Master's thesis

Ergodic theory and Pesin theory

Janne Junnila

June 26, 2012

Contents

1. Basic ergodic theory

1.1. Introduction

Let α be a real number. We'll start our journey into ergodic theory by first inspecting the distribution of the numbers $\alpha, 2\alpha, 3\alpha, \ldots$ modulo 1 on the unit interval. Later on in Section 1.3 we will take a second look at the same example by using Birkhoff's Ergodic Theorem (Theorem 1.3.1).

Definition 1.1.1. A sequence $(x_n)_{n\in\mathbf{N}}$ of real numbers is **equidistributed** modulo 1 if and only if for every interval $I \subset [0,1]$ we have

$$\lim_{n\to\infty} \frac{|\{k \in \{1,2,\ldots,n\} : x_k \bmod 1 \in I\}|}{n} = |I|. \qquad \square$$

Example 1.1.2. We'll prove that $(\alpha n)_{n\in\mathbf{N}}$ is equidistributed modulo 1 if and only if α is irrational.

Suppose first that $\alpha = \frac{p}{q}$ is rational. Let $x_n = \alpha n = \frac{pn}{q}$. Then $x_{q+m} \equiv x_m \pmod 1$, so that the sequence $(x_n)_{n\in\mathbf{N}}$ is periodic with period q, and thus consists of only finitely many different points. By taking an interval that doesn't contain any of these points, we see that $(\alpha n)_{n\in\mathbf{N}}$ is not equidistributed.

Suppose now that α is irrational and write $x_n = \alpha n$ for all $n \in \mathbf{N}$. We shall use a classical argument introduced by Weyl in [Wey16]. Let f be any continuous 1-periodic function on $\mathbf{R}$ and fix $\varepsilon > 0$. Then there exists a trigonometric polynomial

$$T(x) = \sum_{j=0}^{m} (a_j \sin(2\pi jx) + b_j \cos(2\pi jx)),$$

such that $|T - f| < \varepsilon$. (See for example [Rud87] Thm. 4.25.) Now

$$\begin{aligned}
\frac{1}{n}\sum_{k=1}^{n} f(x_k) &\le \frac{1}{n}\sum_{k=1}^{n} (T(x_k) + \varepsilon) \\
&= \frac{1}{n}\sum_{k=1}^{n}\sum_{j=0}^{m} a_j \sin(2\pi jx_k) + \frac{1}{n}\sum_{k=1}^{n}\sum_{j=0}^{m} b_j \cos(2\pi jx_k) + \varepsilon \\
&= \frac{1}{n}\sum_{j=0}^{m} a_j \sum_{k=1}^{n} \sin(2\pi jx_k) + \frac{1}{n}\sum_{j=0}^{m} b_j \sum_{k=1}^{n} \cos(2\pi jx_k) + \varepsilon.
\end{aligned}$$

Because α is irrational $e^{2\pi ij\alpha} \neq 1$ and we have the identity

$$\sum_{k=1}^{n} (\cos(2\pi jx_k) + i\sin(2\pi jx_k)) = \sum_{k=1}^{n} e^{2\pi ij\alpha k} = \frac{e^{2\pi ij(n+1)\alpha} - e^{2\pi ij\alpha}}{e^{2\pi ij\alpha} - 1},$$

the right-hand side of which we can estimate by

$$\left|\frac{e^{2\pi i j(n+1)} - e^{2\pi i j\alpha}}{e^{2\pi i j\alpha} - 1}\right| \leq \frac{2}{|e^{2\pi i j\alpha} - 1|} = \frac{1}{|\sin(\pi j\alpha)|}.$$

In particular we now have

$$\left|\sum_{k=1}^{n} \sin(2\pi j x_k)\right| \leq \frac{1}{|\sin(\pi j\alpha)|} \quad \text{and} \quad \left|\sum_{k=1}^{n} \cos(2\pi j x_k)\right| \leq \frac{1}{|\sin(\pi j\alpha)|}.$$

Thus for all large enough n we have

$$\left|\sum_{j=1}^{m} a_j \frac{1}{n} \sum_{k=1}^{n} \sin(2\pi j x_k)\right| < \varepsilon, \text{and}$$

$$\left|\sum_{j=1}^{m} b_j \frac{1}{n} \sum_{k=1}^{n} \cos(2\pi j x_k)\right| < \varepsilon.$$

For these n we then have

$$\frac{1}{n} \sum_{k=1}^{n} f(x_k) \leq b_0 + 3\varepsilon,$$

and in a similar way one may prove that for large enough n

$$\frac{1}{n} \sum_{k=1}^{n} f(x_k) \geq b_0 - 3\varepsilon.$$

Finally we get by integration that

$$\varepsilon \geq \left|\int_0^1 (T(x) - f(x))\, dx\right| = \left|b_0 - \int_0^1 f(x)\, dx\right|,$$

so that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(x_k) = \int_0^1 f(x)\, dx.$$

Let now $I = [a, b]$ be a closed subinterval of $[0, 1]$. Fix $\varepsilon > 0$ and let $f\colon [0, 1] \to \mathbf{R}$ be a continuous approximation of χ_I such that $\chi_I \leq f$ and

$$\int_0^1 f(x)\, dx \leq |I| + \varepsilon.$$

Then for large enough n we have

$$\frac{1}{n} \sum_{k=1}^{n} \chi_I(x_k \bmod 1) \leq \frac{1}{n} \sum_{k=1}^{n} f(x_k \bmod 1) \leq \int_0^1 f(x)\, dx + \varepsilon \leq |I| + 2\varepsilon.$$

Similarly if g is a continuous approximation of χ_I such that $\chi_I \geq g$ and

$$\int_0^1 g(x)\, dx \geq |I| - \varepsilon,$$

then for large enough n we have

$$\frac{1}{n}\sum_{k=1}^{n}\chi_I(x_k \bmod 1) \geq \frac{1}{n}\sum_{k=1}^{n} g(x_k \bmod 1) \geq \int_0^1 g(x)\,dx - \varepsilon \geq |I| - 2\varepsilon.$$

Therefore

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\chi_I(x_k \bmod 1) = |I|,$$

and the equidistribution of $(\alpha n)_{n\in\mathbf{N}}$ follows. □

The above example generalizes to the case where $\alpha \in \mathbf{R}^d$ is a vector. We have the following theorem by Weyl. (See [Wey16].)

The definition of equidistribution given in the beginning generalizes easily: Just use d-intervals.

Theorem 1.1.3. *Let $(\alpha_n)_{n=1}^{\infty} \subset [0,1]^d$ be a sequence of vectors and for each $n \in \mathbf{N}$ denote $\alpha_n = (\alpha_n^1, \alpha_n^2, \ldots, \alpha_n^d)$. Then α_n is equidistributed in $[0,1]^d$ if and only if for each $m \in \mathbf{Z}^d \setminus \{0\}$ we have*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} e^{2\pi i(m\cdot\alpha_k)} = 0. \tag{1.1.1}$$

Proof. Suppose first that the equation (1.1.1) holds for every $m \in \mathbf{Z}^d \setminus \{0\}$. Like in the example above, let $f\colon [0,1]^d \to \mathbf{R}$ be a continuous function, 1-periodic to each direction. Let us enumerate the points of $\mathbf{Z}^d$ as $0 = p_0, p_1, p_2, \ldots$. Then by the Stone-Weierstraß Theorem (See for example [Rud91], where it is a special case of Theorem 5.7.) for every fixed $\varepsilon > 0$ there exists a trigonometric polynomial $T\colon [0,1]^d \to \mathbf{R}$,

$$T(x) = \sum_{j=0}^{m} c_j e^{i(p_j\cdot x)},$$

so that $|T(x) - f(x)| < \varepsilon$ for all $x \in [0,1]^d$. Let $f_1 = T + \varepsilon$ and $f_2 = T - \varepsilon$. Then

$$\frac{1}{n}\sum_{k=1}^{n} f(\alpha_k) \leq \frac{1}{n}\sum_{k=1}^{n} f_1(\alpha_k) = \sum_{j=0}^{m} c_j \frac{1}{n}\sum_{k=1}^{n} e^{2\pi i(l_j\cdot\alpha_k)} + \varepsilon,$$

and from the assumption we get that for large enough n we have

$$\frac{1}{n}\sum_{k=1}^{n} f(\alpha_k) \leq c_0 + 2\varepsilon.$$

By a similar argument involving f_2 we get that

$$\frac{1}{n}\sum_{k=1}^{n} f(\alpha_k) \geq c_0 - 2\varepsilon.$$

The conclusion of the theorem follows again by approximating the characteristic function of an arbitrary d-interval $I \subset [0,1]^d$ by a continuous function.

On the other hand if α_n is equidistributed in $[0,1]^d$, we may first consider a characteristic function χ of some d-interval to get that

$$\frac{1}{n}\sum_{k=1}^{n}\chi(\alpha_k) \to \int_{[0,1]^d}\chi$$

as $n \to \infty$. Thus we have the result for finite linear combinations of characteristic functions. Now let f be a continuous function on $[0,1]^d$. Then we can find a sequence of simple functions f_n uniformly converging to f in L^∞. Fix $\varepsilon > 0$ and let $m \in \mathbf{N}$ be such that for all $n \geq m$ we have $|f_n - f_m| < \varepsilon$. Then we have the following

$$\left|\frac{1}{n}\sum_{k=1}^{n} f(\alpha_k) - \int f\right| \leq \left|\frac{1}{n}\sum_{k=1}^{n}(f(\alpha_k) - f_n(\alpha_k))\right| + \left|\frac{1}{n}\sum_{k=1}^{n} f_n(\alpha_k) - \int f_n\right| + \left|\int f_n - \int f\right|.$$

The first as well as the last term on the right hand side clearly go to 0 as $n \to \infty$. For the middle term we note that

$$\left|\frac{1}{n}\sum_{k=1}^{n} f_n(a_k) - \int f_n\right| \leq \left|\frac{1}{n}\sum_{k=1}^{n} f_m(\alpha_k) - \int f_m\right| + \frac{1}{n}\sum_{k=1}^{n}|f_n(\alpha_k) - f_m(\alpha_k)| + \int |f_n - f_m|.$$

Here again the first term goes to 0 as $n \to \infty$, the second and third terms are not larger than ε. Thus

$$\frac{1}{n}\sum_{k=1}^{n} f(\alpha_k) \to \int f.$$

By taking $f(x) = e^{2\pi i(m\cdot x)}$ we get the result. ∎

Corollary 1.1.4. *For every $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbf{R}^d$ let $\{\alpha\}$ denote the vector $(\{\alpha_1\}, \ldots, \{\alpha_d\})$ where $\{x\}$ is the fractional part of the real number x, defined by $\{x\} = x - \lfloor x \rfloor$. Then the sequence $\{1\alpha\}, \{2\alpha\}, \ldots$ is equidistributed in $[0,1]^d$ if and only if there doesn't exist $m \in \mathbf{Z}^d \setminus \{0\}$ such that $m \cdot \alpha \in \mathbf{Z}$.* □

Proof. This is now evident since if $m \in \mathbf{Z}^d \setminus \{0\}$, then

$$\left|\sum_{k=1}^{n} e^{2\pi i(m\cdot k\alpha)}\right| = \left|\frac{e^{2\pi i(n+1)(m\cdot\alpha)} - e^{2\pi i(m\cdot\alpha)}}{e^{2\pi i(m\cdot\alpha)} - 1}\right| \leq \frac{2}{|e^{2\pi i(m\cdot\alpha)} - 1|} \leq \frac{1}{|\sin(\pi(m\cdot\alpha))|}.$$ ∎

1.2. Measure-preserving transformations and ergodicity

We shall study measure-preserving transformations and mappings that are invariant with respect to them. The basic definitions and results are taken from the book by Einsiedler and Ward [EW11].

Definition 1.2.1. Let $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$ be probability spaces.

- A measurable map $T\colon X \to Y$ is said to be **measure-preserving** if and only if $\mu(T^{-1}E) = \nu(E)$ for every $E \in \mathcal{C}$.
- A measure-preserving map from the space to itself is a **measure-preserving transformation**.
- The quadruple $(X, \mathcal{B}, \mu, T)$ is called a **measure-preserving system**.
- If T is a measure-preserving transformation for the space $(X, \mathcal{B}, \mu)$, then μ is called T-invariant.
- A measurable function $f\colon X \to Y$ is called T-invariant if and only if $f \circ T = f$ almost everywhere.
- Finally a measurable subset $A \in \mathcal{B}$ is called T-invariant if and only if $\mu(T^{-1}A \Delta A) = 0$. □

Here $A \Delta B$ is the symmetric difference, $A \Delta B = (A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$.

It will be convenient to fix some notation before looking at the basic lemmas and theorems. Let $(X, \mathcal{B}, \mu)$ be a probability space. For $1 \le p < \infty$ we'll denote by $\mathcal{L}^p(X, \mu) = \mathcal{L}^p(X)$ the space of measurable functions $f\colon X \to \mathbf{R}$ that satisfy

$$\int |f|^p < \infty.$$

The space of measurable bounded functions will be denoted by $\mathcal{L}^\infty(X)$. Finally for $1 \le p \le \infty$ we'll denote by $L^p(X, \mu) = L^p(X)$ the quotient space of $\mathcal{L}^p(X)$ where almost everywhere equal functions are identified.

The following lemma gives a characterization of T-invariant measures.

Lemma 1.2.2. *Let $(X, \mathcal{B}, \mu)$ be a probability space and $T\colon X \to X$ a measurable map. A measure μ is T-invariant if and only if for every $f \in \mathcal{L}^\infty(X)$ we have*

$$\int f \, d\mu = \int (f \circ T) \, d\mu. \tag{1.2.1}$$

Moreover if μ is T-invariant, then the equation (1.2.1) holds for every $f \in L^1(X)$. □

Proof. Suppose first that the equation (1.2.1) holds for every $f \in \mathcal{L}^\infty(X)$. Then in particular it holds for characteristic functions, and thus for every $E \in \mathcal{B}$ we have

$$\mu(E) = \int \chi_E \, d\mu = \int (\chi_E \circ T) \, d\mu = \int \chi_{T^{-1}E} \, d\mu = \mu(T^{-1}E).$$

Suppose then that μ is T-invariant. Then the equation (1.2.1) holds for linear combinations of characteristic functions by the above calculation. Therefore (1.2.1) holds for all simple functions. Now if $f \in \mathcal{L}^1(X)$ is non-negative, we may choose an increasing sequence of simple functions $(f_k)_{k=1}^{\infty}$ converging to f. Then $(f_k \circ T)_{k=1}^{\infty}$ is an increasing sequence of simple functions converging to $f \circ T$ and we get

$$\int f \, d\mu = \lim_{k\to\infty} \int f_k \, d\mu = \lim_{k\to\infty} \int f_k \circ T \, d\mu = \int f \circ T \, d\mu.$$

Finally if $f \in \mathcal{L}^1(X)$, then the result follows by using the formula above for f^+ and f^-:

$$\begin{aligned} \int f \, d\mu &= \int f^+ \, d\mu - \int f^- \, d\mu \\ &= \int f^+ \circ T \, d\mu - \int f^- \circ T \, d\mu \\ &= \int f \circ T \, d\mu. \end{aligned}$$

Notice that since $\mathcal{L}^\infty(X) \subset \mathcal{L}^1(X)$, we get the result also for $f \in \mathcal{L}^\infty(X)$. ∎

Before going further, we'd like to present the following important result. The proof is a slightly modified version of the one found in [EW11].

Theorem 1.2.3 (Poincaré Recurrence Theorem). *If $(X, \mathcal{B}, \mu, T)$ is a measure-preserving system and $A \subset X$ is a measurable subset, then for almost every $x \in A$ the sequence $x, Tx, T^2x, \ldots$ visits A infinitely many times.* □

By visiting *we mean that infinitely many of the points $x, Tx, T^2x, \ldots$ belong to A.*

Proof. Consider the set

$$N = \{x \in X : x \in A, T^n x \notin A \text{ for all } n \in \mathbf{N}\}.$$

Now

$$T^{-j}N = \{x \in X : T^j x \in A, T^{n+j}x \notin A \text{ for all } n \in \mathbf{N}\}.$$

Clearly $x \in T^{-j}N$ implies $x \notin T^{-j-n}N$ for all $n \geq 1$, which implies that the sets $N, T^{-1}N, T^{-2}N, \ldots$ are disjoint. Because T is measure-preserving, this implies $\mu(N) = 0$. Let $E = \bigcup_{j\in\mathbf{N}} T^{-j}N$. Then clearly $\mu(E) = 0$. Now if $x \in A$ is such that $x, Tx, T^2x, \ldots$ visits A only finitely many times, then $x \in E$. The claim follows. ∎

We'll next define what we mean by an ergodic transformation. The basic idea is, that a transformation $T: X \to X$ is called ergodic if you can't find a measurable subset A of X, which is invariant under T (that is a set $A \in \mathcal{B}$ such that $T^{-1}A = A$), unless A is either almost the whole space or almost empty. Stated precisely, we have the following definition.

Definition 1.2.4. Let $T: X \to X$ be a measure-preserving transformation. It is called **ergodic** if and only if for every $E \in \mathcal{B}$ that satisfies $T^{-1}E = E$ we have either $\mu(E) = 0$ or $\mu(E) = 1$. □

REMARK 1.2.5. *In some contexts we wish to let the measure vary keeping the transformation fixed. In that case it is convenient to call a* measure *ergodic if the condition in the definition holds for the fixed transformation and the given measure.* □

We make the following note, which shows that we could have defined ergodicity by considering T-invariant sets instead of the **strictly invariant** sets in the definition.

LEMMA 1.2.6. *A measure-preserving transformation $T\colon X \to X$ is ergodic if and only if for every T-invariant $E \in \mathcal{B}$ we have either $\mu(E) = 0$ or $\mu(E) = 1$.* □

Proof. Clearly if the second condition is satisfied, then T is ergodic, so suppose that T is ergodic. Let $E \in \mathcal{B}$ be T-invariant. Form the set

$$E' = \bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} T^{-k}E.$$

Then $T^{-1}E' = \bigcap_{n=0}^{\infty} \bigcup_{k=n+1}^{\infty} T^{-k}E = E'$, so that E' is strictly T-invariant. Hence $\mu(E') = 0$ or $\mu(E') = 1$ by assumption. It remains to show that $\mu(E') = \mu(E)$. Notice that for all $n \in \mathbf{N}$ we have

$$E \Delta \bigcup_{k=n}^{\infty} T^{-k}E \subset \bigcup_{k=n}^{\infty} E \Delta T^{-k}E.$$

Moreover $\mu(E \Delta T^{-k}E) = 0$ for all k, because

$$E \Delta T^{-k}E \subset \bigcup_{i=0}^{k-1} T^{-i}E \Delta T^{-(i+1)}E.$$

Since the sets

$$E_n = \bigcup_{k=n}^{\infty} T^{-k}E, \quad n = 1, 2, \ldots,$$

form a decreasing sequence, and because we have $\mu(E_n \Delta E) = 0$ for all n, $\mu(E' \Delta E) = 0$ follows, and therefore $\mu(E) = \mu(E')$. ■

The following characterisation of ergodic transformations will be useful later on.

THEOREM 1.2.7. *A measure-preserving transformation $T\colon X \to X$ is ergodic if and only if every T-invariant function $f\colon X \to \mathbf{R}$ is constant almost everywhere.* □

The proof we give is a combination of the proofs found in [Fal97] and [EW11].

Proof. Suppose first that T is ergodic and let $f\colon X \to \mathbf{R}$ be a T-invariant function. Fix $\lambda \in \mathbf{R}$ and define

$$A_\lambda = \{x \in X : f(x) < \lambda\}.$$

Because $f \circ T = f$ almost everywhere, we have $\mu(T^{-1}A_\lambda \Delta A_\lambda) = 0$, and hence by ergodicity $\mu(A_\lambda) = 0$ or $\mu(A_\lambda) = 1$. This means that either $f(x) < \lambda$ for almost every $x \in X$ or $f(x) \geq \lambda$ for almost every x. Because λ is arbitrary, the result follows.

Suppose then that every T-invariant function f is constant almost everywhere. Let $A \subset X$ be a measurable subset with $T^{-1}A = A$. Then the characteristic function χ_A is a T-invariant function. Hence $\chi_A = 1$ or $\chi_A = 0$ almost everywhere, and thus $\mu(A) = 0$ or $\mu(A) = 1$. ∎

The next useful lemma will be used without referencing to throughout the thesis.

LEMMA 1.2.8. *Suppose that $(X, \mathcal{B}, \mu, T)$ is a measure-preserving system, $(Y, \mathcal{C}, \nu)$ is some probability space and $\lambda \colon X \to Y$ is T-invariant. Then for almost every $x \in X$*

$$\lambda(T^k x) = \lambda(x)$$

for every $k \in \mathbf{N}$. □

Proof. Consider the set

$$A = \{x \in X : \lambda(T^k x) = \lambda(x) \text{ for all } k \in \mathbf{N}\}.$$

Since for each $x \in X \setminus A$ there is the smallest $k(x) \in \mathbf{N}$ that satisfies

$$\lambda(T^{k(x)} x) \neq \lambda(x),$$

we have

$$\begin{aligned}
\mu(X \setminus A) &= \mu\left(\bigcup_{k=1}^{\infty} \{x \in X : \lambda(T^k x) \neq \lambda(x)\}\right) \\
&= \mu\left(\bigcup_{k=1}^{\infty} \{x \in X : \lambda(T^k x) \neq \lambda(T^{k-1} x)\}\right) \\
&\leq \sum_{k=1}^{\infty} \mu(\{x \in X : \lambda(T^k x) \neq \lambda(T^{k-1} x)\}) \\
&= \sum_{k=1}^{\infty} \mu(T^{-k+1}\{x \in X : \lambda(Tx) \neq \lambda(x)\}) = 0.
\end{aligned}$$ ∎

1.3. Birkhoff's ergodic theorem

Intuitively it would seem that under an ergodic transformation, almost all points would have to move somehow uniformly when the transformation is iterated. This is the content of Birkhoff's Ergodic Theorem, the proof of which we borrow from [Mic06].

THEOREM 1.3.1 (BIRKHOFF'S ERGODIC THEOREM). *Let $f \in L^1(X)$ be an integrable function on a measure-preserving system $(X, \mathcal{B}, \mu, T)$. Then the limit*

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

exists for almost every $x \in X$. Moreover

$$\frac{1}{n}\sum_{k=0}^{n-1}(f \circ T^k) \xrightarrow{L^1} f^*,$$

and f^* is T-invariant. We have $\int f = \int f^*$, and in the case that T is ergodic, $f^* = \int f$ almost everywhere. □

Before proving Theorem 1.3.1 we'll prove the following Maximal Ergodic Theorem, which is used in the proof. Let us first fix some notation. Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system. For every $f \in L^1(X)$ define

$$A_n f = \frac{1}{n}\sum_{k=0}^{n-1}(f \circ T^k), \quad S_N f = \sup_{1\le n\le N} A_n f, \quad Sf = \sup_N S_N f,$$

and

$$f^* = \limsup_{n\to\infty} A_n f, \quad f_* = \liminf_{n\to\infty} A_n f.$$

Theorem 1.3.2. *Let $f \in L^1(X)$, and suppose that λ is a T-invariant function on X with $\lambda^+ \in L^1(X)$. Define $E_\lambda = \{x \in X : Sf(x) > \lambda(x)\}$, then*

$$\int_{E_\lambda}(f - \lambda) \ge 0.$$

□

The reason for the name "Maximal Ergodic Theorem" is justified by the following remark, which connects this theorem to the Hardy-Littlewood maximal function/inequality.

Remark 1.3.3. *Recall that the Hardy-Littlewood maximal function Mf for a function $f\colon \mathbf{R}^n \to \mathbf{R}$ is defined by*

$$Mf(x) = \sup_{x\in B}\frac{1}{|B|}\int_B |f|,$$

where the supremum is taken over all balls $B \subset \mathbf{R}^n$ containing x. Here $|B|$ is the Lebesgue-measure of the ball. In the case that $f \in L^1(\mathbf{R}^n)$ the Hardy-Littlewood maximal inequality states that

$$|\{x \in \mathbf{R}^n : Mf(x) > \lambda\}| < \frac{C}{\lambda}\|f\|_{L^1}$$

for some constant C depending on the dimension n.

Now in the setting of a measure-preserving system we have an analogous result as an immediate corollary of Theorem 1.3.2. Namely: Here Sf plays the role of Mf and if we let $\lambda(x) = \lambda$ be a constant in the theorem, then

$$E_\lambda = \{x \in X : Sf(x) > \lambda\}$$

and the theorem gives us

$$\int |f| \ge \int_{E_\lambda} f \ge \lambda\mu(E_\lambda),$$

or

$$\mu(\{x \in X : Sf(x) > \lambda\}) \le \frac{1}{\lambda}\|f\|_{L^1}.$$

□

We'll now prove Theorem 1.3.2.

Proof. First notice that the theorem is clearly satisfied if $\lambda \notin L^1(E_\lambda)$, for then (since $\lambda^+, f \in L^1(E_\lambda)$)

$$\int_{E_\lambda} (f - \lambda) = \infty \geq 0.$$

So we may assume that $\lambda \in L^1(E_\lambda)$. But then in fact $\lambda \in L^1(X)$, since on $X \setminus E_\lambda$ we have $f \leq \lambda$, so that $\lambda^- \leq \lambda^+ - f$, which gives that λ^- is integrable.

We'll first assume that $f \in L^\infty(X)$. Consider the set

$$E_N = \{x \in X : S_N f(x) > \lambda(x)\}$$

for some fixed $N \in \mathbf{N}$. Then we have

$$(f - \lambda)\chi_{E_N} \geq f - \lambda, \tag{1.3.1}$$

since if $x \notin E_N$, then $f(x) \leq \lambda(x)$.

Let

$$\Lambda = \{x \in E_N : \lambda(x) = \lambda(T^k x) \text{ for all } k \in \mathbf{N}\}.$$

Because λ is T-invariant, we have $\mu(E_N \setminus \Lambda) = 0$. Assume now that $x \in \Lambda$ and $T^m x \in E_N$ for some $m \in \mathbf{N}$. Then

$$\sup\{(A_n f)(T^m x) : 1 \leq n \leq N\} > \lambda(T^m(x))$$

and there exists $n \in \{1, 2, \ldots, N\}$ such that

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^{m+k}x) > \lambda(T^m(x)),$$

or equivalently (because $x \in \Lambda$),

$$\sum_{k=m}^{m+n-1} (f - \lambda)(T^k x) > 0. \tag{1.3.2}$$

Consider now an arbitrary $M > N$ and $x \in \Lambda$. By (1.3.1) and (1.3.2), we can rearrange the sum

$$\sum_{k=0}^{M-1} (f - \lambda)\chi_{E_N}(T^k x)$$

into $j + 1$ sums

$$\sum_{k=0}^{m_1-1} (f-\lambda)\chi_{E_N}(T^k x) + \sum_{k=m_1}^{m_2-1} (f-\lambda)\chi_{E_N}(T^k x) + \cdots + \sum_{k=m_j}^{M-1} (f-\lambda)\chi_{E_N}(T^k x),$$

where $1 \leq m_1 < m_2 < \cdots < m_j \leq M - 1$, and each of the j subsums apart from the last one either contains only zero terms (when the starting term doesn't belong to E_N), or it is positive and has at most N terms. Moreover we can arrange it so that the last subsum also either consists of only zeros, or it has at most N terms. It's not necessarily positive, however. Hence we have the following inequality

$$\sum_{k=0}^{M-1}(f-\lambda)\chi_{E_N}(T^k x) \geq \sum_{k=m_j}^{M-1}(f-\lambda)\chi_{E_N}(T^k x) \geq -N(\|f\|_\infty + \lambda^+(x)),$$

which holds for all $x \in \Lambda$ and hence for almost every $x \in E_N$. Integrating both sides we get by Lemma 1.2.2 that

$$M\int_{E_N}(f-\lambda) \geq -N(\|f\|_\infty + \|\lambda^+\|_1).$$

Now dividing by M and letting $M \to \infty$ gives us

$$\int_{E_N}(f-\lambda) \geq 0.$$

Finally by letting $N \to \infty$ and using the Dominated Convergence Theorem (with dominant $|f| + |\lambda|$) we get

$$0 \leq \int (f-\lambda)\chi_{E_N} \to \int (f-\lambda)\chi_{E_\lambda} = \int_{E_\lambda}(f-\lambda).$$

Suppose now that $f \in L^1(X)$. For each $k \in \mathbf{N}$ define the function $\phi_k \colon X \to \mathbf{R}$ by $\phi_k = f\chi_{\{x \in X : |f(x)| \leq k\}}$. Then $\phi_k \in L^\infty$, so in particular what was done above applies. We have $\phi_k \to f$ almost everywhere. Because $\mu(|f| > k) \to 0$ as $k \to \infty$, we also have that

$$\mu(\{x \in X : f(x) \neq \phi_k(x)\}) \to 0.$$

Now if $x \in X$ satisfies

$$S_N\phi_k(x) \neq S_N f(x),$$

then one easily sees that $S_N|f-\phi_k|(x) \geq |S_N f(x) - S_N\phi_k(x)| > 0$ and hence

$$\{x \in X : S_N\phi_k(x) \neq S_N f(x)\} \subset \{x \in X : S_N|f-\phi_k|(x) > 0\}.$$

Because T is measure-preserving,

$$\lim_{k\to\infty}\mu(\{x \in X : f(T^n x) \neq \phi_k(T^n x)\}) = 0$$

for all $n \in \mathbf{N}$, and we obtain (for a fixed N)

$$\mu(\{x \in X : S_N\phi_k(x) \neq S_N f(x)\}) \to 0.$$

Write

$$F_{N,k} = \{x \in X : S_N\phi_k(x) > \lambda(x)\}, \quad E_N = \{x \in X : S_N f(x) > \lambda(x)\}.$$

Now

$$\mu(F_{N,k}\Delta E_N) \to 0, \text{ as } k \to \infty,$$

and $\chi_{F_{N,k}} \xrightarrow{L^1} \chi_{E_N}$. Therefore there is a subsequence k_i that satisfies $\chi_{F_{N,k_i}} \to \chi_{E_N}$ almost everywhere. Thus by the Dominated Convergence Theorem we have as $k \to \infty$ that

$$0 \le \int_{F_{N,k_i}} (\phi_{k_i} - \lambda) = \int (\phi_{k_i} - \lambda)\chi_{F_{N,k_i}} \to \int (f - \lambda)\chi_{E_N}.$$

The result follows finally by letting $N \to \infty$ and using the Dominated Convergence Theorem once more. ∎

We are now ready to prove Birkhoff's Ergodic Theorem. We start with the following lemma.

Lemma 1.3.4. *The functions f^* and f_* are T-invariant.* □

Proof. We can write

$$\frac{n+1}{n}\left(\frac{1}{n+1}\sum_{k=0}^{n} f(T^k x)\right) = \frac{1}{n}\sum_{k=0}^{n-1} f(T^k(Tx)) + \frac{1}{n}f(x).$$

Now if we consider a subsequence n_k such that the left hand side tends to $\limsup A_n f$, we get $f^* \le f^* \circ T$. Similarly if we consider a subsequence n_k such that the right hand side tends to $\limsup A_n f \circ T$, we get $f^* \ge f^* \circ T$. A similar calculation holds for f_*. ∎

Proof. We'll first prove the almost everywhere convergence of $A_n f$. It is enough to show that f^* is integrable and $\int f^* \le \int f$ because then by applying this to $-f$ we have

$$\begin{aligned} -\int f_* \le -\int f \Rightarrow \int f^* \le \int f \le \int f_* \le \int f^* \\ \Rightarrow \int (f^* - f_*) = \int |f^* - f_*| = 0, \end{aligned}$$

and thus $f^* = f_*$ almost everywhere. This proves the claim.

We note that $\lambda_N = \min\{(f^+)^*, N\} - \frac{1}{N}$ is T-invariant for all N since $(f^+)^*$ is T-invariant by Lemma 1.3.4. Now

$$\{x \in X : (Sf^+)(x) > \lambda_N(x)\} = X$$

so that by Theorem 1.3.2 we get

$$\int f^+ \ge \int \lambda_N \to \int (f^+)^*, \quad N \to \infty.$$

Because $(f^+)^* \ge (f^*)^+$, we conclude that $(f^*)^+$ is integrable. In a similar manner we get that $(f^*)^-$ is integrable (because $(f^-)^* \ge (f^*)^-$), so that also f^* is integrable. Let now $\varepsilon > 0$ be arbitrary and apply Theorem 1.3.2 to $\lambda = f^* - \varepsilon$ to conclude that

$$\int f \ge \int \lambda \to \int f^*, \quad \varepsilon \to 0.$$

For L^∞ functions the L^1 convergence follows from the the Dominated Convergence Theorem, since if $|f| \le M$ for some $M > 0$, then also

$$|A_n f| \le A_n|f| \le M$$

so that $|A_n f - f^*|$ has an integrable dominant $M + |f^*|$. Now if $f \in L^1$ and $\varepsilon > 0$, we can find $g \in L^\infty$ such that $\int |f - g| \le \varepsilon$. It follows that

$$\int |g^* - f^*| = \int \lim_{n\to\infty} |A_n(g - f)| \le \liminf_{n\to\infty} \int A_n|g - f| \le \varepsilon,$$

and hence

$$\begin{aligned}\int |A_n f - f^*| &\le \int (|A_n f - A_n g| + |A_n g - g^*| + |g^* - f^*|) \\ &\le \int A_n|f - g| + \int |A_n g - g^*| + \int |g^* - f^*| \\ &\le 2\varepsilon + \int |A_n g - g^*|.\end{aligned}$$

Since ε is arbitrary and $\int |A_n g - g^*| \to 0$ as $n \to \infty$, we get the L^1 convergence for $A_n f$.

Since f^* is T-invariant (Again by Lemma 1.3.4), it follows from Theorem 1.2.7 that f^* is constant almost everywhere if T is ergodic. ∎

Let's return to Example 1.1.2 and prove it by using Theorem 1.3.1.

Example 1.3.5. For all $\alpha \in \mathbf{R}$ define the measure-preserving transformation $R_\alpha x = x + \alpha$ on the unit circle $S^1 = \mathbf{R}/\mathbf{Z}$ with the usual Lebesgue measure normalized to 1. We shall show that R_α is ergodic if α is irrational. Indeed suppose that $A \subset S^1$ is an invariant set so that $\chi_A \circ R_\alpha = \chi_A$. Let $\sum_{n=-\infty}^{\infty} a_n e^{2\pi i n x}$ be the Fourier series of χ_A. Since χ_A and $\chi_A \circ R_\alpha$ are in L^2, the Fourier series are unique and thus the Fourier series $\sum_{n=-\infty}^{\infty} a_n e^{2\pi i n\alpha} e^{2\pi i n x}$ of $\chi_A \circ R_\alpha$ must equal that of χ_A. I.e. we must have $a_n e^{2\pi i n\alpha} = a_n$ for all $n \in \mathbf{Z}$. This implies that either $a_n = 0$ or $e^{2\pi i n\alpha} = 1$. The latter can only happen if $n = 0$. Hence $a_n = 0$ for all $n \ne 0$ and thus χ_A is constant almost everywhere, $\chi_A = 0$ or $\chi_A = 1$, meaning that $|A| = 0$ or $|A| = 1$. This proves the ergodicity of R_α.

Now by applying Theorem 1.3.1 to a characteristic function of a measurable subset A of S^1, we see that for almost all x the points $x + \alpha, x + 2\alpha, \ldots$ satisfy

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} \chi_A(x + k\alpha) \to |A|. \tag{1.3.3}$$

Lemma 1.3.6. *In the case where $A = (a, b)$ is an interval, (1.3.3) holds for all x. Hence $\alpha, 2\alpha, \ldots$ are equidistributed in S^1.* □

Proof. Without loss of generalization we'll prove the claim for $x = 0$. Choose $\varepsilon > 0$ and consider the interval $A' = (a + \varepsilon, b - \varepsilon)$. By Theorem 1.3.1 there exists some number $x_\varepsilon < \varepsilon$ that satisfies

$$\frac{1}{n} \sum_{k=0}^{n} \chi_E(x_\varepsilon + k\alpha) \to |E|$$

for every measurable set E. Moreover $x_\varepsilon + k\alpha \in (a+\varepsilon, b-\varepsilon)$ implies $k\alpha \in (a,b)$ and thus

$$\frac{1}{n}\sum_{k=0}^{n}\chi_A(k\alpha) \geq \frac{1}{n}\sum_{k=0}^{n}\chi_{A'}(x_\varepsilon + k\alpha).$$

Similarly approximating with a larger interval $A'' = (a-\varepsilon, b+\varepsilon)$ we get

$$\frac{1}{n}\sum_{k=0}^{n}\chi_A(k\alpha) \leq \frac{1}{n}\sum_{k=0}^{n}\chi_{A''}(x_\varepsilon + k\alpha).$$

Letting $n \to \infty$ and $\varepsilon \to 0$ we get what was claimed. ∎

One also sees easily that if α is not irrational, then R_α is not ergodic. Indeed let $\alpha = \frac{p}{q}$ and look at the set $[0, \frac{1}{2q}] \cup [\frac{1}{q}, \frac{3}{2q}] \cup \cdots \cup [\frac{q-1}{q}, \frac{2q-1}{2q}]$. It is clearly invariant under R_α but its measure is not 0 or 1. □

1.4. Conditional expectation and the invariant σ-algebra

To go further, we will first need some results on conditional expectation. The proofs are again mainly taken from [EW11]. Let us first look at the definitions.

Definition 1.4.1. Given a probability space $(X, \mathcal{B}, \mu)$, a function $f \in L^1$ and a σ-algebra $\mathcal{A} \subset \mathcal{B}$, we define a function $\mathbf{E}(f|\mathcal{A})\colon X \to \mathbf{R}$ called the **conditional expectation** by requiring that it is $\mathcal{A}$-measurable and the following equation holds:

$$\int_A \mathbf{E}(f|\mathcal{A}) = \int_A f, \quad \text{for all } A \in \mathcal{A}. \tag{1.4.1}$$

Theorem 1.4.2. *An $\mathcal{A}$-measurable function satisfying* (1.4.1) *exists and is unique up to a set of measure* 0. □

Proof. We will use the Radon-Nikodym Theorem. (See [Rud87], Theorem 6.9.) Indeed, define a measure ν on $\mathcal{A}$ by setting

$$\nu(A) = \int_A f d\mu.$$

Then $\nu \ll \mu|\mathcal{A}$ and hence by Radon-Nikodym theorem there exists an $\mathcal{A}$-measurable function $E(f|\mathcal{A})$ that satisfies

The notation $\nu \ll \mu$ means that ν is absolutely continuous with respect to μ.

$$\int_A f d\mu = \nu(A) = \int_A \mathbf{E}(f|\mathcal{A}) d\mu|\mathcal{A} = \int_A \mathbf{E}(f|\mathcal{A}) d\mu.$$

To prove the uniqueness, suppose that $E'(f|\mathcal{A})$ is also $\mathcal{A}$-measurable and satisfies

$$\int E'(f|\mathcal{A}) d\mu = \int_A f d\mu$$

for all $A \in \mathcal{A}$. Then the set

$$A = \{x \in X : E(f|\mathcal{A})(x) < E'(f|\mathcal{A})\}$$

is in $\mathcal{A}$ and

$$\int_A E(f|\mathcal{A}) d\mu = \int_A f d\mu = \int_A E'(f|\mathcal{A}) d\mu.$$

This shows that $\mu(A) = 0$ and similarly we have

$$\mu(\{x \in X : E(f|\mathcal{A})(x) > E'(f|\mathcal{A})\}) = 0. \qquad \blacksquare$$

The following theorem is used in the next chapter, when we prove the Subadditive Ergodic Theorem (Theorem 2.2.1).

THEOREM 1.4.3. *The map*

$$\mathbf{E}(\cdot|\mathcal{A}): L^1(X, \mathcal{B}, \mu) \to L^1(X, \mathcal{A}, \mu)$$

is a bounded linear operator with norm equal to 1. □

Proof. The linearity follows by uniqueness: Let $f, g \in L^1(X, \mathcal{A}, \mu)$, $c \in \mathbf{R}$ and notice that

$$\int_A E(cf + g|\mathcal{A}) = c\int_A f + \int_A g = \int_A cE(f|\mathcal{A}) + E(g|\mathcal{A}) d\mu.$$

Moreover if we write $f = f^+ - f^-$, then

$$\begin{aligned}\int |\mathbf{E}(f|\mathcal{A})| d\mu &= \int |\mathbf{E}(f^+|\mathcal{A}) - \mathbf{E}(f^-|\mathcal{A})| d\mu \\ &\leq \int \mathbf{E}(f^+|\mathcal{A}) + \mathbf{E}(f^-|\mathcal{A}) d\mu \\ &= \int f^+ + f^- d\mu = \int |f| d\mu,\end{aligned} \qquad \blacksquare$$

so $\mathbf{E}(\cdot|\mathcal{A})$ has norm 1.

We'd also like to present a few equivalents of the classical convergence theorems in the conditional expectation setting. Starting with Monotone Convergence Theorem.

THEOREM 1.4.4. *Suppose that $f_k \geq 0$ is an increasing sequence of integrable functions converging almost everywhere to $f \in L^1(X, \mathcal{B}, \mu)$. Then*

$$\mathbf{E}(f_k|\mathcal{A}) \to \mathbf{E}(f|\mathcal{A})$$

almost everywhere. □

Proof. Since f_k form an increasing sequence, so do $\mathbf{E}(f_k|\mathcal{A})$: Suppose that $k \leq n$. Then we have

$$\int_A \mathbf{E}(f_k|\mathcal{A})\, d\mu = \int_A f_k\, d\mu \leq \int_A f_n\, d\mu = \int_A \mathbf{E}(f_n|\mathcal{A})\, d\mu$$

for all $A \in \mathcal{A}$, which implies that $\mathbf{E}(f_k|\mathcal{A}) \le \mathbf{E}(f_n|\mathcal{A})$ almost everywhere.

Because the sequence $\mathbf{E}(f_k|\mathcal{A})$ is monotonous, there exists a measurable function g so that $\mathbf{E}(f_k|\mathcal{A}) \to g$ almost everywhere. Hence

$$\int_A g(x)\,d\mu = \lim_{k\to\infty}\int_A \mathbf{E}(f_k|\mathcal{A})\,d\mu = \lim_{k\to\infty}\int_A f_k\,d\mu = \int_A f\,d\mu$$

for all $A \in \mathcal{A}$ by the Monotone Convergence Theorem. But by the uniqueness of the conditional expectation we must now have $g = \mathbf{E}(f|\mathcal{A})$ almost everywhere. ∎

Next we'll prove Fatou's lemma for conditional expectations.

THEOREM 1.4.5. *Let $f_k \ge 0$ be integrable functions so that* $\liminf_{k\to\infty} f_k$ *is also integrable. Then*

$$\mathbf{E}(\liminf_{k\to\infty} f_k|\mathcal{A}) \le \liminf_{k\to\infty} \mathbf{E}(f_k|\mathcal{A})$$

almost everywhere. □

Proof. Define $F_n = \inf_{k\ge n} f_k$. Then F_n form an increasing sequence converging monotonously to $\liminf_{k\to\infty} f_k$. Hence by the monotonous convergence theorem we have

$$\mathbf{E}(F_n|\mathcal{A}) \to \mathbf{E}(\liminf_{k\to\infty} f_k|\mathcal{A})$$

as $n \to \infty$. In addition the monotonicity implies almost everywhere that

$$\mathbf{E}(F_n|\mathcal{A}) \le \inf_{k\ge n} \mathbf{E}(f_k|\mathcal{A}) \le \liminf_{k\to\infty} \mathbf{E}(f_k|\mathcal{A}),$$

which proves the result. ∎

Theorem 1.4.3 shows that if $f_k \in L^1(X,\mathcal{B},\mu)$ is a sequence of functions so that $f_k \xrightarrow{L^1} f \in L^1(X,\mathcal{B},\mu)$, then $\mathbf{E}(f_k|\mathcal{A}) \xrightarrow{L^1} \mathbf{E}(f|\mathcal{A})$. The following theorem (Dominated Convergence Theorem for Conditional Expectations) shows that actually the convergence also holds almost everywhere if $f_k \to f$ almost everywhere and the sequence f_k has an integrable dominant.

THEOREM 1.4.6. *Suppose that $f_k \in L^1(X,\mathcal{B},\mu)$ is a sequence of functions that converges to $f \in L^1(X,\mathcal{B},\mu)$ almost everywhere. If there exists a function $g \in L^1(X,\mathcal{B},\mu)$ so that $|f_k| \le g$ almost everywhere, then*

$$\lim_{k\to\infty} \mathbf{E}(f_k|\mathcal{A}) = \mathbf{E}(f|\mathcal{A})$$

almost everywhere. □

Proof. We'll use Fatou's Lemma. First apply it to the sequence $g - f_k$ to get

$$\mathbf{E}(g - \liminf_{k\to\infty} f_k|\mathcal{A}) \le \liminf_{k\to\infty} \mathbf{E}(g - f_k|\mathcal{A}),$$

or equivalently

$$\mathbf{E}(\limsup_{k\to\infty} f_k|\mathcal{A}) \geq \limsup_{k\to\infty} \mathbf{E}(f_k|\mathcal{A}).$$

Now just note that the following holds almost everywhere

$$\begin{aligned}\mathbf{E}(f|\mathcal{A}) = \mathbf{E}(\liminf_{k\to\infty} f_k|\mathcal{A}) &\leq \liminf_{k\to\infty} \mathbf{E}(f_k|\mathcal{A}) \leq \limsup_{k\to\infty} \mathbf{E}(f_k|\mathcal{A}) \\ &\leq \mathbf{E}(\limsup_{k\to\infty} f_k|\mathcal{A}) = \mathbf{E}(f|\mathcal{A}).\end{aligned}$$

This implies that

$$\lim_{k\to\infty} \mathbf{E}(f_k|\mathcal{A})$$

exists almost everywhere and equals $\mathbf{E}(f|\mathcal{A})$. ∎

The next definition gives us a connection between the theory discussed in this section and ergodic theory.

Definition 1.4.7. Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system. The T-invariant subsets $A \in \mathcal{B}$ form a σ-algebra and we'll denote this **invariant σ-algebra** by $\mathfrak{Inv}$. □

We make the following remark on Birkhoff's Ergodic Theorem (Theorem 1.3.1).

Remark 1.4.8. *In Theorem 1.3.1 we have $f^* = \mathbf{E}(f|\mathfrak{Inv})$ (where f^* was the limit function of the Birkhoff averages).* □

Proof. Let $A \in \mathfrak{Inv}$. Then since f^* is invariant it is clearly $\mathfrak{Inv}$-measurable and

$$\begin{aligned}\int_A f^* &= \lim_{n\to\infty} \int_A \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \int_A (f \circ T^k) = \int_A f.\end{aligned}$$

(Remember that we have proved in Theorem 1.3.1 that $\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \xrightarrow{L^1} f^*$.) ∎

Finally we will prove some theorems about conditional measures. The existence of conditional measures will require some restrictions to be imposed on the measure space, and the following definitions are therefore helpful.

Definition 1.4.9. A space $(X, \mathcal{B})$ is a **Borel space**, if X is a Borel subset of a compact metric space $\overline{X}$ and $\mathcal{B}$ is the restriction of the Borel σ-algebra of $\overline{X}$ to X. □

Note that $\mathcal{B}$ is actually the Borel σ-algebra $\mathcal{B}(X)$ of X. Indeed $\mathcal{B}$ trivially contains the subsets of X open in the relative topology and thus $\mathcal{B}(X) \subset \mathcal{B}$. On the other hand one can easily prove that

$$\Gamma = \{A \subset \overline{X} : A \cap X \in \mathcal{B}(X)\}$$

is a σ-algebra. Moreover it clearly contains the open subsets of $\overline{X}$, so $\mathcal{B}(\overline{X}) \subset \Gamma$. Hence if $A \cap X \in \mathcal{B} = \mathcal{B}(\overline{X})|X$, then $A \in \Gamma$ and thus $A \cap X \in B(X)$.

Definition 1.4.10. A space $(X, \mathcal{B}, \mu)$ is called a **Borel probability space**, if $(X, \mathcal{B})$ is a Borel space, X is dense in $\overline{X}$ and μ is a probability measure defined on $\mathcal{B}$. □

Definition 1.4.11. If $\mathcal{A}$ and $\mathcal{B}$ are σ-algebras, then $\mathcal{A} \overset{\mu}{\subset} \mathcal{B}$ means that for every $A \in \mathcal{A}$ there is $B \in \mathcal{B}$ such that $\mu(A \Delta B) = 0$. We define $\mathcal{A} \overset{\mu}{=} \mathcal{B}$ to mean that $\mathcal{A} \overset{\mu}{\subset} \mathcal{B}$ and $\mathcal{B} \overset{\mu}{\subset} \mathcal{A}$. □

Definition 1.4.12. A σ-algebra $\mathcal{B}$ on a set X is **countably-generated** if there exists a countable set $\{A_1, A_2, \dots\}$ of subsets of X such that

$$\mathcal{B} = \sigma(\{A_1, A_2, \dots\}).$$ □

Theorem 1.4.13. *Let $(X, \mathcal{B}, \mu)$ be a Borel probability space. Suppose that we are given a σ-algebra $\mathcal{A} \subset \mathcal{B}$. Then there exists an $\mathcal{A}$-measurable set $X' \subset X$ with $\mu(X') = 1$ and for each $x \in X'$ a measure $\mu_x^{\mathcal{A}}$, which we call* ***conditional measures****, so that the following properties are satisfied:*

(1) The measure $\mu_x^{\mathcal{A}} \colon \mathcal{B} \to \mathbf{R}$ is a probability measure on X for every $x \in X'$. In addition for all $f \in \mathcal{L}^1(X, \mu)$ the function f is $\mu_x^{\mathcal{A}}$-integrable for almost every $x \in X$, and for almost every $x \in X$ we have

$$\mathbf{E}(f|\mathcal{A})(x) = \int_X f \, d\mu_x^{\mathcal{A}}.$$

Here f can't be an equivalence class of functions, since $\mu_x^{\mathcal{A}}$ might be singular with respect to μ.

(2) $\mu_x^{\mathcal{A}}$ is uniquely determined by (1) for almost every $x \in X$. Moreover, it's enough to require that (1) holds for a countable dense subset of $C(\overline{X})|X$.

(3) If $\mathcal{A}$ is countably-generated and $x \in X'$, then we define the ***atom*** *$[x]_{\mathcal{A}} \subset X$ containing x by setting*

$$[x]_{\mathcal{A}} = \bigcap_{x \in A \in \mathcal{A}} A.$$

It is $\mu_x^{\mathcal{A}}$-measurable and $\mu_x^{\mathcal{A}}([x]_{\mathcal{A}}) = 1$. Moreover if $x, y \in X'$ and $[x]_{\mathcal{A}} = [y]_{\mathcal{A}}$, then $\mu_x^{\mathcal{A}} = \mu_y^{\mathcal{A}}$.

Technically one can define the atom also in the setting where $\mathcal{A}$ is not countably-generated, but then we can't prove the measurability, for example.

(4) If $\mathcal{A}'$ is any σ-algebra with $\mathcal{A}' \overset{\mu}{=} \mathcal{A}$ then $\mu_x^{\mathcal{A}} = \mu_x^{\mathcal{A}'}$ almost everywhere. □

Proof. First note that we may assume that X itself is a compact metric space: X is contained in a compact metric space $\overline{X}$ and if we can prove the theorem for $\overline{X}$, we get it for X as well. Namely, let $\mathcal{A}' = \sigma(\mathcal{A})$ be the extension of the σ-algebra $\mathcal{A}$ to the space $\overline{X}$. Define μ on $\overline{\mathcal{B}}$, the Borel σ-algebra of $\overline{X}$, by setting $\mu(\overline{X} \setminus X) = 0$. Now $N = \overline{X} \setminus X$ has measure zero, and hence $\mu_x^{\mathcal{A}'}(N) = 0$ for almost all x. (Consider taking $f = \chi_N$ in (1).) For these x set $\mu_x^{\mathcal{A}} = \mu_x^{\mathcal{A}'}|\mathcal{B}$. Then $\mu_x^{\mathcal{A}}$ is a probability measure on X and if we define $\mathbf{E}(f|\mathcal{A})$ to be zero on N, then it is $\mathcal{A}'$ measurable and

$$\int_B \mathbf{E}(f|\mathcal{A})d\mu = \int_B f d\mu = \int_B \mathbf{E}(f|\mathcal{A}')d\mu,$$

for all $B \in \mathcal{A}'$ and $f \in L^1(\overline{X}, \mu)$. Thus if we fix $f \in \mathcal{L}^1(\overline{X}, \mu)$, then $\mathbf{E}(f|\mathcal{A}) = \mathbf{E}(f|\mathcal{A}')$ almost everywhere and therefore

$$\mathbf{E}(f|\mathcal{A})(x) = \mathbf{E}(f|\mathcal{A}')(x) = \int f d\mu_x^{\mathcal{A}}$$

for almost every $x \in X$. This shows (1). The rest of the properties are straightforward to check. Now to the proof.

Part 1 (uniqueness): Suppose that μ_x and ν_x are families of measures that satisfy (1) for $f \in \{f_1, f_2, \dots\} \subset C(X)$, where $\{f_n\}_{n \in \mathbf{N}}$ is dense. That such a dense subset of functions exists is proved in A.2. Then for all $n \in \mathbf{N}$ and almost every x we have

$$\int f_n d\mu_x = \mathbf{E}(f_n|\mathcal{A})(x) = \int f_n d\nu_x.$$

Let's call the null set outside of which this holds N. Now if $f \in C(X)$, we may approximate it uniformly with a sequence f_{n_k}, and the Dominated Convergence Theorem gives us

$$\int f d\mu_x = \int f d\nu_x$$

for all $x \notin N$. Hence $\mu_x = \nu_x$. (Remember that the probability measures on X correspond to linear functionals on $C(X)$ by Riesz representation theorem [Rud87].) This proves (3).

Part 2 (claim (4)): Suppose that $\mathcal{A}' \overset{\mu}{=} \mathcal{A}$. Let $\overline{\mathcal{A}}$ be the smallest σ-algebra containing both $\mathcal{A}'$ and $\mathcal{A}$. Now if $f \in C(X)$ then for all $B \in \overline{\mathcal{A}}$ there exists $B' \in \mathcal{A}$ so that $\mu(B' \Delta B) = 0$ and thus

$$\int_B \mathbf{E}(f|\mathcal{A})\, d\mu = \int_{B'} \mathbf{E}(f|\mathcal{A})\, d\mu = \int_{B'} f\, d\mu = \int_B f\, d\mu.$$

By the uniqueness of the conditional expectation $\mathbf{E}(f|\mathcal{A}) = \mathbf{E}(f|\overline{\mathcal{A}})$ almost everywhere. A similar argument holds for $\mathbf{E}(f|\mathcal{A}')$ and thus

$$\mathbf{E}(f|\mathcal{A}) = \mathbf{E}(f|\mathcal{A}')$$

almost everywhere. The claim now follows from the uniqueness of the measures $\mu_x^{\mathcal{A}}$ that satisfy (1).

Part 3 (existence): We'll start by proving the existence of measures $\mu_x^{\mathcal{A}}$ by first showing (1) for continuous functions. We let $\mathcal{F} = \{f_0 = 1, f_1, f_2, \dots\} \subset C(X)$ be a countable dense vector space over $\mathbf{Q}$. Such a vector space exists because $C(X)$ is separable (See A.2.), so one can pick a countable dense subset of $C(X)$ and consider the vector space spanned by this subset. The vector space will then itself be countable, since the scalars are taken from $\mathbf{Q}$. Now for each f_i, $i \geq 1$ we can choose a representative $g_i \in \mathcal{L}^1(X)$ of $\mathbf{E}(f_i|\mathcal{A})$. Set $g_0 = 1$. By the properties of $\mathbf{E}(\cdot|\mathcal{A})$ we can find a set N of measure zero, outside of which the following properties hold:

- $g_i(x) \geq 0$ if $f_i \geq 0$;
- $|g_i(x)| \leq \|f_i\|_\infty$;
- if $f_i = af_j + bf_k$, $a, b \in \mathbf{Q}$, then $g_i(x) = ag_j(x) + bg_k(x)$.

For all $x \notin N$ define now a mapping $\Lambda_x\colon \mathcal{F} \to \mathbf{R}$ by setting $\Lambda_x(f_i) = g_i(x)$. This mapping is linear and bounded for all $x \notin N$ and hence continuous. Because $\mathcal{F}$ is dense in $C(X)$, it extends uniquely to a continuous linear mapping $C(X) \to \mathbf{R}$. Now for all $x \notin N$ the Riesz representation theorem gives us a measure $\mu_x^{\mathcal{A}}$ so that

$$\Lambda_x(f) = \int f \, d\mu_x^{\mathcal{A}}$$

for all $f \in C(X)$. In addition since Λ_x is a positive functional and $\Lambda_x(1) = 1$, $\mu_x^{\mathcal{A}}$ is a probability measure. We know by construction that the mapping $x \mapsto \Lambda_x(f_i) = g_i(x)$ is $\mathcal{A}$-measurable for $i \in \mathbf{N}$. Now if $f \in C(X)$, we can find a sequence $f_{n_k} \in \mathcal{F}$ that converges to f uniformly. Thus by the Dominated Convergence Theorem

$$\int f_{n_k} \, d\mu_x^{\mathcal{A}} \to \int f \, d\mu_x^{\mathcal{A}}$$

and hence $x \mapsto \int f \, d\mu_x^{\mathcal{A}}$ is $\mathcal{A}$-measurable. (The pointwise limit of measurable functions is measurable.) Moreover

$$\int_A \int f_{n_k} \, d\mu_x^{\mathcal{A}} \, d\mu(x) = \int_A g_{n_k} \, d\mu = \int_A f_{n_k} \, d\mu$$

and again by the Dominated Convergence Theorem

$$\int_A \int f \, d\mu_x^{\mathcal{A}} \, d\mu(x) = \int_A f \, d\mu$$

for all $A \in \mathcal{A}$. Thus we have shown (1) for continuous functions.

Next we'll extend the result to cover all $f \in \mathcal{L}^1(X)$. If $U \subset X$ is open, then there exists an increasing sequence of continuous functions $f_{n_k} \in \mathcal{F}$ converging to χ_U. Thus the Monotone Convergence Theorem implies (1) for χ_U. If $G = \bigcap_{i=1}^\infty U_i$ is a G_δ-set, then by considering the sequence $\chi_n = \prod_{i=1}^n \chi_{U_i}$, we get by the Dominated Convergence Theorem that (1) holds for χ_G. By looking at $1 - \chi_G = \chi_{X \setminus G}$ we see that (1) holds also for characteristic functions of F_σ-sets. Motivated by this we define the family

$$\Gamma = \{B \in \mathcal{B} : f = \chi_B \text{ satisfies (1) and } x \mapsto \int f d\mu_x^{\mathcal{A}} \text{ is } \mathcal{A}\text{-measurable}\}.$$

The Monotone Convergence Theorem shows that if $B_1 \subset B_2 \subset \dots$ are sets in Γ, then also $\bigcup B_i \in \Gamma$. Similarly if $B_1 \supset B_2 \supset \dots$ are sets in Γ, then also $\bigcap B_i \in \Gamma$. This means that Γ is a monotone class. Next define

$$\mathcal{R} = \left\{ \bigsqcup_{i=1}^{n} (U_i \cap K_i) : U_i \subset X \text{ open, } K_i \subset X \text{ closed.} \right\}$$

We'd like to show that $\mathcal{R}$ is an algebra. To see this, notice that an algebra generated by a finite number of sets is finite and each of the sets in the algebra is a finite union of *atoms* of the algebra that partition the space. Because the union or intersection of two sets of type $U \cap K$ is again of the same type, as is the complement, also the atoms are of this type. Thus if $R = \bigsqcup_{i=1}^{n} U_i \cap K_i \in \mathcal{R}$, then we can consider the algebra generated by the sets $U_i \cap K_i$ to deduce that also $X \setminus R$ can be represented in the desired form. We do the union and intersection of two sets similarly. Now a set of form $U \cap K$ is a G_δ-set and hence in Γ. By linearity we deduce that if $R \in \mathcal{R}$, then

$$\chi_R = \sum_{i=1}^{n} \chi_{U_i \cap K_i}$$

satisfies (1) and hence $R \in \Gamma$. Now $\sigma(\mathcal{R}) \subset \Gamma$ because Γ is a monotone class. But clearly also $\mathcal{B} = \sigma(\mathcal{R})$. This shows that χ_B satisfies (1) for all Borel sets $B \in \mathcal{B}$. We immediately get also that simple functions satisfy (1) and by the Monotone Convergence Theorem all non-negative measurable functions f satisfy (1) too. If $f \in \mathcal{L}^1$ is integrable, we can write it as $f = f^+ - f^-$, where f^+ and f^- are integrable. Because (1) holds for f^+, f^-, we get that

$$\int f^+ \, d\mu_x^{\mathcal{A}} < \infty \text{ and } \int f^- \, d\mu_x^{\mathcal{A}} < \infty$$

for almost every x. This means that f is $\mu_x^{\mathcal{A}}$ integrable and $\mathcal{A}$-measurable for almost every x. We get that (1) holds for f.

Part 4 (claim 3): Assume that $\mathcal{A}$ is countably-generated and write it as $\mathcal{A} = \sigma(\{A_1, A_2, \dots\})$. Then for all $i \in \mathbf{N}$

$$\mathbf{E}(\chi_{A_i} | \mathcal{A})(x) = \chi_{A_i}(x) = \mu_x^{\mathcal{A}}(A_i)$$

almost everywhere. In the previous part we showed that the mapping

$$x \mapsto \int \chi_{A_i} d\mu_x^{\mathcal{A}}$$

is $\mathcal{A}$-measurable and hence the sets where the equality doesn't hold are $\mathcal{A}$-measurable and have measure 0. Because there are countably many sets, we can find a $\mathcal{A}$-measurable set N of measure zero outside of which the equation holds for all x and all i and by including the set N in the previous part we may also assume that (1) holds for any $f \in C(X)$. Hence

$$\mu_x^{\mathcal{A}}(A_i) = \begin{cases} 1, & \text{if } x \in A_i \setminus N \\ 0, & \text{if } x \in X \setminus (A_i \cup N). \end{cases}$$

We claim that

$$[x]_{\mathcal{A}} = \bigcap_{x\in A\in\mathcal{A}} A = \bigcap_{x\in A_i} A_i \cap \bigcap_{x\notin A_i} (X \setminus A_i),$$

which implies that $[x]_{\mathcal{A}}$ is $\mathcal{A}$-measurable. Indeed, if $y \in \bigcap_{x\in A\in\mathcal{A}} A$, then y is in every set A_i where x is. Hence $y \in \bigcap_{x\in A_i} A_i$. Also if $x \notin A_i$, then $x \in X \setminus A_i$ and also $y \in X \setminus A_i$. Thus

$$\bigcap_{x\in A\in\mathcal{A}} A \subset \bigcap_{x\in A_i} A_i \cap \bigcap_{x\notin A_i} (X \setminus A_i).$$

To prove the other direction, we'll prove that the sets $A \in \mathcal{A}$ satisfying the condition

$$\begin{cases} \bigcap_{x\in A_i} A_i \cap \bigcap_{x\notin A_i} (X \setminus A_i) \subset A, & \text{if } x \in A \\ \bigcap_{x\in A_i} A_i \cap \bigcap_{x\notin A_i} (X \setminus A_i) \subset X \setminus A, & \text{if } x \notin A \end{cases} \tag{1.4.2}$$

form a σ-algebra

$$\mathcal{M} = \{A \in \mathcal{A} : A \text{ satisfies } (1.4.2)\}$$

and $A_i \in \mathcal{M}$. Then it will follow that $\mathcal{A} = \mathcal{M}$ and we know that

$$\bigcap_{x\in A_i} A_i \cap \bigcap_{x\notin A_i} (X \setminus A_i) \subset \bigcap_{x\in A\in\mathcal{A}} A$$

for all $A \in \mathcal{A}$. Thus $[x]_{\mathcal{A}}$ is $\mathcal{A}$-measurable and we get

$$\mu_x^{\mathcal{A}}([x]_{\mathcal{A}}) = 1$$

if $x \notin N$. It is obvious that $A_i \in \mathcal{M}$, so we just have to show that $\mathcal{M}$ is a σ-algebra. One immediately sees that $X \in \mathcal{M}$ and that $\mathcal{M}$ is closed under taking complements. Now suppose that $B_1, B_2, \dots \in \mathcal{M}$. Then if $x \in \bigcap_{i=1}^{\infty} B_i$, we know that $x \in B_i$ for all $i \in \mathbf{N}$. This means that every set B_i satisfies the first case in (1.4.2), and hence also $\bigcap_{i=1}^{\infty} B_i$ satisfies it. A similar deduction in the case that $x \notin \bigcap_{i=1}^{\infty} B_i$ shows us that $\mathcal{M}$ is closed under countable intersections.

Finally, in $X' = X \setminus N$ the map

$$x \mapsto \int f \, d\mu_x^{\mathcal{A}}, \quad x \in X'$$

is $\mathcal{A}$-measurable (as shown in the previous part) for any $f \in C(X)$ and hence it is constant on atoms. Thus

$$\int f \, d\mu_x^{\mathcal{A}} = \int f \, d\mu_y^{\mathcal{A}}$$

if $x, y \in X'$ and $[x]_{\mathcal{A}} = [y]_{\mathcal{A}}$, which shows that $[x]_{\mathcal{A}} = [y]_{\mathcal{A}}$ implies $\mu_x^{\mathcal{A}} = \mu_y^{\mathcal{A}}$. ∎

Lemma 1.4.14. *If $(X, \mathcal{B}, \mu)$ is a Borel probability space and $\mathcal{A} \subset \mathcal{B}$ is a σ-algebra then there is a countably-generated σ-algebra $\mathcal{A}' \subset \mathcal{A}$ for which $\mathcal{A}' \overset{\mu}{=} A$.* □

Proof. Consider the space $C(\overline{X})$ of continuous functions on $\overline{X}$ endowed with the sup-norm. This space is separable (See A.2.) and hence also the space $C(\overline{X})|X$ of restrictions of continuous functions $C(\overline{X})$ to X is separable. Because $C(X)$ is dense in $L^1(X, \mathcal{B}, \mu)$, also $L^1(X, \mathcal{B}, \mu)$ is separable. Finally as a subset of $L^1(X, \mathcal{B}, \mu)$, the set of functions

$$\{\chi_A : A \in \mathcal{A}\} \subset L^1(X, \mathcal{B}, \mu)$$

is separable. Hence we can pick a sequence $A_1, A_2, \dots$ of $\mathcal{A}$-measurable subsets so that χ_{A_k} form a dense set in $L^1(X, \mathcal{B}, \mu)$ for $\{\chi_A : A \in \mathcal{A}\}$.

We claim that $\mathcal{A}' = \sigma(A_1, A_2, \dots)$ is the countably-generated σ-algebra we are searching for. Clearly $\mathcal{A}' \subset \mathcal{A}$. Let $A \in \mathcal{A}$. For every $k \in \mathbf{N}$ we have some A_{n_k}, such that

$$\mu(A \Delta A_{n_k}) = \|\chi_A - \chi_{A_{n_k}}\|_1 < \frac{1}{k}.$$

Thus the sequence $\chi_{A_{n_k}}$ is Cauchy in $L^1(X, \mathcal{A}', \mu)$. By completeness there is some $f \in L^1(X, \mathcal{A}', \mu) \subset L^1(X, \mathcal{A}, \mu)$ such that $\chi_{A_{n_k}} \to f$ as $k \to \infty$. By the uniqueness of the limit in $L^1(X, \mathcal{A}, \mu)$, $f = \chi_A$ almost everywhere and hence $A' = f^{-1}\{1\} \in \mathcal{A}'$ satisfies $\mu(A' \Delta A) = 0$. ∎

Definition 1.4.15. Let $\overline{X}$ be a compact metric space and $\mathcal{B}$ its Borel σ-algebra. We define $\mathcal{M}(\overline{X})$ to be the space of Borel probability measures on $(\overline{X}, \mathcal{B})$. □

We note that the space of Borel measures on $(\overline{X}, \mathcal{B})$ can be identified with the dual of the space of continuous functions $\overline{X} \to \mathbf{R}$. (See for example [Rud87].) On this dual we can define the weak*-topology, which then carries over to the subset $\mathcal{M}(\overline{X})$. The compactness part in the following theorem is a consequence of the Banach-Alaoglu Theorem.

Theorem 1.4.16. *$\mathcal{M}(\overline{X})$ is a compact metric space (as a subset of the space of Borel measures with weak*-topology).* □

Proof. See for example [Par67]. ∎

Note that the previous theorem allows us to consider $\mathcal{M}(\overline{X})$ as a Borel space.

Theorem 1.4.17. *Let $(X, \mathcal{B}, \mu)$ be a Borel probability space and $\mathcal{A} \subset \mathcal{B}$ a countably-generated σ-algebra. Then there is a subset $X' \subset X$ such that $\mu(X') = 1$ for which the mapping $\phi\colon X' \to \mathcal{M}(\overline{X})$ given by $\phi(x) = \mu_x^{\mathcal{A}}$ is $\mathcal{A}$-measurable and satisfies*

$$\mathcal{A}|_{X'} = \phi^{-1}\mathcal{B}_{\mathcal{M}},$$

where

$$\mathcal{A}|_{X'} = \{A \cap X' : A \in \mathcal{A}\}$$

is the σ-algebra $\mathcal{A}$ restricted to the set X' and $\mathcal{B}_{\mathcal{M}}$ is the Borel σ-algebra of $\mathcal{M}(\overline{X})$. Moreover

$$[x]_{\mathcal{A}} = \phi^{-1}(\phi(x))$$

for $x \in X'$. □

Proof. Write $\mathcal{A} = \sigma(A_1, A_2, \dots)$. Let X' be a set of measure 1 where Theorem 1.4.13 holds. The theorem implies that

$$\chi_{A_i}(x) = \mu_x^{\mathcal{A}}(A_i)$$

almost everywhere and by perhaps removing a null set we may assume that this holds for all $x \in X'$. The theorem also implies that ϕ is measurable, since $x \mapsto \mu_x^{\mathcal{A}}$ is measurable if and only if $x \mapsto \int f \, d\mu_x^{\mathcal{A}}$ is measurable for all $f \in C(\overline{X})$. Now

$$A_i \cap X' = \phi^{-1}\{\nu \in \mathcal{M} : \nu(A_i) = 1\}$$

since $\phi(x) \in \{\nu \in \mathcal{M} : \nu(A_i) = 1\}$ if and only if $\mu_x^{\mathcal{A}}(A_i) = \chi_{A_i}(x) = 1$. Moreover

$$\{\nu \in \mathcal{M} : \nu(A_i) = 1\} \in \mathcal{B}_{\mathcal{M}}$$

because one can make a similar argument as in the proof of Theorem 1.4.13: One first checks that the mapping

$$\nu \mapsto \int \chi_U d\nu$$

is measurable for open sets U by approximating χ_U with continuous functions. This implies that

$$\{\nu \in \mathcal{M} : \nu(U) = 1\} \in \mathcal{B}_{\mathcal{M}}$$

and one gets the same result for closed sets by considering $1 - \chi_U$. One then shows that

$$\{A \subset X : \{\nu \in \mathcal{M} : \nu(A) = 1\} \in \mathcal{B}_{\mathcal{M}}\}$$

is a monotone class, after which one again considers the algebra

$$\mathcal{R} = \{\bigsqcup_{k=1}^{n} U_k \cap F_k : U_k \in \mathcal{B} \text{ is open }, F_k \in \mathcal{B} \text{ is closed}\},$$

shows that it belongs to the monotone class and hence is finally allowed to deduce that

$$\mathcal{B} \subset \{A \subset X : \{\nu \in \mathcal{M} : \nu(A) = 1\} \in \mathcal{B}_{\mathcal{M}}\}.$$

Hence $\mathcal{A}|X' \subset \phi^{-1}\mathcal{B}_{\mathcal{M}}$. To prove the other direction, we note that the sets of form

$$\{\nu : \left|\int f d\nu - r\right| < \varepsilon\}, \quad r, \varepsilon \in \mathbf{Q}, f \in \{f_1, f_2, \dots\} \subset C(\overline{X}),$$

where $\{f_1, f_2, \dots\} \subset C(\overline{X})$ is a countable dense subset, are open in $\mathcal{M}$ and form a subbase. It is enough to prove that the preimages of sets of this form are in $\mathcal{A}|X'$. But by Theorem 1.4.13 the sets

$$\phi^{-1}\{\nu : \left|\int f d\nu - r\right| < \varepsilon\} = \{x : \left|\int f d\mu_x^{\mathcal{A}} - r\right| < \varepsilon\}$$

are measurable, so we are done. Finally because $\mathcal{B}_{\mathcal{M}}$ separates points, $\phi^{-1}(\phi(x))$ is $\mathcal{A}$-measurable and is the smallest set in $\phi^{-1}\mathcal{B}_{\mathcal{M}}$ that contains x, that is

$$\phi^{-1}(\phi(x)) = [x]_{\mathcal{A}}.$$ ∎

Theorem 1.4.18. *Let $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$ be Borel probability spaces, and $\phi: X \to Y$ a measure-preserving map. If $\mathcal{A} \subset \mathcal{C}$ is a σ-algebra, then*

$$\phi_* \mu_x^{\phi^{-1}\mathcal{A}} = \nu_{\phi(x)}^{\mathcal{A}}$$

μ-almost everywhere. Here $\phi_* \mu_x^{\phi^{-1}\mathcal{A}}$ is the push-forward measure defined by

$$\phi_* \mu_x^{\phi^{-1}\mathcal{A}}(A) = \mu_x^{\phi^{-1}\mathcal{A}}(\phi^{-1}(A))$$

for all $A \in \mathcal{C}$. □

Proof. Let $f \in L^1(Y, \mathcal{C}, \nu)$. Then the mapping $\mathbf{E}_\nu(f|\mathcal{A}) \circ \phi$ is $\phi^{-1}\mathcal{A}$-measurable and

$$\int_{\phi^{-1}A} \mathbf{E}_\nu(f|\mathcal{A}) \circ \phi \, d\mu = \int_A \mathbf{E}_\nu(f|\mathcal{A}) \, d\nu = \int_A f \, d\nu = \int_{\phi^{-1}A} f \circ \phi \, d\mu$$

holds for all $A \in \mathcal{A}$. Hence

$$\mathbf{E}_\nu(f|\mathcal{A}) \circ \phi = \mathbf{E}_\mu(f \circ \phi | \phi^{-1}\mathcal{A}).$$

Thus by Theorem 1.4.13 for any $f \in C(\overline{Y})$ we have

$$\begin{aligned} \int f \, d\nu_{\phi(x)}^{\mathcal{A}} &= \mathbf{E}_\nu(f|\mathcal{A})(\phi(x)) = \mathbf{E}_\mu(f \circ \phi | \phi^{-1}\mathcal{A})(x) \\ &= \int f \circ \phi d\mu_x^{\phi^{-1}\mathcal{A}} = \int f d\left(\phi_* \mu_x^{\phi^{-1}\mathcal{A}}\right) \end{aligned}$$

almost everywhere, which by the uniqueness in Theorem 1.4.13 gives us the claim. ∎

1.5. Ergodic decomposition

Let $(X, \mathcal{B})$ be a Borel space and $T: X \to X$ a measurable transformation. In this section we will show that every T-invariant measure μ can be expressed in terms of ergodic measures, the ergodic components of μ. We will denote the set of T-invariant measures by $\mathcal{M}^T$ and the set of ergodic measures by $\mathcal{E}^T$. As subspaces of $\mathcal{M}(\overline{X})$, they are also metric spaces (under the weak*-topology).

First we'll state the following characterization of ergodicity.

Lemma 1.5.1. *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system on a Borel probability space and $f_n \in C(\overline{X})$, $n = 1, 2, \ldots$ a dense subset of $C(\overline{X})$. Then μ is ergodic if and only if*

$$\frac{1}{n} \sum_{k=0}^{n-1} f_i(T^k x) \to \int f_i \, d\mu \tag{1.5.1}$$

for almost every $x \in X$ and every $i \in \mathbf{N}$. □

Proof. If μ is ergodic, then the claim is proven by Birhoff's Ergodic Theorem (Theorem 1.3.1).

Suppose that (1.5.1) holds. Then by Birkhoff's Ergodic Theorem and Remark 1.4.8 which connects it to the conditional expectation we have

$$\mathbf{E}(f_i|\mathfrak{Inv})(x) = \int f_i \, d\mu$$

for almost every x. Hence for every $f_i \in C(\overline{X})$ we have

$$\mu(A) \int f_i \, d\mu = \int_A \mathbf{E}(f_i|\mathfrak{Inv}) \, d\mu = \int_A f_i \, d\mu,$$

whenever A is a T-invariant set. By approximating χ_A with these f_i, we get

$$\mu(A)^2 = \mu(A) \Leftrightarrow \mu(A)(\mu(A) - 1) = 0.$$

■

The measurability of $\mathcal{E}^T$ in $\mathcal{M}(\overline{X})$ follows as a corollary.

Corollary 1.5.2. *The set of ergodic measures $\mathcal{E}^T \subset \mathcal{M}(\overline{X})$ is Borel measurable under the weak*-topology.* □

Proof. Let $f_n \in C(\overline{X})$, $n = 1, 2, \ldots$ be a dense subset of $C(\overline{X})$. Then we can express $\mathcal{E}^T$ as

$$\mathcal{E}^T = \bigcap_{i=1}^{\infty} \{\mu \in \mathcal{M}(\overline{X}) : \frac{1}{n} \sum_{k=0}^{n-1} f_i(T^k x) \to \int f_i \, d\mu\}.$$

Each of the sets

$$\{\mu \in \mathcal{M}(\overline{X}) : \frac{1}{n} \sum_{k=0}^{n-1} f_i(T^k x) \to \int f_i \, d\mu\}$$

where $i \in \mathbf{N}$ can furthermore be written as

$$\bigcap_{N=1}^{\infty} \bigcup_{j=1}^{\infty} \bigcap_{n=j}^{\infty} \{\mu \in \mathcal{M}(\overline{X}) : \left| \frac{1}{n} \sum_{k=0}^{n-1} f_i(T^k x) - \int f_i \, d\mu \right| < 1/N\}.$$

Each of the sets

$$\{\mu \in \mathcal{M}(\overline{X}) : \left| \frac{1}{n} \sum_{k=0}^{n-1} f_i(T^k x) - \int f_i \, d\mu \right| < 1/N\}$$

is just the inverse image of the measurable set

$$(-1/N + \sum_{k=0}^{n-1} f_i(T^k x), 1/N + \sum_{k=0}^{n-1} f_i(T^k x))$$

under the mapping $\mu \mapsto \int f_i \, d\mu$ and hence measurable. ■

Theorem 1.5.3. *Let $\mu \in \mathcal{M}^T$. Then there exists a Borel probability measure λ on $\mathcal{E}^T$ such that $\mu = \int \nu d\lambda(\nu)$, i.e.*

$$\int_X f\,d\mu = \int_{\mathcal{E}^T} \int_X f\,d\nu\,d\lambda(\nu), \quad \text{for all } f \in \mathcal{L}^1(X,\mu).$$

In the integral above $\int_X f\,d\nu$ is defined for almost every $\nu \in \mathcal{E}^T$. □

Proof. By Lemma 1.4.14 we can pick a countably generated sub-σ-algebra $\mathfrak{Inv}' \subset \mathfrak{Inv}$ so that for every $A \in \mathfrak{Inv}$ there is a set $A' \in \mathfrak{Inv}'$ with the property that $\mu(A\Delta A') = 0$. Write $\mathfrak{Inv}'$ as

$$\mathfrak{Inv}' = \sigma(E_1, E_2, \dots), \quad E_k \in \mathfrak{Inv}\ (k \in \mathbf{N}).$$

Let

$$N_1 = \bigcup_{i=1}^{\infty} T^{-1}E_i \Delta E_i,$$

and let N_2 be a set of measure 0, in the complement of which Theorem 1.4.13 (with both $\mathcal{A} = \mathfrak{Inv}$ and $\mathcal{A} = \mathfrak{Inv}'$), Theorem 1.4.17 (with $\mathcal{A} = \mathfrak{Inv}$), and Theorem 1.4.18 (with $\mathcal{A} = \mathfrak{Inv}$, $\phi = T$) hold. Now let

$$N = \bigcup_{n=0}^{\infty} T^{-n}(N_1 \cup N_2),$$

to get a set of measure 0, out of which the above mentioned theorems hold, and which satisfies

$$T^{-n}N \subset N$$

for all $n \in \mathbf{N}$. If $x \in X \setminus N$, then Theorem 1.4.18 shows that

$$T_* \mu_x^{\mathfrak{Inv}} = \mu_{Tx}^{\mathfrak{Inv}},$$

since $T^{-1}\mathfrak{Inv} \stackrel{\mu}{=} \mathfrak{Inv}$. Since $Tx \in X \setminus N$, we have $[x]_{\mathfrak{Inv}'} = [Tx]_{\mathfrak{Inv}'}$ so that by Theorem 1.4.13 we have $\mu_{Tx}^{\mathfrak{Inv}} = \mu_x^{\mathfrak{Inv}}$. This means that $\mu_x^{\mathfrak{Inv}} \circ T^{-1} = \mu_x^{\mathfrak{Inv}}$, so that $\mu_x^{\mathfrak{Inv}}$ is T-invariant for almost all $x \in X$.

Let $f_k \in C(\overline{X})$, $k \in \mathbf{N}$, form a dense subset of $C(\overline{X})$, and let N' be the union of N and those points $x \in X$ for which

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f_i(T^k x) = \mathbf{E}(f_i|\mathfrak{Inv})(x)$$

doesn't hold for some $i \in \mathbf{N}$. Then N' is still a set of measure 0 by Birkhoff's Ergodic Theorem. Now for all $x \in X \setminus N'$ we have $\mathbf{E}(f_i|\mathfrak{Inv})(x) = \int f_i\,d\mu_x^{\mathfrak{Inv}}$. Let

$$N'' = N' \cup \{x \in X : \mu_x^{\mathfrak{Inv}}(N') > 0\}.$$

By applying Theorem 1.4.13 to $\chi_{N'}$ we see that

$$\begin{aligned} 0 &= \int \chi_{N'}\,d\mu = \int \mathbf{E}(\chi_{N'}|\mathfrak{Inv})\,d\mu = \int\int \chi_{N'}\,d\mu_x^{\mathfrak{Inv}}\,d\mu(x) \\ &= \int \mu_x^{\mathfrak{Inv}}(N')\,d\mu(x), \end{aligned}$$

which implies that $\mu_x^{\mathfrak{Inv}}(N') = 0$ for almost every $x \in X$. Hence $\mu(N'') = 0$. If $x \in X \setminus N''$, $y \in X \setminus N'$ and $[x]_{\mathfrak{Inv}'} = [y]_{\mathfrak{Inv}'}$ then again by Theorem 1.4.13

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f_i(T^k y) = \int f_i \, d\mu_y^{\mathfrak{Inv}} = \int f_i \, d\mu_x^{\mathfrak{Inv}}.$$

Since $\mu_x^{\mathfrak{Inv}}([x]_{\mathfrak{Inv}'} \setminus N') = 1$, the above holds for $\mu_x^{\mathfrak{Inv}}$-almost every $y \in X$ and hence Lemma 1.5.1 shows that $\mu_x^{\mathfrak{Inv}}$ is ergodic.

Apply now Theorem 1.4.13 to deduce that for all $f \in \mathcal{L}^1(X, \mu)$

$$\int f \, d\mu = \int_X \int_X f \, d\mu_x^{\mathfrak{Inv}} \, d\mu(x) = \int_X \int_X f \, d(\phi(x)) \, d\mu(x),$$

where ϕ is given by $\phi\colon x \mapsto \mu_x^{\mathfrak{Inv}}$. Finally by Theorem 1.4.17 the mapping $\phi\colon X \to \mathcal{M}^T$ (defined μ-almost everywhere) is measurable and therefore letting $\lambda = \phi_* \mu$ we have

$$\int f \, d\mu = \int_{X\setminus N''} \int_X f \, d(\phi(x)) \, d\mu(x) = \int_{\mathcal{E}^T} \int_X f \, d\nu \, d\lambda(\nu).$$ ∎

Remark 1.5.4. *If we assume that the transformation T is continuous and X is compact metric space, then there is an easier proof of the above result for continuous functions $f \in C(X)$. The proof utilizes Choquet's Theorem, which lets one to express a point in a compact convex space as an integral over the extreme points of the space with respect to a measure depending on the point. It can be shown that $\mathcal{E}^T$ are the extreme points of $\mathcal{M}^T$. See [Phe01] for more information.* □

2. Smooth ergodic theory and Pesin theory

2.1. Lyapunov exponents

In this first section of the second part of the thesis we look at some basic properties of Lyapunov exponents (defined below), following mainly [BP99]. Let M be a compact smooth Riemannian manifold and $f\colon M \to M$ a diffeomorphism.

Definition 2.1.1. Let $x \in M$. The function $\chi(x,\cdot)\colon T_xM \to \mathbf{R} \cup \{-\infty, \infty\}$ given by

$$\chi(x,v) = \limsup_{n\to\infty} \frac{1}{n} \log \|T_x f^n v\|$$

is called a **Lyapunov exponent**. □

Lemma 2.1.2. *For all $x \in M$ and $v, w \in T_xM$ the following properties hold.*

1. *For all $c \in \mathbf{R} \setminus \{0\}$ we have $\chi(x, cv) = \chi(x,v)$;*
2. *$\chi(x, v+w) \le \max\{\chi(x,v), \chi(x,w)\}$;*
3. *$\chi(x,0) = -\infty$.* □

Proof. The first property is true since

$$\begin{aligned}\chi(x,cv) &= \limsup_{n\to\infty} \frac{1}{n} \log \|T_x f^n(cv)\| \\ &= \limsup_{n\to\infty} \left(\frac{\log|c|}{n} + \frac{1}{n} \log \|T_x f^n v\| \right) \\ &= \limsup_{n\to\infty} \frac{1}{n} \log \|T_x f^n v\| = \chi(x,v).\end{aligned}$$

For the second property assume that $\chi(x,v) \le \chi(x,w)$, so that

$$\limsup_{n\to\infty} \frac{1}{n} \log \|T_x f^n v\| \le \limsup_{n\to\infty} \frac{1}{n} \log \|T_x f^n w\|.$$

Now in the sequence on the left hand side replace each term satisfying $\|T_x f^n v\| < \|T_x f^n w\|$ by $\|T_x f^n w\|$. By doing this one is increasing the left hand side, but it can't exceed the right hand side. Hence we get that

$$\limsup_{n\to\infty} \frac{1}{n} \log \|T_x f^n v\| \le \limsup_{n\to\infty} \frac{1}{n} \log \max(\|T_x f^n v\|, \|T_x f^n w\|) \le$$
$$\limsup_{n\to\infty} \frac{1}{n} \log \|T_x f^n w\| \le \limsup_{n\to\infty} \frac{1}{n} \log \max(\|T_x f^n v\|, \|T_x f^n w\|)$$

and thus

$$\limsup_{n\to\infty} \frac{1}{n} \log \|T_x f^n w\| = \limsup_{n\to\infty} \frac{1}{n} \log \max(\|T_x f^n v\|, \|T_x f^n w\|).$$

Therefore

$$\limsup_{n\to\infty} \frac{1}{n} \log \|T_x f^n (v+w)\| \le \limsup_{n\to\infty} \frac{1}{n} \log(\|T_x f^n v\| + \|T_x f^n w\|) \le$$
$$\limsup_{n\to\infty} \frac{1}{n} \log(2\max(\|T_x f^n v\|, \|T_x f^n w\|)) = \limsup_{n\to\infty} \frac{1}{n} \log(\|T_x f^n w\|).$$

The third property is trivial. ∎

Theorem 2.1.3. *For a fixed $x \in M$ the following statements hold.*

1. *If $\chi(x,v) \neq \chi(x,w)$ then $\chi(x,v+w) = \max(\chi(x,v), \chi(x,w))$.*
2. *We have the inequality*

   $$\chi(x, v_1 + \cdots + v_n) \le \max(\chi(x,v_1), \ldots, \chi(x,v_n))$$

   and if there is an index i such that $\chi(x,v_i) > \chi(x,v_j)$ for all $j \neq i$ then

   $$\chi(x, v_1 + \cdots + v_n) = \chi(x, v_i).$$

3. *Let $v_1, \ldots, v_n \in T_x M \setminus \{0\}$. If $\chi(x,v_1), \ldots, \chi(x,v_n)$ are distinct, then $v_1, \ldots, v_n$ are linearly independent.*
4. *The function $v \mapsto \chi(x,v)$ $(v \in T_x M)$ takes only finitely many values.* □

Proof. To prove the first statement, suppose without loss of generality that $\chi(x,v) < \chi(x,w)$. Then by Lemma 2.1.2 we have

$$\chi(x, v+w) \le \chi(x,w) = \chi(x, v+w-v) \le \max(\chi(x,v+w), \chi(x,v)).$$

Now if $\chi(x,v+w) < \chi(x,v)$ then $\chi(x,w) \le \chi(x,v)$, which is a contradiction, so that $\chi(x,v+w) \ge \chi(x,v)$ and therefore $\chi(x,v+w) = \chi(x,w)$.

For the second statement notice that by induction

$$\begin{aligned}\chi(x, v_1 + \cdots + v_n) &\le \max(\chi(x, v_1 + \cdots + v_{n-1}), \chi(x,v_n)) \\ &\le \max(\chi(x,v_1), \ldots, \chi(x,v_n)).\end{aligned}$$

Also if $1 \le i \le n$ is such that $\chi(x,v_i)$ is larger than the other $\chi(x,v_k)$, then by the second statement

$$\chi(x, v_1 + \cdots + v_{i-1} + v_{i+1} + \cdots + v_n) \le$$
$$\max(\chi(x,v_1), \ldots, \chi(x,v_{i-1}), \chi(x,v_{i+1}), \ldots, \chi(v_n)) < \chi(x,v_i),$$

so that

$$\chi(x, v_1 + \cdots + v_n) = \chi(x, v_i)$$

by the first statement.

To prove the third statement suppose the contrary: There exist real numbers $\alpha_1, \ldots, \alpha_n$ ($n \geq 2$), not all zero, so that

$$\alpha_1 v_1 + \cdots + \alpha_k v_n = 0.$$

We may suppose without loss of generality that $\alpha_1 \neq 0$. Then

$$-\infty = \chi(x, \alpha_1 v_1 + \cdots + \alpha_n v_k) = \max(\chi(x, v_1), \ldots, \chi(x, v_n)) > -\infty,$$

which is a contradiction.

The fourth statement is clear by the third statement because there are only finitely many (the number is bounded by the dimension of M) linearly independent vectors in $T_x M$. ∎

Definition 2.1.4. The **Lyapunov exponents** of a point $x \in M$ are the $s(x)$ distinct values

$$\chi_1(x) < \cdots < \chi_{s(x)}(x)$$

of $\chi(x, v)$, $v \in T_x M$. We also define the subspaces

$$V_i(x) = \{v \in T_x M : \chi(x, v) \leq \chi_i(x)\}, \quad V_0(x) = \{0\}$$

of $T_x M$. They satisfy

$$\{0\} = V_0(x) \subsetneq V_1(x) \subsetneq \cdots \subsetneq V_{s(x)}(x) = T_x M.$$

We call the collection $\mathcal{V}_x = \{V_i : i = 1, \ldots, s(x)\}$ a **filtration** of $T_x M$. □

2.2. Oseledets' Multiplicative Ergodic Theorem

In this section we will prove Oseledets' theorem (Theorem 2.2.11) for a measure-preserving system $(M, \mathcal{B}, \mu, f)$, where M is a compact Riemannian manifold, $\mathcal{B}$ is the Borel σ-algebra on M, and $f \colon M \to M$ is a diffeomorphism. The theorem states (among other things) that for almost all $x \in M$ the tangent space $T_x M$ splits into (is a direct sum of) subspaces corresponding to the vectors with a common Lyapunov exponent.

2.2.1. *Kingman's Subadditive Ergodic Theorem*

We will start with the following subadditive ergodic theorem due to Kingman [Kin68]. Our proof follows the proof given by Steele in [Ste89].

Theorem 2.2.1 (Subadditive Ergodic Theorem). *Let $f_n \in L^1(X)$, $n \in$ **N**, be functions on a measure-preserving system $(X, \mathcal{B}, \mu, T)$ that satisfy the condition*

$$f_{n+m}(x) \leq f_n(x) + f_m(T^n x) \tag{2.2.1}$$

almost everywhere. Then for almost every $x \in X$ we have

$$\lim_{n\to\infty} \frac{f_n(x)}{n} = f_*(x) \geq -\infty,$$

where $f_*(x)$ is a T-invariant function. Moreover $f_*^+ \in L^1$,

$$\int f_* = \lim_{n\to\infty} \frac{1}{n}\int f_n = \inf_{n\in\mathbf{N}} \frac{1}{n}\int f_n,$$

and if T is ergodic then

$$f_* = \inf_{n\in\mathbf{N}} \frac{1}{n}\int f_n$$

almost everywhere. □

Proof. We start with the following lemma that gives us two easy corollaries of the condition (2.2.1).

Lemma 2.2.2. *Suppose that f_n is a sequence of functions satisfying (2.2.1). Then we have the inequality*

$$f_n(x) \le \sum_{k=0}^{n-1} f_1(T^k x).$$

Moreover if $a_1, \dots, a_m \in \mathbf{N}$, then

$$f_{a_1+\dots+a_m}(x) \le f_{a_1}(x) + f_{a_2}(T^{a_1}x) + \dots + f_{a_m}(T^{a_1+\dots+a_{m-1}}x). \quad \square$$

Proof. The first inequality follows by induction. The claim is clear for $n = 1$ and if it holds for n, then

$$f_{n+1}(x) \le f_1(x) + f_n(Tx) \le f_1(x) + \sum_{k=0}^{n-1} f_1(T^{k+1}x) = \sum_{k=0}^{n} f_1(T^k x).$$

For the second inequality we can also use induction on m. For $m = 1$ the inequality is clear. Supposing that it holds for some $m \in \mathbf{N}$ we have

$$\begin{aligned} f_{a_1+\dots+a_m+a_{m+1}}(x) &\le f_{a_1}(x) + f_{a_2+\dots+a_{m+1}}(T^{a_1}x) \\ &\le f_{a_1}(x) + f_{a_2}(T^{a_1}x) + \dots + f_{a_{m+1}}(T^{a_1+\dots+a_m}x). \quad \blacksquare \end{aligned}$$

Now let's go to the proof of the theorem itself.

It is enough to prove the theorem when $f_n \le 0$ for all n. This is seen by defining

$$g_n(x) = f_n(x) - \sum_{k=0}^{n-1} f_1(T^k x),$$

and noticing that by Lemma 2.2.2 we have $g_n(x) \le 0$. In addition by using (2.2.1) we see that the functions g_n also satisfy (2.2.1):

$$\begin{aligned} g_{n+m}(x) &= f_{n+m}(x) - \sum_{k=0}^{n+m-1} f_1(T^k x) \\ &\le f_n(x) + f_m(T^n x) - \sum_{k=0}^{n-1} f_1(T^k x) - \sum_{k=0}^{m-1} f_1(T^{n+k}x) \\ &= g_n(x) + g_m(T^n x). \end{aligned}$$

Once we have proven the theorem in the case $f_n \leq 0$ it follows from Birkhoff's Ergodic Theorem (Theorem 1.3.1) that

$$\frac{f_n(x)}{n} = \frac{g_n(x)}{n} + \frac{1}{n}\sum_{k=0}^{n-1} f_1(T^k x) \xrightarrow{a.e.} g_*(x) + f_1^*(x),$$

which is a T-invariant function as required. Since $g_*^+ = 0$, $(g_* + f_1^*)^+ \leq (f_1^*)^+ \in L^1$ and hence $(g_* + f_1^*)^+ \in L^1$. In addition (because μ is T-invariant) we have

$$\begin{aligned}\int (g_* + f_1^*) &= \int f_1 + \lim_{n\to\infty} \frac{1}{n} \int g_n \\ &= \lim_{n\to\infty} \frac{1}{n} \int \left(g_n + \sum_{k=0}^{n-1} (f_1 \circ T^k) \right) \\ &= \lim_{n\to\infty} \frac{1}{n} \int f_n\end{aligned}$$

and a similar calculation shows that

$$\int (g_* + f_1^*) = \inf_{n\in\mathbf{N}} \frac{1}{n} \int f_n.$$

Assume then that $f_n \leq 0$ for $n \in \mathbf{N}$ and let

$$f_* = \liminf_{n\to\infty} \frac{f_n}{n}.$$

Then clearly $f_*^+ = 0$. We'll start by proving that f_* is T-invariant. Notice that from the condition (2.2.1) it follows that

$$\frac{n+1}{n}\frac{f_{n+1}(x)}{n+1} \leq \frac{f_1(x)}{n} + \frac{f_n(Tx)}{n}$$

and hence by taking lim inf we see that $f_*(x) \leq f_*(Tx)$. This means that for all $\alpha \in \mathbf{R}$,

$$\{x \in X : f_*(x) > \alpha\} \subset \{x \in X : f_*(Tx) > \alpha\}.$$

Because μ is T-invariant, the difference of these sets

$$\Delta_\alpha = T^{-1}\{x \in X : f_*(x) > \alpha\} \setminus \{x \in X : f_*(x) > \alpha\}$$

has measure 0. Now

$$\mu(\{x \in X : f_*(Tx) > f_*(x)\}) = \mu\Big(\bigcup_{\alpha\in\mathbf{Q}} \Delta_\alpha \Big) = 0,$$

because $\mathbf{Q}$ is dense in $\mathbf{R}$. Hence

$$f_*(Tx) = f_*(x)$$

almost everywhere. In the rest of the proof we may without loss of generality assume that $f_*(T^k x) = f_*(x)$ for every $x \in X$ and for every $k \in \mathbf{N}$.

Suppose now that we are given $\varepsilon > 0$, $N \geq 2$ and $M \geq 1$. We set $F_M(x) = \max\{-M, f_*(x)\}$ and define

$$B(N, M) = \{x \in X : f_n(x) > n(F_M(x) + \varepsilon) \text{ for all } 1 \leq n \leq N\},$$
$$A(N, M) = X \setminus B(N, M).$$

Let $x \in X$ and $m \geq N$ be arbitrary, and partition the set $\{0, 1, \ldots, m-1\}$ into intervals by the following algorithm:

We let k be the first integer in $\{0, 1, 2, \ldots, m-1\}$ that doesn't yet belong to a constructed interval. (If such k doesn't exist, we stop.) There are two cases: First if $T^k x \in A(N, M)$ then there is a smallest number n, $1 \leq n \leq N$, such that $f_n(T^k x) \leq n(F_M(T^k x) + \varepsilon) = n(F_M(x) + \varepsilon)$. Add a blue interval $[k, k+n-1]$ if $k + n \leq m$ and a white singleton $\{k\}$ if $k + n > m$. Repeat. Second if $T^k x \in B(N, M)$ then we add a red $\{k\}$ and start over.

After partitioning we are left with u blue intervals $[a_k, a_k + n_k - 1]$, v red singletons $\{b_k\}$ with $T^{b_k} x \in B(N, M)$ and w white singletons $\{c_k\}$ with $c_k + N > m$. Now by Lemma 2.2.2 we have

$$f_m(x) \leq \sum_{k=1}^{u} f_{n_k}(T^{a_k} x) + \sum_{k=1}^{v} f_1(T^{b_k} x) + \sum_{k=1}^{w} f_1(T^{c_k} x).$$

Since $f_k(x) \leq 0$, we get that

$$f_m(x) \leq \sum_{k=1}^{u} f_{n_k}(T^{a_k} x) \leq \sum_{k=1}^{u} n_k(F_M(x) + \varepsilon) \leq F_M(x) \left(\sum_{k=1}^{u} n_k \right) + m\varepsilon. \tag{2.2.2}$$

We also have the inequality

$$m - \sum_{k=0}^{m-1} \chi_{B(N,M)}(T^k x) - N \leq \sum_{k=1}^{u} n_k,$$

because of the construction of the intervals: $T^k x \in B(N, M)$ always, when k belongs to a red interval. The total number of blue points $\sum_{k=1}^{u} n_k$ is given by subtracting from m the number of red points and the number of white points, but the number of white points doesn't exceed N.

Therefore by Theorem 1.3.1 and Remark 1.4.8 we have

$$\liminf_{m\to\infty} \frac{1}{m} \sum_{k=1}^{u} n_k \geq \lim_{m\to\infty} (1 - \frac{1}{m} \sum_{k=0}^{m-1} \chi_{B(N,M)}(T^k x)) = 1 - \mathbf{E}(\chi_{B(N,M)} | \mathfrak{Inv})$$

for almost every x. This combined (notice that $F_M \leq 0$) with the inequality (2.2.2) gives for an arbitrary $K \leq M$ that almost everywhere

$$\limsup_{m\to\infty} \frac{f_m(x)}{m} \leq F_M(x)(1 - \mathbf{E}(\chi_{B(N,M)} | \mathfrak{Inv})) + \varepsilon$$
$$\leq F_K(x)(1 - \mathbf{E}(\chi_{B(N,M)} | \mathfrak{Inv})) + \varepsilon.$$

Because $\chi_{B(N,M)} \to 0$ almost everywhere as $N \to \infty$, we get by Theorem 1.4.3 that

$$\mathbf{E}(\chi_{B(N,M)}|\mathfrak{Inv}) \xrightarrow{L^1} 0$$

as $N \to \infty$ and thus we can pick a strictly increasing sequence N_k such that

$$\mathbf{E}(\chi_{B(N_k,M)|\mathfrak{Inv}}) \to 0 \quad \text{as } k \to \infty$$

almost everywhere. Hence we have

$$\limsup_{m\to\infty} \frac{f_m(x)}{m} \le F_K(x) + \varepsilon$$

for almost every x. Because this holds for every K and ε, we get that

$$\limsup_{m\to\infty} \frac{f_m(x)}{m} \le \liminf_{m\to\infty} \frac{f_m(x)}{m},$$

and hence we get the convergence.

Now by the Monotone Convergence Theorem we have

$$\int f_* = \lim_{M\to\infty} \int \max(f_*, -M),$$

which by Dominated Convergence Theorem equals

$$\lim_{M\to\infty} \lim_{n\to\infty} \int \max(\frac{f_n}{n}, -M).$$

Next we note that the sequence $\int \max(\frac{f_{2^n}}{2^n}, -M)$, $n = 1, 2, \ldots$, is decreasing since by (2.2.1) we have

$$\frac{f_{2^{n+1}}(x)}{2^{n+1}} \le \frac{f_{2^n}(x) + f_{2^n}(T^{2^n} x)}{2^{n+1}}$$

so that

$$\max(\frac{f_{2^{n+1}}}{2^{n+1}}, -M) \le \frac{1}{2} \max(\frac{f_{2^n}}{2^n}, -M) + \frac{1}{2} \max(\frac{f_{2^n} \circ T^{2^n}}{2^n}, -M)$$

and integrating both sides gives

$$\int \max(\frac{f_{2^{n+1}}}{2^{n+1}}, -M) \le \int \max(\frac{f_{2^n}}{2^n}, -M).$$

Thus

$$\begin{aligned} \int f_* &= \lim_{M\to\infty} \lim_{n\to\infty} \int \max(\frac{f_{2^n}}{2^n}, -M) \\ &= \lim_{n\to\infty} \lim_{M\to\infty} \int \max(\frac{f_{2^n}}{2^n}, -M) = \lim_{n\to\infty} \int \frac{f_{2^n}}{2^n}, \end{aligned}$$

where we could exchange the order of the limits because $\int \max(\frac{f_{2^n}}{2^n}, -M)$ is decreasing both in M and n. Next note that the sequence $\int f_n$, $n = 1, 2, \ldots$, satisfies

$$\int f_{n+m} \leq \int f_n + \int f_m,$$

so in particular the constant functions $g_n = \int f_n$ satisfy the subadditivity condition (2.2.1). We have already proven that this implies that the sequence g_n/n converges almost everywhere, which in particular means that

$$\lim_{n\to\infty} \int \frac{f_n}{n}$$

exists. Hence it follows that

$$\int f_* = \lim_{n\to\infty} \int \frac{f_{2^n}}{2^n} = \lim_{n\to\infty} \int \frac{f_n}{n}.$$

In the end we claim that

$$\lim_{n\to\infty} \frac{1}{n} \int f_n = \inf_{n\in\mathbf{N}} \frac{1}{n} \int f_n.$$

Suppose in contrary that there is some $N \in \mathbf{N}$ for which

$$\frac{1}{n} \int f_n > \frac{1}{N} \int f_N$$

for all $n > N$. But this gives a contradiction, since by (2.2.1)

$$\frac{1}{2N} \int f_{2N} \leq \frac{1}{2N} \int f_N(x) + f_N(T^N x)\, d\mu(x) = \frac{1}{N} \int f_N.$$

Finally we note that if T is ergodic then f_* is constant almost everywhere. ∎

We make the following remark, which lets us loosen the assumptions of the Subadditive Ergodic Theorem a bit. The proof is taken from [Rue79].

Remark 2.2.3. *In the assumptions of Theorem 2.2.1 it is enough to assume that $f_1^+ \in L^1(X)$ instead of $f_n \in L^1(X)$.* □

Proof. From $f_1^+ \in L^1(X)$ it follows that

$$\begin{aligned} f_n^+ = f_n + f_n^- &\leq f_1 + f_1 \circ T + \cdots + f_1 \circ T^{n-1} + f_n^- \\ &\leq f_1^+ + f_1^+ \circ T + \cdots + f_1^+ \circ T^{n-1} + f_n^-. \end{aligned}$$

Since $f_n^- = 0$ when $f_n^+ > 0$, we have

$$f_n^+ \leq f_1^+ + f_1^+ \circ T + \cdots + f_1^+ \circ T^{n-1} \in L^1(X),$$

for all $n \in \mathbf{N}$. For every $N \in \mathbf{N}$ define $f_n^N(x) = \max\{f_n(x), -nN\}$. One easily checks that for a fixed N the functions $f_n^N(x)$ satisfy the subadditivity condition (2.2.1):

$$\begin{aligned} f_{n+m}^N(x) &= \max\{f_{n+m}(x), -(n+m)N\} \\ &\leq \max\{f_n(x) + f_m(T^n x), -nN - mN\} \\ &\leq \max\{f_n(x), -nN\} + \max\{f_m(T^n x), -mN\} \\ &= f_n^N(x) + f_m^N(T^n x). \end{aligned}$$

Now because $f_n^+ \in L^1(X)$, also $(f_n^N)^+ \in L^1(X)$. Since $(f_n^N)^-$ is bounded, $f_n^N \in L^1(X)$. Therefore Theorem 2.2.1 holds for f_n^N. For every $N \in \mathbf{N}$ the limit

$$\lim_{n\to\infty} \frac{1}{n} f_n^N(x) = f^N(x)$$

exists almost everywhere, and it easily follows that there is a function $f_*: X \to \mathbf{R} \cup \{-\infty\}$ such that

$$f^N(x) = \max\{f_*(x), -N\}$$

for almost every $x \in X$, with the limit

$$\lim_{n\to\infty} \frac{1}{n} f_n(x) = f_*(x)$$

existing for almost every $x \in X$. Moreover

$$\int f_* = \inf_N \int f^N = \inf_N \inf_n \frac{1}{n} \int f_n^N = \inf_n \frac{1}{n} \int f_n.$$

Finally f_*^+ is in $L^1(X)$, because of the bound

$$\int \frac{f_n^+}{n} \le \int \frac{f_1^+ + f_1^+ \circ T + \cdots + f_1^+ \circ T^{n-1}}{n} = \int f_1^+ < \infty,$$

so that by Fatou's lemma

$$\int f_*^+ = \int \liminf_{n\to\infty} \frac{1}{n} f_n^+ \le \liminf_{n\to\infty} \int f_1^+ < \infty.$$

■

Finally we have the following lemma.

Lemma 2.2.4. *Let $(X, \mathcal{B}, \mu)$ be a Borel probability space, T an invertible measure-preserving transformation, and suppose that the functions $f_n: X \to \mathbf{R}$, $n \in \mathbf{N}$, satisfy the assumptions of Theorem 2.2.1 (or the above remark). Then it follows that the limit*

$$\lim_{n\to\infty} \frac{f_n(T^{-n}x)}{n}$$

exists for almost every $x \in X$ and is equal to

$$\lim_{n\to\infty} \frac{f_n(x)}{n}.$$

□

Proof. Fix first some representative $f_n \in \mathcal{L}^1(X, \mu)$ for each f_n and let $g_n = f_n \circ T^{-n}$. Then g_n is subadditive (with respect to the transformation T^{-1}) since

$$\begin{aligned} g_{n+m}(x) &= f_{n+m}(T^{-n-m}x) \le f_n(T^{-n-m}x) + f_m(T^{-m}x) \\ &= g_m(x) + g_n(T^{-m}x). \end{aligned}$$

Now by Theorem 2.2.1 $\frac{f_n}{n}$ and $\frac{g_n}{n}$ converge almost everywhere to functions f_* and g_* respectively. Let $N = \{x \in X : f_*(x) \neq g_*(x)\}$. We will show that N has measure 0.

Consider the ergodic decomposition (Theorem 1.5.3) λ of μ. Then for λ-almost every ν we have by Theorem 2.2.1 that

$$g_*(x) = \inf_n \int_X \frac{g_n}{n}\, d\nu = \inf_n \int_X \frac{f_n}{n}\, d\nu = f_*(x)$$

ν-almost everywhere. This means that

$$\int_X \chi_N \, d\nu = 0$$

for almost every ν. It follows that

$$\mu(N) = \int_X \chi_N \, d\mu = \int_{\mathcal{E}^T} \int_X \chi_N \, d\nu \, d\lambda(\nu) = 0,$$

and hence $g_*(x) = f_*(x)$ almost everywhere.

Because g_* and f_* will be the same for any chosen representatives f_n up to a set of measure zero, we conclude that this holds also in the case $f_n \in L^1(X, \nu)$. ∎

We'll now start proving Oseledets' theorem. First we'll need the following corollary of the Subadditive Ergodic Theorem. The proof of the Oseledets' theorem follows the proof given by Ruelle in [Rue79], which in turn is based on the proof given in [Rag79].

Corollary 2.2.5. *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system and let $A\colon X \to \mathbf{R}^{m\times m}$ be a measurable mapping from X to the $m \times m$ matrices, which satisfies*

$$x \mapsto \log^+ \|A(x)\| \in L^1(X).$$

Let us write

$$A_x^n = A(T^{n-1}x) \dots A(Tx)A(x).$$

Then the limit

$$\lim_{n\to\infty} \frac{1}{n} \log \|A_x^n\| = \lambda(x)$$

exists for almost every x, $\lambda\colon X \to \mathbf{R} \cup \{-\infty\}$ is a T-invariant function and $\lambda^+ \in L^1(X)$. Moreover

$$\lim_{n\to\infty} \frac{1}{n} \int \log \|A_x^n\| \, \mu(dx) = \int \lambda.$$ □

Proof. Consider $f_n(x) = \log \|A_x^n\|$. The subadditivity condition is satisfied since

$$\begin{aligned} f_{n+m}(x) &= \log \|A_x^{n+m}\| = \log \|A(T^{n+m-1}x) \dots A(x)\| \\ &\leq \log \|A(T^{n-1}x) \dots A(x)\| + \log \|A(T^{n+m-1}x) \dots A(T^n x)\| \\ &= f_n(x) + f_m(T^n x). \end{aligned}$$

Also $f_1^+ \in L^1(X)$ by assumption, so we are done by Theorem 2.2.1 and Remark 2.2.3. ∎

Remark 2.2.6. *The norm for the matrices above is the operator norm given by* $\|A\| = \sup\{\|Av\| : \|v\| \le 1\}$ *where we have endowed the space* $\mathbf{R}^m$ *with the usual Euclidean norm. In practice it doesn't matter, which equivalent norm is picked, however, since the constant multiples vanish when we take the limit.* □

Remark 2.2.7. *By using Lemma 2.2.4 we also see that in Corollary 2.2.5*

$$\lambda(x) = \lim_{n\to\infty} \frac{1}{n} \log \|A^n_{T^{-n}x}\| = \lim_{n\to\infty} \frac{1}{n} \log \|(A_x^{-n})^{-1}\|,$$

where

$$A_x^{-n} = (A(T^{-n}x))^{-1}(A(T^{-n+1}x))^{-1} \dots (A(T^{-1}x))^{-1}.$$ □

2.2.2. *The limit of a product of matrices*

The proof of the next theorem will be a bit long and technical. Some parts of it will be done in the appendix.

Theorem 2.2.8. *Let* $A_n \in \mathbf{R}^{m\times m}$, $n = 1, 2, \dots,$ *be a sequence of matrices and write*

$$A^n = A_n A_{n-1} \dots A_1.$$

Assume that

$$\limsup_{n\to\infty} \frac{1}{n} \log \|A_n\| \le 0,$$

and that the limit

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

exists (we allow it to be $-\infty$*) for every* $q = 1, 2, \dots, m$*. Then the limit*

See Appendix B for the definition of the exterior power $A^{\wedge q}$ *of a linear operator A.*

$$\lim_{n\to\infty} ((A^n)^* A^n)^{\frac{1}{2n}} = \Lambda$$

exists, and the distinct eigenvalues $\exp\lambda^{(1)} < \dots < \exp\lambda^{(s)}$ *of* Λ *are real (we allow* $\lambda^{(1)} = -\infty$*). Let* $U^{(1)}, \dots, U^{(s)}$ *be the corresponding eigenspaces, and let*

$$V^0 = \{0\}, \quad V^{(r)} = U^{(1)} + \dots + U^{(r)}, \text{ for } r = 1, \dots, s.$$

Then if $v \in V^{(r)} \setminus V^{(r-1)}$ *we have*

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n v\| = \lambda^{(r)}.$$ □

Proof. We consider the matrix $(A^{n*}A^n)^{1/2}$ and let $0 \le \lambda_n^{(1)} \le \dots \le \lambda_n^{(m)}$ be its eigenvalues. (See Appendix B, B.8 and B.7.) Now we have assumed that the limit

$$\lim_{n\to\infty} \frac{1}{n} \log \|(A^n)^{\wedge q}\| = \lim_{n\to\infty} \frac{1}{n} \log \prod_{p=m-q+1}^{m} \lambda_n^{(p)}$$

exists for every $q = 1, 2, \dots, m$. Hence also (by setting $q = 1$) the limit

$$\lim_{n\to\infty} \frac{1}{n} \log \lambda_n^{(m)} = \xi^{(m)}$$

exists. If we now set $q = 2$, we get that

$$\lim_{n\to\infty} \frac{1}{n} \log(\lambda_n^{(m)} \lambda_n^{(m-1)}) = \lim_{n\to\infty} \left(\frac{1}{n} \log \lambda_n^{(m)} + \frac{1}{n} \log \lambda_n^{(m-1)} \right)$$

exists. If $\lim_{n\to\infty} \frac{1}{n} \log \lambda_n^{(m)} = -\infty$, then also $\lim_{n\to\infty} \frac{1}{n} \log \lambda_n^{(m-1)} = -\infty$ since $0 \le \lambda_n^{(m-1)} \le \lambda_n^{(m)}$. Otherwise

$$\lim_{n\to\infty} \frac{1}{n} \log \lambda_n^{(m-1)} = \lim_{n\to\infty} \frac{1}{n} \log(\lambda_n^{(m)} \lambda_n^{(m-1)}) - \lim_{n\to\infty} \frac{1}{n} \log \lambda_n^{(m)}$$

exists. Continuing similarly we see that

$$\lim_{n\to\infty} \frac{1}{n} \log \lambda_n^{(p)} = \xi^{(p)}$$

exists for all $p = 1, 2, \dots, m$. We let $\lambda^{(1)} < \dots < \lambda^{(s)}$ be the distinct values $\{\xi^{(p)} : 1 \le p \le m\}$. Now for each $1 \le r \le s$ define

$$P_r = \{p \in \{1, \dots, m\} : \lim_{n\to\infty} \frac{1}{n} \log \lambda_n^{(p)} = \lambda^{(r)}\}$$

and for each $n \in \mathbf{N}$ let $U_n^{(r)}$ be the subspace spanned by the eigenvectors of the eigenvalues $\lambda_n^{(p)}$, $p \in P_r$.

Next we will prove the following technical lemma.

Lemma 2.2.9. *For each $\delta > 0$ there exists $K > 0$ (depending on δ and the eigenvalues $\lambda^{(r)}$, $1 \le r \le s$) such that for all $k > 0$ the following holds for all sufficiently large n and for all $1 \le r, r' \le s$:*

$$\sup\{|\langle u, u' \rangle| : u \in U_n^{(r)}, u' \in U_{n+k}^{(r')}, \|u\| = \|u'\| = 1\} \le K e^{-n(|\lambda^{(r')} - \lambda^{(r)}| - \delta)}.$$

□

Proof. Suppose first that $r < r'$. Let $v_{r,r'}^k$ be the orthogonal projection of $u \in \sum_{t \le r} U_n^{(t)}$ onto $\sum_{t \ge r'} U_{n+k}^{(t)}$. We will attempt to prove that

$$\|v_{r,r'}^k\| \le K \|u\| e^{-n(|\lambda^{(r')} - \lambda^{(r)}| - \delta)}. \tag{2.2.3}$$

The result then follows, since

$$\sup\{|\langle u, u' \rangle| : u \in U_n^{(r)}, u' \in U_{n+k}^{(r')}, \|u\| = \|u'\| = 1\}$$

is the largest norm of the projection of any $u \in U_n^{(r)}$ with $\|u\| = 1$ onto $U_{n+k}^{(r')}$. (See B.11 for clarification.)

We may assume that $\delta < |\lambda^{(r')} - \lambda^{(r)}|$ for all $r \neq r'$. (If the result holds for some δ, it clearly holds also for all larger δ as well.) Let us set $\delta^* = \delta/s$ and let p be the smallest integer in $\{1, 2, \ldots, m\}$ so that

$$\lim_{n\to\infty} \frac{1}{n} \log \lambda_n^{(p)} = \lambda^{(r')}.$$

Now notice that since the eigenspaces $U_{n+1}^{(t)}$, $1 \leq t \leq s$, are orthogonal and since for large enough n

$$\frac{1}{n+1} \log \lambda_n^{(p)} \geq \lambda^{(r')} - \frac{\delta^*}{4},$$

we have

$$\begin{aligned} \|A^{n+1}u\| &= \|(A^{n+1*}A^{n+1})^{1/2}u\| \geq \|(A^{n+1*}A^{n+1})^{1/2}v_{r,r'}^1\| \\ &\geq e^{(n+1)(\lambda^{(r')} - \frac{\delta^*}{4})}\|v_{r,r'}^1\|. \end{aligned}$$

(The last inequality follows from the fact that $(A^{n+1*}A^{n+1})^{1/2}$ stretches the vector $v_{r,r'}^1$ by at least the smallest eigenvalue $\lambda_n^{(p)}$ corresponding to $U_{n+1}^{r'}$.) Similarly we have

$$\|A^n u\| \leq e^{n(\lambda^{(r)} + \frac{\delta^*}{4})}\|u\|.$$

Because we have assumed

$$\limsup_{n\to\infty} \frac{1}{n} \log \|A_n\| \leq 0,$$

there is a $C > 0$ such that for all n

$$\log \|A_n\| \leq C + n\frac{\delta^*}{4},$$

and hence

$$\|A^{n+1}u\| \leq \|A_{n+1}\|\|A^n u\| \leq e^{C+(n+1)\frac{\delta^*}{4}} e^{n(\lambda^{(r)} + \frac{\delta^*}{4})}\|u\|.$$

Combining these we get that

$$\|v_{r,r'}^1\| \leq e^{-n(\lambda^{(r')} - \lambda^{(r)} - \delta^*) + C - \lambda^{(r')} + \frac{\delta^*}{4} - \frac{n\delta^*}{4}}\|u\|.$$

Now if n is so large that $C + \frac{\delta^*}{4} - \lambda^{(r')} \leq \frac{n\delta^*}{4}$, then

$$\|v_{r,r'}^1\| \leq e^{-n(\lambda^{(r')} - \lambda^{(r)} - \delta^*)}\|u\|.$$

Now by induction on k we have

$$
\begin{aligned}
\|v^k_{r,r+1}\| &= \| \operatorname{proj}(u, \sum_{t\geq r+1} U^{(t)}_{n+k})\| \\
&= \| \operatorname{proj}\left(\operatorname{proj}(u, \sum_{t\leq r} U^{(t)}_{n+k-1}) + \operatorname{proj}(u, \sum_{t\geq r+1} U^{(t)}_{n+k-1}), \sum_{t\geq r+1} U^{(t)}_{n+k}\right)\| \\
&\leq \| \operatorname{proj}(u, \sum_{t\leq r} U^{(t)}_{n+k-1})\| e^{-(n+k-1)(\lambda^{(r+1)}-\lambda^{(r)}-\delta^*)} \\
&\quad + \| \operatorname{proj}(u, \sum_{t\geq r+1} U^{(t)}_{n+k-1})\| \\
&\leq \|u\| e^{-(n+k-1)(\lambda^{(r+1)}-\lambda^{(r)}-\delta^*)} + \sum_{i=0}^{k-2} \|u\| e^{-(n+i)(\lambda^{(r+1)}-\lambda^{(r)}-\delta^*)} \\
&= \sum_{i=0}^{k-1} \|u\| e^{-(n+i)(\lambda^{(r+1)}-\lambda^{(r)}-\delta^*)}.
\end{aligned}
$$

And thus

$$
\|v^k_{r,r+1}\| \leq \sum_{i=0}^{\infty} \|u\| e^{-(n+i)(\lambda^{(r+1)}-\lambda^{(r)}-\delta^*)} = K_1 \|u\| e^{-n(\lambda^{(r+1)}-\lambda^{(r)}-\delta^*)},
$$

where

$$
K_1 = \frac{1}{1 - e^{-(\lambda^{(r+1)}-\lambda^{(r)}-\delta^*)}}.
$$

Similarly we deduce again by induction that

$$
\begin{aligned}
\|v^k_{r,r+2}\| &= \| \operatorname{proj}(u, \sum_{t\geq r+2} U^{(t)}_{n+k})\| \\
&= \| \operatorname{proj}\Big(\operatorname{proj}(u, \sum_{t\leq r} U^{(t)}_{n+k-1}) + \\
&\quad \operatorname{proj}(u, U^{(r+1)}_{n+k-1}) + \operatorname{proj}(u, \sum_{t\geq r+2} U^{(t)}_{n+k-1}), \sum_{t\geq r+2} U^{(t)}_{n+k}\Big)\| \\
&\leq \|u\| e^{-(n+k-1)(\lambda^{(r+2)}-\lambda^{(r)}-\delta^*)} + \\
&\quad \|v^{k-1}_{r,r+1}\| e^{-(n+k-1)(\lambda^{(r+2)}-\lambda^{(r+1)}-\delta^*)} + \|v^{k-1}_{r,r+2}\| \\
&\leq \|u\| e^{-(n+k-1)(\lambda^{(r+2)}-\lambda^{(r)}-\delta^*)} + \\
&\quad K_1 \|u\| e^{-n(\lambda^{(r+1)}-\lambda^{(r)}-\delta^*)} e^{-(n+k-1)(\lambda^{(r+2)}-\lambda^{(r+1)}-\delta^*)} + \|v^{k-1}_{r,r+2}\| \\
&\leq \sum_{j=0}^{k-1} \|u\| e^{-(n+j)(\lambda^{(r+2)}-\lambda^{(r)}-\delta^*)} + \\
&\quad \sum_{j=0}^{k-1} K_1 \|u\| e^{-n(\lambda^{(r+2)}-\lambda^{(r)}-2\delta^*)} e^{-j(\lambda^{(r+2)}-\lambda^{(r+1)}-\delta^*)}.
\end{aligned}
$$

And thus

$$
\|v^k_{r,r+2}\| \leq \|u\| \frac{e^{-n(\lambda^{(r+2)}-\lambda^{(r)}-\delta^*)}}{1 - e^{-(\lambda^{(r+2)}-\lambda^{(r)}-\delta^*)}} + K_1 \|u\| \frac{e^{-n(\lambda^{(r+2)}-\lambda^{(r)}-2\delta^*)}}{1 - e^{-(\lambda^{(r+2)}-\lambda^{(r+1)}-\delta^*)}},
$$

from which we gather that

$$\|v^k_{r,r+2}\| \leq K_2 \|u\| e^{-n(\lambda^{(r+2)} - \lambda^{(r)} - 2\delta^*)}$$

for a constant K_2. In a similar fashion we get that

$$\|v^k_{r,r'}\| \leq K_{r'-r} \|u\| e^{-n(\lambda^{(r')} - \lambda^{(r)} - (r'-r)\delta^*)}$$

for some constants $K_{r'-r}$ and since $(r' - r)\delta^* \leq \delta$, we get the result by taking the largest of these constants.

We still have the cases $r = r'$ and $r > r'$ to consider. The case $r = r'$ is trivial, since then $K = 1$ will do: $e^{n\delta}$ is an increasing function and $|\langle u, u' \rangle| \leq 1$ by the Cauchy-Schwarz inequality.

Let's look at the case $r > r'$. Let $u_1, \ldots, u_m$ and $u'_1, \ldots, u'_m$ be orthonormal eigenvectors of $(A^{n*}A^n)^{1/2}$ and $(A^{n+k*}A^{n+k})^{1/2}$ respectively, ordered so that if $u_i \in U_n^{(r_i)}$, then r_i form an increasing sequence, and similarly for $u'_i \in U_{n+k}^{(r_i)}$. We will consider the matrix S with entries

$$S_{ij} = \left\langle u_i, u'_j \right\rangle.$$

Now we have already proven the bound we want for all S_{ij}, $i \leq j$. It remains to prove it for $i > j$. (This is really enough since given any pair of vectors $u \in U_n^{(r)}$ and $u' \in U_{n+k}^{(r')}$ we can choose $u_1, \ldots, u_m$ and $u'_1, \ldots, u'_m$ so that u and u' appear among them.) First notice that S is orthogonal: The inner product between the ith and jth row is given by

$$\begin{aligned}\sum_{k=1}^{m} \langle u_i, u'_k \rangle \left\langle u_j, u'_k \right\rangle &= \sum_{k=1}^{m} \left\langle u_i, \left\langle u_j, u'_k \right\rangle u'_k \right\rangle = \left\langle u_i, \sum_{k=1}^{m} \left\langle u_j, u'_k \right\rangle u'_k \right\rangle \\ &= \left\langle u_i, u_j \right\rangle = \delta_{ij},\end{aligned}$$

and similarly for columns. Therefore $S^* = S^{-1}$ and $\det(S) = \pm 1$. On the other hand we have

$$S^{-1} = \frac{\operatorname{adj}(S)}{\det(S)} = \pm \operatorname{adj}(S),$$

where $\operatorname{adj}(S)$ is the adjugate matrix of S, which is defined as $\operatorname{adj}(S)_{ij} = C_{ji}$, where $C_{ij} = (-1)^{i+j} M_{ij}$ is the (i, j)th cofactor of S and M_{ij} is the determinant of the matrix, which results by deleting row i and column j from S, which we will denote by $S[i, j]$. Now we have

$$S_{ij} = (S^{-1})_{ji} = \pm \operatorname{adj}(S)_{ji} = \pm C_{ij} = \pm M_{ij},$$

from which we get the equation

$$S_{ij} = \pm \sum_{\sigma} \operatorname{sgn}(\sigma) \prod_{k=1}^{m-1} S[i, j]_{k\sigma(k)}, \tag{2.2.4}$$

where σ runs over the permutations of $\{1, 2, \ldots, m-1\}$.

Notice that every entry of S is bounded by 1 by Cauchy-Schwarz inequality. Since we are allowed to choose a δ' for already obtained bounds, we will be done if we can show that in every term of the sum (2.2.4) there exist k factors $S[i,j]_{a_1\sigma(a_1)},\dots,S[i,j]_{a_k\sigma(a_k)}$ $(1\le k\le m)$ that are respectively bounded by

$$S[i,j]_{a_l\sigma(a_l)} \le e^{-n(\lambda^{(p'_l)}-\lambda^{(p_l)}-\delta')},$$

where $p'_l > p_l$, $l=1,2,\dots,k$, satisfy

$$\{r'+1,r'+2,\dots,r\}\subset\bigcup_{l=1}^{k}\{p_l+1,p_l+2,\dots,p'_l\}.$$

Indeed, then there exist numbers $0=a_0<a_1<\cdots<a_q=r-r'$ and distinct intervals $I_h=\{p_{l_h}+1,\dots,p'_{l_h}\}$, $1\le h\le q$, so that

$$\{r'+1,r'+2,\dots,r\}=\bigcup_{h=0}^{q-1}\{r'+a_h+1,\dots,r'+a_{h+1}\},$$

and

$$\{r'+a_h+1,\dots,r'+a_{h+1}\}\subset I_{h+1}$$

for all $0\le h\le q-1$. Thus we get

$$\begin{aligned}\lambda^{(r)}-\lambda^{(r')}&=\sum_{h=0}^{q-1}\left(\lambda^{(r'+a_{h+1})}-\lambda^{(r'+a_h)}\right)\\&\le\sum_{h=1}^{q}\left(\lambda^{(p'_{l_h})}-\lambda^{(p_{l_h})}\right)\le\sum_{l=1}^{k}\left(\lambda^{(p'_l)}-\lambda^{(p_l)}\right),\end{aligned}$$

and hence

$$\prod_{l=1}^{k}e^{-n(\lambda^{(p'_l)}-\lambda^{(p_l)}-\delta')}\le e^{-n(\lambda^{(r)}-\lambda^{(r')}-\delta)}$$

for a suitably chosen δ'.

Let $r'=r_j$ and $r=r_i$ and look at the table below. The entry with coordinates (p,p') contains the intersection of $\{r_{j+1},\dots,r_i\}$ with $\{p+1,p+2,\dots,p'\}$, where $e^{-n(\lambda^{(p')}-\lambda^{(p)}-\delta)}$ is the bound we already have for $S_{p,p'}$. We have to show that for each term of the sum (2.2.4) the union of the corresponding entries in the table below is $\{r_{j+1},\dots,r\}\supset\{r'+1,\dots,r\}$. But this is clear, since when we calculate one of the terms, we choose from each row one entry, according to the permutation σ, so no column gets chosen twice. It's apparent from looking at the table, that for each $1\le k\le i-j$ there are $j+k-1$ rows that contain r_{j+k} in any but $j+k-2$ of their entries, so we see that r_{j+k} belongs to at least one of the chosen sets by Pigeonhole Principle.

	r_1	r_2	$\dots$	r_{j-1}	r_j	r_{j+1}	r_{j+2}	r_{j+3}	$\dots$	r_i	$\dots$	r_m
r_1	–	–	$\dots$	–		r_{j+1}	r_{j+1}, r_{j+2}	$r_{j+1}, \dots, r_{j+3}$	$\dots$	$r_{j+1}, \dots, r_i$	$\dots$	$r_{j+1}, \dots, r_i$
r_2	–	–	$\dots$	–		r_{j+1}	r_{j+1}, r_{j+2}	$r_{j+1}, \dots, r_{j+3}$	$\dots$	$r_{j+1}, \dots, r_i$	$\dots$	$r_{j+1}, \dots, r_i$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
r_{j-1}	–	–	$\dots$	–		r_{j+1}	r_{j+1}, r_{j+2}	$r_{j+1}, \dots, r_{j+3}$	$\dots$	$r_{j+1}, \dots, r_i$	$\dots$	$r_{j+1}, \dots, r_i$
r_j	–	–	$\dots$	–		r_{j+1}	r_{j+1}, r_{j+2}	$r_{j+1}, \dots, r_{j+3}$	$\dots$	$r_{j+1}, \dots, r_i$	$\dots$	$r_{j+1}, \dots, r_i$
r_{j+1}	–	–	$\dots$	–		–	r_{j+2}	r_{j+2}, r_{j+3}	$\dots$	$r_{j+2}, \dots, r_i$	$\dots$	$r_{j+2}, \dots, r_i$
r_{j+2}	–	–	$\dots$	–		–	–	r_{j+3}	$\dots$	$r_{j+3}, \dots, r_i$	$\dots$	$r_{j+3}, \dots, r_i$
r_{j+3}	–	–	$\dots$	–		–	–	–	$\dots$	$r_{j+4}, \dots, r_i$	$\dots$	$r_{j+4}, \dots, r_i$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
r_{i-1}	–	–	$\dots$	–		–	–	–	$\dots$	r_i	$\dots$	r_i
r_i					$\times$							
r_{i+1}	–	–	$\dots$	–		–	–	–	$\dots$	–	$\dots$	–
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
r_m	–	–	$\dots$	–		–	–	–	$\dots$	–	$\dots$	–

This finishes the proof. ∎

We will use the lemma to show that for each r the subspaces $U_n^{(r)}$ form a Cauchy sequence in the Grassmannian $\mathbf{Gr}(d_r, \mathbf{R}^m)$ where d_r is the dimension of $U_n^{(r)}$. (See Appendix C for the basic definitions.) Indeed, since

$$\left(U_{n+k}^{(r)}\right)^{\perp} = \bigcup_{r \neq r'} U_{n+k}^{(r)},$$

we have

$$\begin{aligned} d(U_n^{(r)}, U_{n+k}^{(r)}) &= \sup\{|\langle u, u' \rangle| : u \in U_n^{(r)}, u' \in \left(U_{n+k}^{(r)}\right)^{\perp}, \|u\| = \|u'\| = 1\} \\ &\leq K e^{-nC}, \end{aligned}$$

where C is the smallest of $|\lambda^{(r')} - \lambda^{(r)}| - \delta$, $r' \neq r$. Given that we have chosen δ so small that $\delta < |\lambda^{(r')} - \lambda^{(r)}|$ for all r', this shows that the sequence $U_n^{(r)}$ is Cauchy and therefore $U_n^{(r)}$ converges to some subspace $U^{(r)}$. This suffices to show that the limit

$$\lim_{n \to \infty} (A^{n*} A^n)^{\frac{1}{2n}} = \Lambda$$

exists, since the eigenvalues and eigenspaces of Λ determine Λ (in a continuous way).

We still have to show that for $v \in V^{(r)} \setminus V^{(r-1)}$

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

holds. We can clearly assume that $\|v\| = 1$. Assume first that $v \in U^{(r)} \setminus \{0\}$. Then by Lemma 2.2.9 we have

$$\max\{|\langle v, u' \rangle| : u' \in U_n^{(r')}, \|u'\| = 1\} \leq K e^{-n(|\lambda^{(r')} - \lambda^{(r)}| - \delta)}$$

for every $\delta > 0$. Fix $n \in \mathbf{N}$ and write v as $v = c_1 u_1 + \cdots + c_s u_s$ where u_j, $j = 1, \dots, s$, in $U_n^{(j)}$ are orthonormal. Fix now $\delta > 0$ such that

$$\delta < \frac{1}{4} \min_{1 \leq r'' < r' \leq s} |\lambda^{(r'')} - \lambda^{(r')}|.$$

Then for all large enough n the estimates

$$e^{n\lambda^{(j)}-n\delta} \le \|(A^{n*}A^n)^{1/2}u_j\| \le e^{n\lambda^{(j)}+n\delta}$$

and

$$c_j = \langle v, u_j \rangle \le Ke^{-n(|\lambda^{(j)}-\lambda^{(r)}|-\delta)}$$

hold for all $j = 1, \ldots, s$. Let q_j be the eigenvalues of $(A^{n*}A^n)^{1/2}$ corresponding to the vectors u_j and choose r' so that $|\lambda^{(r')} - \lambda^{(r)}|$ is minimized. Then for large enough n the above estimates give us

$$c_j q_j \le Ke^{-n(\lambda^{(j)}-\lambda^{(r)}-\delta)} e^{n\lambda^{(j)}+n\delta} = Ke^{n(\lambda^{(r)}+2\delta)}, \quad j = 1, \ldots, s$$

and

$$\begin{aligned} c_r q_r &= \left(\sqrt{1 - \sum_{j\neq r} c_j^2}\right) q_r = \sqrt{q_r^2 - \sum_{j\neq r} c_j^2 q_r^2} \\ &\ge \sqrt{e^{2n\lambda^{(r)}-2n\delta} - (s-1)K^2 e^{2n\lambda^{(r)}+4n\delta-2n|\lambda^{(r')}-\lambda^{(r)}|}} \\ &= e^{n(\lambda^{(r)}-2\delta)} \sqrt{e^{2n\delta} - (s-1)K^2 e^{2n(4\delta-|\lambda^{(r')}-\lambda^{(r)}|)}} \\ &\ge e^{n(\lambda^{(r)}-2\delta)}. \end{aligned}$$

Notice now that

$$\|A^n v\| = \|(A^{n*}A^n)^{1/2} v\| = \|\sum_{j=1}^{s} c_j q_j u_j\| = \sqrt{c_r^2 q_r^2 + \sum_{j\neq r} c_j^2 q_j^2}.$$

Thus for large enough n we have

$$\lambda^{(r)} - 2\delta \le \frac{1}{n} \log \|A^n v\| \le \frac{\log(sK^2)}{2n} + \lambda^{(r)} + 2\delta \le \lambda^{(r)} + 3\delta,$$

and therefore

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

If $v \in V^{(r)} \setminus V^{(r-1)}$, then v can be written as $v = v_1 + \cdots + v_r$, where $v_k \in U^{(k)}$, $k = 1, \ldots, r$, and $v_r \neq 0$. By a similar argument as in the proof of Theorem 2.1.3 (because $\lambda^{(r)}$ is the largest among $\lambda^{(j)}, 1 \le j \le r$) we have

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

■

We can also make the following remark about the above theorem.

Remark 2.2.10. *The convergence of*

$$\frac{1}{n} \log \|A^n v\|$$

in Theorem 2.2.8 is uniform when v satisfies $\|v\| = 1$. □

Proof. The estimates made at the end of the proof of Theorem 2.2.8 were based on Lemma 2.2.9, which involves this uniform behaviour. ■

2.2.3. *Oseledets' theorem*

We are now ready to prove the following version of Oseledets' theorem.

THEOREM 2.2.11. *Let M be a compact Riemannian manifold, $f\colon M \to M$ a diffeomorphism, and μ an f-invariant probability (Borel) measure on M. There is a subset $\Gamma \subset M$ such that $\mu(\Gamma) = 1$, $f\Gamma = \Gamma$ and for all $x \in \Gamma$ the following statements hold:*

1. *There is a decomposition*

$$T_x M = \bigoplus_{k=1}^{s(x)} E_k(x)$$

of the tangent space $T_x M$ into subspaces $E_k(x)$, $1 \le k \le s(x)$.

2. *The subspaces $E_k(x)$ are invariant under Tf, that is*

$$T_x f E_k(x) = E_k(f(x)),$$

and they depend measurably on x.

3. *If $v \in E_k(x) \smallsetminus \{0\}$, then*

$$\lim_{n\to\pm\infty} \frac{1}{n} \log \| T_x f^n v \| = \chi_k(x), \tag{2.2.5}$$

where $\chi_1(x) < \cdots < \chi_{s(x)}(x)$ are the Lyapunov exponents of the transformation f at the point x.

4. *Let I_1 and I_2 form a partition of $\{1, 2, \dots, s(x)\}$ and set*

$$E^1(x) = \bigoplus_{k\in I_1} E_k(x), \quad E^2(x) = \bigoplus_{k\in I_2} E_k(x).$$

Then

$$\lim_{n\to\pm\infty} \frac{1}{n} \log \sin(\sphericalangle(E^1(f^n(x)), E^2(f^n(x)))) = 0. \tag{2.2.6}$$

(I.e. the angle between the two subspaces goes to 0 at most exponentially.)

5. *There is a decomposition of the cotangent bundle*

$$T_x^* M = \bigoplus_{k=1}^{s(x)} E_k^*(x),$$

where the subspaces $E_k^*(x)$ are invariant under the codifferential

$$(T_x^* f) v_x^* : v_{f(x)} \mapsto v_x^*((T_x f)^{-1}(v_{f(x)}))$$

for all $v_x^* \in T_x^* M$, $v_{f(x)} \in T_{f(x)} M$. If

$$(e_1^1, \dots, e_{d_1}^1, \dots, e_1^{s(x)}, \dots, e_{d_{s(x)}}^{s(x)})$$

is a basis for T_xM such that $e_1^i, \dots, e_{d_i}^i$ span the subspace $E_i(x)$, then in the dual basis $((e_1^1)^*, \dots, (e_{d_1}^1)^*, \dots, (e_1^{s(x)})^*, \dots, (e_{d_{s(x)}}^{s(x)})^*)$ the covectors $(e_1^i)^*, \dots, (e_{d_i}^i)^*$ span the subspace $E_i^*(x)$. Moreover, for $v^* \in E_i^*(x)$ we have

$$\lim_{n\to\pm\infty} \frac{1}{n} \log \|T_x^* f^n v^*\| = -\chi_k(x). \tag{2.2.7}$$

(Compare (2.2.5).) □

Proof. We will do the proof in parts. In the first part the claims 1–3 are proved.

Part 1. Because M is compact, we may cover it with a finite number of coordinate charts. Fix some such covering $(C_1', x_1), \dots, (C_l', x_l)$. Now the boundaries of the sets C_k' are compact subsets of M. Hence for any point x on a boundary of some sets, there exists some largest distance r, so that all the points y on the boundaries to which the point x doesn't belong satisfy $d(x, y) \geq r$. Thus we may shrink the charts $(C_1', x_1), \dots, (C_l', x_l)$ by removing some small neighbourhoods of the boundaries to obtain charts $(C_1, x_1), \dots, (C_l, x_l)$ so that the coordinate functions x_k are still well-defined in $\overline{C_k}$. Starting with C_1 use the coordinate vector fields to fix a basis for T_xM for every $x \in C_1$. Do the same for $C_2 \setminus C_1$, $(C_3 \setminus C_1) \setminus C_2$ and so on, to get a fixed basis for T_xM for all $x \in M$. Let $A(x)$ be a matrix representing T_xf with respect to the bases chosen for T_xM and $T_{f(x)}M$. For all $x \in M$ and $n \in \mathbf{N}$ let

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

Then since

$$(T_{f^{n-1}(x)}f) \cdots (T_{f(x)}f)(T_xf) = T_xf^n,$$

we see that A_x^n represents the map T_xf^n. Moreover $x \mapsto A(x)$ and $x \mapsto A_x^n$ are measurable functions.

Now $x \mapsto \log^+ \|A(x)\|$ is in $L^1(M, \mu)$, since $A(x)$ is bounded in each of the charts C_k. (Because $A(x)$ is definable and attains its maximum value in the closures of these charts.) Birkhoff's Ergodic Theorem (Theorem 1.3.1) gives us a subset $\Gamma_1 \subset M$ with $\mu(\Gamma_1) = 1$ such that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \log^+ \|A(f^k x)\|$$

converges for all $x \in \Gamma_1$. Therefore for all $x \in \Gamma_1$

$$\lim_{n\to\infty} \frac{1}{n} \log^+ \|A(f^{n-1}x)\| = 0,$$

for otherwise the sequence $\frac{1}{n}\sum_{k=0}^{n-1} \log^+ \|A(f^k x)\|$ wouldn't be Cauchy.

Notice that for $q = 1, \dots, m$ the mapping

$$x \mapsto \log^+ \|A^{\wedge q}(x)\|$$

is in $L^1(M,\mu)$. This follows from the fact that $\|A^{\wedge q}(x)\|$ is the product of the q largest eigenvalues of $\sqrt{A^*(x)A(x)}$ and $A(x)$ is bounded. By Corollary 2.2.5 there is thus a subset $\Gamma_2 \subset M$ with $\mu(\Gamma_2) = 1$ such that

$$\lim_{n\to\infty} \frac{1}{n} \log \|(A_x^n)^{\wedge q}\| = \lim_{n\to\infty} \frac{1}{n} \log \|(A^{\wedge q})_x^n\|$$

exists for all $x \in \Gamma_2$ and $q = 1, \dots, m$, and the function

$$x \mapsto \lim_{n\to\infty} \frac{1}{n} \log \|(A_x^n)^{\wedge q}\|$$

is f-invariant.

Let $\Gamma = \Gamma_1 \cap \Gamma_2$. By applying Theorem 2.2.8 to $A_n = A(f^{n-1}x)$ we get that the limit

$$\lim_{n\to\infty} (A_x^{n*} A_x^n)^{\frac{1}{2n}} = \Lambda_x$$

exists for all $x \in \Gamma$. Let $\exp \lambda_x^{(1)} < \dots < \exp \lambda_x^{(s(x))}$ be the eigenvalues of Λ_x and $U_x^{(1)}, \dots, U_x^{(s(x))}$ the corresponding eigenspaces. Let $V_x^{(k)} = U_x^{(1)} + \dots + U_x^{(k)}$ for $k = 1, 2, \dots, s(x)$ and $V_x^{(0)} = \{0\}$. Then if $v \in V_x^{(k)} \setminus V_x^{(k-1)}$, again by Theorem 2.2.8

$$\lim_{n\to\infty} \frac{1}{n} \log \|A_x^n v\| = \lambda_x^{(k)}$$

for all $k = 1, 2, \dots, s(x)$. Because $A(x)$ is invertible for all x and by compactness of M, there is a smallest number m such that $\|A(x)v\| \geq m\|v\|$ for all $x \in M$, $v \in T_xM$. Hence in particular for any $v \in V_x^{(1)}$ $(v \neq 0)$ we have

$$\frac{1}{n} \log \|A_x^n v\| \geq \frac{1}{n} \log(m^n \|v\|) \to \log(m),$$

so that $\lambda_x^{(1)} \geq \log(m) > -\infty$.

The function $x \mapsto \lambda_x^{(k)}$ is f-invariant since if $v \in V_x^{(k)} \setminus V_x^{(k-1)}$ then

$$\begin{aligned}
\lambda_x^{(k)} &= \lim_{n\to\infty} \frac{1}{n} \log \|A_x^n v\| = \lim_{n\to\infty} \frac{1}{n} \log \|A_{f(x)}^{n-1} A(x) v\| \\
&= \lim_{n\to\infty} \frac{1}{n+1} \log \|A_{f(x)}^n (A(x)v)\| = \lim_{n\to\infty} \frac{1}{n} \log \|A_{f(x)}^n (A(x)v)\|,
\end{aligned}$$

and hence $\lambda_x^{(k)} = \lambda_{f(x)}^{(i)}$ for some i. This holding for all k and taking the ordering into account, we have $\lambda_x^{(k)} = \lambda_{f(x)}^{(k)}$ for all k. The above calculation shows also that if $v \in V_x^{(k)} \setminus V_x^{(k-1)}$ then $A(x)v \in V_{f(x)}^{(k)} \setminus V_{f(x)}^{(k-1)}$. From this it follows that the dimensions of $U_x^{(1)}, \dots, U_x^{(s(x))}$ are also invariant under f. Since $A(x)$ is invertible, we get

$$A(x)(V_x^{(k)} \setminus V_x^{(k-1)}) = V_{f(x)}^{(k)} \setminus V_{f(x)}^{(k-1)}.$$

Let now

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

for all $n \in \mathbf{N}$ and look at the eigenvalues $\exp(\ell_x^{(1)}) < \cdots < \exp(\ell_x^{(s)})$ of

$$\lim_{n\to\infty} (A_x^{-n*} A_x^{-n})^{\frac{1}{2n}} = \tilde{\Lambda}_x.$$

By Remark 2.2.7 (See also the beginning of the proof of Theorem 2.2.8 for the definition of $\xi^{(k)}$. The numbers $\xi'^{(k)}$ are defined analogously, i.e. $\xi'^{(k)} = \lim_{n\to\infty} \frac{1}{n} \log \ell_n^{(k)}$ where $\ell_n^{(k)}$ are eigenvalues of $A_x^{-n*} A_x^{-n}$.) it follows that for all $q = 1, 2, \ldots, m$,

$$\sum_{k=m-q+1}^{m} \xi_x'^{(k)} = \lim_{n\to\infty} \frac{1}{n} \log \|(A_x^{-n})^{\wedge q}\| = \lim_{n\to\infty} \frac{1}{n} \log \|((A_{f^{-n}(x)}^{n})^{-1})^{\wedge q}\|$$

$$= \lim_{n\to\infty} \frac{1}{n} \log \|((A_x^{n})^{-1})^{\wedge q}\| = \sum_{k=1}^{q} -\xi^{(k)}.$$

Now by setting $q = 1, 2, \ldots, m$, we see (as in the proof of Theorem 2.2.8) that

$$\ell^{(s)} = -\lambda_x^{(1)}, \ell^{(s-1)} = -\lambda_x^{(2)}, \ldots, \ell^{(1)} = -\lambda_x^{(s)}$$

and also the multiplicities of eigenvalues $\ell^{(k)}$ and $\lambda^{(k)}$ agree. Let

$$V_x^{(-s(x))} \subset \cdots \subset V_x^{(-1)}$$

be the filtration associated with the eigenvalues

$$-\lambda_x^{(s(x))} < \cdots < -\lambda_x^{(1)}$$

of $\tilde{\Lambda}_x$.

We'll prove the following lemma:

Lemma 2.2.12. *For $r = 2, 3, \ldots, s(x)$ and almost all $x \in M$ we have*

$$V_x^{(r-1)} \cap V_x^{(-r)} = \{0\} \tag{2.2.8}$$

and

$$V_x^{(r-1)} \oplus V_x^{(-r)} = \mathbf{R}^m. \tag{2.2.9}$$

Proof. Let S be the set of those $x \in M$ for which

$$V_x^{(r-1)} \cap V_x^{(-r)} \neq \{0\}.$$

Let $\delta > 0$ be arbitrary. Let S_n consist of those points $x \in S$ for which

$$\|A_x^n u\| \le \|u\| e^{n(\lambda_x^{(r-1)} + \delta)}$$

and

$$\|A_x^{-n} u\| \le \|u\| e^{n(-\lambda_x^{(r)} + \delta)}$$

for all $u \in V_x^{(r-1)} \cap V_x^{(-r)}$. Now if $x \in f^{-n} S_n$, then

$$\|(A_x^n)^{-1} v\| = \|A_{f^n(x)}^{-n} v\| \le \|v\| e^{n(-\lambda_x^{(r)} + \delta)}.$$

for all $v \in V_{f^n(x)}^{(r-1)} \cap V_{f^n(x)}^{(-r)}$. Setting $v = A_x^n u$ we get

$$\|A_x^n u\| \geq \|u\| e^{n(\lambda_x^{(r)} - \delta)}$$

for all $u \in V_x^{(r-1)} \cap V_x^{(-r)}$. If $x \in S_n \cap f^{-n} S_n$, then the first and last inequality can be combined to yield $\lambda_x^{(r)} - \lambda_x^{(r-1)} \leq 2\delta$. Notice now that almost every $x \in S$ belongs to S_n for large enough n. Moreover because of invariance, $f^{-1} S \subset S$ (possibly disregarding a null set). Because f is measure preserving, this implies that $\mu(S_n \cap f^{-n} S_n) \to \mu(S)$ and thus we have $\lambda_x^{(r)} - \lambda_x^{(r-1)} \leq 2\delta$ for almost all $x \in S$. Because δ is arbitrary, this implies $\mu(S) = 0$. Thus we have shown (2.2.8). The equation (2.2.9) follows since $\dim V_x^{(r-1)} + \dim V_x^{(-r)} = m$. ∎

Let now $B_k(x) = V_x^{(k)} \cap V_x^{(-k)}$ for all $k = 1, 2, \ldots, s(x)$. Then by the lemma and since $V_x^{(-k-1)} \subset V_x^{(-k)}$ for all $1 \leq k < s(x)$, we have

$$\begin{aligned}
\mathbf{R}^m &= V_x^{(-1)} \cap (V_x^{(1)} + V_x^{(-2)}) \cap (V_x^{(2)} + V_x^{(-3)}) \cap \cdots \cap V_x^{(s)} \\
&= (V_x^{(1)} \cap V_x^{(-1)} + V_x^{(-2)}) \cap (V_x^{(2)} + V_x^{(-3)}) \cap \cdots \cap V_x^{(s)} \\
&= B_1(x) \oplus (V_x^{(-2)} \cap (V_x^{(2)} + V_x^{(-3)}) \cap \cdots \cap V_x^{(s)}) \\
&= B_1(x) \oplus B_2(x) \oplus (V_x^{(-3)} \cap (V_x^{(3)} + V_x^{(-4)}) \cap \cdots \cap V_x^{(s)}) \\
&= B_1(x) \oplus B_2(x) \oplus \cdots \oplus B_{s(x)}(x).
\end{aligned}$$

Now letting $E_k(x) \subset T_x M$ to be the subspace corresponding to $B_k(x)$ we have proven the first part of the theorem. (I.e. if x has the local coordinates that were chosen in the beginning of the proof, then for every $v = \sum_{i=1}^m v^i \frac{\partial}{\partial x^i}\big|_x \in T_x M$ we have the canonical mapping $v \mapsto (v^1, \ldots, v^m)$. Take $E_k(x)$ as the preimage of $B_k(x)$ under this mapping.) The second part is also clear since we know that $V_x^{(k)}$ are invariant under Tf, and since $A(x)$ was a measurable function, it follows from the way the subspaces $V_x^{(k)}$ were constructed that $x \mapsto V_x^{(k)}$ are measurable functions. Hence $x \mapsto B_k(x)$ is measurable and finally also $x \mapsto E_k(x)$ is measurable.

To prove the third claim, let $x \in \Gamma$ and $v \in E_k(x) \setminus \{0\}$ and let

$$v^* = [v^1, \ldots, v^m]^* \in B_k(x) \subset \mathbf{R}^m.$$

Because $A_x^n v^*$ gives the coordinates of $T_x f^n v$, because there are finitely many charts and because the mappings $v \mapsto (v^1, \ldots, v^m)$ in each chart of the tangent bundle are linear isomorphisms (so that there exists constants $\alpha, \beta > 0$ such that $\alpha \|v\| \leq \|(v^1, \ldots, v^m)\| \leq \beta \|v\|$) we have

$$\lim_{n \to \infty} \frac{1}{n} \log \|T_x f^n v\| = \lim_{n \to \infty} \frac{1}{n} \log \|A_x^n v^*\| = \lambda_x^{(k)}.$$

Part 2. In this second part we'll show the 4. claim. For this we will first prove two lemmas. In this part we follow the presentation by Jairo Bochi given in [Boc08].

Lemma 2.2.13. *Let $\phi\colon M \to \mathbf{R}$ be a measurable function such that $\phi \circ f - \phi$ is integrable. Then*

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

for almost every $x \in M$. □

Proof. Let $\psi = \phi \circ f - \phi$. We get by Birkhoff's Ergodic Theorem (Theorem 1.3.1) that the limit

$$\lim_{n\to\pm\infty} \frac{\phi(f^n(x))}{n} = \lim_{n\to\pm\infty}\left(\frac{\phi(x)}{n} + \frac{1}{n}\sum_{k=0}^{n-1}\psi(f^k(x))\right) = \psi_\pm^*(x)$$

exists for μ-almost every $x \in M$. We have thus almost everywhere the implication that if $\psi_\pm^*(x) \neq 0$ then

$$|\phi(f^n(x))| \to \infty, \quad \text{as } n \to \infty. \tag{2.2.10}$$

Let $A \subset M$ be the set of points for which (2.2.10) holds. Set

$$E_k = \phi^{-1}\{r \in \mathbf{R} : k-1 \le |r| < k\}, \quad k \in \mathbf{N}.$$

For all $k \in \mathbf{N}$, let $E_k' \subset E_k$ be the set of points which returns to E_k infinitely many times. By Poincaré Recurrence Theorem (Theorem 1.2.3),

$$\mu(E_k \setminus E_k') = 0.$$

Now let $x \in A$. Since the sets E_i ($i \in \mathbf{N}$) form a partition of M, we get that $x \in E_k$ for some $k \in \mathbf{N}$. On the other hand (2.2.10) implies for all large enough n that

$$f^n(x) \in \bigcup_{i=k+1}^{\infty} E_i,$$

so that

$$x \in E_k \setminus E_k' \subset \bigcup_{i=1}^{\infty}(E_i \setminus E_i').$$

Hence

$$\mu(A) \le \sum_{i=1}^{\infty}\mu(E_i \setminus E_i') = 0,$$

which shows that $\psi_\pm^*(x) = 0$ almost everywhere. ■

Lemma 2.2.14. *Let $L\colon \mathbf{R}^d \to \mathbf{R}^d$ be an invertible linear transformation and let $v, w \in \mathbf{R}^d \setminus \{0\}$. Then*

$$\frac{1}{\|L\|\cdot\|L^{-1}\|} \le \frac{\sin \sphericalangle(Lv, Lw)}{\sin \sphericalangle(v,w)} \le \|L\|\cdot\|L^{-1}\|. \tag{2.2.11}$$

Proof. By B.12 we have the inequality

$$\|w + \alpha v\| \geq \|w\| \sin \sphericalangle(v, w), \quad \text{for all } v, w \in \mathbf{R}^d \text{ and } \alpha \in \mathbf{R},$$

and the equality occurs exactly when $\alpha = -\frac{\langle v,w\rangle}{\|v\|^2}$. Now set $u = w + \alpha v$ and choose $\alpha = -\frac{\langle Lv,Lw\rangle}{\|Lv\|^2}$ to get

$$\sin \sphericalangle(Lv, Lw) = \frac{\|Lu\|}{\|Lw\|} \geq \frac{\|L^{-1}\|^{-1} \cdot \|u\|}{\|L\| \cdot \|w\|} \geq \frac{\sin \sphericalangle(v, w)}{\|L\| \cdot \|L^{-1}\|}. \tag{2.2.12}$$

On the other hand by choosing $\alpha = -\frac{\langle v,w\rangle}{\|v\|^2}$ we similarly have

$$\sin \sphericalangle(Lv, Lw) \leq \frac{\|Lu\|}{\|Lw\|} \leq \|L\| \cdot \|L^{-1}\| \cdot \sin \sphericalangle(v, w). \tag{2.2.13}$$

We get the result by combining the equations (2.2.12) and (2.2.13). ∎

Consider now the measurable function $\phi\colon M \to \mathbf{R}$ given by the condition $\phi(x) = \log \sin \sphericalangle(E^1(x), E^2(x))$. By Lemma 2.2.14 we have

$$\begin{aligned} |\phi(f(x)) - \phi(x)| &= \left|\log \frac{\sin \sphericalangle(E^1(f(x)), E^2(f(x)))}{\sin \sphericalangle(E^1(x), E^2(x))}\right| \\ &\leq |\log \|A(x)\| + \log \|A(x)^{-1}\||. \end{aligned}$$

Since $\log \|A(x)\|$ and $\log \|A(x)^{-1}\|$ are integrable, we get

$$\lim_{n\to\pm\infty} \frac{1}{n} \log \sin \sphericalangle(E^1(f^n(x)), E^2(f^n(x))) = 0$$

by Lemma 2.2.13.

Part 3. Finally in the third part we'll show that the 5. claim holds. Fix $1 \leq i \leq s(x)$ and suppose that $e_1, \dots, e_{d_i}, \dots, e_m$ is a basis for $T_x M$ such that $e_1, \dots, e_{d_i}$ form a basis for $E_i(x)$. From the first part of the proof it follows that $T_x f e_1, \dots, T_x f e_{d_i}$ is a basis for $E_i(f(x))$. Now define $E_i^*(x)$ to be the space spanned by the covectors $e_1^*, \dots, e_{d_i}^*$. By definition $E_i^*(f(x))$ is spanned by $(T_x f e_1)^*, \dots, (T_x f e_{d_i})^*$. For all $1 \leq j, k \leq d_i$ we have

We can extend the basis $T_x f e_1, \dots, T_x f e_{d_i}$ of $E_i(f(x))$ to a basis of $T_{f(x)}M$.

$$(T_x^* f e_j^*)(T_x f e_k) = e_j^*(e_k) = \delta_{j,k} = (T_x f e_j)^*(T_x f e_k)$$

(where $\delta_{j,k}$ is the Kronecker delta function), so that

$$T_x^* f e_j^* = (T_x f e_j)^*$$

for $1 \leq j \leq d_i$, and hence

$$T_x^* f E_i^*(x) = E_i^*(f(x)).$$

Let now $v^* \in E_i^*(x)$. When proving the limit formula (2.2.7), we can without loss of generality assume that $\|v^*\| = 1$. Then for all $n \in \mathbf{Z}$ we have

$$\begin{aligned}\frac{1}{n}\log\|T_x^* f^n v^*\| &= \frac{1}{n}\log \sup_{\substack{u\in E_i(f^n(x))\\ \|u\|=1}} |v^*((T_x f^n)^{-1}u)| \\ &\le \frac{1}{n}\log \sup_{\substack{u\in E_i(f^n(x))\\ \|u\|=1}} \|(T_x f^n)^{-1}u\| \\ &= \frac{1}{n}\log \sup_{\substack{v\in E_i(x)\\ \|v\|=1}} \frac{\|(T_x f^n)^{-1}(T_x f^n v)\|}{\|T_x f^n v\|},\end{aligned}$$

which is equal to

$$-\frac{1}{n}\inf_{\substack{v\in E_i(x)\\ \|v\|=1}} \log\|T_x f^n v\|. \tag{2.2.14}$$

If $n > 0$, then (2.2.14) becomes

$$-\inf_{\substack{v\in E_i(x)\\ \|v\|=1}} \frac{1}{n}\log\|T_x f^n v\|,$$

and by uniform convergence (See Remark 2.2.10.) this has limit $-\lambda_i$ as $n \to \infty$. Similar calculation holds in the case $n < 0$. On the other hand if $v \in E_i(x)$ satisfies $\|v\| = 1$ and $|v^*(v)| = 1$, we have

$$\begin{aligned}\frac{1}{n}\log\|T_x^* f^n v^*\| &\ge \frac{1}{n}\log|v^*((T_x f^n)^{-1}\frac{T_x f^n v}{\|T_x f^n v\|}| \\ &= -\frac{1}{n}\log\|T_x f^n v\| \to -\lambda_i.\end{aligned}$$

This proves (2.2.7). ∎

2.3. Nonuniform hyperbolicity

2.3.1. *The nonuniformly hyperbolic set*

Let M be a compact Riemannian manifold and $f\colon M \to M$ a diffeomorphism. Define the set $\Lambda \subset M$ by setting

$$\begin{aligned}\Lambda = \{x \in M : \text{There exists } k(x) \in \mathbf{N} \text{ such that}\\ 1 \le k(x) < s(x) \text{ and}\\ \chi_{k(x)}(x) < 0 < \chi_{k(x)+1}(x)\},\end{aligned} \tag{2.3.1}$$

where $s(x)$ is the number of Lyapunov exponents at point x and $\chi_i(x)$, $1 \le i \le s(x)$, are the Lyapunov exponents at x. Notice that all points in Λ have both positive and negative Lyapunov exponents, which is the so called *hyperbolicity condition*. In addition, if we have defined an f-invariant probability measure on M then Λ is f-invariant, since by Oseledets' theorem (Theorem 2.2.11) the functions s, k and the Lyapunov exponents χ_i are invariant functions for almost every $x \in M$.

The content of this section follows mainly the lecture notes [BP99] by L. Barreira and Ja. Pesin, as well as the article [Pes76] by the second author.

Before looking at nonuniformly hyperbolic subsets of a compact Riemannian manifold, it's elucidating to look at the definition (uniformly) hyperbolic subsets.

Definition 2.3.1. A measurable f-invariant subset $\mathcal{R} \subset M$ is **hyperbolic** if there exist

- constants $0 < \lambda < 1$ and $c > 0$,
- for each $x \in \mathcal{R}$ subspaces $E^s(x), E^u(x) \subset T_x M$

with the following properties holding for every $x \in \mathcal{R}$:

1. The subspaces $E^s(x)$ and $E^u(x)$ depend measurably on x and satisfy
$$T_x M = E^s(x) \oplus E^u(x),$$
$$T_x f E^s(x) = E^s(f(x)), \quad T_x f E^u(x) = E^u(f(x)).$$
2. The subspace $E^s(x)$ is stable, meaning that for every $v \in E^s(x)$ and $n > 0$ we have
$$\|T_x f^n v\| \leq c\lambda^n \|v\|.$$
3. The subspace $E^u(x)$ is unstable, meaning that for every $v \in E^u(x)$ and $n > 0$ we have
$$\|T_x f^{-n} v\| \leq c\lambda^n \|v\|.$$
□

Below is the definition for a nonuniformly hyperbolic subset.

Definition 2.3.2. A measurable f-invariant subset $\mathcal{R} \subset M$ is **nonuniformly hyperbolic** if there exist

- functions $\lambda_1, \lambda_2 : \mathcal{R} \to \mathbf{R}^+$ such that $0 < \lambda_1(x) < 1 < \lambda_2(x)$ for all $x \in \mathcal{R}$,
- a real number $\varepsilon_0 > 0$ and an f-invariant function $\varepsilon : \mathcal{R} \to (0, \varepsilon_0)$ satisfying
$$\lambda_1(x) e^{2\varepsilon(x)} < 1, \quad \lambda_2(x)^{-1} e^{2\varepsilon(x)} < 1,$$
- functions $C, K : \mathcal{R} \to \mathbf{R}^+$ and
- for each $x \in \mathcal{R}$ subspaces $E^s(x), E^u(x) \subset T_x M$

with the following properties holding for every $x \in \mathcal{R}$:

(H1) The subspaces $E^s(x)$ and $E^u(x)$ depend measurably on x and satisfy
$$T_x M = E^s(x) \oplus E^u(x),$$
$$T_x f E^s(x) = E^s(f(x)), \quad T_x f E^u(x) = E^u(f(x)).$$

(H2) The subspace $E^s(x)$ is **stable**, meaning that for every $v \in E^s(f^m(x))$, $m \in \mathbf{Z}$ and $n > 0$ we have

$$\|T_{f^m(x)} f^n v\| \leq C(f^m(x))\lambda_1(x)^n e^{\varepsilon(x)n}\|v\|.$$

(H3) The subspace $E^u(x)$ is **unstable**, meaning that for every $v \in E^u(f^m(x))$, $m \in \mathbf{Z}$ and $n < 0$ we have

$$\|T_{f^m(x)} f^n v\| \leq C(f^m(x))\lambda_2(x)^n e^{\varepsilon(x)|n|}\|v\|.$$

(H4) The angle between $E^s(f^n(x))$ and $E^u(f^n(x))$ satisfies

$$\sin \sphericalangle(E^s(f^n(x)), E^u(f^n(x))) \geq K(f^n(x)) \quad \text{for all } n \in \mathbf{Z}.$$

(H5) The functions C and K are restricted by

$$C(f^n(x)) \leq C(x)e^{\varepsilon(x)|n|}, \quad K(f^n(x)) \geq K(x)e^{-\varepsilon(x)|n|},$$

for all $n \in \mathbf{Z}$. □

If we compare the two definitions, we see that in the case of a nonuniformly hyperbolic subset the rate at which the norm of $v \in E^s(f^m(x))$ goes to zero upon iteration by f depends on m. (And similarly for the unstable subspaces.) This dependence appears in the form of the term $C(f^m(x))$, which in turn is restricted by the condition (H5). The condition (H5) ensures that the rate at which the estimates get worse as m grows is at most exponential.

If $\mathcal{R}$ is nonuniformly hyperbolic, one can take it apart by defining the subsets

$$\mathcal{R}_m = \{x \in \mathcal{R} : C(x) \leq m, K(x) \geq 1/m\}, \quad m \in \mathbf{N}.$$

Then $\mathcal{R} = \bigcup_{m \in \mathbf{N}} \mathcal{R}_m$. The sets $\mathcal{R}_m$ resemble uniformly hyperbolic subsets, the difference being that they are no longer f-invariant and that the constant λ in the definition has to be replaced by an f-invariant function.

Example 2.3.3. Take as our manifold M the torus $\mathbf{R}^2/\mathbf{Z}^2$ and let $f \colon M \to M$ be defined by

$$f(x,y) = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \pmod 1.$$

This mapping is known as the Arnold's cat map. We can divide the torus into the four triangles

$$\begin{aligned}
T_1 &= \{(x,y) \in [0,1]^2 : 0 \leq x+y \leq 1 \text{ and } 0 \leq 2x+y \leq 1\} \\
T_2 &= \{(x,y) \in [0,1]^2 : 0 \leq x+y \leq 1 \text{ and } 1 \leq 2x+y \leq 2\} \\
T_3 &= \{(x,y) \in [0,1]^2 : 1 \leq x+y \leq 2 \text{ and } 1 \leq 2x+y \leq 2\} \\
T_4 &= \{(x,y) \in [0,1]^2 : 1 \leq x+y \leq 2 \text{ and } 2 \leq 2x+y \leq 3\}
\end{aligned}$$

and see that the restrictions of f to these triangles can be written as

$$\begin{aligned} f|T_1(x,y) &= (2x+y, x+y) \\ f|T_2(x,y) &= (2x+y-1, x+y) \\ f|T_3(x,y) &= (2x+y-1, x+y-1) \\ f|T_4(x,y) &= (2x+y-2, x+y-1). \end{aligned}$$

In any case we see that the tangent map is given in the local coordinates inherited from $\mathbf{R}^2$ by

$$T_{(x,y)} f v = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} v.$$

This matrix has two eigenvalues, namely $\lambda_1 = \frac{3-\sqrt{5}}{2} < 1$ and $\lambda_2 = \frac{3+\sqrt{5}}{2} = \lambda_1^{-1} > 1$. One can also check that the inverse of f is given by

$$f^{-1}(x,y) = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \pmod 1.$$

The map is hyperbolic. As $E^s(x)$ we can take the eigenspace corresponding to λ_1 and as $E^u(x)$ the eigenspace corresponding to λ_2. Because the matrix is symmetric, these eigenspaces are orthogonal and hence

$$T_x M = \mathbf{R}^2 = E^s(x) \oplus E^u(x).$$

Clearly any $v \in E^s(x)$ satisfies

$$\|T_x f^m v\| = \lambda_1^m \|v\|$$

and any $v \in E^u(x)$ satisfies

$$\|T_x f^{-m} v\| = \lambda_2^{-m} \|v\|.$$

One can now pick $\lambda_1 = \lambda_2^{-1}$ as the λ in the definition of a hyperbolic set. Or one can take λ_1 and λ_2 as the corresponding functions in the definition of the non-uniformly hyperbolic set, set $C = K = 1$ and pick a small enough constant function $\varepsilon(x) = \varepsilon > 0$ to get a non-uniformly hyperbolic structure. □

The following theorem gives us a large set like the one described above, given that “many” points in M satisfy the hyperbolicity condition.

Theorem 2.3.4. *Let M be a compact Riemannian manifold, μ a probability measure defined on its Borel sets and $f\colon M \to M$ a measure-preserving diffeomorphism. Then there exists a subset $\widetilde{\Lambda} \subset \Lambda$ with $\mu(\widetilde{\Lambda}) = \mu(\Lambda)$ that is nonuniformly hyperbolic for all $\varepsilon_0 > 0$ (see the definition above). Here Λ is the set defined at the beginning of the section by the equation* (2.3.1). □

Proof. Let $\widetilde{\Lambda}$ consist of the points in Λ which satisfy Oseledets' theorem (Theorem 2.2.11). Then $\mu(\widetilde{\Lambda}) = \mu(\Lambda)$ is automatically satisfied and if we set

$$E^s(x) = \bigoplus_{i=1}^{k(x)} E_i(x), \quad E^u(x) = \bigoplus_{i=k(x)+1}^{s(x)} E_i(x)$$

(where $E_i(x)$, $1 \le i \le s(x)$ are the subspaces of T_xM given by Theorem 2.2.11, $s(x)$ is the number of Lyapunov exponents at point x and $k(x)$ is the index of the largest negative Lyapunov exponent), then the condition (H1) follows immediately from Theorem 2.2.11. The obvious choices for $\lambda_1(x)$ and $\lambda_2(x)$ are $e^{\chi_{k(x)}}$ and $e^{\chi_{k(x)+1}}$ respectively. The rest of the proof will show that these choices actually do work.

We will first prove the following lemma that is used to construct the functions K and C.

Lemma 2.3.5. *Suppose that $X \subset M$ is a strictly f-invariant measurable subset, $M_1, M_2 \colon X \times \mathbf{R}^+ \to \mathbf{R}^+$ are measurable functions and that $A \colon X \times \mathbf{R}^+ \to \mathbf{R}^+$ is a measurable function satisfying*

$$M_1(x,\varepsilon)e^{-\varepsilon|m|} \le A(f^m(x),\varepsilon) \le M_2(x,\varepsilon)e^{\varepsilon|m|}, \tag{2.3.2}$$

for all $x \in X$, $m \in \mathbf{Z}$ and $\varepsilon > 0$. Then there exist measurable functions $B_1, B_2 \colon X \times \mathbf{R}^+ \to \mathbf{R}^+$, such that for all $x \in X$ and $\varepsilon > 0$,

$$B_1(x,\varepsilon) \le A(x,\varepsilon) \le B_2(x,\varepsilon), \tag{2.3.3}$$

and in addition for all $m \in \mathbf{Z}$,

$$\begin{aligned} B_1(x,\varepsilon)e^{-2\varepsilon|m|} &\le B_1(f^m(x),\varepsilon), \\ B_2(x,\varepsilon)e^{2\varepsilon|m|} &\ge B_2(f^m(x),\varepsilon). \end{aligned} \tag{2.3.4}$$

Proof. The condition (2.3.2) implies that

$$\frac{\log M_1(x,\varepsilon)}{|m|} - \varepsilon \le \frac{1}{|m|}\log A(f^m(x),\varepsilon) \le \frac{\log M_2(x,\varepsilon)}{|m|} + \varepsilon,$$

from which we see that for any fixed x and ε there exists $m(x,\varepsilon) \in \mathbf{N}$ such that for all $m > m(x,\varepsilon)$ we have

$$-2\varepsilon \le \frac{1}{|m|}\log A(f^m(x),\varepsilon) \le 2\varepsilon.$$

Let

$$B_1(x,\varepsilon) = \min(\{1\} \cup \{A(f^k(x),\varepsilon)e^{2\varepsilon|k|} : -m(x,\varepsilon) \le k \le m(x,\varepsilon)\}),$$
$$B_2(x,\varepsilon) = \max(\{1\} \cup \{A(f^k(x),\varepsilon)e^{-2\varepsilon|k|} : -m(x,\varepsilon) \le k \le m(x,\varepsilon)\}).$$

The functions B_1 and B_2 are clearly measurable. Moreover we have

$$B_1(x,\varepsilon)e^{-2\varepsilon|n|} \le A(f^n(x),\varepsilon) \le B_2(x,\varepsilon)e^{2\varepsilon|n|}, \quad \text{for all } n \in \mathbf{Z}. \tag{2.3.5}$$

The equation (2.3.3) follows from (2.3.5) by setting $n = 0$. In addition if $b \le 1$ is such that for all $n \in \mathbf{Z}$ we have $be^{-2\varepsilon|n|} \le A(f^n(x), \varepsilon)$ for some fixed x, then $b \le B_1(x, \varepsilon)$, and hence

$$B_1(x,\varepsilon) = \sup\{b \le 1 : be^{-2\varepsilon|n|} \le A(f^n(x),\varepsilon) \text{ for all } n \in \mathbf{Z}\}. \qquad (2.3.6)$$

In a similar manner we see that

$$B_2(x,\varepsilon) = \inf\{b \ge 1 : be^{2\varepsilon|n|} \ge A(f^n(x),\varepsilon) \text{ for all } n \in \mathbf{Z}\}. \qquad (2.3.7)$$

Now, from (2.3.5) it follows that

$$A(f^{n+m}(x),\varepsilon) \le B_2(x,\varepsilon)e^{2\varepsilon|n+m|} \le B_2(x,\varepsilon)e^{2\varepsilon|n|+2\varepsilon|m|},$$
$$A(f^{n+m}(x),\varepsilon) \ge B_1(x,\varepsilon)e^{-2\varepsilon|n+m|} \ge B_1(x,\varepsilon)e^{-2\varepsilon|n|-2\varepsilon|m|},$$

and hence for all $n \in \mathbf{Z}$,

$$B_1(x,\varepsilon)e^{-2\varepsilon|m|}e^{-2\varepsilon|n|} \le A(f^n(f^m(x)),\varepsilon).$$

Thus the equation (2.3.6) written at the point $f^m(x)$ gives us

$$B_1(x,\varepsilon)e^{-2\varepsilon|m|} \le B_1(f^m(x),\varepsilon).$$

A similar calculation for B_2 and applying (2.3.7) yields the second equation in (2.3.4). ∎

Consider the function $\gamma\colon \widetilde{\Lambda} \to [0,1]$ given by

$$\gamma(x) = \sin \sphericalangle(E^s(x), E^u(x)).$$

By Theorem 2.2.11 we know that

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

Therefore for any $\varepsilon > 0$ there exists $n(\varepsilon) \in \mathbf{N}$ such that for all $n \in \mathbf{N}$, $|n| > n(\varepsilon)$, we have

$$e^{-\varepsilon|n|} \le \gamma(f^n(x)) \le e^{\varepsilon|n|}.$$

Thus by choosing

$$M_1(x,\varepsilon) = \min_{-n(\varepsilon)\le n\le n(\varepsilon)} \frac{\gamma(f^n(x))}{e^{-\varepsilon|n|}} \quad \text{and}$$
$$M_2(x,\varepsilon) = \max_{-n(\varepsilon)\le n\le n(\varepsilon)} \frac{\gamma(f^n(x))}{e^{\varepsilon|n|}},$$

we see that $\gamma(x,\varepsilon) = \gamma(x)$ satisfies the conditions of Lemma 2.3.5, which gives us the function $B_1(x,\varepsilon)$. Picking $K(x) = B_1(x, \frac{\varepsilon}{2})$ we see that (H4) and (H5) hold for K.

We will use the next lemma to construct the function $C(x)$. Let

$$T_{x,i}f^n = T_xf^n|E_i(x), \quad T^*_{x,i}f^n = T^*_xf^n|E^*_i(x), \quad 1 \le i \le s(x).$$

Lemma 2.3.6. *There exists a measurable function* $D: \widetilde{\Lambda} \times \mathbf{R}^+ \to \mathbf{R}^+$ *such that if* $m \in \mathbf{Z}$ *and* $1 \le i \le s(x)$, *then*

$$D(f^m(x), \varepsilon) \le D(x, \varepsilon)^2 e^{2\varepsilon|m|} \tag{2.3.8}$$

and

$$\|T_{x,i} f^n\| \le D(x,\varepsilon) e^{(\chi_i+\varepsilon)n}, \quad \|T_{x,i} f^{-n}\| \ge D(x,\varepsilon)^{-1} e^{(-\chi_i-\varepsilon)n}$$

for all $n \ge 0$. □

Proof. Let $x \in \widetilde{\Lambda}$ and $\varepsilon > 0$. By Theorem 2.2.11 there exists $n(x,\varepsilon) \in \mathbf{N}$ such that for all $n > n(x,\varepsilon)$ we have the following inequalities:

$$\begin{aligned}
\chi_i(x) - \varepsilon \le \frac{1}{n} \log \|T_{x,i} f^n\| &\le \chi_i(x) + \varepsilon, \\
-\chi_i(x) - \varepsilon \le \frac{1}{n} \log \|T_{x,i} f^{-n}\| &\le -\chi_i(x) + \varepsilon, \\
-\chi_i(x) - \varepsilon \le \frac{1}{n} \log \|T^*_{x,i} f^n\| &\le -\chi_i(x) + \varepsilon, \\
\chi_i(x) - \varepsilon \le \frac{1}{n} \log \|T^*_{x,i} f^{-n}\| &\le \chi_i(x) + \varepsilon.
\end{aligned}$$

Henceforth let us denote $\chi_i(x) = \chi_i$. Set

$$\begin{aligned}
D_1^+(x,\varepsilon) &= \min_{\substack{1 \le i \le s(x) \\ 0 \le j \le n(x,\varepsilon)}} \{1, \|T_{x,i} f^j\| e^{(-\chi_i+\varepsilon)j}, \|T^*_{x,i} f^j\| e^{(\chi_i+\varepsilon)j}\}, \\
D_1^-(x,\varepsilon) &= \min_{\substack{1 \le i \le s(x) \\ -n(x,\varepsilon) \le j \le 0}} \{1, \|T_{x,i} f^j\| e^{(-\chi_i-\varepsilon)j}, \|T^*_{x,i} f^j\| e^{(\chi_i-\varepsilon)j}\}, \\
D_2^+(x,\varepsilon) &= \max_{\substack{1 \le i \le s(x) \\ 0 \le j \le n(x,\varepsilon)}} \{1, \|T_{x,i} f^j\| e^{(-\chi_i-\varepsilon)j}, \|T^*_{x,i} f^j\| e^{(\chi_i-\varepsilon)j}\}, \\
D_2^-(x,\varepsilon) &= \max_{\substack{1 \le i \le s(x) \\ -n(x,\varepsilon) \le j \le 0}} \{1, \|T_{x,i} f^j\| e^{(-\chi_i+\varepsilon)j}, \|T^*_{x,i} f^j\| e^{(\chi_i+\varepsilon)j}\},
\end{aligned}$$

and collect these together to define

$$\begin{aligned}
D_1(x,\varepsilon) &= \min\{D_1^+(x,\varepsilon), D_1^-(x,\varepsilon)\}, \\
D_2(x,\varepsilon) &= \max\{D_2^+(x,\varepsilon), D_2^-(x,\varepsilon)\}, \\
D(x,\varepsilon) &= \max\{D_1(x,\varepsilon)^{-1}, D_2(x,\varepsilon)\}.
\end{aligned}$$

We claim that $D(x,\varepsilon)$ is the function we seek. First notice that it satisfies the following inequalities for all $n \ge 0$:

$$\begin{aligned}
D(x,\varepsilon)^{-1} e^{(\chi_i-\varepsilon)n} &\le \|T_{x,i} f^n\| \le D(x,\varepsilon) e^{(\chi_i+\varepsilon)n}, \\
D(x,\varepsilon)^{-1} e^{(-\chi_i-\varepsilon)n} &\le \|T_{x,i} f^{-n}\| \le D(x,\varepsilon) e^{(-\chi_i+\varepsilon)n}, \\
D(x,\varepsilon)^{-1} e^{(-\chi_i-\varepsilon)n} &\le \|T^*_{x,i} f^n\| \le D(x,\varepsilon) e^{(-\chi_i+\varepsilon)n}, \\
D(x,\varepsilon)^{-1} e^{(\chi_i-\varepsilon)n} &\le \|T^*_{x,i} f^{-n}\| \le D(x,\varepsilon) e^{(\chi_i+\varepsilon)n}.
\end{aligned} \tag{2.3.9}$$

When $0 \le n \le n(x,\varepsilon)$, these follow straight from the construction. Outside of that range these are consequences of the limit behaviour and the fact that $D(x,\varepsilon) \ge 1$.

Moreover, one sees that

$$D(x,\varepsilon) = \inf\{d \geq 1 : d \text{ satisfies the inequalities (2.3.9) for all } n \in \mathbf{N} \text{ when it's replaced for } D(x,\varepsilon)\}. \tag{2.3.10}$$

Consider now a fixed x and pick an orthonormal basis

$$v^n = (v_1^n, \dots, v_{d_i}^n)$$

for $E_i(f^n(x))$ and let

$$w^n = (w_1^n, \dots, w_{d_i}^n)$$

be its dual basis. Let $A^i_{n,m}$ be the matrix of the map $T_{f^m(x),i} f^n$ with respect to the bases v^m and v^{n+m}, and similarly let $B^i_{n,m}$ be the matrix of the map $T^*_{f^m(x),i} f^n$ with respect to the bases w^m and w^{n+m}. Notice that for all $x \in M$, $m \in \mathbf{Z}$, $v \in T_x M$ and $\varphi \in T_x^* M$ we have

$$(T_x^* f^m \varphi)(T_x f^m v) = \varphi((T_x f^m)^{-1}(T_x f^m v)) = \varphi(v). \tag{2.3.11}$$

Hence if v' and w' are the coordinates of v and φ in the bases v^0 and w^0 respectively, then from (2.3.11) it follows that $(B^i_{m,0} w')^* (A^i_{m,0} v') = w'^* v'$, which gives us that $w'^* (B^i_{m,0})^* A^i_{m,0} v' = w'^* v'$ for all w' and v'. This implies that for all $m \in \mathbf{Z}$,

$$(B^i_{m,0})^* A^i_{m,0} = I, \tag{2.3.12}$$

where I is the identity matrix.

The chain rule gives us for all $n, m \in \mathbf{Z}$ that

$$\begin{aligned} T_{f^m(x),i} f^n &= (T_{x,i} f^{n+m})(T_{x,i} f^m)^{-1}, \\ T^*_{f^m(x),i} f^n &= (T^*_{x,i} f^{n+m})(T^*_{x,i} f^m)^{-1}, \end{aligned}$$

so that by (2.3.12) we have

$$\begin{aligned} A^i_{n,m} &= A^i_{n+m,0}(A^i_{m,0})^{-1} = A^i_{n+m,0}(B^i_{m,0})^*, \\ B^i_{n,m} &= B^i_{n+m,0}(B^i_{m,0})^{-1} = B^i_{n+m,0}(A^i_{m,0})^*. \end{aligned}$$

From this we obtain the inequalities

$$\begin{aligned} \|A^i_{n,m}\| &\leq \|A^i_{n+m,0}\| \cdot \|B^i_{m,0}\|, \\ \|A^i_{n,m}\| &\geq \frac{\|A^i_{n+m,0}\|}{\|A^i_{m,0}\|}, \\ \|B^i_{n,m}\| &\leq \|B^i_{n+m,0}\| \cdot \|A^i_{m,0}\|, \\ \|B^i_{n,m}\| &\geq \frac{\|B^i_{n+m,0}\|}{\|B^i_{m,0}\|}, \end{aligned}$$

which we can write as

$$\begin{aligned}
\|T_{f^m(x),i}f^n\| &\le \|T_{x,i}f^{n+m}\| \cdot \|T^*_{x,i}f^m\|,\\
\|T_{f^m(x),i}f^n\| &\ge \frac{\|T_{x,i}f^{n+m}\|}{\|T_{x,i}f^m\|},\\
\|T^*_{f^m(x),i}f^n\| &\le \|T^*_{x,i}f^{n+m}\| \cdot \|T_{x,i}f^m\|,\\
\|T^*_{f^m(x),i}f^n\| &\ge \frac{\|T^*_{x,i}f^{n+m}\|}{\|T^*_{x,i}f^m\|}.
\end{aligned} \tag{2.3.13}$$

Now the inequalities (2.3.9) together with (2.3.13) can be used to obtain the following estimates when $m \ge 0$:

- For $n \ge 0$:

$$D(x,\varepsilon)^{-2}e^{-2\varepsilon m}e^{(\chi_i-\varepsilon)n} \le \|T_{f^m(x),i}f^n\| \le D(x,\varepsilon)^2 e^{2\varepsilon m}e^{(\chi_i+\varepsilon)n},$$
$$D(x,\varepsilon)^{-2}e^{-2\varepsilon m}e^{(-\chi_i-\varepsilon)n} \le \|T^*_{f^m(x),i}f^n\| \le D(x,\varepsilon)^2 e^{2\varepsilon m}e^{(-\chi_i+\varepsilon)n}.$$

- For $n \ge 0$, $m-n \ge 0$:

$$\|T_{f^m(x),i}f^{-n}\| \le D(x,\varepsilon)^2 e^{2\varepsilon m}e^{(-\chi_i-\varepsilon)n} \le D(x,\varepsilon)^2 e^{2\varepsilon m}e^{(-\chi_i+\varepsilon)n},$$
$$\|T_{f^m(x),i}f^{-n}\| \ge D(x,\varepsilon)^{-2}e^{-2\varepsilon m}e^{(-\chi_i-\varepsilon)n} \ge D(x,\varepsilon)^{-2}e^{-2\varepsilon m}e^{(-\chi_i-\varepsilon)n},$$

  and similarly

$$D(x,\varepsilon)^{-2}e^{-2\varepsilon m}e^{(\chi_i-\varepsilon)n} \le \|T^*_{f^m(x),i}f^{-n}\| \le D(x,\varepsilon)^2 e^{2\varepsilon m}e^{(\chi_i+\varepsilon)n}.$$

- For $n \ge 0$, $m-n \le 0$:

$$\|T_{f^m(x),i}f^{-n}\| \le D(x,\varepsilon)^2 e^{(-\chi_i+\varepsilon)n} \le D(x,\varepsilon)^2 e^{2\varepsilon m}e^{(-\chi_i+\varepsilon)n},$$
$$\|T_{f^m(x),i}f^{-n}\| \ge D(x,\varepsilon)^{-2}e^{(-\chi_i-\varepsilon)n} \ge D(x,\varepsilon)^{-2}e^{-2\varepsilon m}e^{(-\chi_i-\varepsilon)n},$$

  and similarly

$$D(x,\varepsilon)^{-2}e^{-2\varepsilon m}e^{(\chi_i-\varepsilon)n} \le \|T^*_{f^m(x),i}f^{-n}\| \le D(x,\varepsilon)^2 e^{2\varepsilon m}e^{(\chi_i+\varepsilon)n}.$$

From these inequalities and (2.3.10) one obtains for $m \ge 0$ that

$$D(f^m(x),\varepsilon) \le D(x,\varepsilon)^2 e^{2\varepsilon m}. \tag{2.3.14}$$

Now everything that was done above can be done for f^{-1} in place of f. One readily checks that the obtained function D will be the same. Hence for all $m \ge 0$ we have also

$$D(f^{-m}(x),\varepsilon) \le D(x,\varepsilon)^2 e^{2\varepsilon m}, \tag{2.3.15}$$

and by combining the equations (2.3.14) and (2.3.15) one obtains the equation (2.3.8) and the lemma is proved. ∎

Let's move back to the construction of C. If we replace m by $-m$ and x by $f^m(x)$ in (2.3.8), we get

$$D(x,\varepsilon) \le D(f^m(x),\varepsilon)^2 e^{2\varepsilon|m|},$$

which gives us the inequality

$$D(f^m(x),\varepsilon) \ge \sqrt{D(x,\varepsilon)}e^{-\varepsilon|m|}. \tag{2.3.16}$$

Let now I_1 and I_2 be two disjoint subsets of $\{1,2,\ldots,s(x)\}$, and set

$$L_1(x) = \bigoplus_{i\in I_1} E_i(x), \quad L_2(x) = \bigoplus_{i\in I_2} E_i(x)$$

for all $x \in \widetilde{\Lambda}$.

Exactly as we constructed the function $K(x)$ above, we can construct a function $K_{I_1,I_2}(x)$ that satisfies

$$\sin \sphericalangle (L_1(x), L_2(x)) \ge K_{I_1,I_2}(x), \quad K_{I_1,I_2}(f^m(x)) \ge K_{I_1,I_2}(x)e^{-\varepsilon|m|}.$$

Set now

$$T(x) = \min_{\substack{I_1,I_2\subset\{1,2,\ldots,s(x)\}\\ I_1\cap I_2=\varnothing}} K_{I_1,I_2}(x)$$

and we see that T satisfies the condition

$$T(f^m(x)) \ge T(x)e^{-\varepsilon|m|}$$

as well. Let $v \in E^s(x)$ and write it as $v = \sum_{i=1}^{k(x)} v_i$, where $v_i \in E_i(x)$. Then since

$$v \mapsto \sum_{i=1}^{k(x)} \|v_i\|$$

is a norm, by the equivalence of norms and since $T(x) \le 1$, we can find a constant $L > 1$ such that

$$\|v\| \le \sum_{i=1}^{k(x)} \|v_i\| \le LT^{-1}(x)\|v\|.$$

Let $C'(x) = LD(x,\varepsilon)T^{-1}(x)$ and consider the functions M_1 and M_2 given by

$$\begin{aligned} M_1(x,\varepsilon) &= L\sqrt{D(x,\varepsilon)}, \\ M_2(x,\varepsilon) &= LD(x,\varepsilon)^2 T(x,\varepsilon)^{-1}. \end{aligned}$$

Then by (2.3.16) (and since $T(f^m(x)) \le 1$) we have

$$\begin{aligned} M_1(x,\varepsilon)e^{-\varepsilon|m|} &= L\sqrt{D(x,\varepsilon)}e^{-\varepsilon|m|} \\ &\le T^{-1}(f^m(x))LD(f^m(x),\varepsilon) \\ &= C'(f^m(x),\varepsilon), \end{aligned}$$

and (2.3.8) gives us the upper bound

$$\begin{aligned} C'(f^m(x),\varepsilon) &= T^{-1}(f^m(x))LD(f^m(x),\varepsilon) \\ &\le T^{-1}(x)e^{-\varepsilon|m|}LD^2(x,\varepsilon)e^{2\varepsilon|m|} \\ &= M_2(x,\varepsilon)e^{\varepsilon|m|}. \end{aligned}$$

Thus C' satisfies the condition (2.3.2) of Lemma 2.3.5. Hence there exists a function $C_1(x,\varepsilon)$ which satisfies the equations

$$\begin{aligned} C_1(x,\varepsilon) \ge C'(x,\varepsilon) \quad &\text{and} \\ C_1(x,\varepsilon)e^{2\varepsilon|m|} \ge C_1(f^m(x),\varepsilon)&. \end{aligned}$$

The first equation and Lemma 2.3.6 give us

$$C_1(x,\varepsilon) \ge T^{-1}(f^m(x))LD(f^m(x),\varepsilon) \ge \|T_{f^m(x),i}f^n\|e^{(-\chi_i-\varepsilon)n},$$

which implies that (H2) holds with the function C_1.

Finally, by going through the above construction with $E^s(x)$ and f replaced by $E^u(x)$ and f^{-1}, we get a function C_2 for which (H3) holds. Setting

$$C(x,\varepsilon) = \max\{C_1(x,\varepsilon), C_2(x,\varepsilon)\}$$

gives us the required function.

Finally notice that the existence of the function $\varepsilon(x)$ follows from what was done above. Indeed, we may set

$$\varepsilon(x) = \min(-\frac{1}{2}\log(\max(\lambda_1(x),\lambda_2(x)^{-1})),\varepsilon_0/2)$$

to get the wanted invariant function. The function C is then given by

$$C(x) = C(x,\varepsilon(x)).$$

■

2.3.2. *Local stable and unstable manifolds*

Let us first recall the following definition.

Definition 2.3.7. A function $f\colon \mathbf{R}^n \to \mathbf{R}^n$ is of class $C^{1+\alpha}$, $0<\alpha\le 1$, if its partial derivatives $\frac{\partial f_i}{\partial x_j}$, $1\le i,j\le n$, are Hölder continuous, i.e. they satisfy

$$\|\frac{\partial f_i}{\partial x_j}(x) - \frac{\partial f_i}{\partial x_j}(y)\| \le C\|x-y\|^\alpha$$

for some constant $C>0$. This is the same as requiring

$$\|Df(x) - Df(y)\| \le C\|x-y\|^\alpha$$

for some constant $C>0$. Here $Df(x)$ is the Jacobian matrix of f at point $x\in\mathbf{R}^n$. □

In this section we'll construct for each $x \in \mathcal{R}$ local stable and unstable manifolds $V^s(x)$ and $V^u(x)$. These manifolds have the property that the trajectories of points in $V^s(x)$ will approach the trajectory of x exponentially. The same holds for $V^u(x)$ when time is reversed. For these constructions to be possible, we will have to require that f is of class $C^{1+\alpha}$ for some $0 < \alpha \leq 1$. By this we mean that there exists a constant $C > 0$ such that for all $x \in M$ there exist charts (U_x, ϕ_x), (V_x, ψ_x) with $x \in U$, $x \in V$ and $f(U_x) \subset V_{f(x)}$ such that the map

By "have to require" we mean that there are counterexamples in the case where the condition $f \in C^{1+\alpha}$ is not satisfied and the map is just C^1. See [Pug84].

$$\tilde{f}_x = \psi \circ f \circ \phi^{-1}$$

is $C^{1+\alpha}$ with Hölder constant C. We note that since the manifold is compact, it is enough to have some constants C_x for each such chart. One can then take an open subcover of these neighbourhoods and choose the largest C_x as C. Moreover we note that if ϕ'_x are coordinates in some neighbourhood of x and $\psi'_{f(x)}$ are coordinates in some neighbourhood of $f(x)$, then

$$\psi' \circ f \circ \phi'^{-1} = \psi' \circ \psi^{-1} \circ \psi \circ f \circ \phi^{-1} \circ \phi \circ \phi'^{-1}$$

and since $\psi' \circ \psi^{-1}$ and $\phi \circ \phi'^{-1}$ are C^∞-mappings, we know that

$$\psi' \circ f \circ \phi'^{-1}$$

is $C^{1+\alpha}$ in some neighbourhood of $\phi'(x)$. (Take a neighbourhood A of $\phi'(x)$ such that $\overline{A}$ is compact and is a subset of the domain of definition of ϕ'^{-1}.) Hence if f is of class $C^{1+\alpha}$ for some coordinates, it's $C^{1+\alpha}$ for other coordinates as well, although we might have to shrink the charts a little bit.

To begin, let $\mathcal{R} \subset M$ be a nonuniformly hyperbolic set and introduce the notation used in Definition 2.3.2. We'll first define a new inner product on T_xM for all $x \in \mathcal{R}$, which has several nice properties such as the orthogonality of the stable and unstable subspaces of T_xM, $x \in \mathcal{R}$.

Definition 2.3.8. Let $x \in \mathcal{R}$. Given invariant functions $0 < \lambda'_1(x) < \lambda'_2(x) < \infty$ that satisfy $\lambda'_1(x) > \lambda_1(x)e^{\varepsilon(x)}$ and $\lambda'_2(x) < \lambda_2(x)e^{-\varepsilon(x)}$, we define a **Lyapunov inner product** on T_xM by setting

$$\langle v, w \rangle_x^L = \sum_{k=0}^{\infty} \left\langle T_x f^k v, T_x f^k w \right\rangle_{f^k(x)} \lambda_1'^{-2k}(x),$$

if $v, w \in E^s(x)$,

$$\langle v, w \rangle_x^L = \sum_{k=0}^{\infty} \left\langle T_x f^{-k} v, T_x f^{-k} w \right\rangle_{f^{-k}(x)} \lambda_2'^{2k}(x),$$

if $v, w \in E^u(x)$ and

$$\langle v, w \rangle_x^L = 0,$$

if $v \in E^s(x)$ and $w \in E^u(x)$. Finally we extend it to the whole of T_xM by linearity. (Remember, that we have $T_xM = E^s(x) \oplus E^u(x)$.) □

Remark 2.3.9. *The above definition is well-posed, i.e. both of the series converge. Indeed, by Cauchy-Schwarz inequality and (H2) we have*

$$\sum_{k=0}^{\infty} \left\langle T_x f^k v, T_x f^k w \right\rangle_{f^k(x)} \lambda_1'^{-2k}(x) \leq$$
$$\sum_{k=0}^{\infty} C(x)^2 \lambda_1^{2k}(x) \lambda_1'^{-2k}(x) e^{2\varepsilon(x)k} \|v\| \|w\| \leq$$
$$C(x)^2 (1 - \lambda_1^2(x) \lambda_1'^{-2}(x) e^{2\varepsilon(x)})^{-1} \|v\| \|w\| < \infty$$

for all $v, w \in E^s(x)$, and similarly by Cauchy-Schwarz and (H3) we have

$$\sum_{k=0}^{\infty} \left\langle T_x f^{-k} v, T_x f^{-k} w \right\rangle_{f^k(x)} \lambda_2'^{2k}(x) \leq$$
$$\sum_{k=0}^{\infty} C(x)^2 \lambda_2^{-2k}(x) \lambda_2'^{2k}(x) e^{2\varepsilon(x)k} \|v\| \|w\| \leq$$
$$C(x)^2 (1 - \lambda_2^{-2}(x) \lambda_2'^{2}(x) e^{2\varepsilon(x)})^{-1} \|v\| \|w\| < \infty$$

for all $v, w \in E^u(x)$.

We also note that the Lyapunov inner product so defined depends on the choice of the numbers $\lambda_1'(x)$ and $\lambda_2'(x)$. □

Before proving some further properties of the Lyapunov inner product, we'll have to introduce the following map $\tilde{f}_x$, which can be thought of as a small perturbation of the tangent map $T_x f$. For fixed $x \in \mathcal{R}$ we define $\tilde{f}_x$ by

$$\tilde{f}_x = \exp_{f(x)}^{-1} \circ f \circ \exp_x : B^s(r_0) \times B^u(r_0) \to T_{f(x)} M,$$

where $B^s(r_0) = B^s(0, r_0)$ and $B^u(r_0) = B^u(0, r_0)$ are balls of radius $r_0 > 0$ in $E^s(x)$ and $E^u(x)$ respectively centered at 0. Here $\exp_x : T_x M \to M$ is the *exponential map*, which takes a tangent vector $v \in T_x M$, constructs a unit-speed geodesic $\gamma: I \to M$ with $\gamma'(0) = \frac{v}{\|v\|}$ and returns the point $\gamma(\|v\|)$.

Remark 2.3.10. *$\tilde{f}_x$ is well-defined and C^1 if r_0 is sufficiently small.* □

Proof. It is known that the map $\exp_x^{-1}$ is well-defined in some open ball of radius s around x. Moreover, it is known that the maximal radius s for such balls (the so called *injectivity radius*) varies continuously with respect to x when our Riemannian manifold is compact. Thus it attains its lowest value s_0 at some point.

Now we see that the map $\exp_{f(x)}^{-1}$ exists in a ball of radius s_0 around $f(x)$. Consider a tangent vector $v \in T_x M$ and the curve $\exp_x(tv)$, $0 \leq t \leq 1$. The length of this curve is $\|v\|$, and because M is compact its image under f has length at most $D\|v\|$, where $D > 0$ is a constant depending on the norm of Tf. Therefore the distance between $f(x)$ and $f(\exp_x(v))$ is at most $D\|v\|$. By imposing the restriction $v \in B^s(r_0) \times B^u(r_0)$, where r_0 is chosen so small that necessarily $D\|v\| < s_0$, we see that $\exp_{f(x)}^{-1} \circ f \circ \exp_x$ is well-defined. ■

A restriction of the exponential map $\exp_x: T_x M \to M$ gives us a one-to-one correspondence between a small neighbourhood of 0 in the tangent space and a small neighbourhood of the point x on the manifold.

Now for each $x \in \mathcal{R}$ identify the tangent space $T_xM = E^s(x) \oplus E^u(x)$ with $\mathbf{R}^k \times \mathbf{R}^{d-k}$ – where d is the dimension of M and k is the dimension of $E^s(x)$ – by a linear isomorphism τ_x that satisfies $\tau_x(E^s(x)) = \mathbf{R}^k$ and $\tau_x(E^u(x)) = \mathbf{R}^{d-k}$. Moreover we can assume that for each x we have $\tau_x(B^s(r_0) \times B^u(r_0)) = U$ for some fixed neighbourhood U of 0 in $\mathbf{R}^d$. The Lyapunov inner product naturally carries over τ_m to $\mathbf{R}^d$, and by abuse of notation we will also denote the induced inner product and norm in $\mathbf{R}^d$ by $\langle \cdot, \cdot \rangle^L$ and $\| \cdot \|^L$ respectively. Finally let $\widetilde{F}_x = \tau_{f(x)} \circ \tilde{f}_x \circ \tau_x^{-1}$.

Lemma 2.3.11. *The mapping* $\widetilde{F}_x \colon \mathbf{R}^d \to \mathbf{R}^d$ *may be written in the form*

$$\widetilde{F}_x(v, w) = (A_x v + g_x(v, w), B_x w + h_x(v, w))$$

where A_x *and* B_x *are linear maps and the maps* $g_x \colon U \to \mathbf{R}^k$ *and* $h_x \colon U \to \mathbf{R}^{d-k}$ *satisfy*

$$g_x(0) = h_x(0) = Dg_x(0) = Dh_x(0) = 0. \qquad \square$$

Proof. By applying the chain rule and using the fact that $T_0 \exp_x$ is the identity function we obtain

$$T_0 \tilde{f}_x = T_0 \exp_{f(x)}^{-1} \circ T_x f \circ T_0 \exp_x = T_x f,$$

so that by Taylor expansion we get

$$\begin{aligned} \tilde{f}_x(v, w) &= T_x f(v, w) + F_x(v, w) \\ &= \tilde{A}_x v + \tilde{B}_x w + \tilde{g}_x(v, w) + \tilde{h}_x(v, w), \end{aligned}$$

where we have used the decompositions $T_x f(v, w) = \tilde{A}_x v + \tilde{B}_x w$ and $F_x = \tilde{g}_x + \tilde{h}_x$, where $\tilde{A}_x \colon E^s(x) \to E^s(f(x))$ and $\tilde{B}_x \colon E^u(x) \to E^u(f(x))$ are linear mappings and $\tilde{g}_x, \tilde{h}_x \colon B^s(r_0) \times B^u(r_0) \to T_{f(x)}M$ satisfy

$$\tilde{g}_x(0) = \tilde{h}_x(0) = 0.$$

We also note that from the expansion it follows that

$$T_{(v,w)} F_x = T_{(v,w)} \tilde{f}_x - T_{(v,w)}(T_x f) = T_{(v,w)} \tilde{f}_x - T_x f$$

and plugging in $(v, w) = 0$ gives us

$$T_0 \tilde{g}_x = T_0 \tilde{h}_x = 0.$$

Now we can let $A_x = \tau_{f(x)} \circ \tilde{A} \circ \tau_x^{-1}$ and similarly for B_x, g_x and h_x. ∎

We are now able to prove the following crucial properties of the Lyapunov inner product.

Theorem 2.3.12. *Let* $\| \cdot \|_x^L$ *be the norm associated to the Lyapunov inner product* $\langle \cdot, \cdot \rangle_x^L$. *Then* $\langle \cdot, \cdot \rangle_x^L$ *and* $\| \cdot \|_x^L$ *have the following properties.*

(L1) The spaces $E^s(x)$ and $E^u(x)$ are orthogonal under $\langle\cdot,\cdot\rangle_x^L$ for all $x \in \mathcal{R}$.

(L2) The maps A_x and B_x given by Lemma 2.3.11 are bounded by $\lambda_1'(x)$ and $\lambda_2'(x)$, i.e. we have

$$\|A_x\|^L \leq \lambda_1'(x), \quad \|B_x^{-1}\|^L \leq \lambda_2'^{-1}(x).$$

(L3) The norms induced by the Riemannian inner product and the Lyapunov inner product satisfy

$$\frac{1}{\sqrt{2}}\|v\|_x \leq \|v\|_x^L \leq D(x)\|v\|_x,$$

where $v \in T_x\mathcal{R}$ and

$$D(x) = C(x)K(x)^{-1}\sqrt{(1 - \frac{\lambda_1(x)e^{\varepsilon(x)}}{\lambda_1'(x)})^{-1} + (1 - \frac{\lambda_2'(x)}{\lambda_2(x)e^{-\varepsilon(x)}})^{-1}}.$$

Moreover, the mapping $D\colon \mathcal{R} \to \mathbf{R}^+$ satisfies

$$D(f^m(x)) \leq D(x)e^{2\varepsilon(x)|m|}, \quad \text{for all } m \in \mathbf{Z}.$$

□

Proof. First we note that (L1) is clear from the definition. To prove (L2), let $v \in E^s(x)$. Then $T_x f v \in E^s(f(x))$ and we obtain

$$\begin{aligned}\|T_x f v\|^L &= \sqrt{\sum_{k=0}^{\infty} \langle T_x f^{k+1} v, T_x f^{k+1} v\rangle_{f^{k+1}(x)}\, \lambda_1'^{-2k}(x)} \\ &= \sqrt{(\|v\|_x^L)^2 \lambda_1'^2(x) - \|v\|_x^2 \lambda_1'^2(x)} \leq \|v\|_x^L \lambda_1'(x).\end{aligned}$$

Hence also

$$\|A\tau_x(v)\|^L = \|\tau_{f(x)}(T_x f v)\|^L = \|T_x f v\|^L \leq \|v\|_x^L \lambda_1'(x) = \|\tau_x(v)\|^L \lambda_1'(x).$$

The corresponding calculation for $v \in E^u(x)$ shows the claim for B_x. Finally let $v \in T_x\mathcal{R}$ and write $v = v^s + v^u$, where $v^s \in E^s(x)$ and $v^u \in E^u(x)$. Then by (L1) and the definition of the Lyapunov inner product we have

$$\begin{aligned}(\|v\|_x^L)^2 &= \langle v^s, v^s\rangle_x^L + \langle v^u, v^u\rangle_x^L \\ &= \sum_{k=0}^{\infty} \|T_x f^k v^s\|_{f^k(x)}^2 \lambda_1'^{-2k}(x) + \sum_{k=0}^{\infty} \|T_x f^{-k} v^u\|_{f^{-k}(x)} \lambda_2'^{2k}(x).\end{aligned} \tag{2.3.17}$$

From (2.3.17) we see that

$$\begin{aligned}2(\|v\|_x^L)^2 &\geq 2\|v^s\|_x^2 + 2\|v^u\|_x^2 \\ &= \|v^s\|_x^2 + 2\langle v^s, v^u\rangle_x + \|v^u\|_x^2 + \|v^s\|_x^2 - 2\langle v^s, v^u\rangle_x + \|v^u\|_x^2 \\ &= \|v^s + v^u\|_x^2 + \|v^s - v^u\|^2 \geq \|v^s + v^u\|^2,\end{aligned}$$

which yields the first inequality in (L3). For the second one the calculations in Remark 2.3.9, (H4) and B.12 give us

$$
\begin{aligned}
&K(x)^2(\|v\|_x^L)^2 \le K(x)^2C(x)^2[(1-\lambda_1^2(x)\lambda_1'^{-2}(x)e^{2\varepsilon(x)})^{-1}\|v^s\|^2+\\
&\quad(1-\lambda_2^{-2}(x)\lambda_2'^2(x)e^{2\varepsilon(x)})^{-1}\|v^u\|^2]\\
&\le C(x)^2[(1-\lambda_1^2(x)\lambda_1'^{-2}(x)e^{2\varepsilon(x)})^{-1}\|v^s\|^2\sin^2(\sphericalangle(v^s,v^u))+\\
&\quad(1-\lambda_2^{-2}(x)\lambda_2'^2(x)e^{2\varepsilon(x)})^{-1}\|v^u\|^2\sin^2(\sphericalangle(v^s,v^u))]\\
&\le C(x)^2\|v\|^2[(1-\lambda_1^2(x)\lambda_1'^{-2}(x)e^{2\varepsilon(x)})^{-1}+(1-\lambda_2^{-2}(x)\lambda_2'^2(x)e^{2\varepsilon(x)})^{-1}]\\
&\le C(x)^2\|v\|^2[(1-\lambda_1(x)\lambda_1'^{-1}(x)e^{\varepsilon(x)})^{-1}+(1-\lambda_2^{-1}(x)\lambda_2'(x)e^{\varepsilon(x)})^{-1}].
\end{aligned}
$$

Finally the inequality

$$D(f^m(x))\le D(x)e^{2\varepsilon(x)|m|}$$

is an immediate corollary of (H5) and noticing that

$$\sqrt{(1-\lambda_1(x)e^{\varepsilon(x)}/\lambda_1'(x))^{-1}+(1-\lambda_2'(x)/(\lambda_2(x)e^{-\varepsilon(x)}))^{-1}}$$

is constant on the orbit of x. ∎

We will now fix $x\in\mathcal{R}$ and denote $\widetilde{F}_m=\widetilde{F}_{f^m(x)}$ and $\tau_m=\tau_{f^m(x)}$. By Lemma 2.3.11 we may write $\widetilde{F}_m$ in the form

$$\widetilde{F}_m(v,w)=(A_mv+g_m(v,w),B_mw+h_m(v,w)),$$

where A_m and B_m are linear maps and the maps g_m and h_m are defined on the fixed neighbourhood U chosen above.

Since from now on we'll only be interested on the points on the orbit of x, we will drop the suffix (x) from the invariant functions ε, λ_1, λ_2, λ_1' and λ_2'. We also let $\gamma=e^{-2\varepsilon}$ so that $\lambda_1'^{\alpha}<\gamma<1$. Then we have the following lemma.

Lemma 2.3.13. *The maps g_m and h_m satisfy*

$$g_m(0,0)=h_m(0,0)=0,\quad Dg_m(0,0)=Dh_m(0,0)=0,$$

and

$$
\begin{aligned}
&\|Dg_m(v_1,w_1)-Dg_m(v_2,w_2)\|^L\le\\
&\quad B\gamma^{-m}(\|v_1-v_2\|^L+\|w_1-w_2\|^L)^{\alpha}\\
&\|Dh_m(v_1,w_1)-Dh_m(v_2,w_2)\|^L\le\\
&\quad B\gamma^{-m}(\|v_1-v_2\|^L+\|w_1-w_2\|^L)^{\alpha},
\end{aligned}
\tag{2.3.18}
$$

where $B=B(x)=D(x)C_12^{\frac{\alpha}{2}}>0$ for some constant $C_1>0$. □

Proof. We'll derive these inequalities from the fact that f is of class $C^{1+\alpha}$. The equalities were shown in Lemma 2.3.11. Let $F_m=g_m+h_m$ and notice that

$$DF_m(v,w)=D\widetilde{F}_m(v,w)-D(D\widetilde{F}_m(0))(v,w)=D\widetilde{F}_m(v,w)-D\widetilde{F}_m(0).$$

From here it follows by Hölder continuity and (L3) that

$$
\begin{aligned}
&\|DF_m(v_1,w_1) - DF_m(v_2,w_2)\|^L = \|D\widetilde{F}_m(v_1,w_1) - D\widetilde{F}_m(v_2,w_2)\|^L \\
&= \|\tau_{m+1}(T_{\tau_m^{-1}(v_1,w_1)}\tilde{f}_{f^m(x)} - T_{\tau_m^{-1}(v_2,w_2)}\tilde{f}_{f^m(x)})\|^L \\
&= \|T_{\tau_m^{-1}(v_1,w_1)}\tilde{f}_{f^m(x)} - T_{\tau_m^{-1}(v_2,w_2)}\tilde{f}_{f^m(x)}\|^L \\
&\le D(f^m(x))\|T_{\tau_m^{-1}(v_1,w_1)}\tilde{f}_{f^m(x)} - T_{\tau_m^{-1}(v_2,w_2)}\tilde{f}_{f^m(x)}\| \\
&\le D(x)e^{2\varepsilon(x)|m|}C_1\|\tau_m^{-1}(v_1,w_1) - \tau_m^{-1}(v_2,w_2)\|^\alpha \\
&\le D(x)e^{2\varepsilon(x)|m|}2^{\frac{\alpha}{2}}C_1\|(v_1-v_2,w_1-w_2)\|^L,
\end{aligned}
$$

where $C_1 > 0$ is a constant. The claim follows since

$$
\|Dg_m(v_1,w_1) - Dg_m(v_2,w_2)\|^L \le \|DF_m(v_1,w_1) - DF_m(v_2,w_2)\|^L
$$

and similarly for h_m. ∎

We are now ready to prove the following Local Stable Manifold Theorem. It states, roughly, that for any $x \in \mathcal{R}$ there exists a certain local submanifold $V^s(x)$, so called **local stable manifold**, the points of which get closer to each other upon iteration by f in an exponential rate. It is a central theorem in Pesin theory and as such one of the main theorems of this thesis. The local stable manifolds for the inverse mapping f^{-1} are called local unstable manifolds for the mapping f.

Theorem 2.3.14. *Suppose that $\mathcal{R}$ is a nonuniformly hyperbolic set for a diffeomorphism $f \in C^{1+\alpha}$. For every $x \in \mathcal{R}$ there exists a local stable manifold $V^s(x) \subset M$ that satisfies*

$$
x \in V^s(x), \quad T_xV^s(x) = E^s(x),
$$

and for all $y \in V^s(x)$ we have

$$
\rho(f^n(x), f^n(y)) \to 0 \quad \text{exponentially}. \tag{2.3.19}
$$

Moreover, let $\kappa \in \mathbf{R}$ satisfy $\lambda_1' < \kappa < \min(\lambda_2', \gamma^{\frac{1}{\alpha}})$. There exists a function $\psi^s \colon B^s(r) \to \mathbf{R}^{d-k}$ and numbers r and D_0 such that

$$
V^s(x) = \exp_x(\tau_0^{-1}\{(v,\psi^s(v)) : v \in B^s(r)\}),
$$

which means that $V^s(x)$ is given as the graph of the function ψ^s, and the following claims are satisfied:

1. *$\psi^s \in C^{1+\alpha}$, $\psi^s(0) = 0$ and $D\psi^s(0) = 0$.* **Note:** *This tells about the smoothness of $V^s(x)$ and also rephrases the earlier statement that $x \in V^s(x)$ and $T_xV^s(x) = E^s(x)$.*
2. *$\|D\psi^s(v) - D\psi^s(w)\|^L \le D_0(\|v-w\|^L)^\alpha$ for all $v, w \in B^s(r)$.*
3. *If $m \ge 0$ and $v \in B^s(r)$ then*

$$
\left(\prod_{i=0}^{m-1}\widetilde{F}_i\right)(v,\psi^s(v)) \in B^s(r) \times B^u(r)
$$

and

$$\left\| \left(\prod_{i=0}^{m-1} \widetilde{F}_i \right) (v, \psi^s(v)) \right\|^L \leq D_0 \kappa^m \|(v, \psi^s(v))\|^L,$$

where $\prod_{i=0}^{m-1} \widetilde{F}_i = \widetilde{F}_{m-1} \circ \widetilde{F}_{m-2} \circ \cdots \circ \widetilde{F}_0$*. (One should notice that* $\prod_{i=0}^{m-1} \widetilde{F}_{m-1} = \tau_m \circ \exp^{-1}_{f^m(x)} \circ f^m \circ \exp_x \circ \tau_0^{-1}$*.)* **Note:** *This claim encompasses the fact that for points in the stable manifold the corresponding vectors in the tangent space approach each other upon iteration.*

4. *Fix* $v \in B^s(r)$ *and* $w \in B^u(r)$*. If there exists* $K > 0$ *such that*

$$\left(\prod_{i=0}^{m-1} \widetilde{F}_i \right) (v, w) \in B^s(r) \times B^u(r)$$

and

$$\left\| \left(\prod_{i=0}^{m-1} \widetilde{F}_i \right) (v, w) \right\|^L \leq K \kappa^m$$

for every $m \geq 0$*, then* $\psi^s(v) = w$*.* **Note:** *This claim is a uniqueness result. It says that if some point* $y = \exp_x(\tau_0^{-1}(v, w))$ *has the property that* $\rho(f^m(x), f^m(y)) \to 0$ *fast enough as* $m \to \infty$*, it must belong to the stable manifold.*

5. *The numbers* D_0 *and* r *depend only on* $\lambda_1', \lambda_2', \gamma, \alpha, \kappa$ *and* B*.* □

Proof. We will first show that the first part of the theorem, i.e. the existence of $V^s(x)$ satisfying the equation (2.3.19), follows from the second part. Indeed, define

$$V^s(x) = \exp_x\{(v, \psi^s(v)) : v \in B^s(x)\},$$

where $B^s(x) = \tau_0^{-1} B^s(r)$. Then since $\psi^s(0) = 0$ we have $x \in V^s(x)$, and because $D\psi^s(0) = 0$ we have $T_x V^s(x) = E^s(x)$. Now if $y \in V^s(x)$, then $y = \exp_x(\tau_0^{-1}(v, \psi^s(v)))$ for some $v \in B^s(x)$. Hence by claim 3. we have

$$\begin{aligned} \|\exp^{-1}_{f^m(x)}(f^m(y))\|^L &= \|(\tau_m \circ \exp^{-1}_{f^m(x)} \circ f^m \circ \exp_x \circ \tau_0^{-1})(v, \psi^s(v))\|^L \\ &\leq D_0 \kappa^m \|(v, \psi^s(v))\|^L. \end{aligned}$$

Therefore the vectors representing $f^m(y)$ in the one-to-one correspondence between $T_{f^m(x)}$ and the small neighbourhood of $f^m(x)$ go to zero exponentially (since $\kappa < 1$). Because the function $\exp_{f^m(x)}$ is Lipschitz, this means that also

$$\rho(f^m(x), f^m(y)) \to 0 \quad \text{exponentially.}$$

We'll now proceed with the proof of the rest of the theorem. Consider the vector space of sequences of vectors

$$\Gamma_\kappa = \{(z(m))_{m=0}^\infty \in \mathbf{R}^d : \sup_{m \geq 0} (\kappa^{-m} \|z(m)\|^L) < \infty\}.$$

The formula

$$\|z\|_\kappa = \sup_{m\geq 0} (\kappa^{-m} \|z(m)\|^L)$$

defines a norm on Γ_κ.

Lemma 2.3.15. *Γ_κ is a Banach space under the norm $\|\cdot\|_\kappa$.* □

Proof. We'll prove the completeness of Γ_κ. Let $z_n \in \Gamma_\kappa$, $n \geq 0$, be Cauchy. Then for every $\varepsilon > 0$ there exists $M \in \mathbf{Z}$ such that for all $n, N \geq M$ we have $\|z_n - z_N\|_\kappa < \varepsilon$. Therefore for any fixed $m \geq 0$ we have

$$\kappa^{-m} \|z_n(m) - z_N(m)\|^L < \varepsilon.$$

This suffices to show that $z_n(m)$, $n \geq 0$, is Cauchy for any fixed m and hence converges to some vector $z(m) \in \mathbf{R}^d$. Now we will show that $z = (z(m))_{m=1}^\infty \in \Gamma_\kappa$ is the limit for the original sequence z_n, $n \geq 0$. For all $m \geq 0$ and $N \geq 0$ we have

$$\kappa^{-m} \|z_n(m) - z(m)\|^L \leq \kappa^{-m} \|z_n(m) - z_N(m)\|^L + \kappa^{-m} \|z_N(m) - z(m)\|^L.$$

For large enough n, N the Cauchy condition tells us that $\kappa^{-m}\|z_n(m) - z_N(m)\|^L < \frac{\varepsilon}{2}$, and by picking still larger N if needed, we can assume $\kappa^{-m}\|z_N(m) - z(m)\|^L < \frac{\varepsilon}{2}$. Therefore

$$\kappa^{-m} \|z_n(m) - z(m)\|^L < \varepsilon$$

for large enough n. Since this holds for all m, also $\|z_n - z\|_\kappa < \varepsilon$, which proves that the sequence converges. ■

One easily sees that the set

$$W = \{z \in \Gamma_\kappa : z(m) \in B^s(r_0) \times B^u(r_0) \text{ for all } m \geq 0\}$$

is open. We would like to define the mapping $\Phi_\kappa \colon B^s(r_0) \times W \to \Gamma_\kappa$ by

$$\Phi_\kappa(y, z)(m) = -z(m) + \left(\left(\prod_{i=0}^{m-1} A_i \right) y, 0 \right) \tag{2.3.20}$$
$$+ \left(\sum_{n=0}^{m-1} \left(\prod_{i=n+1}^{m-1} A_i \right) g_n(z(n)), - \sum_{n=0}^{\infty} \left(\prod_{i=0}^{n} B_{i+m} \right)^{-1} h_{n+m}(z(n+m)) \right),$$

for $m \geq 0$. (Here we assume the usual conventions for the empty product and sum and the product is formed from right to left, that is $\prod_{i=0}^{m-1} A_i = A_{m-1}A_{m-2}\ldots A_0$.) This definition might seem somewhat arbitrary, but it works and gives us the right function after we apply the implicit function theorem to it below. We have to first show that the map is well-defined. Notice that for all $z \in B^s(r_0) \times B^u(r_0)$ and $n \geq 0$ we have by the mean value theorem and Lemma 2.3.13 that

$$\begin{aligned} \|g_n(z)\|^L &= \|g_n(z) - g_n(0)\|^L \leq \|Dg_n(\xi)\|^L \cdot \|z\|^L \\ &= \|Dg_n(\xi) - Dg_n(0)\|^L \cdot \|z\|^L \leq B\gamma^{-n} (\|\xi\|^L)^\alpha \|z\|^L \\ &\leq B\gamma^{-n} (\|z\|^L)^{1+\alpha}, \end{aligned}$$

where ξ lies on the line segment between 0 and z. A similar calculation yields

$$\|h_n(z)\|^L \leq B\gamma^{-n}(\|z\|^L)^{1+\alpha}$$

for all $n \geq 0$. Therefore we have

$$\begin{aligned}\|\Phi_\kappa(y,z)\|_\kappa =& \sup_{m\geq 0}(\kappa^{-m}\|\Phi_\kappa(y,z)(m)\|^L)\\ \leq& \sup_{m\geq 0}(\kappa^{-m}\|z(m)\|^L) + \sup_{m\geq 0}\kappa^{-m}\left\|\left(\left(\prod_{i=0}^{m-1}A_i\right)y,0\right)\right\|^L +\\ & \sup_{m\geq 0}\left(\kappa^{-m}\left\|\left(\sum_{n=0}^{m-1}\left(\prod_{i=n+1}^{m-1}A_i\right)g_n(z(n)),0\right)\right\|^L\right) +\\ & \sup_{m\geq 0}\left(\kappa^{-m}\left\|\left(0,-\sum_{n=0}^{\infty}\left(\prod_{i=0}^{n}B_{i+m}\right)^{-1}h_{n+m}(z(n+m))\right)\right\|^L\right),\end{aligned}$$

which by the bounds just established can be further estimated to be less than

$$\begin{aligned}&\|z\|_\kappa + \sup_{m\geq 0}\left(\kappa^{-m}\left(\prod_{i=0}^{m-1}\|A_i\|^L\right)\|y\|^L\right)+\\ &\sup_{m\geq 0}\left(\kappa^{-m}\sum_{n=0}^{m-1}\left(\prod_{i=n+1}^{m-1}\|A_i\|^L\right)B\gamma^{-n}(\|z(n)\|^L)^{1+\alpha}\right)+\\ &\sup_{m\geq 0}\left(\kappa^{-m}\sum_{n=0}^{\infty}\left(\prod_{i=0}^{n}\|B_{i+m}^{-1}\|^L\right)B\gamma^{-n-m}(\|z(n+m)\|^L)^{1+\alpha}\right).\end{aligned}$$

By using (L2) we see that this is less than

$$\begin{aligned}&\|z\|_\kappa + \sup_{m\geq 0}(\kappa^{-m}\lambda_1'^m)\cdot\|y\|^L +\\ &\sup_{m\geq 0}\left(\kappa^{-m}B\|z\|_\kappa^{1+\alpha}\sum_{n=0}^{m-1}\lambda_1'^{m-n-1}\gamma^{-n}\kappa^{(1+\alpha)n}\right)+\\ &\sup_{m\geq 0}\left(\kappa^{-m}B\|z\|_\kappa^{1+\alpha}\sum_{n=0}^{\infty}\lambda_2'^{-n-1}\gamma^{-n-m}\kappa^{(1+\alpha)(n+m)}\right).\end{aligned}$$

Because $\lambda' < \kappa$, we see that $\sup_{m\geq 0}(\kappa^{-m}\lambda_1'^m) = 1$. The maximum of the function $x \mapsto xa^x$ (where $0 < a < 1$) is at $\frac{-1}{\log a}$, and its value there is $\frac{-1}{e\log a}$. Therefore we have

$$\begin{aligned}&\sup_{m\geq 0}\left(\kappa^{-m}\lambda_1'^{m-1}\sum_{n=0}^{m-1}(\lambda_1'^{-1}\gamma^{-1}\kappa^{1+\alpha})^n\right)\\ \leq& \sup_{m\geq 0}\begin{cases}\lambda_1'^{-1}(\kappa^{-1}\lambda_1')^m m, & \text{if } \lambda_1'^{-1}\gamma^{-1}\kappa^{1+\alpha}\leq 1\\ \lambda_1'^{-1}(\gamma^{-1}\kappa^{\alpha})^m m, & \text{if } \lambda_1'^{-1}\gamma^{-1}\kappa^{1+\alpha}\geq 1\end{cases}\\ \leq& \sup_{m\geq 0}\begin{cases}-\lambda_1'^{-1}e^{-1}(\log(\kappa^{-1}\lambda_1'))^{-1}, & \text{if } \lambda_1'^{-1}\gamma^{-1}\kappa^{1+\alpha}\leq 1\\ -\lambda_1'^{-1}e^{-1}(\log(\gamma^{-1}\kappa^{\alpha}))^{-1}, & \text{if } \lambda_1'^{-1}\gamma^{-1}\kappa^{1+\alpha}\geq 1\end{cases}\\ \leq& \lambda_1'^{-1}e^{-1}\max\left(|\log(\lambda_1'\kappa^{-1})|^{-1}, |\log(\gamma^{-1}\kappa^{\alpha})|^{-1}\right).\end{aligned}$$

We'll call the bound obtained in the end M_1. Moreover, because

$$\lambda_2'^{-1}\gamma^{-1}\kappa^{1+\alpha} = (\kappa^\alpha\gamma^{-1})(\kappa\lambda_2'^{-1}) < 1,$$

we also have

$$\begin{aligned}
&\sup_{m\geq 0}\left(\kappa^{-m}\lambda_2'^{-1}\gamma^{-m}\kappa^{(1+\alpha)m}\sum_{n=0}^{\infty}(\lambda_2'^{-1}\gamma^{-1}\kappa^{1+\alpha})^n\right)\\
&=\sup_{m\geq 0}\left(\kappa^{-m}\lambda_2'^{-1}\gamma^{-m}\kappa^{(1+\alpha)m}\frac{1}{1-\lambda_2'^{-1}\gamma^{-1}\kappa^{1+\alpha}}\right)\\
&=\sup_{m\geq 0}\frac{\gamma^{-m}\kappa^{\alpha m}}{\lambda_2'-\gamma^{-1}\kappa^{1+\alpha}} \leq \frac{1}{\lambda_2'-\gamma^{-1}\kappa^{1+\alpha}},
\end{aligned}$$

and we'll call the bound obtained in the end M_2. Now set $M = M_1 + M_2$. Then

$$\|\Phi_\kappa(y,z)\|_\kappa \leq \|z\|_\kappa + \|y\|_\kappa + BM\|z\|_\kappa^{1+\alpha},$$

which implies that Φ_κ is well-defined.

Next we'd like to show that Φ_κ is actually a C^1-map for any fixed $m \geq 0$. Let's start by assuming $y \in B^s(r_0)$ and $t \in \mathbf{R}^k$ are such that $y + t \in B^s(r_0)$. Now for any $w \in W$ and $m \geq 0$ we have

$$\Phi_\kappa(y+t,z)(m) - \Phi_\kappa(y,z)(m) = \left(\left(\sum_{i=0}^{m-1}A_i\right)t, 0\right),$$

so that

$$\frac{\partial}{\partial y}\Phi_\kappa(y,z)(m) = \left(\left(\sum_{i=0}^{m-1}A_i\right), 0\right). \tag{2.3.21}$$

On the other hand let $y \in B^s(r_0)$ be arbitrary, $z \in W$ and $t \in \Gamma_\kappa$ be such that $z + t \in W$. Then

$$\Phi_\kappa(y,z+t) - \Phi_\kappa(y,z) = (A(z) - I)t + o(z,t),$$

where

$$(A(z))t(m) = \left(\sum_{n=0}^{m-1} \left(\prod_{i=n+1}^{m-1} A_i \right) Dg_n(z(n))t(n), \right. \tag{2.3.22}$$

$$\left. - \sum_{n=0}^{\infty} \left(\prod_{i=0}^{n} B_{i+m} \right)^{-1} Dh_{n+m}(z(n+m))t(n+m) \right)$$

$$o(z,t)(m) = \left(\sum_{n=0}^{m-1} \left(\prod_{i=n+1}^{m-1} A_i \right) (g_n((z+t)(n)) - g_n(z(n)) \right. \tag{2.3.23}$$

$$- Dg_n(z(n))t(n)),$$

$$- \sum_{n=0}^{\infty} \left(\prod_{i=0}^{n} B_{i+m} \right)^{-1} (h_{n+m}((z+t)(n+m))$$

$$- h_{n+m}(z(n+m))$$

$$\left. - Dh_{n+m}(z(n+m))t(n+m)) \right).$$

Now if $z, w \in W$ and $t \in \Gamma_\kappa$, then by (2.3.18)

$$\|(A(z) - A(w))t\|_\kappa \le$$

$$\sup_{m \ge 0} \kappa^{-m} \left(\sum_{n=0}^{m-1} \left(\prod_{i=n+1}^{m-1} \|A_i\|^L \right) B\gamma^{-n} (\|z(n) - w(n)\|^L)^\alpha \|t(n)\|^L + \right.$$

$$\left. \sum_{n=0}^{\infty} \left(\prod_{i=0}^{n} \|B_{i+m}^{-1}\| \right) B\gamma^{-n-m} (\|z(n+m) - w(n+m)\|^L)^\alpha \|t(n+m)\|^L \right) \le$$

$$\sup_{m \ge 0} \kappa^{-m} B \|z - w\|_\kappa^\alpha \|t\|_\kappa \left(\sum_{n=0}^{m-1} \lambda_1'^{m-n-1} \gamma^{-n} \kappa^{(1+\alpha)n} + \right.$$

$$\left. \sum_{n=0}^{\infty} \lambda_2'^{-n-1} \gamma^{-n-m} \kappa^{(n+m)(1+\alpha)} \right),$$

which results in

$$\|(A(z) - A(w))t\|_\kappa \le BM \|z - w\|_\kappa^\alpha \|t\|_\kappa. \tag{2.3.24}$$

By the mean value theorem we have

$$\begin{aligned}
&\|g_n((z+t)(n)) - g_n(z(n)) - Dg_n(z(n))t(n)\|^L = \\
&\|\left(\int_0^1 Dg_n((z+st)(n))ds\right) t(n) - Dg_n(z(n))t(n)\|^L \leq \\
&\|\int_0^1 (Dg_n((z+st)(n)) - Dg_n(z(n)))\, ds\|^L \cdot \|t(n)\|^L \leq \\
&\int_0^1 \|Dg_n((z+st)(n)) - Dg_n(z(n))\|\, ds \cdot \|t(n)\|^L \leq \\
&B\gamma^{-n}(\|t(n)\|^L)^{1+\alpha},
\end{aligned}$$

which together with a similar calculation for h_n gives us

$$\begin{aligned}
\|o(z,t)\|_\kappa &\leq \sup_{m\geq 0} \kappa^{-m}\Bigg(\sum_{n=0}^{m-1}\left(\prod_{i=n+1}^{m-1}\|A_i\|^L\right)B\gamma^{-n}(\|t(n)\|^L)^{1+\alpha}+ \\
&\qquad \sum_{n=0}^{\infty}\left(\prod_{i=0}^{n}\|B_i^{-1}\|^L\right)B\gamma^{-n-m}(\|t(n+m)\|^L)^{1+\alpha}\Bigg) \\
&\leq \sup_{m\geq 0} B\kappa^{-m}\|t\|_\kappa^{1+\alpha}\Bigg(\sum_{n=0}^{m-1}\lambda_1'^{m-n-1}\gamma^{-n}\kappa^{(1+\alpha)n}+ \\
&\qquad \sum_{n=0}^{\infty}\lambda_2'^{-n-1}\gamma^{-n-m}\kappa^{(1+\alpha)(n+m)}\Bigg) \\
&\leq BM\|t\|_\kappa^{1+\alpha}.
\end{aligned}$$

From all this we gather that

$$\frac{\partial}{\partial z}\Phi_\kappa(y,z) = A(z) - I \tag{2.3.25}$$

and (2.3.24) tells us that the derivative is continuous.

In the next phase we would like to apply the implicit function theorem to Φ_κ. However, since we want to show that the numbers D_0 and r in the theorem depend only on λ_1', λ_2', γ, κ, α and B, we will have to use the following spiced up version of the theorem.

Lemma 2.3.16 (Implicit Function Theorem). *Let X, Y and Z be Banach spaces, $A \subset X$ and $B \subset Y$ balls centered at the origin with radii $r_A, r_B > 0$ and $f\colon A \times B \to Z$ a C^1 map. Suppose that $f(0,0) = 0$ and that $\partial_2 f(0,0)\colon Y \to Z$ is a linear isomorphism. In addition assume that the derivative Df is Hölder continuous in $X \times Y$ with Hölder constant a and Hölder exponent α. Let $C \subset A$ be a ball of radius r_C centered in the origin, where*

$$r_C = \min\{r_A, r_B, \frac{r_B}{3bc}, \frac{1}{(1+3bc)(2ca)^{\frac{1}{\alpha}}}\},$$

$$b = \max_{x\in A}\|\partial_1 f(x,0)\|, \quad c = \|(\partial_2 f(0,0))^{-1}\|.$$

Then there exists a unique map $u: C \to B$ such that $f(x, u(x)) = 0$ for all $x \in C$. In addition $u \in C^{1+\alpha}$, $u(0) = 0$, and it satisfies

$$\left\| \frac{du}{dx}(x) - \frac{du}{dx}(y) \right\| \le 8ca(1+2bc)^2 \|x-y\|^\alpha, \text{ for all } x, y \in C,$$

and

$$\left\| \frac{du}{dx}(x) \right\| \le 1 + 2bc$$

for all $x \in C$. □

Proof. In this proof we'll follow [Pes76], which in turn refers to the book by J. Dieudonné [Die60].

Write $T = \partial_2 f(0,0)$ for the linear isomorphism $Y \to Z$ and define $g(x,y) = y - T^{-1}f(x,y)$. Suppose that $(x, y_1), (x, y_2) \in A \times B$. Then

$$\begin{aligned} g(x,y_1) - g(x,y_2) &= y_1 - y_2 + T^{-1}f(x,y_2) - T^{-1}f(x,y_1) \\ &= T^{-1}(Ty_1 - Ty_2 + f(x,y_2) - f(x,y_1)). \end{aligned}$$

Choose $r = 3bcr_C$. Then by definition of r_C we have that $r \le r_B$ and $r_C \le r_A$. We let $U \subset X$ and $V \subset Y$ be balls of radii r_C and r respectively, centered at 0. If $x \in U$ and $y_1, y_2 \in V$, then by the mean value theorem (See [Die60], Theorem 8.6.2.) we have

$$\begin{aligned} \|f(x,y_2) - f(x,y_1) - T(y_2 - y_1)\| &= \\ \|f(x,y_2) - f(x,y_1) - Df(0,0)\cdot((x,y_2)-(x,y_1))\| &\le \\ \|y_2 - y_1\| \sup_{(x',y')\in U\times V} \|Df(x',y') - Df(0,0)\| &\le \\ a(r_C + r)^\alpha \|y_2 - y_1\| \le \frac{1}{2c}\|y_2 - y_1\|. \end{aligned}$$

The last inequality follows from the definitions of r_C and r since we have the identity

$$a(r_C + r)^\alpha = a(r_C(1+3bc))^\alpha \le \frac{1}{2c}.$$

Now this implies that for all $x \in U$, $y_1, y_2 \in V$ we have

$$\|g(x,y_1) - g(x,y_2)\| \le \frac{1}{2}\|y_1 - y_2\|.$$

In addition for all $x \in U$ we have

$$\|g(x,0)\| = \|-T^{-1}f(x,0)\| \le c\|f(x,0)\| \le cbr_C = \frac{1}{3}r.$$

This means that the function $g: U \times V \to Y$ satisfies the assumptions of the fixed point theorem A.3. Thus there exists a unique continuous function $u: U \to V$ such that $u(0) = 0$ and $f(x, u(x)) = 0$ for all $x \in S$.

We will next prove that u is continuously differentiable in U. Suppose that $x, h \in U$ and write

$$t = u(x+h) - u(x).$$

Now $f(x+h, u(x)+t) = 0$ and $t \to 0$ as $h \to 0$. Let

$$D_1(x) = \partial_1 f(x, u(x)), \quad D_2(x) = \partial_2 f(x, u(x)),$$

and notice that for any $\varepsilon > 0$ there exists $\delta > 0$ such that for all $h < \delta$ we have

$$\|f(x+h, u(x)+t) - f(x, u(x)) - D_1(x) \cdot h - D_2(x) \cdot t\| \le \varepsilon(\|h\| + \|t\|).$$

This implies that

$$\|D_1(x) \cdot h + D_2(x) \cdot t\| \le \varepsilon(\|h\| + \|t\|).$$

Now notice that $\partial_2 f(x, y)$ is actually invertible in $(x, y) \in U \times V$. Indeed for all $v \in V$ we have

$$\begin{aligned}
\|\partial_2 f(x, y)v\| &\ge \|\partial_2 f(0,0)v\| - \|\partial_2 f(x, y)v - \partial_2 f(0,0)v\| \\
&\ge \frac{\|v\|}{c} - \|\partial_2 f(x, y) - \partial_2 f(0,0)\| \|v\| \\
&\ge \|v\| \left(\frac{1}{c} - a(r + r_C)^\alpha \right),
\end{aligned}$$

which leads to the inequality

$$\|\partial_2 f(x, y)v\| \ge \frac{\|v\|}{2c}. \tag{2.3.26}$$

Since $(x, y) \mapsto \partial_2 f(x, y)$ is a continuous mapping and $U \times V$ is connected, it follows from A.5 that $D_2(x)$ has an inverse and we get

$$\|D_2^{-1}(x) D_1(x) \cdot h + t\| \le \|D_2^{-1}(x)\| \varepsilon(\|h\| + \|t\|).$$

Suppose that ε is chosen so small that $\varepsilon \|D_2^{-1}(x)\| \le \frac{1}{2}$ and set

$$p = 2\|D_2^{-1}(x) D_1(x)\| + 1.$$

Then

$$\|t\| - \frac{p-1}{2}\|h\| \le \frac{1}{2}(\|t\| + \|h\|)$$

and thus $\|t\| \le p\|h\|$. This gives us

$$\|t + D_2^{-1}(x) D_1(x) \cdot h\| \le \varepsilon(p+1)\|D_2^{-1}(x)\| \|h\|,$$

which by definition of t proves the differentiability of u.

Now the inequality (2.3.26) leads us to

$$\|(\partial_2 f(x, y))^{-1}\| \le 2c, \tag{2.3.27}$$

and in addition we have

$$\|\partial_1 f(x, y)\| \le \|\partial_1 f(x, y) - \partial_1 f(x, 0)\| + \|\partial_1 f(x, 0)\| \le \frac{1}{2c} + b. \tag{2.3.28}$$

By differentiating the equation $f(x,u(x))=0$, we get

$$\partial_1 f(x,u(x)) + \partial_2 f(x,u(x))\frac{du}{dx}(x) = 0.$$

The inequalities (2.3.28) and (2.3.27) imply that

$$\|\frac{du}{dx}(x)\| \le 1 + 2bc.$$

Moreover if $x_1, x_2 \in U$, then

$$\begin{aligned}\partial_1 f(x_1,u(x_1)) - \partial_1 f(x_2,u(x_2)) + \\ (\partial_2 f(x_1,u(x_1)) - \partial_2 f(x_2,u(x_2)))\frac{du}{dx}(x_1) + \\ \partial_2 f(x_2,u(x_2))\left(\frac{du}{dx}(x_1) - \frac{du}{dx}(x_2)\right) = 0.\end{aligned}$$

Thus

$$\begin{aligned}\|\frac{du}{dx}(x_1) - \frac{du}{dx}(x_2)\| \le \\ \|(\partial_2 f(x_2,u(x_2)))^{-1}\|\|\partial_1 f(x_1,u(x_1)) - \partial_1 f(x_2,u(x_2))\| \\ +\|(\partial_2 f(x_2,u(x_2)))^{-1}\|\|\partial_2 f(x_1,u(x_1)) - \partial_2 f(x_2,u(x_2))\|\|\frac{du}{dx}(x_1)\| \le \\ 2ca(\|x_1-x_2\| + \|u(x_1)-u(x_2)\|)^\alpha \\ +2ca(1+2bc)(\|x_1-x_2\| + \|u(x_1)-u(x_2)\|)^\alpha \le \\ 2ca(\|x_1-x_2\| + (1+2bc)\|x_1-x_2\|)^\alpha \\ +2ca(1+2bc)(\|x_1-x_2\| + (1+2bc)\|x_1-x_2\|)^\alpha \le \\ 4ca(1+bc)\|x_1-x_2\|^\alpha + 4ca(1+2bc)(1+bc)\|x_1-x_2\|^\alpha \le \\ 8ca(1+2bc)^2\|x_1-x_2\|^\alpha,\end{aligned}$$

which proves the lemma. ■

We'll now show that Φ_κ satisfies the assumptions of the lemma above. First notice that by (2.3.24) we have

$$\begin{aligned}\|\partial_2\Phi_\kappa(y_1,z_1) - \partial_2\Phi_\kappa(y_2,z_2)\|_\kappa \le &\|\partial_2\Phi_\kappa(y_1,z_1) - \partial_2\Phi_\kappa(y_1,z_2)\|_\kappa + \\ &\|\partial_2\Phi_\kappa(y_1,z_2) - \partial_2\Phi_\kappa(y_2,z_2)\|_\kappa \\ = &\|A(z_1) - A(z_2)\|_\kappa \le BM\|z_1-z_2\|_\kappa^\alpha,\end{aligned}$$

so that $\partial_2\Phi_\kappa$ is Hölder continuous. Thus in the lemma we will have

$$\begin{aligned} a &= BM \\ b &= \max_{y\in B^s(r_0)} \|\partial_1\Phi_\kappa(y,0)\|_\kappa = \max_{y\in B^s(r_0)} \sup_{\|t\|^L=1} \|\partial_1\Phi_\kappa(y,0)t\|_\kappa \\ &= \max_{y\in B^s(r_0)} \sup_{\|t\|^L=1} \sup_{m\ge 0} \kappa^{-m}\|(\partial_1\Phi_\kappa(y,0)t)(m)\|^L \\ &= \max_{y\in B^s(r_0)} \sup_{\|t\|^L=1} \sup_{m\ge 0} \kappa^{-m}\|\left(\left(\prod_{i=0}^{m-1} A_i\right)t, 0\right)\|^L = 1 \\ c &= \|(\partial_2\Phi_\kappa(0,0))^{-1}\|_\kappa = \|-I\|_\kappa = 1.\end{aligned}$$

Remember that $\|A_i\| \le \lambda_1' < \kappa$.

Therefore the lemma gives us a map $\varphi\colon B^s(r) \to W$ with Hölder continuous derivative, where r depends only on r_0, B, κ, γ, λ_1', λ_2' and α, and which satisfies

$$\varphi(0) = 0, \quad \Phi_\kappa(y, \varphi(y)) = 0 \quad (y \in B^s(r)). \tag{2.3.29}$$

When we differentiate the second equation we get

$$\frac{d}{dy}\Phi_\kappa(y, \varphi(y)) = \partial_1\Phi_\kappa(y, \varphi(y)) + \partial_2\Phi_\kappa(y, \varphi(y))\frac{d}{dy}\varphi(y),$$

which gives us

$$\frac{d}{dy}\varphi(y) = -(\partial_2\Phi_\kappa(y, \varphi(y)))^{-1}(\partial_1\Phi_\kappa(y, \varphi(y))).$$

If we set $y = 0$ in the above formula, we get by (2.3.22), (2.3.25) and (2.3.21), that

$$\frac{d}{dy}\varphi(0)(m) = \left(\prod_{i=0}^{m-1} A_i, 0\right). \tag{2.3.30}$$

Let us now write $\varphi(y)(m) \in \mathbf{R}^d$ as $(\varphi_1(y)(m), \varphi_2(y)(m)) \in \mathbf{R}^k \times \mathbf{R}^{d-k}$. Then by (2.3.29) and (2.3.20) we see that

$$\varphi_1(y)(m) = \left(\prod_{i=0}^{m-1} A_i\right) y + \sum_{n=0}^{m-1}\left(\prod_{i=n+1}^{m-1} A_i\right) g_n(\varphi(y)(n)) \tag{2.3.31}$$

$$\varphi_2(y)(m) = -\sum_{n=0}^{\infty}\left(\prod_{i=0}^{n} B_{i+m}\right)^{-1} h_{n+m}(\varphi(y)(n+m)). \tag{2.3.32}$$

Therefore it follows that

$$\begin{aligned}
\varphi_1(y)(m+1) =& A_m\left(\prod_{i=0}^{m-1} A_i\right) y + \sum_{n=0}^{m-1} A_m\left(\prod_{i=n+1}^{m-1} A_i\right) g_n(\varphi(y)(n)) + \\
& g_m(\varphi(y)(m)) \\
=& A_m\varphi_1(y)(m) + g_m(\varphi_1(y)(m), \varphi_2(y)(m)) \\
\varphi_2(y)(m+1) =& -\sum_{n=0}^{\infty}\left(\prod_{i=1}^{n+1} B_{i+m}\right)^{-1} h_{n+m+1}(\varphi(y)(n+m+1)) \\
=& -\sum_{n=1}^{\infty}\left(\prod_{i=1}^{n} B_{i+m}\right)^{-1} h_{n+m}(\varphi(y)(n+m)) \\
=& -\sum_{n=1}^{\infty} B_m\left(\prod_{i=0}^{n} B_{i+m}\right)^{-1} h_{n+m}(\varphi(y)(n+m)) \\
=& B_m\varphi_2(y)(m) + h_m(\varphi_1(y)(m), \varphi_2(y)(m)).
\end{aligned}$$

From these relations we immediately obtain

$$\widetilde{F}_m(\varphi(y)(m)) = \varphi(y)(m+1). \tag{2.3.33}$$

We will now define the map ψ^s by

$$\psi^s(v) = \varphi_2(v)(0) = -\sum_{n=0}^{\infty}\left(\prod_{i=0}^{n} B_i\right)^{-1} h_n(\varphi(v)(n))$$

for all $v \in B^s(r)$, the second equality coming from (2.3.32). Then it follows by (2.3.30) and the fact that φ is a $C^{1+\alpha}$ map that the claim 1. of the theorem holds for the map ψ^s. By using (2.3.33), we obtain

$$\begin{aligned}\prod_{i=0}^{m-1}\widetilde{F}_i(v,\psi^s(v)) &= \prod_{i=0}^{m-1}\widetilde{F}_i(\varphi_1(v)(0),\varphi_2(v)(0))\\ &= \prod_{i=0}^{m-1}\widetilde{F}_i(\varphi(v)(0)) = \varphi(v)(m) \in B^s(r)\times B^u(r),\end{aligned}$$

which in turn can be used together with the mean value theorem and Lemma 2.3.16 to get

$$\begin{aligned}\left\|\prod_{i=0}^{m-1}\widetilde{F}_i(v,\psi^s(v))\right\|^L &\le \kappa^m\|\varphi(v)\|_\kappa = \kappa^m\|\varphi(v)-\varphi(0)\|_\kappa\\ &\le \kappa^m\|\frac{d}{d\xi}\varphi(\xi)\|_\kappa\|v\|^L \le 3\kappa^m\|v\|^L\\ &\le 3\kappa^m\|(v,\psi^s(v))\|^L.\end{aligned}$$

This proves the 3. claim of the theorem. Now if $v,w \in B^s(r)$, then again by Lemma 2.3.16 we have

$$\begin{aligned}\|D\psi^s(v)-D\psi^s(w)\|^L &= \|D(\phi_2(v)(0))(v)-D(\phi_2(w)(0))(w)\|^L\\ &\le \|D(\phi(v)(0))(v)-D(\phi(w)(0))(w)\|^L\\ &\le 72BM\|v-w\|_\kappa^\alpha,\end{aligned}$$

which proves the 2. and 5. claim. (We have taken $D_0 = \max\{3, 72BM\}$.)

In order to finish and prove the 4. claim, suppose that $v \in B^s(r)$ and $w \in B^u(r)$ and that there exists a number $K > 0$ such that

$$\left(\prod_{i=0}^{m-1}\widetilde{F}_i\right)(v,w) \in B^s(r)\times B^u(r)$$

and

$$\left\|\left(\prod_{i=0}^{m-1}\widetilde{F}_i\right)(v,w)\right\|^L \le K\kappa^m$$

for all $m \ge 0$. Define $z \in \Gamma_\kappa$ by

$$z(m) = \left(\prod_{i=0}^{m-1}\widetilde{F}_i\right)(v,w),$$

for all $m \ge 0$. We see that z is well-defined (i.e. in Γ_κ), since

$$\|z\|_\kappa = \sup_{m\ge0}\kappa^{-m}\|\left(\prod_{i=0}^{m-1}\widetilde{F}_i\right)(v,w)\|^L \le K.$$

If we can prove that $\Phi_\kappa(v,z) = 0$, then by the uniqueness of φ we have $\varphi(v) = z$ and thus $\psi^s(v) = \varphi_2(v)(0) = w$, which is the 4. claim. We'll prove $\Phi_\kappa(v,z)(m) = 0$ for all $m \geq 0$. Let us first note that if we write $z(m)$ as $(z_1(m), z_2(m)) \in B^s(r) \times B^u(r)$, then we have the following two relations:

$$\begin{aligned} z_1(m+1) &= A_m z_1(m) + g_m(z(m)) \\ z_2(m+1) &= B_m z_2(m) + h_m(z(m)). \end{aligned}$$

We'll first prove by induction that

$$\Phi_1(v,z)(m) = -z_1(m) + \left(\prod_{i=0}^{m-1} A_i\right) v + \sum_{n=0}^{m-1} \left(\prod_{i=n+1}^{m-1} A_i\right) g_n(z(n)) = 0.$$

Notice that when $m = 0$, we have

$$\Phi_1(v,z)(0) = -v + v = 0.$$

Suppose that the claim holds for some $m \geq 0$. Then

$$\begin{aligned} \Phi_1(v,z)(m+1) = &- A_m z_1(m) - g_m(z(m)) \\ &+ \left(\prod_{i=0}^{m} A_i\right) v + \sum_{n=0}^{m} \left(\prod_{i=n+1}^{m} A_i\right) g_n(z(n)) \\ =& A_m(\Phi_1(v,z)(m)) - g_m(z(m)) + g_m(z(m)) = 0. \end{aligned}$$

Next we shall prove that

$$\Phi_2(v,z)(m) = -z_2(m) - \sum_{n=0}^{\infty} \left(\prod_{i=0}^{n} B_{i+m}\right)^{-1} h_{n+m}(z(n+m)) = 0.$$

Suppose that $N > m$. Then

$$\begin{aligned} z_2(N) &= B_{N-1} z_2(N-1) + h_{N-1}(z(N-1)) \\ &= B_{N-1} B_{N-2} z_2(N-2) + B_{N-1} h_{N-2}(z(N-2)) + h_{N-1}(z(N-1)) \\ &= \cdots \\ &= \left(\prod_{i=m}^{N-1} B_i\right) z_2(m) + \sum_{n=m}^{N-1} \left(\prod_{i=n+1}^{N-1} B_i\right) h_n(z(n)). \end{aligned}$$

From this one obtains

$$\begin{aligned} z_2(m) &= \left(\prod_{i=m}^{N-1} B_i\right)^{-1} z_2(N) - \sum_{n=m}^{N-1} \left(\prod_{i=m}^{N-1} B_i\right)^{-1} \left(\prod_{i=n+1}^{N-1} B_i\right) h_n(z(n)) \\ &= \left(\prod_{i=m}^{N-1} B_i\right)^{-1} z_2(N) - \sum_{n=m}^{N-1} \left(\prod_{i=m}^{n} B_i\right)^{-1} h_n(z(n)). \end{aligned}$$

Next we notice that

$$\left\| \left(\prod_{i=m}^{N-1} B_i\right)^{-1} z_2(N) \right\|^L \leq \lambda_2'^{-(N-m)} \|z_2(N)\|^L \leq \lambda_2'^{-(N-m)} r,$$

so that as $N \to \infty$, we get

$$z_2(m) = -\sum_{n=m}^{\infty} \left(\prod_{i=m}^{n} B_i \right)^{-1} h_n(z(n)).$$

Thus $\Phi_\kappa(v,z)(m) = (\Phi_1(v,z)(m), \Phi_2(v,z)(m)) = 0$, which finishes the proof. ∎

2.3.3. *Some further results*

We'd like to describe some futher results on local stable manifolds and their uses. This is where one can use the extra information on the map ψ^s given by Theorem 2.3.14. Proofs are given in [BP99] as well as the original articles by Pesin: [Pes76] and [Pes77].

Recall that Theorem 2.3.4 shows that under suitable conditions a nonuniformly hyperbolic set $\mathcal{R}$ exists. After the definition of the nonuniformly hyperbolic set we made a remark that one can take such a set apart by defining

$$\mathcal{R}^\ell = \{x \in \mathcal{R} : C(x) \le \ell, K(x) \ge \frac{1}{\ell}\}$$

for all $\ell > 0$. One immediately notices that $\mathcal{R}^\ell \subset \mathcal{R}^{\ell'}$ when $0 < \ell < \ell'$. In addition by (H5), for all $m \in Z$ we have $f^m(\mathcal{R}^\ell) \subset \mathcal{R}^{\ell'}$, where $\ell' = \ell \exp(|m|\varepsilon(x))$.

Now, one can prove that the sizes of local stable manifolds are bounded from below on $\mathcal{R}^\ell$. I.e. there exists a constant r_ℓ such that

$$r(x) \ge r_\ell > 0$$

for all $x \in \mathcal{R}^\ell$, where $r(x)$ is the r in the statement of Theorem 2.3.14. Moreover one can prove that the local stable manifolds $V(x)$ depend continuously on $x \in \overline{\mathcal{R}^\ell}$ in the C^1-topology.

Using properties of local stable manifolds and with the help of so called *holonomy map* one can also prove Pesin's Ergodic Component Theorem, which states that if $f: M \to M$ is a $C^{1+\alpha}$ diffeomorphism on a smooth Riemannian manifold M and μ is a f-invariant measure equivalent to the Riemannian volume, then the set Λ divides into countably many ergodic components excluding a set of measure zero.

A. Analysis and topology

Theorem A.1. *Every compact metric space X admits a countable basis for its topology.* □

Proof. For each $k \in \mathbf{N}$ the open balls $B(x, \frac{1}{k})$ $(x \in X)$ form a cover of X. Because X is compact, there is a countable (actually even finite) subcover of balls $B(x_n^k, \frac{1}{k})$, $x_n^k \in X$, $n \in \mathbf{N}$. We will show that the set

$$\{B(x_n^k, \frac{1}{k}) : n, k \in \mathbf{N}\}$$

forms a basis. Indeed suppose that $U \subset X$ is open and that $x \in U$. Let $r = d(x, X \setminus U) > 0$ be the distance from x to $X \setminus U$. Now there is a large enough k for which $\frac{1}{k} < \frac{r}{2}$. Moreover there is $n \in \mathbf{N}$ such that $x \in B(x_n^k, \frac{1}{k}) \subset U$. ■

Theorem A.2. *For a compact metric space X, the space $C(X)$ of continuous functions $X \to \mathbf{R}$ endowed with* sup*-norm is separable.* □

Proof. Because X has a countable basis, call the sets in it B_n $(n = 1, 2, \dots)$, we can choose a sequence $(x_n)_{n=1}^\infty$ of points, where $x_n \in B_n$ for all $n \in \mathbf{N}$. Let $\mathcal{F}$ be the set of functions consisting of $f_n : X \to \mathbf{R}$ given by $f_n(x) = d(x, x_n)$, $n = 1, 2, \dots$, as well as the constant function 1. Let $\mathcal{A}$ be the algebra (over $\mathbf{C}$) generated by functions f_n. Then it separates points because if $x, y \in X$, then we may pick some x_n and x_k close enough to x and y respectively to see that $f_n(y) > f_n(x)$. Stone-Weierstraß Theorem (See the corollary of [Rud91] Theorem 5.7.) now tells us that $\mathcal{A}$ is dense in $C_{\mathbf{C}}(X) = \{f : f : X \to \mathbf{C}\}$.

Now let $f : X \to \mathbf{R}$ be continuous and $\varepsilon > 0$. Then there exists a function g of form

$$g = \sum_{k=1}^{n} c_k \prod_{i=1}^{m_n} f_{k,i}, \quad c_k \in \mathbf{C}, f_{k,i} \in \mathcal{F}$$

satisfying $d(f, g) < \varepsilon$. Now

$$\frac{1}{2}(g + \overline{g}) = \sum_{k=1}^{n} \frac{1}{2}(c_k + \overline{c_k}) \prod_{i=1}^{m_n} f_{k,i}$$

is in $C(X)$ and satisfies $d(f, \frac{1}{2}(g + \overline{g})) < \varepsilon$. Because $\mathbf{Q} \subset \mathbf{R}$ is dense, we get that the functions of form

$$\sum_{k=1}^{n} q_k \prod_{i=1}^{m_n} f_{k,i}, \quad q_k \in \mathbf{Q}, f_{k,i} \in \mathcal{F}$$

are dense in $C(X)$. There are a countable number of them. ■

The following theorem is a version of Banach fixed point theorem. The statement and proof are taken from [Die60].

Theorem A.3. *Let X, Y be two Banach spaces and $U \subset X$, $V \subset Y$ open balls of radius r and s respectively, centered at the origin. Suppose that ψ is a continuous mapping $U \times V \to Y$, which is a contraction in the second variable, i.e.*

$$\|\psi(x, y_1) - \psi(x, y_2)\| \leq k\|y_1 - y_2\|$$

for $x \in U$, $y_1, y_2 \in V$ and $0 \leq k < 1$. Now if $\|\psi(x, 0)\| < s(1-k)$ for all $x \in U$, then there exists a unique continuous mapping $f\colon U \to V$ such that

$$f(x) = \psi(x, f(x))$$

for any $x \in U$. □

Remark A.4. *The theorem implies Banach fixed point theorem. Just apply it to the mapping $\psi(x, y) = g(y)$, where $g\colon V \to Y$ is the given contraction.* □

Proof. Fix some $x \in U$. We recursively construct a sequence of points $0 = y_0, y_1, y_2, \dots \in V$ by defining

$$y_n = \psi(x, y_{n-1}), \quad n \geq 1$$

For this definition to make sense, we have to show that $\psi(x, y_n) \in V$ if y_i is defined as above for all $1 \leq i \leq n$. Now for $2 \leq i \leq n+1$ we have

$$\|y_i - y_{i-1}\| = \|\psi(x, y_{i-1}) - \psi(x, y_{i-2})\| \leq k\|y_{i-1} - y_{i-2}\|,$$

which by induction gives us $\|y_i - y_{i-1}\| \leq k^{i-1}\|y_1\|$. Thus

$$\|y_i\| \leq k^{i-1}\|y_1\| + \|y_{i-1}\| \leq (1 + \dots + k^{i-1})\|y_1\| \leq \frac{\|y_1\|}{1-k} < s,$$

as was claimed. In addition by induction we can also write $y_n = f_n(x)$, where f_n is a continuous mapping $U \to V$ for all $n \geq 0$. These mappings satisfy

$$\|f_n(x) - f_{n-1}(x)\| \leq k^{n-1}s(1-k),$$

and thus $f_n - f_{n-1}$ converges in norm. Because Y is complete, this implies that f_n converges to some continuous function $f\colon U \to Y$. In addition f satisfies

$$\|f(x)\| \leq \frac{\|\psi(x, 0)\|}{1-k} < s$$

for all $x \in U$, which means that the image of f is contained in V. Because

$$f_n(x) = \psi(x, f_{n-1}(x)),$$

we get as $n \to \infty$ that

$$f(x) = \psi(x, f(x))$$

for all $x \in U$.

Suppose finally that $g\colon U \to V$ is another mapping satisfying $g(x) = \psi(x, g(x))$. Then

$$\|g(x) - f(x)\| = \|\psi(x, g(x)) - \psi(x, f(x))\| \le k\|g(x) - f(x)\|,$$

which implies $g(x) = f(x)$ because $k < 1$. ∎

Another lemma that we'll need is the following about the invertibility of bounded linear operators on a connected set.

Lemma A.5. *Suppose that X and Y are Banach spaces and let $\mathcal{B}(X, Y)$ denote the space of bounded linear operators $X \to Y$ with the usual operator norm. Suppose that Γ is a connected subset of $\mathcal{B}(X, Y)$ and there exists a constant $C > 0$ such that for all $T \in \Gamma$ we have $\|T\| \le C$ and $\|Tx\| \ge \frac{\|x\|}{C}$ for all $x \in X$. Then the set*

$$\mathcal{S} = \{T \in \Gamma : T \text{ is invertible}\}$$

is either $\varnothing$ or Γ. In particular if Γ contains one invertible operator, then all the operators in Γ are invertible. □

Proof. Notice that the condition $\|Tx\| \ge \frac{\|x\|}{C}$ readily implies that every $T \in \Gamma$ is an injection, it is hence enough to check surjectivity. Moreover because Γ is connected, it is enough to show that $\mathcal{S}$ is both open and closed in Γ. We'll start by showing it's open.

Suppose that $T \in \mathcal{S}$. We wish to find a small neighbourhood of T such that it is contained in $\mathcal{S}$ and will prove that an open ball of radius $\frac{1}{C}$ works. Suppose that $A \in \Gamma$ satisfies $\|T - A\| < \frac{1}{C}$. We wish to show that A is surjection, so let $y \in Y$ be arbitrary and consider the sequence $x_n \in X$, $n = 1, 2, \dots$, defined by

$$x_1 = T^{-1}y, \quad x_{n+1} = T^{-1}(T - A)x_n.$$

Then one easily sees that

$$A(x_1 + x_2 + \cdots + x_n) = (A - T)x_n + y$$

and

$$x_n = \left(T^{-1}(T - A)\right)^{n-1} T^{-1}y.$$

From the second equation it follows that

$$\|x_n\| \le \|T^{-1}\|\|y\|(\|T^{-1}\|\|T - A\|)^{n-1},$$

and because

$$\|T^{-1}\|\|T - A\| < \frac{\|T^{-1}\|}{C} \le 1,$$

we get

$$\sum_{n=1}^{\infty} \|x_n\| = \frac{\|T^{-1}\|\|y\|}{1 - \|T^{-1}\|\|T - A\|} < \infty.$$

Hence $\sum_{n=1}^{\infty} x_n = x$ exists and we have

$$Ax = \lim_{n\to\infty} A(x_1 + \cdots + x_n) = \lim_{n\to\infty} (A - T)x_n + y = y.$$

Thus $A \in \mathcal{S}$ and in particular $\mathcal{S}$ is open in Γ.

Next we wish to show that $\mathcal{S}$ is closed in Γ. So suppose that $T_n \in \mathcal{S}$, $n = 1, 2, \ldots$, is a sequence of operators converging to $T \in \Gamma$. We want to show that T is a surjection, so let $y \in Y$. Then for each $n \in \mathbf{N}$ there exists $x_n \in X$ such that $T_n x_n = y$. Hence

$$\begin{aligned}\|x_n - x_m\| &= \|T_n^{-1}y - T_m^{-1}y\| \le \|T_n^{-1} - T_m^{-1}\|\|y\| \\ &= \|T_n^{-1}(T_m - T_n)T_m^{-1}\|\|y\| \le C^2\|T_m - T_n\|\|y\|,\end{aligned}$$

which shows that the sequence x_n is Cauchy. Thus it has a limit $x \in X$ and one gets

$$\|Tx - y\| = \|Tx - T_n x_n\| \to 0,$$

so that $\|Tx - y\| = 0$. Hence T is a surjection and we are done. ∎

Finally we have the following results, which are used in a perturbation argument when proving B.3.

Lemma A.6. *Suppose that X and Y are topological spaces, X is compact and $f\colon X \times Y \to \mathbf{R}$ is continuous. Then for any $y_0 \in Y$ and $\varepsilon > 0$ there exists an open neighbourhood U of y_0 such that*

$$\sup_{x\in X, y\in U} |f(x,y) - f(x,y_0)| \le \varepsilon.$$

□

Proof. For all $(x, y_0) \in X \times Y$ there exists a neighbourhood $A_x \times B_x$, $x \in A_x \subset X$, $y_0 \in B_x \subset Y$ are open, such that

$$|f(a,b) - f(x,y_0)| < \varepsilon$$

for all $(a,b) \in A_x \times B_y$. The neighbourhoods A_x cover X and by compactness we may pick a finite subcover $A_{x_1}, \ldots, A_{x_n}$. Now take $U = \bigcap_{i=1}^n B_{x_i}$. Then

$$|f(x,y) - f(x,y_0)| < \varepsilon$$

for all $x \in X$, $y \in U$. ∎

Theorem A.7. *Let $f(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$ be a polynomial and write $a = (a_0, \ldots, a_{n-1})$. Then for any $\varepsilon > 0$ there exists $\delta > 0$ such that for any vector $b = (b_0, b_1, \ldots, b_{n-1})$ satisfying $|b - a| < \delta$ the roots of the polynomial*

$$g(z) = z^n + b_{n-1}z^{n-1} + \cdots + b_1 z + b_0$$

can be bijectively mapped on the roots of f so that the pairs of roots so obtained are within a distance of at most ε from each other. □

Proof. Suppose that we have been given $\varepsilon > 0$. Let $z_1, \ldots, z_k$ be the roots of $f(z)$ with multiplicities $\alpha_1, \ldots, \alpha_k$ respectively. For each $1 \leq i \leq k$ we may choose a radius $0 < r_i < \frac{\varepsilon}{2}$ such that the only zero of f inside the z_i-centered closed disk of radius r_i is z_i. Let r be the largest of $r_i, 1 \leq i \leq k$.

The finite union $F = \bigcup_{i=1}^{k} \partial B(z_i, r_i)$ of compact sets is compact. Consider the mapping $G\colon F \times \mathbf{C}^n \to \mathbf{R}$ given by

$$G(z, b) = \left| \frac{nz^{n-1} + (n-1)b_{n-1}z^{n-2} + \cdots + b_1}{z^n + b_{n-1}z^{n-1} + \cdots + b_1 z + b_0} - \frac{f'(z)}{f(z)} \right|.$$

By using the above lemma to the function

$$(z, b) \mapsto z^n + b_{n-1}z^{n-1} + \cdots + b_1 z + b_0,$$

we see that it doesn't vanish (and hence G is well-defined) when b is in some neighbourhood of a. By using the above lemma to G, we see that there exists $\delta > 0$ such that

$$\max_{z \in F} G(z, b) < \frac{1}{r}$$

when $\|b - a\| < \delta$.

Suppose that $b \in B(a, \delta)$ and write

$$g(z) = z^n + b_{n-1}z^{n-1} + \cdots + b_1 z + b_0.$$

Then

$$\begin{aligned} \left| \frac{1}{2\pi i} \int_{\partial B(z_i, r_i)} \frac{g'(z)}{g(z)} - \frac{f'(z)}{f(z)} dz \right| &\leq \frac{1}{2\pi} \int_{\partial B(z_i, r_i)} \left| \frac{g'(z)}{g(z)} - \frac{f'(z)}{f(z)} \right| |dz| \\ &< \frac{r_i}{r} \leq 1. \end{aligned}$$

Because by argument principle the left hand side is an integer, it must be zero. Thus there are as many roots of f as there are of g in the disk $B(z_i, r_i)$. This proves the theorem since now we may arbitrarily pair the roots inside each disk and the distance between any pair of roots will be at most ε. ∎

Theorem A.8. *Let A be a $m \times m$ matrix. Then the eigenvalues of A depend continuously on A. For any $\varepsilon > 0$ there exists $\delta > 0$ such that for all matrices B with $\|A - B\| < \delta$ the eigenvalues of B can be paired (taking algebraic multiplicity into account) with the eigenvalues of A so that the distance between every pair of eigenvalues is at most ε.* □

Proof. We simply note that the eigenvalues are given by the characteristic polynomial $\det(A - \lambda I)$. The coefficient of λ^m will be 1 if m is even and -1 if m is odd. In the odd case the above theorem can still be used because multiplying the polynomial by -1 won't alter the roots. ∎

B. Linear algebra

Definition B.1. We define an inner product on $\wedge^q \mathbf{R}^m$ by setting

$$\langle v_1 \wedge \cdots \wedge v_q, w_1 \wedge \cdots \wedge w_q \rangle = \det(\langle v_i, w_j \rangle_{i,j}).$$

□

One can prove that under the above defined inner product, the vectors of form $e_{i_1} \wedge \cdots \wedge e_{i_q}$, $1 \leq i_1 < \cdots < i_q \leq m$, form an orthonormal basis for $\wedge^q \mathbf{R}^m$. Here $e_1, \ldots, e_m$ is the usual orthonormal basis for $\mathbf{R}^m$.

Definition B.2. Let A be a linear operator on the finite dimensional vector space V. Then $A^{\wedge q}$ is the linear operator that acts on basis vectors $v_1 \wedge \cdots \wedge v_q$ of $\wedge^q V$ by

$$A^{\wedge q}(v_1 \wedge \cdots \wedge v_q) = Av_1 \wedge \cdots \wedge Av_q.$$

□

Theorem B.3. *Let A be a $m \times m$ matrix with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_m$. Then the eigenvalues of $A^{\wedge q}$ are of the form*

$$\prod_{k=1}^{q} \lambda_{\sigma(k)},$$

where σ is a choice of q (distinct) values from $\{1, 2, \ldots, m\}$. □

Proof. Consider first the case where A has m distinct eigenvalues. Then there exist also m linearly independent eigenvectors $v_1, \ldots, v_m$. They form a basis in $\mathbf{R}^m$ and hence the vectors $v_{\sigma(1)} \wedge \cdots \wedge v_{\sigma(q)}$ form a basis in $\wedge^q \mathbf{R}^m$. Here σ ranges over the choices of q distinct values from $\{1, 2, \ldots, m\}$.

Now

$$A^{\wedge q} v_{\sigma(1)} \wedge \cdots \wedge v_{\sigma(q)} = \lambda_{\sigma(1)} \ldots \lambda_{\sigma(q)} A^{\wedge q} v_{\sigma(1)} \wedge \cdots \wedge v_{\sigma(q)},$$

so we have found the eigenvalues and eigenvectors of $A^{\wedge q}$.

Next we'll proceed by invoking continuity. If A is an arbitrary $m \times m$ matrix, we still find the above eigenvalues, but there might be others. We'll show that this is not the case. We may write A in its Jordan normal form as $A = PJP^{-1}$ where P is a change of basis and J is an upper triangular matrix. The mapping $B \mapsto B^{\wedge q}$ ($B \in \mathbf{R}^{m \times m}$) is continuous and the eigenvalues of $B^{\wedge q}$ also depend continuously on $B^{\wedge q}$. Now suppose that $A^{\wedge q}$ had an eigenvalue that isn't of the form

$$\prod_{k=1}^{q} \lambda_{\sigma(k)},$$

and call it ξ. Let

$$\epsilon = \frac{1}{2} \min_{\sigma} |\xi - \prod_{k=1}^{q} \lambda_{\sigma(k)}|.$$

Because of continuity (See A.8.) in every neighbourhood of J there exists a matrix J' such that

$$A' = PJ'P^{-1}$$

has distinct eigenvalues and there is an eigenvalue of $A'^{\wedge q}$ within a ball of radius ϵ of ξ. This is a contradiction since if J' is close enough to J, all eigenvalues of $A'\wedge q$ are closer than ϵ to some eigenvalue of form

$$\prod_{k=1}^{q} \lambda_{\sigma(k)}.$$

■

Theorem B.4. *Let A be a $m \times m$ matrix. Then*

1. $(A^{\wedge q})^* = (A^*)^{\wedge q}$ *and*
2. *if A is a diagonal and non-negative then $\sqrt{A}^{\wedge q} = \sqrt{A^{\wedge q}}$.* □

Proof. Let's start by proving the 1. claim.

Write the $\binom{m}{q} \times \binom{m}{q}$ matrix $A^{\wedge q}$ in the basis $(d_j), 1 \le j \le \binom{m}{q}$, given by elements of form $e_{i_1} \wedge \cdots \wedge e_{i_q}, 1 \le i_1 < \cdots < i_q \le m$. Then if we write B^j for the row and B_j for the column of an arbitrary matrix B, we have

$$(A^*)_j^{\wedge q} = (A^*)^{\wedge q} d_j = [A^* e_{i_1} \wedge \cdots \wedge A^* e_{i_q}] = [A^{i_1} \wedge \cdots \wedge A^{i_q}],$$

where $[\dots]$ denotes the coordinates with respect to the standard basis. Now

$$[A^{i_1} \wedge \cdots \wedge A^{i_q}] = [\sum_{j_1,\dots,j_q} A^{i_1}_{j_1} \dots A^{i_q}_{j_q} e_{j_1} \wedge \cdots \wedge e_{j_q}]$$

and its coordinate w.r.t. the basis element $e_{k_1} \wedge \cdots \wedge e_{k_q}$ is

$$\sum_{(j_1,\dots,j_q) \in S(k_1,\dots,k_q)} \sigma(J) A^{i_1}_{j_1} \dots A^{i_q}_{j_q}.$$

Here the sum is over permutations of $(k_1, \dots, k_q)$ and $\sigma(J)$ is the sign of the permutation.

On the other hand we also have

$$(A^{\wedge q})^* d_j = (A^{\wedge q})^*_j = (A^{\wedge q})^j = \begin{bmatrix} (A^{\wedge q} d_1)^j \\ \vdots \\ (A^{\wedge q} d_m)^j \end{bmatrix},$$

the coordinate of which w.r.t. the basis element $e_{k_1} \wedge \cdots \wedge e_{k_q}$ is

$$\begin{aligned}
[Ae_{k_1} \wedge \cdots \wedge Ae_{k_q}]^j &= [(A^1_{k_1} e_1 + \cdots + A^m_{k_1} e_m) \wedge \cdots \wedge (A^1_{k_q} e_1 + \cdots + A^m_{k_q} e_m)]^j \\
&= [\sum_{j_1,\dots,j_q} A^{j_1}_{k_1} \dots A^{j_q}_{k_q} e_{j_1} \wedge \cdots \wedge e_{j_q}]^j \\
&= \sum_{j_1,\dots,j_q \in S(i_1,\dots,i_q)} \sigma(J) A^{j_1}_{k_1} \dots A^{j_q}_{k_q}.
\end{aligned}$$

This is easily seen to be equal to the sum given above.

Now let's prove the second part. If A is diagonal, then the column corresponding to the jth basis vector $e_{i_1} \wedge \cdots \wedge e_{i_q}$ in the matrix $A^{\wedge q}$ is

$$[Ae_{i_1} \wedge \cdots \wedge Ae_{i_q}] = a_{i_1} \ldots a_{i_q} d_j,$$

where $a_1, \ldots, a_m$ are the diagonal elements of A. In the same way the column corresponding to the jth basis vector in the matrix $\sqrt{A}^{\wedge q}$ is

$$\sqrt{a_{i_1} \ldots a_{i_q}} d_j,$$

which is the same as the column in the vector $\sqrt{A^{\wedge q}}$. ∎

Definition B.5. A symmetric $n \times n$ matrix A is **positive-semidefinite** if and only if $v^* A v \geq 0$ for all $v \in \mathbf{R}^n$. □

Theorem B.6. *A symmetric $n \times n$ matrix A is positive-semidefinite if and only if the eigenvalues of A are all non-negative.* □

Proof. Since A is symmetric, it may be diagonalized as $A = PDP^*$, where $P = [P_1, \ldots, P_n]$ is an orthogonal matrix consisting of the eigenvectors of A as columns and $D = \text{diag}(d_1, \ldots, d_n)$ is a diagonal matrix with the corresponding eigenvalues on the diagonal. First if D has non-negative entries, then clearly

$$v^* PDP^* v = (P^* v)^* DP^* v \geq 0,$$

so that A is positive-semidefinite. On the other hand if $v^* PDP^* v \geq 0$ for all $v \in \mathbf{R}^n$, then letting $v = P_k$ for $1 \leq k \leq n$ we get

$$0 \leq P_k^* PDP^* P_k = e_k^* D e_k = e_k^* d_k e_k = d_k,$$

so that D has non-negative entries. ∎

Theorem B.7. *Let A be a symmetric positive-semidefinite matrix. Then there exists a positive-semidefinite matrix $A^{1/2}$ such that $A^{1/2} A^{1/2} = A$.* □

Proof. Since A is symmetric and positive-semidefinite, it is diagonalizable as $A = PDP^*$ with a diagonal matrix $D = \text{diag}(d_1, \ldots, d_n)$ with non-negative entries. (By B.6.) We get the result by letting

$$A^{1/2} = P \,\text{diag}(\sqrt{d_1}, \ldots, \sqrt{d_n}) P^{-1}.$$ ∎

Theorem B.8. *Let A be a $n \times n$ real matrix. Then $A^* A$ is symmetric and positive-semidefinite and $\|(A^* A)^{1/2} v\| = \|Av\|$ for all $v \in \mathbf{R}^n$. Moreover the eigenspaces of $(A^* A)^{1/2}$ are orthogonal.* □

Proof. It is well-known that for $n \times n$ matrices A and B we have $(AB)^* = B^* A^*$ and hence $A^* A$ is symmetric. $A^* A$ is positive-semidefinite since $v^* A^* A v = (Av)^* (Av)$.

Furthermore if we let $A^*A = PDP^*$ as in B.7, then

$$\|(A^*A)^{1/2}v\| = \|P\sqrt{D}P^{-1}v\| = \sqrt{v^*P\sqrt{D}P^*P\sqrt{D}P^*v}$$
$$= \sqrt{v^*PDP^*v} = \sqrt{v^*A^*Av} = \|Av\|.$$

The orthogonality of the eigenspaces follows from the factorization

$$(A^*A)^{1/2} = P\sqrt{D}P^*,$$

since it is well-known that the columns of P are eigenvectors of $(A^*A)^{1/2}$ and they are orthogonal. ■

Theorem B.9. *Let A be a $n \times n$ matrix and $q \in \{1, \dots, n\}$. Then*

$$\|A^{\wedge q}\| = \lambda_n \lambda_{n-1} \dots \lambda_{n-q+1},$$

where $\lambda_1 \le \dots \le \lambda_n$ are the eigenvalues of $\sqrt{A^*A}$. □

Proof. Notice first that if we write $A^*A = PDP^*$ as in the previous theorem, then for all basis vectors $v_1 \wedge \dots \wedge v_q$ we have

$$\sqrt{A^*A}^{\wedge q}(v_1 \wedge \dots \wedge v_q) = (P\sqrt{D}P^*)v_1 \wedge \dots \wedge (P\sqrt{D}P^*)v_q$$
$$= P^{\wedge q}\sqrt{D^{\wedge q}}(P^{\wedge q})^*(v_1 \wedge \dots \wedge v_q)$$
$$= \sqrt{(A^{\wedge q})^*A^{\wedge q}}(v_1 \wedge \dots \wedge v_q).$$

Thus $\sqrt{A^*A}^{\wedge q} = \sqrt{(A^{\wedge q})^*A^{\wedge q}}$ and $\sqrt{(A^{\wedge q})^*A^{\wedge q}}$ has the eigenvalues $\lambda_{i_1} \dots \lambda_{i_q}$ where $1 \le i_1 < \dots < i_q \le n$. The largest of these is $\lambda_n \dots \lambda_{n-q+1}$ and the result follows since $\|A^{\wedge q}\| = \|\sqrt{(A^{\wedge q})^*A^{\wedge q}}\|$. ■

Definition B.10. We let $\operatorname{proj}(v, U)$ denote the orthogonal projection of a vector v onto a subspace U. □

Theorem B.11. *Suppose W is a subspace of a vector space V. Then the norm of the orthogonal projection of $u \in V$ with $\|u\| = 1$ onto W is given by*

$$\|\operatorname{proj}(u, W)\| = \sup\{|\langle u, w\rangle| : w \in W, \|w\| = 1\}.$$

Moreover, if $W \subset W'$ and W' is a subspace, then for every $u \in U$ we have

$$\|\operatorname{proj}(u, W)\| \le \|\operatorname{proj}(u, W')\|.$$ □

Proof. Choose an orthonormal basis $v_1, \dots, v_n$ of V, where $v_1, \dots, v_m$ constitute an orthonormal basis of W. The projection of u in W is given by the formula

$$\sum_{k=1}^{m} \langle v_k, u\rangle v_k$$

and its norm is given by

$$\sqrt{\sum_{k=1}^{m} \langle v_k, u \rangle^2}.$$

Now every $u \in V$ can be written as $u = a_1 v_1 + \cdots + a_n v_n$ and every $w \in W$ can be written as $w = b_1 v_1 + \cdots + b_m v_m$. If we assume $\|u\| = \|w\| = 1$, we get also that $a_1^2 + \cdots + a_n^2 = b_1^2 + \cdots + b_m^2 = 1$. Hence we have

$$\sum_{k=1}^{m} \langle v_k, u \rangle^2 \geq |\langle w, u \rangle|^2 \Leftrightarrow \sum_{k=1}^{m} a_k^2 \geq |\sum_{k=1}^{m} a_k b_k|^2,$$

which follows from the Cauchy-Schwarz inequality. Thus

$$\|\operatorname{proj}(u, W)\| \geq \sup\{|\langle u, w \rangle : w \in W, \|w\| = 1\}.$$

On the other hand, by setting

$$w = \frac{\sum_{k=1}^{m} \langle v_k, u \rangle v_k}{\|\sum_{k=1}^{m} \langle v_k, u \rangle v_k\|}$$

we get

$$|\langle u, w \rangle| = \frac{\sum_{k=1}^{m} \langle v_k, u \rangle^2}{\sqrt{\sum_{k=1}^{m} \langle v_k, u \rangle^2}} = \sqrt{\sum_{k=1}^{m} \langle v_k, u \rangle^2},$$

so that

$$\|\operatorname{proj}(u, W)\| \leq \sup\{|\langle u, w \rangle : w \in W, \|w\| = 1\}.$$

Suppose now that $W \subset W'$. Then

$$\begin{aligned}\|\operatorname{proj}(u, W)\| &= \sup\{|\langle u, w \rangle| : w \in W, \|w\| = 1\} \\ &\leq \sup\{|\langle u, w' \rangle| : w' \in W', \|w'\| = 1\} = \|\operatorname{proj}(u, W')\|. \quad \blacksquare\end{aligned}$$

We'd like to also state the following geometric inequality here.

Lemma B.12. *Let $\mathcal{H}$ be a Hilbert space and let $v, w \in \mathcal{H}$, $\alpha \in \mathcal{R}$ be arbitrary. Then we have*

$$\|w + \alpha v\| \geq \|w\| \sin \sphericalangle(v, w)$$

and the equality occurs exactly when $\alpha = -\frac{\langle v, w \rangle}{\|v\|^2}$. □

Proof. Consider the triangle $(0, w, w + \alpha v)$. By the law of sines we have

$$\frac{\|w + \alpha v\|}{\sin \sphericalangle(v, w)} = \frac{\|w\|}{\sin \sphericalangle(v, w + \alpha v)} \geq \|w\|,$$

with equality exactly when $\sphericalangle(v, w + \alpha v) = \frac{\pi}{2}$, i.e. when

$$\langle v, w + \alpha v \rangle = 0 \Leftrightarrow \langle v, w \rangle + \alpha \|v\|^2 = 0. \quad \blacksquare$$

C. Grassmannians

Definition C.1. The Grassmannian $\mathbf{Gr}(d, \mathbf{R}^n)$ consists of the d-dimensional linear subspaces of $\mathbf{R}^n$. We make the Grassmannian a metric space by giving it a metric d defined by $d(U, V) = \| \operatorname{proj}_U - \operatorname{proj}_V \|$ where proj_U and proj_V are the orthogonal projections on the corresponding subspaces U and V and $\| \cdot \|$ is the operator norm. □

The next lemma will be needed in proving the subsequent theorem, which gives us another way of expressing the distance between two subspaces in a Grassmannian.

Lemma C.2. *If U and V are subspaces of $\mathbf{R}^n$ with the same dimension, then*

$$\sup\{|\langle u^\perp, v\rangle| : u^\perp \in U^\perp, v \in V, \|u^\perp\| = \|v\| = 1\} =$$
$$\sup\{|\langle u, v^\perp\rangle| : u \in U, v^\perp \in V^\perp, \|u\| = \|v^\perp\| = 1\}.$$ □

Proof. Notice that it is enough to find an orthogonal linear transformation $A\colon \mathbf{R}^n \to \mathbf{R}^n$ which satisfies $AU = V$ and $A^2 = I$. Namely if we have such an operator, then $A^T = A$ and $\langle Av^\perp, u\rangle = \langle v^\perp, Au\rangle = 0$ for all $v^\perp \in V^\perp$ and $u \in U$, so that $AV^\perp = U^\perp$. Hence for all $u \in U$ and $v^\perp \in V^\perp$ we have

$$\langle u, v^\perp\rangle = \langle Au, Av^\perp\rangle = \langle v, u^\perp\rangle,$$

where $v := Au \in V$ and $u^\perp := Av^\perp \in U^\perp$, so that the two supremums in the statement of the lemma will be equal.

Let's now construct such a transformation A. We may assume that U and V span the whole space since if this is not the case, we can set A to be the identity in the orthocomplement of their span. We note that it's enough to find orthonormal bases $u_1, \dots, u_k$ for U and $v_1, \dots, v_k$ for V such that $\langle u_i, v_j\rangle = 0$ when $i \neq j$ and such that $u_i = v_i \in U \cap V$ for $1 \leq i \leq m$ where m is the dimension of $U \cap V$. Now we can write $\mathbf{R}^n$ as the direct sum

$$(U \cap V) \oplus \operatorname{span}(u_{m+1}, v_{m+1}) \oplus \cdots \oplus \operatorname{span}(u_k, v_k)$$

and define A to be identity in $U \cap V$ and set $Au_i = v_i$, $Av_i = u_i$ in $\operatorname{span}(u_i, v_i)$, $m + 1 \leq i \leq k$. Then A clearly satisfies $A^2 = I$ and $AU = V$. Moreover A is orthogonal since the spaces in the direct sum are pairwise orthogonal and A is orthogonal in each of the summands.

It remains to find the two bases. Consider the operators

$$A = \operatorname{proj}_U \operatorname{proj}_V \operatorname{proj}_U \quad \text{and} \quad B = \operatorname{proj}_V \operatorname{proj}_U \operatorname{proj}_V .$$

Because orthogonal projections are self-adjoint, also $A^* = A$ and $B^* = B$. This implies that A and B are orthogonally diagonalizable. Notice first that the eigenspace corresponding to the eigenvalue 0 contains $U^\perp$. Because eigenvectors corresponding to different eigenvalues are orthogonal, we may choose orthonormal eigenvectors $u_1, \dots, u_k \in U$ that form a basis for U. Let λ_i, $1 \le i \le k$ denote the corresponding eigenvalues. We may assume that

$$\{i \in \{1, \dots, k\} : \operatorname{proj}_V u_i = 0\} = \{1, \dots, l\}$$

for some $0 \le l \le k$. We see that for all $1 \le i \le k$,

$$B \operatorname{proj}_V u_i = B \operatorname{proj}_V \operatorname{proj}_U u_i = \lambda_i \operatorname{proj}_V u_i,$$

so that $\operatorname{proj}_V u_i$ is an eigenvector of B with eigenvalue λ_i. Moreover because proj_V and proj_U are self-adjoint,

$$\langle \operatorname{proj}_V u_i, \operatorname{proj}_V u_j \rangle = \langle u_i, \operatorname{proj}_U \operatorname{proj}_V \operatorname{proj}_U u_j \rangle = \langle u_i \lambda_j u_j \rangle = 0$$

if $i \ne j$. Hence we can choose an orthonormal basis for V of form

$$v_1, \dots, v_l, v_{l+1} = \frac{\operatorname{proj}_V u_{l+1}}{\| \operatorname{proj}_V u_{l+1} \|}, \dots, v_k = \frac{\operatorname{proj}_V u_k}{\| \operatorname{proj}_V u_k \|},$$

where $v_1, \dots, v_l$ are eigenvectors of B. (Like before for A, we utilize here the fact that the eigenspaces of B that correspond to nonzero eigenvalues are contained in V.) Let's check that $\langle u_i, v_j \rangle = 0$ if $i \ne j$. Assume first that $1 \le i \le l$. Then

$$\langle u_i, v_j \rangle = \langle \operatorname{proj}_V u_i, v_j \rangle = 0.$$

If $l + 1 \le i \le k$, then

$$\langle u_i, v_j \rangle = \langle v_i \| \operatorname{proj}_V u_i \|, v_j \rangle = 0.$$

Finally we'll have to check that some of the vectors u_j form a basis for $U \cap V$. So suppose that

$$w = c_1 u_1 + \cdots + c_k u_k = d_1 v_1 + \cdots + d_k v_k$$

is a vector in $U \cap V$. Then if we take the inner product of this vector with u_i, $1 \le i \le l$, we see that $c_i = 0$. Similarly if we take the inner product with v_i, $1 \le i \le l$, we see that $d_i = 0$ and thus

$$w = c_{l+1} u_{l+1} + \cdots + c_k u_k = d_{l+1} v_{l+1} + \cdots + d_k v_k.$$

Now apply proj_V to both sides to get

$$c_{l+1} v_{l+1} \| \operatorname{proj}_V u_{l+1} \| + \cdots + c_k v_k \| \operatorname{proj}_V u_k \| = d_{l+1} v_{l+1} + \cdots + d_k v_k,$$

which implies that $c_{l+1} \| \operatorname{proj}_V u_{l+1} \| = d_{l+1}$. On the other hand if we apply proj_U, we get

$$c_{l+1} u_{l+1} + \cdots + c_k u_k = \frac{d_{l+1}}{\| \operatorname{proj}_V u_{l+1} \|} \lambda_{l+1} u_{l+1} + \cdots + \frac{d_k}{\| \operatorname{proj}_V u_{l+1} \|} \lambda_k u_k.$$

This gives us $d_i = \lambda_i d_i$ for $l+1 \le i \le k$. Hence either $d_i = 0$ or $\lambda_i = 1$. If $\lambda_i = 1$, then

$$\| \operatorname{proj}_V u_i \|^2 = \lambda_i \langle u_i, u_i \rangle = \|u_i\|^2,$$

and hence $u_i \in V$. Because w was arbitrary, this implies that some of the u_i form a basis for $U \cap V$ and we have $v_i = u_i$ for these i. ∎

Theorem C.3. *For all $U, V \in \mathbf{Gr}(d, \mathbf{R}^n)$ we have*

$$\begin{aligned} d(U,V) &= \sup\{|\langle u, v\rangle| : u \in U, v \in V^\perp, \|u\| = \|v\| = 1\} \\ &= \sup\{|\langle u, v\rangle| : u \in U^\perp, v \in V, \|u\| = \|v\| = 1\}. \end{aligned}$$ □

Proof. By definition

$$d(U,V) = \| \operatorname{proj}_U - \operatorname{proj}_V \| = \sup_{\|v\|=1} \| \operatorname{proj}_U(v) - \operatorname{proj}_V(v)\|.$$

If we now restrict us to look at only those $v^\perp$ that reside in $V^\perp$, we see that

$$\begin{aligned} d(U,V) &\ge \sup_{v^\perp \in V^\perp, \|v^\perp\|=1} \| \operatorname{proj}_U(v^\perp)\| \\ &= \sup\{|\langle u, v^\perp\rangle| : u \in U, v^\perp \in V^\perp, \|u\| = \|v^\perp\| = 1\}. \end{aligned}$$

It remains to show the opposite inequality. Suppose that $w \in \mathbf{R}^n$ and write $w = v + v^\perp$, where $v \in V$ and $v^\perp \in V^\perp$. Then

$$\begin{aligned} \| \operatorname{proj}_U(w) - \operatorname{proj}_V(w)\|^2 &= \| \operatorname{proj}_U(v) - v + \operatorname{proj}_U(v^\perp)\|^2 \\ &= \| \operatorname{proj}_U(v) - v\|^2 + \| \operatorname{proj}_U(v^\perp)\|^2, \end{aligned}$$

since $\operatorname{proj}_U(v) - v \in U^\perp$. Moreover, we have $\operatorname{proj}_U(v) - v = -\operatorname{proj}_{U^\perp}(v)$ and thus

$$\begin{aligned} &\| \operatorname{proj}_U(v) - v\|^2 + \| \operatorname{proj}_U(v^\perp)\|^2 = \\ &\|v\|^2 \sup\{|\langle u^\perp, v/\|v\|\rangle|^2 : u \in U^\perp, \|u\| = 1\} + \\ &\|v^\perp\|^2 \sup\{|\langle u, v^\perp/\|v^\perp\|\rangle|^2 : u \in U, \|u\| = 1\}. \end{aligned}$$

Taking sup over $\|w\| = 1$ we see that

$$\begin{aligned} &\sup_{\|w\|=1} \| \operatorname{proj}_U(w) - \operatorname{proj}_V(w)\|^2 \le \\ &\sup_{\|w\|=1} \Big(\|v\|^2 \sup\{|\langle u^\perp, v'\rangle| : u \in U^\perp, v' \in V, \|u\| = \|v'\| = 1\} + \\ &\qquad \|v^\perp\|^2 \sup\{|\langle u, v'\rangle| : u \in U, v' \in V^\perp, \|u\| = \|v'\| = 1\}\Big) \end{aligned}$$

and by the previous lemma this becomes

$$\begin{aligned} &\sup\{|\langle u^\perp, v'\rangle| : u \in U^\perp, v' \in V, \|u\| = \|v'\| = 1\} \sup_{\|w\|=1} (|v|^2 + |v^\perp|^2) = \\ &\sup\{|\langle u^\perp, v'\rangle| : u \in U^\perp, v' \in V, \|u\| = \|v'\| = 1\}. \end{aligned}$$ ∎

Theorem C.4. *The space* $\mathbf{Gr}(d, \mathbf{R}^m)$ *is compact (and hence complete).* □

Proof. Let U_n, $n = 1, 2, \ldots$, be a sequence of d-dimensional subspaces of $\mathbf{R}^m$. For each $n \in \mathbf{N}$ choose an orthonormal basis of U_n and denote it by

$$b_n^{(1)}, \ldots, b_n^{(d)}.$$

The set of d-tuples of orthonormal vectors in $\mathbf{R}^m$ can be identified with a compact subset of $\mathbf{R}^{m\times d}$ and hence there is a subsequence n_k so that

$$b_{n_k}^{(j)} \to b^{(j)}$$

for all $j \in \{1, 2, \ldots, d\}$. By continuity $b^{(1)}, \ldots, b^{(d)}$ are orthonormal and hence form a basis for some d-dimensional subspace U. We claim that $U_{n_k} \to U$ as $k \to \infty$. In other words we have to prove that

$$\| \operatorname{proj}_U - \operatorname{proj}_{U_{n_k}} \| \to 0, \quad \text{as } k \to \infty.$$

Suppose thus that $v \in \mathbf{R}^n$ is such that $\|v\| = 1$. Then

$$\begin{aligned}
&\| \operatorname{proj}_U(v) - \operatorname{proj}_{U_{n_k}}(v)\| = \\
&\| \sum_j \left\langle v, b_{n_k}^{(i)} \right\rangle b_{n_k}^{(j)} - \sum_j \left\langle v, b^{(j)} \right\rangle b^{(j)} \| = \\
&\| \sum_j \left(\left\langle v, b^{(j)} \right\rangle (b_{n_k}^{(j)} - b^{(j)}) + \left\langle v, b_{n_k}^{(j)} - b^{(j)} \right\rangle b_{n_k}^{(j)} \right) \| \le \\
&\sum_j \|b_{n_k}^{(j)} - b^{(j)}\| + \sum_j \|b_{n_k}^{(j)} - b^{(j)}\| = \\
&2 \sum_j \|b_{n_k}^{(j)} - b^{(j)}\| \to 0,
\end{aligned}$$

and the convergence is uniform with respect to v since the last formula doesn't depend on v. Thus $\| \operatorname{proj}_U - \operatorname{proj}_{U_{n_k}} \| \to 0$. ■

Bibliography

[Boc08] Jairo Bochi.
"The multiplicative theorem of Oseledets".
[A note by Jairo Bochi on Oseledets' Multiplicative Ergodic Theorem. Available from his website at http://www.mat.puc-rio.br/ jairo/].
2008.

[BP99] L. Barreira and Y. Pesin.
"Lectures on Lyapunov exponents and smooth ergodic theory".
In: *Smooth Ergodic Theory and Its Applications: Proceedings of the AMS Summer Research Institute on Smooth Ergodic Theory and Its Applications, July 26-August 13, 1999, University of Washington, Seattle.*
Volume 69.
Amer Mathematical Society. 1999,
Page 3.

[Die60] J.A. Dieudonné.
Foundations of modern analysis.
Volume 286.
Academic Press New York,, London, 1960.

[EW11] Manfred Einsiedler and Thomas Ward.
Ergodic Theory with a view towards Number Theory.
Springer London, 2011.

[Fal97] Kenneth Falconer.
Techniques in fractal geometry.
Chichester: John Wiley & Sons Ltd., 1997.

[Kin68] J. F. C. Kingman.
"The Ergodic Theory of Subadditive Stochastic Processes".
In: *Journal of the Royal Statistical Society. Series B (Methodological)* 30.3 (1968), pages 499–510.

[Par67] K.R. Parthasarathy.
Probability measures on metric spaces.
Chelsea Pub Co, 1967.

[Pes76] J.B. Pesin.
"Families of invariant manifolds corresponding to nonzero characteristic exponents".
In: *Mathematics of the USSR-Izvestiya* 10 (1976), page 1261.

[Pes77] Y.B. Pesin.
"Characteristic Lyapunov exponents and smooth ergodic theory".

In: *Russian Mathematical Surveys* 32.4 (1977), pages 55–114.

[Phe01] R.R. Phelps.
Lectures on Choquet's theorem.
Volume 1757.
Springer Verlag, 2001.

[Pug84] C.C. Pugh.
"The C 1+ α hypothesis in Pesin theory".
In: *Publications Mathématiques de l'IHÉS* 59.1 (1984), pages 143–161.

[Rag79] M. Raghunathan.
"A proof of Oseledec's multiplicative ergodic theorem".
In: *Israel Journal of Mathematics* 32.4 (1979), pages 356–362.

[Rud87] Walter Rudin.
Real and complex analysis.
3rd edition.
New York: McGraw-Hill Book Co., 1987.

[Rud91] Walter Rudin.
Functional analysis.
2nd edition.
International Series in Pure and Applied Mathematics.
New York: McGraw-Hill Inc., 1991.

[Rue79] David Ruelle.
"Ergodic theory of differentiable dynamical systems".
In: *Publications Mathématiques de L'IHÉS* 50.1 (1979), pages 27–58.

[Ste89] J. Michael Steele.
"Kingman's subadditive ergodic theorem".
In: *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques* 25.1 (1989), pages 93–98.

[Wey16] Hermann Weyl.
"Über die Gleichverteilung von Zahlen mod. Eins".
In: *Mathematische Annalen* 77.3 (1916), pages 313–352.

[Mic06] Michael, Keane and Karl Petersen.
"Easy and nearly simultaneous proofs of the Ergodic Theorem and Maximal Ergodic Theorem".
In: *IMS Lecture Notes–Monograph Series* 48 (2006).