Notes on Furstenberg Theorem

Ajorda Jiao

Contents

§1 Introduction

Recall strong law of large numbers.

Theorem 1.1 (Strong Law of Large Numbers)
$X_0, X_1, \cdots, X_n, \cdots$ a sequence of i.i.d. random values, $\mathbb{E}|X_0| < \infty$, then

$$\frac{1}{n}(X_0 + X_1 + \cdots + X_{n-1}) \to \mathbb{E}\, X_0 \text{ a.s. }.$$

Remark 1.2 — It can be regarded as a corollary of Birkhoff's ergodic theorem.

Corollary 1.3
$X_0 > 0, \mathbb{E} \log X_0 > 0$, then

$$X_0 X_1 \cdots X_{n-1} \to \infty, \quad \text{exponentially fast a.s. }.$$

We want to generalized this result to some non-commutative case. Let ν be a probability measure on $\mathrm{SL}(d, \mathbb{R})$, let $A_0, A_1, \cdots, A_n, \cdots$ be a sequence of i.i.d. random matrices with connen distribution ν. Let

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

we want to show that $\|A^n\| \sim e^{\lambda n}$ under some assumptions. A natural integrable condition is

$$\int \log \|A_0\| \, \mathrm{d}\nu < \infty.$$

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 the lower Lyapunov exponent, respectively.

Definition 1.5. Let $\widetilde{A} := \lim_{n\to\infty}(A^{n*}A^n)^{\frac{1}{2n}}$, assume the eigenvalues of $\widetilde{A}$ are

$$e^{\lambda_1} \geqslant e^{\lambda_2} \geqslant \cdots \geqslant e^{\lambda_d}.$$

The set $\{\lambda_1, \lambda_2, \cdots, \lambda_d\}$ is called **Lyapunov spectrum**.

Remark 1.6 — We use the Lyapunov exponents to measure the increasing speed. Our aim is to proof $\lambda_+ > 0$ under some assumptions.

§2 Cocycles and ergodic theorems

Definition 2.1. Let $(X, \mathcal{F}, \mu)$ be a probability space and $f : X \to X$ be a measure-preserving map. Let $A : X \to \mathrm{GL}(d, \mathbb{R})$ be a measurable function. The **linear cocycle** defined by A over f is the transformation:

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

Definition 2.2. Let $F : X \times \mathbb{R}^d \to X \times \mathbb{R}^d$ be a linear cocycle. The **projective cocycle** associated with it is defined as

$$\mathbb{P}F : X \times \mathbb{RP}^{d-1} \to X \times \mathbb{RP}^{d-1}, \quad (x, [v]) \mapsto (f(x), [A(x)v]).$$

Example 2.3

Take $X = \mathrm{SL}(d, \mathbb{R})^{\mathbb{N}}$ with probability measure $\mu = \nu^{\mathbb{N}}$. Let $f : X \to X$ be the shift map. The measurable function $A : X \to \mathrm{GL}(d, \mathbb{R})$ is defined as $x = (A_0, A_1, \cdots) \mapsto A_0$. Let $A^n(x) = A_{n-1} \cdots A_1 A_0$, consider the linear cocycle defined by A over f, denoted by $F : X \times \mathbb{R}^d \to X \times \mathbb{R}^d$. Then $F^n(x, v) = (f^n x, A^n(x)v)$.

For general linear cocycle F, assume $F^n(x, v) = (f^n x, A^n(x)v)$, we can defined the Lyapunov exponents of F by $A^n(x)$.

Then the Lyapunov exponents of a random matrices sequence is identified with the Lyapunov exponents of the linear cocyle F constructed in the example above.

The following two ergodic theorems guarantee the existence of Lyapunov exponents and Lyapunov spectrum. Moreover, (X, f) is ergodic with respect to μ, hence the f-invariance of the Lyapunov exponents implies that they are constants almost everywhere.

Theorem 2.4 (Kingman's Sub-additive Ergodic Theorem)
Let (X, μ) be a probability space and $f : X \to X$ be a measure preserving map. Let $(g_n)_{n=1}^{\infty}$ be a sequence of measurable functions such that $g_1^+ \in L^1(X, \mu)$, satisfying the subadditivity condition

$$g_{n+m} \leqslant g_m + g_n \circ f^m \quad \text{for all } m, n \geqslant 1.$$

Then there exists an f-invariant function $g : X \to \mathbb{R} \cup \{-\infty\}$, such that

$$\frac{1}{n} g_n \to g \quad \mu\text{-a.s.}.$$

Moreover,

$$\frac{1}{n} \int g_n \mathrm{d}\mu \to \int g \mathrm{d}\mu = \inf_{n \geqslant 1} \frac{1}{n} \int g_n \mathrm{d}\mu.$$

Theorem 2.5 (Oseledets' Multiplicative Ergodic Theorem)
Let F be a linear cocycle on $(X, \mathcal{F}, \mu)$ defined by $A : X \to \mathrm{GL}(d, \mathbb{R})$ over $f : X \to X$ satisfying the integrablity condition $\log^+ \|A(\cdot)\| \in L^1(X, \mu)$. Then there exists a forward invariant set $\tilde{X} \in \mathcal{F}$ with full measure such that for each $x \in \tilde{X}$, the following statements hold:

(i) $\bar{A}(x) \coloneqq \lim_{n\to\infty} (A^n(x)^* A^n(x))^{\frac{1}{2n}}$ exists.

(ii) Let $e^{\lambda_{p(x)}(x)} < \cdots < e^{\lambda_2(x)} < e^{\lambda_1(x)}$ be the different eigenvalues of $\bar{A}(x)$ and let $U_{p(x)}(x), \cdots, U_2(x), U_1(x)$ be the corresponding eigenspaces with multiplicities $d_i(x) \coloneqq \dim U_i(x)$. Then

$$p(x) = p(fx), \quad \lambda_i(x) = \lambda_i(fx), \quad d_i(x) = d_i(fx).$$

(iii) Put $V_{p(x)+1}(x) \coloneqq \{0\}$, and for $i = 1, 2, \cdots, p(x)$, $V_i(x) = U_{p(x)} \oplus \cdots U_i(x)$, so that

$$V_{p(x)}(x) \subset \cdots \subset V_i(x) \subset \cdots \subset V_1(x) = \mathbb{R}^d$$

defined a filtration of $\mathbb{R}^d$. For each $v \in V_i(x) \setminus V_{i+1}(x)$, the Lyapunov exponent of v exists and coincides with $\lambda_i(x)$, i.e.,

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)v\| = \lambda_i(x).$$

(iv) For all $i = 1, 2, \cdots, p(x)$, $A(x)V_i(x) \subset V_i(fx)$.

Moreover, the maps $x \mapsto p(x)$, $x \mapsto \lambda_i(x)$, $x \mapsto d_i(x)$, $x \mapsto U_i(x)$, $x \mapsto V_i(x)$(the last two convergences are in the sense of $X \to \cup_{k=1}^{d} G_k(d)$, where $G_k(d)$ is the Grassmannian manifold of k-dimensional subspaces of $\mathbb{R}^d$) are measurable.

§3 Main results and some examples

Definition 3.1. We call ν is **irreducible**, if there is no proper subspace $V \subseteq \mathbb{R}^d$, such that $A(V) \subseteq V$ for ν-a.e. A.

Definition 3.2. We call ν is **strongly irreducible**, if there is no proper subspace $V \subseteq \mathbb{R}^d$, such that $A(V) \subseteq V$ for ν-a.e. A.

Definition 3.3. We call ν is **non-compact**, if the support of ν is not contained in a compact subgroup of $\mathrm{SL}(d, \mathbb{R})$.

Remark 3.4 — ν is compact if and only if there exists $P \in \mathrm{SL}(d, \mathbb{R})$ such that $\nu(P^{-1}\mathrm{SO}(d, \mathbb{R})P) = 1$.

Theorem 3.5 (Furstenberg)

Strongly irreducible + non-compact $\implies \lambda_+ > 0 > \lambda_-$.

Let T_ν be the semigroup generated by $\operatorname{supp} \nu$.

Definition 3.6. We call ν is **contracting**, if $\exists \{B_n\} \subseteq T_\nu$ such that $\|B_n\|^{-1} B_n \to B$ with $\operatorname{rank} B = 1$.

Remark 3.7 — Contracting is stronger than non-compact, because non-compact just guarantee that $\operatorname{rank} B \leqslant d - 1$.

Definition 3.8. We call ν is ***p*-strongly irreducible** or ***p*-contracting**, if the action of $(\mathrm{SL}(d, \mathbb{R}), \nu)$ on $\wedge^p(\mathbb{R}^d)$ is strongly irreducible or contracting, respectively.

Theorem 3.9 (Furstenberg)

Strongly irreducible + contracting $\implies \lambda_1 > 0 > \lambda_2$.

Theorem 3.10 (Furstenberg)

p-strongly irreducible + p-contracting $\implies$ Lyapunov spectrum is simple.

Another result is proved by Gol'dsheid and Margulis, which conditions are much easier to verify.

Theorem 3.11 (Gol'dsheid, Margulis)

Assume ν is a probability measure on $\mathrm{GL}(d, \mathbb{R})$ satisfies the integrability condition and T_ν is Zariski dense in $\mathrm{GL}(d, \mathbb{R})$. Then the Lyapunov spectrum is simple.

Some examples

We show some counter examples for $d = 2$.

Example 3.12

If ν supports on $P^{-1}\mathrm{SO}(2, \mathbb{R})P$ for some $P \in \mathrm{SL}(2, \mathbb{R})$, then $\|A^n\| \leqslant \left\|P^{-1}\right\| \|P\|$ is bounded almost everywhere.

Example 3.13

If ν admits an invariant direction:

Consider ν supports on $\left\{\begin{bmatrix}1 & s\\ 0 & 1\end{bmatrix} : s \in \mathbb{R}\right\}$. Assume $A_0 = \begin{bmatrix}1 & X\\ 0 & 1\end{bmatrix}$ where X is a positive random value with $0 < \mathbb{E}\, X < \infty$. Then $\|A^n\| \to \infty$ but just with linear speed.

Example 3.14

If two directions preserved by ν-a.e. A :

Let $M_1 = \begin{bmatrix}e & 0\\ 0 & e^{-1}\end{bmatrix}, M_2 = \begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}$, let $\nu = \frac{1}{2}\delta_{M_1} + \frac{1}{2}\delta_{M_2}$. As we can let t_k be the k-th n with $A_n = M_2$ and $\xi_k = t_k - t_{k-1} - 1$. Then

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

§4 Stationary measures

For every element $A \in \mathrm{GL}(d, \mathbb{R})$ and probability measure η on $\mathbb{RP}^{d-1}$. The linear map $A : \mathbb{RP}^{d-1} \to \mathbb{RP}^{d-1}, [v] \mapsto [Av]$ induces a probability measure $A_*\eta$ on $\mathbb{RP}^{d-1}$.

Let ν be a probability measure on $\mathrm{GL}(d, \mathbb{R})$(or $\mathrm{SL}(d, \mathbb{R})$) and η be a probability measure on $\mathbb{RP}^{d-1}$. Then we define the convolution $\nu * \eta$ be the probability measure on $\mathbb{RP}^{d-1}$ such that for any continuous function f on $\mathbb{RP}^{d-1}$,

$$\int f([v])\mathrm{d}(\nu * \eta)([v]) = \iint f([Av])\mathrm{d}\nu(A)\mathrm{d}\eta([v]).$$

Or we can write as

$$\nu * \eta = \int A_*\eta \mathrm{d}\nu(A).$$

Now, we define the ν-stationary measure on $\mathbb{RP}^{d-1}$. That is if we consider $\mathrm{GL}(n, d)$ acting on $\mathbb{RP}^{d-1}$ with law ν, it gives a random walk on the projective space. The stationary measure is the probability measure on $\mathbb{RP}^{d-1}$ which is invariant under the random walk.

Definition 4.1. A probability measure η on $\mathbb{RP}^{d-1}$ is called a **ν-stationary measure** if $\nu * \eta = \nu$.

As an analogue to the existence of invariant probability measure of the continuous map on a compact metric space, the following proposition tells us the stationary measure always exist on the projective space.

Proposition 4.2

For any probability measure ν on $\mathrm{GL}(n, d)$, there are some ν-stationary measure on $\mathbb{RP}^{d-1}$.

Proof. Let ξ be an arbitrary probability measure on $\mathbb{RP}^{d-1}$, and let

$$\xi_n = \frac{1}{n}\sum_{k=0}^{n-1} \nu^k * \xi.$$

Then ξ_n is a sequence of probability measure on $\mathbb{RP}^{d-1}$. Because the space of probability measures on a compact metric space is compact with respect to the weak-$*$ topology. Let η be a limit point of $(\xi_n)_{n=0}^{\infty}$, then η is a ν-stationary measure on $\mathbb{RP}^{d-1}$. $\square$

Properties of stationary measures

Now we focus on the probability measure ν on $\mathrm{SL}(d,\mathbb{R})$ which satisfies two conditions of Furstenberg's theorem. Let F be the linear cocycle on $\mathrm{SL}(d,\mathbb{R})^{\mathbb{N}} \times \mathbb{R}^d$ which we construct in the example before.

Proposition 4.3

Let η be a ν-stationary measure on $\mathbb{RP}^{d-1}$. If the cocycle F is strongly irreducible then $\eta(V) = 0$ for any proper projective subspace V.

Proof. Otherwise, let d_0 be the smallest dimension such there are some subspaces V with positive measure. Let c be the largest measure of those subspaces. Let

$$\mathcal{M} = \{V \subseteq \mathbb{RP}^{d-1} : \dim V = d_0, \eta(V) = c\}.$$

Then for every $V_1 \neq V_2 \in M$, we have $\eta(V_1 \cap V_2) = 0$, hence $\mathcal{M}$ is finite. Because η is ν-invariant, then for every $V \in \mathcal{M}$, $A^{-1}V \in \mathcal{M}$ for ν-a.e. A. Hence $\mathcal{M}$ is a finite collection of proper subspaces of $\mathbb{RP}^{d-1}$ which is invariant under ν almost every $A \in \mathrm{SL}(d,\mathbb{R})$. $\square$

Remark 4.4 — The probability measure on the projective space with this property is said to be **proper**.

For any proper probability measure ζ on the projective space, we can define $B_*\zeta$ for any $B \neq 0 \in M(d,\mathbb{R})$. Note that for $B_n, B \in M(d,\mathbb{R}) \setminus \{0\}$, if $B_n \to B$, then $(B_n)_*\zeta \to B_*\zeta$ in the weak* topology.

Proposition 4.5

Let ζ be a probability measure on $\mathbb{RP}^{d-1}$ such that the measure of each proper subspace is zero. Then the stabilizer $H(\zeta) := \{A \in \mathrm{SL}(d,\mathbb{R}) : A_*\zeta = \zeta\}$ is a compact subgroup in $\mathrm{SL}(d,\mathbb{R})$.

Proof. The fact that $H(\zeta)$ is a subgroup of $\mathrm{SL}(d,\mathbb{R})$ follows from the definition directly. It is also closed in the weak-$*$ topology. It suffices to show that the norm $\|B\|$ is bounded for $B \in H(\zeta)$. Otherwise, let $(B_n)_{n=1}^{\infty} \subseteq H(\zeta)$ be a sequence of matrices such that $\|B_n\| \to \infty$. Let $\widetilde{B_n} = B_n/\|B_n\|$, by passing to a subsequence, without lost of generality, suppose that $\widetilde{B_n} \to \widetilde{B}$. We have $\widetilde{B_n}_*\zeta = B_{n*}\zeta = \zeta \to \zeta$, but $\widetilde{B}$ does not have full rank which contradicts with the condition of ζ. $\square$

§5 Original proof of $d = 2$

Now we focus on $d = 2$ and show the original proof of Furstenberg. Let $(X,\mu) = (\mathrm{SL}(2,\mathbb{R})^{\mathbb{N}}, \nu^{\mathbb{N}})$, $F : X \times \mathbb{R}^2 \to X \times \mathbb{R}^2$ be the linear cocycle and $\mathbb{P}F : X \times \mathbb{RP}^1 \to X \times \mathbb{RP}^1$ be the associated projective cocycle. Let η be a ν-stationary measure on $\mathbb{RP}^1$ which is proper by the previous discussion. Then $\mu \times \eta$ is invariant under $\mathbb{P}F$.

By Oseledets' multiplicative ergodic theorem, for μ-almost every $x \in X$, there exists an one dimensional subspace $E(x) \in \mathbb{R}^2$ such that for all $v \in \mathbb{R}^2 \setminus E(x)$, it holds $\frac{1}{n}\log\|A^n(x)v\| \to \lambda_+$. Because η is proper, we conclude that for $m = \mu \times \eta$ all $(x, [v]) \in X \times \mathbb{RP}^1$, $\frac{1}{n}\log\|A^n(x)v\| \to \lambda_+$.

Now, we define $\Phi : X \times \mathbb{RP}^1$ given by $\Phi(x, [v]) = \log \frac{\|A(x)v\|}{\|v\|}$. Then,

$$\frac{1}{n}\sum_{k=0}^{n} \Phi \circ F^k(x, [v]) = \frac{1}{n}\log\frac{\|A^n(x)v\|}{\|v\|} \to \lambda_+ \quad m\text{-a.e.}.$$

Note that the left hand side also tends to Birkhoff average of Φ. Hence we get an identity

$$\lambda_+ = \int \Phi \mathrm{d}m = \iint \frac{\log\|A(x)v\|}{\|v\|}\mathrm{d}\nu\mathrm{d}\eta.$$

This identity tells us by applying the stationary measure, we can regard the Lyapunov exponent as a Birkhoff average. The following proposition shows that if suffices to prove the Birkhoff sum divergent almost everywhere.

Proposition 5.1

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

$$\sum_{k=0}^{n-1} \varphi \circ T^k \to +\infty \quad m\text{-a.s.},$$

then $\int \varphi \mathrm{d}m > 0$.

Now, we reduce the problem to show that $\|A^n(x)v\| \to \infty (n \to \infty)$ for m-a.e. $(x, [v])$.

Convergence of measures

For $A \in \mathrm{SL}(2, \mathbb{R})$, let A^* be the transpose of A. For ν be a probability measure on $\mathrm{SL}(2, \mathbb{R})$, let ν^* be the probability measure on $\mathrm{SL}(2, \mathbb{R})$ push forward of ν under the transpose. Let ζ be a ν^*-stationary measure on $\mathbb{RP}^1$, which is also proper.

Lemma 5.2

For μ-a.e. $x \in X$, there exists a probability measure ζ_x on $\mathbb{RP}^1$ such that

$$(A^n(x)^*)_*\zeta \to \zeta_x.$$

Moreover, for ν^*-a.e. $B \in \mathrm{SL}(2, \mathbb{R})$, $(A^n(x)^*B)_*\zeta \to \zeta_x$ in the weak* sense.

Proof. Fix $f \in C(\mathbb{RP}^1)$, consider the function $P : \mathrm{SL}(n, \mathbb{R}) \to \mathbb{R}$ given by

$$P(A) = \int_{\mathbb{RP}^1} f([Av])\mathrm{d}\zeta([v]).$$

Suppose $\mathcal{F}$ is the σ-algebra associated with the probability space (X, μ). Let $\mathcal{F}_n$ be the sub σ-algebra of $\mathcal{F}$ generated by the cylinders of length n. Then we have

$$\begin{aligned}\mathbb{E}(P(A^{n+1}(x)^*)|\mathcal{F}_n) &= \int P(A^n(x)^* B^*)\mathrm{d}\nu(B)\\ &= \iint f([A^n(x)^* Bv])\mathrm{d}\nu^*(B)\mathrm{d}\zeta([v])\\ &= \int f([A^n(x)v])\mathrm{d}(\nu^* * \zeta)([v])\\ &= \int f([A^n(x)v])\mathrm{d}\nu^*([v]) = P(A^n(x)^*).\end{aligned}$$

Besides, $\|P(A^n(x)^*)\|_2 \leqslant \|f\|_\infty < \infty$, hence $(P(A^n(\cdot)^*))_{n=1}^\infty$ is a L^2 bounded martingale. By the martingale convergence theorem, there is an $Lf \in L^2$ such that $P(A^n(x)^*) \to Lf(x)$ μ-a.e. and in L^2. Then

$$\mathbb{E}(|P(A^{n+1}(x)^*) - P(A^n(x)^*)|^2) \to 0.$$

Where we have

$$\mathbb{E}\left(\iint |f([A^n(x)^* Bv]) - f([A^n(x)^* v])|^2 \mathrm{d}\nu^*(B)\mathrm{d}\zeta([v])\right) \to 0,$$

this shows that for ν^*-a.e. $B \in \mathrm{SL}(2,\mathbb{R})$, $P(A^n(x)^* B) \to Lf(x)$ μ-a.e..

Take a countable dense set of f in $C(\mathbb{RP}^1)$, then there is a μ-full measure set of x such that $Lf(x)$ exists for all f. Then the functional $f \mapsto Lf(x)$ gives a probability measure ζ_x on $\mathbb{RP}^1$. These ζ_x satisfy the condition. □

Lemma 5.3

The limit measure ζ_x is a Dirac measure.

Proof. Fix a generic point x, we know that $A^n(x)^*\zeta \to \zeta_x$ and $A^n(x)^* B\nu \to \nu_x$ for ν^*-a.e. B. Choose a limit point of $\|A^n(x)^*\|^{-1}A^n(x)^*$, denoted by A. Then $A_*\zeta = \zeta_x = (AB)_*\zeta$ for ν^*-a.e. $B \in \mathrm{SL}(2,\mathbb{R})$. If A is invertible, by proposition 4.5, ν^* must supports on a compact subgroup of $\mathrm{SL}(2,\mathbb{R})$, contradiction. Then A must be non-invertible, which shows that ζ_x is a Dirac measure. □

Remark 5.4 — Denote this Dirac measure by $\delta_z = \delta_{z(x)}$, the proof of lemma shows that the $z(x)$ is independent of the choice of stationary measure. Moreover, the distribution of $z(x)$ on $\mathbb{RP}^1$ is same as ζ, hence we can prove the uniqueness of the stationary measure.

Proof of Furstenberg Theorem of $d = 2$.

Firstly, for μ-a.e. $x \in X$, we have $(A^n(x)^*)_*\zeta \to \delta_z$. Given generic $x \in X$, there must $\|A^n(x)\| \to \infty$ otherwise $A^n(x)$ have a limit point in $\mathrm{SL}(2,\mathbb{R})$ and the limit measure can't be Dirac.

Then we consider a limit point of $\|A^n(x)\|^{-1}A^n(x)^*$, denote by $A(x)$. Note that $\operatorname{rank} A(x) = 1$ and $\operatorname{Range} A(x) = z(x)\cdot\mathbb{R}$ where $\|z(x)\| = 1$. As $n \to \infty$, we have

$$\frac{\|A_n v\|}{\|A_n\|} = \sup_{\|u\|=1}\left\langle \frac{A^n v}{\|A_n\|}, u\right\rangle = \sup_{\|u\|=1}\left\langle v, \frac{(A^n) * u}{\|A_n\|}\right\rangle \to \sup_{\|u\|=1}\langle v, Au\rangle = |\langle v, z\rangle|.$$

In particular, $\|A_n(x)v\| \to \infty$ otherwise $v \perp z(x)$. Let η be the stationary measure of ν which is proper, the former discussion shows that $\|A_n(x)v\| \to \infty$ for $m = \mu \times \eta$ almost every $(x, [v])$. The theorem follows. □

§6 Invariance Principle

Definition 6.1. Let m be an probability measure on the product space $X \times Y$ that projects to the probability measure μ on X. A **disintegration** of m along vertical fibers is a measurable family $\{m_x : x \in X\}$ of probability measures on Y satisfying

$$m(E) = \int_X m_x\{v : (x, v) \in E\} \mathrm{d}\mu \quad \text{for any measurable } E \subseteq X \times Y.$$

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

Fact 6.2. A disintegration along a vertical fiber does exist. Moreover, the disintegration is unique up to a full μ-measure set.

Assume that X is a separable complete metric space, let μ be an invariant probability measure on X with respect to $f : X \to X$. Let $F : X \times \mathbb{R}^d \to X \times \mathbb{R}^d$ be a linear cocycle defined over f with extremal Lyapunov exponents $\lambda_+(x), \lambda_-(x)$. The following theorem of Ledrappier shows that if the upper Lyapunov exponent and the lower Lyapunov exponent coincide almost everywhere, then any $\mathbb{P}F$-invariant measure must have some invariance property on the conditional probabilities on the fiber.

Theorem 6.3 (Ledrappier)
Assume that $\lambda_-(x) = \lambda_+(x)$ for μ-almost every $x \in X$. Then

$$m_{f(x)} = A(x)_* m_x \quad \text{for } \mu\text{-almost every } x \in X,$$

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 μ.

Proof of Ledrappier's theorem

Let m be a probability measure on $X \times \mathbb{RP}^{d-1}$ invariant under the projective cocycle $\mathbb{P}F$ which projects to μ and let $\{m_x : x \in X\}$ be a disintegration of m. For each $x \in X$, let

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

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

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

Remark 6.4 — 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 6.5. The **fibered entropy** of m is defined by

$$h(m) = -\int \log J \mathrm{d}m.$$

Proposition 6.6

The fibered entropy $h(m)$ is always non-negative. If $h(m) = 0$ then $A(x)_* m_x = m_{f(x)}$ holds for μ-a.e. x.

Proof. By Jensen's inequality, we have

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

When the equality holds, there will be: $m\{J=0\} = 0, \log J$ is a constant m-almost everywhere and $\int J\mathrm{d}m = 1$. The last equality shows that $\xi_x = 0$ for μ-almost $x \in X$. Hence $J \equiv 1$ holds m-almost everywhere. Combining those discussions, we proofs the claim. □

Besides, we have another estimate of the fibered entropy. The difference of the extremal Lyapunov exponents reflects the contraction on $\mathbb{RP}^{d-1}$ with respect to the projective metric. And the fibered entropy measures the contraction on $\mathbb{RP}^{d-1}$ with respect to the conditional probability. As an analogues of the Ruelle inequality in the smooth ergodic theory , it is not surprising that the fibered entropy is bounded by the differences. The following proposition shows this relationship between the fibered entropy and the extremal Lyapunov exponents.

Assume, in addition, m is ergodic with respect to $(X \times \mathbb{RR}^{d-1}, \mathbb{P}F)$.

Proposition 6.7

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

Remark 6.8 — The constant d can be replaced by $d-1$ but doesn't matter.

Note that Ledrappier's theorem follows from proposition 6.6 and proposition 6.7 immediately. It suffices to show proposition 6.7.

Consider any $\Delta > \lambda_+ - \lambda_-$, for any $\varepsilon > 0$, let $J_\varepsilon = J + \varepsilon$ and $h_\varepsilon(m) = -\int J_\varepsilon$. Suppose, for contradiction, $h_\varepsilon(m) > d\Delta + 4\varepsilon$ for some Δ, where ε is small enough.

Lemma 6.9

Each fiber $\{x\} \times \mathbb{RP}^{d-1}$ admits partitions $\mathscr{P}_n(x)$ defined for n large enough, such that

(i) $\#\mathscr{P}_n(x) \leqslant e^{n(d\Delta+2\varepsilon)}$,

(ii) $\operatorname{diam}\mathscr{P}_n(x) \leqslant e^{-n(\Delta+2\varepsilon)}$,

(iii) $m_x(\partial\mathscr{P}_n(x,v)) = 0$ for all $v \in \mathbb{RP}^{d-1}$, where $\mathscr{P}_n(x,v)$ denote the atom of $\mathscr{P}_n(x)$ that contains the point v.

For each $0 \leqslant k \leqslant n$, let $\mathscr{P}_{n,k}(x)$ be a partition of $\{x\} \times \mathbb{RP}^{d-1}$ given by the pull back of $\mathscr{P}_n(f^k(x))$ under $A^k(x)$. That is $\mathscr{P}_{n,k}(x,v) = A^{-k}(x)\mathscr{P}_n(f^k(x,v))$ for each $(x,v) \in X \times \mathbb{RP}^{d-1}$. Consider the function

$$J_{n,k,\varepsilon}(x,v) = J_{n,k}(x,v) + \varepsilon = \frac{m_{f(x)}(\mathscr{P}_{n,k}(F(x,v)))}{m_x(\mathscr{P}_{n,k+1}(x,v))} + \varepsilon.$$

Lemma 6.10

$\sup_{0\leqslant k\leqslant n} \|\log J_{n,k,\varepsilon} - \log J_\varepsilon\|_{L^1(m)} \to 0,\ n\to\infty.$

Proof of Proposition 6.7. Let $J_{n,\varepsilon}(x,v) = \prod_{k=0}^{n-1} J_{n,n-1-k,\varepsilon}\circ F^k(x,v)$, and let

$$J_n(x,v) = \prod_{k=0}^{n-1} J_{n,n-1-k}\circ F^k(x,v) = \frac{m_{f^n(x)}(\mathscr{P}_n(F^n(x,v)))}{m_x(\mathscr{P}_{n,n}(x,v))} \leqslant J_{n,\varepsilon}(x,v).$$

We have

$$\frac{1}{n}\log J_{n,\varepsilon} = \frac{1}{n}\sum_{k=0}^{n-1}\log J_{n,n-1-k,\varepsilon}\circ F^k(x,v).$$

Because we assume that m is ergodic,

$$\frac{1}{n}\sum_{k=0}^{n-1}\log J_\varepsilon\circ F^k(x,v) \to \int \log J_\varepsilon \mathrm{d}m = -h_\varepsilon(m) \quad \text{in } L^1(m).$$

The previous lemma shows when n tends to infinity, $\frac{1}{n}\sum_{k=0}^{n-1}\log J_{n,n-1-k,\varepsilon}\circ F^k(x,v)$ is closed to $\frac{1}{n}\sum_{k=0}^{n-1}\log J_\varepsilon\circ F^k(x,v)$ in $L^1(m)$, hence $\frac{1}{n}\log J_{n,\varepsilon}\to -h_\varepsilon(m)$ in $L^1(m)$. By passing to a subsequence, we can get a sequence $n_j\to\infty$ such that

$$\frac{1}{n_j}\log J_{n_j,\varepsilon}(x,v)\to -h_\varepsilon(m) \quad \text{for } m\text{-a.e. } (x,v)\in X\times\mathbb{RP}^{d-1}.$$

Then,

$$\limsup_j m_{f^{n_j}(x)}(\mathscr{P}_{n_j}(F^{n_j}(x,v))) \leqslant -h_\varepsilon(m).$$

For each large j, there is $E_j\subseteq X\times\mathbb{RP}^{d-1}$, such that $m(E_j)>\frac{1}{2}$ and

$$m_{f^{n_j}(x)}(\mathscr{P}_{n_j}(F^{n_j}(x,v))) \leqslant e^{-n_j(h_\varepsilon(m)-\varepsilon)} \quad \text{for all } (x,v)\in E_j.$$

Hence $m_{f^{n_j}(x)}(F^{n_j}(E_j)\cap(f^{n_j}(x)\times\mathbb{RP}^{d-1})) \leqslant e^{-n_j(h_\varepsilon(m)-\varepsilon)}\cdot e^{n_j(d\Delta+2\varepsilon)} \leqslant e^{-n_j\varepsilon}$ by the assumption $h_\varepsilon(m)>d\Delta+4\varepsilon$. This follows that $m(F^{n_j}(E_j))\leqslant e^{-n_j\varepsilon}\to 0$, as $j\to\infty$. Which contradicts with $m(E_j)>\frac{1}{2}$ and F preserves the measure m. $\square$