STATISTICAL PROPERTIES FOR CERTAIN DYNAMICAL SYSTEMS

AO CAI, PEDRO DUARTE, AND SILVIUS KLEIN

Abstract. This is a collection of notes for the course "Topics on Dynamical Systems" given by the third author at the Pontifical Catholic University of Rio de Janeiro (PUC-Rio) in 2022, with the assistance of the first author. The third chapter of this text consists in part of the notes written by the second author for the mini-course he gave at PUC-Rio in 2020. The mathematical content of this manuscript is to a large extent based upon various papers of the authors.

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 $\xi_1, \dots, \xi_n$ are independent if for any Borel measurable sets $E_1, \dots, E_n \subset \mathbb{R}$,

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

Theorem 1.1 (The law of large numbers - LLN)**.** *Given i.i.d. sequence $\xi_0, \xi_1, \dots, \xi_{n-1}, \xi_n, \dots$ 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, $\forall \, \epsilon > 0$,

$$\mathbb{P} \left\{ \left| \frac{1}{n} S_n - \mathbb{E}\xi_0 \right| > \epsilon \right\} \to 0 \quad \text{ as } \quad n \to \infty.$$

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 $\xi_0, \xi_1, \dots, \xi_{n-1}, \xi_n, \dots$ satisfies a certain growth condition and is non-trivial. Then $\forall\, \epsilon > 0$,*

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

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 $|\xi_0| \le C$ a.s. Then $\forall\, \epsilon > 0$ the following holds* for all $n \in \mathbb{N}$*:*

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

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 $\mathbb{E}\left(\log\|g\|\right) d\mu(g) < \infty$, 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*

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

or

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

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 $\forall\, v \in \mathbb{R}^2$, $v \neq 0$,*

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

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 $\nu \in \text{Prob}(M)$.

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

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

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.** 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

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

where $E \subset \mathbb{R}$ is Borel measurable. In other words, $\mu_\xi = \xi_*\mathbb{P}$. Given a random variable ξ and $\mu \in \mathrm{Prob}(\mathbb{R})$, 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 $\mathbb{E}X \geq \int_E X d\mathbb{P} \geq \lambda \mathbb{P}(E)$. □

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 $\xi_0, \xi_1, \ldots, \xi_{n-1}, \xi_n, \ldots$ of real random variables, if $\mathbb{E}\xi_0^2 < \infty$ then $\forall\, \epsilon > 0$,*

$$\mathbb{P}\left\{\left|\frac{1}{n}S_n - \mathbb{E}\xi_0\right| > \epsilon\right\} \to 0 \quad \text{as} \quad n \to \infty.$$

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

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* $\xi_0, \xi_1, \ldots, \xi_{n-1}, \xi_n, \ldots$ *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

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

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* $\xi_0, \xi_1, \ldots, \xi_{n-1}, \xi_n, \ldots$ *satisfies a certain growth condition and* $\sigma^2 > 0$*. Then* $\forall\, \epsilon > 0$,

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

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

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

By Markov's inequality,

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

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

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

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

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

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

then we would have

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

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,

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

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

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

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

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

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

$$\frac{1}{3}\log\|g_2 g_1 g_0 v\| = \frac{1}{3}[\log\frac{\|g_2 g_1 g_0 v\|}{\|g_1 g_0 v\|} + \log\frac{\|g_1 g_0 v\|}{\|g_0 v\|} + \log\|g_0 v\|].$$

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

$$\frac{1}{3}\log\|g_2 g_1 g_0 v\| = \frac{1}{3}[\varphi(Z_2^v(\omega)) + \varphi(Z_1^v(\omega)) + \varphi(Z_0^v(\omega))].$$

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 $P = \{p_{ij}\}_{1\leq i,j\leq n}$ be a stochastic matrix. Namely,

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

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 $q_i \geq 0$, $\forall 1 \leq i \leq n$ and $\sum_{i=1}^n q_i = 1$.

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

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

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 $P^n_{ij} > 0, \forall\, 1 \leq i, j \leq n$, then $\exists!$ stationary vector q and $P^n_{ij} \to q_j$ exponentially fast for any $1 \leq i \leq n$.

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

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

$\mathcal{B}^+ = \sigma$- 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 $\{Z_n\}_{n\geq 0}$ 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 $\{Z_n\}_{n\geq 0}$ 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 kernel)**.** A Markov kernel $K(x, E)$ (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 $x \mapsto K_x = K(x, \cdot) \in \text{Prob}(M)$ is continuous which in particular impies (2). In other words, we can think of a Markov kernel as a continuous function $K : M \to \text{Prob}(M)$ where we interpret K_x as the probability of transitioning from x to somewhere.

Definition 2.3 (Stationary measure)**.** $\mu \in \text{Prob}(M)$ 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 $\pi \in \text{Prob}(M)$ and K a Markov kernel, 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 $\mathbb{E}_{\delta_x} := \mathbb{E}_x$. 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.

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 kernel 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, $\forall\, \varphi \in L^\infty(M)$,

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

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

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

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 $K : M \to \mathrm{Prob}(M), x \mapsto K_x$ be a continuous kernel. 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 $K_x^n = \int_M K_y^{n-1} dK_x(y)$.

$\mu \in \mathrm{Prob}(M)$ 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, $M \to \mathrm{Prob}(M), x \to \delta_{f(x)}$.

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

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

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 $\mathrm{Prob}(M)$. By definition, we have $\langle \varphi, Q^*\nu \rangle = \langle Q\varphi, \nu \rangle$ for any $\varphi \in C^0(M)$ and $\nu \in \mathrm{Prob}(M)$. In other words, $Q^*\nu$ is the probability on M s.t.

$$\int_M \varphi dQ^*\nu = \int_M Q\varphi d\nu, \quad \forall\, \varphi \in C^0(M).$$

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

In practice, the assumptions in Definition 2.5 is unreasonably strong, primarily because of $\varphi \in L^\infty(M)$ where L^∞ is too big. 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 Markov system $(M, K, \mu, \mathcal{E})$ is strongly mixing if $\forall\, n \in \mathbb{Z}^+$ (for $n = 0$ it holds trivially),

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

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

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

Theorem 2.1 (Cai, Duarte, Klein 2022)**.** *If $(M, K, \mu, \mathcal{E})$ is a strongly mixing Markov system with mixing rate $r_n = \frac{1}{n^p}, p > 0$, then $\forall\, x_0 \in M$ and $\forall\, \epsilon > 0$*

$$\mathbb{P}_{x_0} \left\{ \left| \frac{1}{n} S_n\varphi - \int_M \varphi d\mu \right| > \epsilon \right\} \lesssim e^{-c(\epsilon)n}$$

holds for all $n \in \mathbb{Z}^+$, for all $\varphi \in \mathcal{E}$ and for some $c(\epsilon) = C(C_0, L, p)\epsilon^{2+\frac{1}{p}}$ where $C(C_0, L, p) = (3C_0L)^{-(2+\frac{1}{p})} > 0$ is a constant depending only on *the mixing coefficient* C_0*, the mixing exponent* p *and the upper bound* L *of* $\|\varphi\|_{\mathcal{E}}$.

Remark 2.3. For fixed x_0, we need C^0 norm (we write $\|\cdot\|_\infty$ but we actually mean $\|\cdot\|_0$) in the left hand side of the strongly mixing condition. If we replace $\mathbb{P}_{x_0}$ by $\mathbb{P}_\mu$ since $\mathbb{P}_\mu = \int_M \mathbb{P}_{x_0} d\mu(x_0)$, then L^∞ norm w.r.t μ is enough. In any case, this will not affect any of our applications.

Proof. Without loss of generality, we assume $\mathbb{E}\varphi = 0$, otherwise we consider $\varphi - \mathbb{E}\varphi$. Moreover, it is enough to consider $\mathbb{P}_{x_0}\{S_n\varphi \geq n\epsilon\}$. Using Bernstein's trick, for any $t > 0$ we have

$$\mathbb{P}_{x_0}\{S_n\varphi \geq n\epsilon\} = \mathbb{P}_{x_0}\{e^{tS_n\varphi} \geq e^{tn\epsilon}\} \leq e^{-tn\epsilon}\mathbb{E}_{x_0}(e^{tS_n\varphi}).$$

So our goal in the following is to estimate $\mathbb{E}_{x_0}(e^{tS_n\varphi})$ by relating it to $Q^{n_0}(\varphi)$ for some suitable choice of $n_0 \leq n$.

Note that

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

Take $n_0 \leq n$ such that $n = n_0 m + r$. In order to show the strategy, we may assume $r = 0$ (which is actually without loss of generality because the remainder is bounded by some constant). 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 the generalized Hölder inequality, we have

$$\mathbb{E}_{x_0}(\prod_{j=0}^{n-1} f_j) = \mathbb{E}_{x_0}(F_0 F_1 \cdots F_{n_0-1}) \leq \prod_{k=0}^{n_0-1} [\mathbb{E}_{x_0}(F_k^{n_0})]^{\frac{1}{n_0}}.$$

Thus, it is enough to estimate each $\mathbb{E}_{x_0}(F_k^{n_0})$. In fact, we are going to relate them to some powers of the Markov operator, which we formulate as the following lemma.

Lemma 2.1. *Let* $\varphi \in C^0(M), \|\varphi\|_{\mathcal{E}} \leq L < \infty$*. Let* $n \geq n_0$ *be two integers and denote by* $m := \lfloor \frac{n}{n_0} \rfloor$*. Then* $\forall\, t > 0, \forall\, x_0 \in M$,

$$\mathbb{E}_{x_0}(e^{tS_n\varphi}) \leq e^{2tn_0L} \left\|Q^{n_0}(e^{tn_0\varphi})\right\|_\infty^{m-1}.$$

Proof. By assumption we have $0 < f_j = f(x_j) = e^{t\varphi(x_j)} \leq e^{tL}$, $\forall j \in \mathbb{N}$. We rewrite $f_0 \cdots f_{n-1}$ as $F_0 F_1 \cdots F_{n_0-1} F_{n_0}$ where

$$F_j = f_j f_{n_0+j} \cdots f_{(m-1)n_0+j},\ 0 \leq j \leq n_0 - 1$$

and

$$F_{n_0} = f_{mn_0} f_{mn_0+1} \cdots f_{mn_0+r-1}.$$

Then $F_{n_0} < e^{tn_0 L}$ as $r < n_0$. Therefore, we have

$$\mathbb{E}_{x_0}(\prod_{j=0}^{n-1} f_j) \le e^{tn_0 L} \mathbb{E}_{x_0}(F_0 \cdots F_{n_0-1}) \le e^{tn_0 L} \prod_{k=0}^{n_0-1} [\mathbb{E}_{x_0}(F_k^{n_0})]^{\frac{1}{n_0}}.$$

We will show that $\mathbb{E}_{x_0}(F_k^{n_0}) \le e^{tn_0 L} \|Q^{n_0}(e^{tn_0\varphi})\|_\infty^{m-1}$, $\forall\, k = 0, \cdots, n_0 - 1$, which implies the result of this lemma.

Note that

$$F_k^{n_0} = f_k^{n_0} f_{n_0+k}^{n_0} \cdots f_{(m-1)n_0+k}^{n_0} = e^{tn_0\varphi(x_k)} e^{tn_0\varphi(x_{n_0+k})} \cdots e^{tn_0\varphi(x_{(m-1)n_0+k})}.$$

For convenience, we denote $G(x) = F_k^{n_0}$ and $g(x_k) = e^{tn_0\varphi(x_k)}$ etc. Then by assumption $0 < g(x_k) \le e^{tn_0 L}$ for each k. It remains to estimate the integral of a function of the type:

$$G(x) = g(x_k) g(x_{n_0+k}) \cdots g(x_{(m-1)n_0+k}),\ 0 < g \le e^{tn_0 L}.$$

It is an integral w.r.t. a Markov measure, of a function $G(x)$ depends on a finite and **sparse** set of coordinates. We will show that

$$\mathbb{E}_{x_0}(G) \le e^{tn_0 L} \|Q^{n_0} g\|_\infty^{m-1}.$$

For simplicity, let us prove by showing an example when $k = 1, n_0 = 3, m - 1 = 2$ and $n = 2 \times 3 + 1 = 7$. The general case, which is identically the same, is left to the readers.

By direct computation, we have

$$\begin{aligned}
&\int g(x_1) g(x_4) g(x_7) d\mathbb{P}_{x_0}(x) \\
&= \int \cdots \int g(x_1) g(x_4) g(x_7) dK_{x_6}(x_7) \cdots dK_{x_0}(x_1) d\delta_{x_0}(x_0) \\
&= \int \cdots \int g(x_1) g(x_4) [\int g(x_7) dK_{x_6}(x_7) dK_{x_5}(x_6) dK_{x_4}(x_5)] \cdots
\end{aligned}$$

Note that

$$\int g(x_7) dK_{x_6}(x_7) dK_{x_5}(x_6) dK_{x_4}(x_5) = Q^3 g(x_4) \le \left\|Q^3 g\right\|_\infty.$$

Similarly,

$$\int g(x_4) dK_{x_3}(x_4) dK_{x_2}(x_3) dK_{x_1}(x_2) = Q^3 g(x_1) \le \left\|Q^3 g\right\|_\infty.$$

Thus

$$\mathbb{E}_{x_0}(g(x_1) g(x_4) g(x_7)) \le \left\|Q^3 g\right\|_\infty^2 \int g(x_1) dK_{x_0}(x_1) \le e^{3tL} \left\|Q^3 g\right\|_\infty^2.$$

This finishes the lemma. □

By the strongly mixing assumption of the theorem, we know that (we may insert the mixing coefficient C_0 into L)

$$\|Q^{n_0}\varphi\|_\infty \le C_0 \|\varphi\|_{\mathcal{E}} \frac{1}{n_0^p} \le L\frac{1}{n_0^p}.$$

By the lemma above, we have $\forall\, n \ge n_0$, $\forall\, x_0 \in M$,

$$\mathbb{E}_{x_0}(e^{tS_n\varphi}) \le e^{2tn_0L} \left\|Q^{n_0}(e^{tn_0\varphi})\right\|_\infty^{m-1}.$$

However, $\varphi \in \mathcal{E}$ does not necessarily imply $e^{tn_0\varphi} \in \mathcal{E}$, so if we want to make use of the strongly mixing condition, we have to do one more step.

Note that the following inequality holds for all $x \in \mathbb{R}$

$$e^x \le 1 + x + \frac{x^2}{2} \cdot e^{|x|}.$$

Hence we can write

$$e^x = 1 + x + \frac{x^2}{2} \cdot \psi(x)$$

for some $|\psi(x)| \le e^{|x|}$. Therefore,

$$e^{tn_0\varphi} = 1 + tn_0\varphi + \frac{1}{2}t^2n_0^2\varphi^2\psi(tn_0\varphi)$$

where $|\psi(tn_0\varphi)| \le e^{tn_0\|\varphi\|_\infty} \le e^{tn_0L} \le 2$ if $tn_0L \le \frac{1}{2}$, namely $t \le \frac{1}{2Ln_0}$. Then we have

$$Q^{n_0}(e^{tn_0\varphi}) = 1 + tn_0Q^{n_0}\varphi + \frac{1}{2}t^2n_0^2Q^{n_0}(\varphi^2\psi(tn_0\varphi))$$

which shows

$$\left\|Q^{n_0}(e^{tn_0\varphi})\right\|_\infty \le 1 + tn_0L\frac{1}{n_0^p} + t^2n_0^2L^2 \le 1 + 2t^2n_0^2L^2$$

if we have $tn_0L\frac{1}{n_0^p} \le t^2n_0^2L^2 \Leftrightarrow t \ge \frac{1}{Ln_0^{1+p}}$. Note that we can choose $t \in \mathbb{R}$ satisfying $\frac{1}{Ln_0^{1+p}} \le t \le \frac{1}{2Ln_0}$ since n_0 can be chosen sufficiently large such that $n_0^p \ge 2$.

By the inequality $(1+y)^{\frac{1}{y}} \le e, y > 0$, we have

$$\left\|Q^{n_0}(e^{tn_0\varphi})\right\|_\infty^{\frac{n}{n_0}} \le (1 + 2t^2n_0^2L^2)^{\frac{1}{2t^2n_0^2L^2}\cdot 2t^2n_0^2L^2\cdot\frac{n}{n_0}} \le e^{2t^2n_0L^2n}.$$

By the lemma above, we have

$$\mathbb{E}_{x_0}(e^{tS_n\varphi}) \le e^{2tn_0L}e^{2t^2n_0L^2n} \le 4e^{2t^2n_0L^2n}.$$

By Bernstein's trick, we have

$$\mathbb{P}_{x_0}\{S_n\varphi \geq n\epsilon\} \leq e^{-tn\epsilon}\mathbb{E}_{x_0}(e^{tS_n\varphi}) \leq 4e^{-tn\epsilon}e^{2t^2n_0L^2n}.$$

It remains to estimate $-t\epsilon + 2t^2n_0L^2$ with some proper choice. For our purpose, we can choose $n_0 = (\frac{3L}{\epsilon})^{\frac{1}{p}}$ and $t = \frac{1}{Ln_0^{1+p}}$ such that $-t\epsilon + 2t^2n_0L^2 < -C(L,p)\epsilon^{2+\frac{1}{p}} := -c(\epsilon)$ where $C(L,p) = (3L)^{-(2+\frac{1}{p})} > 0$ is a constant depending only on L, p and the strongly mixing coefficient C_0 as we already insert it into L.

This finishes the whole proof of the theorem. □

We now recall an abstract central limit theorem of Gordin and Livšic (see [4] and [5]).

Theorem 2.2 (Gordin-Livšic)**.** *Let (M,K,ν) be an ergodic Markov system, let $\varphi \in L^2(\nu)$ with $\int \varphi d\nu = 0$ and assume that*

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

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

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

Recall that a Markov system (M,K,ν) is ergodic if the measure ν is an extremal point in the convex space of K-stationary probability measures on M. This is equivalent to the ergodicity of the shift map on the product space X^+ relative to the Markov measure $\mathbb{P} = \mathbb{P}_\nu$. Evidently, if K admits a unique stationary measure, then the corresponding Markov system is ergodic.

As a consequence of the above result we obtain the following.

Proposition 2.2. *Let $(M,K,\nu,\mathcal{E})$ be a strongly mixing Markov system (relative to the uniform norm) with mixing rate $r_n = \frac{1}{n^p}$ with $p > 1$, where $\mathcal{E}$ is a dense subset of $C_b(M)$.*

Assume that for any open set $U \subset M$ with $\nu(U) > 0$ there exists $\phi \in \mathcal{E}$ such that $0 \leq \phi \leq \mathbb{1}_U$ and $\int_M \phi d\nu > 0$. For any observable $\varphi \in \mathcal{E}$, if φ is not ν-a.e. constant then Theorem 2.2 is applicable and the CLT holds.

Proof. The strong mixing condition and the density of $\mathcal{E}$ in $C_b(M)$ imply the uniqueness of the K-stationary measure, which in turn imply the ergodicity of the Markov system. Indeed, if $\tilde{\nu}$ is a K-stationary measure, then for any $\varphi \in C_b(M)$ we have $\int \mathcal{Q}^n\varphi \, d\tilde{\nu} = \int \varphi \, d\tilde{\nu}$ for all $n \in \mathbb{N}$. By strong mixing, for any $\varphi \in \mathcal{E}$ we have that $\mathcal{Q}^n\varphi \to \int \varphi d\nu$ uniformly. Integrating with respect to $\tilde{\nu}$ we conclude that $\int \varphi \, d\tilde{\nu} = \int \varphi \, d\nu$ for all $\varphi \in \mathcal{E}$, so for all $\varphi \in C_b(M)$, which shows that $\tilde{\nu} = \nu$.

Let $\varphi \in \mathcal{E}$ be a non ν-a.e. constant observable. We may of course assume that $\int \varphi d\nu = 0$, otherwise we consider $\varphi - \int \varphi d\nu$.

Let $\psi := \sum_{n=0}^{\infty} \mathcal{Q}^n\varphi$. Since $\varphi \in C_b(M)$, the strong mixing assumption on $\mathcal{Q}$ implies (via the Weierstrass M-test) that $\psi \in C_b(M)$ as well. It remains to show that $\sigma^2(\varphi) > 0$ which ensures the applicability of Theorem 2.2.

Assume by contradiction that $\sigma^2(\varphi) = \|\psi\|_2^2 - \|\mathcal{Q}\psi\|_2^2 = 0$. Then

$$\begin{aligned} 0 &\leq \int ((\mathcal{Q}\psi)(x) - \psi(y))^2 \, dK_x(y) \, d\nu(x) \\ &= \int \left\{ ((\mathcal{Q}\psi)(x))^2 + \psi(y)^2 - 2\,\psi(y)\,(\mathcal{Q}\psi)(x) \right\} dK_x(y)\, d\nu(x) \\ &= \int \left\{ \psi(y)^2 - ((\mathcal{Q}\psi)(x))^2 \right\} dK_x(y)\, d\nu(x) \\ &= \int \psi(y)^2 \, dK_x(y)\, d\nu(x) - \int ((\mathcal{Q}\psi)(x))^2 \, d\nu(x) \\ &= \|\psi\|_2^2 - \|\mathcal{Q}\psi\|_2^2 = 0 \quad (\text{since } \nu \text{ is } K-\text{stationary}). \end{aligned}$$

Therefore, $\psi(y) = \mathcal{Q}\psi(x)$ for ν-a.e. $x \in M$ and K_x-a.e. $y \in M$. By induction we obtain that for all $n \geq 1$,

$$\psi(y) = (\mathcal{Q}^n\psi)(x) \quad \text{for } \nu\text{-a.e. } x \in M \text{ and for } K_x^n\text{-a.e. } y \in M,$$

which implies that for all $n \geq 1$ and for ν-a.e. $x \in M$, the function ψ is K_x^n-a.e. constant. Let us show that in fact ψ is ν-a.e. constant.

If ψ is not ν-a.e constant, then there exist two disjoint open subsets U_1 and U_2 of M such that $\nu(U_1), \nu(U_2) > 0$ and $\psi|_{U_1} < \psi|_{U_2}$. By the assumption, there are two observables $\phi_1, \phi_2 \in \mathcal{E}$ such that $0 \leq \phi_i \leq \mathbb{1}_{U_i}$ and $\displaystyle\int \phi_i \, d\nu > 0$ for $i = 1, 2$. Moreover, for all $x \in M$ and $n \geq 1$,

$$K_x^n(U_i) = (\mathcal{Q}^n \mathbb{1}_{U_i})(x) \geq (\mathcal{Q}^n\phi_i)(x) \to \int \phi_i d\nu > 0,$$

where the above convergence as $n \to \infty$ is uniform in $x \in M$.

Thus for a large enough integer n and for all $x \in M$, both sets U_1 and U_2 have positive K_x^n measure. However, $\psi|_{U_1} < \psi|_{U_2}$, which contradicts the fact that ψ is K_x^n-a.e. constant for ν-a.e. $x \in M$.

We conclude that ψ is ν-a.e constant. Since ν is K-stationary it follows that $\varphi = \psi - \mathcal{Q}\psi = 0$ ν-a.e, which is a contradiction. $\square$

We note that Theorem 2.2 holds not only for the probability $\mathbb{P} = \mathbb{P}_\nu$, but also for the probability $\mathbb{P}_{x_0}$ corresponding to the Markov chain starting from ν-a.e. point $x_0 \in M$ (see the comments after Definition 1.1 in [5]). Then Proposition 2.2 and all of its consequences, also hold w.r.t. these measures.

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 2.4. 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 kernel $K : \mathbb{T} \to \text{Prob}(\mathbb{T})$ 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 polynomial 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 satisfies 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 $A : X \to \mathrm{GL}_2(\mathbb{R})$ 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 $\varphi_n(x) := \log \|A^n(x)\|$, 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,

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

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 a compact metric space 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 assumptions, 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 $\{l_j\}_{1\leq j\leq n}$ 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 $\{h_n\}_{n\geq 1} \subset \mathcal{H}_A$ 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 $\#\mathrm{supp}(\mu) \geq 2$ (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 kernel $\bar{K}$ 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 if A is quasi-irreducible, then 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 kernel 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 kernel $K : \mathbb{P} \to \text{Prob}(\mathbb{P})$:

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

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.*

$$\begin{array}{ccc} C^0(\Sigma \times \mathbb{P}) & \xrightarrow{\bar{Q}} & C^0(\Sigma \times \mathbb{P}) \\ {\scriptstyle \Pi}\downarrow & & \downarrow{\scriptstyle \Pi} \\ C^0(\mathbb{P}) & \xrightarrow{Q} & C^0(\mathbb{P}) \end{array}$$

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. ☐

Recall that given a Markov kernel $K : M \to \text{Prob}(M)$, a measure $\eta \in \text{Prob}(M)$ 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 $\eta \in \text{Prob}_K(M)$. 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 $\eta \in \text{Prob}(\mathbb{P})$, 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

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

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. □

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. $D := \text{Prob}(M)$ 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 $D := \text{Prob}_K(M)$ 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 $\text{Prob}_K(\mathbb{P})$ 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

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

That is, $|\varphi| \le Q\,|\varphi|$, 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 $\psi \equiv \mathrm{const}$, $\mu^{\mathbb{N}} \times \eta$-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 $\varphi \equiv \mathrm{const}$, 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.

□

As a corollary, we have

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

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

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

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

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. □

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. □

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,

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

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

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

Equivalently,

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

and thus

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

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 $\eta \in \text{Prob}_K(\mathbb{P}) \Leftrightarrow \mu \times \eta \in \text{Prob}_{\bar{K}}(\Sigma \times \mathbb{P})$. In fact, we shall see that if $m \in \text{Prob}_{\bar{K}}(\Sigma \times \mathbb{P})$, then $\exists \eta \in \text{Prob}_K(\mathbb{P})$ s.t. $m = \mu \times \eta$. Let us acknowledge this for now and later we will prove it as a lemma.

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

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

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

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

(1) We have

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

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 $\text{Prob}_K(\mathbb{P})$. Then necessarily, $\alpha(\eta) \equiv \beta, \forall\, \eta \in \text{Prob}_K(\mathbb{P})$.

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

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

Thus $\mathcal{M} = \text{Prob}_K(\mathbb{P})$. This shows $\alpha(\eta) = \beta$, $\forall\, \eta \in \text{Prob}_K(\mathbb{P})$. 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

$$\begin{aligned} \|A^n(\omega)(av_1 + bv_2)\| &\le |a|\, \|A^n(\omega)v_1\| + |b|\, \|A^n(\omega)v_2\| \\ &\le \max\{|a|\, \|A^n(\omega)v_1\|, |b|\, \|A^n(\omega)v_2\|\}. \end{aligned}$$

Take "$1/n \log$" on both sides and let $n \to \infty$ (taking lim sup),

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

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 $\eta_- \in \text{Prob}_K(\mathbb{P})$ 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 $\alpha(\eta) \equiv \beta$, $\forall\, \eta \in \text{Prob}_K(\mathbb{P})$.

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

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

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,

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

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

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

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,

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

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

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. □

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. □

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. □

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 $\mathcal{K}_\alpha(A^n, \mu^n) \leq \kappa < 1$. □

Remark 3.6. In practice, the advantage of our method here is that we can give a precise lower bound of the Hölder exponent α for any specific admissible model (A, μ). To achieve this, just choose the first $n \in \mathbb{N}$ such that

$$\log \frac{|\det A^n(\omega)|}{\|A^n(\omega)v\|^2} \leq -1.$$

for any starting point $v \in \mathbb{S}^1$ (this arbitrary choice of v is ensured by uniform convergence) and then use n to determine α by making the desired term smaller than 1. Namely, our α is exactly computable!

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

$$v_\alpha(Q_A^m(\varphi)) \leq [\mathcal{K}_\alpha(A^n, \mu^n)]^k \mathcal{K}_\alpha(A^r, \mu^r) v_\alpha(\varphi) \leq \kappa^k \cdot L v_\alpha(\varphi).$$

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

$$v_\alpha \left(Q_A^n(\varphi) - \int \varphi d\eta \right) \leq v_\alpha \left(Q_A^n(\varphi) \right) \leq C\sigma^n v_\alpha(\varphi) \leq C\sigma^n \left\| \varphi \right\|_\alpha .$$

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 $\eta \in \mathrm{Prob}_K(\mathbb{P})$ 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. □

By Lemma 5, we get that $\bar{Q}$ is strongly mixing on $\mathcal{E} = \mathcal{H}_\alpha(\Sigma \times \mathbb{P}) := \{\varphi \in C^0(\Sigma \times \mathbb{P}) : v_\alpha(\Pi\varphi) < \infty\}$. 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 $\mu \in \mathrm{Prob}(\mathbb{T}^d)$. 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 kernel 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,

$$\varphi(x) \approx \sum_{k=-\infty}^{+\infty} \hat{\varphi}(k) e^{2\pi i k x}$$

where the r.h.s. 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) If φ is Hölder continuous, then

$$\varphi(x) = \sum_{k=-\infty}^{\infty} \hat{\varphi}(k) e^{2\pi i k x}, \quad \forall\, x \in \mathbb{T}.$$

(4) If $\varphi \in L^p(\mathbb{T})$ with $p > 1$, then by Carleson's Theorem

$$\varphi(x) = \sum_{k=-\infty}^{\infty} \hat{\varphi}(k) e^{2\pi i k x}, \quad m\text{-a.e.}\, x \in \mathbb{T}.$$

We also recall estimates on the size of the Fourier coefficients. For example, it is clear that $|\hat{\varphi}(k)| \leq \|\varphi\|_1, \forall\, k \in \mathbb{Z}$ and $\hat{\varphi}(0) = \int_0^1 \varphi(x)dx$. We also have the Riemann-Lebesgue lemma showing that if $\varphi \in L^1$, then $\hat{\varphi}(k) \to 0$ as $|k| \to \infty$. Moreover, if φ is α-Hölder, then $|\hat{\varphi}(k)| \leq C \cdot \frac{1}{|k|^\alpha}$ where $C \sim \|\varphi\|_\alpha$.

There are other facts which concentrate on the approximation property. By Weierstrass approximation theorem, every continuous function is uniformly approximated by trigonometric polynomials of the form

$$p = \sum_{k=-n}^{n} c_k e^{2\pi i k x}, \quad c_k \in \mathbb{C}, \quad \deg p \leq n.$$

If φ is Hölder, we have the following theorem.

Theorem 4.2. *If φ is α-Hölder continuous, then $\forall\, n \in \mathbb{N}$, $\exists\, p_n$ which are trigonometric polynomials with $\deg p_n \leq n$ such that*

$$\|\varphi - p_n\|_\infty \lesssim \|\varphi\|_\alpha \frac{1}{n^\alpha}.$$

Moreover, $\forall\, k \in \mathbb{Z}, |\hat{p}_n(k)| \leq |\hat{\varphi}(k)|$.

4.2. **Mixing measures.** Let $\mu \in \mathrm{Prob}(\mathbb{T})$. We consider the Markov chain on $\mathbb{T}$:

$$\theta \to \theta + \omega_0 \to \theta + \omega_0 + \omega_1 \to \cdots$$

The corresponding Markov kernel K is

$$K_\theta = \int_{\mathbb{T}} \delta_{\theta+\omega_0} d\mu(\omega_0).$$

The corresponding Markov operator is

$$Q = Q_\mu : L^1(\mathbb{T}) \to L^1(\mathbb{T}), \quad Q\varphi(\theta) = \int_{\mathbb{T}} \varphi(\theta + \omega_0) d\mu(\omega_0).$$

Note that Q is bounded on $L^1(\mathbb{T}, m), L^2(\mathbb{T}, m)$ and $L^\infty(\mathbb{T}, m)$ because m is translation invariant, which also ensures that m is K-stationary. Hence $(\mathbb{T}, K, m)$ is a Markov system.

Our first goal is to show that the Markov operator Q of the Markov system $(\mathbb{T}, K, m)$ is strongly mixing under certain assumptions on μ and for an appropriate space of observables.

Let us make some preparations first. Recall that the Fourier coefficients of a measure $\mu \in \text{Prob}(\mathbb{T})$ is defined by

$$\hat{\mu}(k) := \int_{\mathbb{T}} e^{2\pi ikx} d\mu(x).$$

It is equivalent if we put "$e^{-2\pi ikx}$" in the definition. Observe that if $\mu \ll m$, namely $d\mu = hdm$ with $h \geq 0$ and $\int hdm = 1$, then

$$\hat{\mu}(k) = \int_{\mathbb{T}} e^{-2\pi ikx} h(x)dm(x) = \hat{h}(-k).$$

We call $e_k(x) := e^{2\pi ikx} : \mathbb{T} \to \mathbb{C}$, $k \in \mathbb{Z}$ "characters". It is clear that they are group homomorphisms.

Lemma 4.1. *The characters $\{e_k, k \in \mathbb{Z}\}$ form a complete basis of eigenvectors for the Markov operator $Q : L^2(\mathbb{T}) \to L^2(\mathbb{T})$. That is, $Qe_k = \hat{\mu}(k)e_k, \forall\, k \in \mathbb{Z}$ and if $\varphi = \sum_{k=-\infty}^{\infty} \hat{\varphi}(k)e_k$ in $L^2(\mathbb{T}, m)$, then*

$$Q\varphi = \sum_{k=-\infty}^{\infty} \hat{\mu}(k)\hat{\varphi}(k)e_k, \text{ in } L^2(\mathbb{T}, m).$$

Proof. By the linearity of Q, it is enough to prove the first equality. For any $\theta \in \mathbb{T}$ and any $k \in \mathbb{Z}$, we have

$$\begin{aligned} Qe_k(\theta) &= \int_{\mathbb{T}} e_k(\theta + \omega_0)d\mu(\omega_0) \\ &= \int_{\mathbb{T}} e_k(\theta)e_k(\omega_0)d\mu(\omega_0) \\ &= e_k(\theta) \int e_k d\mu = e_k(\theta)\hat{\mu}(k). \end{aligned}$$

Thus the result follows. □

Remark 4.2. It turns out that the mixed model $(\Sigma^{\mathbb{Z}} \times \mathbb{T}, f, \mu^{\mathbb{Z}} \times m)$ is ergodic if and only if $\hat{\mu}(k) \neq 1, \forall\, k \in \mathbb{Z}\backslash\{0\}$, if and only if $\forall\, \varphi \in C^0(\mathbb{T})$

$$\frac{1}{n}\sum_{j=0}^{n-1} Q^j\varphi(\theta) \to \int \varphi dm \quad \text{as} \quad n \to \infty, \forall\, \theta \in \mathbb{T}.$$

For completion, we borrow all the equivalent conditions of ergodicity from [1]. To make it consistent with our pedagogical context, one can simply let $d = 1$ in the following theorem.

Theorem 4.3. *Let $\mu \in \mathrm{Prob}(\Sigma)$ where $\Sigma = \mathbb{T}^d$, and consider the skew product map on $\Sigma^{\mathbb{Z}} \times \mathbb{T}^d$ given by $f(\{\beta_i\}, \theta) = (\sigma\{\beta_i\}, \theta + \beta_0)$. The following statements are equivalent:*

(1) f is ergodic w.r.t. $\mu^{\mathbb{Z}} \times m$;
(2) f is ergodic w.r.t. $\mu^{\mathbb{N}} \times m$;
(3) Every m-stationary observable $\varphi \in L^\infty(\mathbb{T}^d)$ is constant m-a.e.;
(4) $\hat{\mu}(k) \neq 1$ for every $k \in \mathbb{Z}^d \setminus \{0\}$;
(5) For every $k \in \mathbb{Z}^d \setminus \{0\}$ there exists $\alpha \in \Sigma$ such that $\langle k, \alpha \rangle \notin \mathbb{Z}$;
(6) $\mathbb{T}^d = \overline{\cup_{n \geq 1} \Sigma^n}$ where $\Sigma = \mathrm{supp}(\mu)$ and $\Sigma^n := \Sigma + \Sigma^{n-1} \quad \forall n \geq 2$;
(7) m is the unique μ-stationary measure in $\mathrm{Prob}(\mathbb{T}^d)$,
(8) $\lim_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} (\mathcal{Q}_\mu^j \varphi)(\theta) = \int_{\mathbb{T}^d} \varphi \, dm, \quad \forall\, \theta \in \mathbb{T}^d \quad \forall \varphi \in C^0(\mathbb{T}^d)$.

Proof. (1) $\Rightarrow$ (2) holds trivially because f in (2) is a factor f in (1), i.e., because of the commutativity of the following diagram of measure preserving transformations.

$$\begin{array}{ccc} X & \xrightarrow{\;f\;} & X \\ {\scriptstyle\pi}\big\downarrow & & \big\downarrow{\scriptstyle\pi} \\ X^+ & \xrightarrow{\;f\;} & X^+ \end{array}$$

Conversely, (2) $\Rightarrow$ (1) holds by Lemma 5.3.1 in [6].
The equivalence (2) $\Leftrightarrow$ (3) follows from Proposition 5.13 in [7].

Given a bounded measurable function $\varphi : \mathbb{T}^d \to \mathbb{C}$, we have $\varphi \in L^2(\mathbb{T}^d, m)$. Consider its Fourier series

$$\varphi = \sum_{k \in \mathbb{Z}^d} \hat{\varphi}(k)\, e_k \quad \text{with} \quad e_k(\theta) := e^{2\pi i \langle k, \theta \rangle}.$$

A simple calculation shows that

$$\mathcal{Q}_\mu \varphi = \sum_{k \in \mathbb{Z}^d} \hat{\mu}(k)\, \hat{\varphi}(k)\, e_k.$$

(3) $\Rightarrow$ (4): If $\hat{\mu}(k) = 1$ for some $k \in \mathbb{Z}^d \setminus \{0\}$, then e_k is a non constant m-stationary observable. In other words, if (4) fails then so does (3).

(4) $\Rightarrow$ (3): Given φ m-stationary, comparing the two Fourier developments above, for all $k \in \mathbb{Z}^d$ $\hat{\mu}(k)\, \hat{\varphi}(k) = \hat{\varphi}(k) \Leftrightarrow \hat{\varphi}(k)\, (\hat{\mu}(k) - 1) = 0$. By (4) we then get $\hat{\varphi}(k) = 0$ for all $k \in \mathbb{Z}^d \setminus \{0\}$, which implies that $\varphi = \hat{\varphi}(0)$ is m-a.e. constant. This proves (3).

Since $\hat{\mu}(k)$ is an average of a continuous function with values on the unit circle, we have

$$\hat{\mu}(k) = 1 \quad \Leftrightarrow \quad e^{2\pi i \langle k, \alpha \rangle} = 1, \ \forall \alpha \in \Sigma \quad \Leftrightarrow \quad \langle k, \alpha \rangle \in \mathbb{Z}, \ \forall \alpha \in \Sigma.$$

This proves that (4) $\Leftrightarrow$ (5).

(5) $\Rightarrow$ (6): Let $H = \overline{\cup_{n\geq 1} \Sigma^n}$ and assume that $H \neq \mathbb{T}^d$. By definition H is a subsemigroup of $\mathbb{T}^d$. By Poincaré recurrence theorem, H is also a group. By Pontryagin's duality for locally compact abelian groups, there exists a non trivial character $e_k : \mathbb{T}^d \to \mathbb{C}$ which contains H in its kernel. In particular this implies that there exists $k \in \mathbb{Z}^d \setminus \{0\}$ such that $\langle k, \beta \rangle \in \mathbb{Z}$ for all $\beta \in \Sigma$. This argument shows that if (6) fails then so does (5).

(6) $\Rightarrow$ (5): Assume that (5) does not hold, i.e., for some $k \in \mathbb{Z}^d \setminus \{0\}$ we have $\langle k, \alpha \rangle \in \mathbb{Z}$ for all $\alpha \in \Sigma$. Then e_k is a non trivial character of $\mathbb{T}^d$ and $H := \{\theta \in \mathbb{T}^d \colon e_k(\theta) = 1\}$ is a proper sub-torus, i.e. a compact subgroup of $\mathbb{T}^d$. The assumption implies that $\Sigma \subset H$, and since H is a group, $S^n \subset H$, $\forall n \geq 1$. This proves that (6) fails.

Since the adjoint operator $\mathcal{Q}_\mu^* : \text{Prob}(\mathbb{T}^d) \to \text{Prob}(\mathbb{T}^d)$ satisfies $\mathcal{Q}_\mu^* \pi = \mu * \pi$, denoting by $\mu^{*j} := \mu * \cdots * \mu$ the j-th convolution power of μ, we have $(\mathcal{Q}_\mu^*)^n \delta_0 = \mu^{*n} \ \ \forall n \in \mathbb{N}$.

Lemma 4.2. *Any sublimit of the sequence* $\pi_n := \frac{1}{n} \sum_{j=0}^{n-1} \mu^{*j}$ *is a* μ-*stationary measure.*

Proof. Given $\varphi \in C^0(\mathbb{T}^d)$,

$$\begin{aligned}
\langle \mathcal{Q}_\mu \varphi - \varphi, \pi_n \rangle &= \frac{1}{n} \sum_{j=0}^{n-1} \langle \mathcal{Q}_\mu \varphi - \varphi, (\mathcal{Q}_\mu^*)^j \delta_0 \rangle \\
&= \frac{1}{n} \sum_{j=0}^{n-1} (\mathcal{Q}_\mu^{j+1} \varphi)(0) - (\mathcal{Q}_\mu^j \varphi)(0) \\
&= \frac{1}{n} \left((\mathcal{Q}_\mu^n \varphi)(0) - \varphi(0) \right) = \mathcal{O}(\frac{1}{n}).
\end{aligned}$$

Hence, if $\pi \in \text{Prob}(\mathbb{T}^d)$ is a sublimit of π_n, taking the limit along the corresponding subsequence of integers we have

$$\langle \varphi, \mathcal{Q}_\mu^* \pi - \pi \rangle = \langle \mathcal{Q}_\mu \varphi - \varphi, \pi \rangle = 0,$$

which implies that $\mathcal{Q}_\mu^* \pi = \pi$. $\square$

(2) $\Rightarrow$ (8): By ergodicity of f w.r.t. $\mu^{\mathbb{N}} \times m$ and Birkhoff Ergodic Theorem, given $\varphi \in C^0(\mathbb{T}^d)$ there exists a full measure set of $(\omega, \theta) \in \Sigma^{\mathbb{N}} \times \mathbb{T}^d$ with

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(\theta + \tau^j(\omega)) = \int \varphi \, dm,$$

where $\tau^j(\omega) = \omega_0 + \cdots + \omega_{j-1}$ and $\omega = \{\omega_j\}_{j\in\mathbb{N}}$. Hence there exists a Borel set $\mathcal{B} \subset \mathbb{T}^d$ with $m(\mathcal{B}) = 1$ such that, applying the Dominated Convergence Theorem, we have for all $\theta \in \mathcal{B}$,

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} (\mathcal{Q}_\mu^j \varphi)(\theta) = \int \varphi \, dm.$$

The set $\mathcal{B}$ depends on the continuous function φ, but since the space $C^0(\mathbb{T}^d)$ is separable we can choose this Borel set $\mathcal{B}$ so that the previous limit holds for every $\theta \in \mathcal{B}$ and $\varphi \in C^0(\mathbb{T}^d)$. This implies the following weak* convergence in $\mathrm{Prob}(\mathbb{T}^d)$:

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} (\mathcal{Q}_\mu^*)^j \delta_\theta = m.$$

Given any $\theta' \notin \mathcal{B}$ take $\theta \in \mathcal{B}$. Convolving both sides on the right by $\delta_{\theta'-\theta}$ we get

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} (\mathcal{Q}_\mu^*)^j \delta_{\theta'} = \lim_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu^{*j} * \delta_\theta * \delta_{\theta'-\theta} = m * \delta_{\theta'-\theta} = m,$$

which proves (8).

(8) $\Rightarrow$ (7): If there exists $\eta \neq m$ in $\mathrm{Prob}_\mu(\mathbb{T}^d)$, then there exists at least one more ergodic measure $\zeta \neq m$ such that ζ is an extreme point of $\mathrm{Prob}_\mu(\mathbb{T}^d)$. Choosing $\varphi \in C^0(\mathbb{T}^d)$ such that $\int \varphi \, d\zeta \neq \int \varphi \, dm$, by Birkhoff Ergodic Theorem there exists $\theta \in \mathbb{T}^d$ such that

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} (\mathcal{Q}_\mu^j \varphi)(\theta) = \int \varphi \, d\zeta \neq \int \varphi \, dm.$$

which contradicts (8).

(7) $\Rightarrow$ (6): Consider the compact subgroup $H := \overline{\cup_{n\geq 1} \Sigma^n}$. If (6) fails then $H \neq \mathbb{T}^d$ and by Lemma 4.2 we can construct a stationary measure $\pi \in \mathrm{Prob}_\mu(\mathbb{T}^d)$ with $\mathrm{supp}(\pi) \subset H$. This shows that $\pi \neq m$ and hence there is more than one stationary measure. $\square$

In fact, we need stronger conditions than ergodicity of μ to prove that Q is strongly mixing.

We move forward step by step. One concept stronger than ergodicity is mixing.

Definition 4.1. (Q, m) is called mixing if $\forall\, \varphi \in C^0(\mathbb{T})$,

$$Q^n\varphi(\theta) \to \int \varphi dm \quad \text{as} \quad n \to \infty, \quad \forall\, \theta \in \mathbb{T}.$$

It is clear that the mixing of (Q, m) implies the ergodicity of f, like the similar statement in dynamical systems. Moreover, mixing has some equivalent characterizations as follows.

Theorem 4.4. *The following statements are equivalent.*

(1) (Q, m) *is mixing.*
(2) $|\hat{\mu}(k)| < 1, \forall\, k \in \mathbb{Z}\backslash\{0\}$.
(3) $\forall\, k \in \mathbb{Z}\backslash\{0\}$, $\exists\, \alpha \neq \beta \in \operatorname{supp}(\mu)$ *such that* $k(\alpha - \beta) \notin \mathbb{Z}$.

Corollary 4.3. *If* $\exists\, \alpha \neq \beta \in \operatorname{supp}(\mu)$ *such that* $\beta - \alpha \notin \mathbb{Q}$, *then* (Q, m) *is mixing.*

Proof. (1) $\Rightarrow$ (2). If $\exists\, k \in \mathbb{Z}\backslash\{0\}$ such that $|\hat{\mu}(k)| = 1$, then since $Q^n e_k = \hat{\mu}(k)^n e_k, \forall\, n \in \mathbb{N}$, we have

$$|Q^n e_k| = |\hat{\mu}(k)^n e_k| = 1 \nrightarrow 0 = \int e_k dm \;\text{ as }\; n \to \infty.$$

This contradicts the mixing condition.

(2) $\Rightarrow$ (1). First step. Let $p = \sum_{k=-N}^{N} c_k e_k$ be a trigonometric polynomial. Note that $\int pdm = c_0$ and $\hat{\mu}(0) = 1$, so we have

$$Q^n p - \int pdm = \sum_{0<|k|\leq N} c_k \hat{\mu}(k)^n e_k.$$

Hence

$$\left\| Q^n p - \int pdm \right\|_\infty \leq \sum_{0<|k|\leq N} |c_k|\, |\hat{\mu}(k)|^n$$

Let $\sigma = \max\{\hat{\mu}(k) : 0 < |k| \leq N\} < 1$. Then

$$\left\| Q^n p - \int pdm \right\|_\infty \leq 2N \left\| p \right\|_\infty \sigma^n \to 0 \;\text{ as }\; n \to 0.$$

Second step. Given any $\varphi \in C^0(\mathbb{T}), \epsilon > 0$, by Weierstrass approximation theorem $\exists\, p$ a trigonometric polynomial such that $\|\varphi - p\|_\infty < \epsilon$. Moreover, by the first step $\exists\, n \in \mathbb{N}$ such that $\left\|Q^n p - \int pdm\right\|_\infty < \epsilon$.

Writing $\varphi = p + \varphi - p$, then

$$Q^n\varphi = Q^n p + Q^n(\varphi - p)$$

and

$$\int \varphi dm = \int pdm + \int (\varphi - p)dm.$$

Therefore,

$$\begin{aligned}\left\| Q^n\varphi \int \varphi dm \right\|_\infty \leq& \left\| Q^n p - \int pdm \right\|_\infty + \|\varphi - p\|_\infty + \|Q^n(\varphi - p)\|_\infty \\ \leq& \epsilon + \epsilon + \epsilon = 3\epsilon.\end{aligned}$$

This proves mixing.

(2) $\Leftrightarrow$ (3). It holds if and only if $\exists\, k \in \mathbb{Z}\backslash\{0\}$, $|\hat{\mu}(k)| = 1 \Leftrightarrow \exists\, k \in \mathbb{Z}\backslash\{0\}$, $\forall\, \alpha \neq \beta \in \mathrm{supp}(\mu)$ s.t. $e^{2\pi ik\alpha} = e^{2\pi ik\beta}$. It is further equivalent to $\left|\int e^{2\pi ik\alpha} d\mu(\alpha)\right| = 1$ if and only if $e^{2\pi ik\alpha}$ is constant for μ-a.e. α, which this is obvious. □

For the sake of the readers, we give a more general lemma here clarifying the last part of the proof above. In fact, it will also be used later.

Lemma 4.4. *Let (Ω, ρ) be a probability space. Assume $f : \Omega \to \mathbb{C}$ is Lebesgue integrable. If $\left|\int_\Omega f d\rho\right| = \int_\Omega |f|\, d\rho$, then $\arg f$ is constant ρ-a.e. In other words, $\exists\, \theta_0 \in \mathbb{R}$ such that $f(x) = e^{i\theta_0} |f(x)|$ for ρ-a.e. $x \in \Omega$.*

Proof. By assumptions, $\int_\Omega f d\rho \in \mathbb{C}$. Let $\theta_0 := \arg(\int_\Omega f d\rho)$. Then

$$\int_\Omega f d\rho = e^{i\theta_0} \left| \int_\Omega f d\rho \right|.$$

Then,

$$\begin{aligned}0 =& \left| \int_\Omega f d\rho \right| - \int_\Omega f d\rho \\ =& e^{-i\theta_0} \int_\Omega f d\rho - \int_\Omega |f|\, d\rho \\ =& \int [e^{-i\theta_0} f - \left|e^{-i\theta_0} f\right|] d\rho \\ =& \Re \int_\Omega [e^{-i\theta_0} f - \left|e^{-i\theta_0} f\right|] d\rho \\ =& \int_\Omega [\Re(e^{-i\theta_0} f) - \left|e^{-i\theta_0} f\right|] d\rho \\ \leq& 0.\end{aligned}$$

This implies $\Re(e^{-i\theta_0}f) = \left|e^{-i\theta_0}f\right| \geq 0$, ρ-a.e. In particular, $\Im(e^{-i\theta_0}f) = 0$, ρ-a.e. Therefore,

$$e^{-i\theta_0}f = \Re(e^{-i\theta_0}f) = \left|e^{-i\theta_0}f\right| = |f|\,,\ \rho\text{-a.e.}$$

which gives $f = e^{i\theta_0}\,|f|$, ρ-a.e. □

In the following, we are going to prove two propositions in which Q is strongly mixing with different rates under different assumptions of the measure $\mu \in \text{Prob}(\mathbb{T})$. It turns out that μ being mixing is also not enough for our purpose. So we will consider the so-called mixing Diophantine measures (to be defined later) and absolutely continuous measures. In fact, absolute continuity implies mixing Diophantine, but to warm-up, let us first assume that $\mu \ll m \Leftrightarrow d\mu = hdm$ with $h \geq 0$ and $\int hdm = 1$.

Lemma 4.5. *If $\mu \ll m$ then $\exists\, \sigma_0 \in (0,1)$ such that $|\hat{\mu}(k)| \leq \sigma_0, \forall\, k \in \mathbb{Z}\backslash\{0\}$.*

Proof. By definition, $\hat{\mu}(k) = \int e^{2\pi ikx}d\mu(x)$. Then by Lemma 4.4, we have $\hat{\mu}(k) < 1, \forall\, k \in \mathbb{Z}\backslash\{0\}$. Otherwise $e^{2\pi ikx} = e^{2\pi ik\theta_0}$ for some $\theta_0 \in \mathbb{T}, \mu$-a.e. which contradicts $\mu \ll m$ (by Riemann-Lebesgue).

Since $\mu \ll m$, by Riemann-Lebesgue we have

$$|\hat{\mu}(k)| = \left|\hat{h}(-k)\right| \to 0 \text{ as } |k| \to \infty.$$

Hence $\exists\, N \in \mathbb{N}$ such that $|\hat{\mu}(k)| \leq \frac{1}{2}, \forall\, |k| > N$. Let

$$\sigma_0 := \max\left\{\frac{1}{2}, |\hat{\mu}(j)| : 0 < |j| \leq N\right\} < 1.$$

Then $\hat{\mu}(k) \leq \sigma_0, \forall\, k \in \mathbb{Z}\backslash\{0\}$. □

With this lemma in hand, we can prove the following proposition.

Proposition 4.6. *If $\mu \ll m$, then Q is strongly mixing with exponential rate on any space of Hölder continuous functions $\mathcal{H}_\alpha(\mathbb{T}), \forall\, \alpha \in (0,1)$. That is, $\exists\, C < \infty, \sigma \in (0,1)$ such that*

$$\left\|Q^n\varphi - \int \varphi dm\right\|_\infty \leq C\,\|\varphi\|_\alpha\,\sigma^n, \quad \forall\, \varphi \in \mathcal{H}_\alpha(\mathbb{T}).$$

In fact, $\sigma = {\sigma_0}^\alpha$.

Before we formally give the proof, we first describe the **attempt** to prove it.

Recall that $e_k(\theta) = e^{2\pi i k\theta}$, hence $Qe_k = \hat{\mu}(k)e_k$ and $Q^n e_k = \hat{\mu}(k)^n e_k$. If we write $\varphi \in \mathcal{H}_\alpha(\mathbb{T})$ as a Fourier series

$$\varphi = \sum_{k=-\infty}^{\infty} \hat{\varphi}(k) e_k.$$

Then

$$Q^n \varphi = \sum_{k=-\infty}^{\infty} \hat{\varphi}(k)\hat{\mu}(k)^n e_k = \sum_{k\neq 0} \hat{\varphi}(k)\hat{\mu}(k)^n e_k + \int \varphi dm.$$

Then (we hope to have)

$$\left\| Q^n \varphi - \int \varphi dm \right\|_\infty \leq \sum_{k\neq 0} |\hat{\varphi}(k)| \, |\hat{\mu}(k)|^n .$$

In general this is not allowed because the infinity sum is not necessarily absolutely convergent.

Since $\mu \ll m$, $|\hat{\mu}(k)| \leq \sigma_0 < 1, \forall\, k \neq 0$. We have

$$\left\| Q^n \varphi - \int \varphi dm \right\|_\infty \leq \sigma_0^n \sum_{k\neq 0} |\hat{\varphi}(k)| .$$

There is still a problem of an infinite sum which may not converge unless we assume some sufficient conditions like $\varphi \in C^{1+\epsilon}$. However, we want to deal with less regularities like Lipschitz and Hölder. Therefore, we really have to change our minds and use approximations of trigonometric polynomials. Let us start the formal proof.

Proof. Fix n to be the number of "iterations". Let N be the degree of approximation which will be chosen later. Since $\varphi \in \mathcal{H}_\alpha(\mathbb{T})$, by Theorem 4.2, there exists p_N trigonometric polynomial of $\deg p_N \leq N$ such that $\|\varphi - p_N\|_\infty \lesssim \|\varphi\|_\alpha \frac{1}{N^\alpha}$. In fact, p_N is the convolution of φ with the Jackson kernel,

$$p_N = \sum_{k=-N}^{N} c_k e_k, \; |c_k| \leq |\hat{\varphi}(k)| \lesssim \|\varphi\|_\alpha \frac{1}{|k|^\alpha}.$$

Therefore, we can write $\varphi = p_N + (\varphi - p_N) := p_N + r_N$. So by linearity we have

$$Q^n \varphi = Q^n p_N + Q^n r_N$$

and

$$\int \varphi dm = \int p_N dm + \int r_N dm.$$

Thus

$$Q^n\varphi - \int \varphi dm = Q^n p_N - \int p_N dm + Q^n r_N - \int r_N dm,$$

which shows that

$$\left\| Q^n\varphi - \int \varphi dm \right\|_\infty \le \left\| Q^n p_N - \int p_N dm \right\|_\infty + \|Q^n r_N\|_\infty + \int |r_N|\, dm.$$

Due to the estimates on r_N, the second and third term in the r.h.s. are smaller than $C\|\varphi\|_\alpha \frac{1}{N^\alpha}$ for some constant $C > 0$. So let us estimate the first term. Since $p_N = \sum_{k=-N}^{N} c_k e_k$,

$$Q^n p_N = \sum_{k=-N}^{N} c_k \hat{\mu}(k)^n e_k.$$

This implies

$$\left\| Q^n p_N - \int p_N dm \right\|_\infty \le \sum_{0<|k|\le N} |c_k|\, |\hat{\mu}(k)|^n \lesssim \|\varphi\|_\alpha \sigma_0^n N^{1-\alpha}.$$

Combining the estimates above, we have

$$\left\| Q^n\varphi - \int \varphi dm \right\|_\infty \lesssim \|\varphi\|_\alpha [\sigma_0^n N + \frac{1}{N^\alpha}].$$

Choose $N \in \mathbb{N}$ such that $\sigma_0^n N^{1-\alpha} = \frac{1}{N^\alpha}$. Thus

$$N = (\frac{1}{\sigma_0})^n,$$

hence

$$N^{-\alpha} = [\sigma_0^\alpha]^n$$

Take $\sigma := \sigma_0^\alpha$, we conclude the proof. □

Remark 4.3. If we consider $\mathbb{T}^d, \forall\, d \ge 1$, then similar computation yields $\sigma = {\sigma_0}^{\frac{\alpha}{d}}$.

Up to now, we know that

(1) if $\mu \ll m$, then $|\hat{\mu}(k)| \le \sigma_0 < 1, \forall\, k \in \mathbb{Z}\backslash\{0\}$.
(2) if (Q, m) is mixing, then $|\hat{\mu}(k)| < 1, \forall\, k \in \mathbb{Z}\backslash\{0\}$.

For item (1), we already proved Proposition 4.6. However, it turns out that in order to obtain a similar but weaker proposition, (2) is not enough. For our purpose, we shall introduce the notion of mixing Diophantine.

Definition 4.2. We say $\mu \in \mathrm{Prob}(\mathbb{T})$ satisfies a mixing Diophantine Condition (mixing DC) if

$$|\hat{\mu}(k)| \leq 1 - \frac{\gamma}{|k|^\tau}, \; \forall\, k \in \mathbb{Z}\backslash\{0\},$$

for some $\gamma, \tau > 0$.

This is inspired by the concept of the Diophantine Condition (DC) for numbers. We say that $\alpha \in [0, 1)$ satisfies the Dionphantine Condition $\mathrm{DC}(\gamma, \tau)$ if

$$\inf_{j \in \mathbb{Z}} |k\alpha - j| \geq \frac{\gamma}{|k|^\tau}, \; \forall\, k \in \mathbb{Z}\backslash\{0\}.$$

Note that here $\gamma > 0$ but $\tau > 1$. This is because when $\tau = 1$, $\mathrm{DC}(\gamma, \tau)$ is of Lebesgue measure zero. If $\tau < 1$, $\mathrm{DC}(\gamma, \tau)$ is empty. For the mixing DC, any $\tau > 0$ is fine (because the space of probability measures on $\mathbb{T}$ is infinite dimensional).

We give some examples regarding mixing DC measures.

(1) If $\mu \ll m$, then μ is mixing DC with any $\tau \geq 0$.
(2) If $\mu = \delta_\alpha$ then $\hat{\mu}(k) = \int e^{2\pi i k x} d\delta_\alpha(x) = e^{2\pi i k \alpha}$. This shows $|\hat{\mu}(k)| \equiv 1, \forall\, k \in \mathbb{Z}$ which implies that δ_α is not mixing (DC).
(3) If $\mu = t\delta_\alpha + (1-t)\delta_\beta$ with $t \in (0, 1)$ and $\beta - \alpha \in \mathrm{DC}$, then μ is mixing DC.
(4) If $\mu \in \mathrm{Prob}(\mathbb{T})$ is finitely supported such that $\exists\, \alpha, \beta \in \mathrm{supp}(\mu)$ such that $\beta - \alpha \in \mathrm{DC}$, then μ is mixing DC.
(5) If μ_1 is mixing DC, for any $t \in (0, 1]$ and $\mu_2 \in \mathrm{Prob}(\mathbb{T})$, $\mu := t\mu_1 + (1-t)\mu_2$ is mixing DC.

Note that (5) implies that mixing DC measures are prevalent.

Now we can formulate our second proposition.

Proposition 4.7. *If μ is mixing DC with parameters γ and τ, then Q is strongly mixing with power rate on any space of Hölder continuous functions $\mathcal{H}_\alpha(\mathbb{T})$. More precisely, $\exists\, C < \infty, p > 0$ such that*

$$\left\| Q^n \varphi - \int \varphi dm \right\|_\infty \leq C \left\| \varphi \right\|_\alpha \frac{1}{n^p}, \; \forall\, \varphi \in \mathcal{H}_\alpha(\mathbb{T}), n \in \mathbb{Z}^+.$$

In fact, p can be chosen as close as we want to $\frac{\alpha}{\tau}$ from below.

Proof. The proof is exactly along the same line as that of Proposition 4.6. So we borrow the same notations: n, N, p_n, r_n from there. Moreover, given the other parameters γ, τ and α, we just need to consider the case when n and N are sufficiently large.

Write $\varphi = p_n + (\varphi - p_n)$, the same argument yields

$$\left\| Q^n\varphi - \int \varphi dm \right\|_\infty \leq \left\| Q^n p_N + \int p_N dm \right\|_\infty + \|Q^n r_N\|_\infty + \int |r_N|\, dm.$$

The last two terms are again bounded by $C\,\|\varphi\|_\alpha \frac{1}{N^\alpha}$ for some constant $C > 0$. The difference is on the first term. Now we have

$$\left\| Q^n p_N - \int p_N dm \right\|_\infty \leq \sum_{0<|k|\leq N} |c_k|\, |\hat{\mu}(k)|^n \lesssim \|\varphi\|_\alpha \sum_{0<|k|\leq N} \frac{1}{|k|^\alpha}(1-\frac{\gamma}{|k|^\tau})^n.$$

Using the inequality $(1-x)^{\frac{1}{x}} \leq e^{-1}, x > 0$, we have

$$(1-\frac{\gamma}{|k|^\tau})^n \leq (1-\frac{\gamma}{N^\tau})^n \leq e^{-\frac{n\gamma}{N^\tau}}.$$

We have to make $\frac{1}{N^\alpha} = N^{1-\alpha} \cdot e^{-\frac{n\gamma}{N^\tau}}$, thus we can take for example

$$N = (n\gamma)^{\frac{9}{10\tau}}$$

such that

$$\frac{1}{N^\alpha} \approx n^{-\frac{9\alpha}{10\tau}}, \quad N \cdot e^{-\frac{n\gamma}{N^\tau}} \approx n^{\frac{9}{10\tau}} \cdot e^{-(n\gamma)^{\frac{1}{10}}} \ll n^{-\frac{9\alpha}{10\tau}}.$$

In the end we have

$$\left\| Q^n p_N - \int p_N dm \right\|_\infty \leq C\,\|\varphi\|_\alpha \frac{1}{n^p}$$

with $p = \frac{9\alpha}{10\tau}$. Adjusting the parameter $\frac{9}{10}$ closer and closer to 1, we get $p \nearrow \frac{\alpha}{\tau}$. □

Remark 4.4. If we consider $\mathbb{T}^d, \forall\, d \geq 1$, then similar computation yields the same estimate $p \nearrow \frac{\alpha}{\tau}$ as polynomial growth is nothing compared with exponential decay! This shows that in any dimension d, our result is strictly stronger than Bence Borda's Theorem 4. That is, for any discrete measure μ, we can get central limit theorem if we assume $\alpha > \tau = \frac{d}{r} \Leftrightarrow \alpha r > d$ where $r \in \mathbb{Z}^+$ is the number of badly approximated elements in the support of μ, while he requires $\alpha r > 2d$.

4.3. **Statistical properties for mixed systems.** Let $\mu \in \text{Prob}(\mathbb{T})$ and let $K : \mathbb{T} \to \text{Prob}(\mathbb{T}), K_\theta = \int \delta_{\theta+\omega_0} d\mu(\omega_0)$ be the Markov kernel.

$Q : C^0(\mathbb{T}) \to C^0(\mathbb{T}), Q\varphi(\theta) = \int \varphi(\theta+\omega_0) d\mu(\omega_0)$ is the corresponding Markov operator.

$\theta \to \theta + \omega_0 \to \theta + \omega_0 + \omega_1 \to \cdots$ is the K-Markov chain where we can denote $Z_0 = \theta, Z_j = \theta + \omega_0 + \cdots + \omega_{j-1}, j \in \mathbb{Z}^+$ in which each ω_i is chosen independently w.r.t. μ.

We have proved that if μ is mixing DC, then $(\mathbb{T}, K, m, C^\alpha(\mathbb{T}))$ is a strongly mixing Markov system with decaying rate $r_n = \frac{1}{n^p}, n \in \mathbb{Z}^+$ and $p = \frac{\alpha}{\tau}-$.

By the abstract LDT Theorem 2.1 of CDK, we have

Corollary 4.8. *$\forall\, \varphi \in C^\alpha(\mathbb{T})$, $\forall\, \theta \in \mathbb{T}$ and $\forall\, \epsilon > 0$, we have*

$$\mu^{\mathbb{N}} \left\{ \left| \frac{1}{n}[\varphi(\theta) + \cdots + \varphi(\theta + \cdots + \omega_{n-1})] - \int \varphi dm \right| > \epsilon \right\} < e^{-c(\epsilon)n}.$$

for some constant $c(\epsilon) > 0$.

By the abstract CLT Theorem 2.2 of Gordin-Livšic, we have

Corollary 4.9. *If $\alpha > \tau$ so that $p > 0$ then $\forall\, \varphi \in C^\alpha(\mathbb{T})$ with zero mean, then if $\sigma(\varphi) > 0$ then*

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

Recall that we also have a slightly more general setup. Let $\Sigma := \mathbb{T}$, on $\Sigma \times \mathbb{T}$ consider the Markov kernel

$$\bar{K}_{(\omega_0,\theta)} := \int \delta_{(\omega_1, \theta+\omega_0)} d\mu(\omega_1)$$

and the corresponding Markov operator $\bar{Q} : C^0(\Sigma \times \mathbb{T}) \to C^0(\Sigma \times \mathbb{T})$

$$\bar{Q}\varphi(\omega_0, \theta) = \int \varphi(\omega_1, \theta + \omega_0) d\mu(\omega_1).$$

The $\bar{K}$-Markov chain is

$$(\omega_0, \theta) \to (\omega_1, \theta + \omega_0) \to (\omega_2, \theta + \omega_0 + \omega_1) \to \cdots$$

Define $\Pi : C^0(\Sigma \times \mathbb{T}) \to C^0(\mathbb{T})$, $\Pi\varphi(\theta) = \int \varphi(\omega_0, \theta) d\mu(\omega_0)$. It is clear that $\bar{Q}\varphi(\omega_0, \theta) = \Pi\varphi(\theta + \omega_0)$. By induction,

$$\bar{Q}^n \varphi(\omega_0, \theta) = Q^{n-1}(\Pi\varphi)(\theta + \omega_0).$$

Define the space $C^{0,\alpha}(\Sigma \times \mathbb{T}))$ as follows:

$$C^{0,\alpha}(\Sigma \times \mathbb{T})) := \left\{ \varphi \in C^0(\Sigma \times \mathbb{T}) : v_\alpha^{\mathbb{T}}(\varphi) < \infty \right\}$$

where

$$v_\alpha^{\mathbb{T}}(\varphi) := \sup_{\omega_0 \in \Sigma} \sup_{\theta \neq \theta'} \frac{|\varphi(\omega_0, \theta) - \varphi(\omega_0, \theta')|}{|\theta - \theta'|}.$$

The corresponding α-norm is defined by $\|\varphi\|_\alpha = \|\varphi\|_\infty + v_\alpha^{\mathbb{T}}(\varphi)$. Then $(\Sigma \times \mathbb{T}, \bar{K}, \mu \times m, C^{0,\alpha}(\Sigma \times \mathbb{T}))$ is a Markov system (simple exercise). Since Q is strongly mixing on $C^\alpha(\mathbb{T})$, then $\bar{Q}$ is strongly mixing on $C^{0,\alpha}(\Sigma \times \mathbb{T})$ with the same decaying rate $r_n = \frac{1}{n^p}, n \in \mathbb{Z}^+$, because

$\Pi C^{0,\alpha}(\Sigma \times \mathbb{T}) \subset C^{\alpha}(\mathbb{T})$. So we get LDT and (if $\alpha > \tau$) CLT. One can formulate them mimicing the previous two corollaries.

Nevertheless, our ultimate goal is to prove LDT and CLT for the mixed random-quasiperiodic system with Hölder observables. Let $\Sigma := \mathbb{T}$, $\mu \in \text{Prob}(\mathbb{T})$, $X := \Sigma^{\mathbb{Z}}$, $\mu^{\mathbb{Z}} \in \text{Prob}(X)$. Define

$$f : X \times \mathbb{T} \to X \times \mathbb{T}, \quad f(\omega, \theta) = (\sigma\omega, \theta + \omega_0).$$

The triple $(X \times \mathbb{T}, f, \mu^{\mathbb{Z}} \times m)$ is called a mixed random-quasiperiodic system.

We say $\varphi \in \mathcal{H}_\alpha(X \times \mathbb{T})$ if $v_\alpha(\varphi) = v_\alpha^X(\varphi) + v_\alpha^{\mathbb{T}}(\varphi) < \infty$ where

$$v_\alpha^X(\varphi) := \sup_{\theta \in \mathbb{T}} \sup_{\omega \neq \omega'} \frac{|\varphi(\omega, \theta) - \varphi(\omega', \theta)|}{d(\omega, \omega')^\alpha}.$$

and

$$v_\alpha^{\mathbb{T}}(\varphi) := \sup_{\omega \in X} \sup_{\theta \neq \theta'} \frac{|\varphi(\omega, \theta) - \varphi(\omega, \theta')|}{|\theta - \theta'|^\alpha}.$$

Note that v_α is a semi-norm. Here $\omega, \omega' \in X$ and

$$d(\omega, \omega') := 2^{-\min\{|j| : \omega_j \neq \omega'_j\}}.$$

Remark 4.5. Note that in general this metric does not make (X, d) a compact metric space unless μ is finitely supported. The essential reason is that this metric only shows the information of where two points differ without telling how much they differ, which does not metrize the product topology. However, results under this metric are stronger because they also hold for functions which are not necessarily Hölder with respect to the standard compactified metric.

Then we can define the norm

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

Then $(\mathcal{H}_\alpha(X \times \mathbb{T}), \|\cdot\|_\alpha)$ is a Banach space.

Consider the corresponding Markov chain on $X \times \mathbb{T}$

$$(\omega, \theta) \to (\sigma\omega, \theta + \omega_0) \to (\sigma^2\omega, \theta + \omega_0 + \omega_1) \to \cdots$$

This is not strongly mixing because it is determined. So we can not derive LDT and CLT directly from the two abstract theorems by strongly mixing condition. We have to find a way around.

To proceed, first we consider observables that are future independent

$$\varphi(\cdots, \omega_{-1}, \omega_0, \cdots) = \varphi(\cdots, \omega_{-1}, \omega_0), \, \forall \, \omega \in X.$$

Namely, $\varphi \in \mathcal{H}_\alpha(X^- \times \mathbb{T})$ which is similarly defined as $\mathcal{H}_\alpha(X \times \mathbb{T})$ above. With an appropriate kernel K^-, the corresponding Markov kernel will be strongly mixing on $\mathcal{H}_\alpha(X^- \times \mathbb{T})$ (essentially because

$(X^- \times \mathbb{T}, K^-, \mu^{-N} \times m, \mathcal{H}_\alpha(X^- \times \mathbb{T}))$ is an extension of $(\Sigma \times \mathbb{T}, \bar{K}, \mu \times m, \mathcal{H}_\alpha(\Sigma \times \mathbb{T}))$.

Moreover, If $\varphi \in \mathcal{H}_\alpha(X \times \mathbb{T})$, it turns out that it is cohomologous to a $\varphi^- \in \mathcal{H}_\alpha(X^- \times \mathbb{T})$ in the sense that

$$\varphi - \varphi^- \circ f = \eta - \eta \circ f$$

for some $\eta \in \mathcal{H}_\alpha^\beta(X \times \mathbb{T})$ with some $\beta < \alpha$. This will allow us to lift the LDT estimates.

So let us first focus on the future independent mixed system. Let $\Sigma := \mathbb{T}$, $X^- := \Sigma^{-\mathbb{N}} = \{\omega^- = \{\omega_j\}_{j \leq 0} : \omega_j \in \Sigma\}$ endowed with the distance d defined before. Denote by $\mu^{-\mathbb{N}}$ the product measure.

The Markov kernel K^- on $X^- \times \mathbb{T}$ is defined by

$$K^-_{(\omega^-, \theta)} = \int \delta_{(\omega^- \omega_1, \theta + \omega_0)} d\mu(\omega_1)$$

and the corresponding Markov operator Q^- on $C^0(X^- \times \mathbb{T})$ is

$$Q^- \varphi(\omega^-, \theta) = \int \varphi(\omega^- \omega_1, \theta + \omega_0) d\mu(\omega_1).$$

The associated Markov chain is

$$(\omega^-, \theta) \to (\omega^- \omega_1, \theta + \omega_0) \to (\omega^- \omega_1 \omega_2, \theta + \omega_0 + \omega_1) \to \cdots$$

The space of α-Hölder observables, denoted by $\mathcal{H}_\alpha(X^- \times \mathbb{T})$, is defined as

$$\mathcal{H}_\alpha(X^- \times \mathbb{T}) := \left\{ \varphi \in C^0(X^- \times \mathbb{T}) : v_\alpha(\varphi) := v_\alpha^{X^-}(\varphi) + v_\alpha^{\mathbb{T}}(\varphi) < \infty \right\}$$

where

$$v_\alpha^{X^-}(\varphi) := \sup_{\theta \in \mathbb{T}} \sup_{\omega^- \neq \omega'^-} \frac{|\varphi(\omega^-, \theta) - \varphi(\omega'^-, \theta)|}{d(\omega^-, \omega'^-)^\alpha}.$$

and

$$v_\alpha^{\mathbb{T}}(\varphi) := \sup_{\omega^- \in X^-} \sup_{\theta \neq \theta'} \frac{|\varphi(\omega^-, \theta) - \varphi(\omega^-, \theta')|}{|\theta - \theta'|^\alpha}.$$

The norm is defined by $\|\varphi\|_\alpha := \|\varphi\|_\infty + v_\alpha(\varphi)$. Moreover, $(\mathcal{H}_\alpha(X^- \times \mathbb{T}), \|\cdot\|_\alpha)$ is a Banach space.

For the sake of convenience, we recall the definition of Markov systems which was introduced previously in Section 2.

Definition 4.3. A Markov system is a tuple $(M, K, \mu, \mathcal{E})$ where

(1) M is a compact metric space,
(2) $K : M \to \text{Prob}(M)$ is an SDS,
(3) $\mu = K * \mu := \int K_x \, d\mu(x) \in \text{Prob}(M)$ is a stationary measure,

(4) $\mathcal{E} = (\mathcal{E}, \|\cdot\|_{\mathcal{E}})$ is a Banach subspace of $C^0(M)$ such that action of $\mathcal{Q}_K$ on $\mathcal{E}$ and the inclusion $\mathcal{E} \subset L^\infty(M)$ are both continuous. In other words there are constants $M_1 < \infty$ and $M_2 < \infty$ such that $\|\varphi\|_\infty \le M_1 \|\varphi\|_{\mathcal{E}}$ and $\|\mathcal{Q}\varphi\|_{\mathcal{E}} \le M_2 \|\varphi\|_{\mathcal{E}}$, for all $\varphi \in \mathcal{E}$.

Remark 4.6. When $\mathcal{E} = C^0(M)$ condition (4) follows from (1)-(3).

Proposition 4.10. *$(X^- \times \mathbb{T}, K^-, \mu^{-\mathbb{N}} \times m, \mathcal{H}_\alpha(X^- \times \mathbb{T}))$ is a Markov system.*

Proof. Given $\varphi \in C^0(X^- \times \mathbb{T})$, it is indeed uniformly continuous due to the compactness of $X^- \times \mathbb{T}$. Therefore, for any $\epsilon > 0$, we can take $\delta > 0$ to be sufficiently small such that $d((x,\theta),(y,\eta)) < \delta$ implies $|\varphi(x,\theta) - \varphi(y,\eta)| < \epsilon$.

Therefore, if $d((x,\theta),(y,\eta)) < \delta/4$ we have

$$\begin{aligned} &|(Q_{K^-}\varphi)(x,\theta) - (Q_{K^-}\varphi)(y,\eta)| \\ &\qquad \le \left|\int \varphi(xs, \theta + x_0) d\mu(s) - \int \varphi(ys, \eta + y_0) d\mu(s)\right| \le \epsilon, \end{aligned}$$

which proves that $Q_{K^-}\varphi \in C^0(X^- \times \mathbb{T})$ and K^- is an SDS.

Every measurable set $V \subset X^- \times \mathbb{T}$ can be approximated in measure by countable unions of cylinders, i.e., measurable sets having the form $A_i \times B_i$ with $A_i \subset X^-$ and $B_i \subset \mathbb{T}$. Direct computation shows

$$\begin{aligned} \left(K^- * (\mu^{-\mathbb{N}} \times m)\right)(A_i \times B_i) &= \int K^-_{(x,\theta)}(A_i \times B_i)\, d(\mu^{-\mathbb{N}} \times m)(x,\theta) \\ &= \iint K^-_{(x,\theta)}(A_i \times B_i)\, d\mu^{-\mathbb{N}}(x) dm(\theta) \\ &= \iint \delta_{(xs,\theta+x_0)}(A_i \times B_i)\, d\mu(s) d\mu^{-\mathbb{N}}(x) dm(\theta) \\ &= \left(\int \delta_y(A_i)\, d\mu^{-\mathbb{N}}(y)\right)\left(\int \delta_{\theta+x_0}(B_i)\, dm(\theta)\right) \\ &= \left(\mu^{-\mathbb{N}} \times m\right)(A_i \times B_i) \end{aligned}$$

holds for any $A_i \times B_i$. Thus it also holds for any measurable set $V \subset X^- \times \mathbb{T}$. This proves that $\mu^{-\mathbb{N}} \times m$ is a K^--stationary measure.

Item (4) of Definition 4.3 is straightforward to check and it holds with $M_1 = M_2 = 1$. $\square$

Our current goal is to prove that $(X^- \times \mathbb{T}, K^-, \mu^{-\mathbb{N}} \times m, \mathcal{H}_\alpha(X^- \times \mathbb{T}))$ is strongly mixing with rate $\frac{1}{n^p}$, $p = \frac{\alpha}{\tau}^-$. We showed that $(\Sigma \times \mathbb{T}, \bar{K}, \mu \times m, C^\alpha(\Sigma \times \mathbb{T}))$ is strongly mixing with rate $\frac{1}{n^p}$, $p = \frac{\alpha}{\tau}^-$. We will prove that $(\Sigma \times \mathbb{T}, \bar{K}, \mu \times m, C^\alpha(\Sigma \times \mathbb{T}))$ is a contracting factor of $(X^- \times \mathbb{T}, K^-, \mu^{-\mathbb{N}} \times m, \mathcal{H}_\alpha(X^- \times \mathbb{T}))$, which allows us to lift the strongly mixing property from $\bar{K}$ to K^-.

We first introduce the definition of *factor*.

Definition 4.4. Given two Markov systems $(M, K, \mu, \mathcal{E})$ and $(\tilde{M}, \tilde{K}, \tilde{\mu}, \tilde{\mathcal{E}})$, the first is called a *factor* of the second if there exists a continuous projection $\pi : \tilde{M} \to M$ such that

(1) $\pi_* \tilde{\mu} = \mu$,
(2) $K_{\pi(\tilde{x})} = \pi_* \tilde{K}_{\tilde{x}}$ for all $\tilde{x} \in \tilde{M}$,
(3) there exists $\eta : M \to \tilde{M}$ continuous with $\pi \circ \eta = \mathrm{id}_M$ such that $\eta^*(\tilde{\mathcal{E}}) \subseteq \mathcal{E}$ and $\|\varphi \circ \eta\|_{\mathcal{E}} \leq M_1 \|\varphi\|_{\tilde{\mathcal{E}}}$ for some constant $M_1 < \infty$ and all $\varphi \in \tilde{\mathcal{E}}$,
(4) $\pi^*(\mathcal{E}) \subseteq \tilde{\mathcal{E}}$ and $\|\varphi \circ \pi\|_{\tilde{\mathcal{E}}} \leq M_2 \|\varphi\|_{\mathcal{E}}$ for some constant $M_2 < \infty$ and all $\varphi \in \mathcal{E}$.

Factors have the following properties.

Proposition 4.11. *Let $(M, K, \mu, \mathcal{E})$ be a factor of $(\tilde{M}, \tilde{K}, \tilde{\mu}, \tilde{\mathcal{E}})$.*

(1) $\pi^* \circ Q_K = Q_{\tilde{K}} \circ \pi^*$, i.e. the following commutative diagram holds

$$\begin{array}{ccc} C^0(\tilde{M}) & \xrightarrow{Q_{\tilde{K}}} & C^0(\tilde{M}) \\ {\scriptstyle \pi^*}\Big\uparrow & & \Big\uparrow{\scriptstyle \pi^*} \\ C^0(M) & \xrightarrow[Q_K]{} & C^0(M) \end{array} .$$

(2) The bounded linear map $\pi^* : \mathcal{E} \to \pi^*(\mathcal{E})$ is an isomorphism onto the closed linear subspace $\pi^*(\mathcal{E}) \subseteq \tilde{\mathcal{E}}$.

Proof. Since $K_{\pi(\tilde{x})} = \pi_* \tilde{K}_{\tilde{x}}$,

$$(Q_{\tilde{K}} \circ \pi^* \varphi)(\tilde{x}) = \int \varphi \circ \pi d\tilde{K}_{\tilde{x}} = \int \varphi dK_{\pi(\tilde{x})} = (\pi^* \circ Q_K \varphi)(\tilde{x})$$

which gives item (1).

Let us prove item (2). By definition, $\pi^* \varphi = \varphi \circ \pi$ is a bounded linear operator. Since π is surjective, then π^* is one to one. Thus $\pi^* : \mathcal{E} \to \pi^*(\mathcal{E})$ is a linear bijection. Define the closed linear subspace of $\tilde{\mathcal{E}}$

$$\mathcal{V} := \{\varphi \in \tilde{\mathcal{E}} : \forall\, x, y \in \tilde{M}, \pi(x) = \pi(y) \Rightarrow \varphi(x) = \varphi(y)\}.$$

The linearity is clear. For the closedness, assume that $\tilde{\varphi}_n \in \mathcal{V}$ and $\tilde{\varphi}_n \to \tilde{\varphi}$ pointwise in $\tilde{\mathcal{E}}$. If $\pi(x) = \pi(y)$, then $\tilde{\varphi}_n(\tilde{x}) = \tilde{\varphi}_n(\tilde{y})$. Therefore, let $n \to \infty$ we get $\tilde{\varphi}(\tilde{x}) = \tilde{\varphi}(\tilde{y})$ with $\tilde{\varphi} \in \tilde{\mathcal{E}}$, which shows that $\tilde{\varphi} \in \mathcal{V}$.

Clearly $\pi^*(\mathcal{E}) \subseteq \mathcal{V}$. Conversely, given $\varphi \in \mathcal{V}$ consider the function $\psi := \varphi \circ \eta \in \mathcal{E}$ where $\eta : M \to \tilde{M}$ is defined in Definition 4.4. Since $\pi(x) = \pi(\eta(\pi(x)))$, by definition of $\mathcal{V}$ we have $\varphi(x) = \varphi(\eta(\pi(x)))$ for all $x \in \tilde{M}$, which proves $\varphi = \psi \circ \pi \in \pi^*(\mathcal{E})$. Therefore, $\mathcal{V} = \pi^*(\mathcal{E})$ is a closed linear subspace of $\tilde{\mathcal{E}}$, thus also a Banach (sub) space with the prescribed norm on $\tilde{\mathcal{E}}$. Finally, by the Banach open mapping theorem π^* is an open map. Namely, the inverse map $(\pi^*)^{-1} : \pi^*(\mathcal{E}) \to \mathcal{E}$ is continuous, thus also a bounded linear map. This proves that π^* is an isomorphism. □

We introduce the notion of contracting factors.

Definition 4.5. We call $(M, K, \mu, \mathcal{E})$ a contracting factor of $(\tilde{M}, \tilde{K}, \tilde{\mu}, \tilde{\mathcal{E}})$ with contracting rate τ if additionally we have the following: $\exists\, C > 0$ such that $\forall\, \tilde{\varphi} \in \tilde{\mathcal{E}}$, $\exists\, \psi_n \in \mathcal{E}, n \in \mathbb{N}$ satisfying

$$\|\psi_n\|_\infty \leq \|\tilde{\varphi}\|_\infty\,, \;\; \|\psi_n\|_{\mathcal{E}} \leq C\, \|\tilde{\varphi}\|_{\tilde{\mathcal{E}}}$$

and

$$\left\| \tilde{Q}^n \tilde{\varphi} - \pi^* \psi_n \right\|_\infty \leq C\, \|\tilde{\varphi}\|_{\tilde{\mathcal{E}}}\, \tau(n)$$

for all $n \in \mathbb{N}$.

We have the following abstract theorem.

Theorem 4.5. *Assume that $(M, K, \mu, \mathcal{E})$ is strongly mixing with rate r and that $(M, K, \mu, \mathcal{E})$ is a contracting factor of $(\tilde{M}, \tilde{K}, \tilde{\mu}, \tilde{\mathcal{E}})$ with contracting rate τ. Then $(\tilde{M}, \tilde{K}, \tilde{\mu}, \tilde{\mathcal{E}})$ is strongly mixing with rate $r^*(n) = \max\{r(\frac{n}{2}), \tau(\frac{n}{2})\}$.*

Proof. Fix $\tilde{\varphi} \in \tilde{\mathcal{E}}$ and $n \in \mathbb{N}$. We may assume that n is even. Otherwise since $\tilde{Q}^n \tilde{\varphi} = \tilde{Q}^{n-1}(\tilde{Q}\tilde{\varphi})$, we can work with $\tilde{Q}\tilde{\varphi}$ instead of $\tilde{\varphi}$. For this $\tilde{\varphi}$ and $\frac{n}{2}$, consider $\psi_{\frac{n}{2}} =: \psi \in \mathcal{E}$ such that

$$\|\psi\|_\infty \leq \|\tilde{\varphi}\|_\infty\,, \;\; \|\psi\|_{\mathcal{E}} \lesssim \|\tilde{\varphi}\|_{\tilde{\mathcal{E}}}$$

and

$$\left\| \tilde{Q}^{\frac{n}{2}} \tilde{\varphi} - \pi^* \psi \right\|_\infty \lesssim \|\tilde{\varphi}\|_{\tilde{\mathcal{E}}}\, \tau(\frac{n}{2}).$$

Since $\tilde{\mu}$ is $\tilde{K}$-stationary, then

$$\int \tilde{Q}^j \tilde{\varphi} d\tilde{\mu} = \int \tilde{\varphi} d\tilde{\mu}, \forall\, j \in \mathbb{N}.$$

As $\pi_* \tilde{\mu} = \mu$, we have

$$\int \pi^* \psi d\tilde{\mu} = \int \psi d\mu.$$

Then by an integration on both sides of the previous inequality, we obtain

$$\left|\int \tilde{\varphi} d\tilde{\mu} - \int \psi d\mu\right| \lesssim \|\tilde{\varphi}\|_{\tilde{\mathcal{E}}}\, \tau(\frac{n}{2}).$$

Using that $(M, K, \mu, \mathcal{E})$ is strongly mixing with rate r, we have

$$\left\|Q^{\frac{n}{2}}\psi - \int \psi d\mu\right\|_\infty \lesssim \|\psi\|_{\mathcal{E}}\, r(\frac{n}{2}) \lesssim \|\tilde{\varphi}\|_{\tilde{\mathcal{E}}}\, r(\frac{n}{2}).$$

On the other hand, by the commutative diagram we have

$$\tilde{Q}^{\frac{n}{2}}\pi^*\psi = \pi^* Q^{\frac{n}{2}}\psi.$$

Treating $\int \psi d\mu$ as a constant function, we have $\pi^*(\int \psi d\mu) = \int \psi d\mu$. Thus

$$\left\|\tilde{Q}^{\frac{n}{2}}\pi^*\psi - \int \psi d\mu\right\|_\infty = \left\|\pi^*\tilde{Q}^{\frac{n}{2}}\psi - \pi^*(\int \psi d\mu)\right\|_\infty \lesssim \|\tilde{\varphi}\|_{\tilde{\mathcal{E}}}\, r(\frac{n}{2}).$$

Finally, note that

$$\tilde{Q}^n\tilde{\varphi} - \int \tilde{\varphi} d\tilde{\mu} = \tilde{Q}^n\tilde{\varphi} - \tilde{Q}^{\frac{n}{2}}(\pi^*\psi) + \tilde{Q}^{\frac{n}{2}}(\pi^*\psi) - \int \psi d\mu + \int \psi d\mu - \int \tilde{\varphi} d\tilde{\mu}.$$

Thus by triangle inequality,

$$\left\|\tilde{Q}^n\tilde{\varphi} - \int \tilde{\varphi} d\tilde{\mu}\right\| \lesssim \|\tilde{\varphi}\|_{\tilde{\mathcal{E}}} \left(\tau(\frac{n}{2}) + r(\frac{n}{2}) + \tau(\frac{n}{2})\right).$$

The result follows. □

We will apply this abstract result to the following setting: the factor $(M, K, \mu, \mathcal{E})$ will be $(\Sigma \times \mathbb{T}, \bar{K}, \mu \times m, C^{0,\alpha}(\Sigma \times \mathbb{T}))$. We have proved that this system is strongly mixing with rate $r(n) = \frac{1}{n^p}, p > 0$ provided that the measure μ is mixing DC. Actually we proved this for $(\mathbb{T}, K, m, C^\alpha(\mathbb{T}))$. On the other hand, the system $(\tilde{M}, \tilde{K}, \tilde{\mu}, \tilde{\mathcal{E}})$ will be $(X^- \times \mathbb{T}, K^-, \mu^{-\mathbb{N}} \times m, \mathcal{H}_\alpha(X^- \times \mathbb{T}))$.

Theorem 4.6. $(\Sigma \times \mathbb{T}, \bar{K}, \mu \times m, C^{0,\alpha}(\Sigma \times \mathbb{T}))$ *is a contracting factor with exponential rate of* $(X^- \times \mathbb{T}, K^-, \mu^{-\mathbb{N}} \times m, \mathcal{H}_\alpha(X^- \times \mathbb{T}))$. *Therefore, the second is strongly mixing with rate* $\frac{1}{n^p}$.

Proof. Define $\pi : X^- \times \mathbb{T} \to \Sigma \times \mathbb{T}$, $\pi(\omega^-, \theta) = (\omega_0, \theta)$. Fix $a \in \Sigma$, define $\eta : \Sigma \times \mathbb{T} \to X^- \times \mathbb{T}$, $\eta(\omega_0, \theta) = (\cdots aa\omega_0, \theta)$. It is straightforward to check items (1)-(4) in Definition 4.4, thus the first system is a factor of the second. Let us prove that it is indeed a contracting factor.

Fix $n \in \mathbb{N}$, denote by $\mathcal{H}_{\alpha,n}(X^- \times \mathbb{T})$ the functions in $\mathcal{H}_\alpha(X^- \times \mathbb{T})$ that only depend on last n random coordinates $\omega_{-n+1}, \cdots, \omega_{-1}, \omega_0$ and θ.

If $\varphi \in \mathcal{H}_{\alpha,n}(X^- \times \mathbb{T})$, then

$$(Q^-)^n\varphi(\omega^-,\theta) = \int \cdots \int \varphi(\omega^-\omega_1\cdots\omega_n, \theta+\omega_0+\cdots+\omega_{n-1})d\mu(\omega_n)\cdots d\mu(\omega_1)$$

only depends on (ω_0, θ), so $(Q^-)^n\varphi(\omega^-,\theta) \in C^{0,\alpha}(\Sigma \times \mathbb{T})$.

Fix any $\varphi \in \mathcal{H}_\alpha(X^- \times \mathbb{T})$ and $n \in \mathbb{N}$, we construct $\psi_n \in C^{0,\alpha}(\Sigma \times \mathbb{T})$ in the following. Let

$$\varphi_n(\omega^-,\theta) = \varphi(\cdots aa\omega_{-n+1}\cdots\omega_0, \theta).$$

and denote $i_n(\omega^-,\theta) = (\cdots aa\omega_{-n+1}\cdots\omega_0,\theta)$. So $\varphi_n \in \mathcal{H}_{\alpha,n}(X^- \times \mathbb{T})$. We have proved that the bounded linear map $\pi^* : \mathcal{E} \to \pi^*(\mathcal{E})$ is an isomorphism onto the closed linear subspace $\pi^*(\mathcal{E}) \subseteq \tilde{\mathcal{E}}$. So let $\psi_n \in C^{0,\alpha}(\Sigma \times \mathbb{T})$ be such that

$$\pi^*\psi_n = \psi_n \circ \pi = (Q^-)^n\varphi_n.$$

Then

$$\begin{aligned}
&\left\|(Q^-)^n(\varphi) - \psi_n \circ \pi\right\|_\infty \\
=&\left\|(Q^-)^n(\varphi) - (Q^-)^n\varphi_n\right\|_\infty \\
\le&\left\|\varphi - \varphi_n\right\|_\infty \\
=&\sup_{(\omega^-,\theta)\in X^-\times\mathbb{T}} \left|\varphi(\omega^-,\theta) - \varphi(i_n(\omega^-,\theta))\right| \\
\le& v_\alpha^{X^-}(\varphi) d\left(\omega^-, (\cdots aa\omega_{-n+1}\cdots\omega_0)\right)^\alpha \\
\le& 2^{-n\alpha}\left\|\varphi\right\|_\alpha .
\end{aligned}$$

Let $\sigma = 2^{-\alpha} < 1$, we conclude the proof by applying Theorem 4.5. □

Fix any $\varphi \in \mathcal{H}_\alpha(X^- \times \mathbb{T})$ and $(\omega^-,\theta) \in X^- \times \mathbb{T}$, consider the K^- Markov chain $\{Z_n\}_{n\ge0}$ such that

$$Z_0 = (\omega^-,\theta),\ Z_1 = (\omega^-\omega_1, \theta+\omega_0), \cdots,$$
$$Z_n = (\omega^-\omega_1\cdots\omega_n, \theta+\omega_0+\cdots+\omega_{n-1}), \cdots$$

Denote by $S_n\varphi = \varphi(Z_0) + \cdots + \varphi(Z_{n-1})$.

By the abstract LDT Theorem 2.1, we have

$$\mathbb{P}_{(\omega^-,\theta)}\left\{\left|\frac{1}{n}S_n\varphi - \int_{X^-\times\mathbb{T}} \varphi d\mu^{-\mathbb{N}} \times m\right| > \epsilon\right\} \lesssim e^{-c(\epsilon)n}.$$

It implies that

$$\mathbb{P}_{\mu^{-\mathbb{N}}\times m}\left\{\left|\frac{1}{n}S_n\varphi - \int_{X^-\times\mathbb{T}} \varphi d\mu^{-\mathbb{N}} \times m\right| > \epsilon\right\} \lesssim e^{-c(\epsilon)n}.$$

On the other hand, consider the probability space $(X \times \mathbb{T}, \mu^{\mathbb{Z}} \times m)$. Define $f : X\times\mathbb{T} \to X\times\mathbb{T}$, $f(\omega,\theta) = (\sigma\omega, \theta+\omega_0)$ and define $\pi : X\times\mathbb{T} \to$

$X^- \times \mathbb{T}, \pi(\omega,\theta) = (\omega^-, \theta)$. Then for every $Z_0 = (\omega^-, \theta) \in X^- \times \mathbb{T}$ chosen according to $\mu^{-\mathbb{N}} \times m$, we have $Z_0 = \pi(\omega,\theta)$ with $(\omega,\theta) =: \bar{Z}_0$ chosen according to $\mu^{\mathbb{Z}} \times m$. In fact $\pi_*(\mu^{\mathbb{Z}} \times m) = \mu^{-\mathbb{N}} \times m$.

Therefore, for every $Z_j, j \geq 0$ which takes value in $X^- \times \mathbb{T}$, we have $Z_j = \pi[f^j(\bar{Z}_0)]$. This shows that the previous inequality is equivalent to the following:

$$\mu^{\mathbb{Z}} \times m \left\{ (\omega,\theta) : \left| \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ \pi[f^j(\omega,\theta)] - \int_{X^- \times \mathbb{T}} \varphi d\mu^{-\mathbb{N}} \times m \right| > \epsilon \right\} \lesssim e^{-c(\epsilon)n}$$

If $\alpha > \tau$, we also get the central limit theorem by the abstact CLT Theorem 2.1.

We want to extend these statistical properties to observables that also depend on the future. In fact, we want to obtain similar results of the Markov system $(X \times \mathbb{T}, f, \mu^{\mathbb{Z}} \times m, \mathcal{H}_\alpha(X \times \mathbb{T}))$ (defined before already) in which the Markov operator is exactly the Koopman operator: $Q_K(\varphi) = \varphi \circ f$ and K is the dirac delta. This Koopman operator is no longer strongly mixing.

Given $\varphi \in C^0(X \times \mathbb{T})$ and $(\omega,\theta) \in X \times \mathbb{T}$, the Birkhoff sums are

$$S_n\varphi(\omega,\theta) = \varphi(\omega,\theta) + \varphi \circ f(\omega,\theta) + \cdots + \varphi \circ f^{n-1}((\omega,\theta)).$$

Remark 4.7. We will identify $\mathcal{H}_\alpha(X^- \times \mathbb{T})$ with the subspace of observables in $\mathcal{H}_\alpha(X \times \mathbb{T})$ that are future independent. φ is called future independent if $\varphi(x,\theta) = \varphi(y,\theta)$ whenever $x^- = y^-$.

The idea of proving LDT, CLT for the DDS $(X \times \mathbb{T}, f)$ is to "reduce" an observable $\varphi \in \mathcal{H}_\alpha(X \times \mathbb{T})$ to an observable $\varphi^- \in \mathcal{H}_\beta(X^- \times \mathbb{T})$. More precisely, we will prove the following proposition.

Proposition 4.12. *$\forall\, \varphi \in \mathcal{H}_\alpha(X \times \mathbb{T})$, there are $\varphi^- \in \mathcal{H}_\beta(X^- \times \mathbb{T})$ and $\eta \in \mathcal{H}_\beta(X \times \mathbb{T})$ with $\beta = \frac{\alpha}{3}$ such that*

$$\varphi - \varphi^- \circ f = \eta - \eta \circ f. \tag{4.1}$$

Moreover, the map $\varphi \to \varphi^-$ is a bounded linear operator, $\|\varphi^-\|_\beta \lesssim \|\varphi\|_\alpha$.

Let us assume for now that this proposition is valid and then we can derive the LDT and CLT for $(X \times \mathbb{T}, f)$.

Integrating both sides of the equation (4.1) w.r.t. $\mu^{\mathbb{Z}} \times m$, we have

$$\int \varphi d\mu^{\mathbb{Z}} \times m - \int \varphi^- \circ f d\mu^{\mathbb{Z}} \times m = \int \eta d\mu^{\mathbb{Z}} \times m - \int \eta \circ f d\mu^{\mathbb{Z}} \times m.$$

Since $\mu^{\mathbb{Z}} \times m$ is f-invariant, the right hand side equals zero. This shows

$$\int \varphi d\mu^{\mathbb{Z}} \times m = \int \varphi^- \circ f d\mu^{\mathbb{Z}} \times m = \int \varphi^- d\mu^{\mathbb{Z}} \times m = \int \varphi^- d\mu^{-\mathbb{N}} \times m.$$

At the same time, equation (4.1 is equivalent to

$$\varphi = \varphi^- \circ f + \eta - \eta \circ f.$$

For LDT, taking the Birkhoff sums on both sides, we obtain

$$S_n\varphi = S_n(\varphi^- \circ f) + \eta - \eta \circ f^n,$$

which further implies that

$$\begin{aligned}\frac{1}{n}S_n\varphi - \int \varphi d\mu^{\mathbb{Z}} \times m &= \frac{1}{n}S_n(\varphi^- \circ f) - \int \varphi^- d\mu^{-\mathbb{N}} \times m + \frac{\eta - \eta \circ f^n}{n} \\ &\approx \frac{1}{n}S_n(\varphi^-) - \int \varphi^- d\mu^{-\mathbb{N}} \times m\end{aligned}$$

as n is large. This shows the equivalence of the LDT estimates between $(X\times\mathbb{T}, f, \mu^{\mathbb{Z}}\times m, \mathcal{H}_\alpha(X\times\mathbb{T}))$ and $(X^-\times\mathbb{T}, K^-, \mu^{-\mathbb{N}}\times m, \mathcal{H}_\beta(X^-\times\mathbb{T}))$. In particular, we can lift the LDT estimates from the second system to the first one.

For CLT (we need $\alpha > 3\tau$ such that $\beta/\tau > 1$), since we have

$$S_n\varphi - \int \varphi d\mu^{\mathbb{Z}} \times m = S_n(\varphi^- \circ f) - \int \varphi^- d\mu^{-\mathbb{N}} \times m + \eta - \eta \circ f^n,$$

dividing at both sides by $\sigma\sqrt{n}$ with $\sigma > 0$ we have

$$\frac{S_n\varphi - \int \varphi d\mu^{\mathbb{Z}} \times m}{\sigma\sqrt{n}} = \frac{S_n(\varphi^- \circ f) - \int \varphi^- d\mu^{-\mathbb{N}} \times m}{\sigma\sqrt{n}} + \frac{\eta - \eta \circ f^n}{\sigma\sqrt{n}}.$$

This shows the equivalence of the CLT between $(X\times\mathbb{T}, f, \mu^{\mathbb{Z}}\times m, \mathcal{H}_\alpha(X\times\mathbb{T}))$ and $(X^- \times \mathbb{T}, K^-, \mu^{-\mathbb{N}} \times m, \mathcal{H}_\beta(X^- \times \mathbb{T}))$. Namely,

$$\frac{S_n(\varphi^-) - \int \varphi^- d\mu^{-\mathbb{N}} \times m}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1),$$

if and only if

$$\frac{S_n\varphi - \int \varphi d\mu^{\mathbb{Z}} \times m}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1).$$

Therefore, it remains to prove Proposition 4.12. Before that, let us make some preparations regarding the concepts of continuous disintegration and unstable holonomy.

Let $\pi : X \times \mathbb{T} \to X^- \times \mathbb{T}, \pi(\omega,\theta) := (\omega^-,\theta)$ be the standard projection. For $\omega \in X$, we will write $\omega = (\omega^-;\omega^+)$ where $\omega^- \in X^-$ and $\omega^+ \in X^+ := \Sigma^{\mathbb{N}^+}$.

Obviously, we have $\pi_*(\mu^{\mathbb{Z}} \times m) = \mu^{-\mathbb{N}} \times m$.

Definition 4.6. Let $(\tilde{M}, \tilde{\mu})$ and (M, μ) be measurable spaces. Assume that $\tilde{M}$ and M are compact metric spaces and $\pi : \tilde{M} \to M$ is continuous with $\pi_* \tilde{\mu} = \mu$. A continuous disintegration of $\tilde{\mu}$ over π is a family of probability measures $\{\tilde{\mu}_a\}_{a \in M}$ such that

(1) $\tilde{\mu}_a \in \text{Prob}(\tilde{M})$ and $\tilde{\mu}_a(\pi^{-1}\{a\}) = 1$,
(2) $M \ni a \mapsto \tilde{\mu}_a \in \text{Prob}(\tilde{M})$ is continuous,
(3) $\forall \varphi \in C^0(\tilde{M})$,

$$\int_{\tilde{M}} \varphi d\tilde{\mu} = \int_M (\int_{\pi^{-1}\{a\}} \varphi d\tilde{\mu}_a) d\mu(a).$$

For any $(\omega^-, \theta) \in X^- \times \mathbb{T}$, let

$$\mathbb{P}_{(\omega^-,\theta)} := \delta_{\omega^-} \times \mu^{\mathbb{N}^+} \times \delta_\theta \in \text{Prob}(X \times \mathbb{T}).$$

Then clearly we have $\{\mathbb{P}_{(\omega^-,\theta)}\}_{(\omega^-,\theta) \in X^- \times \mathbb{T}}$ is a continuous disintegration of $\mathbb{P} = \mu^{\mathbb{Z}} \times m$ along π. Moreover, for $(\omega^-, \theta) \in X^- \times \mathbb{T}$,

$$\pi^{-1}\left\{(\omega^-, \theta)\right\} = \left\{(\omega^-, \omega^+, \theta) : \omega^+ \in X^+\right\} =: W^u_{loc}(\omega^-, \theta)$$

are the local unstable sets of the partially hyperbolic dynamical system $f : X \times \mathbb{T} \to X \times \mathbb{T}$. We clarify this in the following.

Let $x, y \in X$ with $x^- = y^-$. Namely,

$$x = (\cdots, x_{-1}, x_0; x_1, \cdots), \quad y = (\cdots, x_{-1}, x_0; y_1, \cdots).$$

Then

$$\sigma^{-1}x = (\cdots, x_{-1}; x_0, x_1, \cdots), \quad \sigma^{-1}y = (\cdots, x_{-1}; x_0, y_1, \cdots)$$

which gives $d(\sigma^{-1}x, \sigma^{-1}y) \leq 2^{-1}$. If $(x, \theta), (y, \theta)$ belong to the same fiber $W^u_{loc}(x^-, \theta)$, then $x^- = y^-$ and

$$f^{-1}(x, \theta) = (\sigma^{-1}x, \theta + x_{-1}), \quad f^{-1}(y, \theta) = (\sigma^{-1}y, \theta + x_{-1})$$

are still in the same fiber $W^u_{loc}(x^-, \theta + x_{-1})$ with

$$d_0(f^{-1}(x, \theta), f^{-1}(y, \theta)) \leq 2^{-1}.$$

So f^{-1} contracts the fibers. By induction,

$$d_0(f^{-n}(x, \theta), f^{-n}(y, \theta)) \leq 2^{-n}.$$

Backward contracting means they are unstable sets.

Let $\varphi \in \mathcal{H}_\alpha(X \times \mathbb{T})$, we may define the unstable holonomies between two points $(x, \theta), (y, \theta) \in W^u_{loc}(x^-, \theta)$ by

$$h^u_\varphi((x, \theta), (y, \theta)) := \sum_{n=1}^{\infty} [\varphi(f^{-n}(y, \theta)) - \varphi(f^{-n}(x, \theta))] \leq v_\alpha(\varphi) \sum_{n=1}^{\infty} 2^{-n\alpha} < \infty.$$

Proposition 4.13. *Given $(x,\theta),(y,\theta),(z,\theta)$ on the same fiber, the following properties hold (the last one holds if $f(y,\theta)$ and $f(x,\theta)$ belong to the same fiber):*

(a) $h^u_\varphi((x,\theta),(x,\theta))=0$,
(b) $h^u_\varphi((x,\theta),(y,\theta))=-h^u_\psi((y,\theta),(x,\theta))$,
(c) $h^u_\varphi((x,\theta),(z,\theta))=h^u_\varphi((x,\theta),(y,\theta))+h^u_\varphi((y,\theta),(z,\theta))$,
(d) $h^u_\varphi((x,\theta),(y,\theta))+\varphi(y,\theta)=\varphi(x,\theta)+h^u_\varphi(f(x,\theta),f(y,\theta))$.

In the following, we are going to prove Proposition 4.12, which reduces $\varphi\in\mathcal{H}_\alpha(X\times\mathbb{T})$ to

$$\varphi^-\in\mathcal{H}_\beta(X^-\times\mathbb{T})\equiv\{\psi\in\mathcal{H}_\beta(X\times\mathbb{T}):\psi\,\text{is future independent}\}$$

in the sense that φ and $\varphi^-\circ f$ are cohomologous through some $\eta\in\mathcal{H}_\beta(X\times\mathbb{T})$. Moreover, the map $\varphi\to\varphi^-$ is a bounded linear operator.

We will try to guess what η and the φ^- should be. Actually η determines φ^- simply by

$$\varphi^-\circ f=\eta\circ f-\eta+\varphi$$

which is equivalent to

$$\varphi^-=\eta-\eta\circ f^{-1}+\varphi\circ f^{-1}.$$

This is still related to the cohomological equation.

So, let us recall some basic results on the cohomological equations. Let (M,f) be a dynamical system. For an observable $\varphi:M\to\mathbb{R}$, the goal is to find an $\eta:M\to\mathbb{R}$ such that $\varphi=\eta-\eta\circ f$. In fact, we want more in the sense that if φ has some regularity, the solution η should have the same or almost the same regularity.

Theorem 4.7 (Gottschalk-Hedlund)**.** *Let M be a compact metric space. Assume $f:M\to M$ is a minimal homeomorphism and the observable $\varphi:M\to\mathbb{R}$ is continuous. If $\{S_n\varphi(x)\}_{n\geq 0}$ is uniformly bounded in $n\in\mathbb{N}$ and $x\in M$, then there is a continuous function $\eta:M\to\mathbb{R}$ satisfying $\varphi=\eta-\eta\circ f$.*

We give a hint of the proof here. Let

$$\eta(x):=\sup_{n\geq 0}S_n\varphi(x)=\varphi(x)+\sup_{n\geq 0}S_n(\varphi\circ f(x))=\varphi(x)+\eta\circ f(x).$$

And then prove η is continuous.

Remark 4.8. Instead of the $\sup_{n\geq 0}$, one can take any other "intrinsic characteristics" of the sequence $S_n\varphi$ that makes sense. For example, $\inf_{n\geq 0}, \lim_{n\geq 0}$ etc.

Actually, there is a more subtle result.

Theorem 4.8 (Livšic). *Let f be a diffeomorphism on a Riemannian manifold $\Lambda \subset M$ which is compact, topological transitive and is a hyperbolic set. Let $\varphi : \Lambda \to \mathbb{R}$ be α-Hölder continuous. Assume that whenever $f^p(x) = x$ we have $S_p\varphi(x) = 0$. Then $\exists!$ (up to an additive constant) $\eta : \Lambda \to \mathbb{R}$ α-Hölder continuous such that $\varphi = \eta - \eta \circ f$.*

We also give a rough idea of the proof. By topological transitivity, $\exists x_0 \in \Lambda$ s.t. $\mathcal{O}_f^+(x_0)$ is dense in Λ. We define η on this orbit and extend it to Λ by continuity:

$$\eta(f^n(x_0)) = S_n\varphi(x_0) + \eta(x_0).$$

The thing left is to prove Hölder continuity using hyperbolicity.

As was already mentioned, we are going to guess how we should define η and φ^- (which is future independent) such that

$$\varphi - \varphi^- \circ f = \eta - \eta \circ f$$

which is equivalent to

$$\varphi \circ f^{-1} - \varphi^- = \eta \circ f^{-1} - \eta. \tag{4.2}$$

If we could solve the cohomological equation for $\varphi \circ f^{-1}$ (solution η_1) and for φ^- (solution η_2), then $\eta_1 - \eta_2$ would be a solution of (4.2).

For $\varphi \circ f^{-1}$, η_1 could be

$$\eta_1(a) = -\sum_{n=0}^{\infty} \varphi \circ f \circ f^{-n}(a) = -\sum_{n=1}^{\infty} \varphi \circ f^{-n}(a),\ a \in X \times \mathbb{T}.$$

Then formally we have

$$\varphi \circ f^{-1} = \eta_1 \circ f^{-1} - \eta_1.$$

For φ^-, let us say

$$\varphi^- = \eta_2 \circ f^{-1} - \eta_2$$

where we do not know φ^- but it is related to φ and it is future independent.

The current question is how to get observables that are future independent. Fix the future $p^+ \in X^+$, define

$$P : X \times \mathbb{T} \to X \times \mathbb{T},\ P(\omega^-; \omega^+; \theta) = (\omega^-; p^+; \theta).$$

Then given any $\psi : X \times \mathbb{T} \to \mathbb{R}$, $\psi \circ P$ is future independent.

Knowing this, let formally

$$\eta_2(a) := -\sum_{n=1}^{\infty} \varphi \circ f^{-n} \circ P(a),\ a \in X \times \mathbb{T}.$$

Then

$$\varphi^- := \eta_2 \circ f^{-1} - \eta_2 = \varphi \circ f^{-1} \circ P$$

is future independent.

Let $\eta = \eta_1 - \eta_2$,

$$\eta(\omega,\theta) = \sum_{n=1}^{\infty} \varphi \circ f^{-n}(P(\omega,\theta)) - \varphi \circ f^{-n}(\omega,\theta) = h^u_\varphi((\omega,\theta), P(\omega,\theta)).$$

Note that $(\omega,\theta), P(\omega,\theta) \in W^u_{loc}(\omega^-,\theta)$ since they share the same past x^- and have the same θ.

Remark 4.9. $\sigma^{-n}\omega$ and $\sigma^{-n}(\omega^-,p^+)$ share the same coordinates at least until n, thus $d(\sigma^{-n}\omega, \sigma^{-n}(\omega^-,p^+)) \leq 2^{-(n+1)}$. This implies

$$d_0(f^{-n}(\omega,\theta), f^{-n}(P(\omega,\theta))) \leq 2^{-(n+1)} \leq 2^{-n}.$$

Now, let us show the well-definedness of $\eta = \eta_\varphi$. For any $\varphi \in \mathcal{H}_\alpha(X \times \mathbb{T})$, we have

$$\left|\varphi(f^{-n}(\omega,\theta)) - \varphi(f^{-n}P(\omega,\theta))\right| \leq v^X_\alpha(\varphi) \cdot 2^{-n\alpha}$$

which implies

$$\|\eta_\varphi\|_\infty \leq v^X_\alpha(\varphi) \sum_{n=1}^{\infty} 2^{-n\alpha} \lesssim v^X_\alpha(\varphi).$$

Define φ^- s.t. the homological equation holds:

$$\varphi^- = \eta_\varphi - \eta_\varphi \circ f^{-1} + \varphi \circ f^{-1}.$$

We separate several steps.

Step 1. Note that $\varphi \to \eta_\varphi$ is linear, which gives that $\varphi \to \varphi^-$ is linear also.

Step 2. φ^- is indeed future independent.

$$\varphi^-(a) = h^u_\varphi(a, P(a)) - h^u_\varphi(f^{-1}(a), P(f^{-1}(a)) + \varphi \circ f^{-1}(a).$$

Note that

$$h^u_\varphi(a, P(a)) + \varphi \circ f^{-1}(a) = h^u_\varphi(f^{-1}(a), f^{-1}(P(a))) + \varphi \circ f^{-1}(P(a)).$$

This implies

$$\varphi^-(a) = h^u_\varphi(Pf^{-1}(a), f^{-1}(P(a))) + \varphi \circ f^{-1}(P(a)).$$

Since P fixes the future and P is in every term (where $Pf^{-1}(a)$ does not depend on non-negative coordinates which is even better), φ^- is future independent.

Step 3. $\|\varphi^-\|_\beta \lesssim \|\varphi\|_\alpha$. This is equivalent to saying that $\varphi^- \in \mathcal{H}_\beta$ and $\varphi \to \varphi^-$ is bounded. Write again

$$\varphi^- = \eta_\varphi - \eta_\varphi \circ f^{-1} + \varphi \circ f^{-1}.$$

Since f^{-1} is Lipschitz w.r.t. d_0, it is enough to show that

$$\|\eta_\varphi\|_\beta \lesssim \|\varphi\|_\alpha \,.$$

By definition,

$$\|\eta_\varphi\|_\beta = \|\eta_\varphi\|_\infty + v_\beta^X(\eta_\varphi) + v_\beta^{\mathbb{T}}(\eta_\varphi).$$

But we already know $\|\eta_\varphi\|_\infty \lesssim v_\alpha^X(\varphi) \leq \|\varphi\|_\alpha$.

It is enough to show that respectively $v_\beta^X(\eta_\varphi) \lesssim \|\varphi\|_\alpha$ and $v_\beta^{\mathbb{T}}(\eta_\varphi) \lesssim \|\varphi\|_\alpha$. We prove them one by one.

Rewrite

$$\eta_\varphi(\omega,\theta) = \sum_{n=1}^{\infty} \varphi \circ f^{-n} \circ P(\omega,\theta) - \varphi \circ f^{-n}(\omega,\theta) =: \sum_{n=1}^{\infty} g_n(\omega,\theta).$$

Fix $x, y \in X$, assume that $x_j = y_j$ until $|j| \leq k$, then $d(x,y) = 2^{-(k+1)} \leq 2^{-k}$. We want to show that for any $\theta \in \mathbb{T}$ and any $k \in \mathbb{N}$

$$|\eta_\varphi(x,\theta) - \eta_\varphi(y,\theta)| \lesssim v_\alpha^X(\varphi) 2^{-k\beta},$$

which will imply $v_\beta^X(\eta_\varphi) \lesssim v_\alpha^X(\varphi) \leq \|\varphi\|_\alpha$.

By triangle inequality, we have (without loss of generality, assume k is even, otherwise just take the integer part of $k/2$)

$$|\eta_\varphi(x,\theta) - \eta_\varphi(y,\theta)| \leq \sum_{n=1}^{\frac{k}{2}} |g_n(x,\theta) - g_n(y,\theta)| + \sum_{n>\frac{k}{2}} |g_n(x,\theta) - g_n(y,\theta)| \,.$$

We analyze the right hand side separately.

$$\sum_{n>\frac{k}{2}} |g_n(x,\theta) - g_n(y,\theta)| \leq \sum_{n>\frac{k}{2}} |g_n(x,\theta)| + \sum_{n>\frac{k}{2}} |g_n(y,\theta)| \lesssim \sum_{n>\frac{k}{2}} v_\alpha^X(\varphi) 2^{-n\alpha}.$$

Since $n > \frac{k}{2}$, then $2n > k$. Let $\beta = \frac{\alpha}{3}$, we have

$$n\alpha = 3n\beta = 2n\beta + n\beta > k\beta + n\beta,$$

and then

$$2^{-n\alpha} < 2^{-k\beta} \cdot 2^{-n\beta}.$$

Hence

$$\sum_{n>\frac{k}{2}} |g_n(x,\theta) - g_n(y,\theta)| \lesssim v_\alpha^X(\varphi) 2^{-k\beta} \sum_{n\geq 0} 2^{-n\beta} \lesssim v_\alpha^X(\varphi) 2^{-k\beta}.$$

On the other hand, for $\sum_{n=1}^{\frac{k}{2}} |g_n(x,\theta) - g_n(y,\theta)|$, we know $2n \leq k$,

$$2^{-(k+n)} \leq d(\sigma^{-n}x, \sigma^{-n}y) \leq 2^{-(k-n)}$$

and

$$2^{-(k+n)} \leq d_0(f^{-n}(x,\theta), f^{-n}(y,\theta)) \leq 2^{-(k-n)}.$$

Therefore,

$$\left|\varphi \circ f^{-n}(x,\theta) - \varphi \circ f^{-n}(y,\theta)\right| \leq v_\alpha^X(\varphi) 2^{-(k-n)\alpha}$$

as $\varphi \in \mathcal{H}_\alpha$. The same estimate holds with $P(x,\theta), P(y,\theta)$ instead of $(x,\theta), (y,\theta)$, which implies

$$|g_n(x,\theta) - g_n(y,\theta)| \lesssim v_\alpha^X(\varphi) 2^{-(k-n)\alpha}.$$

Hence, because of $k \geq 2n$ thus $(k-n)\alpha \geq n\alpha$, we have

$$\sum_{n > \frac{k}{2}} |g_n(x,\theta) - g_n(y,\theta)| \lesssim v_\alpha^X(\varphi) \sum_{n=1}^{\frac{k}{2}} 2^{-(n-k)\alpha} \lesssim v_\alpha^X(\varphi) \sum_{n=1}^{\frac{k}{2}} 2^{-k\beta} \cdot 2^{-n\beta}$$

which again shows

$$\sum_{n > \frac{k}{2}} |g_n(x,\theta) - g_n(y,\theta)| \lesssim v_\alpha^X(\varphi) 2^{-k\beta}.$$

Combining the previous estimate, we have

$$|\eta_\varphi(x,\theta) - \eta_\varphi(y,\theta)| \lesssim v_\alpha^X(\varphi) 2^{-k\beta}$$

for every $\theta \in \mathbb{T}$ and every $k \in \mathbb{N}$, which further implies

$$v_\beta^X(\eta_\varphi) \lesssim v_\alpha^X(\varphi) \leq \|\varphi\|_\alpha .$$

Following the same strategy, we also have $v_\beta^{\mathbb{T}}(\eta_\varphi) \lesssim \|\varphi\|_\alpha$. More precisely, fix any $x \in X$ and for $\theta, \theta' \in \mathbb{T}$, let $N \in \mathbb{N}$ such that $|\theta - \theta'| \approx 2^{-N}$, so $N \approx \log \frac{1}{|\theta - \theta'|}$.

Like before, we have

$$\sum_{n \geq N} |g_n(x,\theta)| \leq \sum_{n \geq N} v_\alpha^X(\varphi) 2^{-n\alpha} \lesssim v_\alpha^X(\varphi) \left|\theta - \theta'\right|^\alpha .$$

The same estimate holds for (x,θ'). Therefore,

$$\sum_{n \geq N} |g_n(x,\theta) - g_n(x,\theta')| \lesssim v_\alpha^X(\varphi) \left|\theta - \theta'\right|^\alpha \leq \|\varphi\|_\alpha \left|\theta - \theta'\right|^\alpha .$$

On the other hand,

$$\left|\varphi \circ f^{-n}(x,\theta) - \varphi \circ f^{-n}(x,\theta')\right| \leq v_\alpha^{\mathbb{T}}(\varphi) \left|\theta - \theta'\right|^\alpha$$

and the same estimate holds for $P(x,\theta), P(x,\theta')$. So we get

$$|g_n(x,\theta) - g_n(x,\theta')| \lesssim v_\alpha^{\mathbb{T}}(\varphi) \left|\theta - \theta'\right|^\alpha .$$

This implies

$$\sum_{0 \leq n < N} |g_n(x,\theta) - g_n(x,\theta')| \lesssim v_\alpha^{\mathbb{T}}(\varphi) \left|\theta - \theta'\right|^\alpha \log \frac{1}{|\theta - \theta'|} \lesssim \|\varphi\|_\alpha \left|\theta - \theta'\right|^{\alpha^-}$$

and thus

$$|\eta_\varphi(x,\theta) - \eta_\varphi(x,\theta')| \lesssim \|\varphi\|_\alpha |\theta-\theta'|^{\alpha^-} + \|\varphi\|_\alpha |\theta-\theta'|^\alpha \leq \|\varphi\|_\alpha |\theta-\theta'|^\beta .$$

This shows $v_\beta^{\mathbb{T}}(\eta_\varphi) \lesssim \|\varphi\|_\alpha$. Note that actually here $\beta = \alpha^-$ but we do not make use of it.

To conclude, $\|\eta_\varphi\|_\beta \lesssim \|\varphi\|_\alpha$ for $\beta = \frac{\alpha}{3}$. Thus $\eta_\varphi \in \mathcal{H}_\beta(X\times\mathbb{T})$ and $\varphi^- \in \mathcal{H}_\beta(X\times\mathbb{T})$ which is future independent, namely $\varphi^- \in \mathcal{H}_\beta(X^-\times\mathbb{T})$. Also $\varphi \to \varphi^-$ is bounded.

This finishes the whole proof of Proposition 4.12.

Remark 4.10. For $v_\beta^{\mathbb{T}}$ semi-norm, it almost has no loss since we can choose $\beta = \alpha^-$. Note that our choice of $\beta = \frac{\alpha}{3}$ is not optimal. The sharp estimate of β is left to the readers.

5. Statistical properties for certain dynamical systems via the transfer operator

By "certain dynamical systems" we mean (mostly) the simplest possible models, the objective being to illustrate the method.

Recall that we already obtain statistical properties like LDT and CLT for certain systems via the Markov operator in the previous section. We managed to sort of fit certain deterministic dynamical systems (DDS) into the abstract probability scheme. For instance,

(1) Mixed random-quasiperiodic base dynamics.
(2) Random linear cocycles (locally constant).

Other models that sort of fit the scheme:

(3) Cocycles over a uniformly hyperbolic base dynamics and the fiber dynamics are partially hyperbolic when projectivized (Duarte-Klein-Poletti).
(4) Cocycles over mixed random-quasiperiodic base dynamics (Cai-Duarte-Klein).

An important aspect in all of these models is the use of coding (symbolic dynamics), the shift.

We will present a brief introduction to the use of the transfer operator via the functional approach to the study of statistical properties for certain DDS without coding.

Transfer operator encodes the action of a DS on mass densities of initial conditions. Let $(M, \mathcal{B}, m)$ be a Borel probability space where m is the reference measure. Let $f : M \to M$ be continuous (non-invertible) and non-singular in the sense that $m(E) = 0 \Rightarrow m(f^{-1}(E)) = 0, \forall\, E \in \mathcal{B}$.

Example. $(\mathbb{T}, \mathcal{B}, m)$ where m is the Lebesgue measure. $f(x) = 2x \bmod 1$ which is the doubling map.

Start with a density function $h \in L^1(dm)$, consider the measure $dm_h = hdm$ and $f_* m_h = m_h \circ f^{-1}$.

Remark 5.1. Note that $f_* m_h \ll m$. Because if $m(E) = 0$, then $m(f^{-1}(E)) = 0$ and $f_* m_h(E) = m_h(f^{-1}(E)) = \int_{f^{-1}(E)} hdm = 0$.

Then we naturally have the following definition.

Definition 5.1. $\mathcal{L} : L^1(dm) \to L^1(dm)$,

$$\mathcal{L}h = \frac{df_* m_h}{dm} = \text{Radon-Nikodym derivative of } f_* m_h \text{ w.r.t. } m.$$

In fact, we have an equivalent characterization of $\mathcal{L}$.

Proposition 5.1. *Given $h \in L^1(dm)$, $\mathcal{L}h$ is characterized by*

$$\forall \varphi \in L^\infty(dm), \quad \int \varphi \cdot \mathcal{L}hdm = \int (\varphi \circ f)hdm$$

in the sense that $\mathcal{L}h$ is the unique function in $L^1(dm)$ such that the equation holds.

Proof.

$$\mathcal{L}h = \frac{df_* m_h}{dm} \Rightarrow \mathcal{L}hdm = df_* m_h.$$

This shows $\forall \varphi \in L^\infty(dm)$,

$$\int \varphi \cdot \mathcal{L}hdm = \int \varphi df_* m_h = \int \varphi \circ f dm_h = \int (\varphi \circ f)hdm.$$

For uniqueness, if $\psi_1, \psi_2 \in L^1(dm)$ s.t.

$$\int (\varphi \circ f)hdm = \int \varphi \psi_1 dm = \int \varphi \psi_2 dm, \quad \forall \varphi \in L^\infty(M),$$

then by simple measure theory, $\psi_1 = \psi_2$ m-a.e. □

Here are some properties of $\mathcal{L}$.

Proposition 5.2. *$\mathcal{L}$ is a linear operator. It is positive: if $h \geq 0$ then $\mathcal{L}h \geq 0$. It is also bounded with norm 1 on $L^1(dm)$.*

Proof. Linearity follows from the characterization. Positivity is also clear because it is the Radon-Nikodym derivative of two positive measures.

We want to show that $\|\mathcal{L}h\|_1 \leq \|h\|_1$. For our purpose, we first show that if $h \geq 0$, then $\|\mathcal{L}h\|_1 = \|h\|_1$.

Fix $h \geq 0$, by the characterization, if $\varphi = \mathbb{1}$ is a constant function, then

$$\|\mathcal{L}h\|_1 = \int \mathcal{L}h dm = \int \mathbb{1} \cdot \mathcal{L}h dm = \int (\mathbb{1} \circ f) h dm = \int h dm = \|h\|_1 .$$

In general, let $h \in L^1(dm)$, then $|h| \in L^1(dm)$ and $|h| \geq 0$. So by linearity and positivity we have both $\mathcal{L}h \leq \mathcal{L}|h|$ and $\mathcal{L}(-h) \leq \mathcal{L}|h|$. This implies $|\mathcal{L}h| \leq \mathcal{L}|h|$. Hence

$$\|\mathcal{L}h\|_1 = \int |\mathcal{L}h| \, dm \leq \int \mathcal{L}|h| \, dm = \int |h| \, dm = \|h\|_1 .$$

This finished the proof. □

Example. Doubling map $f(x) = 2x \bmod 1$ on $[0,1]$, $h \in L^1(dm)$,

$$\mathcal{L}h(x) = \frac{1}{2}\left[h(\frac{x}{2}) + h(\frac{x+1}{2})\right].$$

Proof. For $\varphi \in L^\infty(dm)$,

$$\begin{aligned}
\int_0^1 \varphi(x)\mathcal{L}h(x)dx &= \int_0^1 \varphi(2x \bmod 1)h(x)dx \\
&= \int_0^{\frac{1}{2}} \varphi(2x)h(x)dx + \int_{\frac{1}{2}}^1 \varphi(2x-1)h(x)dx \\
&= \frac{1}{2}\int_0^1 \varphi(y)h(\frac{y}{2})dy + \frac{1}{2}\int_0^1 \varphi(y)h(\frac{y+1}{2})dy \\
&= \int_0^1 \varphi(y)\left[\frac{1}{2}\left(h(\frac{y}{2}) + h(\frac{y+1}{2})\right)\right] dy.
\end{aligned}$$

□

Let us proceed with the smooth expanding maps of the torus.

Let $f \in C^r(\mathbb{T}, \mathbb{T})$ with $r \geq 2$. Assume that $|f'(x)| \geq \lambda_* > 1, \forall\, x \in \mathbb{T}$. As before, denote by m the Lebesgue measure. (Or more generally, consider M a compact, connected Riemannian manifold, $f : M \to M$ smooth. $\forall\, x \in M, \forall\, v \in T_xM$, $|Df_x(v)| \geq \lambda_* \|v\|, \lambda_* > 1$).

By the derivative assumption, all $x \in \mathbb{T}$ have the same (via inverse function theorem, connectedness and compactness of $\mathbb{T}$) finite number n of preimages. Moreover, there is an open partition of $\mathbb{T}$:

$$\{I_1, \cdots, I_n\} \text{ s.t. } \bigcup_{j=1}^n \bar{I}_j = \mathbb{T}$$

with each I_j open, such that every

$$f|_{I_j} : I_j \to \mathbb{T}\backslash\{0\}$$

is a bijection. Let

$$g_j := f|_{I_j}^{-1} : \mathbb{T}\backslash\{0\} \to I_j.$$

For any test function $\varphi \in L^\infty(dm)$,

$$\begin{aligned}\int_{\mathbb{T}} \varphi \mathcal{L} h dm &= \int_{\mathbb{T}} (\varphi \circ f) h dm \\ &= \sum_{j=1}^{n} \int_{I_j} (\varphi \circ g_j^{-1})(y) h(y) dy \\ &= \sum_{j=1}^{n} \int_{\mathbb{T}} \varphi(x) \frac{h(g_j(x))}{|f'(g_j(x))|} dx \\ &= \int_{\mathbb{T}} \varphi(x) \sum_{j=1}^{n} \frac{h(g_j(x))}{|f'(g_j(x))|} dx.\end{aligned}$$

Here in the third equality we used the change of variables $x = g_j^{-1}(y)$. Then $y = g_j(x)$ and $dy = \left|g_j'(x)\right| dx = \frac{1}{|f'(g_j(x))|} dx$. Moreover, for every $x \in \mathbb{T}\backslash\{0\}$, $f^{-1}(x) = \{g_j(x) : j = 1, \cdots, n\}$. Hence, we have proved that

$$\mathcal{L}h(x) = \sum_{y:f(y)=x} \frac{1}{|f'(y)|} h(y).$$

We are interested in finding invariant measures for $(\mathbb{T}, f)$ where $f : \mathbb{T} \to \mathbb{T}$ is a differentiable topological dynamical system.

We would like to see it as an MPDS, so we need to consider an f-invariant Borel probability measure on $\mathbb{T}$. There are plenty of such measure, e.g. if $f^m(p) = p$, then $\frac{1}{m}\sum_{j=0}^{m-1} \delta_{f^j(p)}$ is f-invariant.

Certainly, one type of interesting f-invariant measure is the measure which is absolutely continuous with respect to the Lebesgue measure m. The following proposition gives a characterization of a.c. measures.

Proposition 5.3. *$d\mu_0 := h_0 dm$ is an f-invariant measure $\Leftrightarrow \mathcal{L}h_0 = h_0$. In other words, the a.c. f-invariant measures correspond to the eigenvector of $\mathcal{L}$ with eigenvalue 1.*

Proof. μ_0 is f-invariant $\Leftrightarrow \int \varphi \circ f d\mu_0 = \int \varphi d\mu_0, \forall \varphi \in L^\infty(dm)$. This is equivalent to saying that

$$\int (\varphi \circ f) h_0 dm = \int \varphi h_0 dm$$

$\Leftrightarrow$

$$\int \varphi \cdot \mathcal{L} h_0 dm = \int \varphi h_0 dm$$

$\Leftrightarrow$

$$\mathcal{L}h_0 = h_0.$$

$\square$

Our goal is to study the spectral properties of the transfer operator $\mathcal{L}$. It turns out that the spectrum of $\mathcal{L}$ on L^1 is

$$\bar{\mathbb{D}} = \{z \in \mathbb{C} : |z| \leq 1\}.$$

For this purpose, let us introduce the Lasota-Yorke inequalities.
Recall that

$$\mathcal{L}h(x) = \sum_{y:f(y)=x} \frac{h(y)}{|f'(y)|}.$$

Moreover, $f^{-1}\{x\} = \{y_1, \cdots, y_m\}$, $g_j = f|_{I_j}^{-1} : \mathbb{T}\backslash\{0\} \to I_j$ such that $g_j(x) = y_j$. Since

$$g_j'(x) = \frac{1}{f'(g_j(x))},$$

we have

$$\mathcal{L}h(x) = \sum_{j=1}^{m} \frac{h(g_j(x))}{|f'(g_j(x))|}.$$

Assume that the derivative of h, h' exists, we want to find $(\mathcal{L}h)'$. Direct computation shows that

$$(\mathcal{L}h)' = \mathcal{L}(h' \cdot \frac{1}{f'}) - \mathcal{L}(h \cdot \frac{f''}{(f')^2}),$$

provided $h \in L^1$, h' exists and $h' \in L^1$.

Recall that if $\varphi \in L^1$, $\psi \in L^\infty \Rightarrow \varphi\psi \in L^1$ and $\|\varphi\psi\|_1 \leq \|\varphi\|_1 \|\psi\|_\infty$. As $|f'(x)| \geq \lambda_* > 1, \forall\, x \in \mathbb{T}$, we have

$$\left|\frac{1}{f'}\right| \leq \frac{1}{\lambda_*} < 1, \quad \sup_{\mathbb{T}} \left|\frac{f''}{(f')^2}\right| =: D \text{ is called the distortion of f.}$$

In particular, for the doubling map $f(x) = 2x \bmod 1$, $D = 0$.

Let

$$\mathcal{W}^{1,1}(\mathbb{T}) := \left\{h : \mathbb{T} \to \mathbb{R}, h \in L^1, h' \text{ exists a.e. and } h' \in L^1\right\}$$

be the Sobolev space.

We already know that $\|\mathcal{L}g\|_1 \leq \|g\|_1$, so

$$\left\|\mathcal{L}(h' \cdot \frac{1}{f'})\right\|_1 \leq \left\|h' \cdot \frac{1}{f'}\right\|_1 \leq \lambda_*^{-1} \|h'\|_1 < \|h'\|_1.$$

Note that $\mathcal{W}^{1,1}(\mathbb{T}) \ni h \to \|h'\|_1$ is a seminorm. Therefore, on $\mathcal{W}^{1,1}(\mathbb{T})$ we consider the norm

$$\|h\|_{1,1} = \|h'\|_1 + a\|h\|_1\,,\ a > 0.$$

The inequalities

$$\|\mathcal{L}h\|_1 \le \|h\|_1\,,$$
$$\|(\mathcal{L}h)'\|_1 \le \lambda_*^{-1}\|h'\|_1 + D\|h\|_1\,,$$

are called Lasota-Yorke type inequalities. In particular, they show that $\mathcal{L}$ is bounded on the Sobolev space $\mathcal{W}^{1,1}(\mathbb{T})$. If $h \in \mathcal{W}^{1,1}$,

$$\begin{aligned}
\|\mathcal{L}h\|_{1,1} &= \|(\mathcal{L}h)'\|_1 + a\|\mathcal{L}h\|_1 && (5.1)\\
&\le \lambda_*^{-1}\|h'\|_1 + (D+a)\|h\|_1 && (5.2)\\
&\le \lambda_*^{-1}\|h\|_{1,1} + \frac{D+a}{a}\|h\|_{1,1} && (5.3)\\
&\le C\|h\|_{1,1}\,, && (5.4)
\end{aligned}$$

where $C = 2\max\left\{\lambda_*^{-1}, \frac{D+a}{a}\right\}$.

Let us summarize: $(\mathcal{W}^{1,1}, \|\cdot\|_{1,1}) \hookrightarrow (L^1, \|\cdot\|_1)$ and $\mathcal{L}$ is bounded on both spaces. The Lasota-Yorke inequalities are

$$\begin{cases} \|\mathcal{L}h\|_1 \le \|h\|_1\,, \forall\, h \in L^1, \\ \|\mathcal{L}h\|_{1,1} \le \lambda_*^{-1}\|h'\|_{1,1} + C'\|h\|_1\,. \end{cases}$$

By induction, $\forall\, n \in \mathbb{N}$ we have

$$\begin{cases} \|\mathcal{L}^n h\|_1 \le \|h\|_1\,, \\ \|\mathcal{L}^n h\|_{1,1} \le (\lambda_*^{-1})^n\|h'\|_{1,1} + C''\|h\|_1\,. \end{cases}$$

By a theorem of Hennion (Ionescu-Tulcea & Marinescu), it turns out that $\mathcal{L}$ is quasi-compact on $\mathcal{W}^{1,1}$ with essential spectral radius less than or equal to $\lambda_*^{-1} < 1$.

Let us consider a particular situation when we do not need the force of Hennion: $\lambda_*^{-1} + D < 1$. For doubling map, $\lambda_* = 2$ and $D = 0$ so it is satisfied.

We take $\|h\|_{1,1} = \|h'\|_1 + a\|h\|_1$ where $a > 0$ such that

$$\lambda_*^{-1} + D + a < 1.$$

Then $\mathcal{L}$ will be a strict contraction on some subspace

$$\mathcal{V} = \left\{ h \in \mathcal{W}^{1,1} : \int h dm = 0 \right\}.$$

Note that $\mathcal{V}$ is $\mathcal{L}$-invariant.

If $h \in \mathcal{V}$, then $\int hdm = 0$. Since h is continuous, then $\exists\, x_0 \in \mathbb{T}$ s.t. $h(x_0) = 0$. Hence, if $h \in \mathcal{V}$,

$$\begin{aligned} \|h\|_1 &= \int_{\mathbb{T}} |h(x)|\, dx = \int_{\mathbb{T}} \left| \int_{x_0}^{x} h'(y)dy \right| dx \\ &\leq \int_{\mathbb{T}} \left(\int_{x_0}^{x} |h'(y)|\, dy \right) dx \leq \int_{\mathbb{T}} \|h'\|_1\, dx \leq \|h'\|_1 . \end{aligned}$$

Then, for $h \in \mathcal{V}$,

$$\begin{aligned} \|\mathcal{L}h\|_{1,1} &\leq \lambda_*^{-1} \|h'\|_1 + (D + a) \|h\|_1 \\ &\leq (\lambda_*^{-1} + D + a) \|h'\|_1 \\ &\leq (\lambda_*^{-1} + D + a) \|h\|_{1,1} \\ &\leq \sigma \|h\|_{1,1} \end{aligned}$$

where $\sigma < 1, \forall\, h \in \mathcal{V}$. This shows the contracting of $\mathcal{L}$ on $\mathcal{V}$.

Theorem 5.1. *There exists a unique absolutely continuous f-invariant probability measure $d\mu_* = h_* dm$. Moreover, if $h \in \mathcal{W}^{1,1}$ then $\exists\, \sigma < 1$ such that $\forall\, n \in \mathbb{N}$*

$$\left\| \mathcal{L}^n h - \left(\int hdm \right) h_* \right\|_{1,1} \leq \sigma^n \left\| h - \left(\int hdm \right) h_* \right\|_{1,1} \leq C\sigma^n \|h\|_{1,1} .$$

Proof. Consider the dual operator $\mathcal{L}^*$, $\mathcal{L}^* m = m \Leftrightarrow$

$$\langle \varphi, \mathcal{L}^* m \rangle = \langle \varphi, m \rangle, \forall\, \varphi \in C^0(\mathbb{T})$$

$\Leftrightarrow$

$$\langle \mathcal{L}\varphi, m \rangle = \langle \varphi, m \rangle, \forall\, \varphi \in C^0(\mathbb{T})$$

$\Leftrightarrow$

$$\int \mathcal{L}\varphi dm = \int \varphi dm, \forall\, \varphi \in C^0(\mathbb{T})$$

which is true by the characterization of $\mathcal{L}$. This shows that 1 is an eigenvalue of $\mathcal{L}^*$ which is equivalent to saying that 1 is an eigenvalue of $\mathcal{L}$ (in finite dimensional case this is true without any assumption but here it is indeed true because of the quasi-compactness of $\mathcal{L}$, we will clarify this in detail later). Namely, $\exists\, h_* \in L^1$ s.t. $\mathcal{L}h_* = h_*$, which implies that $h_* dm$ is f-invariant. This shows the existence of f-invariant absolutely continuous measure.

For the uniqueness, assume that the inequality in the theorem holds. Then if $d\mu = \varphi dm$ is another f-invariant probability measure, then $\int \varphi dm = 1$ and $\mathcal{L}\varphi = \varphi$. Moreover, by the inequality,

$$\left\| \mathcal{L}^n \varphi - (\int \varphi dm) h_* \right\|_{1,1} \leq C\sigma^n \|h\|_{1,1} \to 0$$

as $n \to \infty$. This shows $\|\varphi - h_*\|_{1,1} \to 0$ which gives $\varphi = h_*$.

Finally, let us prove the inequality. Consider any $h \in \mathcal{W}^{1,1}$, let

$$\mathcal{N}h := \left(\int hdm\right) h_*, \quad \mathcal{R}h := \mathcal{L}(h - \mathcal{N}h).$$

Note that $h - \mathcal{N}h \in \mathcal{V}$ because

$$\int h - \mathcal{N}hdm = \int hdm - \int \mathcal{N}hdm = 0.$$

Then

$$\mathcal{L}h = \mathcal{R}h + \mathcal{N}h = \mathcal{R}h + (\int hdm)\mathcal{L}h_* = \mathcal{R}h + (\int hdm)h_*.$$

Since $\mathcal{R} \circ \mathcal{N} = \mathcal{N} \circ \mathcal{R} = 0$, then $\mathcal{L} = \mathcal{R} \oplus \mathcal{N}$ where $\mathcal{N}^2 = \mathcal{N}$. So

$$\mathcal{L}^n h = \mathcal{R}^n h + (\int hdm)h_*.$$

Hence

$$\mathcal{R}^n h = \mathcal{L}^n h - (\int hdm)h_*.$$

It follows that

$$\left\|\mathcal{L}^n h - (\int hdm)h_*\right\|_{1,1} = \|\mathcal{R}^n h\|_{1,1} = \left\|\mathcal{L}^n \left(h - (\int hdm)h_*\right)\right\|_{1,1}.$$

As $h - (\int hdm)h_* \in \mathcal{V}$, we have

$$\left\|\mathcal{L}^n h - (\int hdm)h_*\right\|_{1,1} \leq \sigma^n \left\|h - (\int hdm)h_*\right\|_{1,1}$$

This finishes the proof. □

Remark 5.2. $\mathcal{L}^n \mathbb{1} \to h_*$ as $n \to \infty$.

5.1. **Hennion's Theorem.** In the following, we present the theorem of Hennion as well as the proof of it.

Theorem 5.2 (Hennion 1993)**.** *Let $\mathcal{B} \subset \mathcal{B}_\omega$ be two Banach spaces, $\|\cdot\|$ and $\|\cdot\|_\omega$ being the respective norm, satisfying $\|\cdot\|_\omega \leq \|\cdot\|$. In addition, let $T : \mathcal{B} \to \mathcal{B}$ be a linear operator s.t. $\exists\, m, C, \theta > 0, \theta < m$ and $n_0 \in \mathbb{N}$ s.t. $T^{n_0} : \mathcal{B} \to \mathcal{B}_\omega$ is a compact operator and for each $n \in \mathbb{N}$ and $v \in \mathcal{B}$,*

$$\|T^n v\|_\omega \leq Cm^n \|v\|_\omega,$$

$$\|T^n v\| \leq C\theta^n \|v\| + Cm^n \|v\|_\omega.$$

Then the spectral radius of T is bounded by m and its essential spectral radius is bounded by θ.

Before the proof, there are various lemmas and definitions to be given.

Denote by $\mathcal{L}(\mathcal{B},\mathcal{B})$ the set of bounded linear operators from $\mathcal{B}$ to itself. For each $A \in \mathcal{L}(\mathcal{B},\mathcal{B})$, we denote by $\mathcal{R}(A)$ the range of A and by $\mathcal{N}(A)$ the kernal of A.

Definition 5.2. We say that an operator $P \in \mathcal{L}(\mathcal{B},\mathcal{B})$ is a projection if and only if $P^2 = P$.

Lemma 5.4. *If P is a projection, then $\mathcal{B} = \mathcal{N}(P) \oplus \mathcal{R}(P)$.*

Definition 5.3. An operator $K \in \mathcal{L}(\mathcal{B},\mathcal{B})$ is compact if and only if for any bounded set D, the closure of $k(D)$ is compact.

Definition 5.4. Given $A \in \mathcal{L}(\mathcal{B},\mathcal{B})$, we define the resolvent set of A as

$$\rho(A) := \{z \in \mathbb{C} : z\mathbb{1} - A \,\text{has bounded inverse}\}\,,$$

and the spectrum of A as $\sigma(A) = \mathbb{C}\backslash\rho(A)$. For simplicity, in the following we omit "$\mathbb{1}$".

We define the spectral radius of A as

$$r(A) = \lim_{n\to\infty} \|A^n\|^{\frac{1}{n}}\,.$$

Definition 5.5 (Essential spectrum)**.** Let $T \in \mathcal{L}(\mathcal{B},\mathcal{B})$, the essential spectrum of T, denoted by $\sigma_{ess}(T)$ is the set of $\lambda \in \sigma(T)$ such that at least one of the following conditions holds:

(1) $\mathcal{R}(\lambda - T)$ is not closed.
(2) $\cup_{n\geq 1}\mathcal{N}(\lambda - T)^n$ is infinite dimensional.
(3) λ is a limit point of $\sigma(T)\backslash\{\lambda\}$.

Lemma 5.5. *Let $\mathcal{B}$ be a Banach space, $\mathcal{V} \subset \mathcal{B}$ a proper closed subspace. Then for every $\epsilon > 0$, there exists $x_0 \in \mathcal{B}$ with $\|x_0\| = 1$ and* $\operatorname{dist}(x_0, \mathcal{V}) \geq 1 - \epsilon$.

Definition 5.6 (Proper)**.** A continuous map $F : U \subset X \to Y$ between topological spaces is called proper if $F^{-1}(M)$ is compact whenever $M \subset Y$ is compact.

Theorem 5.3. *Every locally compact space X has finite dimension.*

Lemma 5.6. *Let X and Y be Banach spaces and $S \in \mathcal{L}(X, Y)$. If S restricted to closed bounded set is proper, then $\mathcal{N}(S)$ is finite dimensional and $\mathcal{R}$ is closed.*

A more detailed version of this lemma with proof will appear later.
let $\mathcal{B}$ be a Banach space and A be a bounded set of $\mathcal{B}$.

Definition 5.7 (Measure of noncompactness)**.** The measure of noncompactness of $A \subset \mathcal{B}$ is defined as the infimum of $d > 0$ such that there exists a finite number of sets $S_1, \cdots, S_n$ with $\mathrm{diam}(S_i) \leq d, \forall\, i = 1, \cdots, n$ and $A \subset \cup_{i=1}^n S_i$. Moreover, we denote the measure of noncompactness of A by $r(A)$.

Definition 5.8 (Ball measure of noncompactness)**.** The ball measure of noncompactness, $\tilde{r}(A)$ is the infimum of $d > 0$ such that there exists a finite number of balls $B_1, \cdots, B_n$ with centers in B with radius d and $A \subset \cup_{i=1}^n B_i$.

Remark 5.3. It is obvious that $r(A) \leq \tilde{r}(A)$.

Definition 5.9 (K-set(-ball) contraction)**.** Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be Banach spaces. We say that $T \in \mathcal{L}(\mathcal{B}_1, \mathcal{B}_2)$ is a K-set contraction if $\forall\, A \subset \mathcal{B}_1$ bounded,

$$r_{\mathcal{B}_2}(T(A)) \leq K r_{\mathcal{B}_1}(A).$$

T is called a K-set-ball contraction if

$$\tilde{r}_{\mathcal{B}_2}(T(A)) \leq K \tilde{r}_{\mathcal{B}_1}(A).$$

We define

$$r(T) = \inf\{K > 0 : T \text{ is a K-set contraction}\}.$$

$$\tilde{r}(T) = \inf\{K > 0 : T \text{ is a K-set-ball contraction}\}.$$

Lemma 5.7. *We have the following properties.*

(1) *If $A \subset \mathcal{B}$, then $\bar{A}$ is compact iff $\tilde{r}(A) = 0$. Also, $\bar{A}$ is compact iff $r(A) = 0$.*
(2) *$T \in \mathcal{L}(\mathcal{B}, \mathcal{B})$ is compact iff $\tilde{r}(T) = 0$. Also, T is compact iff $r(T) = 0$.*
(3) *$r(T) \leq \|T\|$.*
(4) *For bounded sets $A, B \in \mathcal{B}$, we have*

$$r(A + B) \leq r(A) + r(B)$$

and

$$\tilde{r}(A + B) \leq \tilde{r}(A) + \tilde{r}(B).$$

Lemma 5.8 (B3, Liverani)**.** *Let X, Y be Banach spaces, $S \in \mathcal{L}(X, Y)$, then the following are equivalent:*

(1) *For any $B \subset X$ closed and bounded, $S|_B$ is proper (i.e. $K \subset Y$ compact implies that $S^{-1}(K) \cap B \subset X$ is compact).*
(2) *If $\{x_n\} \subset X$ is bounded and $Sx_n \to y$, then $\exists\, \{x_{n_k}\}$ which converges.*
(3) *$\mathcal{N}(S) = \ker(S)$ is finite dimensional and $\mathcal{R}(S) = range(S)$ is closed.*

Proof. We give the proof in the following order.

(1) $\Rightarrow$ (2).

$Sx_n \to y$ implies that $A := \{Sx_n : n \geq 1\} \cup \{y\} \subset Y$ is compact, which by properness further implies that $\{x_n\} \subset S^{-1}(A) \cap B$ is compact. Thus $\exists \{x_{n_k}\}$ which converges.

(2) $\Rightarrow$ (1).

Fix $B \subset X$ closed and bounded, $K \subset Y$ compact. Take any $\{x_n\} \subset S^{-1}(K) \cap B \subset X$, then $\{x_n\}$ is bounded. Note also that $\{Sx_n\} \subset K$ which is compact, thus $\exists\, Sx_{n_k} \to y$. Then by (2), $\exists \{x_{n_{k_l}}\}$ which converges. This proves $S^{-1}(K) \cap B \subset X$ is compact, hence (1) holds.

(3) $\Rightarrow$ (2).

If $F \subset X$ is finite dimensional, then there exists a closed complement C of F in X such that $X = F \oplus C$ (because if P is the projection on F, then $C = \ker P$). This is only valued when F has finite dimension.

By assumption, $\mathcal{N}(S)$ having finite dimensional implies that $\exists\, C$ closed s.t. $X = \mathcal{N}(S) \oplus C$. $\mathcal{R}(S)$ is closed in Y which is a Banach space, so $\mathcal{R}(S)$ is also Banach with the induced norm from Y. Also, C is Banach as it is closed. Then

$$S|_C : C \to R(S)$$

is bounded linear and surjective. By Banach open mapping theorem, $S|_C^{-1}$ is continuous and bounded.

Let $\{x_n\} \subset X$ be bounded and $Sx_n \to y$. Write

$$x_n = a_n + c_n, \quad a_n \in \mathcal{N}(S),\, c_n \in C.$$

Then

$$Sx_n = 0 + Sc_n \to y \in \mathcal{R}(S),$$

which implies $c_n \to S^{-1}y$ and c_n is bounded.

On the other hand,

$$a_n = x_n - c_n$$

with x_n bounded and c_n bounded. So a_n is also bounded $\forall\, n \in \mathbb{N}$. Since $\{a_n\} \subset \mathcal{N}(S)$ is finite dimensional, then $\exists \{a_{n_k}\}$ which converges to $a \in \mathcal{N}(S)$. Therefore, we have

$$X_{n_k} \to a + S^{-1}y.$$

This proves (2).

(2) $\Rightarrow$ (3).

$\mathcal{N}(S)$ is a closed linear subspace of X so it is Banach. We are going to show that $\mathcal{N}(S)$ is locally compact thus is finite dimensional. Note that $\mathcal{N}(S)$ being locally compact is equivalent to saying that $\forall \{x_n\}$ bounded in $\mathcal{N}(S)$, it has a convergent subsequence. Note that

$x_n \in \mathcal{N}(S)$ means $Sx_n \equiv 0 \to 0$. Thus by hypothesis, $\exists\, x_{n_k}$ converges. We conclude that $\mathcal{N}(S)$ is finite dimensional.

Since $\mathcal{N}(S)$ is finite dimensional, $\exists\, C \subset X$ closed s.t. $X = \mathcal{N}(S) \oplus C$. We want to show that $\mathcal{R}(S)$ is closed.

Take $\{Sx_n\} \subset \mathcal{R}(S)$ such that $Sx_n \to y$, we want to prove $y = Sx$ for some $x \in X$, which implies closedness of $\mathcal{R}(S)$. Like before, let

$$x_n = a_n + c_n$$

then

$$Sx_n = 0 + Sc_n \to y.$$

We claim that $\{c_n\}$ is bounded. By hypothesis, $\exists\, c_{n_k} \to c$. Thus $Sc_{n_k} \to Sc$ and also $Sc_{n_k} = Sx_{n_k} \to y$. By uniqueness of limit, $y = Sc$ for some $c \in C \subset X$.

Therefore, it remains to prove the claim. To prove $\{c_n\}$ is bounded, assume by contradiction that $\{c_n\}$ is unbounded. Then up to passing to a subsequence,

$$\|c_n\| \to \infty, \quad c_n \in C.$$

Let

$$z_n := \frac{c_n}{\|c_n\|} \in C, \quad \|z_n\| = 1.$$

Then

$$Sz_n = \frac{Sc_n}{\|c_n\|} \to 0$$

since $Sc_n \to y$ and $\|c_n\| \to \infty$.

By hypothesis, $\exists\, \{z_{n_k}\}$ coverges to $z \in X$ which implies $\|z\| = 1$. But $Sz_{n_k} \to Sz$ with $Sz_{n_k} \to 0$, so $Sz = 0$. This shows that $z \in \mathcal{N}(S) \cap C = \{0\}$ which gives $z = 0$. This contradicts to $\|z\| = 1$.

This finishes the whole proof. □

Denote $r_e = \sup\{|\lambda| : \lambda \in \sigma_{ess}(T)\}$. We have the following lemma.

Lemma 5.9. *Let X be a Banach space and $T \in \mathcal{L}(X, X)$. Define*

$$r'_e := \inf_{n \geq 0} \left(\tilde{r}(T^n)\right)^{\frac{1}{n}}.$$

Then $\lim_{n\to\infty} \left(\tilde{r}(T^n)\right)^{\frac{1}{n}}$ and $\lim_{n\to\infty} \left(r(T^n)\right)^{\frac{1}{n}}$ exist and are equal to r'_e. Furthermore, if $|\lambda| > r'_e$, then $\mathcal{N}(\lambda - T)^r$ is finite dimensional for any $r \geq 1$ and $\mathcal{R}(\lambda - T)$ is closed.

Proof. We will prove that

$$\limsup_{n\to\infty} \left(\tilde{r}(T^n)\right)^{\frac{1}{n}} \leq r'_e.$$

Take any $\epsilon > 0$ and choose $m \in \mathbb{N}$ s.t.

$$(\tilde{r}(T^m))^{\frac{1}{m}} \leq r'_e + \epsilon.$$

We take n large enough s.t. $n = pm + q$ with $0 \leq q \leq m - 1$. As a fact, for $S \in \mathcal{L}(X, X)$ and $A \subset X$, we have

$$\tilde{r}(S(A)) \leq \tilde{r}(S)\tilde{r}(A).$$

For $S, T \in \mathcal{L}(X, X)$,

$$\tilde{r}(ST(A)) \leq \tilde{r}(S)\tilde{r}(T)\tilde{r}(A)$$

and

$$\tilde{r}(ST) \leq \tilde{r}(S)\tilde{T}$$

which is multiplicative. Hence

$$\begin{aligned}(\tilde{r}(T^n))^{\frac{1}{n}} &= \tilde{r}(T^{pm+q})^{\frac{1}{n}} \\ &\leq \tilde{r}(T^m)^{\frac{mp}{mn}}\tilde{r}(T)^{\frac{q}{n}} \\ &\leq (r'_e + \epsilon)^{\frac{mp}{n}}\tilde{r}(T)^{\frac{q}{n}}.\end{aligned}$$

Take $\limsup$ in n on both sides, it implies

$$\limsup_{n\to\infty} (\tilde{r}(T^n))^{\frac{1}{n}} \leq r'_e = \inf_n \leq \liminf_n \leq \limsup_n .$$

Thus $\lim_{n\to\infty} (\tilde{r}(T^n))^{\frac{1}{n}} = r'_e$. Note that $r(T) \leq \tilde{r}(T)$ for any T, arguing similarly for $r(T)$ we get the first part of the lemma.

For the second part, choose λ s.t. $|\lambda| > r'_e$ and n such that

$$(\tilde{r}(T^n))^{\frac{1}{n}} < |\lambda| .$$

Let $T_1 := \frac{1}{|\lambda|}T$, then $\tilde{r}(T_1^n) = \frac{1}{|\lambda|}$. By Lemma 5.8, we will conclude the proof if we prove the following lemma

Lemma 5.10. *If for some $n \in \mathbb{N}, \tilde{r}(T^n) < 1$, then $(\mathbb{1} - T)^r$ restricted to closed and bounded sets is proper for any $r \geq 1$.*

Proof. Let $A \subset X$ be closed and bounded and $M \subset X$ be compact. Define

$$M_1 := \{x \in A : (\mathbb{1} - T)x \in M\} .$$

We claim that M_1 is compact which implies $\tilde{r}(M_1) = 0$.

For $x \in M_1$, $\exists m \in M$ s.t. $m = x - Tx \Leftrightarrow x = Tx + m$. Thus

$$x = T(Tx + m) + m = T^2x + Tm + m.$$

By iteration, we get

$$x = T^nx + \sum_{i=0}^{n-1} T^i(m).$$

As T is bounded thus continuous, $m_* := \sum_{i=0}^{n-1} T^i(m)$ is compact. We know that

$$M_1 \subset T^n(M_1) + M_*,$$

which implies

$$\tilde{r}(M_1) \leq \tilde{r}(T^n(M_1)) + \tilde{r}(M_*) = \tilde{r}(T^n(M_1)).$$

By submultiplicativity, we know

$$\tilde{r}(M_1) \leq \tilde{r}(T^n(M_1)) \leq \tilde{r}(T^n)\tilde{r}(M_1).$$

This implies that $\tilde{r}(M_1) = 0$ as $\tilde{r}(T^n) < 1$, thus M_1 is compact. This proves that $(\mathbb{1} - T)$ is proper. Suppose that $(\mathbb{1} - T)^{r-1}$ is proper. Let M be compact, then $(\mathbb{1} - T)^{-(r-1)}(M)$ is compact. Since

$$(\mathbb{1} - T)^{-r}(M) = (\mathbb{1} - T)^{-1}[(\mathbb{1} - T)^{-(r-1)}(M)],$$

then $(\mathbb{1} - T)^{-r}(M)$ is compact thus $(\mathbb{1} - T)^r$ is proper for any $r \geq 1$. □

The whole proof is thus finished. □

The following lemma implies that $r'_e \geq r_e$.

Lemma 5.11. *If $|\lambda_0| > r'_e$, then λ_0 is not a limit point of $\sigma(T)\backslash\{\lambda_0\}$.*

Proof. We claim that there exists a neighborhood of B of λ_0, such that $\forall\, \lambda \neq \lambda_0, \lambda \in B$, we have $\lambda \in \rho(T)$. Then this will show that λ_0 is not a limit point of $\sigma(T)\backslash\{\lambda_0\}$.

If $\lambda_0 \in \rho(T)$, this is trivial. Now let us assume that $\lambda_0 \in \sigma(T)$. We are going to show that either

$$\mathcal{N}(\lambda_0 - T) \neq 0 \quad \text{or} \quad \mathcal{N}(\lambda_0 - T^*) \neq 0.$$

Suppose both are equal to zero, we denote $D := \mathcal{R}(\lambda_0 - T)$. Then $(\lambda_0 - T)^{-1} : D \to X$ exists. Moreover, using the previous lemma, D is closed.

Assume $D \neq X$, then D is a closed proper subspace. Hence, $\exists\, u \in X$ with $\|u\| = 1$ and $\|u - w\| \geq \frac{1}{2}, \forall\, w \in D$. Let $V := \text{span}\{u, D\}$. If $v \in V$, then

$$v = \alpha u + w, \quad \alpha \in \mathbb{R},\ \omega \in D.$$

Define a linear functional $l : v \to \mathbb{R}$ s.t. $l(v) = \alpha$. Then

$$\|v\| = \|\alpha u + w\| = |\alpha| \left\|u - (-\alpha^{-1} w)\right\| \geq |\alpha| \cdot \frac{1}{2},$$

which implies

$$|l(v)| \leq 2\,\|v\| =: \|p(v)\|.$$

Applying Hahn-Banach, l can be extended to the whole X and since $l(u) = 1 \neq 0$, $l \neq 0$.

Let $v \in X$, we have

$$(\lambda_0 - T^*)l(v) = l((\lambda_0 - T)v) = 0,$$

This contradicts to $\mathcal{N}(\lambda_0 - T^*) = 0$. So $D = X$ and $(\lambda_0 - T)$ is invertible on X. Thus $(\lambda_0 - T)^{-1}$ is a bounded operator. This shows $\lambda \notin \sigma(T)$ which contradicts to $\lambda_0 \in \sigma(T)$. Therefore, either $\mathcal{N}(\lambda_0 - T) \neq 0$ or $\mathcal{N}(\lambda_0 - T^*) \neq 0$.

Now suppose that $\exists \{\tilde{\lambda}_n\}_{n\geq 0} \subset \sigma(T)\backslash\{\lambda_0\}$ which accumulates to λ_0. Then there are either infinitely many $\tilde{u}_n \in \mathcal{N}(\tilde{\lambda}_n - T)$ or $\tilde{l}_n \in \mathcal{N}(\tilde{\lambda}_n - T)^*$ by the claim above.

Thus, given $\epsilon > 0$, $\exists \bar{n}$ s.t. $\forall n > \bar{n}$

$$\left|\tilde{\lambda}_n - \lambda_0\right| < \epsilon \left|\lambda_0\right|.$$

Moreover, let M_K be the subspace spanned by the eigenvectors

$$\tilde{u}_{\bar{n}}, \tilde{u}_{\bar{n}+1}, \cdots, \tilde{u}_{\bar{n}+K}$$

and denote $u_K := \tilde{u}_{\bar{n}+K}$, $\lambda_K := \tilde{\lambda}_{\bar{n}+K}$.

Since $u_1, \cdots, u_K$ are linearly independent, each M_{K-1} is a proper closed subspace of M_K. Then $\exists\, v_K \in M_K$ with $\|v_K\| = 1$ s.t. $d(v_K, M_{K-1}) > 1 - \epsilon$. In addition,

$$v_K = \alpha_K u_K + w_K, \quad w_K \in M_{K-1}.$$

Take $S > K \in \mathbb{N}$, $r \in \mathbb{N}$, then

$$\begin{aligned}
\|T^r v_S - T^r v_K\| &= \|T^r(\alpha_S u_S) + T^r w_S - T^r v_K\| \\
&= \|\alpha_S \lambda_S^r u_S + T^r w_S - T^r v_K\| \\
&= |\lambda_S^r| \left\|v_S - (w_S - \lambda_S^{-r} T^r w_S + \lambda_S^{-r} T^r v_K)\right\| \\
&\geq |\lambda_S^r| (1 - \epsilon) \\
&= |(\lambda_S - \lambda_0 + \lambda_0)^r| (1 - \epsilon) \\
&= |\lambda_0|^r \left|1 + \frac{\lambda_S - \lambda_0}{\lambda_0}\right|^r (1 - \epsilon) \\
&\geq |\lambda_0|^r (1 - \epsilon)^{r+1}.
\end{aligned}$$

This shows that $T\{\|v\| \leq 1\}$ can not be covered by a finite number of sets of diameter $\frac{1}{4} |\lambda_0|^r (1 - \epsilon)^{r+1}$.

As ϵ is arbitrary, we get that

$$\tilde{r}(T^r) \geq r(T^r) \geq \frac{1}{4} |\lambda_0|^r.$$

In the second case, we get

$$\tilde{r}((T^*)^r) \geq r((T^*)^r) \geq \frac{1}{4} |\lambda_0|^r.$$

Note that $r(T^*) \le \tilde{r}(T)$, so in either case, we have $\tilde{r}(T^r) \ge \frac{1}{4} |\lambda_0|^r$.

Since $r'_e = \inf_n (\tilde{r}(T^n))^{\frac{1}{n}} \ge |\lambda_0|$, this contradicts $|\lambda_0| > r'_e$. Then λ_0 is not a limit point of $\sigma(T) \backslash \{\lambda_0\}$.

This finishes the proof. □

Finally, we can give the proof of Theorem 5.2.

Proof of Hennion's theorem. Recall that our goal is to prove $r(\mathcal{L}) \le m$ and $r_e \le \theta$ with $\theta < m$. By assumption, we have

$$\|\mathcal{L}^n v\| \le C\theta^n \|v\| + Cm^n \|v\|_\omega \le 2Cm^n \|v\|_\omega .$$

Thus $r(\mathcal{L}) = \lim_{n\to\infty} \|\mathcal{L}^n\|^{\frac{1}{n}} \le m$ which easily gives the first result we want.

Now, let us prove $r_e \le \theta$.

Define $B_1 := \{v \in \mathcal{B} : \|v\| \le 1\}$. Note that

$$r_e \le r'_e = \inf_n (\tilde{r}(\mathcal{L}^n))^{\frac{1}{n}} \le \lim_{n\to\infty} (\tilde{r}(\mathcal{L}^n))^{\frac{1}{n}}.$$

We claim $\mathcal{L}^n B_1$ can be covered by a finite number of balls with radius less than $C\theta^n$ which will finish the proof.

Without loss of generality, we prove for $n_0 = 1$. Other cases are the same. $\mathcal{L} : \mathcal{B} \to \mathcal{B}_\omega$ is a compact operator, so $\mathcal{L}B_1$ is relatively compact. Let

$$B_\epsilon(v) = \{w \in \mathcal{B}_\omega : \|v - w\|_\omega \le \epsilon\} .$$

Then obviously $\{B_\epsilon(v)\}_{v\in\mathcal{L}B_1}$ covers $\mathcal{L}B_1$. By relative compactness, there exist a finite number of sets

$$\{V(v_i)\}_{i=1}^N := \{B_\epsilon(v_i) \cap \mathcal{L}B_1\}_{i=1}^N$$

that cover $\mathcal{L}B_1$.

Now we will show that $\operatorname{diam}(\mathcal{L}^n(V(v_i))) \le C\theta^n$ for any $n \in \mathbb{N}$. Let $v \in \mathcal{L}B_1$, $v = \mathcal{L}(w), w \in B_1$. Moreover, let $v_i = \mathcal{L}(w_i)$. Direct computation shows

$$\begin{aligned}
\|\mathcal{L}^n(w) - \mathcal{L}^n(w_i)\| &= \left\|\mathcal{L}^{n-1}(v - v_i)\right\| \\
&\le C\theta^{n-1} \|v - v_i\| + Cm^{n-1} \|v - v_i\|_\omega \\
&\le C\theta^{n-1}(\|v\| + \|v_i\|) + Cm^{n-1}\epsilon \\
&\lesssim \theta^n
\end{aligned}$$

if $\epsilon = \frac{\theta^n}{m^{n-1}}$. This shows that $\mathcal{L}^n B_1$ can be covered by a finite number of balls with radius less than $C\theta^n$.

The proof is thus finished. □

5.2. **The spectral property.** Let X, Y be Banach spaces and $T : X \to Y$ be a bounded linear operator. The normed spaces are respectively $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$.

Let us recall the definition of compact operators.

Definition 5.10. T is a compact operator if $T(B) \subset Y$ is relatively compact whenever B is bounded. Equaivalently, $\forall \{x_n\}_{n\geq 0} \subset X$ a $\|\cdot\|_X$-bounded sequence, $\{Tx_n\}_{n\geq 0} \subset Y$ has a $\|\cdot\|_Y$-convergent subsequence.

The following are some facts about compact operators.

(1) If $z \in \sigma(T), z \neq 0$, then z is an eigenvalue of finite multiplicity.
(2) $\forall r > 0$, the set of $z \in \sigma(T)$ with $|z| \geq r$ is finite. So the spectrum of a compact operator on an infinite dimensional space consists of "0, some other eigenvalues of finite multiplicity with modulus less than $r(T)$, and a finite number of eigenvalues of finite multiplicity with modulus $r(T)$".
(3) If $z \in \sigma(T)$ and $z \neq 0$, then $\mathcal{N}(z - T)^r$ stabilizes, i.e. $\exists\, n \geq 1$, $\ker(z - T)^n = \ker(z - T)^r$ for all $r \geq n$. Moreover, if $\exists\, n_0 \geq 1$ such that T^{n_0} is compact, then this item still holds for T.

Theorem 5.4 (Riesz operator)**.** *Let $T : X \to X$ be a bounded linear operator and denote by $\sigma(T)$ its spectrum. If $\tau \subset \sigma(T)$ is an isolated part of the spectrum in the sense that τ and $\tau' = \sigma(T)\backslash\tau$ are both closed. Then there is a projection $P_\tau : X \to X, P_\tau^2 = P_\tau$ which commutes with T, $P_\tau \circ T = T \circ P_\tau$ (same for $P_{\tau'}$). If we put $M = \Im P_\tau, L = \ker P_\tau$, then $X = M \oplus L$ is a T-invariant decomposition and $\sigma(T|_M) = \tau$, $\sigma(T|_L) = \tau'$.*

The proof uses holomorphic functional calculus. Moreover, $P_\tau + P_{\tau'} = \mathrm{id}, P_\tau P_{\tau'} = P_{\tau'} P_\tau = 0$.

Definition 5.11 (Quasi-compact operator)**.** $T : X \to X$ is called quasi-compact if there is a T-invariant decomposition $X = F \oplus H$ such that $r(T|_H) < r(T)$, $\dim F < \infty$ and each eigenvalue of $T|_F$ has modulus equal to $r(T)$. Moreover, if $\dim F = 1$, then T is called quasi-compact and simple.

A quasi-compact operator has the spectral gap property. In fact, using Riesz projectors, they are equivalent.

Definition 5.12 (Discrete and essential spectrum, Browder)**.** Let $T : X \to X$ be a bounded linear operator. We say $\lambda \in \sigma_d(T)$ if

(1) λ is an isolated point of $\sigma(T)$.
(2) The Riesz projector P_λ has finite rank.

Clearly, the discrete spectrum σ_d is at most countable. In addition, the essential spectrum is $\sigma_{ess}(T) = \sigma(T) \backslash \sigma_d(T)$.

We denote $r_{ess}(T) = \sup \{|\lambda| : \lambda \in \sigma_{ess}(T)\}$.

Theorem 5.5 (Ionescu-Tulcea and Marinescu, Hennion)**.** *Let $X \subset Y$ be two Banach spaces with $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ and $\|\cdot\|_Y \leq \|\cdot\|_X$. Let $T : X \to X$ be a bounded linear operator. Assume that for some $\sigma_0 \in (0,1)$, $C < \infty$, $n_0 \in \mathbb{N}$, the following hold:*

(1) $T^{n_0} : (X, \|\cdot\|_X) \to (Y, \|\cdot\|_Y)$ is compact.

(2) $\forall n \in \mathbb{N}, \forall x \in X$,

$$\|T^n x\|_Y \leq C \|x\|_Y .$$

(3) $\forall n \in \mathbb{N}, \forall x \in X$,

$$\|T^n x\|_X \leq C \sigma_0^n \|x\|_X + C \|x\|_Y .$$

Then $r(T) \leq 1$ and $r_{ess}(T) \leq \sigma_0$.

Corollary 5.12. *Under the assumptions of this theorem, T is quasi-compact. In fact, $\exists \sigma_1 \in (\sigma_0, 1)$ s.t. $\sigma(T)$ consists of a finite number of eigenvalues of modulus 1: namely τ, and the other part τ' where $r(T|_{\ker P_\tau}) < \sigma_1$.*

The complication of having σ_1 instead of σ_0 is because we only know T^{n_0} is compact instead of T. Let

$$\sigma = \max \{\sigma_0, |\lambda| : \lambda < 1, \lambda \in \sigma_d(T)\} .$$

Using Riesz projectors, we have moreover

$$I = P_\tau + P_{\tau'} = \sum_{\theta \in \mathcal{F}} e^{i\theta} P_{e^{i\theta}} + P_{\tau'}$$

where $\mathcal{F} \subset [0, 2\pi)$ is finite. Apply T on both sides, we get

$$T = \sum_{\theta \in \mathcal{F}} e^{i\theta} T P_{e^{i\theta}} + T P_{\tau'}$$

and we denote $TP_{e^{i\theta}} = \Pi_\theta$ and $TP_{\tau'} = S$. Namely,

$$T = \sum_{\theta \in \mathcal{F}} e^{i\theta} \Pi_\theta + S$$

where $r(S) \leq \sigma < 1$, $\Pi_\theta^2 = e^{i\theta}\Pi_\theta$, $\Pi_{\theta'}\Pi_\theta = 0$ if $\theta' \neq \theta$ and $\Pi_\theta S = S\Pi_\theta = 0, \forall \theta \in \mathcal{F}$. Therefore, for any $n \in \mathbb{N}^+$,

$$T^n = \sum_{\theta \in \mathcal{F}} e^{in\theta} \Pi_\theta + S^n.$$

So $\left\| T^n - \sum_{\theta \in \mathcal{F}} e^{in\theta} \Pi_\theta \right\|_X \to 0$ as $r(S) < 1$.

Let us go back to talk about the expanding maps of the circle. We will prove that its transfer operator is quasi-compact using Ionescu-Tulcea and Marinescu or Hennion's theorem. Recall that we have $(\mathbb{T}, f)$ where $f \in C^2$ with $|f'(y)| \geq \lambda_* > 1, \forall\, y \in \mathbb{T}$. We define the transfer operator on $L^1(\mathbb{T}, dm)$ to itself by

$$\mathcal{L}h = \frac{df_* m_h}{dm}$$

where $dm_h = hdm$. It turns out that

$$\mathcal{L}h = \sum_{y \in f^{-1}\{x\}} \frac{1}{|f'(y)|} h(y)$$

and $\mathcal{L}h$ is uniquely determined by

$$\int \varphi \cdot (\mathcal{L}h) dm = \int (\varphi \circ f) \cdot hdm, \quad \forall\, L^\infty(dm).$$

By the properties of a transfer operator, $\mathcal{L}$ is linear, bounded ($\|\mathcal{L}h\|_1 \leq \|h\|_1$ thus $\|\mathcal{L}^n h\|_1 \leq \|h\|_1$ so (2) of Ionescu-Tulcea and Marinescu or Hennion is satisfied) and positive (if $h \geq 0$ then $\mathcal{L}h \geq 0$). Moreover, by linearity and positivity we have that $\mathcal{L}\,|h| \geq |\mathcal{L}h|$.

We consider the Sobolev space:

$$\mathcal{W}^{1,1}(\mathbb{T}) := \left\{ h : \mathbb{T} \to \mathbb{R}, h \in L^1, h' \, \text{exists a.e. and } h' \in L^1 \right\}$$

which equals the space of absolutely continuous functions on $\mathbb{T}$. More precisely, h is absolutely continuous on $\mathbb{T}$ means $\forall\, \epsilon > 0$, $\exists\, \delta > 0$ s.t. if $\sum_{i=1}^n (b_i - a_i) < \delta, a_i, b_i \in \mathbb{T}, n \in \mathbb{N}^+$, then $\sum_{i=1}^n |h(b_i) - h(a_i)| < \epsilon$.

Clearly, $\mathcal{W}^{1,1}(\mathbb{T})$ is a linear space endowed with the Sobolev norm

$$\|h\|_{1,1} := \|h\|_1 + \|h'\|_1 .$$

This is a Banach space (actually a Banach algebra) and $\mathcal{W}^{1,1}(\mathbb{T}) \hookrightarrow C^0(\mathbb{T})$ is a bounded inclusion.

Remember that we have the Lasota-Yorke inequality:

$$\begin{cases} \|\mathcal{L}h\|_1 \leq \|h\|_1 , \\ \|(\mathcal{L}h)'\|_1 \leq \lambda_*^{-1} \|h'\|_1 + D \|h\|_1 . \end{cases}$$

Thus $\mathcal{L}$ is a bounded linear operator on $\mathcal{W}^{1,1}(\mathbb{T})$. By induction, (3) of Ionescu-Tulcea and Marinescu or Hennion is also satisfied. So it remains to check that

$$\mathcal{L} : (\mathcal{W}^{1,1}(\mathbb{T}), \|\cdot\|_{1,1}) \to (L^1, \|\cdot\|_1)$$

is compact. For this purpose, we just need to prove that if $\{h_n\}_{n\geq 1} \subset \mathcal{W}^{1,1}(\mathbb{T})$ satisfies $\|h_n\|_{1,1} \leq C, \forall\, n \geq 1$, then $\{\mathcal{L}h_n\}_{n\geq 1}$ contains a convergent subsequence in $L^1(\mathbb{T}, dm)$.

We will use Frechét-Kolmogorov theorem which is an L^p version of Arzelà-Ascoli theorem, saying that "uniform boundedness" and "equicontinuity" in $L^p(\Omega)$ implies pre-compactness in $L^p(\Omega)$. For us, $p = 1$ and $\Omega = \mathbb{T}$.

Definition 5.13 (Uniform boundedness)**.** We say $\{\varphi_n\}_{n\geq 1} \subset L^1(\mathbb{T}, dm)$ is uniformly bounded if there exists some $C < \infty$ such that $\|\varphi_n\|_1 \leq C$ for all $n \geq 1$.

Definition 5.14 (Equicontinuity)**.** We say $\{\varphi_n\}_{n\geq 1} \subset L^1(\mathbb{T}, dm)$ is equicontinuous if

$$\|\varphi_n(\cdot + t) - \varphi_n(\cdot)\|_1 \to 0 \text{ as } t \to 0$$

uniformly in n. More precisely, $\forall\, \epsilon > 0, \exists\, \delta > 0$ s.t. if $|t| < \delta$, then

$$\|\varphi_n(\cdot + t) - \varphi_n(\cdot)\|_1 < \epsilon, \forall\, n \geq 1.$$

For the sequence $\{h_n\}$, we define $\varphi_n := \mathcal{L}h_n, n \geq 1$, then by Lasota-Yorke and the assumption we have

$$\|\mathcal{L}h_n\|_1 \leq \|h_n\|_1 \leq \|h_n\|_{1,1} \leq C.$$

So $\{\varphi_n\}_{n\geq 1}$ is uniformly bounded. Thus it remains to check its equicontinuity. Note that it is sufficient to prove the following:

$$\|\mathcal{L}h_n(\cdot + t) - \mathcal{L}h_n(\cdot)\|_1 \lesssim |t|\,, \forall\, n \geq 1.$$

By the Newton-Leibniz formula (Fundamental theorem of calculus), we have

$$\mathcal{L}h_n(x + t) - \mathcal{L}h_n(x) = \int_x^{x+t} (\mathcal{L}h_n)'(s)ds.$$

This implies

$$\begin{aligned}
\|\mathcal{L}h_n(\cdot + t) - \mathcal{L}h_n(\cdot)\|_1 &\leq \int_0^1 \int_x^{x+t} |(\mathcal{L}h_n)'(s)|\, dsdx \\
&\leq \int_0^1 \int_0^1 |(\mathcal{L}h_n)'(u + x)| \cdot \mathbb{1}_{[0,t]}(u)dudx \\
&= \int \|(\mathcal{L}h_n)'\|_1 \cdot \mathbb{1}_{[0,t]}(u)du \\
&\leq C\,|t|\,.
\end{aligned}$$

Here we have used Fubini's theorem to exchange the order of the integral and also applied the second inequality in Lasota-Yorke to ensure the final step.

To conclude, applying Frechét-Kolmogorov theorem, we obtain that $\mathcal{L}$ is a compact operator. Thus all conditions of Ionescu-Tulcea and

Marinescu or Hennion's theorem are satisfied, so we get that $\mathcal{L}$ is quasi-compact.

In the following, our remaining goal is to prove that $\mathcal{L}$ is not only just quasi-compact, but also simple with 1 as the only peripheral eigenvalue. We formulate this result as a theorem.

Theorem 5.6. *The transfer operator $\mathcal{L}$ of the expanding maps of the circle is quasi-compact and simple.*

Proof. It is left to prove that $\mathcal{L}$ is simple. Let us first show that 1 is an eigenvalue of $\mathcal{L}$. Indeed, by the same argument as before,

$$\mathcal{L}^* m = m \Leftrightarrow \langle \varphi, \mathcal{L}^* m\rangle = \langle \varphi, m\rangle, \forall \varphi \in C^0(\mathbb{T}).$$

$$\Leftrightarrow \langle \mathcal{L}\varphi, m\rangle = \langle \varphi, m\rangle \Leftrightarrow \int \mathcal{L}\varphi dm = \int \varphi dm$$

which is true. Thus 1 is an eigenvalue of $\mathcal{L}^*$, in particular $1 \in \sigma(\mathcal{L}^*)$. Since the spectrum of a bounded linear operator and its adjoint on a Banach space are the same, $1 \in \sigma(\mathcal{L})$. As $\mathcal{L}$ is quasi-compact (actually compact from $(\mathcal{W}^{1,1}, \|\cdot\|_{1,1})$ to $(L^1, \|\cdot\|_1)$), 1 is an eigenvalue of finite multiplicity. So $0 \in \mathcal{F} \subset [0, 2\pi)$ which is finite. Recall that we have

$$\mathcal{L} = \sum_{\theta' \in \mathcal{F}} e^{i\theta'} \Pi_{\theta'} + S,$$

and

$$\mathcal{L}^k = \sum_{\theta' \in \mathcal{F}} e^{ik\theta'} \Pi_{\theta'} + S^k, \forall k \in \mathbb{N}^+.$$

Fix any $\theta \in \mathbb{T}$, multiply $e^{-ik\theta}$ on both sides and we get

$$e^{-ik\theta}\mathcal{L}^k = \sum_{\theta' \in \mathcal{F}} e^{ik(\theta'-\theta)} \Pi_{\theta'} + e^{-ik\theta} S^k.$$

Take $\frac{1}{n}\sum_{k=1}^n$ on both sides and let $n \to \infty$, we get

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n e^{-ik\theta}\mathcal{L}^k = \Pi_\theta, \forall \theta \in \mathcal{F}.$$

This is because when $n \to \infty$,

$$\frac{1}{n}\sum_{k=1}^n e^{-ik\theta} S^k \to 0$$

and

$$\frac{1}{n}\sum_{k=1}^n \sum_{\theta' \in \mathcal{F}} e^{ik(\theta'-\theta)} = \begin{cases} 1, & \text{if } \theta' = \theta, \\ 0, & \text{otherwise.} \end{cases}$$

In particular, $\Pi_0 = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathcal{L}^k$ and it is a positive operator because $\mathcal{L}$ is positive.

Note that for $h \geq 0$, we have $\mathcal{L}h \geq 0$ as $\mathcal{L}$ is positive. Let $g := \Pi_0 \mathbb{1}$ which is continuous. Then we claim that $\mathcal{L}g = g$, $g > 0$ and $\int gdm = 1$.

By the expression of $\mathcal{L}$, $\mathcal{L}g = g$ is trivial. Therefore, if $d\mu = gdm$ then μ is f-invariant by equivalence. $\int gdm = 1$ is also obvious by the expression of Π_0 and Lebesgue's dominated convergence theorem. By positivity of Π_0, we know that $g = \Pi_0 \mathbb{1} \geq 0$, so it remains to prove that g is indeed strictly positive.

If $g(x) = 0$ for some $x \in \mathbb{T}$, then since $\mathcal{L}g = g$, we have

$$\mathcal{L}g(x) = \sum_{y\in f^{-1}\{x\}} \frac{1}{|f'(y)|} g(y) = 0.$$

So $g(y) = 0, \forall\, y \in f^{-1}\{x\}$. By induction, $g(y) = 0, \forall\, y \in f^{-n}\{x\}$ for all $n \geq 1$. But the set of preimages of any point is dense in $\mathbb{T}$, because for any interval $I \subset \mathbb{T}$, $\exists\, n \in \mathbb{N}$ such that $f^n(I) = \mathbb{T} \ni x$, then $f^{-n}\{x\} \in I$ for this n.

Moreover, since g is continuous then $g \equiv 0$ which is a contradiction to $\int gdm = 1$. Therefore, $g(x) > 0, \forall\, x \in \mathbb{T}$

Remark 5.4. We can conclude that if $\varphi \geq 0$ is continuous, $\mathcal{L}\varphi = \varphi$ and $\varphi(x_0) = 0$ for some $x_0 \in \mathbb{T}$, then $\varphi \equiv 0$.

Next, we are going to prove that 1 is the only peripheral eigenvalue and $\mathcal{L}$ is simple. Let h be an eigenvector of $\mathcal{L}$ with eigenvalue $e^{i\theta}, \theta \in \mathcal{F}$. Namely,

$$\mathcal{L}h = e^{i\theta} h \Leftrightarrow \Pi_\theta h = e^{i\theta} h.$$

We need to prove that $\theta = 0$ and $h = \lambda g$ for some $\lambda \in \mathbb{C}$.

Since we have $|\mathcal{L}h| \leq \mathcal{L}\,|h|$ and $|\mathcal{L}h| = \left|e^{i\theta} h\right| = |h|$, then $|h| \leq \mathcal{L}h$. By induction, $|h| \leq \mathcal{L}^k\,|h|\,, \forall\, k \geq 1$. This implies

$$|h| \leq \frac{1}{n}\sum_{k=1}^{n} \mathcal{L}^k\,|h|\,, \forall\, n \geq 1.$$

Let $n \to \infty$, we get $|h| \leq \Pi_0\,|h|$. Using dominated convergence theorem, we have

$$\begin{aligned}\int \Pi_0 |h| - |h| \, dm &= \int \Pi_0 |h| \, dm - \int |h| \, dm \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \int \mathcal{L}^k |h| \, dm - \int |h| \, dm \\ &= \int |h| \, dm - \int |h| \, dm = 0\end{aligned}$$

Thus $\Pi_0 |h| = |h|$ because they are continuous. Since Π_θ is perpendicular to each other and also to S, this shows that $\mathcal{L} |h| = |h|$.

Now we consider

$$\beta := \min \frac{|h|}{g} = \frac{|h(x_0)|}{g(x_0)} \quad \text{for some } x_0 \in \mathbb{T}.$$

Then $|h| - \beta g$ has a zero for x_0 and it is non-negative. So

$$\mathcal{L}(|h| - \beta g) = \mathcal{L} |h| - \beta \mathcal{L} g = |h| - \beta g.$$

By the remark before, we get $|h| - \beta g \equiv 0$ thus $|h| = \beta g$. So $h = e^{i\varphi} \beta g$ where $\varphi \in C[0, 2\pi)$.

We will show that $\varphi - \varphi \circ f \equiv \theta$. This would imply

$$0 = \int \varphi - \varphi \circ f d\mu = \theta$$

which gives $\theta = 0$ and $\varphi = \varphi \circ f$. Using the expanding condition again, φ is constant. Thus $h = \lambda g$ for some $\lambda \in \mathbb{C}$.

Let us show that $\varphi - \varphi \circ f \equiv \theta$. By $\mathcal{L} h = e^{i\theta} h$ and $h = e^{i\varphi} \beta g$ we have

$$\mathcal{L} h = \beta \mathcal{L}(e^{i\varphi} g) \;\Leftrightarrow\; e^{i\theta} e^{i\varphi} \beta g = \beta \mathcal{L}(e^{i\varphi} g)$$

which shows

$$\mathcal{L}(e^{i\varphi} g) = e^{i(\theta+\varphi)} g.$$

By direct computation, we have

$$\begin{aligned}\mathcal{L}\left(e^{i(\varphi - \varphi\circ f - \theta)} g\right)(x) &= \sum_{f(y)=x} \frac{1}{|f'(y)|} e^{i(\varphi(y) - \varphi(x) - \theta)} g(y) \\ &= e^{-i(\varphi(x)+\theta)} \sum_{f(y)=x} \frac{1}{|f'(y)|} e^{i\varphi(y)} g(y) \\ &= e^{-i(\varphi(x)+\theta)} \cdot \mathcal{L}(e^{i\varphi})(x) \\ &= e^{-i(\varphi(x)+\theta)} \cdot e^{i(\varphi(x)+\theta)} g(x) \\ &= g(x).\end{aligned}$$

Therefore, we have

$$\mathcal{L}\left((e^{i(\varphi-\varphi\circ f-\theta)}-1)g\right)=0,$$

which implies

$$\int \mathcal{L}\left((e^{i(\varphi-\varphi\circ f-\theta)}-1)g\right)dm=\int (e^{i(\varphi-\varphi\circ f-\theta)}-1)gdm=0.$$

Taking the real part, we have

$$\int [1-\cos(\varphi-\varphi\circ f-\theta)]gdm=0$$

where $g>0$ and $1-\cos(\varphi-\varphi\circ f-\theta)\geq 0$. Using the previous remark, we get $\cos(\varphi-\varphi\circ f-\theta)=1$ which implies $\varphi-\varphi\circ f-\theta\in 2\pi\mathbb{Z}$. Since it is continuous on $[0,2\pi)$, we get $\varphi-\varphi\circ f-\theta\equiv 2\pi k$ for some $k\in\mathbb{Z}$. Integrating w.r.t. μ, we have $0-\theta=2\pi k$. As $\theta\in[0,2\pi)$, $k=0\Rightarrow\theta=0\Rightarrow\varphi-\varphi\circ f=0$.

This finishes the whole proof. □

We have proved that on $\mathcal{W}^{1,1}$

$$\mathcal{L}=\Pi_0+S,\quad r(S)\leq\sigma<1.$$

Then

$$\mathcal{L}^n=\Pi_0+S^n.$$

Hence,

$$\|\mathcal{L}^n h-\Pi_0 h\|_{1,1}=\|S^n h\|_{1,1}\leq\sigma^n\|h\|_{1,1}.$$

Moreover, $\Pi_0 h=\lambda g$ for some $\lambda\in\mathbb{C}$ and

$$\int \Pi_0 hdm=\int \lambda gdm=\lambda\int gdm=\lambda.$$

On the other hand,

$$\Pi_0 h=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\mathcal{L}^k h,$$

so by dominated convergence theorem,

$$\int \Pi_0 hdm=\int hdm.$$

Thus we get $\Pi_0 h=(\int hdm)g$, which shows that there is $\sigma\in(0,1)$ such that for all $h\in\mathcal{W}^{1,1}(\mathbb{T})$,

$$\left\|\mathcal{L}^n h-\left(\int h\,dm\right)g\right\|_{1,1}\leq\sigma^n\|h\|_{1,1}\tag{5.5}$$

for some $g\in\mathcal{W}^{1,1}(\mathbb{T})$ satisfying $g>0$, $\mathcal{L}g=g$ and $\int gdm=1$.

Using this, one can derive a large deviations principle and a central limit theorem for $(\mathbb{T}, f)$ with observables in $\mathcal{W}^{1,1}(\mathbb{T})$, see for instance [3]. The arguments of course are quite involved. We propose a different strategy, using the Markov operator and the abstract results derived above, which will provide an effective LDT estimate and a CLT.

Note that $\mathcal{L}\mathbb{1} = \mathbb{1}$ if and only if the reference measure m is f-invariant. This holds for instance when f is the doubling map, but does not hold in general, hence the transfer operator is usually not a Markov operator.

However, changing the reference measure for $d\mu = g dm$, where g was described above, since $\mathcal{L}g = g$ we have that μ is f-invariant. Consider the transfer operator relative to this reference measure

$$\mathcal{Q}h := \frac{df_*\mu_h}{d\mu},$$

where $d\mu_h := h\, d\mu$.

Then $\mathcal{Q}\mathbb{1} = \mathbb{1}$ so this transfer operator is also a Markov operator and μ is a stationary measure (since $\int \mathcal{Q}h\, d\mu = \int h\, d\mu$ for all h). Moreover, $\mathcal{Q}$ is related to $\mathcal{L}$ by

$$(\mathcal{Q}h)g = \mathcal{L}(hg),$$

so

$$\mathcal{Q}h(x) = \frac{1}{g(x)} \sum_{y \in f^{-1}\{x\}} \frac{g(y)}{|f'(y)|} h(y)$$

which shows that the Markov kernel is given by $K \colon \mathbb{T} \to \mathrm{Prob}(\mathbb{T})$,

$$K_x(y) = \sum_{y \in f^{-1}\{x\}} \frac{g(y)}{g(x)} \frac{1}{|f'(y)|} \delta_y .$$

That is, K_x is a convex combination of Dirac delta measures supported on the pre-images of the point x. Since, moreover, for all $n \in \mathbb{N}$, $(\mathcal{Q}^n h)g = \mathcal{L}^n(hg)$, we have that

$$\mathcal{Q}^n h - \int h d\mu = \frac{1}{g} \left(\mathcal{L}^n(hg) - g \int hg\, dm \right).$$

Therefore, using (5.5),

$$\begin{aligned} \left\| \mathcal{Q}^n h - \int h d\mu \right\|_{C^0} &\lesssim \left\| \frac{1}{g} \left(\mathcal{L}^n(hg) - g \int hg\, dm \right) \right\|_{1,1} \\ &\le C(g) \left\| \mathcal{L}^n(hg) - g \int hg dm \right\|_{1,1} \\ &\le C(g)\, \sigma^n \left\| hg \right\|_{1,1} \le C'(g)\, \sigma^n \left\| h \right\|_{1,1} , \end{aligned}$$

where $C(g), C'(g) < \infty$ depend only on the $\mathcal{W}^{1,1}$-norms of g and $\frac{1}{g}$.

This shows that the observed Markov system $(\mathbb{T}, K, \mu, \mathcal{W}^{1,1}(\mathbb{T}))$ is strongly mixing with exponential rate. By Theorem 2.1 and Theorem 2.2 effective LDT estimates and a CLT hold for this stochastic dynamical system, which then easily translate to effective LDT estimates and a CLT for the deterministic dynamical system $(\mathbb{T}, f, \mu)$ with observables in $\mathcal{W}^{1,1}(\mathbb{T})$. More precisely, we obtain the following (compare with Theorem 1.22 and Theorem 1.32 in [3]).

Theorem 5.7. *Let $\varphi \in \mathcal{W}^{1,1}(\mathbb{T})$. Given any $\epsilon > 0$ there are $n(\epsilon) \in \mathbb{N}$ and $c(\epsilon) > 0$ such that for all $n \geq n(\epsilon)$ we have*

$$\mu \left\{ x \colon \left| \frac{\varphi(x) + \varphi \circ f(x) + \cdots + \varphi \circ f^{n-1}(x)}{n} - \int_{\mathbb{T}} \varphi d\mu \right| > \epsilon \right\} \leq e^{-c(\epsilon) n} .$$

Besides ϵ, the parameters $n(\epsilon)$ and $c(\epsilon)$ only depend (explicitly) on the $\mathcal{W}^{1,1}$-norms of φ, g and $\frac{1}{g}$.

Moreover, if $\int \varphi \, d\mu = 0$ and if φ is not a coboundary, which in this setting means that there is no function $\eta \in C^0(\mathbb{T})$ such that $\varphi(\theta) = \eta(\theta) - \eta \circ f(\theta)$ for all $\theta \in \mathbb{T}$, then there is $\sigma = \sigma(\varphi) > 0$ such that

$$\frac{\varphi + \varphi \circ f + \cdots + \varphi \circ f^{n-1}}{\sigma \sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1) .$$

Proof. Let $X^+ := \mathbb{T}^{\mathbb{N}}$ and let $\mathbb{P}$ be the Markov measure on X^+ with initial distribution μ and transition kernel K defined above. For an observable $\varphi \colon \mathbb{T} \to \mathbb{R}$, we reserve the notation $S_n\varphi$ for the stochastic Birkhoff sums $S_n\varphi \colon X^+ \to \mathbb{R}$,

$$S_n\varphi(\omega) = \varphi(\omega_0) + \varphi(\omega_1) + \cdots + \varphi(\omega_{n-1}),$$

while the expression of the (deterministic) Birkhoff sums relative to de dynamics f will be written explicitly.

For every $x \in \mathbb{T}$, K_x is supported on the pre-images of x via f. Then the set Ω of "admissible words" consists of sequences $\omega = \{\omega_n\}_{n \in \mathbb{N}} \in X^+$ that satisfy $f(\omega_{j+1}) = \omega_j$ for all $j \in \mathbb{N}$. This is, as it should be, a full $\mathbb{P}$-measure set. Indeed,

$$\Omega = \bigcap_{j \in \mathbb{N}} \left\{ \omega \in X^+ \colon f(\omega_{j+1}) = \omega_j \right\} = \bigcap_{j \in \mathbb{N}} \sigma^{-j} \left\{ \omega \in X^+ \colon f(\omega_1) = \omega_0 \right\}$$

and

$$\mathbb{P} \left\{ \omega \in X^+ \colon f(\omega_1) = \omega_0 \right\} = \int K_{\omega_0} \left\{ \omega_1 \colon f(\omega_1) = \omega_0 \right\} d\mu(\omega_0) = 1.$$

Since μ is K-stationary, $\mathbb{P}$ is σ-invariant, so $\mathbb{P}(\Omega) = 1$.

As already established, $(\mathbb{T}, K, \mu, \mathcal{W}^{1,1}(\mathbb{T}))$ is strongly mixing, so applying the abstract LDT in Theorem 2.1, for $n \geq n(\epsilon)$ we have that

$$\mathbb{P}\left\{\omega \in X^+ \colon \left|\frac{1}{n} S_n\varphi(\omega) - \int \varphi \, d\mu\right| > \epsilon\right\} \leq e^{-c(\epsilon)n} .$$

On the other hand, let

$$E := \left\{x \in \mathbb{T} \colon \left|\frac{\varphi(x) + \varphi \circ f(x) + \cdots + \varphi \circ f^{n-1}(x)}{n} - \int_{\mathbb{T}} \varphi d\mu\right| > \epsilon\right\}.$$

Then since μ is K-stationary,

$$\begin{aligned}
\mu(E) &= \int K_x^n(E)\, d\mu(x) = \int \mathbb{P}_x \left\{\omega \in X^+ \colon \omega_n \in E\right\}\, d\mu(x) \\
&= \mathbb{P}\left\{\omega \in X^+ \colon \omega_n \in E\right\} = \mathbb{P}\left\{\omega \in \Omega \colon \omega_n \in E\right\} \\
&= \mathbb{P}\left\{\omega \in \Omega \colon \left|\frac{\varphi(\omega_n) + \varphi(\omega_{n-1}) + \cdots + \varphi(\omega_1)}{n} - \int_{\mathbb{T}} \varphi d\mu\right| > \epsilon\right\} \\
&= \mathbb{P}\left\{\omega \colon \left|\frac{1}{n} S_n\varphi(\omega) - \int \varphi \, d\mu\right| > \epsilon\right\} \leq e^{-c(\epsilon)n} ,
\end{aligned}$$

which proves the LDT estimate for the dynamical system $(\mathbb{T}, f, \mu)$.

Now let us assume that φ has μ-mean zero and it is not a coboundary. We first show that the abstract CLT given by Theorem 2.2 is applicable to the Markov system $(\mathbb{T}, K, \mu, \mathcal{W}^{1,1}(\mathbb{T}))$. Indeed, its ergodicity is derived exactly as in Proposition 2.2. Moreover, for $\psi := \sum_{n=0}^{\infty} \mathcal{Q}^n \varphi$, if, by contradiction, $\sigma^2(\varphi) := \|\psi\|_2^2 - \|\mathcal{Q}\psi\|_2^2 = 0$, then exactly as in the proof of Proposition 2.2 we obtain that $\psi(y) = \mathcal{Q}\psi(x)$ for μ-a.e. $x \in \mathbb{T}$ and K_x-a.e. $y \in \mathbb{T}$. This immediately implies that $\psi(y) = \mathcal{Q}\psi(f(y))$ for μ-a.e. $y \in \mathbb{T}$. But since ψ and $\mathcal{Q}\psi \circ f$ are continuous and $d\mu = g\, dm$, where g is continuous and bounded away from zero, we conclude that $\psi = \mathcal{Q}\psi \circ f$ everywhere. But then

$$\varphi = \psi - \mathcal{Q}\psi = \mathcal{Q}\psi \circ f - \mathcal{Q}\psi ,$$

showing that φ is a coboundary, which is a contradiction.

Then the abstract CLT is applicable, so for all $\lambda \in \mathbb{R}$ we have

$$\mathbb{P}\left\{\omega \in X^+ \colon \frac{S_n\varphi(\omega)}{\sigma\sqrt{n}} \leq \lambda\right\} \to \int_{-\infty}^{\lambda} e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} .$$

As before, we can show that

$$\mu\left\{x \colon \frac{\varphi(x) + \cdots + \varphi \circ f^{n-1}(x)}{\sigma\sqrt{n}} \leq \lambda\right\} = \mathbb{P}\left\{\omega \colon \frac{S_n\varphi(\omega)}{\sigma\sqrt{n}} \leq \lambda\right\} ,$$

which establishes the CLT for the dynamical system $(\mathbb{T}, f, \mu)$. □

5.3. **Toral automorphism.** Let $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ be the corresponding matrix action of Arnold's cat map. Obviously $\det A = 1$. Let the cat map $f : \mathbb{T}^2 \to \mathbb{T}^2$ be such that $f(x) = Ax \bmod 1$. Simple calculation gives $\lambda_\pm = \frac{3\pm\sqrt{5}}{2}$ with $\lambda_+\lambda_- = 1$. In the following we denote $\lambda := \lambda^+$. We have that

$$Av^u = \lambda v^u, \quad Av^s = \lambda^{-1}v^s,$$

where v^u, v^s are the corresponding eigenvectors of λ and λ^{-1} and thus $\langle v^u, v^s\rangle = 0$.

Remark 5.5. In general, we may consider symmetric matrix $A \in \mathrm{SL}(2,\mathbb{Z})$ with $a_{ij} > 0, i, j = 1, 2$.

The transfer operator corresponding to f is

$$\mathcal{L}h = h \circ f^{-1}$$

where $\mathcal{L}$ is uniquely determined by

$$\int_{\mathbb{T}^2} \varphi \cdot \mathcal{L}h dm = \int_{\mathbb{T}^2} \varphi \circ f \cdot h dm, \quad \forall\, \varphi \in L^\infty(dm).$$

In fact, $(f, \mathbb{T}^2, m)$ is mixing. Moreover, without loss of generality we assume $\int_{\mathbb{T}^2} hdm = 1$.

We will prove that we can get exponential mixing. The main tool we use is Fourier Analysis. By direct computation,

$$(\widehat{\mathcal{L}^n h})_k = \int_{\mathbb{T}^2} e^{-2\pi i\langle k,x\rangle}\mathcal{L}^n h(x)dm(x)$$

where $k = (k_1, k_2) \in \mathbb{Z}^2$, $x = (x_1, x_2) \in \mathbb{T}^2$ and $\langle k, x\rangle = k_1x_1 + k_2x_2$.

For simplicity, we assume that $h \in C^r(\mathbb{T}^2)$ with $r > 2$, then

$$\left|\widehat{h}(k)\right| \le \frac{\|k\|_{C^r}}{(\|k\| + 1)^r}.$$

Using the unique representation and the symmetry of A, we have

$$(\widehat{\mathcal{L}^n h})_k = \int_{\mathbb{T}^2} e^{-2\pi i\langle k,A^n x\rangle} h(x)dm(x) = \int_{\mathbb{T}^2} e^{-2\pi i\langle A^n k,x\rangle} h(x)dm(x).$$

which shows that $(\widehat{\mathcal{L}^n h})_k = (\widehat{h})_{A^n k}$.

Hence, $\forall \varphi, h \in C^r(\mathbb{T}^2)$ and $\forall n \in \mathbb{N}$, we have (using Paserval's Theorem in the first inequality)

$$\begin{aligned}\left|\int \varphi \cdot \mathcal{L}^{2n} h dm - \int \varphi dm \cdot \int h dm\right| &\leq \sum_{k \neq (0,0)} |\widehat{\varphi}_k| \left|\widehat{h}_{A^{2n}k}\right| \\ &\leq \sum_{k \in \mathbb{Z}^2} \frac{\|\varphi\|_{C^r} \cdot \|h\|_{C^r}}{(\|k\|+1)^r (\|A^{2n}k\|+1)^r} \\ &\leq \sum_{k \in \mathbb{Z}^2} \frac{\|\varphi\|_{C^r} \cdot \|h\|_{C^r}}{(\|A^{-n}k\|+1)^r (\|A^n k\|+1)^r}.\end{aligned}$$

To obtain further estimates, note that since $k = a v^u + b v^s$ with $\|k\|^2 = a^2 + b^2$, then

$$A^n k = a\lambda^n v^u + b\lambda^{-n} v^s$$

and

$$A^{-n} k = a\lambda^{-n} v^u + b\lambda^n v^s.$$

Therefore,

$$\|A^n k\|^2 + \left\|A^{-n} k\right\|^2 \geq (a^2 + b^2)\lambda^{2n} = \|k\|^2 \lambda^{2n},$$

so

$$(\|A^n k\| + 1)(\left\|A^{-n} k\right\| + 1) \geq \|k\| \lambda^n$$

which gives

$$\left|\int \varphi \cdot \mathcal{L}^{2n} h dm - \int \varphi dm \cdot \int h dm\right| \leq C_r \|\varphi\|_{C^r} \|h\|_{C^r} \lambda^{-nr}$$

where C_r is a constant depending only on r.

This proves the exponential mixing (decay of correlation) of the transfer operator.

Now let us introduce the Sobolev norm. Consider $C^\infty(\mathbb{T}^2, \mathbb{C})$, we define

$$\|h\|_p^2 := \sum_{k \in \mathbb{Z}^2} \langle k \rangle^p \left|\widehat{h}_k\right|^2, \quad \langle k \rangle := 1 + \|k\|^2$$

and

$$\|h\|_{p\alpha}^2 := \sum_{k \in \mathbb{Z}^2} \langle k \rangle^{p\alpha(k)} \left|\widehat{h}_k\right|^2$$

where $\alpha \in C^0(\mathbb{P}, [-1, 1])$ is defined in the following way.

Fix some $\sigma \in (\lambda^{-1}, 1)$, then $\exists K > 0$ such that

$$\langle Av \rangle \geq \sigma^{-2} \langle v \rangle, \quad \forall v \in I_+, \|v\| > K,$$

$$\langle Av \rangle \leq \sigma^{2} \langle v \rangle, \quad \forall v \in I_-, \|v\| > K,$$

where $I_\pm$ are neighborhoods of v^u, v^s and we denote $I = A^{\pm} I_\pm \subset I_\pm$. Then α is defined as

$$\alpha(k) = \begin{cases} 1, & \text{if } k \in I_+, \\ -1, & \text{if } k \in I_-, \\ \text{others}, & \text{monotonic.} \end{cases}$$

For $v \in \mathbb{P}\backslash\{I_+ \cup I_-\}$, there exists $c > 0$ such that $d(v, Av) \geq c$. Therefore, $\exists\, \gamma > 0$ such that

$$\alpha(v) - \alpha(A^{-1}v) \geq \gamma \Leftrightarrow \alpha(v) - \gamma \geq \alpha(A^{-1}v), \quad \forall\, v \in \mathbb{P}\backslash\{I_+ \cup I_-\}.$$

We want to obtain Lasota-Yorke type inequality for the Sobolev norm. By computation, we have

$$\|\mathcal{L}h\|_{p\alpha}^2 = \sum_{k\in\mathbb{Z}^2} \langle k\rangle^{p\alpha(k)} \left|\widehat{h}_{Ak}\right|^2 = \sum_{k\in\mathbb{Z}^2} \frac{\langle A^{-1}k\rangle^{p\alpha(A^{-1}k)}}{\langle k\rangle^{p\alpha(k)}} \cdot \langle k\rangle^{p\alpha(k)} \left|\widehat{h}_k\right|^2.$$

If $k \in I_+$ and $k \geq K$, we have

$$\frac{\langle A^{-1}k\rangle^{\alpha(A^{-1}k)}}{\langle k\rangle^{\alpha(k)}} = \frac{\langle A^{-1}k\rangle}{\langle k\rangle} \leq \sigma^2.$$

If $k \in I_-$ and $k \geq K$, we have

$$\frac{\langle A^{-1}k\rangle^{\alpha(A^{-1}k)}}{\langle k\rangle^{\alpha(k)}} = \frac{\langle k\rangle}{\langle A^{-1}k\rangle} \leq \sigma^2.$$

If $k \notin I_+ \cup I_-$, let $B := \|A^{-1}\|$, then

$$\frac{\langle A^{-1}k\rangle^{\alpha(A^{-1}k)}}{\langle k\rangle^{\alpha(k)}} \leq \frac{\langle A^{-1}k\rangle^{\alpha(k)-\gamma}}{\langle k\rangle^{\alpha(k)}} \leq B \cdot \langle k\rangle^{-\gamma}.$$

Hence, denote

$$\Gamma := \left\{ k \in \mathbb{Z}^2 : \langle k\rangle \leq \max\{(\sigma^{-2}B)^{\frac{1}{\gamma}}, K\} := L \right\}$$

which is a finite set. Then

$$\sup_{k\notin\Gamma} \frac{\langle A^{-1}k\rangle^{\alpha(A^{-1}k)}}{\langle k\rangle^{\alpha(k)}} \leq \sigma^2.$$

Therefore, denote $\|h\|_{\omega}^{2} := \sum_{k\in\Gamma} \left|\widehat{h}_k\right|^2$

$$\begin{aligned}
\|\mathcal{L}h\|_{p\alpha} &= \sqrt{\sum_{k\in\Gamma} \frac{\langle A^{-1}k\rangle^{p\alpha(A^{-1}k)}}{\langle k\rangle^{p\alpha(k)}} \cdot \langle k\rangle^{p\alpha(k)} \left|\widehat{h}_k\right|^2 + \sum_{k\notin\Gamma} \cdots} \\
&\leq \sqrt{L^p \|h\|_{\omega}^2 + \sigma^{2p} \|h\|_{p\alpha}^2} \\
&\leq C \|h\|_{\omega} + \sigma^p \|h\|_{p\alpha} .
\end{aligned}$$

This is the Lasota-Yorke inequality that we want.

The following goal is to prove that for toral automorphism, the transfer operator is quasi-compact on $C^\infty(\mathbb{T}^2, \mathbb{C}, \|\cdot\|_{p\alpha})$. In fact, it is possible to prove that there is a Lasota-Yorke also on $\|\cdot\|_{p\beta}$ where $\exists\, c > 0$ s.t. $\beta + c \leq \alpha$. Moreover, $\|\cdot\|_{p\beta}$ is weakly compact on the space $\|\cdot\|_{p\alpha}$. By Hennion, we get quasi-compactness.

We finish this subsection with a remark.

Remark 5.6. If we consider C^r, C^α, the transfer operator will have spectral radius larger than 1. Compared with our Markov operator, controlling the Sobolev space here is like controlling the future when we consider observables depending on all coordinates.

6. Limit laws for hyperbolic systems

7. Partially hyperbolic systems

These notes may eventually become a book. Any suggestions for improvement are welcome. A more up-to-date version of certain parts of this manuscript can be found in our recent preprint [2]. The sections on transfer operators and uniformly hyperbolic (or even partially hyperbolic) maps may expand. A proof of the abstract CLT will be included (if we manage to come up with an argument ourselves, or find someone to help us with the translation from Russian of [4]). Well, we have a very long way to go.

References

1. Ao Cai, Pedro Duarte, and Silvius Klein, *Mixed random-quasiperiodic cocycles*, Bull. Braz. Math. Soc. (N.S.) **53** (2022), no. 4, 1469–1497. MR 4502841
2. Ao Cai, Pedro Duarte, and Silvius Klein, *Statistical properties for mixing markov chains with applications to dynamical systems*, 2023.
3. Mark F. Demers, Niloofar Kiamari, and Carlangelo Liverani, *Transfer operators in hyperbolic dynamics—an introduction*, 33 º Colóquio Brasileiro de Matemática, Instituto Nacional de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, 2021. MR 4472820
4. M. I. Gordin and B. A. Lifšic, *Central limit theorem for stationary Markov processes*, Dokl. Akad. Nauk SSSR **239** (1978), no. 4, 766–767.
5. Mikhail Gordin and Hajo Holzmann, *The central limit theorem for stationary Markov chains under invariant splittings*, Stoch. Dyn. **4** (2004), no. 1, 15–30.
6. Michael Hochman, *Notes on ergodic theory*, (2013).
7. M. Viana, *Lectures on Lyapunov exponents*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2014.

School of Mathematical Sciences, Soochow University, Soochow 215006, China and Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Brazil
E-mail address: godcaiao@gmail.com

Departamento de Matemática and CMAFcIO, Faculdade de Ciências, Universidade de Lisboa, Portugal
E-mail address: pmduarte@fc.ul.pt

Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Brazil
E-mail address: silviusk@puc-rio.br