Universidad del Rosario

Escuela de Ciencias e Ingeniería
Programa de Matemáticas Aplicadas y Ciencias de la Computación

Lyapunov Exponents to Predict the Behavior of the Product of Random Matrices

Juliana Bermúdez Guzmán

Advisor: Mauro Artigiani
Co-advisor: Cristian Martínez

Bogotá, Colombia
May 2025

Contents

Abstract

In this thesis, we investigate the asymptotic behavior of products of random matrices through Lyapunov exponents. Our theoretical framework is grounded in Kingman's Subadditive Ergodic Theorem, from which we derive the Furstenberg-Kesten Theorem and Oseledets' Theorem in two dimensions. These results provide the tools to quantify exponential growth rates and directional behavior in random matrix products. To visualize our theoretical conclusions, we present a series of simulations that illustrate the emergence of Lyapunov exponents and their predictive power in practical settings.

1 Introduction

The behavior of random matrix products has attracted much attention due to its relevance in many fields, such as statistical physics, dynamical systems, and even economics. In particular, understanding the asymptotic growth rate of such products allows us to study the stability of systems affected by noise or uncertainty. This thesis is dedicated to the study of the Lyapunov exponents, which quantify such asymptotic behavior.

The goal is to analyze the conditions under which the products of random matrices exhibit regular long-term behavior. We will develop the necessary theoretical tools, such as Kingman's Subadditive Ergodic Theorem and its consequences, like the Furstenberg-Kesten and Oseledets theorems, and support our results with numerical simulations. These tools enable us to describe both the growth rate of the norm of matrix products and the directions in which vectors expand or contract under repeated application of random matrices.

To gain better intuition for the asymptotic behavior of random matrix products, consider a simple example where matrices are drawn randomly from the generators of $\mathrm{SL}(2, \mathbb{Z})$:

$$\mathrm{SL}(2, \mathbb{Z}) = \left\langle \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \right\rangle,$$

each chosen independently with probability $\frac{1}{2}$. Although these matrices are individually unipotent and seemingly simple, their random products exhibit exponential growth in norm due to accumulated shearing effects. This growth is quantified by a positive Lyapunov exponent, illustrating the kind of long-term behavior we aim to characterize and predict throughout this thesis.

This work is structured as follows: we begin with a review of measure theory, convergence theorems, and foundational probability. We then introduce concepts from dynamical systems necessary to define linear cocycles, which model the behavior of sequences of matrices. The core theoretical results are presented next, followed by their consequences. Finally, simulations illustrate how these theorems manifest.

2 Measure Theory

Through this subsection, we will present certain concepts of measure theory which are vital to understand the main proof as well as some theorems used in there. Standard references are [1, 2, 3, 4, 5].

The fundamental concept is that we want to work with "well-behaved" spaces. By that, we mean that we intend to find spaces that can be measurable. Without an underline measure, it is impossible to find averages such as the one proposed by Birkhoff's Ergodic Theorem, which states that a time averages converges to a space average. Hence, we will define σ-algebra to later on define the measurable space.

Definition 2.1. A family $\mathbf{X}$ of subsets of a set X is said to be a σ-algebra or σ-field if:

1. $\emptyset, X$ belong to $\mathbf{X}$.
2. If A belongs to $\mathbf{X}$, then $A^c = X - A$ belongs to $\mathbf{X}$.
3. If (A_n) is a sequence of sets in $\mathbf{X}$, then $\cup_{n=1}^{\infty} A_n$ belongs to $\mathbf{X}$.

Now, we can say that an ordered pair $(X, \mathbf{X})$, in which X is a space and $\mathbf{X}$ a σ-algebra over X, is a **measurable space**. We will see an example and counterexample of a σ-algebra.

Example 2.2. Let $X = \mathbb{N}$ and $\mathbf{X} = \{\emptyset, \{1, 3, 5, \dots\}, \{2, 4, 6, \dots\}, X\}$. $\mathbf{X}$ is a σ-algebra because:

1. $\emptyset, X$ belong to $\mathbf{X}$.
2. The even numbers belong to $\mathbf{X}$, and their complement, odd numbers, as well.
3. Closed over countable unions because: $\{1,3,5,\dots\} \cup \{2,4,6,\dots\} = \mathbb{N}$, $\{1,3,5,\dots\} \cup \emptyset = \{1,3,5,\dots\}$, $\mathbb{N} \cup \{1,3,5,\dots\} = \mathbb{N}$ and $\mathbb{N} \cup \{2,4,6,\dots\} = \mathbb{N}$.

Counterexample 2.3. Let $X = \mathbb{N}$ and $\mathbf{X} = \{\emptyset, \{1\}, X\}$. It is not a σ-algebra because $\{1\}$ belongs to $\mathbf{X}$, but the complement $\{2,3,4,\dots\}$ does not belong to it.

We are interested in functions, so we are going to define what does being **X-measurable** means.

Definition 2.4. A function $f : X \to \mathbb{R}$ is said to be **X-measurable** if, for every real number α, the set

$$\{x \in X : f(x) > \alpha\}$$

belongs to $\mathbf{X}$.

To understand better measurable functions, we will observe one that will be particularly representative further on, the **characteristic function**.

Example 2.5. If $E \in \mathbf{X}$, then the **characteristic function**

$$\chi_E(x) = \begin{cases} 1, & \text{if } x \in E \\ 0, & \text{if } x \notin E \end{cases}$$

is measurable. It is true because $\{x \in \mathbf{X} : \chi_E(x) > \alpha\}$ is either X, E or $\emptyset$ which all belong to $\mathbf{X}$.

It is important to note that the product, the addition, and the absolute value of measurable functions are also measurable. On that note, we will define a way of decomposing a function into its positive and negative parts such that both being measurable implies the original function is too.

Definition 2.6. Let $f : X \to \mathbb{R}$ where f^- and f^+ are *nonnegative* functions on X defined as

$$f^+(x) = \sup\{f(x), 0\} \quad f^-(x) = \sup\{-f(x), 0\}$$

such that f^+ is known as the **positive part** and f^- is the **negative part**. Hence, it is clear that

$$f = f^+ - f^- \quad \text{and} \quad |f| = f^+ + f^-.$$

Moreover, these nonnegative functions can be approximated by an increasing sequence of functions called **simple** which take finite values. These functions are relevant because defining an integral over them is easier to compute.

Definition 2.7. A real-valued function is **simple** if it has only a finite number of values. More specifically, a simple measurable function φ can be represented as

$$\varphi = \sum_{j=1}^{n} a_j \chi_{E_j}$$

where $a_j \in \mathbb{R}$ and χ_{E_j} is the characteristic function of a set E_j in $\mathbf{X}$. The unique **standard representation** of φ guarantees that a_j are distinct and E_j are disjoint nonempty subsets of X such that $X = \cup_{j=1}^{n} E_j$.

These simple functions serve a specific role when defining integrals. We must keep in mind that $\overline{\mathbb{R}}$ is the representation of the extended reals, which is equivalent to $\mathbb{R} \cup \{-\infty, \infty\}$. The collection of all nonnegative $\mathbf{X}$-measurable functions from X to $\overline{\mathbb{R}}$ is denoted as $M^+(X, \mathbf{X})$. Now, we proceed to define the integral with respect to $M^+(X, \mathbf{X})$.

Definition 2.8. If f belongs to $M^+(X, \mathbf{X})$, we define the **integral of f with respect to μ**, with μ being a measure over the space, to be the extended real number

$$\int f \mathrm{d}\mu = \sup \int \varphi \,\mathrm{d}\mu$$

where the supremum is extended over all simple functions φ such in $M^+(X, \mathbf{X})$ such that $0 \leq \varphi(x) \leq f(x)$ for all $x \in X$.

Now that these concepts have been explained, it is possible to understand the monotone convergence theorem and Fatou's lemma. These two results will be used throughout the proofs of the main results, and they will be introduced in the next section. However, before we do that, we must define L^1 spaces.

Definition 2.9. The collection $L^1(\mu) = L^1(X, \mathbf{X}, \mu)$ of **integrable functions** consists of all real-valued $\mathbf{X}-$measurable functions f defined on X, such that both positive and negative parts of f have finite integrals with respect to μ. In this case, we define the **Lebesgue integral of f with respect to μ** to be

$$\int f \mathrm{d}\mu = \int f^+ \mathrm{d}\mu - \int f^- \mathrm{d}\mu.$$

If E belongs to $\mathbf{X}$, we define

$$\int_E f \mathrm{d}\mu = \int_E f^+ \mathrm{d}\mu - \int_E f^- \mathrm{d}\mu.$$

Now, we will work on a result known as the *property of absolute integrability* of the Lebesgue integral. We know that the absolute value of a proper Riemann integrable function is Riemann integrable, but that is not the case for certain functions with an improper Riemann integral, such as $f(x) = x^{-1} \sin x$ on the interval $1 \leq x \leq +\infty$. Nonetheless, when working with measurable functions, their absolute value is always integrable, as we are about to see.

Theorem 2.10. *A measurable function f belongs to $L^1(\mu)$ if and only if $|f|$ belongs to $L^1(\mu)$. In this case*

$$\left| \int f \mathrm{d}\mu \right| \leq \int |f| \mathrm{d}\mu.$$

Proof. We know that $f \in L^1(\mu)$ if and only if f^+ and f^- are in M^+ and have finite integrals. Aside from that, we also know that f is integrable if and only if $|f|$ is integrable. By definition, we know that $|f|^+ = |f| = f^+ + f^-$ and $|f|^- = 0$. Hence,

$$\left| \int f \,\mathrm{d}\mu \right| = \left| \int f^+ \,\mathrm{d}\mu - \int f^- \,\mathrm{d}\mu \right| \leq \int f^+ \,\mathrm{d}\mu + \int f^- \,\mathrm{d}\mu = \int |f| \,\mathrm{d}\mu.$$

$\square$

Lemma 2.11. *i. If f and g belong to $M^+(X, \mathbf{X})$ and $f \leq g$, then*

$$\int f \mathrm{d}\mu \leq \int g \mathrm{d}\mu.$$

ii. If f belongs to $M^+(X, \mathbf{X})$, if E, F belong to $\mathbf{X}$, and if $E \subseteq F$, then

$$\int_E f \mathrm{d}\mu \leq \int_F f \mathrm{d}\mu.$$

Proof. i. If φ is a simple function in M^+ such that $0 \leq \varphi \leq f$, then $0 \leq \varphi \leq g$ and the inequality holds.

ii. We know that $f\chi_E \leq f\chi_F$ so

$$\int_E f \mathrm{d}\mu = \int f\chi_E \mathrm{d}\mu \leq \int f\chi_F \mathrm{d}\mu = \int_F f \mathrm{d}\mu$$

and the inequality holds from the previous part as $f\chi_E, f\chi_F \in M^+(X, \mathbf{X})$.

□

Now, we will see how the inequality between two functions is preserved with the integrals. The proof follows directly from Lemma 2.11 and Theorem 2.10.

Corollary 2.12. *If f is measurable, g is integrable, and $|f| \leq |g|$, then f is integrable, and*

$$\int |f| \mathrm{d}\mu \leq \int |g| \mathrm{d}\mu.$$

Now, we shall see that the integral is a linear operator on the space $L^1(\mu)$.

Theorem 2.13. *A constant multiple αf and a sum $f+g$ of functions in $L^1(\mu)$ belongs to $L^1(\mu)$ and*

$$\int \alpha f \mathrm{d}\mu = \alpha \int f \mathrm{d}\mu, \quad \int (f+g) \mathrm{d}\mu = \int f \mathrm{d}\mu + \int g \mathrm{d}\mu.$$

We will leave the proof to the reader.

3 Convergence Theorems

We begin by presenting two foundational convergence theorems that will be instrumental in establishing our main results. The first is Fatou's Lemma, and the second is the Monotone Convergence Theorem, which plays a crucial role in defining integrals as limits of integrals of increasing sequences of functions. Before stating these theorems, we introduce a few preliminary lemmas.

Lemma 3.1. *Let μ be a measure defined on a σ-algebra $\mathbf{X}$. If (E_n) is an increasing sequence in $\mathbf{X}$, then*

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} \mu(E_n).$$

Proof. First, we will consider when $\mu(E_n) = +\infty$ for some n. In this case, both sides are $+\infty$, so the equality holds. Now, we will assume that $\mu(E_n) < +\infty$ for all n.

Let $A_1 = E_1$ and $A_n = E_n - E_{n-1}$ for $n > 1$. Then (A_n) is a disjoint sequence of sets in $\mathbf{X}$. Consequently,

$$E_n = \bigcup_{j=1}^{n} A_n \quad \text{and} \quad \bigcup_{n=1}^{\infty} E_n = \bigcup_{n=1}^{\infty} A_n.$$

Since μ is countably additive on disjoint sets,

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(A_n) = \lim_{m\to\infty} \sum_{n=1}^{m} \mu(A_n).$$

Given that $\mu(A_n) = \mu(E_n) - \mu(E_{n-1})$ for $n > 1$, the finite series on the right side is telescoping and

$$\sum_{n=1}^{m} \mu(A_n) = \mu(E_m).$$

□

Lemma 3.2. *If φ is a simple function in $M^+(X, \mathbf{X})$ and λ is defined for E in $\mathbf{X}$ by*

$$\lambda(E) = \int \varphi \chi_E \mathrm{d}\mu,$$

then λ is a measure on $\mathbf{X}$.

Proof. Note that

$$\varphi\chi_E = \sum_{j=1}^{n} a_j \chi_{E_j \cap E}, \quad \text{where} \quad \varphi = \sum_{j=1}^{n} a_j \chi_{E_j}.$$

Hence, we know that

$$\lambda(E) = \int \varphi\chi_E \mathrm{d}\mu = \sum_{j=1}^{n} a_j \int \chi_{E_j \cap E} \mathrm{d}\mu = \sum_{j=1}^{n} a_j \mu(E_j \cap E).$$

We have written λ as a linear combination of measures because $E \to \mu(E_j \cap E)$ is a measure. Consequently, λ is a measure of $\mathbf{X}$. □

Theorem 3.3 (Monotone Convergence Theorem)**.** *If (f_n) is a monotonically increasing sequence of functions in $M^+(X, \mathbf{X})$ that converges to f, then*

$$\int f \,\mathrm{d}\mu = \int \lim_{n\to\infty} f_n \,\mathrm{d}\mu = \lim_{n\to\infty} \int f_n \,\mathrm{d}\mu.$$

Proof. We know that (f_n) is a sequence in $M^+(X, \mathbf{X})$ which converges to f on X, so $f \in M^+(X, \mathbf{X})$. Hence, f is measurable. Also, as $f_n \leq f_{n+1} \leq f$, we know that

$$\int f_n \mathrm{d}\mu \leq \int f_{n+1} \mathrm{d}\mu \leq \int f \mathrm{d}\mu \quad \text{for all} \quad n \in \mathbb{N}.$$

Therefore,

$$\lim_{n\to\infty} \int f_n \mathrm{d}\mu \leq \int f \mathrm{d}\mu.$$

Now we want to get the other inequality. Let α be a real number such that $0 < \alpha < 1$. In addition to that, consider a simple measurable function φ such that $0 \leq \varphi \leq f$. Let

$$A_n = \{x \in X : f_n(x) \geq \alpha\varphi(x)\}$$

such that $A_n \in \mathbf{X}, A_n \subseteq A_{n+1}$, and $X = \bigcup A_n$.

Lemma 2.11 implies that

$$\int_{A_n} \alpha\varphi \mathrm{d}\mu \leq \int_{A_n} f_n \mathrm{d}\mu \leq \int f_n \mathrm{d}\mu. \tag{1}$$

Given that the sequence (A_n) increases monotonically and its union is X, from Lemma 3.2 we have that

$$\int \varphi \mathrm{d}\mu = \lim_{n\to\infty} \int_{A_n} \varphi \mathrm{d}\mu.$$

If we take the limit with respect to n of the equation 1, we get

$$\alpha \int \varphi \mathrm{d}\mu = \lim_{n\to\infty} \int_{A_n} \alpha\varphi \mathrm{d}\mu \leq \lim_{n\to\infty} \int f_n \mathrm{d}\mu.$$

Given that this holds for all $0 < \alpha < 1$, we can have that

$$\int \varphi \mathrm{d}\mu \leq \lim_{n\to\infty} \int f_n \mathrm{d}\mu.$$

Aside from this, as φ is a simple function in M^+ such that $0 \leq \varphi \leq f$, we conclude that

$$\int f \mathrm{d}\mu = \sup_{\varphi} \int \varphi \mathrm{d}\mu \leq \lim_{n\to\infty} \int f_n \mathrm{d}\mu.$$

Combining this inequality with the opposite inequality we got before, we have that

$$\int f \mathrm{d}\mu = \lim_{n\to\infty} \int f_n \mu.$$

□

We have successfully proven the equality of the integral of a function with the limit of a monotonically increasing sequence that converges to that function. Nonetheless, not all functions have this behavior, so we will state a consequence of Theorem 3.3 that helps us work with functions that are not monotone.

Lemma 3.4 (Fatou's Lemma)**.** *If* (f_n) *belongs to* $M^+(X, \mathbf{X})$*, then*

$$\int (\liminf_{n\to\infty} f_n) \mathrm{d}\mu \leq \liminf_{n\to\infty} \int f_n \mathrm{d}\mu.$$

Proof. We will define $g_m = \inf\{f_m, f_{m+1}, \dots\}$ such that $g_m \leq f_n$ whenever $m \leq n$. Consequently,

$$\int g_m \mathrm{d}\mu \leq \int f_n \mathrm{d}\mu, \quad m \leq n$$

and

$$\int g_m \mathrm{d}\mu \leq \liminf_{n\to\infty} \int f_n \mathrm{d}\mu.$$

Given that the sequence (g_m) is increasing and $(g_m) \to \liminf_{n\to\infty} f_n$, Theorem 3.3 implies that

$$\int (\liminf_{n\to\infty} f_n) \,\mathrm{d}\mu = \lim_{n\to\infty} \int g_m \,\mathrm{d}\mu \tag{2}$$

$$\leq \liminf_{n\to\infty} \int f_n \,\mathrm{d}\mu \tag{3}$$

Hence, we have the desired inequality whenever $f_n \geq 0$. □

The final result of this section is Lebesgue Dominated Convergence Theorem.
Theorem 3.5 (Dominated Convergence Theorem)**.** *Let (f_n) be a sequence of complex measurable functions on X such that*

$$f(x) = \lim_{n\to\infty} f_n(x)$$

exists for every $x \in X$. If there is a function $g \in L^1(\mu)$ such that

$$|f_n(x)| \leq g(x) \quad (n = 1, 2, \dots; x \in X)$$

then $f \in L^1(\mu)$,

$$\lim_{n\to\infty} \int_X |f_n - f| \, \mathrm{d}\mu = 0, \tag{4}$$

and

$$\lim_{n\to\infty} \int_X f_n \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu \tag{5}$$

Proof. Since $|f| \leq g$ and f is measurable, $f \in L^1(\mu)$. Now, we have that $|f_n - f| \leq 2g$, so Fatou's lemma applies to the functions $2g - |f_n - f|$ and yields

$$\begin{aligned}
\int_X 2g \, \mathrm{d}\mu &\leq \liminf_{n\to\infty} \int_X (2g - |f_n - f|) \, \mathrm{d}\mu \\
&= \int_X 2g \, \mathrm{d}\mu + \liminf_{n\to\infty} \left(- \int_X |f_n - f| \, \mathrm{d}\mu \right) \\
&= \int_X 2g \, \mathrm{d}\mu - \limsup_{n\to\infty} \int_X |f_n - f| \, \mathrm{d}\mu.
\end{aligned} \tag{6}$$

Given that $\int 2g \, \mathrm{d}\mu$ is finite, we may subtract and obtain that

$$\limsup_{n\to\infty} \int_X |f_n - f| \, \mathrm{d}\mu \leq 0. \tag{7}$$

If a sequence of nonnegative real numbers fails to converge to 0, then its upper limit is positive. Consequently, 3 implies 4; and applying Theorem 2.13 to $f_n - f$, 4 implies 5. □

4 Probability

Given that our main results rely heavily on probabilistic arguments, we must first define some fundamental concepts in probability theory. These concepts provide the basis for understanding random matrices, stochastic processes, and their convergence behaviors. The definitions and examples below come from the following sources [6, 7, 8, 9].
Definition 4.1. The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is a **probability space** where:

- Ω is a nonempty set of all possible outcomes known as *sample space*;
- $\mathcal{F}$ is a σ-algebra over Ω with the collection of *measurable events*;
- $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a non-negative measure on the measurable space $(\Omega, \mathcal{F})$ satisfying $\mathbb{P}(\Omega) = 1$ (the total mass is 1).

Now that we have defined the space, we can define the building blocks of our interest: random variables.

Definition 4.2 (Random variable ([1])). A function $X : \Omega \to \mathbb{R}$ is called a **random variable** if it is $\mathcal{F}$-measurable, which means that for every $\alpha \in \mathbb{R}$,

$$\{\omega \in \Omega : X(\omega) \leq \alpha\} \in \mathcal{F}.$$

If this variable took on either a finite or a countable number of possible values, it is considered *discrete* and a probability mass function models it. If it takes on a continuum of possible values it is *continous*, and its behavior is defined by a cumulative distribution function.

Now, we will define a property of these random variables that must be checked in order to apply almost all probabilistic theorems: independence [10]. However, first we must know that *Borel sets* are sets that can be constructed from open or closed sets by repeatedly taking countable unions and intersections.

Definition 4.3 (Independence([11])). Let $X_1, X_2, \ldots, X_n$ be random variables. We say they are **independent** if, for any selection of Borel sets $B_1, B_2, \ldots, B_n$,

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i).$$

As a matter of fact, this concept can be extended to families of σ-algebras which become relevant when it comes to constructing product probability spaces (such as in sequences of random matrices). Hence, we will also define independent σ-algebras as it was done by Kolmogorov and presented in [6].

Definition 4.4. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and, for each i from an index set $\mathcal{I}$, we have that $\mathcal{F}_i$ is a sub-σ-algebra of $\mathcal{F}$. We say that the σ-algebras $\mathcal{F}_i$, $i \in \mathcal{F}_i$, are **mutually $\mathbb{P}$-independent** or **$\mathbb{P}$-independent** if for every finite subset $\{i_1, \ldots, i_n\}$ of distinct elements of $\mathcal{I}$ and every choice of $A_{i_m} \in \mathcal{F}_{i_m}, 1 \leq m \leq n$,

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_n}) = \mathbb{P}(A_{i_1}) \ldots \mathbb{P}(A_{i_n})$$

As we can see, this definition is a generalization of the classical notion of independent pairs of sets encountered in non-measure theoretic presentations. Since our main goal is the product of random matrices, we have to define the product probability space.

Definition 4.5 (Product probability space ([12])). Let $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ be a probability space for each $i \in \mathbb{N}$. The **product probability space** is defined as follows:

- The *sample space* is the cartesian product, so it can be defined as

$$\Omega = \prod_{i=1}^{\infty} \Omega_i.$$

- The σ-algebra is the product of σ-algebras defined as follows:

$$\mathcal{F} = \bigotimes_{i=1}^{\infty} \mathcal{F}_i,$$

  is the smallest σ-algebra such that all projection maps $\pi_j : \Omega \to \Omega_j$ are measurable for every $j \in \mathbb{N}$.

- The **product measure** $\mathbb{P}$ is the probability measure such that for all cylinder sets, that is, sets of the form

$$C = A_{i_1} \times A_{i_2} \times \cdots \times A_{i_n} \times \prod_{i \neq i_n} \Omega_i, \text{ with } A_i \in \mathcal{F}_i,$$

we have that

$$\mathbb{P}(C) = \prod_{i=1}^{n} \mathbb{P}_i(A_i).$$

We must note that $\mathcal{F}_1 \times \mathcal{F}_2 \neq \mathcal{F}_1 \otimes \mathcal{F}_2$, because the first one may not be closed under complements or unions, so we need the second operation to guarantee that we have a σ-algebra.

Now that we know the construction of product probability spaces and independence, we can see a special notation for independent and identically distributed (i.i.d) spaces. In this case, we write $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega_0^{\mathbb{N}}, \mathcal{F}_0^{\otimes \mathbb{N}}, \mathbb{P}_0^{\otimes \mathbb{N}})$, and the random variables form an i.i.d. sequence. Now that we are familiarized with these basic concepts, we are going to define something extremely important for most ergodic theory constructions: a stochastic process.

Definition 4.6. A **stochastic process** is a sequence $\{X_t : t \in T\}$, with T an index set, of random variables defined on some fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Hence, X_t represents the value of a random quantity at the time t.

Example 4.7. We will keep working with the coin flip. In this case, we know that $\Omega = \{\text{H,T}\}$; $\mathcal{F}$ is going to be a collection of subsets of Ω, which given our finite sample space we will assume is the power set of Ω; $\mathbb{P}$ will be $\mathbb{P}(\text{H}) = \mathbb{P}(\text{T}) = 0.5$.

This way, each coin flip is a random variable (X_t) and the sequence of flips is indexed over T forming a stochastic process.

Nonetheless, what we are truly interested in are stationary processes. In this case, we are going to consider *time shifts*, defined as follows. If there's a Δ such that $t \in T$, then $t + \Delta \in T$.

Definition 4.8. A **stationary process** is a stochastic process where the joint distributions are invariant under fixed time shifts (Δ), meaning that

$$\mathbb{P}((X_{t_1}, X_{t_2}, \ldots, X_{t_n}) \in A) = \mathbb{P}((X_{t_1+\Delta}, X_{t_2+\Delta}, \ldots, X_{t_n+\Delta}) \in A)$$

for any set $A \in \mathbb{R}^n$.

Example 4.9. Consider a discrete-time random process that is $\{X_n : n \in \mathbb{Z}\}$, in which all X_n's are i.i.d. and their cumulative density function (CDF) is $F_{X_n}(x) = F(x)$.

Now, we have that

$$\begin{aligned} F_{X_{n_1} X_{n_2} \ldots X_{n_r}}(x_1, x_2, \ldots, x_r) &= F_{X_{n_1}}(x_1) F_{X_{n_2}}(x_2) \ldots F_{X_{n_r}}(x_r) \quad \text{(due to independence)} \\ &= F(x_1) F(x_2) \ldots F(x_r) \quad \text{(applying the CDF)} \end{aligned}$$

Aside from this, we also have that

$$
\begin{aligned}
F_{X_{n_1+\Delta}X_{n_2+\Delta}\ldots X_{n_r+\Delta}}(x_1, x_2, \ldots, x_r) &= F_{X_{n_1+\Delta}}(x_1)F_{X_{n_2+\Delta}}(x_2)\ldots F_{X_{n_r+\Delta}}(x_r) \quad \text{(due to independence)} \\
&= F(x_1)F(x_2)\ldots F(x_r) \quad \text{(applying the CDF)}
\end{aligned}
$$

We can see that a time shift does not affect the outcome, so it is stationary.

This concept is extremely important because the construction of cocycles over a shift map (the case of random matrix products) relies on the assumption that they are stationary processes. Turning our focus to fundamental probabilistic results, we will define two lemmas that are key for establishing almost sure convergence in probability: the Borel-Cantelli Lemma [6, 13].

Lemma 4.10 (Borel-Cantelli)**.** *Let* $\{E_n : n \in \mathbb{Z}^+\} \subseteq \mathcal{F}$ *be given. Then*

$$
\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty \implies \mathbb{P}(\limsup_{n\to\infty} E_n) = 0.
$$

Moreover, if the E_n *are* $\mathbb{P}$*-independent sets, then*

$$
\sum_{n=1}^{\infty} \mathbb{P}(E_n) = \infty \implies \mathbb{P}(\limsup_{n\to\infty} E_n) = 1.
$$

Proof. The first assertion is an application of countable subadditivity, showing that

$$
\mathbb{P}(\limsup_{m\to\infty} E_n) = \lim_{m\to\infty} \mathbb{P}(\bigcup_{n\geq m} E_n) \leq \lim_{m\to\infty} \sum_{n\geq m} \mathbb{P}(E_n)
$$

if $\sum_{n=1}^{\infty} \mathbb{P}(E_n) > \infty$.

For the second assertion, we will consider E_n's are independent, so by countable additivity, $\mathbb{P}(\limsup_{n\to\infty} E_n) = 1$ if and only if

$$
\lim_{m\to\infty} \mathbb{P}\left(\bigcap_{n\geq m} E_n^c\right) = \mathbb{P}\left(\bigcup_{m=1}^{\infty} \bigcap_{n\geq m} E_n^c\right) = \mathbb{P}\left(\left(\lim_{n\to\infty} E_n\right)^c\right) = 0.
$$

By independence and using countable additivity, for any $m \geq 1$ we have that

$$
\mathbb{P}\left(\bigcap_{n=m}^{\infty} E_n^c\right) = \lim_{N\to\infty} \prod_{n=m}^{N} (1 - \mathbb{P}(E_n)) \leq \lim_{N\to\infty} \exp\left[-\sum_{n=m}^{N} \mathbb{P}(E_n)\right] = 0
$$

if $\sum_{n=1}^{\infty} \mathbb{P}(E_n) = \infty$.

□

Now, we will work on a result that Kingman's Subadditive Ergodic Theorem generalizes: the Strong Law of Large Numbers (SLLN) following [14]. This theorem shows that the partial averages of a sequence of random i.i.d. variables converge almost surely to that finite mean.

Theorem 4.11 (Strong Law of Large Numbers). *Let $X_1, X_2, \dots, X_n$ be a sequence of independent random variables, each having the same finite mean μ, and having $\mathbf{E}[(X_i - \mu)^4] \le a \le \infty$. Then*

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \mu\right) = 1$$

Proof. Since $(X_i - \mu)^2 \le (X_i - \mu)^4 + 1$ (we must keep in mind that we will have two cases: $(X_i - \mu)^2 \le 1$ and $(X_i - \mu)^2 > 1$), we know that each X_i must have variance $\sigma^2 \le a + 1 \equiv v < \infty$. Without loss of generality, we will assume that $\mu = 0$.

Let $S_n = X_1 + X_2 + \cdots + X_n$ and consider $\mathbf{E}(S_n^4)$. We know that the expected value will have terms of the form $X_i X_j X_k X_l$ for i, j, k, l distinct. All these terms will disappear because each one has an expected value of zero. Likewise, all the terms of the form $X_i X_j (X_k)^2$ and $X_i (X_j)^3$ have an expected value of zero. hence, the only terms that will not vanish are the n terms of the form $(X_i)^4$, and the $3n(n-1)$ (from $\binom{n}{2}\binom{4}{2}$) of the form $(X_i)^2 (X_j)^2$ with $i \ne j$. We have that $\mathbf{E}((X_i)^4) \le a$. Aside from this, we know that X_i^2 and X_j^2 are independent, and since $\mu = 0$, we have that $\mathbf{E}((X_i^2)(X_j^2)) = \mathbf{E}((X_i^2))\mathbf{E}((X_j^2)) = \mathbf{Var}(X_i^2)\mathbf{Var}(X_j^2) \le v^2$. Thanks to this process, we know that $\mathbf{E}(S_n^4) \le na + 3n(n-1)v^2 \le Kn^2$ for $K = a + 3v^2$.

Now, we must note that for any $\varepsilon > 0$, we can apply Markov's inequality and get that

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{1}{n} S_n\right| \ge \epsilon\right) &= \mathbb{P}(|S_n| \ge n\varepsilon) = \mathbb{P}\left(|S_n|^4 \ge n^4 \varepsilon^4\right) \\
&\le \frac{\mathbf{E}(S_n^4)}{n^4 \varepsilon^4} \le \frac{Kn^2}{n^4 \varepsilon^4} = K\varepsilon^{-4} \cdot \frac{1}{n^2}.
\end{aligned}$$

Since $\sum_{n=1}^{\infty} \frac{1}{n^2} \le \infty$, we can see that $\frac{1}{n} S_n$ converges to zero almost surely. □

5 Dynamics

Standard references are [15, 5, 16]. We recall that, given a probability space $(M, \mathcal{B}, \mu)$, a map $T : M \to M$ is *measure-preserving* if $\mu(T^{-1}(B)) = \mu(B)$ for all $B \in \mathcal{B}$. Now, we will start defining a fundamental concept: linear cocycles.

Let $(M, \mathcal{B}, \mu)$ be a probability space and $f : M \to M$ a measure-preserving map. Also, we will consider $A : M \to \mathrm{GL}(d, \mathbb{R})$ a measurable function with values in the linear group of invertible $d \times d$ matrices with real coefficients. We will consider the Borel $\sigma-$algebra in $\mathrm{GL}(d)$. Nonetheless, we must keep in mind that sometimes A can take values in the special linear group $\mathrm{SL}(d, \mathbb{R})$ of real $d \times d$ matrices with determinant ± 1.

Definition 5.1. The *linear cocycle* defined by A over f is the transformation

$$\begin{aligned}
F : M \times \mathbb{R}^d &\to M \times \mathbb{R}^d \\
(x, v) &\to (f(x), A(x)v)
\end{aligned}$$

We observe that $F^n(x, v) = (f^n(x), A^n(x))$ for every $n \ge 1$, where

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

and if f is invertible so is F. Moreover, we can see the same property with the inverse, because $F^{-n}(x,v) = (f^{-n}(x), A^{-n}(x))$ for all $n \geq 1$, where

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

If we want to be more general, it is possible for A to take values in the group $\mathrm{GL}(d,\mathbb{C})$ of invertible $d \times d$ matrices with complex coefficients, or the subgroup $\mathrm{SL}(d,\mathbb{C})$ or matrices with determinant in the unit circle. This generates the complex linear cocycles in dimension d, which is also a real cocycle in dimension $2d$. Conversely, every d-dimensional real linear cocycle defines a d-dimensional complex linear cocycle. We will focus on the real case, but they are very similar. We will use the product of random matrices as an example.

Example 5.2. Let $X = \mathrm{GL}(d)$ and $M = X^{\mathbb{Z}}$ and the shift map on X

$$\begin{aligned} &f : M \to M, \\ &(\alpha_k)_k \to (\alpha_{k+1})_k. \end{aligned}$$

Now, we will consider the function

$$\begin{aligned} &A : M \to \mathrm{GL}(d), \\ &(\alpha_k)_k \to \alpha_0 \end{aligned}$$

and let $F : M \times \mathbb{R}^d \to M \times \mathbb{R}^d$ be the linear cocycle defined by A over f. The kth iterate of F would be

$$F^n((\alpha_k)_k, v) = ((\alpha_{k+n})_k, \alpha_{n-1} \dots \alpha_1 \alpha_0 v).$$

Now, we will consider a probability measure p in the space $\mathrm{GL}(d)$ and consider the product measure $\mu = p^{\mathbb{Z}}$, which is characterized by

$$\mu(\{(\alpha_k)_k : \alpha_i \in E_i, \dots, \alpha_j \in E_j\}) = p(E_i) \dots p(E_j)$$

for every $i \leq j$ and any measurable serts $E_i, \dots, E_j \subset X$. Hence, one can see that μ is invariant under the shift map.

Aside from this example, which is closely related to our main goal, we are going to observe an application of this concept to construct a particular cocycle: the derivative cocycle.

Example 5.3. Let $f : S \to S$ with S a smooth surface embedded in $\mathbb{R}^3$ (more generally, S could be a smooth manifold, but we consider surfaces for simplicity of exposition). We will construct a smooth vector field $X_1, \dots, X_d$ on S such that $\{X_1(x), \dots, X_d(x)\}$ is a basis of the tangent space $T_x S$ for every $x \in M$ (in this case $d = 2$). The *derivative cocycle* of f is

$$\begin{aligned} F : S \times \mathbb{R}^d &\to S \times \mathbb{R}^d \\ (x, v) &\mapsto (f(x), A(x)v) \end{aligned} \tag{8}$$

where $A(x) \in \mathrm{GL}(d,\mathbb{R})$ is the matrix, with respect to these bases, of the derivative $Df(x) : T_x S \to T_{f(x)} S$.

6 Kingman's Subadditive Ergodic Theorem

We will begin by proving Kingman's Subadditive Ergodic Theorem, which implies the Furstenberg-Kesten Theorem and the Oseledets Theorem in two dimensions. The reference for the construction of this proof was [15], who was inspired by [17] and the original proof [18]. Our main result is going to be the following.

Theorem 6.1 (Kingman's Subadditive Ergodic Theorem)**.** *Let $\varphi_n : M \to [-\infty, +\infty)$, $n \geq 1$ be a subadditive sequence of measurable functions such that $\varphi_1^+ \in L^1(\mu)$. Then $(\varphi_n/n)_n$* converges μ-almost everywhere to some invariant function $\varphi : M \to [-\infty, +\infty)$ *with respect to the dynamic. Moreover, the positive part φ^+ is integrable and*

$$\int \varphi \, d\mu = \lim_{n\to\infty} \frac{1}{n} \int \varphi_n \, d\mu = \inf_n \frac{1}{n} \int \varphi_n \, d\mu \in [-\infty, +\infty).$$

We will construct the proof in the next sections. Subadditivity is defined first in Subsection 6.1. Let us consider the probability space $(M, \mathcal{B}, \mu)$ and $f : M \to M$ a measure-preserving transformation.

6.1 Subadditivity

Definition 6.2. A sequence $(a_n)_n$ in $[-\infty, \infty)$ is said to be **subadditive** if $a_{m+n} \leq a_m + a_n$ holds for any $m, n \geq 1$.

Also, a measurable function φ is called essentially invariant if $\varphi(f(x)) = \varphi(x)$ for μ-almost all $x \in X$ with f being a measurable transformation. By definition, a measurable subset of X is invariant if its characteristic function is invariant. Now, we can define it properly.

Definition 6.3. A sequence $\varphi_n : X \to [-\infty, \infty)$, $n \geq 1$ of measurable functions is **subadditive**, relative to f, if

$$\varphi_{m+n} \leq \varphi_m + \varphi_n \circ f^m \quad \text{for all} \quad m, n \geq 1.$$

In order to understand this concept better, we will give an example that includes some concepts of ergodic theory as well.

Example 6.4. If we consider any measurable function $\Psi : X \to \mathbb{R}$ and its orbital sum as $\varphi_n = \sum_{j=0}^{n-1} \Psi \circ f^j$. Then, $\varphi_{m+n} = \varphi_m + \varphi_n \circ f^m$ for every m, n, showing that not only $(\varphi_n)_n$ is a subadditive sequence but also an additive sequence since the equality holds.

Example 6.5. Given any measurable function $A : M \to \mathrm{GL}(d)$, consider the sequence $\varphi_n(x) = \log \|A^n(x)\|$, where $A^n(x) = A(f^{n-1}(x)) \dots A(f(x))A(x)$. Now, as $\|B_1 B_2\| \leq \|B_1\| \|B_2\|$ for every $B_1, B_2 \in \mathrm{GL}(d, \mathbb{R})$, the sequence $(\varphi_n)_n$ is subadditive.

Lemma 6.6. *If $(a_n)_n$ is a subadditive sequence then*

$$\lim_n \frac{a_n}{n} = \inf_n \frac{a_n}{n} \in [-\infty, \infty). \tag{9}$$

Proof. If, for some m, $a_m = -\infty$, we have by subadditivity that $a_n = -\infty \quad \forall n > m$. Then, both sides are $-\infty$ and the equality holds for this case.

Now, let us assume that $a_n \in \mathbb{R}$ for all n. We will consider $L = \inf_n(\frac{a_n}{n}) \in [-\infty, +\infty)$ and L' any real number bigger than L. Then, it is possible to get a $k \geq 1$ such that $\frac{a_k}{k} < L'$. Now, any $n > k$ can be written as $n = kp + q$ with $p \geq 1$ and $1 \leq q \leq k$. Hence, applying subadditivity

$$a_n = a_{kp+q} \leq a_{kp} + a_q \leq pa_k + a_q \leq pa_k + \alpha$$

where $\alpha = \max\{a_i : 1 \leq i \leq k\}$ (remember that q is within that same interval). Then,

$$\frac{a_n}{n} \leq \frac{pa_k + \alpha}{n} = \frac{pa_k}{n} + \frac{\alpha}{n} = \frac{pk}{n}\frac{a_k}{k} + \frac{\alpha}{n}.$$

Observe that when $n \to \infty$, $\frac{pk}{n}$ converges to 1 and $\frac{\alpha}{n}$ to zero. Therefore,

$$L \leq \frac{a_n}{n} < L'$$

for any large enough n. Consequently, $L' \to L$ and

$$\lim_n \frac{a_n}{n} = L = \inf_n \frac{a_n}{n}.$$

□

Now, we will consider $(\varphi_n)_n$ where $\varphi_n : M \to [-\infty, \infty)$ with $n \geq 1$ a subadditive sequence of measurable functions such that $\varphi_1^+ \in L^1(\mu)$. By subadditivity,

$$\varphi_n \leq \varphi_1 + \varphi_1 \circ f + \cdots + \varphi_1 \circ f^{n-1}$$

which also holds if we replace the functions with their positive parts. In fact, the hypothesis that $\varphi_1^+ \in L^1(\mu)$ implies that $\varphi_n^+ \in L^1(\mu)$ for all n.

Aside from that, the subadditivity of $(\varphi_n)_n$ implies that

$$a_n = \int \varphi_n d\mu, \quad n \geq 1$$

is a subadditive sequence in $[-\infty, \infty)$. Then, applying Lemma 6.6,

$$\lim_n \frac{1}{n}\int \varphi_n d\mu = \lim_n \frac{a_n}{n} = \inf_n \frac{a_n}{n} = \inf_n \frac{1}{n}\int \varphi_n d\mu = L \in [-\infty, \infty)$$

exists. In fact, if we define $\varphi_- : M \to [-\infty, +\infty]$ and $\varphi_+ : M \to [-\infty, +\infty]$ as

$$\varphi_-(x) = \liminf_n \frac{\varphi_n}{n}(x) \quad \varphi_+(x) = \limsup_n \frac{\varphi_n}{n}(x)$$

it is clear that $\varphi_-(x) \leq \varphi_+(x)$ for every $x \in M$. Hence, we are going to prove that if every function φ_n is bounded away from $-\infty$ then

$$\int \varphi_- d\mu \geq L \geq \int \varphi_+ d\mu \tag{10}$$

which means that φ_- and φ_+ coincide at $\mu-$almost every point and their integrals are equal to L.

6.2 Bounding the functions

We will proceed to prove (10) by working on both sides of the inequality.

We will assume that $\varphi_- > -\infty$ at every point, fix an $\varepsilon > 0$, and, for each $k \in \mathbb{N}$, define

$$E_k = \{x \in M : \varphi_j(x) \leq j(\varphi_-(x) + \varepsilon) \quad \text{for some} \quad j \in 1, \dots, k\}$$

such that $E_k \subset E_{k+1}$ for any k. Also, the definition of $\varphi_-(x)$ implies that $M = \cup_k E_k$. We will also define

$$\Psi_k(x) = \begin{cases} \varphi_-(x) + \varepsilon, & \text{if } x \in E_k \\ \varphi_1(x), & \text{if } x \in E_k^c \end{cases}$$

Given how we defined E_k, we will have that $\varphi_1(x) > \varphi_-(x) + \varepsilon$ for every $x \in E_k^c$. Hence, if we apply the monotonic convergence theorem,

$$\int \Psi_k d\mu \to \int (\varphi_- + \varepsilon) d\mu \qquad \text{as} \quad k \to \infty.$$

Now, we will state the following estimate that will help us further on.

Lemma 6.7. *For any $n > k \geq 1$ at μ-almost every point $x \in M$,*

$$\varphi_n(x) \leq \sum_{i=0}^{n-k-1} \Psi_k(f^i(x)) + \sum_{i=n-k}^{n-1} \max\{\Psi_k, \varphi_1\}(f^i(x)).$$

This lemma is showing that the sequence $(\varphi_n)_n$ is bounded from above by an orbital which is the sum of the $\Psi_k(x)$ and an *error term*. Through this process, we are reducing our subadditive sequence to an additive one because orbital sums are additive sequences. Also, as $\Psi_k(x) \to \psi_-(x) + \varepsilon$ when $k \to \infty$, the sum on the right hand is negligible as n might be much larger than k.

Proof. Let us take $x \in M$ such that $\varphi_-(x) = \varphi_-(f^j(x))$ for any $j \geq 1$. Also, consider a, possibly finite, sequence of integer numbers

$$m_0 \leq n_1 < m_1 \leq n_2 < m_2 < \dots \tag{11}$$

defined inductively as shown in Figure 1.

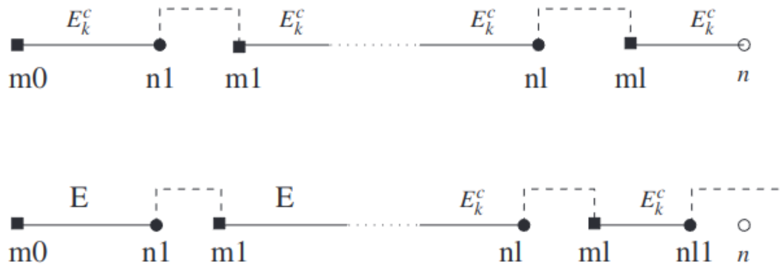

Figure 1: Splitting the trajectory of a point [15].

Take $m_0 = 0$. Given $j \geq 1$, let n_j be the smallest integer greater or equal than m_{j-1} that satisfies $f^{n_j}(x) \in E_k$ (if it exists). Then, by definition of E_k, there exists an m_j such that $1 \leq m_j - n_j \leq k$ and

$$\varphi_{m_j - n_j}(f^{n_j}(x)) \leq (m_j - n_j)(\varphi_-(f^{n_j}(x)) + \varepsilon). \tag{12}$$

Given any $n \geq k$, let $l \geq 0$ be the largest such that $m_j \leq n$. By subadditivity,

$$\varphi_{n_j - m_{j-1}}\left(f^{m_{j-1}}(x)\right) \leq \sum_{i=m_{j-1}}^{n_j - 1} \varphi_1\left(f^i(x)\right) \tag{13}$$

for any $j = 1, \ldots, l$ such that $m_{j-1} \neq n_j$, and similarly for $\varphi_{n-m_l}(f^{m_l}(x))$. Thus,

$$\varphi_n(x) \leq \sum_{i \in I} \varphi_1(f^i(x)) + \sum_{j=1}^{l} \varphi_{m_j - n_j}(f^{n_j}(x)) \tag{14}$$

where $I = \bigcup_{j=1}^{l}[m_{j-1}, n_j) \cup [m_l, n)$. Observe that

$$\varphi_1(f^i(x)) = \Psi_k(f^i(x)) \quad \text{for any} \quad i \in \bigcup_{j=1}^{l}[m_{j-1}, n_j) \cup [m_l, \min\{n_{l+1}, n\}),$$

because $f^i(x) \in E_k^c$ in all of these cases. Furthermore, as φ_- is a constant on orbits (see 6.5)and $\Psi_k \geq \varphi_- + \varepsilon$, so 6.2 implies that

$$\varphi_{m_j - n_j}(f^{n_j}(x)) \leq \sum_{i=n_j}^{m_j - 1} (\varphi_-(f^i(x)) + \varepsilon) \leq \sum_{i=n_j}^{m_j - 1} \Psi_k(f^i(x))$$

for every $j = 1, \ldots, l$. Then, using 14 we can conclude that

$$\varphi_n(x) \leq \sum_{i=0}^{\min\{n_{l+1}, n\} - 1} \Psi_k(f^i(x)) + \sum_{i=n_{l+1}}^{n-1} \varphi_1(f^i(x)).$$

The lemma is proved as $n_{l+1} > n - k$. $\square$

6.2.1 Bounding from below

We will proceed to prove the left side of (10).

Lemma 6.8. $\int \varphi_- d\mu = L$.

Proof. In first place, we will assume that there is a $\kappa > 0$ such that $\frac{\varphi_n}{n} \geq -\kappa$ for every n (this means that the sequence is uniformly bounded from below). Then $\varphi_- \geq -\kappa > -\infty$. Given that inequality, we get that $\frac{\varphi_n}{n} + \kappa$ is a sequence of non-negative functions, and applying Fatou's lemma it is possible to say that φ_- is integrable and

$$\int \varphi_- d\mu \leq \lim \int \frac{\varphi_n}{n} d\mu = L.$$

Now, we use Lemma 6.7 to prove the opposite inequality, as we get that

$$\frac{1}{n} \int \varphi_n d\mu \leq \frac{n-k}{n} \int \Psi_k d\mu + \frac{k}{n} \int \max\{\Psi_k, \varphi_1\} d\mu.$$

Given what we know about Ψ_k, $\max\{\Psi_k, \varphi_1\} \leq \max\{\varphi_- + \varepsilon, \varphi_1^+\}$, and this last function is integrable. Hence, the $\limsup_n$ of the final term of the inequality is non-positive. Now, taking $n \to \infty$ we obtain that

$$L \leq \int \Psi_k d\mu$$

for any k. Also, if we take into consideration that $k \to \infty$, we have that

$$L \leq \int \varphi_- d\mu + \varepsilon.$$

which when $\varepsilon \to 0$ is equal to

$$L \leq \int \varphi_- d\mu$$

and we get our desired result assuming the sequence is uniformly bounded form below.

We now remove that assumption, so, for each $\kappa > 0$, we will define

$$\varphi_n^k = \max\{\varphi_n, -\kappa n\} \quad \text{and} \quad \varphi_-^k = \max\{\varphi_-, -\kappa\}.$$

Note that $(\varphi_n^k)_n$ is a subadditive sequence of measurable functions such that $\varphi_1^+ \in L^1(\mu)$ and it is subadditive. Then, $\varphi_-^\kappa = \liminf_n (\frac{1}{n}) \varphi_n^\kappa$ and, by the previous part of this proof:

$$\int \varphi_-^\kappa d\mu = \inf_n \frac{1}{n} \int \varphi_n^\kappa d\mu.$$

By the monotone convergence theorem we also have that

$$\int \varphi_n \, d\mu = \inf_\kappa \int \varphi_n^\kappa \, d\mu \quad \text{and} \quad \int \varphi_- \, d\mu = \inf_\kappa \int \varphi_-^\kappa \, d\mu.$$

Combining these equalities,

$$\int \varphi_- \, d\mu = \inf_\kappa \int \varphi^\kappa \, d\mu = \inf_\kappa \inf_n \frac{1}{n} \int \varphi_n^\kappa \, d\mu = \inf_n \frac{1}{n} \int \varphi_n \, d\mu = L.$$

$\square$

6.2.2 Bounding from above

We will now proceed to prove that $\int \varphi_+ d\mu \le L$ if every φ_n is bounded from $-\infty$. However, we must introduce some results first.

Lemma 6.9. *If $\phi : M \to \mathbb{R}$ is integrable with respect to μ then*

$$\lim_n \frac{1}{n} \phi(f^n(x)) = 0 \quad \textit{for} \quad \mu - \textit{almost all} \quad x \in M.$$

Proof. We will fix an arbitrary $\varepsilon > 0$, and, since μ is invariant under f, we get that

$$\begin{aligned}\mu\left(\{x \in M : |\phi(f^n(x))| \ge n\varepsilon\}\right) &= \mu\left(\{x \in M : |\phi(x)| \ge n\varepsilon\}\right) \\ &= \sum_{k=n}^{\infty} \mu\left(\left\{x \in M : k \le \frac{|\phi(x)|}{\varepsilon} < k+1\right\}\right).\end{aligned}$$

Adding these inequalities over $n \in \mathbb{N}$, we get that

$$\begin{aligned}\sum_{n=1}^{\infty} \mu\left(\{x \in M : |\phi(f^n(x))| \ge n\varepsilon\}\right) &= \sum_{k=1}^{\infty} k\mu\left(\left\{x \in M : k \le \frac{|\phi(x)|}{\varepsilon} < k+1\right\}\right) \\ &\le \int \frac{|\phi|}{\varepsilon} \, d\mu.\end{aligned}$$

Given that ϕ is integrable, the right-hand is finite and the Theorem 4 applies. In fact, it helps us conclude that the set $B(\varepsilon)$ of points x such that $|\phi(f^n(x))| \ge n\varepsilon$ has measure zero for infinitely many vaues of n. Given what we know of $B(\varepsilon)$, for every $x \notin B(\varepsilon)$ there is a $p \ge 1$ such that $|\phi(f^n(x))| < n\varepsilon$ for every $n \ge p$.

Consequently, if we consider $B = \cup_{i=1}^{\infty} B(\frac{1}{i})$ (the measure of a countable union of measure zero sets) we can conclude it has measure zero and $\lim_n (\frac{1}{n}) \phi(f^n(x)) = 0$ for every $x \notin B$.

$\square$

Lemma 6.10. *For any fixed k,*

$$\limsup_n \frac{\varphi_{kn}}{n} = k \limsup_n \frac{\varphi_n}{n}$$

Proof. We know that $\limsup_n \frac{\varphi_{kn}}{n} \le k \limsup_n \frac{\varphi_n}{n}$ because $\frac{\varphi_{kn}}{kn}$ is a subsequence of $\frac{\varphi_n}{n}$. Now, to get the other inequality, we will write $n = kq_n + r_n$ with $r_n \in \{1, \dots, k\}$. Given that (φ_n) is subadditive, we get that

$$\varphi_n \le \varphi_{kq_n} + \varphi_{r_n} \circ f^{kq_n} \le \varphi_{kq_n} + \Psi \circ f^{kq_n}$$

where $\Psi = \max\{\varphi_1^+, \dots, \varphi_k^+\}$. Note that $\frac{n}{q_n} \to k$ as $n \to \infty$ and $\Psi \in L^1(\mu)$. Hence, using Lemma 6.9 we have that $\frac{\Psi \circ f^n}{n}$ converges to zero at μ-almost every point. Hence, we can take the previous inequality, divide it by n and take the lim sup as $n \to \infty$ to get

$$\limsup_n \frac{1}{n}\varphi_n \leq \limsup_n \frac{1}{n}\varphi_{kq_n} + \limsup_n \frac{1}{n}\Psi \circ f^{kq_n} = \frac{1}{k}\limsup_q \frac{1}{q}\varphi_{kq}$$

which is equal to

$$k \limsup_n \frac{1}{n}\varphi_n \leq \limsup_q \frac{1}{q}\varphi_{kq} = \limsup_n \frac{1}{n}\varphi_{kn}$$

and we get the other side of the inequality necessary to prove the equality stated above. $\square$

After proving this results, we are finally ready to bound the function from above and prove the left side of (10).

Lemma 6.11. *Suppose that* $\inf \varphi_n > -\infty$ *for any* n*. Then,* $\int \varphi_+ d\mu \leq L$*.*

Proof. We will consider a fixed k and $n \geq 1$ and define $\theta_n = -\sum_{j=0}^{n-1} \varphi_k \circ f^{jk}$. Given that f^k is a measure preserving function,

$$\int \theta_n d\mu = -n \int \varphi_k d\mu \quad \text{for every} \quad n. \tag{15}$$

As $(\varphi_n)_n$ is a subadditive sequence, for any n we will have $\theta_n \leq -\varphi_{kn}$. Hence, applying Lemma 6.10 we have that

$$\theta_- = \liminf_n \frac{\theta_n}{n} \leq -\limsup_n \frac{\varphi_{kn}}{n} = -k \limsup_n \frac{\varphi_n}{n} = -k\varphi_+$$

and that implies that

$$\int \theta_- d\mu \leq -k \int \varphi_+ d\mu.$$

Given that $(\theta_n)_n$ is the sum of additive functions (as we saw in Lemma 6.7) we have that $\theta_{m+n} = \theta_m + \theta_n \circ f^{km}$ for any $m, n \geq 1$. As $\theta_1 = -\varphi_k$ is bounded from above by $-\inf \varphi_k$, the function θ_1^+ is bounded and consequently integrable. Hence, it is possible to apply Lemma 6.8 and 6.2.2 to conclude that

$$\int \theta_- d\mu = \lim_n \int \frac{\theta_n}{n} d\mu = -\int \varphi_k d\mu.$$

Joining both inequalities above, we obtain

$$\int \varphi_+ d\mu \leq \frac{1}{k} \int \varphi_k d\mu$$

which yields $\int \varphi_+ d\mu \leq L$ when taking the infimum on k. $\square$

After much preparation, we are ready to prove the Kingman's Subadditive Ergodic Theorem.

Proof. (Kingman's Subadditive Ergodic Theorem(6.1)) Using Lemma-6.8 and Lemma 6.11, we already proved that $(\frac{\varphi_n}{n})_n$ converges μ-almost everywhere to an invariant function when $\inf \varphi_n > -\infty$ for any n. Hence, we need to prove it for the general case, so we will define the following

$$\varphi_n^\kappa = \max\{\varphi_n, -\kappa n\}, \quad \varphi_-^\kappa = \max\{\varphi_-, -\kappa n\} \quad \text{and} \quad \varphi_+^\kappa = \max\{\varphi_+, -\kappa n\}$$

for any constant $\kappa > 0$.

Hence, the same arguments we have used can be applied to the sequence $(\varphi_n^\kappa)_n$ for any fixed $\kappa > 0$. Therefore, $\varphi_+^\kappa = \varphi_-^\kappa$ for any $\kappa > 0$ at μ-almost every point. Given that $\varphi_-^\kappa \to \varphi_-$ and $\varphi_+^\kappa \to \varphi_+$ when $\kappa \to \infty$, we can extend our previous statement to say that $\varphi_+ = \varphi_-$ at μ-almost every point. Officially, we have proved the convergence of the sequence and the equivalence of the integral with the infimum of the sequence. □

7 Consequences of Kingman's Subadditive Ergodic Theorem

Our main results are corollaries of Kingman's Subadditive Ergodic Theorem, so we will proceed to prove them in the following order: Birkhoff's Ergodic Theorem, Furstenberg-Kesten Theorem and Oseledets Theorem in 2 Dimensions.

Theorem 7.1 (Birkhoff Ergodic Theorem)**.** *Let $\varphi : M \to \mathbb{R}$ be a μ-integrable function. Then*

$$\tilde{\varphi}(x) = \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x))$$

exists at μ-every point. Moreover, the function $\tilde{\varphi}$ is invariant and μ-integrable, with $\int \tilde{\varphi} \mathrm{d}\mu = \int \varphi \mathrm{d}\mu$.

Proof. Thanks to Example 6.4, we have that this is a particular case of Theorem 6.1. □

7.1 Furstenberg-Kesten Theorem

Now, we will proceed to prove our main result. In order to do that, we will take $F : M \times \mathbb{R}^d \to M \times \mathbb{R}^d$ defined as $F(x, v) = (f(x), A(x)v)$ for some measurable function $A : M \to \mathrm{GL}(d)$ and f measure preserving.

Furstenberg and Kesten Theorem. *If $\log^+ \|A^{\pm 1}\| \in L^1(\mu)$ then*

$$\lambda_+(x) = \lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)\| \quad \text{and} \quad \lambda_-(x) = \lim_{n\to\infty} \frac{1}{n} \log \|(A^n(x))^{-1}\|^{-1}$$

exist for μ-almost everywhere $x \in M$. *Moreover, the functions $\lambda_\pm$ are* invariant and μ-integrable*, with*

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

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

Proof. This is a direct consequence of Kingman's Theorem. Let us call

$$\varphi_n(x) = \log \|A^n(x)\| \quad \text{and} \quad \Psi_n(x) = \log \|(A^n(x))^{-1}\|.$$

Thanks to the hypothesis, we know that $\varphi_1^+, \Psi_1^+ \in L^1(\mu)$ so $\varphi_1, \Psi_1 \in [-\infty, +\infty)$ for μ-almost every x. Given that the norm of linear operators is sub-multiplicative (see 6.5), the sequences φ_n and Ψ_n are subadditive.

Now, we have all the requirements to apply Theorem 6.1 and get that

$$\int \lambda_+(x) d\mu = \lim_{n\to\infty} \frac{1}{n} \int \log \|A^n(x)\| d\mu$$

and

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

□

This theorem provides a fundamental result on the asymptotic growth rate of the product of random matrices. Given an independent identically distributed sequence of random matrices in $\mathrm{GL}(d, \mathbb{R})$, the general linear group of invertible $d \times d$ random matrices, that follow certain conditions of integrability and irreducibility, we are able to find the so-called **Lyapunov exponents** which are λ_+ and λ_-. The exponents found constitute a quantifiable ergodic characterization of the rates of exponential growth or decay of vectors under the action of a sequence of matrices.

7.2 Oseledets in Two Dimensions Theorem

Through this theorem, we are going to prove a version of the multiplicative ergodic theorem for invertible and non-invertible 2-dimensional cocycles. The reason why it is exposed as a separate case is because it can be handled with simpler methods, while still being useful for our main purpose of analyzing the product of random matrices. Oseledets' multiplicative ergodic theorem is more powerful than Furstenberg's as it is capable of providing the whole Lyapunov spectrum, their multiplicities, and the decomposition of the space into invariant subspaces. This way, we have the directions of growth and decay, not just the maximal rate at which all vectors tend to the unstable or stable direction.

First, we will introduce a result that will be useful further on and is a consequence of Theorem 7.1.

Corollary 7.2. *Let $\phi : M \to \mathbb{R}$ be a measurable function such that the function $\Psi = \phi \circ f - \phi$ is integrable with respect to μ. Then for $\mu-$almost all $x \in M$*

$$\lim_n \frac{1}{n} \phi(f^n(x)) = 0.$$

In particular, this holds if $\phi \in L^1(\mu)$.

Proof. Note that $\phi(f^n(x)) = \phi(x) + \sum_{j=0}^{n-1} \Psi(f^j(x))$ for every x and n. Hence, if we apply Birkhoff's ergodic theorem to the integrable function Ψ,

$$\lim_{n\to\infty} \frac{1}{n} \phi(f^n(x)) = \lim_{n\to\infty} \frac{1}{n} \phi(x) + \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \Psi(f^j(x)) \tag{16}$$

exists at μ-almost every point. Aside from this, since μ is f-invariant,

$$\mu\left(\left\{x : \left|\frac{1}{n}\phi(f^n(x))\right| \geq c\right\}\right) = \mu\left(\{y : |\phi(y)| > nc\}\right) \to 0 \quad \text{as } n \to \infty.$$

Which means that the sequence $\frac{1}{n}\phi \circ f^n$ converges to zero in measure. Thus, the limit (16) must be zero at μ-almost every point. $\square$

Let us define $F : M \times \mathbb{R}^2 \to M \times \mathbb{R}^2$ and $F(x, v) = (f(x), A(x)v)$, for some measurable function $A : M \to \mathrm{GL}(2)$ satisfying $\log^+ \|A^{\pm 1}\| \in L^1(\mu)$. Also, we will take $s_n(x)$ and $u_n(x)$ as the unit vectors most contracted and most expanded under $A^n(x)$, such that $\|A^n s_n\| = \|A^n\|^{-1}$ and $\|A^n u_n\| = \|A^n\|$. With this in mind, we will start defining some lemmas that will contribute to the proof of the main theorem.

Lemma 7.3. *The angle $\angle(s_n(x), s_{n+1}(x))$ decreases exponentially such that*

$$\limsup_n \frac{1}{n} \log |\sin \angle(s_n(x), s_{n+1}(x))| \leq -2\lambda(x)$$

with $\lambda = \lambda_+ - \lambda_-$.

Proof. We will denote $\alpha_n = \angle(s_n(x), s_{n+1}(x))$. Then, if we project $s_n(x)$ onto the direction of $s_{n+1}(x)$ and $u_{n+1}(x)$ which is orthonormal to $s_{n+1}(x)$ (we are decomposing a vector into components along an orthonormal basis). Then,

$$s_n(x) = \sin \alpha_n u_{n+1}(x) + \cos \alpha_n s_{n+1}(x).$$

With this in mind, we want to bound $\|A^{n+1}(x)s_n(x)\|$, taking into consideration that it is the norm of the product of the cocycle product times the direction of greatest contraction. First, we have that

$$A^{n+1}(x)s_n(x) = A^{n+1}(x) \sin \alpha_n u_{n+1}(x) + A^{n+1}(x) \cos \alpha_n s_{n+1}(x)$$

and if we apply the norm, thanks to the triangular inequality with only one of the summands, we get that

$$\|A^{n+1}(x)s_n(x)\| \geq \|\sin \alpha_n A^{n+1}(x)u_{n+1}(x)\| = |\sin \alpha_n| \|A^{n+1}(x)\|.$$

Now, to get an upper bound, we use the definition of cocycles, which states that

$$A^{n+1}(x) = A(f^n(x))A^n(x)$$

and get that

$$\|A^{n+1}(x)s_n(x)\| = \|A(f^n(x))A^n(x)s_n(x)\|.$$

Applying the submultiplicative property of norms [19], which is that $\|ABv\| \leq \|A\|\|Bv\|$, we obtain

$$\|A^{n+1}(x)s_n(x)\| \leq \|A(f^n(x))\|\|A^n(x)s_n(x)\| = \|A(f^n(x))\|\|A^n(x)\|^{-1}.$$

Hence,

$$|\sin\alpha_n| \le \frac{\|A(f^n(x))\|}{\|A^{n+1}(x)\|\ \|A^n(x)\|}. \tag{17}$$

Corollary 7.2 ensures that

$$\lim_n \frac{1}{n}\log\|A(f^n(x))\| = 0.$$

So, the limit on both sides of the expression 17 leads to

$$\limsup_n \frac{1}{n}\log|\sin\alpha_n| \le -2\lambda(x).$$

□

Now, we want to know where does the limit goes to. Hence, we will show that the sequence we are interested in is Cauchy in projective space. We use projective spaces because they are equipped to deal with points at infinity and regular points in a uniform way without having to make distinctions [20].

Lemma 7.4. *The sequence $(s_n(x))_n$ is Cauchy in projective space.*

Proof. Let us consider any $\varepsilon > 0$ such that $-2\lambda(x)+\varepsilon < 0$. Thanks to Lemma 7.3, we have that

$$|\sin\alpha_n| \le e^{n(-2\lambda(X)+\varepsilon)}$$

for every n large enough. Then, replacing $s_j(x)$ by $-s_j(x)$ on the orthonormal decomposition exposed before and taking the norm, we have that

$$\|s_n(x) - s_{n+1}(x)\| \le 2e^{n(-2\lambda(x)+\varepsilon)}$$

for every n large enough. Consequently, there exists a $C > 0$ such that

$$\|s_{n+k}(x) - s_n(x)\| \le Ce^{n(-2\lambda(x)+\varepsilon)}$$

for every $k \ge 1$ and n large enough. In particular, the sequence is Cauchy. □

Define $s(x) = \lim s_n(x)$.

Lemma 7.5. *The vector $s(x)$ is contracted at the rate $-\lambda(x)$, that is,*

$$\lim_n \frac{1}{n}\log\|A^n(x)s(x)\| = -\lambda(x).$$

Proof. Let $\beta_n = \angle(s(x), s_n(x))$. Then, we can do orthogonal decomposition as well and get that $s(x) = \cos\beta_n s_n(x) + \sin\beta_n u_n(x)$ and using triangular inequality over the norm we have that

$$\|A^n(x)s(x)\| \le |\cos\beta_n|\|A^n(x)s_n(x)\| + |\sin\beta_n|\|A^n(x)u_n(x)\|.$$

Then, applying $\limsup_n \frac{1}{n}$ we have that

$$
\begin{aligned}
&\limsup_n \frac{1}{n} \log \|A^n(x)s(x)\| \\
&\leq \max\left\{\limsup_n \frac{1}{n} \log|\cos\beta_n|\|A^n(x)s_n(x)\|, \limsup_n \frac{1}{n} \log|\sin\beta_n|\|A^n(x)u_n(x)\|\right\} \\
&\leq \max\left\{\limsup_n \frac{1}{n} \log\|A^n(x)\|^{-1}, \limsup_n \frac{1}{n} \log|\sin\beta_n| + \limsup_{n\to\infty} \frac{1}{n} \log\|A^n(x)u_n(x)\|\right\} \\
&\leq \max\{-\lambda(x), -2\lambda(x) + \lambda(x)\} = -\lambda(x).
\end{aligned}
$$

completing the proof. $\square$

Now, we will check an important property about the asymptotic growth rate of the norm of vectors under linear cocycle iterations.

Lemma 7.6. *If $v \in \mathbb{R}^2$ is not collinear with $s(x)$ then*

$$\lim_n \frac{1}{n} \log\|A^n(x)v\| = \lambda(x).$$

Proof. Denote $\gamma_n = \angle(v, s_n(x))$. Then $v = \cos\gamma_n s_n(x) + \sin\gamma_n u_n(x)$ decomposing the vector in orthonormal components. Hence, applying the triangular inequality over the norm after multiplying it by the cocycle $A^n(x)$ we have that

$$\|A^n(x)v\| \geq |\sin\gamma_n|\|A^n(x)u_n(x)\| - |\cos\gamma_n|\|A^n(x)s_n(x)\|.$$

Note that $|\sin\gamma_n|$ is bounded from zero for all large enough n because $s_n(x) \to s(x)$ and v is not collinear to $s(x)$. Taking into consideration Lemma 7.3, we have that

$$\|A^n(x)u_n(x)\| \approx e^{n\lambda(x)} \quad \text{and} \quad \|A^n(x)s_n(x)\| \approx e^{-n\lambda(x)}$$

and substituting on the previous inequality we have that

$$\|A^n(x)v\| \geq |\sin\gamma_n|e^{n\lambda(x)} - |\cos\gamma_n|e^{-n\lambda(x)}.$$

If we apply the logarithm, multiply $\frac{1}{n}$ times and take the inferior limit, we have that whenever $n \to \infty$,

$$\liminf_n \frac{1}{n} \log\|A^n(x)v\| \geq \lambda(x).$$

We get the opposite inequality immediately from the triangular inequality applied over $\|A^n(X)v\|$. Hence,

$$\limsup_n \frac{1}{n} \log\|A^n(x)v\| \leq \lim_n \frac{1}{n} \log\|A^n(x)\| = \lambda(x)$$

and the proof is complete. $\square$

Now, we will work on a lemma that shows that applying the linear operator $A(x)$ to the unit vector that points towards the direction of most significant contraction, $s(x)$, preserves its direction when mapped forward by f.

Lemma 7.7. *$A(x)s(x)$ is collinear to $s(f(x))$.*

Proof. Applying Lemma 7.5, we have that

$$\lim_n \frac{1}{n+1} \log \|A^{n+1}(x)s(x)\| = -\lambda(x)$$

and the left side is equivalent to $\lim_n \frac{1}{n} \log \|A^n(f(x))A(x)s(x)\|$. Now, using Lemma 7.6, for every v not collinear to $s(f(x))$, we have that

$$\lim_n \frac{1}{n} \log \|A^n(f(x))v\| = \lambda(f(x)) = \lambda(x).$$

Hence, we must have $A(x)s(x)$ collinear to $s(f(x))$. □

After working on these lemmas, we will apply them to prove the one-sided theorem.

Oseledets One-sided Theorem. *For μ-almost every $x \in M$,*

(i) either $\lambda_-(x) = \lambda_+(x)$ and

$$\lim_n \frac{1}{n} \log \|A^n(x)v\| = \lambda_\pm(x), \text{ for all } v \in \mathbb{R}^2;$$

(ii) or $\lambda_+(x) > \lambda_-(x)$ and there exists a vector line $E^s_x \subset \mathbb{R}^2$ such that

$$\lim_n \frac{1}{n} \log \|A^n(x)v\| = \begin{cases} \lambda_-(x) & \text{if } v \in E^s_x \setminus \{0\}, \\ \lambda_+(x) & \text{if } v \in \mathbb{R}^2 \setminus E^s_x. \end{cases}$$

Moreover, $A(x)E^s_x = E^s_{f(x)}$ for every x in the second case.

Proof. We will consider $A \in \mathrm{SL}(2)$ and any x as stated in the conclusion of Theorem 7.1 such that $\lambda_+(x) + \lambda_-(x) = 0$ (because the determinant is 1, so the log of the determinant is 0). We will begin considering the case $\lambda(x) = \lambda_+(x) = -\lambda_-(x)$ which suggests that for $x \in M$, $\lambda(x) = 0$. For any $v \in \mathbb{R}^2$,

$$\|A^n(x)\|^{-1}\|v\| = \|A^n(x)^{-1}\|^{-1}\|v\| \leq \|A^n(x)v\| \leq \|A^n(x)\|\|v\|$$

where we are applying the standard submultiplicative property of operator norms and the property of operator norms for invertible matrices [19]. Through this process, we are bounding the norm of the transformed vector $\|A^n(x)v\|$ from above by the norm of the operator times $\|v\|$, and from below by the inverse operator norm acting on $\|v\|$. We have that

$$\frac{1}{n} \log \left(\|A^n(x)\|^{-1}\|v\|\right) \leq \frac{1}{n} \log \|A^n(x)v\| \leq \frac{1}{n} \log \left(\|A^n(x)\|\|v\|\right).$$

Consequently, when $n \to \infty$, the left side tends to $-\lambda(x) = 0$ and the right side to $\lambda(x) = 0$. Hence, we have already proved the first case.

Now, we will suppose that $\lambda(x) > 0$, so $\|A^n(x)\| \approx e^{n\lambda(x)}$ which is larger than 1 for every n large enough. Now, we know that given $A \in \mathrm{SL}(2)$ such that $\|A\| \neq 1$, there exists unit vectors s and u such that $\|A(u)\| = \|A\|$ and $\|A(s)\| = \|A^{-1}\|^{-1} = \|A\|^{-1}$. These vectors are unique, up to multiplication by -1, they are orthogonal, and their images are orthogonal as well. Hence, we will take $s_n(x)$ as the unit vector most contracted under $A^n(x)$, and $u_n(x)$ as the unit vector most expanded under $A^n(x)$. With this vectors we have that

$$\|A^n(x)s_n(x)\| = \|A^n(x)\|^{-1} \quad \text{and} \quad \|A^n(x)u_n(x)\| = \|A^n(x)\|.$$

Hence, if we take E_x^s as the real line generated by $s(x)$, we have shown the theorem through the previous lemmas. We proved that the angle of $s_n(x)$ decreases exponentially and converges throughout multiple iterations in Lemma 7.3 and Lemma 7.4. Later on, through Lemma 7.6 we showed that whenever vectors are not collinear with our E_x^s (meaning they are in $\mathbb{R}^2 - E_x^s$), $\lim_n \frac{1}{n} \log \|A^n(x)v\| = \lambda_+(x)$ because we are taking the vector of greatest expansion. Considering the other case, exposed in Lemma 7.5, whenever they are collinear (meaning they are in $E_x^s - \{0\}$), we are taking the vector of greatest contraction and consequently $\lim_n \frac{1}{n} \log \|A^n(x)v\| = \lambda_-(x)$. Finally, through Lemma 7.7 we showed that $A(x)E_x^s = E_{f(x)}^s$. □

8 Simulations

To illustrate the behavior of the product of random matrices and the convergence towards their Lyapunov exponents, we are going to perform a series of numerical simulations using the programming language Julia and guided by [21]. These experiments intend to visualize the asymptotic behavior described in Theorem 7.1 and Theorem 7.2. For these cases, we will assume that f is the identity matrix.

Given the criteria for the theorems to be valid, we are going to simulate a sequence (A_n) of i.i.d. random matrices in $\mathrm{GL}(2, \mathbb{R})$ and compute the norm of the product over time. In this case, each entry will be taken from a normal distribution, and in the function that creates the new matrices, we will add a condition to avoid near-singular matrices. Now, we will start simulating the theorems.

8.1 Furstenberg and Kesten

8.1.1 Top Lyapunov Exponent

We know that the Lyapunov exponent is approximated numerically by

$$\lambda_n = \frac{1}{n} \log \|A_n A_{n-1} \dots A_1\|,$$

so we will calculate these value for 10000 iterations. This is the way they are calculated because we assume they are exponential, so by applying the logarithm and getting the average, we are isolating the growth rate (the Lyapunov exponents). As we can see in A.1, we are going to define the size of the matrix, the number of iterations, and a seed to guarantee the reproducibility of the results obtained. Then, we are going to create a function called `random_matrix()` that generates a matrix of the dimensions specified in which each entry has been randomly selected from a normal standard distribution. This distribution was chosen because all the values are between 0 and 1, but not too small or too large. This is satisfactory because values too close to zero tend to get too small fast, and large values lead to memory leakage. Aside from that, this function checks the determinant and if its absolute value is below `1e-3`, a new one is calculated. This way, we are avoiding near-singular matrices. Finally, we have a function called `simulate_lyapunov(d,N)` that for each iteration generates a random matrix, multiplies it times the matrix that is the result of the previous multiplications, and calculates the Lyapunov exponent. After these calculations, we plot the values of the Lyapunov exponents for each iteration, and the result is displayed in 2.

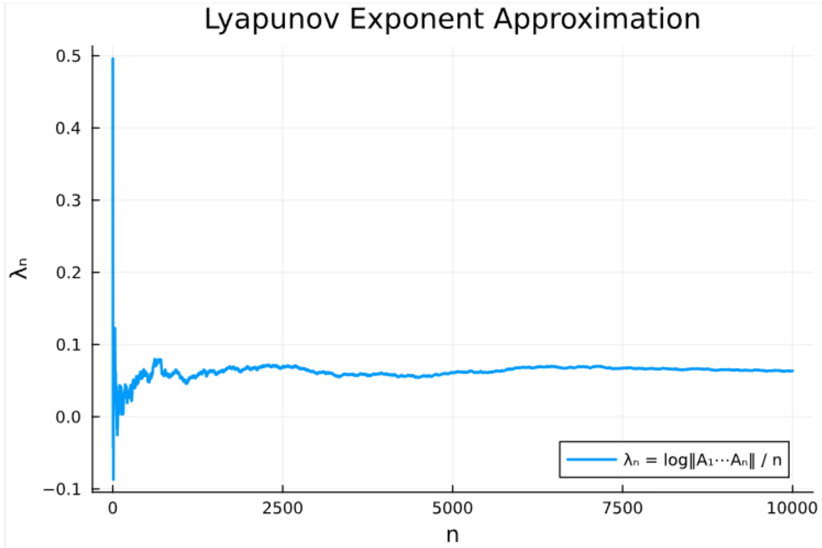

Figure 2: Asymptotic behavior of top Lyapunov exponent in 10000 iterations.

As we can see in the graph, during the first 1250 iterations, the values fluctuate immensely. However, as the number of iterations increases, we can see that it stabilizes and the sequence (λ_n) seems to converge to a deterministic limit, which is interpreted as the **top lyapunov exponent** λ_+.

8.1.2 Histogram of Top Lyapunov Exponents

As we want to observe the distribution of the top Lyapunov exponent across 1000 independent trials, we run a simulation `M` A.2 times and compute the Lyapunov exponent from the product of random matrices.

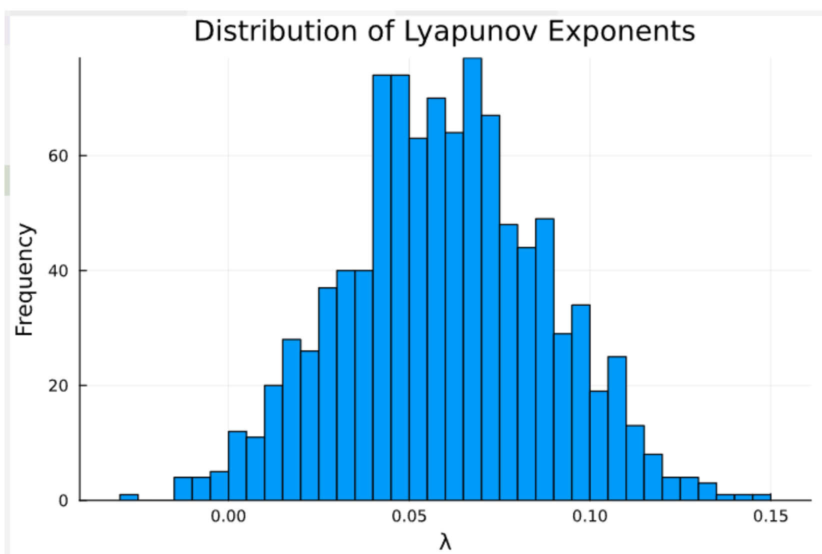

Figure 3: Distribution of top Lyapunov Exponent over 1000 independent trials.

From Figure 3, we can see that most computed exponents concentrate around a typical value close to 0.06, supporting the fact that the exponent exists almost surely and is deterministic. Aside from that, the width of the distribution reflects the effect of having a finite sample size. Having a larger `n` would yield a tighter concentration around the value. Aside from this, the shape resembles a normal distribution, which makes sense given the central limit behavior for subadditive sequences.

8.2 Oseledets

This theorem is stronger when it comes to understanding asymptotic behavior as it gives us an almost sure existence of all Lyapunov exponents (one per dimension) and their corresponding invariant subspaces (Oseledets splitting) that describe directional behavior. Hence, we are going to run simulations over those two main features: calculation of all Lyapunov exponents and directional behavior.

8.2.1 Lyapunov exponents

We are going to show how this theorem helps us find all λ_i through A.3. Nonetheless, to do the numerical simulation, we are forced to do the QR decomposition before. We have to do this because the exponential growth or decay of the norm may cause underflow or overflow of memory. Also, after many multiplications, the product is dominated by the top Lyapunov direction, so we might lose information about the other directions. Doing this process "re-normalizes" the product at each iteration by rewriting the matrix as an orthogonal matrix times a positive diagonal matrix [22]. This can be seen as a computational implementation of the multiplicative ergodic theory. With this in mind, we are going to calculate the main Lyapunov exponents for different dimensions with the following code (if we change the value of `d` in the main function we are altering the number of dimension).

Now, can see in Figure 4 that as the two Lyapunov exponents converge as time passes to specific values.

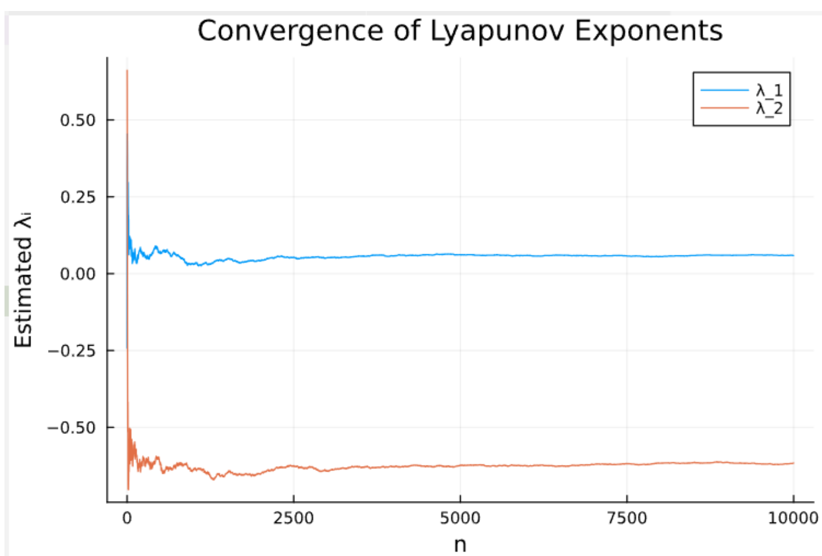

Figure 4: Lyapunov exponents of two main directions.

In this case, we can see that λ_1 is the unstable growth rate and λ_2 is the stable growth rate.

This means that as time passes, all vectors will go to the unstable direction on a rate of λ_1, and the stable direction will tend to zero with a rate of λ_2.

8.2.2 Directional Behavior

We know that one of the strengths of Oseledets' Theorem is its ability to get the invariant subspaces associated with each Lyapunov exponent. Hence, we will start by graphing the directions in three dimensions and interpreting its results obtained through A.4.

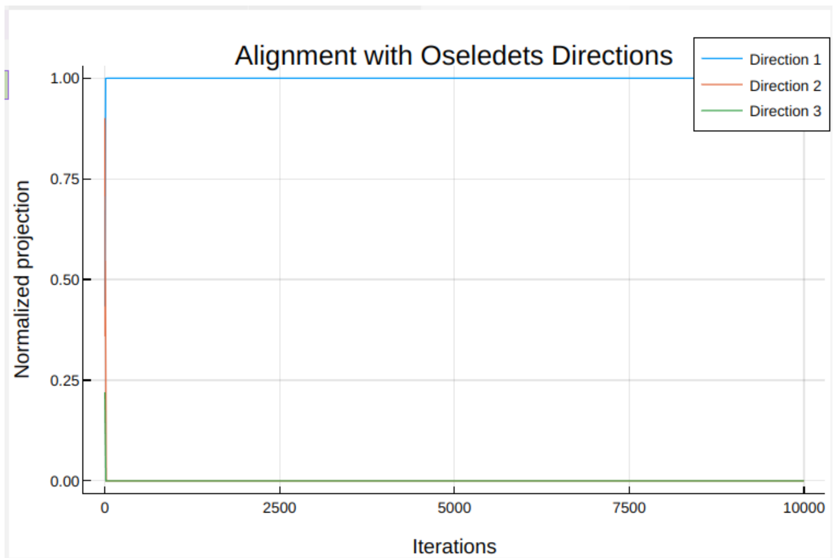

Figure 5: Evolution of directions on a three dimensional matrix

In this case, we can see that the Direction 1 is associated to the largest Lyapunov exponent and is the top Oseledet's direction. Also, the other two directions go almost immediately to zero, meaning that the vector becomes orthogonal to those subspaces. The reason why the convergence is so fast is because we are using Gaussian matrices, to the dominant selection is found extremely fast.

To get a more visual understanding of what is going on, we can see in Figure 6 a two, two-dimensional representation of the directions found through the Oseledets theorem obtained through A.5. We can see (with a bit of noise) how the two directions oscillate as time passes, within two subregions, until they converge into the two main directions that can be seen with the straight lines.

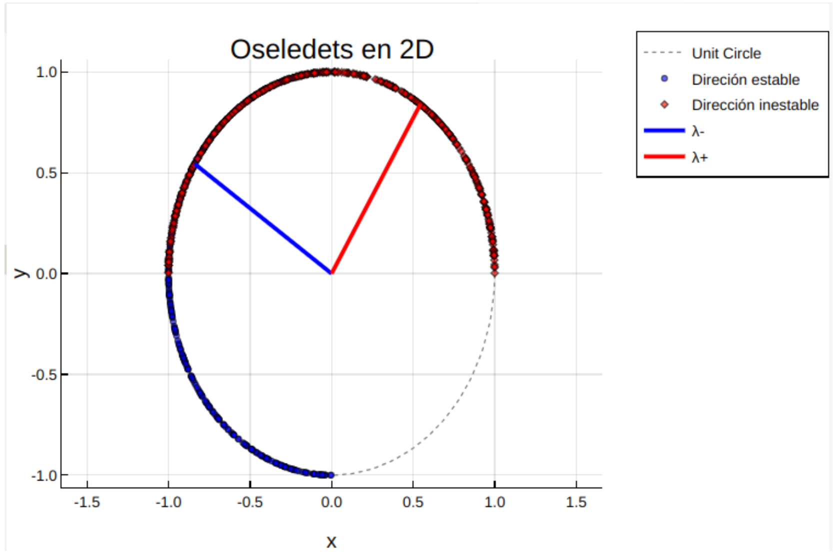

Figure 6: Principal directions through Oseledets theorem

A Appendix

A.1 Top Lyapunov Exponent Approximation

```
using LinearAlgebra, Random, Plots

# Parameters
d = 2                        # Dimension
N = 10_000                   # Number of iterations
Random.seed!(42)             # Reproducibility

# Function to generate random matrix in GL(2)
function random_matrix()
    A = randn(d, d)
    while abs(det(A)) < 1e-3  # Avoid near-singular matrices
        A = randn(d, d)
    end
    return A
end

# Function to simulate Lyapunov exponent
function simulate_lyapunov(d, N)
    P = Matrix{Float64}(I, d, d)  # Identity matrix
    lyap_exp = zeros(N)

    for n in 1:N
        A = random_matrix()
        P *= A
        lyap_exp[n] = log(norm(P)) / n
    end

    return lyap_exp
end

# Run simulation
lyap_exp = simulate_lyapunov(d, N)

# Plotting
plot(1:N, lyap_exp, lw=2, xlabel="n", ylabel="     ", title="Lyapunov Exponent Approximation",
     label="     =    l   o   g   AA    / n", legend=:bottomright)
```

A.2 Histogram of Top Lyapunov Exponents

```
using LinearAlgebra, Random, Plots

d = 2                        # Dimension
N = 10_000                   # Number of iterations
Random.seed!(42)

M = 1000      # Number of trials
n = 500       # Length of each product
lyap_exps = zeros(M)

function random_matrix()
    A = randn(d, d)
    while abs(det(A)) < 1e-3  # Avoid near-singular matrices
```

```
        A = randn(d, d)
    end
    return A
end

for i in 1:M
    P = Matrix{Float64}(I, 2, 2)
    for _ in 1:n
        P *= random_matrix()
    end
    lyap_exps[i] = log(norm(P)) / n
end

# Plot histogram
histogram(lyap_exps, bins=40, title="Distribution of Lyapunov Exponents",
          xlabel="  ", ylabel="Frequency", legend=false)
```

A.3 Main Lyapunov Exponents

```
using LinearAlgebra
using Random
using Plots

function generate_random_matrix(d)
    return randn(d, d)  # standard Gaussian entries
end

function estimate_lyapunov_exponents(d::Int, n::Int)
    Q = Matrix(I, d, d)  # initial orthonormal basis
    log_diag_accum = zeros(d)

    for i in 1:n
        A = generate_random_matrix(d)
        Q, R = qr(A * Q).Q, qr(A * Q).R
        log_diag_accum .+= log.(abs.(diag(R)))  # Accumulate log of
            scaling factors
    end

    return log_diag_accum ./ n  # Approximated Lyapunov exponents
end

# Parameters
d = 3          # dimension
n = 10_000     # number of steps

exponents = estimate_lyapunov_exponents(d, n)
println("Estimated Lyapunov exponents: ", exponents)

function convergence_plot(d::Int, n::Int)
    Q = Matrix(I, d, d)
    log_history = zeros(d, n)

    for i in 1:n
        A = generate_random_matrix(d)
        Q, R = qr(A * Q).Q, qr(A * Q).R
        log_history[:, i] .= log.(abs.(diag(R)))
    end
```

```
    cumulative = cumsum(log_history, dims=2)
    estimates = cumulative ./ reshape(1:n, 1, :)

    plt = plot(xlabel="n", ylabel="Estimated      ", title="Convergence of Lyapunov Exponents")
    for i in 1:d
        plot!(plt, 1:n, estimates[i, :], label=" _$i ")
    end
    display(plt)
end



convergence_plot(d, n)
```

A.4 Oseledets Directions in 3D

```
using LinearAlgebra
using Random
using Plots

function generate_random_matrix(d::Int)
    return randn(d, d)
end

function directional_convergence(d::Int, n::Int)
    Q = Matrix(I, d, d)
    v = randn(d)
    v /= norm(v)

    projection_history = zeros(d, n)

    for i in 1:n
        A = generate_random_matrix(d)
        Q, _ = qr(A * Q)
        v = A * v
        v /= norm(v)

        for j in 1:d
            projection_history[j, i] = abs(dot(v, Q[:, j]))
        end
    end

    # Create plot object
    plt = plot(1:n, projection_history[1, :],
               label="Direction 1",
               xlabel="Iterations",
               ylabel="Normalized projection",
               title="Alignment with Oseledets Directions",
               legend=:topright)

    for j in 2:d
        plot!(plt, 1:n, projection_history[j, :], label="Direction $j")
    end

    display(plt)  # Ensure plot is shown
```

```
end

# Run it
directional_convergence(3, 10000)
```

A.5 3D Representation of the Evolution of Main Directions

```
using LinearAlgebra
using Random
using Plots
plotlyjs()  # Use 3D-capable backend

function generate_random_matrix(d::Int)
    return randn(d, d)
end

function oseledets_3d_visualization(d::Int, n::Int; skip=100)
    Q = Matrix(I, d, d)
    directions_over_time = []

    for i in 1:n
        A = generate_random_matrix(d)
        Q, _ = qr(A * Q)

        # Save every `skip` steps
        if i % skip == 0
            push!(directions_over_time, copy(Q))
        end
    end

    # Extract points for each direction
    dir1 = [Q[:,1] for Q in directions_over_time]
    dir2 = [Q[:,2] for Q in directions_over_time]
    dir3 = [Q[:,3] for Q in directions_over_time]

    # Function to extract x, y, z components
    extract_xyz(vecs) = (getindex.(vecs, 1), getindex.(vecs, 2), getindex
        .(vecs, 3))

    x1, y1, z1 = extract_xyz(dir1)
    x2, y2, z2 = extract_xyz(dir2)
    x3, y3, z3 = extract_xyz(dir3)

    plt = plot(x1, y1, z1, label="Direction 1", lw=2, c=:blue)
    plot!(plt, x2, y2, z2, label="Direction 2", lw=2, c=:red)
    plot!(plt, x3, y3, z3, label="Direction 3", lw=2, c=:green)

    plot!(title="Evolution of Oseledets Directions in 3D",
          xlabel="x", ylabel="y", zlabel="z", legend=:outertopright)

    display(plt)
end

# Run it
oseledets_3d_visualization(3, 5000, skip=20)
```

A.6 3D Representation of Main Directions

```julia
using LinearAlgebra
using Random
using Plots
gr()

function random_matrix(d)
    return randn(d, d)
end

function animate_oseledets(d::Int=3, n::Int=300, step::Int=10)
    Q = Matrix(I, d, d)
    frames = []

    for i in 1:n
        A = random_matrix(d)
        Q, _ = qr(A * Q)

        # Save frame every `step` iterations
        if i % step == 0
            push!(frames, copy(Q))
        end
    end

    anim = @animate for Qframe in frames
        plot3d(; xlim=(-1,1), ylim=(-1,1), zlim=(-1,1), legend=:topright,
               title="Oseledets Directions (iteration frame)", xlabel="X", ylabel="Y", zlabel="Z")

        for i in 1:d
            quiver!([0.0], [0.0], [0.0],
                    quiver=([Qframe[1,i]], [Qframe[2,i]], [Qframe[3,i]]),
                    color=[:blue :red :green][i], label="Direction $i")
        end
    end

    gif(anim, "oseledets_directions.gif", fps=10)
end

animate_oseledets()
```

References

[1] Robert G. Bartle. *The Elements of Integration and Lebesgue Measure*. 1st ed. Wiley Classics Library v.92. Hoboken: John Wiley & Sons, Incorporated, 1995. 1 p. ISBN: 9780471042228 9781118164488.

[2] Walter Rudin. *Real and complex analysis*. 3. ed., internat. ed., [Nachdr.] McGraw-Hill international editions Mathematics series. New York, NY: McGraw-Hill, 2013. 416 pp. ISBN: 978-0-07-100276-9 978-0-07-054234-1.

[3] Donald L. Cohn. *Measure Theory: Second Edition*. Birkhäuser Advanced Texts Basler Lehrbücher. New York, NY: Springer New York, 2013. ISBN: 978-1-4614-6955-1 978-1-4614-6956-8. DOI: 10.1007/978-1-4614-6956-8. URL: https://link.springer.com/10.1007/978-1-4614-6956-8 (visited on 08/14/2024).

[4] Elias M. Stein and Rami Shakarchi. *Real analysis: measure theory, integration, and Hilbert spaces*. Princeton lectures in analysis v. 3. Princeton, N.J: Princeton University Press, 2005. 402 pp. ISBN: 9780691113869.

[5] Manfred Einsiedler and Thomas Ward. *Ergodic Theory: with a view towards Number Theory*. London: Springer London, 2011. ISBN: 978-0-85729-020-5 978-0-85729-021-2. DOI: 10.1007/978-0-85729-021-2. URL: https://link.springer.com/10.1007/978-0-85729-021-2 (visited on 08/14/2024).

[6] Daniel W. Stroock. *Probability theory: an analytic view*. 2nd ed. Cambridge New York: Cambridge University Press, 2011. 1 p. ISBN: 978-0-521-76158-1 978-0-511-97424-3 978-1-139-01188-4.

[7] Sheldon M. Ross. *Introduction to probability models*. Tenth edition. Amsterdam Boston: Academic Press, an imprint of Elsevier, 2010. 1 p. ISBN: 978-0-12-375686-2 978-0-12-375687-9.

[8] Dimitri P. Bertsekas and John N. Tsitsiklis. *Introduction to probability*. 2nd ed. Optimization and computation series. Belmont: Athena scientific, 2008. ISBN: 978-1-886529-23-6.

[9] Patrick Billingsley. *Probability and measure*. 3rd ed. Wiley series in probability and mathematical statistics. New York: Wiley, 1995. 593 pp. ISBN: 978-0-471-00710-4.

[10] Paul E. Pfeiffer. *Probability for Applications*. Springer Texts in Statistics. New York, NY: Springer New York, 1990. 679 pp. ISBN: 978-1-4615-7676-1. DOI: 10.1007/978-1-4615-7676-1.

[11] Queen Mary University of London. *Lecture 2: Second Language Acquisition*. https://qmplus.qmul.ac.uk/pluginfile.php/2470175/mod_resource/content/13/L2-2022.pdf. Accessed: 2025-04-21. 2022.

[12] Alessandro Rinaldo. *Lecture Notes: February 20, 2018*. https://www.stat.cmu.edu/~arinaldo/Teaching/36752/S18/Scribed_Lectures/Feb20.pdf. Scribed lecture notes for 36-752: Advanced Probability Overview, Spring 2018, Carnegie Mellon University. 2018.

[13] David A. Stephens. *The Borel-Cantelli Lemma*. https://www.math.mcgill.ca/dstephens/OldCourses/556-2006/Math556-BorelCantelli.pdf. Lecture notes for Math 556: Mathematical Statistics I, McGill University. 2006.

[14] Jeffrey S. Rosenthal. *A first look at rigorous probability theory*. 2. ed., reprinted. New Jersey: World Scientific, 2010. 219 pp. ISBN: 978-981-270-371-2.

[15] Marcelo Viana. *Lectures on Lyapunov exponents*. Cambridge studies in advanced mathematics 145. Cambridge ; New York: Cambridge University Press, 2014. 202 pp. ISBN: 978-1-107-08173-4.

[16] Peter Walters. *An introduction to ergodic theory*. 1. softcover printing. Graduate texts in mathematics 79. New York: Springer, 2000. 250 pp. ISBN: 9780387951522.

[17] Artur Avila and Jairo Bochi. “On the Subadditive Ergodic Theorem”. In: *Ergodic Theory and Dynamical Systems* 29.1 (Jan. 2009). Publisher: Cambridge University Press,

pp. 1–16. DOI: 10.1017/S014338570800067X. URL: https://doi.org/10.1017/S014338570800067X.

[18] Hillel Furstenberg and Harry Kesten. "Products of Random Matrices". In: *The Annals of Mathematical Statistics* 31.2 (June 1960), pp. 457–469. ISSN: 0003-4851. DOI: 10.1214/aoms/1177705909. URL: http://projecteuclid.org/euclid.aoms/1177705909 (visited on 08/14/2024).

[19] Walter Rudin. *Functional analysis*. 2nd ed. International series in pure and applied mathematics. New York: McGraw-Hill, 1991. 424 pp. ISBN: 9780070542365.

[20] Jean H. Gallier. *Geometric methods and applications: for computer science and engineering*. 2nd ed. Texts in applied mathematics 38. New York: Springer, 2011. 680 pp. ISBN: 9781441999603 9781441999610.

[21] Geon Ho Choe. *Computational ergodic theory*. Algorithms and computation in mathematics v. 13. Berlin: Springer, 2005. 453 pp. ISBN: 978-3-540-23121-9.

[22] Walter Gander. *Algorithms for the QR-Decomposition*. Tech. rep. 80-021. Re-typed in LaTeX in October 2003, In memory of Prof. H. Rutishauser. Zürich, Switzerland: Seminar für Angewandte Mathematik, Eidgenössische Technische Hochschule, 1980. URL: https://people.inf.ethz.ch/gander/papers/qrneu.pdf.