LYAPUNOV EXPONENTS AND THE MULTIPLICATIVE ERGODIC THEOREM

ALINA ZHU

Abstract. This paper gives a proof of the Oseledets multiplicative ergodic theorem for $d \times d$ linear cocycles, and mention some applications of this theorem to products of random matrices and Schrödinger cocycles.

Contents

1. Introduction

This expository paper aims to provide an overview of a classical result from dynamical systems and the theory of Lyapunov exponents. In particular, we focus on proving the Oseledets multiplicative ergodic theorem for $d \times d$ linear cocycles, which establishes the existence of Lyapunov exponents (also called Lyapunov characteristic numbers). Then, at the end of the paper, we discuss some preliminary applications of this theorem to products of random matrices and Schrödinger cocycles.

Throughout this paper, we take $(M, \mathscr{B}, \mu)$ to be a probability space and $f : M \to M$ a **measure preserving** transformation, i.e., a measurable map satisfying

$$\mu(f^{-1}(B)) = \mu(B) \ , \text{ for every } B \in \mathscr{B} \ .$$

We define a **dynamical system** to be the quadruple $(M, \mathscr{B}, \mu, f)$. If A is a $d \times d$ matrix, we take $\|A\|$ to be the standard operator norm

$$\|A\| := \sup_{|v|=1} |Av| \ .$$

The version of multiplicative ergodic theorem proven in this paper is concerned with describing the asymptotic behavior of a product of matrices chosen using an arbitrary dynamical system $(M, \mathscr{B}, \mu, f)$. In particular, if $A : M \to GL(d)$ is a measurable map with values in the general linear group of $d \times d$ matrices with real entries, which also satisfies some integrability assumptions, we show that for $\mu-$almost every $x \in M$, there exist $k = k(x) \in \mathbb{N}$, real numbers $\lambda_1(x) >$

$\cdots > \lambda_k(x)$ (called **Lyapunov exponents**), and a family of decreasing linear subspaces of $\mathbb{R}^d$ $\mathbb{R}^d = V_x^i \supsetneq \cdots \supsetneq V_x^k \supsetneq \{0\}$ (called a **flag**), such that for all $1 \leq i \leq k$, we have

$$\lim_{n\to\infty} \frac{1}{n} \log \left|A(f^{n-1}(x)) \cdots A(f(x))A(x)v\right| = \lambda_i(x) \ , \ \text{for all } v \in V_x^i \backslash V_x^{i+1} \ . \tag{1.1}$$

A full statement of the theorem will be introduced in Section 1.1.

We remark that there are several proofs of the multiplicative ergodic theorem. It was first proven by Oseledets in 1968 [1], based on previous results by Furstenberg and Kesten [2]. A different approach to prove the theorem is due to [6], which is based on exterior algebra and singular value decomposition. Another proof can be found at [7], which uses an approach called the Avalanche Principle. The approach used in this paper is primarily based on the proofs by Viana, Walters, and Bochi (see [3], [4], [5], respectively).

In terms of proof strategy, we will first establish a weaker version of the theorem by replacing the limit in (1.1) with a limsup. Then, we show that the limit exists using an inductive argument on the number of subspaces in the flag $\mathbb{R}^d = V_x^i \supsetneq \cdots \supsetneq V_x^{k(x)} \supsetneq \{0\}$.

1.1. **Motivation.** In this subsection, we briefly describe some motivation for proving the multiplicative ergodic theorem, from the point of view of products of random matrices.

We begin by introducing a concrete model. Let $(p_1, ..., p_m)$ be a probability vector, so that $p_j \geq 0$ and $p_1 + ... + p_m = 1$. Take a subset $\{B_1, ..., B_m\} \subseteq GL(d)$ of the general linear group of $d \times d$ invertible matrices with real entries. We define a sequence of independent, identically distributed random variables $A_0, A_1, ...$ taking values in $\{B_1, ..., B_m\}$, such that for each $i \geq 0$, $1 \leq j \leq m$, one has

$$\mathbb{P}\{A_i = B_j\} = p_j \ . \tag{1.2}$$

For $n \geq 0$, let $A^n := A_{n-1} \cdots A_0$ be the matrix product of the A_i's.

We are interested in studying the limiting behavior of A^n as $n \to \infty$. In 1960, Furstenberg and Kesten showed that with probability 1, there are real numbers $\lambda_\pm$ such that for large N, the operator norms of A^N and $(A^N)^{-1}$ exhibit the exponential growth rates

$$\left\|A^N\right\| \sim \mathrm{e}^{N\lambda_+} \ , \ \left\|\left(A^N\right)^{-1}\right\|^{-1} \sim \mathrm{e}^{N\lambda_-} \ . \tag{1.3}$$

The numbers $\lambda_\pm$ are called the **extremal Lyapunov exponents**, which are defined precisely as the limits

$$\lambda_+ := \lim_{n\to\infty} \frac{1}{n} \log \|A^n\| \ , \ \lambda_- := \lim_{n\to\infty} \frac{1}{n} \log \left\|(A^n)^{-1}\right\|^{-1} \ . \tag{1.4}$$

This result is a special case of their general theorem, which is proven in [2] for all stationary stochastic processes under some integrability assumptions.

Example 1.5. For example, consider the case in which our sequence of random matrices take values in the set

$$\left\{B_1 = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 2 \end{pmatrix} \ , \ B_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\right\} \subseteq GL(2) \ ,$$

with probability given by $(p, 1-p)$, for $p \in [0, 1]$. Then, we have

$$\lambda_+ = \begin{cases} \log 2 & \text{if } p = 1 \\ 0 & \text{if } p \in (0,1) \end{cases} \quad \text{and} \quad \lambda_- = \begin{cases} \log \frac{1}{2} & \text{if } p = 1 \\ 0 & \text{if } p \in (0,1) \ . \end{cases}$$

See [17, Section 4] for more examples.

We see that the growth rates of the operator norms are completely determined by the extremal Lyapunov exponents. By the definition of the operator norm, this implies that for large N, we have the following upper and lower bounds on the length of the vector $A^N v$:

$$\mathrm{e}^{N\lambda_-} |v| \leq \left|A^N v\right| \leq \mathrm{e}^{N\lambda_+} |v| \ , \ \text{for all } v \in \mathbb{R}^d \ .$$

However, the extremal Lyapunov exponents are not sufficient to completely determine the expansion rate of $|A^n v|$ (where v is chosen arbitrarily form $\mathbb{R}^d$), because we can find subspaces of $\mathbb{R}^d$ in which $|A^n v|$ admit different growth rates. In the following, we illustrate how such subspaces arise for powers of one matrix. This can also be interpreted as taking $m = 1$, i.e., a constant sequence of matrices, from the random matrix model in (1.2).

Suppose $B \in GL(d)$ is a normal matrix (i.e., $B^*B = BB^*$, where $B^* := \overline{B}^T$). For $v \in \mathbb{R}^d \backslash \{0\}$, we consider the limit

$$\lim_{n\to\infty} \frac{1}{n} \log |B^n w| \ , \text{ for } w \in \mathbb{R}^d \ . \tag{1.6}$$

It turns out that if $\nu_1, ..., \nu_d$ are the (not necessarily distinct) eigenvalues of B, indexed in a way such that $|\nu_1| \geq ... \geq |\nu_d|$, and $w_1, ..., w_d$ are the distinct, orthonormal eigenvectors of B (which exist because B is normal), then

$$\lim_{n\to\infty} \frac{1}{n} \log |B^n w| = \log |\nu_i| \text{ , for all } w \in \text{span}\{w_i, ..., w_d\} \backslash \text{span}\{w_{i+1}, ..., w_d\} \ .$$

See Section 3 for more details and a proof. It follows that we have

$$|B^n w| \sim |v_i|^n \text{ , for all } w \in \text{span}\{w_i, ..., w_d\} \backslash \text{span}\{w_{i+1}, ..., w_d\} \ .$$

Therefore, there is a decomposition of $\mathbb{R}^d$ according to the eigenspaces of B such that vectors in different subspaces expand at different rates under iterations of B.

If $d > 2$ and B has more than 2 distinct eigenvalues, then there must be some vector subspace of $\mathbb{R}^d$ in which $|B^n w|$ expands at a rate different to those given by the extremal Lyapunov exponents.

Example 1.7. For example, suppose we consider the following normal matrix

$$B = \begin{pmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{pmatrix} \ .$$

B is unitarily diagonizable, with $B = PDP^T$,

$$P = (w_1 \ w_2 \ w_3) = \begin{pmatrix} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{pmatrix} , \ D = \begin{pmatrix} 8 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{pmatrix} \ .$$

For arbitrary $w \in \mathbb{R}^d$, we write $w = a_1 w_1 + a_2 w_2 + a_3 w_3$. If $a_1 \neq 0$, then we have the rate

$$\lim_{n\to\infty} \frac{1}{n} \log |B^n w| = \log 8 \ .$$

But, if $a_1 = 0$ and $a_2 \neq 0$, then the rate is $\log 6$. If $a_1 = a_2 = 0$ and $a_3 \neq 0$, then the rate is $\log 3$. Thus, we see that there are 3 distinct expansion rates.

In the above example, we observed that for powers of 1 matrix, there is a decomposition of $\mathbb{R}^d$ into subspaces that admit different growth rates. A natural question is whether the same is true for nontrival models from products of random matrices, for example, the model introduced in (1.2) for $m > 1$.

It turns out that linear cocycles are useful tools for us to analyze this generalization. Note that if $(M, \mathscr{B}, \mu, f)$ is a dynamical system and $A : M \to GL(d)$ is a measurable map, then a **linear cocycle** defined by A over f is a map $F : M \times \mathbb{R}^d \to M \times \mathbb{R}^d$,

$$F(x, v) := F(f(x), A(x)v) \ .$$

Iterating F on itself yields $F^n(x, v) = (f^n(x), A^n(x)v)$, where $A^n(x)$ denotes the matrix product

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

The advantage of considering linear cocycles is that (a) they allow us to formally define general matrix products (see the second coordinate of F^n), and (b) they allow us to use tools from dynamical systems and ergodic theory to prove that the desired decomposition of $\mathbb{R}^d$ exists. We are thus interested in the following question about linear cocycles:

Question 1.1. Suppose $(M, \mathscr{B}, \mu, f)$ is a dynamical system and $A : M \to GL(d)$ is measurable. Let $F : M \times \mathbb{R}^d \to M \times \mathbb{R}^d$ be a linear cocycle defined by f over A. Suppose we take the second coordinate of $F^n(x, v)$, where $x \in M, v \in \mathbb{R}^d$ are chosen arbitrarily. Can we find a decomposition of $\mathbb{R}^d$ for the following limit that is similar to the trivial example for powers of 1 matrix?

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

This leads us to the statement of the main theorem of this paper:

Theorem 1.8 (Oseledets Multiplicative Ergodic Theorem)**.** *Suppose* $\log^+ \left\|A^{\pm 1}\right\|$ *are integrable with respect to* μ*, where* $\log^+ t := \max\{\log t, 0\}$*. Then, for* $\mu-$*almost every* $x \in M$*, there exist* $k = k(x) \in \mathbb{N}$*, real numbers* $\lambda_1(x) > \cdots > \lambda_k(x)$*, and a flag* $\mathbb{R}^d = V_x^1 \supsetneq \cdots \supsetneq V_x^k \supsetneq \{0\}$ *such that for all* $1 \leq i \leq k$*, the following hold:*

(a) $k(f(x)) = k(x)$*,* $\lambda_i(f(x)) = \lambda_i(x)$*, and* $A(x) \cdot V_x^i = V_{f(x)}^i$ *,*
(b) *the maps* $x \mapsto k(x)$*,* $x \mapsto \lambda_i(x)$*, and* $x \mapsto V_x^i$ *are measurable,*
(c) *one has*

$$\lim_{n\to\infty} \frac{1}{n} \log |A^n(x)v| = \lambda_i(x) \ , \ \textit{for all } v \in V_x^i \backslash V_x^{i+1} \ .$$

The numbers $\lambda_1(x), ..., \lambda_k(x)$ are called the **Lyapunov exponents** of the linear cocycle F at x, and the set consisting of all Lyapunov exponents is called the **Lyapunov spectrum**.

In particular, we may consider all products of independent, identically distributed random matrices as represented by special kinds of linear cocycles (see Example 2.4), for which the underlying dynamical system is a **Bernoulli scheme** (see Example 2.1). Thus, as a consequence of Theorem 1.8, under some integrability assumptions, we can find a decomposition for all products of i.i.d., invertible random matrices. In Section 6, we will discuss this implication in more detail.

1.2. **Structure of the Paper.** In Section 2, we introduce the formal definition of linear cocycles as well as some examples. In Section 3, we state the proof for the trivial example concerning powers of one matrix mentioned in Section 1.1.

In Section 4, we state Kingman's subadditive ergodic theorem (Theorem 4.1). Then, we use it to deduce a version of Furstenberg and Kesten's theorem for linear cocycles (Theorem 4.2) and the Birkhoff ergodic theorem (Theorem 4.7), since these results will be utilized later in the proof of the multiplicative ergodic theorem (Theorem 1.8).

In Section 5, we state the proof of Theorem 1.8. In particular, Subsections 5.1 and 5.2 prove the claim of the theorem when the limit in part (c) is replaced by limsup; Subsections 5.3 and 5.4 discuss two useful lemmas for induction; Subsection 5.5 shows that the limit in part (c) exists via an inductive argument.

In Section 6, we mention some simple applications of the multiplicative ergodic theorem to examples introduced in Section 2.

2. Definition of Linear Cocycles and Examples

In this section, we discuss the basic set up of the proof of the multiplicative ergodic theorem. In particular, we state the definition of linear cocycles, and introduce some examples.

In the rest of this paper, we take $(M, \mathscr{B}, \mu)$ to be a complete separable probability space. Recall that **complete** means that if $U \subseteq B$ and $B \in \mathscr{B}$ with $\mu(B) = 0$, then $U \in \mathscr{B}$. **Separable** means that there exists a countable family $\mathscr{C} \subseteq \mathscr{B}$ such that for any $\epsilon > 0$ and $B \in \mathscr{B}$ there exists $E \in \mathscr{C}$ with $\mu(B\Delta E) < \epsilon$, where $B\Delta E$ denotes the set difference $B\backslash E \cup E\backslash B$.

To define linear cocycles, we first need to define measure preserving transformations. Recall that a transformation $f : M \to M$ is **measure preserving** (also called $\mu-$invariant) if f is measurable and

$$\mu(f^{-1}(B)) = \mu(B) \ , \text{ for all } B \in \mathscr{B} \ ,$$

and that the quadruple $(M, \mathscr{B}, \mu, f)$ is called a **dynamical system**. In the following, we mention two examples of dynamical systems.

Example 2.1 (Bernoulli Scheme)**.** For $m \in \mathbb{N}$, define $X := \{1, ..., m\}$, and let $(p_1, ..., p_m)$ be a probability vector. Consider the $\sigma-$algebra $\mathscr{C}$ defined by the power set of X, and a measure ρ over X defined by

$$\rho(\{i\}) := p_i \ .$$

The space $(X, \mathscr{C}, \rho)$ is a probability space, and we shall use it to construct a Bernoulli scheme.

For each $i \in \mathbb{Z}$, let $(M_i, \mathscr{B}_i, \mu_i) := (X, \mathscr{C}, \rho)$, and consider the countable product

$$(M, \mathscr{B}, \mu) := \prod_{n=-\infty}^{+\infty} (M_i, \mathscr{B}_i, \mu_i) \ .$$

Then, M is called a **shift space** and a point in M is a $X-$valued sequence $\{x_i\}_{i\in\mathbb{Z}}$. The $\sigma-$algebra $\mathscr{B}$ of subsets of M is the countable product of $\mathscr{C}$ with itself, i.e., the smallest $\sigma-$algebra containing all sets (called **measurable rectangles**) of the form

$$R = \{\{x_i\}_{i\in\mathbb{Z}} \mid x_j = a_j,\ |j| \leq n\}, \text{ for some } n \geq 0 \text{ and } a_j \in X\ .$$

The measure μ on M is the product measure defined by

$$\mu(R) := \Pi_{j=-n}^{n}\rho(\{a_j\})\ ,$$

and $(M, \mathscr{B}, \mu)$ is a probability space.

We further define the **shift map** $f : M \to M$ as the transformation

$$f\left(\{x_i\}_{i\in\mathbb{Z}}\right) = \{x_{i+1}\}_{i\in\mathbb{Z}}\ .$$

In particular, f preserves the product measure μ (see, e.g., [13, pp. 20-21]). The dynamical system $(M, \mathscr{B}, \mu, f)$ is called a **two-sided Bernoulli scheme**.

Note that f is also measure preserving if the product is taken over the non-negative integers:

$$(M, \mathscr{B}, \mu) := \prod_{n=0}^{+\infty}(M_i, \mathscr{B}_i, \mu_i)\ ,$$

in which case the dynamical system $(M, \mathscr{B}, \mu, f)$ is called a **one-sided Bernoulli scheme**.

A simple concrete example of a Bernoulli scheme is the fair coin flip, in which case we take $X = \{0, 1\}$ and $p = \left(\frac{1}{2}, \frac{1}{2}\right)$, and consider the the space $M = X^{\mathbb{Z}}$ of sequences of 0's and 1's, endowed with the product measure.

We remark that the above construction may also be generalized by taking the base space $(X, \mathscr{C}, \rho)$ to be any arbitrary probability space, not necessarily discrete.

Example 2.2 (Irrational Rotation)**.** Let $M = \mathbb{T}^d = \mathbb{R}^d\backslash\mathbb{Z}^d$ be the $d-$dimensional torus. Let $\mathscr{B}$ be the Borel $\sigma-$algebra on M, and let μ be the normalized Lebesgue measure on M. Then, $(M, \mathscr{B}, \mu)$ forms a probability space.

Let $\alpha \in \mathbb{T}^d$ be rationally independent (i.e., none of the coordinates of α can be written as a linear combination of the others with rational coefficients). We define the **irrational rotation map** $f_\alpha : M \to M$ as

$$f_\alpha(x) := x + \alpha \mod 1\ .$$

Then, one can prove using properties of the Haar measure that f preserves the measure μ (see, e.g., [13, p.20]). Thus, $(M, \mathscr{B}, \mu, f)$ is a dynamical system.

Now, let's recall the definition of a linear cocycle:

Definition 2.3. Let $(M, \mathscr{B}, \mu, f)$ be a dynamical system and $A : M \to GL(d)$ be a measurable map. The **linear cocycle** defined by A over f is the function $F : M \times R^d \to M \times \mathbb{R}^d$,

$$F(x, v) := (f(x), A(x)v)\ .$$

In the following, we introduce two examples of linear cocycles.

Example 2.4 (Random Transformations)**.** In general, if $(M, \mathscr{B}, \mu, f)$ is a Bernoulli scheme and $A(x)$ depends only on the first coordinate of $x \in M$, then F is called a **random transformation**. In the following, we show that products of independent, identically distributed random matrices are equivalent to special kinds of random transformations.

Suppose $A_0, A_1, ...$ is a sequence of i.i.d., invertible, $d\times d$ random matrices. Since all matrices are identically distributed, all of A_i are formally given by a measurable function from some probability space $(\Omega, \mathscr{F}, \eta)$ to some measurable space $(X, \mathscr{C}) \subseteq GL(d)$. We think of X as a probability space endowed with the distribution measure of A_i, i.e., $\rho = A_{i*}\eta$, and use $(X, \mathscr{C}, \rho)$ to construct a one-sided Bernoulli scheme $(M, \mathscr{B}, \mu, f)$ in Example 2.1.

For instance, for the model in (1.2), we can take $X := \{B_1, ..., B_m\} \subseteq GL(d)$, and consider the probability space defined by X and the measure $\sigma = p_1\delta_{B_1} + ... + p_m\delta_{B_m}$, where $p_i \geq 0$ is taken from a probability vector $(p_1, ..., p_m)$.

Now, let $A : M \to GL(d)$ be the map

$$A\left(\{A_n\}_{n\in\mathbb{Z}_{\geq 0}}\right) = A_0\ ,$$

and let $F : M \times \mathbb{R}^d \to M \times \mathbb{R}^d$ be the linear cocycle defined by A over f (recall that f is the shift map). Then, for a given point $(\{A_n\}_n, v) \in M \times \mathbb{R}^d$, iterating F for k times yields

$$F^k(\{A_n\}_n, v) = \left(f^k(\{A_n\}_n), A^k(f^n(\{A_n\}_n))v\right) = (\{A_{n+k}\}_n, A_{k-1} \cdots A_0 v) \ .$$

We see that the second coordinate looks exactly like a product of random matrices.

Here is how we relate results about the second coordinate of F^k to the original product of i.i.d. random matrices. We may think of a sequence of random matrices as a single random variable taking value in the shift space $M = X^{\mathbb{Z}_{\geq 0}}$, with distribution given by the product measure μ. Then, since the random variables are independent, a property holds for $\mu-$almost every sequences in M if and only if the same property holds for the sequence of random matrices with probability 1.

Example 2.5 (Schrödinger Cocycles)**.** Let $l^2(\mathbb{Z}) := \left\{ u = \{u_n\}_{n \in \mathbb{Z}} : \sum_n |u_n|^2 < \infty \right\}$. Given an $\mathbb{R}-$valued sequence $\{v_n\}_n$, we define the one dimensional Schrödinger operator $H : l^2(\mathbb{Z}) \to l^2(\mathbb{Z})$ associated with the sequence $\{v_n\}_n$ as the following:

$$H(\{u_n\}_n) := \{u_{n-1} + u_{n+1} + v_n u_n\}_n \ .$$

We further suppose that the sequence $\{v_n\}_n$ is generated by a function $V : M \to \mathbb{R}$ and a dynamical system $(M, \mathscr{B}, \mu, f)$ via the composition $v_n := V(f^n(x))$, for some $x \in M$.

In particular, the eigenvalue equation of H, i.e.,

$$Hu = Eu \ , \text{ where } E \in \mathbb{R} \ , \tag{2.6}$$

can be expressed using the trajectory of a 2×2 linear cocycle. In the below, we explain the construction of this linear cocycle. Note that if $u = \{u_n\}_n$ is a solution to (2.6), then for each $n \in \mathbb{Z}$, we can write $u_{n-1} + u_{n+1} + v_n u_n = E u_n$. Thus, equivalently, we have

$$\begin{bmatrix} u_{n+1} \\ u_n \end{bmatrix} = \begin{bmatrix} E - V(f^n(x)) & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} u_n \\ u_{n-1} \end{bmatrix} \ .$$

Therefore, we may define a 2×2 linear cocycle $F_E : M \times \mathbb{R}^2 \to M \times \mathbb{R}^2$ of the function $A_E : M \to GL(2)$ over f, where

$$A_E(y) := \begin{bmatrix} E - V(y) & -1 \\ 1 & 0 \end{bmatrix} \ .$$

It follows that $u = \{u_n\}_n$ is a solution to the eigenvalue equation of H if and only if the trajectory of F_E given by plugging in $v = (u_0, u_1)$ coincides with u.

If the base system $(M, \mathscr{B}, \mu, f)$ is a Bernoulli scheme and $V : M \to \mathbb{R}$ only depends on the first coordinate of $x \in M$, then F_E is called a **random Schrödinger cocycle**. If the base system is an irrational rotation on $M = \mathbb{T}^d$ and $V : M \to \mathbb{R}$ is analytic (i.e., its Taylor series converges in a neighborhood around every point), then F_E is called a **quasi-periodic Schrödinger cocycle**.

3. Trivial Example

In this section, we discuss the trivial case of powers of one matrix mentioned in Section 1.1. We prove the following theorem:

Theorem 3.1. *Suppose $B \in GL(d)$ is normal. Let $\nu_1, ..., \nu_d$ be its eigenvalues satisfying*

$$|\nu_1| \geq ... \geq |\nu_d| \ .$$

Let $w_1, ..., w_d$ be the corresponding distinct orthonormal eigenvectors of B. For $1 \leq i \leq d$, denote $E_i := span\{w_i, ..., w_d\}$. Then, there is a decreasing family of subspaces $\mathbb{R}^d = E_1 \supsetneq \cdots \supsetneq E_d \supsetneq \{0\}$ such that for all $1 \leq i \leq d$,

$$\lim_{n\to\infty} \frac{1}{n} \log |B^n w| = \log |\nu_i| \ , \textit{ for all } w \in E_i \backslash E_{i+1} \ .$$

The following Lemma is fundamental to the proof of Theorem 3.1:

Lemma 3.2. *Let B be as in Theorem 3.1. Then for all $w \in E_i$ such that $|w| = 1$, for all $n \in \mathbb{N}$, one has $|B^n w| \leq |\nu_i|^n$.*

Assuming this lemma, we prove Theorem 3.1:

Proof of Theorem 3.1. Suppose $w \in E_i \backslash E_{i+1}$. Then, we can write w as the orthogonal projection $w = cw_i + \widetilde{w}$, where $c \neq 0$ and $\widetilde{w} \in E_{i+1}$.

If $\widetilde{w} = 0$, then the proof is complete. Thus, without loss of generality, we assume that $\widetilde{w} \neq 0$. It follows that

$$\left| \frac{1}{n} \log |B^n w| - \log |\nu_i| \right| = \frac{1}{n} \left| \log \left| cw_i + \frac{B^n \widetilde{w}}{\nu_i^n} \right| \right| .$$

To show that this difference converges to zero, we need to verify that for all $n \geq 1$, one has

$$0 < \left| cw_i + \frac{B^n \widetilde{w}}{\nu_i^n} \right| < \infty . \tag{3.3}$$

Note that by Lemma 3.2, for all $n \geq 1$ we have

$$\left| \frac{B^n \widetilde{w}}{\nu_i^n} \right| \leq |\widetilde{w}| < \infty,$$

and so it follows that $\left| cw_i + \frac{B^n \widetilde{w}}{\nu_i^n} \right| < \infty$. Moreover, note that since $\widetilde{w} \in E_{i+1}$, we can write

$$\widetilde{w} = c_{i+1} w_{i+1} + ... + c_d w_d .$$

Thus, for any $n \geq 1$, we have $w_i, B^n \widetilde{w} \neq 0$, and $B^n \widetilde{w} \cdot w_i = 0$. Therefore, by positivity of the Euclidean inner product, $B^n \widetilde{w} \neq a w_i$ for any $a \in \mathbb{R}$. It follows that $\left| cw_i + \frac{B^n \widetilde{w}}{\nu_i^n} \right| > 0$.

In conclusion, by (3.3), we have that as $n \to \infty$,

$$\left| \frac{1}{n} \log |B^n w| - \log |\nu_i| \right| = \frac{1}{n} \left| \log \left| cw_i + \frac{B^n \widetilde{w}}{\nu_i^n} \right| \right| \longrightarrow 0 .$$

□

We conclude this section by the proof of Lemma 3.2.

Proof of Lemma 3.2. Suppose $w \in E_i$ and $|w| = 1$. Then for all $j > i$ we have $w \cdot w_j = 0$.

Since B is normal, by the spectral theorem, it is unitarily diagonizable. Thus, we may represent $B = PDP^*$, where $D = diag(\nu_1, ..., \nu_d)$ and P is a unitary matrix whose columns are orthonormal eigenvectors of B. It follows that

$$|Bw|^2 = w^T B^T B w = w^T P |D|^2 \underbrace{P^T w}_{=:y} = y^T |D|^2 y = y_1 |\nu_1|^2 + ... + y_d |\nu_d|^2 .$$

Since P is unitary, it is an isometry of $\mathbb{R}^d$, and so we have $1 = |w| = |y|$. Since $w \cdot w_j = 0$ for all $j > i$ we have $y_1 = ... = y_{i-1} = 0$. It follows that

$$|Bw|^2 = y_i |\nu_i|^2 + ... + y_d |\nu_d|^2 \leq 1 \cdot \max_{i \leq k \leq d} |\nu_k|^2 = |\nu_i|^2 .$$

Note that for all $n \geq 1$, if B is normal then so is any polynomial of B. Moreover, if B has eigenvalue λ then B^n has eigenvalue λ^n. Thus, we may apply the same argument to B^n to obtain

$$|B^n w| \leq |\nu_i|^n .$$

□

4. Extremal Lyapunov Exponents

In this section, we state Kingman's subadditive ergodic Theorem (Theorem 4.1), and use it to deduce a version of Furstenberg and Kesten's theorem (Theorem 4.2) and the Birkhoff ergodic theorem (Theorem 4.7) as corollaries. These results are crucial for the proof of the multiplicative ergodic theorem, in Section 5.

Let $(M, \mathscr{B}, \mu, f)$ be a dynamical system. We write $g \in L^1(\mu)$ if a function g on M is $\mu-$integrable, i.e.,

$$\int_M |g| \ d\mu < +\infty .$$

We call a measurable function $\varphi : M \to [-\infty, +\infty)$ (essentially) $f-$**invariant** if $\varphi(f(x)) = \varphi(x)$ for $\mu-$almost every $x \in M$. Moreover, we call a sequence of measurable functions, $\varphi_n : M \to [-\infty, +\infty)$, $n \geq 1$, **subadditive** relative to f if

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

The sequence is called **super-additive** if

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

Theorem 4.1 (Kingman's Subadditive Ergodic Theorem)**.** *Let $\varphi_n : M \to [-\infty, +\infty)$, $n \geq 1$ be a subadditive sequence of measurable functions such that the positive part $\varphi_1^+ \in L^1(\mu)$. Then, the sequence $\{\varphi_n/n\}_{n\in\mathbb{N}}$ converges $\mu-$almost everywhere to some $f-$invariant function $\varphi : M \to [-\infty, +\infty)$. Moreover, the positive part φ^+ of φ is integrable and*

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

Proof. A proof can be found in [3, Theorem 3.3]. □

Now, we use Theorem 4.1 to obtain a version of Furstenberg and Kesten's Theorem for $d \times d$ linear cocycles. Let $(M, \mathscr{B}, \mu, f)$ be a dynamical system and $A : M \to GL(d)$ be measurable. We take F to be the linear cocycle defined by A over f.

Theorem 4.2 (Furstenberg and Kesten's Theorem)**.** *Suppose $\log^+ \left\| A^{\pm 1} \right\|$ are integrable with respect to μ, where $\log^+ t := \max\{\log t, 0\}$. Then, the limits*

$$\lambda_+(x) := \lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)\| \ , \lambda_-(x) := \lim_{n\to\infty} \frac{1}{n} \log \left\| (A^n(x))^{-1} \right\|^{-1}$$

exist for $\mu-$almost every $x \in M$. Moreover, the functions $\lambda_\pm(x)$ are $\mu-$integrable and $f-$invariant, with

$$\int \lambda_+ \ \mathrm{d}\mu = \lim_{n\to\infty} \frac{1}{n} \int \log \|A^n(x)\| \ \mathrm{d}\mu \ , \ \int \lambda_+ \ \mathrm{d}\mu = \lim_{n\to\infty} \frac{1}{n} \int \log \left\| (A^n(x))^{-1} \right\|^{-1} \ \mathrm{d}\mu \ . \tag{4.3}$$

In particular, $\lambda_\pm(x)$ are called the **extremal Lyapunov exponents** of F at x.

Proof. Given a function $A : M \to GL(d)$, since the operator norm $\|\cdot\|$ is sub-multiplicative (i.e., $\|XY\| \leq \|X\| \|Y\|$), the sequences $\{\varphi_n\}_{n\geq 1}$, $\{\psi\}_{n\geq 1}$ defined by

$$\begin{aligned} \varphi_n(x) &:= \log \|A^n(x)\| = \log \left\| A(f^{n-1}(x)) \cdots A(x) \right\| \ , \\ \psi_n(x) &:= \log \left\| (A^n(x))^{-1} \right\|^{-1} = -\log \left\| (A(f^{n-1}(x)) \cdots A(x))^{-1} \right\| \ . \end{aligned}$$

are subadditive and super-additive respectively.

We first treat $\{\varphi_n\}_n$. By the assumption $\log^+ \left\| A^{\pm 1} \right\| \in L^1(\mu)$, we know that $\varphi_1^+ \in L^1(\mu)$, and that for all $n \geq 1$, $\varphi_n \in [-\infty, +\infty)$ for $\mu-$almost every x. Therefore, by Theorem 4.1, the limit

$$\varphi(x) := \lim_{n\to\infty} \frac{1}{n} \varphi_n(x) \in [-\infty, +\infty)$$

exists for $\mu-$almost every $x \in M$, the positive part φ^+ is $\mu-$integrable, and

$$\int \varphi \ \mathrm{d}\mu = \lim_{n\to\infty} \frac{1}{n} \int \log \|A^n(x)\| \ \mathrm{d}\mu \in [-\infty, \infty) \ . \tag{4.4}$$

We now show that $\varphi(\cdot) \in L^1(\mu)$. Since f is measure preserving and $\|B\| \geq \left\| B^{-1} \right\|^{-1}$ for every invertible matrix B, for any n we have

$$\begin{aligned} \frac{1}{n} \int \log \|A^n(x)\| \ \mathrm{d}\mu &\geq \frac{1}{n} \int \log \left\| (A^n(x))^{-1} \right\|^{-1} \ \mathrm{d}\mu \\ &\geq \frac{1}{n} \int \log \left\| (A(x))^{-1} \right\|^{-1} \cdots \left\| (A(f^{n-1}(x)))^{-1} \right\|^{-1} \ \mathrm{d}\mu \\ &= \int \log \left\| (A(x))^{-1} \right\|^{-1} \ \mathrm{d}\mu \ , \end{aligned}$$

so by integrability of $\log^+ \|A^\pm(\cdot)\|$ and (4.4), we have

$$-\infty < \int \varphi \ \mathrm{d}\mu < +\infty \ . \tag{4.5}$$

Thus, (4.5) and the fact that $\varphi^+ \in L^1(\mu)$ implies $\varphi^- \in L^1(\mu)$. Finally, we have

$$\int |\varphi| \ \mathrm{d}\mu = \int \varphi^+ \ \mathrm{d}\mu + \int \varphi^- \ \mathrm{d}\mu < \infty \ ,$$

which proves that $\lambda_+(\cdot) = \varphi(\cdot) \in L^1(\mu)$. We can use a similar argument for λ_- and ψ.

□

Remark 4.6. Note that $\mu-$integrability of $\lambda_{\pm}(\cdot)$ implies that the functions $\lambda_{\pm}(\cdot)$ are finite $\mu-$almost everywhere. This property is important for the proof of the multiplicative ergodic theorem (particularly for Lemma 5.3 below).

Theorem 4.7 (Birkhoff Ergodic Theorem)**.** *Let $\varphi : M \to \mathbb{R}$ be a $\mu-$integrable function. Then the limit*

$$\varphi^*(x) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \varphi(f^j(x))$$

exists for $\mu-$almost every x. Moreover, $\varphi^*(\cdot)$ is $f-$invariant and $\mu-$integrable with

$$\int \varphi^* \, \mathrm{d}\mu = \int \varphi \, \mathrm{d}\mu \ .$$

Proof. Consider the orbital sum

$$\varphi_n := \sum_{j=0}^{n-1} \varphi \circ f^j \ .$$

Observe that for every $x \in M$, this function satisfies

$$\varphi_{n+m}(x) = \sum_{j=0}^{n+m-1} \varphi \circ f^j = \varphi_m(x) + \varphi_n(f^m(x)) \ .$$

Thus, the conclusion of the theorem follows directly from Theorem 4.1. □

We conclude this section by stating a useful (particularly for the proof of Lemma 5.19 below) corollary of the Birkhoff ergodic theorem (Theorem 4.7):

Corollary 4.8. *Let $\varphi : M \to \mathbb{R}$ be a measurable function such that $\psi := \varphi\circ f - \varphi$ is $\mu-$integrable. Then, for $\mu-$almost every $x \in M$,*

$$\lim_{n\to\infty} \frac{1}{n}\varphi(f^n(x)) = 0 \ .$$

Proof. A proof can be found in [3, Corollary 3.11]. □

5. Oseledets Multiplicative Ergodic Theorem

In this section, we prove the multiplicative ergodic theorem. For an arbitrary dynamical system $(M, \mathscr{B}, \mu, f)$, we consider the linear cocycle $F : M \times \mathbb{R}^d \to M \times \mathbb{R}^d$ defined by a measurable map $A : M \to GL(d)$ over $f : M \to M$.

Recall that a **flag** in $\mathbb{R}^d$ is a decreasing family $R^d = V^1 \supsetneq \cdots \supsetneq V^k \supsetneq \{0\}$ of linear subspaces of the $d-$dimensional Euclidean space. For instance, the containment $\mathbb{R}^d = E_1 \supsetneq E_2 \supsetneq \cdots \supsetneq E_d \supsetneq \{0\}$ in Section 3 is an example of a flag.

Theorem 5.1 (Oseledets Multiplicative Ergodic Theorem)**.** *Suppose $\log^+ \left\|A^{\pm 1}\right\|$ are integrable with respect to μ, where $\log^+ t := \max\{\log t, 0\}$. Then, for $\mu-$almost every $x \in M$, there exist $k = k(x) \in \mathbb{N}$, real numbers $\lambda_1(x) > \cdots > \lambda_k(x)$, and a flag $\mathbb{R}^d = V_x^1 \supsetneq \cdots \supsetneq V_x^k \supsetneq \{0\}$ such that for all $1 \le i \le k$, the following hold:*

(a) *$k(f(x)) = k(x)$, $\lambda_i(f(x)) = \lambda_i(x)$, and $A(x)\cdot V_x^i = V_{f(x)}^i$,*
(b) *the maps $x \mapsto k(x)$, $x \mapsto \lambda_i(x)$, and $x \mapsto V_x^i$ are measurable,*
(c) *one has*

$$\lim_{n\to\infty} \frac{1}{n}\log|A^n(x)v| = \lambda_i(x) \ , \ \textit{for all } v \in V_x^i \backslash V_x^{i+1} \ .$$

To prove this theorem, we first replace the limit in item (c) by a limsup, i.e.,

(c)′ one has

$$\limsup_{n\to\infty} \frac{1}{n}\log|A^n(x)v| = \lambda_i(x) \ , \ \text{for all } v \in V_x^i \backslash V_x^{i+1} \ ,$$

and find functions $k(x)$, V_x^i, and $\lambda_i(x)$ that satisfy the claims of parts (a), (b), and (c)$'$. We then show that the limsup is actually a limit (part (c)), via an inductive argument on the number of subspaces in the flag $\mathbb{R}^d = V_x^1 \supsetneq \cdots \supsetneq V_x^k \supsetneq \{0\}$.

In Subsection 5.1, we find the functions $k(x)$, V_x^i, and $\lambda_i(x)$ that satisfy parts (a) and (c)$'$. In Subsection 5.2 we show that these functions are also measurable (part (b)). Finally, in Subsection 5.3 and 5.4, we prove two important lemmas to prepare for the induction. In Subsection 5.5, we use an inductive argument on $k = k(x)$ to show that the limit in part (c) exists.

5.1. **Existence and Invariance of $k(x)$, V_x^i, and $\lambda_i(x)$.** In this subsection, we replace the limit by limsup and show existence of measurable functions $k(x)$, V_x^i, and $\lambda_i(x)$ for the limsup (part(c)$'$). We then check that the functions are invariant as claimed in part (a).

For the ease of notation, in the rest of this paper, we denote

$$\lambda(x,v) := \limsup_{n\to\infty} \frac{1}{n} \log |A^n(x)v| \ .$$

We take $\log 0 := -\infty$, which means $\lambda(x,0) = -\infty$.

Proposition 5.2. *For $\mu-$almost every $x \in M$, there exists $k = k(x) \in \mathbb{N}$, numbers $\lambda_1(x) > \cdots > \lambda_k(x)$, and a flag $R^d = V_x^1 \supsetneq \cdots \supsetneq V_x^k \supsetneq \{0\}$ such that for all $1 \le i \le k$, parts (c)$'$ and (a) hold.*

To prove Proposition 5.2, we will use the following lemmas.

Lemma 5.3. *For $\mu-$almost every $x \in M$ and any $v, v' \in \mathbb{R}^d \backslash \{0\}$, one has*

(i) *$\lambda(x,v)$ is well-defined and finite, and $\lambda_-(x) \le \lambda(x,v) \le \lambda_+(x)$;*
(ii) *$\lambda(x,cv) = \lambda(x,v)$ for all $c \neq 0$;*
(iii) *$\lambda(x,v+v') \le \max\{\lambda(x,v), \lambda(x,v')\}$;*
(iv) *$\lambda(x,v) = \lambda(f(x), A(x)v)$.*

Lemma 5.4. *Suppose a function $g : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ satisfies the following assumptions for all $v, w \in \mathbb{R}^d$:*

(i) *$g(v+w) \le \max\{\lambda(v), \lambda(w)\}$;*
(ii) *$g(cv) = \lambda(v)$, for all $c \neq 0$;*
(iii) *$g(0) = -\infty$.*

Then, the following hold:

(a) *if $g(v) \neq \lambda(w)$ for some $v, w \in \mathbb{R}^d$, then $g(v+w) = \max\{g(v), g(w)\}$;*
(b) *if $g(v_1), ..., g(v_m)$ are distinct for some $v_1, ..., v_m \in \mathbb{R}^d \backslash \{0\}$, then $v_1, ..., v_m$ are linearly independent;*
(c) *g attains at most d distinct finite values.*

Lemma 5.5. *If $\{a_n\}_n, \{b_n\}_n$ are sequences such that $a_n, b_n > 0$, then one has*

$$(a)\ \limsup_{n\to\infty} \frac{1}{n} \log(a_n + b_n) = \max\left\{ \limsup_{n\to\infty} \frac{1}{n} \log a_n,\ \limsup_{n\to\infty} \frac{1}{n} \log b_n \right\} \ ;$$

$$(b)\ \limsup_{n\to\infty} \frac{1}{n} \log \sqrt{a_n^2 + b_n^2} = \max\left\{ \limsup_{n\to\infty} \frac{1}{n} \log a_n,\ \limsup_{n\to\infty} \frac{1}{n} \log b_n \right\} \ ;$$

$$(c)\ \liminf_{n\to\infty} \frac{1}{n} \log(a_n + b_n) \ge \max\left\{ \liminf_{n\to\infty} \frac{1}{n} \log a_n,\ \liminf_{n\to\infty} \frac{1}{n} \log b_n \right\} \ ;$$

$$(d)\ \liminf_{n\to\infty} \frac{1}{n} \log \sqrt{a_n^2 + b_n^2} \ge \max\left\{ \liminf_{n\to\infty} \frac{1}{n} \log a_n,\ \liminf_{n\to\infty} \frac{1}{n} \log b_n \right\} \ .$$

Assuming these lemmas, we prove Proposition 5.2.

Proof of Proposition 5.2. To find the desired functions, we first show that for $\mu-$almost every $x \in M$, the set

$$K_x := \left\{ \lambda(x,v) \mid v \in \mathbb{R}^d \backslash \{0\} \right\}$$

contains only finitely many elements, all of which are finite.

By Lemma 5.3 (i), we know that for $\mu-$almost every x, the limit $\lambda(x,v)$ exists and is finite for every $v \in \mathbb{R}^d \backslash \{0\}$. Moreover, recall that $\lambda(x,0) = -\infty$. Thus, combining this fact with Lemma 5.3 (ii) and (iii), we see that the function $\lambda(x,\cdot) : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ is well-defined for $\mu-$almost every x and satisfies all three of the hypotheses of Lemma 5.4. Therefore, it follows from Lemma 5.4 (c) that $K_x \subseteq \mathbb{R}$ and K_x contains at most d distinct elements.

Next, we proceed to construct the Oseledets' flag. Let $k = k(x)$ be the number of distinct elements in K_x, and denote by $\lambda_i(x) > ... > \lambda_k(x)$ those elements. For $1 \le i \le k$, we define

$$V_x^i := \left\{v \in \mathbb{R}^d \backslash \{0\} \mid \lambda(x,v) \le \lambda_i(x)\right\} \cup \{0\} \ .$$

Then, by Lemma 5.3 (ii) and (iii), we see that V_x^i forms a vector subspace of $\mathbb{R}^d$ for each i, and by construction of V_x^i, we have a flag $\mathbb{R}^d = V_x^1 \supsetneq \cdots \supsetneq V_x^k \supsetneq \{0\}$.

In particular, if $v \in V_x^i \backslash V_x^{i+1}$, then one has

$$\lambda_{i+1}(x) < \lambda(x,v) \le \lambda_i(x) \ .$$

If $\lambda(x,v) \neq \lambda_i(x)$, then K_x contains $k+1$ distinct elements. Contradiction. Thus, this forces

$$\lambda(x,v) = \lambda_i(x) \ , \text{ for all } v \in V_x^i \backslash V_x^{i+1} \ ,$$

which proves part (c)$'$ in the proposition.

We proceed to show that the functions $x \mapsto k(x)$, $x \mapsto \lambda_i(x)$, and $x \mapsto V_x^i$ are $f-$invariant (part (a)). Note that by Lemma 5.3 (iv) we know that for almost every $x \in M$ and all $v \in \mathbb{R}^d \backslash \{0\}$, $\lambda(x,v) = \lambda(f(x), A(x)v)$. Therefore, since $A(x)$ is an invertible linear transformation of $\mathbb{R}^d$, we have

$$\begin{aligned} K_x &= \left\{\lambda(x,v) \mid v \in \mathbb{R}^d \backslash \{0\}\right\} \\ &= \left\{\lambda(f(x), A(x)v) \mid v \in \mathbb{R}^d \backslash \{0\}\right\} = \left\{\lambda(f(x), w) \mid w \in \mathbb{R}^d \backslash \{0\}\right\} = K_{f(x)} \ . \end{aligned}$$

It follows that $k(f(x)) = k(x)$, $\lambda_i(f(x)) = \lambda_i(x)$, and so by construction of V_x^i,

$$\begin{aligned} A(x)V_x^i &= \left\{A(x)v \mid \lambda(x,v) \le \lambda_i(x),\ v \in \mathbb{R}^d \backslash \{0\}\right\} \cup \{0\} \\ &= \left\{w \in \mathbb{R}^d \backslash \{0\} \mid \lambda\left(x, A(x)^{-1}w\right) \le \lambda_i(f(x))\right\} \cup \{0\} \\ &= \left\{w \in \mathbb{R}^d \backslash \{0\} \mid \lambda\left(f(x), w\right) \le \lambda_i(f(x))\right\} \cup \{0\} \\ &= V_{f(x)}^i \ . \end{aligned}$$

□

We now proceed to prove the mentioned lemmas.

Proof of Lemma 5.3. To verify part (i), we begin with finding a bound for the Euclidean norm $|A^n(x)v|$. Note that for all $v \in \mathbb{R}^d \backslash \{0\}$, we have

$$\frac{|A^n(x)v|}{|v|} \le \sup_{|v|=1} |A^n(x)v| = \|A^n(x)\| \ .$$

If we take $A^n(x)v = w$, then we also have

$$\frac{|A^n(x)v|}{|v|} = \frac{|w|}{|(A^n(x))^{-1}w|} \ge \left\|(A^n(x))^{-1}\right\|^{-1} \ .$$

Thus, for all $x \in M$ and $v \in \mathbb{R}^d \backslash \{0\}$, we obtain

$$\left\|(A^n(x))^{-1}\right\|^{-1} |v| \le |A^n(x)v| \le \|A^n(x)\| \, |v| \ .$$

By Theorem 4.2, since $\log^+ \|(A(\cdot))^{\pm}\| \in L^1(\mu)$, we know that the extremal Lyapunov exponents $\lambda_-(x)$ and $\lambda_+(x)$ exist and are real-valued for $\mu-$almost every $x \in M$. Therefore, this forces $\lambda(x,v)$ to exist, and so

$$-\infty < \lambda_-(x) \le \lambda(x,v) \le \lambda_+(x) < +\infty \ ,$$

for $\mu-$almost every $x \in M$ and $v \in \mathbb{R}^d \backslash \{0\}$.

For (ii), observe that if $v \neq 0$, then for any $c \neq 0$, we have

$$\begin{aligned} \lambda(x, cv) &= \limsup_{n\ to\infty} \frac{1}{n} \log |A^n(x)cv| \\ &= \limsup_{n\to\infty} \frac{1}{n} \log |A^n(x)v| + \underbrace{\limsup_{n\to\infty} \frac{1}{n} \log |c|}_{=0} = \lambda(x,v) \ . \end{aligned}$$

For (iii), observe that if $v+v' \neq 0$, then by part (i), the limit $\lambda(x, v+v')$ exists and is finite for almost every $x \in M$. By Lemma 5.5, we have

$$\begin{aligned}\lambda(x, v+v') &= \limsup_{n\to\infty} \frac{1}{n} \log |A^n(x)(v+v')| \\ &\leq \limsup_{n\to\infty} \frac{1}{n} \log\left(|A^n(x)v| + |A^n(x)v'|\right) \\ &= \max\left\{\limsup_{n\to\infty} \frac{1}{n} \log\left(|A^n(x)v|\right) \ , \limsup_{n\to\infty} \frac{1}{n} \log\left(|A^n(x)v'|\right)\right\} \\ &= \max\left\{\lambda(x,v),\ \lambda(x,v')\right\}\ .\end{aligned}$$

To prove (iv), note that by construction, $A^n(x) = A(f^{n-1}(x)) \cdots A(x)$. Therefore,

$$\begin{aligned}\lambda(f(x), A(x)v) &= \limsup_{n\to\infty} \frac{1}{n} \log |A^n(f(x))A(x)v| \\ &= \limsup_{n\to\infty} \frac{1}{n} \log \left|A^{n+1}(x)v\right| \\ &= \limsup_{n\to\infty} \underbrace{\frac{n+1}{n}}_{\sim 1} \frac{1}{n+1} \log\left|A^{n+1}(x)v\right| = \lambda(x,v)\ .\end{aligned}$$

□

Proof of Lemma 5.4. The proof of this lemma is based on [8, Theorem 2.1.2].

We begin with proving part (a). Let $v, w \in \mathbb{R}^d$. Without loss of generality, we assume that $g(v) < g(w)$. Then, by part (i) of the hypothesis,

$$\begin{aligned}g(v+w) &\leq \max\left\{g(v),\ g(w)\right\} = g(w) \\ &= g(v+w-v) \leq \max\left\{g(v+w),\ g(v)\right\}\ .\end{aligned} \tag{5.6}$$

If $g(v+w) < g(v)$, then we will have

$$g(w) = g(w+v-v) \leq \max\{g(v+w),\ g(v)\} = g(v)\ ,$$

contradicting our assumption that $g(v) < g(w)$. Thus, we must have $g(v+w) \geq g(v)$, but this implies

$$g(w) \leq \max\left\{g(v+w),\ g(v)\right\} = g(v+w)\ .$$

Combing with (5.6), we have $g(v+w) = g(w)$, which proves (a).

Now we proceed to prove (b). By contradiction, suppose $g(a_1), ..., g(a_m)$ are distinct and $a_1, ..., a_m \in \mathbb{R}\backslash\{0\}$ are linearly dependent. Then there exists $c_1, ..., c_m \in \mathbb{R}$, not all zero, such that

$$c_1 a_1 + ... + c_m a_m = 0\ .$$

Since $g(a_1), ..., g(a_m)$ are distinct, they cannot all take on the value $-\infty$, so we have

$$\begin{aligned}-\infty = g(0) &= g(c_1 a_1 + ... + c_m a_m) \\ &= \max\left\{g(c_i a_i) \mid 1 \leq i \leq m\right\} \\ &= \max\left\{g(a_i) \mid 1 \leq i \leq m,\ c_i \neq 0\right\} \neq -\infty\ .\end{aligned}$$

Contradiction.

For (c), observe that if g attains more than d distinct values on $\mathbb{R}^d\backslash\{0\}$, then there will be more than d linearly independent vectors in $\mathbb{R}^d$. Thus, this forces g to attain at most d distinct finite values.

□

Proof of Lemma 5.5. Suppose we set $a'_n := \max\{a_n, b_n,\}$, $b'_n := \min\{a_n, b_n\}$. Then $b'_n/a'_n \leq 1$ and for all $n \geq 1$, one has

$$\log\left(a_n + b_n\right) = \log\left(a'_n + b'_n\right) = \underbrace{\log\left(1 + \frac{b'_n}{a'_n}\right)}_{\leq \log 2} + \log a'_n\ .$$

Therefore, we obtain

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log\left(a_n + b_n\right) &= \underbrace{\limsup_{n\to\infty} \frac{1}{n} \log\left(1 + \frac{b'_n}{a'_n}\right)}_{=0} + \limsup_{n\to\infty} \frac{1}{n} \log a'_n \\
&= \limsup_{n\to\infty} \frac{1}{n} \log\left(\max\{a_n, b_n, \}\right) \\
&= \max\left\{\limsup_{n\to\infty} \frac{1}{n} \log a_n,\ \limsup_{n\to\infty} \frac{1}{n} \log b_n\right\} \ .
\end{aligned}$$

A similar argument can be used to prove the case for the liminf and $\sqrt{a_n^2 + b_n^2}$.

□

5.2. **Measurability.** In this subsection, we show that the functions $x \mapsto k(x)$, $x \mapsto \lambda_i(x)$, and $x \mapsto V_x^i$ are measurable (part (b) of Theorem 5.1). Note that for each $x \in M$, V_x^i is a linear subspace of $\mathbb{R}^d$. Thus, before proving part (b), we first give a brief characterization of how measurability is defined for set and space valued functions.

Suppose (Y, d) is a complete separable metric space. We use $\mathcal{H}_c(Y)$ to denote the space consisting of all non-empty compact subsets of Y, endowed with the Hausdorff metric $d_{\text{Hausdorff}}$: for $A, B \in \mathcal{H}_c(Y)$,

$$d_{\text{Hausdorff}}(A, B) := \max\left\{\sup_{a\in A} d(a, B)\ ,\ \sup_{b\in B} d(b, A)\right\}\ ,$$

where

$$d(a, B) = \inf_{b\in B} d(a, b)$$

denotes the distance between the point a and the set B. The metric $d_{\text{Hausdorff}}$ induces a topology and thus a Borel $\sigma-$algebra $\mathscr{B}(\mathcal{H}_c(Y))$ on $\mathcal{H}_c(Y)$. See [10, Chapter 11] or [9, Chapter 18] for details about the construction of this topology. A set-valued function $g : M \to \mathcal{H}_c(Y)$ is measurable if $g^{-1}(B) \in \mathscr{B}$ for all $B \in \mathscr{B}(\mathcal{H}_c(Y))$.

Let $Gr(d)$ be the **Grassmannian** of $\mathbb{R}^d$, i.e., the disjoint union of all Grassmannian manifolds $Gr(l, d)$, $1 \le l \le d$. Each of $Gr(l, d)$ is the collection of all $l-$dimensional linear subspaces of $\mathbb{R}^d$. We define a metric d_{Grass} on $Gr(d)$ via

$$d_{\text{Grass}}(V, W) := d_{\text{Hausdorff}}(S^{d-1} \cap V, S^{d-1} \cap W)\ ,$$

where S^{d-1} is the unit sphere in $\mathbb{R}^d$ and $V, W \in Gr(d)$. The metric d_{Grass} also induces a Borel $\sigma-$algebra on $Gr(d)$, and the measurability of a function $g : M \to Gr(d)$ is defined with respect to this $\sigma-$algebra.

The following Lemmas are used to characterize the measurability of functions with values in $\mathcal{H}_c(Y)$ and $Gr(d)$.

Lemma 5.7. *Let $(M, \mathscr{B}, \mu)$ be a complete probability space, and (Y, d) be a complete separable metric space. Denote by $\mathscr{B}(Y)$ the Borel $\sigma-$algebra of Y and $\mathcal{H}_c(Y)$ the collection of compact subsets of Y. Then, the following statements are equivalent:*

(i) *A map $g : M \to \mathcal{H}_c(Y)$, $g(x) = K_x$ is measurable.*
(ii) *The graph of g, i.e., the set $\{(x, y) \in M \times Y : y \in K_x\}$ is in the product $\sigma-$algebra $\mathscr{B} \times \mathscr{B}(Y)$ on $M \times Y$.*
(iii) *$\{x \in M \mid K_x \cap U \ne \emptyset\} \in \mathscr{B}$ for any open set $U \subseteq \mathbb{R}^d$.*

*Moreover, any of these conditions implies that there is a measurable map $\sigma : M \to \mathbb{R}^d$ (called a **measurable selection**) such that $\sigma(x) \in K_x$ for every $x \in M$.*

Proof. See a proof of (ii)$\Longleftrightarrow$(iii) in [10, Theorem III.30]. See a proof of $(i)$$\Longleftrightarrow$(iii) in [9, Theorem 19.2], and a proof of the measurable selection in [9, Theorem 19.6]. □

Lemma 5.8. *Let $(M, \mathscr{B}, \mu)$ be a complete probability space. Denote by $\mathscr{B}(\mathbb{R}^d)$ the Borel $\sigma-$algebra of $\mathbb{R}^d$ and $Gr(d)$ the Grassmannian of $\mathbb{R}^d$. Then, the following statements are equivalent:*

(i) *The map $g : M \to Gr(d)$, $g(x) = V_x$ is measurable.*
(ii) *The graph of g, i.e., the set $\{(x, y) \in M \times \mathbb{R}^d : y \in V_x\}$ is in the product $\sigma-$algebra $\mathscr{B} \times \mathscr{B}(\mathbb{R}^d)$ on $M \times \mathbb{R}^d$.*

(iii) *For each* $1 \le l \le d$, *the set* $M_l := \{x \in M \mid \dim V_x = l\}$ *is measurable and there exist measurable vector fields* $v_i, : M_l \to \mathbb{R}^d$, $1 \le i \le l$, *such that* $\{v_1(x), ..., v_l(x)\}$ *is a basis of* V_x *for every* $x \in M_l$.

Proof. See a proof of this lemma in [4, Theorem 7]. □

We further mention a useful fact about the canonical projection map $\pi : M \times Y \to M$:

Fact 5.9. *Let* $\mathscr{B} \times \mathscr{B}(Y)$ *be the product* $\sigma-$*algebra on* $M \times Y$, *and* $\pi : M \times Y \to M$ *be the projection map* $\pi(x, y) := x$. *Then, one has* $\pi(E) \in \mathscr{B}$ *for every* $E \in \mathscr{B} \times \mathscr{B}(Y)$.

Proof. See a proof of this fact in [10, Theorem III.23]. □

Assuming the above results, we prove the following proposition:

Proposition 5.10. *The functions* $x \mapsto k(x)$, $x \mapsto \lambda_i(x)$ *and* $x \mapsto V_x^i$ *found in Proposition 5.2 are measurable.*

Proof. We use a recursive argument. First note that it is trivial that the set

$$k^{-1}\left((-\infty, 1]\right) = \{x \in M \mid k(x) \ge 1\} = M$$

is measurable. Moreover, note that for almost all $x \in M$, we have $V_x^1 = \mathbb{R}^d$, so for each $1 \le l \le d$,

$$\left\{x \in M \mid \dim V_x^1 = l\right\}$$

is a measurable set and any arbitrary basis $\{e_1, ..., e_d\}$ of $\mathbb{R}^d$ is a basis for $V_x^1 = \mathbb{R}^d$. Therefore, by Lemma 5.8, we know that the map $x \mapsto V_x^1$ is measurable.

Now, let $\{e_1, ..., e_d\}$ be an arbitrary basis of $V_x^1 = \mathbb{R}^d$. Then, at least one of e_i is in the set $V_x^1 \backslash V_x^2$, since V_x^2 is of a dimension strictly lower by the construction of the Oseledets flag. Thus, we have

$$\max\{\lambda(x, e_i) \mid 1 \le i \le d\} = \lambda_1(x) \ .$$

Since $(x, v) \mapsto \lambda(x, v)$ is measurable, we have that $x \mapsto \lambda_1(x)$ is a measurable function on M.

Thus, we showed that $k^{-1}\left((-\infty, 1]\right)$ is measurable and, $x \mapsto \lambda_1(x)$, $x \mapsto V_x^1$ are measurable functions on M. Next, we show that the same holds for $i = 2$. Observe that since $\lambda(x, v), \lambda_1(x)$ are measurable, the function $g(x, v) := \lambda(x, v) - \lambda_1(x)$ is measurable. Thus, the set

$$\begin{aligned} V_*^2 :&= \left\{(x, v) \in M \times \mathbb{R}^d \backslash \{0\} \mid \lambda(x, v) < \lambda_1(x)\right\} \\ &= \left\{(x, v) \in M \times \mathbb{R}^d \backslash \{0\} \mid g(x, v) < 0\right\} \\ &= g^{-1}((-\infty, 0)) \end{aligned}$$

is a measurable subset of $M \times \mathbb{R}^d$. By Fact 5.9, we see that the set

$$\begin{aligned} M \supseteq \pi\left\{V_*^2\right\} &= \left\{x \in M \mid \lambda(x, v) < \lambda_1(x) \text{ for some } v \in \mathbb{R}^d \backslash \{0\}\right\} \\ &= \{x \in M \mid k(x) \ge 2\} \\ &= k^{-1}((-\infty, 2]) \end{aligned}$$

is measurable. Also, by Lemma 5.8, since the set

$$\begin{aligned} \left\{(x, v) \in M \times \mathbb{R}^d \mid v \in V_x^2\right\} &= \left\{(x, v) \in M \times \mathbb{R}^d \backslash \{0\} \mid \lambda(x, v) \le \lambda_2(x)\right\} \cup (M \times \{0\}) \\ &= V_*^2 \cup (M \times \{0\}) \end{aligned}$$

is a measurable subset of $M \times \mathbb{R}^d$, the function $x \mapsto V_x^2$ is a measurable function on $\pi\left(V_*^2\right)$.

Since $x \mapsto V_x^2$ is a measurable, it follows from Lemma 5.8 that each

$$M_l^2 := \left\{x \in \pi\left(V_*^2\right) \mid \dim V_x^2 = l\right\} \ , \ 1 \le l \le d$$

is a measurable subset of M and for each l there are measurable functions $v_j : M_l^2 \to \mathbb{R}^d$, $1 \le j \le l$ such that $\{v_1(x), ..., v_l(x)\}$ forms a basis of V_x^2 for every $x \in M_l^2$. Thus, for each x, at least one of $v_j(x)$ is in $V_x^2 \backslash V_x^3$, since V_x^3 is of a dimension strictly lower. Thus, we have

$$\max\{\lambda(x, v_j(x)) \mid 1 \ \le j \le l\} = \lambda_2(x)$$

is a measurable function on M_l^2, for all $1 \le l \le d$. Since $\pi(V_*^2) = \bigcup_{1 \le l \le d} M_l^2$, the map $x \mapsto \lambda_2(x)$ is a measurable function on $\pi(V_*^2)$.

Therefore, we showed that $k^{-1}((-\infty,2])$ is measurable, and $x \mapsto \lambda_1(x)$, $x \mapsto V_x^1$ are measurable functions on $\pi(V_*^2)$. To show measurability for $i=3$ we may proceed to define

$$V_*^3 := \left\{(x,v) \in M \times \mathbb{R}^d \backslash \{0\} \mid \lambda(x,v) < \lambda_2(x)\right\} \ .$$

and use the same argument. Continuing this recursive construction yields that

(i) $k^{-1}((-\infty,i])$ is measurable for all $i \geq 1$, so $x \mapsto k(x)$ is measurable;
(ii) $x \mapsto \lambda_i(x)$, $x \mapsto V_x^i$ are measurable functions on $\pi(V_*^i) = \{x \in M \mid k(x) \geq i\}$.

□

5.3. **Lemma for the Base Case.** In the previous subsections we showed that a weaker version of Theorem 5.1 holds. In this subsection, we prove a useful lemma for the base case of the inductive argument.

We say that a map $x \mapsto V_x$ is a **measurable sub-bundle** of $M \times \mathbb{R}^d$ if one, and hence all, of the statements in Lemma 5.8 hold. A measurable sub-bundle $x \mapsto V_x$ is called **invariant** if $A(x)V_x = V_{f(x)}$ for $\mu-$almost every x.

Lemma 5.11. *Consider the linear cocycle $F : M \times \mathbb{R}^d \to M \times \mathbb{R}^d$. Let $x \mapsto V_x$ be a measurable invariant sub-bundle of $M \times \mathbb{R}^d$, then for $\mu-$almost every x, one has*

$$(a)\ \lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)|V_x\| = \max\left\{\lambda(x,v) \mid v \in V_x \backslash \{0\}\right\} \ ,$$

$$(b)\ \lim_{n\to\infty} \frac{1}{n} \log \left\|(A^n(x)|V_x)^{-1}\right\|^{-1} = \min\left\{\lambda(x,v) \mid v \in V_x \backslash \{0\}\right\} \ .$$

Remark 5.12. A direct consequence of Lemma 5.11 is that $\lambda_1(x) = \lambda_+(x)$ and $\lambda_k(x) = \lambda_-(x)$, where $\lambda_\pm$ are the extremal Lyapunov exponents of F. This is obtained by taking $V_x = \mathbb{R}^d$.

To prove Lemma 5.11, we will need tools from the dynamics of skew products (Theorem 5.14 and Corollary 5.15 below). In the following, we state the needed results.

Let P be a compact metric space. Let $\mathcal{C}^0(P)$ denote the space of continuous real-valued functions on P, endowed with the norm

$$\|g\|_0 := \sup_{x\in P} |g(x)| \ .$$

Denote by $\mathscr{F}$ the space of all measurable functions $\Psi : M \times P \to \mathbb{R}$ such that $\Psi(x,\cdot) \in \mathcal{C}^0(P)$ for $\mu-$almost every $x \in M$ and the function $x \mapsto \|\Psi(x,\cdot)\|_0$ is $\mu-$integrable. Then,

$$\|\Psi\|_1 = \int \|\Psi(x,\cdot)\|_0 \ \mathrm{d}\mu(x)$$

defines a complete norm on $\mathscr{F}$.

Let $\mathscr{M}(\mu)$ be the space of probability measures on $M \times P$ such that $\pi_*\eta = \mu$, where $\pi : M \times P \to M$ is the canonical projection map. The weak* topology on $\mathscr{M}(\mu)$ is the smallest topology such that the operator $\phi : \mathscr{M}(\mu) \to \mathbb{R}$, defined as

$$\phi(\eta) := \int \Psi \ \mathrm{d}\eta, \text{ is continuous for all } \Psi \in \mathscr{F} \ .$$

We take for granted the following fact about $\mathscr{M}(\mu)$:

Fact 5.13. *The weak* topology on $\mathscr{M}(\mu)$ is compact and two probability measures $\eta, \xi \in \mathscr{M}(\mu)$ are equal if and only if $\int \Psi \ \mathrm{d}\eta = \int \Psi \ \mathrm{d}\xi$ for all $\Psi \in \mathscr{F}$.*

Next, we proceed to state the needed claims.

Theorem 5.14. *Let $\mathscr{G} : M \times P \to M \times P$ be a measurable map of the form*

$$\mathscr{G}(x,v) = (f(x), \mathscr{G}_x(v)) \ ,$$

where $\mathscr{G}_x : P \to P$ is continuous for $\mu-$almost every $x \in M$. Given any $\Phi \in \mathscr{F}$, define

$$I(x) := \lim_{n\to\infty} \frac{1}{n} \inf_{v\in P} \sum_{j=0}^{n-1} \Phi(\mathscr{G}^j(x,v)) \ , \ S(x) := \lim_{n\to\infty} \frac{1}{n} \sup_{v\in P} \sum_{j=0}^{n-1} \Phi(\mathscr{G}^j(x,v)) \ .$$

The limits $I(x), S(x)$ exist for $\mu-$almost every x and there exist $\mathscr{G}-$invariant measures $\eta_I, \eta_S \in \mathscr{M}(\mu)$ such that

$$\int \Phi \ \mathrm{d}\eta_I = \int I \ \mathrm{d}\mu \ , \ \int \Phi \ \mathrm{d}\eta_S = \int S \ \mathrm{d}\mu \ .$$

Corollary 5.15. *For $\mu-$almost every $x \in M$, there are $v_I(x), v_S(x) \in P$ such that*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \Phi(\mathscr{G}^j(x, v_I(x))) = I(x) \ , \ \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \Phi(\mathscr{G}^j(x, v_S(x))) = S(x) \ .$$

Assuming these results, we first prove Lemma 5.11. The proofs of Theorem 5.14 and Corollary 5.15 will be given at the end of this subsection.

Proof of Lemma 5.11. Note that $\{x \ : \ \dim V_x = l\}$ is measurable by Lemma 5.8, and $x \mapsto V_x$ is invariant. Thus, without loss of generality, we may assume that $\dim V_x$ is constant for all $x \in M$ by restricting to measurable subsets of M. Suppose $\dim V_x = l$ and let $x \in M$.

We want to construct a new linear cocycle whose extremal Lyapunov exponents correspond to the limits given by the claim of the theorem. By Lemma 5.8 and Gram-Schmidt, there exist measurable functions v_j such that $\{v_1(x), ..., v_l(x)\}$ is an orthonormal basis of V_x. Using an isometry $T_x : V_x \to \mathbb{R}^l$, $T_x(v_i(x)) = e_i$, we may identify V_x with $\mathbb{R}^l \subseteq \mathbb{R}^d$. Therefore, we may assume $V_x = \mathbb{R}^l$ and $A(x)|V_x \in GL(l)$.

Let $D(x) := A(x)|V_x$, and let $G : M \times \mathbb{R}^l \to M \times \mathbb{R}^l$ be the linear cocycle defined by D over f. Since

$$\left\|(D(x))^{\pm}\right\| = \sup_{v \in V_x \setminus \{0\}} \frac{|(A(x))^{\pm} v|}{|v|} \leq \left\|(A(x))^{\pm}\right\| \ ,$$

we know that $\log^+ \left\|(D(\cdot))^{\pm 1}\right\| \in L^1(\mu)$, and so by Furstenberg and Kesten's theorem (Theorem 4.2), the extremal Lyapunov exponents of G exist for $\mu-$almost every x. Denote by $u_{\pm}(x)$ the extremal Lyapunov exponents of G, then by our construction,

$$\lim_{n\to\infty} \frac{1}{n} \log \left\|(A^n(x)|V_x)^{\pm}\right\|^{\pm} = \lim_{n\to\infty} \frac{1}{n} \log \left\|(D^n(x))^{\pm}\right\|^{\pm} = u_{\pm}(x) \ .$$

We want to show that $u_{\pm}(x)$ is equal to the max/min in the claim respectively by applying Corollary 5.15 to the linear cocycle G, but $\mathbb{R}^l$ is not compact. Thus, we projectivize G by considering the function $\mathscr{G} : M \times \mathbb{PR}^l \to M \times \mathbb{PR}^l$ (see [11, Lemma 5.1] for a proof that $\mathbb{PR}^l$ is a compact metric space),

$$\mathscr{G}(x, [v]) = (f(x), [D(x)v]) \ .$$

Note that an element $[v] \in \mathbb{PR}^l$ is equivalent to a line $\{tv \mid t \in \mathbb{R}\} \subseteq \mathbb{R}^l$ for some $v \in \mathbb{R}^l \setminus \{0\}$. Now, define a function $\Phi : M \times \mathbb{PR}^l \to \mathbb{R}$ as

$$\Phi(x, [v]) := \log \frac{|D(x)v|}{|v|} \ .$$

Then, we have $\Phi \in \mathscr{F}$, since Φ is measurable, $\Phi(x, \cdot) \in \mathcal{C}^0(\mathbb{PR}^l)$ for every $x \in M$, and $\log^+ \|(D(x))^{\pm}\| \leq \log^+ \|(A(x))^{\pm}\| \in L^1(\mu)$.

Thus, for every $n \geq 0$, $v \in \mathbb{R}^l \setminus \{0\}$, we have

$$\sum_{j=0}^{n} \Phi(\mathscr{G}^j(x, [v])) = \sum_{j=0}^{n} \Phi\left(x, [D^j(x)v]\right) = \sum_{j=0}^{n} \log \frac{|D^{j+1}(x)v|}{|D^j(x)v|} = \log \frac{|D^n(x)v|}{|v|} \ .$$

It follows that

$$S_n(x) = \sup_{v \in \mathbb{R}^l \setminus \{0\}} \sum_{j=0}^{n} \Phi(\mathscr{G}^j(x, [v])) = \sup_{v \in \mathbb{R}^l \setminus \{0\}} \log \frac{|D^n(x)v|}{|v|} = \log \|D^n(x)\| \ ,$$

$$I_n(x) = \inf_{v \in \mathbb{R}^l \setminus \{0\}} \sum_{j=0}^{n} \Phi(\mathscr{G}^j(x, [v])) = \inf_{v \in \mathbb{R}^l \setminus \{0\}} \log \frac{|D^n(x)v|}{|v|} = \log \left\|(D^n(x))^{-1}\right\|^{-1} \ ,$$

and so $I(x) = u_-(x)$, $S(x) = u_+(x)$, and the functions I, S have properties as in Theorem 5.14.

Moreover, for any $v \in V_x \backslash \{0\} = \mathbb{R}^l \backslash \{0\}$, we have

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \Phi(\mathscr{G}^j(x,[v])) &= \limsup_{n\to\infty} \frac{1}{n} \log \frac{|D^n(x)v|}{|v|} \\ &= \limsup_{n\to\infty} \frac{1}{n} \log \frac{|A^n(x)v|}{|v|} = \lambda(x,v) \ .\end{aligned}$$

Thus, by Corollary 5.15, there exists $v_S(x), v_I(x) \in \mathbb{R}^l \backslash \{0\}$ such that

$$\lambda(x, v_S(x)) = S(x) = u_+(x) \ , \ \lambda(x, v_I(x)) = I(x) = u_-(x) \ .$$

Note that if $v \in \mathbb{R}^l \backslash \{0\}$, by Lemma 5.3 we have $u_-(x) \le \lambda(x,v) \le u_+(x)$ for $\mu-$almost every x. Thus, in summary,

$$\begin{aligned} u_+(x) &= \lambda(x, v_S(x)) = \max\{\lambda(x,v) \mid v \in \mathbb{R}^l \backslash \{0\}\} \ , \\ u_-(x) &= \lambda(x, v_I(x)) = \min\{\lambda(x,v) \mid v \in \mathbb{R}^l \backslash \{0\}\} \ . \end{aligned}$$

□

We conclude this subsection by proofs of Theorem 5.14 and Corollary 5.15.

Proof of Theorem 5.14. The roles of supremum and infimum can be interchanged if we replace Φ by $-\Phi$, so it suffices to only prove the claim for $I(x)$.

To begin with, we want to use the subadditive ergodic theorem (Theorem 4.1) to show that the limit $I(x)$ exists for $\mu-$almost every $x \in M$. For $x \in M$, define

$$I_n(x) := \inf_{v\in P} \sum_{j=0}^{n-1} \Phi(\mathscr{G}^j(x,v)) \ .$$

We first verify that each $I_n(x)$ is measurable. Note that P is separable since it is compact, and for a continuous function $g : P \to \mathbb{R}$, one always has $g(\overline{A}) \subseteq \overline{g(A)}$. Thus, if A is a countable dense subset of P, then

$$I_n(x) = \inf_{v\in P} \sum_{j=0}^{n-1} \Phi(\mathscr{G}^j(x,v)) = \inf_{v\in A} \sum_{j=0}^{n-1} \Phi(\mathscr{G}^j(x,v)) \ .$$

Moreover, note that the operation of taking an infimum over a countable set of measurable functions is measurable: for $g_\alpha : P \to \mathbb{R}$, $\alpha \ge 0$, measurable, the set

$$\left(\inf_{\alpha\ge0} g_\alpha\right)^{-1}((-\infty,c)) = \left\{t \in P \mid \inf_{\alpha\ge0} g_\alpha(t) < c\right\} = \bigcup_{\alpha\ge0} \{t \in P \mid g_\alpha(t) < c\}$$

is a countable union of measurable sets and so it is measurable. It follows that each I_n is measurable since they are compositions of measurable functions.

We then verify that $I_n(x)$ is superadditive (so that $-I_n(x)$ is subadditive). For $m, n \ge 1$, for all $x \in M$,

$$\begin{aligned} I_{m+n}(x) = \inf_{v\in P} \sum_{j=0}^{m+n-1} \Phi(\mathscr{G}^j(x,v)) &= \inf_{v\in P} \sum_{j=0}^{m-1} \Phi(\mathscr{G}^j(x,v)) + \sum_{j=m}^{n+m-1} \Phi(\mathscr{G}^j(x,v)) \\ &\ge \inf_{v\in P} \sum_{j=0}^{m-1} \Phi(\mathscr{G}^j(x,v)) + \inf_{v\in P} \sum_{j=0}^{n-1} \Phi(\mathscr{G}^{j+m}(x,v)) \\ &\ge I_m(x) + I_n(f^m(x)) \ . \end{aligned}$$

Finally, note that I_1 is integrable since $\|\Phi(x,\cdot)\|_0 \in L^1(\mu)$. Thus, we can apply Theorem 4.1 to $\{-I_n\}_{n\in\mathbb{N}}$ to show that the limit

$$I(x) = \lim_{n\to\infty} \frac{1}{n} I_n(x)$$

exists for $\mu-$almost every $x \in M$.

Next, we proceed to construct a measure η_I which satisfies the condition in the claim. For $n \geq 0$, consider measurable subsets of $M \times P$ defined by

$$\Gamma_n := \left\{ (x,v) \in M \times P \mid \sum_{j=0}^{n-1} \Phi(\mathscr{G}^j(x,v)) = I_n(x) \right\} \in \mathscr{B} \times \mathscr{B}(P) \ .$$

Additionally, for $x \in M$, define

$$\Gamma_n(x) := \{v \in P \mid (x,v) \in \Gamma_n\} = \left\{ v \in P \mid v \text{ is a minimum of} \sum_{j=0}^{n-1} \Phi(\mathscr{G}^j(x,v)) \right\} \ .$$

Since P is compact and the map $v \mapsto \Phi(\mathscr{G}^j(x,v))$ is continuous for every j, $\Gamma_n(x)$ is non-empty and compact for $\mu-$almost every $x \in M$. Therefore, $x \mapsto \Gamma_n(x)$ is a map with values in the set of compact subsets of P, and its graph is measurable, so by Lemma 5.7, there exists a measurable selection $v_n : M \to P$ such that $v_n(x) \in \Gamma_n(x)$ for $\mu-$almost every x.

Now, for $n \geq 0$, we define the following probability measures on $M \times P$:

$$\xi_n(A) := \int_A \delta_{(x,v_n(x))} \, \mathrm{d}\mu(x) \ , \ \eta_n(A) := \frac{1}{n} \sum_{j=0}^{n-1} \mathscr{G}_*^j \xi_n(A) \ ,$$

for arbitrary measurable $A \subseteq M \times P$. Note that for $B \subseteq M$ measurable, we have

$$\begin{aligned} \pi_* \xi_n(B) = \xi_n(\pi^{-1}(B)) &= \int_{\pi^{-1}(B)} \delta_{(x,v_n(x))} \, \mathrm{d}\mu(x) \\ &= \int_B 1 \, \mathrm{d}\mu(x) \\ &= \mu(B) \ , \end{aligned}$$

so $\xi_n \in \mathscr{M}(\mu)$ for each n. Since f preserves the measure μ and $\mathscr{G}(x,v) = (f(x), \mathscr{G}_x(v))$, we also have $\pi_* \eta_n = \mu$ for each n. It follows that $\eta_n \in \mathscr{M}(\mu)$ for all $n \in \mathbb{N}$.

By compactness of $\mathscr{M}(\mu)$ (Fact 5.13), there exists a subsequence $\{\eta_{n_k}\}_{k \in \mathbb{N}}$ that converges to some $\eta_I \in \mathscr{M}(\mu)$ in the weak* topology.

We first verify $\mathscr{G}-$invariance of η_I. Observe that for any $\Psi \in \mathscr{F}$, by $\mu-$invariance of f, we have

$$\begin{aligned} \left| \int \Psi \circ \mathscr{G} \, \mathrm{d}\eta_{n_k} - \int \Psi \, \mathrm{d}\eta_{n_k} \right| &= \frac{1}{n_k} \left| \int \Psi \circ \mathscr{G}^{n_k} - \Psi \, \mathrm{d}\eta_{n_k} \right| \\ &\leq \frac{1}{n_k} \int \left| \Psi \circ \mathscr{G}^{n_k}(x, v_{n_k}(x)) - \Psi(x, v_{n_k}(x)) \right| \, \mathrm{d}\mu(x) \\ &\leq \frac{2}{n_k} \int \|\Psi(x,\cdot)\|_0 \, \mathrm{d}\mu(x) \\ &= \frac{2}{n_k} \|\Psi\|_1 \ . \end{aligned}$$

By the definition of weak* topology on $\mathscr{M}(\mu)$, the left side converges to $\left| \int \Psi \circ \mathscr{G} \, \mathrm{d}\mu_I - \int \Psi \, \mathrm{d}\mu_I \right|$. At the same time, the right side converges to 0, so we obtain

$$\int \Psi \circ \mathscr{G} \, \mathrm{d}\eta_I = \int \Psi \, \mathrm{d}\eta_I \ . \tag{5.16}$$

Fact 5.13 and (5.16) implies that $\mathscr{G}_* \eta_I = \eta_I$, so η_I is $\mathscr{G}-$invariant. Finally, by the subadditive ergodic theorem (Theorem 4.1), we have

$$\begin{aligned} \int \Phi \, \mathrm{d}\eta_I = \lim_{k \to \infty} \int \Phi \, \mathrm{d}\eta_{n_k} &= \lim_{k \to \infty} \int \frac{1}{n_k} \sum_{j=0}^{n_k-1} \Phi(\mathscr{G}^j(x, v_{n_k}(x))) \, \mathrm{d}\mu(x) \\ &= \lim_{k \to \infty} \frac{1}{n_k} \int I_{n_k}(x) \, \mathrm{d}\mu(x) \\ &= \int I \, \mathrm{d}\mu \ . \end{aligned}$$

□

Proof of Corollary 5.15. The roles of I, S can be interchanged if we replace Φ by $-\Phi$, so it suffices to only prove the claim for $I(x)$. First note that is obvious that

$$I(x) \leq \liminf_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \Phi(\mathscr{G}^j(x,v)) \leq \limsup_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \Phi(\mathscr{G}^j(x,v)) \leq S(x) \ .$$

Also, note that by the Birkhoff ergodic theorem (Theorem 4.7), given any $\mathscr{G}-$invariant measure η, the limit

$$\widetilde{\Phi}(x,v) = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \Phi(\mathscr{G}^j(x,v))$$

exists for $\eta-$almost every point (x,v), and satisfies $\int \widetilde{\Phi}\ \mathrm{d}\eta = \int \Phi\ \mathrm{d}\eta$.

Let $\eta = \eta_I$ as in Theorem 5.14. We obtain

$$\int \widetilde{\Phi}\ \mathrm{d}\eta_I = \int \Phi\ \mathrm{d}\eta_I = \int I\ \mathrm{d}\mu \ .$$

Note that $\eta_I \in \mathscr{M}(\mu)$, so $\pi_*\eta_I = \mu$. Thus, we can think of I as a function $I(x,v)$ constant for every $v \in P$. Then, we have $\int \widetilde{\Phi}\ \mathrm{d}\eta_I = \int I\ \mathrm{d}\eta_I$. Therefore, it follows that the set

$$E := \left\{(x,v) \in M \times P \mid \widetilde{\Phi}(x,v) = I(x)\right\}$$

is measurable and has η_I-full measure. By Fact 5.9, the projection $\pi(E)$ is measurable and since $\pi_*\eta_I = \mu$,

$$\mu(\pi(E)) = \eta_I(\pi^{-1}(\pi(E))) \geq \eta_I(E) = 1 \ .$$

Hence, for all $x \in \pi(E)$, there exists $v \in P$ such that $(x,v) \in E$, which proves the corollary. $\square$

5.4. **Lemma for the Inductive Step.** In this subsection, we prove a useful lemma for the inductive step of the inductive argument.

Throughout this subsection, we take $x \mapsto V_x$ to be a measurable invariant sub-bundle and $\alpha(x) < \beta(x)$ to be $f-$invariant, $\mu-$integrable functions such that for $\mu-$almost every $x \in M$, one has

(i) $\lambda(x,v) \leq \alpha(x)$ for every $v \in V_x\backslash\{0\}$,
(ii) $\lambda(x,u) \geq \beta(x)$ for every $u \in \mathbb{R}^d\backslash V_x$.

For $x \in M$, let $V_x^\perp$ denote the orthogonal complement of V_x. Note that since $x \mapsto V_x$ is measurable and the orthogonal complement map $\perp\colon Gr(l,d) \to Gr(d-l,d)$ is a diffeomorphism for every l, the map $x \mapsto V_x^\perp$ is also measurable.

We think of $A(x)$ as a linear map $A(x) : (V_x \oplus V_x^\perp = \mathbb{R}^d) \to (\mathbb{R}^d = V_{f(x)} \oplus V_{f(x)}^\perp)$. Recall that by invariance we have $A(x)V_x = V_{f(x)}$ for $\mu-$almost every x (part (a)). Let

$$A(x) = \begin{pmatrix} B(x) & 0 \\ C(x) & D(x)\end{pmatrix} \tag{5.17}$$

denote the expression of $A(x)$ relative to the direct sum decomposition $\mathbb{R}^d = V_x \oplus V_x^\perp$. Then, $D(x) : V_x \to V_{f(x)}$ is the restriction of $A(x)$ to V_x, $B(x) : V_x^\perp \to V_{f(x)}^\perp$ gives a new linear cocycle, and $C(x)$ is a linear transformation $C(x) : V_x^\perp \to V_{f(x)}$. Since $\log^+ \left\|A^{\pm 1}\right\|$ is $\mu-$integrable, we know that

$$\log^+ \left\|B^{\pm 1}\right\| \ , \ \log^+ \|C\| \ , \ \log^+ \left\|D^{\pm 1}\right\| \in L^1(\mu) \ . \tag{5.18}$$

Based on the above set-up, we prove the following Lemma:

Lemma 5.19. *For $\mu-$almost every $x \in M$, for all $u \in V_x^\perp\backslash\{0\}$ and $v \in V_x$, we have*

(a) $\limsup_n \frac{1}{n}\log|B^n(x)u| = \limsup_n \log|A^n(x)(u+v)|$;
(b) *if* $\lim_n \frac{1}{n}\log|B^n(x)u|$ *exists, then* $\lim_n \frac{1}{n}|A^n(x)(u+v)|$ *exists and the two limits are equal.*

To prove this lemma, we need the following fact, which will be proven at the end of this subsection.

Fact 5.20. *Given any $\epsilon > 0$, there exists a measurable function $d_\epsilon(x) > 0$, finite $\mu-$almost everywhere, such that for all $m, n \geq 0$.*

$$\|D^n(f^m(x))\| \leq d_\epsilon(x) e^{\alpha(x)n+(m+n)\epsilon} \ .$$

Proof of Lemma 5.19. The the following, we first show that it suffices to prove the claim of part (a) for $v = 0$. Let $u \in V_x^\perp \backslash \{0\}$ and $v \in V_x$. Observe that Lemma 5.5 and assumptions (i) and (ii) imply

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log |A^n(x)(u+v)| &\leq \limsup_{n\to\infty} \frac{1}{n} \log(|A^n(x)u| + |A^n(x)v|) \\ &= \max\left\{\limsup_{n\to\infty} \frac{1}{n} \log |A^n(x)u| \ , \ \limsup_{n\to\infty} \frac{1}{n} \log |A^n(x)v|\right\} \\ &= \limsup_{n\to\infty} \frac{1}{n} \log |A^n(x)u| \ .\end{aligned}$$

Similarly, since $u + v \in \mathbb{R}^d \backslash V_x$,

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log |A^n(x)u| &\leq \limsup_{n\to\infty} \frac{1}{n} \log(|A^n(x)(u+v)| + |A^n(x)v|) \\ &= \max\left\{\limsup_{n\to\infty} \frac{1}{n} \log |A^n(x)(u+v)| \ , \ \limsup_{n\to\infty} \frac{1}{n} \log |A^n(x)v|\right\} \\ &= \limsup_{n\to\infty} \frac{1}{n} \log |A^n(x)(u+v)| \ .\end{aligned}$$

Therefore, it suffices to prove (a) for $u \in V_x^\perp$ and $v = 0$, since for all $u \in V_x^\perp \backslash \{0\}$, $v \in V_x$, we have

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

Recall that $A^n(x) = A(f^{n-1}(x)) \cdots A(x)$, so according to the decomposition in (5.17), we have

$$A^n(x) = \begin{pmatrix} B^n(x) & 0 \\ C_n(x) & D^n(x) \end{pmatrix} \ ,$$

where

$$C_n(x) := \sum_{j=0}^{n-1} D^{n-j-1}(f^{j+1}(x)) C(f^j(x)) B^j(x) \ .$$

For $x \in M$ and $u \in V_x^\perp$, define

$$\gamma := \max\left\{\alpha(x), \limsup_{n\to\infty} \frac{1}{n} \log |B^n(x)u|\right\} \ .$$

We want to find an estimate for $|C_n(x)u|$ in terms of γ. To do so, we need to bound $\left|B^j(x)u\right|$, $\left\|C(f^j(x))\right\|$, and $\|D^n(f^m(x))\|$ respectively.

Let $\epsilon > 0$. Since $\limsup_{n\to\infty} \frac{1}{n} \log |B^n(x)u| \leq \gamma$, by definition of the limsup, there is a constant $b_\epsilon \in \mathbb{R}$ such that

$$\left|B^j(x)u\right| \leq b_\epsilon e^{j(\gamma+\epsilon)} \ , \text{ for all } j \geq 0 \ .$$

Now, observe that by (5.18), the function $x \mapsto \log \|C(f(x))\| - \log \|C(x)\|$ is $\mu-$integrable. Thus, by Corollary 4.8, for $\mu-$almost every x, we have

$$\lim_{n\to\infty} \frac{1}{n} \log \|C(f^n(x))\| = 0 \ .$$

It follows that there exists a measurable function $c_\epsilon > 0$ such that

$$\left\|C(f^j(x))\right\| \leq c_\epsilon(x) e^{j\epsilon} \ , \text{ for all } j \geq 0 \ .$$

Moreover, by Fact 5.20, there exists a measurable function $d_\epsilon > 0$ such that for all $m, n \geq 0$.

$$\|D^n(f^m(x))\| \leq d_\epsilon(x) e^{\alpha(x)n+(m+n)\epsilon} \ .$$

Therefore, we have

$$\begin{aligned} |C_n(x)u| &\le \sum_{j=0}^{n-1} \left| D^{n-j-1}(f^{j+1}(x)) C(f^j(x)) B^j(x) u \right| \\ &\le \sum_{j=0}^{n-1} \underbrace{b_\epsilon c_\epsilon(x) d_\epsilon(x)}_{=:a_\epsilon(x) < +\infty} \exp\{(n-j-1)\alpha(x) + (n+j)\epsilon + j(\gamma+\epsilon)\} \\ &\le a_\epsilon(x) e^{n(\gamma+3\epsilon)} \ . \end{aligned}$$

Taking a limsup yields

$$\limsup_{n\to\infty} \frac{1}{n} \log |C_n(x)u| \le \gamma + 3\epsilon \ .$$

Now, observe that according to our decomposition of the matrix $A(x)$ in (5.17), we can express the vector $A^n(x)u \in \mathbb{R}^d$ as $A^n(x)u = (B^n(x)u, C_n(x)u)$, so

$$|A^n(x)u|^2 = |B^n(x)u|^2 + |C_n(x)u|^2 \ .$$

Thus, by Lemma 5.5, we obtain

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n} \log |A^n(x)u| &= \max \left\{ \limsup_{n\to\infty} \frac{1}{n} \log |B^n(x)u| \ , \ \limsup_{n\to\infty} \frac{1}{n} \log |C_n(x)u| \right\} \\ &\le \max\{\gamma, \gamma + 3\epsilon\} \\ &= \gamma + 3\epsilon \ . \end{aligned}$$

Since $\epsilon > 0$ is arbitrary, follows that

$$\limsup_{n\to\infty} \frac{1}{n} \log |B^n(x)u| \le \limsup_{n\to\infty} \frac{1}{n} \log |A^n(x)u| \le \gamma \tag{5.21}$$

But assumption (i) and (ii) implies that

$$\alpha(x) < \beta(x) \le \limsup_{n\to\infty} \frac{1}{n} \log |A^n(x)u| \le \gamma \ ,$$

so $\alpha(x)$ is strictly smaller than γ. Hence, by definition of γ, we must have

$$\limsup_{n\to\infty} \frac{1}{n} \log |B^n(x)u| = \gamma.$$

Therefore, combining with (5.21), this forces

$$\limsup_{n\to\infty} \frac{1}{n} \log |B^n(x)u| = \limsup_{n\to\infty} \frac{1}{n} \log |A^n(x)u| \ .$$

We proceed to prove part (b). Note that if $u \in V_x^\perp \backslash \{0\}$, $v \in V_x$, we can write the vector $A^n(x)(u+v)$ as

$$A^n(x)(u+v) = A^n(x)u + A^n(x)v = (B^n(x)u, C_n(x)u + D^n(x)v) \ ,$$

which implies

$$|A^n(x)(u+v)|^2 = |B^n(x)u|^2 + |C_n(x)u + D^n(x)v|^2 \ .$$

If the limit $\lim_n \frac{1}{n} \log |B^n(x)u|$ exists, then by Lemma 5.5,

$$\begin{aligned} &\liminf_{n\to\infty} \frac{1}{n} \log |A^n(x)(u+v)| \\ &\qquad \ge \max \left\{ \liminf_{n\to\infty} \frac{1}{n} \log |B^n(x)u| \ , \ \liminf_{n\to\infty} \frac{1}{n} \log \{|C_n(x)u| + |D^n(x)u|\} \right\} \\ &\qquad \ge \liminf_{n\to\infty} \frac{1}{n} \log |B^n(x)u| = \limsup_{n\to\infty} \frac{1}{n} \log |B^n(x)u| \ . \end{aligned}$$

Combining with part (a), this proves part (b).

□

Proof of Fact 5.20. Define

$$b_\epsilon(x) := \sup_{n\geq 0} \|D^n(x)\| \, e^{-n(\alpha(x)+\epsilon)} \ .$$

We want to bound the function $b_\epsilon(f^m(x))$ for arbitrary $m \geq 0$ by applying Corollary 4.8.

By assumption (i) and Lemma 5.11, we have

$$\lim_{n\to\infty} \frac{1}{n} \log \|D^n(x)\| = \lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)|V_x\| \leq \alpha(x) \ ,$$

so by definition of the limsup we must have $1 \leq b_\epsilon(x) < +\infty$, for $\mu-$almost every $x \in M$.

We proceed to verify that $b_\epsilon \circ f - b_\epsilon$ is $\mu-$integrable. First, observe that for all $x \in M$,

$$\begin{aligned} b_\epsilon(f(x)) &= \sup_{n\geq 0} \|D^n(f(x))\| \, e^{-n(\alpha(x)+\epsilon)} \\ &\leq \sup_{n\geq 0} \|D(f^n(x)) \cdots D(x)\| \, e^{-(n+1)(\alpha(x)+\epsilon)} \left\|(D(x))^{-1}\right\| e^{\alpha(x)+\epsilon} \\ &= \left\|(D(x))^{-1}\right\| e^{\alpha(x)+\epsilon} \sup_{n\geq 0} \left\|D^{n+1}(x)\right\| e^{-(n+1)(\alpha(x)+\epsilon)} \\ &\leq \left\|(D(x))^{-1}\right\| e^{\alpha(x)+\epsilon} b_\epsilon(x) \ , \end{aligned}$$

and so we have

$$\log b_\epsilon(f(x)) - \log b_\epsilon(x) \leq \log^+ \left\|(D(x))^{-1}\right\| + \alpha(x) + \epsilon \ . \tag{5.22}$$

Next, observe that since the operator norm is sub-multiplicative (i.e., $\|XY\| \leq \|X\| \, \|Y\|$), we have

$$\begin{aligned} b_\epsilon(f(x)) &= \sup_{n\geq 0} \|D^n(f(x))\| \, e^{-n(\alpha(x)+\epsilon)} \\ &= \sup_{n\geq 1} \left\|D(f^{n-1}(x)) \cdots D(f(x))\right\| e^{-n(\alpha(x)+\epsilon)} \frac{\|D(x)\|}{\|D(x)\|} e^{\alpha(x)+\epsilon} \\ &\geq \|D(x)\|^{-1} \, e^{\alpha(x)+\epsilon} \cdot \sup_{n\geq 1} \|D^n(x)\| \, e^{n(\alpha(x)+\epsilon)} \ . \end{aligned} \tag{5.23}$$

Therefore, there are two possibilities:

(1) If $b_\epsilon(x) = \sup_{n\geq 1} \|D^n(x)\| \, e^{-n\alpha(x)+\epsilon}$, then by (5.23) we have

$$b_\epsilon(f(x)) \geq \|D(x)\|^{-1} \, e^{\alpha(x)+\epsilon} \cdot b_\epsilon(x) \ ,$$

which implies

$$\log b_\epsilon(f(x)) \geq \log \|D(x)\|^{-1} + \log b_\epsilon(x) + \alpha(x) + \epsilon \ .$$

(2) If $b_\epsilon = 1$ (i.e., the supremum is attained at $n = 0$), then we have $b_\epsilon(f(x)) \geq 1 = b_\epsilon(x)$, so

$$\log b_\epsilon(f(x)) \geq \log b_\epsilon(x) \ .$$

Combining the two cases, we have

$$\begin{aligned} \log b_\epsilon(f(x)) - \log b_\epsilon(x) &\geq \min \left\{ \log \|D(x)\|^{-1} + \alpha(x) + \epsilon, 0 \right\} \\ &\geq \min \left\{ -\log^+ \|D(x)\| + \alpha(x) + \epsilon, 0 \right\} \ . \end{aligned} \tag{5.24}$$

Therefore, integrability of $\|D^{\pm}(\cdot)\|$ and $\alpha(x)$, equations (5.24) and (5.22) implies that $b_\epsilon \circ f - b_\epsilon$ is $\mu-$integrable. By Corollary 4.8, we see that for $\mu-$almost every x,

$$\lim_{m\to\infty} \frac{1}{m} \log b_\epsilon(f^m(x)) = 0 \ ,$$

which implies that for $\mu-$almost every x,

$$0 < d_\epsilon(x) := \sup_{m\geq 0} b_\epsilon(f^m(x)) e^{-\epsilon m} < +\infty \ .$$

It follows that for $\mu-$almost every x, there is a measurable function d_ϵ such that for all $m, n \geq 0$,

$$\|D^n(f^m(x))\| \leq b_\epsilon(x) e^{n(\alpha(x)+\epsilon)} \leq d_\epsilon e^{n\alpha+(n+m)\epsilon} \ .$$

∎

5.5. **Induction.** In this subsection, we prove part (c) of Theorem 5.1. Recall that we proved in propositions 5.2 and 5.10 that parts (a), (b), (c)$'$ hold. We now use the results we obtained in the previous subsections (Lemmas 5.11 and 5.19) to show that for $\mu-$almost every $x \in M$, for $1 \le i \le k$, one has

$$\lim_{n\to\infty} \frac{1}{n} \log |A^n(x)v| = \lambda_i(x) \ , \ \text{for all } v \in V_x^i \backslash V_x^{i+1} \ . \tag{5.25}$$

Proposition 5.26. *Part* (c) *of Theorem 5.1 holds.*

Proof. Note that the functions $V^i(x), k(x), \lambda_i(x)$ are invariant (Lemma 5.2). Thus, without loss of generality, we may assume that $k = k(x)$ is independent of x, and the dimension l of the linear subspace $V_x = V_x^k$ is constant, by restricting measurable invariant subsets of M.

We will prove the claim using a recursive argument. To do so, we first show that (5.25) holds for $i = k$ (base case), by applying Lemma 5.11 to $V_x = V_x^k$. Observe that since $x \mapsto V_x$ is a measurable invariant sub-bundle, by Lemma 5.11, we have

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \log \left\| (A^n(x)|V_x)^{-1} \right\|^{-1} &= \min\left\{\lambda(x,v) \mid v \in V_x \backslash \{0\}\right\} \\
&= \lambda_k(x) \\
&= \max\left\{\lambda(x,v) \mid v \in V_x \backslash \{0\}\right\} = \lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)|V_x\| \,.
\end{aligned}$$

Thus, this forces

$$\lim_{n\to\infty} \frac{1}{n} \log |A^n(x)v| = \lambda_k(x) \ , \ \text{for all } v \in V_x \backslash \{0\} \ .$$

Now, we proceed to show that the limit exists for $i = k-1$ by applying Lemma 5.19 (inductive step). Let $\alpha(x) := \lambda_k(x)$, $\beta(x) := \lambda_{k-1}(x)$. Then, by construction, conditions (i) and (ii) in the set-up of Subsection 5.4 are satisfied. Consider $A(x)$ relative to the direct sum decomposition $\mathbb{R}^d = V_x \oplus V_x^\perp$, as in (5.17):

$$A(x) = \begin{pmatrix} B(x) & 0 \\ C(x) & D(x) \end{pmatrix} \ .$$

Since $x \mapsto V_x^\perp$ is measurable, by Lemma 5.8, we may choose a measurable orthonormal basis $\{w_1(x), ..., w_{d-l}(x)\}$ of $V_x^\perp$ and assume that $V_x^\perp = \mathbb{R}^{d-l}$ via an isometry. Thus, we may consider $B(x)$ as an element of $GL(d-l)$.

For each $1 \le i \le k$, we define $U_x^i := V_x^\perp \cap V_x^i$. Observe that for each $1 \le i \le k-1$ and $u \in U_x^i \backslash U_x^{i+1}$, we have $u \in (V_x^i \backslash V_x^{i+1}) \cap V_x^\perp$, so by Lemma 5.19(a), we obtain

$$\limsup_{n\to\infty} \frac{1}{n} \log |B^n(x)u| = \limsup_{n\to\infty} \frac{1}{n} \log |A^n(x)u| = \lambda_i(x) \ .$$

In general, recall that for all $1 \le i \le k-1$, if $w \in V_x^i \backslash V_x^{i+1}$, we can write $w = u + v$ where $0 \ne u \in U_x^i \backslash U_x^{i+1}$ and $v \in V_x$. Via Lemma 5.19(a), we will get

$$\limsup_{n\to\infty} \frac{1}{n} \log |A^n(x)w| = \limsup_{n\to\infty} \frac{1}{n} \log |B^n(x)u| = \lambda_i(x) \ . \tag{5.27}$$

Therefore, there is a new linear cocycle defined by B over f, such that $V_x^\perp = \mathbb{R}^{d-l} = U_x^1 \supsetneq \cdots \supseteq U_x^{k-1} \supseteq \{0\}$ is the Oseledets flag of B. This is a linear cocycle with one less subspace in its flag than the linear cocycle defined by A over f, and we showed in (5.27) that the former completely determines the latter.

Note that $x \mapsto U_x^{k-1}$ is a measurable invariant sub-bundle since $x \mapsto V_x, V_x^{k-1}$ are measurable invariant. So we may apply Lemma 5.11 to the linear cocycle defined by B and U_x^{k-1} to show that for all $u \in U_x^{k-1} \backslash \{0\}$,

$$\lim_{n\to\infty} \frac{1}{n} \log |B^n(x)u| = \lambda_{k-1}(x) \ . \tag{5.28}$$

Using Lemma 5.19(b), if the limit in (5.28) exists, then the corresponding limit for the linear cocycle defined by $A(x)$ exists, so for all $v \in V_x^{k-1} \backslash V_x^k$, we have

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

We can show that the limit for $i = k-2$ exists by repeating the same argument and decomposing B relative the the direct sum $\mathbb{R}^{d-l} = U_x^{k-1} \oplus (U_x^{k-1})^{\perp}$. Recursively applying this argument yields that for all $1 \le i \le k$, for almost every $x \in M$, all $v \in V_x^i \backslash V_x^{i+1}$,

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

which concludes the proof of Theorem 5.1. $\square$

6. Applications

In this subsection, we discuss some applications of the multiplicative ergodic theorem to product of random matrices and Schrödinger cocycles. To do so, we first introduce some elementary notions from ergodic theory.

Definition 6.1. Let $(M, \mathscr{B}, \mu)$ be an arbitrary probability space. A measure preserving transformation $f : M \to M$ is called **ergodic** if for all $B \in \mathscr{B}$ such that $f^{-1}(B) = B$, one has $\mu(B) = 0$ or $\mu(B) = 1$.

There are several equivalent characterizations of ergodicity, but the one most important to us is the following:

Theorem 6.2. *Let $(M, \mathscr{B}, \mu)$ be a probability space. If a map $f : M \to M$ is measure preserving, then the following statements are equivalent:*

(1) *f is ergodic;*
(2) *Whenever a measurable map $g : M \to \mathbb{R}$ satisfies $g \circ f(x) = g(x)$ for $\mu-$almost every $x \in M$, g is constant $\mu-$almost everywhere.*

Proof. See a proof in [13, Theorem 1.6]. $\square$

Thus, for a linear cocycle F defined by A over f, it follows directly from Theorem 6.2 that if f is ergodic, then the functions $x \mapsto k(x), x \mapsto \lambda_i(x)$ found in Theorem 5.1 are constant, and so are the dimensions of the subspaces V_x^i. Therefore, we have the following corollary:

Corollary 6.3. *Suppose $f : M \to M$ is ergodic and $\log^+ \|A^{\pm 1}\|$ are integrable with respect to μ. Then, there are constants $1 \le k \le d$, $\lambda_1 > ... > \lambda_k$ such that for $\mu-$almost every $x \in M$, there is a flag $\mathbb{R}^d = V_x^1 \supsetneq \cdots \supsetneq V_x^k \supsetneq \{0\}$ such that for all $1 \le i \le k$, the following holds:*

(a) *the map $x \mapsto V_x^i$ is measurable and satisfies $A(x) \cdot V_x^i = V_{f(x)}^i$,*
(b) *one has*

$$\lim_{n\to\infty} \frac{1}{n} \log |A^n(x)v| = \lambda_i \ , \ \textit{for all } v \in V_x^i \backslash V_x^{i+1} \ .$$

6.1. **Products of Random Matrices.** Now, we discuss some implications of Theorem 5.1 for product of random matrices (see Example 2.4).

In particular, as we mentioned in Example 2.4, products of i.i.d. random matrices models are equivalent to random transformations for which $A : M \to GL(d)$ is the function $\{A_n\}_n \mapsto A_0$. Moreover, the shift map $f : M \to M$ in any Bernoulli scheme is ergodic (see [13, Theorem 1.12] for proof). Thus, Corollary 6.3 applies to all products of i.i.d. random matrices, and in the language of probability, we have:

Corollary 6.4. *Suppose the sequence $A_0, A_1, \ldots$ is a sequence of invertible, i.i.d, $d \times d$ random matrices. Suppose $\mathbb{E}\left[\log^+ \|A_0^{\pm}\|\right] < \infty$. Then, there are constants $1 \le k \le d$, $\lambda_1 > ... > \lambda_k$ and random subspaces $\mathbb{R}^d = V^1 \supsetneq \cdots V^k \supsetneq \{0\}$ such that for all $1 \le i \le k$, for all $v \in V^i \backslash V^{i+1}$, one has*

$$\lim_{n\to\infty} \frac{1}{n} \log |A_{n-1} \cdots A_0 v| = \lambda_i \ .$$

Therefore, the multiplicative ergodic theorem allows us to establish the existence of Lyapunov exponents for products of random matrices.

We conclude this subsection by mentioning some further results about the Lyapunov exponents of products of random matrices. These results provide examples of responses to three key questions that have been historically important to the theory of Lyapunov exponents [16, Part I]. The first one is **non-triviality**: when is it the case that there are more than one Lyapunov exponents, i.e., $k \neq 1$? The second is **simplicity**: when do we have d distinct Lyapunov exponents, i.e., $k = d$?

The third question is concerned with **continuity**: how do the Lyapunov exponents depend on their underlying linear cocycle?

- (Conditions for distinct Lyapunov exponents) Note that a $2d \times 2d$ **symplectic matrix** is a matrix M with real entries satisfying $M^T \Omega M = \Omega$, for
$$\Omega = \begin{pmatrix} 0 & I_d \\ -I_d & 0 \end{pmatrix},$$
where I_d is the $d \times d$ identity matrix. In 1970, Virster proved the following theorem for products of symplectic random matrices:

Theorem 6.5 (Virster, 1970)**.** *For a sequence of i.i.d.* $2d \times 2d$ *symplectic random matrices, we have the following:*
  (i) *the Lyapunov spectrum is nondegenerate (i.e.,* $k = 2d$*, so there are* $2d$ *distinct Lyapunov exponents);*
  (ii) $\lambda_1 > ... > \lambda_d > 0$*;*
  (iii) $\lambda_{2d-i+1} = -\lambda_i$ *.*

  See [14, pp. 22] for more details.
- (Generalized law of large numbers) In his monograph, Bougerol proved that for a sequence of i.i.d. random matirces $A_0, A_1, ...$, under some irreducibility and integrability assumptions, there is a real number λ_1 such that with probability 1,
$$\lim_{n\to\infty} \frac{1}{n} \log |A^n v| = \lambda_1 \ , \text{ for all } v \in \mathbb{R}^d \backslash \{0\} \ .$$
See [15, Theorem 3.4] for more details.
- (Continuity of Lyapunov exponents) Let $X = \{B_0, ..., B_m\} \subseteq GL(d)$ and $(p_0, ..., p_m)$ be an element of the open simplex
$$\Delta_m^o := \left\{ (p_0, p_1, ..., p_m) \in \mathbb{R}^{m+1} \mid 0 < p_i < 1 \text{ and } \sum_{i=0}^{m} p_i = 1 \right\} \ .$$
Consider a sequence of i.i.d. random matrices $\{A_0, A_1, ...\}$ taking values in X with probability given by
$$\mathbb{P}(A_i = B_j) = p_j \ .$$
Let $\lambda_1 \geq ... \geq \lambda_d$ be the Lyapunov exponents of the random matrix product formed by the sequence $\{A_0, A_1, ...\}$. Avila, Eskin, and Viana proved in [16] that these Lyapunov exponents depend continuously on the underlying linear cocycle:

Theorem 6.6. *For each* $1 \leq j \leq d$*, the number* λ_j *depends continuously on the* B_i *and and* p_i *at every point in the domain* $GL(d)^m \times \Delta_m^o$*.*

  In particular, Example 1.5 in the introduction demonstrates how continuity cannot be extended to the closed simplex.

6.2. **Schrödinger Cocycles.** In this subsection, we mention some preliminary applications of the multiplicative ergodic theorem to Schrödinger cocycles.

Note that irrational rotation map and the shift map are both ergodic (see, e.g., [12, Theorem 4.2.2] and [13, Theorem 1.12] for proof). Therefore, Corollary 6.3 applies to both random and quasi-periodic Schrodinger cocycles. In both cases, if the Lyapunov exponents exist, they must be constant.

Let $F_E : M \times \mathbb{R}^2 \to M \times \mathbb{R}^2$ be a Schrödinger cocycle as defined in Example 2.5, satisfying the desired integrability conditions of the multiplicative ergodic theorem. Then:

Corollary 6.7. *For* $\mu-$*almost every* $x \in M$*,*

(i) *Either there is a constant* $\lambda_\pm$ *such that*
$$\lim_{n\to\infty} \frac{1}{n} \log |A_E^n(x) v| = \lambda_\pm \quad \textit{for all } v \in \mathbb{R}^2 \backslash \{0\} \ ,$$
(ii) *or there are constants* $\lambda_+ > \lambda_-$*, and a line* E_x *such that*
$$\lim_{n\to\infty} \frac{1}{n} \log |A_E^n(x) v| = \begin{cases} \lambda_- & \textit{if } v \in E_x \backslash \{0\} \\ \lambda_+ & \textit{if } v \in \mathbb{R}^2 \backslash E_x \ . \end{cases}$$

The Lyapunov exponents of F_E tell us a lot about the the spectral properties of the operator $H : l^2 \to l^2$ in Example 2.5. For instance, if $\lambda_\pm \neq 0$, then E cannot be an eigenvalue of H. See [3, Chapter 2.1.3].

7. Acknowledgement

I really want to thank my mentor, Meg Doucette, for suggesting the readings, helping me to work through the details of the proofs, and always encouraging me to learn new materials that interest me the most. She is fantastic and I am really lucky to have her as my mentor. I also really want to thank Peter May for giving the algebraic topology lectures and organizing this wonderful REU. I am very grateful for his generous support throughout the program, and I hold a deep appreciation for the positive influence he has had on my professional and personal growth. I also really want to thank Professor Lawler for giving the analysis and probability lectures, arranging the weekly lunches, and helping me to think about probabilistic aspects of the multiplicative ergodic theorem.

References

[1] V. I. Oseledec, *A multiplicative ergodic theorem and Lyapunov characteristic numbers for dynamical systems,* Transactions of the Moscow mathematical Society 19, pp. 197-231, 1968.

[2] H. Furstenberg and H. Kesten. *Products of random matrices*, Annals of Mathematical Statistics, 1960.

[3] M. Viana, *Lectures on Lyapunov exponents*, Cambridge university press, Studies in advanced mathematics, Cambridge, 2014.

[4] P. Walters, *A dynamical proof of the multiplicative ergodic theorem*, Transactions of the American mathematical society, Volume 335, Number 1, 1993.

[5] J. Bochi, *The multiplicative ergodic theorem of Oseledets*, https://www.mat.uc.cl/~jairo.bochi/docs/oseledets.pdf (online lecture notes), 2008.

[6] M. S. Raghunathan, *A proof of Oseledec's multiplicative ergodic theorem*, Israel J. Math. 32, pp. 356-362, 1979.

[7] P. Duarte and S. Klein, *Lyapunov exponents of linear cocycles*, Atlantis studies in dynamical systems Vol. 3, Atlantis press, 2016.

[8] L. Barreira, *Lyapunov exponents*, Birkhäuser, Springer International Publishing AG, 2017.

[9] A. V. Arutyunov and V. Obukhovskii, *Convex and set-Valued analysis: selected topics*, De Gruyter, Berlin and Boston, https://doi.org/10.1515/9783110460308, 2017.

[10] C. Castaing and M. Valadier, *Convex analysis and measurable multifunctions*, Lecture Notes in Math., Vol. 580, Springer-Verlag, Berlin and New York, 1975.

[11] J. W. Milnor and J. D. Stasheff, *Characteristic classes*, TEXromancers, 2022.

[12] A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Cambridge University Press, Cambridge, doi:10.1017/CBO9780511809187, 1995.

[13] P. Walters, *An Introduction to ergodic theory*, Springer-Verlag, Newyork, 1982.

[14] A. Crisanti, G. Paladin, and A. Vulpiani, *Products of random matrices in statistical physics*, Springer-Verlag, Berlin and Heidelberg, 1993.

[15] P. Bougerol and J. Lacroix, *Products of Random Matrices with Applications to Schrödinger Operators*, Progress in probability and statistics Vol. 8, Boston, 1985.

[16] A. Avila, A. Eskin and M. Viana, *Continuity of the Lyapunov exponents of random matrix products*, arXiv preprint arXiv:2305.06009, 2023.

[17] D. Sutter, F. Omar, and R. Renner, *Bounds on Lyapunov exponents via entropy accumulation*, IEEE Transactions on Information Theory 67, no. 1, 2020.