STATISTICAL PROPERTIES FOR CERTAIN DYNAMICAL SYSTEMS

Contents

1. Introduction to the main topics of the course

1.1. **Additive random processes.** Let $\xi_0, \xi_1, \dots, \xi_{n-1}, \xi_n, \dots$ be a sequence of independent and identically distributed (i.i.d.) real random variables. Let

$$S_n := \xi_0 + \xi_1 + \cdots + \xi_{n-1}$$

be the partial sum process and let

$$\frac{1}{n} S_n = \frac{1}{n} \, (\xi_0 + \xi_1 + \cdots + \xi_{n-1})$$

be the average partial sum process.

Question. What is the behavior of these averages when $n \to \infty$?

Remark 1.1. Recall that two random variables ξ_1 and ξ_2 are identically distributed if $\mathbb{P}\{\xi_1 \in E\} = \mathbb{P}\{\xi_2 \in E\}$ for any Borel measurable set $E \subset \mathbb{R}$. In this case $\mathbb{E}\xi_1 = \mathbb{E}\xi_2$ and in fact $\mathbb{E}\phi(\xi_1) = \mathbb{E}\phi(\xi_2)$ for any integrable function $\phi \colon \mathbb{R} \to \mathbb{R}$.

Recall also that the random variables $\xi_1, \ldots, \xi_n$ are independent if for any Borel measurable sets $E_1, \ldots, E_n \subset \mathbb{R}$,

$$\mathbb{P}\{\xi_1 \in E_1 \wedge \ldots \wedge \xi_n \in E_n\} = \mathbb{P}\{\xi_1 \in E_1\} \cdots \mathbb{P}\{\xi_n \in E_n\}.$$

Theorem 1.1 (The law of large numbers - LLN)**.** *Given i.i.d. sequence $\xi_0, \xi_1, \ldots, \xi_{n-1}, \xi_n, \ldots$ of real random variables, if $\mathbb{E}\xi_0 < \infty$ then*

$$\frac{1}{n} S_n \to \mathbb{E}\xi_0 \quad a.s.$$

In particular, convergence in probability also holds. That is, $\forall\, \epsilon > 0$,

$$\mathbb{P}\left\{ \left| \frac{1}{n} S_n - \mathbb{E}\xi_0 \right| > \epsilon \right\} \to \infty \qquad as \qquad n \to \infty.$$

Question. It is natural to ask if there is a rate of convergence to 0 of the probability of the tail event above. It turns out that there is, as shown by the large deviations principle (LDP) below.

Theorem 1.2 (LDP of Cramér)**.** *Assume that the common distribution of the i.i.d. sequence of real random variables $\xi_0, \xi_1, \ldots, \xi_{n-1}, \xi_n, \ldots$ satisfies a certain growth condition and is non-trivial. Then $\forall\, \epsilon > 0$,*

$$\mathbb{P}\left\{ \left| \frac{1}{n} S_n - \mathbb{E}\xi_0 \right| > \epsilon \right\} \asymp e^{-c(\epsilon) n} \qquad as \qquad n \to \infty,$$

where $c(\epsilon) \asymp c_0 \epsilon^2$ for some $c_0 > 0$.

More precisely, assuming that the common distribution has finite exponential moments:

$$M(t) := \mathbb{E}\left(e^{t\xi_0}\right) < \infty \quad \forall t \in \mathbb{R},$$

it follows that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left\{ \left| \frac{1}{n} S_n - \mathbb{E}\xi_0 \right| > \epsilon \right\} = -c(\epsilon)$$

where

$$c(\epsilon) = \sup_{t \in \mathbb{R}} \left(t\epsilon - \log M(t)\right)$$

is the Legendre transform of $\log M(t)$.

This rate function $c(\epsilon)$ is strictly convex near $\epsilon = 0$, with $c(0) = 0$, $c'(0) = 0$ and $c''(0) > 0$, so that $c(\epsilon) \asymp c_0 \epsilon^2$.

Remark 1.2. The LDP is a very precise but asymptotic result. We are usually more interested in *finitary*, albeit less precise results, which will be referred to as large deviations type (LDT) estimates. A typical such result is the folllowing.

Theorem 1.3 (Hoeffding's Inequality)**.** *Assume the much stronger growth condition* $|\xi_0| \leq C$ *a.s. Then* $\forall\, \epsilon > 0$ *the following holds* for all $n \in \mathbb{N}$*:*

$$\mathbb{P}\left\{ \left| \frac{1}{n} S_n - \mathbb{E}\xi_0 \right| > \epsilon \right\} \leq 2 e^{-(2C)^{-2} \epsilon^2 n} .$$

Question. What is the typical size of the sum process $S_n - n\mathbb{E}\xi_0$? Note that by the LLN, almost surely we have

$$\frac{S_n - n\mathbb{E}\xi_0}{n} \to 0,$$

which implies that $S_n - n\mathbb{E}\xi_0 \ll n$. It turns out that from a certain point of view, $S_n - n\mathbb{E}\xi_0 \asymp \sqrt{n}$. More precisely, the following central limit theorem (CLT) holds.

Theorem 1.4 (CLT of Lindeberg-Lévy)**.** *Consider an i.i.d. sequence* $\xi_0, \xi_1, \dots, \xi_{n-1}, \xi_n, \dots$ *of real random variables and assume that the variance* $\sigma^2 = \mathbb{E}\xi_0^2 - (\mathbb{E}\xi_0)^2 \in (0, \infty)$*. Then for all* $[a, b] \subset \mathbb{R}$*,*

$$\mathbb{P}\left\{ \frac{S_n - n\mathbb{E}\xi_0}{\sigma \sqrt{n}} \in [a, b] \right\} \to \int_a^b e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \quad as \quad n \to \infty.$$

In other words, with the appropriate scaling we have the convergence in distribution to the standard normal distribution

$$\frac{S_n - n\mathbb{E}\xi_0}{\sigma\, \sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1) .$$

1.2. **Multiplicative random processes.** Let μ be a probability measure on the group of matrices $\mathrm{GL}_2(\mathbb{R})$. Given $g_0, g_1, \dots, g_{n-1}, g_n, \dots$ an i.i.d. sequence of random matrices chosen according to the probability μ, consider

$$\Pi_n := g_{n-1} \cdots g_1\, g_0$$

the corresponding multiplicative process.

Recall that for a matrix $g \in \mathrm{GL}_2(\mathbb{R})$, the norm is its maximal expansion

$$\|g\| = \max_{\|v\|=1} \|gv\|$$

while the co-norm is its minimal expansion

$$m(g) = \min_{\|v\|=1} \|gv\| = \left\| g^{-1} \right\|^{-1} .$$

The LLN for additive random processes has the following analog for multiplicative random processes.

Theorem 1.5 (Furstenberg-Kesten)**.** *Assuming the integrability condition* $\mathbb{E}\,(\log\|g\|)\;d\mu(g) < \infty$*, there are two numbers* $L^+(\mu) \geq L^-(\mu)$ *called the maximal respectively the minimal Lyapunov exponents of* μ *such that*

$$\frac{1}{n}\log\|\Pi_n\| \to L^+(\mu), \quad a.s.$$

and

$$\frac{1}{n}\log\|\Pi_n^{-1}\|^{-1} \to L^-(\mu), \quad a.s.$$

In particular we also have convergence in probability: $\forall\,\epsilon > 0$,

$$\mathbb{P}\left\{\left|\frac{1}{n}\log\|\Pi_n\| - L^+(\mu)\right| > \epsilon\right\} \to 0 \quad as \quad n \to \infty\,.$$

Instead of the maximal (or minimal) expansion of the random matrix products, we may consider the expansion of any vector. That is, given $v \in \mathbb{R}^2$, $v \neq 0$ consider the random walk $\{g_{n-1}\cdots g_1 g_0 v \colon n \geq 0\}$.

Theorem 1.6 (Furstenberg-Kifer's non-random filtration)**.** *For any given vector* $v \in \mathbb{R}^2$, $v \neq 0$, *either*

$$\frac{1}{n}\log\|\Pi_n v\| \to L^+(\mu) \quad as \quad n \to \infty\,,$$

or

$$\frac{1}{n}\log\|\Pi_n v\| \to L^-(\mu) \quad as \quad n \to \infty\,.$$

Remark 1.3. It turns out that under certain generic conditions to be defined in the future (namely the irreducibility of the measure μ), we have that $\forall\,v \in \mathbb{R}^2, v \neq 0$ the almost sure limit is the maximal Lyapunov exponent:

$$\frac{1}{n}\log\|\Pi_n v\| \to L^+(\mu) \quad \text{a.s.}$$

Moreover, if $L^+(\mu) > L^-(\mu)$ then

$$\mathbb{E}\left(\frac{1}{n}\log\|\Pi_n v\|\right) \to L^+(\mu)$$

uniformly in v.

Question. It is natural to ask if in this multiplicative random setting there are analogues of the LDP, LDT and CLT from the additive setting. As shown below, the answer is affirmative, at least in the generic setting. The precise statements will be provided later.

Theorem 1.7 (LDP - Le Page)**.** *Under generic assumptions, if* $L^+(\mu) > L^-(\mu)$*, then* $\forall\, v \in \mathbb{R}^2$*,* $v \neq 0$ *and* $\forall\, \epsilon > 0$*,*

$$\mathbb{P}\left\{ \left| \frac{1}{n} \log\|\Pi_n\| - L^+(\mu) \right| > \epsilon \right\} \asymp e^{-c(\epsilon)n} \quad \textit{as} \quad n \to \infty.$$

Theorem 1.8 (LDT - Duarte, Klein)**.** *Under generic assumptions, if* $L^+(\mu) > L^-(\mu)$*, then* $\forall\, v \in \mathbb{R}^2$*,* $v \neq 0$*,* $\forall\, \epsilon > 0$ *and* $\forall\, n \in \mathbb{N}$*,*

$$\mathbb{P}\left\{ \left| \frac{1}{n} \log\|\Pi_n\| - L^+(\mu) \right| > \epsilon \right\} \leq C e^{-c(\epsilon)n}$$

for some constant $C < \infty$ *and* $c(\epsilon) > 0$*.*

Theorem 1.9 (CLT - Le Page)**.** *Under generic assumptions, there is* $\sigma \in (0, \infty)$ *such that* $\forall\, v \in \mathbb{R}^2$*,* $v \neq 0$*,*

$$\mathbb{P}\left\{ \frac{\log\|\Pi_n v\| - nL^+(\mu)}{\sigma\sqrt{n}} \in [a, b] \right\} \to \int_a^b e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \quad \textit{as} \quad n \to \infty.$$

1.3. **Observed dynamical systems.** Let (M, f) be a dynamical system where M is a compact metric space and $f : M \to M$ is continuous. Consider an appropriate f-invariant measure $\nu \in \mathrm{Prob}(M)$.

Remark 1.4. Recall that v is called f-invariant if $f_*\nu = \nu$, which is equivalent to saying that $\nu(E) = \nu(f^{-1}(E))$ for all Borel measurable $E \subset M$.

Moreover, ν is called ergodic w.r.t. f if all f-invariant sets (i.e. E such that $E = f^{-1}(E)$) are of ν measure 0 or 1. Note that ergodic measures are extremal points in the the space of f-invariant measures (this space is convex and weak-$*$ compact).

The triple (M, f, ν) is called a measure-preserving dynamical system (MPDS). Given an observable $\xi : M \to \mathbb{R}$ in an appropriate space of functions, the quadruple (M, f, ν, ξ) is called an observed MPDS.

For all iterates j, consider the real-valued random variable on M

$$\xi_j := \xi \circ f^j\,.$$

Since ν is f-invariant, and hence f^j-invariant for all j, the sequence $\xi_0, \xi_1 \ldots, \xi_{n-1}, \xi_n, \ldots$ is identically distributed. However, in general this sequence is *not* independent.

Consider the sum process, that is, the Birkhoff sums

$$S_n\xi := \xi + \xi \circ f + \cdots + \xi \circ f^{n-1} = \xi_0 + \xi_1 + \cdots + \xi_{n-1}\,.$$

Birkhoff's ergodic theorem is a generalization of the LLN in this setting.

Theorem 1.10 (Birkhoff's ergodic theorem)**.** *Assume that ν is ergodic w.r.t. f and that $\int_M |\xi| \, d\nu < \infty$. Then*

$$\frac{1}{n} S_n \xi \to \int \xi d\nu \quad \nu\text{-a.e.}$$

In particular the convergence in measure also holds: $\forall \epsilon > 0$,

$$\nu \left\{ x \in M \colon \left| \frac{1}{n} S_n \xi(x) - \int_M \xi \, d\nu \right| > \epsilon \right\} \to 0 \quad \text{as} \quad n \to \infty.$$

Question. A fundamental problem in ergodic theory is to establish statistical properties like LDP, LDT, CLT for various kinds of observed dynamical systems.

In other words, the question is to determine for which dynamical system (M, f), for which appropriate choice of f-invariant measures ν and for which kinds of observables ξ one has an LDT estimate

$$\nu \left\{ x \in M \colon \left| \frac{1}{n} S_n \xi(x) - \int_M \xi \, d\nu \right| > \epsilon \right\} \le C \, e^{-c(\epsilon) n}$$

or a CLT

$$\nu \left\{ x \in M : \frac{S_n \xi(x) - n \int \xi \, d\nu}{\sigma \sqrt{n}} \in [a, b] \right\} \to \int_a^b e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \, .$$

A short but vague answer is that systems with *some hyperbolicity* should satisfy such statistical properties. The question is extremely far reaching, and for now it only has a very incomplete rigorous answer.

Some of the main tools used to address it, which will make their entry in this course in due time, are the transition (or Markov) operator and the transfer (or Ruelle) operator.

1.4. **The moment method and Bernstein's trick.** You decide where to put this aula 2.

Let ξ be a random variable on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The distribution (or law) of ξ is the probability measure μ_ξ on $\mathbb{R}$ given by

$$\mu_\xi(E) = \mathbb{P}\{\xi \in E\} = \mathbb{P}\left\{\xi^{-1} E\right\}$$

where $E \subset \mathbb{R}$ is Borel measurable. In other words, $\mu_\xi = \xi_* \mathbb{P}$. Given a random variable ξ and $\mu \in \operatorname{Prob}(\mathbb{R})$, we write $\xi \sim \mu$ when $\mu_\xi = \mu$.

Example 1. The continuous uniform distribution on some interval $[a, b] \subset \mathbb{R}$ is

$$\mu_{unif} = \frac{1}{b-a} \mathbb{1}_{[a,b]} dm$$

which is absolutely continuous to the Lebesgue measure m on $\mathbb{R}$.

Example 2. The standard normal distribution

$$\mathcal{N}(0,1) = G(t)dm$$

where $G(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ is the Gaussian.

Remark 1.5. The distribution of a random variable ξ determines its expectation, standard deviation, moments, etc. For example, its expectation satisfies

$$\mathbb{E}\xi = \int_{\mathbb{R}} x d\mu_\xi(x).$$

More generally, if $\varphi : \mathbb{R} \to \mathbb{R}$ is Lebesgue integrable, then

$$\mathbb{E}\varphi(\xi) = \int_{\mathbb{R}} \varphi(x) d\mu_\xi(x).$$

In fact, by the change of variables formula we have

$$\begin{aligned}\mathbb{E}\varphi(\xi) &= \int_{\Omega} \varphi(\xi(\omega)) d\mathbb{P}(\omega) \\ &= \int_{\mathbb{R}} \varphi(x) d\xi_* \mathbb{P}(x) \\ &= \int_{\mathbb{R}} \varphi(x) d\mu_\xi(x).\end{aligned}$$

We recall the meaning of random variables being identically distributed and independent in the following.

Definition 1.1. ξ_1 and ξ_2 are identically distributed if $\mu_{\xi_1} = \mu_{\xi_2}$. $\xi_1, \xi_2, \cdots, \xi_n$ are independent if

$$\mu_{(\xi_1 \cdots \xi_n)} = \mu_{\xi_1} \times \cdots \times \mu_{\xi_n}.$$

Namely, the joint distribution is precisely the product measure.

From now on, let us fix some notations as follows.
ξ is the real random variable.
$\mu = \mathbb{E}\xi$ is the expectation of ξ.
$\sigma^2 = \mathbb{E}(\xi - \mu)^2 = \mathbb{E}\xi^2 - \mu^2 \in [0, \infty]$ is the variance of ξ.
$\mathbb{E}\xi^n$ is called the n-th moment of ξ. By the Hölder inequality we have $\mathbb{E}\xi \lesssim (\mathbb{E}\xi^2)^{\frac{1}{2}}$ and $\mathbb{E}\xi^2 \lesssim (\mathbb{E}\xi^4)^{\frac{1}{2}}$ etc. Note that working with even moments avoids negativity.

The following lemma is trivial but extremely useful throughout probability theory.

Lemma 1.1 (Markov's inequality)**.** *If $X \geq 0$ and $\lambda > 0$ then*

$$\mathbb{P}\{X \geq \lambda\} \leq \frac{\mathbb{E}X}{\lambda}.$$

Proof. Denote $E = \{X \geq \lambda\}$, then we have $\mathbb{E}X \geq \int_E X d\mathbb{P} \geq \lambda \mathbb{P}(E)$. $\square$

We will use Markov's inequality to prove weak LLN and strong LLN respectively under some minor additional conditions.

Theorem 1.11 (Weak LLN)**.** *Given i.i.d. sequence* $\xi_0, \xi_1, \dots, \xi_{n-1}, \xi_n, \dots$ *of real random variables, if* $\mathbb{E}\xi_0^2 < \infty$ *then* $\forall\, \epsilon > 0$,

$$\mathbb{P}\left\{\left|\frac{1}{n}S_n - \mathbb{E}\xi_0\right| > \epsilon\right\} \to \infty \quad \text{as} \quad n \to \infty.$$

Proof. Without loss of generality, we may assume that $\mu = 0$. Then it is enough to show $\mathbb{P}\left\{\frac{S_n^2}{n^2} > \epsilon^2\right\} = \mathbb{P}\left\{S_n^2 > n^2\epsilon^2\right\} \to 0$ as $n \to \infty$.

By Markov's inequality, we have

$$\mathbb{P}\left\{S_n^2 > n^2\epsilon^2\right\} \leq \frac{\mathbb{E}S_n^2}{n^2\epsilon^2}.$$

Note that $S_n^2 = (\sum_{j=0}^{n-1} \xi_j)^2 = \sum_{j=0}^{n-1} \xi_j^2 + \sum_{j\neq k} \xi_j\xi_k$. Taking expectations on both sides, we obtain

$$\mathbb{E}S_n^2 = \sum_{j=0}^{n-1} \mathbb{E}\xi_j^2 + \sum_{j\neq k} \mathbb{E}(\xi_j\xi_k) = \sum_{j=0}^{n-1} \mathbb{E}\xi_j^2 = n\mathbb{E}\xi_0^2.$$

Here the second equality uses the independence of the random variables.

This shows that

$$\mathbb{P}\left\{S_n^2 > n^2\epsilon^2\right\} \leq \frac{\mathbb{E}S_n^2}{n\epsilon^2} \to 0 \quad \text{as} \quad n \to \infty.$$

This finishes the proof of weak LLN. $\square$

Remark 1.6. If $X_n \to X$ a.s. then $X_n \to X$ in probability. In general, the converse is not true. However, if $\forall\, \epsilon > 0$ we have

$$\sum_{n=0}^{\infty} \mathbb{P}\left\{|X_n - X| > \epsilon\right\} < \infty,$$

then $X_n \to X$ a.s. This is ensured by Borel-Cantelli Lemma.

Theorem 1.12 (Strong LLN)**.** *Given i.i.d. sequence* $\xi_0, \xi_1, \dots, \xi_{n-1}, \xi_n, \dots$ *of real random variables, if* $\mathbb{E}\xi_0^4 < \infty$ *then* $\frac{S_n}{n} \to \mu$ *a.s.*

Proof. Without loss of generality, we may again assume $\mu = 0$. By Remark 1.6, it is enough to show $\mathbb{P}\left\{S_n^4 > n^4\epsilon^4\right\} \leq \frac{c}{n^2}$ where c is a constant.

By Markov's inequality, we have

$$\mathbb{P}\left\{S_n^4 > n^4\epsilon^4\right\} \le \frac{\mathbb{E}S_n^4}{n^4\epsilon^4}.$$

By direct computations and use the independence condition we get $\mathbb{E}S_n^4 = O(n^2)$. Therefore,

$$\mathbb{P}\left\{S_n^4 > n^4\epsilon^4\right\} \lesssim \frac{1}{n^2} \to 0 \quad \text{as} \quad n \to \infty.$$

□

In the following, we are going to prove the following LDT estimates.

Theorem 1.13 (Cramér's inequality)**.** *Assume that the common distribution of the i.i.d. sequence of real random variables* $\xi_0, \xi_1, \dots, \xi_{n-1}, \xi_n, \dots$ *satisfies a certain growth condition and* $\sigma^2 > 0$*. Then* $\forall\, \epsilon > 0$,

$$\mathbb{P}\left\{\left|\frac{1}{n}S_n - \mu\right| > \epsilon\right\} \le 2e^{-\hat{C}(\epsilon)n} \quad \textit{as} \quad n \to \infty,$$

where $\hat{C}(\epsilon) \approx C_0\epsilon^2 > 0$ *with constant* $C_0 > 0$*.*

We introduce the Bernstein's trick first.

Let X be a random variable and $\lambda \in \mathbb{R}$. Then

$$X \ge \lambda \Leftrightarrow e^{tX} \ge e^{t\lambda}, \forall\, t > 0.$$

By Markov's inequality,

$$\mathbb{P}\left\{X \ge \lambda\right\} = \mathbb{P}\left\{e^{tX} \ge e^{t\lambda}\right\} \le \frac{\mathbb{E}(e^{tX})}{e^{t\lambda}},$$

which gives $\mathbb{P}\left\{X \ge \lambda\right\} \le e^{-t\lambda}\mathbb{E}(e^{tX})$.

Definition 1.2. The function $M : \mathbb{R} \to (0, \infty)$ defined by $M(t) = \mathbb{E}(e^{tX})$ is called the moment generating function of X while $c(t) = \log M(t)$ is called the cumulant generating function of X.

Proof of Theorem1.13 . Without loss of generality, assume $\mu = 0$. Note that it is enough to estimate $\mathbb{P}\left\{S_n > n\epsilon\right\}$, the other part $\mathbb{P}(-S_n > n\epsilon)$ is the same. This is why the coefficient 2 appears in the r.h.s. of the inequality.

By Bernstein's trick, we have

$$\mathbb{P}\left\{S_n > n\epsilon\right\} \le e^{-tn\epsilon}\mathbb{E}(e^{tS_n}).$$

Typically, $\mathbb{E}(e^{tS_n})$ can be exponentially large. But if we can prove something like

$$\mathbb{E}(e^{tS_n}) \le e^{nLt^2},$$

then we would have

$$\mathbb{P}\{S_n > n\epsilon\} \le e^{-nt\epsilon} e^{nLt^2} = e^{-n(t\epsilon - Lt^2)} = e^{-nc(\epsilon)}$$

It is easy to check that $c(\epsilon) = \frac{1}{4L}\epsilon^2$ is the maximum value of $t\epsilon - Lt^2$. Thus it is enough to estimate $\mathbb{E}(e^{tS_n})$.

Using the independence condition, we have

$$\mathbb{E}(e^{tS_n}) = \mathbb{E}(e^{t\xi_0}) \cdots \mathbb{E}(e^{t\xi_{n-1}}) = (\mathbb{E}(e^{t\xi_0}))^n = e^{nC_{\xi_0}(t)}.$$

Therefore,

$$\mathbb{P}\{S_n > n\epsilon\} \le e^{-nt\epsilon} e^{nC_{\xi_0}(t)} = e^{-n(t\epsilon - C_{\xi_0}(t))}.$$

Let $\hat{C}_{\xi_0}(\epsilon) := \sup_{t\in\mathbb{R}}(t\epsilon - C_{\xi_0}(t))$. This is called the Legendre transform of $C_{\xi_0}(t)$. Thus we have

$$\mathbb{P}\{S_n > n\epsilon\} \le e^{-n\hat{C}_{\xi_0}(\epsilon)}.$$

Since $\mu = 0$ and $\sigma^2 > 0$, it is straightforward to check that $C_{\xi_0}(t)$ satisfies $C_{\xi_0}(0) = 0, C'_{\xi_0}(0) = 0$ and $C''_{\xi_0}(0) = \sigma^2 > 0$. So $C_{\xi_0}(t) \approx \frac{\sigma^2}{2}t^2$ when $|t| \ll 1$. This gives us $\hat{C}_{\xi_0}(\epsilon) \approx C_0\epsilon^2 > 0$ with constant $C_0 > 0$.

This finishes the proof. □

In the rest of this section, we are going to prove the CLT of Lindeberg-Lévy.

We first recall some definitions.

Definition 1.3 (Convergence in distribution)**.** $X_n \xrightarrow{d} X$ if $\mu_{X_n} \to \mu_X$ in the weak* topology. More precisely, $\int_{\mathbb{R}} g d\mu_{X_n} \to \int_{\mathbb{R}} g d\mu_X$, $\forall\, g \in C_c(\mathbb{R})$.

Remark 1.7. Almost sure convergence implies convergence in probability, which further implies convergence in distribution. In general, the inverse directions are not true.

Definition 1.4. The cumulative distribution function (CDF) of a random variable X is

$$F_X(t) = \mathbb{P}(X \le t), F_X : \mathbb{R} \to [0, 1]$$

which is non-decreasing. This implies that F_X is continuous almost everywhere.

We list a useful Proposition below without proof.

Proposition 1.2. *$X_n \xrightarrow{d} X \Leftrightarrow F_{X_n}(t) \to F_X(t)$ for all t which is a continuous point of F_X.*

For convenience, we recall the CLT below

Theorem 1.14 (CLT of Lindeberg-Lévy)**.** *Consider an i.i.d. sequence* $\xi_0, \xi_1, \ldots, \xi_{n-1}, \xi_n, \ldots$ *of real random variables and assume that the variance* $\sigma^2 = \mathbb{E}\xi_0^2 - (\mathbb{E}\xi_0)^2 \in (0, \infty)$*. Then for all* $[a, b] \subset \mathbb{R}$*,*

$$\mathbb{P}\left\{\frac{S_n - n\mathbb{E}\xi_0}{\sigma\sqrt{n}} \in [a, b]\right\} \to \int_a^b e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \quad as \quad n \to \infty.$$

In other words, with the appropriate scaling we have the convergence in distribution to the standard normal distribution

$$\frac{S_n - n\mathbb{E}\xi_0}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Proof. Without loss of generality, we can assume that $\mu = 0$ and $\sigma^2 = 1$ so that we just need to prove

$$\frac{S_n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

The proof follows from Lévy and we will use Fourier analysis.

Define the characteristic function of a random variable X by

$$\varphi_X : \mathbb{R} \to \mathbb{C}, \quad \varphi_X(t) = \mathbb{E}(e^{itX}) = \int_{\mathbb{R}} e^{itx} d\mu_X(x).$$

This is the Fourier transform of μ_X.

Recall that Lévy's continuity theorem says the following:

$$X_n \xrightarrow{d} X \Longleftrightarrow \varphi_{X_n}(t) \to \varphi_X(t), \forall\, t \in \mathbb{R}.$$

This indicates the phenomenon that μ_{X_n} converges in the weak* topology if and only if its Fourier transform $\hat{\mu}_{X_n}$ converges for all t.

Moreover, we list some properties of $\varphi_X(t) = \mathbb{E}(e^{itX})$.

- $\varphi_X(0) = 1$,
- If X subjects to $\mathcal{N}(0, 1)$, then $\mu_X = G(t)dm$ and $\varphi_X(t) = \hat{G}(t) = e^{-\frac{t^2}{2}}$,
- $\varphi_{cX} = \varphi_X(ct)$,
- If X, Y are independent, then $\varphi_{X+Y} = \varphi_X \cdot \varphi_Y$.

By the properties, we get

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = \varphi_{S_n}(\frac{t}{\sqrt{n}}) = \prod_{j=0}^{n-1} \varphi_{\xi_j}(\frac{t}{\sqrt{n}}) = [\varphi_{\xi_0}(\frac{t}{\sqrt{n}})]^n.$$

By direct computations, we have $\varphi_{\xi_0}(0) = 1, \varphi'_{\xi_0}(0) = i\mu = 0$ and $\varphi''_{\xi_0}(0) = -\sigma^2 = -1$.

Therefore, by Taylor expansion we obtain

$$\varphi_{\xi_0}(\frac{t}{\sqrt{n}}) = e^{-(\frac{t}{\sqrt{n}})^2/2 + o((\frac{t}{\sqrt{n}})^3)}.$$

This proves that

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = e^{-\frac{t^2}{2}+o(\frac{t^3}{\sqrt{n}})},$$

which implies

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) \to e^{-\frac{t^2}{2}} \quad \text{as} \quad n \to \infty.$$

The proof is finished by using Lévy's continuity theorem. □

2. Stochastic dynamical systems

2.1. **Strongly mixing Markov chains.** To prove LDT and CLT for dynamical systems, we have to work with certain types of Markov chains (which are non-independent processes in general).

Example. LDT for multiplicative random processes.

Given $\mu \in \mathrm{Prob}_c(\mathrm{GL}_2(\mathbb{R}))$ and assume some generic condition, for an i.i.d. sequence $\{g_n\}_{n\geq 0}$ and $\Pi_n := g_{n-1}\cdots g_1 g_0$, we have that $\forall\, v \in \mathbb{R}^2, v \neq 0$

$$\mathbb{P}\left\{\left|\frac{1}{n}\log\|\Pi_n v\| - L^+(\mu)\right| > \epsilon\right\} \leq Ce^{-c(\epsilon)n}$$

for some $C < \infty$ and $c(\epsilon) > 0$.

In fact, we may relate the multiplicative process to a Markov chain in the following sense. For simplicity, let us try $n = 3$ first. For any $v \in \mathrm{S}^1$, we have

$$\frac{1}{3}\log\|g_2g_1g_0v\| = \frac{1}{3}[\log\frac{\|g_2g_1g_0v\|}{\|g_1g_0v\|} + \log\frac{\|g_1g_0v\|}{\|g_0v\|} + \log\|g_0v\|].$$

Denote $\Sigma = \mathrm{supp}(\mu) \subset \mathrm{GL}_2(\mathbb{R})$. Define $\varphi : \Sigma \times \mathrm{S}^1 \to \mathbb{R}$ by $\varphi(g, v) = \log\|gv\|$. Let $\omega \in \Omega = \Sigma^{\mathbb{N}}$ and $\omega = \{g_i\}_{i\geq 0}$. Define $Z_j^v : \Omega \to \Sigma \times \mathrm{S}^1$ by $Z_j^v(\omega) = (g_j, \frac{g_{j-1}\cdots g_0 v}{\|g_{j-1}\cdots g_0 v\|})$, $j \geq 1$ and $Z_0^v(\omega) = (g_0, v)$. Then we obtain

$$\frac{1}{3}\log\|g_2g_1g_0v\| = \frac{1}{3}[\varphi(Z_2^v(\omega)) + \varphi(Z_1^v(\omega)) + \varphi(Z_0^v(\omega))].$$

In general,

$$\frac{1}{n}\log\|\Pi_n v\| = \frac{1}{n}\sum_{j=0}^{n-1}\varphi(Z_j^v(\omega)).$$

where $\varphi(g, v) = \log\|gv\|$ and $\{Z_n\}_{n\geq 0}$ is a Markov chain with values in $\Sigma \times \mathrm{S}^1$ and transition $(g_0, v) \to (g_1, \frac{g_0 v}{\|g_0 v\|})$ which is precisely the underlying fiber projective dynamics of the multiplicative random process.

Therefore, in order to prove LDT and CLT for multiplicative processes or other types of dynamical systems, we need to study appropriate Markov chains. Let us begin with a simple model.

Model: subshift of finite type.

Let $\Sigma = \{1, \cdots, n\}$ be a finite space of symbols and let $P = \{p_{ij}\}_{1 \le i,j \le n}$ be a stochastic matrix. Namely,

$$\forall 1 \le i \le n, \sum_{j=1}^{n} p_{ij} = 1; \quad p_{ij} \ge 0, \forall\, 1 \le i, j \le n.$$

P can be seen as a transition matrix giving the transition probability from i to j by p_{ij}.

Let $q = (q_1, \cdots, q_n)$ be a probability vector satisfying $q_i \ge 0$, $\forall 1 \le i \le n$ and $\sum_{i=1}^{n} q_i = 1$.

Definition 2.1. q is P-stationary if $qP = q$. That is

$$q_j = \sum_{i=1}^{n} q_i p_{ij}, \forall\, 1 \le j \le n.$$

Remark 2.1. Every stochastic matrix P has at least one stationary measure. Moreover, if P is primitive which means that $\exists m \in \mathbb{Z}^+$ such that $P^n_{ij} > 0, \forall\, 1 \le i, j \le n$, then $\exists!$ stationary vector q and $P^n_{ij} \to q_j$ exponentially fast for any $1 \le i \le n$.

In the following, we are going to define the Markov measure. Let us begin with some notations.

$X^+ = \Sigma^{\mathbb{N}} = \{\{x_n\}_{n \ge 0} : x_n \in \Sigma\}$.

$\mathcal{B}^+ = \sigma$- algebra generated by cylinders of the form

$$C[i_0, \cdots, i_n] = \{x \in X^+ : x_0 = i_0, \cdots, x_n = i_n\}.$$

Given q a probability vector and P a stochastic matrix, define

$$\mathbb{P}_{(q,P)}(C[i_0, \cdots, i_n]) := q_{i_0} P_{i_0 i_1} \cdots P_{i_{n-1} i_n}.$$

This is a pre-measure. By Carathéodory's extension theorem, this pre-measure has a unique extension to a measure on $\mathcal{B}^+$ called Markov measure.

Let $\sigma : X^+ \to X^+$ be the forward shift. Note that if q is P-stationary, then $\mathbb{P}_{(q,P)}$ is σ-invariant. Therefore, $(X^+, \sigma, \mathbb{P}_{(q,P)})$ is an MPDS called a subshift of finite type. Moreover, if P is primitive, then $(X^+, \sigma, \mathbb{P}_{(q,P)})$ is exponentially mixing (hence ergodic).

A Markov chain with values in Σ is a sequence of random variables $\{Z_n\}_{n \ge 0}$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $Z_n : \Omega \to \Sigma$ satisfying the Markov property.

$$\mathbb{P}\left\{Z_{n+1} = j \middle| Z_n = i_n, \cdots, Z_0 = i_0\right\} = \mathbb{P}\left\{Z_{n+1} = j \middle| Z_n = i_n\right\}.$$

A Markov chain $\{Z_n\}_{n \ge 0}$ is said to have an initial distribution q and a transition P if

$\mathbb{P}\{Z_0 = i\} = q_i$,
$\mathbb{P}\left\{Z_{n+1} = j \middle| Z_n = i\right\} = p_{ij}$.

By Kolmogorov, there are such Markov chains on $(X^+, \mathcal{B}^+, \mathbb{P}_{(q,P)})$. In fact, the Markov chain $\{Z_j\}_{j\geq 0}$ is precisely the projection $Z_j : X^+ \to \Sigma$ defined by $Z_j(x) = x_j$ for any $j \geq 0$.

Now let us consider a more general setting.

Let $(M, \mathcal{F})$ be a measurable space. M is a compact metric space and $\mathcal{F}$ is the Borel σ-algebra on M.

Definition 2.2 (Markov kernal)**.** A Markov kernal $K(x, E)$ (which can be interpreted as the probability of x transitioning to E) is a function $K : M \times \mathcal{F} \to [0, 1]$ such that

(1) $\forall\, x \in M, E \mapsto K_x(E)$ is a probability measure on $\mathcal{F}$,
(2) $\forall\, E \in \mathcal{F}$, $K(\cdot, E)$ is $\mathcal{F}$-measurable.

Remark 2.2. In practice, we may assume that $x \mapsto K_x = K(x, \cdot) \in \mathrm{Prob}(M)$ is continuous which in particular impies (2). In other words, we can think of a Markov kernal as a continuous function $K : M \to \mathrm{Prob}(M)$ where we interpret K_x as the probability of transitioning from x to somewhere.

Definition 2.3 (Stationary measure)**.** $\mu \in \mathrm{Prob}(M)$ is called K-stationary if $\mu = \int_M K_x d\mu(x)$ in the sense that $\mu(E) = \int_M K_x(E) d\mu(x)$, $\forall\, E \in \mathcal{F}$.

Now we can define the Markov measure.

Given $\pi \in \mathrm{Prob}(M)$ and K a Markov kernal, by Kolmogorov there exists a unique probability measure $\mathbb{P}_\pi = \mathbb{P}_{(\pi,K)}$ on $X^+ = M^{\mathbb{N}}$, $\mathcal{B}^+ =$ σ-algebra generated by the cylinders of the form:

$$C[A_0, \cdots, A_n] = \left\{x = \{x_n\}_{n\geq 0} \in X^+ : x_j \in A_j, \forall\, 0 \leq j \leq n\right\},$$

where all $A_j \in \mathcal{F}$. It is easy to check that

$$\mathbb{P}_{(\pi,K)}(C[A_0, \cdots, A_n]) = \int_{A_0} \int_{A_n} \cdots \int_{A_1} 1 dK_{x_0}(x_1) \cdots dK_{x_{n-1}}(x_n) d\pi(x_0).$$

Note that If $\varphi : X^+ \to \mathbb{R}$, then $\mathbb{E}_\pi(\varphi) = \int_{X^+} \varphi d\mathbb{P}_\pi$.

A Markov chain $\{Z_n : \Omega \to M\}$ is a sequence of random variables with values in M on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ satisfying the following Markov property.

$$\mathbb{P}\left\{Z_{n+1} \in E \middle| Z_n, \cdots, Z_0\right\} = \mathbb{P}\left\{Z_{n+1} \in E \middle| Z_n\right\}.$$

The Markov chain $\{Z_n\}_{n\geq 0}$ is said to have initial distribution π and transition K if

$\mathbb{P}\{Z_0 \in E\} = \pi(E)$,
$\mathbb{P}\left\{Z_{n+1} \in E \middle| Z_n = x\right\} = K_x(E)$.

Example: $\Omega = X^+ = M^{\mathbb{N}}$, $\mathcal{F} = \mathcal{B}^+ = \sigma$-algebra generated by cylinders. $\mathbb{P} = \mathbb{P}_{(\pi,K)}$, $Z_n : X^+ \to M, Z_n(x) = x_n \, \forall \, n \geq 0$.

Note that if $\pi = \delta_x$ then we write $\mathbb{P}_{\delta_x} := \mathbb{P}_x$ and $\mathbb{E}_{\delta_x} := \mathbb{E}_x$.

If μ is a K-stationary measure, then a (μ, K) Markov chain $\{Z_n\}_{n\geq 0}$ is stationary. More precisely,

$$\mathbb{P}\{Z_0 \in E_0, \cdots, Z_n \in E_n\} = \mathbb{P}\{Z_j \in E_0, \cdots, Z_{j+n} \in E_n\}, \, \forall \, j, n \geq 0$$

where $E_j \in \mathcal{F}$ is arbitrary. Moreover, the Markov shift (X^+, σ) is $\mathbb{P}_\mu$-invariant so $(X^+, \sigma, \mathbb{P}_\mu)$ is an MPDS.

The powers of the Markov kernal can be derived inductively:

$$K^{n+1}(x, E) = \int_M K(y, E) dK_x^n(y).$$

Definition 2.4 (Markov system and strongly mixing)**.** If μ is K-stationary, then (M, K, μ) is called a Markov system. It is strongly mixing if $K_x^n \to \mu$ exponentially fast $\forall \, x \in M$ in the weak* topology. More precisely, $\forall \, \varphi \in L^\infty(M)$,

$$\left\| \int_M \varphi(y) dK_x^n(y) - \int_M \varphi(y) d\mu(y) \right\|_\infty \leq C\rho^n \|\varphi\|_\infty \tag{2.1}$$

holds $\forall \, n \in \mathbb{N}$ where $C < \infty$ and $\rho \in (0, 1)$.

We may also consider the same concept from a different perspective, as we shall see below.

Definition 2.5 (Markov operator)**.** Given a Markov system (M, K, μ), the Markov operator $Q = Q_K : L^\infty(M) \to L^\infty(M)$ defined by

$$(Q\varphi)(x) = \int_M \varphi(y) dK_x(y).$$

The n-th iterates are

$$(Q^n\varphi)(x_0) = \int_M \cdots \int_M \varphi(x_n) dK_{x_{n-1}}(x_n) \cdots dK_{x_0}(x_1) = \int_M \varphi dK_{x_0}^n.$$

Therefore, (2.1) is equivalent to

$$\left\| (Q^n\varphi)(x) - \int_M \varphi d\mu \right\|_\infty \leq C\rho^n \|\varphi\|_\infty, \, \forall \, n \in \mathbb{N}.$$

2.2. **Large deviations for strongly mixing Markov chains.** We first recall some definitions. We begin with

Deterministic dynamical systems (DDS) (M, f).

Let M be a metric space and let $f : M \to M$ be a continuous map. Once the initial state of the system $x_0 = x$ is fixed, then $x_n = f^n(x_0), n \geq 0$ are all determined.

A probability measure $\mu \in \text{Prob}(M)$ is f-invariant if $f_*\mu = \mu$. Equivalently, $\int_M \delta_{f(x)} d\mu(x) = \mu$ or $\forall\, \varphi \in C_c(M), \int_M \varphi(f(x))d\mu(x) = \int \varphi d\mu$. The triple (M, f, μ) is called a measure preserving dynamical system (as a convention, we omit the Borel σ-algebra on M).

A subset $E \subset M$ is called f-invariant if $f^{-1}(E) = E$. Equivalently, $x \in E \Leftrightarrow f(x) \in E$ or $x \in E \Leftrightarrow \delta_{f(x)}(E) = 1$. μ is f-ergodic if E is f-invariant $\Rightarrow \mu(E) = 0$ or 1. Given an observable $\varphi : M \to \mathbb{R}$, then $\varphi(f^n(x))$ is the observed n-th state of the system which is to be considered.

Stochastic dynamical system (SDS) (M, K).

Let M be a compact metric space and let $K : M \to \text{Prob}(M), x \to K_x$ be a continuous kernal. If $x_0 = x$ is the initial state of the system, the next state x_1 is not determined like in the DDS case by a transition law f. It is known only with a certain probability: $\mathbb{P}\{x_1 \in E\} = K_{x_0}(E)$. The iterates of K are $K_x^n = \int_M K_y^{n-1} dK_x(y)$.

$\mu \in \text{Prob}(M)$ is called K-stationary if $K * \mu = \mu$ in the sense that $\int_M K_x d\mu(x) = \mu$. The triple (M, K, μ) is a Markov system.

$E \subset M$ is K-invariant if $x \in E \Leftrightarrow K_x(E) = 1$. A K-stationary measure μ is ergodic if whenever E is K-invariant, we have $\mu(E) = 0$ or 1. If $\varphi : M \to \mathbb{R}$ is an observable, we will consider

$$(Q^n\varphi)(x) = \int_M \varphi(y) dK_x^n(y) = \int_M \cdots \int_M \varphi(y) dK_{x_{n-1}}(y) \cdots dK_{x_0}(x_1).$$

Example 1. Any DDS (M, f) is itself an SDS. That is, $M \to \text{Prob}(M), x \to \delta_{f(x)}$.

Example 2. $\mu \in \text{Prob}_c(\text{GL}_2(\mathbb{R})), \Sigma = \text{supp}(\mu)$ and $\{g_n\}_{n\geq 0}$ is a sequence of i.i.d. matrices chosen with law μ. We may consider the kernal K on $\Sigma \times \text{S}^1$ as follows $K : \Sigma \times \text{S}^1 \to \text{Prob}(\Sigma \times \text{S}^1)$ such that $K_{(g_0, \hat{v})} = \mu \times \delta_{g_0 \hat{} v}$. Then $(\Sigma \times \text{S}^1, K)$ is an SDS.

Let us formally talk about the Kolmogorov extension. Let (M, K, μ) be a Markov system. Denote $X^+ = M^{\mathbb{N}} = \{x = \{x_n\}_{n\geq 0} : x_n \in M\}$. If $\pi \in \text{Prob}(M)$, then $\exists! \mathbb{P}_\pi \in \text{Prob}(X^+)$ s.t.

$\mathbb{P}_\pi(C[E_0]) = \pi(E_0)$.

$\mathbb{P}_\pi(C[E_0, E_1]) = \int_{E_0} \int_{E_1} 1 dK_{x_0}(x_1) d\pi(x_0)$.

If $f : X^+ \to \mathbb{R}$, then

$$\mathbb{E}_\pi(f) = \int_{X^+} f(x_0, \cdots, x_n, \cdots) dK_{x_0}(x_1) \cdots K_{x_{n-1}}(x_n) \cdots d\pi(x_0).$$

When $\pi = \delta_x$, we simply write $\mathbb{E}_x$ and $\mathbb{P}_x$. When $\pi = \mu$ which is K-stationary, we write $\mathbb{E}_\mu = \mathbb{E}$ and $\mathbb{P}_\mu = \mathbb{P}$.

Things here are repeated, see Aula 3. You decide which to be deleted.

Recall that a Markov chain $\{Z_n : \Omega \to M\}$ is a sequence of random variables with values in M on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ satisfying the following Markov property.

$$\mathbb{P}\left\{Z_{n+1} \in E \middle| Z_n, \cdots, Z_0\right\} = \mathbb{P}\left\{Z_{n+1} \in E \middle| Z_n\right\}.$$

The Markov chain $\{Z_n\}_{n\geq 0}$ is said to have initial distribution π and transition K if

$\mathbb{P}\{Z_0 \in E\} = \pi(E)$,

$\mathbb{P}\left\{Z_{n+1} \in E \middle| Z_n = x\right\} = K_x(E)$.

Example: $\Omega = X^+ = M^{\mathbb{N}}$, $\mathcal{F} = \mathcal{B}^+ = \sigma$-algebra generated by cylinders. $\mathbb{P} = \mathbb{P}_{(\pi,K)}$, $Z_n : X^+ \to M, Z_n(x) = x_n \,\forall\, n \geq 0$.

Note that any K-stationary Markov chain can be realized as the example because $Z_n : \Omega \to M$ can always be written as $Z_n = e_n \circ Z$, when $Z(\omega) = \{Z_n(\omega)\}_n \in X^+$ and $e_n(\{Z_n(\omega)\}_{n\geq 0}) = Z_n(\omega)$ is the standard projection on to the n-th coordinates.

We have already defined the Markov operator in Definition 2.5. Here we consider Q defined on the space of continuous functions on M.

Example 1. Let (M, f) be a DDS, then $K_x = \delta_{f(x)}$. The corresponding Markov operator

$$(Q\varphi)(x) = \int \varphi(y) d\delta_{f(x)}(y) = \varphi(f(x)).$$

Namely $Q\varphi = \varphi \circ f$ is the Koopman operator.

Example 2. $K_{(g_0,\hat{v})} = \mu \times \delta_{\widehat{g_0 v}}$, then the corresponding Markov operator is

$$(Q\varphi)(g_0, \hat{v}) = \int \varphi(g_1, \widehat{g_0 v}) d\mu(g_1).$$

There are some basic properties of the Markov operator Q. Let us give two examples.

(1) $Q1 = 1$,
(2) $\|Q\varphi\|_\infty \leq \|\varphi\|_\infty$,
(3) If $\varphi \geq 0$, then $Q\varphi \geq 0$.

The dual of Q, denoted by Q^*, acts on the space of probabilities $\text{Prob}(M)$. By definition, we have $\langle \varphi, Q^*\nu\rangle = \langle Q\varphi, \nu\rangle$ for any $\varphi \in C(M)$ and $\nu \in \text{Prob}(M)$. In other words, $Q^*\nu$ is the probability on M s.t.

$$\int_M \varphi dQ^*\nu = \int_M Q\varphi d\nu, \quad \forall\, \varphi \in C(M).$$

Note that μ is K-stationary $\Leftrightarrow Q^*\mu = \mu$.

In practice, the assumptions in Definition 2.5 is unreasonably strong, primarily because of $\varphi \in C(M)$. We are going to replace it by something weaker.

let $(\mathcal{E}, \|\cdot\|_{\mathcal{E}})$ be a Banach space where $\mathcal{E} \subset C^0(M)$ is Q-invariant in the sense that $\varphi \in \mathcal{E} \Leftrightarrow Q\varphi \in \mathcal{E}$. Moreover, we assume the constant function $1 \in \mathcal{E}$ and the inclusion of $\mathcal{E} \subset C^0(M)$ is continuous, namely $\|\varphi\|_\infty \leq C_1 \|\varphi\|_{\mathcal{E}}$ for some constant $C_1 < \infty$. We also assume that Q is bounded (or continuous) on $(\mathcal{E}, \|\cdot\|_{\mathcal{E}})$, i.e. $\|Q\varphi\|_{\mathcal{E}} \leq C_2 \|\varphi\|_{\mathcal{E}}$ with $C_2 < \infty$.

Definition 2.6 (Weaker version of strongly mixing)**.** The system $(M, K, \mu, \mathcal{E})$ is strongly mixing if $\forall\, n \geq 0$,

$$\left\|(Q^n\varphi)(x) - \int_M \varphi d\mu\right\|_\infty \leq C \|\varphi\|_{\mathcal{E}}\, r_n,\ \forall\, \varphi \in \mathcal{E}$$

for some $C < \infty$ and for some mixing rate $\{r_n\}_{n\geq 0}$ (e.g. $r_n = \rho^n$ with $\rho \in (0,1)$ or $r_n = \frac{1}{(1+n)^p}$ with $p > 0$).

Let $\{Z_n\}_{n\geq 0}$ be the K-Markov chain, $Z_n : X^+ \to M, Z_n(x) = x_n$, for $\varphi : M \to \mathbb{R}$, we denote

$$S_n\varphi = \varphi(Z_0) + \cdots \varphi(Z_{n-1}) := \varphi_0 + \cdots + \varphi_{j-1}.$$

Theorem 2.1 (Cai, Duarte, Klein 2022)**.** *If $(M, K, \mu, \mathcal{E})$ is a strongly mixing Markov system with mixing rate $r_n = \frac{1}{(1+n)^p}, p > 0$, then $\forall\, x \in M$ and $\forall\, \epsilon > 0$*

$$\mathbb{P}_x\left\{\left|\frac{1}{n}S_n\varphi - \int_M \varphi d\mu\right| > \epsilon\right\} \leq e^{-c(\epsilon)n}$$

holds for all $n \geq 0$, for all $\varphi \in \mathcal{E}$ and for some $c(\epsilon) > 0$.

Note that $\mathbb{P}_x$ can be replaced by $\mathbb{P}_\mu$ simply because $\mathbb{P}_\mu = \int_M \mathbb{P}_x d\mu(x)$. One should be careful with the notation x because it is both used as the element in M and the element in X^+ respectively.

Proof. Without loss of generality, we assume $\mathbb{E}\varphi = 0$, otherwise we consider $\varphi - \mathbb{E}\varphi$. Moreover, it is enough to consider $\mathbb{P}_x\{S_n\varphi \geq n\epsilon\}$. Using Bernstein's trick, for any $t > 0$ we have

$$\mathbb{P}_x\{S_n\varphi \geq n\epsilon\} = \mathbb{P}_x\{e^{tS_n\varphi} \geq e^{tn\epsilon}\} \leq e^{-tn\epsilon}\mathbb{E}_x(e^{tS_n\varphi}).$$

So our goal in the following is to estimate $\mathbb{E}_x(e^{tS_n\varphi})$ by relating it to $Q^{n_0}(\varphi)$ for some suitable choice of $n_0 \leq n$.

Note that

$$e^{tS_n\varphi} = \prod_{j=0}^{n-1} e^{t\varphi_j} := \prod_{j=0}^{n-1} f_j = f_0 \cdots f_{n-1}.$$

Take $n_0 \le n$ such that $n = n_0 m + r$ and we first treat the case when $r = 0$. The key trick that we use is the following. We rewrite $f_0 \cdots f_{n-1}$ as

$$(f_0 f_{n_0} \cdots f_{(m-1)n_0})(f_1 f_{n_0+1} \cdots f_{(m-1)n_0+1}) \cdots (f_{n_0-1} f_{2n_0-1} \cdots f_{mn_0-1}).$$

We denote $F_j = f_j f_{n_0+j} \cdots f_{(m-1)n_0+j}$. By using Hölder inequality, we have

$$\mathbb{E}_x(\prod_{j=0}^{n-1} f_j) = \mathbb{E}_x(F_0 F_1 \cdots F_{n_0-1}) \le \prod_{r=0}^{n_0-1} [\mathbb{E}_x(F_r^{n_0})]^{\frac{1}{n_0}}.$$

Thus, it is enough to estimate each $\mathbb{E}_x(F_r^{n_0})$.

Aula 5 is basically the complete proof of this theorem, you told me to leave to you so I will jump to Aula 6 □

We also state an abstract CLT here, the proof will be given later.

Theorem 2.2 (Abstract CLT, Gordin-Livšic). *Let $\psi \in L^2(\mu)$, $\int \psi d\mu = 0$. Assume that*

$$\sum_{n=0}^{\infty} \|Q^n \psi\|_2 < \infty.$$

Then, denoting $\varphi = \sum_{n=0}^{\infty} Q^n \psi$, we have $\varphi \in L^2(\mu)$ and $\psi = \varphi - Q\varphi$. If $\sigma^2 = \|\varphi\|_2^2 - \|Q\varphi\|_2^2 > 0$, then the following CLT holds:

$$\frac{S_n \psi}{\sigma \sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1).$$

Corollary 2.1. *If (M, K, μ) is strongly mixing on $\mathcal{E}$ with mixing rate $r_n = \frac{1}{(1+n)^p}, p > 1$, then CLT holds for any observable $\psi \in \mathcal{E}$ provided $\sigma(\psi) > 0$.*

In the next subsections, we will introduce examples of dynamical systems that fit this abstract framework.

2.3. **Applications of the abstract LDT.** We will mainly study two certain skew-products.

Mixed random-quasiperiodic systems. Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ be the one dimensional torus. Assume $\alpha \notin \mathbb{Q}$, let $\tau_\alpha : \mathbb{T} \to \mathbb{T}$ be the torus translation by α such that $\tau_\alpha(\theta) = \theta + \alpha (\mod 1)$. Therefore, $(\mathbb{T}, \tau_\alpha, m)$ is an ergodic MPDS.

Remark 2.3. The (Markov) Koopman operator of this system is not strongly mixing, so the torus translation cannot be studied in this abstract framework. This is simply because $Q^n \varphi(\theta) = \varphi(\theta + n\alpha) \nrightarrow \int \varphi d\mu$ as $n \to \infty$ for any non-constant $\varphi \in C^0(\mathbb{T})$.

Therefore, instead of torus translation, we are going to consider an iterated functions system (IFS) of rotations.

Let $\mu \in \text{Prob}(\mathbb{T})$, denote $\{\alpha_n\}_{n\geq 0}$ an i.i.d. sequence of translations with distribution μ. We consider the iterates

$$\theta \mapsto \theta + \alpha_0 \mapsto \theta + \alpha_0 + \alpha_1 \mapsto \cdots$$

Then given $\varphi : \mathbb{T} \to \mathbb{R}$ an observable, we may consider

$$Q\varphi(\theta) = \int \varphi(\theta + \alpha) d\mu(\alpha).$$

Obviously, the corresponding kernal $K : \mathbb{T} \to \text{Prob}(\mathbb{T})$ is $K_\theta = \mu * \delta_\theta$.

It turns out that the system $(\mathbb{T}, K, m)$ is strongly mixing with a certain rate r_n having either polynomical or exponential decay, provided μ satisfies some general arithmetic properties (to be defined later) and φ is Hölder continuous. The proof will use some Fourier Analysis.

Note that the observable φ above only depends on one variable. In fact, we will consider a more complex system which allows φ to depend on infinite coordinates.

Regard $\Sigma := \mathbb{T}$ as the space of symbols with the measure μ. Let $X := \Sigma^{\mathbb{Z}}$ and consider the shift system $(X, \sigma, \mu^{\mathbb{Z}})$ where σ is the two sided Bernoulli shift. Then the skew product dynamical system is defined by

$$f : X \times \mathbb{T} \to X \times \mathbb{T},\ f(\alpha, \theta) = (\sigma\alpha, \theta + \alpha_0).$$

The n-iterates are $f^n(\alpha, \theta) = (\sigma^n\alpha, \theta + \alpha_0 + \cdots + \alpha_{n-1})$.

The triple $X \times \mathbb{T}, f, \mu^{\mathbb{Z}} \times m$ is called a mixed random-quasiperiodic system. Under certain general assumptions on μ, it is ergodic and it satisfied LDT and CLT for certain types of observables.

Certain types of linear cocycles. Examples are Random, Markov, Fiber-bunched and Mixed cocycles.

We first recall the definition of linear cocycles. For more details, see [Viana] and [DK-CBM].

Let (X, f, μ) be an ergodic MPDS. A linear cocycle over (X, f, μ) is a skew-product map

$$F : X \times \mathbb{R}^2 \to X \times \mathbb{R}^2,\ F(x, v) = (f(x), A(x)v),$$

where $A : X \to \text{GL}_2(\mathbb{R})$ is a measurable function. We ususally call f the base dynamics and A the fiber dynamics. We may also consider the projective cocycle.

$$\hat{F} : X \times \mathbb{P} \to X \times \mathbb{P},\ \hat{F}(x, \hat{v}) = (f(x), \widehat{A(x)v}).$$

The n-th iterates of the cocycle are $F^n(x, v) = (f^n(x), A^n(x)v)$ where

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

are called transfer matrices in Mathematical Physics.

We will always assume a mild integrability condition:

$$\int_X \log \|A(x)\| \, d\mu(x) < \infty.$$

Denote by $\varphi_x(x) := \log \|A^n(x)\|$, then the sequence $\{\varphi_n\}_{n\geq 0}$ is f-subadditive in the sense that

$$\varphi_{n+m} \leq \varphi_n \circ f^m + \varphi_m, \, \forall \, m, n \in \mathbb{N}$$

By Kingman's subadditive ergodic theorem,

$$\frac{1}{n}\varphi \to L, \quad \mu\text{-a.e.}$$

That is

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)\| = L^+(A), \quad \mu\text{-a.e.}\, x \in X$$

where $L^+(A)$ is called the maximal Lyapunov exponent of A. Moreover, for μ-a.e. $x \in X$,

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)\| = \lim_{n\to\infty} \int \frac{1}{n} \log \|A^n\| \, d\mu = \inf_{n\geq 1} \int \frac{1}{n} \log \|A^n\| \, d\mu.$$

By a similar argument, we have

$$\lim_{n\to\infty} \frac{1}{n} \log \left\|A^n(x)^{-1}\right\|^{-1} = L^-(A), \quad \mu\text{-a.e.}\, x \in X$$

Note that $\forall \, g \in \mathrm{GL}_2(\mathbb{R})$, $\|g^{-1}\|^{-1} \leq \|g\|$, so $L^-(A) \leq L^+(A)$.

We recall the Oseledets multiplicative ergodic theorem.

Theorem 2.3. *Let $F_A : X \times \mathbb{R}^2 \to X \times \mathbb{R}^2$ be a μ-integrable cocycle given by $A : X \to \mathrm{GL}_2(\mathbb{R})$ over an ergodic MPDS (X, f, μ), then*

(1) If $L^+(A) = L^-(A)$, then $\forall \, v \in \mathbb{R}^2$ non-zero,

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)v\| = L^+(A), \quad \mu\text{-a.e.}\, x \in X$$

(2) If $L^+(A) > L^-(A)$, then there is a measurable map

$$x \mapsto V_x \subset \mathbb{R}^2$$

where V_x is a one dimensional subspace of $\mathbb{R}^2$, such that

$$A(x)V_x = V_{f(x)}$$

i.e. V_x is an F- invariant section. Moreover, if $v \notin V_x$, then

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

Otherwise, if $v \in V_x$*, then*

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

Moreover, if f *is invertible then there exists a measurable splitting of the fiber: for* μ*-almost every* $x \in X$, $\mathbb{R}^2 = E_x^+ \oplus E_x^-$ *such that*

(1) $A(x)E_x^\pm = E_{f(x)}^\pm$.
(2) $\lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)v\| = L^\pm(A)$, $v \in E_x^\pm, v \neq 0$.
(3) $\lim_{n\to\infty} \frac{1}{n} \log \left|\sin \angle(E_{f^n(x)}^+, E_{f^n(x)}^-)\right| = 0$

Examples of linear cocycles are quasi-periodic cocycles over a torus translation τ_α (which does not fit our framework) and random cocycles over a Bernoulli shift σ.

3. Large deviations for random linear cocycles

We begin with the definition of a random linear cocycle.

Setup. Let (Σ, μ) be a probability space (Σ is always assumed to be compact throughout this section). Denote $X := \Sigma^{\mathbb{Z}}$ and let σ be the two sided (Bernoulli) shift. $\mu^{\mathbb{Z}}$ is the product measure on the infinite product space X. The triple $(X, \sigma, \mu^{\mathbb{Z}})$ is called a Bernoulli shift. This is the base dynamics.

Let $A : X \to \mathrm{GL}_2(\mathbb{R})$ be a continuous random cocycle. Moreover, assume that A is locally constant, namely, $A(\omega) = A(\omega_0)$ where $\omega = \{\omega_n\}_{n\in\mathbb{Z}}$. Given the Bernoulli shift, A determines a random linear cocycle

$$F = F_A : X \times \mathbb{R}^2 \to X \times \mathbb{R}^2,\ F(\omega, v) = (\sigma\omega, A(\omega_0)v).$$

The n-th iterates of the cocycle are

$$F^n(\omega, v) = (\sigma^n\omega, A^n(\omega)v)$$

where $A^n(\omega) = A(\omega_{n-1}) \cdots A(\omega_1)A(\omega_0)$. As before, we may also consider the projective cocycle $\hat{F}$ that is similarly defined.

We say that F satisfies a fiber LDT (or A satisfies an LDT) if $\forall\, v \neq 0, v \in \mathbb{R}^2, \forall\, \epsilon > 0$

$$\mu^{\mathbb{Z}} \left\{\omega \in X : \left|\frac{1}{n} \log \|A^n(\omega)v\| - L^+(A)\right| > \epsilon\right\} < e^{-c(\epsilon)n}$$

for all $n \geq n(\epsilon, A)$ and for some $c(\epsilon) > 0$.

We will prove this LDT under certain "generic assumptions" on A and μ. Under the same assumption, we will also get a CLT:

$$\frac{\log \|A^n(\omega)v\| - nL^+(A)}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1)\,.$$

Remark 3.1. Note that since $A : \Sigma \to \mathrm{GL}_2(\mathbb{R})$ is continuous, then $\nu = A_*\mu \in \mathrm{Prob}_c(\mathrm{GL}_2(\mathbb{R}))$. Therefore, we can start with a compactly supported probability measure ν in $\mathrm{Prob}_c(\mathrm{GL}_2(\mathbb{R}))$ and consider the multiplicative process associated to an i.i.d. sequence of random matrices $\{g_n\}_{n\in\mathbb{Z}}, g_n \in \mathrm{GL}_2(\mathbb{R})$ with distribution ν. These two settings are essentially equivalent.

Generic assumptions. Let (Σ, μ) be a probability space, $A \in \Sigma \to \mathrm{GL}_2(\mathbb{R})$.

Definition 3.1. A line $l \subset \mathbb{R}^2$ is A-invariant if $A(x)l = l$ for μ-a.e. $x \in \Sigma$.

Let $\mathcal{H}_A$ be the group generated by the support of $A_*\mu$. Note that if l is A-invariant, then l is $\mathcal{H}_A$-invariant.

Definition 3.2. A cocycle A is called irreducible if there is no A-invariant line.

Definition 3.3. A cocycle A is called strongly irreducible if there is no finite union of lines which is A-invariant. Namely, $\forall\, n \in \mathbb{Z}^+$, there exist no lines $\{l_j\}_{1\le j\le n}$ such that $A(x)\bigcup_{j=1}^n l_j = \bigcup_{j=1}^n l_j$ for a.e. $x \in \Sigma$.

Definition 3.4. A (or $A_*\mu$) is called non-compact if there exists a sequence of matrices $\{h_n\}_{n\ge1} \subset \mathcal{H}_A$ such that $\|h_n\| \cdot \|h_n^{-1}\| \to \infty$.

We introduce a profound theorem of Furstenberg.

Theorem 3.1 (Furstenberg's Theorem)**.** *If A is non-compact and strongly irreducible, then $L^+(A) > L^-(A)$.*

Example 1. Triangular matrices $A : \Sigma \to \mathrm{GL}_2(\mathbb{R})$:

$$A(x) = \begin{pmatrix} a(x) & b(x) \\ 0 & c(x) \end{pmatrix}$$

is reducible because the line l of the direction $(1, 0)$ is A-invariant.

Example 2. Random Schrödinger cocycles. Let $\Sigma \subset \mathbb{R}$ be compact and $\mu \in \mathrm{Prob}(\Sigma)$. Then $S : \Sigma \to \mathrm{GL}_2(\mathbb{R})$ is defined by

$$S(a) = \begin{pmatrix} a & -1 \\ 1 & 0 \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}).$$

Assume that $\#\mathrm{supp}(\mu) \ge 2$ (μ is not a single Dirac), then S is non-compact and strongly irreducible. We leave the proof to the readers. Hint: play with $S(a)S(b)^{-1}$ and $S(a)^{-1}S(b)$, show that $(1, 0) \cup (0, 1)$ is the only potential candidate for S-invariance and prove it is actually not S-invariant. Thus strongly irreducible condition is fulfilled. As for non-compactness, take n-th power of either $S(a)S(b)^{-1}$ or $S(a)^{-1}S(b)$.

Let $l \subset \mathbb{R}^2$ be an A-invariant line, namely $A(x)l = l$ for μ-a.e. $x \in \Sigma$, then we can restrict A to l. We denote it by $A|_l$. Fix a unit vector $v \in l$, then $A(\omega)v = \lambda(\omega)v$ for some $\lambda : X \to \mathbb{R}$ which is in fact also locally constant. Let $A^n(\omega)v = A(\sigma^{n-1}\omega)\cdots A(\sigma\omega)A(\omega)v$, then by Birkhoff ergodic theorem

$$\frac{1}{n}\log\|A^n(\omega)v\| = \frac{1}{n}\sum_{j=0}^{n-1}\log\left|\lambda(\sigma^j\omega)\right| = \int_X \log|\lambda(\omega)|\, d\mu^{\mathbb{Z}}(\omega) = L(A|_l).$$

Definition 3.5. A is called quasi-irreducible if either there is no A-invariant line or $L(A|_l) = L^+(A)$.

We will prove the following theorem.

Theorem 3.2 (Le-Page, Duarte-Klein)**.** *If A is quasi-irreducible and $L^+(A) > L^-(A)$, then A satisfies LDT: $\forall\, \epsilon > 0$*

$$\mu^{\mathbb{Z}}\left\{\omega \in X : \left|\frac{1}{n}\log\|A^n(\omega)v\| - L^+(A)\right| > \epsilon\right\} < e^{-c(\epsilon)n}$$

holds $\forall\, v \neq 0, v \in \mathbb{R}^2$, $\forall\, n \geq n(\epsilon, A)$ and for some $c(\epsilon) > 0$.

General strategy for the proof. Consider the projective cocycle

$$\hat{F}_A : X \times \mathbb{P} \to X \times \mathbb{P},\ \hat{F}_A(\omega, \hat{v}) = (\sigma\omega, \hat{A}(\omega_0)\hat{v}).$$

The corresponding Markov chain on $M := \Sigma \times \mathbb{P}$ is

$$(\omega_0, \hat{v}) \to (\omega_1, \hat{A}(\omega_0)\hat{v}) \to (\omega_2, \hat{A}(\omega_1)\hat{A}(\omega_0)\hat{v}) \to \cdots$$

where we denote $(\omega_n, \hat{A}(\omega_{n-1})\cdots\hat{A}(\omega_0)\hat{v}) =: x_n$.

The associated SDS is

$$\bar{K} : \Sigma \times \mathbb{P} \to \mathrm{Prob}(\Sigma \times \mathbb{P}),\ \bar{K}_{(\omega_0,\hat{v})} = \mu \times \delta_{\hat{A}(\omega_0)\hat{v}}.$$

This kernal $\bar{K}$ defines a Markov operator

$$\bar{Q} : C^0(\Sigma \times \mathbb{P}) \to C^0(\Sigma \times \mathbb{P}),$$

$$\bar{Q}\varphi(\omega_0, \hat{v}) = \int_\Sigma \varphi(\omega_1, \hat{A}(\omega_0)\hat{v})d\mu(\omega_1).$$

We will consider a special observable $\xi = \xi_A : \Sigma \times \mathbb{P} \to \mathbb{R}$ such that

$$\xi_A(\omega_0, \hat{v}) = \log\|A(\omega_0)v\|$$

where v is a unit representative of $\hat{v}$.

Recall that $m \in \mathrm{Prob}(\Sigma \times \mathbb{P})$ is $\bar{K}$-stationary if and only if $\bar{Q}^*m = m$ where Q^* is the dual of Q. Then by Furstenberg's Formula, we have

$$L^+(A) = \max_{m \in \mathrm{Prob}_{\bar{K}}(\Sigma\times\mathbb{P})}\left\{\int_{\Sigma\times\mathbb{P}} \xi_A(\omega_0, \hat{v})dm(\omega_0, \hat{v})\right\}$$

Let $x = \{x_n\}_{n\geq 0} \in M^{\mathbb{N}}$, then if we start with an initial $\hat{v}$ which is a unit vector. we have

$$S_n\xi_A(x) = \xi_A(x_0) + \xi_A(x_1) + \cdots + \xi_A(x_{n-1}) = \log \|A^n(\omega)v\| .$$

Thus

$$\frac{1}{n}S_n\xi_A(x) = \frac{1}{n}\log \|A^n(\omega)v\| . \tag{3.1}$$

Note that the l.h.s. will converge to $\int \xi_A dm = L^+(A)$ by Birkhoff, so intuitively the LDT should follow.

3.1. **Stationary measures.** Equation (3.1) shows that in order to prove fiber-LDT for A, it would be enough to prove the corresponding Markov chain with observable ξ_A. For this purpose, it would be enough to show (because of the abstract LDT) that the Markov operator $\bar{Q}$ is strongly mixing on some appropriate space $(\mathcal{E}, \|\cdot\|_{\mathcal{E}})$ which contains the special observable ξ_A.

A priori, this space will be $\mathcal{H}^\alpha(\Sigma\times\mathbb{P})$ of α-Hölder continuous functions in $\mathbb{P}$ with some appropriate norm with some $\alpha > 0$. On this space, $\bar{Q}$ will be shown to be quasi-compact and simple in which case $r_n = \sigma^n$ with $\sigma \in (0,1)$. In fact, it will be convenient to work with a simpler kernal and the associated Markov operator.

Let $Q : C(\mathbb{P}) \to C(\mathbb{P})$ such that

$$Q\psi(\hat{v}) = \int \psi(\hat{A}(\omega_0)\hat{v})d\mu(\omega_0).$$

Then Q is the Markov operator corresponding to the kernal $K : \mathbb{P} \to \mathrm{Prob}(\mathbb{P})$:

$$K_{\hat{v}} = \int \delta_{\hat{A}(\omega_0)\hat{v}} d\mu(\omega_0).$$

Consider the projection $\Pi : C(\Sigma\times\mathbb{P}) \to C(\mathbb{P})$ defined by

$$\Pi\varphi(\hat{v}) = \int \varphi(\omega_0, \hat{v})d\mu(\omega_0).$$

Lemma 3.1. *The following diagram is commutative.*

$$\begin{tikzcd}
C^0(\Sigma\times\mathbb{P}) \arrow[r, "\bar{Q}"] \arrow[d, "\Pi"'] & C^0(\Sigma\times\mathbb{P}) \arrow[d, "\Pi"] \\
C^0(\mathbb{P}) \arrow[r, "Q"] & C^0(\mathbb{P})
\end{tikzcd}$$

Namely, $\Pi\circ\bar{Q} = Q\circ\Pi$.

Proof. A simple calculation. ∎

Lemma 3.2. *$\forall \varphi \in C^0(\Sigma \times \mathbb{P})$, $\forall n \geq 1$, we have*

$$\bar{Q}^n \varphi(\omega_0, \hat{v}) = Q^{n-1}(\Pi\varphi)(\hat{A}(\omega_0)\hat{v}).$$

This shows that in order to prove that $\bar{Q}$ is strongly mixing on $\mathcal{E}$, it is enough to show that Q is strongly mixing on $\Pi(\mathcal{E})$.

Proof.

$$\begin{aligned}\bar{Q}^n \varphi(\omega_0, \hat{v}) &= \int_{\Sigma^n} \varphi(\omega_n, \hat{A}^n(\omega)\hat{v}) d\mu(\omega_n) \cdots d\mu(\omega_1) \\ &= \int_{\Sigma^n} \varphi(\omega_n, \hat{A}(\omega_{n-1}) \cdots \hat{A}(\omega_0)\hat{v}) d\mu(\omega_n) \cdots d\mu(\omega_1).\end{aligned}$$

On the other hand, for any $\psi \in C(\mathbb{P})$ and $\hat{p} \in \mathbb{P}$, we have

$$Q^{n-1}\psi(\hat{p}) = \int_{\Sigma^{n-1}} \psi(\hat{A}(\omega_{n-1}) \cdots \hat{A}(\omega_1)\hat{p}) d\mu(\omega_{n-1}) \cdots d\mu(\omega_1). \quad (3.2)$$

If we take $\hat{p} = \hat{A}(\omega_0)\hat{v}$ and $\psi = \Pi\varphi$, then (3.2) equals to

$$\int_{\Sigma^{n-1}} \int_{\Sigma} \varphi(\omega_n, \hat{A}(\omega_{n-1}) \cdots \hat{A}(\omega_0)\hat{v}) d\mu(\omega_{n-1}) \cdots d\mu(\omega_1).$$

This finished the proof. $\square$

Recall that given a Markov kernal $K : M \to \mathrm{Prob}(M)$, a measure $\eta \in \mathrm{Prob}(M)$ is called K-stationary if $Q^*\eta = \eta$ where Q^* is the dual of the Markov operator Q associated with K. In this case, we will denote $\eta \in \mathrm{Prob}_K(M)$. In fact, there are several equivalent definitions as follows:

(1) $Q^*\eta = \eta$
(2) $\forall \varphi \in C^0(M)$, $\int_M Q\varphi d\eta = \int_M \varphi d\eta$.
(3) $\forall \varphi \in C^0(M)$, $\int_M \int_M \varphi(y) dK_x(y) d\eta(x) = \int_M \varphi d\eta$.
(4) $K * \eta = \eta$ where $K * \eta = \int K_x d\eta(x)$.

Proposition 3.3. *Given $\eta \in \mathrm{Prob}(\mathbb{P})$, the following are equivalent (TFAE):*

(1) η is K-stationary.
(2) $\mu \times \eta$ is $\bar{K}$-stationary.
(3) $\mu^{\mathbb{N}} \times \eta$ is $\hat{F}^+$-invariant where $\hat{F}^+ : X^+ \times \mathbb{P} \to X^+ \times \mathbb{P}$. Namely, $(X^+ \times \mathbb{P}, \hat{F}^+, \mu^{\mathbb{N}} \times \eta)$ is an MPDS.

Proof. We will first prove (1) $\Leftrightarrow$ (2). It is enough to show that $K * \eta = \eta \Leftrightarrow \bar{K} * (\mu \times \eta) = \mu \times \eta$. In fact, we will show that

$$\bar{K} * (\mu \times \eta) = \mu \times (K * \eta). \quad (3.3)$$

This will conclude the proof because if $K * \eta = \eta$, then $\bar{K} * (\mu \times \eta) = \mu \times \eta$. If $\bar{K} * (\mu \times \eta) = \mu \times \eta$, then $\mu \times (K * \eta) = \mu \times \eta$ which gives $K * \eta = \eta$.

Note that (3.3) is equivalent to saying that: $\forall \varphi \in C^0(\Sigma \times \mathbb{P})$,

$$\int \varphi d[\bar{K} * (\mu \times \eta)] = \int \varphi d\mu d(K * \eta).$$

We first look at the left hand side. By the definition of convolution,

$$\bar{K} * (\mu \times \eta) = \int \bar{K}_{(\omega_0, \hat{v})} d\mu(\omega_0) d\eta(\hat{v}) = \int \mu \times \delta_{\hat{A}(\omega_0)\hat{v}} d\mu(\omega_0) d\eta(\hat{v}).$$

Thus we have

$$\int \varphi d[\bar{K} * (\mu \times \eta)] = \int \varphi(\omega_1, \hat{A}(\omega_0)\hat{v}) d\mu(\omega_1) d\mu(\omega_0) d\eta(\hat{v}).$$

Now let us focus on the r.h.s.

$$K * \eta = \int K_{\hat{v}} d\eta(\hat{v}) = \int \int \delta_{\hat{A}(\omega_0)\hat{v}} d\mu(\omega_0) d\eta(\hat{v}).$$

Thus we have

$$\int \varphi d\mu d(K * \eta) = \int \int \int \varphi(\omega_1, \hat{A}(\omega_0)\hat{v}) d\mu(\omega_1) d\mu(\omega_0) d\eta(\hat{v}).$$

This proves (1) $\Leftrightarrow$ (2).

In the following, we are going to prove (1) $\Leftrightarrow$ (3). Recall that $\mu^{\mathbb{N}} \times \eta$ is $\hat{F}^+$-invariant if and only if $\forall \varphi \in C^0(X^+ \times \mathbb{P})$,

$$\int \varphi d\mu^{\mathbb{N}} \times \eta = \int \varphi \circ \hat{F}^+ d\mu^{\mathbb{N}} \times \eta.$$

More precisely,

$$\int \varphi(\omega, \hat{v}) d\mu^{\mathbb{N}}(\omega_0, \omega_1, \cdots) d\eta(\hat{v}) = \int \varphi(\sigma\omega, \hat{A}(\omega_0)\hat{v}) d\mu^{\mathbb{N}}(\omega_1, \omega_2, \cdots) d\eta(\hat{v}).$$

If we denote $\psi := \int \varphi(\omega, \hat{v}) d\mu^{\mathbb{N}}(\omega) \in C^0(\mathbb{P})$ which is arbitrary since φ is arbitrary, then the l.h.s. becomes $\int \psi d\eta$ and the r.h.s. becomes

$$\begin{aligned} \int \varphi(\sigma\omega, \hat{A}(\omega_0)\hat{v}) d\mu^{\mathbb{N}}(\omega_1, \omega_2, \cdots) d\eta(\hat{v}) &= \int \int \psi(\hat{A}(\omega_0)\hat{v}) d\mu(\omega_0) d\eta(\hat{v}) \\ &= \int Q\psi(\hat{v}) d\eta(\hat{v}). \end{aligned}$$

Thus it is clear that (1) $\Leftrightarrow$ (3). This finishes the proof. $\square$

Before we proceed, we recall some convex analysis concepts.

Let $\mathfrak{X}$ be a topological vector space that is Hausdorff and locally convex. Given $D \subset \mathfrak{X}$, $p \in D$ is an extreme point of D if it is not between any two different points in D. That is, there are no $x, y \in D$ with $x \neq y$ such that for some $t \in (0,1)$,

$$p = tx + (1-t)y.$$

Theorem 3.3 (Krein-Milman)**.** *If $D \subset \mathfrak{X}$ is compacy, convex and non-empty, then D has at least one extreme point, i.e. extreme$(D) \neq \emptyset$. Moreover, the closed convex hull of extreme(D) is D.*

Here the closed convex hull $\overline{Co}(S)$ is the smallest closed convex set containing S.

Example 1. Let M be a compact metric space. $D := \mathrm{Prob}(M)$ with the weak* topology is compact convex and non-empty. So Krein-Milman applies. In this case, $\mathfrak{X}$ is the space of signed measures which is metrizable with the weak* topology.

Example 2. Under the same settings as in Ex 1, let $D := \mathrm{Prob}_K(M)$ which is closed. Thus D is compact, convex and non-empty. So Krein-Milman also applies.

Stationary measures, continuation.

To be more precise, let us rewrite the three levels of objects.

(1) DDS, projective linear cocycle:

$$\hat{F}^+ : X^+ \times \mathbb{P} \to X^+ \times \mathbb{P},\ \hat{F}^+(\omega, \hat{v}) = (\sigma\omega, \hat{A}(\omega_0)\hat{v}).$$

(2) SDS on $\Sigma \times \mathbb{P}$:

$$\bar{K}_{(\omega,\hat{v})} = \mu \times \delta_{\hat{A}(\omega_0)\hat{v}},$$

with the corresponding Markov operator $\bar{Q}$:

$$\bar{Q} : C^0(\Sigma \times \mathbb{P}) \to C^0(\Sigma \times \mathbb{P}),$$

$$\bar{Q}\varphi(\omega_0, \hat{v}) = \int_\Sigma \varphi(\omega_1, \hat{A}(\omega_0)\hat{v}) d\mu(\omega_1).$$

$\bar{K}$-Markov chain $\{Z_n\}_{n\geq 0}$ where $Z_n : X^+ \times \mathbb{P} \to \Sigma \times \mathbb{P}$ such that

$$Z_0(\omega, \hat{v}) = (\omega_0, \hat{v}), \quad Z_n(\omega, \hat{v}) = (\omega_n, \hat{A}^n(\omega)\hat{v})$$

with initial distribution $\mu \times \delta_{\hat{v}}$ (non-stationary case) or $\mu \times \eta$ (stationary case).

(3) SDS on $\mathbb{P}$:

$$K_{\hat{v}} = \int_\Sigma \delta_{\hat{A}(\omega_0)\hat{v}} d\mu(\omega_0),$$

with the corresponding Markov operator Q:

$$Q : C^0(\mathbb{P}) \to C^0(\mathbb{P}), \quad Q\varphi(\hat{v}) = \int_{\mathbb{P}} \varphi(\hat{A}(\omega_0)\hat{v}) d\mu(\omega_0).$$

We will mainly consider the special observable on $\Sigma \times \mathbb{P}$:

$$\bar{\xi} : \Sigma \times \mathbb{P} \to \mathbb{R}, \quad \bar{\xi}(\omega_0, \hat{v}) = \log \|A(\omega_0)v\|$$

where $v \in \hat{v}$ with $\|v\| = 1$.

The corresponding observable on $\mathbb{P}$ is $\xi = \Pi\bar{\xi} : \mathbb{P} \to \mathbb{R}$ where $\Pi\varphi(\hat{v}) = \int_\Sigma \varphi(\omega_0, \hat{v}) d\mu(\omega_0)$ for any $\varphi \in C^0(\Sigma \times \mathbb{P})$.

The corresponding observable on $X^+ \times \mathbb{P}$ is

$$\Phi : X^+ \times \mathbb{P} \to \mathbb{R}, \quad \Phi(\omega, \hat{v}) = \bar{\xi}(\omega_0, \hat{v}).$$

Remark 3.2. We emphasize that all the places where $\varphi, \psi \in C^0$ above can be replaced by $\varphi, \psi \in L^\infty$ simply because we can define the Markov operator not only on the continuous function space, but also on the space of essentially bounded functions.

In the following, we will prove that η is an extremal point of $\mathrm{Prob}_K(\mathbb{P})$ if and only if $\mu^\mathbb{N} \times \eta$ is $\hat{F}^+$-invariant.

Definition 3.6. An observable $\varphi \in L^\infty(\mathbb{P})$ is called η-stationary if $Q\varphi(\hat{v}) = \varphi(\hat{v})$ for η-a.e. $\hat{v} \in \mathbb{P}$. A Borel set $E \subset \mathbb{P}$ is called η-stationary if $\mathbb{1}_E$ is η-stationary. Or equivalently, η-a.e. $\hat{v} \in E \Leftrightarrow \hat{A}(\omega_0)\hat{v} \in E$ for μ-a.e. $\omega_0 \in \Sigma$.

The equivalence statement is due to the condition:

$$Q\mathbb{1}_E(\hat{v}) = \int_\Sigma \mathbb{1}_E(\hat{A}(\omega_0)\hat{v}) d\mu(\omega_0) = \mathbb{1}_E(\hat{v}),\ \eta\text{-a.e.}\ \hat{v} \in \mathbb{P}.$$

Proposition 3.4. *Let $\eta \in \mathrm{Prob}_K(\mathbb{P})$, the following are equivalent:*

(1) η is an extremal point of $\mathrm{Prob}_K(\mathbb{P})$.
(2) If $F \subset \mathbb{P}$ is η-stationary, then $\eta(F) = 0$ or 1.
(3) If $\varphi \in L^\infty(\mathbb{P})$ is η-stationary, then $\varphi \equiv const$, η-a.e.
(4) $(X^+ \times \mathbb{P}, \hat{F}^+, \mu^\mathbb{N} \times \eta)$ is an ergodic MPDS.

Proof. We prove by this order: (1) $\Rightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (4) $\Rightarrow$ (1).

(1) $\Rightarrow$ (2). Assume by contradiction the existence of $F \subset \mathbb{P}$ which is η-stationary with $t = \eta(F) \in (0, 1)$. The same holds for F^c. Namely, F^c is also η-stationary and $\eta(F^c) = 1 - t \in (0, 1)$.

Let

$$\eta_F \in \mathrm{Prob}(\mathbb{P}), \quad \eta_F(E) = \frac{\eta(E \cap F)}{\eta(F)}$$

which is the conditional probability. Then by the Law of total probability,

$$\eta = t\eta_F + (1 - t)\eta_{F^c}.$$

Moreover, since $\eta_F(F) = 1$ and $\eta_{F^c}(F) = 0$, we have $\eta_F \neq \eta_{F^c}$.

If we can show that $\eta_F \in \mathrm{Prob}_K(\mathbb{P})$(then so does η_{F^c}), we will get a contradiction because η is assumed to be an extremal point of $\mathrm{Prob}_K(\mathbb{P})$.

Let $\varphi \in L^\infty(\mathbb{P})$, direct computation shows

$$\begin{aligned}
\int_{\mathbb{P}} Q\varphi d\eta_F &= \frac{1}{\eta(F)} \int_F Q\varphi d\eta \\
&= \frac{1}{\eta(F)} \int_F \int_\Sigma \varphi(\hat{A}(\omega_0)\hat{v}) d\mu(\omega_0) d\eta(\hat{v}) \\
&= \frac{1}{\eta(F)} \int_{\mathbb{P}} \int_\Sigma \varphi(\hat{A}(\omega_0)\hat{v}) \mathbb{1}_F(\hat{v}) d\mu(\omega_0) d\eta(\hat{v}) \\
&= \frac{1}{\eta(F)} \int_{\mathbb{P}} \int_\Sigma \varphi(\hat{A}(\omega_0)\hat{v}) \mathbb{1}_F(\hat{A}(\omega_0)\hat{v}) d\mu(\omega_0) d\eta(\hat{v}) \\
&= \frac{1}{\eta(F)} \int_{\mathbb{P}} \int_\Sigma (\varphi|_F)(\hat{A}(\omega_0)\hat{v}) d\mu(\omega_0) d\eta(\hat{v}) \\
&= \frac{1}{\eta(F)} \int_{\mathbb{P}} \int_\Sigma (\varphi|_F)(\hat{v}) d\mu(\omega_0) d\eta(\hat{v}) \\
&= \frac{1}{\eta(F)} \int_{\mathbb{P}} Q(\varphi|_F)(\hat{v}) d\eta(\hat{v}) \\
&= \frac{1}{\eta(F)} \int_{\mathbb{P}} (\varphi|_F)(\hat{v}) d\eta(\hat{v}) \\
&= \frac{1}{\eta(F)} \int_F \varphi(\hat{v}) d\eta(\hat{v}) \\
&= \int_{\mathbb{P}} \varphi d\eta_F.
\end{aligned}$$

This proves that η_F is K-stationary, so is η_{F^c}.This contradicts that η is extremal, so (2) holds.

(2) $\Rightarrow$ (3). Let $\varphi \in L^\infty(\mathbb{P})$ be η-stationary. We will use the following useful fact from measure theory.

Exercise. φ is constant η-a.e. iff the sub-level sets $\{\varphi > c\} = \{\hat{v} : \varphi(\hat{v}) < c\}$ have η measure either 1 or 0, $\forall\ c \in \mathbb{R}$.

Fix $c \in \mathbb{R}$, let $E = \{\hat{v} : \varphi(\hat{v}) < c\}$. We will show that $\mathbb{1}_E$ is η-stationary. Namely, E is η-stationary and by (2) we obtain that $\eta(E)$ is either 0 or 1. Since c is arbitrary, by the Exercise above, we get φ is constant η-a.e.

Let $\mathcal{S} := \{\varphi \in L^\infty(\mathbb{P}) : \varphi \text{ is } \eta\text{-stationary}\}$. We will show that $\mathbb{1}_E \in \mathcal{S}$. We list two properties of $\mathcal{S}$ below:

(1) $\mathcal{S}$ is a linear space,
(2) $\mathcal{S}$ is a lattice.

Item (1) is obvious. For Item (2), being a lattice means

- $\phi \in \mathcal{S} \Rightarrow |\varphi| \in \mathcal{S}$.
- If $\varphi, \psi \in \mathcal{S}$, then $\min\{\varphi, \psi\}, \max\{\varphi, \psi\} \in \mathcal{S}$.

We prove the first item. Since $\eta \in \mathrm{Prob}_K(\mathbb{P})$, we have

$$\int_{\mathbb{P}} Q\,|\varphi| - |\varphi|\, d\eta = \int_{\mathbb{P}} Q\,|\varphi|\, d\eta - \int_{\mathbb{P}} |\varphi|\, d\eta = 0.$$

$\varphi \in \mathcal{S} \Rightarrow Q\varphi = \varphi$, η-a.e. This implies

$$\begin{aligned} |\varphi(\hat{v})| &= |Q\varphi(\hat{v})| \\ &= \left| \int \varphi(\hat{A}(\omega_0)\hat{v}) d\mu(\omega_0) \right| \\ &\leq \int \left| \varphi(\hat{A}(\omega_0)\hat{v}) \right| d\mu(\omega_0) \\ &= Q\,|\varphi|\,(\hat{v}). \end{aligned}$$

That is, $|\varphi| \leq Q\,|\varphi|$, η-a.e. Therefore, $|\varphi| = Q\,|\varphi|$, η-a.e. which shows $|\varphi| \in \mathcal{S}$.

The second item follows simply from the first item and the linearity because

$$\min\{\varphi, \psi\} = \frac{\varphi + \psi}{2} - \frac{|\varphi - \psi|}{2},$$

$$\max\{\varphi, \psi\} = \frac{\varphi + \psi}{2} + \frac{|\varphi - \psi|}{2}.$$

Now, let $\varphi_n(\hat{v}) = \min\{1, n \cdot \{c - \varphi(\hat{v}), 0\}\}$. Clearly, $\varphi_n \to \mathbb{1}_E$ as $n \to \infty$. Moreover, by the properties of $\mathcal{S}$ and the definition of φ_n, we have $\varphi_n \in \mathcal{S}$. Thus $Q\varphi_n = \varphi_n$, η-a.e. Then we have $Q\varphi_n = \varphi_n \to \mathbb{1}_E$ and also $Q\varphi_n \to Q\mathbb{1}_E$, η-a.e. Finally, by the uniqueness of limit, we have $Q\mathbb{1}_E = \mathbb{1}_E$, η-a.e. which proves $\mathbb{1}_E \in \mathcal{S}$. This gives (3).

(3) $\Rightarrow$ (4). To prove that $\mu^{\mathbb{N}} \times \eta$ is $\hat{F}^+$-ergodic, it is equivalent to showing that if $\psi \in L^\infty(X^+ \times \mathbb{P})$ satisfies $\psi \circ \hat{F}^+ = \psi$, $\mu^{\mathbb{N}} \times \eta$-a.e., then $\psi \equiv \text{const}$, $\mu^{\mathbb{N}} \times \eta$-a.e.

Let $\varphi : \mathbb{P} \to \mathbb{R}$, $\varphi(\hat{v}) = \int_{X^+} \psi(\omega, \hat{v}) d\mu^{\mathbb{N}}(\omega)$. We will first show that $\varphi \equiv \text{const}$, η-a.e. For this, it is enough to show that φ is η-stationary because of (3). By direct computation,

$$
\begin{aligned}
Q\varphi(\hat{v}) &= \int_\Sigma \varphi(\hat{A}(\omega_0)\hat{v})d\mu(\omega_0)\\
&= \int_\Sigma \int_{X^+} \Psi(\omega', \hat{A}(\omega_0)\hat{v})d\mu^{\mathbb{N}}(\omega')d\mu(\omega_0)\\
&= \int_{X^+} \psi(\sigma\omega, \hat{A}(\omega_0)\hat{v})d\mu^{\mathbb{N}}(\omega)\\
&= \int_{X^+} \psi \circ \hat{F}^+(\omega, \hat{v})d\mu^{\mathbb{N}}(\omega)\\
&= \int_{X^+} \psi(\omega, \hat{v})d\mu^{\mathbb{N}}(\omega)\\
&= \varphi(\hat{v})
\end{aligned}
$$

for η-a.e. $\hat{v} \in \mathbb{P}$.

It is left to show that ψ does not depend on $\omega = (\omega_0, \cdots, \omega_{k-1}, \cdots)$. Fix $k \geq 1$, it is enough to show ψ does not depend on $(\omega_0, \cdots, \omega_{k-1})$. By assumption, we have

$$\psi = \psi \circ \hat{F}^+ = \cdots = \psi \circ (\hat{F}^+)^k, \quad \mu^{\mathbb{N}} \times \eta\text{-a.e.}$$

Namely,

$$\psi(\omega, \hat{v}) = \psi(\sigma^k\omega, \hat{A}^k(\omega)\hat{v}).$$

Therefore, we have

$$
\begin{aligned}
\int_{X^+} \psi(\omega, \hat{v})d\mu^{\mathbb{N}}(\omega_k, \omega_{k+1}, \cdots) &= \int_{X^+} \psi(\sigma^k\omega, \hat{A}^k(\omega)\hat{v})d\mu^{\mathbb{N}}(\omega_k, \omega_{k+1}, \cdots)\\
&= \varphi(\hat{A}^k(\omega)\hat{v})\\
&= \text{const}
\end{aligned}
$$

for η-a.e. $\hat{v} \in \mathbb{P}$.

So ψ is constant in $(\omega_0, \cdots, \omega_{k-1})$, $\forall\, k \geq 1$. Thus ψ is constant in ω (one can also consider in terms of conditional expectation w.r.t. sub-algebras generated by cylinders). This proves (4).

(4) $\Rightarrow$ (1). Assume by contradiction that η is not extremal, then $\exists t \in (0,1), \eta_1 \neq \eta_2 \in \mathrm{Prob}_K(\mathbb{P})$ such that $\eta = t\eta_1 + (1-t)\eta_2$. In particular,

$$\mu^{\mathbb{N}} \times \eta = t\mu^{\mathbb{N}} \times \eta_1 + (1-t)\mu^{\mathbb{N}} \times \eta_2.$$

Since $\mu^{\mathbb{N}} \times \eta_i, i = 1, 2$ are $\hat{F}^+$-invariant, then $\mu^{\mathbb{N}} \times \eta$ is not ergodic because ergodic measures are extremal points of the space of invariant measures. This contradicts (4), thus (1) holds.

The proof is thus finished.

$\square$

As a corollary, we have

Corollary 3.5. *If $\eta \in \mathrm{Prob}_K(\mathbb{P})$ is extremal, then*

$$\frac{1}{n}\log\|A^n(\omega)v\| \to \int_{\mathbb{P}}\int_{\Sigma}\log\|A(\omega_0)\hat{v}\|\, d\mu(\omega_0)d\eta(\hat{v}),\ \textit{as } n\to\infty$$

for $\mu^{\mathbb{N}}\times\eta$-a.e. $(\omega,\hat{v})$, where $v\in\hat{v}$ with $\|v\|=1$.

3.2. **Conditional expectation.** Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space. Let $\xi\in L^1(\Omega,\mathcal{F},\mathbb{P})$ and denote $\mathbb{E}\xi=\int_\Omega \xi d\mathbb{P}$. Take $\mathcal{F}_0\subset\mathcal{F}$ a sub-σ-algebra.

$\mathbb{E}(\xi|\mathcal{F}_0)$ is the conditional expectation of ξ w.r.t. $\mathcal{F}_0$. Intuitively, it is the best prediction of ξ given the information $\mathcal{F}_0$. Formally, we have

Definition 3.7. $\mathbb{E}(\xi|\mathcal{F}_0)$ is the "unique" random variable $\tilde{\xi}:\Omega\to\mathbb{R}$ such that

(1) $\tilde{\xi}$ is $\mathcal{F}_0$-measurable.
(2) $\forall\, E\in\mathcal{F}_0$, $\int_E\tilde{\xi}d\mathbb{P}=\int_E\xi\mathbb{P}$.

The existence and uniqueness of $\mathbb{E}(\xi|\mathcal{F}_0)$ are given by Lebesgue-Radon-Nikodym. More precisely, consider the map

$$\mathcal{F}_0\ni E\mapsto\int_E\xi d\mathbb{P}\in\mathbb{R},$$

which is a signed measure. We denote it by ν. Morever, it is clear that $\nu\ll\mathbb{P}|_{\mathcal{F}_0}$ since if $\mathbb{P}(E)=0, E\in\mathcal{F}_0$ then $\nu(E)=\int_E\xi d\mathbb{P}=0$. Then by Lebesgue-Radon-Nikodym, $\exists!\tilde{\xi}\in L^1(\Omega,\mathcal{F}_0,\mathbb{P})$ s.t. $\frac{d\nu}{d\mathbb{P}|_{\mathcal{F}_0}}=\tilde{\xi}$, which gives (2).

Remark 3.3. If $Y_1,\cdots,Y_k$ are random variables on Ω, then

$$\mathbb{E}(\xi|Y_1,\cdots,Y_k):=\mathbb{E}(\xi|\sigma(Y_1,\cdots,Y_k))$$

which is the conditional expectation of ξ w.r.t. the σ-algebra generated by the random variables $Y_1,\cdots,Y_k$.

We may think of $\mathbb{E}(\xi|\mathcal{F}_0)$ as a "pixelation" of ξ where the resolution of the pixels is determined by how fine $\mathcal{F}_0$ is.

Example 1. $([0,1],\mathcal{B}[0,1],Leb)$. For $n\geq 0$, let

$$\begin{aligned}\mathcal{D}_n:=&\sigma\{\text{dyadic integrals of generation n}\}\\=&\sigma\left\{[\frac{j}{2^n},\frac{j+1}{2^n}),j=0,\cdots,2^n-1\right\}.\end{aligned}$$

If $\xi:[0,1]\to\mathbb{R}$ is Borel measurable, $\mathbb{E}(\xi|\mathcal{D}_n)$ is a function constant on the dyadic intervals of length $\frac{1}{2^n}$, where the value of the constant on such an interval J is $\frac{1}{|J|}\int_J \xi$. It is clear that $\mathcal{D}_n \subset \mathcal{D}_{n+1}$, so $\{\mathcal{D}_n\}_{n\geq 0}$ is a "filtration" of $\mathcal{B}[0,1]$ and $\sigma(\cup_{n\geq 0}\mathcal{D}_n) = \mathcal{B}[0,1]$.

Example 2. Let $(\Sigma, \mathcal{B}, \mu)$ be a metric space of symbols. Denote $X^+ = \Sigma^{\mathbb{N}}, \mathcal{F} = \sigma\{\text{cylinders}\}$ and let $\mu^{\mathbb{N}}$ be the corresponding measure on X. For $n \geq 0$, let

$$\begin{aligned}\mathcal{F}_n :=& \sigma\,\{\text{cylinders in at most } n \text{ variables}, C[A_0, \cdots, A_{n-1}], A_i \subset \Sigma\}\\ =& \sigma\,\{\text{random variables depending only on} : \omega_0, \cdots, \omega_{n-1}, \omega_i \in \Sigma\}\end{aligned}$$

Given $\xi : X^+ \to \mathbb{R}$ an L^1-function.

$$\mathbb{E}(\xi|\mathcal{F}_n) = \int_X \xi(\omega) d\mu^{\mathbb{N}}(\omega_n, \cdots).$$

It is clear that $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ and $\sigma(\cup_{n\geq 0}\mathcal{F}_n) = \mathcal{F}$. Thus $\{\mathcal{F}_n\}_{n\geq 0}$ is a filtration of $\mathcal{F}$.

In the following, we list some basic properties of the conditional expectation.

Proposition 3.6. *Let $\mathcal{F}_0 \subset \mathcal{F}$ be a sub-σ-algebra. The map $L^1(\Omega, \mathcal{F}, \mathbb{P}) \ni \xi \mapsto \mathbb{E}(\xi|\mathcal{F}_0) \in L^1(\Omega, \mathcal{F}_0, \mathbb{P})$ has the following properties:*

1. *linear: $\mathbb{E}(a\xi_1 + b\xi_2|\mathcal{F}_0) = a\mathbb{E}(\xi_1|\mathcal{F}_0) + b\mathbb{E}(\xi_2|\mathcal{F}_0)$, $\forall\, a, b \in \mathbb{R}$.*
2. *positive: $\xi \geq 0$-a.s. $\Rightarrow \mathbb{E}(\xi|\mathcal{F}_0) \geq 0$-a.s.*
3. *monotone: if $\xi_1 \leq \xi_2$-a.s. then $\mathbb{E}(\xi_1|\mathcal{F}_0) \leq \mathbb{E}(\xi_2|\mathcal{F}_0)$-a.s.*
4. *Jensen's inequality. Assume that $\varphi : \mathbb{R} \to \mathbb{R}$ is convex and $\varphi(\xi) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, then*
$$\varphi(\mathbb{E}(\xi|\mathcal{F}_0)) \leq \mathbb{E}(\varphi(\xi)|\mathcal{F}_0).$$
5. *If $\xi_n \nearrow \xi$ with $\xi \geq 0$ and $\mathbb{E}\xi < \infty$, then $\mathbb{E}(\xi_n|\mathcal{F}_0) \nearrow \mathbb{E}(\xi|\mathcal{F}_0)$.*
6. *If $\mathcal{F}_1 \subset \mathcal{F}_2$, then*
$$\mathbb{E}(\mathbb{E}(\xi|\mathcal{F}_2)|\mathcal{F}_1) = \mathbb{E}(\xi|\mathcal{F}_1)$$
and
$$\mathbb{E}(\mathbb{E}(\xi|\mathcal{F}_1)|\mathcal{F}_2) = \mathbb{E}(\xi|\mathcal{F}_1).$$

Definition 3.8. We say that a random variable ξ is independent of $\mathcal{F}_0$ if $\sigma(\xi)$ and $\mathcal{F}_0$ are independent. That is, if $E \in \sigma(\xi)$ and $F \in \mathcal{F}_0$, then $\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$.

Proposition 3.7. *We have the following two properties:*

1. *If ξ is $\mathcal{F}_0$-measurable, then $\mathbb{E}(\xi|\mathcal{F}_0) = \xi$. Moreover, if $f \in L^1$ is any other random variable, then*
$$\mathbb{E}(\xi f|\mathcal{F}_0) = \xi\mathbb{E}(f|\mathcal{F}_0).$$
2. *If ξ is independent of $\mathcal{F}_0$, then $\mathbb{E}(\xi|\mathcal{F}_0) = \mathbb{E}(\xi)$.*

Proof. We only prove (2) as (1) can be derived in the same way.

Let $\tilde{\xi} = \mathbb{E}(\xi)$. It is $\mathcal{F}_0$-measurable because it is a constant. It is enough to show that $\forall\, E \in \mathcal{F}_0$,

$$\int_E \xi d\mathbb{P} = \int_E \tilde{\xi} d\mathbb{P} = \mathbb{E}(\xi)\mathbb{P}(E).$$

Step 1. Let $\xi = \sum_{i=1}^k c_i \mathbb{1}_{E_i}$ be independent of $\mathcal{F}_0$, $\forall\, E \in \mathcal{F}_0$

$$\begin{aligned}\int_E \xi d\mathbb{P} &= \sum_{i=1}^k c_i \int_E \mathbb{1}_{E_i} d\mathbb{P} \\ &= \sum_{i=1}^k c_i \mathbb{P}(E \cap E_i) \\ &= \sum_{i=1}^k c_i \mathbb{P}(E)\mathbb{P}(E_i) \\ &= \mathbb{E}(\xi)\mathbb{P}(E).\end{aligned}$$

Step 2. Let $\xi \geq 0, \xi \in L^1$ be independent of $\mathcal{F}_0$, then by the Simple Function Approximation Theorem, $\exists\, \{\xi_n\}_{n\geq 0}$ a sequence of pointwise increasing simple functions which are also independent of $\mathcal{F}_0$ such that $\xi_n \nearrow \xi$. Moreover, $\sigma(\xi_n) \subset \sigma(\xi)$. Therefore, by Step 1 we have that $\forall\, E \in \mathcal{F}_0$,

$$\mathbb{E}(\xi_n|\mathcal{F}_0) = \mathbb{E}(\xi_n)\mathbb{P}(E)$$

Let $n \to \infty$, by item (5) of Proposition 3.6, the l.h.s. converges to $\mathbb{E}(\xi|\mathcal{F}_0)$. Moreover, by the Monotone Convergence Theorem, the r.h.s. converges to $\mathbb{E}(\xi)\mathbb{P}(E)$. Thus by the uniqueness of limit, we obtain

$$\mathbb{E}(\xi|\mathcal{F}_0) = \mathbb{E}(\xi)\mathbb{P}(E).$$

Step 3. Let $\xi \in L^1$ be independent of $\mathcal{F}_0$, we may rewrite $\xi = \xi^+ - \xi^-$ with $\xi^\pm \geq 0$ being also independent of $\mathcal{F}_0$. Moreover, $\sigma(\xi^\pm) \subset \sigma(\xi)$. By item (1) of Proposition 3.6, we have

$$\mathbb{E}(\xi|\mathcal{F}_0) = \mathbb{E}(\xi^+|\mathcal{F}_0) - \mathbb{E}(\xi^-|\mathcal{F}_0).$$

This implies $\forall\, E \in \mathcal{F}_0$,

$$\begin{aligned}\int_E \mathbb{E}(\xi|\mathcal{F}_0) d\mathbb{P} &= \int_E \mathbb{E}(\xi^+|\mathcal{F}_0) - \mathbb{E}(\xi^-|\mathcal{F}_0) d\mathbb{P} \\ &= \int_E \mathbb{E}(\xi^+|\mathcal{F}_0) d\mathbb{P} - \int_E \mathbb{E}(\xi^-|\mathcal{F}_0) d\mathbb{P} \\ &= \mathbb{E}(\xi^+)\mathbb{P}(E) - \mathbb{E}(\xi^-)\mathbb{P} \\ &= \mathbb{E}(\xi)\mathbb{P}(E).\end{aligned}$$

This finishes the proof. $\square$

Let $\mathcal{H}$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and let $\mathcal{H}_0 \subset \mathcal{H}$ be a closed subspace. Take any $v \in \mathcal{H}$, we may define the orthogonal projection of v to the subspace $\mathcal{H}_0$ by $u =: \mathrm{Proj}_{\mathcal{H}_0} v$ satisfying $u \in \mathcal{H}_0$ and $v - u \perp \mathcal{H}_0$.

In particular, $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a Hilbert space with inner product $\langle \xi, f \rangle = \mathbb{E}(\xi f)$. If $\mathcal{F}_0 \subset \mathcal{F}$ is a sub-σ-algebra, then $\mathcal{H}_0 = L^2(\Omega, \mathcal{F}_0, \mathbb{P})$.

The following proposition says that we may regard the conditional expectation as an orthogonal projection.

Proposition 3.8. *If $\xi \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, then $\mathbb{E}(\xi|\mathcal{F}_0)$ is the orthogonal projection of ξ to the subspace $L^2(\Omega, \mathcal{F}_0, \mathbb{P})$.*

Proof. We first verify that $\mathbb{E}(\xi|\mathcal{F}_0) \in L^2(\mathcal{F}_0)$.

By Jensen's equality, we have

$$|\mathbb{E}(\xi|\mathcal{F}_0)|^2 \leq |\mathbb{E}(|\xi|\,|\mathcal{F}_0)|^2 \leq \mathbb{E}(|\xi|^2\,|\mathcal{F}_0).$$

This implies

$$\int_\Omega |\mathbb{E}(\xi|\mathcal{F}_0)|^2\, d\mathbb{P} \leq \int_\Omega \mathbb{E}(|\xi|^2\,|\mathcal{F}_0) d\mathbb{P} = \int_\Omega |\xi|^2\, d\mathbb{P} = \mathbb{E}\,|\xi|^2 < \infty.$$

Thus we have $\mathbb{E}(\xi|\mathcal{F}_0) \in L^2(\mathcal{F}_0)$.

Then we are going to verify that $\xi - \mathbb{E}(\xi|\mathcal{F}_0) \perp f$, $\forall\, f \in L^2(\mathcal{F}_0)$ which is equivalent to $\langle \xi, f \rangle = \langle \mathbb{E}(\xi|\mathcal{F}_0), f \rangle$. Namely, $\mathbb{E}(\xi f) = \mathbb{E}(\mathbb{E}(\xi|\mathcal{F}_0) f)$. Since $f \in L^2(\mathcal{F}_0)$, we have

$$\mathbb{E}(\mathbb{E}(\xi|\mathcal{F}_0) f) = \mathbb{E}(\mathbb{E}(\xi f|\mathcal{F}_0)) = \mathbb{E}(\xi f).$$

This finishes the proof. $\square$

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A filtration is a sequence of σ-algebras $\{\mathcal{F}_n\}_{n \geq 0}$ with $\mathcal{F}_n \subset \mathcal{F}_{n+1}$.

Definition 3.9 (Martingale)**.** A martingale is a sequence $\{(\xi_n, \mathcal{F}_n)\}_{n \geq 0}$ such that

(1) $\mathbb{E}\,|\xi_n| < \infty, \forall\, n \geq 0$,
(2) $\{\mathcal{F}_n\}_{n \geq 0}$ is a filtration,
(3) ξ_n is $\mathcal{F}_n$-measurable,
(4) $\mathbb{E}(\xi_{n+1}|\mathcal{F}_n) = \xi_n$.

Example 1. (Standard random walk). $X_n : \Omega \to \mathbb{R}, n \geq 1$ are i.i.d. random variables with $\mathbb{E}X_1 = 0$ and $\mathbb{E}\,|X_1| < \infty$. $S_n = X_1 + \cdots + X_n$ and $\mathcal{F}_n = \sigma\{X_1, \cdots, X_n\}$. Clearly, S_n is $\mathcal{F}_n$-measurable. Then $\{(S_n, \mathcal{F}_n)\}_{n \geq 1}$ is a martingale. Note that $S_{n+1} = S_n + X_{n+1}$. Thus

$$\mathbb{E}(S_{n+1}|\mathcal{F}_n) = \mathbb{E}(S_n|\mathcal{F}_n) + \mathbb{E}(X_{n+1}|\mathcal{F}_n) = S_n + 0 = S_n.$$

Example 2. (Doob's martingale). Let $X_n : \Omega \to \mathbb{R}, n \geq 1$ be random variables. $\mathcal{F}_n = \sigma\{X_1, \cdots, X_n\}$. Assume $\xi : \Omega \to \mathbb{R}$ is a random variable with $\mathbb{E}\,|\xi| < \infty$. Let $\xi_n = \mathbb{E}(\xi|\mathcal{F}_n)$, then $\{(\xi_n, \mathcal{F}_n)\}_{n\geq 1}$ is a martingale. Note that

$$\mathbb{E}(\xi_{n+1}|\mathcal{F}_n) = \mathbb{E}(\mathbb{E}(\xi|\mathcal{F}_{n+1})|\mathcal{F}_n) = \mathbb{E}(\xi|\mathcal{F}_n) = \xi_n.$$

Theorem 3.4 (Martingale convergence theorem)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\{(\xi_n, \mathcal{F}_n)\}_{n\geq 0}$ be a martingale. Then there exists $\xi_\infty \in L^1(\omega, \mathcal{F}, \mathbb{P})$ such that*

(1) $\xi_n \to \xi_\infty$-a.s. as $n \to \infty$,
(2) $\mathbb{E}(\xi_\infty|\mathcal{F}_n) = \xi_n$-a.s. $\forall\, n \geq 0$,
(3) ξ_∞ is $\mathcal{F}_\infty$ measurable where $\mathcal{F}_\infty = \sigma\{\cup_{n\geq 0}\mathcal{F}_n\}$.

3.3. **Furstenberg formula.** We begin with an abstract result.

Theorem 3.5 (Furstenberg-Kifer)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let M be a compact metric space and let $K : M \to \mathrm{Prob}(M)$ be an SDS. Given a K-Markov chain $\{Z_n : \Omega \to M\}_{n\geq 0}$, for any $f \in C(M)$, with probability one the following hold*

(1) $\limsup_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(Z_j) \leq \sup\left\{\int_M f d\eta : \eta \in \mathrm{Prob}_K(M)\right\}$.
(2) If $\mathrm{Prob}_K(M) \ni \eta \mapsto \int_M f d\eta$ is constant equal to β, then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(Z_j) = \beta.$$

In fact, by compactness of $\mathrm{Prob}_K(M)$ and continuity of f, we may replace "sup" in item (1) by "max".

Proof. For (1), we first consider the first case when $f = Qg - g$ for some $g \in C(M)$. In this case, we prove the following lemma.

Lemma 3.9. *If $f = Qg - g$, $g \in C(M)$, then*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(Z_j(\omega)) = 0, \quad \mathbb{P}\text{-}a.s.$$

Proof. Consider the random variables $W_n : \Omega \to M, n \geq 1$,

$$W_n := \sum_{j=1}^{n} \frac{Qg(Z_{j-1}) - g(Z_j)}{j}.$$

Then W_n depends on $Z_0, \cdots, Z_n$. Let $\mathcal{F}_n = \sigma\{Z_0, \cdots, Z_n\}$. We claim that $\{(W_n, \mathcal{F}_n)\}_{n\geq 1}$ is a Martingale. By definition,

$$W_{n+1} = W_n + \frac{1}{n+1}(Qg(Z_n) - g(Z_{n+1})).$$

This implies

$$\mathbb{E}(W_{n+1}|\mathcal{F}_n) = W_n + \frac{1}{n+1}\mathbb{E}(Qg(Z_n)|\mathcal{F}_n) - \frac{1}{n+1}\mathbb{E}(g(Z_{n+1})|\mathcal{F}_n).$$

It is clear that $\mathbb{E}(Qg(Z_n)|\mathcal{F}_n) = Qg(Z_n)$. On the other hand, by Markov property we have

$$\mathbb{E}(g(Z_{n+1})|\mathcal{F}_n) = \mathbb{E}(g(Z_{n+1})|Z_0, \cdots, Z_n) = \mathbb{E}(g(Z_{n+1})|Z_n).$$

Moreover, by the definition of the K-Markov chain,

$$\mathbb{P}(Z_{n+1} \in E|Z_n = x) = K_x(E).$$

Thus

$$\mathbb{E}(g(Z_{n+1})|Z_n = x) = \int_M g dK_x = Qg(Z_n).$$

Therefore,

$$\mathbb{E}(W_{n+1}|\mathcal{F}_n) = W_n.$$

There other properties of being a martingale are straightforward.Thus we prove that $\{(W_n, \mathcal{F}_n)\}_{n\geq 1}$ is a martingale. By Martingale convergence theorem, $W_n \to W_\infty < \infty$ almost surely.

Recall that Kronecker's lemma says if $\sum_{n=1}^{\infty} a_n < \infty$, then $\frac{1}{n}\sum_{j=1}^{n} ja_j \to 0$ as $n \to \infty$. Then by this lemma, we have that when $n \to \infty$,

$$\frac{1}{n}\sum_{j=1}^{n} j \cdot \frac{Qg(Z_{j-1}) - g(Z_j)}{j} = \frac{1}{n}\sum_{j=1}^{n}[Qg(Z_{j-1}) - g(Z_j)] \to 0,$$

namely,

$$\frac{1}{n}\sum_{j=1}^{n}[f(Z_{j-1}) + g(Z_{j-1}) - g(Z_j)] \to 0.$$

Note that for g this is a telescoping sum. Since g is bounded, when divided by n the second and third terms in the sum disappear as $n \to \infty$, which gives

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(Z_j) = 0.$$

Note that all the statements are in almost sure sense as W_∞ is. This proves the lemma. □

As M is compact, the space of continuous functions on M denoted by $C(M)$ is separable. Then $\exists g_1, \cdots, g_k, \cdots$ which are dense in $C(M)$. Apply the previous lemma to $f_k := Qg_k - g_k$, so $\exists \Omega_k \subset \Omega, \mathbb{P}(\Omega_k) = 1$ s.t. $\forall\, \omega \in \Omega_k$,

$$\frac{1}{n}\sum_{j=1}^{n}[Qg_k(Z_j(\omega)) - g_k(Z_j(\omega))] \to 0, \quad \text{as} \quad n \to \infty.$$

Let $\Omega_* = \cap_{k\geq 1}\Omega_k$, then $\mathbb{P}(\Omega_*) = 1$. Fix an arbitrary $\omega \in \Omega_*$, then

$$\frac{1}{n}\sum_{j=1}^{n}[Qg_k(Z_j(\omega)) - g_k(Z_j(\omega))] \to 0, \quad \text{as} \quad n \to \infty.$$

For any $n \geq 1$, consider the measure on M

$$\eta_n := \frac{1}{n}\sum_{j=0}^{n-1}\delta_{Z_j(\omega)} \in \mathrm{Prob}(M).$$

Then the previous statement is equivalent to

$$\int_M [Qg_k - g_k]d\eta_n \to 0, \quad \text{as} \quad n \to \infty.$$

Since $\{g_k\}_{k\geq 1}$ is dense in $C(M)$, $\forall\, g \in C(M)$,

$$\int_M [Qg - g]d\eta_n \to 0, \quad \text{as} \quad n \to \infty.$$

Let η_* be any weak* limit of $\{\eta_n\}_{n\geq 1}$. Then since $g, Qg \in C(M)$,

$$\int_M (Qg - g)d\eta_* = 0.$$

Namely,

$$\int_M Qg d\eta_* = \int_M g\eta_*, \quad \forall\, g \in C(M),$$

which shows that $\eta_* \in \mathrm{Prob}_K(M)$. Note that f is bounded, by the definition and existence of limsup, there exists a sequence $\{n_k\}_{k\geq 1}$ such that

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}f(Z_j(\omega)) = \lim_{k\to\infty}\frac{1}{n_k}\sum_{j=0}^{n_k-1}f(Z_j(\omega)) = \lim_{k\to\infty}\int_M f d\eta_{n_k} < \infty.$$

Besides, since M is compact, then $\mathrm{Prob}(M)$ is weak* compact. So we can choose a subsequence $\{n_{k_i}\}$ such that

$$\int_M f d\eta_{n_{k_i}} \to \int_M f\eta_0, \quad \text{as} \quad i \to \infty,$$

where $\eta_0 \in \mathrm{Prob}_K(M)$ by the previous argument. This proves

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}f(Z_j) \leq \sup\left\{\int_M f d\eta : \eta \in \mathrm{Prob}_K(M)\right\}$$

with probability one. Moreover, we may replace "sup" by "max" because of the compactness of $\mathrm{Prob}_K(M)$ and the continuity of f.

For (2), by assumption we have, with probability one,

$$\int_M -f d\eta = -\int_M f d\eta = -\beta, \, \forall \, \eta \in \mathrm{Prob}_K(M).$$

Apply (1) to $-f$, we have

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} -f(Z_j) \le \beta.$$

Equivalently,

$$-\liminf_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(Z_j) \le -\beta,$$

and thus

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(Z_j) \ge \beta.$$

Combining (1), we have with probability one,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(Z_j) = \beta.$$

□

We will apply this theorem to the DDS: projective linear cocycle as described before. Recall that we have proved $\eta \in \mathrm{Prob}_K(\mathbb{P}) \Leftrightarrow \mu \times \eta \in \mathrm{Prob}_{\bar{K}}(\Sigma \times \mathbb{P})$. In fact, we shall see that if $m \in \mathrm{Prob}_{\bar{K}}(\Sigma \times \mathbb{P})$, then $\exists \eta \in \mathrm{Prob}_K(\mathbb{P})$ s.t. $m = \mu \times \eta$. Let us acknowledge this for now and later we will prove it as a lemma.

Define $\alpha : \mathrm{Prob}_K(\mathbb{P}) \to \mathbb{R}$ as

$$\alpha(\eta) := \int_{\Sigma\times\mathbb{P}} \xi(\omega_0, \hat{v}) d\mu(\omega_0) d\eta(\hat{v}),$$

where $\xi(\omega_0, \hat{v}) = \log \|A(\omega_0)v\|$ with $v \in \hat{v}$ a unit representative. It is clear that α is a continuous linear functional. We define

$$\beta := \max \{\alpha(\eta) : \eta \in \mathrm{Prob}_K(\mathbb{P})\}.$$

The maximum is attained again because $\mathrm{Prob}_K(\mathbb{P})$ is compact.

Theorem 3.6 (Furstenberg-Kifer). *$\forall \, v \in \mathbb{R}^2$ non-zero,*

(1) We have

$$\limsup_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)v\| \le \beta$$

for $\mu^{\mathbb{N}}$-a.e. $\omega \in X^+$.

(2) If $\alpha(\eta) = \beta$, $\forall\, \eta \in \mathrm{Prob}_K(\mathbb{P})$, *then*

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)v\| = \beta$$

for $\mu^{\mathbb{N}}$*-a.e.* $\omega \in X^+$.

Proof. Note that

$$\max\left\{\int \xi dm : m \in \mathrm{Prob}_{\bar{K}}(\Sigma \times \mathbb{P})\right\} \tag{3.4}$$

$$= \max\left\{\int \xi d\mu d\eta : \eta \in \mathrm{Prob}_K(\mathbb{P})\right\} = \beta. \tag{3.5}$$

Consider the $\bar{K}$-Markov chain, $Z_n : X^+ \times \mathbb{P} \to \Sigma \times \mathbb{P}$ with $Z_n(\omega, \hat{v}) = (\omega_n, \hat{A}^n(\omega)\hat{v})$. Recall that we have by direct computation

$$\frac{1}{n} \sum_{j=0}^{n-1} \xi(Z_n(\omega, \hat{v})) = \frac{1}{n} \log \|A^n(\omega)v\|$$

for $\mu^{\mathbb{N}} \times \delta_{\hat{v}}$-a.e. $(\omega, \hat{v})$. Thus by Theorem 3.5, it remains to prove the following:

Lemma 3.10. *If* $m \in \mathrm{Prob}_{\bar{K}}(\Sigma \times \mathbb{P})$ *then* $\exists\, \eta \in \mathrm{Prob}_K(\mathbb{P})$ *such that* $m = \mu \times \eta$.

Proof. To define a measure η, it is enough to define its corresponding integral.

For any $\psi \in C(\mathbb{P})$, let

$$I(\psi) := \int_{\Sigma\times\mathbb{P}} \varphi dm$$

where $\pi\varphi = \psi$. Here $\pi : C(\Sigma \times \mathbb{P}) \to C(\mathbb{P})$ is defined by

$$\pi\varphi(\hat{p}) = \int_{\Sigma} \varphi(\omega_0, \hat{p}) d\mu(\omega_0).$$

For I to make sense, we should have that if $\pi\varphi_1 = \pi\varphi_2$, then $\int_{\Sigma\times\mathbb{P}} \varphi_1 dm = \int_{\Sigma\times\mathbb{P}} \varphi_2 dm$. Note that

$$\bar{Q}\varphi(\omega_0, \hat{v}) = \int \varphi(\omega_1, \hat{A}(\omega_0)\hat{v}) d\mu(\omega_1) = \pi\varphi(\hat{A}(\omega_0)\hat{v}).$$

Then if $\pi\varphi_1 = \pi\varphi_2$, then $\bar{Q}\varphi_1 = \bar{Q}\varphi_2$, which shows

$$\int_{\Sigma\times\mathbb{P}} \bar{Q}\varphi_1 dm = \int_{\Sigma\times\mathbb{P}} \bar{Q}\varphi_2 dm.$$

Since m is $\bar{K}$-stationary, we have

$$\int_{\Sigma\times\mathbb{P}} \varphi_1 dm = \int_{\Sigma\times\mathbb{P}} \varphi_2 dm.$$

Therefore, I is well defined positive linear functional and $I(1) = 1$. By Riesz-Markov-Kakutani representation theorem, there exists a unique Radon measure $\eta \in \mathrm{Prob}(\mathbb{P})$ such that

$$I(\psi) = \int \psi d\eta.$$

Thus we have $\eta \in \mathrm{Prob}(\mathbb{P})$ such that

$$\int \pi\varphi d\eta = \int \varphi dm, \, \forall \, \varphi \in C(\Sigma \times \mathbb{P}).$$

Namely,

$$\int_{\Sigma\times\mathbb{P}} \varphi d\mu d\eta = \int_{\Sigma\times\mathbb{P}} \varphi dm, \, \forall \, \varphi \in C(\Sigma \times \mathbb{P}).$$

This shows $m = \mu \times \eta$. □

Thus the whole proof is finished. □

Next we are going to prove the Furstenberg's formula which is particularly useful in proving modulus of continuity of the first Lyapunov exponent.

Theorem 3.7 (Furstenberg's formula)**.** *Given a probability space* (Σ, μ) *and given a random linear cocycle* A*, its maximal Lyapunov exponent* $L^+(A)$ *satisfies the following equation:*

$$L^+(A) = \max\left\{\int_{\Sigma\times\mathbb{P}} \log\|A(\omega_0)v\|\, d\mu(\omega_0)d\eta(\hat{v}) : \eta \in \mathrm{Prob}_K(\mathbb{P})\right\}.$$

where $v \in \hat{v}$ *is a unit representative.*

Proof. For $g \in \mathrm{GL}_2(\mathbb{R})$, we can alternatively define its norm

$$\|g\|' := \max\left\{\|ge_1\|, \|ge_2\| : \{e_1, e_2\} \text{ is a basis of } \mathbb{R}^2\right\}.$$

Note that all the norms in finite dimension are equivalent.

Let $\alpha : \mathrm{Prob}_K(\mathbb{P}) \to \mathbb{R}$ be the continuous linear functional

$$\alpha(\eta) = \int_{\Sigma\times\mathbb{P}} \xi(\omega_0, \hat{v}) d\mu(\omega_0) d\eta(\hat{v}).$$

Then $\max\{\alpha(\eta) : \eta \in \mathrm{Prob}_K(\mathbb{P})\} =: \beta$ is attained since $\mathrm{Prob}_K(\mathbb{P})$ is weak* compact. Then it is enough to prove $L^+(A) = \beta$.

Let $\mathcal{M} := \{\eta \in \mathrm{Prob}_K(\mathbb{P}) : \alpha(\eta) = \beta\}$. Then $\mathcal{M}$ is non-empty, convex and closed (hence compact). By Krein-Milman, $\mathcal{M}$ has at least one extreme point. Moreover, the closed convex hull of *extreme*$(\mathcal{M})$ is $\mathcal{M}$.

Let η_0 be such an extremal point of $\mathcal{M}$, then it is easy to see that η_0 is also an extremal point in $\mathrm{Prob}_K(\mathbb{P})$ (one can prove it easily by contradiction that all the admissible extremal points of $\mathcal{M}$ must belong to the extremal points of $\mathrm{Prob}_K(\mathbb{P})$). Then by Proposition 3.4, $\mu^{\mathbb{N}} \times \eta_0$ is $\hat{F}^+$-ergodic.

Then by Birkhoff ergodic theorem, we have for $\mu^{\mathbb{N}} \times \eta_0$-a.e. $(\omega, \hat{v})$,

$$\begin{aligned}
\beta = \alpha(\eta_0) &= \int_{\Sigma\times\mathbb{P}} \xi(\omega_0, \hat{v}) d\mu(\omega_0) d\eta_0(\hat{v}) \\
&= \int_{X^+\times\mathbb{P}} \Phi(\omega, \hat{v}) d\mu^{\mathbb{N}}(\omega) d\eta_0(\hat{v}) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \Phi \circ (\hat{F}^+)^j(\omega, \hat{v}) \\
&= \lim_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)v\| \\
&\leq \lim_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)\| \\
&= L^+(A) \\
&= \lim_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)\|' \\
&\leq \max_{1,2} \limsup_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)e_i\| \\
&\leq \beta.
\end{aligned}$$

Here the last inequality is due to Theorem 3.6. So $L^+(A) = \beta$.

This finishes the proof. □

3.4. **Furstenberg-Kifer non-random filtration.** In order to make a better comparison, we first recall the Oseledets multiplicative ergodic theorem (it is called "random" because the subspace depends on the base point).

Theorem 3.8 (Oseledets). *Let $F = F_A : \Omega \times \mathbb{R}^2 \to \Omega \times \mathbb{R}^2, F(\omega, v) = (f(\omega), A(\omega)v)$ be a μ-integrable cocycle given by $A : X \to \mathrm{GL}_2(\mathbb{R})$ over an ergodic MPDS (Ω, f, ν), then*

(1) If $L^+(A) = L^-(A)$, then $\forall\, v \in \mathbb{R}^2$ non-zero,

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)v\| = L^+(A), \quad \nu\text{-a.e.}\, \omega \in \Omega.$$

(2) If $L^+(A) > L^-(A)$, then there is a measurable map

$$\omega \mapsto V_\omega \subset \mathbb{R}^2$$

where V_ω is a one dimensional subspace of $\mathbb{R}^2$, such that

$$A(\omega)V_\omega = V_{f(\omega)}$$

i.e. V_ω is an F- invariant section. Moreover, if $v \notin V_\omega$, then

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)v\| = L^-(A).$$

Otherwise, if $v \in V_\omega$, then

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)v\| = L^+(A)$$

Moreover, if f is invertible then there exists a measurable splitting of the fiber: for ν-almost every $\omega \in \Omega$, $\mathbb{R}^2 = E^+_\omega \oplus E^-_\omega$ such that

(1) $A(\omega)E^\pm_\omega = E^\pm_{f(\omega)}$.
(2) $\lim_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)v\| = L^\pm(A)$, $v \in E^\pm_\omega, v \neq 0$.
(3) $\lim_{n\to\infty} \frac{1}{n} \log \left|\sin \angle(E^+_{f^n(\omega)}, E^-_{f^n(\omega)})\right| = 0$.

Note that given any $v \in \mathbb{R}^2\backslash\{0\}$,

$$\frac{1}{n} \log \|A^n(\omega)v\| \to L^-(A) \text{ or } L^+(A),\ \nu\text{-a.e.}\, \omega \in \Omega.$$

But it could be that for some ω's, the convergence is to $L^-(A)$ and for other ω's to L^+. Namely, given v, where the limit goes depend on the base point $\omega \in \Omega$. This holds for any cocycle over any ergodic base dynamics. However, for random linear cocycles, we will show that the filtration is non-random: $\exists\, V \subsetneq \mathbb{R}^2$ a linear subspace, such that

(1) $A(\omega)V = V$ ν-a.e. $\omega \in \Omega$,
(2) if $v \in V\backslash\{0\}$, $\frac{1}{n} \log \|A^n(\omega)v\| \to L^-(A)$, ν-a.e. $\omega \in \Omega$,
(3) if $v \notin V$, $\frac{1}{n} \log \|A^n(\omega)v\| \to L^+(A)$, ν-a.e. $\omega \in \Omega$.

This ensures that the limit is independent of the base point. Moreover, if A is quasi-irreducible, then $V = \{0\}$, so $\forall\, v \in \mathbb{R}^2\backslash\{0\}$,

$$\frac{1}{n} \log \|A^n(\omega)v\| \to L^+(A),\ \nu\text{-a.e. } \omega \in \Omega.$$

In particular. by Lebesgue's dominated convergence theorem, $\forall\, v \in \mathbb{R}^2\backslash\{0\}$,

$$\mathbb{E}(\frac{1}{n} \log \|A^n(\omega)v\|) \to L^+(A).$$

Furthermore, the convergence is indeed uniform in $v \in \mathbb{S}^1$. This is the main ingredient in proving the strong mixing of the Markov operator.

Before we introduce the Furstenberg-Kifer non-random filtration, let us make some preparation.

By Furstenberg's formula, we know that

$$L^+(A) = \max\left\{\int_{\Sigma\times\mathbb{P}} \log\|A(\omega_0)v\|\, d\mu(\omega_0)d\eta(\hat{v}) : \eta \in \mathrm{Prob}_K(\mathbb{P})\right\} = \beta$$

where we denote $\alpha(\eta) := \int_{\Sigma\times\mathbb{P}} \log\|A(\omega_0)v\|\, d\mu(\omega_0)d\eta(\hat{v})$. Let

$$\mathcal{E} := \{\alpha(\eta) : \eta \text{ is an extreme point of } \mathrm{Prob}_K(\mathbb{P})\}.$$

Thus we have $\max \mathcal{E} = \beta$ (one can prove it by contradiction easily).

Lemma 3.11. *We have $L^+(A) \in \mathcal{E} \subset \{L^+(A), L^-(A)\}$. In other words, $\max\mathcal{E} = L^+(A)$ and if there are other elements in $\mathcal{E}$, they are just $L^-(A)$.*

Proof. If η is an extreme point in $\mathrm{Prob}_K(\mathbb{P})$, then $\mu^{\mathbb{N}}\times\eta$ is $\hat{F}^+$-ergodic. So by Birkhoff ergodic theorem, for $\mu^{\mathbb{N}}\times\eta$-a.e. $(\omega,\hat{v})$, we have

$$\begin{aligned}\lim_{n\to\infty}\sum_{j=0}^{n-1}\Psi\circ(\hat{F}^+)^j(\omega,\hat{v}) &= \int_{X^+\times\mathbb{P}}\Psi(\omega,\hat{v})d\mu^{\mathbb{N}}(\omega)d\eta(\hat{v})\\ &= \int_{\Sigma\times\mathbb{P}}\xi(\omega_0,\hat{v})d\mu(\omega_0)d\eta(\hat{v}) = \alpha(\eta).\end{aligned}$$

Note that the l.h.s. equals $\lim_{n\to\infty}\frac{1}{n}\log\|A^n(\omega)v\|, v\in\hat{v}, \|v\| = 1$ which is either $L^+(A)$ or $L^-(A)$ (here it is a bit subtle in the sense that we already know the limit exists by Birkhoff for $\mu^{\mathbb{N}}\times\eta$-a.e. $(\omega,\hat{v})$, and at the same time, by Oseledets we know $\forall\, v\in\mathbb{R}^2\backslash\{0\}$, depending on the base point $\omega\in X^+$ which belongs to a full measure set, the limit is either $L^+(A)$ or $L^-(A)$. Therefore, combining these two conditions we obtain that the limit is either $L^+(A)$ or $L^-(A)$ for $\mu^{\mathbb{N}}\times\eta$-a.e. $(\omega,\hat{v})$). Thus $\mathcal{E}\subset\{L^+(A), L^-(A)\}$. Since we already have $\max\mathcal{E} = L^+(A)$, the lemma follows. ∎

Now we can formulate the main theorem in this subsection.

Theorem 3.9 (Furstenberg-Kifer non-random filtration)**.** *There is a linear subspace $V \subsetneq \mathbb{R}^2$ such that*

(1) V is A-invariant, $A(\omega_0)V = V$ for μ-a.e. $\omega_0\in\Sigma$.

(2) If η is an extreme point in $\mathrm{Prob}_K(\mathbb{P})$ and $\alpha(\eta) = L^-(A)$, then $\eta(\hat{v}) = 1$ where $v\in V$.

(3) If $v\in V\backslash\{0\}$ then

$$\lim_{n\to\infty}\frac{1}{n}\log\|A^n(\omega)v\| = L^-(A),\ \mu^{\mathbb{N}}\text{-a.e.}$$

(4) If $v\notin V$, then

$$\lim_{n\to\infty}\frac{1}{n}\log\|A^n(\omega)v\| = L^+(A),\ \mu^{\mathbb{N}}\text{-a.e.}$$

Proof. **Case 1.** $\#\mathcal{E} = 1$, i.e. $\mathcal{E} = \{L^+(A)\}$. We will show that in this case $V = \{0\}$.

By assumption, we have $\alpha(\eta) = \beta$ for any η being an extreme point of $\mathrm{Prob}_K(\mathbb{P})$. Then necessarily, $\alpha(\eta) \equiv \beta, \forall\, \eta \in \mathrm{Prob}_K(\mathbb{P})$.

Indeed, let again $\mathcal{M} := \{\eta \in \mathrm{Prob}_K(\mathbb{P}) : \alpha(\eta) = \beta\}$ which is non-empty, convex and compact. So by Krein-Milman,

$$\mathcal{M} = \overline{Co}(\mathcal{M}) \supset \overline{Co}(\mathrm{extreme}(\mathrm{Prob}_K(\mathbb{P}))) = \mathrm{Prob}_K(\mathbb{P}).$$

Thus $\mathcal{M} = \mathrm{Prob}_K(\mathbb{P})$. This shows $\alpha(\eta) = \beta$, $\forall\, \eta \in \mathrm{Prob}_K(\mathbb{P})$. Since α is constant, by Theorem 3.6 we have $\forall\, v \in \mathbb{R}^2 \backslash \{0\}$,

$$\frac{1}{n} \log \|A^n(\omega)v\| \to \beta = L^+(A), \text{ as } n \to \infty$$

for $\mu^{\mathbb{N}}$-a.e. $\omega \in X^+$. Therefore if we put $V = \{0\}$, then the theorem holds.

Case 2. $\#\mathcal{E} = 2$, i.e. $\mathcal{E} = \{L^+(A), L^-(A)\}$ and $L^+(A) > L^-(A)$. Let

$$V := \left\{ v \in \mathbb{R}^2 : \limsup_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)v\| \le L^-(A),\, \mu^{\mathbb{N}}\text{-a.e.}\, \omega \in X^+ \right\}.$$

We are going to prove V satisfies (1)-(4). We do several steps.

(1) V is a linear subspace. Let $v_1, v_2 \in V$, $a, b \in \mathbb{R}$. Then

$$\begin{aligned}\|A^n(\omega)(av_1 + bv_2)\| &\le |a|\, \|A^n(\omega)v\| + |b|\, \|A^n(\omega)v_2\| \\ &\le \max \{|a|\, \|A^n(\omega)v\|, |b|\, \|A^n(\omega)v_2\|\}.\end{aligned}$$

Take $1/n \log$ on both sides and let $n \to \infty$ (taking $\limsup$), we

$$\begin{aligned}&\limsup_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)(av_1 + bv_2)\| \\ \le& \max \left\{ \limsup_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)v_1\|, \limsup_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)v_2\| \right\} \\ &\le L^-(A),\end{aligned}$$

for $\mu^{\mathbb{N}}$-a.e. $\omega \in X^+$. Thus $av_1 + bv_2 \in V$ which shows V is a linear subspace.

(2) If $\eta_- \in \mathrm{Prob}_K(\mathbb{P})$ which is extreme such that $\alpha(\eta_-) = L^-(A)$ (in case 2 there are such measures), then we have $\eta_-(\hat{v}) = 1, v \in V$. Indeed,

$$L^-(A) = \alpha(\eta_-) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \Psi \circ (\hat{F}^+)^j(\omega, \hat{v}) = \lim_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)v\|$$

holds for $\mu^{\mathbb{N}} \times \eta_-$-a.e. $(\omega, \hat{v})$. By Fubini's theorem, for η_--a.e. $\hat{v} \in \mathbb{P}$ we have

$$\lim_{n\to\infty} \log \|A^n(\omega)v\| = L^-(A), \ \mu^{\mathbb{N}}\text{-a.e. } \omega \in X^+.$$

This shows for such $\hat{v}$'s, $v \in V$. Thus $\eta_-(\hat{v}) = 1$ where $v \in V$. In particular $V \neq \{0\}$, otherwise $\hat{v} = \emptyset \Rightarrow \eta(\hat{v}) = 0 \neq 1$.

(3) V is a proper subspace. We already have $V \neq \{0\}$, so it is enough to show that $V \neq \mathbb{R}^2$.

$$\exists \eta_+, \text{ s.t. } \alpha(\eta_+) = L^+(A) > L^-(A).$$

and

$$\alpha(\eta_+) = \lim_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)v\|, \ \mu^{\mathbb{N}} \times \eta_+\text{-a.e. } (\omega, \hat{v}).$$

Such v's are not in V, which shows $V \neq \mathbb{R}^2$.

(4) V is A-invariant. Let η_- be an extreme point in $\mathrm{Prob}_K(\mathbb{P})$ s.t. $\alpha(\eta_-) = L^-(A)$, then we know $\eta_-(\hat{v}) = 1$ with $v \in V$. Since η_- is K-stationary, we have $\forall\, \varphi \in L^\infty(\mathbb{P})$

$$\int \varphi d\eta_- = \int Q\varphi d\eta_- = \int \varphi(\hat{A}(\omega_0)\hat{v}))d\mu(\omega_0)d\eta_-(\hat{v}).$$

Take $\varphi = \mathbb{1}_{\hat{v}}$, then

$$\begin{aligned} 1 = \eta_-(\hat{v}) = \int \mathbb{1}_{\hat{v}} d\eta_- &= \int \mathbb{1}_{\hat{v}}(\hat{A}(\omega_0)\hat{v})d\eta_-(\hat{v})d\mu(\omega_0) \\ &= \int \mathbb{1}_{\widehat{A(\omega_0)^{-1}v}}(\hat{v})d\eta_-(\hat{v})d\mu(\omega_0) \\ &= \int \eta_-(\widehat{A(\omega_0)^{-1}v})d\mu(\omega_0). \end{aligned}$$

This shows $\eta_-(\widehat{A(\omega_0)^{-1}v}) = 1$ for $\mu^{\mathbb{N}}$-a.e. $\omega_0 \in \Sigma$. Therefore $\forall\, v \in V$, $A(\omega_0)v = v$ for $\mu^{\mathbb{N}}$-a.e. $\omega_0 \in \Sigma$. Namely, V is A-invariant.

(5) If $v \in V\backslash\{0\}$, then $\hat{V} = \{\hat{v}\}$, $\eta_-(\hat{V}) = \eta_-(\hat{v}) = 1$ where η_- is extreme such that $\alpha(\eta_-) = L^-(A)$. Moreover,

$$\alpha(\eta_-) = L^-(A) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \Psi \circ (\hat{F}^+)^j(\omega, \hat{v}) = \lim_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)v\|$$

for $\mu^{\mathbb{N}} \times \eta_-$-a.e. $(\omega, \hat{v})$. This implies

$$L^-(A) = \lim_{n\to\infty} \frac{1}{n} \log \|A^n(\omega)v\|$$

for $\mu^{\mathbb{N}}$-a.e. $\omega \in X^+$.

(6) Let $v \notin V$, V is a one-dimensional linear subspace which is A-invariant. By a change of variables, we can assume that $v = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Then

$$A(\omega) = \begin{pmatrix} b(\omega) & c(\omega) \\ 0 & d(\omega) \end{pmatrix}$$

and

$$A^n(\omega) = \begin{pmatrix} b_n(\omega) & c_n(\omega) \\ 0 & d_n(\omega) \end{pmatrix}$$

It is easy to see that for $\mu^{\mathbb{N}}$-a.e. $\omega \in X^+$

$$L^+(A) = \max\left\{ \frac{1}{n} \log |b_n(\omega)|, \frac{1}{n} \log |d_n(\omega)| \right\}.$$

Moreover,

$$\left\| A^n(\omega) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\| = \left\| \begin{pmatrix} b_n(\omega) \\ 0 \end{pmatrix} \right\| = |b_n(\omega)|.$$

So we have

$$\frac{1}{n} \log |b_n(\omega)| = \frac{1}{n} \log \left\| A^n(\omega) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\| \to L^-(A), \text{ as } n \to \infty.$$

Thus

$$\frac{1}{n} \log |d_n(\omega)| \to L^+(A), \text{ as } n \to \infty.$$

almost surely.

Now take any $v \notin V$, then

$$v = \begin{pmatrix} t \\ 1 \end{pmatrix}$$

for some $t \in \mathbb{R}$. Therefore,

$$A^n(\omega)v = \begin{pmatrix} tb_n(\omega) + c_n(\omega) \\ d_n(\omega) \end{pmatrix}$$

which implies that

$$\|A^n(\omega)v\| \geq |d_n(\omega)|.$$

Thus

$$\frac{1}{n} \log \|A^n(\omega)v\| \geq \frac{1}{n} \log |d_n(\omega)| \to L^+(A).$$

Combining with the Furstenberg-Kifer theorem, we have

$$\frac{1}{n} \log \|A^n(\omega)v\| \to L^+(A)$$

for $\mu^{\mathbb{N}}$-a.e. $\omega \in X^+$.

This finishes the whole proof. □

3.5. **Uniform convergence of the directional Lyapunov exponent.** Recall that the definitions of irreducibility and quasi-irreducibility is defined as follows:

Definition 3.10. A linear cocycle A is irreducible if there is no invariant proper subspace (which is a line). Namely, $\nexists\, l \subset \mathbb{R}^2$ s.t. $A(\omega_0)l = l$, μ-a.e. $\omega_0 \in \Sigma$. A is quasi-irreducible if $\nexists\, l \subset \mathbb{R}^2$ such that l is A-invariant and $L(A|_l) < L^+(A)$.

Remark 3.4. A is quasi-irreducible if and only if the Furstenberg-Kifer non-random filtration is trivial: $V = \{0\}$, i.e. $\forall\, v \in \mathbb{R}^2\backslash\{0\}$,

$$\frac{1}{n}\log\|A^n(\omega)v\| \to L^+(A),\ \mu^{\mathbb{N}}\text{-a.e.}\,\omega \in X^+,\ \text{ as } n \to \infty.$$

Moreover, it is also equivalent to saying that $\alpha(\eta) \equiv \beta$, $\forall\, \eta \in \mathrm{Prob}_K(\mathbb{P})$.

Theorem 3.10. *Assume that A is quasi-irreducible and $L^+(A) > L^-(A)$, then*

$$\mathbb{E}(\frac{1}{n}\log\|A^n(\omega)v\|) \to L^+(A),\ \ \textit{as } n \to \infty$$

uniformly in $v \in \mathbb{S}^1$.

Proof. Since A is irreducible, by the previous remark and Lebesgue dominated convergence theorem, we have the pointwise convergence:

$$\mathbb{E}(\frac{1}{n}\log\|A^n(\omega)v\|) \to L^+(A),\ \forall\, v \in \mathbb{S}^1,\ \text{ as } n \to \infty.$$

Assume by contradiction that the convergence is not uniform, then $\exists\, \delta > 0$ and a sequence of $\{v_{n_k}\}_{k\geq 1} \subset \mathbb{S}^1$ such that

$$\left|\mathbb{E}(\frac{1}{n_k}\log\|A^{n_k}(\omega)v_{n_k}\|) - L^+(A)\right| \geq \delta,\ \forall\, k \geq 1.$$

To simplify the notation, we just write n standing for n_k but we should bear in mind that from now on $\{n\}$ is actually a subsequence. Moreover, we may assume that $v_n \to v_0 \in \mathbb{S}^1$ by compactness of the circle.

Note that for $n \geq N$ with N large enough, it can not happen that

$$\mathbb{E}(\frac{1}{n}\log\|A^n(\omega)v_n\|) \geq L^+(A) + \delta$$

because

$$\mathbb{E}(\frac{1}{n}\log\|A^n(\omega)v_n\|) \leq \mathbb{E}(\frac{1}{n}\log\|A^n(\omega)\|) < L^+(A) + \frac{\delta}{2}.$$

Thus we only need to consider the case when

$$\mathbb{E}(\frac{1}{n}\log\|A^n(\omega)v_n\|) \leq L^+(A) - \delta.$$

We are going to prove it actually can not happen either. To achieve that, we give a claim first and prove it later.

We claim that

$$\liminf_{n\to\infty} \frac{\|A^n(\omega)v_n\|}{\|A^n(\omega)\|} = c(\omega) > 0,\ \mu^{\mathbb{N}}\text{-a.e.}\,\omega \in X^+.$$

Accepting it for now, we get

$$\frac{1}{n}\log\frac{\|A^n(\omega)v_n\|}{\|A^n(\omega)\|} \to 0$$

almost surely as $n \to \infty$. By Lebesgue dominated convergence theorem,

$$\mathbb{E}(\frac{1}{n}\log\frac{\|A^n(\omega)v_n\|}{\|A^n(\omega)\|}) \to 0, \qquad \text{as } n \to \infty.$$

However, the l.h.s. is equal to

$$\mathbb{E}(\frac{1}{n}\log\|A^n(\omega)v_n\|) - \mathbb{E}(\frac{1}{n}\log\|A^n(\omega)\|) \le -\frac{\delta}{2}, \qquad \text{as } n \to \infty.$$

This is a contradiction. So we prove the uniform convergence in $v \in \mathbb{S}^1$.

Before giving the proof of the claim, we recall the concept of singular values and singular directions. These are some ingredients in the proof of Oseledets.

Let $g \in \mathrm{GL}_2(\mathbb{R})$, the singular values of $g : s_+(g) \ge s_-(g) \ge 0$ are the eigenvalues of $(g^*g)^{\frac{1}{2}}$ It turns out that

$$s_+(g) = \max_{v\in\mathbb{S}^1}\|gv\| = \|g\|$$

is the maximum expansion of g.

$$s_-(g) = \min_{v\in\mathbb{S}^1}\|gv\| = \|g\|$$

is the minimum expansion of g.

If $s_+(g) > s_-(g)$, we can define (up to a sign) the singular directions $v_+(g), v_-(g) \in \mathbb{S}^1$ as the eigendirections of $(g^*g)^{\frac{1}{2}}$ corresponding to the eigenvalues $s_+(g), s_-(g)$. Note that $v_+(g) \perp v_-(g)$. We can do the some for the transpose g^* and we have $s_\pm(g) = s_\pm(g^*)$, $gv_\pm(g) = s_\pm(g)v_\pm(g^*)$.

For any $w \in \mathbb{S}^1$, we have $w = av_+(g) + bv_-(g)$ where $a = \langle w, v_+(g)\rangle$ and $b = \langle w, v_-(g)\rangle$. Applying g on both sides of the equation, we get

$$g\omega = a\|g\|\,v_+(g^*) + b\left\|g^{-1}\right\|^{-1} v_-(g^*).$$

Thus $\|g\omega\| \ge |a|\,\|g\|$

By assumption, we have $A^n(\omega) \in \mathrm{GL}_2(\mathbb{R})$ and $L^+(A) > L^-(A)$. Moreover,

$$L^+(A) = \lim_{n\to\infty}\frac{1}{n}\log s_+(A^n(\omega)),$$

and

$$L^-(A) = \lim_{n\to\infty} \frac{1}{n} \log s_-(A^n(\omega))$$

for $\mu^{\mathbb{N}}$-a.e. $\omega \in X^+$. Therefore, for almost every ω, $\exists\, N_\omega$ such that $\forall\, n \geq N_\omega$, we have $s_+(A^n(\omega)) > s_-(A^n(\omega))$. Thus in this case, $v_\pm(A^n(\omega))$ are well defined.

For $n \geq 1$, let $u_n(\omega) = v_+(A^n(\omega))$ when it makes sense (e.g. n is large enough). Then $u_n : X^+ \to \mathbb{S}^1$ is the most expanding direction of each n-th iterates. We also write $\hat{u}_n : X^+ \to \mathbb{P}$ as the corresponding projective version. We list two facts of u_n and $\hat{u}_n$ below:

- Fact 1. $\{\hat{u}_n\}_{n\geq 1}$ converges as $n \to \infty$ for $\mu^{\mathbb{N}}$ a.e. $\omega \in X^+$ and we call the limit respectively $\hat{u}_\infty : X^+ \to \mathbb{P}$ and $u_\infty : X^+ \to \mathbb{S}^1$.
- Fact 2. $u_\infty(\omega)^\perp = E^-(\omega)$ for $\mu^{\mathbb{N}}$ a.e. $\omega \in X^+$.

Now we can prove the claim, by direct computation

$$\frac{\|A^n(\omega)v_n\|}{\|A^n(\omega)\|} \geq |\langle v_n, v_+(A^n(\omega))\rangle| = |\langle v_n, u_n\rangle| \,.$$

Take lim inf on both sides, we have

$$\liminf_{n\to\infty} \frac{\|A^n(\omega)v_n\|}{\|A^n(\omega)\|} \geq |\langle v_0, u_\infty(\omega)\rangle|$$

for $\mu^{\mathbb{N}}$ a.e. $\omega \in X^+$.

Note that if $\langle v_0, u_\infty(\omega)\rangle = 0$, then $v \in u_\infty(\omega)^\perp = E^-(\omega)$. However, $v_0 \in E^-(\omega)$ happens for a set of ω's of probability zero because of quasi-irreducibility. This shows

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

for $\mu^{\mathbb{N}}$ a.e. $\omega \in X^+$, which proves the claim.

This finishes the whole proof. □

3.6. **The strong mixing of the Markov operator.** Our setup is the following. (Σ, μ) is a probability space. $A : \Sigma \to \mathrm{GL}_2(\mathbb{R})$ is continuous, quasi-irreducible and $L^+(A) > L^-(A)$. Moreover, there is some constant C such that $\|A\| \leq C$ (a consequence by being continuous on a compact set) and $\|A^{-1}\| \leq C$ (extra assumption). The Markov operator $Q = Q_A : L^\infty(\mathbb{P}) \to L^\infty(\mathbb{P})$ is defined as

$$Q\varphi(\hat{v}) = \int_\Sigma \varphi(\hat{A}(\omega_0)\hat{v}) d\mu(\omega_0).$$

Define the metric on $\mathbb{P}$ by $\delta(\hat{p}, \hat{q}) = \sin\angle(p, q) = \frac{\|p\wedge q\|}{\|p\|\|q\|}, p \in \hat{p}, q \in \hat{q}$. On $L^\infty(\mathbb{P})$, the infinity norm is defined by $\|\varphi\|_\infty = \sup_{\hat{p}\in\mathbb{P}} |\varphi(\hat{p})|$. For

$\alpha \in (0,1)$, we define the α seminorm on $L^\infty(\mathbb{P})$ as

$$v_\alpha(\varphi) := \sup_{\hat{p}\neq\hat{q}} \frac{|\varphi(\hat{p}) - \varphi(\hat{q})|}{\delta(\hat{p},\hat{q})^\alpha}.$$

This is not a norm as $v_\alpha(\varphi) = 0 \Rightarrow \varphi = \text{const}$. We call it α-Hölder seminorm. Then we can define the α-Hölder norm by

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

Denote $\mathcal{H}_\alpha(\mathbb{P}) := \{\varphi \in L^\infty(\mathbb{P}) : v_\alpha(\varphi) < \infty\}$. Then $(\mathcal{H}_\alpha(\mathbb{P}), \|\cdot\|_\alpha)$ is a normed space. For the observable

$$\psi_A(\hat{v}) = \int_\Sigma \log\|A(\omega_0)v\|\, d\mu(\omega_0),$$

it is easy to see that $\psi_A \in \mathcal{H}_\alpha(\mathbb{P})$.

Our goal is to show that $\exists\, \alpha \in (0,1)$ s.t. Q_A is strongly mixing on $\mathcal{H}_\alpha(\mathbb{P})$. That is,

$$\left\| Q_A^n(\varphi) - \int \varphi d\eta \right\|_\infty \le c\sigma^n \|\varphi\|_\alpha, \forall\, \varphi \in \mathcal{H}_\alpha(\mathbb{P})$$

with constants $c > 0$ and $\sigma \in (0,1)$.

Define

$$\mathcal{K}_\alpha(A,\mu) := \sup_{\hat{p}\neq\hat{q}} \int_\Sigma \left[\frac{\delta(\hat{A}(\omega_0)\hat{p}, \hat{A}(\omega_0)\hat{q})}{\delta(\hat{p},\hat{q})}\right]^\alpha d\mu(\omega_0).$$

It measures the average Hölder constant of $\hat{p} \to \hat{A}(\omega_0)\hat{p}$. We will prove several propositions about $\mathcal{K}_\alpha$.

Proposition 3.12. $\forall\, \varphi \in \mathcal{H}_\alpha(\mathbb{P})$, $v_\alpha(Q_A(\varphi)) \le \mathcal{K}_\alpha(A,\mu) v_\alpha(\varphi)$.

Proof. Given $\varphi \in \mathcal{H}_\alpha(\mathbb{P}), \forall\, \hat{p}, \hat{q} \in \mathbb{P}$, we have

$$\begin{aligned}
&\frac{|Q_A(\varphi)(\hat{p}) - Q_A(\varphi)(\hat{q})|}{\delta(\hat{p},\hat{q})^\alpha} \\
=&\frac{\left|\int_\Sigma \varphi(\hat{A}(\omega_0)\hat{p}) - \varphi(\hat{A}(\omega_0)\hat{q}) d\mu(\omega_0)\right|}{\delta(\hat{p},\hat{q})^\alpha} \\
\leq&\frac{\int_\Sigma \left|\varphi(\hat{A}(\omega_0)\hat{p}) - \varphi(\hat{A}(\omega_0)\hat{q})\right| d\mu(\omega_0)}{\delta(\hat{p},\hat{q})^\alpha} \\
\leq&\int_\Sigma \frac{\left|\varphi(\hat{A}(\omega_0)\hat{p}) - \varphi(\hat{A}(\omega_0)\hat{q})\right|}{\delta(\hat{A}(\omega_0)\hat{p}, \hat{A}(\omega_0)\hat{q})^\alpha} \cdot \frac{\delta(\hat{A}(\omega_0)\hat{p}, \hat{A}(\omega_0)\hat{q})^\alpha}{\delta(\hat{p},\hat{q})^\alpha} d\mu(\omega_0) \\
\leq&v_\alpha(\varphi) \cdot \int_\Sigma \frac{\delta(\hat{A}(\omega_0)\hat{p}, \hat{A}(\omega_0)\hat{q})^\alpha}{\delta(\hat{p},\hat{q})^\alpha} d\mu(\omega_0).
\end{aligned}$$

Take the supremum in $\hat{p} \neq \hat{q}$ on both sides, we get exactly

$$v_\alpha(Q_A(\varphi)) \leq \mathcal{K}_\alpha(A,\mu) v_\alpha(\varphi).$$

□

Proposition 3.13. *The sequence $\{\mathcal{K}_\alpha(A^n, \mu^n)\}_{n\geq 0}$ is sub-multiplicative: $\forall\, n, m \in \mathbb{N}$,*

$$\mathcal{K}_\alpha(A^{n+m}, \mu^{n+m}) \leq \mathcal{K}_\alpha(A^n, \mu^n)\mathcal{K}_\alpha(A^m, \mu^m).$$

Note that for $n = 0$, $\mathcal{K}_\alpha(A^n, \mu^n) = 1$.

Proof. Direct computation shows

$$\begin{aligned}
&\mathcal{K}_\alpha(A^{n+m}, \mu^{n+m}) = \sup_{\hat{p}\neq\hat{q}} \int_{\Sigma^{n+m}} \left[\frac{\delta(\hat{A}^{n+m}(\omega)\hat{p}, \hat{A}^{n+m}(\omega)\hat{q})}{\delta(\hat{p},\hat{q})}\right]^\alpha d\mu^{n+m}(\omega) \\
=&\sup_{\hat{p}\neq\hat{q}} \int_{\Sigma^{n+m}} \left[\frac{\delta(\hat{A}^{n+m}(\omega)\hat{p}, \hat{A}^{n+m}(\omega)\hat{q})}{\delta(\hat{A}^m(\omega)\hat{p}, \hat{A}^m(\omega)\hat{q})}\right]^\alpha \left[\frac{\delta(\hat{A}^m(\omega)\hat{p}, \hat{A}^m(\omega)\hat{q})}{\delta(\hat{p},\hat{q})}\right]^\alpha d\mu^{n+m}(\omega) \\
=&\sup_{\hat{p}\neq\hat{q}} \int_{\Sigma^m} \left[\frac{\delta(\hat{A}^m(\omega)\hat{p}, \hat{A}^m(\omega)\hat{q})}{\delta(\hat{p},\hat{q})}\right]^\alpha \int_{\Sigma^n} \left[\frac{\delta(\hat{A}^{n+m}(\omega)\hat{p}, \hat{A}^{n+m}(\omega)\hat{q})}{\delta(\hat{A}^m(\omega)\hat{p}, \hat{A}^m(\omega)\hat{q})}\right]^\alpha d\mu^n d\mu^m \\
\leq&\mathcal{K}_\alpha(A^n, \mu^n)\mathcal{K}_\alpha(A^m, \mu^n).
\end{aligned}$$

Note that the last equality holds because A takes value in $\mathrm{GL}_2(\mathbb{R})$ which never maps a line to zero. □

Remark 3.5. As A and A^{-1} are assumed to be bounded by some constant $C > 0$, we have that given any $n \in \mathbb{N}$, for $0 < \alpha < \frac{1}{4n}$, we have $\mathcal{K}_\alpha(A^n, \mu^n) \leq e^C =: L$.

Proposition 3.14. $\forall\, n \in \mathbb{N}$, $Q_{A^n} = (Q_A)^n$.

Proof. By definition,

$$Q_A(\varphi)(\hat{v}) = \int_\Sigma \varphi(\hat{A}(\omega)\hat{v})d\mu(\omega_0).$$

Then

$$\begin{aligned}(Q_A)^2(\varphi)(\hat{v}) &= \int_\Sigma (Q_A\varphi)(\hat{A}(\omega_0)\hat{v})d\mu(\omega_0)\\ &= \int_\Sigma\int_\Sigma \varphi(\hat{A}(\omega_1)\hat{A}(\omega_0)\hat{v})d\mu(\omega_0)d\mu(\omega_1)\\ &= (Q_{A^2})(\varphi)(\hat{v}).\end{aligned}$$

The proof follows by induction. $\square$

Proposition 3.15. *Given $\alpha > 0$ and two point $\hat{p} \neq \hat{q} \in \mathbb{P}$, we have*

$$\left[\frac{\delta(\hat{A}(\omega_0)\hat{p}, \hat{A}(\omega_0)\hat{q})}{\delta(\hat{p},\hat{q})}\right]^\alpha \leq \frac{|\det A(\omega_0)|^\alpha}{2}\left[\frac{1}{\|A(\omega_0)p\|^{2\alpha}} + \frac{1}{\|A(\omega_0)q\|^{2\alpha}}\right].$$

for any $\omega_0 \in \Sigma$.

Proof. By the property of exterior product, we have

$$\|A(\omega_0)p \wedge A(\omega_0)q\| = \|\wedge_2 A(\omega_0)(p\wedge q)\| = |\det A(\omega_0)|\,\|p\wedge q\|.$$

Hence,

$$\begin{aligned}\left[\frac{\delta(\hat{A}(\omega_0)\hat{p}, \hat{A}(\omega_0)\hat{q})}{\delta(\hat{p},\hat{q})}\right]^\alpha &\leq \left[\frac{\|A(\omega_0)p\wedge A(\omega_0)q\|}{\|A(\omega_0)p\|\,\|A(\omega_0)q\|}\cdot\frac{\|p\|\,\|q\|}{\|p\wedge q\|}\right]^\alpha\\ &= \left[\frac{|\det A(\omega_0)|}{\|A(\omega_0)p\|\,\|A(\omega_0)q\|}\right]^\alpha\\ &\leq \frac{|\det A(\omega_0)|^\alpha}{2}\left[\frac{1}{\|A(\omega_0)p\|^{2\alpha}} + \frac{1}{\|A(\omega_0)q\|^{2\alpha}}\right].\end{aligned}$$

Here the last inequality uses $\sqrt{ab} \leq \frac{1}{2}(a+b)$ for non-negative a and b. $\square$

Proposition 3.16. *Given a cocycle $(A,\mu) \in L^\infty(\Sigma, \mathrm{GL}_2(\mathbb{R}))\times\mathrm{Prob}(\Sigma)$, we have that*

$$\mathcal{K}_\alpha(A,\mu) \leq \sup_{\hat{p}\in\mathbb{P}}\int_\Sigma \frac{|\det A(\omega_0)|^\alpha}{\|A(\omega_0)p\|^{2\alpha}}d\mu(\omega_0) = \sup_{\hat{p}\in\mathbb{P}}\mathbb{E}(\left[\frac{|\det A(\omega_0)|}{\|A(\omega_0)p\|^2}\right]^\alpha)$$

holds $\forall\,\alpha > 0$. Note that $|\det A(\omega_0)| = s_1(A(\omega_0))s_2(A(\omega_0))$.

Proof. It follows from the definition of $\mathcal{K}_\alpha$ and the previous proposition by taking integral and supremum in $\hat{p} \neq \hat{q}$ on both sides. $\square$

Proposition 3.17. *Let $(A, \mu) \in L^\infty(\Sigma, \mathrm{GL}_2(\mathbb{R})) \times \mathrm{Prob}(\Sigma)$ be a quasi-irreducible cocycle with $L^+(A) > L^-(A)$. There are numbers $\alpha \in (0, 1)$, $\kappa \in (0, 1)$ and $n \in \mathbb{N}$ s.t. $\mathcal{K}_\alpha(A^n, \mu^n) \leq \kappa$.*

Proof. We know by Theorem 3.10 that as $n \to \infty$

$$\mathbb{E}(\frac{1}{n} \log \|A^n(\omega)v\|) \to L^+(A)$$

uniformly in $v \in \mathbb{S}^1$. Thus

$$\mathbb{E}(\frac{1}{n} \log \|A^n(\omega)v\|^{-2}) \to -2 \cdot L^+(A)$$

uniformly in $v \in \mathbb{S}^1$. Therefore, $\forall\, \epsilon > 0$, $\forall\, v \in \mathbb{S}^1$, $\exists N = N(\epsilon) \in \mathbb{N}$ which does not depend on v, such that $\forall\, n > N$ we have

$$-2L^+(A) - \epsilon \leq \mathbb{E}(\frac{1}{n} \log \|A^n(\omega)v\|^{-2}) \leq -2L^+(A) + \epsilon$$

Therefore, by choosing ϵ sufficiently small e.g. $\epsilon < \frac{1}{4}[L^+(A) - L^-(A)]$ and n large enough, we have

$$\mathbb{E}(\log \|A^n(\omega)v\|^{-2}) \leq n(-2L^+(A) + \epsilon)$$

Moreover, we have

$$\log |\det A^n(\omega)| = \log |s_1(A^n(\omega))| + \log |s_2(A^n(\omega))| \leq n(L^+(A) + L^-(A) + \epsilon).$$

Combining these two estimates, we have

$$\begin{aligned}
\log \frac{|\det A^n(\omega)|}{\|A^n(\omega)v\|^2} &\leq n(L^+(A) + L^-(A) + \epsilon) + n(-2L^+(A) + \epsilon) \\
&= n(L^-(A) - L^+(A) + 2\epsilon) \\
&\leq n \cdot \frac{1}{2}(L^-(A) - L^+(A)) \\
&\leq -1
\end{aligned}$$

as n is sufficiently large and $L^+(A) > L^-(A)$.

Making use of the inequality

$$e^x \leq 1 + x + \frac{x^2}{2} e^{|x|},$$

we have $\forall\, v \in \mathbb{S}^1$,

$$\begin{aligned}
&\mathbb{E}(\left[\frac{|\det A^n(\omega)|}{\|A^n(\omega)v\|^2}\right]^\alpha) \\
=&\mathbb{E}(e^{\alpha\log\frac{|\det A^n(\omega)|}{\|A^n(\omega)v\|^2}}) \\
\leq&\mathbb{E}(1+\alpha\log\frac{|\det A^n(\omega)|}{\|A^n(\omega)v\|^2}+\frac{[\alpha\log\frac{|\det A^n(\omega)|}{\|A^n(\omega)v\|^2}]^2}{2}e^{\left|\alpha\log\frac{|\det A^n(\omega)|}{\|A^n(\omega)v\|^2}\right|}) \\
\leq&1-\alpha+\mathrm{O}(\alpha^2) \\
\leq&\kappa<1
\end{aligned}$$

as we take α sufficiently small. Therefore, by the previous proposition we get $\mathcal{K}_\alpha(A^n, \mu^n) \leq \kappa < 1$. □

Now we can easily prove that Q_A is strongly mixing.

Theorem 3.11. *Q_A is strongly mixing on $\mathcal{H}_\alpha(\mathbb{P})$ where $\alpha \in (0,1)$ is given by Proposition 3.17. In fact, we can prove a stronger statement. That is, for $\varphi \in \mathcal{H}_\alpha(\mathbb{P})$, $\forall\, n \in \mathbb{N}$*

$$\left\|Q_A^n(\varphi) - \int \varphi d\eta\right\|_\alpha \leq C_0\sigma^n \|\varphi\|_\alpha$$

where $C_0 > 0$ and $\sigma \in (0,1)$ are constants.

Proof. By Proposition 3.12 and 3.14, we have $\forall\, \varphi \in \mathcal{H}_\alpha(\mathbb{P})$,

$$v_\alpha(Q_A^s(\varphi)) = v_\alpha(Q_{A^s}(\varphi)) \leq \mathcal{K}_\alpha(A^s, \mu^s)v_\alpha(\varphi), \quad \forall\, s \in \mathbb{N}.$$

Choose the parameter n from Proposition 3.17. For any $m = kn+r \in \mathbb{N}$ with $k, r \in \mathbb{N}$ and $r < n$, we have

$$v_\alpha(Q_A^m(\varphi)) \leq [\mathcal{K}_\alpha(A^n, \mu^n)]^k\mathcal{K}_\alpha(A^r, \mu^r)v_\alpha(\varphi) \leq \kappa^k \cdot Lv_\alpha(\varphi).$$

Then if we denote $\sigma = \kappa^{\frac{1}{n}} < 1$, then

$$v_\alpha(Q_A^m(\varphi)) \leq C\sigma^m v_\alpha(\varphi),$$

where C is a constant.

$\forall\, \varphi \in \mathcal{H}_\alpha(\mathbb{P})$, we have $\forall\, n \in \mathbb{N}$

$$\left\|Q_A^n(\varphi) - \int \varphi d\eta\right\|_\alpha = \left\|Q_A^n(\varphi) - \int \varphi d\eta\right\|_\infty + v_\alpha\left(Q_A^n(\varphi) - \int \varphi d\eta\right).$$

Note that

$$v_\alpha\left(Q_A^n(\varphi) - \int \varphi d\eta\right) \leq v_\alpha\left(Q_A^n(\varphi)\right) \leq C\sigma^n v_\alpha(\varphi) \leq C\sigma^n \|\varphi\|_\alpha.$$

Since $v_\alpha\left(Q_A^n(\varphi)\right) \leq C\sigma^n v_\alpha(\varphi)$, then $Q_A^n(\varphi)$ is almost constant in $\hat{v} \in \mathbb{P}$: $\forall\, \hat{p} \neq \hat{q} \in \mathbb{P}$

$$|Q_A^n(\varphi)(\hat{p}) - Q_A^n(\varphi)(\hat{q})| \leq C\sigma^n v_\alpha(\varphi).$$

Thus $\forall\, \hat{p} \in \mathbb{P}$,

$$\left| Q_A^n(\varphi)(\hat{p}) - \int Q_A^n(\varphi) d\eta \right| \leq C\sigma^n v_\alpha(\varphi).$$

Note that $\eta \in \operatorname{Prob}_K(\mathbb{P})$ is Q_A-invariant. So $\int Q_A^n(\varphi) d\eta = \int \varphi d\eta$. Thus

$$\left\| Q_A^n(\varphi) - \int \varphi d\eta \right\|_\infty \leq C\sigma^n v_\alpha(\varphi).$$

To conclude,

$$\left\| Q_A^n(\varphi) - \int \varphi d\eta \right\|_\alpha \leq 2C\sigma^n v_\alpha(\varphi) \leq 2C\sigma^n \left\|\varphi\right\|_\alpha.$$

This finishes the proof. $\square$

By Lemma 3.2, we get that $\bar{Q}$ is strongly mixing on $\mathcal{E} = \mathcal{H}_\alpha(\Sigma \times \mathbb{P}) := \{\varphi \in L^\infty(\Sigma \times \mathbb{P}) : v_\alpha(\Pi\varphi) < \infty\}$. Note that $\Pi\mathcal{H}_\alpha(\Sigma \times \mathbb{P}) = \mathcal{H}_\alpha(\mathbb{P})$.

Denote $(M, K, \mu, \mathcal{E}) = (\Sigma \times \mathbb{P}, \bar{K}, \mu \times \eta, \mathcal{H}_\alpha(\Sigma \times \mathbb{P}))$, apply Theorem 2.1, we obtain Theorem 3.2.

4. Mixed random-quasiperiodic dynamics

We will derive statistical properties for the following skew-product dynamical system.

Let $\Sigma = \mathbb{T}^d$ and $\mu \in \operatorname{Prob}(\mathbb{T}^d)$. Consider the map

$$f : \Sigma^\mathbb{Z} \times \mathbb{T}^d \to \Sigma^\mathbb{Z} \times \mathbb{T}^d, \quad f(\omega, \theta) = (\sigma\omega, \theta + \omega_0).$$

We will consider the MPDS $(\Sigma^\mathbb{Z} \times \mathbb{T}^d, f, \mu^\mathbb{Z} \times m)$ where m is the Haar measure on the torus $\mathbb{T}^d$. For simplicity, from now on we set $d = 1$. Things are the same in the higher dimensional torus. For $d = 1$, m is just the Lebesgue measure on the circle.

Theorem 4.1. *$(\Sigma^\mathbb{Z} \times \mathbb{T}, f, \mu^\mathbb{Z} \times m)$ is ergodic if and only if $\forall\, k \neq 0 \in \mathbb{Z}$, $\exists\, \alpha \in \operatorname{supp}(\mu)$ such that $k\alpha \notin \mathbb{Z}$. In particular, if $\exists\, \alpha \in \operatorname{supp}(\mu)$ with $\alpha \notin \mathbb{Q}$, then f is ergodic.*

Remark 4.1. A simple example of not having any irrational number in the $\operatorname{supp}(\mu)$ but still having ergodicity is $\operatorname{supp}(\mu) = \{\frac{1}{n}\}_{n\in\mathbb{N}}$.

Consider the Markov chain on $\Sigma \times \mathbb{T}$:

$$(\omega_0, \theta) \to (\omega_1, \theta + \omega_0) \to (\omega_2, \theta + \omega_0 + \omega_1) \to \cdots$$

Its Markov kernal on $\Sigma \times \mathbb{T}$ is defined as

$$\bar{K}_{(\omega_0,\theta)} = \int_\Sigma \delta_{(\omega_1,\theta+\omega_0)} d\mu(\omega_1).$$

The corresponding Markov operator $\bar{Q}$ on $L^\infty(\Sigma \times \mathbb{T})$ is

$$\bar{Q}\varphi(\omega_0,\theta) = \int_\Sigma \varphi(\omega_1, \theta + \omega_0) d\mu(\omega_1).$$

Our goal is to prove that $\bar{Q}$ is strongly mixing on an appropriate space of observables. To achieve this, we will make some preparations.

4.1. **Some basic Fourier analysis concepts.** Let $\varphi \in L^1(\mathbb{T})$, its Fourier coefficients are

$$\hat{\varphi}(k) := \int_0^1 \varphi(x) e^{-2\pi i k x} dx, \quad \forall\, k \in \mathbb{Z}.$$

Note that roughly speaking,

$$\varphi(x) \approx \sum_{k=-\infty}^{+\infty} \hat{\varphi}(k) e^{2\pi i k x},$$

which is the Fourier series of φ. Moreover, for $N \in \mathbb{N}$ let

$$S_N\varphi(x) := \sum_{k=-N}^{N} \hat{\varphi}(k) e^{2\pi i k x}$$

be the N-th partial series.

The Fourier series "represents" the function in certain sense.

(1) if $\hat{\varphi}_1(k) = \hat{\varphi}_2(k)$, $\forall\, k \in \mathbb{Z}$, then $\varphi_1 = \varphi_2$ m-a.e.
(2) if $\varphi \in L^2(\mathbb{T})$, then

$$\varphi(x) = \sum_{k=-\infty}^{+\infty} \hat{\varphi}(k) e^{2\pi i k x},$$

in $L^2(\mathbb{T})$. That is $\|S_N\varphi - \varphi\|_2 \to 0$ as $N \to \infty$.
(3)
(4)

5. Limit laws for multiplicative random systems

6. Limit laws for hyperbolic systems

7. Partially hyperbolic systems