{SL}_2(\mathbb{R})$, we also have $(A^n(x)^t g^t)_*\widetilde{\nu} \to \eta_x$

Proof. For every continuous function ϕ on $\mathbb{P}(\mathbb{R}^2)$, define

$$P = P_\phi : \mathrm{SL}_2(\mathbb{R}) \to \mathbb{P}(\mathbb{R}), \quad g \mapsto \int_{\mathbb{P}(\mathbb{R}^1)} \phi([gv])\, \mathrm{d}\widetilde{\nu}([v]).$$

Recall that $X = (\mathrm{SL}_2(\mathbb{R}))^{\mathbb{N}}$ is endowed with the Borel σ-algebra $\mathscr{B}$. Let $\mathscr{B}_n$ denote the sub-σ-algebra consisting of all sets of the form $C \times (\mathrm{SL}_2(\mathbb{R}))^{\mathbb{N}}$, where C is a Borel measurable subset of $(\mathrm{SL}_2(\mathbb{R}))^n$. Then we have

$$\begin{aligned}
\mathbf{E}(P(A^{n+1}(x)^t)|\mathscr{B}_n) &= \int P(A^n(x)^t g^t)\, \mathrm{d}\mu(g) \\
&= \iint \phi([A^n(x)^t gv])\, \mathrm{d}\mu^t(g)\mathrm{d}\widetilde{\nu}([v]) \\
&= \int \phi([A^n(x)v])\, \mathrm{d}(\mu^t * \widetilde{\nu})([v]) \\
&= \int \phi([A^n(x)v])\, \mathrm{d}\widetilde{\nu}([v]) = P(A^n(x)^t).
\end{aligned}$$

Hence $(P_\phi(A^n(x)^t))_{n=1}^\infty$ is a martingale with respect to $(\mathscr{B}_n)_{n=1}^\infty$. Moreover, we have

$$\mathbf{E}\, P(A^n(x)^t)^2 = \int_X \left(\int_{\mathbb{P}(\mathbb{R}^1)} \phi([A^n(x)^t v])\, \mathrm{d}\widetilde{\nu}([v]) \right)^2 \mathrm{d}\mathbf{P}(x) \leqslant \|\phi\|_\infty^2.$$

Thus, $(P\phi(A^n(x)^t))n = 1^\infty$ is in fact an L^2-bounded martingale. By Doob's martingale convergence theorem, there exists $\mathcal{L}\phi \in L^2(X, \mathbf{P})$ such that $P_\phi(A^n(x)^t) \to \mathcal{L}\phi(x)$ both almost surely and in L^2. In particular, $(P_\phi(A^n(x)^t))_{n=1}^\infty$ is an L^2-Cauchy sequence. This implies that

$$\mathbf{E}\left(\iint |\phi([A^n(x)^t g^t v]) - \phi([A^n(x)^t v])|^2\, \mathrm{d}\mu(g)\mathrm{d}\zeta([v]) \right) \to 0,$$

which shows that $P_\phi(A^n(x)^t g^t) \to \mathcal{L}\phi$ for μ-almost every g.

By choosing a countable dense subset of $C(\mathbb{P}(\mathbb{R}^2))$, we may find a $\mathbb{P}$-full measure subset of X such that $\mathcal{L}\phi(x)$ exists for all $\phi \in C(\mathbb{P}(\mathbb{R}^2))$ whenever x belongs to this subset. For every such x, define $\mathcal{L}_x : \phi \mapsto \mathcal{L}\phi(x)$, which is a positive linear functional on $C(\mathbb{P}(\mathbb{R}^2))$ satisfying $\mathcal{L}_x \mathbb{1} = 1$. Therefore, $\mathcal{L}_x$ defines a probability measure η_x on $C(\mathbb{P}(\mathbb{R}^2))$, as desired. □

Lemma 4.3 The limit measure η_x is a Dirac measure $\delta_{[u(x)]}$.

Proof. Fix a typical point x, we know that $(A^n(x)^t)_*\widetilde{\nu} \to \zeta_x$ and $A^n(x)^t g^t \widetilde{\nu} \to \eta_x$ for μ-almost every g. Choose a limit point of $\|A^n(x)^t\|^{-1} A^n(x)^t$, denoted by B. Then

$$B_*\widetilde{\nu} = (Bg)_*\widetilde{\nu} = \eta_x$$

for μ^t-almnost every $g \in \mathrm{SL}_2(\mathbb{R})$. If B is invertible, then $\widetilde{\nu} = g_*\widetilde{\nu}$ for μ^t-almost every g. This contradicts Proposition 3.9 since μ is non-compact. Therefore B is non-invertible. Then $\operatorname{rank} B = 1$ since $\|B\| = 1$. Let $u(x) \in B(\mathbb{R}^2)$ be a unit vector. Then $\eta_x = \delta_{[u(x)]}$. □

Remark 4.4 The proof also implicitly shows that $[u(x)]$ is independent of the choice of stationary measures. Moreover, by the construction of η_x, we have

$$\int \delta_{[u(x)]}\, \mathrm{d}\mathbf{P}(x) = \int \eta_x\, \mathrm{d}\mathbf{P}(x) = \lim \int (A^n(x)^t)_*\widetilde{\nu}\, \mathrm{d}\mathbf{P}(x) = \lim \widetilde{\nu} = \widetilde{\nu},$$

which implies that $\widetilde{\nu}$ is the unique μ^t-stationary measure.

Exercise 4.5. Use the dominated convergence theorem to give a precise proof of the above equality.

Replacing μ^t by μ, we conclude that every non-compact and strongly irreducible measure μ on $\mathrm{SL}_2(\mathbb{R})$ admits a unique stationary measure on $\mathbb{P}(\mathbb{R}^2)$. For general d-dimensional cases, the non-compactness assumption should be replaced by the proximality condition.

Proof of Theorem 2.9 for $d = 2$. Let $x \in X$ be a typical point. Then $|A^n(x)| \to \infty$. Otherwise, the sequence $(A^n(x)^t)$ would have a limit point $B \in \mathrm{SL}_2(\mathbb{R})$, which would imply $B_*\widetilde{\nu} = \delta_{[u(x)]}$, a contradiction.

Now consider a limit point of $\|A^n(x)\|^{-1}A^n(x)^t$, denoted by $B(x)$. As in the proof of Lemma 4.3, we have $\operatorname{rank} B(x) = 1$ and $B(\mathbb{R}^2) = \mathbb{R} \cdot u(x)$. Then

$$\frac{\|A^n v\|}{\|A^n\|} = \sup_{\|w\|=1} \left\langle \frac{A^n v}{\|A_n\|}, w \right\rangle = \sup_{\|w\|=1} \left\langle v, \frac{(A^n)^t w}{\|A_n\|} \right\rangle \to \sup_{\|w\|=1} \langle v, Bw \rangle = |\langle v, u(x) \rangle|.$$

In particular, $\|A^n(x)v\| \to \infty$ unless $v \perp u(x)$. Let ν be the stationary measure of μ, which is proper. Then the previous discussion shows that $\|A_n(x)v\| \to \infty$ for $(\mathbf{P} \times \nu)$-almost every $(x, [v])$. The theorem follows. □

§5 Invariance Principles

We now present another approach to Theorem 2.9, based on the invariance principle. The general philosophy is that if the extremal Lyapunov exponents λ_+ and λ_- coincide, then the entropy of the corresponding projective action must vanish, which in turn forces the existence of additional invariant structures for the measure. This type of invariance principle appears in several contexts, such as Ledrappier (LNM, 1986), Viana (Ann. of Math., 2008), and Avila–Viana (Invent. Math., 2010). We refer the reader to Chapter 7 of Viana's textbook *Lectures on Lyapunov Exponents* for further details.

Theorem 5.1 (Invariance principle for random products, Furstenberg)
If $\lambda_+ = 0$ then every μ-stationary measure ν on $\mathbb{P}(\mathbb{R}^d)$ is invariant for every $g \in \operatorname{supp} \mu$.

Under the assumptions in Theorem 2.9, we know that there is no common invariant measure on $\mathbb{P}(\mathbb{R}^d)$ for all $g \in \operatorname{supp} \mu$ by Corollary 3.10. Hence Theorem 2.9 follows directly from this invariance principle.

Now we explain the idea of the proof for the invariance principle. Recall that $X = (\mathrm{SL}_d(\mathbb{R}))^{\mathbb{N}}$ and $\mathbf{P} = \mu^{\otimes \mathbb{N}}$ as above. The linear cocycle A leads to the skew-product $\mathbb{P}F : X \times \mathbb{P}(\mathbb{R}^d) \to$

$\mathbb{P}(\mathbb{R}^d)$. Every μ-stationary measure ν on $\mathbb{P}(\mathbb{R}^d)$ corresponds to a $\mathbb{P}F$-invariant measure $\mathbf{P} \times \nu$. We may consider a general $\mathbb{P}F$-invariant measure m on $X \times \mathbb{P}(\mathbb{R}^d)$ that projects to $\mathbf{P}$.

Definition 5.2. Let m be an probability measure on the product space $X \times \mathbb{P}(\mathbb{R}^d)$ that projects to the probability measure $\mathbf{P}$ on X. A **disintegration** of m along vertical fibers is a measurable family $\{m_x : x \in X\}$ of probability measures on $\mathbb{P}(\mathbb{R}^d)$ satisfying

$$m(E) = \int_X m_x\{v : (x, [v]) \in E\} \, \mathrm{d}\mathbf{P} \quad \text{for any measurable } E \subset X \times \mathbb{P}(\mathbb{R}^d).$$

The measures m_x on each fiber are called the **conditional probabilities** of m.

Fact 5.3. A disintegration along a vertical fiber does exist. Moreover, the disintegration is unique up to a full $\mathbf{P}$-measure set.

We have the following invariance principle for linear cocycles. It in fact applies to much more general base dynamical systems, not only to the Bernoulli shifts $(X, \mathbf{P}, \sigma)$.

Theorem 5.4 (Invariance principle, Ledrappier)

Assume that $\lambda_-(x) = \lambda_+(x)$ for $\mathbf{P}$-almost every $x \in X$. Then for any disintegration $\{m_x : x \in X\}$ of any $\mathbb{P}F$-invariant probability measure m on $X \times \mathbb{RP}^{d-1}$ that projects down to $\mathbf{P}$, we have

$$m_{\sigma(x)} = A(x)_* m_x \quad \text{for } \mathbf{P}\text{-almost every } x \in X.$$

To establish this invariance principle, we may consider the entropy along fibers. For each $x \in X$, let

$$A(x)_*^{-1} m_{\sigma(x)} = \zeta_x + \xi_x,$$

where $\zeta_x \ll m_x$ and $\xi_x \perp m_x$. Let $J(x, \cdot)$ be the Radon-Nikodym derivate of ζ_x with respect to m_x, then we have

$$\mathrm{d}A(x)_*^{-1} m_{\sigma(x)} = J(x, \cdot) \, \mathrm{d}m_x + \mathrm{d}\xi_x.$$

Remark 5.5 The Radon-Nikodym derivate J reflects the contraction of $A(x)$ on the projective space with respect to the conditional probability on each fiber.

Definition 5.6. The **fibered entropy** of m is defined by

$$h(m) := -\int \log J(x, [v]) \, \mathrm{d}m(x, [v]).$$

Lemma 5.7

The fibered entropy $h(m)$ is always non-negative. If $h(m) = 0$ then $A(x)_* m_x = m_{\sigma(x)}$ holds for $\mathbf{P}$-almost every $x \in X$.

Proof. By Jensen's inequality, we have

$$h(m) = \int_{\{J>0\}} -\log J \, \mathrm{d}m + \infty \cdot m\{J = 0\} \geqslant -\log \int_{\{J>0\}} J \, \mathrm{d}m + \infty \cdot m\{J = 0\} \geqslant 0.$$

Equality holds only if all of the following conditions are satisfied:

- $m\{J = 0\} = 0$;
- $\log J$ is a constant m-almost everywhere;
- $\int J \mathrm{d}m = 1$.

The last condition implies that $\xi_x = 0$ for $\mathbf{P}$-almost every $x \in X$. Together with the first two conditions, we conclude that $J \equiv 1$ for $\mathbf{P}$-almost everywhere. This establishes the additional invariance of conditional measures along fibers. □

Besides, we have an upper bound for the fibered entropy in terms of the Lyapunov exponents. The difference between the extremal Lyapunov exponents controls the rate of contraction on $\mathbb{P}(\mathbb{R}^d)$ with respect to the projective metric (see the exercise below), while the fibered entropy measures the average contraction on $\mathbb{P}(\mathbb{R}^d)$ with respect to the conditional probabilities.

Exercise 5.8. We fix an $\mathrm{SO}_d(\mathbb{R})$-invariant metric on $\mathbb{P}(\mathbb{R}^d)$, defined by

$$d([u],[v]) := \frac{\|u \wedge v\|}{\|u\| \cdot \|v\|}, \quad \forall [u],[v] \in \mathbb{P}(\mathbb{R}^d).$$

For every $[v] \in \mathbb{P}(\mathbb{R}^d)$, we can identify the tangent space $T_{[v]}\mathbb{P}(\mathbb{R}^d)$ with the orthogonal complement $v^\perp \subset \mathbb{R}^d$.

(1) Assume that $g \in \mathrm{SL}_d(\mathbb{R})$ is a diagonal matrix $g = \mathrm{diag}(a_1, \cdots, a_d)$ with $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_d > 0$. Show that

$$\|D_{[v]}g(w)\| \geqslant \frac{a_d}{a_1}\|w\|, \quad \forall [v] \in \mathbb{P}(\mathbb{R}^d), w \in T_{[v]}\mathbb{P}(\mathbb{R}^d).$$

(2) Using the singular value decomposition (Cartan decomposition), show that for every $g \in \mathrm{SL}_d(\mathbb{R})$,

$$\|D_{[v]}g(w)\| \geqslant \frac{\|w\|}{\|g\| \cdot \|g^{-1}\|}, \quad \forall [v] \in \mathbb{P}(\mathbb{R}^d), w \in T_{[v]}\mathbb{P}(\mathbb{R}^d).$$

As an analogue of Ruelle's inequality in smooth ergodic theory, it is therefore natural that the fibered entropy is bounded above by this difference. The following proposition makes this relationship precise.

Proposition 5.9 $0 \leqslant h(m) \leqslant d(\lambda_+ - \lambda_-)$.

The proof is more technical and we refer to Viana's textbook *Lectures on Lyapunov Exponents* for detailed proofs. In particular, if $\lambda_+ = \lambda_-$ then Proposition 5.9 forces $h(m) = 0$. Together with Lemma 5.7, we obtain Theorem 5.4.