MŰEGYETEM 1782

Thesis

Random walk on projective circe

Tamás Havas

Supervisor: Balázs Bárány

associate professor

BME Institute of Mathematics,

Department of Stochastics

BME

2018

Contents

Introduction

In the center of this thesis is the random walk on the projective circle, which simply consists of the directions of the real plain. The random walk can be expressed with products of random matrices, therefore it is a Markov-chain with a continuous state space. As in other cases of Markov-processes we are interessted in finding the recurrent states, therefore we are looking for measures on the set of directions, which will not change under the effect of the matrices. The support of these measures will contain all the recurrent states, and every other direction on the plain is transient. Since our distributions have a continuous domain, the recurrency should be considered to mean the revisiting of small neighbourhoods, instead of specific states.
More precisely for a discreate Markov-chain the distribution on the matrices would be

$$\mu(\{Y\}) = \mathbb{P}(X_n = Y\underline{x} | X_{n-1} = \underline{x}). \tag{1}$$

Where Y is a matrix of order two and $\underline{x}$ is a two dimensional vector. In our case however

$$\mu(\{Y | Y\underline{x} \in U\}) = \mathbb{P}(X_n \in U | X_{n-1} = \underline{x}) \tag{2}$$

for some open set U. And the questions are can find a probability measure on the plain for which $Y\underline{x}$ and $\underline{x}$ has the same distribution? Would it be continuous?

The Ergodic theorem which will be stated first is the work of Michael Keane and Karl Petersen [2]. Every other piece of theory in this work is borrowed from the first two chapter of the book *Products of Random Matrices with Applications to Schrödinger Operators* by Philippe Bougerol and Jean Lacroix, furthermore the example presented here was proposed there as an exercise.

Notations and basics

- $M(d, \mathbb{R})$: set of the d-dimensonal square matrices over the real numbers.
- $G1(d, \mathbb{R})$: set of the invertible elements of $M(d, \mathbb{R})$.
- $S1(d, \mathbb{R})$: subset of $G1(d, \mathbb{R})$, its elements have determinant one.
- $\|.\|$ for vectors: Euclidean norm
- $\|.\|$ for matrices: supremum norm
- $f^+(x) = \sup(f(x), 0)$ and $f^-(x) = \sup(-f(x), 0)$
- $[n] = 0, 1, 2, \ldots, n$

Definition 0.1. A *topological semigroup* is a topoligical set with an associative product on it.

Definition 0.2. A *topological group* is a group where $(g, h) \to gh$ and $g \to g^{-1}$ are continuous.

0.1 An ergodic theorem

Before we start to build our theorotical background, we state a theorem, the corollary of which will be of use later.

First of all some notations. Let $(\Omega, \mathcal{F}, \mu)$ be a probability space, $\theta : \Omega \to \Omega$ a measurepreserving transformation, and $f \in L^1(\Omega, \mathcal{F}, \mu)$. Furthermore let

$$A_k f = \frac{1}{k} \sum_{j=0}^{k-1} f\theta^j, \quad f_N^* = \sup_{1 \le k \le N} A_k f, \quad f^* = \sup_N f_N^* \text{ and } \overline{A} = \limsup_{k \to \infty} A_k f. \tag{3}$$

Theorem 0.1. *Let λ be an invariant function on X with $\lambda^+ \in L^1$ and $\lambda \circ \theta = \lambda$ a.e. Then*

$$\int_{\{f^* > \lambda\}} (f - \lambda) d\mu > 0. \tag{4}$$

Proof. If $\lambda \notin L^1(\{f^* > \lambda\})$, then

$$\int\limits_{\{f^*>\lambda\}} (f - \lambda^+) d\mu + \int\limits_{\{f^*>\lambda\}} (f + \lambda^-) d\mu = \infty > 0, \tag{5}$$

so we are done. Now assume that $\lambda \in L^1(\{f^* > \lambda\})$. Actually this implies that $\lambda \in L^1(X)$, because on $\{f^* \leq \lambda\}$ $f \leq \lambda$ must hold, therefore on this set $\lambda^- \leq -f + \lambda^+$, where the left-hand side is integrable.

For now suppose that $f \in L^\infty$, and for any positive N let

$$E_N = \{f_N^* > \lambda\}. \tag{6}$$

This gives us the upperbound

$$(f - \lambda)\mathbb{1}_{E_N} \geq (f - \lambda), \tag{7}$$

since from $x \notin E_N$ $(f - \lambda)(x) \leq 0$ follows. For the main part of this argument we take a large index $m \gg N$, and break done this sum

$$\sum_{k=0}^{m-1} (f - \lambda)\mathbb{1}_{E_N}(\theta^k x) \tag{8}$$

into short strings. One type of string only consists of zeros (meaning so far we only had x's outside of E_N). Then at some point the indicator switches to one, this non-zero string stops when it reaches the value of f_N^* (so its length is no more than N). We iterate this through the sum (so there may be consequtive non-zero strings). Formally, there is an $M \leq N$, for which

$$\sum_{k=k'}^{M+k'} (f - \lambda) \circ \mathbb{1}_{E_N}(\theta^k x) = \sum_{k=1}^{M} (f - \lambda) \circ \mathbb{1}_{E_N} \circ \theta^k(\theta^{k'} x) \geq M(f_N^* - \lambda)(x) > 0. \tag{9}$$

Silently we used the upper bound shown in (7). During this iteration we leave all these non-negative terms and at the end we are left with a string which is shorter than N. So there is a $j \in [m - N + 1, m]$, for which

$$\sum_{k=0}^{m-1} (f - \lambda)\mathbb{1}_{E_N}(\theta^k x) \geq \sum_{k=j}^{m-1} (f - \lambda)\mathbb{1}_{E_N}(\theta^k x) \geq -N(\|f\|_\infty + \lambda^+(x)). \tag{10}$$

By integrating the inequality, dividing by m and letting m go to ∞ we get

$$\begin{aligned} m\int\limits_{E_N}(f-\lambda)d\mu &\geq -N(\|f\|_\infty + \left\|\lambda^+\right\|_1), \\ \int\limits_{E_N}(f-\lambda)d\mu &\geq \frac{-N}{m}(\|f\|_\infty + \left\|\lambda^+\right\|_1), \\ \int\limits_{E_N}(f-\lambda)d\mu &\geq 0. \end{aligned} \tag{11}$$

Taking the limit to infinity in N and switching the integral and the limes with the Dominated Convergence Theorem gives back the theorem, for $f \in L^\infty$.

To prove for $f \in L^1$, we cut off f at some $s \in \mathbb{N}^+$. Let $\phi_s = f\mathbb{1}_{\{|f| \leq s\}}$, thus $\phi_s \in L^\infty$ and $\phi_s \to f$ a.s. and in L^1. Similary for a fixed N the following hold

$$(\phi_s)^*_N \to f^*_N \tag{12}$$

a.s. and in L^1 as well, moreover

$$\mu(\{(\phi_s)^*_N > \lambda\}\Delta\{f^*_N > \lambda\}) \to 0. \tag{13}$$

Therefore by the Dominated Convergence Theorem

$$0 \leq \int\limits_{\{(\phi_s)^*_N > \lambda\}} (\phi_s - \lambda)d\mu \to \int\limits_{\{f^*_N > \lambda\}} (f-\lambda)d\mu. \tag{14}$$

Again letting N go to infinity concludes the proof. □

Corollary 0.1. *The sequence $(A_k)f$ converges a.e.*

Proof. It is enough to show that

$$\int \overline{A} \leq \int f. \tag{15}$$

Because then, applying it to $-f$ gives

$$-\int \underline{A} \leq -\int f, \tag{16}$$

where $\underline{A} = \liminf A_k f$. Therefore

$$\int \overline{A} \leq \int f \leq \int \underline{A} \leq \int \overline{A}. \tag{17}$$

Which means that $\underline{A} = \overline{A}$ a.e.

In the last part of the argument ε denotes a positive real number, and convergences hold by letting $\varepsilon \to 0$.

To use the previous theorem we need a well chosen λ and for that we need $(\overline{A})^+$ to be integrable. Consider the $\overline{A}$ associated with f^+ (call it $\overline{A}_{f^+}$). Let $\lambda_1 = \overline{A}_{f^+} - \varepsilon$ be an invariant function (that is the property assumed in the Theorem 0.1). We have $\{(f^+)^* > \lambda_1\} = X$, thus

$$\int f^+ \geq \int \lambda_1 \to \int \overline{A}_{f^+}. \tag{18}$$

Therefore $(\overline{A})^+ < \overline{A}_{f^+}$ integrable. Now let $\lambda_2 = \overline{A}_f - \varepsilon$, by the previous theorem

$$\int f \geq \int \lambda_2 \to \int \overline{A}_f. \tag{19}$$

□

1 The upper Lyapunov exponent

In this section we will introduce a value associated with a spesific random walk in the same manner it was done in the first chapter of [1]. The importance of this mysterious upper Lyapunov exponent will be shown later.

Let $\{Y_i \mid i \geq 1\}$ be i.i.d. random matrices and $S_n = Y_n \dots Y_1$. According to Cauchy and Schwarz

$$\|S_n\| \leq \|Y_n\| \dots \|Y_1\|. \tag{20}$$

So if $\mathbb{E}(\log^+\|Y_1\|) < \infty$, then $\log^+\|S_n\|$ is integrable. Furthermore for any $n, m \in \mathbb{N}^+$

$$\mathbb{E}(\log\|S_{n+m}\|) \leq \mathbb{E}(\log\|Y_{n+m} \dots Y_{n+1}\| + \log\|S_n\|) = \mathbb{E}(\log\|S_m\|) + \mathbb{E}(\log\|S_n\|). \tag{21}$$

Hence the sequence $a_n = \mathbb{E}(\log\|S_n\|)$ is subadditive. Given a fixed $m \in \mathbb{N}^+$ any $n \in \mathbb{N}^+$ can be expressed as $n = mp + q$ for some $p \in \mathbb{N}$ and $q \in [m-1]$. Using the subadditivity

$$\frac{a_n}{n} = \frac{a_{mp+q}}{mp+q} \leq p\frac{a_m}{mp+q} + \frac{a_q}{mp+q}. \tag{22}$$

The right-hand side converges to $\frac{a_m}{m}$ as $n \to \infty$, therefore $\overline{\lim_{n\to\infty}} \frac{a_n}{n} \leq \frac{a_m}{m}$ for any $m \in \mathbb{N}^+$, thus $\frac{1}{n}\mathbb{E}(\log\|S_n\|)$ converges to $\inf_{m\in\mathbb{N}^+} \frac{1}{m}\mathbb{E}(\log\|S_m\|)$ (note that this can take values from $\mathbb{R} \cup \{-\infty\}$).

Definition 1.1. Let $\{Y_i \mid i \geq 1\}$ be i.i.d. random matrices. If $\mathbb{E}(\log^+\|Y_1\|) < \infty$, then we call the following value *the upper Lyapunov exponent*:

$$\gamma = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}(\log\|Y_n \ldots Y_1\|). \tag{23}$$

1.1 Cocycles

Let G be a topological semigroup and B a topological space.

Definition 1.2. We say that G *acts on* B, if there is a continuous function $\bullet$: $G \times B \to B$, which has the following propety for any $g_1, g_2 \in G$ and $x \in B$:

$$(g_1 g_2) \bullet x = g_1 \bullet (g_2 \bullet x). \tag{24}$$

Furthermore if G is a group with unit e, for which $e \bullet x = x$ is true for any $x \in B$, then we call B a *G-space*.

From now on suppose that G is acting on B.

Definition 1.3. A countinuous map $\sigma : G \times B \to \mathbb{R}$ is called *an additive cocycle*, if

$$\sigma(g_1 g_2, x) = \sigma(g_1, g_2 \bullet x) + \sigma(g_2, x) \tag{25}$$

for any $g_1, g_2 \in G$ and $x \in B$.

Definition 1.4. Let μ be a probability measure on G, and similarly ν on B. We denote by $\mu *^\bullet \nu$ the distribution on B which satisfies

$$\int\limits_B f(x) d(\mu *^\bullet \nu)(x) = \int\limits_G \int\limits_B f(g \bullet x) d\mu(g) d\nu(x) \tag{26}$$

for any Borel-function f.

Definition 1.5. Having the same setting as in the previous definition we call ν $\mu - invariant$, if $\mu *^\bullet \nu = \nu$.

Let μ and λ be two probability measures on G. The convolution product $\mu * \lambda$ is defined by

$$\int f(k) d(\mu * \lambda)(k) = \int \int f(gh) d\mu(g) d\lambda(h) \tag{27}$$

for any f on G, and $\mu^i = \mu^{i-1} * \mu$ for $i > 1$. Note that $(\mu * \lambda) *^\bullet \nu = \mu *^\bullet (\lambda *^\bullet \nu)$ for a distribution ν on B.

Theorem 1.1. *Let σ be an additive cocycle on $G \times B$, and $\{Y_n \mid n \geq 1\}$ i.i.d. elements of G with distribution μ. If ν is a μ-invariant distribution on B, for which $\int\int |\sigma(g,x)| d\mu(g) d\nu(x) < \infty$ holds, then*

$$\lim_{n\to\infty} \frac{1}{n}\sigma(Y_n(\omega)\dots Y_1(\omega), x) \tag{28}$$

exists for $\mathbb{P} \otimes \nu$-almost all (ω, x), where ω takes the possible realizations of the sequence $(Y_n \mid n \in \mathbb{N}^+)$, $Y_i(\omega)$ is just the i^{th} coordinate of ω, and $\mathbb{P} = \bigotimes_1^\infty \mu$.

Proof. Let us define $\theta^p((Y_n)_{n\in\mathbb{N}^+}, x) = ((Y_{n+p})_{n\in\mathbb{N}^+}, (Y_p \dots Y_1) \bullet x)$ for $p \in \mathbb{N}$, and $F(\omega, x) = \sigma(Y_1, x)$. Let A_0 be a Borel subset of B, and $A_1, \dots$ Borel subsets of G. Then

$$\begin{aligned}(\mathbb{P}\otimes\nu)\{(\omega,x) \mid \theta^1(\omega,x) \in (A_1\times\dots)\times A_0\} &= (\mathbb{P}\otimes\nu)\{(\omega,x) \mid Y_2 \in A_1, \dots, Y_1 \bullet x \in A_0\}\\ &= (\mathbb{P}\otimes\nu)\{(\omega,x) \mid Y_2 \in A_1, \dots, x \in A_0\}\\ &= (\mathbb{P}\otimes\nu)\{(\omega,x) \mid Y_1 \in A_1, \dots, x \in A_0\}.\end{aligned} \tag{29}$$

Which means that θ^p preserves $(\mathbb{P}\otimes\nu)$. At the first equality we used the definition of θ. The second holds, since the μ-invariance of ν impies $\int \mathbb{1}_{A_0} d(\mu *^\bullet \nu) = \int \mathbb{1}_{A_0} d\nu$. And at last we used the fact that the Y_i-s are i.i.d-s to shift the indices.

Now look at σ for a part of the sequence:

$$\begin{aligned}\sigma(Y_n\dots Y_1, x) &= \sigma(Y_n, (Y_{n-1}\dots Y_1)\bullet x) + \sigma(Y_{n-1}\dots Y_1, x)\\ &= \sum_{p=1}^n \sigma(Y_p, (Y_{p-1}\dots Y_1)\bullet x)\\ &= \sum_{p=1}^n F(\theta^{p-1}(\omega, x)).\end{aligned} \tag{30}$$

We used that σ is an additive cocycle and the definition of F. Therefore by the Corrolary 0.1

$$\frac{1}{n}\sigma(Y_n\dots Y_1, x) = \frac{1}{n}\sum_{p=1}^n F(\theta^{p-1}(\omega, x)) \tag{31}$$

converges $(\mathbb{P}\otimes\nu)$ a.s. □

In this theorem we assumed that there is a μ-invariant distribution, but now we will show that sometimes it does exist.

Lemma 1.1. *Let B be a compact separable G-space and μ be a distribution on G. For any distribution m on B each limit point of $\{\frac{1}{n}\sum_{i=1}^{n}\mu^i *^\bullet m \mid n \in \mathbb{N}^+\}$ is a μ-invariant distribution on B.*

Proof. Let $\nu_n = \frac{1}{n}\sum_{i=1}^{n}\mu^i *^\bullet m$. Since B is separable and compact (ν_n) has a weakly convergent subsequence. Denote its limit by ν, which is a probability measure on B. Now for any $n \in \mathbb{N}^+$

$$\begin{aligned}\mu *^\bullet \nu_n &= \frac{1}{n}\sum_{i=1}^{n}\mu^{i+1} *^\bullet m \\ &= \frac{1}{n}\sum_{i=1}^{n}\mu^{i} *^\bullet m + \frac{1}{n}(\mu^{n+1} *^\bullet m - \mu *^\bullet m) \\ &= \nu_n + \frac{1}{n}(\mu^{n+1} *^\bullet m - \mu *^\bullet m).\end{aligned} \tag{32}$$

By letting $n \to \infty$ we get $\mu *^\bullet \nu = \nu$. □

1.2 The theorem of Furstenberg and Kesten

With this background we are able to prove the following useful theorem.

Theorem 1.2. *Let $Y_1, \dots$ be i.i.d. matrices in $G1(d,\mathbb{R})$ with distribution μ. If $\mathbb{E}(log^+\|Y_1\|)$ is finite, then with probability one*

$$\lim_{n\to\infty}\frac{1}{n}\log\|Y_1 \dots Y_n\| = \gamma, \tag{33}$$

where γ is the upper Lyapunov exponent associated with μ.

Proof. We will prove this in two parts, first we check it for $\gamma = -\infty$, then for the finite case.

Fix an integer m. Every $n \in \mathbb{N}^+$ can be written as $n = pm + q$ for some $p \in \mathbb{N}$ and $q \in [m-1]$.

$$\frac{1}{n}\log\|Y_n \dots Y_1\| \le \frac{1}{n}\sum_{i=1}^{q}\log\|Y_i\| + \frac{1}{pm+q}\sum_{j=0}^{p-1}\log\big\|Y_{(j+1)m}\dots Y_{jm+1}\big\|. \tag{34}$$

By the Strong Law of Large Numbers for any $m \in \mathbb{N}^+$

$$\limsup_{n\to\infty}\frac{1}{n}\log\|Y_n \dots Y_1\| \le \frac{1}{m}\mathbb{E}(\log\|Y_m \dots Y_1\|). \tag{35}$$

Therefore with probability one

$$\limsup_{n\to\infty} \frac{1}{n} \log\|Y_n \dots Y_1\| \le \lim_{m\to\infty} \frac{1}{m}\mathbb{E}(\log\|Y_m \dots Y_1\|) = \gamma. \tag{36}$$

Now we will deal with the case when $\gamma \in \mathbb{R}$. Let B be the subset of $M(d, \mathbb{R})$, where the norm is strictly one. Take a $Y \in G1(d, \mathbb{R})$ and $M \in B$, define

$$Y \bullet M = \frac{YM}{\|YM\|}, \tag{37}$$

then B is a $G1(d, \mathbb{R})$-space. Set m as the Dirac-mass at the identity matrix on B and μ the distribution of Y_1. For an integer n look at the distribution ν_n on B defined by

$$\nu_n = \frac{1}{n} \sum_{i=1}^{n} \mu^i *^{\bullet} m. \tag{38}$$

From Lemma 1.1 we know, that for a convergent subsequence (ν_{n_i}) the limit distribution ν is μ-invariant. Define σ: $G1(d, \mathbb{R}) \times B \to \mathbb{R}$ by

$$\sigma(Y, M) = \log\|YM\|, \quad Y \in G1(d, \mathbb{R}), \; M \in B. \tag{39}$$

For any Y_1, Y_2 in $G1(d, \mathbb{R})$ and M in B

$$\begin{aligned} \sigma(Y_1 Y_2, M) &= \log\|Y_1 Y_2 M\| \\ &= \log\left\|Y_1 \frac{Y_2 M}{\|Y_2 M\|}\right\| + \log\|Y_2 M\| \\ &= \sigma(Y_1, Y_2 \bullet M) + \sigma(Y_2, M), \end{aligned} \tag{40}$$

which means that σ is an additive cocycle.

We know that $\sigma(Y, M) \le \log\|Y\|\|M\| = \log\|Y\|$, therefore we have

$$\sigma^+(Y, M) \le \log^+\|Y\|. \tag{41}$$

We had the condition $\mathbb{E}(\log^+\|Y_1\|) < \infty$, thus $\int \sigma^+(Y, M) d\mu(Y)$ is a bounded continuous function on B and

$$\lim_{i\to\infty} \int\int \sigma^+ d\mu d\nu_{n_i} = \int\int \sigma^+ d\mu d\nu. \tag{42}$$

Now look at σ^-. For any constant $k \in \mathbb{N}^+$

$$\liminf_{i\to\infty} \int\int \sigma^- d\mu d\nu_{n_i} \ge \liminf_{i\to\infty} \int\int \inf\{k, \sigma^-\} d\mu d\nu_{n_i} = \int\int \inf\{k, \sigma^-\} d\mu d\nu. \tag{43}$$

Since it is true for all k we can take the limes-inferior of the right-hand side, and then use Fatou's lemma, or formally

$$\liminf_{i\to\infty} \int\int \sigma^- d\mu d\nu_{n_i} \geq \liminf_{k\to\infty} \int\int \inf\{k, \sigma^-\} d\mu d\nu \geq \int\int \sigma^- d\mu d\nu. \tag{44}$$

Using the definition of σ^+ and σ^- combined with the results of (42) and (44) we obtain

$$\limsup_{i\to\infty} \int\int \sigma d\mu d\nu_{n_i} \leq \int\int \sigma d\mu d\nu. \tag{45}$$

By the definition of ν_n

$$\begin{aligned} \int\int \sigma(Y, M) d\mu(Y) d\nu_n(M) &= \frac{1}{n}\sum_{i=1}^{n} \int\int \sigma(Y, M) d(\mu^i *^{\bullet} m) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(\sigma(Y_{i+1}, (Y_i \dots Y_1) \bullet id)) \\ &= \frac{1}{n}\mathbb{E}(\sigma(Y_n \dots Y_1, id)) \\ &= \frac{1}{n}\mathbb{E}(\log\|Y_n \dots Y_1\|). \end{aligned} \tag{46}$$

Where we used the definition of $\mathbb{E}(.)$, the property of the additive cocycle and the definition of σ. This yields that

$$\int\int \sigma d\mu d\nu \geq \lim_{i\to\infty} \int\int \sigma d\mu d\nu_{n_i} = \gamma. \tag{47}$$

Since the left-hand side of (42) and (44) add up to a finite value, σ is in $L^1(\mu \otimes \nu)$. By Theorem 1.1 for some $f : \Omega \to \mathbb{R}$, as $n \to \infty$,

$$\frac{1}{n}\log\|Y_n(\omega)\dots Y_1(\omega)M\| = \frac{1}{n}\sigma(Y_n(\omega)\dots Y_1(\omega)M) \to f(\omega, M) \tag{48}$$

with probability one. Using the Cauchy-Schwarz inequality and (36)

$$f(\omega, M) \leq \liminf_{n\to\infty} \frac{1}{n}(\log\|Y_n \dots Y_1\| + \log\|M\|) \leq \gamma. \tag{49}$$

We want to switch the integrals with the limes, so we look at

$$\begin{aligned} \int\int \left|\frac{1}{n}\sigma(Y_n \dots Y_1, M)\right| d\mathbb{P} d\nu &\leq \int\int \frac{1}{n}\sum_{i=1}^{n} |\sigma(Y_i, Y_{i-1} \dots Y_1 M)| d\mathbb{P} d\nu \\ &= \int\int |\sigma(Y_1, M)| d\mu d\nu < \infty. \end{aligned} \tag{50}$$

We used the properties of the additive cocycle and the i.i.d. property of Y_i with the μ-invariance of ν inside the sum. Therefore by the Dominated Convergence Theorem

$$\gamma \leq \lim_{n\to\infty} \int\int \frac{1}{n}\sigma(Y_n \dots Y_1, M) d\mathbb{P}d\nu = \int\int f(\omega, M) d\mathbb{P}d\nu, \tag{51}$$

where the inequality holds by ergodicity and (47). (49) and (51) together imply that $f \equiv \gamma$ almost surely. So (36) and the first statement of (49) together conclude the theorem. $\square$

2 Matrices of order two

In this section we will talk about the property that under broad conditions any pair of non-zero starting vectors tends to "line up" quite fast during the random walk. Also this section correspondes to the second chapter of [1].

We say that two vectors ($x, y \in \mathbb{R}^d$) have the same direction, if there is a constant $c \in \mathbb{R}$ for which $cx = y$. This is an equivalence relation (Γ) on $\mathbb{R}^d - \{0\}$. Formally

Definition 2.1. *The projective cicrle* is the set of directions on $\mathbb{R}^d$. We denote it by $P(\mathbb{R}^d)$, which is defined as the quotient space $\mathbb{R}^d - \{0\}/\Gamma$.

For an $x \in \mathbb{R}^d$ $\overline{x}$ will be its direction (i.e. its class).

For $M \in G1(d, \mathbb{R})$ we set $M \bullet \overline{x} = \overline{Mx}$.

We introduce a metric so we can talk about this tendency of lining up. Let x, y be unit vectors in $\mathbb{R}^d$

$$\delta(\overline{x}, \overline{y}) = (1- < x, y >^2)^{\frac{1}{2}}. \tag{52}$$

This is basicly the sine of the angle between the vectors. In case of $d = 2$ for any x, y in $\mathbb{R}^2 - \{0\}$

$$\delta(\overline{x}, \overline{y}) = \frac{|x_1 y_2 - x_2 y_1|}{\|x\|\|y\|} = \frac{1}{\|x\|\|y\|}|det([x|y])|. \tag{53}$$

For some $A \in G1(d, \mathbb{R}^2)$

$$\begin{aligned} \delta(\overline{Ax}, \overline{Ay}) &= \frac{1}{\|Ax\|\|Ay\|}|det([Ax|Ay])| = \frac{1}{\|Ax\|\|Ay\|}|det(A \cdot [x|y])| \\ &= \frac{|det(A)|}{\|Ax\|\|Ay\|}|det([x|y])| = \frac{|det(A)|\|x\|\|y\|}{\|Ax\|\|Ay\|}\delta(\overline{x}, \overline{y}) \end{aligned} \tag{54}$$

Now let $\{Y_i\}_{i \in \mathbb{N}^+}$ be i.i.d. random matrices with determinant one and $S_n = Y_n \dots Y_1$. We can make this assumption without loss of generality, because one can devide every matrix with the squareroot of its determinant, and the random walk on $P(\mathbb{R}^2)$ would remain the same. Therefore we are looking for conditions for which the term

$$\delta(S_n \overline{x}, S_n \overline{y}) = \frac{\|x\|\|y\|}{\|S_n x\|\|S_n y\|}\delta(\overline{x}, \overline{y}) \tag{55}$$

goes to zero exponentially fast, i.e. that the upper Lyapunov exponent associated with the distribution is positive.

2.1 The two lemmas

We will need two lemmas to state Furstenberg's theorem.

Lemma 2.1. *Let G be a topological semigroup acting on a compact separable space B. Let $\{X_i\}_{i\in\mathbb{N}^+}$ be independent random elements of G with distribution μ, and let ν be a μ-invariant distribution on B. Then for almost all ω there exists a probability measure ν_ω on B, such that*

$$\{X_1(\omega)\dots X_n(\omega)g\nu \mid n\in\mathbb{N}^+\} \tag{56}$$

converges weakly to ν_ω as $n\to\infty$ for almost all $g\in G$ with respect to $\lambda = \sum_{i=1}^{\infty} 2^{-n-1}\mu^n$, and for any bounded Borel-function f on B

$$\int f d\nu = \mathbb{E}\Big(\int f d\nu_\omega\Big). \tag{57}$$

Proof. Let f be a bounded countinuous real Borel-function on B. Define $F: G\to\mathbb{R}$ as

$$F(g) = \int f(g\bullet x)d\nu(x). \tag{58}$$

Let $M_n = X_1\dots X_n$ and $\mathcal{F}_n$ be the $\sigma-algebra$ generated by $X_1\dots X_n$. Now

$$\begin{aligned}\mathbb{E}(F(M_{n+1})|\mathcal{F}_n) &= \int F(M_n g)d\mu(g)\\ &= \int\int f(M_n g\bullet x)d\mu(g)d\nu(x)\\ &= \int f(M_n\bullet x)d(\mu *^{\bullet} \nu)(x)\\ &= \int f(M_n\bullet x)d\nu(x)\\ &= F(M_n).\end{aligned} \tag{59}$$

We have used the properties of conditional expectation, the definition of $F(.)$ and the fact that ν is μ-invariant. Thus $\{F(M_n)\}_{n\in\mathbb{N}^+}$ is a bounded martingale. Therefore $F(M_n)$ conferges a.s. to some Φ_f and

$$\mathbb{E}(\Phi_f) = \mathbb{E}(F(M_1)) = \int f d\nu. \tag{60}$$

For any indices k and r,

$$\begin{aligned}
\mathbb{E}(|F(M_{k+r}) - F(M_k)|^2) &= \mathbb{E}(F(M_{k+r})^2) + \mathbb{E}(F(M_k)^2) - 2\mathbb{E}(F(M_{k+r})F(M_k)) \\
&= \mathbb{E}(F(M_{k+r})^2) + \mathbb{E}(F(M_k)^2) - 2\mathbb{E}(\mathbb{E}(F(M_{k+r})F(M_k)|\mathcal{F}_k)) \\
&= \mathbb{E}(F(M_{k+r})^2) + \mathbb{E}(F(M_k)^2) - 2\mathbb{E}(F(M_k)^2) \\
&= \mathbb{E}(F(M_{k+r})^2) - \mathbb{E}(F(M_k)^2).
\end{aligned} \tag{61}$$

Where the first equation holds by the linearity of the expectation, the second by the tower property and the third by the martingale property. Using the cancellation in the summation, for any p,

$$\begin{aligned}
\sum_{k=1}^{p} \mathbb{E}(|F(M_{k+r}) - F(M_k)|^2) &= \sum_{k=1}^{r} \mathbb{E}(F(M_{k+p})^2) - \sum_{k=1}^{r} \mathbb{E}(F(M_k)^2) \\
&\leq \sum_{k=1}^{r} \mathbb{E}(F(M_{k+p})^2) + \sum_{k=1}^{r} \mathbb{E}(F(M_k)^2) \\
&\leq 2r \sup_{g\in G} |F(g)|^2.
\end{aligned} \tag{62}$$

From this

$$\begin{aligned}
\sum_{k=1}^{\infty} \mathbb{E}\Big(\int |F(M_k g) - F(M_k)|^2 d\lambda(g)\Big) &= \sum_{k=1}^{\infty}\sum_{r=0}^{\infty} \frac{1}{2^{r+1}} \mathbb{E}\Big(\int |F(M_k g) - F(M_k)|^2 d\mu^r(g)\Big) \\
&= \sum_{k=1}^{\infty}\sum_{r=0}^{\infty} \frac{1}{2^{r+1}} \mathbb{E}\Big(|F(M_{k+r}) - F(M_k)|^2\Big) \\
&\leq \sum_{r=0}^{\infty} r \frac{1}{2^r} \sup_{g\in G} |F(g)|^2.
\end{aligned} \tag{63}$$

Which is summable, therefore $F(M_k g)$ converges to Φ_f $\mathbb{P}\otimes\lambda$-almost surely as well.

Choose a dense sequence $(f_q \mid f_q \in \mathcal{C}_0(B))_{q\in\mathbb{N}^+}$ of continuous functions on B (it is possible, because B is separable and compact, thus $\{f \mid f : B \to \mathbb{R}\}$ is separable as well). There is a subset A of $\Omega \times G$ with measure one (wrt $\mathbb{P}\otimes\lambda$) such that if $(\omega, g) \in A$, then

$$\int f_q(M_n(\omega)g \bullet x) d\nu(x) \to \Phi_{f_q}(\omega). \tag{64}$$

If $\nu_{\omega,g}$ is a limit point of $W = \{M_n(\omega)g\nu\}_{n\in\mathbb{N}^+}$, then for all q

$$\int f_q d\nu_{\omega,g} = \Phi_{f_q}(\omega). \tag{65}$$

Where the left-hand side does not depend on the measure of the integration. With our dense sequence of functions we can approximate the integral of any continuous Borel-function upto an arbitrary small error, and similarly with continuous functions we can get the integral of indicator functions upto a negligible error. If we do this approximations for different limit points of W, we get that the integrals of indicators can get closer to the same value than any positive real number. From this we obtain, that there is only one limit point $\nu_{\omega,g}$, furthermore it does not depend on g. Therefore we can denote it by ν_ω. From the martingale convergence we know that

$$\int f_q d\nu = \mathbb{E}(\Phi_{f_q}) = \mathbb{E}\Big(\int f_q d\nu_\omega\Big). \tag{66}$$

And since (f_q) is dense and its domain is a compact set the previous equation holds for any bounded Borel-function. □

Lemma 2.2. *Let B be a G-space, $\{Y_i\}_{i\in\mathbb{N}^+}$ independent random elements of G with distribution μ and σ an additive cocycle on $G \times B$. Suppose that ν is a μ-invariant distribution on B, for which*

(i)

$$\int\int \sigma^+(g,x)d\mu(g)d\nu(x) < \infty, \text{ and} \tag{67}$$

(ii) for $\mathbb{P}\otimes\nu$-almost all (ω,x), $\lim_{n\to\infty} \sigma(Y_n(\omega)\dots Y_1(\omega),x) = \infty$.

Then σ is in $L^1(\mathbb{P}\otimes\nu)$ and $\int\int \sigma(g,x)d\mu(x)d\nu(x) > 0$

Proof. This follows immediately once we understand the next lemma. □

Lemma 2.3. *Let $(E,\mathcal{F},\lambda)$ be a probability space and $\theta : E \to E$ a measurable transformation which preserves λ. If $f : E \to \mathbb{R}$ is such that $\int f^+ d\lambda < \infty$ and $\lim_{n\to\infty} \sum_{i=1}^{n} f\theta^i = \infty$ almost everywhere, then $f \in L^1(\lambda)$ and $\int f d\lambda > 0$.*

Proof. By Corrolary 0.1 we know that $\frac{1}{n}\sum_{i=1}^n f\theta^i$ converges to some ϕ as n goes to ∞. Define $\mathcal{J} = \{A \in \mathcal{F} \mid \lambda(A\Delta\theta^{-1}A) = 0\}$, where Δ denotes the symmetric difference. It is easy to show that $\mathcal{J}$ is a σ-algebra, and that the limit $\phi = \mathbb{E}(f|\mathcal{J})$. Since $\sum_{i=1}^n f\theta^i \to \infty$, $\mathbb{E}(f|\mathcal{J})$ is non-negative, and

$$\mathbb{E}(f^+) - \mathbb{E}(f^-) = \mathbb{E}(f) = \mathbb{E}(\mathbb{E}(f|\mathcal{J})) \geq 0. \tag{68}$$

So $\int f^- d\lambda \leq \int f^+ d\lambda < \infty$, thus f is in L^1.

Suppose that $\int f d\lambda = 0$, then using that $\mathbb{E}(f|\mathcal{J}) \geq 0$ with probability one

$$\frac{1}{n}\sum_{i=1}^{n} f\theta^i \to 0 \tag{69}$$

For any $\varepsilon > 0$ define $I_\varepsilon(t) = [t-\varepsilon, t+\varepsilon]$ and $S_n(x) = \left(\sum_{i=1}^{n} f\theta^i\right)(x)$. Choosing m to be the Lebesque-measure on $\mathbb{R}$, define

$$R_n^\varepsilon(x) = m\left(\bigcup_{i=1}^{n} I_\varepsilon(S_i(x))\right). \tag{70}$$

By (69) for almost all x and any δ there is an index n_0, above which (meaning $k > n_0$) $|S_k(x)| \leq k\delta$. Therefore for all n $R_n^\varepsilon(x) \leq R_{n_0}^\varepsilon(x) + 2(n\delta + \varepsilon)$. From this we obtain $\limsup_{n\to\infty} \frac{1}{n} R_n^\varepsilon(x) \leq 2\delta$ for any $\delta > 0$, which means that $\frac{1}{n} R_n^\varepsilon(x) \to 0$ a.s. and by dominated convergence

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}(R_n^\varepsilon(x)) = 0. \tag{71}$$

Using the fact, that $S_n \circ \theta = S_{n+1} - S_1$, it can be shown

$$\begin{aligned} R_{n+1}^\varepsilon(x) - R_n^\varepsilon(x)\circ\theta &= m\left(\bigcup_{i=1}^{n+1} I_\varepsilon(S_i(x))\right) - m\left(\bigcup_{i=1}^{n} I_\varepsilon(S_{i+1}(x) - S_1(x))\right) \\ &= m\left(\bigcup_{i=1}^{n+1} I_\varepsilon(S_i(x))\right) - m\left(\bigcup_{i=2}^{n+1} I_\varepsilon(S_i(x))\right). \end{aligned} \tag{72}$$

So there is a simple lower bound, namely

$$R_{n+1}^\varepsilon(x) - R_n^\varepsilon(x)\circ\theta \geq 2\varepsilon\, \mathbb{1}_{\{|S_i - S_1| > 2\varepsilon \ | \ i=2,\ldots,n+1\}}. \tag{73}$$

Integrating the inequality we arrive at

$$\mathbb{E}(R_{n+1}^\varepsilon(x)) - \mathbb{E}(R_n^\varepsilon(x)\circ\theta) \geq 2\varepsilon\lambda(\{x \mid |S_i(x) - S_1(x)| > 2\varepsilon,\ i = 2,\ldots,n+1\}). \tag{74}$$

Since θ preserves λ,

$$\mathbb{E}(R_{n+1}^\varepsilon(x)) - \mathbb{E}(R_n^\varepsilon(x)) \geq 2\varepsilon\lambda(\{x \mid |S_i(x)| > 2\varepsilon,\ i = 1,\ldots,n\}). \tag{75}$$

Which implies

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}(R_n^\varepsilon(x)) \geq 2\varepsilon\lambda(\{x \mid |S_i(x)| > 2\varepsilon,\ i \in \mathbb{N}^+\}). \tag{76}$$

By (69), for all $\varepsilon > 0$

$$\lambda(\{x \mid |S_i(x)| > \varepsilon,\ i \in \mathbb{N}^+\}) = 0 \tag{77}$$

Since $S_i \sim (S_{i+p} - S_p)$ holds for any $p \in \mathbb{N}^+$,

$$\lambda(\{x \mid |S_{i+p}(x) - S_p(x)| > \varepsilon,\ i \in \mathbb{N}^+\}) = 0. \tag{78}$$

But this contradicts the assumption, that $\lim_{n\to\infty} S_n = \infty$ a.s.

□

2.2 Contraction properties

Take a sequence (Y_i) of i.i.d. random matrices of order two and with $|\det(Y_i)| = 1$. We want to show that under some conditions for any non-zero $x \in \mathbb{R}^2$, $\|Y_n \dots Y_1 x\|$ goes to infinity with probability one. By the following result (Theorem 2.1) we can see, that we should look at the action of the transposed walk.

First of all we introduce a notation. Let $M \in G1(d, \mathbb{R})$ and m be a probability measure on $P(\mathbb{R}^d)$, we denote by Mm the probability measure on $P(\mathbb{R}^d)$ defined by

$$\int f d(Mm) = \int f(M \bullet \overline{x}) dm(\overline{x}) \tag{79}$$

for any bounded Borel function f.

Lemma 2.4. *Let A be a non-zero matrix (not necessarily invertible) of order two, and m be a continuous distribution on $P(\mathbb{R}^2)$. Then the equation*

$$\int f d(Am) = \int f(A \bullet \overline{x}) dm(\overline{x}), \tag{80}$$

valid for all bounded Borel functions, defines a probability measure Am on $P(\mathbb{R}^2)$. And if (A_n) is a sequence of non-zero matrices which converges to A, then $A_n m$ converges to Am.

Proof. Take any $x \in \mathbb{R}^2$. If $Ax \neq 0$, then $A \bullet \overline{x} = \overline{Ax}$ is well defined. Since $A \neq 0$ there is atmost one direction $\overline{y}$, for which if y has this direction, $Ay = 0$. Since m is continuous, $m(\{\overline{y}\}) = 0$. This means that Am can be defined m-almost everywhere, furthermore $\int \mathbb{1}_{\overline{x}} dAm(\overline{x}) = \int \mathbb{1}_{A\bullet\overline{x}} dm(\overline{x}) = 1$.

If $A_n \to A$, then for all $\overline{x}$ outside a countable set (which consists of the directions of the kernels of the sequence) $A_n \bullet \overline{x}$ exists and converges to $A \bullet \overline{x}$. $\square$

With this we can prove the following.

Theorem 2.1. *Let $(A_n)_{n\in\mathbb{N}^+}$ be a sequence of 2×2 matrices with determinant one. Suppose that there exists a continuous distribution m on $P(\mathbb{R}^2)$ such that $A_n m$ converges weakly to a Dirac-measure $\delta_{\overline{z}}$. Then*

$$\lim_{n\to\infty} \|A_n\| = \lim_{n\to\infty} \left\|A_n^T\right\| = \infty, \tag{81}$$

and if z is a unit vector with direction $\overline{z}$, then for any $x \in \mathbb{R}^2$

$$\lim_{n\to\infty} \frac{\left\|A_n^T x\right\|}{\|A_n^T\|} = |< x, z >|. \tag{82}$$

Proof. Suppose that $\|A_n\|^{-1} A_n$ converges to some matrix A. This implies that $\|A\| = 1$, and therefore $A \neq 0$. By the previous lemma and our assumption

$$Am = \delta_{\overline{z}}. \tag{83}$$

We know that $\det A$ can only be zero, otherwise m would be a Dirac-measure at $A^{-1} \bullet \overline{z}$, which is not continuous. From this

$$0 = |\det A| = \lim_{n\to\infty} \left| \frac{\det A_n}{\|A_n\|^2} \right| = \lim_{n\to\infty} \frac{1}{\|A_n\|^2} \tag{84}$$

proving (81).

Moreover the range of A is a line with the direction $\overline{z}$. So take a unit vector z with direction $\overline{z}$, and let $\{e_1, e_2\}$ be the ortonormal basis, for which $Kernel(A) = span(e_1)$. Then $Ae_1 = 0$ and similarly $Ae_2 = \pm\|Ae_2\| z$ for some sign. For any $x \in \mathbb{R}^2$

$$\begin{aligned} \left\|A^T x\right\|^2 &=< A^T x, e_1 >^2 + < A^T x, e_2 >^2 \\ &=< x, Ae_1 >^2 + < x, Ae_2 >^2 \\ &= \|Ae_2\|^2 < x, z >^2 . \end{aligned} \tag{85}$$

Where we used the Pithagorian-theorem and the properties of the inner product. Now we show that $\|Ae_2\|^2 = 1$. Since $\|A\| = 1$, we have $\|Ae_2\| \leq 1$. On the other hand for some $\alpha, \beta \in \mathbb{R}$, which satisfy $\alpha^2 + \beta^2 = 1$ (so $\beta^2 \leq 1$),

$$1 = \|A\| = \|A(\alpha e_1 + \beta e_2)\| = |\beta| \|Ae_2\|. \tag{86}$$

Therefore

$$\lim_{n\to\infty} \frac{\|A_n^T x\|^2}{\|A_n^T\|^2} = \frac{\|A^T x\|^2}{1} =< x, z >^2 . \tag{87}$$

□

The only thing left to show is that $Y_1^T \dots Y_n^T m$ converges to a Dirac-measure for some continuous distribution m.

Theorem 2.2. *Let $X_1, \dots$ be independent random matrices of order two, with $|\det X_i| = 1$ and with distribution μ. Suppose that there exists a continuous μ-invariant distribution ν. Then if the support of μ is not contained in a compact subgroup of $G1(d, \mathbb{R})$, there exists with probability one a direction $\overline{Z}_\omega$ such that $X_1 \dots X_n \nu$ converges to $\delta_{\overline{Z}_\omega}$.*

Moreover the distribution of $\overline{Z}$ is ν and it is the unique μ-invariant distribution on $P(\mathbb{R}^2)$.

Proof. Let $A_n = X_1 \dots X_n$. From Lemma 2.1 we know that for almost all ω there exists a probbility measure ν_ω for which $A_n \nu$ and $A_n M \nu$ convegres weakly to ν_ω. (In the second case of convergence we mean for μ-almost all M.)

Now fix an ω. By Lemma 2.4 for the limit point $A(\omega)$ of $(\|A_n(\omega)\|^{-1} A_n(\omega))$

$$A(\omega)\nu = A(\omega)M\nu = \nu_\omega \tag{88}$$

holds for μ-almost all M.

Note that $H = \{M \mid M \in G1(2, \mathbb{R}),\ M\nu = \nu,\ |\det(M)| = 1\}$ must be compact. Otherwise there would exist a sequence (M_k) in H with $\|M_k\| \to \infty$, such that $\|M_k\|^{-1} M_k$ congerges to some matrix C. So $C\nu = \nu$ and

$$|\det(C)| = \lim \left| \det \left(\frac{M_k}{\|M_k\|} \right) \right| = \lim \frac{1}{\|M_k\|^2} = 0. \tag{89}$$

Thus ν would be discrete.

Suppose that $A(\omega)$ is invertible, then by (88) $M\nu = \nu$ for almost all M, which means that $\mu(H) = 1$, and that contradicts our assumption. Therefore $A(\omega)$ has a rank one. Denote the direction of its range by $\overline{Z}(\omega)$. Now from (88) we have $\delta_{\overline{Z}(\omega)} = \nu_\omega$. This proves the weak convergence stated in the theorem.

By Lemma 2.1 for any Borel-function

$$\int f d\nu = \mathbb{E}\Big(\int f d\nu_\omega\Big) = \mathbb{E}\Big(\int f d\delta_{\overline{Z}(\omega)}\Big) = \mathbb{E}(f(\overline{Z})) \tag{90}$$

So ν is the distribution of $\overline{Z}$ and it is unique. □

Now we are able to prove the basics of the Furstenberg theorem. Which will show the importance of γ, namely to have a continuous invariant measure we need the γ to be positive.

Theorem 2.3. *Let $(Y_i)_{i\in\mathbb{N}^+}$ be a sequence of independent random matrices of order two with the distriution μ. Suppose that:*

(a) $|\det Y_i| = 1$ *a.s.*

(b) The support of μ is not contained in a compact subgroup of $G1(2,\mathbb{R})$.

(c) There exists a continuous μ^T-invariant distribution of $P(\mathbb{R}^2)$, where μ^T denotes the distribution of Y_i^T.

Then:

(i) For any $\overline{x}, \overline{y} \in P(\mathbb{R}^2)$, with probability one

$$\lim_{n\to\infty} \delta(Y_n \dots Y_1 \bullet \overline{x}, Y_n \dots Y_1 \bullet \overline{y}) = 0. \tag{91}$$

(ii) If $\mathbb{E}(\log^+\|Y_1\|) < \infty$, there exists a unique $\gamma \in \mathbb{R}^+$ such that if $x \neq 0$, then

$$\lim_{n\to\infty} \frac{1}{n} \log\|Y_n \dots Y_1 x\| = \lim_{n\to\infty} \frac{1}{n} \log\|Y_n \dots Y_1\| = \gamma. \tag{92}$$

a.s., and for the unique μ-invariant distribution ν on $P(\mathbb{R}^2)$

$$\gamma = \int\int \log \frac{\|Mx\|}{\|x\|} d\mu(M) d\nu(\overline{x}) > 0. \tag{93}$$

Proof. Set $S_n = Y_n \dots Y_1$. Let m be a continuous μ^T-invariant distribution. By the previous theorem there exists a measuralbe set Ω_0 with $\mathbb{P}(\Omega_0) = 1$ and a random direction $\overline{Z}$ such that for $\omega \in \Omega_0$

$$\lim_{n\to\infty} S_n^T(\omega) m = \delta_{\overline{Z}(\omega)} \tag{94}$$

Let Z be a unit vector with direction $\overline{Z}$. By Theorem 2.1 for all $\omega \in \Omega_0$

$$\lim_{n\to\infty} \|S_n(\omega)\| = \infty \tag{95}$$

and

$$\lim_{n\to\infty} \frac{\|S_n(\omega)x\|}{\|S_n(\omega)\|} = |< Z(\omega), x >|. \tag{96}$$

For any fixed $x \neq 0$, the event that $Z(\omega)$ is orthogonal to x has a measure zero, therefore

$$\lim_{n\to\infty} \|S_n(\omega)x\| = \infty. \tag{97}$$

and

$$\lim_{n\to\infty} \frac{\|S_n(\omega)x\|}{\|S_n(\omega)\|} > 0. \tag{98}$$

Both of which happen almost surely. Thus we have for almost all $x, y \in \mathbb{R}^2$

$$\lim_{n\to\infty} \delta(S_n(\omega) \bullet \overline{x}, S_n(\omega) \bullet \overline{y}) = \lim_{n\to\infty} \frac{\|x\|\|y\|\delta(\overline{x},\overline{y})}{\|S_n(\omega)x\|\|S_n(\omega)y\|} = 0 \tag{99}$$

$\mathbb{P}$-almost surely.

Suppose that $\mathbb{E}(\log\|Y_1\|) < \infty$. We define σ as

$$\sigma(Y, \overline{x}) = \log \frac{\|Yx\|}{\|x\|}, \tag{100}$$

for $Y \in G1(2, \mathbb{R})$ and $\overline{x} \in P(\mathbb{R}^2)$. Note that σ is an additive cocycle, and by (97)

$$\lim_{n\to\infty} \sigma(S_n(\omega), \overline{x}) = \infty \tag{101}$$

$\mathbb{P} \otimes \nu$-almost surely.

And by definition $\int\int \sigma^+(Y, \overline{x}) d\mu(Y) d\nu(\overline{x}) \leq \int \log\|Y\| d\mu(Y) < \infty$. Thus we can apply Lemma 2.2, Therefore $\sigma \in L^1(\mathbb{P} \otimes \nu)$, and its integral wrt μ and ν is positive. So define γ as

$$\gamma = \int\int \log \frac{\|Yx\|}{\|x\|} d\mu(Y) d\nu(\overline{x}). \tag{102}$$

Furthermore by Theorem 1.1, for some $\Phi : \Omega \times P(\mathbb{R}^2) \to \mathbb{R}$

$$\lim_{n\to\infty} \frac{1}{n} \sigma(S_n(\omega), \overline{x}) = \Phi(\omega, \overline{x}) \tag{103}$$

$\mathbb{P}\otimes\nu$-almost surely with $\int\int \Phi d\mathbb{P}d\nu = \gamma$. Fix a direction $\overline{x}$ such that (103) holds $\mathbb{P}$-a.s. Now take any $m < n$

$$\lim_{n\to\infty}\frac{1}{n}\log\left\|S_n\frac{x}{\|x\|}\right\| = \lim_{n\to\infty}\frac{1}{n}\left(\log\left\|Y_n\dots Y_{m+1}\frac{S_m x}{\|S_m x\|}\right\| + \log\left\|S_m\frac{x}{\|x\|}\right\|\right). \tag{104}$$

This means that the event $\Phi(\omega, x) = \rho$ for $\rho\in\mathbb{R}$, is in the tail σ-algebra of the independent random variables $Y_1,\dots$. Hence by Kolmogorov 0-1 it equals to a specific constant with probability one, which must be the upper Lyapunov-exponent.

□

2.3 Furstenberg's theorem

In this section we give sufficient conditions for Theorem 2.3.

Theorem 2.4 (First form). *Let μ be a probability measure on $G1(d,\mathbb{R})$ such that if G_μ is the smallest closed subgroup of $G1(d,\mathbb{R})$, which contains the support of μ, and the following hold:*

(i) For all M in G_μ, $|\det M| = 1$.

(ii) G_μ is not compact.

(iii) There does not exist a subset L of $\mathbb{R}^2$ which is a finite union of one dimensional subspaces for which $M(L) = L$ for any $M\in G_\mu$.

Then the conditions of the Theorem 2.3 are satisfied for matrices with distribution μ, so the conclusions hold.

Proof. This is the immediate consequence of the following lemma. □

Lemma 2.5. *Let μ be a probability measure on $G1(d,\mathbb{R})$, for which the condition (iii) of the previous theorem holds. Then any μ-invariant and μ^T-invariant distribution on $P(\mathbb{R}^2)$ is continuous.*

Proof. Let ν be a μ-invariant probability distribution on $P(\mathbb{R}^2)$. For any $\alpha\in(0,1]$, $\{\overline{x}\in P(\mathbb{R}^2)\mid \nu(\{\overline{x}\})>\alpha\}$ is finite. Therefore if ν is not continuous, there is a

$\beta \in (0,1]$, for which $\nu(\{\overline{x}\}) \leq \beta$ (for any $\overline{x} \in P(\mathbb{R}^2)$) and the set of directions with measure β (call it F) is non-empty.

Let $\overline{x}_0 \in F$, using $\mu *^{\bullet} \nu = \nu$

$$\beta = \nu(\{\overline{x}_0\}) = \int \int \mathbb{1}_{\{\overline{x}_0\}}(M \bullet \overline{x}) d\mu(M) d\nu(\overline{x}) = \int \nu(M^{-1} \bullet \overline{x}_0) d\mu(M) \leq \beta. \quad (105)$$

This means that for almost all M, $M^{-1} \bullet \overline{x}_0$ is in F. So if we define $L = \{0\} \cup \{x \in \mathbb{R}^2 - \{0\} \mid \overline{x} \in F\}$, then L is the finite union of one dimensional subspaces, for which $M(L) = L$ for any M in G_μ. And this contradicts our assumption, thus ν is continuous.

To prove the continuity for a μ^T-invariant distribution, we show that (iii) holds for it aswell. Let G be a subset of $M(2, \mathbb{R})$, and let $V_1, \ldots, V_r$ be a number of one dimensional subspaces, and $W_1, \ldots, W_r$ their ortogonals respectively. Then for any $M \in G$ and any pair of indices $M(V_i) = V_j$ is equivalent to $M^T(W_j) = W_i$. Therefore (iii) holds for $G_\mu \Leftrightarrow$ (iii) holds for G_{μ^T}. ☐

One can use the condition (iii) in a different way.

Theorem 2.5. *If μ satisfies the conditions (i) and (ii) of the Theorem 2.4, then the condition (iii) is equivalent to*

(iii)' For any direction $\overline{x}$, $K = \{M \bullet \overline{x} \mid M \in G_\mu\}$ has more than two elements.

Proof. $(iii) \Rightarrow (iii)'$: assume that for a specific $\overline{x}$ there is only two element in the set K defined above. Let $k_1, k_2 \in K$, then

$$M \bullet \{k_1, k_2, \overline{x}\} = \{k_1, k_2, \overline{x}\}. \quad (106)$$

But (iii) should hold. With a similar argument for $|K| = 1$ proves the implication.

$(iii)' \Rightarrow (iii)$: we prove the converse by contradiction as well. Let $\overline{x}_1, \ldots, \overline{x}_r$ be distinct directions with the property, that for any $M \in G1(2, \mathbb{R})$

$$M \bullet \{\overline{x}_1, \ldots, \overline{x}_r\} = \{\overline{x}_1, \ldots, \overline{x}_r\}. \quad (107)$$

Then every M corresponds to a permutation on $\{\overline{x}_1, \ldots, \overline{x}_r\}$, denote this relation by the function $\Phi(.)$ (so $\Phi(M)$ is a permutation). The kernel of Φ ($H = \{M \in$

$G_\mu \mid M \bullet \overline{x}_i = \overline{x}_i,\ for\ all\ i\}$) is a closed subgroup of G_μ and G_μ/H is finite, since it is isomorph to the permutation group on r elements. By (ii) G_μ is infinite therefore H must be infinite aswell.

Now suppose that $r \geq 3$. Consider the non-zero vectors x_1, x_2, x_3 with the directions $\overline{x_1}, \overline{x_2}, \overline{x_3}$. We can write

$$x_3 = \alpha x_1 + \beta x_2,\ for\ some\ \alpha, \beta \in \mathbb{R} - \{0\}, \tag{108}$$

and for each $M \in H$, for some $\lambda_i \neq 0$

$$Mx_i = \lambda_i x_i. \tag{109}$$

This yields

$$\alpha\lambda_3 x_1 + \beta\lambda_3 x_2 = \lambda_3 x_3 = Mx_3 = \alpha M x_1 + \beta M x_2 = \alpha\lambda_1 x_1 + \beta\lambda_2 x_2. \tag{110}$$

Which means $\lambda_1 = \lambda_3 = \lambda_2$ and $M = \lambda_3 \cdot id$. From (i) we know that $|\det M| = 1$. Therefore $H = \{id, -id\}$, but H should be infinite, so $r \leq 2$. □

And in the other form.

Theorem 2.6 (Second form). *Let $Y_1, \ldots$ be independent invertible matrices with the distribution μ. Suppose that there is no distribution m on $P(\mathbb{R}^2)$ such that $Mm = m$ for all M in the group generated by the support of μ. Then if $|\det Y_1| = 1$ a.s., then the assumptions of the Theorem 2.4 are satisfied.*

Proof. (i): This is between the new conditions.

(ii): Suppose that *supp*(G_μ) is compact, then there exist a Haar-measure on it (call it ψ). Now for any $\overline{x} \in P(\mathbb{R}^2)$ and $M \in G_\mu$

$$M(\psi *^\bullet \delta_{\overline{x}}) = (\delta_M * \psi) *^\bullet \delta_{\overline{x}} = (\psi *^\bullet \delta_{\overline{x}}), \tag{111}$$

and it contradicts the conditions.

(iii): Suppose this does not hold. Now we have a set of directions $\{\overline{x}_1, \ldots, \overline{x_r}\}$, for which for any $M \in G_\mu$

$$M \bullet \{\overline{x}_1, \ldots, \overline{x}_r\} = \{\overline{x}_1, \ldots, \overline{x}_r\} \tag{112}$$

This means that

$$M(\frac{1}{r}\sum_{i=1}^{r}\delta_{\overline{x}_i}) = \frac{1}{r}\sum_{i=1}^{r}\delta_{\overline{x}_i}, \tag{113}$$

and it contradicts the conditions. □

3 Example

Our example is about a subset of the upper triangular matrices. This particular process can also be understood as a random walk on the real line. To see this relation think of $P(\mathbb{R}^2)$ as the set $\{[x,1]^T \mid x \in \mathbb{R} \cup \{\infty\}\}$.

We will find the μ-invariant measures for different cases of this distribution, and then calculate the upper Lyapunov exponent associated with it.

Let $\{Y_i \mid i \geq 1\}$ be independent matrices with distribution μ, and of the form of

$$Y_i = \begin{bmatrix} a_i & b_i \\ 0 & 1 \end{bmatrix}, \; a_i \neq 0 \; a.s. \tag{114}$$

In any case $\underline{e} = [1,0]^T$ is an eigenvector of Y_i, and therefore $\delta_{\overline{e}}$ is a μ-invariant distribution.

Invariant measures

Case 1 Let us assume that $a_i x + b_i = x$ a.s. for some $x \in \mathbb{R}$. In this case $\underline{x}_1 = [x,1]^T$ is also an eigenvector of the matrices, and Y_i can be written in the form of $Q\Lambda_i Q^{-1}$, where $\Lambda_i = \begin{bmatrix} 1 & 0 \\ 0 & a_i \end{bmatrix}$ and $Q = \begin{bmatrix} x & 1 \\ 1 & 0 \end{bmatrix}$. And the product evolves as:

$$Y_n \cdot \ldots \cdot Y_1 = Q \cdot \Lambda_n \cdot \ldots \cdot \Lambda_1 \cdot Q^{-1} = Q \cdot \begin{bmatrix} 1 & 0 \\ 0 & \prod_{i=1}^{n} a_i \end{bmatrix} \cdot Q^{-1}. \tag{115}$$

Therefore $\delta_{\overline{x}_1}$ and $\delta_{\overline{e}}$ are μ-invariant, and for any other distribution ν on $P(\mathbb{R}^2)$:

$$\nu \to \delta_{\overline{x}_1}, \; if \; \prod_{i=1}^{n} a_i \to 0, \tag{116}$$

$$\nu \to \delta_{\overline{e}}, \; if \; \prod_{i=1}^{n} a_i \to \infty. \tag{117}$$

In this paper we will not discuss the cases where the product has a different limit point from the ones above.

We have started with a really strict condition, and now we look at the other possibilities.

Case 2 From now on we will assume, that for any Q, $Q^{-1}Y_iQ$ is not diagonal a.s. and that $\mathbb{E}(\log\|Y_i\|) < \infty$.

(a) If $\mathbb{E}(\log|a_i|) < 0$, then $u = \sum_{n=1}^{\infty} a_1 \dots a_{n-1}b_n$ is convergent with probality one. This is true, since by Cauchy's criterion we need:

$$\mathbb{P}(\limsup_{n\to\infty} \sqrt[n]{|a_1 \dots a_n|} < 1) = 1. \tag{118}$$

This is equivalent to

$$\mathbb{P}(\limsup_{n\to\infty} \frac{1}{n} \log|a_1 \dots a_n| < 0) = 1, \tag{119}$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \log|a_1 \dots a_n| = \limsup_{n\to\infty} \frac{1}{n}(\log|a_1| + \dots + \log|a_n|) = \mathbb{E}(\log|a_1|), \tag{120}$$

where the first equality holds by the properties of the absolute value, and the second by fact that $\{a_i \mid i \geq 1\}$ are i.i.d. and the Strong Law of Large Numbers. So the convergence holds.

With this information we can create a μ-invariant distribution. Let Y has the distribution μ, and let ν be the distribution of the direction of $\underline{x}_2 = [u, 1]^T$, then ν is μ-invariant, since for any f

$$\begin{aligned}\int f(\overline{x})d(\mu *^{\bullet} \nu) &= \int f(Y \bullet \overline{x})d\mu d\mathbb{P} \\ &= \int f\left(\overline{\begin{bmatrix} a \sum_{n=1}^{\infty} a_1 \dots a_{n-1}b_n + b \\ 1 \end{bmatrix}}\right) d\mu d\mathbb{P} \\ &= \int f\left(\overline{\begin{bmatrix} u \\ 1 \end{bmatrix}}\right) d\mathbb{P} = \int f(\overline{x})d\nu.\end{aligned} \tag{121}$$

In words $Y \cdot \underline{x}_2$ has the same distribution as $\underline{x}_2$, thus it is μ-invariant.

For the conclusion suppose that there is a μ-invariant distribution ν', which can not be written as $\alpha \cdot \delta_{\overline{e}} + (1 - \alpha) \cdot \nu$, where $\alpha \in [0, 1]$, we will use state space of ν' in the form of $[s, 1]^T$. Let f be a continuous function on $P(\mathbb{R}^2)$.

$$I := \int f(\overline{x})d\nu' = \int f(\overline{x})d\mu^n *^{\bullet} \nu' = \int f\left(\overline{\begin{bmatrix} a_1 \dots a_n s + \sum_{i=1}^{n} a_1 \dots a_{i-1} b_i \\ 1 \end{bmatrix}}\right) d\mu^n d\nu'. \tag{122}$$

And for any $\kappa \in \mathbb{R}^+$ and $s \in \mathbb{R}$, if n is large enough

$$\delta\left(\overline{\begin{bmatrix} \sum_{i=1}^{n} a_1 \dots a_{i-1} b_i \\ 1 \end{bmatrix}}, \overline{\begin{bmatrix} \sum_{i=1}^{n} a_1 \dots a_{i-1} b_i \\ 1 \end{bmatrix} + \begin{bmatrix} a_1 \dots a_n s \\ 0 \end{bmatrix}}\right) < \kappa \tag{123}$$

Therefore by using the uniform continuity of f: for any $\varepsilon \in \mathbb{R}^+$ there is an $n_0 \in \mathbb{N}$, such that if $n > n_0$, then

$$\left| I - \int f\left(\overline{\begin{bmatrix} \sum_{i=1}^{n} a_1 \dots a_{i-1} b_i \\ 1 \end{bmatrix}}\right) d\mu^n \right| < \varepsilon \tag{124}$$

And similarly for $J := \int f(\overline{x})d\nu$ for any $\varepsilon \in \mathbb{R}^+$ there is an $n_1 \in \mathbb{N}$, such that if $n > n_1$, then

$$\left| J - \int f\left(\overline{\begin{bmatrix} \sum_{i=1}^{n} a_1 \dots a_{i-1} b_i \\ 1 \end{bmatrix}}\right) d\mu^n \right| < \varepsilon \tag{125}$$

From these we conclude that $|I - J| < 2\varepsilon$ for any continuous f, which means that $\nu' = \nu$. And this contradicts our choice of measure.

To sum it up with these conditions all the μ-invariant measures have the form $\alpha \cdot \delta_{\overline{e}} + (1 - \alpha) \cdot \nu$, where $\alpha \in [0, 1]$.

(b) If $\mathbb{E}(\log|a_i|) \geq 0$, then $\delta_{\overline{e}}$ is the only μ-invariant distribution on $P(\mathbb{R}^2)$.

There can not be any non-continuous distribution, because starting from any vector $[x, 1]^T$:

$$(Y_n \cdot \dots \cdot Y_1) \bullet \overline{\begin{bmatrix} x \\ 1 \end{bmatrix}} = \overline{\begin{bmatrix} a_1 \dots a_n x + \sum_{i=2}^{n} a_2 \dots a_{i-1} b_i \\ 1 \end{bmatrix}} \to \overline{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}, \tag{126}$$

where the convergence comes from the fact, that the sum is divergent.

And there can not be a continuous distribution (prooving by contradiction): Let ν' be this distribution and $X_i = \frac{1}{\sqrt{a_i}} Y_i$.

* $\det(X_i) = 1$
* $\|.\|$ is a continuous function and $\|X_n \cdot \ldots \cdot X_1\| \to \infty$, therefore by Weierstrass $supp(\mu)$ is not contained in a compact subgroup.
* We just assumed that there is a continuous measure.

So by the Theorem 2.3 for any $\overline{x}, \overline{y} \in P(\mathbb{R}^2)$:

$$\lim_{n\to\infty} \delta(X_n \cdot \ldots \cdot X_1 \bullet \overline{x}, X_n \cdot \ldots \cdot X_1 \bullet \overline{y}) = 0. \tag{127}$$

And since $\delta_{\overline{e}}$ is μ-invariant, for any $\overline{x} \in P(\mathbb{R}^2)$ and $\varepsilon_0 \in \mathbb{R}^+$ there is an $n_0 \in \mathbb{N}$, such that if $m > n_0$, then

$$\delta(X_m \cdot \ldots \cdot X_1 \bullet \overline{x}, X_m \cdot \ldots \cdot X_1 \bullet \overline{e}) < \varepsilon_0. \tag{128}$$

So for any continuous function f and $\varepsilon \in \mathbb{R}^+$ there will be an $n_0 \in \mathbb{N}$, such that if $m > n_0$, then

$$\left| \int f(\overline{x}) d\delta_{\overline{e}} - \int f(\overline{x}) d\mu^m *^\bullet \nu'(\overline{x}) \right| < \varepsilon. \tag{129}$$

And since

$$\int f(\overline{x}) d\nu'(\overline{x}) = \int f(\overline{x}) d\mu^m *^\bullet \nu'(\overline{x}), \tag{130}$$

$\nu' = \delta_{\overline{e}}$, which can not be true, since ν' is continuous. So the only μ-invariant distribution is $\delta_{\overline{e}}$.

Upper Lyapunov exponent

In the first case $\mathbb{E}(\log\|Y_1\|) < \infty$ is true ($\|Y_1\| = \max(1, a_1)$), and in the second case we have assumed it. So by Theorem 1.2 it is enough to calculate the $\lim \frac{1}{n} \log\|Y_n \ldots Y_1\|$, we will denote this value by γ.

First of all let $\gamma_a := \lim\limits_{n\to\infty} \frac{1}{n} \log|a_n \ldots a_1| = \mathbb{E}(\log|a_1|)$, the equality holds by the properties of absolute value and logarithm, and the Strong Law of Large Numbers. Furthermore let

$$Y_n \cdot \ldots \cdot Y_1 = \begin{bmatrix} R_{00}^n & R_{01}^n \\ 0 & 1 \end{bmatrix}, \tag{131}$$

and similarly

$$Y_{2n} \cdot \ldots \cdot Y_{n+1} = \begin{bmatrix} L_{00}^n & L_{01}^n \\ 0 & 1 \end{bmatrix} \tag{132}$$

Using γ_a there is a lower bound for the Lyapunov exponent. Let e_1, e_2 be $[1,0]^T$ and $[0,1]^T$ respectively. Then by the definition of the supremum norm $\gamma_a = \lim \frac{1}{n} \log\|Y_n \ldots Y_1 e_1\| \leq \gamma$. Similarly $0 = \lim \frac{1}{n} \log\left\|Y_1^T \ldots Y_n^T e_2\right\| \leq \gamma$. In one expression $\gamma \geq \max(\mathbb{E}(\log|a_1|), 0)$.

Now we will search for some upper bounds. By the definition of the limit we know that for any $\varepsilon \in \mathbb{R}^+$ there exist an $n_0 \in \mathbb{N}$ such that for any $n > n_0$

$$\left|\frac{1}{n} \log|a_n \ldots a_1| - \gamma_a\right| < \varepsilon. \tag{133}$$

From this we obtain

$$|a_{2n} \ldots a_{n+1}| = |L_{00}^n| \sim |R_{00}^n| = |a_n \ldots a_1| < e^{n(\varepsilon+\gamma_a)}. \tag{134}$$

And for the other non-constant cells we choose a large enough n, for which

$$\frac{1}{n} \log|R_{01}^n| < \frac{1}{n} \log\|Y_n \ldots Y_1 e_2\| \leq \frac{1}{n} \log\|Y_n \ldots Y_1\| < \varepsilon + \gamma \tag{135}$$

is true. And we obtain the upper bound

$$|L_{01}^n| \sim |R_{01}^n| < e^{n(\varepsilon+\gamma)}. \tag{136}$$

$$\begin{aligned}
\|Y_{2n} \ldots Y_1\| = \|L^n R^n\| &= \left\| \begin{matrix} L_{00}^n R_{00}^n & L_{00}^n R_{01}^n + L_{01}^n \\ 0 & 1 \end{matrix} \right\| \\
&= \left\| \begin{bmatrix} L_{00}^n R_{00}^n & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & L_{00}^n R_{01}^n + L_{01}^n \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\| \\
&\leq e^{2n(\varepsilon+\gamma_a)} + e^{n(\varepsilon+\gamma+\gamma_a)} + e^{n(\varepsilon+\gamma)} + 1 \\
&\leq 4e^{n(\varepsilon+\gamma+\max(0,\gamma_a))}
\end{aligned} \tag{137}$$

This yields for all $\varepsilon \in \mathbb{R}^+$

$$\gamma \leq \frac{\varepsilon + \gamma + \max(0, \gamma_a)}{2} \tag{138}$$

Thus we have: $\gamma = \max(0, \gamma_a)$.

References

[1] Philippe Bougerol and Jean Lacroix. *Products of random matrices with applications to Schrödinger operators*, volume 8 of *Progress in Probability and Statistics*. Birkhäuser Boston, Inc., Boston, MA, 1985.

[2] Michael Keane and Karl Petersen. Easy and nearly simultaneous proofs of the ergodic theorem and maximal ergodic theorem. In *Dynamics & stochastics*, volume 48 of *IMS Lecture Notes Monogr. Ser.*, pages 248–251. Inst. Math. Statist., Beachwood, OH, 2006.