

## CONTENTS

<b>Introduction</b>	<b>ix</b>
<b>Preliminaries</b>	<b>xi</b>
1. Conventions	xi
2. Notation	xi
3. Pre-requisites in point set topology (Chapters 1-6)	xi
4. Pre-requisites in measure theory (Chapters 7-12)	xii
5. Subadditive sequences	xiii
6. References	xiii
<b>Chapter 1. Examples and basic properties</b>	<b>1</b>
1.1. Examples	1
1.2. Transitivity	2
1.3. Other characterizations of transitivity	4
1.4. Transitivity for subshifts of finite type	4
1.5. Minimality and the Birkhoff recurrence theorem	6
1.6. Commuting homeomorphisms	8
1.7. Comments and references	8
<b>Chapter 2. An application of recurrence to arithmetic progressions</b>	<b>11</b>
2.1. Van der Waerden's theorem	11
2.2. A dynamical proof	12
2.3. The proofs of Sulemma 2.2.2 and Sublemma 2.2.3	15
2.4. Comments and references	17
<b>Chapter 3. Topological entropy</b>	<b>19</b>
3.1. Definitions	19
3.2. The Perron-Frobenius theorem and subshifts of finite type	22
3.3. Other definitions and examples	26
3.4. Conjugacy	30
3.5. Comments and references	32
<b>Chapter 4. Interval maps</b>	<b>33</b>
4.1. Fixed points and periodic points	33
4.2. Topological entropy of interval maps	37
4.3. Markov maps	39

4.4. Comments and references	44
<b>Chapter 5. Hyperbolic toral automorphisms</b>	<b>47</b>
5.1. Definitions	47
5.2. Entropy for Hyperbolic Toral Automorphisms	49
5.3. Shadowing and semi-conjugacy	52
5.4. Comments and references	55
<b>Chapter 6. Rotation numbers</b>	<b>57</b>
6.1. Homeomorphisms of the circle and rotation numbers	57
6.2. Denjoy's theorem	60
6.3. Comments and references	63
<b>Chapter 7. Invariant measures</b>	<b>65</b>
7.1. Definitions and characterization of invariant measures	65
7.2. Borel sigma-algebras for compact metric spaces	65
7.3. Examples of invariant measures	67
7.4. Invariant measures for other actions	69
7.5. Comments and references	71
<b>Chapter 8. Measure theoretic entropy</b>	<b>73</b>
8.1. Partitions and conditional expectations	73
8.2. The entropy of a partition	76
8.3. The entropy of a transformation	78
8.4. The increasing martingale theorem	81
8.5. Entropy and sigma algebras	83
8.6. Conditional entropy	85
8.7. Proofs of Lemma 8.7 and Lemma 8.8	87
8.8. Isomorphism	88
8.9. Comments and references	89
<b>Chapter 9. Ergodic measures</b>	<b>91</b>
9.1. Definitions and characterization of ergodic measures	91
9.2. Poincaré recurrence and Kac's theorem	91
9.3. Existence of ergodic measures	93
9.4. Some basic constructions in ergodic theory	94
9.4.1. Skew products	95
9.4.2. Induced transformations and Rohlin towers	95
9.4.3. Natural extensions	96
9.5. Comments and references	97
<b>Chapter 10. Ergodic theorems</b>	<b>99</b>
10.1. The Von Neumann ergodic Theorem	99
10.2. The Birkhoff theorem (for ergodic measures)	102
10.3. Applications of the ergodic theorems	106
10.4. The Birkhoff theorem (for invariant measures)	111
10.5. Comments and references	112
<b>Chapter 11. Mixing</b>	<b>113</b>

11.1. Weak mixing	113
11.2. A density one convergence characterization of weak mixing	114
11.3. A generalization of the Von Neumann Ergodic Theorem	116
11.4. The spectral viewpoint	118
11.5. Spectral characterization of weak mixing	120
11.6. Strong mixing	122
11.7. Comments and reference	123
<b>Chapter 12. Statistical properties in ergodic theory</b>	<b>125</b>
12.1. Exact endomorphisms	125
12.2. Statistical properties of expanding Markov maps	126
12.3. Rohlin's entropy formula	132
12.4. The Shannon-McMillan-Breiman theorem	134
12.5. Comments and references	136
<b>Chapter 13. Fixed points for homeomorphisms of the annulus</b>	<b>137</b>
13.1. Fixed points for the annulus	137
13.2. Outline proof of Brouwer's plane translation theorem	141
13.3. Comments and references	143
<b>Chapter 14. The variational principle</b>	<b>145</b>
14.1. The variational principle for entropy	145
14.2. The proof of the variational principle	145
14.3. Comments and references	149
<b>Chapter 15. Invariant measures for commuting transformations</b>	<b>151</b>
15.1. Furstenburg's conjecture and Rudolph's theorem	151
15.2. The proof of Rudolph's theorem	151
15.3. Comments and references	156
<b>Chapter 16. Multiple recurrence and Szemerédi's theorem</b>	<b>159</b>
16.1. Szemerédi's theorem on arithmetic progressions	159
16.2. An ergodic proof of Szemerédi's theorem	160
16.3. The proof of Theorem 16.2	161
16.3.1.(UMR) for weak-mixing systems, weak-mixing extensions and compact systems	161
16.3.2.The non-weak-mixing case	162
16.3.3.(UMR) for compact extensions	163
16.3.4.The last step	163
16.4. Appendix to section 16.3	164
16.4.1.The proofs of Propositions 16.3 and 16.4	164
16.4.2.The proof of Proposition 16.5	168
16.4.3.The proof of Proposition 16.6	169

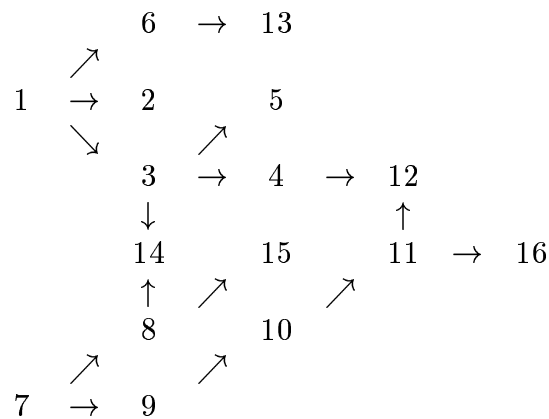
16.4.4. The proof of Proposition 16.7	170
16.4.5. The proof of Proposition 16.8	171
16.4.6. The proof of Proposition 16.9	173
16.5. Comments and references	152
<b>Index</b>	<b>177</b>

## INTRODUCTION

This book is intended as an introduction to both dynamical systems and ergodic theory. Our aim is to give a direct and detailed introduction to the basic theory, suitable as a text for advanced undergraduate students or beginning graduate students in mathematics.

The notes divide naturally into three parts. The first part (chapters 1-6) concentrates on topological dynamics. The second part (chapters 7-12) deals with ergodic theory and measurable dynamics. The third part (chapters 13-16) consists of more advanced material to supplement the two earlier parts.

Each of the first two parts is intended to be essentially self-contained, as is illustrated by the following diagram of the relationships between chapters:



The areas of dynamical systems and ergodic theory are rich in connections with other subjects (e.g. number theory, geometry, statistics, mathematical physics, biology, etc.). In the course of these notes we have tried to motivate the general theory with some important applications (particularly to number theory, in chapter 2 and chapter 16).

There are already a number of excellent books on dynamical systems and ergodic theory (e.g. Devaney's *Introduction to Chaotic Dynamical Systems*) Walters' *Introduction to Ergodic Theory* and *Introduction to the Modern Theory of Dynamical Systems* by Katok and Hasselblatt).

To pre-empt any comparison with these fine texts, we should emphasize that this book is intended to be a more modest introduction to the subject. The reader who would like to find out more about dynamical systems and ergodic theory will find much more in these books.

Mark Pollicott  
Department of Mathematics  
Manchester University

Michiko Yuri  
Department of Business Administration  
Sapporo University

## PRELIMINARIES

**1. Conventions.** The book is divided into 16 chapters, each subdivided into sections numbered in order (e.g. chapter 12, section 3 is numbered **12.3**).

Within each chapter results (Theorems, Propositions or Lemmas) are labelled by the chapter and then the order of occurrence (e.g. the fifth result in chapter 3 is **Proposition 3.5**). The exceptions to this rule are: sublemmas which are presented within the context of the proof of a more important result (e.g. the proof of Theorem 2.2 contains Sublemmas 2.2.1 and 2.2.2); and corollaries (the corollary to Theorem 5.5 is Corollary 5.5.1).

We denote the end of a proof by **■**.

Finally, equations are numbered by the chapter and their order of occurrence (e.g. the fourth equation in chapter 5 is labelled (5.4))

**2. Notation.** We shall use the standard notation:  $\mathbb{R}$  to denote the *real numbers*;  $\mathbb{Q}$  to denote the *rational numbers*;  $\mathbb{Z}$  to denote the *integer numbers*;  $\mathbb{N}$  to denote the *natural numbers*; and  $\mathbb{Z}^+$  to denote the non-negative integers. We use the convenient convention that:  $\mathbb{R}/\mathbb{Z} = \{x + \mathbb{Z} : x \in \mathbb{R}\}$  (which is homeomorphic to the standard unit circle);  $\mathbb{R}^2/\mathbb{Z}^2 = \{(x_1, x_2) + \mathbb{Z}^2 : (x_1, x_2) \in \mathbb{R}^2\}$  (which is homeomorphic to the standard 2-torus); etc. However, for  $x \in \mathbb{R}$  we denote the corresponding element in  $\mathbb{R}/\mathbb{Z}$  by  $x \pmod{1}$  (and similarly for  $\mathbb{R}^2/\mathbb{Z}^2$ , etc.).

We denote the interior of a subset  $A$  of a metric space by  $\text{int}(A)$ , and we denote its closure by  $\text{cl}(A)$ .

If  $T : X \rightarrow X$  denotes a continuous map on a compact metric space then  $T^n$  ( $n \geq 1$ ) denotes the composition with itself  $n$  times.

If  $T : I \rightarrow I$  is a  $C^1$  map on the unit interval  $I = [0, 1]$  then  $T'$  denotes its derivative.

**3. Prerequisites in point set topology (chapters 1-6).** The first six chapters consist of various results in topological dynamics for which the only prerequisite is a working knowledge of point set topology for metric spaces. For example:

**THEOREM A (BAIRE).** *Let  $X$  be a compact metric space; then if  $\{U_n\}_{n \in \mathbb{N}}$  is a countable family of open dense sets then  $\bigcap_{n \in \mathbb{N}} U_n \subset X$  is dense.*

**THEOREM B (SEQUENTIAL COMPACTNESS).** *Let  $X$  be a metric space; then  $X$  is compact if and only if every sequence  $(x_n)_{n \in \mathbb{N}}$  in  $X$  contains a convergent subsequence.*

**THEOREM C (ZORN'S LEMMA).** *Let  $Z$  be a set with a partial ordering. If every totally ordered chain has a lower bound in  $Z$  then there is a minimal element in  $Z$ .*

Two good references for this material are [4] and [5]

**4. Pre-requisites in measure theory (chapters 7-12).** Chapters 7-12 form an introduction to ergodic theory, and suppose some familiarity (if not expertise) with abstract measure theory and harmonic analysis. The following results will be required.

**THEOREM D (RIESZ REPRESENTATION).** *There is a bijection between*

- (1) *probability measures  $\mu$  on a compact metric space  $X$  (with the Borel sigma algebra),*
- (2) *Continuous linear functionals  $c : C^0(X) \rightarrow \mathbb{R}$ ,*

*given by  $c(f) = \int f d\mu$ .*

**THEOREM E.** *Let  $(X, \mathcal{B}, \mu)$  be a measure space. For every linear functional  $\alpha : L^1(X, \mathcal{B}, \mu) \rightarrow L^1(X, \mathcal{B}, \mu)$  there exists  $k \in L^\infty(X, \mathcal{B}, \mu)$  such that  $\alpha(f) = \int f \cdot k d\mu, \forall f \in L^1(X, \mathcal{B}, \mu)$  [3, p.121].*

In proving invariance of measures in examples the following basic result will sometimes be assumed.

**THEOREM F (KOLMOGOROV EXTENSION).** *Let  $\mathcal{B}$  be the Borel sigma-algebra for a compact metric space  $X$ . If  $\mu_1$  and  $\mu_2$  are two measures for the Borel sigma-algebra which agree on the open sets of  $X$  then  $\mu_1 = \mu_2$  [3, p. 310].*

The following terminology will be used in the chapter on ergodic measures. Given two probability measures  $\mu, \nu$  we say that  $\mu$  is *absolutely continuous* with respect to  $\nu$  if for every set  $B \in \mathcal{B}$  for which  $\nu(B) = 0$  we have that  $\mu(B) = 0$ . We write  $\mu \ll \nu$  and then we have the following result.

**THEOREM G (RADON-NIKODYM).** *If  $\mu$  is absolutely continuous with respect to  $\nu$  then there exists a (unique) function  $f \in L^1(X, \mathcal{B}, d\nu)$  such that for any  $A \in \mathcal{B}$  we can write  $\mu(A) = \int_A f d\nu$ .*

We usually write  $f = \frac{d\mu}{d\nu}$  and call this the *Radon-Nikodym derivative* of  $\mu$  with respect to  $\nu$ .

We call two measures  $\mu, \nu$  *mutually singular* if there exists a set  $B \in \mathcal{B}$  such that  $\mu(A) = 0$  and  $\nu(A) = 1$ . We then write  $\mu \perp \nu$ .

In chapter 8 we shall need a passing reference to Lebesgue spaces. A *Lebesgue space* is a measure space which is measurably equivalent to the



the union of unit intervals (with the usual Lebesgue measure) with at most countably many points (with non-zero measure).

In chapter 11 we shall use the following result.

**THEOREM H (DOMINATED CONVERGENCE).** *Let  $h \in L^1(X, \mathcal{B}, \mu)$  and let  $(f_n)_{n \in \mathbb{Z}^+} \subset L^1(X, \mathcal{B}, \mu)$ , with  $|f_n(x)| \leq h(x)$ , converge (almost everywhere) to  $f(x)$ ; then  $\int f_n d\mu \rightarrow \int f d\mu$  as  $n \rightarrow +\infty$ .*

Good general references for this material are [1], [2], [3].

**5. Subadditive sequences.** A simple result which proves its worth several times in these notes is the following.

**THEOREM F (SUBADDITIVE SEQUENCES).** *Let  $(a_n)_{n \in \mathbb{N}}$  be a sequence of real numbers such that  $a_{n+m} \leq a_n + a_m$ ,  $\forall n, m \in \mathbb{N}$  (i.e. a subadditive sequence); then  $a_n \rightarrow a$ , as  $n \rightarrow +\infty$ , where  $a = \inf\{a_n/n: n \geq 1\}$*

**PROOF.** First note that  $a_n \leq a_1 + a_{n-1} \leq \dots \leq na_1$ , and so  $a \leq a_1$ . For  $\epsilon > 0$  we choose  $N > 0$  with  $a_N < N(a + \epsilon)$ . For any  $n \geq 1$  we can write  $n = kN + r$ , where  $k \geq 0$  and  $1 \leq r \leq N - 1$ . Then

$$a_n \leq a_{kN} + a_r \leq ka_N + a_r \leq ka_N + \sup_{1 \leq r \leq N} a_r$$

and we see that

$$\limsup_{n \rightarrow +\infty} \frac{a_n}{n} \leq \limsup_{k \rightarrow +\infty} \frac{ka_N + \sup_{1 \leq r \leq N} a_r}{kN} = \frac{a_N}{N} \leq a + \epsilon.$$

This shows that  $\frac{a_n}{n} \rightarrow a$ , as required. ■

### References

1. P. Halmos, *Measure Theory*, Van Nostrand, Princeton N.J., 1950.
2. K. Partasarathy, *An Introduction to Probability and Measure Theory*, Macmillan, New Delhi, 1977.
3. H. Roydon, *Real Analysis*, Macmillan, New York, 1968.
4. G. Simmons, *Introduction to Topology and Modern Analysis*, McGraw-Hill, New York, 1963.
5. W. Sutherland, *Introduction to Topological and Metric spaces*, Clarendon Press, Oxford, 1975.

## CHAPTER 1

### EXAMPLES AND BASIC PROPERTIES

In this chapter we shall introduce some of the basic dynamical properties associated to continuous maps  $T : X \rightarrow X$  on compact metric spaces.

#### 1.1 Examples

To set the stage, we begin with some standard examples of continuous maps (transformations) which will be used to illustrate different properties.

**EXAMPLE 1 (DOUBLING MAP).** Let  $X$  denote the unit interval with its endpoints identified,  $X = \mathbb{R}/\mathbb{Z}$ . Define a continuous map  $T : X \rightarrow X$  by  $T(x) = 2x \pmod{1}$ , i.e.

$$Tx = \begin{cases} 2x & \text{if } 0 \leq x < \frac{1}{2}, \\ 2x - 1 & \text{if } \frac{1}{2} \leq x \leq 1. \end{cases}$$

This is usually called the “doubling map”, since it doubles distances on  $X$ . An equivalent formulation would be if we let  $X = K := \{z \in \mathbb{C} : |z| = 1\}$  and then define  $T : X \rightarrow X$  by  $T(e^{2\pi i\theta}) = e^{2\pi i2\theta}$ , where  $0 < \theta < 1$ . This is equivalent in the sense that there is a homeomorphism  $\rho : \mathbb{R}/\mathbb{Z} \rightarrow K$  given by  $\rho(\theta + \mathbb{Z}) = e^{2\pi i\theta}$  which relates the two transformations. We shall return to this notion of equivalence (or conjugacy) in chapter 3.

**EXAMPLE 2 (ROTATIONS ON THE CIRCLE).** Let  $X = \mathbb{R}/\mathbb{Z}$  and fix a number  $\alpha \in [0, 1)$ . We define a homeomorphism  $T : X \rightarrow X$  by  $T(x) = x + \alpha \pmod{1}$ , i.e.

$$Tx = \begin{cases} x + \alpha & \text{if } 0 \leq x + \alpha \leq 1, \\ x + \alpha - 1 & \text{if } x + \alpha > 1. \end{cases}$$

(An equivalent formulation would be if we let  $X = K$  and then define  $T : X \rightarrow X$  by  $T(e^{2\pi i\theta}) = e^{2\pi i(\theta + \alpha)}$ . This is equivalent in the sense that the homeomorphism  $\rho : \mathbb{R}/\mathbb{Z} \rightarrow K$  given by  $\rho(t + \mathbb{Z}) = e^{2\pi it}$  relates the two transformations.)

**EXAMPLE 3 (SHIFT MAP).** For  $k \geq 2$  let  $X_k = \prod_{n \in \mathbb{Z}} \{1, 2, \dots, k\}$  denote the space of all sequences taking values  $\{1, 2, \dots, k\}$  indexed by  $\mathbb{Z}$ . In order to define a metric we first associate to two sequences  $x = (x_n)_{n \in \mathbb{Z}}$  and

$y = (y_n)_{n \in \mathbb{Z}}$  an integer  $N(x, y) = \min\{N \geq 0 : x_N \neq y_N \text{ or } x_{-N} \neq y_{-N}\}$ . We define a metric on  $X_k$  by

$$d(x, y) = \begin{cases} \left(\frac{1}{2}\right)^{N(x, y)} & \text{if } x \neq y, \\ 0 & \text{otherwise.} \end{cases}$$

LEMMA 1.1.  $X_k$  is a compact space.

PROOF. We shall actually show that  $X_k$  is sequentially compact. Let  $x^{(m)} = (x_n^{(m)})_{n \in \mathbb{Z}}$  ( $m = 1, 2, 3, \dots$ ) be a sequence in  $X_k$ ; then we need to show that there exist a point  $x \in X_k$  and a sub-sequence  $x^{(m_l)} \rightarrow x$  ( $l = 1, 2, 3, \dots$ ).

First observe that the zeroth terms  $x_0^{(m)}$  ( $m = 1, 2, 3, \dots$ ) must take some value in  $\{1, 2, \dots, k\}$  infinitely often. Choose such an  $x_0 \in \{1, 2, \dots, k\}$  with  $x_0^{(m)} = x_0$ , for infinitely many  $m$ . We continue inductively: For  $l > 0$ , choose  $x_l \in \{1, 2, \dots, k\}$  and  $x_{-l} \in \{1, 2, \dots, k\}$  such that  $x_{-l}^{(m)} = x_{-l}, \dots, x_0^{(m)} = x_0, \dots, x_l^{(m)} = x_l$ , say, for infinitely many  $m$ . Finally, we define  $x = (x_l)_{l \in \mathbb{Z}}$ . For each  $l \geq 0$  we choose  $m_l := m$  such that  $x_{-l}^{(m)} = x_{-l}, \dots, x_0^{(m)} = x_0, \dots, x_l^{(m)} = x_l$ ; then  $d(x^{(m_l)}, x) \leq \frac{1}{2^l}$  and so  $d(x^{(m_l)}, x) \rightarrow 0$  as  $l \rightarrow +\infty$ . ■

DEFINITION. We can define a map  $\sigma : X_k \rightarrow X_k$  by  $(\sigma x)_n = x_{n+1}$ ,  $\forall n \in \mathbb{Z}$ , i.e.

$$\sigma : (\dots, x_{-2}, x_{-1}, x_0, x_1, x_2, \dots) \longmapsto (\dots, x_{-1}, x_0, x_1, x_2, x_3, \dots).$$

Since this map shifts sequences by one place it is called the *shift map*.

LEMMA 1.2. The map  $\sigma : X_k \rightarrow X_k$  is a homeomorphism.

PROOF. To show continuity we observe that if  $x \neq y$  and  $d(x, y) = \left(\frac{1}{2}\right)^N$  then we know that  $x_i = y_i$  for  $-N \leq i \leq N$ . Thus we have that  $(\sigma x)_i = x_{i+1} = y_{i+1} = (\sigma y)_i$  for  $i = -(N+1), \dots, N-1$ . This means that  $d(\sigma x, \sigma y) \leq \left(\frac{1}{2}\right)^{N-1} = \frac{1}{2}d(x, y)$  and we see that  $\sigma$  is continuous.

Clearly  $\sigma : X_k \rightarrow X_k$  is invertible (since the inverse transformation  $\sigma^{-1} : X_k \rightarrow X_k$  simply shifts sequences back one place). Finally, the inverse map  $\sigma^{-1} : X_k \rightarrow X_k$  is continuous by the same sort of argument as above. ■

## 1.2 Transitivity

In this section we shall introduce some basic properties of continuous maps  $T : X \rightarrow X$  on compact metric spaces  $X$ .

DEFINITION. We say that a homeomorphism  $T : X \rightarrow X$  of a compact metric space  $X$  is *transitive* if there exists a point  $x \in X$  such that its orbit  $\{T^n x : n \in \mathbb{Z}\} = \{\dots, T^{-2}x, T^{-1}x, x, Tx, T^2x, \dots\}$  is dense in  $X$ . We call such a point  $x \in X$  a *transitive point*.

We say that a continuous map  $T : X \rightarrow X$  of a compact metric space  $X$  is (*forward*) *transitive* if there exists a point  $x \in X$  such that its *orbit*  $\{T^n x : n \in \mathbb{Z}^+\} = \{x, Tx, T^2x, \dots\}$  is dense in  $X$ . We call such a point  $x \in X$  a (*forward*) *transitive point*.

We can check each of the examples in section 1.1 for this property.

EXAMPLE 1. We shall show that this example is forward transitive when  $k = 2$ , other cases being similar. Consider the sequence 1, 2, 11, 12, 21, 22, 111, 112, 121, 122, 221,  $\dots$ , 222, 1111,  $\dots$ . We can write down  $x_n \in \{1, 2\}$ ,  $n \geq 0$ , as the  $n$ th term in the sequence

$$1211122122111112121122221 \dots 2221111 \dots$$

Finally, consider the point  $x \in [0, 1]$  given by the series  $x = \sum_{n=0}^{+\infty} \frac{(x_n - 1)}{2^{n+1}}$ . We claim that the point  $x$  is a (forward) transitive point. Observe that

$$\begin{aligned} Tx &= 2 \left( \sum_{n=0}^{+\infty} \frac{(x_n - 1)}{2^{n+1}} \right) \pmod{1} = x_0 + \sum_{n=0}^{+\infty} \frac{(x_{n+1} - 1)}{2^{n+1}} \pmod{1} \\ &= \sum_{n=0}^{+\infty} \frac{(x_{n+1} - 1)}{2^{n+1}}. \end{aligned}$$

Similarly,  $T^k x = \sum_{n=0}^{+\infty} \frac{x_{n+k} - 1}{2^{n+1}} \pmod{1}$ .

To show that the set  $\{T^n x : n \geq 0\}$  is dense it suffices to show that for each interval of the form  $[\frac{p}{2^l}, \frac{p+1}{2^l}]$ , with  $0 \leq p \leq 2^l - 1$ , we can find  $N \geq 0$  with  $T^N x \in [\frac{p}{2^l}, \frac{p+1}{2^l}]$ . Given  $p$  we can write it in binary form as  $i_0 \dots i_{n-1}$ , with  $i_0, \dots, i_{n-1} \in \{0, 1\}$ . But for some  $N$  we can find  $x_N = i_0, x_{N+1} = i_1, \dots, x_{N+n-1} = i_{n-1}$ . This means that  $T^N x \in [\frac{p}{2^l}, \frac{p+1}{2^l}]$ , as required.

EXAMPLE 2. There are two different cases, depending on whether or not  $\alpha$  is irrational.

First assume that  $\alpha$  is irrational, then the map  $T : X \rightarrow X$  can be shown to be transitive (and even forward transitive) where  $x = 0$ , say. It suffices to show that the orbit  $\{T^n 0\}_{n \in \mathbb{Z}^+}$  is dense. Since this is an infinite set in  $\mathbb{R}/\mathbb{Z}$  we can choose  $x \in \mathbb{R}/\mathbb{Z}$  and a sub-sequence  $n_i \rightarrow +\infty$  with  $T^{n_i} 0 = n_i \alpha \pmod{1} \rightarrow x$ . For any sufficiently small  $\epsilon > 0$  we can choose  $n_i > n_j$  with  $|T^{n_i} 0 - x| < \frac{\epsilon}{2}$  and  $|T^{n_j} 0 - x| < \frac{\epsilon}{2}$  and thus  $|T^{n_i} 0 - T^{n_j} 0| = |T^{n_i - n_j} 0| < \epsilon$ . Moreover,  $T^{n_i} 0 \neq T^{n_j} 0$ , since if not this would contradict  $\alpha$  being irrational. Thus the points  $T^{(n_i - n_j)k} 0$ ,  $k \geq 1$ , form an  $\epsilon$ -dense subset of  $\mathbb{R}/\mathbb{Z}$ . Since  $\epsilon$  can be chosen arbitrarily small this completes the proof of transitivity.

Now assume that  $\alpha = \frac{p}{q}$  with  $p, q \in \mathbb{Z}$  having no common divisors and  $q \neq 0$ . For any  $x \in X$  the orbit  $\{T^n x : n \in \mathbb{Z}\}$  would be a finite set  $\{x, x + \frac{1}{q}, \dots, x + \frac{q-1}{q} \pmod{1}\}$ . In particular,  $T$  is *not* transitive.

EXAMPLE 3. We shall show that this example is forward transitive. The sequence

$$\{x_n\}_{n \in \mathbb{N}} = \{1, 2, \dots, k, 1, 1, \dots, 1, k, 2, 1, \dots, 2, k, \dots, \underbrace{z_0, z_1, \dots, z_{N-1}}_{\text{All strings appear}} \dots\}$$

in  $X_k$  (in which all finite strings appear once) is a forward transitive point. To see this choose any point  $z \in X_k$  and for any  $\epsilon > 0$  choose  $N > 0$  sufficiently large that  $(\frac{1}{2})^N < \epsilon$ . If we choose  $r$  such that  $x_r = z_0, \dots, x_{r+N-1} = z_{N-1}$  then we see that  $(\sigma^r x)_0 = x_r = z_0, \dots, (\sigma^r x)_{N-1} = x_{r+N-1} = z_{r+N-1}$  and so  $d(\sigma^r x, z) \leq (\frac{1}{2})^N < \epsilon$ .

### 1.3 Other characterizations of transitivity

The following result gives equivalent conditions for a homeomorphism of a compact metric space to be transitive.

THEOREM 1.3. *The following are equivalent.*

- (i)  $T : X \rightarrow X$  is transitive.
- (ii) If  $U$  is an open set with  $TU = U$  then either  $U$  is dense or  $U = \emptyset$ .
- (iii) If  $U, V$  are non-empty open sets then for some  $n \in \mathbb{Z}$  we have that  $T^n U \cap V \neq \emptyset$ .
- (iv) The set  $\{x \in X : \text{the orbit } \{T^n x\}_{n \in \mathbb{Z}} \text{ is dense in } X\}$  is a dense  $G_\delta$  set (i.e. the intersection of a countable collection of open dense sets).

PROOF. (i)  $\implies$  (ii). Assume  $x \in X$  has a dense orbit. Assume that  $TU = U \neq \emptyset$ . We can choose  $n \in \mathbb{Z}$  such that  $T^n x \in U$ . Moreover, for any  $m \in \mathbb{Z}$  we have that  $T^m x \in T^{m-n}U = U$ . Since the orbit of  $x$  is dense (i.e.  $\cup_{m \in \mathbb{Z}} T^m x \subset X$  is dense) we see that  $U$  is dense.

(ii)  $\implies$  (iii). The  $T$ -invariant union  $\cup_{n \in \mathbb{Z}} T^n U$  is dense in  $X$  by assumption (ii). Thus  $\cup_{n \in \mathbb{Z}} T^n U \cap V \neq \emptyset$  and so  $\exists n \in \mathbb{Z}$  with  $T^n U \cap V \neq \emptyset$ .

(iii)  $\implies$  (iv). Consider a dense set  $\{x_n\}_{n \in \mathbb{N}}$  and consider the balls of radius  $\frac{1}{k}, k \geq 1$ , denoted by  $B(x_n, \frac{1}{k})$ . We can identify

$$\{x \in X : \{T^m x\}_{m \in \mathbb{Z}} \text{ is dense in } X\} = \cap_{n=0}^{+\infty} \cap_{k=1}^{+\infty} \cup_{m=-\infty}^{+\infty} T^m B\left(x_n, \frac{1}{k}\right)$$

(i.e.  $\forall n \geq 0, \forall k \geq 1, \exists m \in \mathbb{Z}$  with  $T^m x \in B(x_n, \frac{1}{k})$ ).

(iv)  $\implies$  (i). This is immediate. ■

REMARK. There is a similar result giving equivalent conditions for forward transitivity [4, p. 128]

### 1.4 Transitivity for subshifts of finite type

In section 1.1 we defined the shift transformation  $\sigma : X_k \rightarrow X_k$  on  $X_k = \prod_{n \in \mathbb{Z}} \{1, \dots, k\}$ . For any closed  $\sigma$ -invariant subset  $X \subset X_k$  (i.e.  $\sigma(X) = X$ ) we consider the restriction  $\sigma|_X$ . We can use the same notation  $\sigma : X \rightarrow X$ .

DEFINITION. Let  $A$  be a  $k \times k$  matrix with entries 0 or 1. We call the matrix *irreducible* if  $\forall 1 \leq i, j \leq k, \exists N > 0$  such that  $A^N(i, j) > 0$ .

EXAMPLES. When  $k = 3$  the matrix  $A = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}$  is irreducible. However, the matrix  $A' = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$  is not irreducible. (These properties are readily checked).

DEFINITION. Given a  $k \times k$  matrix  $A$  with entries 0 or 1 we define

$$X_A = \{(x_n)_{n \in \mathbb{Z}} \in \prod_{n=-\infty}^{\infty} \{1, \dots, k\} : A(x_n, x_{n+1}) = 1, n \in \mathbb{Z}\}.$$

We define the *subshift of finite type*  $\sigma : X_A \rightarrow X_A$  to be the restriction  $\sigma|_{X_A}$ .

The following gives necessary and sufficient conditions for  $\sigma : X_A \rightarrow X_A$  to be transitive.

THEOREM 1.4. *A subshift of finite type  $\sigma : X_A \rightarrow X_A$  is transitive if and only if  $A$  is irreducible.*

PROOF. Assume that  $\sigma$  is transitive. Consider the sets

$$[i]_0 := \{(x_n)_{n \in \mathbb{Z}} \in X_A : x_0 = i\}$$

for  $i = 1, \dots, k$ . These sets are open. Given  $1 \leq i, j \leq k$  we know that there exists  $N > 0$  such that  $\sigma^{-N}[j]_0 \cap [i]_0 \neq \emptyset$ . Choose  $(x_n)_{n \in \mathbb{Z}} \in \sigma^{-N}[j]_0 \cap [i]_0$ ; then we know that  $x_0 = i$  and  $x_N = j$ . Notice that

$$A^N(i, j) = \sum_{r_1=1}^k \dots \sum_{r_{N-1}=1}^k A(i, r_1)A(r_1, r_2) \dots A(r_{N-2}, r_{N-1})A(r_{N-1}, j).$$

But since  $A(i, x_1) = A(x_1, x_2) = \dots = A(x_{N-1}, j) = 1$  we see that  $A^N(i, j) \geq 1$ .

Conversely, assume that for  $1 \leq i, j \leq k$  we have that  $A^N(i, j) \geq 1$ . Given  $U, V \neq \emptyset$  open sets we can choose  $(i_n)_{n \in \mathbb{Z}} \in U$  and  $(j_n)_{n \in \mathbb{Z}} \in V$  such that for  $M > 0$  sufficiently large

$$U \supset [i_{-M}, i_{-M-1}, \dots, i_M]_{-M}^M := \{(x_n)_{n \in \mathbb{Z}} \in X_A : x_k = i_k, -M \leq k \leq M\},$$

$$V \supset [j_{-M}, j_{-M-1}, \dots, j_M]_{-M}^M := \{(x_n)_{n \in \mathbb{Z}} \in X_A : x_k = j_k, -M \leq k \leq M\}.$$

By hypothesis we can find  $N > 0$  such that  $A^N(i_M, j_{-M}) \geq 1$ . This means that we can find a string  $x'_1, \dots, x'_{N-1}$  such that  $A(i_M, x'_1) = A(x'_1, x'_2) = \dots = A(x'_{N-1}, j_{-M}) = 1$  and then define

$$x_n = \begin{cases} i_n & \text{if } n \leq M, \\ x'_{n-M} & \text{if } M+1 \leq n \leq M+N-1, \\ j_{n-(2M+N)} & \text{if } M+N \leq n; \end{cases}$$

then we have that  $x \in U \cap \sigma^N V$  i.e.  $U \cap \sigma^N V \neq \emptyset$ . ■

### 1.5 Minimality and the Birkhoff recurrence theorem

In this section we want to present a simple but important recurrence result, called the *Birkhoff recurrence theorem*. Our starting point is to define the following property.

**DEFINITION.** A homeomorphism  $T : X \rightarrow X$  is *minimal* if for every  $x \in X$  the orbit  $\{T^n x : n \in \mathbb{Z}\}$  is dense in  $X$ .

The following is obvious from the definitions

**PROPOSITION 1.5.** *A minimal homeomorphism is necessarily transitive.*

We can now consider each of the examples from section 1.1 and ask which of these are minimal. Since Example 1 is not a homeomorphism we begin with Example 2.

**EXAMPLE 2.**

**LEMMA 1.6.** *When  $\alpha$  is irrational then  $T(x) = x + \alpha$  is minimal.*

**PROOF.** It suffices to show that for every  $x \in \mathbb{R}/\mathbb{Z}$  and every neighbourhood  $(y - \epsilon, y + \epsilon)$  ( $y \in \mathbb{R}/\mathbb{Z}, \epsilon > 0$ ) we can find  $n \geq 1$  such that  $T^n(x) \in (y - \epsilon, y + \epsilon)$ .

We already know that  $T$  is transitive (i.e. there exists at least one transitive point  $x_0 \in \mathbb{R}/\mathbb{Z}$  with dense orbit). Fix  $y \in \mathbb{R}/\mathbb{Z}$  and use the transitivity to choose a sub-sequence  $n_i$  with  $T^{n_i} x_0 \rightarrow (y - x + x_0)$  as  $i \rightarrow +\infty$ . Thus

$$\begin{aligned} T^{n_i} x &= n_i \alpha + x \pmod{1} \\ &= n_i \alpha + x_0 + (x - x_0) \pmod{1} \\ &= T^{n_i}(x_0) + (x - x_0) \pmod{1} \\ &\rightarrow y + (x_0 - x) + (x - x_0) = y \pmod{1}. \end{aligned}$$
■

**EXAMPLE 3.** The shift map is not minimal since it contains a fixed point (e.g.  $x = (\dots, 1, 1, 1, \dots)$ ).

The following theorem gives equivalent definitions.

**THEOREM 1.7.** *Let  $T : X \rightarrow X$  be a homeomorphism of a compact metric space. The following properties are equivalent.*

- (i)  $T$  is minimal.
- (ii) If  $TE = E$  is a closed  $T$ -invariant set, then either  $E = \emptyset$  or  $E = X$ .
- (iii) If  $U \neq \emptyset$  is an open set then  $X = \cup_{n \in \mathbb{Z}} T^n U$ .

**PROOF.** (i)  $\implies$  (ii) Assume that  $TE = E \neq \emptyset$  and choose  $x \in E$ . Hypothesis (i) gives that  $X = \text{cl}(\{T^n x\}_{n \in \mathbb{Z}}) \subset E \subset X$ .

(ii)  $\implies$  (iii) Given a non-empty open set  $U$  let  $E = X - (\cup_{n \in \mathbb{Z}} T^n U)$ . By construction  $TE = E$  and  $E \neq X$  (since  $U \neq \emptyset$ ) and so by hypothesis (ii) we have that  $E = \emptyset$ . Thus  $X = \cup_{n \in \mathbb{Z}} T^n U$ .

(iii)  $\implies$  (i) Fix  $x \in X$  and an open neighbourhood  $U \subset X$ . Since  $x \in T^n U$  for some  $n \in \mathbb{Z}$  (by hypothesis (iii)) we have that  $T^{-n}x \in U$ . This shows that the orbit  $\{T^n x\}_{n \in \mathbb{Z}}$  is dense in  $X$ . ■

Using property (ii) we get the following surprising result that every homeomorphism contains a minimum homeomorphism.

**THEOREM 1.8.** *Let  $T : X \rightarrow X$  be a homeomorphism of a compact metric space  $X$ . There exists a non-empty closed set  $Y \subset X$  with  $TY = Y$  and  $T : Y \rightarrow Y$  is minimal.*

**PROOF.** This follows from an application of Zorn's Lemma. Let  $\mathcal{E}$  denote the family of all closed  $T$ -invariant subsets of  $X$  with the partial ordering by inclusion, i.e.  $Z_1 \leq Z_2$  iff  $Z_1 \subset Z_2$ .

Every totally ordered subset (or "chain")  $\{Z_\alpha\}$  has a least element  $Z = \cap_\alpha Z_\alpha$  (which is non-empty by compactness of  $X$ ). Thus by Zorn's lemma there exists a minimal element  $Y \subset X$  (i.e.  $Y \in \mathcal{E}$  and  $Y' \in \mathcal{E}$  with  $Y' \leq Y$  implies that  $Y = Y'$ ). By property (ii) of Theorem 1.7 this can be re-interpreted as saying that  $T : Y \rightarrow Y$  is minimal. ■

As a corollary we get the following simple but elegant result.

**COROLLARY 1.8.1 (BIRKHOFF RECURRENCE THEOREM).** *Let  $T : X \rightarrow X$  be a homeomorphism of a compact metric space  $X$ . We can find  $x \in X$  such that  $T^{n_i} x \rightarrow x$  for a sub-sequence of the integers  $n_i \rightarrow +\infty$ .*

**PROOF.** By Theorem 1.8 we can choose a  $T$ -invariant subset  $Y \subset X$  such that  $T : Y \rightarrow Y$  is minimal. For any  $x \in Y \subset X$  we have the required property. ■

**EXAMPLE 2.** Consider the case  $X = \mathbb{R}/\mathbb{Z}$  and  $T : X \rightarrow X$  defined by  $Tx = x + \alpha \pmod{1}$ , where  $\alpha$  is an irrational number.



Let  $\epsilon > 0$ ; then we can find  $n > 0$  (by Birkhoff's theorem) such that  $|\alpha n \pmod{1}| \leq \epsilon$ , i.e. there exists  $p \in \mathbb{N}$  such that  $-\epsilon \leq \alpha n - p \leq \epsilon$ . Rewriting this, we have that for any irrational  $\alpha$ ,  $\exists p, n \in \mathbb{N}$  such that  $|\alpha - \frac{p}{n}| \leq \frac{\epsilon}{n}$ . This is a (marginal) improvement on the most obvious estimate.

### 1.6 Commuting homeomorphisms

Let  $T_1, \dots, T_N : X \rightarrow X$  be commuting homeomorphisms on a compact metric space  $X$ , i.e.  $T_i T_j = T_j T_i$  for  $1 \leq i, j \leq N$ . In this section we shall briefly consider how some of the ideas from the previous section might be modified for such families of maps.

EXAMPLE 4. Consider the simple example of two rotations on the torus  $X = \mathbb{R}^n / \mathbb{Z}^n$  of the form

$$\begin{aligned} T_1(x_1, \dots, x_n) &= (x_1 + a_1^{(1)}, \dots, x_n + a_n^{(1)}) \pmod{1}, \\ &\vdots \\ T_N(x_1, \dots, x_n) &= (x_1 + a_1^{(N)}, \dots, x_n + a_n^{(N)}) \pmod{1}, \end{aligned}$$

where  $(a_1^{(1)}, \dots, a_n^{(1)}), \dots, (a_1^{(N)}, \dots, a_n^{(N)}) \in \mathbb{R}^n$ .

We can consider all closed *simultaneously invariant* sets  $A \subset X$ , i.e.  $T_i A = A$ ,  $i = 1, \dots, N$ . By a similar argument to that before, we can consider the partial order by inclusion on all such closed sets and by applying Zorn's lemma (just as in the proof of Theorem 1.8) we can deduce that there exists a closed set  $X_0 \subset X$  such that

- (i)  $T_i X_0 = X_0$ ,  $i = 1, \dots, N$ .
- (ii) whenever  $A \subset X_0$  with  $A$  closed and  $T_i A = A$  for  $i = 1, \dots, N$  then necessarily  $A = X_0$ .

The following lemma will prove useful in chapter 2.

LEMMA 1.9. *For each open set  $U \subset X_0$  we can choose a finite number  $M$  and  $n_{ij} \in \mathbb{Z}$  with  $1 \leq i \leq N$ ,  $1 \leq j \leq M$  with  $X_0 = \cup_{j=1}^M (T_1^{n_{1j}} \circ \dots \circ T_N^{n_{Nj}})U$ .*

PROOF. Clearly  $X_0 = \cup_{n_1 \in \mathbb{Z}} \dots \cup_{n_N \in \mathbb{Z}} (T_1^{n_1} \circ \dots \circ T_N^{n_N})U$  (since otherwise the difference  $X_0 - (\cup_{n_1 \in \mathbb{Z}} \dots \cup_{n_N \in \mathbb{Z}} (T_1^{n_1} \circ \dots \circ T_N^{n_N})U)$  is a closed (non-empty) set invariant under  $T_1, \dots, T_N$ , contradicting property (ii) above). Now by compactness we can choose a *finite* subcover. This completes the proof. ■

To formulate a generalisation of the Birkhoff recurrence theorem to a family of commuting maps is a more substantial exercise, and will be a principal part of chapter 2.

### 1.7 Comments and references

A wealth of interesting examples can be found in the literature (cf. [1], [2], [3], [4], [5]).

The simple Birkhoff recurrence theorem has a version for commuting homeomorphisms (the multiple Birkhoff recurrence theorem) which we shall describe in Chapter 2. The corresponding result to the Birkhoff recurrence theorem in ergodic theory is the Poincaré recurrence theorem, which appears in section 9.2.

#### References

1. P. Billingsley, *Ergodic Theory and Information*, Addison-Wesley, New York, 1965.
2. R. Devaney, *An Introduction to Chaotic Dynamical Systems*, Addison-Wesley, New York, 1989.
3. A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*, C.U.P., Cambridge, 1995.
4. W. Szlenk, *An Introduction to the Theory of Smooth Dynamical Systems*, Wiley, New York, 1984.
5. P. Walters, *An Introduction to Ergodic Theory*, Graduate Texts in Mathematics, 79, Springer, Berlin, 1982.

## CHAPTER 2

### AN APPLICATION OF RECURRENCE TO ARITHMETIC PROGRESSIONS

In this chapter we shall describe a particularly nice application of the recurrence ideas from chapter 1 to a result in number theory.

#### 2.1 Van der Waerden's theorem

We begin with a simple idea from number theory.

**DEFINITION.** An *arithmetic progression* is a sequence of integers  $\{a + jb\}_{j=0}^{N-1}$  for  $a, b \in \mathbb{Z}$  ( $b \neq 0$ ),  $N \geq 1$ . We call  $N$  the *length* of the arithmetic progression.

**EXAMPLES.**

- (1) The sequence 10, 13, 16, 19, 22 is an arithmetic progression with  $a = 10, b = 3, N = 5$ .
- (2) The sequence  $-4, 0, 4, 8$  is an arithmetic progression with  $a = -4, b = 4, N = 4$ .

Consider a *partition* of the integers  $\mathbb{Z} = B_1 \cup \dots \cup B_l$  where

- (i)  $B_i \neq \emptyset$ ,
- (ii)  $B_i \cap B_j = \emptyset$  for  $i \neq j$ .

The main result we want to prove is the following.

**THEOREM 2.1 (VAN DER WAERDEN).** *Consider a finite partition  $\mathbb{Z} = B_1 \cup \dots \cup B_k$ . At least one element  $B_r$  in the partition will contain arithmetic progressions of arbitrary length (i.e.  $\exists 1 \leq r \leq k, \forall N > 0, \exists a, b \in \mathbb{Z}$  ( $b \neq 0$ ) such that  $a + jb \in B_r$  for  $j = 0, \dots, N - 1$ ).*

Since an arithmetic progression of length  $N$  contains arithmetic progressions of all shorter lengths, this is equivalent to:  $\exists N_i \rightarrow +\infty, \exists a_i, b_i \in \mathbb{Z}$  such that  $a_i + jb_i \in B_r$  for  $j = 0, \dots, N_i - 1$ .

We give below some simple examples.

**EXAMPLES.**

- (1) If the sets  $B_2, \dots, B_k$ , say, in the partition are finite then it is easy to see that  $B_1$  is the element with arithmetic progressions of arbitrary length.

- (2) If  $\mathbb{Z} = B_1 \cup B_2$  where  $B_1 = \{\text{odd numbers}\}$  and  $B_2 = \{\text{even numbers}\}$  then both contain arithmetic progressions of arbitrary length.
- (3) If  $B_1 = \{\text{prime numbers}\}$  and  $B_2 = \{\text{non-prime numbers}\}$  then  $B_2$  contains arithmetic progressions of arbitrary length. However, it is an unsolved problem as to whether  $B_1$  contains arithmetic progressions of arbitrary length.

**HISTORICAL NOTE.** This result was originally conjectured by Baudet and proved by Van der Waerden in 1927 [6, 7]. The theorem gained a wider audience when it was included in Khintchine's famous book *Three pearls in number theory* [4]. The dynamical proof we give is due to Furstenberg and Weiss [3] (from 1978).

## 2.2 A dynamical proof

The key to proving Van der Waerden's theorem is the following generalization of Birkhoff's theorem.

**THEOREM 2.2.** *Let  $T_1, \dots, T_N : X \rightarrow X$  be homeomorphisms of a compact metric space such that  $T_i T_j = T_j T_i$  for  $1 \leq i, j \leq N$ . There exist  $x \in X$  and  $n_j \rightarrow +\infty$  such that  $d(T_i^{n_j} x, x) \rightarrow 0$  for each  $i = 1, \dots, N$ .*

We shall first prove Theorem 2.1 assuming Theorem 2.2 and then return to the proof of Theorem 2.2.

**PROOF OF THEOREM 2.1 (ASSUMING THEOREM 2.2).** We want to begin by associating to the partition  $\mathbb{Z} = B_1 \cup \dots \cup B_k$  a suitable homeomorphism  $T : X \rightarrow X$  (and then we set  $T_j = T^j$ ,  $j = 1, \dots, N$ ).

Let  $\Omega = \prod_{n \in \mathbb{Z}} \{1, \dots, k\}$  and then we can associate to the partition  $\mathbb{Z} = B_1 \cup \dots \cup B_k$  a sequence  $z = (z_n)_{n \in \mathbb{Z}} \in \Omega$  by  $z_n = i$  if and only if  $n \in B_i$ .

Let  $\sigma : \Omega \rightarrow \Omega$  be the shift introduced in Example 3 of section 1.1 (i.e.  $(\sigma x)_n = x_{n+1}$ ,  $n \in \mathbb{Z}$ ). Consider the orbit  $\{\sigma^n z : n \in \mathbb{Z}\}$  and its closure  $X = \text{cl}(\cup_{n \in \mathbb{Z}} \sigma^n z)$ . Finally, we define  $T_i := T^i = \sigma \circ \dots \circ \sigma$  ( $T$  composed with itself  $i$  times).

By Theorem 2.2 (with  $\epsilon = \frac{1}{4}$ ) we can find  $x \in X$  and  $b \geq 1$  with

$$d(T_1^b x, x) < \frac{1}{4}, d(T_2^b x, x) < \frac{1}{4}, \dots, d(T_N^b x, x) < \frac{1}{4}.$$

Since  $X = \text{cl}(\cup_{n \in \mathbb{Z}} \sigma^n z)$  we can choose  $a \in \mathbb{Z}$  such that

$$d(x, T^a z) < \frac{1}{4}, d(T_1^b x, T^a T_1^b z) < \frac{1}{4}, \dots, d(T_N^b x, T^a T_N^b z) < \frac{1}{4}.$$

Thus, for each  $i = 1, \dots, N$  we have that

$$d(T^a T_i^b x, T^a z) \leq d(T^a T_i^b x, T_i^b x) + d(T_i^b x, x) + d(x, T^a z) < \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}.$$

Since  $d(x, y) = \left(\frac{1}{2}\right)^{N(x, y)}$  (where  $N(x, y) = \min\{|N| \geq 0 : x_N \neq y_N, \text{ or } x_{-N} \neq y_{-N}\}$ ) we see that  $(T^b T_i^a x)_0 = x_{b+ia} = z_a \in \{1, \dots, k\}$  for  $i = 1, \dots, N$ . This means that  $b + ia \in B_{z_a}$ , for  $i = 1, \dots, N$ , and completes the proof of Theorem 2.1. ■

All that remains is to prove Theorem 2.2. This is a fairly detailed proof and to help clarify matters we shall divide it into sublemmas.

PROOF OF THEOREM 2.2. We shall use a proof by induction.

CASE  $N = 1$ . For  $N = 1$  the multiple Birkhoff recurrence theorem reduces to the (usual) Birkhoff recurrence theorem (Corollary 1.8.1).

INDUCTIVE STEP. Assume that the result is known for  $N - 1$  commuting homeomorphisms. We need to show that it holds for  $N$  commuting homeomorphisms.

SIMPLIFYING FACT. We can assume that  $X$  is the *smallest* closed set invariant under each of  $T_1, \dots, T_N$ . If this is not the case we can restrict to such a set (using Zorn's lemma as in section 1.6).

In order to establish the Birkhoff multiple recurrence theorem for these  $N$  commuting homeomorphisms, the following simple alternative formulation of this result is useful.

ALTERNATIVE FORMULATION. Let  $\mathcal{X}_N = X \times \dots \times X$  be the  $N$ -fold cartesian product of  $X$  and let  $\mathcal{D}_N = \{(x, \dots, x) \in \mathcal{X}_N\}$  be the diagonal of the space. Let  $S : \mathcal{X}_N \rightarrow \mathcal{X}_N$  be given by  $S(x_1, \dots, x_N) = (T_1 x_1, \dots, T_N x_N)$ . Then the following are equivalent:

- (i)<sub>N</sub> the Birkhoff multiple recurrence holds for  $T_1, \dots, T_N$ ;
- (ii)<sub>N</sub>  $\exists \underline{z} = (z, \dots, z) \in \mathcal{D}_N$  such that  $d_{\mathcal{X}_N}(S^{n_i} \underline{z}, \underline{z}) \rightarrow 0$  as  $n_i \rightarrow +\infty$   
(where  $d_{\mathcal{X}_N}(\underline{z}, \underline{w}) = \sup_{1 \leq i \leq N} d(z_i, w_i)$ ).

We can apply the inductive hypothesis to the  $(N - 1)$  commuting homeomorphisms  $T_1 T_N^{-1}, \dots, T_{N-1} T_N^{-1}$  and using the equivalence of (i)<sub>N-1</sub> and (ii)<sub>N-1</sub> above we have that for the map  $R := T_1 T_N^{-1} \times \dots \times T_{N-1} T_N^{-1} : \mathcal{X}_{N-1} \rightarrow \mathcal{X}_{N-1}$  defined by

$$R : (x_1, \dots, x_{N-1}) \mapsto (T_1 T_N^{-1} x_1, \dots, T_{N-1} T_N^{-1} x_{N-1})$$

there exists  $\underline{z} = (z, \dots, z) \in \mathcal{D}_{N-1} \subset \mathcal{X}_{N-1}$  with  $d_{\mathcal{X}_{N-1}}(R^{n_i} \underline{z}, \underline{z}) \rightarrow 0$  as  $n_i \rightarrow +\infty$ . In particular,  $d_{\mathcal{X}_N}(S^{n_i} \underline{z}', z) \rightarrow 0$  as  $n_i \rightarrow +\infty$  where  $\underline{z} = (z, \dots, z), \underline{z}' = (T_N^{-n_i} z, \dots, T_N^{-n_i} z) \in \mathcal{D}_N$ .

Thus we have proved the following result.

SUBLEMMA 2.2.1.  $\forall \epsilon > 0, \exists z, z' \in \mathcal{D}_N, \exists n \geq 1$  such that  $d_{\mathcal{X}_N}(S^n z, z') < \epsilon$ .

Unfortunately, this is not quite in the form of  $(ii)_N$  we need for the inductive step. (For example, we would like to take  $z = z'$ .) To get a stronger result, we break the argument up into steps represented by the following sublemmas.

SUBLEMMA 2.2.2.  $\forall \epsilon > 0, \forall x \in \mathcal{D}_N, \exists y \in \mathcal{D}_N$  and  $\exists n \geq 1$  such that  $d(S^n y, x) < \epsilon$ .

(This changes one of the quantifiers  $\exists$  to  $\forall$ .)

SUBLEMMA 2.2.3.  $\forall \epsilon > 0, \exists z \in \mathcal{D}_N$  and  $n \geq 1$  such that  $d(S^n z, z) < \epsilon$

(This is almost the Birkhoff multiple recurrence theorem, except that  $z$  might still depend on the choice of  $\epsilon > 0$ .)

We will now complete the proof of the Birkhoff multiple recurrence theorem assuming Sublemma 2.2.3. (We shall then return to the proofs “Sublemma 2.2.1  $\implies$  Sublemma 2.2.2” and “Sublemma 2.2.2  $\implies$  Sublemma 2.2.3” in the next section.)

Consider the function  $F : \mathcal{D}_N \rightarrow \mathbb{R}^+ = [0, +\infty)$  defined by  $F(x) = \inf_{n \geq 1} d(S^n x, x)$ . It is easy to see that to complete the proof of Theorem 2.2 we need only show there exists a point  $x_0 \in \mathcal{D}_N$  with  $F(x_0) = 0$ . To show this fact, the following properties of  $F$  are needed.

SUBLEMMA 2.2.4.

- (i)  $F : \mathcal{D}_N \rightarrow \mathbb{R}^+$  is upper semi-continuous (i.e.  $\forall x \in \mathcal{D}_N, \forall \epsilon > 0, \exists \delta > 0$  such that  $d(x, y) < \delta \implies F(y) \leq F(x) + \epsilon$ ).
- (ii)  $\exists x_0 \in \mathcal{D}_N$  such that  $F : \mathcal{D}_N \rightarrow \mathbb{R}^+$  is continuous at  $x_0$ .

PROOF.

(i) This is an easy exercise from the definition of  $F$ .

(ii) For  $\epsilon > 0$  we can define  $A_\epsilon = \{x \in \mathcal{D}_N : \forall \eta > 0, \exists y$  such that  $d(y, x) < \eta$  and  $F(y) \leq F(x) - \epsilon\}$  (i.e.  $\exists$  points  $y$  arbitrarily close to  $x$  with  $F(y) \leq F(x) - \epsilon$ ). Notice that

- (a)  $A_\epsilon$  is closed,
- (b)  $A_\epsilon$  has empty interior.

(To see part (b) observe that if  $\text{int}(A_\epsilon) \neq \emptyset$  we could choose a sequence of pairs  $x, x_1 \in \text{int}(A_\epsilon)$  with  $F(x_1) \leq F(x) - \epsilon$ ,  $x_1, x_2 \in \text{int}(A_\epsilon)$  with  $F(x_2) \leq F(x_1) - \epsilon$ , etc. Together these inequalities give  $F(x_n) \leq F(x) - n\epsilon < 0$  for  $n$  arbitrarily large. But this contradicts  $F \geq 0$ ).

The set of points at which  $F$  is continuous is

$$\{x \in \mathcal{D}_N : x \notin A_\epsilon, \epsilon > 0\} = \bigcap_{n=1}^{\infty} \left( \mathcal{D}_N - A_{\frac{1}{n}} \right).$$

Since this is a countable intersection of open dense sets, it is still dense (by Baire's theorem). Thus there exists at least one point of continuity for  $F : \mathcal{D}_N \rightarrow \mathbb{R}^+$  (in fact, infinitely many). This completes the proof of Sublemma 2.2.4. ■

Let  $x_0$  be such a point of continuity.

Assume for a contradiction that  $F(x_0) > 0$ . We can then choose  $\delta > 0$  and an open neighbourhood  $U \ni x_0$  such that  $F(x) > \delta > 0$  for  $x \in U$ . However, we also know that for the diagonal actions  $T_i : (x_1, \dots, x_N) \mapsto (T_i x_1, \dots, T_i x_N)$

$$\mathcal{D}_N \subset \cup_{j=1}^M (T_1^{n_{1j}} \circ \dots \circ T_N^{n_{Nj}})^{-1} U$$

(since by the simplifying assumption  $X$  is the smallest closed set invariant under  $T_1, \dots, T_N$  and so we may apply Lemma 1.9 from Chapter 1).

By (uniform) continuity of the family  $\{T_1^{n_{1j}} \circ \dots \circ T_N^{n_{Nj}}\}_{j=1}^M$  there exists  $\eta > 0$  such that

$$d(x, y) < \eta \implies d(T_1^{n_{1j}} \circ \dots \circ T_N^{n_{Nj}} x, T_1^{n_{1j}} \circ \dots \circ T_N^{n_{Nj}} y) < \delta \quad (2.1)$$

(for  $1 \leq j \leq M$ ). Observe that for  $y \in (T_1^{n_{1j}} \circ \dots \circ T_N^{n_{Nj}})^{-1} U$  ( $j = 1, \dots, M$ ) we have that  $F(y) \geq \eta$ . If this were not the case then there would exist  $n \geq 1$  with  $d(y, S^n y) < \eta$ , from the definition of  $F$ . This then implies that  $d(T_1^{n_{1j}} \circ \dots \circ T_N^{n_{Nj}} y, T_1^{n_{1j}} \circ \dots \circ T_N^{n_{Nj}} S^n y) < \delta$  by (2.1). Choosing  $x := T_1^{n_{1j}} \circ \dots \circ T_N^{n_{Nj}} y \in U$  gives  $F(x) = \inf_{n \geq 1} d(x, S^n x) < \delta$  which contradicts our hypothesis.

Finally we see that by (2.1) we have  $F(y) \geq \eta$  for all  $y \in \mathcal{D}_N$ . However, this contradicts Sublemma 2.2.3 and we conclude that  $F(x_0) = 0$ .

The proof of Theorem 2.3 is finished (given the proofs of Sublemma 2.2.2 and Sublemma 2.2.3). ■

### 2.3. The proofs of Sublemma 2.2.2 and Sublemma 2.2.3

We now supply the missing proofs of Sublemma 2.2.2 and Sublemma 2.2.3.

PROOF OF SUBLEMMA 2.2.2 (ASSUMING SUBLEMMA 2.2.1). Consider the  $N$  commuting maps  $\hat{T}_1, \hat{T}_2, \dots, \hat{T}_N : \mathcal{D}_N \rightarrow \mathcal{D}_N$  defined by

$$\left\{ \begin{array}{l} \hat{T}_1 = T_1 \times \dots \times T_1 : \mathcal{D}_N \rightarrow \mathcal{D}_N, \\ \hat{T}_2 = T_2 \times \dots \times T_2 : \mathcal{D}_N \rightarrow \mathcal{D}_N, \\ \dots \\ \hat{T}_N = T_N \times \dots \times T_N : \mathcal{D}_N \rightarrow \mathcal{D}_N. \end{array} \right.$$

We want to apply Lemma 1.9 to these commuting maps with the choice of open set  $U = \{w \in \mathcal{D}_N : d_{\mathcal{D}_N}(x, w) < \frac{\epsilon}{2}\}$ . This allows us to conclude that there exist  $n_{1j}, \dots, n_{Nj}$  ( $j = 1, \dots, M$ ) such that

$$\mathcal{D}_N = \cup_{j=1}^N \hat{T}^{-n_{1j}} \dots \hat{T}^{-n_{Nj}} U.$$

Thus for any  $z \in \mathcal{D}_N$  we have some  $1 \leq j \leq M$  such that

$$d_{\mathcal{D}_N}(\hat{T}^{n_{1j}} \dots \hat{T}^{n_{Nj}} z, x) < \frac{\epsilon}{2}. \quad (2.2)$$

Next we can use (uniform) continuity of  $\hat{T}^{n_{1j}} \circ \dots \circ \hat{T}^{n_{Nj}}$  to say that there exists  $\delta > 0$  such that whenever  $d_{\mathcal{D}_N}(z, z') < \delta$  for  $z, z' \in \mathcal{D}_N$  then we have that

$$d_{\mathcal{X}_N}(\hat{T}^{n_{1j}} \circ \dots \circ \hat{T}^{n_{Nj}} z, \hat{T}^{n_{1j}} \circ \dots \circ \hat{T}^{n_{Nj}} z') < \frac{\epsilon}{4}. \quad (2.3)$$

By Sublemma 2.2.1  $\exists z, z' \in \mathcal{D}_N$  and  $\exists n \geq 1$  such that  $d_{\mathcal{X}_N}(S^n z, z') < \delta$ . Therefore by inequality (2.3) we have that

$$d_{\mathcal{X}_N}\left(S^n\left(\hat{T}^{n_{1j}} \circ \dots \circ \hat{T}^{n_{Nj}} z\right), \hat{T}^{n_{1j}} \circ \dots \circ \hat{T}^{n_{Nj}} z'\right) < \frac{\epsilon}{4}. \quad (2.4)$$

Writing  $y = \hat{T}^{n_{1j}} \dots \hat{T}^{n_{Nj}} z$  and comparing (2.2), (2.3) and (2.4) gives that

$$\begin{aligned} d_{\mathcal{X}_N}(S^n y, x) &\leq d_{\mathcal{X}_N}(S^n y, \hat{T}^{n_{1j}} \circ \dots \circ \hat{T}^{n_{Nj}} z') + d_{\mathcal{X}_N}(\hat{T}^{n_{1j}} \circ \dots \circ \hat{T}^{n_{Nj}} z', x) \\ &\quad + d_{\mathcal{X}_N}(\hat{T}^{n_{1j}} \circ \dots \circ \hat{T}^{n_{Nj}} z, \hat{T}^{n_{1j}} \circ \dots \circ \hat{T}^{n_{Nj}} z') \\ &\quad + d_{\mathcal{X}_N}(\hat{T}^{n_{1j}} \circ \dots \circ \hat{T}^{n_{Nj}} z, x) \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{4} + \frac{\epsilon}{4} = \epsilon. \end{aligned}$$

This completes the proof of Sublemma 2.2.2. ■

PROOF OF SUBLEMMA 2.2.3 (ASSUMING SUBLEMMA 2.2.2). Fix  $z_0 \in \mathcal{D}_N$  and let  $\epsilon_1 = \frac{\epsilon}{2}$ . By Sublemma 2.2.2 we can choose  $n_1 \geq 1$  and  $z_1 \in \mathcal{D}_N$  with  $d(T^{n_1} z_1, z_0) < \epsilon_1$ .

By continuity of  $T^{n_1}$  we can find  $\epsilon_1 > \epsilon_2 > 0$  such that  $d(z, z_1) < \epsilon_2$  implies that  $d(T^{n_1} z, z_0) < \epsilon_1$ .

We can now continue inductively (for  $k \geq 2$ ):

- (a) By Sublemma 2.2 we can choose  $n_k \geq 1$  and  $z_k \in \mathcal{D}_N$  with  $d(T_k^{n_k} z_k, z_{k-1}) < \epsilon_k$ .
- (b) By continuity of  $T^{n_k}$  we can find  $\epsilon_k > \epsilon_{k+1} > 0$  such that  $d(z, z_k) < \epsilon_{k+1}$  implies that  $d(T^{n_k} z, z_{k-1}) < \epsilon_k$ .



This results in sequences

$$z_0, z_1, z_2, \dots \in \mathcal{D}_N, \\ n_0, n_1, n_2 \dots \in \mathbb{N}, \text{ such that } \begin{cases} d(T^{n_k} z_k, z_{k-1}) < \epsilon_k, k \leq 1, \\ d(z, z_i) < \epsilon_{k+1} \implies d(T^{n_k} z, z_{k-1}) < \epsilon_k. \end{cases} \\ \epsilon_0 > \epsilon_1 > \epsilon_2 > \dots$$

In particular we get that whenever  $j < i$  then

$$d(T^{n_i+n_{i-1}+\dots+n_{j+2}+n_{j+1}} z_i, z_j) < \epsilon_{i+1} \leq \frac{\epsilon}{2}. \quad (2.5)$$

By compactness of  $\mathcal{D}_N$  we can find  $d(z_i, z_j) < \frac{\epsilon}{2}$  for some  $j < i$ .

By the triangle inequality we have that for  $N = n_i + n_{i-1} + \dots + n_{j+1}$

$$d(T^N z_i, z_i) \leq d(T^N z_i, z_j) + d(z_j, z_i) < \epsilon.$$

Thus the choice  $z = z_i$  completes the proof of Sublemma 2.2.3. ■

## 2.4 Comments and references

A treatment of Van der Waerden's theorem (and many other related applications of dynamics to number theory) can be found in [1]. The proof originally appeared in the article [3] and the survey [2]. An account also appears in [5].

Sublemma 2.2.2 was originally proved by Bowen.

Finally, there is a stronger version of this result due to Szemerédi. In chapter 16 we shall present Furstenberg's proof of this using ergodic theory.

### References

1. H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, P.U.P., Princeton N.J., 1981.
2. H. Furstenberg, *Poincaré recurrence and number theory*, Bull. Amer. Math. Soc. **5** (1981), 211-234.
3. H. Furstenberg and B. Weiss, *Topological dynamics and combinatorial number theory*, J. d'Analyse Math. **34** (1978), 61-85.
4. A. Khintchine, *Three pearls of number theory*, Graylock Press, New York, 1948.
5. K. Petersen, *Ergodic Theory*, C.U.P., Cambridge, 1983.
6. B.L. Van der Waerden, *Beweis einer Baudet'schen Vermutung*, Nieuw. Arch. Wisk. **15** (1927), 212-16.
7. B.L. Van der Waerden, *How the proof of Baudet's conjecture was found*, Studies in Pure Mathematics presented to Richard Rado (L. Mirsky, ed.), Academic Press, London, 1971, pp. 251-260.

## TOPOLOGICAL ENTROPY

In this chapter we shall introduce an important numerical quantity called topological entropy. This is an important quantifier of their dynamical behaviour (as we shall see in chapters 4 and 5). It also plays an important rôle as an invariant for the classification of continuous maps up to conjugacy.

### 3.1 Definitions

We begin with some simple ideas on finite open covers for  $X$ .

DEFINITIONS. Let  $X$  be a compact metric space. If  $\alpha = \{A_i\}, \beta = \{B_j\}$  are (finite) open covers of  $X$ , define the *refinement*  $\alpha \vee \beta = \{A_i \cap B_j : A_i \cap B_j \neq \emptyset\}$ . More generally, if  $\alpha^r = \{A_1^r, \dots, A_{N_r}^r\}$ ,  $r = 1, \dots, k$ , are open covers of  $X$  (of cardinality  $N_r$ ) then we define their *refinement*

$$\bigvee_{r=1}^k \alpha^r = \{A_{i_1}^1 \cap A_{i_2}^2 \cap \dots \cap A_{i_k}^k : i_j \in \{1, \dots, N_r\}, j = 1, \dots, k\}.$$

To draw the map  $T : X \rightarrow X$  into the definition we make the following definitions.

DEFINITION. Let  $\alpha = \{A_1, \dots, A_n\}$  be an open cover for a compact metric space  $X$ ; then for a continuous map  $T : X \rightarrow X$  we define

$$T^{-1}\alpha = \{T^{-1}A_1, \dots, T^{-1}A_n\}.$$

For  $k \geq 1$  we define

$$\begin{aligned} \bigvee_{i=0}^{k-1} T^{-i}\alpha &= \alpha \vee T^{-1}\alpha \vee \dots \vee T^{-(k-1)}\alpha \\ &= \{A_{i_0} \cap T^{-1}A_{i_1} \cap \dots \cap T^{-(k-1)}A_{i_{k-1}} : 1 \leq i_0, i_1, \dots, i_{k-1} \leq n\}. \end{aligned}$$

Given a cover  $\alpha = \{A_1, \dots, A_n\}$  we call  $\beta \subset \alpha$  a *subcover* if it is still a cover for  $X$  (i.e.  $X = \cup_{B \in \beta} B$ ). We can now define the topological entropy of a cover  $\alpha$  for  $X$  as follows.

DEFINITION. The *topological entropy* of the cover  $\alpha$  is defined to be the logarithm  $H(\alpha) = \log N(\alpha)$  of the smallest number  $N(\alpha)$  of sets that can be used in a subcover of  $\alpha$ .

It is useful to illustrate this with a little example.

EXAMPLE. Let  $X = \mathbb{R}/\mathbb{Z}$  and consider the open cover

$$\alpha = \left\{ \left( \frac{1}{4}, 1 \right), \left( 0, \frac{3}{4} \right), \left( \frac{1}{2}, 1 \right] \cup \left( 0, \frac{1}{4} \right), \left( \frac{3}{4}, 1 \right] \cup \left( 0, \frac{1}{2} \right) \right\}$$

consisting of four open sets. Here  $N(\alpha) = 2$  and so  $H(\alpha) = \log 2$ .

The next lemma gives some of the basic properties of the topological entropy of covers.

LEMMA 3.1.

- (i)  $H(\alpha) \geq 0$ .
- (ii) If  $\beta \subset \alpha$  is a subcover then  $H(\alpha) \geq H(\beta)$ .
- (iii) If  $\alpha$  and  $\beta$  are two finite covers for  $X$  then  $H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$ .
- (iv) If  $T : X \rightarrow X$  is continuous and  $T(X) = X$  then  $H(\alpha) \geq H(T^{-1}\alpha)$ .  
If  $T : X \rightarrow X$  is a homeomorphism then  $H(\alpha) = H(T^{-1}\alpha)$ .

PROOF. Parts (i) and (ii) follow immediately from the definitions. For (iii) we let  $\alpha \supset \alpha' = \{A_1, \dots, A_n\}$  and  $\beta \supset \beta' = \{B_1, \dots, B_m\}$  be subcovers of  $\alpha$  and  $\beta$  of minimal cardinality. We can then write  $\alpha' \vee \beta' = \{A_i \cap B_j : 1 \leq i \leq n, 1 \leq j \leq m, A_i \cap B_j \neq \emptyset\}$  which is a subcover of  $\alpha \vee \beta$  of cardinality at most  $n \times m$ . Thus  $H(\alpha \vee \beta) \leq \log(nm) = \log(n) + \log(m) = \log H(\alpha) + \log H(\beta)$ .

For (iv), we note that if  $\alpha \supset \alpha' = \{A_1, \dots, A_n\}$  is a subset of  $\alpha$  of least cardinality then  $T^{-1}\alpha' = \{T^{-1}A_1, \dots, T^{-1}A_n\}$  is a subcover of  $T^{-1}\alpha$  of cardinality  $n$ . Thus  $N(T^{-1}\alpha) \leq n = N(\alpha)$ .

If  $T$  is surjective, then for a minimal subcover  $T^{-1}\alpha' = \{T^{-1}A_1, \dots, T^{-1}A_m\} \subset T^{-1}\alpha$  we have that  $\alpha' = \{A_1, \dots, A_m\}$  is a subcover of  $\alpha$ . Thus  $N(T^{-1}\alpha) = m \geq N(\alpha)$ . Together these two inequalities give an equality. ■

We can define the topological entropy of a transformation relative to a cover as follows.

DEFINITION. Let  $T : X \rightarrow X$  be a continuous map on  $X$ ; then we define the *topological entropy of  $T$  relative to a cover  $\alpha$*  by

$$h(T, \alpha) = \limsup_{n \rightarrow +\infty} \frac{1}{n} H(\vee_{i=0}^{n-1} T^{-i}\alpha).$$

We can see that  $h(T, \alpha) < +\infty$  by the following lemma.

LEMMA 3.2.  $\frac{1}{n} H(\vee_{i=0}^{n-1} T^{-i}\alpha) \leq H(\alpha)$  for  $n \geq 1$ .

PROOF. We can write  $H(\vee_{i=0}^{n-1} T^{-i}\alpha) \leq \sum_{i=0}^{n-1} H(T^{-i}\alpha) \leq nH(\alpha)$  by parts (iii) and (iv) of Lemma 3.1. ■

Finally, we can define the topological entropy of a transformation as follows.

DEFINITION. If  $T : X \rightarrow X$  is a continuous map on a compact metric space then we define the *topological entropy* by

$$h(T) = \sup\{h(T, \alpha) : \alpha \text{ is a finite cover for } X\}.$$

REMARK. In fact the limsup in the definition of the topological entropy of  $T$  relative to a cover  $\alpha$  can be replaced by a straightforward limit.

If we write  $a_n = H(\bigvee_{i=0}^{n-1} T^{-i} \alpha)$  then for  $n, m \geq 1$  we have that

$$\begin{aligned} a_{n+m} &= H(\bigvee_{i=0}^{n+m-1} T^{-i} \alpha) \\ &\leq H(\bigvee_{i=0}^{n-1} T^{-i} \alpha) + H(\bigvee_{i=n}^{n+m-1} T^{-i} \alpha) \\ &\leq H(\bigvee_{i=0}^{n-1} T^{-i} \alpha) + H(\bigvee_{i=0}^{m-1} T^{-i} \alpha) \\ &= a_n + a_m \end{aligned}$$

(i.e. this sequence is *subadditive*). In particular, if  $a = \inf\{\frac{a_n}{n} : n \geq 1\}$ , then we know that  $\frac{a_n}{n} \rightarrow a$ , as  $n \rightarrow +\infty$ .

Although the definition of topological entropy we have given is not particularly convenient for computations, we shall take the opportunity to show that the trivial identity map has zero topological entropy.

TRIVIAL EXAMPLE. Let  $T = id : X \rightarrow X$  be the identity transformation on a space  $X$ . For any cover  $\alpha$  we have that  $T^{-i} \alpha = \alpha$  and so  $\alpha = \bigvee_{i=0}^{n-1} T^{-i} \alpha$ . This means that  $H(\alpha) = H(\bigvee_{i=0}^{n-1} T^{-i} \alpha)$  and so

$$h(T, \alpha) = \limsup_{n \rightarrow +\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i} \alpha) = \limsup_{n \rightarrow +\infty} \frac{1}{n} H(\alpha) = 0.$$

Since this holds for any cover  $\alpha$ , we see that  $h(T) = 0$ .

We shall describe a method of computing topological entropy.

DEFINITION. We call a finite cover  $\alpha$  a *generator* for a homeomorphism  $T : X \rightarrow X$  if  $\forall \epsilon > 0 \exists N > 0$  the cover  $\bigvee_{n=-N}^N T^{-n} \alpha = \{B_1, \dots, B_m\}$  consists of open sets each of which has diameter at most  $\epsilon$ , i.e.  $\sup_i \{\text{diam}(B_i)\} \leq \epsilon$ .

We call a finite cover  $\alpha$  a (*strong*) *generator* for a continuous map  $T : X \rightarrow X$  if  $\forall \epsilon > 0 \exists N > 0$  the cover  $\bigvee_{n=0}^N T^{-n} \alpha = \{B_1, \dots, B_m\}$  consists of open sets each of which has diameter at most  $\epsilon$ , i.e.  $\sup_i \{\text{diam}(B_i)\} < \epsilon$ .

This brings us to the following useful result.

PROPOSITION 3.3. *If  $\alpha$  is a strong generating cover for a continuous map  $T : X \rightarrow X$  (or a generator for a homeomorphism  $T : X \rightarrow X$ ) then  $h(T, \alpha) = h(T)$*

PROOF. We prove the result for a strong generator (the proof for a generator being similar). Let  $\beta$  be an arbitrary cover. Let  $\delta$  be the Lebesgue number for  $\beta$  (i.e. every ball of diameter  $\delta$  is contained inside some open set in the cover  $\beta$ ). For sufficiently large  $N$  we get that  $\forall A \in \bigvee_{n=0}^{N-1} T^{-n} \alpha, \exists B \in \beta$  with  $A \subset B$ , from the definition of  $\alpha$  being generating. This means that  $N(\bigvee_{i=0}^{k-1} T^{-i} (\bigvee_{n=0}^{N-1} T^{-n} \alpha)) \leq N(\bigvee_{i=0}^{k-1} T^{-i} \beta)$ , for  $k \geq 1$ .

We claim that  $h(T, \alpha) = h(T, \bigvee_{n=0}^N T^{-n} \alpha)$ . In fact, the equality

$$H(\bigvee_{i=0}^{k-1} T^{-i} (\bigvee_{n=0}^N T^{-n} \alpha)) = H(\bigvee_{i=0}^{k+N-1} T^{-i} \alpha), \quad k \geq 1,$$

gives

$$\begin{aligned} h(T, \bigvee_{n=0}^N T^{-n} \alpha) &= \limsup_{k \rightarrow +\infty} \frac{1}{k} H(\bigvee_{i=0}^{k-1} T^{-i} (\bigvee_{n=0}^N T^{-n} \alpha)) \\ &= \limsup_{k \rightarrow +\infty} \frac{1}{k} H(\bigvee_{i=0}^{k+N-1} T^{-i} \alpha) = h(T, \alpha). \end{aligned}$$

Altogether we see that  $h(T, \alpha) \geq h(T, \beta)$ . Since this holds for all open covers  $\beta$  we see that  $h(T) = h(T, \alpha)$ . ■

EXAMPLE (FULL SHIFT ON  $k$  SYMBOLS). Let  $X = \prod_{n \in \mathbb{Z}} \{1, \dots, k\}$  and  $\sigma : X \rightarrow X$  be the shift map. We can choose a cover by open sets of the form

$$\alpha = \{[1]_0, \dots, [k]_0\}, \text{ where } [i]_0 = \{x = (x_n) \in X : x_0 = i\},$$

for  $i = 1, \dots, k$ . Observe that

$$\bigvee_{n=-N}^N \sigma^{-n} \alpha = \{[i_{-N}, \dots, i_0, \dots, i_N]_{-N}^N : i_{-N}, \dots, i_0, \dots, i_N \in \{1, \dots, k\}\}$$

where we write

$$[i_{-N}, \dots, i_0, \dots, i_N]_{-N}^N = \{x = (x_n) \in X : x_j = i_j \text{ for } -N \leq j \leq N\}.$$

For any  $\epsilon > 0$  we can choose  $N > 0$  sufficiently large that  $\frac{1}{2^N} \leq \epsilon$ ; then for any  $[i_{-N}, \dots, i_0, \dots, i_N]_{-N}^N \in \bigvee_{n=-N}^N \sigma^{-n} \alpha$  we have that

$$\text{diam}([i_{-N}, \dots, i_0, \dots, i_N]_{-N}^N) \leq \epsilon$$

(i.e. the open cover  $\alpha$  is a generator). Moreover, each of the sets in  $\bigvee_{n=-N}^N \sigma^{-n} \alpha$  is disjoint and  $\bigvee_{n=-N}^N \sigma^{-n} \alpha$  contains  $k^{2N+1}$  elements. This means that  $N(\bigvee_{n=-N}^N \sigma^{-n} \alpha) = k^{2N+1}$ . Thus  $h(T) = h(T, \alpha) = \log k$ .

### 3.2 The Perron-Frobenius theorem and subshifts of finite type

We want to consider a subshift of finite type  $\sigma : X_A \rightarrow X_A$  where  $A$  is a  $k \times k$  matrix satisfying the following property.

DEFINITION. The matrix  $A$  is called *aperiodic* if  $\exists N > 0, \forall 1 \leq i, j \leq k, A^N(i, j) \geq 1$ .

This property is stronger than that of irreducibility introduced in chapter 1.

In order to understand the topological entropy we need the following.

LEMMA 3.4 (PERRON-FROBENIUS). *Let  $A$  be an aperiodic  $k \times k$  matrix; then there exists a positive eigenvalue  $\lambda_1 > 0$  such that all other eigenvalues  $\lambda_i$  satisfy  $|\lambda_i| < \lambda_1, i = 2, \dots, k$ . Moreover, there is a unique positive vector  $v = (v_1, \dots, v_k)$  such that  $Av = \lambda_1 v$ .*

PROOF. We denote the *positive cone* in  $\mathbb{R}^k$  by

$$\mathcal{C} = \{x = (x_1, \dots, x_k) \in \mathbb{R}^k : x_i \geq 0, 1 \leq i \leq k\}.$$

Observe that  $A : \mathcal{C} \rightarrow \mathcal{C}$  where

$$A(v_1, \dots, v_k)^T = (v'_1, \dots, v'_k)^T = \left( \sum_{i=1}^k A(i, 1)v_i, \dots, \sum_{i=1}^k A(i, k)v_i \right)$$

and so  $v'_i \geq 0, i = 1, \dots, k$  (i.e.  $Av \in \mathcal{C}$ ). We denote the *standard simplex* in  $\mathbb{R}^k$  by

$$\mathcal{S} = \{x = (x_1, \dots, x_k) \in \mathbb{R}^k : x_i \geq 0, 1 \leq i \leq k, \text{ and } \sum_{i=1}^k x_i = 1\} \subset \mathcal{C}$$

and we can define  $T : \mathcal{S} \rightarrow \mathcal{S}$  by

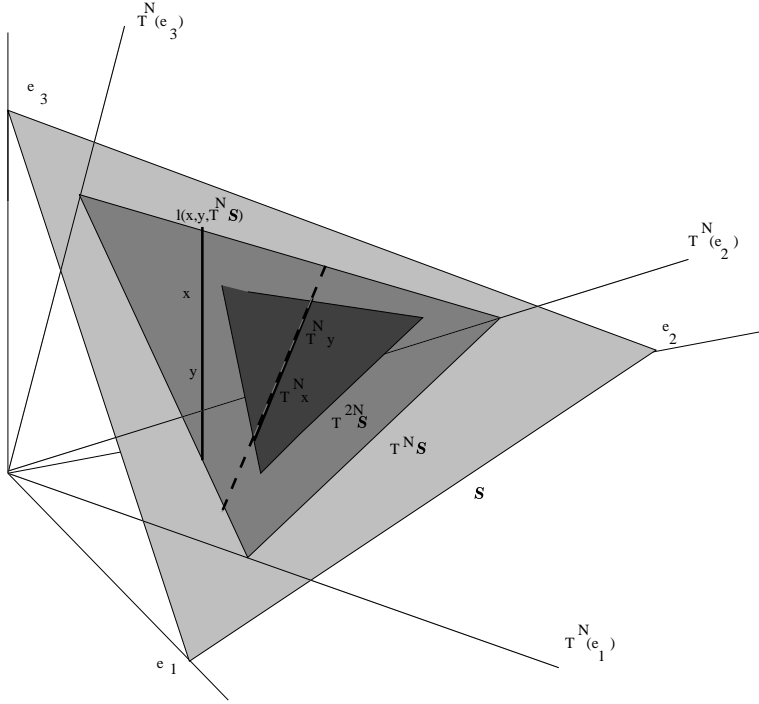
$$T(v_1, \dots, v_k) = \left( \frac{\sum_{i=1}^k A(i, 1)v_i}{\sum_{i=1}^k \sum_{j=1}^k A(i, j)v_i}, \dots, \frac{\sum_{i=1}^k A(i, k)v_i}{\sum_{i=1}^k \sum_{j=1}^k A(i, j)v_i} \right).$$

Observe that for each  $n \geq 0$ , we can write

$$T^n \mathcal{S} = \left\{ \sum_{i=1}^n x_i T^n(e_i) : x_i \geq 0, 1 \leq i \leq k, \text{ and } \sum_{i=1}^k x_i = 1 \right\}$$

where  $e_i, 1 \leq i \leq k$ , are the standard basis vectors for  $\mathbb{R}^n$ . Notice that:

- (1) because of the assumption that  $A$  is aperiodic, we know that for some  $N > 0$  we have  $A^N(i, j) = 1$ , for all  $1 \leq i, j \leq k$ . In particular, we see that  $T^N \mathcal{S} \subset \text{int}(\mathcal{S})$  (see figure 3.1); and
- (2)  $T^N(\mathcal{S})$  is convex.

FIGURE 3.1. The image  $T(\mathcal{S}) \subset \text{int}(\mathcal{S})$ 

We would like to show that  $\mathcal{S} \supset T^N(\mathcal{S}) \supset T^{2N}(\mathcal{S}) \supset \dots \supset T^{nN}(\mathcal{S}) \supset \dots$  converges to a single point by using the contraction mapping theorem applied to a suitable metric.

Given two points  $x, y \in \text{int}(T^N(\mathcal{S}))$  let  $l(x, y, T^N(\mathcal{S}))$  be the line segment given by the intersection of the convex set  $T^N(\mathcal{S})$  with the unique line in  $\mathcal{S}$  through  $x$  and  $y$ . We can always translate each such segment  $l(x, y, T^N(\mathcal{S}))$  to the standard one-dimensional simplex  $\Delta = \{(x_1, x_2) \in \mathbb{R}^2 : x_1, x_2 \geq 0, x_1 + x_2 = 1\}$  by a surjective affine transformation  $L = L_{x, y, T^N} : l(x, y, T^N(\mathcal{S})) \rightarrow \Delta$ . Furthermore, we can identify  $\Delta$  with  $\mathbb{R}^+$  by  $(x_1, x_2) \mapsto \frac{x_1}{x_2}$ .

If we write  $L(x) = (x_1, x_2)$  and  $L(y) = (y_1, y_2)$  then we can introduce a metric on  $\text{int}(T^N \mathcal{S})$  by

$$d(x, y) = \left| \log \left( \frac{x_2 y_1}{x_1 y_2} \right) \right|.$$

We claim that  $T^N : T^N(\mathcal{S}) \rightarrow T^{2N}(\mathcal{S}) \subset T^N(\mathcal{S})$  is a contraction with respect to this metric (i.e. there exists  $0 < C < 1$  such that  $d(T^N x, T^N y) \leq Cd(x, y)$ ).

We can affinely identify each of the line segments  $l(x, y, T^N(\mathcal{S}))$  and  $l(T^N x, T^N y, T^N(\mathcal{S}))$  with  $\Delta$ , and hence  $\mathbb{R}^+$ , as above. The transformation  $P : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  corresponding to  $T^N : l(x, y, T^N(\mathcal{S})) \rightarrow l(T^N x, T^N y, T^N(\mathcal{S}))$  is a linear fractional transformation  $P(z) = \frac{az+b}{cz+d}$  ( $a, b, c, d \in \mathbb{R}^+$ ). The cor-

responding metric on  $\mathbb{R}$  is given by

$$\theta(z_1, z_2) = \left| \log \left( \frac{z_1}{z_2} \right) \right|.$$

To show that  $T^N$  is a contraction it suffices to show that each such corresponding  $P$  is a contraction, and that the contraction constant  $C < 1$  is independent of the choices  $x, y$ . Using the usual euclidean distance on  $\mathbb{R}^+$  we see that  $P'(z) = \frac{ad-bc}{(cz+d)^2}$  and since the ratio of the  $\theta$  distance relative to the euclidean distance at  $z \in \mathbb{R}^+$  is given by  $\frac{1}{z}$  we conclude that  $P$  is Lipschitz with constant

$$C = \sup_{z \in \mathbb{R}} \left\{ z \cdot \left( \frac{ad-bc}{(cz+d)^2} \right) \left( \frac{cz+d}{az+b} \right) \right\}.$$

However, elementary calculus shows that the supremum is realised where  $z = \left(\frac{bd}{ac}\right)^{1/2}$  and thus

$$C < \frac{\nu^{1/2} - \nu^{-1/2}}{\nu^{1/2} + \nu^{-1/2}} < 1,$$

where  $\nu := \frac{ad}{bc} > 1$ .

Finally, since  $\log(\nu) = \log\left(\frac{ad}{bc}\right) = \theta(P(0), P(+\infty))$  we conclude from the fact that  $T^N(T^N\mathcal{S}) \subset \text{int}(T^N\mathcal{S})$  that  $\nu$  is uniformly bounded over all choices  $l(x, y, T^N\mathcal{S})$ . In particular, we can choose a value  $C < 1$  valid for all such choices.

This completes the proof of the lemma. ■

This lemma has immediate applications to the topological entropy of mixing subshifts of finite type.

**PROPOSITION 3.5.** *Let  $A$  be an aperiodic  $k \times k$  matrix with entries 0 or 1. Let  $\sigma : X_A \rightarrow X_A$  be the associated subshift of finite type and let  $\lambda_1$  be the maximal positive eigenvalue; then  $h(\sigma) = \log \lambda_1$ .*

**PROOF.** We can choose a cover for  $X_A$  by the disjoint open sets of the form  $\alpha = \{[1]_0, \dots, [k]_0\}$  where  $[i]_0 = \{x = (x_n) \in X_A : x_0 = i\}$ ,  $i = 1, \dots, k$ . We see that  $\alpha$  is a generating cover.

Since all of the open sets in the cover  $\bigvee_{n=0}^{N-1} \sigma^{-n} \alpha$  are disjoint we see that

$$\begin{aligned} H\left(\bigvee_{n=0}^{N-1} \sigma^{-n} \alpha\right) &= \log \text{Card}\left(\bigvee_{n=0}^{N-1} \sigma^{-n} \alpha\right) \\ &= \log \left( \sum_{i,j} A^{N-1}(i, j) \right). \end{aligned}$$

From elementary linear algebra there are an invertible matrix  $U$  and a matrix

$$D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & B_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & B_l \end{pmatrix},$$



with Jordan block matrices  $B_2, \dots, B_l$ , such that  $A = UDU^{-1}$ . In particular, we can write

$$\begin{aligned} \sum_{i,j} A^N(i,j) &= \sum_{i,j} (UD^N U^{-1})(i,j) \\ &= C\lambda_1^N + E(n) \end{aligned}$$

where

- (i)  $C = \left( \sum_{i,j} U(i,1)U^{-1}(1,j) \right)$ , and
- (ii)  $\lim_{N \rightarrow +\infty} \frac{|E(N)|}{\lambda_1^N} = 0$ .

Thus we see that

$$\lim_{N \rightarrow +\infty} \frac{1}{N} H(\vee_{i=0}^{N-1} \sigma^{-i} \alpha) = \log \lambda_1.$$

■

### 3.3 Other definitions and examples

There are equivalent definitions of topological entropy which, in some examples, are easier to apply. In this section we shall describe two such definitions.

**DEFINITION.** Let  $T : X \rightarrow X$  be a continuous map on a compact metric space.

- (1) For  $n \geq 1$  and  $\epsilon > 0$  we call a finite set  $S \subset X$  an  $(n, \epsilon)$ -separated set if for distinct points  $x, y \in S$  we have that  $d(T^i x, T^i y) \geq \epsilon$  for some  $0 \leq i \leq n-1$ .

Let  $s(n, \epsilon) \in \mathbb{N}$  denote the maximal cardinality of any  $(n, \epsilon)$ -separated set.

- (2) For  $n \geq 1$  and  $\epsilon > 0$  we call a finite set  $R \subset X$  an  $(n, \epsilon)$ -spanning set if  $\forall x \in X, \exists y \in R$  such that  $d(T^i x, T^i y) < \epsilon$  for all  $i = 0, \dots, n-1$ .

Let  $r(n, \epsilon) \in \mathbb{N}$  denote the least cardinality of any  $(n, \epsilon)$ -spanning set.

**LEMMA 3.6.**

- (i) For  $\epsilon > \epsilon'$  we have that  $s(n, \epsilon') \geq s(n, \epsilon)$  and  $r(n, \epsilon') \geq r(n, \epsilon)$ . In particular,

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log(s(n, \epsilon')) \geq \limsup_{n \rightarrow +\infty} \frac{1}{n} \log(s(n, \epsilon))$$

and

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log(r(n, \epsilon')) \geq \limsup_{n \rightarrow +\infty} \frac{1}{n} \log(r(n, \epsilon)).$$

- (ii) For  $\epsilon > 0$  and  $n \geq 1$  we have that

$$r(n, \epsilon) \leq s(n, \epsilon) \text{ and } s(n, 2\epsilon) \leq r(n, \epsilon).$$

PROOF. Part (i) follows directly from the definitions.

For part (ii) we first observe from the definitions that  $s(n, \epsilon) \geq r(n, \epsilon)$  since an  $(n, \epsilon)$ -separated set  $S$  of maximum cardinality  $s(n, \epsilon)$  must also be an  $(n, \epsilon)$ -spanning set.

If  $R$  is an  $(n, \epsilon)$ -spanning set of least cardinality  $r(n, \epsilon)$  then we know that  $X = \cup_{x \in R} D(x, n, \epsilon)$  where

$$D(x, n, \epsilon) = \{y \in X : d(T^i x, T^i y) < \epsilon, \quad i = 1, \dots, n\}.$$

If  $S$  is an  $(n, 2\epsilon)$ -separated set, then for each  $y \in S$  we can choose a *distinct* open set  $D(x, n, \epsilon) \ni y$ , with  $x \in R$  (since if  $y, y' \in S$  and  $x \in R$  with  $y, y' \in D(x, n, \epsilon)$  then  $d(T^i y, T^i y') \leq d(T^i y, T^i x) + d(T^i x, T^i y') < \epsilon + \epsilon = 2\epsilon$  which, since  $S$  is  $(n, 2\epsilon)$ -separated, implies that  $y = y'$ ). Thus  $\text{Card}(S) \leq r(n, \epsilon)$  for all such sets  $S$ , i.e.  $s(n, 2\epsilon) \leq r(n, \epsilon)$ . ■

The values  $r(n, \epsilon)$  and  $s(n, \epsilon)$  tend to be more tractable than, say,  $N(\bigvee_{i=0}^{n-1} T^{-i} \alpha)$ . These quantities are compared by the following lemma.

LEMMA 3.7.

(i) *If  $\alpha$  is a finite open cover for  $X$  with Lebesgue number  $\delta$  then*

$$N(\bigvee_{i=0}^{n-1} T^{-i} \alpha) \leq r(n, \delta)$$

*for all  $n \geq 1$ .*

(ii) *If  $\epsilon > 0$  and  $\gamma = \{B_1, \dots, B_k\}$  is an open cover with  $\max_{1 \leq i \leq k} \text{diam}(B_i) < \epsilon$ , then*

$$s(n, \epsilon) \leq N(\bigvee_{i=0}^{n-1} T^{-i} \gamma)$$

PROOF. (i) If  $R$  is an  $(n, \delta)$ -spanning set of maximum cardinality  $r(n, \delta)$  then  $X = \cup_{x \in R} D(x, n, \delta)$ . However, for each  $x \in R$  we have  $B(T^j x, \delta) \subset A_{i_j}$ , where  $A_{i_j} \in \alpha$ , for  $0 \leq j \leq n-1$ , i.e.  $D(x, n, \delta) \subset A_{i_0} \cap T^{-1} A_{i_1} \cap \dots \cap T^{-(n-1)} A_{i_{n-1}} \in \bigvee_{i=0}^{n-1} T^{-i} \alpha$ . In particular, since these sets form a subcover for  $X$  we see that  $N(\bigvee_{i=0}^{n-1} T^{-i} \alpha) \leq r(n, \delta)$ .

(ii) Let  $S$  be a  $(n, \epsilon)$ -separated set of cardinality  $s(n, \epsilon)$ . Each point  $x \in S$  must lie in a different element of  $\bigvee_{i=0}^{n-1} T^{-i} \gamma$  (since for  $x, y \in B_{i_0} \cap T^{-1} B_{i_1} \cap \dots \cap T^{-(n-1)} B_{i_{n-1}} \in \bigvee_{i=0}^{n-1} T^{-i} \gamma$  we see that  $d(T^r x, T^r y) \leq \text{diam}(B_{i_r}) < \epsilon$ ).

In particular,  $s(n, \epsilon) \leq N(\bigvee_{i=0}^{n-1} T^{-i} \gamma)$ . ■

This gives us the following equivalent definitions of topological entropy

PROPOSITION 3.8. *The topological entropy of a continuous map  $T : X \rightarrow X$  on a compact metric space is given by*

$$h(T) = \lim_{\epsilon \rightarrow 0} \limsup_{n \rightarrow +\infty} \frac{1}{n} \log (r(n, \epsilon))$$

and

$$h(T) = \lim_{\epsilon \rightarrow 0} \limsup_{n \rightarrow +\infty} \frac{1}{n} \log (s(n, \epsilon)).$$

PROOF. Since  $h(T) = \sup_{\alpha} h(T, \alpha)$  we can choose for any  $\eta > 0$  a finite cover  $\alpha$  such that

$$h(T, \alpha) + \eta \geq h(T) \geq h(T, \alpha).$$

Let  $\delta > 0$  be the Lebesgue number for the open cover  $\alpha$ .

By Lemma 3.6 we see that the two limits are the same, i.e.

$$\lim_{\epsilon \rightarrow 0} \limsup_{n \rightarrow +\infty} \frac{1}{n} \log (r(n, \epsilon)) = \lim_{\epsilon \rightarrow 0} \limsup_{n \rightarrow +\infty} \frac{1}{n} \log (s(n, \epsilon)). \quad (3.1)$$

Observe that we have the following inequalities:

$$\begin{aligned} & \lim_{\epsilon \rightarrow 0} \limsup_{n \rightarrow +\infty} \frac{1}{n} \log (r(n, \epsilon)) \\ & \geq \limsup_{n \rightarrow +\infty} \frac{1}{n} \log (r(n, \delta)) \quad (\text{by Lemma 3.6 (i)}) \\ & \geq \limsup_{n \rightarrow +\infty} \frac{1}{n} \log N (\vee_{i=0}^{n-1} T^{-i} \alpha) \quad (\text{by Lemma 3.7 (i)}) \\ & = h(T, \alpha) \geq h(T) - \eta. \end{aligned} \quad (3.2)$$

For any  $\epsilon > 0$  we can choose an open cover  $\beta = \{B_1, \dots, B_k\}$  for  $X$  with  $\max_{1 \leq i \leq k} \{\text{diam}(B_i)\} < \epsilon$ . This gives the following inequalities:

$$\begin{aligned} & \limsup_{n \rightarrow +\infty} \frac{1}{n} \log (s(n, \epsilon)) \\ & \leq \limsup_{n \rightarrow +\infty} \frac{1}{n} \log N (\vee_{i=0}^{n-1} T^{-i} \beta) \quad (\text{by Lemma 3.7 (ii)}) \\ & \leq h(T, \beta) \leq h(T). \end{aligned}$$

Letting  $\epsilon \rightarrow 0$  now gives that

$$\lim_{\epsilon \rightarrow 0} \limsup_{n \rightarrow +\infty} \frac{1}{n} \log (s(n, \epsilon)) \leq h(T). \quad (3.3)$$

Comparing the estimates (3.1),(3.2) and (3.3) (and recalling that  $\eta > 0$  can be chosen arbitrarily small) completes the proof. ■

EXAMPLE (ROTATIONS). Let  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  by  $Tx = x + \alpha \pmod{1}$ . We claim that  $h(T) = 0$ .

Observe that for  $\epsilon > 0$  and  $n = 1$  we choose a finite cover for  $[0, 1]$  consisting of intervals  $(x_i - \frac{\epsilon}{2}, x_i + \frac{\epsilon}{2})$ ,  $i = 1, \dots, N$ , say. Thus  $R = \{x_i : i = 1, \dots, N\}$  is a  $(1, \epsilon)$ -spanning set and  $r(1, \epsilon) \leq N$ .

We claim that  $R$  is also an  $(n, \epsilon)$ -spanning set for each  $n \geq 1$ . For any  $x \in [0, 1]$  we can choose  $d(x, x_i) \leq \epsilon$  then  $d(T^r x_i, T^r x) = d(x, x_i) < \epsilon$ , for  $r = 0, \dots, N - 1$ . In particular,  $r(n, \epsilon) \leq N$ . Thus we see that

$$\begin{aligned} 0 \leq h(T) &= \lim_{\epsilon \rightarrow 0} \left( \limsup_{n \rightarrow +\infty} \frac{1}{n} \log r(n, \epsilon) \right) \\ &\leq \lim_{\epsilon \rightarrow 0} \left( \limsup_{n \rightarrow +\infty} \frac{1}{n} \log N \right) \\ &= 0. \end{aligned}$$

REMARK. This proof works equally well for any isometry  $T : X \rightarrow X$  on a compact metric space.

EXAMPLE (DOUBLING MAP). Consider the map  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  defined by  $Tx = 2x \pmod{1}$ .

For  $k \geq 1$  we can define the set

$$F_k = \left\{ \frac{m}{2^k} : m = 0, \dots, 2^k - 1 \right\}$$

(whose cardinality is  $2^k$ ). For  $\epsilon > 0$  choose  $k \geq 1$  such that  $\frac{1}{2^k} \leq \epsilon < \frac{1}{2^{k-1}}$ .

For  $n \geq 1$  we claim that  $F_{n+k-2}$  is an  $(n, \epsilon)$ -separated set. This is clear since for distinct points  $\frac{m_1}{2^{k+n-2}}, \frac{m_2}{2^{k+n-2}} \in F_{n+k-2}$  we have that  $d(T^{n-1}(\frac{m_1}{2^{k+n}}), T^{n-1}(\frac{m_2}{2^{k+n}})) = d(\frac{m_1}{2^{k-1}}, \frac{m_2}{2^{k-1}}) \geq \frac{1}{2^{k-1}} > \epsilon$ . Thus  $s(n, \epsilon) \geq 2^{k+n-2}$ .

For  $n \geq 1$  we claim that  $F_{n+k-1}$  is an  $(n, \epsilon)$ -spanning set. For any  $x \in \mathbb{R}/\mathbb{Z}$  we can choose  $\frac{m}{2^{k+n-1}} \in F_{n+k-1}$  with  $d(x, \frac{m}{2^{k+n-1}}) < \frac{1}{2^{k+n-1}}$ . Then for  $r = 0, 1, \dots, n - 1$  we have that  $d(T^r \frac{m}{2^{k+n-1}}, T^r x) < \frac{1}{2^{k+1}} < \epsilon$ . Thus  $r(n, \epsilon) \leq 2^{k+n-1}$ .

We now know that

$$h(T) = \lim_{\epsilon \rightarrow 0} \left( \limsup_{n \rightarrow +\infty} \frac{1}{n} \log s(n, \epsilon) \right) \geq \log 2$$

and

$$h(T) = \lim_{\epsilon \rightarrow 0} \left( \limsup_{n \rightarrow +\infty} \frac{1}{n} \log r(n, \epsilon) \right) \leq \log 2.$$

Together these inequalities show that  $h(T) = \log 2$ .

Finally, we give a result that will be useful later on.

**COROLLARY 3.8.1 (ABRAMOV'S THEOREM).** *Let  $T : X \rightarrow X$  be a continuous map on a compact metric space  $X$ ; then for any  $m \geq 1$  we have  $h(T^m) = mh(T)$ .*

**PROOF.** Let  $T : X \rightarrow X$  be a continuous map on a compact metric space  $X$ . For  $m \geq 1$ , let  $s_{T^m}(n, \epsilon)$  denote the maximal cardinality of an  $(n, \epsilon)$ -separated set for  $T^m$  and let  $r_{T^m}(n, \epsilon)$  denote the least cardinality of an  $(n, \epsilon)$ -spanning set for  $T^m$ . We shall show that  $s_{T^m}(n, \epsilon) \geq s_T(nm, \epsilon)$  and  $r_{T^m}(n, \epsilon) \leq r_T(nm, \epsilon)$ . Let  $S$  be an  $(n, \epsilon)$ -separated set for  $T^m$  of cardinality  $s_{T^m}(n, \epsilon)$ . Then  $\forall$  (distinct)  $x, x' \in S$ ,  $\exists 0 \leq i \leq n-1$  with  $d(T^{mi}x, T^{mi}x') \geq \epsilon$ . In particular,  $S$  is also an  $(nm, \epsilon)$ -separated set for  $T$ . Thus  $s_{T^m}(n, \epsilon)$  is a lower bound on the maximal cardinality of  $(n, \epsilon)$ -separated sets for  $T$ :  $s_{T^m}(n, \epsilon) \geq s_T(nm, \epsilon)$

Similarly, let  $R$  be an  $(nm, \epsilon)$ -spanning set for  $T$  of maximal cardinality  $r_T(nm, \epsilon)$ ; then  $\forall x \in X$ ,  $\exists z \in R$  such that  $d(T^i x, T^i z) < \epsilon$  for  $0 \leq i \leq nm-1$ . In particular,  $R$  is also an  $(n, \epsilon)$ -spanning set for  $T^m$  and so  $r_{T^m}(n, \epsilon) \leq r_T(nm, \epsilon)$ .

Finally, we have that

$$h(T^m) = \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow +\infty} \frac{1}{n} r_{T^m}(n, \epsilon) \leq \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow +\infty} \frac{1}{n} r_T(nm, \epsilon) = mh(T) \text{ and}$$

$$h(T^m) = \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow +\infty} \frac{1}{n} s_{T^m}(n, \epsilon) \geq \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow +\infty} \frac{1}{n} s_T(nm, \epsilon) = mh(T).$$

which completes the proof. ■

### 3.4 Conjugacy

We begin with the following definition.

**DEFINITION.** Two homeomorphisms (or continuous maps)  $T_1 : X_1 \rightarrow X_1$  and  $T_2 : X_2 \rightarrow X_2$  are *conjugate* if there exists a homeomorphism  $h : X_1 \rightarrow X_2$  such that  $h \circ T_1 = T_2 \circ h$ .

**REMARKS.**

- (i) The existence of such a homeomorphism requires  $X_1$  and  $X_2$  to be homeomorphic (before any consideration is made of the maps  $T_1$  and  $T_2$ ).
- (ii) The condition  $T_2 \circ h = h \circ T_1$  means that  $h$  respects the orbits of  $T_1$  and  $T_2$ , i.e. for any  $x \in X_1$  and  $n \geq 0$  (or  $n \in \mathbb{Z}$  for homeomorphisms) we have that  $h(x) \in X_2$  and  $h(T_1^n x) = T_2^n(hx)$
- (iii) Conjugacy is an equivalence relation on the set of all homeomorphisms on compact metric spaces.

The following is easy to prove from the definitions.

PROPOSITION 3.9. *Assume that  $T_1 : X_1 \rightarrow X_1$  and  $T_2 : X_2 \rightarrow X_2$  are conjugate maps then;*

- (1)  $T_1 : X_1 \rightarrow X_1$  is transitive if and only if  $T_2 : X_2 \rightarrow X_2$  is transitive,
- (2)  $T_1 : X_1 \rightarrow X_1$  is minimal if and only if  $T_2 : X_2 \rightarrow X_2$  is minimal.

PROOF. This is immediate from the definitions and the easy observation that the homeomorphic image of a dense set is dense. ■

DEFINITION. Denote by  $\text{Fix}(T^n)$  the set of fixed points for the  $n$ th iterate  $T^n : X \rightarrow X$  of a map  $T : X \rightarrow X$  (also called periodic points). We let  $\text{Card}(\text{Fix}(T^n))$  denote the cardinality of this set (which might be empty, finite or infinite).

Remark (ii) above now gives us that the number of fixed points for  $T^n$  gives us a simple set of invariants for conjugacy.

PROPOSITION 3.10. *Conjugate homeomorphisms  $T_1$  and  $T_2$  have the property that  $h(\text{Fix}(T_1^n)) = \text{Fix}(T_2^n)$  and  $\text{Card}(\text{Fix}(T_1^n)) = \text{Card}(\text{Fix}(T_2^n))$ , for  $n \geq 1$ .*

PROOF. This is immediate from Remark (ii). ■

EXAMPLES.

- (i) Let  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  be the homeomorphism given by  $Tx = x + \alpha \pmod{1}$ .

If  $\alpha$  is irrational then there are no fixed points or periodic points. To see this, assume that  $T^n x = x$ ; then  $x + n\alpha = x + m$  with  $n, m \in \mathbb{Z}$ . Then  $\alpha = \frac{m}{n}$ , contradicting the fact that  $\alpha$  is irrational.

If  $\alpha$  is rational ( $\alpha = \frac{p}{q}$ , where  $p, q \geq 1$  are co-prime, i.e. they have no common divisors) then  $\text{Fix}(T^n) = \mathbb{R}/\mathbb{Z}$  if  $n$  is a multiple of  $q$  and the empty set otherwise.

- (ii) Let  $\sigma : X \rightarrow X$  be a subshift of finite type described by a  $k \times k$  matrix with entries either 0 or 1. A fixed point will be of the form

$$x = (\dots, x_0, x_0, x_0, x_0, x_0, \dots)$$

where  $A(x_0, x_0) = 1$ . We see that  $\text{Card}(\text{Fix}(T)) = \text{trace}(A)$  and similarly  $\text{Card}(\text{Fix}(T^n)) = \text{trace}(A^n)$ .

DEFINITION. Let  $T_1 : X_1 \rightarrow X_1$  and  $T_2 : X_2 \rightarrow X_2$  be continuous maps on compact metric spaces. A continuous map  $h : X_1 \rightarrow X_2$  is called a *semi-conjugacy* if  $h \circ T_1 = T_2 \circ h$  and  $h(X_1) = X_2$ .

If  $h : X_1 \rightarrow X_2$  is homeomorphism with  $h \circ T_1 = T_2 \circ h$  is conjugacy.

The following relates the entropies of semi-conjugate and conjugate maps.

PROPOSITION 3.11. *If  $T_1 : X_1 \rightarrow X_1$  is semi-conjugate to  $T_2 : X_2 \rightarrow X_2$  then  $h(T_1) \geq h(T_2)$ . If  $T_1 : X_1 \rightarrow X_1$  and  $T_2 : X_2 \rightarrow X_2$  are conjugate then  $h(T_1) = h(T_2)$*

PROOF. Let  $\alpha$  be an open cover for  $X_2$ . Then for any  $n \geq 1$  we have that

$$\begin{aligned} H(\bigvee_{i=0}^{n-1} T_2^{-i} \alpha) &= H(h^{-1}(\bigvee_{i=0}^{n-1} T_2^{-i} \alpha)) \\ &= H(\bigvee_{i=0}^{n-1} (h^{-1} T_2^{-i} \alpha)) \\ &= H(\bigvee_{i=0}^{n-1} T_1^{-i} (h^{-1} \alpha)). \end{aligned}$$

In particular,

$$\begin{aligned} h(T_2, \alpha) &= \limsup_{n \rightarrow +\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T_2^{-i} \alpha) \\ &= \limsup_{n \rightarrow +\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T_1^{-i} (h^{-1} \alpha)) \\ &= h(T_1, h^{-1} \alpha). \end{aligned}$$

This gives us that

$$h(T_2) = \sup_{\alpha} h(T_2, \alpha) = \sup_{h^{-1} \alpha} h(T_1, h^{-1} \alpha) \leq \sup_{\beta} h(T_1, \beta) = h(T_1).$$

If  $h : X_1 \rightarrow X_2$  is a homeomorphism then we can interchange  $T_1 : X_1 \rightarrow X_1$  and  $T : X_2 \rightarrow X_2$  in the above argument (and replace  $h : X_1 \rightarrow X_2$  by the homeomorphism  $h^{-1} : X_2 \rightarrow X_1$ ) to get  $h(T_1) \leq h(T_2)$ . ■

### 3.5 Comments and references

A more detailed treatment of topological entropy and its properties can be found in Walters' book [3] (or in Bowen's notes [2]). The original reference for this material is [1].

#### References

1. R. Adler, A. Konheim and M. McAndrew, *Topological entropy*, Trans. Amer. Math. Soc. **114** (1965), 61-85.
2. R. Bowen, *Equilibrium States and Ergodic Theory of Anosov Diffeomorphisms*, Lecture Notes in Mathematics, 470, Springer, Berlin, 1975.
3. P. Walters, *An Introduction to Ergodic Theory*, Graduate Text in Mathematics, 79, Springer, Berlin, 1982.

## INTERVAL MAPS

In this chapter we shall concentrate on the special case of continuous maps on the closed interval  $I = [0, 1]$ . This level of specialization allows us to prove some particularly striking results on periodic points and topological entropy.

## 4.1 Fixed points and periodic points

Let  $T : I \rightarrow I$  be a continuous map of the interval  $I = [0, 1]$  to itself. Recall that a fixed point  $x \in I$  satisfies  $Tx = x$  and that a periodic point (of period  $n$ ) satisfies  $T^n x = x$ . We say that  $x$  has *prime period*  $n$  if  $n$  is the smallest positive integer with this property (i.e.  $T^k x \neq x$  for  $k = 1, \dots, n - 1$ ).

For interval maps a very simple visualization of fixed points exists. We can draw the graph  $\mathcal{G}_T$  of  $T : I \rightarrow I$  and the diagonal  $\mathcal{D} = \{(x, x) : x \in I\}$ .

LEMMA 4.1. *The fixed points  $Tx = x$  occur at the intersection points  $(x, x) \in \mathcal{G}_T \cap \mathcal{D}$  (see figure 4.1).*

*Similarly, if for  $n \geq 2$  we look for intersections of the graph  $\mathcal{G}_{T^n}$  (of  $n$ -compositions  $T^n : I \rightarrow I$ ) with the diagonal  $\mathcal{D}$  then the intersection points  $(x, x) \in \mathcal{G}_{T^n} \cap \mathcal{D}$  are periodic points of period  $n$ .*

LEMMA 4.2. *Assume that we have an interval  $J \subset I$  with  $T(J) \supset J$ ; then there exists a fixed point  $Tx = x \in J$ .*

PROOF. We see that showing that there exists  $Tx = x \in J$  is equivalent to showing the restriction of the graph  $\mathcal{G}_T$  to the portion above  $J$  intersects the diagonal  $\mathcal{D}$ . This is obvious by the intermediate value theorem and figure 4.1. ■

The following simple lemma will prove useful.

LEMMA 4.3. *If  $T : I \rightarrow I$  is a continuous map and  $J_1, J_2 \subset I$  are (closed) sub-intervals with  $T(J_1) \supset J_2$  then we can choose a sub-interval  $J_0 \subset J_1$  with  $T(J_0) = J_2$ .*

PROOF. Let  $J_2 = [a, b]$  and introduce the disjoint closed sets  $A = \{x \in J_1 : T(x) = a\}$  and  $B = \{y \in J_1 : T(y) = b\}$ . Choose  $a' \in A$ ,  $b' \in B$



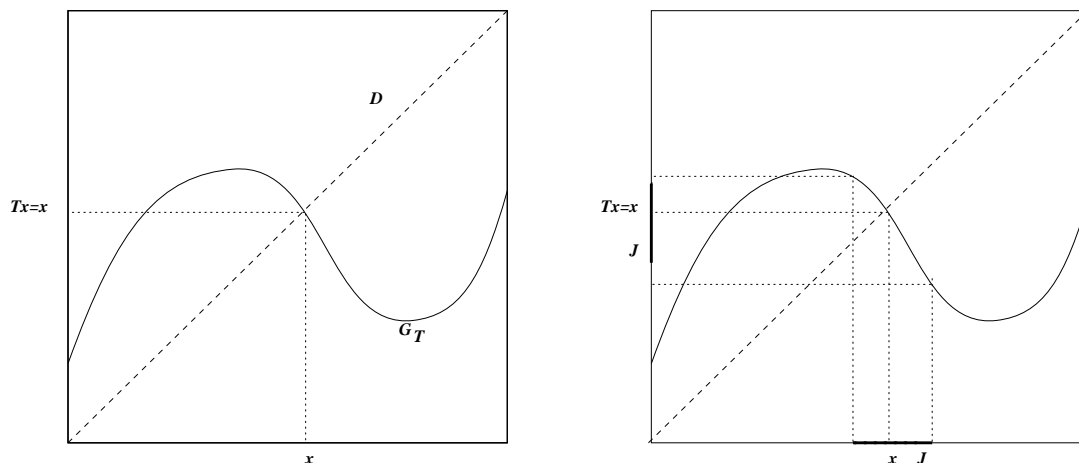


FIGURE 4.1. Characterizing fixed points (Lemma 4.1) and the existence of fixed points in  $J \subset T(J)$  (Lemma 4.2)

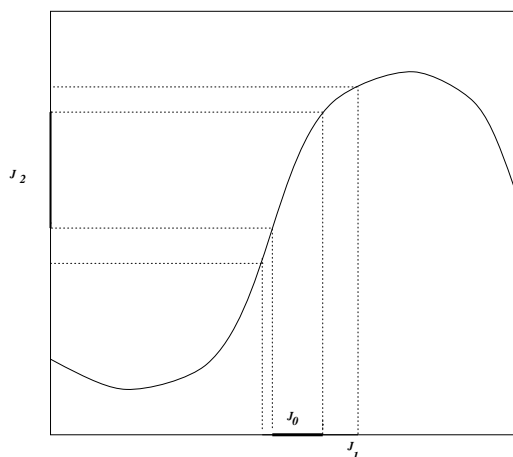


FIGURE 4.2. Choosing a subinterval  $J_0 \subset J_1$  for which  $T(J_0) = J_2$

such that  $|a' - b'| = \inf\{|x - y| : x \in A, y \in B\}$ ; then with  $J_0 = [a', b']$  or  $J_0 = [b', a']$  the results follows. ■

Our first theorem shows the significance of having a periodic point of prime period 3.

**THEOREM 4.4.** *Let  $T : I \rightarrow I$  be a continuous map and suppose there exists a periodic point  $x$  of prime period 3. Then for all  $n \geq 1$  there exists a periodic point of prime period  $n$  (i.e.  $\forall n \geq 1, \exists z \in I$  with  $T^n z = z$ ).*

**PROOF.** We shall do the simpler case  $n = 1$  and the trickier case  $n \geq 2$  separately.

**(I) Existence of a fixed point (i.e.  $n = 1$ ).** Let  $\{x, Tx, T^2x\}$  be the three distinct points in the orbit of  $x$ . Let us assume (for simplicity) that  $x < Tx < T^2x$ . (The five other permutations are easy to derive from this case either by replacing  $x$  by  $Tx$  or  $T^2x$ , or by reversing the horizontal axis of the graph of  $T$ .)

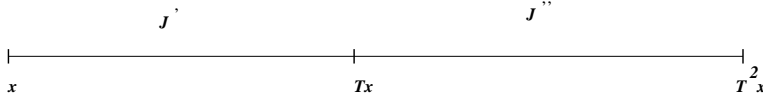


FIGURE 4.3. Intervals  $J'$  and  $J''$  defined with endpoints containing the orbit of  $x$

Let  $J = [x, T^2x]$ ; then we can write  $J = J' \cup J''$  with  $J' = [x, Tx]$  and  $J'' = [Tx, T^2x]$  (see figure 4.3). With these choices we have

- (a)  $T(J') \supset J''$ , and
- (b)  $T(J'') \supset J$

since the endpoints of the intervals  $J'$  and  $J''$  are mapped to the endpoints of  $J'$  and  $J$ , respectively, and the continuous image of an interval is again an interval.

The existence of a fixed point in  $J''$  now follows immediately from Lemma 4.2 and (b) since  $T(J'') \supset J \supset J''$ .

**(II) Existence of a point of period  $n \geq 2$ .** Since by the hypothesis of the theorem we already have a point  $x$  of prime period 3 we shall assume henceforth that  $n \neq 3$ .

**SUBLEMMA 4.4.1.** *There exists a nested sequence of intervals*

$$J'' = I_0 \supset I_1 \supset I_2 \supset \dots \supset I_{n-2} \supset I_{n-1}$$

*with the following properties:*

- (i)  $I_k = T(I_{k+1})$  for  $k = 0, \dots, n-3$ ;
- (ii)  $T^{n-1}(I_{n-1}) = J'$ ; and
- (iii)  $T^n(I_{n-1}) \supset J''$ .

To see that Sublemma 4.4.1 implies Theorem 4.4 we first observe that by part (iii) we have that  $T^n(I_{n-1}) \supset J'' \supset I_{n-1}$  and so applying Lemma 4.2 (with  $T$  replaced by the  $n$ -fold composition  $T^n$ ) shows the existence of a fixed point  $z = T^n z \in I_{n-1}$  for  $T^n$  (i.e.  $z$  is a point of period  $n$  for  $T : I \rightarrow I$ ). However, we still have to show that this is a periodic point of *prime* period  $n$ . We see from Sublemma 4.4.1 that

$$\begin{aligned} z, Tz, T^2z, \dots, T^{n-2}z &\in J'' \text{ (by part (i) since } T^i z \in T^i(I_{n-1}) = I_{n-i-2}), \\ T^{n-1}z &\in J' \text{ (by part (ii)).} \end{aligned} \tag{4.1}$$

To proceed, we want to eliminate the possibility that  $z = Tx (\in J' \cap J'')$ . Assume for a contradiction that  $z = Tx$ , then we immediately have  $T^3z = z$  (since  $T^3x = x$ ) and  $z$  is a periodic point of prime period 3 (since the same is true of  $x$ ). However, this contradicts our assumption that  $n \neq 3$ .

In particular, this means that in (4.1) we can “improve” the second conclusion to  $T^{n-1}z \notin J''$ . We are now in a position to see that  $n$  is the prime period of  $z$ . If this were not the case, then  $T^kz = z$  for some  $1 \leq k \leq n-1$  (which must divide  $n$ ). But this would mean, in particular, that  $T^{k-1}z = T^{n-1}z \notin J''$  which is inconsistent with the first line of (4.1).

The only thing that remains in order to complete the proof of Theorem 4.4 is to prove Sublemma 4.4.1.

PROOF OF SUBLEMMA 4.4.1.

(i) We know from (b) that  $T(J'') \supset J \supset J''$  and so by Lemma 4.3 we can choose  $I_1 \subset J''$  with  $T(I_1) = J''$ . Similarly, since  $T(I_1) = J'' \supset I_1$  we can apply Lemma 4.3 again to choose  $I_2 \subset I_1$  with  $T(I_2) = I_1$ .

Proceeding inductively, we can construct a sequence  $J'' \supset I_1 \supset I_2 \supset \dots \supset I_{n-2}$  with  $T(I_k) = I_{k-1}$  for  $k = 1, 2, \dots, n-2$  (which, in particular, implies that  $T^k(I_k) = J''$  for  $k = 1, 2, \dots, n-2$ ).

(ii) To construct  $I_{n-1}$ , observe that  $T^{n-1}(I_{n-2}) = T(J'') \supset J'$  by (b). Applying Lemma 4.3 we can find  $I_{n-1} \subset I_{n-2}$  with  $T^{n-1}(I_{n-1}) = J'$ .

(iii) Finally, we observe that  $T^n(I_{n-1}) = T(J') \supset J''$  (by (a)).

This completes the proof of the sublemma (and consequently of Theorem 4.4). ■

This result that a point of prime period 3 implies points of all possible prime periods is a special case of a more general result due to Sharkovski.

We can introduce a new ordering on the natural numbers  $\mathbb{N}$  by

$$\begin{aligned}
& 3 \prec 5 \prec 7 \prec 9 \prec 11 \prec \dots \prec 2m+1 \prec \dots \\
& \dots \prec 6 \prec 10 \prec 14 \prec 18 \prec 22 \prec \dots \prec 2(2m+1) \prec \dots \\
& \dots \prec 12 \prec 20 \prec 28 \prec 36 \prec 44 \prec \dots \prec 4(2m+1) \prec \dots \\
& \quad \dots \quad \dots \quad \dots \\
& \dots \prec 2^r \cdot 3 \prec 2^r \cdot 5 \prec 2^r \cdot 7 \prec 2^r \cdot 9 \prec 2^r \cdot 11 \prec \dots \prec 2^r(2m+1) \prec \dots \\
& \quad \dots \quad \dots \quad \dots \\
& \dots \prec 2^{r+1} \prec 2^r \prec 2^{r-1} \prec \dots \prec 16 \prec 8 \prec 4 \prec 2 \prec 1
\end{aligned}$$

This ordering is clearly somewhat different from the usual ordering on the natural numbers. For example, the ordering of the first dozen natural numbers becomes  $3 \prec 5 \prec 7 \prec 9 \prec 11 \prec 6 \prec 10 \prec 12 \prec 8 \prec 4 \prec 2 \prec 1$ .

**THEOREM (SHARKOVSKI).** *Let  $T : I \rightarrow I$  be a continuous map and assume that  $T$  has a point of prime period  $n$ . Then for each  $m > n$  (with respect to the above ordering) there exist periodic points of prime period  $m$ .*

The proof of Sharkovski's theorem runs along similar lines to that of Theorem 4.4.

## 4.2 Topological entropy of interval maps

Given a continuous map  $T : I \rightarrow I$  we want to express the topological entropy in terms of the growth of the number of monotone intervals for the  $n$ th iterate  $T^n$ .

**DEFINITION.** We let  $\mathcal{N}(T)$  denote the number of intervals of monotonicity for the map  $T : I \rightarrow I$  (i.e. the number of disjoint maximal subintervals  $I_1, I_2, \dots, I_k$  for which each restriction  $T : I_j \rightarrow I$  is strictly monotone).

**LEMMA 4.5.** *For continuous maps  $S_1 : I \rightarrow I$  and  $S_2 : I \rightarrow I$  we have that  $\mathcal{N}(S_1 \circ S_2) \leq \mathcal{N}(S_2)\mathcal{N}(S_1)$ . In particular, for a continuous map  $T : I \rightarrow I$*

- (i)  $\mathcal{N}(T^{n+m}) \leq \mathcal{N}(T^n)\mathcal{N}(T^m)$  for  $n, m \geq 1$ ,
- (ii) the sequence  $\mathcal{N}(T^n)$  is monotone increasing.

**PROOF.** Assume that  $\{I_i\}_{i=1}^n$  are disjoint intervals of monotonicity for  $S_1 : I \rightarrow I$  and  $\{J_j\}_{j=1}^m$  are disjoint intervals of monotonicity for  $S_2 : I \rightarrow I$ . The intervals of monotonicity for  $S_1 \circ S_2 : I \rightarrow I$  take the form  $J_j \cap S_2^{-1}(I_i)$ . Thus  $\mathcal{N}(S_1 \circ S_2) \leq \text{Card}\{(i, j) : J_j \cap S_2^{-1}(I_i) \neq \emptyset\} \leq \mathcal{N}(S_1) \cdot \mathcal{N}(S_2)$ . ■

**REMARK.** If we were to also consider  $C^1$  maps  $T : I \rightarrow I$  then the endpoints of the intervals of monotonicity would be the critical points  $\{c : T'(c) = 0\}$ . The above results become even more transparent using the chain rule  $(T^n)'(x) = \prod_{i=0}^{n-1} T'(T^i x)$ .

Lemma 4.5 (i) shows that the sequence  $\log \mathcal{N}(T^n)$  is subadditive and so  $\frac{1}{n} \log \mathcal{N}(T^n)$  converges (to  $\inf \{\frac{1}{n} \log \mathcal{N}(T^n) : n \geq 1\}$ ).

**THEOREM 4.6.** *Assume that  $T : I \rightarrow I$  is a continuous map with  $\mathcal{N}(T) < +\infty$ ; then*

$$h(T) = \lim_{n \rightarrow +\infty} \frac{1}{n} \log \mathcal{N}(T^n).$$

**PROOF.** For each  $n \geq 1$  let  $E_n = \{x_1, x_2, \dots, x_N\}$  be a maximal  $(n, \epsilon)$ -separated set for  $T : I \rightarrow I$  (i.e.  $N = s(n, \epsilon)$ ). By definition, for each  $1 \leq i \leq N$  there exists some  $0 \leq r_i \leq n - 1$  with  $|T^{r_i}(x_i) - T^{r_i}(x_{i+1})| > \epsilon$ . In particular, we can choose  $0 \leq r \leq n - 1$  with  $|T^r(x_{i_j}) - T^r(x_{i_{j+1}})| > \epsilon$  for a subset  $\{x_{i_0} < x_{i_1} < \dots < x_{i_{m-1}}\} \subset E_n$  with cardinality  $m \geq \frac{N}{n}$ .

If  $C = \sup_{x \in I} |T(x)|$  then this property implies there are at least  $\frac{m\epsilon}{C}$  intervals of monotonicity for  $T^r$ , i.e.  $\mathcal{N}(T^r) \geq \frac{m\epsilon}{C} \geq \frac{N\epsilon}{Cn}$ . Thus

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{n} \log \mathcal{N}(T^n) &\geq \lim_{n \rightarrow +\infty} \frac{1}{n} (\log s(n, \epsilon) - (\log C + \log n - \log \epsilon)) \\ &= \lim_{n \rightarrow +\infty} \frac{1}{n} \log s(n, \epsilon) \\ &= h(T). \end{aligned} \tag{4.2}$$

The opposite inequality is slightly more complicated. We begin with some preliminary estimates. We can fix  $m \geq 1$  and consider  $S := T^m : I \rightarrow I$ . Observe that

$$\begin{aligned} m \left( \lim_{n \rightarrow +\infty} \frac{1}{n} \log \mathcal{N}(T^n) \right) &= m \left( \limsup_{n \geq 1} \frac{1}{n} \log \mathcal{N}(T^n) \right) \\ &\geq \limsup_{k \geq 1} \frac{1}{k} \log \mathcal{N}(S^k) \\ &= \lim_{k \rightarrow \infty} \frac{1}{k} \log \mathcal{N}(S^k) \end{aligned}$$

(since  $\left(\frac{m \log \mathcal{N}(S^k)}{k}\right)_{k \in \mathbb{N}}$  is a sub-sequence of  $\left(\frac{\log \mathcal{N}(T^n)}{n}\right)_{n \in \mathbb{N}}$ ). Moreover, we can estimate

$$\begin{aligned} &m \left( \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathcal{N}(T^n) \right) \\ &\leq \lim_{n \rightarrow \infty} \frac{m}{n} \log \left( \mathcal{N}(S^{\lfloor \frac{n}{m} \rfloor}) \mathcal{N}(T^{n - \lfloor \frac{n}{m} \rfloor m}) \right) \quad (\text{by Lemma 4.5}) \\ &\leq \lim_{n \rightarrow \infty} \frac{m}{n} \log \left( \mathcal{N}(S^{\lfloor \frac{n}{m} \rfloor}) + \max_{0 \leq i \leq m-1} \{ \log \mathcal{N}(T^i) \} \right) \\ &= \lim_{k \rightarrow \infty} \frac{1}{k} \log \mathcal{N}(S^k). \end{aligned}$$

Comparing these last two inequalities we see that

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log \mathcal{N}(S^k) = m \left( \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathcal{N}(T^n) \right).$$

We shall now concentrate on the interval map  $S : I \rightarrow I$ . We denote by  $\{J_r\}_{r=0}^M$  the intervals of monotonicity for  $S : I \rightarrow I$  (with neighbouring intervals labelled consecutively).

Let  $\alpha = \{U_r\}_{r=1}^m$  denote the open cover for  $I$  whose elements are open intervals formed from neighbouring intervals from  $\{J_r\}$  (i.e.  $U_1 = \text{int}(J_1 \cup J_2)$ ,  $U_2 = \text{int}(J_1 \cup J_2 \cup J_3), \dots$ ,  $U_r = \text{int}(J_{r-1} \cup J_r \cup J_{r+1}), \dots$ ,  $U_m = \text{int}(J_{m-1} \cup J_m)$ ).

For  $n \geq 1$ , each non-degenerate interval  $J_{i_0} \cap S^{-1}J_{i_1} \cap \dots \cap S^{-(n-1)}J_{i_{(n-1)}}$  is now an interval of monotonicity for  $S^n$  and corresponds to at most  $3^n$  elements of the cover

$$\begin{aligned} & \bigvee_{r=0}^{n-1} S^{-r} \alpha \\ & = \left\{ U_{j_0} \cap S^{-1}U_{j_1} \cap \dots \cap S^{-(n-1)}U_{j_{(n-1)}} : j_0, \dots, j_{(n-1)} \in \{1, \dots, N\} \right\} \end{aligned}$$

(given by the at most three choices  $\text{int}J_{(i_r-1)}$ ,  $\text{int}J_{i_r}$ , and  $\text{int}J_{(i_r+1)}$  contained in  $U_{i_r}$ , for  $0 \leq r \leq n-1$ ). In particular, we see that  $\mathcal{N}(S^k) \leq 3^k N(\bigvee_{r=0}^{k-1} S^{-r} \alpha)$ . By the definition of topological entropy in section 3.2 we have that  $h(S) \geq h(S, \alpha)$  and then we may write

$$\begin{aligned} h(S) & \geq h(S, \alpha) \\ & = \lim_{k \rightarrow +\infty} \frac{1}{k} H(\bigvee_{i=0}^{k-1} S^{-i} \alpha) \\ & \geq \limsup_{k \rightarrow +\infty} \frac{1}{k} \log(\mathcal{N}(S^k)/3^k) \quad (\text{since } \mathcal{N}(S^k) \leq 3^k N(\bigvee_{r=0}^{k-1} S^{-r} \alpha)) \\ & = \limsup_{k \rightarrow +\infty} \frac{1}{k} \log \mathcal{N}(S^k) - \log 3 \end{aligned}$$

Finally, by Corollary 3.8.1 we know that  $h(S) = mh(T)$  and so we have

$$\begin{aligned} h(T) & = \frac{h(S)}{m} \\ & \geq \frac{1}{m} \left( \limsup_{k \rightarrow +\infty} \frac{1}{k} \log \mathcal{N}(S^k) \right) - \frac{\log 3}{m} \\ & = \limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mathcal{N}(T^n) - \frac{\log 3}{m}. \end{aligned}$$

Since  $m \geq 1$  can be arbitrarily large, we get

$$h(T) \geq \limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mathcal{N}(T^n). \quad (4.3)$$

Comparing (4.2) and (4.3) completes the proof. ■

### 4.3 Markov maps

Consider a division of the interval  $I = [0, 1]$  into a finite number of closed sub-intervals  $I_i = [x_i, x_{i+1}]$  ( $i = 0, \dots, k-1$ ) with endpoints  $0 = x_0 < x_1 < \dots < x_k = 1$ .

DEFINITION. We are interested in surjective maps  $T : I \rightarrow I$  which are  $C^1$  and monotone on each of the (open) intervals  $\text{int}I_i = (x_{i-1}, x_i)$  and satisfy the following additional properties.

- (i) (*Piecewise expanding*) There exists  $\beta > 1$  such that  $|T'(x)| \geq \beta$ ,  $\forall x \in I_i$  ( $i = 1, \dots, k$ );
- (ii) (*Markov property*) If  $T(\text{int}I_i) \cap \text{int}I_j \neq \emptyset$  then  $T(\text{int}I_i) \supset \text{int}I_j$  (for  $i, j = 1, \dots, k$ ).

(see figure 4.4)

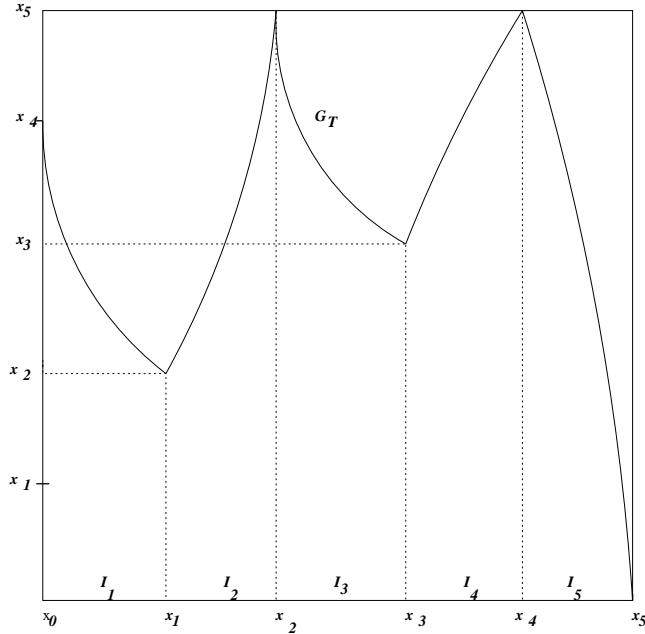


FIGURE 4.4.  $\mathcal{G}_T$  for a Markov map

EXAMPLE. Consider the interval  $I = [0, 1]$  and the map  $T : I \rightarrow I$  defined by

$$T(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq \frac{1}{2}, \\ 2x - 1 & \text{if } \frac{1}{2} < x \leq 1. \end{cases}$$

In this case we take  $I_1 = [0, \frac{1}{2}]$  and  $I_2 = [\frac{1}{2}, 1]$ . It is easy to see that this map is piecewise expanding (where we can take  $\beta = 2$  in (i)). Also, since  $T(0, \frac{1}{2}) = T(\frac{1}{2}, 1) = (0, 1) \supset (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$  it is a simple matter to see that the Markov property holds.

TECHNICAL POINT FOR THE CAUTIOUS. These maps  $T : I \rightarrow I$  may not be continuous at the points  $x_1 < \dots < x_{k-1}$ , which sits a little uncomfortably with our earlier definitions (of topological entropy, etc.) for continuous maps. This can be rectified by the simple device of looking at the map  $T$  on the

*disjoint* union  $\cup_{i=1}^{k-1} I_i$  (i.e. treating the endpoints of different intervals as different points). To avoid this concern, we invite the reader to restrict his attention to continuous maps.

The following is a useful property for such maps to have.

DEFINITION. We say that  $T : I \rightarrow I$  is *locally eventually onto* if for every interval  $J \subset I$  there exists  $n \geq 1$  with  $T^n J = I$ .

We want to associate to a piecewise expanding Markov interval map a  $k \times k$  matrix  $A$  with entries either 0 or 1.

DEFINITION. We define a *transition matrix*  $A$  by

$$A(i, j) = \begin{cases} 1 & \text{if } T(\text{int}I_i) \supset \text{int}I_j, \\ 0 & \text{if } T(\text{int}I_i) \cap \text{int}I_j = \emptyset. \end{cases}$$

We can now define a *one-sided subshift of finite type* on the shift space of sequences,

$$X_A^+ = \left\{ x = (x_n)_{n \in \mathbb{Z}^+} \in \prod_{\mathbb{Z}^+} \{1, \dots, k\} : A(x_n, x_{n+1}) = 1, \text{ for } n \geq 0 \right\},$$

by  $\sigma : X_A^+ \rightarrow X_A^+$  with  $\sigma((x_n)_{n \in \mathbb{Z}^+}) = (x_{n+1})_{n \in \mathbb{Z}^+}$  (i.e. all the terms in the sequence  $x$  are shifted one place to the left, except the zeroth term  $x_0$  which is thrown away).

For  $x = (x_n)_{n \in \mathbb{Z}^+}$  and  $y = (y_n)_{n \in \mathbb{Z}^+}$  we define

$$N^+(x, y) = \min\{N \geq 0 : x_N \neq y_N\}.$$

A natural metric on this space is defined by  $d(x, y) = \left(\frac{1}{2}\right)^{N^+(x, y)}$  for  $x, y \in X_A^+$  ( $x \neq y$ ).

LEMMA 4.7. *The space  $X_A^+$  is compact and the map  $\sigma : X_A^+ \rightarrow X_A^+$  is continuous. Moreover, every point  $x \in X_A^+$  has at most  $k$  pre-images*

PROOF. The proofs of the first two assertions are almost identical to those for  $\sigma : X_A \rightarrow X_A$  in chapter 1, and so we shall omit them.

For the last part, we observe that if  $x = (x_n)_{n \in \mathbb{Z}^+} \in X_A^+$  and  $y = (y_n)_{n \in \mathbb{Z}^+}$  satisfies  $\sigma(y) = x$  then the values  $y_1 = x_0, y_2 = x_1, \dots, y_r = x_{r-1}, \dots$  are all determined, and so  $y$  is completely specified up to the at most  $k$  possible choices  $y_0 \in \{1, \dots, k\}$  (such that  $A(y_0, x_0) = 1$ ). ■

The next proposition shows that  $\sigma : X_A^+ \rightarrow X_A^+$  gives a lot of information about  $T : I \rightarrow I$ .



PROPOSITION 4.8. *There is continuous map  $\pi : X_A^+ \rightarrow I$  such that*

- (i)  $\pi$  is surjective and  $\pi \circ \sigma = T \circ \pi$  (i.e.  $\pi$  is a semi-conjugacy),
- (ii) points  $z \in I$  have exactly one or two pre-images in  $X_A^+$  (i.e.  $\forall z \in I$  the set  $E(z) = \{x \in X_A^+ : \pi(x) = z\}$  consists of either one or two points),
- (iii) the set of points  $z \in I$  such that  $E(z)$  consists of more than one point is contained in the countable set  $\cup_{n \in \mathbb{Z}^+} T^{-n} \{x_0, \dots, x_k\}$ ,
- (iv) if  $T : I \rightarrow I$  is locally eventually onto then  $A$  is aperiodic (i.e.  $\exists n \geq 1, \forall 1 \leq i, j \leq k, A^n(i, j) > 0$ ).

PROOF. We want to define the map  $\pi : X_A^+ \rightarrow I$  by

$$\pi(w) = \bigcap_{n=0}^{\infty} \text{cl} (T^{-n} \text{int}(I_{w_n}))$$

where  $w = (w_n)_{n \in \mathbb{Z}^+} \in X_A^+$  (i.e.  $\pi(w)$  should correspond to a point  $x \in I$  such that  $T^n x \in I_{w_n}$  for  $n \geq 0$ ). The use of  $\text{cl}$  (closure) and  $\text{int}$  (interior) helps to avoid tiresome problems with the endpoints of intervals.

We first show that this map is well-defined. For any  $N \geq 0$  we have that  $J_N(w) = \bigcap_{n=0}^N \text{cl} (T^{-n} \text{int}(I_{w_n}))$  is non-empty and closed. Since  $J_0(w) \supset J_1(w) \supset \dots \supset J_k(w) \supset \dots$  is a nested sequence of closed sets we have by compactness that  $\bigcap_{k \in \mathbb{Z}^+} J_k(w) \neq \emptyset$ . Moreover, since  $T : I \rightarrow I$  is piecewise expanding we see that

$$\begin{aligned} \text{diam}(J_n(x)) &\leq \frac{1}{\beta} \text{diam}(J_{n-1}(x)) \leq \dots \\ &\dots \leq \frac{1}{\beta^{n-1}} \text{diam}(J_1(x)) \leq \frac{1}{\beta^n} \rightarrow 0 \end{aligned}$$

as  $n \rightarrow +\infty$ . In particular, this intersection consists of a single point, which we take to be  $\pi(w)$ .

To see that  $\pi : X_A^+ \rightarrow I$  is continuous, we begin by choosing for each  $\epsilon > 0$  an integer  $n \geq 1$  such that  $\frac{1}{\beta^n} < \epsilon$ . If  $d(x, y) \leq \frac{1}{2^n}$  then  $x_i = y_i$  for  $i = 0, \dots, n-1$  and  $\pi(x), \pi(y) \in J_n(x)$  (by definition of  $J_n(x)$ ). In particular,  $|\pi(x) - \pi(y)| \leq \text{diam} J_n(x) \leq \frac{1}{\beta^n} < \epsilon$ .

(i) To see that  $\pi$  is surjective, we need only observe that for every point  $x$  in the complement of the dense set  $\cup_{n \in \mathbb{Z}^+} T^{-n} \{x_0, \dots, x_k\}$  we have a unique choice  $(w_n)_{n \in \mathbb{Z}^+} \in X_A^+$  with  $T^n(x) \in \text{int}(I_{w_n})$  for  $n \geq 0$ , and thus  $\pi((w_n)_{n \in \mathbb{Z}^+}) = x$ . Since the image of  $\pi$  is compact and contains a dense set, we conclude that  $\pi$  is surjective.

To see that  $\pi \circ \sigma = T \circ \pi$  observe that

$$\begin{aligned} \pi \circ \sigma(w_n)_{n \in \mathbb{Z}^+} &= \pi((w_{n+1})_{n \in \mathbb{Z}^+}) \\ &= \bigcap_{n \in \mathbb{Z}^+} \text{cl} (T^{-n} \text{int}(I_{w_{n+1}})) \\ &= (T \circ \pi)((w_n)_{n \in \mathbb{Z}^+}). \end{aligned}$$

(ii) Assume that  $z = \pi(w) = \pi(y)$  with  $w = (w_n)_{n \in \mathbb{Z}^+} \neq y = (y_n)_{n \in \mathbb{Z}^+}$ . If  $w_0 \neq y_0$  then  $z \in I_{w_0} \cap I_{y_0} \neq \emptyset$ , i.e.  $z$  must be one of the endpoints  $x_0, \dots, x_n$ . However, if  $z = x_i$ , say, then the choice of either  $w_0 = i$  or  $w_0 = i + 1$  uniquely determines the rest of the sequence  $w = (w_n)_{n \in \mathbb{Z}^+}$  with  $\pi(w) = z$ . Thus  $E(z)$  can have cardinality at most two.

More generally, if  $w_i = y_i$  for  $i = 0, \dots, r - 1$  but  $w_r \neq y_r$  then  $T^r z \in I_{w_r} \cap I_{y_r}$ . Thus,  $T^r z = x_j$ , say. Moreover, the choice of  $w_r = j$  or  $w_r = j + 1$  uniquely determines the rest of the sequence  $w = (w_n)_{n \in \mathbb{Z}^+}$  with  $\pi(w) = z$ .

(iii) From the proof of (ii) we see that we require that  $z \in \cup_{n=0}^{\infty} T^{-n}\{x_0, \dots, x_k\}$ . Since  $T$  is Markov this is a countable set.

(iv) For each  $1 \leq i \leq k$  we can use the definition of locally eventually onto to choose  $n_i \geq 1$  with  $T^{n_i} I_i = I$ . In particular, for  $n \geq \max\{n_i : 1 \leq i \leq k\}$  we see that for any  $1 \leq i \leq k$  we have  $T^n I_i = T^{n-n_i}(T^{n_i} I_i) = T^{n-n_i}(I) = I$ . In particular, this means that  $\forall 1 \leq i, j \leq k$  we can find  $x \in I_i \cap T^{-n} I_j$ . If we choose  $i_1, \dots, i_{n-1} \in \{1, \dots, k\}$  such that  $T^r x \in I_{i_r}$  ( $r = 1, \dots, n - 1$ ) then we see that

$$A^n(i, j) \geq A(i, i_1)A(i_1, i_2) \dots A(i_{n-1}, j) = 1.$$

This shows that  $A$  is aperiodic. ■

The usefulness of Proposition 4.8 lies in converting relatively easy results for  $\sigma : X_A^+ \rightarrow X_A^+$  into corresponding results for  $T : I \rightarrow I$ .

**THEOREM 4.9.** *If  $T : I \rightarrow I$  is locally eventually onto then  $h(T) = \log \lambda_1$  where  $\lambda_1$  is the unique maximal eigenvalue for the matrix  $A$ .*

**PROOF OF THEOREM 4.9.** Since  $\pi : X_A^+ \rightarrow I$  is a surjective semi-conjugacy between  $\sigma : X_A^+ \rightarrow X_A^+$  and  $T : I \rightarrow I$  we have by Proposition 3.11 that  $h(T) \leq h(\sigma)$ . To get the reverse inequality we need only argue as in the proof of Theorem 4.6. This shows that  $h(T) = h(\sigma)$ .

To complete the proof, we only need the following sublemma.

**SUBLEMMA 4.9.1.**  $h(\sigma) = \log \lambda_1$ .

**PROOF OF SUBLEMMA 4.9.1.** The proof is identical to that of Proposition 3.5. ■

An algebraic number is one which is the root of a polynomial with integer entries. There are only countably many algebraic numbers.

COROLLARY 4.9.1. *The value  $e^{h(T)}$  is an algebraic number.*

PROOF. Observe that  $\lambda_1 = e^{h(\sigma)}$  is an eigenvalue for  $A$  and so, in particular, a root of the characteristic polynomial  $p(z) = \det(zI - A)$  which has integer coefficients (since the matrix  $A$  has integer entries). By Theorem 4.9 we have  $e^{h(\sigma)} = e^{h(T)}$ . ■

DEFINITION. For each  $n \geq 1$  we let  $\text{Fix}(T^n) = \{x \in I : T^n x = x\}$  denote the set of fixed point for  $T^n$  and let  $\text{Card}(\text{Fix}(T^n))$  be the cardinality of this set.

THEOREM 4.10.  $\lim_{n \rightarrow +\infty} \frac{\text{Card}(\text{Fix}(T^n))}{e^{nh(T)}} = 1$ .

PROOF OF THEOREM 4.10. We first establish the analogous result in the easier context of  $\sigma : X_A^+ \rightarrow X_A^+$ . By Sublemma 4.9.1. the topological entropy of this map is  $\log \lambda_1$ .

SUBLEMMA 4.10.1.  $\lim_{n \rightarrow +\infty} \frac{\text{CardFix}(\sigma^n)}{\lambda_1^n} = 1$

PROOF OF SUBLEMMA 4.10.1. A fixed point  $x = (x_k)_{k \in \mathbb{Z}^+}$  for  $\sigma^n : X_A^+ \rightarrow X_A^+$  is a sequence for which  $x_k = x_{k+n} = x_{k+2n} = \dots$ , for  $k = 0, 1, \dots, n-1$ . Thus  $\text{Card}(\text{Fix}(\sigma^n))$  is exactly the number of strings  $(x_0, x_1, \dots, x_{n-1})$  with  $A(x_0, x_1) = A(x_1, x_2) = \dots = A(x_{n-2}, x_{n-1}) = A(x_{n-1}, x_0) = 1$ . However, this is given by  $\text{trace}(A^n) = \lambda_1^n + \lambda_2^n + \dots + \lambda_k^n$ , where  $\lambda_i$  ( $i = 1, \dots, k$ ) are the eigenvalues for  $A$ .

Since by the Perron-Frobenius theorem we have  $|\lambda_i| < \lambda_1$  for  $i = 2, \dots, k$  we see that

$$\lim_{n \rightarrow +\infty} \frac{\text{Card}(\text{Fix}(\sigma^n))}{\lambda_1^n} = \lim_{n \rightarrow +\infty} \frac{\lambda_1^n + \dots + \lambda_k^n}{\lambda_1^n} = 1. \quad (4.4)$$
■

By Proposition 4.8 (iii) there is a bijection between  $\text{Fix}(\sigma^n)$  and  $\text{Fix}(T^n)$ , with the possible exception of the finite set  $\text{Fix}(T^n) \cap \{x_1, \dots, x_k\}$ . Thus  $|\text{Fix}(\sigma^n) - \text{Fix}(T^n)| \leq k$  and so by (4.4)

$$\begin{aligned} 1 &= \lim_{n \rightarrow +\infty} \frac{\text{Card}(\text{Fix}(\sigma^n)) - k}{e^{nh(\sigma)}} \leq \liminf_{n \rightarrow +\infty} \frac{\text{Card}(\text{Fix}(T^n))}{e^{nh(T)}} \\ &\leq \limsup_{n \rightarrow +\infty} \frac{\text{Card}(\text{Fix}(T^n))}{e^{nh(T)}} \leq \lim_{n \rightarrow +\infty} \frac{\text{Card}(\text{Fix}(\sigma^n)) + k}{e^{nh(\sigma)}} = 1. \end{aligned}$$

This completes the proof. ■

#### 4.4 Comments and references

Sharkovski's theorem on periodic points can be found in [8], [9], [2] or [5] (for period 3 orbits).

The interpretation of the topological entropy as the growth rate of monotone intervals is due to Milnor and Thurston [6].

The use of subshifts of finite type to study Markov interval maps is a simple analogue of Markov Partitions and symbolic dynamics for Axiom A diffeomorphisms (cf. references to chapter 5). If we drop the Markov assumption then there is a construction due to Hofbauer [4] of Markov extensions.

Other important aspects of interval maps which we have omitted are dealt with in [2] (kneading theory), [1] (renormalization) and [3] (period doubling).

#### References

1. P. Collet and J.-P. Eckmann, *Iterated Maps of the Interval as Dynamical Systems*, Birkhäuser, Boston Mass., 1980.
2. R. Devaney, *An Introduction to Chaotic Dynamical Systems*, Addison-Wesley, New York, 1989.
3. W. de Melo and S. van Strien, *One Dimensional Dynamics*, Springer, Berlin, 1993.
4. F. Hofbauer, *On the intrinsic ergodicity of piecewise monotonic transformations*, Israel J. Math. **34** (1979), 213-237.
5. T. Li and J. Yorke, *Period three implies chaos*, Amer. Math. Monthly **82** (1975), 985-992.
6. J. Milnor and W. Thurston, *On iterated maps of the interval*, Dynamical Systems (J. Alexander, ed.), Lecture Notes in Mathematics 1342, Springer, Berlin, 1988, pp. 465-563.
7. F. Schweiger, *Ergodic Theory of Fibred Markov Systems and Metric Number Theory*, O.U.P., Oxford, 1995.
8. Sharkovski, *Coexistence of cycles of a continuous map of the line into itself*, Ukrain. Mat. Zh. **16** (1964), 61-71.
9. P. Stefan, *A theorem of Sharkovski on the existence of periodic orbits of continuous endomorphisms on the real line*, Commun. Math. Phys. **54** (1977), 237-248.

## HYPERBOLIC TORAL AUTOMORPHISMS

We want to consider a simple class of homeomorphisms whose dynamical behaviour illustrates many of the ideas we have discussed.

## 5.1 Definitions

Let  $GL(2, \mathbb{Z})$  be the set of all  $2 \times 2$  matrices  $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ , where  $a, b, c, d \in \mathbb{Z}$  and  $\text{Det}(A) = ad - bc = \pm 1$ . Each such matrix  $A$  gives a linear map on  $\mathbb{R}^2$  by  $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ .

Let  $\mathbb{T}^2 = \mathbb{R}^2 / \mathbb{Z}^2$  be the two-dimensional torus and we define a *linear toral automorphism*  $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  by  $T(x_1, x_2) = (ax_1 + bx_2, cx_1 + dx_2) \pmod{1}$  (see figure 5.1). We say that  $T : \mathbb{T}^2 / \mathbb{Z}^2 \rightarrow \mathbb{T}^2 / \mathbb{Z}^2$  is *hyperbolic* if  $A$  does not have eigenvalues of modulus 1.

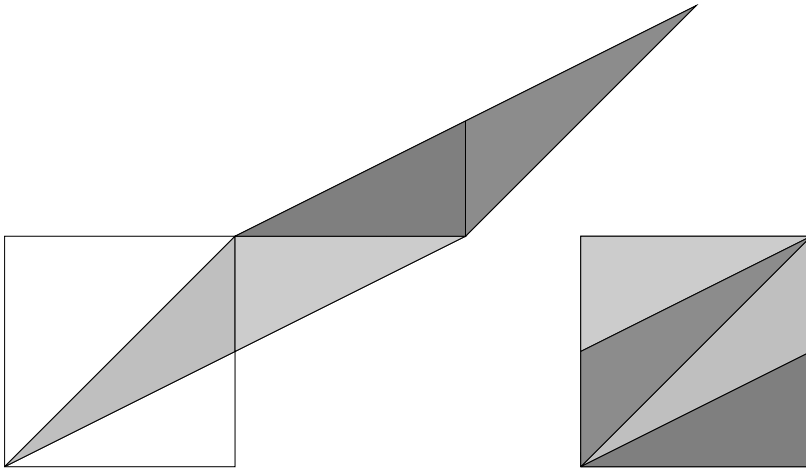


FIGURE 5.1. The map  $A : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  projects down to a map  $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$

**PROPOSITION 5.1.** *A hyperbolic linear toral automorphism  $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  is a homeomorphism*

**PROOF.** The map  $T$  is clearly continuous since if  $|x_1 - y_1|, |x_2 - y_2| < \epsilon$  then  $|(T(x_1, x_2))_1 - (T(y_1, y_2))_1| \leq (|a| + |b|)\epsilon$  and  $|(T(x_1, x_2))_2 - (T(y_1, y_2))_2| \leq (|c| + |d|)\epsilon$ .

To see that  $T$  is invertible we note that if we write the inverse matrix

$$A^{-1} = \begin{pmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix}$$

then since  $ad - bc = \pm 1$  we see that  $A^{-1} \in GL(2, \mathbb{Z})$ . The inverse to  $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  is then the linear toral automorphism associated to  $A^{-1}$ , i.e.

$$T^{-1}(x_1, x_2) = \left( \frac{d}{ad-bc}x_1 + \frac{-b}{ad-bc}x_2, \frac{-c}{ad-bc}x_1 + \frac{a}{ad-bc}x_2 \right) \pmod{1},$$

■

We give some examples.

EXAMPLES.

- (i) Let  $A_1 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ ; then  $\text{Det}(A_1) = 1$  and the associated map  $T_1 : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  takes the form  $T_1(x_1, x_2) = (2x_1 + x_2, x_1 + x_2) \pmod{1}$ .
- (ii) Let  $A_2 = \begin{pmatrix} 57 & 2 \\ 85 & 3 \end{pmatrix}$ ; then  $\text{Det}(A_2) = 1$  and the associated map  $T_2 : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  takes the form  $T_2(x_1, x_2) = (57x_1 + 2x_2, 85x_1 + 3x_2) \pmod{1}$ .
- (iii) Let  $A_3 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$  then  $\text{Det}(A_3) = 1$  and the associated map  $T_3 : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  takes the form  $T_3(x_1, x_2) = (x_1 + x_2, x_2) \pmod{1}$ .

DEFINITION. Let  $A \in GL(2, \mathbb{Z})$  have eigenvalues  $\lambda_1, \lambda_2$  then if  $|\lambda_1| > 1 > |\lambda_2| (= \frac{1}{|\lambda_1|})$  we call the matrix  $A$  *hyperbolic* and the associated linear toral transformation  $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  is called *hyperbolic*.

The matrix  $A_1 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$  has eigenvalues  $\frac{3 \pm \sqrt{5}}{2}$ , and so is hyperbolic. The matrix  $A_2 = \begin{pmatrix} 57 & 2 \\ 85 & 3 \end{pmatrix}$  has eigenvalues  $30 \pm \frac{1}{2}\sqrt{3596}$ , and so is hyperbolic. The matrix  $A_3 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$  has both eigenvalues 1, and so is *not* hyperbolic.

The following results shows that for these simple homeomorphisms much information on the periodic points can be easily obtained.

PROPOSITION 5.2. *Let  $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  be a hyperbolic toral automorphism associated to  $A \in GL(2, \mathbb{Z})$  (with eigenvalues  $\lambda_1, \lambda_2$ ).*

- (i) *The periodic points have rational coordinates in  $\mathbb{T}^2$  (i.e. they are of the form  $(\frac{p_1}{q}, \frac{p_2}{q})$  with natural numbers  $0 \leq p_1, p_2 < q$ ).*
- (ii) *We have  $\text{Card}\{x \in \mathbb{T}^2 : T^n x = x\} = |\lambda_1^n + \lambda_2^n - 2|$ .*

PROOF. (i) If  $(x_1, x_2) = (\frac{p_1}{q}, \frac{p_2}{q})$  is a point with rational co-ordinates then we can write  $T^n(x_1, x_2) + \mathbb{Z}^2 = \left( \frac{p_1^{(n)}}{q}, \frac{p_2^{(n)}}{q} \right) + \mathbb{Z}^2$ , where the integers

$0 \leq p_1^{(n)}, p_2^{(n)} \leq q - 1$  are given by  $\begin{pmatrix} p_1^{(n)} \\ p_2^{(n)} \end{pmatrix} = A^n \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$ . Since there are at most  $q^2$  distinct choices  $p_1^{(n)}, p_2^{(n)}$  we can choose  $q^2 + 1 \geq n_1 > n_0 > 0$  with  $p_1^{(n_0)} = p_1^{(n_1)}$  and  $p_2^{(n_0)} = p_2^{(n_1)}$ . In particular, this means that  $A^{n_1} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = A^{n_0} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$ . This corresponds to the identity  $T^{n_1}(x_1, x_2) = T^{n_0}(x_1, x_2)$  and so we conclude that  $T^{n_1 - n_0}(x_1, x_2) = (x_1, x_2)$ , i.e.  $(x_1, x_2)$  is periodic.

Conversely, assume that  $T^n(x_1, x_2) = (x_1, x_2)$  is a periodic point; then

$$A^n \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \end{pmatrix}, \tag{5.1}$$

for some  $n_1, n_2 \in \mathbb{Z}$ . Since  $A^n$  does not have 1 as an eigenvalue, the matrix  $A^n - I$  is invertible and solutions to (5.1) are of the form  $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (A^n - I)^{-1} \begin{pmatrix} n_1 \\ n_2 \end{pmatrix}$ . Since  $A^n - I$  has entries in  $\mathbb{Z}$  the matrix  $(A^n - I)^{-1}$  has entries in  $\mathbb{Q}$ , and we can conclude the same for the values  $x_1, x_2$ .

This completes the proof of part (i).

(ii) If we write  $A^n = \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix}$  then we can define a map  $S = (A^n - I) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  by  $S : (x_1, x_2) \rightarrow ((a_n - 1)x_1 + b_n x_2, c_n x_1 + (d_n - 1)x_2)$ . This maps the unit square  $[0, 1) \times [0, 1)$  onto the parallelogram  $\mathcal{R} = \{\alpha u + \beta v : 0 \leq \alpha, \beta < 1\}$  where  $u = S(0, 1)$  and  $v = S(1, 0)$  (see figure 5.2).

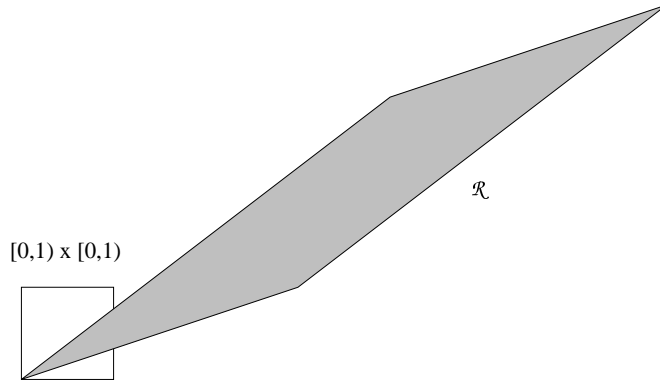


FIGURE 5.2. The number of points in  $\mathbb{Z}^2 \cap \mathcal{R}$  is  $\text{Card}(\text{Fix}(T^n))$

A fixed point  $T^n(x_1, x_2) = (x_1, x_2)$  corresponds to a solution to the identity (5.1) or, equivalently, to solutions  $S(x_1, x_2) \in \mathbb{Z}^2 \cap \mathcal{R}$  (where  $(x_1, x_2) \in [0, 1) \times [0, 1)$ ). Due to the shape of a parallelogram the number of such solutions is precisely the area of  $\mathcal{R}$ , which is  $|\text{Det}(A^n - I)| = |\lambda_1^n + \lambda_2^n - 2|$ . This completes the proof of part (ii). ■

### 5.2 Entropy for hyperbolic toral automorphisms

The main result on the entropy is the following.

**THEOREM 5.3.** *The entropy of the hyperbolic toral automorphism is given by  $h(T) = \log \lambda_1$ .*

**PROOF.** We fix  $\epsilon > 0$ . We can cover  $\mathbb{T}^2$  by  $\epsilon$ -balls centred at a finite set of points  $\{(x_1^1, x_2^1), \dots, (x_1^k, x_2^k)\}$ ,

$$B((x_1^i, x_2^i), \epsilon) = \{(z_1, z_2) \in \mathbb{T}^2 : |(z_1, z_2) - (x_1, x_2)| < \epsilon\}$$

(cf. Figure 5.3.)

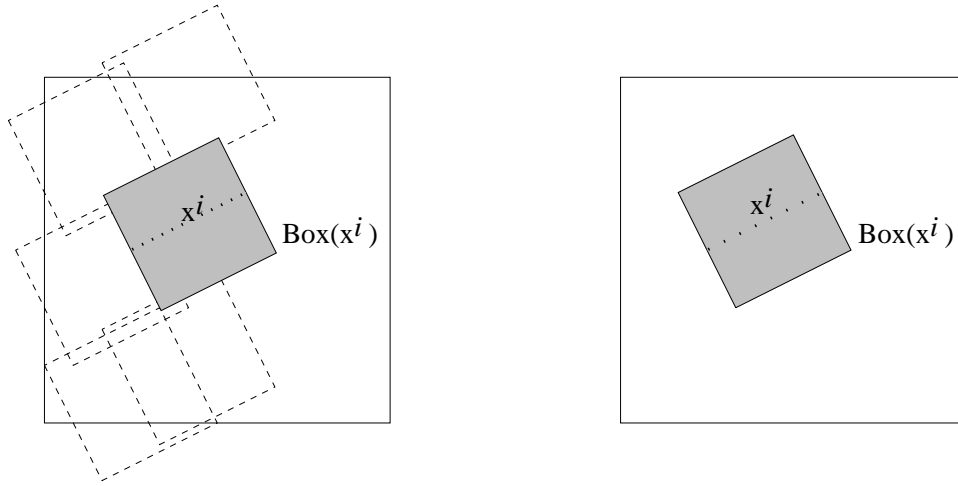
Around each point  $x^i = (x_1^i, x_2^i)$ ,  $i = 1, \dots, k$ , we can describe a “box”

$$\text{Box}(x_1^i, x_2^i) = \{(x_1^i, x_2^i) + \alpha v_1 + \beta v_2 : -\epsilon \leq \alpha, \beta \leq \epsilon\}$$

where  $v_1$  and  $v_2$  are the eigenvectors for  $A$  (with  $|v_1| = |v_2| = 1$ ) corresponding to  $|\lambda_1| > 1$  and  $|\lambda_2| < 1$ , respectively. We can choose boxes such that  $\text{Box}(x_1^i, x_2^i)$ ,  $i = 1, \dots, n$  cover the torus. For each  $n \geq 1$  and  $1 \leq i \leq k$  we can consider the finite subset of  $\text{Box}(x^i)$  consisting of the points

$$R(x_1^i, x_2^i) = \{(x_1^i, x_2^i) + \frac{j\epsilon}{|\lambda_1|^n} v_1 : j = -\lfloor |\lambda_1|^n \rfloor, \dots, \lfloor |\lambda_1|^n \rfloor\}.$$

This set has cardinality  $2\lfloor |\lambda_1|^n \rfloor + 1$ .



**FIGURE 5.3.** Construction an  $(n, 2\epsilon)$ -spanning set and an  $(n, \epsilon)$ -separating set



**SUBLEMMA 5.3.1.**  $R = \cup_{i=1}^k R(x^i)$  is an  $(n, 2\epsilon)$ -spanning set (see figure 5.3).

**PROOF.** For any point  $(z_1, z_2) \in \mathbb{R}^2/\mathbb{Z}^2$  we can choose some  $i = 1, \dots, k$  such that  $(z_1, z_2) \in \text{Box}(x_1^i, x_2^i)$  and so we can write  $(z_1, z_2) = (x_1^i, x_2^i) + \alpha v_1 + \beta v_2$  for some  $-\epsilon \leq \alpha, \beta \leq \epsilon$ .

If we choose  $-[\lambda_1|^n] \leq j \leq [\lambda_1|^n]$  with  $|\alpha - \frac{j\epsilon}{|\lambda_1|^n}| \leq \frac{\epsilon}{2|\lambda_1|^n}$  then we can select  $(w_1, w_2) = (x_1^i, x_2^i) + \frac{j\epsilon}{|\lambda_1|^n} v_1 \in R(x_1^i, x_2^i)$ .

For any  $0 \leq r \leq n$  we can write

$$\begin{aligned} T^r(z_1, z_2) &= T^r(w_1, w_2) + \left( \alpha - \frac{j\epsilon}{|\lambda_1|^n} \right) A^r v_1 + \beta A^r v_2 \\ &= T^r(w_1, w_2) + \left( \alpha - \frac{j\epsilon}{|\lambda_1|^n} \right) \lambda_1^r v_1 + \beta \lambda_2^r v_2 \end{aligned}$$

and, in particular,

$$\begin{aligned} |T^r(z_1, z_2) - T^r(w_1, w_2)| &\leq \left| \left( \alpha - \frac{j\epsilon}{|\lambda_1|^n} \right) \right| \cdot |\lambda_1|^r \cdot |v_1| + \beta |\lambda_2|^r |v_2| \\ &\leq \frac{\epsilon |\lambda_1|^r}{2|\lambda_1|^n} + \epsilon \leq \frac{\epsilon}{2} + \epsilon = 2\epsilon. \end{aligned}$$

This completes the proof of Sublemma 5.3.1. ■

The cardinality of the  $(n, 2\epsilon)$ -spanning set  $R$  is at most  $k(2[\lambda_1|^n] + 1)$  we see that this is an upper bound on the least cardinality  $r(n, \epsilon)$  of  $(n, 2\epsilon)$ -spanning sets. By Proposition 3.11 we can write

$$\begin{aligned} h(T) &= \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow +\infty} \frac{1}{n} \log r(n, \epsilon) \\ &\leq \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow +\infty} \frac{1}{n} \log (k(2[\lambda_1|^n] + 1)) \\ &= \lim_{\epsilon \rightarrow 0} \log |\lambda_1| = \log |\lambda_1|. \end{aligned} \tag{5.2}$$

To get the reverse inequality, we fix a point  $(x_1, x_2) \in \mathbb{T}^2$ . For  $\epsilon > 0$  and  $n \geq 1$  we can consider the subset

$$S = \left\{ (x_1, x_2) + \frac{j2\epsilon}{|\lambda_1|^n} v_1 : j = -[\lambda_1|^n], \dots, [\lambda_1|^n] \right\}.$$

**SUBLEMMA 5.3.2.**  $S$  is an  $(n, 2\epsilon/|\lambda_1|)$ -separated set (see Figure 5.3).

**PROOF.** Any two distinct points in  $S$  will be of the form  $(u_1, u_2) = (x_1, x_2) + \frac{i2\epsilon}{|\lambda_1|^n} v_1$ ,  $(w_1, w_2) = (x_1, x_2) + \frac{j2\epsilon}{|\lambda_1|^n} v_1$ , with  $-[\lambda_1|^n] \leq i, j \leq [\lambda_1|^n]$ .

For  $0 \leq r \leq n - 1$  we have

$$\begin{aligned} |T^r(u_1, u_2) - T^r(w_1, w_2)| &= \left| \frac{(j-i)2\epsilon}{|\lambda_1|^n} T^r(v_1) \right| \\ &= 2|j-i| \frac{\epsilon}{|\lambda_1|^{n-r}}. \end{aligned}$$

In particular, for some  $0 \leq r \leq n-1$  we have that  $|T^n(u_1, u_2) - T^n(w_1, w_2)| > \epsilon$ . This completes the proof of Sublemma 5.3.2. ■

The cardinality of the  $(n, \epsilon)$ -separated set  $S$  is  $2\lceil |\lambda_1|^n \rceil + 1$  and so this gives a lower bound on the maximal cardinality  $s(n, \epsilon)$  of  $(n, \epsilon)$ -separated sets. By Proposition 3.11 we have that

$$\begin{aligned} h(T) &= \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow +\infty} \frac{1}{n} \log s(n, \epsilon) \\ &\geq \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow +\infty} \frac{1}{n} \log (2\lceil |\lambda_1|^n \rceil + 1) \\ &= \lim_{\epsilon \rightarrow 0} (\log |\lambda_1|) = \log |\lambda_1|. \end{aligned} \tag{5.3}$$

The two complementary inequalities (5.2) and (5.3) combine to show that  $h(T) = \log |\lambda_1|$ . This completes the proof of Theorem 5.3. ■

COROLLARY 5.3.1.

$$\lim_{n \rightarrow +\infty} \frac{1}{e^{nh(T)}} \text{Card}(\text{Fix}(T^n)) = 1.$$

PROOF. This is immediate from Proposition 5.2 (ii) and Theorem 5.3 ■

### 5.3 Shadowing and semi-conjugacy

We begin with a very useful property for hyperbolic toral automorphisms.

DEFINITION. Let  $\delta > 0$ . We call a sequence of points  $(x_n)_{n \in \mathbb{Z}} \in \mathbb{T}^2$  a  $\delta$ -pseudo-orbit if  $|T(x_n) - x_{n+1}| < \delta$ , for all  $n \in \mathbb{Z}$ .

THEOREM 5.4 (SHADOWING PROPERTY). *For all  $\epsilon > 0$  there exists  $\delta > 0$  such that any  $\delta$ -pseudo-orbit  $(x_n)_{n \in \mathbb{Z}}$  is  $\epsilon$ -close to a true orbit  $(T^n(x))_{n=-\infty}^{\infty}$  (in the sense that  $\forall n \in \mathbb{Z}$ ,  $|T^n(x) - x_n| < \epsilon$ ).*

*Furthermore, if  $\delta > 0$  is sufficiently small then there is a unique point  $x$  with this property.*

PROOF. Given  $\epsilon > 0$  we choose  $\delta > 0$  sufficiently small that any vector  $v \in \mathbb{R}^2$  with  $|v| \leq \delta$  can be written in the form  $v = \alpha v_1 + \beta v_2$  with  $|\alpha|, |\beta| < \frac{\epsilon}{3} (1 - \frac{1}{|\lambda_1|})$ .

Let  $(x_n)_{n \in \mathbb{Z}}$  be any  $\delta$ -pseudo-orbit. By our choice of  $\delta$ , for each  $n \in \mathbb{Z}$  we can write  $T(x_{n-1}) = x_n - \alpha_n v_1 - \beta_n v_2$ , where  $|\alpha_n|, |\beta_n| \leq \frac{\epsilon}{3}(1 - \frac{1}{|\lambda_1|})$ . Iterating this identity gives that  $T^2(x_{n-2}) = x_n - (\alpha_n + \alpha_{n-1}\lambda_1)v_1 - (\beta_n + \beta_{n-1}\lambda_2)v_2$  and proceeding inductively we get that

$$x_n = \begin{cases} T^n(x_0) + \left(\sum_{i=1}^n \alpha_i \lambda_1^{n-i}\right) v_1 + \left(\sum_{i=1}^n \beta_i \lambda_2^{n-i}\right) v_2 & \text{for } n \geq 1, \\ T^n(x_0) + \left(\sum_{i=-n}^{-1} \alpha_i \lambda_1^{n-i}\right) v_1 + \left(\sum_{i=-n}^{-1} \beta_i \lambda_2^{n-i}\right) v_2 & \text{for } n \leq -1. \end{cases}$$

We can now define  $x = x_0 + \left(\sum_{i=1}^{\infty} \alpha_i \lambda_1^{-i}\right) v_1 + \left(\sum_{i=-\infty}^{-1} \beta_i \lambda_2^{-i}\right) v_2$ . It is easy to see that  $|x - x_0| \leq (\sup_i |\alpha_i| + \sup_i |\beta_i|) / (1 - \frac{1}{|\lambda_1|}) \leq \frac{2}{3}\epsilon$ . We observe that  $T^n(x) = T^n(x_0) + \lambda_1^n \left(\sum_{i=1}^{\infty} \lambda_1^{-i} \alpha_i\right) v_1 + \lambda_2^n \left(\sum_{i=-\infty}^{-1} \lambda_2^{-i} \beta_i\right) v_2$ , for any  $n \in \mathbb{Z}$ .

For  $n \geq 1$  we have that

$$\begin{aligned} T^n(x) - x_n &= \left( \lambda_1^n \sum_{i=1}^{\infty} \alpha_i \lambda_1^{-i} - \sum_{i=1}^n \alpha_i \lambda_1^{n-i} \right) v_1 \\ &\quad + \left( \lambda_2^n \sum_{i=-\infty}^{-1} \beta_i \lambda_2^{-i} - \sum_{i=1}^n \beta_i \lambda_2^{n-i} \right) v_2. \end{aligned} \tag{5.4}$$

The coefficient of  $v_1$  in (5.4) is  $\sum_{i=n+1}^{\infty} \alpha_i \lambda_1^{n-i}$ , which is bounded by  $(\sup_i |\alpha_i|) / (1 - \frac{1}{|\lambda_1|}) \leq \epsilon/3$ . The coefficient of  $v_2$  in (5.4) can be crudely bounded by  $2(\sup_i |\beta_i|) / (1 - \frac{1}{|\lambda_1|}) \leq 2\epsilon/3$ . Combining these estimates shows that  $|T^n(x) - x_n| \leq \frac{2\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$ , for all  $n \geq 1$ .

For  $n \leq -1$  we have that

$$\begin{aligned} T^n(x) - x_n &= \left( \lambda_1^n \sum_{i=1}^{\infty} \alpha_i \lambda_1^{-i} - \sum_{i=n}^{-1} \alpha_i \lambda_1^{n-i} \right) v_1 \\ &\quad + \left( \lambda_2^n \sum_{i=-\infty}^{-1} \beta_i \lambda_2^{-i} - \sum_{i=n}^{-1} \beta_i \lambda_2^{n-i} \right) v_2. \end{aligned} \tag{5.5}$$

The coefficient of  $v_2$  is  $\sum_{i=-\infty}^{n+1} \beta_i \lambda_2^{n-i}$ , which is bounded by  $(\sup_i |\beta_i|) \frac{1}{1 - \frac{1}{|\lambda_1|}} \leq \frac{\epsilon}{3}$ , and the coefficient of  $v_1$  can be crudely dominated by  $(\sup_i |\alpha_i|) \frac{2}{1 - \frac{1}{|\lambda_1|}} \leq 2\epsilon/3$ . Thus  $|T^n(x) - x_n| \leq \frac{\epsilon}{3} + \frac{2\epsilon}{3} = \epsilon$ , for all  $n \leq -1$ .

This shows that the pseudo-orbit  $(x_n)_{n \in \mathbb{Z}}$  is  $\epsilon$ -close to the orbit  $(T^n x)_{n \in \mathbb{Z}}$ .

Consider two points  $x, y \in \mathbb{T}^2$  which are  $\epsilon$ -close to the same pseudo-orbit  $(x_n)_{n \in \mathbb{Z}}$ . To complete the proof of the theorem we need to show that providing  $\delta$  is sufficiently small this property implies  $x = y$ .

By the triangle inequality we have that  $|T^n(y) - T^n(x)| \leq |T^n(y) - x_n| + |x_n - T^n(x)| \leq 2\epsilon$ .

Provided the separation  $|x - y| < 2\epsilon$  is small we can write  $y = x + \alpha v_1 + \beta v_2$  and then  $T^n(y) = T^n(x) + \alpha \lambda_1^n v_1 + \beta \lambda_2^n v_2$ , for any  $n \in \mathbb{Z}$ . If  $\alpha \neq 0$  then  $|T^n(y) - T^n(x)| \geq |\alpha| |\lambda_1|^n - |\beta| |\lambda_2|^n > 2\epsilon$ , for some  $n \geq 1$ , giving a contradiction. If  $\beta \neq 0$  then  $|T^{-n}(y) - T^{-n}(x)| \geq |\beta| |\lambda_1|^n - |\alpha| |\lambda_2|^n > 2\epsilon$ , for some  $n \geq 1$ , giving a contradiction.

This completes the proof. ■

Let  $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  be a hyperbolic toral automorphism and let  $S : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  be a second homeomorphism. We can write these maps in terms of their co-ordinates:

$$\begin{cases} T(x_1, x_2) = (T_1(x_1, x_2), T_2(x_1, x_2)) \text{ and} \\ S(x_1, x_2) = (S_1(x_1, x_2), S_2(x_1, x_2)). \end{cases}$$

DEFINITION. Given  $\delta > 0$  we say that  $S$  and  $T$  are *uniformly*  $\delta$ -close if

$$\sup_{(x_1, x_2) \in \mathbb{T}^2} \{|T_1(x_1, x_2) - S_1(x_1, x_2)|, |T_2(x_1, x_2) - S_2(x_1, x_2)|\} < \delta.$$

The next result tells that any homeomorphism close to a hyperbolic toral automorphism  $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  (in the above sense) must be semi-conjugate to  $T$ .

THEOREM 5.5. *Let  $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  be a hyperbolic toral automorphism then there exists  $\delta > 0$  such that any uniformly  $\delta$ -close  $S : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  is semi-conjugate to  $T$ .*

PROOF. We use Theorem 5.4 to choose  $\epsilon > 0$  sufficiently small that there exists  $\delta > 0$  with every  $\delta$ -pseudo-orbit being  $\epsilon$ -close to the orbit of *exactly* one point in  $\mathbb{T}^2$ .

For any  $x \in \mathbb{T}^2$  it follows from the definition of  $S$  being uniformly  $\delta$ -close to  $T$  that the orbit  $(x_n)_{n \in \mathbb{Z}} := (S^n x)_{n \in \mathbb{Z}}$  under  $S$  gives a  $\delta$ -pseudo-orbit for  $T$ . By Theorem 5.4 there is a *unique* point  $z \in \mathbb{T}^2$  whose  $T$ -orbit  $(T^n(z))_{n \in \mathbb{Z}}$  is  $\epsilon$ -close to  $(x_n)_{n \in \mathbb{Z}}$  (i.e.  $\sup_{n \in \mathbb{Z}} |T^n(z) - S^n(x)| \leq \epsilon$ ). We define  $\pi : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  by  $\pi(x) := z$ .

It is immediate that this is a semi-conjugacy, since from the definitions the orbit  $(S^n(Sx))_{n \in \mathbb{Z}}$  of  $Sx$  (under  $S$ ) is  $\epsilon$ -close to the orbit of  $Tz$  (under  $T$ ). By uniqueness, we see that  $\pi(Sx) = Tz = T(\pi x)$ .

We next show that  $\pi : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  is continuous. Let  $\eta > 0$  be fixed and choose  $N \geq 0$  such that  $2(3\delta + 1)|\lambda_2|^N < \eta$ . The continuity of  $S : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  shows that provided  $x, y \in \mathbb{T}^2$  are sufficiently close then we have  $|x_n - y_n| = |S^n(x) - S^n(y)| \leq \delta$  for  $-N \leq n \leq N$  (where  $x_n = S^n x$  and  $y_n = S^n y$ ). In

particular,  $|T^n(\pi x) - T^n(\pi y)| \leq |T^n(\pi x) - x_n| + |x_n - y_n| + |y_n - T^n(\pi y)| \leq 3\delta$ . However, if we write  $\pi(x) = \pi(y) + \alpha v_1 + \beta v_2$  then  $3\delta \geq |T^n(\pi x) - T^n(\pi y)| = |\alpha \lambda_1^n v_1 + \lambda_2^n \beta v_2|$ . For  $N \geq n \geq 1$  we see that  $3\delta \geq |\lambda_1|^n \alpha - |\lambda_2|^n \beta$  and for  $-N \leq n \leq -1$  we see that  $3\delta \geq |\lambda_2|^n \beta - |\lambda_1|^n \alpha$ . This allows us to see that  $|\alpha|, |\beta| \leq (3\delta + 1)|\lambda_2|^N$  and so  $|\pi(x) - \pi(y)| \leq |\alpha| + |\beta| \leq 2(3\delta + 1)|\lambda_2|^N < \eta$ .

Finally, to see that  $\pi : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  is surjective we first observe from the proof of Theorem 5.4 that  $S \rightarrow \pi_S := \pi$  is continuous and  $\pi_T = I$  (where  $I : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  is the (surjective) identity map). Thus for  $S$  sufficiently close to  $T$  we have that  $\pi$  is surjective. ■

**COROLLARY 5.5.1.** *Let  $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  be a hyperbolic toral automorphism; then there exists  $\epsilon > 0$  such that for any uniformly  $\epsilon$ -close homeomorphism  $S : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  we have  $h(T) \leq h(S)$ .*

**PROOF.** This follows by Theorem 5.4 and Proposition 3.5.

**REMARK.** If we make the strong assumption that  $S : \mathbb{T}^2 \rightarrow \mathbb{T}^2$  is  $C^1$  and  $S$  and  $T$  are uniformly close *and* their derivatives are close then  $T$  and  $S$  are conjugate. This property is called *structural stability* of  $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$ . ■

## 5.4 Comments and references

We have chosen to consider only hyperbolic toral automorphisms in two dimensions, rather than in arbitrary dimensions, to make the proofs more graphic. However, the results and proofs would be the same.

Good accounts of the general hyperbolic theory are in [3], [6], [5], etc. Other treatments of toral automorphisms are in [2] and [7]. For other important related topics see [2], [6], [7] (for structural stability and stable manifolds), [4], [1], [6, chapter 10], [3] (for Markov partitions)

### References

1. R. Adler and B. Weiss, *Similarity of automorphisms of the torus*, *Memoirs Amer. Math. Soc.* **98** (1970).
2. V. Arnol'd, *Ordinary Differential Equations*, M.I.T. Press, Cambridge Mass., 1973.
3. R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, *Lecture Notes in Mathematics*, 470, Springer, Berlin, 1975.
4. R. Bowen, *On Axiom A diffeomorphisms*, *CMBS Reg. Conference series*, 35, Amer. Math. Soc., Providence R.I., 1978.
5. R. Mañé, *Ergodic Theory and Differentiable Dynamics*, Springer, Berlin, 1987.
6. M. Shub, *Global Stability of Dynamical Systems*, Springer, Berlin, 1986.
7. W. Szlenk, *Introduction to the Theory of Smooth Dynamical Systems*, Wiley, New York, 1984.

ROTATION NUMBERS

In this chapter we shall define the useful concept of the rotation number for orientation preserving homeomorphisms of the circle.

6.1 Homeomorphisms of the circle and rotation numbers

Let  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  be an orientation preserving homeomorphism of the circle to itself. There is a canonical projection  $\pi : \mathbb{R} \rightarrow \mathbb{R}/\mathbb{Z}$  given by  $\pi(x) = x \pmod{1}$ . We call a monotone map  $\hat{T} : \mathbb{R} \rightarrow \mathbb{R}$  a *lift* of  $T$  if the canonical projection  $\pi : \mathbb{R} \rightarrow \mathbb{R}/\mathbb{Z}$  is a semi-conjugacy (i.e.  $\pi \circ \hat{T} = T \circ \pi$ ).

For a given map  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  a lift  $\hat{T} : \mathbb{R} \rightarrow \mathbb{R}$  will not be unique.

EXAMPLE. If  $T(x) = (x + \alpha) \pmod{1}$  then for any  $k \in \mathbb{Z}$  the map  $\hat{T} : \mathbb{R} \rightarrow \mathbb{R}$  defined by  $\hat{T}(x) = x + \alpha + k$  is a lift. To see this observe that  $\pi(\hat{T}(x)) = \pi(x + \alpha + k) = x + \alpha \pmod{1}$  and  $T(\pi(x)) = \pi(x) + \alpha \pmod{1} = x + \alpha \pmod{1}$ .

The following lemma summarizes some simple properties of lifts.

LEMMA 6.1.

- (i) *Let  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  be a homeomorphism of the circle; then if  $\hat{T} : \mathbb{R} \rightarrow \mathbb{R}$  is a lift, then any other lift  $\hat{T}' : \mathbb{R} \rightarrow \mathbb{R}$  must be of the form  $\hat{T}'(x) = \hat{T}(x) + k$ , for some  $k \in \mathbb{Z}$ .*
- (ii) *For any  $x, y \in \mathbb{R}$  with  $|x - y| \leq k$  ( $k \in \mathbb{Z}^+$ ) we have  $|\hat{T}(x) - \hat{T}(y)| \leq k$ . Iterating this gives that*

$$|\hat{T}^n(x) - \hat{T}^n(y)| \leq k, \quad \forall n \geq 0.$$

PROOF. These are easily seen from the continuity and the monotonicity of  $\hat{T}$ . ■

DEFINITION. We define the *rotation number*  $\rho(T)$  of the homeomorphism by

$$\rho(T) = \limsup_{n \rightarrow +\infty} \frac{\hat{T}^n(x)}{n} \pmod{1}.$$

(The limsup is independent of the choice  $x \in \mathbb{R}$  by Lemma 6.1. The choice of lift  $\hat{T}$  can only alter the limsup by an integer, which has no bearing since we define  $\rho(T)$  modulo one.)

EXAMPLE. Consider the standard rotation  $R_\rho : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  defined by  $R_\rho(x) = x + \rho \pmod{1}$ , where  $\rho \in [0, 1)$ , say. Any lift  $\hat{R}_\rho : \mathbb{R} \rightarrow \mathbb{R}$  will be of the form  $\hat{R}_\rho(x) = x + \rho + k$ , for some  $k \in \mathbb{Z}$  i.e. translation on the real line by  $\rho + k$ . It is now immediate from the definition that the rotation number for  $R_\rho$  is merely  $\rho \pmod{1}$ .

We can now show some interesting properties of  $\rho(T)$ .

PROPOSITION 6.2.

- (i) For  $n \geq 1$  we have that  $\rho(T^n) = n\rho(T) \pmod{1}$ .
- (ii) If  $T$  has a periodic point (i.e.  $\exists n \geq 1, \exists x \in \mathbb{R}/\mathbb{Z}$  such that  $T^n x = x$ ) then  $\rho(T)$  is rational.
- (iii) If  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  has no periodic points then  $\rho(T)$  is irrational.
- (iv) The limit actually exists and we can write

$$\rho(T) = \lim_{n \rightarrow +\infty} \frac{\hat{T}^n(x)}{n} \pmod{1}.$$

PROOF. (i) Since  $\hat{T}^n$  (the  $n$ th iterate of the lift  $\hat{T}$  for  $T$ ) is itself a lift for  $T^n : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  this is immediate from the definitions.

(ii) Since  $T^n(x + \mathbb{Z}) = x + \mathbb{Z}$  we have that  $\hat{T}^n(x)$  and  $x$  differ by an integer (i.e.  $\hat{T}^n(x) - x = k \in \mathbb{Z}$ ). Then for  $pn + r$  with  $0 \leq r \leq n - 1$  and  $p \geq 0$  we have that  $\hat{T}^{pn+r}(x) = \hat{T}^r(\hat{T}^{pn}x) = \hat{T}^r(x) + pk$  because of Lemma 6.1. Thus  $\rho(T) = \limsup_{p \rightarrow +\infty} \frac{\hat{T}^{pn+r}(x)}{pn+r} = \frac{k}{n} \pmod{1}$ .

(iii) Assume for contradiction that  $\rho(T) = \frac{p}{q}$  is rational. By part (i) we see that for  $S := T^q$  we have  $\rho(S) = 0$  and since  $T$  has no periodic points we conclude that  $S : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  has no fixed points.

If  $\hat{S} : \mathbb{R} \rightarrow \mathbb{R}$  is a lift for  $S$  then the absence of fixed points for  $S$  implies either  $\hat{S}(x) > x, \forall x \in \mathbb{R}$ , or  $\hat{S}(x) < x, \forall x \in \mathbb{R}$ . Assume  $\hat{S}(x) > x, \forall x \in \mathbb{R}$  (the other case being similar), i.e.  $\hat{S}$  is strictly increasing. If  $\exists k > 0$  with  $\hat{S}^k(0) > 1$  then we see that  $\hat{S}^{mk}(0) > m$  and so

$$\rho(S) = \limsup_{n \rightarrow +\infty} \frac{\hat{S}^n(x)}{n} > \frac{1}{k},$$

contradicting that  $\rho(S) = 0$ . This leaves the possibility that  $\hat{S}^k(0) < 1$  for all  $k \geq 1$ . Since  $\hat{S}$  is strictly increasing the sequence  $(\hat{S}^k(0))_{k=1}^\infty$  is monotone increasing and the supremum  $z \in \mathbb{R}$  satisfies  $\hat{S}(z) = z$ . Thus  $S$  has a fixed point  $S(z + \mathbb{Z}) = z + \mathbb{Z}$ , giving a contradiction.

(iv) If  $T$  has a periodic point  $T^n x = x$  then the argument in part (i) actually shows that  $\rho(T) = \lim_{N \rightarrow +\infty} \frac{T^N(x)}{N} = \frac{k}{n} \pmod{1}$ , in particular, showing that the limit exists.

Assume that  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  has no periodic points. Thus for all  $n \geq 1$  there exists  $k_n \in \mathbb{Z}$  such that  $\hat{T}^n(x) - x \in [k_n, k_n + 1]$ ,  $\forall x \in \mathbb{R}$ , and, in particular, observe that  $|\frac{\hat{T}^n(0)}{n} - \frac{k_n}{n}| < \frac{1}{n}$ . Then, for any  $m \geq 1$  we have that

$$\begin{aligned} \hat{T}^{nm}(0) &= \hat{T}^n \left( \hat{T}^{n(m-1)}(0) \right) - \left( \hat{T}^{n(m-1)}(0) \right) \\ &\quad + \hat{T}^n \left( \hat{T}^{n(m-2)}(0) \right) - \left( \hat{T}^{n(m-2)}(0) \right) + \dots \\ &\quad \dots + \hat{T}^n \left( \hat{T}^n(0) \right) - \left( \hat{T}^n(0) \right) + \hat{T}^n(0) \in [mk_n, m(k_n + 1)]. \end{aligned} \quad (6.1)$$

In particular, we see from (6.1) that  $|\frac{\hat{T}^{nm}(0)}{nm} - \frac{k_n}{n}| < \frac{1}{n}$ . The triangle inequality gives that

$$\begin{aligned} \left| \frac{\hat{T}^m(0)}{m} - \frac{\hat{T}^n(0)}{n} \right| &\leq \left| \frac{\hat{T}^m(0)}{m} - \frac{k_m}{m} \right| + \left| \frac{k_m}{m} - \frac{\hat{T}^{mn}(0)}{mn} \right| \\ &\quad + \left| \frac{\hat{T}^{mn}(0)}{mn} - \frac{k_n}{n} \right| + \left| \frac{k_n}{n} - \frac{\hat{T}^n(0)}{n} \right| \\ &\leq \frac{2}{m} + \frac{2}{n} \end{aligned}$$

which shows that the sequence  $\left( \frac{\hat{T}^n(0)}{n} \right)_{n=0}^{\infty}$  is Cauchy, and in particular the limit exists. ■

The next lemma and its corollary will be useful later.

LEMMA 6.3. *Assume that  $\rho$  is irrational.*

- (i) *Let  $n_1, n_2, m_1, m_2 \in \mathbb{Z}$  and  $x, y \in \mathbb{R}$ . If  $\hat{T}^{n_1}(x) + m_1 < \hat{T}^{n_2}(x) + m_2$  then  $\hat{T}^{n_1}(y) + m_1 < \hat{T}^{n_2}(y) + m_2$ ;*
- (ii) *The bijection  $n\rho(T) + m \rightarrow \hat{T}^n(0) + m$  between the sets*

$$\Omega = \{n\rho(T) + m : n, m \in \mathbb{Z}\} \text{ and } \Lambda = \{\hat{T}^n(0) + m : n, m \in \mathbb{Z}\}$$

*preserves the natural ordering on  $\mathbb{R}$ .*

PROOF. (i) If  $\exists x, y \in \mathbb{R}$  for which the ordering is reversed, then by continuity (and the intermediate value theorem) there exists  $z \in \mathbb{R}$  with  $\hat{T}^{n_1}(z) + m_1 = \hat{T}^{n_2}(z) + m_2$ , i.e.  $\hat{T}^{n_1}(z) - \hat{T}^{n_2}(z) \in \mathbb{Z}$ . But then  $T^{n_1-n_2}(z + \mathbb{Z}) = z + \mathbb{Z}$  is a periodic point. This contradicts the assumption that  $T$  has irrational rotation number (and so no periodic points by Proposition 6.2 (ii)).



(ii) Assume that  $\hat{T}^{n_1}(0) + m_1 < \hat{T}^{n_2}(0) + m_2$  with  $n_1 > n_2$ ; then we wish to show that  $n_1\rho + m_1 < n_2\rho + m_2$ . We can rewrite the first inequality as  $\hat{T}^{n_1-n_2}(\hat{T}^{n_2}0) - \hat{T}^{n_2}0 < (m_2 - m_1)$  and we can apply part (i) with  $x = \hat{T}^{n_2}0$  and  $y = 0$  to deduce that

$$\hat{T}^{n_1-n_2}(0) < (m_2 - m_1). \quad (6.2)$$

Next, we can apply part (i) to (6.2) with the choices  $x = \hat{T}^{n_1-n_2}(0)$  and  $y = 0$  to deduce

$$\begin{aligned} \hat{T}^{2(n_1-n_2)}(0) - \hat{T}^{n_1-n_2}(0) &= \hat{T}^{n_1-n_2}(\hat{T}^{n_1-n_2}0) - \hat{T}^{n_1-n_2}(0) \\ &< (m_2 - m_1). \end{aligned} \quad (6.3)$$

Comparing (6.2) and (6.3) gives that  $\hat{T}^{2(n_1-n_2)}(0) < 2(m_2 - m_1)$ . Proceeding inductively shows that for any  $N \geq 1$  we have  $\hat{T}^{N(n_1-n_2)}(0) < N(m_2 - m_1)$ . Finally, we see that

$$\rho(T) = \lim_{n \rightarrow +\infty} \frac{\hat{T}^n(0)}{n} = \lim_{N \rightarrow +\infty} \frac{\hat{T}^{N(n_1-n_2)}(0)}{N(n_1-n_2)} \leq \frac{(m_2 - m_1)}{(n_1 - n_2)}$$

and in fact  $\rho < \frac{(m_2 - m_1)}{(n_1 - n_2)}$  since  $\rho$  is assumed irrational. This is the required inequality, and so this completes the proof. ■

**COROLLARY 6.3.1.** *Let  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  have irrational rotation number  $\rho$ . For any  $x \in \mathbb{R}/\mathbb{Z}$  the orbits of  $x$  under  $T$  and the rotation  $R_\rho : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  have the same ordering.*

**PROOF.** This follows immediately from part (ii) of Lemma 6.3, since a difference in the ordering of  $T^n(x)$  and  $R_\rho^n(x)$  would contradict the conclusion of the lemma. ■

## 6.2 Denjoy's theorem

The following result gives a sufficient condition for a homeomorphism to be conjugate to a rotation.

**PROPOSITION 6.4.** *If  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  is a minimal orientation preserving homeomorphism with irrational rotation number  $\rho$  then  $T$  is topologically conjugate to the standard rotation  $R_\rho : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ .*

**PROOF.** Let  $\hat{T} : \mathbb{R} \rightarrow \mathbb{R}$  be a lift of  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ . Observe that since  $\rho$  is irrational we have that  $\Omega \subset \mathbb{R}$  (as defined in Lemma 6.3 (ii)) is dense. Moreover, since  $T$  is minimal we know that  $\{T^n 0\}$  is dense in  $\mathbb{R}/\mathbb{Z}$ , and so we also have that  $\Lambda \subset \mathbb{R}$  is dense.

The map  $\phi : \Lambda \rightarrow \Omega$  given by  $\phi(\hat{T}^n(0) + m) = n\rho + m$  is order preserving by Lemma 6.3. Thus it extends to a homeomorphism  $\phi : \mathbb{R} \rightarrow \mathbb{R}$ .

Observe that

$$\phi\left(\hat{T}(\hat{T}^n(0) + m)\right) = \phi\left(\hat{T}^{n+1}(0) + m\right) = (n+1)\rho + m$$

and

$$\hat{R}_\rho\phi(\hat{T}^n(0) + m) = \hat{R}_\rho(n\rho + m) = (n+1)\rho + m$$

where  $\hat{R}_\rho : \mathbb{R} \rightarrow \mathbb{R}$  is a lift for  $R_\rho$ . Thus  $\phi \circ \hat{T} = \hat{R}_\rho \circ \phi$ .

Finally, we observe that by construction  $\phi(x+1) = \phi(x) + 1$ . Thus the homeomorphism  $\phi : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  defined by  $\phi(x + \mathbb{Z}) = \phi(x) + \mathbb{Z}$  is well-defined. Moreover, the identity  $\phi \circ \hat{T} = \hat{R}_\rho \circ \phi$  implies the conjugacy relation  $\phi \circ T = R_\rho \circ \phi$ . This completes the proof of the proposition. ■

Given a  $C^1$  map  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  we consider its derivative  $T' : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}$ .

DEFINITION. We define the *variation* of  $\log |T'| : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}$  by

$$\begin{aligned} & \text{Var}(\log |T'|) \\ &= \sup \left\{ \sum_{i=0}^{n-1} |\log |T'|(x_{i+1}) - \log |T'|(x_i)| : 0 = x_0 < x_1 < \dots < x_n = 1 \right\}. \end{aligned}$$

We say that the logarithm of  $|T'|$  has *bounded variation* if this value  $\text{Var}(\log |T'|)$  is finite.

It is easy to see that if  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  is  $C^2$  with  $|T''|/|T'|$  is bounded then  $\text{Var}(\log |T'|)$  is finite.

We now come to the main result of this section which gives sufficient conditions for a homeomorphism to be conjugate to a rotation (and are more readily checked than those of Proposition 6.4).

**THEOREM 6.5 (DENJOY'S THEOREM).** *If  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  is a  $C^1$  orientation preserving homeomorphism of the circle with derivative of bounded variation and irrational rotation number  $\rho = \rho(T)$  then  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  is topologically conjugate to the standard rotation  $R_\rho : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ .*

PROOF. It suffices to show that  $T$  is minimal, then the result follows by applying Proposition 6.4. The proof of minimality will come via two sublemmas.

**SUBLEMMA 6.5.1.** *If  $T$  has irrational rotation number and there are a constant  $C > 0$  and a sequence of integers  $q_n \rightarrow +\infty$  such that the maps  $T^{q_n} : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  satisfy*

$$|(T^{q_n})'(x)| \cdot |(T^{-q_n})'(x)| \geq C$$

*then  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  is minimal.*

**PROOF.** If  $T$  is not minimal we may choose  $x \in \mathbb{R}/\mathbb{Z}$  such that  $Y = \text{cl}(\cup_{n \in \mathbb{Z}} T^n x) \neq X$ . We can choose a (maximal) interval  $I_0 \subset X - Y$ ; then we claim that  $I_n := T^{-n} I_0 \subset X - Y$  are distinct (maximal) intervals. To see this we observe that by maximality  $I_0$  must be of the form  $I_0 = (a, b)$  (with  $a, b \in Y$ ). Thus  $I_n = (T^{-n}a, T^{-n}b)$  and if  $I_n \cap I_m \neq \emptyset$  then again by maximality  $I_n = I_m$  and, in particular,  $T^{-n}a = T^{-m}a$ . But if  $n \neq m$  then this means  $a$  is a periodic point, which contradicts  $T$  having an irrational rotation number (by Lemma 6.1 (ii)).

If  $|I_n|$  denotes the length of the interval  $I_n$ ,  $n \in \mathbb{Z}$ , then by the disjointness we see that  $\sum_{n \in \mathbb{Z}} |I_n| \leq 1$ . In particular,  $|I_n| \rightarrow 0$  as  $n \rightarrow +\infty$ .

However, we see that for all  $n \geq 1$ ,

$$\begin{aligned} |I_{q_n}| + |I_{-q_n}| &= \int_{I_0} (|(T^{q_n})'(x)| + |(T^{-q_n})'(x)|) dx \\ &\geq 2 \int_{I_0} (|(T^{q_n})'(x)| |(T^{-q_n})'(x)|)^{\frac{1}{2}} dx \\ &\geq 2C^{\frac{1}{2}} |I_0| \end{aligned}$$

(since arithmetic averages are larger than geometric averages). This contradicts  $|I_n| \rightarrow 0$ , and so completes the proof of the sublemma. ■

To make use of the assumption of the bounded variation of  $\log |T'|$  we need a second sublemma.

**SUBLEMMA 6.5.2.** *Fix  $x \in \mathbb{R}/\mathbb{Z}$  and write  $x_n = T^n(x)$ , for  $n \in \mathbb{Z}$ . There exists an increasing sequence  $q_n \rightarrow +\infty$  of natural numbers such that the intervals*

$$(x_0, x_{q_n}), (x_1, x_{q_n+1}), (x_2, x_{q_n+2}), \dots, (x_i, x_{q_n+i}), \dots, (x_{q_n}, x_{2q_n})$$

*are all disjoint.*

**PROOF.** By Corollary 6.3.1 we see that the order on  $\mathbb{R}/\mathbb{Z}$  of points in the orbit of  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  is the same as that of the rotation  $R_\rho : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ . Thus it suffices to prove this sublemma with  $R_\rho$  rather than  $T$ .

For each  $n \geq 1$  we want to choose the sequence  $q_n$  to correspond to the successive nearest approaches of  $\{T^m x\}$  to  $x$  (i.e.  $|T^{q_n} x - x| = \delta_n =$

$\inf\{|T^j x - x| : 1 \leq j \leq q_n - 1\}$ . Consider a typical interval  $(x_i, x_{q_n+i})$ ,  $0 \leq i \leq q_n$ , (the case  $(x_{q_n+i}, x_i)$  being similar) and assume for a contradiction that there exists  $x_r \in (x_i, x_{q_n+i})$ ,  $0 \leq r \leq 2q_n$ . There are two cases.

(a) Firstly, assume that  $r < i$ . We then know that

$$x_0 = R_\rho^{-r}(x_r) \in R_\rho^{-r}(x_i, x_{q_n+i}) = (x_{(i-r)}, x_{q_n+(i-r)}).$$

In particular,  $|x_{i-r} - x_0| < |x_{q_n+(i-r)} - x_{(i-r)}| = |x_{q_n} - x_0| = \delta_n$ . But since  $q_n > (i-r) > 0$  this contradicts the definition of  $q_n$ .

(b) Secondly, assume that  $r > i$ . We then know that

$$x_{r-i} = R_\rho^{-i}(x_r) \in R_\rho^{-i}(x_i, x_{q_n+i}) = (x_0, x_{q_n})$$

and so we see that  $r - i > q_n$  (from the definition of  $q_n$ ). But then  $x_{(r-i)-q_n} = R_\rho^{-q_n}(x_{r-i}) \in R_\rho^{-q_n}(x_0, x_{q_n}) = (x_{-q_n}, x_0)$  and, in particular,  $|x_{(r-i)-q_n} - x_0| < |x_{-q_n} - x_0| = |R_\rho^{q_n}(x_{-q_n}) - R_\rho^{q_n}(x_0)| = |x_0 - x_{q_n}| = \delta_n$ . Since  $0 < (r-i) - q_n < q_n$  this contradicts the definition of  $q_n$ .

This completes the proof of the sublemma. ■

Since the intervals in Sublemma 6.5.2 are disjoint we have for any  $n \geq 1$

$$\begin{aligned} \text{Var}(\log |T'|) &\geq \sum_{i=0}^{q_n} |\log |T'(x_i)| - \log |T'(x_{q_n+i})|| \\ &\geq \left| \sum_{i=0}^{q_n} \log |T'(x_i)| - \sum_{i=0}^{q_n} \log |T'(x_{q_n+i})| \right| \\ &= \left| \log \left( \frac{\prod_{i=0}^{q_n-1} |T'(T^i x_0)|}{\prod_{i=0}^{q_n-1} |T'(T^i x_{q_n})|} \right) \right| \tag{6.4} \\ &= \left| \log \left( \frac{|(T^{q_n})'(x_0)|}{|(T^{q_n})'(x_{q_n})|} \right) \right| \\ &= \left| \log |(T^{q_n})'(x_{q_n})(T^{-q_n})'(x_{q_n})| \right| \end{aligned}$$

(where by the chain rule  $(T^{-q_n})'(x_{q_n})(T^{q_n})'(x_0) = 1$ ). Since this holds for arbitrary  $x$  the point  $x_{q_n}$  can be replaced by an arbitrary point on the circle. If we take the exponential in identity (6.4) then Theorem 6.5 now follows from Sublemma 6.5.1 and 6.4. ■

REMARK. We should remark that the assumption that  $\log T'$  has bounded variation is necessary. If we relax this assumption then we have that  $T$  may be merely semi-conjugate to the rotation  $R_\rho$ .

### 6.3 Comments and references

Some basic results about rotation numbers can be found in [2, pp. 102-108], [5, chapter 12] and [1].

The question of when the conjugating map is differentiable is subtler (cf. [4], [3], [9] and [6], [7], [8] for variants).

#### References

1. V. Arnol'd, *Ordinary Differential Equations*, M.I.T. Press, Cambridge Mass., 1973.
2. R. Devaney, *An Introduction to Chaotic Dynamical Systems*, Addison-Wesley, New York, 1989.
3. P. Deligne, *Les difféomorphismes du cercle (d'après M. R. Herman)*, Exposé 477, Semin. Bourbaki 1975/76, Lect. Notes Math. 567, Springer, Berlin, 1977.
4. M. Herman, *Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations*, Publ. Math. (IHES) **49** (1979), 5-234.
5. A. Katok and B. Hasselblatt, *An Introduction to the Modern Theory of Dynamical Systems*, C.U.P., Cambridge, 1995.
6. Y. Katznelson and D. Ornstein, *The differentiability of the conjugation of certain diffeomorphisms of the circle*, Ergod. Th. and Dynam. Sys. **9** (1989), 643-680.
7. K. Khanin and Y. Sinai, *A new proof of M. Herman's theorem*, Commun. Math. Phys. **112** (1987), 89-101.
8. D. Rand, *Universality and renormalisation in dynamical systems*, New Directions in Dynamical Systems (T. Bedford and J. Swift, ed.), C.U.P., Cambridge, 1988, pp. 1-56.
9. J.-C. Yoccoz, *Conjugaison différentiable des difféomorphismes du cercle dont le nombre de rotation vérifie une condition Diophantienne*, Ann. Sci. Ec. Norm. Sup. **17** (1984), 333-361.

## INVARIANT MEASURES

In this chapter we shall introduce some basic definitions in ergodic theory.

## 7.1 Definitions and characterization of invariant measures

Let  $(X, \mathcal{B}, \mu)$  be a measure space. Assume that  $\mu$  is a probability measure, i.e.  $\mu(X) = 1$ .

DEFINITION. A measurable map  $T : X \rightarrow X$  (i.e.  $T^{-1}\mathcal{B} \subset \mathcal{B}$ ) is said to preserve the measure  $\mu$  if for any  $B \in \mathcal{B}$  we have  $\mu(B) = \mu(T^{-1}B)$ . Alternatively we say that  $\mu$  is  $T$ -invariant.

The next lemma gives a characterization in terms of integrable functions.

LEMMA 7.1.  $T$  preserves  $\mu$  iff  $\int f d\mu = \int f \circ T d\mu$  for all  $f \in L^1(X, \mathcal{B}, \mu)$ .

PROOF. This is easy, using indicator functions. ■

## 7.2 Borel sigma-algebras for compact metric spaces

We now specialize to homeomorphisms  $T : X \rightarrow X$  of a compact metric space  $X$ . In this case a natural sigma-algebra is the Borel sigma-algebra  $\mathcal{C}$ , i.e. the smallest sigma-algebra containing all of the open sets in  $X$ .

LEMMA 7.2.  $T$  preserves  $\mu$  iff  $\int f d\mu = \int f \circ T d\mu$  for all  $f \in C^0(X)$ .

PROOF. One way follows from Lemma 7.1. The other follows from the Hahn-Banach and Riesz representation theorems. ■

The next proposition shows that there always exists an invariant measure.

PROPOSITION 7.3 (EXISTENCE OF INVARIANT MEASURES). *Let  $X$  be a compact metric space and  $\mathcal{B}$  be the Borel sigma-algebra. Given any homeomorphism  $T : X \rightarrow X$  (or more generally, a continuous map) there exists at least one probability measure  $\mu$  preserving  $T$ .*

PROOF. Choose

- (1) a countable dense subset  $\{f_k \in C^0(X, \mathbb{R}) : k \geq 0\}$ , in the uniform topology
- (2) a point  $x \in X$ .

We can consider the averages  $\frac{1}{N} \sum_{n=0}^{N-1} f_k(T^n x) \in \mathbb{R}$ , for  $k \geq 0, N \geq 1$ ; then clearly  $-\|f_k\|_\infty \leq \frac{1}{N} \sum_{n=0}^{N-1} f_k(T^n x) \leq \|f_k\|_\infty$ . By compactness of the interval  $[-\|f_0\|_\infty, \|f_0\|_\infty]$  we can choose a sub-sequence  $\mathcal{N}^{(0)} = \{N_r^{(0)}\}_{r=0}^\infty \subset \mathbb{Z}^+$  such that

$$\frac{1}{N_r^{(0)}} \sum_{n=0}^{N_r^{(0)}-1} f_0(T^n x) \rightarrow C(f_0) \text{ as } r \rightarrow +\infty.$$

We repeat the argument with  $f_1$  replacing  $f_0$ . By compactness of the interval  $[-\|f_1\|_\infty, \|f_1\|_\infty]$  we can choose a subsequence  $\mathcal{N}^{(1)} = \{N_r^{(1)}\}_{r=0}^\infty \subset \mathcal{N}^{(0)}$  such that  $\frac{1}{N_r^{(1)}} \sum_{n=0}^{N_r^{(1)}-1} f_1(T^n x) \rightarrow C(f_1)$  as  $r \rightarrow +\infty$ .

Repeating the argument inductively we can arrange subsequences  $\mathcal{N}^{(m)} = \{N_r^{(m)}\}_{r=0}^\infty$  with  $\mathcal{N}^{(m)} \subset \mathcal{N}^{(m-1)} \subset \dots \subset \mathcal{N}^{(1)} \subset \mathcal{N}^{(0)}$  for  $m \geq 0$  such that

$$\frac{1}{N_r^{(m)}} \sum_{n=0}^{N_r^{(m)}-1} f_m(T^n x) \rightarrow C(f_m)$$

as  $r \rightarrow +\infty$ .

Next we choose a diagonal sub-sequence  $N_n = N_n^{(n)}$  and we conclude that for *any*  $k \geq 0$  we have that

$$\frac{1}{N_n^{(n)}} \sum_{n=0}^{N_n^{(n)}-1} f_k(T^n x) \rightarrow C(f_k)$$

as  $n \rightarrow +\infty$ .

We claim that for any  $f \in C^0(X)$  we have that  $\frac{1}{N_n^{(n)}} \sum_{n=0}^{N_n^{(n)}-1} f(T^n x)$ ,  $n \geq 1$ , converges (to a limit we shall call  $C(f)$ ) as  $n \rightarrow +\infty$ .

For any  $\epsilon > 0$  we choose  $\|f - f_k\|_\infty < \epsilon$  and thus

$$\limsup_{n \rightarrow +\infty} \left| \frac{1}{N_n^{(n)}} \sum_{n=0}^{N_n^{(n)}-1} f(T^n x) - C(f_k) \right| \leq \epsilon.$$

Since  $\epsilon > 0$  can be chosen arbitrarily small this is enough to show that the sequence  $\frac{1}{N_n^{(n)}} \sum_{n=0}^{N_n^{(n)}-1} f(T^n x)$  converges.

We can summarize the properties of the limit  $C(f)$  as follows:

- (1)  $\forall f_1, f_2 \in C^0(X)$  and  $\lambda_1, \lambda_2 \in \mathbb{R}$  we have  $C(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 C(f_1) + \lambda_2 C(f_2)$  and  $|C(f)| \leq \|f\|_\infty$ ,
- (2) if  $f(x) \geq 0 \forall x \in X$  we have  $C(f) \geq 0$ ,
- (3)  $C(1) = 1$ , where  $1(x) = 1, \forall x \in X$ ,
- (4)  $C(f \circ T) = C(f)$  (since it is true for each  $f = f_k$  by

■

### 7.3 Examples of invariant measures

It is illuminating to look at some simple examples.

**EXAMPLE 1 (ROTATIONS ON TORI).** Let  $X = \mathbb{T}^n = \mathbb{R}^n / \mathbb{Z}^n$  be a torus. Define a generalized rotation  $T : X \rightarrow X$  (or *Kronecker system*) by fixing a vector  $(a_1, \dots, a_n)$  and defining

$$T((x_1, \dots, x_n) + \mathbb{Z}^n) = (x_1 + a_1, \dots, x_n + a_n) + \mathbb{Z}^n.$$

We let  $\mathcal{B}$  be the Borel sigma-algebra. We let  $\mu$  be the Lebesgue-Haar measure and then naturally  $\mu$  is  $T$ -invariant. When  $n = 1$  this is simply rotation on a circle.

One way to see this is the following. For any  $f(x_1, \dots, x_n) \in C^0(X)$  we write it as a Fourier series:

$$f(x_1, \dots, x_n) = \sum_{k_1, \dots, k_n = -\infty}^{\infty} a_{k_1, \dots, k_n} e^{2\pi i(k_1 x_1 + \dots + k_n x_n)}.$$

Moreover, we see that  $a_0 = \int f dx_1 \dots dx_n$  (just by integrating both sides). The action of the rotation can be written as

$$\begin{aligned} (fT)(x_1, \dots, x_n) &= \sum_{k_1, \dots, k_n = -\infty}^{\infty} a_{k_1, \dots, k_n} e^{2\pi i(k_1(x_1 + a_1) + \dots + k_n(x_n + a_n))} \\ &= \sum_{k_1, \dots, k_n = -\infty}^{\infty} b_{k_1, \dots, k_n} e^{2\pi i(k_1 x_1 + \dots + k_n x_n)} \end{aligned}$$

where  $b_0 = \int fT dx_1 \dots dx_n$ , as before. But by comparing terms we see that  $b_0 = a_0$ .

**REMARK.** More generally this is a property of Haar measure: Assume that  $X$  is a (locally) compact second countable metric space which is also a group with identity element  $e$ , where multiplication between  $x, y \in X$  is denoted by  $xy$  and the inverse to  $x$  is denoted by  $x^{-1}$ . Assume that the map  $X \times X \rightarrow X$  given by  $(x, y) \rightarrow xy$  is continuous. By a general theorem (Haar measure theorem) there exists a (sigma) finite measure  $\mu$  on the Borel sigma-algebra for  $X$  such that

- (1)  $\mu(U) > 0$  for every non-empty open set  $U \subset X$ ,
- (2) for any Borel set  $A$  and  $g \in X$  we have  $\mu(gA) = \mu(A)$ .



More generally, we can take  $X$  to be a compact topological group and for any fixed element  $g \in X$  we define a transformation  $T : X \rightarrow X$  by  $Tx = gx$ . The (normalized) Haar measure is then an invariant probability measure.

EXAMPLE 2 (ADDING MACHINE). A slightly more exotic example is the adding machine transformation defined on  $X = \prod_{n=0}^{\infty} \{0, 1\}$ . This is a topological group under “addition with infinite carry”. In particular, let  $g = (1, 0, 0, 0, \dots)$  and define  $T : X \rightarrow X$  as follows: If  $x = (0, 0, \dots, 0, 1, a_k, a_{k+1}, a_{k+2}, \dots)$  then  $Tx = x + g = (1, 0, \dots, 0, 1, a_k, a_{k+1}, a_{k+2}, \dots)$ , and if  $x = (1, 1, \dots, 1, 0, a_k, a_{k+1}, a_{k+2}, \dots)$  then  $Tx = x + g = (0, 0, \dots, 0, 1, a_k, a_{k+1}, a_{k+2}, \dots)$ . The invariant measure here is given by the measure on cylinders,

$$[z_0, \dots, z_n] = \{x \in X_A : x_i = z_i \text{ for } 0 \leq i \leq n\}$$

by  $\mu([z_0, \dots, z_n]) = \left(\frac{1}{2}\right)^{n+1}$ .

EXAMPLE 3 (DOUBLING MAP). Let  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  be defined by  $T(x) = 2x \pmod{1}$ . Let  $\mathcal{B}$  be the usual Borel sigma-algebra. Let  $\mu$  be the usual Lebesgue-Haar measure. To show that  $\mu$  is invariant, it suffices to show that for each interval  $I = [a, b]$  we have that  $\mu(T^{-1}I) = \mu(I)$ . Writing

$$T^{-1}I = \left[\frac{a}{2}, \frac{b}{2}\right] \cup \left[\frac{a}{2} + \frac{1}{2}, \frac{b}{2} + \frac{1}{2}\right],$$

this becomes apparent.

TECHNICAL REMARK. The reason that it suffices to prove this only on intervals (or more generally on a basis for the topology) and not necessarily all sets is by virtue of the Kolmogorov extension theorem.

EXAMPLE 4 (MARKOV MEASURES). Let  $A$  be a  $k \times k$  matrix with entries either 0 or 1, and let  $\sigma : X_A \rightarrow X_A$  be the associated subshift of finite type.

Let  $P = \{P_{ij}\}$  be any  $k \times k$  stochastic matrix, i.e.

- (1)  $P_{ij} \geq 0$ ,
- (2) The row sums are unity (i.e.  $\sum_{j=1}^k P_{ij} = 1$ ).

In addition we ask that  $P_{ij} = 0$  whenever  $A(i, j) = 0$ . The second condition means that  $(1, \dots, 1)^T$  is a right eigenvector. We let  $p = (p_1, \dots, p_k)$  be the left eigenvector with eigenvalue 1. We can define a measure  $\mu$  on  $X_A$  on the cylinders by

$$\mu([z_{-m}, \dots, z_n]) = p_{z_{-m}} P_{z_{-m}, z_{-m+1}} P_{z_{-m+1}, z_{-m+2}} \cdots P_{z_{n-1}, z_n}.$$

This is enough to define the measure on the entire sigma-algebra (i.e. Kolmogorov extension theorem). Clearly the measure is shift-invariant. The use

of  $p$  in the definition is to ensure that it is consistent, i.e.

$$\begin{aligned}
& \mu([z_{-m+1}, \dots, z_n]) \\
&= p_{z_{-m+1}} P_{z_{-m+1}, z_{-m+2}} P_{z_{-m+1}, z_{-m+2}} \cdots P_{z_{n-1}, z_n} \\
&= \left( \sum_{\substack{1 \leq z_{-m} \leq k \\ A(z_{-m}, z_{m-1})=1}} \frac{p_{z_{-m}}}{p_{z_{-m+1}}} P_{z_{-m}, z_{-m+1}} \right) p_{z_{-m+1}} P_{z_{-m+1}, z_{-m+2}} \cdots P_{z_{n-1}, z_n} \\
&= \sum_{\substack{1 \leq z_{-m} \leq k \\ A(z_{-m}, z_{m-1})=1}} \mu([z_{-m}, \dots, z_n])
\end{aligned}$$

This is called a *Markov measure*.

**SPECIAL CASE: FULL SHIFT.** In the special case that *all* of the entries in the matrix are 1s the shift  $X_A = \prod_{n \in \mathbb{Z}} \{1, \dots, k\}$  is called a full shift on  $k$  symbols. In this case let  $p = (p_1, \dots, p_k)$  be any probability vector (i.e.  $p_1, \dots, p_k \geq 0$  and  $p_1 + \dots + p_k = 1$ ). We then define  $\mu([z_{-m}, \dots, z_n]) = p_{z_{-m}} \cdots p_{z_n}$ . This is called a Bernoulli measure.

**ANOTHER EXAMPLE.** In section 12.2 one can find another example involving expanding maps of the interval.

## 7.4 Invariant measures for other actions

In this section we shall consider the case where the discrete transformation  $T$  is replaced by either a flow (an  $\mathbb{R}$ -action) or a  $\mathbb{Z}^n$ -action. Although this will not play an important role in the sequel, we shall include a few comments here for completeness.

**DEFINITION (INVARIANT MEASURES FOR FLOWS).** A flow  $T_t : X \rightarrow X$  is a one-parameter family of measurable invertible maps (i.e.  $(x, t) \mapsto T_t x$  is measurable) which satisfy

- (1)  $T_t T_s = T_{s+t}$  for all  $s, t \in \mathbb{R}$ ,
- (2)  $T_0$  is the identity map.

**DEFINITION.** A measurable flow  $T_t : X \rightarrow X$  is said to preserve the measure  $\mu$  if for any  $B \in \mathcal{B}$  we have  $\mu(B) = \mu(T_t^{-1} B)$  for all  $t \in \mathbb{R}$ .

Alternatively we say that  $\mu$  is  $T$ -invariant.

**EXAMPLE 5 (SUSPENDED FLOWS).** Consider a measure space  $(X, \mathcal{B}, \mu)$  and a transformation  $T : X \rightarrow X$  which preserves the measure  $\mu$ . Given any function  $r : X \rightarrow \mathbb{R}^+$  with  $\int r d\mu < +\infty$  we define a new space by

$$X^r = \{(x, t) \in X \times \mathbb{R} : 0 \leq t \leq r(x)\}$$

where we identify the points  $(x, r(x)) = (Tx, 0)$ . We associate to this the product sigma algebra (from the sigma-algebra on  $X$  and the sigma-algebra on the real line). We define a (probability) measure  $\mu^r$  on  $X^r$  by  $d\mu^r = \frac{1}{\int r d\mu} d\mu \times dt$ . (Here the factor  $\frac{1}{\int r d\mu}$  appears to normalize this to be a probability measure.)

We define a flow  $T_t : X^r \rightarrow X^r$  by

$$T_t(x, u) = (T^{n-1}x, u + t - \sum_{i=0}^{n-1} r(T^i x))$$

whenever  $t \geq 0$  and  $0 \leq u + t - \sum_{i=0}^{n-1} r(T^i x) \leq r(T^n x)$ , and

$$T_t(x, u) = (T^{-n}x, u + t + \sum_{i=-n}^{-1} r(T^i x))$$

whenever  $t \leq 0$  and  $0 \leq u + t + \sum_{i=-n}^{-1} r(T^i x) \leq r(T^{-n+1}x)$ . It is easy to see that this is a measurable flow, and that the measure  $\mu^r$  is invariant.

EXAMPLE 6 (GEODESIC AND HOROCYCLE FLOWS). Consider the group

$$G = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{R}, \quad ad - bc = 1 \right\}$$

This is a *locally* compact topological group. There is an associated Haar measure defined by  $d\mu(g) = \frac{1}{|d|} (da)(db)(dc)$  on  $G$  (although the total measure of the space  $G$  is infinite!).

Assume that  $\Gamma \subset G$  is a discrete subgroup such that the measure of the quotient space  $G/\Gamma$  is finite (i.e. we can find  $B \subset G$  such that  $B/\Gamma = G/\Gamma$  and  $\mu(B) < +\infty$ ).

We can define two flows on  $G/\Gamma$  as follows:

- (1) the geodesic flow  $\phi_t : g\Gamma \rightarrow g_t g\Gamma$  defined by left multiplication by the matrices  $g_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}$ ;
- (2) the horocycle flow  $\psi_t : g\Gamma \rightarrow h_t g\Gamma$  defined by left multiplication by the matrices  $h_t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$ .

Each of these flows preserves the measure on  $G/\Gamma$  corresponding to  $\mu$ .

REMARK. Geometrically these correspond to flows on the unit tangent bundle of compact surfaces of constant negative curvature. The invariant measure then corresponds to the Liouville measure.

Let  $T_1, \dots, T_k : X \rightarrow X$  be a family of commuting measurable (invertible) transformations on a measurable space  $(X, \mathcal{B})$ . We can define a  $\mathbb{Z}^k$ -action  $A : \mathbb{Z}^k \times X \rightarrow X$  by  $A(n_1, \dots, n_k; x) = T^{n_1} \circ \dots \circ T^{n_k} x$ .

DEFINITION. We say that a measure  $\mu$  is invariant under the action  $A$  if for every  $B \in \mathcal{B}$  we have that  $\mu(B) = \mu(T_1^{-n_1} \dots T_k^{-n_k} B)$  for all  $(n_1, \dots, n_k) \in \mathbb{Z}^k$  (or equivalently,  $\mu(B) = \mu(T_i^{-1} B)$  for all  $i = 1, \dots, k$ ).

### 7.5 Comments and references

More examples of invariant measures for different transformations can be found in [2] and [4], and also particularly in [3].

Details on the theory of suspended flows can be found in [1, pp. 292-295].

#### References

1. I. Cornfeld, S. Fomin and Y. Sinai, *Ergodic Theory*, Springer, Berlin, 1982.
2. A. Katok and B. Hasselblatt, *An Introduction to the Modern Theory of Dynamical Systems*, C.U.P., Cambridge, 1995.
3. F. Schweiger, *The Ergodic Theory of Fibred Systems and Metric Number Theory*, O.U.P., Oxford, 1994.
4. P. Walters, *An Introduction to Ergodic Theory*, Springer, Berlin, 1982.

## MEASURE THEORETIC ENTROPY

In this chapter we shall show how to associate to a measure preserving transformation an important quantity called the measure theoretic entropy. This gives important information on the dynamics of the map (cf. Chapter 12) and is useful in classifying measure preserving transformations.

The essential results are contained in sections 8.1-8.3. If one accepts Sinai's result on strong generators (Lemma 8.8) without proof then sections 8.4-8.8 will only be required again in chapter 12.

## 8.1 Partitions and conditional expectations

Let  $\alpha = \{A_i\}_{i \in I}$  be a countable measurable partition of the probability space  $(X, \mathcal{B}, \mu)$ , i.e.

- (i)  $X = \cup_i A_i$  (up to a set of zero  $\mu$ -measure), and
- (ii)  $A_i \cap A_j = \emptyset$  for  $i \neq j$  (up to a set of zero  $\mu$ -measure).

DEFINITION. We define the information function  $I(\alpha) : X \rightarrow \mathbb{R}$  by

$$I(\alpha)(x) = - \sum_i \log \mu(A_i) \chi_{A_i}(x),$$

i.e.  $I(\alpha)(x) = -\log \mu(A_i)$  if  $x \in A_i$ .

Consider a *sub*-sigma-algebra  $\mathcal{A} \subset \mathcal{B}$ ; then we can define a measure space  $(X, \mathcal{A}, \mu)$  with respect to the smaller sigma-algebra. For any  $f \in L^1(X, \mathcal{B}, d\mu)$  we can define a measure on the measure space  $(X, \mathcal{A}, \mu)$  by  $\mu_{\mathcal{A}}(A) = \int_A f d\mu$ , for  $A \in \mathcal{A}$ . Clearly,  $\mu_{\mathcal{A}} \ll \mu$  (where  $\mu$  is here defined on  $\mathcal{A}$ ).

DEFINITION. By the Radon-Nikodym theorem there is a unique function  $E(f|\mathcal{A}) := \frac{d\mu_{\mathcal{A}}}{d\mu} \in L^1(X, \mathcal{A}, d\mu)$  which is called the *conditional expectation*.

Since in general  $\mathcal{A}$  is *strictly* contained in  $\mathcal{B}$  then  $E(f|\mathcal{A})$  may be very different from  $f$ , since it must be measurable on a smaller sigma-algebra. For example, if  $\mathcal{A} = \{X, \emptyset\}$  then  $E(f|\mathcal{A})$  is the constant function  $\int f d\mu$ .

The main properties of  $E(f|\mathcal{A})$  are

- (i)  $\int_A E(f|\mathcal{A}) d\mu = \int_A f d\mu$  for all  $A \in \mathcal{A}$ ,
- (ii)  $E(f|\mathcal{A}) \in L^1(X, \mathcal{A}, d\mu)$ ,

(The first two properties are just the definition repeated.),

- (iii) if  $f \in L^1(X, \mathcal{B}, \mu)$  and  $g \in L^\infty(X, \mathcal{A}, d\mu)$  then  $E(fg|\mathcal{A}) = gE(f|\mathcal{A})$ ,
- (iv) if  $f \in L^1(X, \mathcal{B}, \mu)$  and  $\mathcal{A}_2 \subset \mathcal{A}_1 \subset \mathcal{B}$  then  $E(E(f|\mathcal{A}_1)|\mathcal{A}_2) = E(f|\mathcal{A}_2)$ ,
- (v) if  $f \in L^1(X, \mathcal{B}, \mu)$  then  $|E(f|\mathcal{A})| \leq E(|f||\mathcal{A})$ , and if  $f, g \in L^2(X, \mathcal{B}, d\mu)$  and  $\frac{1}{p} + \frac{1}{q} = 1$  then  $E(|fg||\mathcal{A}) \leq E(|f|^p|\mathcal{A})^{1/p} E(|g|^q|\mathcal{A})^{1/q}$ ,
- (vi) If  $T$  preserves  $\mu$  then  $E(f|\mathcal{A})T = E(f \circ T|T^{-1}\mathcal{A})$ , where  $T^{-1}\mathcal{A} = \{T^{-1}A : A \in \mathcal{A}\}$ .

Parts (iii)-(vi) are a trivial exercise from the definitions (cf. [5, p. 10]).

DEFINITION. Given any sub-sigma-algebra  $\mathcal{A} \subset \mathcal{B}$  we can define the *conditional information function*  $I(\alpha|\mathcal{A}) : X \rightarrow \mathbb{R}$  by

$$I(\alpha|\mathcal{A})(x) = - \sum_i \log \mu(A_i|\mathcal{A})(x) \chi_{A_i}(x)$$

where we write  $\mu(A_i|\mathcal{A})(x) = E(\chi_{A_i}|\mathcal{A})(x)$  (called the *the conditional measure*).

Assume that we “know” the position of the point  $x$  relative to  $\mathcal{A}$ , then  $I(\alpha|\mathcal{A})$  is an indicator of how much additional information we get from “knowing” the position of the point  $x$  relative to the partition  $\alpha$ .

The following properties all come directly from the definition.

LEMMA 8.1.

- (1) When  $\mathcal{A} = \{\emptyset, X\}$  then  $E(A_i|\{\emptyset, X\})(x) = \mu(A_i)$  and  $I(\alpha|\{\emptyset, X\})(x) = I(\alpha)(x)$ .
- (2) If  $T : X \rightarrow X$  preserves the measure  $\mu$  then  $I(\alpha|\mathcal{A})(Tx) = I(T^{-1}\alpha|T^{-1}\mathcal{A})(x)$  (where  $T^{-1}\alpha = \{T^{-1}A_i\}$ ).
- (3) If  $\alpha \subset \mathcal{A}$  then  $I(\alpha|\mathcal{A}) = 0$  almost everywhere.

We can associate to each partition  $\alpha$  the sigma-algebra  $\hat{\alpha}$  generated by  $\alpha$ .

DEFINITION. Given two partitions  $\alpha, \beta$  we define their *refinement*

$$\alpha \vee \beta = \{A_i \cap B_j : A_i \in \alpha, B_j \in \beta\}.$$

Given two sigma-algebras  $\mathcal{A}, \mathcal{B}$  we denote by  $\mathcal{A} \vee \mathcal{B}$  the sigma-algebra generated by  $\{A \cap B : A \in \alpha, B \in \beta\}$ .

The following lemma will prove very useful throughout this chapter.

LEMMA 8.2 (BASIC IDENTITY FOR INFORMATION). *Given partitions  $\alpha, \beta$  and a third  $\gamma$  with associated sigma-algebra  $\hat{\gamma}$  we have that*

$$I(\alpha \vee \beta | \hat{\gamma})(x) = I(\alpha | \hat{\beta} \vee \hat{\gamma})(x) + I(\beta | \hat{\gamma})(x)$$

(almost everywhere).

PROOF. Observe that for any function  $g \in L^1(X, \mathcal{B}, \mu)$  we have that

$$E(g | \hat{\gamma})(x) = \sum_{C \in \gamma} \chi_C \frac{\int_C g(x) d\mu}{\mu(C)}.$$

In particular, for  $B \in \beta$  we can set  $g(x) = \chi_B(x)$  and then we get

$$\mu(B | \hat{\gamma})(x) = \sum_{C \in \gamma} \chi_C(x) \frac{\mu(B \cap C)}{\mu(C)}$$

and therefore

$$I(\beta | \hat{\gamma})(x) = - \sum_{C \in \gamma, B \in \beta} \chi_{C \cap B}(x) \log \left( \frac{\mu(B \cap C)}{\mu(C)} \right). \quad (8.1)$$

The partition  $\beta \vee \gamma$  (with elements of the form  $B \cap C$  with  $B \in \beta, C \in \gamma$ ) gives that

$$I(\alpha | \hat{\gamma} \vee \hat{\beta})(x) = - \sum_{C \in \gamma, B \in \beta, A \in \alpha} \chi_{A \cap B \cap C}(x) \log \left( \frac{\mu(A \cap B \cap C)}{\mu(B \cap C)} \right). \quad (8.2)$$

Adding (8.1) and (8.2) gives that

$$\begin{aligned} & I(\beta | \hat{\gamma})(x) + I(\alpha | \hat{\gamma} \vee \hat{\beta})(x) \\ &= - \sum_{C \in \gamma, B \in \beta, A \in \alpha} \chi_{A \cap B \cap C}(x) \left( \log \left( \frac{\mu(B \cap C)}{\mu(C)} \right) + \log \left( \frac{\mu(A \cap B \cap C)}{\mu(B \cap C)} \right) \right) \\ &= - \sum_{C \in \gamma, B \in \beta, A \in \alpha} \chi_{A \cap B \cap C}(x) \log \left( \frac{\mu(A \cap B \cap C)}{\mu(C)} \right) \\ &= I(\alpha \vee \beta | \hat{\gamma})(x). \end{aligned}$$

This completes the proof. ■

If  $\alpha$  and  $\beta$  are partitions we write  $\alpha < \beta$  if every element of  $\alpha$  is a union of elements of  $\beta$ . In this case  $\alpha \vee \beta = \beta$ .

COROLLARY 8.2.1. If  $\alpha < \beta$  then  $I(\alpha | \hat{\gamma})(x) \leq I(\beta | \hat{\gamma})(x)$ .

### 8.2 The entropy of a partition

DEFINITION. We define the *entropy* of the partition  $\alpha$  by

$$H(\alpha) = \int I(\alpha)(x) d\mu(x) = - \sum_{A \in \alpha} \mu(A) \log \mu(A).$$

Given a partition  $\alpha$  and a sub-sigma-algebra  $\mathcal{A} \subset \mathcal{B}$  we define the *conditional entropy* by  $H(\alpha|\mathcal{A}) = \int I(\alpha|\mathcal{A})(x) d\mu(x)$ .

LEMMA 8.3.

- (1) When  $\mathcal{A} = \{\emptyset, X\}$  then  $H(\alpha|\{\emptyset, X\}) = H(\alpha)$ .
- (2) If  $T : X \rightarrow X$  preserves the measure  $\mu$  then  $H(\alpha|\mathcal{A}) = H(T^{-1}\alpha|T^{-1}\mathcal{A})$ .
- (3) If  $\alpha \subset \mathcal{A}$  then  $H(\alpha|\mathcal{A}) = 0$ .
- (4) Given  $\alpha, \gamma$  we have  $H(\alpha|\hat{\gamma}) \leq H(\alpha)$ .

PROOF. Parts (1), (2) and (3) follow by integrating the corresponding results for the information function in Lemma 8.1.

For part (4) we have

$$\begin{aligned} H(\alpha|\gamma) &= - \sum_{A \in \alpha, C \in \gamma} \mu(A \cap C) \log \left( \frac{\mu(A \cap C)}{\mu(C)} \right) \\ &= - \sum_{C \in \gamma} \mu(C) \left[ \sum_{A \in \alpha} \frac{\mu(A \cap C)}{\mu(C)} \log \left( \frac{\mu(A \cap C)}{\mu(C)} \right) \right] \\ &\leq - \sum_{A \in \alpha} \left[ \sum_{C \in \gamma} \mu(A \cap C) \right] \log \left[ \sum_{C \in \gamma} \mu(A \cap C) \right] \\ &\leq - \sum_{A \in \alpha} \mu(A) \log \mu(A) = H(\alpha) \end{aligned}$$

since for fixed  $A \in \alpha$  we can bound

$$- \sum_{C \in \gamma} \frac{\mu(A \cap C)}{\mu(C)} \log \left( \frac{\mu(A \cap C)}{\mu(C)} \right) \leq - \left[ \sum_{C \in \gamma} \mu(A \cap C) \right] \log \left[ \sum_{C \in \gamma} \mu(A \cap C) \right]$$

using concavity of  $t \mapsto -t \log t$ . (A more general result appears in Lemma 8.13.)

■

LEMMA 8.4 (BASIC IDENTITY FOR ENTROPY). *Given partitions  $\alpha, \beta$  and a third  $\gamma$  with associated sigma-algebra  $\hat{\gamma}$  we have that*

$$H(\alpha \vee \beta|\hat{\gamma}) = H(\alpha|\hat{\beta} \vee \hat{\gamma}) + H(\beta|\hat{\gamma}).$$



**COROLLARY 8.4.1** (“MONOTONICITY” OF ENTROPY FOR PARTITIONS).  
*Given two partitions  $\alpha, \beta$  with  $\alpha < \beta$  we have that  $H(\alpha|\hat{\gamma}) \leq H(\beta|\hat{\gamma})$  (and, in particular  $H(\alpha) \leq H(\beta)$ )*

With the next definition, we begin to re-introduce measure preserving transformations.

**DEFINITION.** Assume that  $T : X \rightarrow X$  preserves  $\mu$ . Given a partition  $\alpha = \{A_i\}$  we write

$$\bigvee_{i=0}^{n-1} T^{-i} \alpha = \{A_{r_0} \cap T^{-1} A_{r_1} \cap \cdots \cap T^{-(n-1)} A_{r_{n-1}} : A_{r_i} \in \alpha, i = 0, \dots, n-1\}.$$

**NOTATIONAL COMMENT.** Frequently it proves convenient to drop the circumflex (hat) over  $\hat{\gamma}$ . Thus if we write  $H(\alpha|\beta)$ , say, we understand this to mean  $H(\alpha|\hat{\beta})$ .

For  $n \geq 1$  we can write  $H_n(\alpha) = H(\bigvee_{i=0}^{n-1} T^{-i} \alpha)$ . By the above estimates we have that

$$\begin{aligned} H_{n+m}(\alpha) &= H\left(\bigvee_{i=0}^{n+m-1} T^{-i} \alpha\right) \\ &= H\left(\bigvee_{i=0}^{n-1} T^{-i} \alpha\right) + H\left(\bigvee_{i=n}^{n+m-1} T^{-i} \alpha \mid \bigvee_{i=0}^{n-1} T^{-i} \alpha\right) \\ &\leq H\left(\bigvee_{i=0}^{n-1} T^{-i} \alpha\right) + H\left(\bigvee_{i=n}^{n+m-1} T^{-i} \alpha\right) \\ &= H_n(\alpha) + H_m(\alpha). \end{aligned} \tag{8.3}$$

Thus the sequence  $H_n(\alpha)$ ,  $n \geq 1$ , is subadditive (which shows that the limit in the following definition exists).

**DEFINITION.** We define the *entropy of the partition  $\alpha$*  relative to the transformation  $T : X \rightarrow X$  as the limit  $h(T, \alpha) = \lim_{n \rightarrow +\infty} \frac{H_n(\alpha)}{n}$ .

Notice that in particular from (8.3) we have that  $0 \leq h(T, \alpha) \leq H(\alpha)$ . The following result gives an equivalent characterization.

**PROPOSITION 8.5** (ALTERNATIVE DEFINITION OF  $h(T, \alpha)$ ).

$$h(T, \alpha) = \lim_{n \rightarrow +\infty} H(\alpha \mid \bigvee_{i=1}^{n-1} T^{-i} \alpha).$$

(*N.B. Sometimes it is convenient to write this limit as  $H(\alpha \mid \bigvee_{i=1}^{\infty} T^{-i} \alpha)$ .)*

**PROOF.** Using Lemma 8.4 we see that

$$\begin{aligned} H(\bigvee_{i=0}^{n-1} T^{-i} \alpha) &= H(\alpha \mid \bigvee_{i=1}^{n-1} T^{-i} \alpha) + H(\bigvee_{i=0}^{n-2} T^{-i} \alpha) \\ &= H(\alpha \mid \bigvee_{i=1}^{n-1} T^{-i} \alpha) + H(\alpha \mid \bigvee_{i=1}^{n-2} T^{-i} \alpha) + H(\bigvee_{i=0}^{n-3} T^{-i} \alpha) \\ &\dots \\ &= \sum_{r=2}^n H(\alpha \mid \bigvee_{i=1}^{r-1} T^{-i} \alpha) + H(\alpha). \end{aligned}$$

We then see that

$$\lim_{n \rightarrow +\infty} \frac{1}{n} H(\vee_{i=0}^{n-1} T^{-i} \alpha) = \lim_{n \rightarrow +\infty} H(\alpha | \vee_{i=1}^{n-1} T^{-i} \alpha)$$

as required (since if  $a_n \rightarrow a$  for any sequence of real numbers then  $\frac{a_1 + \dots + a_n}{n} \rightarrow a$ ). ■

The entropy of the transformation relative to two different partitions is described by the following inequality.

LEMMA 8.6. *For finite entropy partitions  $\alpha, \beta$  of  $X$  we have that*

$$h(T, \alpha) \leq h(T, \beta) + H(\alpha | \beta).$$

PROOF. Since  $(\vee_{i=0}^{n-1} T^{-i} \alpha) \vee (\vee_{i=0}^{n-1} T^{-i} \beta) > \vee_{i=0}^{n-1} T^{-i} \alpha$  we have that

$$\begin{aligned} H(\vee_{i=0}^{n-1} T^{-i} \alpha) &\leq H((\vee_{i=0}^{n-1} T^{-i} \alpha) \vee (\vee_{i=0}^{n-1} T^{-i} \beta)) \\ &= H(\vee_{i=0}^{n-1} T^{-i} \beta) + H(\vee_{i=0}^{n-1} T^{-i} \alpha | \vee_{i=0}^{n-1} T^{-i} \beta) \end{aligned}$$

(where we use Lemma 8.4, with  $\gamma$  being the trivial partition, for the last line). We next estimate

$$\begin{aligned} &H(\vee_{i=0}^{n-1} T^{-i} \alpha | \vee_{i=0}^{n-1} T^{-i} \beta) \\ &= H(\alpha | \vee_{i=0}^{n-1} T^{-i} \beta) + H(\vee_{i=1}^{n-1} T^{-i} \alpha | \alpha \vee (\vee_{i=0}^{n-1} T^{-i} \beta)) \\ &\leq H(\alpha | \beta) + H(\vee_{i=1}^{n-1} T^{-i} \alpha | \vee_{i=1}^{n-1} T^{-i} \beta) \\ &\leq H(\alpha | \beta) + H(\vee_{i=0}^{n-2} T^{-i} \alpha | \vee_{i=0}^{n-2} T^{-i} \beta). \end{aligned}$$

Proceeding inductively gives us

$$H(\vee_{i=0}^{n-1} T^{-i} \alpha | \vee_{i=0}^{n-1} T^{-i} \beta) \leq nH(\alpha | \beta).$$

Finally, we see that

$$\frac{1}{n} H(\vee_{i=0}^{n-1} T^{-i} \alpha) \leq \frac{1}{n} H(\vee_{i=0}^{n-1} T^{-i} \beta) + H(\alpha | \beta).$$

Letting  $n \rightarrow +\infty$  gives the correct inequality. ■

COROLLARY 8.6.1. *For finite entropy partitions  $\alpha, \beta$  of  $X$  we have that*

$$|h(T, \beta) - h(T, \alpha)| \leq H(\beta | \alpha) + H(\alpha | \beta).$$

PROOF. By interchanging  $\alpha$  and  $\beta$  in Lemma 8.6 we get that  $h(T, \beta) \leq h(T, \alpha) + H(\beta | \alpha)$ .

### 8.3 The entropy of a transformation

Consider a measure preserving transformation  $T : X \rightarrow X$  on a probability space  $(X, \mathcal{B}, \mu)$ . We want to associate to this a numerical invariant. We start from the definition of the entropy relative to a partition  $\alpha$  and then remove the dependence on  $\alpha$  by taking a supremum.

**DEFINITION.** We define the *measure theoretic entropy* of  $T : X \rightarrow X$  for the probability space  $(X, \mathcal{B}, \mu)$  by  $h_\mu(T) = \sup_{\{\alpha: H(\alpha) < +\infty\}} h(T, \alpha)$ .

We write the measure  $\mu$  as a subscript not only to remind us that there is an ambient measure, but also to distinguish the notation from that of topological entropy in chapter 3.

As one might imagine, it can be very difficult to compute the measure theoretic entropy from the definition given. We now want to describe a very important method of *practical* computation. We begin with a result which replaces the supremum in the definition of the measure theoretic entropy with a limit.

**LEMMA 8.7 (ABRAMOV).** *Let  $\beta_1 \subset \beta_2 \subset \dots \subset \beta_k \subset \mathcal{B}$  be an increasing sequence of partitions with  $H(\beta_k) < +\infty$ ,  $\forall k \geq 1$ ; and such that  $\cup_n \beta_k$  generates the sigma-algebra  $\mathcal{B}$ . Then  $h_\mu(T) = \lim_{k \rightarrow +\infty} h(T, \beta_k)$ .*

We shall return to the proof in section 8.7.

The following definition gives us a way to generate the increasing partitions.

**DEFINITION.** We say that a partition  $\alpha$  with  $H(\alpha) < +\infty$  is called a *strong generator* for the probability space  $(X, \mathcal{B}, \mu)$  if  $\vee_{i=0}^{\infty} T^{-i}\alpha = \mathcal{B}$ .

If  $T$  is invertible, then we say that a partition  $\alpha$  with  $H(\alpha) < +\infty$  is called a *generator* for the probability space  $(X, \mathcal{B}, \mu)$  if  $\vee_{i=-\infty}^{\infty} T^{-i}\alpha = \mathcal{B}$ .

**LEMMA 8.8 (SINAI).** *If  $\alpha$  is a (strong) generator then  $h_\mu(T) = h(T, \alpha)$ .*

We shall return to the proof in section 8.7. Before developing the theory needed to prove these two results we shall use them to compute the measure theoretic entropy of some simple examples.

**Example 1 (doubling map).** Let  $X = \mathbb{R}/\mathbb{Z}$ ,  $\mathcal{B}$  denote the Borel sigma-algebra, and  $\mu$  the Haar-Lebesgue measure. We let  $T : X \rightarrow X$  be the doubling map  $T(x) = 2x \pmod{1}$ . Let  $\alpha = \{[0, \frac{1}{2}), [\frac{1}{2}, 1)\}$ ; then observe that

$$\alpha \vee T^{-1}\alpha = \left\{ \left[0, \frac{1}{4}\right), \left[\frac{1}{4}, \frac{1}{2}\right), \left[\frac{1}{2}, \frac{3}{4}\right), \left[\frac{3}{4}, 1\right) \right\}$$

and more generally,

$$\bigvee_{i=0}^{n-1} T^{-i} \alpha = \left\{ \left[ \frac{i}{2^n}, \frac{i+1}{2^n} \right) : i = 0, \dots, 2^n - 1 \right\}.$$

We can now calculate

$$\begin{aligned} H(\bigvee_{i=0}^{n-1} T^{-i} \alpha) &= - \sum_{i=0}^{2^n-1} \mu \left( \left[ \frac{i}{2^n}, \frac{i+1}{2^n} \right) \right) \log \left( \left[ \frac{i}{2^n}, \frac{i+1}{2^n} \right) \right) \\ &= - \sum_{i=0}^{2^n-1} \left( \frac{1}{2^n} \right) \log \left( \frac{1}{2^n} \right) \\ &= -2^n \left( \frac{1}{2^n} \right) \log \left( \frac{1}{2^n} \right) \\ &= n \log 2. \end{aligned}$$

Thus we see that  $\frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i} \alpha) = \log 2$  and thus letting  $n \rightarrow +\infty$  gives that  $h_\mu(T) = \log 2$ .

**Example 2 (rotations on the circle).** Let  $X = \mathbb{R}/\mathbb{Z}$ , let  $\mathcal{B}$  denote the Borel sigma-algebra, and let  $\mu$  be the Haar-Lebesgue measure. We let  $T : X \rightarrow X$  be the rotation  $T(x) = x + a \pmod{1}$  for some fixed values  $a \in \mathbb{R}$ .

First assume that  $a = \frac{p}{q}$  is a rational number. For any partition  $\beta$  we see that  $T^{-q} \beta = \beta$ . Thus

$$\bigvee_{k=0}^{nq-1} T^{-k} \beta = \bigvee_{k=0}^{q-1} T^{-k} \beta$$

and so in particular

$$\begin{aligned} h_\mu(T, \beta) &= \lim_{n \rightarrow +\infty} \frac{1}{qn} H \left( \bigvee_{k=0}^{nq-1} T^{-k} \beta \right) \\ &= \lim_{n \rightarrow +\infty} \frac{1}{qn} H \left( \bigvee_{k=0}^{q-1} T^{-k} \beta \right) \\ &= 0. \end{aligned}$$

Thus the measure theoretic entropy of any partition is zero, and thus the measure theoretic entropy of the transformation,  $h_\mu(T) = \sup_{H(\beta) < +\infty} h(T, \beta) = 0$ .

Next, assume that  $a$  is irrational. We let  $\beta = \{[0, \frac{1}{2}), [\frac{1}{2}, 1)\}$ . Since the sequence  $\frac{1}{2} + na \pmod{1}$  is dense in the unit circle (Weyl's theorem) we see that the partition is (strong) generating. Moreover, we see that  $\mathcal{B} = \bigvee_{k=1}^{\infty} T^{-k} \beta$ . As we observed before

$$h(T, \beta) = \lim_{n \rightarrow +\infty} H(\beta | \bigvee_{k=1}^{n-1} T^{-k} \beta) = 0$$

and therefore  $h_\mu(T) = h(T, \beta) = 0$ .

As one might imagine, a similar method applies to rotation on tori  $\mathbb{R}^n / \mathbb{Z}^n$ ,  $n \geq 2$ .

**Example 3 (Markov measures).** Let  $\sigma : X \rightarrow X$  denote a subshift of finite type

$$X = \{(x_n) \in \prod_{-\infty}^{+\infty} \{1, \dots, k\} : A_{x_n x_{n+1}} = 1\}$$

where  $A = (A_{ij})$  is a  $k \times k$  matrix with entries either zero or unity.

We associate to this a  $k \times k$  stochastic matrix  $P = (P_{ij})$  (cf. Example 4 in section 7.3) and let  $p = (p_1, \dots, p_k)$  be the left eigenvector associated to the left eigenvalue unity.

The partition  $\alpha = \{[1], \dots, [k]\}$  for  $X$  is generating. Let  $\sigma : X \rightarrow X$  denote the shift transformation. The refined partition  $\bigvee_{k=0}^{n-1} T^{-k} \alpha$  consists of “cylinder” sets of the form

$$[z_0, \dots, z_{n-1}] = \{(x_n)_{n \in \mathbb{Z}} \in X : x_i = z_i, 0 \leq i \leq n-1\}$$

where  $z_i \in \{1, \dots, k\}$ .

By the definition of the Markov measure  $\mu$  associated to  $P$  we have that

$$\mu([z_0, \dots, z_{n-1}]) = p_{z_0} P_{z_0 z_1} P_{z_1 z_2} \cdots P_{z_{n-2} z_{n-1}}.$$

We explicitly compute:

$$\begin{aligned} & H(\bigvee_{k=0}^{n-1} T^{-k} \alpha) \\ &= - \sum_{[z_0, \dots, z_{n-1}]} \mu([z_0, \dots, z_{n-1}]) \log \mu([z_0, \dots, z_{n-1}]) \\ &= - \sum_{[z_0, \dots, z_{n-1}]} p_{z_0} P_{z_0 z_1} P_{z_1 z_2} \cdots P_{z_{n-2} z_{n-1}} \log (p_{z_0} P_{z_0 z_1} P_{z_1 z_2} \cdots P_{z_{n-2} z_{n-1}}) \\ &= - \sum_{[z_0, \dots, z_{n-1}]} p_{z_0} P_{z_0 z_1} P_{z_1 z_2} \cdots P_{z_{n-2} z_{n-1}} (\log p_{z_0} + \log P_{z_0 z_1} + \log P_{z_1 z_2} \\ & \qquad \qquad \qquad + \log P_{z_2 z_3} + \cdots + \log P_{z_{n-2} z_{n-1}}) \\ &= - \sum_{i=1}^k p_i \log p_i - (n-1) \sum_{i,j=1}^k p_i P_{ij} \log P_{ij} \end{aligned}$$

(where we use that  $pP = p$  and  $P$  is stochastic).

Therefore we see that

$$h_\mu(T) = \lim_{n \rightarrow +\infty} \frac{1}{n} H(\bigvee_{k=0}^{n-1} T^{-k} \alpha) = - \sum_{i,j=1}^k p_i P_{ij} \log P_{ij}.$$

In the special case that  $X = X_k = \prod_{-\infty}^{+\infty} \{1, \dots, k\}$  we can define a “Bernoulli measure” from a probability vector  $(p_1, \dots, p_k)$  (i.e.  $(p_1 + \dots + p_k = 1)$ ) by

$$\mu([z_0, \dots, z_{n-1}]) = p_{z_0} p_{z_1} p_{z_2} \cdots p_{z_{n-1}}.$$

The measure theoretic entropy in this case is  $h_\mu(T) = \sum_{i=1}^k p_i \log p_i$ . For example, where  $X = X_2$  and  $p = (\frac{1}{2}, \frac{1}{2})$  we have that  $h_\mu(T) = \log 2$ . Where  $X = X_3$  and  $p = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$  we have that  $h_\mu(T) = \log 3$ , etc.

#### 8.4 The increasing martingale theorem

We now begin to develop some of the machinery need to prove the results in section 8.3.

We know that if  $f \in L^1(X, \mathcal{B}, \mu)$  and  $\mathcal{A} \subset \mathcal{B}$  is a sub-sigma-algebra then we can associate the conditional expectation  $E(f|\mathcal{A}) \in L^1(X, \mathcal{A}, \mu)$ . The increasing martingale theorem describes how  $E(f|\mathcal{A})$  depends on the sigma-algebra  $\mathcal{A}$ . This is crucial in understanding the corresponding behaviour of the information function and thus the measure theoretic entropy.

The following simple lemma is very useful.

LEMMA 8.9. *If  $(X, \mathcal{B}, \mu)$  is a probability space and if  $\mathcal{B}_1 \subset \mathcal{B}_2 \subset \dots \subset \mathcal{B}_N \subset \mathcal{B}$  are sigma-algebras and  $\lambda > 0$  then if we let*

$$E = \{x \in X : \max_{1 \leq n \leq N} E(f|\mathcal{B}_n)(x) > \lambda\}$$

*then we have the upper bound on its measure  $\mu(E) \leq \frac{1}{\lambda} \int |f| d\mu$  with  $f \in L^1(X, \mathcal{B}, \mu)$ .*

PROOF. Without loss of generality we can assume that  $f \geq 0$  (otherwise we replace  $f$  by  $\max\{f(x), 0\}$ ). We can partition  $E = E_1 \cup \dots \cup E_N$  where

$$E_n = \{x \in X : E(f|\mathcal{B}_n)(x) > \lambda, E(f|\mathcal{B}_i)(x) \leq \lambda, i = 1, 2, \dots, n-1\}$$

(and observe that  $E_n \in \mathcal{B}_n$ ); then  $E_i \cap E_j = \emptyset$  for  $i \neq j$ . We then write

$$\int_E f d\mu = \sum_{n=1}^N \int_{E_n} f d\mu = \sum_{n=1}^N \int_{E_n} E(f|\mathcal{B}_n) d\mu \geq \sum_{n=1}^N \lambda \mu(E_n) = \lambda \mu(E).$$

Thus  $\mu(E) \leq \frac{1}{\lambda} \int f d\mu = \frac{1}{\lambda} \int |f| d\mu$ . ■

REMARK. This is very similar to the Chebyshev inequality for  $f \in L^1(X, \mathcal{B}, \mu)$  and  $\lambda > 0$  which says that  $\mu\{x \in X : f(x) > \lambda\} \leq \frac{\int |f| d\mu}{\lambda}$ .

This brings us to the main result of this section.

**THEOREM 8.10 (INCREASING MARTINGALE THEOREM).** *Let  $f \in L^1(X, \mathcal{B}, \mu)$ . Assume that  $\mathcal{B}_1 \subset \mathcal{B}_2 \subset \dots \subset \mathcal{B}_n \subset \dots \subset \mathcal{B}$  is an increasing sequence of sigma-algebras and that the union  $\cup_{n=1}^{\infty} \mathcal{B}_n$  generates  $\mathcal{B}$  (written  $\mathcal{B}_n \rightarrow \mathcal{B}$ ). Then  $E(f|\mathcal{B}_n) \rightarrow f$  in  $L^1(X, \mathcal{B}, \mu)$  and  $E(f|\mathcal{B}_n)(x) \rightarrow f(x)$ , almost everywhere.*

**PROOF.** The theorem is clearly true on the subspace  $\cup_{k=1}^{\infty} L^1(X, \mathcal{B}_k, \mu)$  since if  $g \in L^1(X, \mathcal{B}_k, \mu)$  then  $E(g|\mathcal{B}_n) = g$  for  $n \geq k$ . Moreover, this subspace is dense in  $L^1(X, \mathcal{B}, \mu)$  in the  $L^1$  norm.

Given an arbitrary  $f \in L^1(X, \mathcal{B}, \mu)$  we can choose  $\epsilon > 0$  and  $g \in L^1(X, \mathcal{B}_k, \mu)$ , say, with  $\int |f - g|d\mu < \epsilon$ . We then see that for any  $n \geq k$  we have that

$$\begin{aligned} & \int |E(f|\mathcal{B}_n) - f|d\mu \\ & \leq \int |E(f|\mathcal{B}_n) - E(g|\mathcal{B}_n)|d\mu + \int |E(g|\mathcal{B}_n) - g|d\mu + \int |g - f|d\mu \\ & \leq 2 \int |g - f|d\mu \end{aligned}$$

(where  $\int |E(g|\mathcal{B}_n) - g|d\mu = 0$  since  $E(g|\mathcal{B}_n) = g$  and we use that  $E(\cdot|\mathcal{B}_n)$  is a contraction on  $L^1(X, \mathcal{B}, \mu)$ ). In particular,  $\limsup_{n \rightarrow +\infty} \int |E(f|\mathcal{B}_n) - f|d\mu \leq 2\epsilon$ . Since  $\epsilon > 0$  is arbitrary, we see that we have  $L^1$  convergence.

To show that we also have almost everywhere convergence, we argue as follows:

$$\begin{aligned} & \mu\{x \in X : \limsup_{n \rightarrow +\infty} |E(f|\mathcal{B}_n)(x) - f(x)| > \epsilon^{1/2}\} \\ & \leq \mu\{x \in X : \limsup_{n \rightarrow +\infty} (|E((f - g)|\mathcal{B}_n)(x) - (f - g)(x)| \\ & \quad + |E(g|\mathcal{B}_n)(x) - g(x)|) > \epsilon^{1/2}\} \\ & \leq \mu\{x \in X : \limsup_{n \rightarrow +\infty} |E((f - g)|\mathcal{B}_n)(x)| + |(f - g)(x)| > \epsilon^{1/2}\} \\ & \leq \mu\{x \in X : \limsup_{n \rightarrow +\infty} |E((f - g)|\mathcal{B}_n)(x)| > \frac{1}{2}\epsilon^{1/2}\} \\ & \quad + \mu\{x \in X : |(f - g)(x)| > \frac{1}{2}\epsilon^{1/2}\} \\ & \leq 2 \left( \frac{1}{\frac{1}{2}\epsilon^{1/2}} \right) \int |f - g|d\mu \leq 2 \left( \frac{1}{\frac{1}{2}\epsilon^{1/2}} \right) \epsilon \leq 4\epsilon^{1/2} \end{aligned}$$

(where we have used Lemma 8.9 and the Chebyshev inequality). Since  $\epsilon > 0$  is arbitrary this shows almost everywhere convergence. ■

**REMARK.** There is a corresponding “decreasing martingale theorem”, but we shall not need it.

### 8.5 Entropy and sigma-algebras

We want to apply the increasing martingale theorem to the information functions. First we need a simple technical lemma.

LEMMA 8.11. *If  $\alpha$  is a partition with  $H(\alpha) < +\infty$  and we have sub-sigma-algebras  $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \dots \subset \mathcal{B}$  then*

$$\int \left( \sup_{n \geq 1} I(\alpha | \mathcal{A}_n) \right) d\mu \leq H(\alpha) + 1$$

(and, in particular,  $f(x) = \sup_{n \geq 1} I(\alpha | \mathcal{A}_n)(x) \in L^1(X, \mathcal{B}, \mu)$ ).

PROOF. We can write

$$\int f(x) d\mu(x) = \int_0^\infty F(t) dt \quad (8.4)$$

where  $F(t) = \mu \{x \in X : f(x) > t\}$ , provided the right hand side of (8.1) is finite.

We can write

$$\begin{aligned} F(t) &= \mu \{x \in X : \sup_{n \geq 1} I(\alpha | \mathcal{A}_n)(x) > t\} \\ &= \mu \left\{ x \in X : \sup_{n \geq 1} \left( - \sum_{A \in \alpha} \chi_A(x) \log \mu(A | \mathcal{A}_n)(x) \right) > t \right\} \\ &= \sum_{A \in \alpha} \mu \left( A \cap \{x \in X : \sup_{n \geq 1} (-\log \mu(A | \mathcal{A}_n)(x)) > t\} \right) \end{aligned}$$

(since the sets  $A$  are disjoint). However, we can simplify this by writing

$$\begin{aligned} &\{x \in X : \sup_{n \geq 1} (-\log \mu(A | \mathcal{A}_n)(x)) > t\} \\ &= \{x \in X : \inf_{n \geq 1} (\log \mu(A | \mathcal{A}_n)(x)) < -t\} \\ &= \{x \in X : \inf_{n \geq 1} (\mu(A | \mathcal{A}_n)(x)) < e^{-t}\} = \cup_{n \geq 1} A_n \end{aligned}$$

where  $A_n = \{x \in X : \mu(A | \mathcal{A}_n)(x) < e^{-t} \text{ and } \mu(A | \mathcal{A}_i)(x) \geq e^{-t}, \text{ for } i = 1, \dots, n-1\}$  are disjoint sets. If we write

$$F(t) = \sum_{A \in \alpha} \mu(A \cap (\cup_{n \geq 1} A_n)) = \sum_{A \in \alpha} \sum_{n \geq 1} \mu(A \cap A_n)$$

then we can use the estimates

$$\mu(A \cap A_n) = \int_{A_n} \chi_A d\mu = \int_{A_n} E(\chi_A | \mathcal{A}_n) d\mu \leq \int_{A_n} e^{-t} d\mu = e^{-t} \mu(A_n).$$



We now have two possible upper bounds on the same summation:

$$\sum_{n \geq 1} \mu(A \cap A_n) \leq \sum_n e^{-t} \mu(A_n) = e^{-t} \quad \text{and} \quad \sum_{n \geq 1} \mu(A \cap A_n) \leq \mu(A).$$

Therefore  $F(t) \leq \sum_{A \in \alpha} \min\{e^{-t}, \mu(A)\}$ . Finally, we can use this bound to estimate

$$\begin{aligned} \int_0^\infty F(t) dt &\leq \int_0^\infty \left( \sum_{A \in \alpha} \min\{e^{-t}, \mu(A)\} \right) dt \\ &= - \sum_{A \in \alpha} \left( \mu(A) \log \mu(A) - \int_{-\log \mu(A)}^\infty e^{-t} dt \right) \\ &= - \sum_{A \in \alpha} (\mu(A) \log \mu(A) - \mu(A)) \\ &= H(\alpha) + 1. \end{aligned}$$

■

We are now in a position to prove the following crucial result.

**THEOREM 8.12.** *If  $\alpha$  is a partition with  $H(\alpha) < +\infty$  and  $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \dots \rightarrow \mathcal{B}$  is an increasing sequence of sub-sigma-algebra then  $I(\alpha|\mathcal{A}_n)(x) \rightarrow I(\alpha|\mathcal{B})(x)$  almost everywhere and in  $L^1$ . Thus  $H(\alpha|\mathcal{A}_n) \rightarrow H(\alpha|\mathcal{B})$  as  $n \rightarrow +\infty$ .*

**PROOF.** By Theorem 8.10  $\mu(A|\mathcal{A}_n) \rightarrow \mu(A|\mathcal{B})$  almost everywhere, for any  $A \in \alpha$ . This implies that  $I(\alpha|\mathcal{A}_n)(x) \rightarrow I(\alpha|\mathcal{B})(x)$  almost everywhere.

By Lemma 8.11 we have that  $I(\alpha|\mathcal{A}_n)$  are dominated by the integrable function  $\sup_{n \geq 1} I(\alpha|\mathcal{A}_n)(x)$ . Thus by Lebesgue's dominated convergence theorem we have that  $I(\alpha|\mathcal{A}_n)(x) \rightarrow I(\alpha|\mathcal{B})(x)$  in  $L^1$  (i.e.  $\int |I(\alpha|\mathcal{A}_n)(x) - I(\alpha|\mathcal{B})(x)| d\mu(x) \rightarrow 0$  as  $n \rightarrow +\infty$ ).

Integrating shows the corresponding result for measure theoretic entropy i.e.  $H(\alpha|\mathcal{A}_n) \rightarrow H(\alpha|\mathcal{B})$  as  $n \rightarrow +\infty$ .

■

Using Theorem 8.12 we can extend the basic identities (Lemma 8.2 and Lemma 8.4) to arbitrary sub-sigma-algebras  $\mathcal{C} \subset \mathcal{B}$ , i.e.

(1) for the information functions

$$I(\alpha \vee \beta|\mathcal{C}) = I(\alpha|\hat{\beta} \vee \mathcal{C}) + I(\beta|\mathcal{C})$$

(and, in particular,  $I(\alpha|\mathcal{C}) \geq I(\beta|\mathcal{C})$  if  $\alpha > \beta$ ),

(2) for the measure theoretic entropy

$$H_\mu(\alpha \vee \beta|\mathcal{C}) = H_\mu(\alpha|\hat{\beta} \vee \mathcal{C}) + h_\mu(\beta|\mathcal{C})$$

(and, in particular,  $H_\mu(\alpha|\mathcal{C}) \geq H_\mu(\beta|\mathcal{C})$  if  $\alpha > \beta$ ).

This only requires that we choose partitions  $\gamma$  such that  $\hat{\gamma} \rightarrow \mathcal{C}$  and apply the theorem (and this is a basic property of Lebesgue spaces).

### 8.6 Conditional entropy

We want to consider how changing the sigma-algebra (with the same partition) affects the conditional measure theoretic entropy. The following lemmas are useful.

**LEMMA 8.13.** *Assume that  $f \in L^1(X, \mathcal{B}, \mu)$  and  $0 \leq f(x) \leq 1$  a.e. and that  $\mathcal{A} \subset \mathcal{B}$  is a sub-sigma-algebra. Let  $\psi : [0, 1] \rightarrow \mathbb{R}$  be a concave function (i.e.  $\psi(\alpha x + (1-\alpha)y) \geq \alpha\psi(x) + (1-\alpha)\psi(y)$ ); then  $\psi(E(f|\mathcal{A})) \geq E(\psi(f)|\mathcal{A})$ .*

**PROOF.** First consider the case of simple functions  $f(x) = \sum_{i=1}^n b_i \chi_{B_i}$ , where  $\{B_1, \dots, B_n\}$  is a partition for  $X$ .

By linearity of  $E(\cdot|\mathcal{A})$ ,

$$E(f|\mathcal{A})(x) = \sum_{i=1}^n b_i E(\chi_{B_i}|\mathcal{A})(x) = \sum_{i=1}^n b_i \mu(B_i|\mathcal{A})(x)$$

and observe that  $\sum_{i=1}^n \mu(B_i|\mathcal{A})(x) = 1$ . We can compute

$$\psi(E(f|\mathcal{A})) \geq \sum_{i=1}^n \psi(b_i) \mu(B_i|\mathcal{A}) = E\left(\sum_{i=1}^n \psi(b_i) \chi_{B_i}|\mathcal{A}\right) = E(\psi(f)|\mathcal{A}). \quad (8.5)$$

For an arbitrary function  $f \in L^1(X, \mathcal{B}, \mu)$  we can choose a monotonically increasing sequence of step functions  $f_k$  increasing to  $f$  a.e. Since  $E(\cdot|\mathcal{A})$  takes positive functions to positive functions we see that  $E(f_k|\mathcal{A}) \rightarrow E(f|\mathcal{A})$ . We can now take limits in (8.5) to get that  $\psi(E(f|\mathcal{A})) \geq E(\psi(f)|\mathcal{A})$ , as required. ■

The following is a simple application of this result.

**PROPOSITION 8.14.** *If  $\beta$  is a partition and  $\mathcal{A}_2 \subset \mathcal{A}_1 \subset \mathcal{B}$  are sub-sigma-algebras then  $H(\beta|\mathcal{A}_1) \leq H(\beta|\mathcal{A}_2)$ .*

(The corresponding result for information functions may not be true).

**PROOF.** For each  $B \in \beta$  we fix the choice  $f = \mu(B|\mathcal{A}_1)$ . We then fix  $\psi(t) = -t \log(t)$  for  $0 < t \leq 1$  (and  $\psi(0) = 0$ ). We then have that  $\psi(f) = -\mu(B|\mathcal{A}_1) \log(\mu(B|\mathcal{A}_1))$  and the lemma (Jenson's inequality) gives that

$$-E(\mu(B|\mathcal{A}_1) \log(\mu(B|\mathcal{A}_1))|\mathcal{A}_2) \leq -\mu(B|\mathcal{A}_2) \log \mu(B|\mathcal{A}_2).$$

Integrating both sides with respect to  $\mu$  (and summing over  $B \in \beta$ ) gives that  $H(\beta|\mathcal{A}_1) \leq H(\beta|\mathcal{A}_2)$ . ■

### 8.7 Proofs of Lemma 8.7 and Lemma 8.8

We can now use the results from the preceding sections to supply the omitted proofs of Lemmas 8.7 and 8.8.

PROOF OF LEMMA 8.7. We know that  $h(T, \beta_n) \leq h(T, \beta_m) + H(\beta_n | \hat{\beta}_m)$  for  $n, m \geq 1$ . Moreover, if  $m \geq n$  then  $\beta_n \subset \beta_m$  and so  $H(\beta_n | \beta_m) = 0$ . Thus  $h(T, \beta_n)$  is monotonically decreasing and so converges.

For any partition  $\alpha$  with  $H(\alpha) < +\infty$  we can start with the inequality  $h(T, \alpha) \leq h_\mu(T, \beta_n) + H(\alpha | \beta_n)$  (proved in a previous lemma). By a corollary to the increasing martingale theorem we know that  $H(\alpha | \beta_n) \rightarrow H(\alpha | \mathcal{B}) = 0$  (since  $\alpha \subset \mathcal{B}$ ). We conclude that

$$h(T, \alpha) \leq \limsup_{n \rightarrow +\infty} h(T, \beta_n).$$

Taking the supremum over all such  $\alpha$  gives that

$$h_\mu(T) = \sup_{\alpha} h(T, \alpha) \leq \lim_{n \rightarrow +\infty} h(T, \beta_n).$$

Clearly,  $h_\mu(T) \geq h(T, \beta_n)$  for  $n \geq 1$ . Thus

$$h_\mu(T) \geq \sup_n h(T, \beta_n) \geq \lim_{n \rightarrow +\infty} h(T, \beta_n)$$

and the proof is complete. ■

Before moving on to the proof of Lemma 8.8, we recall the following useful fact.

A FACT ABOUT LEBESGUE SPACES. For Lebesgue spaces a necessary and sufficient condition for  $\beta_n \rightarrow \mathcal{B}$  is that there exists a set of zero measure  $N \subset X$  such that for  $x, y \in X - N$  (with  $x \neq y$ ) there exist  $n \geq 1$  and  $B \in \beta_n$  such that  $x \in B$  but  $y \notin B$ .

PROOF OF LEMMA 8.8. This is an application of Abramov's result where we take  $\beta_n = \vee_{i=-n}^n T^{-i} \alpha$  or  $\beta_n = \vee_{i=0}^n T^{-i} \alpha$ , as appropriate. We then have that

$$\begin{aligned} h(T, \beta_k) &= H(\beta_k | \vee_{i=1}^{\infty} T^{-i} \beta_k) \\ &= H(\vee_{i=1}^{k-1} T^{-i} \alpha | \vee_{i=1}^{\infty} T^{-i} \alpha) \\ &= h(T, \alpha) \end{aligned}$$

and we need only let  $k \rightarrow +\infty$ . ■

### 8.8 Isomorphism

Entropy is very important in the classification of measure preserving transformations. We begin with a definition.

DEFINITION. Let  $(X_i, \mathcal{B}_i, \mu_i)$ , for  $i = 1, 2$ , be probability spaces; then an *isomorphism* between measure preserving transformations  $T_1 : X_1 \rightarrow X_1$  and  $T_2 : X_2 \rightarrow X_2$  is a map  $\phi : X_1 \rightarrow X_2$  such that

- (1)  $\phi$  is a bijection (after removing sets of zero measure, if necessary),
- (2) both  $\phi$  and  $\phi^{-1}$  are measurable (i.e.  $\phi^{-1}\mathcal{B}_2 \subset \mathcal{B}_1$  and  $\phi\mathcal{B}_1 \subset \mathcal{B}_2$ ),
- (3)  $\mu_1(\phi^{-1}B) = \mu_2(B)$  for  $B \in \mathcal{B}_2$  (also  $\mu_2(\phi B) = \mu_1(B)$  for  $B \in \mathcal{B}_1$ ),
- (4)  $\phi \circ T_1 = T_2 \circ \phi$ .

THEOREM 8.15. *Entropy is an isomorphism invariant (i.e. if  $T_1 : X_1 \rightarrow X_1$  and  $T_2 : X_2 \rightarrow X_2$  are isomorphic then  $h_{\mu_1}(T_1) = h_{\mu_2}(T_2)$ ).*

PROOF. Let  $\alpha$  be a partition for  $X_2$ ; then clearly  $\phi^{-1}\alpha = \{\phi^{-1}(A) : A \in \alpha\}$  is a partition for  $X_1$ . Moreover, the properties for  $\phi$  imply that

$$\frac{1}{n}H(\bigvee_{i=0}^{n-1}T_2^{-i}\alpha) = \frac{1}{n}H(\bigvee_{i=0}^{n-1}T_1^{-i}\phi^{-1}\alpha).$$

Letting  $n \rightarrow +\infty$  gives that  $h_{\mu_1}(T_1, \phi^{-1}\alpha) = h_{\mu_2}(T_2, \alpha)$ . Taking the supremum over all partitions  $\alpha$  (with  $H(\alpha) < +\infty$ ) gives the result. ■

EXAMPLE (BERNOULLI SHIFTS). Using the formula for the entropy of a Bernoulli measure from section 8.3 we see that the shifts

$$\begin{cases} \sigma : \prod_{n \in \mathbb{Z}} \{1, 2\} \rightarrow \prod_{n \in \mathbb{Z}} \{1, 2\} \text{ with probability vector } (\frac{1}{2}, \frac{1}{2}), \\ \sigma : \prod_{n \in \mathbb{Z}} \{1, 2\} \rightarrow \prod_{n \in \mathbb{Z}} \{1, 2, 3\} \text{ with probability vector } (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) \end{cases}$$

have entropies  $\log 2$  and  $\log 3$ , respectively. Therefore they are *not* isomorphic.

REMARK. Let us consider a slightly different situation where we drop the assumption that  $\phi$  is a bijection. That is, if we consider  $(X_i, \mathcal{B}_i, \mu_i)$ , for  $i = 1, 2$ , to be probability spaces then a *factor map* between  $T_1 : X_1 \rightarrow X_1$  and  $T_2 : X_2 \rightarrow X_2$  satisfies

- (1)  $\phi(X_1)$  is equal to  $X_2$  (after removing a set of zero  $\mu_2$ -measure, if necessary),
- (2)  $\phi$  is measurable,
- (3)  $\mu_1(\phi^{-1}B) = \mu_2(B)$  for  $B \in \mathcal{B}_2$ ,
- (4)  $\phi \circ T_1 = T_2 \circ \phi$ .

In this case it is easy to see that  $h_{\mu}(T_1) \geq h_{\mu}(T_2)$ .

We usually say that  $T_1$  is an *extension* of  $T_2$  or that  $T_2$  is an *factor* of  $T_1$ .

### 8.9 Comments and references

Our development of entropy has followed the lines of Parry's treatment in [2]. It is possible to reduce some of the analysis if we accept working only with countable sigma-algebras [5].

References for some more advanced topics we have omitted include [3, §7.5] (Krieger's generator theorem), [3, §7.6] (Ornstein's isomorphism theorem), and [1, §10.7] (Keane-Smorodinsky finitary Isomorphism Theorem).

#### References

1. I. Cornfold, S. Fomin and Y. Sinai, *Ergodic Theory*, Springer, Berlin, 1982.
2. W. Parry, *Topics in Ergodic Theory*, C.U.P., Cambridge, 1981.
3. D. Rudolph, *Fundamentals of Measurable Dynamics*, O.U.P., Oxford, 1990.
4. Y. Sinai, *Topics in Ergodic Theory*, P.U.P., Princeton N.J., 1994.
5. P. Walters, *An Introduction to Ergodic Theory*, Springer, Berlin, 1982.

## ERGODIC MEASURES

In this chapter we shall consider the stronger property of ergodicity for an invariant probability measure  $\mu$ . This property is more appropriate (amongst other things) for understanding the “long term” average behaviour of a transformation.

## 9.1 Definitions and characterization of ergodic measures

DEFINITION. Given a probability space  $(X, \mathcal{B}, \mu)$ , a transformation  $T : X \rightarrow X$  is called *ergodic* if for every set  $B \in \mathcal{B}$  with  $T^{-1}B = B$  we have that either  $\mu(B) = 0$  or  $\mu(B) = 1$ .

Alternatively we say that  $\mu$  is  $T$ -ergodic.

The following lemma gives a simple characterization in terms of functions.

LEMMA 9.1.  *$T$  is ergodic with respect to  $\mu$  iff whenever  $f \in L^1(X, \mathcal{B}, \mu)$  satisfies  $f = f \circ T$  then  $f$  is a constant function.*

PROOF. This is an easy observation using indicator functions. ■

## 9.2 Poincaré recurrence and Kac's theorem

We begin with one of the most fundamental results in ergodic theory.

THEOREM 9.2 (POINCARÉ RECURRENCE THEOREM). *Let  $T : X \rightarrow X$  be a measurable transformation on a probability space  $(X, \mathcal{B}, \mu)$ . Let  $A \in \mathcal{B}$  have  $\mu(A) > 0$ ; then for almost points  $x \in A$  the orbit  $\{T^n x\}_{n \geq 0}$  returns to  $A$  infinitely often.*

PROOF. Let  $F = \{x \in A : T^n x \notin A, \forall n \geq 1\}$ , then it suffices to show that  $\mu(F) = 0$ .

Towards this end, we first observe that  $T^{-m}F \cap T^{-n}F = \emptyset$  when  $n > m$ , say. If this were not the case and  $w \in T^{-m}F \cap T^{-n}F$  then  $T^m w \in F$  and  $T^{n-m}(T^m w) \in F \subset A$ , which contradicts the definition of  $F$ .

Thus since the sets  $\{T^{-n}F\}_{n \geq 0}$  are disjoint we see that

$$\sum_{n=0}^{\infty} \mu(T^{-n}F) = \mu(\cup_{n=0}^{\infty} T^{-n}F) \leq \mu(X) = 1$$

and then because  $\mu$  is  $T$ -invariant  $\mu(F) = \mu(T^{-1}F) = \dots = \mu(T^{-n}F) = \dots$  so we can only have that  $\mu(F) = 0$ . ■

DEFINITION. Let  $n_A : A \rightarrow \mathbb{Z}^+ \cup \{+\infty\}$  be the first return time i.e.  $n_A(x) > 0$  is the smallest value for which  $T^{n_A(x)}x \in A$ .

By Theorem 9.2  $n_A(x)$  is finite almost everywhere. The next theorem shows that when  $\mu$  is an ergodic measure then the *average* return time to  $A$  can be calculated explicitly.

THEOREM 9.3 (KAC'S THEOREM). *Let  $T : X \rightarrow X$  be an ergodic transformation on a probability space  $(X, \mathcal{B}, \mu)$ . Let  $A \in \mathcal{B}$  have  $\mu(A) > 0$  then we define the return time function  $n_A : A \rightarrow \mathbb{Z}^+ \cup \{\infty\}$  (which is finite, almost everywhere). The average return time (with respect to the induced probability measure  $\mu_A$ ) is*

$$\int_A n_A(x) d\mu_A(x) = \frac{1}{\mu(A)}.$$

PROOF. By definition of  $\mu_A$  it is equivalent to show that  $\int_A n_A(x) d\mu(x) = 1$ . It is useful to define the following sets.

- (a) For each  $n \geq 1$  we define  $A_n = \{x \in A : n_A(x) = n\}$ , and write  $A = \cup_{n \geq 1} A_n$  (with  $A_i \cap A_j = \emptyset$  for  $i \neq j$ ). In particular,  $\sum_{n=1}^{\infty} \mu(A_n) = \mu(A)$ .
- (b) For  $n \geq 1$  we define  $B_n = \{x \in X : T^j x \notin A \text{ for } 1 \leq j \leq n-1, T^n x \in A\}$ . The sets  $B_n$  are disjoint (i.e.  $B_i \cap B_j = \emptyset$  for  $i \neq j$ ) and by ergodicity  $X = \cup_{n \geq 1} B_n$  (since  $\cup_{n \geq 1} B_n \supset \cup_{n \geq 1} T^{-n}A \supset A$ ) so that  $\sum_{n=1}^{\infty} \mu(B_n) = 1$ .

We can rewrite

$$\int_A n_A(x) d\mu(x) = \sum_{k=1}^{\infty} k \mu(A_k) = \sum_{k=1}^{\infty} \left( \sum_{n=k}^{\infty} \mu(A_n) \right);$$

then if we can show that  $\sum_{n=k}^{\infty} \mu(A_n) = \mu(B_k)$  this will complete the proof.

When  $k = 1$  we have from the definitions that  $B_1 = T^{-1}A$  and so  $\sum_{n=1}^{\infty} \mu(A_n) = \mu(A) = \mu(B_1)$ , as required. For  $k > 1$  we can proceed by induction. We can partition  $T^{-1}B_k = B_{k+1} \cup T^{-1}A_k$  (where  $T^{-1}A_k = T^{-1}B_k \cap T^{-1}A$  and  $B_{k+1} = T^{-1}B_k \cap (X - T^{-1}A)$ ). Thus  $\mu(T^{-1}B_k) = \mu(B_k) = \mu(B_{k+1}) + \mu(T^{-1}A_k) = \mu(B_{k+1}) + \mu(A_k)$  and using the inductive hypothesis we see that  $\mu(B_{k+1}) = \mu(B_k) - \mu(A_k) = \sum_{n=k+1}^{\infty} \mu(A_n)$ . This completes the inductive step and the proof. ■

### 9.3 Existence of ergodic measures

When  $T : X \rightarrow X$  is a continuous map on a compact metric space there is a very simple relationship between ergodic measures and invariant measures which we can now describe.

Let  $\mathcal{M}$  denote the set of invariant probability measures on  $X$ . There is a natural topology on this space called the *weak-star* topology, i.e. the weakest topology such that a sequence  $\mu_n \in \mathcal{M}$  converges to  $\mu \in \mathcal{M}$  iff  $\forall f \in C^0(X)$ ,  $\int f d\mu_n \rightarrow \int f d\mu$ .

The following properties of  $\mathcal{M}$  are well-known (and easily checked):

- (i)  $\mathcal{M}$  is convex (i.e. if  $\mu_1, \mu_2 \in \mathcal{M}$  and  $0 < \alpha < 1$ , then  $\alpha\mu_1 + (1-\alpha)\mu_2 \in \mathcal{M}$ );
- (ii) the set  $\mathcal{M}$  is compact (in the weak-star topology) [5, Theorem 6.10].

LEMMA 9.4. *The extremal points in the convex set  $\mathcal{M}$  are ergodic measures (i.e.  $\mu \in \mathcal{M}$  is ergodic if whenever  $\exists \mu_1, \mu_2 \in \mathcal{M}$  and  $0 < \alpha < 1$  with  $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$  then  $\mu_1 = \mu_2$ ).*

The converse is also true, but we shall not require it.

PROOF. If  $\mu$  is not ergodic then we can find  $B \in \mathcal{B}$  with  $T^{-1}B = B$  and  $0 < \mu(B) < 1$ . But for any set  $A \in \mathcal{B}$  we can write  $A = (A \cap B) \cup (A \cap (X - B))$  and thus

$$\begin{aligned} \mu(A) &= \mu((A \cap B) \cup (A \cap (X - B))) \\ &= \mu(B) \left( \frac{\mu(A \cap B)}{\mu(B)} \right) + \mu(X - B) \left( \frac{\mu(A \cap (X - B))}{\mu(X - B)} \right) \\ &= \alpha\mu_1(A) + (1 - \alpha)\mu_2(A) \end{aligned}$$

where  $\alpha = \mu(B)$  and  $\mu_1(A) = \frac{\mu(A \cap B)}{\mu(B)}$ ,  $\mu_2(A) = \frac{\mu(A \cap (X - B))}{\mu(X - B)}$ . This shows that  $\mu = \alpha\mu_1 + (1 - \alpha)\mu_2$ . ■

PROPOSITION 9.5 (EXISTENCE OF ERGODIC MEASURES). *Let  $X$  be a compact metric space and  $\mathcal{B}$  be the Borel sigma-algebra. Given any continuous map  $T : X \rightarrow X$  there exists at least one  $T$ -ergodic probability measure  $\mu$ .*

PROOF. Choose a dense set of functions  $f_k \in C^0(X)$ ,  $k \geq 0$ . Since the map  $\mu \rightarrow \int f_0 d\mu$  is continuous on  $\mathcal{M}$  there exists by (weak-star) compactness at least one  $\nu \in \mathcal{M}$  such that  $\int f_0 d\nu = \sup_{\mu \in \mathcal{M}} \{ \int f_0 d\mu \}$ . We let

$$\mathcal{M}_0 = \left\{ \nu \in \mathcal{M} : \int f_0 d\nu = \sup_{\mu \in \mathcal{M}} \left\{ \int f_0 d\mu \right\} \right\};$$

then clearly  $\mathcal{M}_0$  is non-empty and closed. Similarly, define

$$\mathcal{M}_1 = \left\{ \nu \in \mathcal{M}_0 : \int f_1 d\nu = \sup_{\mu \in \mathcal{M}_0} \left\{ \int f_1 d\mu \right\} \right\}$$



and the same reasoning shows that  $\mathcal{M}_1 \subset \mathcal{M}_0 \subset \mathcal{M}$  is non-empty and closed.

Proceeding inductively we define

$$\mathcal{M}_k = \left\{ \nu \in \mathcal{M}_{k-1} : \int f_k d\nu = \sup_{\mu \in \mathcal{M}_{k-1}} \left\{ \int f_k d\mu \right\} \right\}$$

and arrive at a nested sequence  $\mathcal{M} \supset \mathcal{M}_0 \supset \mathcal{M}_1 \supset \mathcal{M}_2 \supset \dots \supset \mathcal{M}_k \supset \dots$ . Since the sets are all closed in  $\mathcal{M}$  (and hence compact) we have that the intersection is non-empty. Assume  $\mu \in \bigcap_{k \in \mathbb{Z}^+} \mathcal{M}_k$ . We want to show that  $\mu$  is ergodic by showing that it is an extreme point in  $\mathcal{M}$ .

Assume that  $\mu$  can be written as an affine combination  $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$  (with  $0 < \alpha < 1$ ); then to show that  $\mu$  is ergodic we need to show that  $\mu_1 = \mu_2$ . Thus it suffices to show that for every  $f_k \in C^0(X)$  we have that  $\int f_k f d\mu_1 = \int f_k f d\mu_2$  (since the set  $f_k$  is dense).

We begin with  $k = 0$  and observe that by assumption  $\int f_0 d\mu = \alpha \int f_0 d\mu_1 + (1-\alpha) \int f_0 d\mu_2$ . Since  $\mu \in \mathcal{M}_0$  we see that  $\sup_{m \in \mathcal{M}} \{ \int f_0 dm \} = \int f_0 d\mu$  implies that  $\int f_0 d\mu_1 = \int f_0 d\mu_2 = \sup_{m \in \mathcal{M}} \{ \int f_0 dm \}$ . We thus conclude

- (1) the first identity  $\int f_0 d\mu_1 = \int f_0 d\mu_2$  is proved.
- (2)  $\mu_1, \mu_2 \in \mathcal{M}_0$ .

Continuing inductively, we establish that for arbitrary  $k \geq 0$  we have  $\int f_k d\mu_1 = \int f_k d\mu_2$  and  $\mu_1, \mu_2 \in \mathcal{M}_k$ . This completes the proof (i.e.  $\mu_1 = \mu_2$  and  $\mu$  is an extremal measure). ■

REMARK. The following facts are easy to check.

- (3) If  $\nu, \mu$  are distinct  $T$ -ergodic measures then  $\nu \perp \mu$ .
- (4) If  $\mu$  is ergodic then it is an extremal measure in  $\mathcal{M}$ . (The converse to Lemma 9.4.)

Since  $\mathcal{M}$  is a compact convex metric space there is a general theorem of Choquet that says *every invariant measure  $\mu \in \mathcal{M}$  can be written as a convex combination of extremal measures in  $\mathcal{M}$* . More precisely, we can find a measure  $\rho = \rho_\mu$  on the space  $\mathcal{M}$  (with respect to the Borel sigma-algebra associated to the weak-star topology) such that

- (1) for any function  $f \in C^0(X)$  we have

$$\int f d\mu = \int_{\mathcal{M}} \left( \int f d\nu \right) d\rho(\nu).$$

- (2)  $\rho(\{\nu : \nu \text{ is extremal}\}) = 1$ .

#### 9.4 Some basic constructions in ergodic theory

In this final section of chapter 9 we shall describe two basic constructions in ergodic theory.

**9.4.1. Skew products.** Let  $T : X \rightarrow X$  be a measure preserving transformation of a probability space  $(X, \mathcal{B}, \mu)$ . Let  $(G, \mathcal{B})$  be a compact Lie group with the Borel sigma-algebra  $\mathcal{B}$ . We can consider the product space  $X \times G$  with the product sigma-algebra  $\mathcal{A}$ .

DEFINITION. Given a measure preserving transformation of  $T : X \rightarrow X$  and a measurable map  $\phi : X \rightarrow G$  we define a *skew product* to be the transformation  $S : X \times G \rightarrow X \times G$  defined by  $S(x, g) = (Tx, \phi(x)g)$ . Given any  $T$ -invariant probability measure  $\mu$  we can associate the  $S$ -invariant measure  $\nu$  defined by  $d\nu = d\mu \times dt$ .

A simple example is the following.

EXAMPLE. Let  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  be given by  $T(x) = x + \alpha \pmod{1}$  for some  $\alpha \in \mathbb{R}$ . Let  $G = \mathbb{R}/\mathbb{Z}$  and we define  $\phi : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  by  $\phi(x) = x \pmod{1}$  (i.e. the identity map). The associated skew product is then the map  $S : \mathbb{R}^2/\mathbb{Z}^2 \rightarrow \mathbb{R}^2/\mathbb{Z}^2$  given by  $S(x, y) = (x + \alpha, x + y) \pmod{1}$ .

**9.4.2. Induced transformations and Rohlin towers.** Assume that  $T : X \rightarrow X$  is a measurable transformation on a measurable space  $(X, \mathcal{B})$ . Assume that  $A \subset X$  with  $A \in \mathcal{B}$ .

DEFINITION. The transformation  $T_A : A \rightarrow A$  defined by  $T_A(x) = T^{n_A(x)}x$  is called the *induced transformation* on  $A$ . We denote by  $\mathcal{B}_A = \{B \cap A : B \in \mathcal{B}\}$  the restriction of the sigma-algebra  $\mathcal{B}$  to  $A$ .

If  $\mu$  is a  $T$ -invariant sigma-finite measure on  $(X, \mathcal{B})$  and  $0 < \mu(A) < \infty$  then we can define a  $T_A$ -invariant measure  $\mu_A$  on  $(A, \mathcal{B}_A)$  by  $\mu_A(B) = \frac{\mu(A \cap B)}{\mu(A)}$ .

EXAMPLE (CONTINUED FRACTION TRANSFORMATION). Consider the case where  $(X, \mathcal{B})$  is the positive half-line  $\mathbb{R}^+ = (0, +\infty)$  with the Borel sigma-algebra. We define a transformation  $T : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  by

- (1)  $Tx = x - 1$  if  $x \in [1, +\infty)$ , and
- (2)  $Tx = \frac{1}{x}$  if  $x \in (0, 1)$ .

We can consider the induced transformation  $T_A : A \rightarrow A$  on the interval  $A = (0, 1]$  defined by  $T_Ax = \frac{1}{x} - [\frac{1}{x}]$ .

The measure  $\mu_A$  defined by  $\mu_A(B) = \frac{1}{\log 2} \int_B \frac{1}{1+x} dx$  is  $T_A$ -invariant.

REMARK. We need not be too careful about the definition of  $T$  and  $T_A$  on a countable set of points since they have zero measure.

Consider an (ergodic) transformation  $T : X \rightarrow X$  on a probability space  $(X, \mathcal{B}, \mu)$  and let  $A \in \mathcal{B}$  have  $\mu(A) > 0$ .

DEFINITION. We can define a space

$$A^{n_A} = \{(x, k) \in A \times \mathbb{Z}^+ : 0 \leq k \leq n_A(x)\},$$

where we identify  $(x, n_A(x)) \sim (T^{n_A(x)}x, 0)$ , and introduce the product sigma-algebra  $\bar{\mathcal{B}}$  (i.e. the smallest sigma-algebra containing the products of sets in  $\mathcal{B}_A$  and  $\mathcal{B}_{\mathbb{Z}^+}$ ).

We define a probability measure on the space  $A^{n_A}$  by  $\nu = \frac{\mu \times dn}{\int n_A d\mu}$  (where  $dn$  corresponds to the usual counting measure on  $\mathbb{Z}^+$ ).

Finally, we define a transformation  $T_A^{n_A} : A^{n_A} \rightarrow A^{n_A}$  by

- (1)  $T_A^{n_A}(x, k) = (x, k + 1)$  if  $0 \leq k < n_A(x)$ , and
- (2)  $T_A^{n_A}(x, n_A(x)) = T_A^{n_A}(T_A x, 0) = (T_A x, 1)$ .

This construction is called the *Rohlin tower* over  $A$ .

(N.B. A Rohlin tower is the converse process to induced transformations. We reproduce the original transformation on  $X$  from the induced transformation on  $A$ .) The following lemma tells us the Rohlin tower is a good model of the original transformation.

LEMMA 9.6. *The map  $\phi : (A^{n_A}, \bar{\mathcal{B}}, \nu) \rightarrow (X, \mathcal{B}, \mu)$  defined by  $\phi(x, k) = T^k(x)$  is measurable and satisfies the following:*

- (1)  $\phi$  is a bijection (almost everywhere);
- (2)  $\forall B \in \mathcal{B}$  we have that  $\nu(\phi^{-1}B) = \mu(B)$ ; and
- (3)  $\phi T_A^{n_A} = T\phi$  (almost everywhere).

PROOF. The result follows almost immediately from the definitions. ■

REMARK. The map  $\phi$  is an *isomorphism* which implies that from the point of view of ergodic theory the transformations  $T$  and  $T_A^{n_A}$  are the same.

**9.4.3 Natural extensions.** Given a non-invertible map  $T : (X, \mathcal{B}, \mu) \rightarrow (X, \mathcal{B}, \mu)$  there is a natural way of associating to it an invertible transformation  $\hat{T} : (\hat{X}, \hat{\mathcal{B}}, \hat{\mu}) \rightarrow (\hat{X}, \hat{\mathcal{B}}, \hat{\mu})$  with similar dynamical properties.

We define

$$\hat{X} = \{(x_n)_{n \in \mathbb{Z}^+} \in \prod_{n \in \mathbb{Z}^+} X : T(x_n) = x_{n+1}, n \geq 0\}$$

and associate the sigma-algebra generated by the sets

$$B_m := \left\{ (x_n)_{n \in \mathbb{Z}^+} \in \hat{X} : x_m \in B \right\} \text{ for } B \in \mathcal{B} \text{ and } m \in \mathbb{Z}^+.$$

We next define a probability measure  $\hat{\mu}$  on  $\hat{\mathcal{B}}$  by  $\hat{\mu}(B_m) = \mu(B)$ . Finally, we define the (invertible) transformation  $\hat{T} : \hat{X} \rightarrow \hat{X}$  by

$$\hat{T}(x_0, x_1, x_2, \dots) = (Tx_0, x_0, x_1, x_2, \dots).$$

It is easy to see from the construction that  $\hat{T}$  is measurable and preserves the probability measure  $\hat{\mu}$ .

DEFINITION. We call  $\hat{T} : \hat{X} \rightarrow \hat{X}$  the *natural extension* of  $T$ .

There is a canonical map  $\pi : \hat{X} \rightarrow X$  defined by  $\pi((x_n)_{n \in \mathbb{Z}^+}) = x_0$ . The natural extension  $\hat{T}$  has the following properties:

- (i)  $\hat{T}$  is an extension of  $T$  in the sense that  $\pi \circ \hat{T} = T \circ \pi$ ; and
- (ii) if we denote by  $\hat{\mathcal{B}}^+ \subset \hat{\mathcal{B}}$  the sub-sigma-algebra generated by sets  $\{\pi^{-1}(B) : B \in \mathcal{B}\}$  then

$$\dots \subset \hat{T}^{-1}\hat{\mathcal{B}}^+ \subset \hat{\mathcal{B}}^+ \subset T\hat{\mathcal{B}}^+ \subset \dots \subset \bigcup_{n \in \mathbb{Z}^+} T^n \hat{\mathcal{B}}^+ = \mathcal{B}.$$

REMARK. In fact, any transformation satisfying (i) and (ii) will be isomorphic to the natural extension as we have defined it above [3].

EXAMPLE (SUBSHIFTS OF FINITE TYPE). Let  $\sigma : X_A^+ \rightarrow X_A^+$  be a (one-sided) subshift of finite type, defined by the  $k \times k$  matrix  $A$ . Relative to a Markov measure, say, its natural extension is the shift  $\sigma : X \rightarrow X$ .

## 9.5 Comments and references

More can be found on ergodic measures in [1], [2] and [5].

Important applications of ergodic theory beyond the scope of these notes are Mostow's rigidity theorem [4] and the Margulis super-rigidity theorem [6, §5.1].

The skew product example in subsection 9.4.1 was used by Furstenberg to give a simple proof of a result on diophantine approximation due to Hardy and Littlewood [2].

## References

1. P. Billingsley, *Ergodic Theory and Information*, Wiley, New York, 1965.
2. W. Parry, *Topics in Ergodic Theory*, C.U.P., Cambridge, 1981.
3. V. Rohlin, *Exact endomorphisms of lebesgue space*, Amer. Math. Soc. Transl. (2) **39** (1964), 1-36.
4. W. Thurston, *Topology and geometry of three manifolds*, unpublished notes, Princeton University N.J., Princeton, 1978.
5. P. Walters, *An Introduction to Ergodic Theory*, Springer, New York, 1989.
6. R. Zimmer, *Ergodic Theory and Semi-simple Groups*, Birkhäuser, Basel, 1984.

## ERGODIC THEOREMS

In this chapter we shall describe the ergodic theorems and some of their applications. These simple but elegant theorems are important in many other areas.

Let  $f$  be a function which is an observable for a physical quantity. One of the main themes in ergodic theory is to study the asymptotic behaviour of their time evolution  $\{f \circ T^k\}_{k \in \mathbb{Z}^+}$ . Under the ergodic hypothesis, their averages  $\frac{1}{N} \sum_{k=0}^{N-1} f \circ T^k$  converge to the space average  $\int f d\mu$ . This property also implies the well-known law of large numbers, which is a key concept in statistics (i.e. the distribution of the long term average converges to the Dirac measure supported on  $\int f d\mu$ ).

**10.1 The von Neumann ergodic theorem.** Let  $(X, \mathcal{B}, \mu)$  be a measure space with  $\mu$  a probability measure, and assume that  $T : X \rightarrow X$  preserves  $\mu$ . In general, ergodic theorems describe how the averages of functions  $f : X \rightarrow \mathbb{R}$  along the orbits  $\{T^n x\}_{n=0}^{\infty}$  behave for (almost all) points  $x \in X$ . The simplest theorem of this type is the following “mean” ergodic theorem.

**THEOREM 10.1 (VON NEUMANN MEAN ERGODIC THEOREM).** *Let  $(X, \mathcal{B}, \mu)$  be a probability space, and assume that  $T : X \rightarrow X$  preserves  $\mu$ . Then for  $f, g \in L^2(X, \mathcal{B}, \mu)$  the averages*

$$\frac{1}{N} \sum_{n=0}^{N-1} \int f(T^n x)g(x)d\mu(x)$$

converge as  $N \rightarrow +\infty$ . If  $T$  is ergodic then

$$\frac{1}{N} \sum_{n=0}^{N-1} \int f(T^n x)g(x)d\mu(x) \rightarrow \int f d\mu \int g d\mu$$

as  $N \rightarrow +\infty$ .

(Cf. [6, §2], [8, pp. 23-27].)

**PROOF.** Consider the subspace  $I = \{f \in L^2(X, \mathcal{B}, \mu) : fT = f\}$  of  $T$ -invariant functions. (If  $T$  is ergodic then  $I$  consists of constant functions.) Consider too the subspace

$$B = \{f \in L^2(X, \mathcal{B}, \mu) : \exists h \in L^2(X, \mathcal{B}, \mu) \text{ with } f = hT - h\}$$

of *coboundaries*.

On each of these spaces the averages are easier to study.

On the first space  $I$  we see that for  $f_1 \in I$  and any  $N \geq 1$

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) = f_1(x) \in L^2(X, \mathcal{B}, \mu).$$

On the second space  $B$  we see that if  $f_2 = h \circ T - h \in B$  then

$$\frac{1}{N} \sum_{n=0}^{N-1} f_2(T^n x) = \frac{1}{N} (h(T^N x) - h(x)).$$

By the Cauchy-Schwarz inequality we have that

$$\begin{aligned} \frac{1}{N} \left| \int (h(T^N x) - h(x)) g d\mu(x) \right| &\leq \frac{1}{N} \left( \left| \int h \circ T^N \cdot g d\mu \right| + \left| \int h \cdot g d\mu \right| \right) \\ &\leq \frac{1}{N} (\|h \circ T^N\|_2 \|g\|_2 + \|h\|_2 \|g\|_2) \\ &= \frac{2}{N} \|h\|_2 \|g\|_2 \end{aligned}$$

which converges to zero as  $N \rightarrow +\infty$ .

If we can write  $f = f_1 + f_2$  with  $f_1 \in I$  and  $f_2 \in B$  then

$$\begin{aligned} &\frac{1}{N} \sum_{n=0}^{N-1} \int f(T^n x) g(x) d\mu(x) \\ &= \frac{1}{N} \sum_{n=0}^{N-1} \int f_1(T^n x) g(x) d\mu(x) + \frac{1}{N} \sum_{n=0}^{N-1} \int f_2(T^n x) g(x) d\mu(x) \\ &= \int f_1(x) g(x) d\mu(x) + \frac{1}{N} \sum_{n=0}^{N-1} \int f_2(T^n x) g(x) d\mu(x) \end{aligned}$$

and therefore

$$\frac{1}{N} \sum_{n=0}^{N-1} \int f(T^n x) g(x) d\mu(x) \rightarrow \int f_1(x) g(x) d\mu(x)$$

as  $N \rightarrow +\infty$ . If  $T$  is ergodic, then

$$\frac{1}{N} \sum_{n=0}^{N-1} \int f(T^n x) g(x) d\mu(x) \rightarrow \int f(x) d\mu(x) \int g(x) d\mu(x)$$

as  $N \rightarrow +\infty$ .

More generally, assume that for any  $\epsilon > 0$  we can find  $f_1 \in I$  and  $f_2 \in B$ ; so then  $\int |f - (f_1 + f_2)|^2 d\mu < \epsilon$  then by the Cauchy-Schwarz inequality

$$\begin{aligned} & \frac{1}{N} \sum_{n=0}^{N-1} \int (f - (f_1 + f_2))(T^n x) g(x) d\mu(x) \\ & \leq \frac{1}{N} \sum_{n=0}^{N-1} \| (f - (f_1 + f_2)) T^n \|_2 \|g\|_2 \\ & = \frac{1}{N} \sum_{n=0}^{N-1} \| (f - (f_1 + f_2)) \|_2 \|g\|_2 \\ & \leq \sqrt{\epsilon} \|g\|_2 \end{aligned}$$

(where to get from the second to third line we use the property that  $T$  preserves  $\mu$  to write  $\|h \circ T^n\|_2^2 = \int |h \circ T^n| d\mu = \int |h| \circ T^n d\mu = \int |h| d\mu = \|h\|_2^2$ ). Therefore the limit (points) of the sequence

$$\frac{1}{N} \sum_{n=0}^{N-1} \int f(T^n x) g(x) d\mu(x)$$

lie in the interval  $(\int f d\mu \int g d\mu - \epsilon, \int f d\mu \int g d\mu + \epsilon)$ . Since  $\epsilon > 0$  can be chosen arbitrarily small we see that

$$\frac{1}{N} \sum_{n=0}^{N-1} \int f(T^n x) g(x) d\mu(x) \rightarrow \int f d\mu \int g d\mu$$

as  $N \rightarrow +\infty$ .

So to prove the theorem it suffices to show that  $L^2(X, \mathcal{B}, \mu) = I + \text{cl}(B)$ . To see this, assume that  $f \perp \text{cl}(B)$ , i.e.  $\int f \cdot f_2 d\mu = 0, \forall f_2 \in \text{cl}(B)$  or in particular  $\int f \cdot h \circ T d\mu = \int f \cdot h d\mu, \forall h \in L^2(X, \mathcal{B}, \mu)$ . Taking  $h = f$  gives us that

$$\int |f|^2 d\mu = \int f \circ T \cdot f d\mu. \quad (10.1)$$

We need to show that  $f \in I$  (i.e.  $f \circ T = f$ ) and to this end we can write

$$\begin{aligned} \int |f \circ T - f|^2 d\mu &= \int |f \circ T|^2 d\mu + \int |f|^2 d\mu - 2 \int f \circ T \cdot f d\mu \\ &= 2 \left( \int |f|^2 d\mu - \int f \circ T \cdot f d\mu \right) \\ &= 0 \end{aligned}$$

where we use (10.1) and that  $\int |f \circ T|^2 d\mu = \int |f|^2 d\mu$ . This shows that  $f \circ T = f$  a.e. (i.e.  $f \in I$ ).

This completes the proof. ■

### 10.2 The Birkhoff theorem (for ergodic measures)

Let  $(X, B, \mu)$  be a probability space, and assume that the transformation  $T : X \rightarrow X$  preserves  $\mu$ . The Birkhoff “individual” ergodic theorem gives a strong type of ergodic theorem in that it describes the average of functions along individual “typical” orbits.

We first prove the theorem under the additional assumption that  $\mu$  is ergodic, and then for arbitrary invariant measures.

**THEOREM 10.2 (BIRKHOFF’S THEOREM (ERGODIC VERSION)).** *Consider  $f \in L^1(X, B, \mu)$ . If the measure  $\mu$  is ergodic then for almost all  $x \in X$  we have that the averages*

$$\frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) \rightarrow \int f d\mu$$

as  $N \rightarrow +\infty$  (i.e.  $\mu\{x \in X : \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) \neq \int f d\mu\} = 0$ ).

(Cf. [6 §1.2], [8, p. 30].)

**PROOF.** Assume for convenience that  $\int f d\mu = 0$ . (If this is not the case then replace  $f$  by  $f - \int f d\mu$  in the two sides of the equality in the statement of the theorem.)

For any  $\epsilon > 0$  we can define

$$E_\epsilon(f) = \{x \in X : \limsup_{N \rightarrow +\infty} \frac{1}{N} \left| \sum_{n=0}^{N-1} f(T^n x) \right| \geq \epsilon\}.$$

To prove the theorem it suffices to show that for each  $\epsilon > 0$  we have that  $\mu(E_\epsilon) = 0$ . We do this via two lemmas.

We begin with the following estimate.

**SUBLEMMA 10.2.1.**  $\mu(E_{2\epsilon}(f)) \leq \frac{\int |f| d\mu}{\epsilon}$ .

**PROOF.** We first write  $f = f_+ - f_-$  where  $f_+, f_- \geq 0$  and  $\min(f_+, f_-) = 0$ .

We next define the sets

$$E_\epsilon^M(f_+) = \{x \in X : \exists 1 \leq N \leq M, \sum_{n=0}^{N-1} f_+(T^n x) \geq \epsilon N\}$$

and

$$E_\epsilon^M(f_-) = \{x \in X : \exists 1 \leq N \leq M, \sum_{n=0}^{N-1} f_-(T^n x) \geq \epsilon N\}$$



for any  $M \geq 1$ .

Next observe that for any  $x \in X$  and any  $P \geq M$  we have the lower bounds

$$\sum_{n=0}^{P-1} f_+(T^n x) \geq \epsilon \sum_{j=0}^{P-M} \chi_{E_\epsilon^M(f_+)}(T^j x)$$

and

$$\sum_{n=0}^{P-1} f_-(T^n x) \geq \epsilon \sum_{j=0}^{P-M} \chi_{E_\epsilon^M(f_-)}(T^j x)$$

where we bound  $f_+(T^n x)$  or  $f_-(T^n x)$  from below by  $\epsilon$  or 0 as appropriate. Integrating each side of these identities gives that

$$\int \left( \sum_{n=0}^{P-1} f_+(T^n x) \right) d\mu(x) = P \int f_+ d\mu \geq (P-M)\epsilon \mu(E_\epsilon^M(f_+))$$

and

$$\int \left( \sum_{n=0}^{P-1} f_-(T^n x) \right) d\mu(x) = P \int f_- d\mu \geq (P-M)\epsilon \mu(E_\epsilon^M(f_-))$$

for all  $M \geq 1$ .

Letting  $P \rightarrow +\infty$  gives that

$$\int f_+ d\mu \geq \epsilon \mu(E_\epsilon^M(f_+)) \quad \text{and} \quad \int f_- d\mu \geq \epsilon \mu(E_\epsilon^M(f_-)). \quad (10.2)$$

We finally see that

$$\begin{aligned} \mu(E_{2\epsilon}^M(f)) &\leq \limsup_{M \rightarrow +\infty} \mu(E_\epsilon^M(f_+)) + \limsup_{M \rightarrow +\infty} \mu(E_\epsilon^M(f_-)) \\ &\leq \frac{\int f_+ d\mu}{\epsilon} + \frac{\int f_- d\mu}{\epsilon} \\ &= \frac{\int |f| d\mu}{\epsilon}. \end{aligned}$$

This completes the proof of Sublemma 10.2.1. ■

The inequality in Sublemma 10.2.1 is not effective unless we can control the size of  $\int |f| d\mu$ . The next sublemma allows us to do this (and this is also where the ergodicity of  $\mu$  is used).

**SUBLEMMA 10.2.2.** *Let  $\int f d\mu = 0$ . Given  $\delta > 0$  we can choose a function  $h \in L^\infty(X, \mathcal{B}, \mu)$  with  $\int |f - (hT - h)| d\mu < \delta$ .*

**PROOF.** We need to show that the subspace

$$E := \{hT - h : h \in L^\infty(X, \mathcal{B}, \mu)\}$$

is dense in the subspace

$$B_0 = \{f \in L^1(X, \mathcal{B}, \mu) : \int f d\mu = 0\}.$$

By the Hahn-Banach theorem it suffices to show that all linear functionals which vanish on  $E$  also vanish on  $B_0$ .

As is well-known, for each linear functional  $\alpha$  on  $L^1(X, \mathcal{B}, \mu)$  there exists  $k \in L^\infty(X, \mathcal{B}, \mu)$  such that  $\alpha(f) = \int f k d\mu$ . Given a linear functional, and thus  $k \in L^\infty(X, \mathcal{B}, \mu)$ , the hypothesis that the linear functional vanishes on  $E$  implies that  $\int (hT - h) k d\mu = 0, \forall h \in L^\infty(X, \mathcal{B}, \mu)$ .

With the specific choice  $h = k$  we get that  $\int kT \cdot k d\mu = \int k^2 d\mu$  which implies that

$$\begin{aligned} \int (k \circ T - k)^2 d\mu &= \int (k \circ T)^2 d\mu + \int (k)^2 d\mu - 2 \int k \circ T \cdot k d\mu \\ &= 2 \left( \int (k)^2 d\mu - \int k \circ T \cdot k d\mu \right) \\ &= 0. \end{aligned}$$

In particular,  $k \circ T = k$ . By the ergodicity assumption this implies that  $k$  is constant. So if  $f \in B_0$  then  $\int f d\mu = 0$ , which in turn implies that  $\int f k d\mu = 0$  and so  $f$  is in the kernel of the functional, i.e. the functional vanishes on  $B_0$ . ■

We can assume without loss of generality that  $f \in B_0$  (otherwise we replace  $f$  by  $f - \int f d\mu$ ). For an arbitrary  $\delta > 0$  we can use sublemma 10.2.2 to choose  $h \in L^\infty(X, \mathcal{B}, \mu)$  with  $\int |f - (hT - h)| d\mu < \delta$ .

From the definitions we see that for arbitrary  $\epsilon > 0$  we have  $E_\epsilon(f) \subset E_{\frac{\epsilon}{2}}(f - (hT - h)) \cup E_{\frac{\epsilon}{2}}(hT - h)$  and therefore

$$\mu(E_\epsilon(f)) \leq \mu(E_{\frac{\epsilon}{2}}(f - (hT - h))) + \mu(E_{\frac{\epsilon}{2}}(hT - h)). \quad (10.3)$$

However,

(1) since  $\forall x \in X$

$$\left| \frac{1}{N} \sum_{n=0}^{N-1} (hT - h)(T^n x) \right| = \frac{1}{N} |h(T^N x) - h(x)| \leq \frac{2 \|h\|_\infty}{N}$$

we see that  $E_{\frac{\epsilon}{2}}((hT - h)) = \emptyset$  and so  $\mu(E_{\frac{\epsilon}{2}}((hT - h))) = 0$ .

(2) by sublemma 10.2.1 we have that

$$\mu(E_{\frac{\epsilon}{2}}(f - (hT - h))) \leq \frac{\int |f - (hT - h)| d\mu}{\epsilon/4} \leq \frac{4\delta}{\epsilon}.$$

Since  $\delta > 0$  was arbitrary we deduce from part (2) that  $\mu(E_{\frac{\epsilon}{2}}(f - (hT - h))) = 0$  and so by (10.3) and part (1) that  $\mu(E_{\epsilon}(f)) = 0, \forall \epsilon > 0$ . This suffices to prove the result. ■

The following corollary shows that the Birkhoff Theorem gives a quantitative version of the Poincaré recurrence theorem.

**COROLLARY 10.2.1.** *Let  $(X, B, \mu)$  be a probability space, and assume that the transformation  $T : X \rightarrow X$  preserves  $\mu$  and ergodic.. The proportion of time spent by almost all points in a subset  $B \in \mathcal{B}$  is given by its measure  $\mu(B)$ , i.e.*

$$\lim_{N \rightarrow +\infty} \frac{1}{N} \text{Card}\{0 \leq n \leq N - 1 : T^n x \in B\} = \mu(B)$$

for almost all points  $x \in X$ .

**PROOF.** This follows from applying the ergodic theorem to the indicator function on the set  $A$ . ■

**EXAMPLE (LACK OF CONVERGENCE ON A SET OF ZERO MEASURE).** Consider the map  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  defined by  $T(x) = 2x \pmod{1}$  and the usual Haar-Lebesgue measure  $\mu$ . Consider any continuous function  $f : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}$  such that  $f(0) \neq \int f d\mu$ . Clearly,

$$\frac{1}{N} \sum_{n=0}^{N-1} f(T^n 0) = f(0)$$

since  $T^n 0 = 0$ , for all  $n \geq 0$ . But since  $f(0) \neq \int f d\mu$  we see that at the point 0 the sequence does *not* converge to the integral.

**REMARK.** Given a flow  $\phi_t : X \rightarrow X$  and an ergodic measure  $\mu$  there is a corresponding version of the ergodic theorem. For  $f \in L^1(X, \mathcal{B}, \mu)$  the averages  $\frac{1}{\tau} \int_0^\tau f(\phi_t x) dt \rightarrow \int f(x) d\mu(x)$  as  $\tau \rightarrow +\infty$  for almost all points  $x \in X$ .

### 10.3 Applications of the ergodic theorems

In this section we shall consider some of the beautiful applications of the Birkhoff ergodic theorem.

**Application 1 (Normal numbers).** Let  $k \geq 2$  be a positive integer. For any  $0 < x < 1$  we can expand

$$x = \frac{i_1}{k} + \frac{i_2}{k^2} + \frac{i_3}{k^3} + \dots$$

where  $i_1, i_2, i_3, \dots \in \{0, 1, \dots, k-1\}$ . (This expansion may not be unique.) When  $k = 10$  this is the usual decimal expansion.

We say that  $x$  is normal to base  $k$  if for each  $i \in \{0, 1, \dots, k-1\}$  that term occurs in the expansion for  $x$  with density  $\frac{1}{k}$ , i.e.  $\lim_{N \rightarrow +\infty} \frac{1}{N} \text{Card}\{1 \leq n \leq N : i_n = i\} = \frac{1}{k}$ .

The transformation  $T : [0, 1) \rightarrow [0, 1)$  defined by  $Tx = kx \pmod{1}$  preserves Lebesgue measure  $\mu$  and is  $\mu$ -ergodic. Moreover, we see that  $T^{n-1}(x) \in (\frac{i_n}{k}, \frac{i_n+1}{k})$  and so by Theorem 10.2 we have the following.

**PROPOSITION 10.3.** *For almost all  $0 < x < 1$ ,  $x$  is normal to any base  $k \geq 2$ .*

**Application 2 (First digits of powers of 2).** Consider the sequence  $2, 4, 8, 16, 32, 64, 128, \dots, 2^n$  and consider the sequence of first digits  $2, 4, 8, 1, 3, 6, 1, \dots, r_n \in \{0, 1, \dots, 9\}$ .

**PROPOSITION 10.4.** *The frequency of the occurrence of the symbol  $k$  is given by  $\log_{10}(1 + \frac{1}{k})$ , i.e.  $\lim_{N \rightarrow +\infty} \frac{1}{N} \text{Card}\{1 \leq n \leq N : r_n = k\} = \log_{10}(1 + \frac{1}{k})$ .*

For example, the frequency of the occurrence of the symbol 7 is given by  $\log_{10}(1 + \frac{1}{7})$ .

To prove Proposition 10.4, consider the distribution of the first digits of  $\{2^n\}$ . The first digit of  $2^n$  equals  $k$  iff  $k \cdot 10^r \leq 2^n < (k+1) \cdot 10^r$  for some  $r \geq 0$ . (Equivalently,  $n \cdot \log_{10} 2 \in [\log_{10}(k+1), \log_{10} k) \pmod{1}$ ).

But since  $a = \log_{10} 2$  is irrational the map  $Tx = x + a$  is ergodic with respect to Lebesgue measure. From the ergodic theorem it is easy to deduce that the proportion of the orbit  $T^n 0$  spent in the interval  $[\log_{10}(k+1), \log_{10} k)$  is equal to its length  $|\log_{10}(k+1) - \log_{10} k| = \log_{10}(1 + \frac{1}{k})$ .

**Application 3 (Continued fractions).** The continued fraction transformation  $T : (0, 1) \rightarrow (0, 1)$  defined by  $Tx = \frac{1}{x} - [\frac{1}{x}]$  preserves the *Gauss measure*

$$\mu(B) = \frac{1}{\log 2} \int_B \frac{dx}{1+x}$$

where  $B \in \mathcal{B}$  is in the Borel sigma-algebra.

In particular, for any  $B \in \mathcal{B}$ , if  $l(B)$  denotes its measure with respect to the Haar-Lebesgue measure then

$$\frac{1}{2 \log 2} l(B) \leq \mu(B) \leq \frac{1}{\log 2} l(B)$$

(since  $\frac{1}{2} \leq \frac{1}{1+x} \leq 1$ ). Consider the (measurable) partition of  $(0, 1)$  by the intervals of the form  $I_k := (\frac{1}{k+1}, \frac{1}{k})$  for  $k \geq 1$ . For any  $n \geq 1$  we can define a finer partition into disjoint intervals by  $I_{k_0} \cap T^{-1}I_{k_1} \cap \dots \cap T^{-n}I_{k_n}$ , where  $k_0, k_1, \dots, k_n \geq 1$ .

We want to show that  $\mu$  is ergodic. As is easily seen,

- (i)  $|T'(x)| > 1$  and  $|(T^2)'(x)| \geq 4$ , for any  $x \in (0, 1)$ ,
- (ii) there exists  $D > 0$  such that  $\sup_{x, y, z \in I_k} \frac{|T''(x)|}{|T'(y)T'(z)|} \leq D, \forall 0 < x < 1$ .

The maps  $T^{n+1} : I_{k_0} \cap T^{-1}I_{k_1} \cap \dots \cap T^{-n}I_{k_n} \rightarrow [0, 1]$  have inverses which we denote by  $\psi_{k_0 \dots k_n} : [0, 1] \rightarrow I_{k_0} \cap T^{-1}I_{k_1} \cap \dots \cap T^{-n}I_{k_n}$  and we see that

$$\left| \frac{\psi'_{k_0 \dots k_n}(x)}{\psi'_{k_0 \dots k_n}(y)} \right| \leq \prod_{i=0}^n \left( 1 + \frac{D}{4^{[(n-i-1)/2]}} |T^n x - T^n y| \right) \leq C, \quad \forall x, y \in (0, 1).$$

Fix a Borel-measurable set  $E \in \mathcal{B}$  then by the change of variables formula we get the following two inequalities:

- (a) If  $l(E \cap I_{k_0} \cap T^{-1}I_{k_1} \cap \dots \cap T^{-n}I_{k_n}) > 0$ , then  $l(E \cap I_{k_0} \cap T^{-1}I_{k_1} \cap \dots \cap T^{-n}I_{k_n}) = l(\psi_{k_0 \dots k_n}(T^n E)) = \int_{T^n E} |\psi'_{k_0 \dots k_n}(x)| dx \geq l(T^n E) (\inf_{x \in [0, 1]} |\psi'_{k_0 \dots k_n}(x)|)$ ;
- (b)  $l(I_{k_0} \cap T^{-1}I_{k_1} \cap \dots \cap T^{-n}I_{k_n}) \leq \sup_{x \in [0, 1]} (\psi'_{k_0 \dots k_n}(x))$ .

From the definitions we have

$$(c) \quad l(T^n E) \geq \log 2 \cdot \mu(T^n E) = \log 2 \mu(T^{-n} T^n E) \geq \log 2 \mu(E) \geq \frac{1}{2} l(E).$$

In particular, we see that

$$l(E \cap I_{k_0} \cap T^{-1}I_{k_1} \cap \dots \cap T^{-n}I_{k_n}) \geq \frac{1}{2D} l(E) \cdot l(I_{k_0} \cap T^{-1}I_{k_1} \cap \dots \cap T^{-n}I_{k_n}).$$

Thus we see that for  $E \in \mathcal{B}$  for which  $T^{-1}E = E$  and any  $\epsilon > 0$  we can choose  $n$  large and a cover for the complement  $([0, 1] - E) \subset \cup_i I_i$  by intervals  $I_i$  of the form  $I_{k_0} \cap T^{-1}I_{k_1} \cap \dots \cap T^{-n}I_{k_n}$  such that  $\sum_i l(I_i) \leq$

$l([0, 1] - E) + \epsilon = 1 - l(E) + \epsilon$ . Then

$$\begin{aligned}
& l(E) \cdot l([0, 1] - E) \\
& \leq l(E) \left( \sum_i l(I_i) + \epsilon \right) \\
& \leq 2D \left( \sum_i l(E \cap I_i) \right) + \epsilon \\
& = 2D \left( \sum_i l(E \cap I_i) + \left( \sum_i l(I_i \cap ([0, 1] - E)) - l([0, 1] - E) \right) \right) + \epsilon \\
& = 2D \left( \sum_i l(I_i) - l([0, 1] - E) \right) + \epsilon \leq (2D + 1)\epsilon.
\end{aligned}$$

Since  $\epsilon > 0$  is arbitrary, we see that either  $l(E) = 0$  or  $l([0, 1] - E) = 0$  (or equivalently,  $\mu(E) = 0$  or  $\mu([0, 1] - E) = 0$ ).

For almost all  $0 < x < 1$  we can define a sequence  $a_n = a_n(x) \in \{1, 2, \dots\}$  (for  $n \geq 0$ ) by  $\frac{1}{1+a_n} < T^n(x) < \frac{1}{a_n}$ . We can then write  $x$  in its *continued fraction expansion* as

$$x = \frac{1}{a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \dots}}}.$$

Consider the function  $f : (0, 1) \rightarrow \mathbb{R}$  defined by  $f(x) = \log n$  whenever  $\frac{1}{1+n} < x < \frac{1}{n}$ . Observe that

$$\begin{aligned}
\int f(x) d\mu(x) &= \sum_{n=1}^{\infty} \log n \frac{1}{\log 2} \int_{\frac{1}{1+n}}^{\frac{1}{n}} \frac{1}{1+x} dx \\
&= \sum_{n=1}^{\infty} \frac{\log n}{\log 2} \log \left( \frac{(n+1)^2}{n(n+2)} \right) < +\infty.
\end{aligned}$$

By the Birkhoff theorem we have for almost all points  $x$  that

$$\frac{1}{N} \sum_{n=0}^{N-1} \log(a_n) = \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) \rightarrow \int f(x) d\mu$$

as  $N \rightarrow +\infty$ . Taking exponentials of both sides gives that

$$\lim_{N \rightarrow +\infty} (a_1 a_2 \dots a_N)^{\frac{1}{N}} = \prod_{n=1}^{\infty} \left( \frac{(n+1)^2}{n(n+2)} \right)^{\frac{\log n}{\log 2}}$$

for almost all points  $x$ .

**Application 4 (Diffeomorphisms which preserve geodesics).** Let  $M_1$  and  $M_2$  be two finite volume smooth Riemannian surfaces whose curvatures are negative (and uniformly bounded away from zero). Let  $f : M_1 \rightarrow M_2$  be a  $C^2$  diffeomorphism.

**THEOREM 10.5.** *If the image of each geodesic in  $M_1$  is a geodesic in  $M_2$ , then the diffeomorphism  $f$  is an isometry (up to a homothetic rescaling).*

We denote by  $\phi_t^1 : SM_1 \rightarrow SM_1$  and  $\phi_t^2 : SM_2 \rightarrow SM_2$  the associated geodesic flows on the unit tangent bundles  $SM_1$  and  $SM_2$  of the two manifolds. The diffeomorphism  $f : M_1 \rightarrow M_2$  extends to a map  $F : SM_1 \rightarrow SM_2$  defined by  $F(v_x) = \frac{D_x f(v_x)}{\|D_x f(v_x)\|}$ .

The proof is based on the following sublemmas:

**SUBLEMMA 10.5.1.** *Under the above hypotheses we have the following results.*

- (i)  *$F$  carries orbits of  $\phi^1$  to orbits of  $\phi^2$  (but without necessarily preserving the parameterization).*
- (ii) *Each flow  $\phi_t^i$  ( $i = 1, 2$ ) preserves the smooth Liouville measure  $\nu_i$  on the sphere bundle  $SM_i$ .*
- (iii) *The map  $F$  is nonsingular (i.e. carries sets of zero measure to sets of zero measure).*

**SUBLEMMA 10.5.2.**

- (i) *If  $M_1$  is a surface then for a (geodesic) arc  $C_1$  between two points  $x, y \in M_1$ , the set  $S_{C_1}M_1 = \cup_{x \in C_1} S_x M_1$  has relative Liouville measure  $\mu_1(S_{C_1}M_1) = 2\pi d_{M_1}(x, y)$ .*
- (ii) *If  $M_1$  is of dimension  $n \geq 2$  then for an  $(n-1)$ -dimensional geodesic submanifold  $C_1$  the set  $S_{C_1}M_1 = \cup_{x \in C_1} S_x M_1$  has relative Liouville measure  $\mu_1(S_{C_1}M_1) = \text{Vol}(S^{n-1})\text{Vol}_{(n-1)}(C_1)$ .*

*A similar conclusion holds on replacing  $M_1$  by  $M_2$  and  $C_1 \subset M_1$  by a geodesic arc or submanifold  $C_2 \subset M_2$ .*

By assumption, the image  $C_2 := f(C_1)$  is a geodesic arc from  $f(x)$  to  $f(y)$ . By applying sublemma 10.5.2 on  $M_2$  we see that  $\mu_2(S_{C_2}M_2) = 2\pi d_{M_2}(x, y)$ .

Observe that the subsets  $S_{C_1}M_1$  and  $S_{C_2}M_2$  of the unit tangent bundles  $SM_1$  and  $SM_2$ , respectively, are transverse almost everywhere to the respective flows  $\phi^1$  and  $\phi^2$ . To prove Theorem 10.5 we want to show the slightly more general result that there exists a constant  $C > 0$  with the property that if  $B_1$  is a cross-section for the flow  $\phi^1$  (and thus  $B_2 = F(B_1)$  is a cross-section for the flow  $\phi^2$ ) then  $\mu_1(B_1) = C \cdot \mu_2(B_2)$ . In particular, we want to apply this with  $B_1 = S_{C_1}M_1$ .

We shall use two simple consequences of the well-known Birkhoff ergodic theorem (presented as sublemma 10.5.3 and sublemma 10.5.4 below).

LEMMA 10.5.3 (SIZE AND RETURN TIMES). *Given a point  $x \in SM_i$  let*

$$\pi(x, t) = \text{Card}\{0 \leq u \leq t : \phi_t^i(x) \in B_i\};$$

*then for almost all  $x \in SM_i$  (with respect to Liouville measure) we have that*

$$\lim_{t \rightarrow +\infty} \frac{1}{t} \pi(x, t) = \mu_i(B_i)$$

*for  $i = 1, 2$ .*

This statement and its proof essentially appear on p. 295 of [4].

Observe that since the map  $F$  preserves the orbits of the two flows, it is a conjugacy up to a change of parameterization, i.e. there exists a map  $\alpha : SM_1 \times \mathbb{R} \rightarrow \mathbb{R}$  such that  $F(\phi_t^1(x)) = \phi_{\alpha(x,t)}^2(Fx)$ . Moreover, from the definition of  $F$  we can write  $\alpha(x, t) = \int_0^t \|Df(\phi_u^1 x)\| du$ .

The following is a direct application of the Birkhoff ergodic theorem for flows.

LEMMA 10.5.4 (AVERAGE REPARAMETERIZATION). *For almost all points  $x \in M_1$  we have that*

$$\lim_{t \rightarrow +\infty} \frac{\alpha(x, t)}{t} = \int \|Df(x)\| d\nu_1(x).$$

PROOF OF THEOREM 10.5. For a typical point  $x \in M_1$ , we can write down the (increasing) sequence of times  $t_n$  such that  $\phi_{t_n}^1 x \in B_1$ . For the corresponding point  $f(x) \in M_2$ , we can write down the (increasing) sequence of times  $t'_n$  such that  $\phi_{t'_n}^2 f(x) \in B_2$ . Notice that by construction we have that  $\pi(x, t_n) = \pi(f(x), t'_n)$ . By sublemma 10.5.4 we have that  $t'_n \sim t_n (\int \|Df(x)\| d\nu_1(x))$  as  $n \rightarrow +\infty$  (for almost all points  $x \in M_1$ ).

For almost all  $x \in M_1$  we therefore have that

$$\begin{aligned} \nu_1(B_1) &= \lim_{n \rightarrow +\infty} \frac{1}{t_n} \pi(x, t_n) \\ &= \lim_{n \rightarrow +\infty} \frac{1}{t_n} \pi(Fx, t'_n) \\ &= \left( \int \|Df(x)\| d\nu_1(x) \right) \lim_{n \rightarrow +\infty} \frac{1}{t'_n} \pi(Fx, t'_n) \\ &= \left( \int \|Df(x)\| d\nu_1(x) \right) \nu_2(B_2). \end{aligned}$$

We need only choose one point  $x$  in this set of full measure to see that this identity gives us the desired identity with the choice  $C = (\int \|Df(x)\| d\nu_1(x))$ . ■



### 10.4 The Birkhoff theorem (for invariant measures)

We begin with a definition.

**DEFINITION.** We let  $\mathcal{I} = \{A \in \mathcal{B} : T^{-1}A = A\}$  denote the *invariant sigma-algebra*, i.e. the sigma-algebra consisting of  $T$ -invariant sets.

Notice that  $E(f \circ T|\mathcal{I}) = E(f|\mathcal{I})$  for  $f \in L^1(X, \mathcal{B}, \mu)$  (since  $T^{-1}\mathcal{I} = \mathcal{I}$  and so  $\int_A E(f \circ T|\mathcal{I})d\mu = \int_A f \circ Td\mu = \int_{T^{-1}A} f \circ Td\mu = \int_A fd\mu$ ).

In the case that  $T$  is ergodic this sigma algebra is trivial i.e.  $\mathcal{I} = \{X, \emptyset\}$ .

**THEOREM 10.6 (BIRKHOFF'S ERGODIC THEOREM).** *Consider  $f \in L^1(X, \mathcal{B}, \mu)$ . If the measure  $\mu$  is  $T$ -invariant then for almost all  $x \in X$  we have that the averages*

$$\frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) \rightarrow E(f|\mathcal{I})$$

as  $N \rightarrow +\infty$ , for almost all  $x \in X$ .

**PROOF.** We can assume for convenience that  $E(f|\mathcal{I}) = 0$ . (If this is not the case we can replace  $f$  by  $f - E(f|\mathcal{I})$  on both sides of the identity.)

We need to show that if  $E(f|\mathcal{I}) = 0$  then we can choose  $h \in L^\infty(X, \mathcal{B}, \mu)$  with  $\int |f - (hT - h)|d\mu < \delta$ , i.e.  $\ker(E(\cdot|\mathcal{I})) = \text{cl}(B)$ . Since we are also at liberty to use sublemmas 10.5.1 and 10.5.2 as before (since their proofs did not require ergodicity) the proof will be complete.

Let  $B = \{hT - h : h \in L^\infty(X, \mathcal{B}, \mu)\}$  (denoted by  $E$  in the previous proof of Birkhoff's theorem). First notice that  $B \subset \ker(E(\cdot|\mathcal{I}))$  since if  $f = hT - h$  with  $h \in L^\infty(X, \mathcal{B}, \mu)$  then

$$\begin{aligned} E(hT - h|\mathcal{I}) &= E(hT|\mathcal{I}) - E(h|\mathcal{I}) \\ &= E(h|\mathcal{I})T - E(h|\mathcal{I}) \\ &= 0. \end{aligned}$$

To show that  $\ker(E(\cdot|\mathcal{I})) = \text{cl}(B)$  it suffices to show that any linear functional which vanishes on  $B$  must also vanish on  $\ker(E(\cdot|\mathcal{I}))$ . We know that any linear functional can be written in the form  $f \rightarrow \int fkd\mu$ , where  $k \in L^\infty(X, \mathcal{B}, \mu)$ . If we assume that this functional vanishes on  $B$  then it is equivalent to  $\int gkd\mu = 0$  for all  $g = hT - h \in B$ , i.e.  $\int hTk d\mu = \int hkd\mu$ ,  $\forall h \in L^\infty(X, \mathcal{B}, \mu)$ . With the specific choice  $h = k$  we get that  $\int kT \cdot kd\mu = \int k^2 d\mu$  which implies that  $\int (kT - k)^2 d\mu = 0$  (as in the proof of Theorem 10.2). In particular,  $kT = k$ .

If we assume that  $f \in \ker(E(\cdot|\mathcal{I}))$  then the image under the linear functional is

$$\int fkd\mu = \int E(fk|\mathcal{I})d\mu = \int kE(f|\mathcal{I})d\mu = 0$$

(by property (iii)), i.e. the functional vanishes on  $\ker E(\cdot|\mathcal{I})$ . This completes the proof. ■

### 10.5 Comments and references

Many simple example of applications of the ergodic theorems can be found in the appendices of [1]. A nice treatment of continued fractions and the Birkhoff ergodic theorem is contained in [2]. There exist various alternative proofs of the von Neumann ergodic theorem (cf. [3],[9]) and Birkhoff ergodic theorem (cf. [5]).

Other interesting ergodic theorems we have not considered are discussed in [6], [7] and [8, pp 101-103].

#### References

1. V. Arnol'd and A. Avez, *Ergodic Problems of Classical Mechanics*, Benjamin, New York, 1968.
2. P. Billingsley, *Ergodic Theory and Information*, Wiley, New York, 1965.
3. J. Bourgain, *An approach to pointwise ergodic theorems*, GAFA-seminar, 1987, Lecture Notes in Mathematics, Springer, Berlin, 1987.
4. I. Cornfeld, S. Fomin and Y. Sinai, *Ergodic Theory*, Springer, Berlin, 1982.
5. Y. Katznelson and B. Weiss, *A simple proof of some ergodic theorems*, Israel J. Math. **42** (1982), 291-296.
6. U. Krengel, *Ergodic Theorems*, de Gruyter, Berlin, 1985.
7. W. Parry, *Topics in Ergodic Theory*, C.U.P., Cambridge, 1986.
8. K. Petersen, *Ergodic Theory*, C.U.P., Cambridge, 1983.
9. R. Zimmer, *Essential Results of Functional Analysis*, Chicago University Press, Chicago, 1990.

## MIXING PROPERTIES

We now want to consider two stronger properties than ergodicity. These are weak mixing and strong mixing which are important from the statistical point of view, as we shall see in the next chapter.

## 11.1 Weak mixing

DEFINITION. Let  $T : X \rightarrow X$  be a measure preserving map on a probability space  $(X, \mathcal{B}, \mu)$ ; then we call  $T$  *weak-mixing* if for any  $A, B \in \mathcal{B}$  we have that

$$\frac{1}{N} \sum_{n=0}^{N-1} |\mu(T^{-n}A \cap B) - \mu(A)\mu(B)| \rightarrow 0$$

as  $N \rightarrow +\infty$ .

We have the following equivalent characterization.

LEMMA 11.1. *The following are equivalent.*

- (i)  $T$  is weak-mixing;
- (ii) for  $f, g \in L^2(X, \mathcal{B}, \mu)$  we have that

$$\frac{1}{N} \sum_{n=0}^{N-1} \left| \int fT^n g d\mu - \int f d\mu \int g d\mu \right| \rightarrow 0$$

as  $N \rightarrow +\infty$ .

PROOF. For “(ii) implies (i)” we need only make the choices  $f = \chi_A$  and  $g = \chi_B$ . For “(i) implies (ii)” we can use an argument of approximation by step functions (finite linear combinations of characteristic functions). ■

The following lemma shows that weak-mixing is a stronger property than ergodicity.

LEMMA 11.2. *If a transformation  $T : X \rightarrow X$  on a probability space  $(X, \mathcal{B}, \mu)$  is weak-mixing then it is necessarily ergodic.*

PROOF. If  $T$  is weak-mixing then by definition we have that for any  $A, B \in \mathcal{B}$

$$\frac{1}{N} \sum_{n=0}^{N-1} |\mu(T^{-n}A \cap B) - \mu(A)\mu(B)| \rightarrow 0$$

as  $N \rightarrow +\infty$ . By the triangle inequality we have that

$$\begin{aligned} & \left| \frac{1}{N} \sum_{n=0}^{N-1} \mu(T^{-n}A \cap B) - \mu(A)\mu(B) \right| \\ & \leq \frac{1}{N} \sum_{n=0}^{N-1} |\mu(T^{-n}A \cap B) - \mu(A)\mu(B)| \\ & \rightarrow 0. \end{aligned} \tag{11.1}$$

If we assume (for a contradiction) that  $T$  were not ergodic then there would exist a  $T$ -invariant set  $E \in \mathcal{B}$  with  $T^{-1}E = E$  with  $0 < \mu(E) < 1$ . If we take  $A = E$  and  $B = X - E$  in (11.1) then since  $\mu(T^{-n}E \cap (X - E)) = \mu(E \cap (X - E))$ , for all  $n \geq 0$ , we deduce that  $\mu(E) \cdot \mu(X - E) = 0$  giving the required contradiction. Thus  $T$  is ergodic. ■

The converse is not true: there exist examples of transformations which are ergodic but not weak-mixing, as the following simple example shows.

EXAMPLE (ERGODIC, NOT WEAK-MIXING). Let  $X = \mathbb{R}/\mathbb{Z}$ , let  $\mathcal{B}$  be the Borel sigma-algebra, and let  $\mu$  be the Haar-Lebesgue measure. For any irrational number  $a \in \mathbb{R}$  the transformation  $T : X \rightarrow X$  defined by  $T(x) = x + a \pmod{1}$  is known to be ergodic. We can see that it is not weak-mixing by choosing  $A = B = [0, \frac{1}{2}]$ , and then  $\mu(A)\mu(B) = \frac{1}{4}$ . Since the sequence  $na + \mathbb{Z}$  is uniformly distributed we know that the proportion of the terms in the sub-sequence  $n_i$  for which  $n_i a \in [0, \frac{1}{100}] \pmod{1}$  is  $\frac{1}{100}$ . For these terms we have that  $\mu(T^{n_i}A \cap B) - \mu(A)\mu(B) \geq \frac{49}{100} - \frac{1}{4} = \frac{24}{100}$  which means that

$$\liminf_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=0}^{N-1} |\mu(T^{n_i}A \cap B) - \mu(A)\mu(B)| \geq \frac{1}{100} \cdot \frac{24}{100} > 0.$$

## 11.2 A density one convergence characterization of weak mixing

Using the previous lemma on sequences there is also a characterization for weak mixing which is closer to that of strong mixing. We say that a sequence  $\{n_i\}_{i \in \mathbb{N}}$  of natural numbers has density one if

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \text{Card}\{n_i \in [0, 1, \dots, n-1] : i \in \mathbb{N}\} = 1.$$

PROPOSITION 11.3. *The transformation  $T$  is weak-mixing if there exists a sequence  $\{n_i\}_{i \in \mathbb{N}}$  of density one such that  $\mu(T^{-n_i} A \cap B) \rightarrow \mu(A)\mu(B)$  as  $i \rightarrow +\infty$ .*

The proof requires only the following simple lemma on sequences.

LEMMA 11.4. *The following are equivalent for a bounded sequence of real positive numbers  $\{a_n\}$ .*

- (1)  $\frac{1}{n} \sum_{k=0}^{n-1} a_k \rightarrow 0$  as  $n \rightarrow +\infty$ ; and
- (2)  $\lim_{k \rightarrow +\infty} a_{n_k} = 0$  for some sub-sequence  $\{n_k\} \subset \mathbb{N}$  of density one (i.e.  $\lim_{n \rightarrow +\infty} \frac{1}{n} \text{Card}\{n_i \in [0, 1, \dots, n-1]: i \in \mathbb{N}\} = 1$ ).

PROOF. (2)  $\implies$  (1): Let  $J$  be a sequence of density one. Given  $\epsilon > 0$  we can choose  $N$  such that for  $n \geq N$  we have that

$$\text{Card}\{n_i \in [0, 1, \dots, n-1]\} \geq n(1 - \epsilon)$$

and for  $n_i \geq N$  we have  $a_{n_i} \leq \epsilon$ . In particular,

$$\begin{aligned} \frac{1}{n} \sum_{k=0}^{n-1} a_k &= \frac{1}{n} \left( \sum_{k=0}^N a_k + \sum_{\substack{k \in \{n_i\} \\ N < k \leq n-1}} a_k + \sum_{\substack{k \notin \{n_i\} \\ N < k \leq n-1}} a_k \right) \\ &\leq \frac{1}{n} \sum_{k=0}^N a_k + \epsilon ((1 - \epsilon) + \sup\{a_i\}). \end{aligned}$$

We can choose  $n$  sufficiently large that  $\frac{1}{n} \sum_{k=0}^N a_k < \epsilon$ . Since  $\epsilon > 0$  can be chosen arbitrarily small we have that  $\frac{1}{n} \sum_{k=0}^{n-1} a_k \rightarrow 0$ .

(1)  $\implies$  (2): Assume that  $\frac{1}{n} \sum_{k=0}^{n-1} a_k \rightarrow 0$ . For each  $m \geq 1$  we define

$$J_m = \left\{ n \in \mathbb{N} : a_n \leq \frac{1}{m} \right\}$$

and observe that this has density one, since

$$\frac{1}{m} \left( \frac{1}{n} \sum_{k=0}^{n-1} \chi_{\mathbb{N}-J_m}(k) \right) \leq \frac{1}{n} \sum_{k=0}^{n-1} a_k \rightarrow 0$$

as  $n \rightarrow +\infty$ . Thus for each  $m \geq 1$  we can choose  $n_m$  such that

$$\frac{1}{n} \left( \sum_{k=0}^{n-1} \chi_{\mathbb{N}-J_m}(k) \right) \leq \frac{1}{m}$$

for  $n \geq n_m$ . We then define

$$J = \bigcup_{k=1}^{\infty} J_k \cap [n_k, \dots, n_{k+1}]$$

and it is easy to see that  $J$  has density one and  $\lim_{n \rightarrow \infty: n \in J} a_n = 0$ . ■

The following corollary is quite useful.

COROLLARY 11.4.1.  $\frac{1}{n} \sum_{k=0}^{n-1} a_k \rightarrow 0$  as  $n \rightarrow +\infty$  if and only if  $\frac{1}{n} \sum_{k=0}^{n-1} a_k^2 \rightarrow 0$  as  $n \rightarrow +\infty$ .

### 11.3 A generalization of the von Neumann ergodic theorem

We want to present the following interesting generalization of the Von Neumann ergodic theorem (Theorem 10.1) for weak-mixing transformations. It will only be used in chapter 16 and is not required for the rest of this chapter.

THEOREM 11.5. Assume that  $f_1, \dots, f_k \in L^\infty(X, \mathcal{B}, \mu)$ . If  $T : X \rightarrow X$  is weak-mixing then

$$\frac{1}{N} \sum_{n=1}^N f_1(T^n x) f_2(T^{2n} x) \dots f_k(T^{kn} x) \rightarrow \int f_1 d\mu \int f_2 d\mu \dots \int f_k d\mu$$

(in the  $L^2$  topology) as  $N \rightarrow +\infty$ .

PROOF. The proof is by induction. When  $k = 1$ , this is precisely the Von Neumann ergodic theorem (Theorem 10.1).

Assume that the result has been established for  $k - 1$  functions. We may assume without loss of generality that  $\int f_k d\mu = 0$  (otherwise we need only replace  $f_k$  by  $f_k - \int f_k d\mu$ ). Thus it suffices to show that

$$\int \left| \frac{1}{N} \sum_{n=1}^N f_1(T^n x) f_2(T^{2n} x) \dots f_k(T^{kn} x) \right|^2 d\mu(x) \rightarrow 0 \text{ as } N \rightarrow +\infty.$$

For any  $1 \leq m \leq N$  we can now bound

$$\begin{aligned} & \int \left| \frac{1}{N} \sum_{n=1}^N f_1(T^n x) f_2(T^{2n} x) \dots f_k(T^{kn} x) \right|^2 d\mu(x) \\ & \leq \int \left| \frac{1}{N} \sum_{n=1}^N \left( \frac{1}{m} \sum_{j=0}^{m-1} f_1(T^{n+j} x) f_2(T^{2(n+j)} x) \dots f_k(T^{k(n+j)} x) \right) \right|^2 d\mu(x) \\ & \quad + \left( \frac{2m}{N} + \frac{m^2}{N^2} \right) \left( \max_{1 \leq i \leq k} \|f_i\|_\infty \right)^2 \\ & \leq \frac{1}{N} \sum_{n=1}^N \left( \int \left| \frac{1}{m} \sum_{j=0}^{m-1} f_1(T^{n+j} x) f_2(T^{2(n+j)} x) \dots f_k(T^{k(n+j)} x) \right|^2 d\mu(x) \right) \\ & \quad + \left( \frac{2m}{N} + \frac{m^2}{N^2} \right) \left( \max_{1 \leq i \leq k} \|f_i\|_\infty \right)^2. \end{aligned} \tag{11.2}$$

(The first inequality comes from the observation that

$$\frac{1}{N} \sum_{n=1}^N a_n \leq \frac{1}{N} \sum_{n=1}^{N-(m-1)} \left( \frac{1}{m} \sum_{j=0}^{m-1} a_{n+j} \right) + \frac{m}{N} \left( \max_{1 \leq i \leq m} |a_i| + \max_{N-m \leq i \leq N} |a_i| \right)$$

for any real numbers  $a_1, \dots, a_N$ . The second inequality comes from the observation that  $\left( \frac{1}{N} \sum_{n=1}^N b_n \right)^2 \leq \frac{1}{N} \sum_{n=1}^N |b_n|^2$  for any real numbers  $b_1, \dots, b_N$ .) We next observe that

$$\begin{aligned} & \int \left| \sum_{j=0}^{m-1} f_1(T^{n+j}x) f_2(T^{2(n+j)}x) \dots f_k(T^{k(n+j)}x) \right|^2 d\mu(x) \\ &= \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \int \left( \prod_{l=1}^k f_l(T^{l(n+i)}x) \right) \left( \prod_{l=1}^k f_l(T^{l(n+j)}x) \right) d\mu(x) \quad (11.3) \\ &= \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \int \left( \prod_{l=1}^k (f_l \cdot f_l \circ T^{l(j-i)})(T^{l(n+i)}x) \right) d\mu(x). \end{aligned}$$

By the inductive hypothesis we know that for each  $0 \leq i, j \leq m-1$ ,

$$\frac{1}{N} \sum_{n=1}^N \prod_{l=2}^k (f_l \cdot f_l \circ T^{l(j-i)})(T^{l(n+i)}x) \rightarrow \prod_{l=2}^k \int f_l \cdot f_l \circ T^{l(j-i)} d\mu$$

as  $N \rightarrow +\infty$  (in the  $L^2$  topology) and so

$$\begin{aligned} & \frac{1}{N} \sum_{n=1}^N \int \prod_{l=1}^k (f_l \cdot f_l \circ T^{l(j-i)})(T^{l(n+i)}x) d\mu(x) \\ &= \frac{1}{N} \sum_{n=1}^N \int (f_1 \cdot f_1 \circ T^{l(j-i)})(x) \left( \prod_{l=2}^k (f_l \cdot f_l \circ T^{l(j-i)})(T^{l(n+i)}x) \right) d\mu(x) \\ &\rightarrow \prod_{l=1}^k \int f_l \cdot f_l \circ T^{l(j-i)} d\mu. \end{aligned} \quad (11.4)$$

Comparing (11.2), (11.3) and (11.4) we see that

$$\begin{aligned} & \limsup_{N \rightarrow +\infty} \int \left| \frac{1}{N} \sum_{n=1}^N f_1(T^n x) f_2(T^{2n} x) \dots f_k(T^{kn} x) \right|^2 d\mu(x) \\ &\leq \frac{1}{m^2} \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \left( \prod_{l=1}^k \int f_l \cdot f_l \circ T^{l(j-i)} d\mu(x) \right). \end{aligned} \quad (11.5)$$

Finally, since  $T$  is weak-mixing we know that  $\int f_l \cdot f_l \circ T^{lr_n} d\mu(x) \rightarrow \int f_l d\mu(x) = 0$  where  $r_n \rightarrow +\infty$  through a set of density one. Thus for sufficiently large  $m$  the expression in (11.4) can be made arbitrarily small. ■

### 11.4 The spectral viewpoint

Consider a measure preserving transformation  $T : X \rightarrow X$  on a probability space  $(X, \mathcal{B}, \mu)$ . Consider the Hilbert space  $H = L^2(X, \mathcal{B}, \mu)$  of square integrable functions with the inner product  $\langle f, g \rangle = \int f \bar{g} d\mu$ . We can associate to  $T$  an operator  $U_T : H \rightarrow H$  defined by  $(U_T f)(x) = f(Tx)$  whenever  $f \in H$ . It is easy to see the following.

LEMMA 11.6.  $U_T : H \rightarrow H$  is an isometry. i.e.  $\|U_T f\| = \|f\|$ .

We recall a few elementary observations about operators  $U : H \rightarrow H$  on a Hilbert space  $H$ . We call a linear operator  $U$  an *isometry* if for every pair of vectors  $x, y \in H$  we have that  $\langle Ux, Uy \rangle = \langle x, y \rangle$ . We shall only be interested in isometries.

An eigenvalue for  $U : H \rightarrow H$  is a complex number  $\alpha \in \mathbb{C}$  for which there exists a (non-zero) vector  $x \in H$  (called the eigenvector) such that  $Ux = \alpha x$ .

LEMMA 11.7. *Eigenvalues of isometries must be complex numbers of modulus unity.*

PROOF. Clearly, if  $Ux = \alpha x$  then  $\langle Ux, Ux \rangle = \alpha \bar{\alpha} \langle x, x \rangle$ . But since  $U$  is an isometry we have that  $\langle Ux, Ux \rangle = \langle x, x \rangle$  and so  $\langle x, x \rangle = |\alpha|^2 \langle x, x \rangle$  i.e.  $|\alpha| = 1$ , as claimed.

An important aspect of the spectrum of the operator is the variety (or lack of it) of eigenvectors. Two extreme cases are the following.

DEFINITION. The operator  $U : H \rightarrow H$  has *continuous spectrum* if there are no eigenvectors. The operator  $U : H \rightarrow H$  has *pure point spectrum* if  $H$  is the closure of the linear span of the eigenvectors.

REMARK. Between these two extreme cases we have the possibility of having *mixed spectrum*. We can let  $V \subset H$  denote the subspace spanned by the eigenvectors. We let  $V^\perp \subset H$  denote the orthogonal subspace to  $V$ . From the definitions we see that  $U : V \rightarrow V$  then has pure point spectrum and  $U : V^\perp \rightarrow V^\perp$  has continuous spectrum.

We now recall one of the basic theorems in spectral theory.

DEFINITION. A sequence  $r_n \in \mathbb{C}$ ,  $n \in \mathbb{Z}$ , is called *positive definite* if for each  $N \geq 1$  and each sequence  $a_0, \dots, a_N$  we always have that

$$\sum_{n,m=0}^N r_{n-m} a_n \bar{a}_m \geq 0.$$



**BOCHNER-HERGLOTZ SPECTRAL THEOREM.** *If  $r_n, n \in \mathbb{Z}$ , is positive definite then there is a unique finite Borel measure  $\mu$  on  $\mathbb{R}/\mathbb{Z}$  such that  $r_n = \int_0^1 e^{2\pi i n t} d\mu(t)$ .*

(The proof can be found in the appendix to [1].)

**APPLICATION TO ISOMETRIES.** The Bochner-Herglotz theorem is particularly well suited to isometries  $U : H \rightarrow H$ . Fix  $x \in H$  and then set  $r_n = \langle U^n x, x \rangle$  and  $r_{-n} = \langle x, U^n x \rangle$  and observe that

$$\sum_{n,m=0}^N r_{n-m} a_n \bar{a}_m = \left\langle \sum_{n=0}^N a_n U^n x, \sum_{n=0}^N a_n U^n x \right\rangle \geq 0.$$

The measure  $\mu$  on the unit circle  $\mathbb{R}/\mathbb{Z}$  is called the *spectral measure*.

The choice of point  $x \in H$  affects the resulting spectral measure  $\mu$ . If  $x \in V$  (i.e.  $x$  is in the closure of the span of the eigenvectors) then the associated measure is singular with respect to the Haar-Lebesgue measure on the circle  $\mathbb{R}/\mathbb{Z}$ . If  $x \in V^\perp$  then the associated measure is absolutely continuous with respect to the Haar-Lebesgue measure on the circle  $\mathbb{R}/\mathbb{Z}$ .

**EXAMPLE.** Consider an irrational rotation  $T : X \rightarrow X$  on the unit circle  $X = \mathbb{R}/\mathbb{Z}$  defined by  $T(x + \mathbb{Z}) = (x + a + \mathbb{Z})$ . The functions  $e_n(x) = e^{2\pi i n x}$  are eigenfunctions for the operator  $U : L^2(X, \mathcal{B}, \mu) \rightarrow L^2(X, \mathcal{B}, \mu)$  since

$$U e_n(x) = e^{2\pi i n(x+a)} = e^{2\pi i n a} e_n(x).$$

Since the family  $e_n, n \in \mathbb{Z}$ , spans the space the transformation  $T$  has pure point spectrum.

In applying this to ergodic theory, we consider a measure preserving transformation  $T : X \rightarrow X$  on the Hilbert space  $H = L^2(X, \mathcal{B}, \mu)$  with inner product  $\langle f, g \rangle = \int f \bar{g} d\mu$ .

**REMARK.** The following lemma is also a standard result from spectral theory (although we won't require it).

**RIEMANN-LEBESGUE LEMMA.** *If the spectral measure  $\mu$  on  $\mathbb{R}/\mathbb{Z}$  for the operator  $U : H \rightarrow H$  (and a point  $x \in H$ ) is absolutely continuous then  $\langle U^n x, x \rangle \rightarrow 0$  as  $n \rightarrow +\infty$ .*

### 11.5 Spectral characterization of weak mixing

The Hilbert space  $H = L^2(X, \mathcal{B}, \mu)$  has the obvious one-dimensional subspace consisting of constant functions and denoted by  $\mathbb{C}$ . We let  $\mathbb{C}^\perp$  denote the orthonormal (co-dimension one) subspace. For a measure preserving transformation  $T : X \rightarrow X$  the associated isometry  $U : H \rightarrow H$  preserves both  $\mathbb{C}$  and  $\mathbb{C}^\perp$ .

**PROPOSITION 11.8.** *Let  $T : X \rightarrow X$  be a measure preserving transformation on the probability space  $(X, \mathcal{B}, \mu)$ . The following conditions are equivalent:*

- (1) *for the map  $U : \mathbb{C}^\perp \rightarrow \mathbb{C}^\perp$  has continuous spectrum;*
- (2)  *$T$  is weak-mixing;*
- (3) *the measure preserving transformation  $T \times T : X \times X \rightarrow X \times X$  (on  $X \times X$  with the product measure  $\mu \times \mu$ ) defined by  $(T \times T)(x, y) = (Tx, Ty)$  is weak-mixing (and thus ergodic).*

**PROOF.**

(1)  $\implies$  (2): Assume that  $U : \mathbb{C}^\perp \rightarrow \mathbb{C}^\perp$  has continuous spectrum.  $\bar{\mu}$  denotes the spectral measure. Choose any vector  $f \in \mathbb{C}^\perp$ ; then we estimate that

$$\begin{aligned}
 & \frac{1}{N} \sum_{n=0}^{N-1} \left| \int f \circ T^n \bar{f} d\mu \right|^2 \\
 &= \frac{1}{N} \sum_{n=0}^{N-1} \left( \int f \circ T^n \bar{f} d\mu \right) \left( \int \bar{f} \circ T^n f d\mu \right) \\
 &= \frac{1}{N} \sum_{n=0}^{N-1} \left( \int_0^1 e^{2\pi i n t} d\bar{\mu}(t) \right) \left( \int_0^1 e^{-2\pi i n s} d\mu(s) \right) \\
 &= \frac{1}{N} \sum_{n=0}^{N-1} \int_0^1 \int_0^1 e^{2\pi i n(t-s)} d(\bar{\mu} \times \mu)(t, s) \\
 &= \int_0^1 \int_0^1 \frac{1}{N} \left( \frac{e^{2\pi i N(t-s)} - 1}{e^{2\pi i(t-s)} - 1} \right) d(\bar{\mu} \times \mu)(t, s).
 \end{aligned}$$

Observe that since  $\mu$  has continuous spectrum the product measure  $\bar{\mu} \times \mu$  gives the diagonal  $\{(t, t) : t \in \mathbb{R}/\mathbb{Z}\}$  measure zero. In particular, the last integrand is finite almost everywhere. Since

$$\frac{1}{N} \frac{e^{2\pi i N(t-s)} - 1}{e^{2\pi i(t-s)} - 1} \rightarrow 0$$

for almost all  $(s, t)$  and it is dominated by the constant function 1, we see by the Lebesgue dominated convergence theorem that

$$\frac{1}{N} \sum_{n=0}^{N-1} \left| \int f \circ T^n \bar{f} d\mu \right|^2 \leq \int_0^1 \int_0^1 \left( \frac{1}{N} \frac{e^{2\pi i N(t-s)} - 1}{e^{2\pi i(t-s)} - 1} \right) d(\bar{\mu} \times \bar{\mu})(t, s) \rightarrow 0$$

as  $N \rightarrow +\infty$ . Finally, we observe that

$$\frac{1}{N} \sum_{n=0}^{N-1} \left| \int f \circ T^n \bar{f} d\mu \right|^2 \rightarrow 0$$

implies

$$\frac{1}{N} \sum_{n=0}^{N-1} \left| \int f \circ T^n \bar{f} d\mu \right| \rightarrow 0$$

(by Corollary 11.4.1).

(2)  $\implies$  (3): Consider sets  $E = \sup_i (A_i \times B_i)$ ,  $F = \sup_j (C_j \times D_j)$  (for finite disjoint unions of product sets  $A_i$ ,  $B_i$ ,  $C_j$  and  $D_j$ ). Since  $T$  is weak-mixing we know that

$$\mu(T^{-n_k} A_i \cap C_j) \rightarrow \mu(A_i) \mu(C_j) \quad (11.6)$$

and

$$\mu(T^{-n_k} B_i \cap D_j) \rightarrow \mu(B_i) \mu(D_j) \quad (11.7)$$

for sequences  $n_k \rightarrow +\infty$  of density one (and without loss of generality we can assume that we have the same sequence in each case).

To show  $T \times T$  is weak-mixing we want to show that

$$(\mu \times \mu) \left( (T \times T)^{-n_i} E \cap F \right) \rightarrow (\mu \times \mu)(E) (\mu \times \mu)(F)$$

for a sequence  $n_k \rightarrow +\infty$  of unit density. This follows from (11.5) and (11.6) since

$$\begin{aligned} & (\mu \times \mu) \left( (T \times T)^{-n_i} (A_i \times B_i) \cap (C_j \times D_j) \right) \\ &= \mu(T^{-n_i} A_i \cap C_j) \cdot \mu(T^{-n_i} B_i \cap D_j) \\ &\rightarrow \mu(A_i) \mu(C_j) \mu(B_i) \mu(D_j) \\ &= \mu(T^{-n_i} A_i \cap C_j) \cdot \mu(T^{-n_i} B_i \cap D_j) \end{aligned}$$

for a sequence  $n_k \rightarrow +\infty$  of unit density.

(3)  $\implies$  (1): Assume for a contradiction that there is a *non-constant* eigenfunction  $f \in H$  for  $T : X \rightarrow X$ , i.e.  $Uf = \alpha f$ . We can then define a function  $F : X \times X \rightarrow \mathbb{C}$  by  $F(x, y) = f(x) \bar{f}(y)$ . Observe that

$$F(Tx, Ty) = f(Tx) \bar{f}(Ty) = \alpha \bar{\alpha} f(x) \bar{f}(y) = F(x, y).$$

But then  $F(x, y)$  is a  $(T \times T)$ -invariant function which is non-constant. This contradicts  $T \times T$  being ergodic (and therefore being weak-mixing).

### 11.6 Strong mixing

We now turn to another notion of “mixing”.

DEFINITION. Let  $T : X \rightarrow X$  be a measure preserving transformation on a probability space  $(X, \mathcal{B}, \mu)$ ; then we call  $T$  *strong mixing* if for any  $A, B \in \mathcal{B}$  we have that

$$\mu(T^{-n}A \cap B) \rightarrow \mu(A)\mu(B)$$

as  $n \rightarrow +\infty$ . We have the following equivalent characterization.

LEMMA 11.9. *The following are equivalent:*

- (i)  $T$  is strong-mixing;
- (ii) for  $f, g \in L^2(X, \mathcal{B}, \mu)$ ,  $\int f \circ T^n g d\mu \rightarrow \int f d\mu \int g d\mu$  as  $n \rightarrow +\infty$ .

PROOF. For “(ii)  $\implies$  (i)” we need only make the choices  $f = \chi_A$  and  $g = \chi_B$ . For “(i)  $\implies$  (ii)” we can use an argument of approximation by step functions (finite linear combinations of characteristic functions). ■

The following lemma states the obvious fact that strong mixing is a stronger property than weak mixing.

LEMMA 11.10. *If  $T$  is a strong-mixing transformation on a probability space  $(X, \mathcal{B}, \mu)$  then it is necessarily weak-mixing (and thus also ergodic).*

PROOF. This is immediate from the definitions. ■

EXAMPLE (MARKOV MEASURES AND SHIFTS). Recall that a subshift of finite type  $T : X \rightarrow X$  is defined on a space

$$X = \{x \in \prod_{n \in \mathbb{Z}} \{1, \dots, k\} : A(x_n, x_{n+1}) = 1, n \in \mathbb{Z}\}$$

for some  $k \times k$  matrix  $A$  with entries either zero or unity. We define  $T(x_n) = (x_{n+1})$  (i.e. all terms in the infinite sequences are shifted one place to the left). We shall assume in addition that  $A$  is aperiodic, i.e. there exists  $n \geq 1$  such that for each  $1 \leq i, j \leq k$  we have that  $A^n(i, j) = 1$ .

Let  $P$  denote a  $k \times k$  stochastic matrix with entries  $P(i, j) = 0$  iff  $A(i, j) = 0$ , and let  $p$  be its left eigenvector. Recall that we define the associated Markov measure by

$$\mu[i_0, \dots, i_{l-1}] = p(i_0)P(i_0, i_1) \dots P(i_{l-2}, i_{l-1}).$$

Let  $C = [i_0, \dots, i_{l-1}]$  and  $D = [j_0, \dots, j_{l-1}]$  be two cylinder sets. Observe that

$$C \cap T^{-(n+l)}D = \cup_{x_l, \dots, x_{n+l-1}} [i_0, \dots, i_{l-1}, x_l, \dots, x_{n+l-1}, j_0, \dots, j_{l-1}]$$

and thus

$$\begin{aligned} & \mu(C \cap T^{-(n+l)}D) \\ &= \sum_{x_l, \dots, x_{n+l-1}} p(i_0)P(i_0, i_1) \dots P(i_{l-1}, x_l) \dots P(x_{n+l-1}, j_0) \dots P(j_{l-2}, j_{l-1}) \\ &= \mu(C)\mu(D) \left( \frac{1}{p(j_0)} \sum_{x_l, \dots, x_{n+l-1}} P(i_{l-1}, x_l) \dots P(x_{n+l-1}, j_0) \right) \\ &= \mu(C)\mu(D) \frac{P^n(i_{l-1}, j_0)}{p(j_0)}. \end{aligned}$$

However, we know that  $P^n(i_{l-1}, j_0) \rightarrow p(j_0)$  as  $n \rightarrow +\infty$  (by writing  $P$  in terms of Jordan forms) and so we know that

$$\mu(C \cap T^{-(n+l)}D) \rightarrow \mu(C)\mu(D)$$

as  $n \rightarrow +\infty$ . For arbitrary sets  $A, B \in \mathcal{B}$  we can cover them by unions of disjoint cylinders  $A \subset \cup_{i=1}^n C_i$  and  $B \subset \cup_{j=1}^m D_j$  such that  $\mu((\cup_{i=1}^n C_i - A) \cap B) < \epsilon$  and  $\mu((\cup_{j=1}^m D_j - B) \cap A) < \epsilon$  and then by approximation we see that  $\mu(A \cap T^{-(n+l)}B) \rightarrow \mu(A)\mu(B)$  as  $n \rightarrow +\infty$ .

### 11.7 Comments and reference

We have given only the briefest introduction to the spectral theory associated with measure preserving transformations. A particularly nice introduction is contained in the appendix to [1].

#### Reference

1. W. Parry, *Topics in Ergodic Theory*, C.U.P., Cambridge, 1981.

STATISTICAL PROPERTIES IN ERGODIC THEORY

12.1 Exact endomorphisms

DEFINITION. We call a measure preserving transformation  $T : X \rightarrow X$  on a probability space  $(X, \mathcal{B}, \mu)$  an *exact endomorphism* if  $\bigcap_{n=0}^{\infty} T^{-n}\mathcal{B} = \{X, \emptyset\}$  up to a set of zero measure (i.e. if  $B \in T^{-n}\mathcal{B}$ , for every  $n \geq 0$ , then  $\mu(B) = 0$  or  $\mu(B) = 1$ ).

PROPOSITION 12.1.  $T : X \rightarrow X$  is exact if for any positive measure set  $A$  with  $T^n A \in \mathcal{B}$  ( $n \geq 0$ ),  $\mu(T^n(A)) \rightarrow 1$  as  $n \rightarrow +\infty$ .

It is easy to see that this sufficient condition for exactness is also necessary [2] (although we will not need this here).

PROOF. First we remark that  $T$  is exact if every measurable set  $A$  satisfying for arbitrary  $n$  the relationship  $A = T^{-n}(T^n A)$  is of either measure zero or measure 1. For such a set  $A$ , it is clear that  $\mu(A) = 1$  if  $\mu(A) > 0$ , as  $\mu(T^n A) = \mu(A)$  and so  $\lim_{n \rightarrow \infty} \mu(T^n A) = \mu(A) = 1$  if  $\mu(A) > 0$ . ■

PROPOSITION 12.2. If  $T$  is exact then it is strong-mixing.

PROOF. Consider the sub-sigma-algebras  $\mathcal{B} \supset T^{-1}\mathcal{B} \supset T^{-2}\mathcal{B} \supset \dots \supset \{X, \emptyset\}$ . We can associate the nested subspaces  $L^2(\mathcal{B}) \supset L^2(T^{-1}\mathcal{B}) \supset L^2(T^{-2}\mathcal{B}) \supset \dots \supset \mathbb{C}$  and for each  $n \neq 0$  we can choose an orthonormal basis  $\{k_i \circ T^n\}_{i=0}^{j_n}$  for  $L^2(T^{-n}\mathcal{B}) \ominus L^2(T^{-(n+1)}\mathcal{B})$ . It follows that  $\{k_i \circ T^n\}_{i=0}^{j_n} \infty_{n=0}$  is an orthonormal basis for  $L^2(X, \mathcal{B}, \mu)$ . Two functions  $f, g \in L^2(X, \mathcal{B}, \mu) \ominus \mathbb{R}$  can be written in the form

$$\begin{cases} f = \sum_{n=0}^{\infty} \sum_i a_{n,i} k_i \circ T^n + (\int f d\mu), \\ g = \sum_{n=0}^{\infty} \sum_i b_{n,i} k_i \circ T^n + (\int g d\mu), \end{cases}$$

where  $a_{n,i}, b_{n,i} \in \mathbb{R}$ . In particular,

$$\int f \circ T^N g d\mu = \sum_{n=0}^{\infty} \sum_i a_{n,i} b_{n+N,i} + \int f d\mu \int g d\mu \rightarrow \int f d\mu \int g d\mu$$

as  $N \rightarrow +\infty$ , i.e.  $T$  is strong-mixing. ■

EXAMPLE 1 (ONE-SIDED APERIODIC MARKOV SHIFTS). We can modify the definition of the Markov shift and define

$$X_A^+ = \left\{ x \in \prod_{n \in \mathbb{N}^+} \{0, \dots, k-1\} : A(x_n, x_{n+1}) = 1, n \in \mathbb{Z}^+ \right\}$$

and  $\sigma : X_A^+ \rightarrow X_A^+$  by  $(\sigma x)_n = x_{n+1}$ . For the stochastic matrix  $P$  (with entries  $P(i, j) = 0$  iff  $A(i, j) = 0$ ) letting  $p$  be its left eigenvector we define the measure on a cylinder

$$[i_0, \dots, i_{l-1}] = \{x \in X_A^+ : x_j = i_j, 0 \leq j \leq l-1\},$$

$$\mu[i_0, \dots, i_{l-1}] = p(i_0)P(i_0, i_1) \dots P(i_{l-2}, i_{l-1}).$$

Let  $A$  be aperiodic. Then the argument for the (two sided) Markov shift still applies and we see that  $T$  is strong-mixing; moreover,  $\forall \epsilon > 0, \forall$  cylinders  $C, \exists N > 0$  such that  $\forall n \geq N$  and any cylinder  $D$  we have  $|\mu(C \cap T^{-n}D) - \mu(C)\mu(D)| \leq \epsilon\mu(C)\mu(D)$ . By approximating an arbitrary set  $B \in \mathcal{B}$  by a cylinder  $D$  we see that the same result holds on replacing  $D$  by  $B$ .

Assume that  $E \in \bigcap_{n=0}^{\infty} T^{-n}\mathcal{B}$  and write  $E = T^{-n}E_n$ . For any cylinder  $C$  we see from the above observations that

$$\mu(C \cap E) = \mu(C \cap T^{-n}E_n) \geq (1 - \epsilon)\mu(E_n)\mu(C) = (1 - \epsilon)\mu(E)\mu(C);$$

since  $\epsilon > 0$  is arbitrary we see that  $\mu(C \cap E) \geq \mu(E)\mu(C)$  for all cylinders  $C$ . By approximation by disjoint unions of cylinders we can replace this by  $\mu(B \cap E) \geq \mu(E)\mu(B), \forall B \in \mathcal{B}$ . If we take  $B = X - E$  we see that  $\mu(E)\mu(X - E) = 0$ . This completes the proof that  $T$  is exact.

## 12.2 Statistical properties of piecewise expanding Markov maps

Consider a piecewise expanding  $C^2$  surjective Markov map  $T : I \rightarrow I$  for which there exists  $\beta > 1$  with  $\inf_{x \in I} |T'(x)| \geq \beta$ . We can define an operator  $\mathcal{L} : L^1(I) \rightarrow L^1(I)$  as follows.

DEFINITION. Given  $f \in L^1(I)$  we define the *Perron-Frobenius operator* by

$$\mathcal{L}f(x) = \sum_{y \in T^{-1}x} \frac{f(y)}{|T'(y)|} \left( = \sum_{i=1}^k f(\psi_i x) |\psi_i'(x)| \chi_{TI_i}(x) \right)$$

(where  $\psi_i$  denotes the inverse of  $T|I_i$ ).

LEMMA 12.3. *For any  $f \in L^1(I)$  satisfying  $(\mathcal{L}f)(x) = f(x)$  the measure  $\mu$  defined by  $f = \frac{d\mu}{dx}$  is  $T$ -invariant.*

PROOF. This follows from the change of variables formula since we have  $\mu(T^{-1}A) = \int_{T^{-1}A} f(x)dx = \sum_{i=1}^k \int_{T I_i \cap A} |\psi'_i(x)| f \circ \psi_i(x) dx = \int_A \mathcal{L}f(x) dx = \mu(A)$ . ■

We have the following result.

PROPOSITION 12.4 (SMOOTH INVARIANT MEASURES FOR PIECEWISE EXPANDING MARKOV MAPS). *There exists an invariant probability measure  $\mu$  which is absolutely continuous with respect to the (normalized) Haar-Lebesgue measure  $\lambda$  (i.e. there exists  $f \in L^1(I)$  such that  $\mu(B) = \int_B f(x)d\lambda(x)$  for every Borel set  $B \in \mathcal{B}$ ).*

PROOF. By Lemma 12.3, to construct  $\mu$  it suffices to find such a function  $f$  satisfying  $\mathcal{L}f = f$ . We first choose a point  $x \in I$  and for any  $n \geq 1$  we look at the families  $T^{-n}x$  of all  $n$ -iterate pre-images of  $x$ .

It is easy to see from the chain rule that

$$\mathcal{L}^n 1(x) = \sum_{y \in T^{-1}x} \frac{\mathcal{L}^{n-1} 1(y)}{|T'(y)|} = \sum_{y \in T^{-n}x} \frac{1}{|T^{n'}(y)|}.$$

We denote the inverse of  $T^n|_{\cap_{j=0}^{n-1} T^{-j} I_{i_{j+1}}}$  by  $\psi_{i_1 \dots i_n}$ . Let  $\mathcal{V}$  be the partition generated by  $\{T(I_i) : 1 \leq i \leq k\}$ . Then for  $x, x' \in V \in \mathcal{V}$  we can compare

$$\begin{aligned} |\mathcal{L}^n 1(x) - \mathcal{L}^n 1(x')| &= \left| \sum_{y \in T^{-n}x} \frac{1}{|T^{n'}(y)|} - \sum_{y' \in T^{-n}x'} \frac{1}{|T^{n'}(y')|} \right| \\ &= \sum_{i_1, \dots, i_n} \left| |\psi'_{i_1 \dots i_n}(x)| - |\psi'_{i_1 \dots i_n}(x')| \right| \chi_{T^n I_{i_1 \dots i_n}}(x) \end{aligned}$$

where  $I_{i_1 \dots i_n} = \cap_{j=0}^{n-1} T^{-j} I_{i_{j+1}}$ . Observe that

$$\begin{aligned} \log \left| \frac{\psi'_{i_1 \dots i_n}(x')}{\psi'_{i_1 \dots i_n}(x)} \right| &= \sum_{j=1}^n \log \left| \frac{\psi'_{i_j}(\psi_{i_{j+1} \dots i_n} x')}{\psi'_{i_j}(\psi_{i_{j+1} \dots i_n} x)} \right| \\ &= \sum_{j=1}^n \log \left| \frac{T'(\psi_{i_j \dots i_n} x')}{T'(\psi_{i_j \dots i_n} x)} \right| \\ &\leq \sum_{j=1}^n \log \left( 1 + D \frac{|x - x'|}{\beta^{n-j}} \right) \end{aligned}$$



where  $D$  bounds  $\frac{|T''|}{|T'|}$  on  $I$ . Then we have a constant  $C > 1$  such that

$$\frac{\sup_{x \in TI_{i_n}} |\psi'_{i_1 \dots i_n}(x)|}{\inf_{x \in TI_{i_n}} |\psi'_{i_1 \dots i_n}(x)|} \leq C, \quad \forall i_1, \dots, i_n, n > 0.$$

The property allows us to find a constant  $K < +\infty$  such that

$$\sum_{i_1, \dots, i_n} |\psi'_{i_1 \dots i_n}(x)| \chi_{TI_{i_n}}(x) \leq K, \quad \forall x.$$

We conclude that there exists  $D' > 0$  such that  $|\mathcal{L}^n 1(x) - \mathcal{L}^n 1(x')| \leq K \left( |x - y| \frac{D'}{1 - \beta} \right)$

(where none of the bounds on the right hand side depends on  $n$ ). We conclude that  $\forall n \geq 1$

- (1) the functions  $\mathcal{L}^n 1$  are bounded in the supremum norm,
- (2) the functions  $\mathcal{L}^n 1$  are an equicontinuous family.

We construct a new family of averages

$$F_n(x) = \frac{1}{n} \sum_{k=0}^{n-1} \mathcal{L}^k 1(x), \quad n \geq 0.$$

We again see that

- (1) the functions  $F_n$  are bounded in the supremum norm,
- (2) the functions  $F_n$  are an equicontinuous family.

By the Ascoli theorem, there must be a limit point  $F_{n_r} \rightarrow f$  ( $\geq 0$ ) in the continuous functions on each component of  $[0, 1] - \{x_0, \dots, x_k\}$  and since  $\int \mathcal{L}^n 1 dx = 1$  we have  $\int f d\lambda = \lim_{n \rightarrow \infty} \int F_{n_r} d\lambda = 1$ . Moreover, we see that

$$\begin{aligned} \mathcal{L}F_{n_r}(x) &= \sum_{y \in T^{-1}x} \frac{F_{n_r}(y)}{|T'(y)|} = \sum_{y \in T^{-1}x} \frac{1}{n_r} \sum_{k=0}^{n_r-1} \frac{\mathcal{L}^k 1(y)}{|T'(y)|} \\ &= \frac{1}{n_r} \sum_{k=0}^{n_r-1} \sum_{y \in T^{-1}x} \frac{\mathcal{L}^k 1(y)}{|T'(y)|} \\ &= \frac{1}{n_r} \sum_{k=0}^{n_r-1} \mathcal{L}^{k+1} 1(x) \\ &= F_{n_r}(x) - \frac{1}{n_r} (1(x) - \mathcal{L}^{n_r} 1(x)). \end{aligned}$$

Letting  $r \rightarrow +\infty$  we see that

$$\mathcal{L}f(x) = \sum_{y \in T^{-1}x} \frac{f(y)}{|T'(y)|} = f(x),$$

completing the proof. ■

DEFINITION. We say that  $T$  is aperiodic if there exists a positive number  $m$  such that  $\lambda(T^{-m}I_i \cap I_j) > 0, \forall i, j > 0$ .

THEOREM 12.5. *The absolutely continuous invariant measure  $\mu$  in Proposition 12.4 is exact if  $T : I \rightarrow I$  is aperiodic.*

PROOF. By Proposition 12.1 it suffices to show that for any set  $A \in \mathcal{B}$  with  $\mu(A) > 0$  and for which  $T^n A \in \mathcal{B}$  for all  $n \geq 0$  we have that  $\lim_{n \rightarrow +\infty} \mu(T^n A) = 1$ .

Given  $\underline{i} = (i_1, \dots, i_n)$  we write  $I_{\underline{i}} = \cap_{j=1}^n T^{-j+1}I_{i_j}$  if  $\text{int}(\cap_{j=1}^n T^{-j+1}I_{i_j}) \neq \emptyset$ . As  $T$  is piecewise invertible on each atom  $I_i$ , we know that  $T^n|_{I_{\underline{i}}}$  is a  $C^1$ -diffeomorphism. For all  $I_{\underline{i}}$  and for all  $n > 0$  we write  $(T^n|_{I_{\underline{i}}})^{-1} = \psi_{\underline{i}}$ . Let  $x, y \in T^n I_{\underline{i}} (= TI_{i_n})$ ; then it follows from the mean value theorem that

$$|\psi_{\underline{i}}(x) - \psi_{\underline{i}}(y)| = |\psi'_{\underline{i}}(\theta)| |x - y|$$

for some  $\theta \in I_{\underline{i}}$ . From the above equality and the condition (i) in page 39, the diameter  $\text{diam}(I_{\underline{i}})$  of  $I_{\underline{i}}$  decays exponentially fast (i.e.,  $\text{diam}(I_{\underline{i}}) \leq \frac{1}{\beta^n}$ ). This implies that the partition  $\mathcal{I} = \{I_i\}$  is a “generating partition”. In particular, for any  $\epsilon > 0$  we can choose a finite disjoint set of cylinders  $\{I_{\underline{j}} : \underline{j} = (j_1, \dots, j_l)\}$ , with  $\mu\left(\left(\cup_{\underline{j}} I_{\underline{j}}\right) \Delta A\right) < \epsilon$ . The following estimates will be useful in the rest of the proof.

- (a) Given  $\delta > 0$  there exists at least one cylinder  $I_{\underline{j}}$  (where  $\underline{j} = (j_1 \dots j_l)$ , say) for which

$$\lambda(A \cap I_{\underline{j}}) \geq (1 - \delta) \lambda(I_{\underline{j}}). \tag{12.1}$$

Assume for a contradiction that this is not the case, then for all cylinders  $I_{\underline{j}}$  we would have  $\lambda(A \cap I_{\underline{j}}) \leq (1 - \delta) \lambda(I_{\underline{j}})$ . We can extend this inequality to disjoint unions of cylinders, and then by approximation to arbitrary sets  $B \in \mathcal{B}$  to get  $\mu(A \cap B) \leq (1 - \delta) \mu(B)$ . However, if we take  $B = A$ , then we get  $\mu(A) \leq (1 - \delta) \mu(A)$ , which contradicts  $\mu(A) > 0$ .

- (b) We observe that there is a constant  $C \geq 1$  such that for any cylinder  $I_{\underline{i}}$

$$\sup_{x, y \in T^n I_{\underline{i}}} \frac{|\psi'_{\underline{i}}(x)|}{|\psi'_{\underline{i}}(y)|} \leq C. \tag{12.2}$$

(This is usually referred to as *Renyi's condition*.)

From the change of variables formula we see that

$$\begin{aligned}
\lambda(T^l I_{\underline{j}} \cap (T^l A)^c) &\leq \int_{I_{\underline{j}} \cap A^c} |(T^l)'(x)| d\lambda(x) \\
&\leq \left( \sup_{y \in I_{\underline{j}}} |(T^l)'(y)| \right) \lambda(I_{\underline{j}} \cap A^c) \\
&\leq C \left( \inf_{y \in I_{\underline{j}}} |(T^l)'(y)| \right) \lambda(I_{\underline{j}} \cap A^c) \quad (\text{using (12.2)}) \\
&\leq C \frac{\int_{I_{\underline{j}}} |(T^l)'(x)| d\lambda(x)}{\lambda(I_{\underline{j}})} \lambda(I_{\underline{j}} \cap A^c) \\
&\leq C \delta \lambda(T^l I_{\underline{j}}) \quad (\text{using (12.1)}).
\end{aligned}$$

If  $T^l(I_{\underline{j}}) = I$ , then we could proceed directly to the end of the proof. However, since this need not be the case, we require the following sublemma.

**SUBLEMMA 12.5.1.** *There exist  $S > 0$  and a subset  $I'$  of  $T^l(I_{\underline{j}})$  which is a finite disjoint union of elements of  $\bigvee_{i=0}^{S-1} T^{-i} \{I_1 \dots I_k\}$  and satisfies  $T^S(I') = I$ .*

**PROOF.** Let  $\{U_1, \dots, U_N\} = \{TI_1, \dots, TI_k\}$ , where  $N \leq k$ , denote the collection of images under  $T$  of the original intervals. The aperiodicity assumption implies that for each  $1 \leq j \leq N$  there exists  $0 < s_j < +\infty$  such that each  $U_i$ ,  $i = 1, \dots, N$ , contains a cylinder  $I_{m_1, \dots, m_{s_j}}^{(i,j)}$  satisfying  $T^{s_j} I_{m_1, \dots, m_{s_j}}^{(i,j)} = U_j$ . In particular, we see that  $T^{s_i} U_i \supset T^{s_i} I_{m_1, \dots, m_{s_j}}^{(i,j)} = U_i$ .

Let  $T^l(I_{\underline{j}}) = U_i$ . Setting  $S = \prod_{j=1}^N s_j$  and  $I' = \bigcup_{j=1}^N I_{m_1, \dots, m_{s_j}}^{(i,j)}$  allows us to have that  $I' \subset T^l(I_{\underline{j}})$  and  $T^S I' \supset \bigcup_{j=1}^N U_j = X$ . ■

We need only modify the previous argument to write

$$\lambda(T^S(I' \cap (T^l A)^c)) \leq D\delta$$

for some uniform constant  $D > 0$ . Since  $\lambda(T^S(I' \cap (T^l A)^c)) \geq 1 - \lambda(T^S(I' \cap T^l A))$ , we see that

$$\lambda(T^{l+S} A) > \lambda(T^S(I' \cap T^l A)) > 1 - D\delta.$$

Since  $\mu$  is absolutely continuous with respect to  $\lambda$  we conclude that  $\mu(T^n A) \rightarrow 1$  as  $n \rightarrow +\infty$ . ■

**COROLLARY 12.5.1.** *If  $T : I \rightarrow I$  is aperiodic, then it is strong-mixing with respect to any absolutely continuous invariant measure. In particular, there exists a unique absolutely continuous invariant probability measure.*

**PROOF.** By Proposition 12.1 the exact measure  $\mu$  is also strong mixing. By Proposition 11.2 it is also ergodic, and since no two distinct ergodic measures can be equivalent to Lebesgue measure (and thus each other) uniqueness follows.

**PROPOSITION 12.6.**  *$\mu$  is equivalent to  $\lambda$ .*

**PROOF.** First we show the following fact:

$$\forall \epsilon > 0, \exists N(\epsilon) > 0 \text{ such that for each } x \in I, T^{-N(\epsilon)}x \text{ is } \epsilon\text{-dense in } I. \quad (12.3)$$

As we have already observed in Theorem 12.5, for  $\forall I_{j_1 \dots j_l}$  there exist a set of cylinders  $\{I_{m_1 \dots m_{s_i}}^{(i)} : i = 1, \dots, N\}$  and  $S > 0$  satisfying  $T^S(\cup_{i=1}^N I_{m_1 \dots m_{s_i}}^{(i)}) = I$ . Let  $x \in I_{h_1 \dots h_t}$ . Then  $\exists i$  s.t.  $m_1 \dots m_{s_i} h_1 \dots h_t$  is an admissible sequence and so  $\psi_{m_1 \dots m_{s_i}}(x) \in I_{m_1 \dots m_{s_i}} \subset T^l I_{j_1 \dots j_l}$ . Hence we have that  $T^{-(l+s_i+t)}x \cap I_{j_1 \dots j_l} \neq \emptyset$ . Here we take  $t = S - s_i$ . Let  $l = l(\epsilon)$  be a positive integer such that  $\sup_{I_{j_1 \dots j_l}} \text{diam} I_{j_1 \dots j_l} < \epsilon$ . Then, each  $I_{j_1 \dots j_l}$  contains at least a point belonging to  $T^{-(l+S)}x$ . Choosing  $N(\epsilon) = l(\epsilon) - S$ , we have the fact (12.3).

It remains to show that  $f$  is bounded away from zero. Assume for a contradiction that  $f(x) = 0$ . Then since for all  $n \geq 1$ ,  $\mathcal{L}^n f(x) = \sum_{y \in T^{-n}x} \frac{f(y)}{|T^n'(y)|} = 0$ , we see that  $f(y) = 0$  whenever  $T^n y = x$ . By the property (12.3) the set of such points is dense. The continuity of  $f$  implies that  $f$  is identically zero, contradicting  $\int f d\lambda = 1$ . ■

**PROPOSITION 12.7.** *For irreducible piecewise expanding Markov maps  $T : I \rightarrow I$  the following condition is equivalent to strong mixing:*

$$\lambda \circ T^{-n}(A) \rightarrow \mu(A), \text{ as } n \rightarrow +\infty \quad (\forall A \in \mathcal{B}),$$

where  $\lambda$  is Lebesgue measure.

**PROOF.** It is enough to observe that

$$\begin{aligned} \lambda(T^{-n}A) &= \int_I \chi_{T^{-n}A}(x) d\lambda(x) = \int \chi_A(T^n x) f(x)^{-1} d\mu(x) \\ &\rightarrow \left( \int \chi_A(x) d(\mu(x)) \right) \cdot \left( \int d\lambda(x) \right) = \mu(A). \end{aligned}$$
■

**REMARK.** Under the generating condition we can extend these results to multi-dimensional piecewise expanding Markov maps with countable infinite partitions.

Since the invariant density  $f$  is strictly positive, we can make the following definition.

DEFINITION. We define an operator  $\hat{\mathcal{L}} : L^1(I) \rightarrow L^1(I)$  by  $\hat{\mathcal{L}}(h) = \frac{1}{f} \mathcal{L}(fh)$  where  $h \in L^1(X)$ .

PROPOSITION 12.8.  $\hat{\mathcal{L}}^*(\mu) = \mu$ , i.e. the dual operator  $\hat{\mathcal{L}}^*$  acting on measures (defined by  $(\mathcal{L}^*\mu)(A) = \int \mathcal{L}\chi_A d\mu$ ) fixes  $\mu$ .

PROOF. It is an immediate consequence of Sublemma 14.2.3 and the definition. ■

THEOREM 12.9 (CONVERGENCE TO INVARIANT DENSITY).  $\mathcal{L}^n(h) \rightarrow f \left( \int h d\lambda \right)$  uniformly for  $h \in C^0(I)$ .

PROOF. Define  $g = \frac{f(x)}{f(Tx)|T'(x)|}$ . From Renyi's condition we have that there exists a uniform constant  $D \geq 1$  such that  $\forall x, x' \in U_k$

$$D(x, x') = \sup_{n \geq 1} \sup_{y \in T^{-n}x, y' \in T^{-n}x'} \prod_{i=1}^{n-1} \frac{|g(T^i y)|}{|g(T^i y')|}$$

is bounded above by  $D$  and furthermore

$$D(x, x') \rightarrow 1 \text{ as } |x - x'| \rightarrow 0.$$

An easy calculation shows that  $\{\hat{\mathcal{L}}^n h : n \geq 0\}$  is equicontinuous on each component of  $I - \partial\mathcal{V}$  for  $\forall h \in C_0(I - \partial\mathcal{V})$ . It follows from the definition of  $\hat{\mathcal{L}}^n$  that  $\|\hat{\mathcal{L}}^n h\|_\infty$  is bounded by  $\|h\|_\infty$  and so the closure of  $\{\hat{\mathcal{L}}^n h : n \geq 0\}$  in  $C(I - \partial\mathcal{V})$  is compact. Hence there are a subsequence  $\{n_i\} \rightarrow \infty$  ( $i \rightarrow \infty$ ) and  $h^* \in C^0(I - \partial\mathcal{V})$  such that  $\hat{\mathcal{L}}^{n_i} h \rightarrow h^*$  uniformly.

We can now show that any limit point of the sequence is a constant which, in particular, shows that the limit exists. Notice that  $\min_{x \in I} (\hat{\mathcal{L}}^k h^*(x)) = \min_{x \in I} (h^*(x))$  for all  $k \geq 0$ . For any  $k \geq 0$  choose  $z \in I$  such that  $\hat{\mathcal{L}}^k h^*(z) = \min_{x \in I} h^*(x)$ . Then for all  $y \in T^{-k}z$  we have that  $h^*(y) = \min_{x \in I} h^*(x)$ . In fact,

$$\hat{\mathcal{L}}^k h^*(z) = \sum_{T^k y = z} (g(y) \dots g(T^{k-1}y)) h^*(y) \geq \min_{x \in I} h^*(x)$$

with equality if and only if  $h^*(y) = \min_{x \in I} h^*(x)$ ,  $\forall y \in T^{-k}z$ . By (12.3) we see that the set of  $y$  such that  $\exists k \geq 1$  with  $T^k y = z$  is dense. Thus  $h^*$  is a constant function with value  $\min_{x \in I} h^*(x)$  on a dense set, and thus by piecewise continuity is constant almost everywhere.

Moreover, this constant takes the value  $\lim_{n \rightarrow +\infty} \int \hat{\mathcal{L}}^n h d\mu = \int h d\mu$ . Replacing  $h$  by  $\frac{h}{f}$  for  $h \in C^0(I)$  and appealing to the definition of  $\hat{\mathcal{L}}$  we get that

$$\mathcal{L}^n(h) = \mathcal{L}^n\left(f \cdot \frac{h}{f}\right) = f \cdot \hat{\mathcal{L}}^n\left(\frac{h}{f}\right) \rightarrow f \left( \int \frac{h}{f} \cdot f d\lambda \right) = f \cdot \left( \int h d\lambda \right)$$

uniformly as  $n \rightarrow +\infty$ . ■

### 12.3 Rohlin's entropy formula

In this section we want to give a formula for the entropy of an irreducible Markov piecewise expanding interval map  $T : I \rightarrow I$  with respect to the unique absolutely continuous probability measure  $\mu$ .

**THEOREM 12.10 (ROHLIN ENTROPY FORMULA).**

$$h_\mu(T) = \int \log |T'(T^i x)| d\mu(x).$$

**PROOF.** The proof follows immediately from the string of statements (i)-(iv) below.

- (i) By the chain rule we can write  $\log |(T^N)'(x)| = \sum_{i=0}^{N-1} \log |T'(T^i x)|$  for each  $x \in I$ ,  $N \geq 1$ . Since the measure  $\mu$  is ergodic (even exact) we can apply the Birkhoff ergodic theorem to deduce that

$$\frac{1}{N} \log |(T^N)'(x)| \rightarrow \int \log |T'(x)| d\mu(x) \text{ as } N \rightarrow +\infty.$$

- (ii) Let  $x \in I_{i_1, \dots, i_N} = \cap_{j=1}^N T^{-(j-1)} I_{i_j}$ ; then using Renyi's condition we can estimate

$$\begin{aligned} \lambda(I_{i_1, \dots, i_N}) &= \int_{T^N I_{i_1, \dots, i_N}} \frac{1}{|(T^N)'(\psi_{i_1, \dots, i_N} z)|} d\lambda(z) \\ &\leq C \left( \inf_{x \in I_{i_1, \dots, i_N}} \frac{1}{|(T^N)'(x)|} \right) \lambda(T I_{i_N}) \\ &\leq C \left( \frac{1}{|(T^N)'(x)|} \right) \end{aligned}$$

and

$$\begin{aligned} \lambda(I_{i_1, \dots, i_N}) &\geq \frac{1}{C} \left( \sup_{z \in I_{i_1, \dots, i_N}} \frac{1}{|(T^N)'(z)|} \right) \lambda(T(I_{i_N})) \\ &\geq \frac{1}{C} (\min_{1 \leq i \leq k} \lambda(I_i)) \frac{1}{|(T^N)'(x)|}. \end{aligned}$$

Thus we see that for any  $x \in I$

$$- \lim_{N \rightarrow +\infty} \frac{1}{N} \log \lambda(I_{i_1, \dots, i_N}) = \lim_{N \rightarrow +\infty} \frac{1}{N} \log |(T^N)'(x)|$$

(where  $x \in I_{i_1 \dots i_N}$ ).

- (iii) Since the density  $f$  of the invariant measure is bounded from below and away from infinity, we see that

$$-\lim_{N \rightarrow +\infty} \frac{1}{N} \log \lambda(I_{i_1 \dots i_N}) = -\lim_{N \rightarrow +\infty} \frac{1}{N} \log \mu(I_{i_1 \dots i_N}).$$

- (iv) Finally, we claim that

$$-\lim_{N \rightarrow +\infty} \frac{1}{N} \log \mu(I_{i_1 \dots i_N}) = h_\mu(T).$$

This is an application of the Shannon-McMillan-Brieman theorem to interval maps, whose proof we present in the next section. ■

## 12.4 The Shannon-McMillan-Brieman theorem

We now give an application of entropy to describe the asymptotic size of elements in partitions.

Let  $\alpha = \{A_1, A_2, \dots\}$  be a measurable partition of the space  $(X, \mathcal{B})$ , i.e.  $X = \cup_{i=1}^n A_i$  and  $A_i \cap A_j = \emptyset$  for  $i \neq j$  (up to a set of zero  $\mu$ -measure).

For each  $n \geq 1$  we consider the new partition  $\alpha_n = \vee_{i=0}^{n-1} T^{-i} \alpha$ . For almost all  $x \in X$  we can choose a unique element  $A_n(x) \in \alpha_n$  with  $x \in A_n(x)$ .

**THEOREM 12.11 (SHANNON-MCMILLAN-BREIMAN THEOREM).** *Let  $T : X \rightarrow X$  be a measure preserving transformation of a probability space  $(X, \mathcal{B}, \mu)$ . Let  $\alpha$  be a partition. For almost all  $x \in X$  we have that*

$$-\frac{\log \mu(A_n(x))}{n} \rightarrow E(f|\mathcal{I})(x)$$

as  $n \rightarrow +\infty$ , where  $f(x) = I(\alpha | \vee_{n=1}^{\infty} T^{-n} \alpha)(x)$  and  $\mathcal{I}$  is the sigma-algebra generated by the  $T$ -invariant sets  $T^{-1}B = B$ .

**COROLLARY 12.11.1.** *If  $T$  is ergodic then for almost all  $x \in X$*

$$-\frac{\log \mu(A_n(x))}{n} \rightarrow h(T, \alpha) \text{ as } n \rightarrow +\infty.$$

*If  $\alpha$  is a generating partition then*

$$-\frac{\log \mu(A_n(x))}{n} \rightarrow h_\mu(T) \text{ as } n \rightarrow +\infty.$$

PROOF. Assuming the theorem, the ergodicity of the measure and the  $T$ -invariance of the limit imply that it is a constant. Integrating therefore gives that the limit is

$$\int E(f|\mathcal{I})d\mu = \int fd\mu = H(\alpha|\bigvee_{n=1}^{\infty} T^{-n}\alpha) = h(T, \alpha).$$

PROOF OF THEOREM 12.11. We first observe that

$$I(\bigvee_{i=0}^{n-1} T^{-i}\alpha)(x) = -\log \mu(A_n(x))$$

Using the basic identities for the information function we see that

$$\begin{aligned} & I(\bigvee_{i=0}^{n-1} T^{-i}\alpha) \\ &= I(\alpha|\bigvee_{i=1}^{n-1} T^{-i}\alpha) + I(\bigvee_{i=1}^{n-1} T^{-i}\alpha) \\ &= I(\alpha|\bigvee_{i=1}^{n-1} T^{-i}\alpha) + I(\alpha|\bigvee_{i=1}^{n-2} T^{-i}\alpha)T \\ &\quad + \dots + I(\alpha|T^{-1}\alpha)T^{n-2} + I(\alpha)T^{n-1}. \end{aligned} \tag{12.4}$$

We see from (12.4) that (almost everywhere)

$$\begin{aligned} & \limsup_{n \rightarrow +\infty} \frac{1}{n} |I(\bigvee_{i=0}^{n-1} T^{-i}\alpha) - E(f|\mathcal{I})| \\ & \leq \limsup_{n \rightarrow +\infty} \frac{1}{n} |I(\bigvee_{i=0}^{n-1} T^{-i}\alpha) - \sum_{i=0}^{n-1} fT^i| \\ & \quad + \limsup_{n \rightarrow +\infty} \left| \frac{1}{n} \sum_{i=0}^{n-1} fT^i - E(f|\mathcal{I}) \right| \end{aligned} \tag{12.5}$$

(using the triangle inequality). By the Birkhoff ergodic theorem (Theorem 10.6) we know that

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \left| \sum_{i=0}^{n-1} fT^i - E(f|\mathcal{I}) \right| = 0$$

(almost everywhere) and thus the second term on the right hand side of (12.5) vanishes.

We can next write from (12.5) that

$$\begin{aligned} & \frac{1}{n} \left| I(\bigvee_{i=0}^{n-1} T^{-i}\alpha) - \sum_{i=0}^{n-1} fT^i \right| \\ & \leq \frac{1}{n} \sum_{i=0}^{n-1} |I(\alpha|\bigvee_{j=1}^{n-1-i} T^{-j}\alpha)T^i - I(\alpha|\bigvee_{j=1}^{\infty} T^{-j}\alpha)T^i| \end{aligned}$$



(using also the definition of  $f$ ). For  $N \geq 1$  we define

$$F_N(x) := \sup_{N \leq i \leq n} |I(\alpha | \bigvee_{j=1}^{n-i} T^{-j} \alpha)(x) - I(\alpha | \bigvee_{j=1}^{\infty} T^{-j} \alpha)(x)|$$

and then upon fixing  $N \geq 1$  we see that

$$\begin{aligned} & \frac{1}{n} \left| I(\bigvee_{i=0}^{n-1} T^{-i} \alpha) - \sum_{i=0}^{n-1} f T^i \right| \\ & \leq \left( \frac{F_N T^n + F_N T^{n-1} + \dots + F_N T^{n-N}}{n} \right) \\ & \quad + \left( \frac{\sum_{i=0}^{N-1} |I(\alpha | \bigvee_{j=1}^{n-i} T^{-j} \alpha) T^i - I(\alpha | \bigvee_{j=1}^{\infty} T^{-j} \alpha) T^i|}{n} \right) \end{aligned} \quad (12.6)$$

We can bound the second term on the right hand side of (12.6) by

$$\begin{aligned} & \frac{1}{n} \left| \sum_{i=0}^{N-1} |I(\alpha | \bigvee_{j=1}^{n-i} T^{-j} \alpha) T^i - I(\alpha | \bigvee_{j=1}^{\infty} T^{-j} \alpha) T^i| \right| \\ & \leq \frac{N}{n} \left( \sup_{k \geq N} (I(\alpha | \bigvee_{j=1}^{\infty} T^{-j} \alpha) + I(\alpha | \bigvee_{j=1}^k T^{-j} \alpha)) \right) \end{aligned}$$

which tends to 0 (almost everywhere) as  $n \rightarrow +\infty$ .

We now turn to the first term on the right hand side of (12.6). We observe that by the Birkhoff ergodic theorem

$$\limsup_{n \rightarrow +\infty} \left( \frac{F_N T^n + F_N T^{n-1} + \dots + F_N T^{n-N}}{n} \right) = E(F_N | \mathcal{I}).$$

Notice that  $F_N \geq F_{N+1}$  and so

$$E(F_N | \mathcal{I}) \geq E(F_{N+1} | \mathcal{I}) \geq 0$$

(since  $E(\cdot | \mathcal{I})$  is a positive operator). Since  $E(F_N | \mathcal{I}) \rightarrow 0$  (and is dominated by an integrable function) then

$$\lim_{N \rightarrow +\infty} \int E(F_N | \mathcal{I}) d\mu = \lim_{N \rightarrow +\infty} \int F_N d\mu = 0.$$

This completes the proof. ■

### 12.5 Comments and references

A good reference for more information on exactness is Rohlin's original paper [2].

Without the Markov assumption (but still assuming the uniform expansion property) the existence of an absolutely continuous invariant measure follows from the work of Lasota and Yorke [1].

There is an alternative proof of the Shannon-McMillan-Brieman theorem given in [3, 5.2]

#### References

1. A. Lasota and J. Yorke, *On the existence of invariant measures for piecewise monotonic transformations*, Trans. Amer. Math. Soc. **86** (1973), 481-488.
2. V. Rohlin, *Exact endomorphisms of lebesgue space*, Amer. Math. Soc. Transl. (2) **39** (1964), 1-36.
3. D. Rudolph, *Ergodic Theory on Lebesgue spaces*, O.U.P., Oxford, 1994.

**FIXED POINTS FOR  
HOMEOMORPHISMS OF THE ANNULUS**

**13.1 Fixed points for the annulus**

Let  $A = \mathbb{R}/\mathbb{Z} \times [0, 1]$  be a closed annulus. Assume that  $T : A \rightarrow A$  is a homeomorphism that preserves the two boundary circles (i.e.  $T(\mathbb{R}/\mathbb{Z} \times \{0\}) = \mathbb{R}/\mathbb{Z} \times \{0\}$  and  $T(\mathbb{R}/\mathbb{Z} \times \{1\}) = \mathbb{R}/\mathbb{Z} \times \{1\}$ ).

**DEFINITION.** We say that  $T : A \rightarrow A$  is *area preserving* if the Haar-Lebesgue measure  $\lambda$  is  $T$ -invariant.

**REMARK.** We say that  $T : A \rightarrow A$  is *conservative* if there are no wandering sets of positive measure. This condition will suffice for most of the results of this section.

**EXAMPLE.** For any pair of values  $0 < \alpha, \beta < 1$  consider the map  $T : A \rightarrow A$  given by  $T(x, y) = (x + \alpha y + \beta(1 - y), y)$ . To see that this is area preserving we can write this affine transformation as  $T(x, y) = (\beta, 0) + B(x, y)$  where  $B = \begin{pmatrix} 1 & \alpha - \beta \\ 0 & 1 \end{pmatrix}$ . Since  $\det(B) = 1$  we see that  $T$  is area preserving.

**DEFINITION.** An  $\epsilon$ -chain (for  $T : A \rightarrow A$ ) from  $(x, y)$  to  $(w, z)$  is a sequence of points

$$(x, y) = (x_0, y_0), (x_1, y_1), \dots, (x_n, y_n) = (w, z) \in A$$

such that

$$d(T(x_i, y_i), (x_{i+1}, y_{i+1})) < \epsilon \text{ for } i = 0, \dots, n - 1.$$

We can use the same notation for  $\epsilon$ -chains for the lift  $\hat{T} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \times [0, 1]$ .

These are finite versions of the pseudo-orbits introduced in chapter 5.

**LEMMA 13.1.** *If  $(x, y) = (x_0, y_0), (x_1, y_1), \dots, (x_n, y_n) = (w, z)$  is an  $\epsilon$ -chain from  $(x, y)$  to  $(w, z)$  and  $(w, z) = (w_0, z_0), (w_1, z_1), \dots, (w_m, z_m) = (u, v)$  is an  $\epsilon$ -chain from  $(w, z)$  to  $(u, v)$  then defining  $(x_{n+i}, y_{n+i}) = (w_i, z_i)$  for  $0 \leq i \leq m$  makes  $(x_0, y_0), \dots, (x_n, y_n), (x_{n+1}, y_{n+1}), \dots, (x_{n+m}, y_{n+m})$  an  $\epsilon$ -chain from  $(x, y)$  to  $(u, v)$ .*

PROOF. This is immediate from the definitions. ■

DEFINITION. We say that a point  $(x, y)$  is *chain recurrent* if for each  $\epsilon > 0$  we can find an  $\epsilon$ -pseudo-orbit from  $(x, y)$  to itself.

We say that  $T : A \rightarrow A$  is *chain recurrent* if for every  $(x, y), (w, z) \in A$  and every  $\epsilon > 0$  we can find a finite  $\epsilon$ -pseudo-orbit from  $(x, y)$  to  $(w, z)$ .

LEMMA 13.2. *Let  $T : A \rightarrow A$  be an area preserving homeomorphism; then*

- (i) *every point  $(x, y) \in A$  is chain recurrent,*
- (ii)  *$T : A \rightarrow A$  is chain recurrent.*

PROOF. (i) Fix  $\epsilon > 0$  and then by (uniform) continuity we can choose  $\frac{\epsilon}{2} > \delta > 0$  such that whenever  $|(x, y) - (u, v)| < \delta$  then  $|T(x, y) - T(u, v)| < \frac{\epsilon}{2}$ .

Let us choose a finite cover of  $\delta$ -balls

$$A \subset \cup_{i=1}^N B((z_i, w_i), \delta) \text{ where } (z_1, w_1), \dots, (z_N, w_N) \in A.$$

Since  $T$  is area preserving we have for each  $i = 1, \dots, N$  that we can choose  $n_i$  such that  $T^{-n_i} B((z_i, w_i), \delta) \cap B((z_i, w_i), \delta) \neq \emptyset$ . If we choose  $(u_i, v_i)$  in this intersection then

$$\left\{ \begin{array}{l} (x_0, y_0) = (u_i, v_i), \\ (x_1, y_1) = T(u_i, v_i), \\ \vdots \\ (x_j, y_j) = T^j(u_i, v_i), \\ \vdots \\ (x_{n_i-1}, y_{n_i-1}) = T^{(n_i-1)}(u_i, v_i), \\ (x_{n_i}, y_{n_i}) = (u_i, v_i) \end{array} \right.$$

gives an  $\frac{\epsilon}{2}$ -chain from  $(z_i, w_i)$  to itself since we observe that

- (a)  $|T((x_0, y_0)) - T(z_i, w_i)| < \frac{\epsilon}{2}$  (since  $|(x_0, y_0) - (z_i, w_i)| = |(u_i, v_i) - (z_i, w_i)| < \delta$ ),
- (b)  $T(x_j, y_j) = T(T^j(u_i, v_j)) = T^{j+1}(u_j, v_j)$  for  $j = 1, \dots, n_i - 2$ ; and
- (c)  $T(x_{n_i-1}, y_{n_i-1}) = T^{n_i}(u_j, v_j) \in B((z_i, w_i), \delta) \subset B((z_i, w_i), \frac{\epsilon}{2})$ .

For any  $(z, w) \in A$  we can choose some  $(z_i, w_i)$  ( $i = 1, \dots, N$ ) which is  $\delta$ -close to  $(z, w)$  (i.e.  $|(z_i, w_i) - (z, w)| < \delta$ ). We then see that the above  $\frac{\epsilon}{2}$ -chain from  $(z_i, w_i)$  to itself also serves as an  $\epsilon$ -chain from  $(z, w)$  to itself on replacing both  $(x_0, y_0)$  and  $(x_{n_i}, y_{n_i})$  by  $(z, w)$ . To see this observe that:

- (d)  $|T(z, w) - (x_1, y_1)| \leq |T(z, w) - T(z_i, w_i)| + |T(z_i, w_i) - T(x_1, y_1)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$  (since  $|(z_i, w_i) - (z, w)| < \delta$  implies that  $|T(z_i, w_i) - T(z, w)| < \frac{\epsilon}{2}$ ),
- (e)  $|T(x_{n_i}, y_{n_i}) - (z, w)| \leq |T(x_{n_i}, y_{n_i}) - (z_i, w_i)| + |(z_i, w_i) - (z, w)| \leq \frac{\epsilon}{2} + \delta < \epsilon$ .

(ii) We may choose a sequence  $(z_{i_0}, w_{i_0}), (z_{i_1}, w_{i_1}), \dots, (z_{i_n}, w_{i_n})$  such that

$$\begin{cases} (x, y) \in B((z_{i_0}, w_{i_0}), \delta), \\ B((z_{i_j}, w_{i_j}), \delta) \cap B((z_{i_{j+1}}, w_{i_{j+1}}), \delta) \neq \emptyset \text{ for } 0 \leq j \leq n-1, \\ (u, v) \in B((z_{i_n}, w_{i_n}), \delta). \end{cases}$$

If we write in succession the  $\epsilon$ -pseudo-orbit from  $(z_{i_0}, w_{i_0})$  to  $(z_{i_1}, w_{i_1})$ , and then the  $\frac{\epsilon}{2}$ -pseudo-orbit from  $(z_{i_1}, w_{i_1})$  to  $(z_{i_2}, w_{i_2})$ , etc. until we get to the  $\frac{\epsilon}{2}$ -pseudo-orbit from  $(z_{i_n}, w_{i_n})$  to  $(z_{i_n}, w_{i_n})$  then resulting concatenated sequence is an  $\epsilon$ -chain from  $(x, y)$  to  $(u, v)$  (cf. Lemma 13.1). ■

The following result tells us that there is no distinction between the existence of fixed points and periodic points for the map  $\hat{T} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \times [0, 1]$ .

**BROUWER PLANE THEOREM.** *If  $\hat{T} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \times [0, 1]$  has a periodic point (i.e.  $\exists n \neq 1, \hat{T}^n(x, y) = (x, y)$ ) then  $\hat{T}$  has a fixed point (i.e.  $\exists T(u, v) = (u, v)$ ).*

This is a classical result. We sketch the proof in the final section. (For detailed proofs we refer to [2], [1].)

**THEOREM 13.3 (POINCARÉ-BIRKHOFF).** *Assume that  $T : A \rightarrow A$  is an area preserving homeomorphism and that the rotation numbers  $\rho_0$  and  $\rho_1$  for  $T : \mathbb{R}/\mathbb{Z} \times \{0\} \rightarrow \mathbb{R}/\mathbb{Z} \times \{0\}$  and  $T : \mathbb{R}/\mathbb{Z} \times \{0\} \rightarrow \mathbb{R}/\mathbb{Z} \times \{0\}$ , respectively, satisfy either  $\rho_0 < 0 < \rho_1$  or  $\rho_1 < 0 < \rho_0$ . Then there exists a fixed point for  $T$ .*

**REMARK.** This result is also known as *Poincaré's last geometric theorem*. Usually the statement involves the existence of two distinct fixed points.

Theorem 13.3 is a special case of the following more general result.

**THEOREM 13.4 (FRANKS).** *Assume that  $T : A \rightarrow A$  is chain recurrent and that the rotation numbers  $\rho_0$  and  $\rho_1$  for  $T : \mathbb{R}/\mathbb{Z} \times \{0\} \rightarrow \mathbb{R}/\mathbb{Z} \times \{0\}$  and  $T : \mathbb{R}/\mathbb{Z} \times \{0\} \rightarrow \mathbb{R}/\mathbb{Z} \times \{0\}$ , respectively, satisfy either  $\rho_0 < 0 < \rho_1$  or  $\rho_1 < 0 < \rho_0$ . Then there exists a fixed point for  $T$ .*

**PROOF OF THEOREM 13.3 (ASSUMING THEOREM 13.4).** By Lemma 13.2 the hypothesis that  $T$  is area preserving implies that  $T$  is chain recurrent. The results follows immediately from Theorem 13.4. ■

It remains to prove Theorem 13.4.

**PROOF OF THEOREM 13.4.** The proof of the theorem will be conveniently divided into the following sublemmas.

SUBLEMMA 13.4.1. *Let  $\hat{T} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \times [0, 1]$  be the lift of a chain recurrent homeomorphism  $T : A \rightarrow A$ . There are four possibilities:*

- (i)  $\hat{T} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \times [0, 1]$  has a chain recurrent point,
- (ii) all points move to the right (i.e.  $\forall (x, y) \in \mathbb{R} \times [0, 1]$  if  $(x^{(n)}, y^{(n)}) := \hat{T}^n(x, y)$  then  $\lim_{n \rightarrow +\infty} x^{(n)} = +\infty$ ),
- (iii) all points move to the left (i.e.  $\forall (x, y) \in \mathbb{R} \times [0, 1]$  we have  $\lim_{n \rightarrow +\infty} x^{(n)} = -\infty$ ), or
- (iv)  $\forall M > 0, \exists (x, y), (w, z) \in \mathbb{R} \times [0, 1] \exists n, m \geq 1$  with  $(x^{(n)} - x^{(m)}) < -M$  and  $(w^{(n)} - w) > M$ .

The following result shows that case (iv) is actually redundant.

SUBLEMMA 13.4.2. *Case (iv) implies case (i).*

SUBLEMMA 13.4.3. *Let  $\hat{T} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \times [0, 1]$  be a lift of  $T : A \rightarrow A$ . Assume that  $\exists (x, y) \in \mathbb{R} \times [0, 1], \forall \epsilon > 0, \exists (x_i, y_i), i = 0, \dots, n$ , such that  $(x_0, y_0) = (x_n, y_n) = (x, y)$  and  $|\hat{T}(x_i, y_i) - (x_{i+1}, y_{i+1})| < \epsilon$  for  $i = 0, \dots, n-1$ . Then there exists a fixed point for  $\hat{T} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \times [0, 1]$ .*

Assuming these sublemmas the proof of Theorem 13.4 is now a simple matter. By the area preserving hypothesis  $T : A \rightarrow A$  is chain recurrent and Sublemma 13.4.1 applies.

By the hypotheses on the rotation numbers of the boundaries, points on the two boundaries move in opposite directions. Thus we see that cases (ii) and (iii) are eliminated. If (i) holds then  $T$  has a chain recurrent point. If (iv) holds than by Sublemma 13.4.2 this again leads to the same conclusion, that there exists a chain recurrent point.

Finally, by Sublemma 13.4.3 the existence of a chain recurrent point implies the existence of a fixed point. ■

We are only left with the chore of proving the sublemmas.

PROOF OF SUBLEMMA 13.4.1. Let us assume that (iv) fails, then to prove the sublemma we need to show that either (i), (ii) or (iii) holds.

Let us assume that (iv) fails because  $\exists M > 0, \forall (x, y) \in \mathbb{R} \times [0, 1], \forall n \geq 1$  we have that  $x^{(n)} - x \geq -M$ . We have two possibilities.

Firstly, if for some  $(x, y) \in \mathbb{R} \times [0, 1]$  we have that the sequence  $x^{(n)} - x, n \geq 1$ , is also bounded above then the sequence  $(\hat{T}^i(x, y))_{i=0}^{\infty}$  is confined to a bounded region of  $\mathbb{R} \times [0, 1]$  and so must have an accumulation point  $(x^*, y^*)$ , say. However, for any  $\delta > 0$  we need only choose  $n' > n \geq 1$  with  $|\hat{T}^{n'}(x, y) - (x^*, y^*)| < \frac{\delta}{2}$  and  $|\hat{T}^n(x, y) - (x^*, y^*)| < \frac{\delta}{2}$  and then the sequence  $((x^{(i)}, y^{(i)}))_{i=n}^{n'}$  is a  $\delta$ -chain from  $(x^*, y^*)$  back to  $(x^*, y^*)$ . Thus  $(x^*, y^*)$  is a chain recurrent point and we are in case (i).

The second possibility is that  $\forall (x, y) \in \mathbb{R} \times [0, 1]$  the sequence  $x^{(n)} - x \geq -M$  ( $n \geq 0$ ) is unbounded (i.e.  $\hat{T}^n$  may move points arbitrarily far to the

right, but never to the left). In particular, for any  $C > 0$  we can choose  $N$  with  $x^{(N)} - x \geq C$ . Thus if  $n \geq N$  then  $x^{(n)} - x = (x^{(N)})^{(n-N)} - x^{(N)} + (x^{(N)} - x) \geq -M + C$ . i.e.  $\lim_{n \rightarrow +\infty} x^{(n)} = +\infty$  (thus we are in case (ii)).

If we had assumed that (iv) failed because the second condition in (iv) was not met then a similar argument would have given that we are in either case (i) or case (iii). ■

**PROOF OF SUBLEMMA 13.4.2.** We need a preliminary observation. Consider any two points  $(u, v), (s, t) \in A$ ; then by Lemma 13.2 (i) we can find an  $\epsilon$ -chain  $(u_i, v_i)_{i=0}^n$  from  $(u, v)$  to  $(s, t)$ . Lifting this chain to  $\mathbb{R} \times [0, 1]$  we get that there is an  $\epsilon$ -chain in  $\mathbb{R} \times [0, 1]$  from  $(u, v)$  to  $(s + r, t)$ , say, for some  $r \in \mathbb{Z}$ . In addition,  $n$  can be bounded above by a bound  $D$ , say, depending only on  $\epsilon$  and not on the choice of  $(u, v)$  and  $(s, t)$ .

Returning to the proof of Sublemma 13.4.2, assuming property (iv) let us take  $M > 4D$ , then let  $(x, y), (w, z) \in \mathbb{R} \times [0, 1]$  be the two points described in its statement.

- (a) Given any point  $(u, v) \in \mathbb{R} \times [0, 1]$  we can construct an  $\epsilon$ -chain from  $(u, v)$  to  $(x + r_1, y)$ , for some  $r_1 \in \mathbb{Z}$ , by the above observation (with  $|r_1| \leq D$ ).
- (b) We can construct an  $\epsilon$ -chain in  $\mathbb{R} \times [0, 1]$  from  $(x + r_1, y)$  to  $(x^{(n)} + r_2, y^{(n)})$ , for some  $r_2 \in \mathbb{Z}$ , by taking the lift of the orbit sequence  $(x^{(i)}, y^{(i)})_{i=0}^n$ . By hypotheses,  $r_2 \geq 4D$ .
- (c) We can construct an  $\epsilon$ -chain in  $\mathbb{R} \times [0, 1]$  from  $(x^{(n)} + r_2, y^{(n)})$  to  $(u + r_2, v)$ , for some  $r \in \mathbb{Z}$ , by the above observation (with  $|r_2| \leq D$ ).

Thus by Lemma 13.1 we can concatenate these to get an  $\epsilon$ -chain from  $(u, v)$  to  $(u, v) + (r, 0)$  with  $r > 2D$

A similar argument (using the second part of property (iv)) shows that there is an  $\epsilon$ -chain from  $z$  to  $z - (s, 0)$ , say, for some  $s \in \mathbb{Z}$ .

If  $s = r$  then we can use Lemma 13.1 to combine the  $\epsilon$ -chain from  $z$  to  $z + (r, 0)$  with the  $\epsilon$ -chain from  $z + (r, 0)$  to  $z + (r - s, 0) = z$  to get an  $\epsilon$ -chain from  $z$  to itself. If  $r \neq s$ , we can repeat  $s$  times the  $\epsilon$ -chain (applying Lemma 13.1 repeatedly) from  $z$  to  $z + (r, 0)$  (to get an  $\epsilon$ -chain from  $z$  to  $z + (rs, 0)$ ) followed by  $r$  times the  $\epsilon$ -chain from  $z$  to  $z - (s, 0)$  (applying Lemma 13.1 repeatedly) to get from  $z + (rs, 0)$  to  $z$ . ■

**PROOF OF SUBLEMMA 13.4.3.** We shall first show that for each  $\delta > 0$  we can find a homeomorphism  $S : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \times [0, 1]$  such that  $S$  has a fixed point and  $\sup_{(x,y) \in \mathbb{R} \times [0,1]} |\hat{T}(x, y) - S(x, y)| < \frac{\delta}{2}$ .

By hypothesis, we can choose a  $\frac{\delta}{4}$ -pseudo-orbit  $(x_i, y_i)$ ,  $i = 0, \dots, N$ , from  $(x, y)$  to itself. We introduce a homeomorphism  $h : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \times [0, 1]$  such that  $h(\hat{T}(x_i, y_i)) = (x_{i+1}, y_{i+1})$  and  $\sup_{(x,y) \in \mathbb{R} \times [0,1]} |h(x, y) - (x, y)| < \frac{\delta}{2}$ .

(Intuitively, this seems easy, although in practice it is harder to write down details.)

If we define  $g(x, y) = (h \circ \hat{T})(x, y)$  then we can arrange that

$$\sup_{(x,y) \in \mathbb{R} \times [0,1]} |\hat{T}(x, y) - S(x, y)| < \frac{\delta}{2} \text{ and } S^n(x, y) = (x, y).$$

Thus there exists a periodic point for  $S$  and therefore by Brouwer's theorem there is a fixed point for  $S$ .

We observe that if we assume for a contradiction that  $\hat{T}$  did *not* have a fixed point then (by compactness of  $A$ ) the same would be true for any sufficiently close homeomorphism  $S$ . This contradicts the above construction.

Finally, this fixed point for  $\hat{T}$  projects to a fixed point for  $T : A \rightarrow A$ .

### 13.2 Outline proof of Brouwer's theorem

In the previous section we made use of a classical (but not standard) result of Brouwer. In this section we shall outline the main ideas in the proof.

OUTLINE PROOF. The proof has two distinct parts:

- (i) If  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is a homeomorphism with a periodic point of prime period  $n \geq 3$  then there exists a homeomorphism  $T' : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  with either a fixed point or a periodic point of prime period at most 2 *and the two homeomorphisms have the same set of fixed points.*
- (ii) If  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  has a periodic point of prime period 2 then there exists a fixed point.

Part (i) is proved by an iterative method. Specifically, If  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is a homeomorphism with a periodic point  $T^n x = x$  of prime period  $n \geq 3$  then one shows there exists a homeomorphism  $T' : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  with a periodic point of prime period at most  $n - 1$  and the two homeomorphisms have the same set of fixed points.

To see this, consider the family of balls  $B(x, \epsilon)$  about  $x$ , and their images  $T(B(x, \epsilon))$  as neighbourhoods of  $Tx$  (cf. Figure 13.1). We choose the *smallest*  $\epsilon > 0$  such that  $\text{cl}(B(x, \epsilon)) \cap T(\text{cl}(B(x, \epsilon))) \neq \emptyset$ . We can choose a point  $z \in \text{cl}(B(x, \epsilon)) \cap T(\text{cl}(B(x, \epsilon)))$  and a path  $\gamma$  in  $B(x, \epsilon)$  from  $T^{-1}z$  to  $z$  (passing through  $x$ ). By construction, this path  $\gamma$  has the property that  $\gamma \cap T(\gamma) = \emptyset$ . Since  $x \in \gamma$  we see that  $T^n(\gamma) \cap \gamma \neq \emptyset$  and so the "first" intersection  $y \in T^k(\gamma) \cap \gamma$  ( $n \geq k \geq 2$ ) gives rise to a simple closed curve containing  $T(\gamma) \cup T^2(\gamma) \cup \dots \cup T^{n-1}(\gamma)$ . The homeomorphism  $T$  can be changed in a continuous way, or *isotoped* (but only in a small neighbourhood of this closed curve), to  $T' : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  so that  $z$  becomes a point of period  $k - 1$  for  $T'$ .

Furthermore, since  $T$  can have no fixed points on the simple closed curve then it has some neighbourhood  $U$  in which the same is true. If we arrange



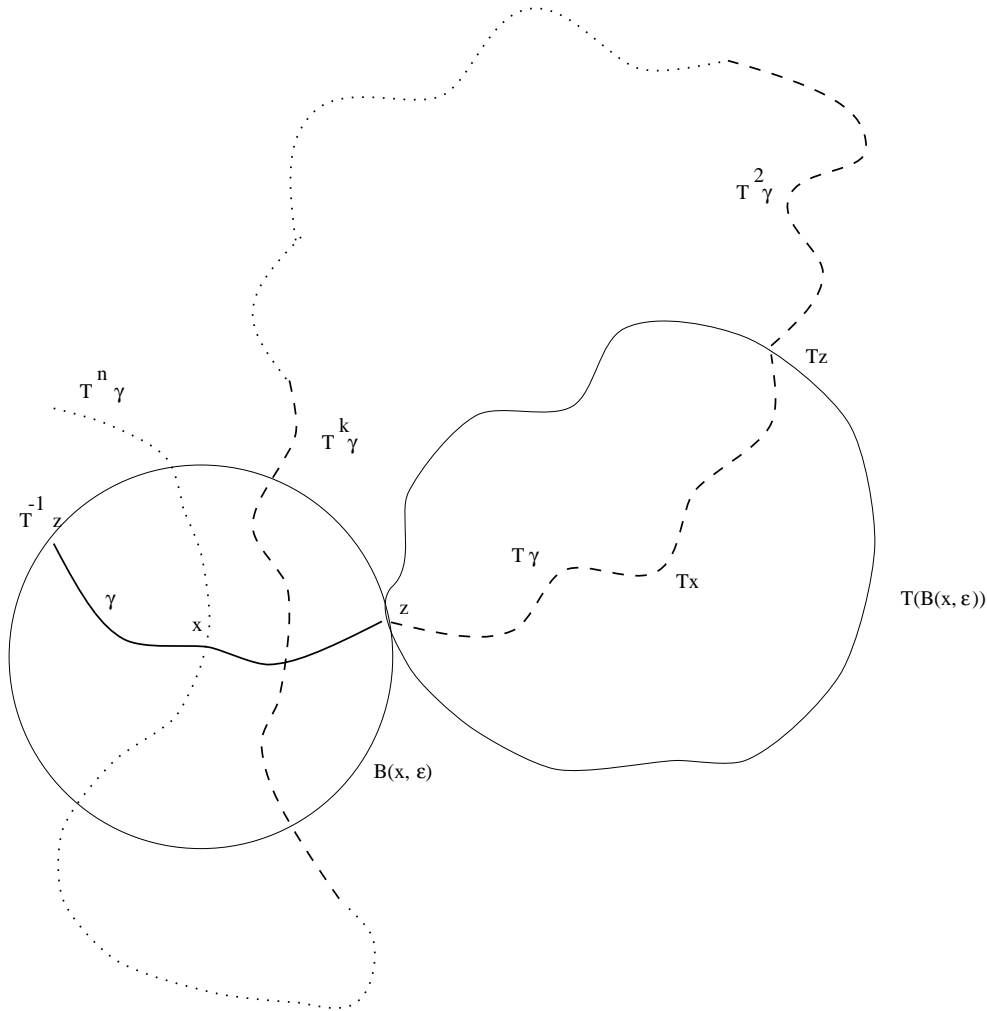


FIGURE 13.1. The proof of Brouwer's theorem

that  $T'$  differs from  $T$  only in this neighbourhood, then they have the same set of fixed points in  $\mathbb{R}^2$ .

Part (ii) is proved using some elementary topology. Assume that  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  has a periodic point  $T^2x = x$  (with  $x$  and  $Tx$  distinct). By “adding the fixed point at infinity”  $\infty$  this corresponds to a homeomorphism on the standard 2-sphere  $S^2$ . By “blowing-up” the points  $x$  and  $Tx$  into circles we finally get a corresponding homeomorphism  $\hat{T} : A \rightarrow A$  on a closed annulus  $A = S^1 \times [0, 1]$ , say, which preserves orientation (since  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  preserved orientation) and interchanges the two boundary components (since  $T$  interchanged  $x$  and  $Tx$ ). Finally, the universal cover of the annulus is  $X = \mathbb{R} \times [0, 1]$ , and the lift  $\tilde{T} : X \rightarrow X$  interchanges the two sides  $\mathbb{R} \times \{0\}$  and  $\mathbb{R} \times \{1\}$ . Let  $\pi : X \rightarrow A$  be the covering projection and let  $g : X \rightarrow X$  be a generator for the covering group (isomorphic to  $\mathbb{Z}$ ).

A simple application of the familiar Brouwer fixed point theorem gives

that there exist fixed points  $\tilde{T}z_0 = z_0$  for  $\tilde{T} : X \rightarrow X$  and  $(\tilde{T}g)z_1 = z_1$  for  $(\tilde{T}g) : X \rightarrow X$ . There are two possibilities: either  $\pi(z_0) \neq \infty$  or  $\pi(z_0) = \infty$ . In the first case,  $\pi(z)$  corresponds to a genuine fixed point for the original map  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ , and the proof is complete. In the second case, we have that  $\pi(z_1)$  is distinct from  $\pi(z_0) = \infty$  and so it corresponds to a genuine fixed point for the original map  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ . ■

### 13.3 Comments and references

The Poincaré-Birkhoff theorem usual guarantees the existence of two fixed points. However, the theorem of Franks has much weaker hypotheses [3].

Modern proofs of the Brouwer plane translation theorem can be found in [1] and [2].

#### References

1. M. Brown, *A new proof of Brouwer's lemma on translation arcs*, Houston J. Math. **10** (1984), 35-41.
2. A. Fathi, *An orbit closing proof of Brouwer's lemma on translation arcs*, L'enseignement Math. **33** (1987), 315-322.
3. J. Franks, *Generalisations of the Poincaré-Birkhoff theorem*, Annals of Math. **128** (1988), 139-151.

**THE VARIATIONAL PRINCIPLE**

We introduced in chapter 3 the topological entropy  $h(T)$  of a continuous map  $T : X \rightarrow X$  of a compact metric space  $X$  and in chapter 8 the entropy  $h_\mu(T)$  of a  $T$ -invariant probability measure  $\mu$ . In this chapter we show that these two notions are closely related.

**14.1 The variational principle for entropy**

The main result of this chapter is the following.

**THEOREM 14.1 (VARIATIONAL PRINCIPLE).** *Let  $T : X \rightarrow X$  be a continuous map on a compact metric space.*

- (1) *For any  $T$ -invariant probability measure  $\mu$  we have that  $h_\mu(T) \leq h(T)$ .*
- (2)  *$h(T) = \sup\{h_\mu(T) : \mu \text{ is a } T\text{-invariant probability measure}\}$ .*

**14.2 The proof of the variational principle**

The proof we give is due to Misiurewicz [1]. Recall that the topological entropy of a cover  $\mathcal{U}$  is  $H(\mathcal{U}) = \log N(\mathcal{U})$  and the entropy of a partition  $\alpha$  with respect to  $\mu$  is  $H_\mu(\alpha) = -\sum_{A \in \alpha} \mu(A) \log \mu(A)$ .

**PROOF OF (1).** Fix a finite Borel measurable partition  $\alpha = \{A_1, \dots, A_k\}$  for  $X$ . Given  $\epsilon > 0$ , say, we want to “improve” this partition by choosing a family of closed sets  $\hat{A}_1, \dots, \hat{A}_k$  such that

- (1)  $\hat{A}_i \subset A_i, i = 1, \dots, k$ , and
- (2)  $\mu(A_i - \hat{A}_i) < \epsilon$ ,

and then defining a new partition  $\hat{\alpha} = \{\hat{A}_1, \dots, \hat{A}_k, V\}$ , where  $V = X - (\cup_{i=1}^k \hat{A}_i)$ .

We can consider an open cover for  $X$  defined by

$$\mathcal{U} = \left\{ \hat{A}_1 \cup V, \dots, \hat{A}_k \cup V \right\}$$

If we compare the open covers  $\vee_{i=0}^{n-1} T^{-i}\mathcal{U}$  and the partitions  $\vee_{i=0}^{n-1} T^{-i}\hat{\alpha}$  then we see that

$$N\left(\vee_{i=0}^{n-1} T^{-i}\hat{\alpha}\right) \leq 2^n N\left(\vee_{i=0}^{n-1} T^{-i}\mathcal{U}\right), \quad n \geq 1 \tag{14.1}$$

(where we recall that  $N(\bigvee_{i=0}^{n-1} T^{-i} \mathcal{U})$  is the number of elements in a minimal subcover for  $\bigvee_{i=0}^{n-1} T^{-i} \mathcal{U}$  and  $N(\bigvee_{i=0}^{n-1} T^{-i} \hat{\alpha})$  is the number of non-trivial elements in  $\bigvee_{i=0}^{n-1} T^{-i} \hat{\alpha}$ ).

SUB-LEMMA 14.1.1.  $H_\mu(\bigvee_{i=0}^{n-1} T^{-i} \hat{\alpha}) \leq \log N(\bigvee_{i=0}^{n-1} T^{-i} \hat{\alpha})$ .

PROOF. Assume that  $\bigvee_{i=0}^{n-1} T^{-i} \hat{\alpha} = \{C_1, \dots, C_N\}$ ; then we can write  $H_\mu(\bigvee_{i=0}^{n-1} T^{-i} \hat{\alpha}) = -\sum_{i=1}^N \mu(C_i) \log \mu(C_i)$ . ■

We can use Sub-lemma 14.7 to bound

$$\begin{aligned} & H_\mu(\bigvee_{i=0}^{n-1} T^{-i} \hat{\alpha}) \\ & \leq \log N(\bigvee_{i=0}^{n-1} T^{-i} \hat{\alpha}) \\ & \leq n \log 2 + \log N(\bigvee_{i=0}^{n-1} T^{-i} \mathcal{U}) \quad (\text{by (14.1)}). \end{aligned}$$

Recalling that

$$h(T) \geq h(T, \mathcal{U}) = \lim_{n \rightarrow +\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i} \mathcal{U})$$

and

$$h_\mu(T, \alpha) = \lim_{n \rightarrow +\infty} \frac{1}{n} H_\mu(\bigvee_{i=0}^{n-1} T^{-i} \alpha)$$

we see that  $h_\mu(T, \hat{\alpha}) \leq \log 2 + h(T)$ . Moreover, by Corollary 8.6.1 we have that

$$\begin{aligned} |h_\mu(T, \hat{\alpha}) - h_\mu(T, \alpha)| & \leq H_\mu(\alpha | \hat{\alpha}) + H_\mu(\hat{\alpha} | \alpha) \\ & = -\sum_{C \in \alpha} \sum_{\hat{C} \in \hat{\alpha}} \mu(C \cap \hat{C}) \log \left( \frac{\mu(C \cap \hat{C})}{\mu(\hat{C})} \right) \\ & \quad - \sum_{C \in \alpha} \sum_{\hat{C} \in \hat{\alpha}} \mu(C \cap \hat{C}) \log \left( \frac{\mu(C \cap \hat{C})}{\mu(C)} \right) < 1, \end{aligned}$$

say, providing  $\epsilon$  was sufficiently small.

Since  $\alpha$  was arbitrary, we see that

$$h_\mu(T) = \sup\{h_\mu(T, \alpha) : \alpha \text{ is a finite partition}\} \leq h(T) + \log 2 + 1.$$

Finally, we can apply the argument to iterates  $T^k$  ( $k \geq 1$ ) to see that  $h_\mu(T^k) \leq h(T^k) + \log 2 + 1$ . By Corollary 3.8.1 we know that  $h(T^k) = kh(T)$ . The following gives the analogous result for measure theoretic entropy.

SUB-LEMMA 14.1.2 (ABRAMOV'S THEOREM). *For  $k \geq 1$ ,  $h_\mu(T^k) = kh_\mu(T)$ .*

PROOF. Given any partition  $\alpha$  we observe that

$$\begin{aligned} h_\mu(T^k, \bigvee_{i=0}^{k-1} T^{-i} \alpha) & = \lim_{n \rightarrow +\infty} \frac{1}{n} H_\mu(\bigvee_{i=0}^{n-1} T^{-ik} (\bigvee_{j=0}^{k-1} T^{-j} \alpha)) \\ & = \lim_{N \rightarrow +\infty} \frac{k}{N} H_\mu(\bigvee_{i=0}^{N-1} T^{-i} \alpha) = kh_\mu(T, \alpha). \end{aligned}$$

Given  $\epsilon > 0$  we can choose  $\alpha$  with  $h(T, \alpha) > h_\mu(T) - \epsilon$  so that we have

$$\begin{aligned} h_\mu(T^k) &\geq h_\mu(T^k, \bigvee_{i=0}^{k-1} T^{-i} \alpha) \\ &\geq kh_\mu(T, \alpha) \geq kh_\mu(T) - k\epsilon. \end{aligned}$$

Since  $\epsilon > 0$  is arbitrary we see that  $h_\mu(T^k) \geq kh_\mu(T)$ .

To get the reverse inequality, notice that  $h_\mu(T^k, \alpha) \leq h_\mu(T^k, \bigvee_{i=0}^{k-1} T^{-i} \alpha)$ , using Lemma 8.6. Given  $\epsilon > 0$  we can choose  $\alpha$  with  $h_\mu(T^k, \alpha) > h_\mu(T^k) - \epsilon$  and then

$$\begin{aligned} kh_\mu(T) &\geq kh_\mu(T, \alpha) = h_\mu(T^k, \bigvee_{i=0}^{k-1} T^{-i} \alpha) \\ &\geq h_\mu(T^k, \alpha) > h_\mu(T^k) - \epsilon. \end{aligned}$$

Since  $\epsilon > 0$  is arbitrary we see that  $h_\mu(T^k) \leq kh_\mu(T)$ . ■

We can now complete the proof of (1) since

$$\begin{aligned} h_\mu(T) &= \lim_{k \rightarrow +\infty} \frac{h_\mu(T^k)}{k} \\ &\leq \lim_{k \rightarrow +\infty} \frac{h(T^k)}{k} + \lim_{k \rightarrow +\infty} \frac{\log 2 + 1}{k} = h(T). \end{aligned}$$
■

PROOF OF (2). It suffices to show that given  $\delta > 0$  there exists a  $T$ -invariant probability measure  $\mu$  with  $h_\mu(T) \geq h(T) - \delta$ . We want to choose  $\epsilon > 0$  sufficiently small that  $\lim_{n \rightarrow +\infty} \frac{1}{n} \log(s(n, \epsilon)) \geq h(T) - \delta$ , where  $s(n, \epsilon)$  is the maximal cardinality of an  $(n, \epsilon)$ -separating set. We can find a sub-sequence  $n_i \rightarrow +\infty$  such that  $\frac{1}{n_i} \log(s(n_i, \epsilon)) = h(T)$ . Let  $S_{n_i}$  be such an  $(n_i, \epsilon)$ -separated set.

For each  $n_i$  we can define a (possibly non-invariant) probability measure

$$\nu_{n_i} = \frac{1}{s(n_i, \epsilon)} \sum_{x \in S_{n_i}} \delta_x.$$

In order to arrive at a  $T$ -invariant probability measure we can consider an accumulation point  $\mu$  (in the weak-star topology) of the measures

$$\mu_{n_i} = \frac{1}{n_i} \sum_{r=0}^{n_i-1} (T^r)^* \nu_{n_i}.$$

By replacing  $\{n_i\}$  by a sub-sequence, if necessary, we can assume that  $\mu_{n_i} \rightarrow \mu$ .

Let want to consider a finite partition  $\alpha = \{A_1, \dots, A_k\}$  such that

- (1)  $\text{diam}(A_i) < \epsilon$ ,  $i = 1, \dots, k$ ; and
- (2)  $\mu(\partial A_i) = 0$ , for  $i = 1, \dots, k$ .

Since  $S_{n_i}$  is an  $(n_i, \epsilon)$ -separated set we know that each set  $C \in \alpha^{(n_i)} := \bigvee_{j=0}^{n_i-1} T^{-j} \alpha$  contains at most one point  $x = x_C \in S_{n_i}$ . Thus of the sets in  $S_{n_i}$  there are  $s(n_i, \epsilon)$  sets with  $\nu_{n_i}$ -measure  $\frac{1}{s(n_i, \epsilon)}$  and the remainder have  $\nu_{n_i}$ -measure zero. In particular, we see that

$$\log(s(n_i, \epsilon)) = - \sum_{C \in \alpha^{(n_i)}} \nu_{n_i}(C) \log \nu_{n_i}(C). \quad (14.2)$$

In order to take limits in a sensible way we fix first  $1 < N < n_i$  and then  $0 \leq j \leq N - 1$ . We can write

$$\alpha^{(n_i)} = \bigvee_{i=0}^{n_i-1} T^{-i} \alpha = \left( \bigvee_{\substack{l=j \pmod{N} \\ 0 \leq l \leq n_i - N}} T^{-l} \left( \bigvee_{i=0}^{N-1} T^{-i} \alpha \right) \right) \vee \left( \bigvee_{i \in E} T^{-i} \alpha \right)$$

where  $E = \{0, 1, \dots, j-1\} \cup \{M_j, M_j+1, \dots, n_i-1\}$ , with  $M_j = N \lfloor \frac{n_i-j}{N} \rfloor$ , has cardinality at most  $2N$ .

SUB-LEMMA 14.1.3. *Given measurable partitions  $\beta$  and  $\gamma$  we have that*

$$H_{\nu_{n_i}}(\beta \vee \gamma) \leq H_{\nu_{n_i}}(\beta) + H_{\nu_{n_i}}(\gamma)$$

PROOF. For *invariant* measures, this would be an immediate consequence of Lemma 8.4 (and Corollary 8.4.1). However, although in chapter 8 we assumed that the ambient measures were invariant, this property was not used at this stage and the result remains true without it. ■

In particular, we have that

$$\begin{aligned} & - \sum_{C \in \alpha^{(n_i)}} \nu_{n_i}(C) \log \nu_{n_i}(C) \\ & \leq \sum_{\substack{l=j \pmod{N} \\ 0 \leq l \leq N - n_i}} \left( - \sum_{C \in T^{-l} \alpha^{(N)}} \nu_{n_i}(C) \log \nu_{n_i}(C) \right) \\ & \quad + \sum_{i \in E} \left( - \sum_{C \in T^{-i} \alpha^{(N)}} \nu_{n_i}(C) \log \nu_{n_i}(C) \right) \quad (14.3) \\ & \leq \sum_{r=0}^{M_j} \left( - \sum_{D \in \alpha^{(N)}} (T^{rN+j})^* \nu_{n_i}(D) \log((T^{rN+j})^* \nu_{n_i}(D)) \right) \\ & \quad + 2N \log k \end{aligned}$$

(where for  $l = rN + j$  there is a natural correspondence between  $D \in \alpha^{(N)}$  and  $C \in T^{-l}\alpha^{(N)}$  with  $(T^l)^*\nu_{n_i}(D) := \nu_{n_i}(T^{-l}D) = \nu_{n_i}(C)$  and  $C = T^{-l}D$ ). Summing the inequalities (14.3) over  $j = 0, \dots, N - 1$  we have by (14.2)

$$\begin{aligned} & N \log(s(n_i, \epsilon)) \\ & \leq \sum_{l=0}^{n_i-1} \left( - \sum_{D \in \alpha^{(N)}} (T^l)^*\nu_{n_i}(D) \log((T^l)^*\nu_{n_i}(D)) \right) + 2N^2 \log k \end{aligned} \quad (14.4)$$

**SUB-LEMMA 14.1.4.** *Let  $\alpha$  be a measurable partition and let  $\nu_1$  and  $\nu_2$  be (not necessarily invariant) probability measures; then given  $0 \leq a \leq 1$  we have that*

$$\begin{aligned} & \sum_{A \in \alpha} [a\nu_1 + (1-a)\nu_2](A) \log[a\nu_1 + (1-a)\nu_2](A) \\ & \leq a \left( \sum_{A \in \alpha} \nu_1(A) \log \nu_1(A) \right) + (1-a) \left( \sum_{A \in \alpha} \nu_2(A) \log \nu_2(A) \right). \end{aligned}$$

**PROOF.** This follows immediately since  $t \mapsto t \log t$  is convex. ■

Dividing (14.4) by  $n_i N$  we get that

$$\begin{aligned} & \frac{1}{n_i} \log(s(n_i, \epsilon)) \\ & \leq \frac{1}{n_i} \sum_{r=0}^{n_i-1} \left( - \frac{1}{N} \sum_{C \in \alpha^{(N)}} (T^r)^*\nu_{n_i}(C) \log((T^r)^*\nu_{n_i}(C)) \right) + \frac{2N \log k}{n_i} \\ & \leq - \frac{1}{N} \sum_{C \in \alpha^{(N)}} \mu_{n_i}(C) \log \mu_{n_i}(C) + \frac{2N \log k}{n_i} \end{aligned}$$

where we have used Sub-lemma 14.1.4 repeatedly for the last line.

Since we have assumed  $\mu(\partial A_i) = 0$ , letting  $n_i \rightarrow +\infty$  (with  $N$  fixed) we have that

$$- \sum_{C \in \alpha^{(N)}} \mu_{n_i}(C) \log \mu_{n_i}(C) \rightarrow H_\mu(\alpha^{(N)}).$$

This means that

$$\begin{aligned} h(T) - \delta & \leq \lim_{n_i \rightarrow +\infty} \frac{1}{n_i} \log(s(n_i, \epsilon)) \\ & \leq \frac{1}{N} H_\mu(\alpha^{(N)}) + \lim_{n_i \rightarrow +\infty} \frac{2N^2 \log k}{n_i} \\ & = \frac{1}{N} H_\mu(\alpha^{(N)}). \end{aligned}$$

Letting  $N \rightarrow +\infty$  we have that

$$h(T) - \delta \leq \lim_{N \rightarrow +\infty} \frac{1}{N} H_\mu(\alpha^{(N)}) = h_\mu(\alpha) \leq h_\mu(T).$$

Since  $\delta > 0$  is arbitrary this completes the proof. ■

### 14.3 Comments and reference

The proof we give is due to Misiurewicz [1]. Theorem 14.1 (1) was originally due to Goodman. Theorem 14.1 (2) was subsequently proved by Walters.

#### References

1. M. Misiurewicz, *A short proof of the variational principle for a  $\mathbb{Z}_+^N$  action on a compact space*, Astérisque **40** (1976), 147-187.



## INVARIANT MEASURES FOR COMMUTING TRANSFORMATIONS

In this chapter we describe an important conjecture of Furstenberg and related work of Rudolph.

### 15.1 Furstenberg's conjecture and Rudolph's theorem

Consider the transformations

- (i)  $S : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  defined by  $S(x) = 2x \pmod{1}$ , and
- (ii)  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  defined by  $T(x) = 3x \pmod{1}$ .

(For a mnemonic aid:  $S$  stands for “second” and  $T$  for “third”.) It is easy to see that these transformations commute, i.e.  $ST = TS$ ).

Recall that the  $S$ -invariant probability measures form a convex weak-star compact set  $\mathcal{M}_S$  (and similarly, the  $T$ -invariant probability measures form a convex weak-star compact set  $\mathcal{M}_T$ ).

We want to describe the probability measures which are both  $T$ -invariant and  $S$ -invariant (i.e. the intersection  $\mathcal{M}_S \cap \mathcal{M}_T$ ). We need only consider the  $(S, T)$ -ergodic measures  $\mu$  in  $\mathcal{M}_S \cap \mathcal{M}_T$  (i.e. those probability measures invariant under both  $S$  and  $T$  for which the only Borel sets  $B$  with  $T^{-n}S^{-m}B = B \forall n, m \geq 0$  have either  $\mu(B) = 0$  or 1, since these are the extremal measures in  $\mathcal{M}_S \cap \mathcal{M}_T$ ).

**FURSTENBERG'S CONJECTURE.** *The only  $(S, T)$ -ergodic measures are the Haar-Lebesgue measure and measures supported on a finite set.*

Notice that the Haar-Lebesgue measure  $\nu$  has entropies  $\log 2$  and  $\log 3$ , respectively, for the transformations  $S$  and  $T$ , and any finitely supported measure always has zero entropy with respect to either  $S$  or  $T$ . The following partial solution is due to D.J. Rudolph.

**THEOREM 15.1 (RUDOLPH).** *The only  $(S, T)$ -ergodic measure  $\mu$  which has non-zero entropy (w.r.t. either  $S$  or  $T$ ) is the Haar-Lebesgue measure.*

### 15.2 The proof of Rudolph's theorem

We begin with a few comments.

- (1) Haar-Lebesgue measure  $\nu$  on the unit circle is characterized as the only probability measure invariant under all rotations on the circle.

Moreover, it is the only measure invariant under all rotations  $x \mapsto x + a \pmod{1}$ , where  $a$  is any rational number  $a = \frac{j}{2^k}$ .

(2) For  $n \geq 1$  and  $f \in L^2(X, \mathcal{B}, \mu)$  we can write that

$$E(f|T^{-n}\mathcal{B})(x) = \sum_{y \in T^{-n}T^n x} \frac{f(y)}{(T^n)'(y)}$$

where  $T'(x) = \frac{d\mu T}{d\mu}$  (and similarly for  $S$ ).

(3) We can write  $(S^n)'(x) = S'(S^{n-1}x) \dots S'(x)$ . If we knew that  $S'$  is constant (almost everywhere) then by the martingale theorem we would have that

$$\begin{aligned} \int f(x+a) d\mu &= \lim_{n \rightarrow +\infty} \int E(f(x+a)|S^{-n}\mathcal{B}) \\ &= \lim_{n \rightarrow +\infty} \int E(f(x)|S^{-n}\mathcal{B}) \\ &= \int f(x) d\mu \end{aligned}$$

and thus we know that  $\nu$  is the Haar-Lebesgue measure.

(4) Since  $ST = TS$  we have that

$$T'(Sx) \cdot S'(x) = (TS)'(x) = (ST)'(x) = S'(Tx) \cdot S'(x).$$

In particular, we can write  $\frac{S'(Tx)}{T'(Sx)} = \frac{S'(x)}{T'(x)}$ .

We begin with the following simple (but fundamental) Sub-lemma.

SUB-LEMMA 15.1.1.  $S'(Tx) = S'(x)$  for almost all  $x$ .

PROOF. We begin by claiming that  $E(S'|T^{-1}\mathcal{B})(x) = S'(Tx)$ . To see this we observe that

$$\begin{aligned} E(S'|T^{-1}\mathcal{B})(x) &= \sum_{y:Ty=Tx} \frac{S'(y)}{T'(y)} \\ &= \sum_{y:Ty=Tx} \frac{S'(Ty)}{T'(Sy)} \text{ (by (4) above)} \\ &= S'(Tx) \left( \sum_{y:Ty=Tx} \frac{1}{T'(Sy)} \right) \\ &= S'(Tx) \end{aligned}$$

where we have used that there is a bijection between  $\{y : Ty = Tx\}$  and  $\{w : Tw = T(Sx)\}$  to write that

$$\sum_{y:Ty=Tx} \frac{1}{T'(Sy)} = \sum_{w:Tw=T(Sx)} \frac{1}{T'(w)} = 1.$$

This proves the claim. To complete the proof of the Sub-lemma we need only show that  $E(S'|T^{-1}\mathcal{B})(x) = S'(x)$ . However, since  $E(\cdot|T^{-1}\mathcal{B})(x) : L^2(X, \mathcal{B}, \mu) \rightarrow L^2(X, \mathcal{B}, \mu)$  is a positive operator which is a contraction and

$$\|E(S'|T^{-1}\mathcal{B})\|_2 = \|S'T\|_2 = \|S'\|_2,$$

we indeed see that  $S'T = E(S'|T^{-1}\mathcal{B})(x) = S'(x)$ . ■

If we knew that  $T$  was ergodic we could now deduce  $S'$  is constant. Unfortunately, we don't know this and a little more work is required.

DEFINITION. We let  $\mathcal{A}_1 \subset \mathcal{B}$  denote the *smallest* sub-sigma-algebra for which  $S'(x)$  is measurable.

In the course of the proof we shall establish that  $\mathcal{A}_1 \subset S^{-1}\mathcal{B}$  (which, by the definition of  $\mathcal{A}_1$ , will imply that  $S'(x)$  is constant).

Similarly, we can introduce the sub-sigma-algebras  $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \dots \subset \mathcal{A}_n \subset \dots \subset \mathcal{B}$  where  $\mathcal{A}_n$  is the smallest sub sigma-algebra for which all of the functions  $S'(x), (S^2)'(x), \dots, (S^n)'(x)$  are measurable.

SUB-LEMMA 15.1.2. *For each  $n \geq 1$  we have that*

- (a)  $S^{-1}\mathcal{A}_n \subset \mathcal{A}_{n+1}$ ,
- (b)  $T^{-1}\mathcal{A}_n = \mathcal{A}_n$ .

PROOF.

- (a) If we write  $(S^{n+1})'(x) = (S^n)'(Sx) \cdot S'(x)$  then, since by hypothesis  $S^n$  and  $S'$  are  $\mathcal{A}_n$ -measurable, the right hand side is measurable with respect to  $S^{-1}\mathcal{A}_n$ .
- (b) Since  $S'(Tx) = S'(x)$  we also see that

$$\begin{aligned} (S^n)'(Tx) &= S'(Tx) \cdot S'(STx) \dots S'(S^{n-1}Tx) \\ &= S'(x) \cdot S'(Sx) \dots S'(S^{n-1}x) \\ &= (S^n)'(x). \end{aligned}$$

But the right hand side is  $\mathcal{A}_n$ -measurable by hypothesis. ■

DEFINITION. We write  $\mathcal{A} = \bigvee_{n=1}^{\infty} \mathcal{A}_n$ . The above lemma guarantees that  $T^{-1}\mathcal{A} = \mathcal{A}$  and  $S^{-1}\mathcal{A} \subset \mathcal{A}$ .

We now move on to entropy considerations. Let  $\gamma = \{[0, \frac{1}{6}], [\frac{1}{6}, \frac{2}{6}], \dots, [\frac{5}{6}, 1]\}$  denote the partition into intervals of length one sixth.

SUB-LEMMA 15.1.3. *There exists a sequence  $s_n \rightarrow +\infty$  such that the partitions*

$$\bigvee_{i=0}^{s_n} S^{-i}\gamma = \left\{ \left[0, \frac{1}{6 \cdot 2^{s_n}}\right], \left[\frac{1}{6 \cdot 2^{s_n}}, \frac{2}{6 \cdot 2^{s_n}}\right], \dots, \left[\frac{6 \cdot 2^{s_n} - 1}{6 \cdot 2^{s_n}}, 1\right] \right\} \text{ and}$$

$$\bigvee_{i=0}^{3^n} T^{-i}\gamma = \left\{ \left[0, \frac{1}{6 \cdot 3^n}\right], \left[\frac{1}{6 \cdot 3^n}, \frac{2}{6 \cdot 3^n}\right], \dots, \left[\frac{6 \cdot 3^n - 1}{6 \cdot 3^n}, 1\right] \right\}$$

have the property that every element of either partition is contained in at most four elements of the other partition.

PROOF. For each  $n \geq 1$  we choose the values  $s_n \geq 1$  such that  $3^{n-1} \leq 2^{s_n} \leq 3^n$ . The lengths of the intervals for each partition are  $\frac{1}{6 \cdot 2^{s_n}}$  and  $\frac{1}{6 \cdot 3^n}$  and thus their ratios are bounded above and below by 3 and  $\frac{1}{3}$ , respectively. This is enough to complete the proof. ■

In what follows we shall make frequent use of the basic identity for entropy:  $H(\alpha \vee \beta | \mathcal{C}) = H(\alpha | \mathcal{C} \vee \beta) + H(\beta | \mathcal{C})$ .

Recall that the *entropies* of the transformations are given by

$$h(T) := \lim_{n \rightarrow +\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\gamma)$$

and

$$h(S) := \lim_{k \rightarrow +\infty} \frac{1}{k} H(\bigvee_{i=0}^{k-1} S^{-i}\gamma).$$

The following sub-lemma shows similar limits involving the sigma-algebra  $\mathcal{A}$ .

SUB-LEMMA 15.1.4. *The following limits exist and are equal to the entropies:*

$$h(T) = \lim_{n \rightarrow +\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\gamma | \mathcal{A})$$

and

$$h(S) = \lim_{n \rightarrow +\infty} \frac{1}{s_n} H(\bigvee_{i=0}^{s_n-1} S^{-i}\gamma | \mathcal{A}).$$

PROOF. We begin with an argument which is borrowed from the standard entropy identities. We see that for any  $n, m \geq 0$  we have that

$$\begin{aligned} & H(\vee_{i=0}^{n+m-1} T^{-i} \gamma | \mathcal{A}) \\ &= H(\vee_{i=0}^{n-1} T^{-i} \gamma | \mathcal{A}) + H(\vee_{i=n}^{n+m-1} T^{-i} \gamma | \mathcal{A} \vee (\vee_{i=0}^{n-1} T^{-i} \gamma)) \\ &\leq H(\vee_{i=0}^{n-1} T^{-i} \gamma | \mathcal{A}) + H(\vee_{i=n}^{n+m-1} T^{-i} \gamma | \mathcal{A}) \\ &= H(\vee_{i=0}^{n-1} T^{-i} \gamma | \mathcal{A}) + H(\vee_{i=0}^{m-1} T^{-i} \gamma | \mathcal{A}) \end{aligned}$$

(where for the last equality we use that  $T^{-1} \mathcal{A} = \mathcal{A}$ ). Thus by subadditivity the limit  $h(T | \mathcal{A}) := \lim_{n \rightarrow +\infty} \frac{1}{n} H(\vee_{i=0}^{n-1} T^{-i} \gamma | \mathcal{A})$  exists. By the basic equalities for entropy we see that

$$\begin{aligned} & H(\vee_{i=0}^{s_n-1} S^{-i} \gamma | \mathcal{A}) - H(\vee_{i=0}^{n-1} T^{-i} \gamma | \mathcal{A}) \\ &= H((\vee_{i=0}^{s_n} S^{-i} \gamma) \vee (\vee_{i=0}^{n-1} T^{-i} \gamma) | \mathcal{A}) + H(\vee_{i=0}^{s_n} S^{-i} \gamma | (\vee_{i=0}^{n-1} T^{-i} \gamma) \vee \mathcal{A}) \\ &\quad - H((\vee_{i=0}^{s_n} S^{-i} \gamma) \vee (\vee_{i=0}^{n-1} T^{-i} \gamma) | \mathcal{A}) - H(\vee_{i=0}^{n-1} T^{-i} \gamma | (\vee_{i=0}^{s_n} S^{-i} \gamma) \vee \mathcal{A}) \end{aligned}$$

and so we can identify the limit as

$$\begin{aligned} h(T | \mathcal{A}) &= \lim_{n \rightarrow +\infty} \frac{1}{n} H(\vee_{i=0}^{n-1} T^{-i} \gamma | \mathcal{A}) \\ &= \lim_{n \rightarrow +\infty} \frac{1}{n} \left( H(\vee_{i=0}^{n-2} T^{-i} \gamma | \mathcal{A}) + H(T^{-(n-1)} \gamma | \mathcal{A} \vee (\vee_{i=0}^{n-2} T^{-i} \gamma)) \right) \\ &= h(T) \end{aligned}$$

since

$$h(T) = \lim_{n \rightarrow +\infty} \frac{1}{n} \left( H(\vee_{i=0}^{n-2} T^{-i} \gamma) \right)$$

and

$$H\left(T^{-(n-1)} \gamma | \mathcal{A} \vee (\vee_{i=0}^{n-2} T^{-i} \gamma)\right) \leq H\left(T^{-(n-1)} \gamma | \mathcal{A}\right) = H(\gamma | \mathcal{A}) < +\infty.$$

By sublemma 15.1.3 we have that the final expression above is bounded (independently of  $n$ ) and thus we have that the following limit exists

$$h(S | \mathcal{A}) := \lim_{n \rightarrow +\infty} \frac{1}{n} H(\vee_{i=0}^{s_n-1} T^{-i} \gamma | \mathcal{A}).$$

Moreover, this argument gives that  $h(T | \mathcal{A}) = \frac{\log 3}{\log 2} h(S | \mathcal{A})$ .

Observe that if we replace  $\mathcal{A}$  by the trivial sigma-algebra then the same argument gives that  $h(T) = \frac{\log 3}{\log 2} h(S)$ . Comparing these identities we see that  $h(S) = h(S | \mathcal{A})$ . ■

We now apply Sub-lemma 15.1.4 to show that  $\mathcal{A} \subset \vee_{i=1}^{\infty} S^{-i} \gamma$ , which is essentially the end of the proof.

SUB-LEMMA 15.1.5.  $H(\mathcal{A} | \bigvee_{i=1}^{\infty} S^{-i}\mathcal{B}) = 0$ .

PROOF. By the basic equality for entropy we have that

$$\begin{aligned} H(\mathcal{A} | \bigvee_{i=1}^{\infty} S^{-i}\gamma) &= H(\gamma \vee \mathcal{A} | \bigvee_{i=1}^{\infty} S^{-i}\beta) - H(\gamma | \bigvee_{i=1}^{\infty} S^{-i}\gamma \vee \mathcal{A}) \\ &= H(\gamma | \bigvee_{i=1}^{\infty} S^{-i}\beta) - H(\gamma | \bigvee_{i=1}^{\infty} S^{-i}\gamma \vee \mathcal{A}) \quad (15.1) \\ &= h(S) - H(\gamma | \bigvee_{i=1}^{\infty} S^{-i}\gamma \vee \mathcal{A}) \end{aligned}$$

(where we have used that  $H(\gamma \vee \mathcal{A} | \bigvee_{i=1}^{\infty} S^{-i}\beta) = H(\gamma | \bigvee_{i=1}^{\infty} S^{-i}\beta) = h(S)$ ). We next observe that

$$\begin{aligned} &H(\bigvee_{i=0}^{n-1} S^{-i}\gamma | \mathcal{A}) \\ &= H(\gamma | \bigvee_{i=1}^{n-1} S^{-i}\gamma \vee \mathcal{A}) + H(\bigvee_{i=1}^{n-1} S^{-i}\gamma | \mathcal{A}) \\ &\leq H(\gamma | \bigvee_{i=1}^{n-1} S^{-i}\gamma \vee \mathcal{A}) + H(\bigvee_{i=0}^{n-2} S^{-i}\gamma | \mathcal{A}) \\ &\dots \\ &\leq H(\gamma | \bigvee_{i=1}^{n-1} S^{-i}\gamma \vee \mathcal{A}) + H(\gamma | \bigvee_{i=1}^{n-2} S^{-i}\gamma \vee \mathcal{A}) + \dots + H(\gamma | \mathcal{A}) \end{aligned}$$

(This argument is a modification of the standard entropy proof that  $h(S) = H(\gamma | \bigvee_{i=1}^{\infty} S^{-i}\gamma)$ .) Thus from the definition of  $h(S|\mathcal{A})$  we have that

$$\begin{aligned} h(S|\mathcal{A}) &:= \lim_{n \rightarrow +\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} S^{-i}\gamma | \mathcal{A}) \\ &= \lim_{n \rightarrow +\infty} \frac{1}{n} H(\gamma | \bigvee_{i=1}^{n-1} S^{-i}\gamma \vee \mathcal{A}) + \lim_{n \rightarrow +\infty} \frac{1}{n} H(\bigvee_{i=1}^{n-1} S^{-i}\gamma | \mathcal{A}) \\ &= H(\gamma | \bigvee_{i=1}^{\infty} S^{-i}\gamma). \end{aligned} \quad (15.2)$$

Comparing (15.1) and (15.2) we see that

$$0 \leq H(\gamma | \bigvee_{i=1}^{\infty} S^{-i}\gamma) \leq h(T) - h(T|\mathcal{A}) = 0.$$

■

To finish off the proof of Theorem 15.1 we need only recall that  $H(\mathcal{A} | \bigvee_{i=1}^{\infty} S^{-i}\gamma) = 0$  implies that  $\mathcal{A} \subset \bigvee_{i=1}^{\infty} S^{-i}\gamma$ .

Repeating the argument with  $S$  replaced by  $S^k$  for  $k = 1, 2, \dots$  we see that  $\mathcal{A} \subset \bigcap_{n=0}^{\infty} S^{-n}\mathcal{B}$ . In particular, this shows that  $(S^n)'(y)$  is constant for  $y \in \{w | S^n x = S^n w\}$  (almost everywhere).

We observe that since  $h(S) > 0$  (equivalently  $h(T) > 0$ ), there must be a set of positive measure on which  $S$  has two pre-images (otherwise  $S$  would be invertible almost everywhere and then have entropy zero). Moreover we claim that the set with two  $S$  pre-images is invariant under  $S$  and  $T$ . By *ergodicity* of  $(S, T)$  we see that almost all points have two pre-images.

This suffices to apply the argument in comment (3).

■

### 15.3 Comments and references

The original proof of Rudolph had a symbolic formulation [2]. The proof we give here is a version due to Parry [1].

#### References

1. W. Parry, *Squaring and cubing the circle*, Ergodic Theory, Proceedings of the Warwick Symposium on  $\mathbb{Z}^d$ -actions (M. Pollicott and K. Schmidt, ed.), C.U.P., Cambridge, 1996, pp. 177-183.
2. D. Rudolph,  *$\times 2$  and  $\times 3$  invariant measures and entropy*, Ergod. Th. and Dynam. Sys. **10** (1990), 395-406.

**MULTIPLE RECURRENCE  
AND SZEMEREDI'S THEOREM**

In this chapter we shall present a famous application of ergodic theory to number theory. The theorem we shall prove is an improvement on Van der Waerden's theorem (Theorem 2.1) originally proved by Szemerédi. The proof we give is due to Furstenberg.

**16.1 Szemerédi's theorem on arithmetic progressions**

To state the theorem we begin with the following definition.

DEFINITION. We say that a set of integers  $\mathcal{N} \subset \mathbb{Z}$  has *positive density* if

$$\delta(\mathcal{N}) := \limsup_{N \rightarrow +\infty} \frac{1}{2N+1} \text{Card}\{-N \leq n \leq N : n \in \mathcal{N}\} > 0.$$

(This quantity is called the (*upper*) *density* of  $\mathcal{N}$ .)

The following result generalizes the result of Van der Waerden in chapter 2.

THEOREM 16.1 (SZEMEREDI). *If  $\mathcal{N} \subset \mathbb{Z}$  has positive density then for all  $k \geq 1$  there exists an arithmetic progression of length  $k$  in  $\mathcal{N}$  (i.e.  $\forall k \geq 1, \exists a, b \in \mathbb{Z}, b \neq 0$  such that  $a + ib \in \mathcal{N}, i = 0, \dots, k-1$ ).*

REMARKS.

- (i) It is clear from the definition of  $\delta$  that in the special case  $\mathcal{N} = \mathbb{Z}$  we have that  $\delta(\mathbb{Z}) = 1$ . Moreover, if we consider a finite partition of the integers  $\mathbb{Z} = \mathcal{N}_1 \cup \dots \cup \mathcal{N}_k$  then  $\sup_{1 \leq i \leq k} \delta(\mathcal{N}_i) \geq \frac{1}{k}$ , and thus at least one of the elements in this partition has positive density. Thus Van der Waerden's theorem can be considered as a corollary to Szemerédi's theorem.
- (ii) Finite sets obviously have density zero. The set of odd (or even) numbers has density  $\frac{1}{2}$ .
- (iii) The prime numbers have zero density (i.e.  $\delta(\{\text{primes}\}) = 0$ ) and so Szemerédi's theorem does not apply. It is an open problem as to whether the conclusion holds for this set.



**A historical comment.** The above result was conjectured by Erdős and Turan in 1936. For the special case of  $k = 3$  the above theorem was proved by Roth in 1952. The special case  $k = 4$  was then proved by Szemerédi in 1969 before proving the general theorem in 1975.

In 1977 Furstenberg gave an alternative proof using ergodic theory.

### 16.2 An ergodic proof of Szemerédi's theorem

The key to the ergodic theory proof of Szemerédi's theorem is the following multi-dimensional version of the Poincaré recurrence theorem.

**THEOREM 16.2.** *Let  $T : (X, \mathcal{B}) \rightarrow (X, \mathcal{B})$  be a measurable transformation and let  $\mu$  be a  $T$ -invariant probability measure. For any  $A \in \mathcal{B}$  with  $\mu(A) > 0$  and  $k \geq 1$  there exists  $n \geq 1$  such that*

$$\mu(A \cap T^{-n}A \cap T^{-2n}A \cap \dots \cap T^{-kn}A) > 0. \quad (MR)$$

**NOTATION.** We can say that the set  $A$  satisfies the *multiple recurrence property* when the condition (MR) holds. A stronger condition, which we refer to as *uniform multiple recurrence*, is when for all  $k \geq 1$  the following condition holds:

$$\liminf_{N \geq 1} \frac{1}{N} \sum_{n=1}^N \mu(A \cap T^{-n}A \cap T^{-2n}A \cap \dots \cap T^{-kn}A) > 0. \quad (UMR)$$

**PROOF OF THEOREM 16.1 ASSUMING THEOREM 16.2.** Let  $X = \prod_{n=-\infty}^{\infty} \{0, 1\}$  and let  $\sigma : X \rightarrow X$  be the associated shift map (i.e.  $(\sigma x)_n = x_{n-1}$ ).

Given  $\mathcal{N} \subset \mathbb{N}$  we can define a sequence  $x = (x_n)_{n \in \mathbb{Z}} \in X$  by

$$x_n = \begin{cases} 1 & \text{if } n \in \mathcal{N}, \\ 0 & \text{if } n \notin \mathcal{N}. \end{cases}$$

For any  $n \geq 1$  we can define a probability measure  $\mu_n$  on  $X$  by

$$\mu_n = \frac{1}{2n+1} \sum_{i=-n}^{i=n} \delta_{\sigma^i x}.$$

Since the space of probability measures on  $X$  is compact (in the weak-star topology) we can choose a limit point  $\mu$  of  $\{\mu_n : n \geq 1\}$ . This will have the following properties:

- (a)  $\mu$  is a  $\sigma$ -invariant probability measure (since for any continuous function  $f : X \rightarrow \mathbb{R}$  we have that

$$\begin{aligned} \int f \sigma d\mu &= \lim_{n \rightarrow +\infty} \frac{1}{2n+1} \sum_{i=-n+1}^{i=n+1} f(\sigma^i x) \\ &= \lim_{n \rightarrow +\infty} \frac{1}{2n+1} \sum_{i=-n}^{i=n} f(\sigma^i x) = \int f d\mu; \end{aligned}$$

- (b) If we choose  $A = \{y = (y_n)_{n \in \mathbb{Z}} : y_0 = 1\}$  then  $\mu(A) > 0$  (since  $\mu(A) = \limsup_{n \rightarrow +\infty} \mu_n(A) = \limsup_{n \rightarrow +\infty} \frac{1}{2n+1} (\text{Card}\{\mathcal{N} \cap [-n, n]\}) = \delta(\mathcal{N}) > 0$ ).

We can now apply Theorem 16.2 to  $\sigma : X \rightarrow X$ , the measure  $\mu$  and the set  $A$ , to deduce that  $\forall k \geq 1, \exists n \geq 1$  with

$$\mu(A \cap \sigma^{-n}A \cap \sigma^{-2n}A \cap \dots \cap \sigma^{-kn}A) > 0.$$

From the construction of  $\mu$  we see that for (open and closed) sets  $A \cap \sigma^{-n}A \cap \sigma^{-2n}A \cap \dots \cap \sigma^{-kn}A$  there are values of  $N \geq 1$  such that  $\mu_N(A \cap \sigma^{-n}A \cap \sigma^{-2n}A \cap \dots \cap \sigma^{-kn}A) > 0$ . In particular, for some  $-N \leq i \leq N$  we have that  $\sigma^i x \in A \cap \sigma^{-n}A \cap \sigma^{-2n}A \cap \dots \cap \sigma^{-kn}A$ .

Finally, from the definition of  $x$  this means that the arithmetic progression  $i + jk \in \mathcal{N}$ , for  $j = 0, \dots, n$ . ■

### 16.3 The proof of Theorem 16.2

In this section we present an overview of the proof of Theorem 16.2. Details of the technical propositions used are presented in the appendix to this section.

**16.3.1: (UMR) for weak-mixing systems, weak-mixing extensions and compact systems.** The (UMR) condition is easier to establish if we assume that  $T : (X, \mathcal{B}) \rightarrow (X, \mathcal{B})$  under additional assumptions. For example:

**PROPOSITION 16.3.** *Let  $T : (X, \mathcal{B}) \rightarrow (X, \mathcal{B})$  be weak-mixing; then  $T$  satisfies (UMR).*

To introduce the definition of a weak-mixing extension we first recall a useful fact on skew products (which can be found in the work of Rohlin [4]).

**LEMMA (ROHLIN).** *Let  $T : (X, \mathcal{B}) \rightarrow (X, \mathcal{B})$  be a measurable map, and let  $\mu$  be a  $T$ -invariant ergodic measure. Given a  $T$ -invariant sub-sigma-algebra  $\mathcal{A} \subset \mathcal{B}$  (i.e. if  $A \in \mathcal{A}$  then  $T^{-1}A \in \mathcal{A}$ ) there exists a skew product*

$$\begin{cases} S : X_1 \times X_2 \rightarrow X_1 \times X_2, \\ S(x_1, x_2) = (S_1(x_1), S_2(x_1, x_2)) \end{cases}$$

(with respect to measure spaces  $(X_1, \mathcal{B}_1, \mu_1)$  and  $(X_2, \mathcal{B}_2, \mu_2)$ ) such that

- (i) there exists an isomorphism  $\psi : (X_1 \times X_2, \mathcal{B}_1 \times \mathcal{B}_2, \mu_1 \times \mu_2) \rightarrow (X, \mathcal{B}, \mu)$  between  $S$  and  $T$ ,
- (ii) the images  $\psi(A \times X_2)$ ,  $A \in \mathcal{B}_1$ , correspond to sets in  $\mathcal{A} \subset \mathcal{B}$ ,
- (iii)  $S_2(x_1, \cdot) : X_2 \rightarrow X_2$  preserves  $\mu_2$  (for a.e.  $(\mu_1) x_1$ ).

The property of sub-sigma-algebras we want to specify is closely related to the definition of weak-mixing (cf. chapter 11) and makes use of the above lemma of Rohlin.

We can identify the quadruple  $\mathcal{A} = (X, \mathcal{A}, \mu, T)$  with the quadruple  $(X_1, \mathcal{B}_1, \mu_1, S_1)$ .

DEFINITION. Let  $T : (X, \mathcal{B}) \rightarrow (X, \mathcal{B})$  be a measurable map and let  $\mu$  be an ergodic  $T$ -invariant probability measure. Given a  $T$ -invariant sub-sigma-algebra  $\mathcal{A} \subset \mathcal{B}$  we define the  $\mathcal{A}$ -cartesian product to be the skew product  $T \times_{\mathcal{A}} T := S_1 \times (S_2 \times S_2)$ , i.e.

$$\begin{cases} S_1 \times (S_2 \times S_2) : X_1 \times (X_2 \times X_2) \rightarrow X_1 \times (X_2 \times X_2), \\ S_1 \times (S_2 \times S_2)(x_1, x_2, y_2) = (S_1(x_1), S_2(x_1, x_2), S_2(x_1, y_2)). \end{cases}$$

with the product sigma-algebra  $\mathcal{B}_1 \times \mathcal{B}_2 \times \mathcal{B}_2$  and product measure  $\mu_1 \times \mu_2 \times \mu_2$ .

We say that  $T : (X, \mathcal{B}, \mu) \rightarrow (X, \mathcal{B}, \mu)$  is *weak-mixing (relative to  $\mathcal{A}$ )* if the  $\mathcal{A}$ -cartesian product  $T \times_{\mathcal{A}} T$  is ergodic.

REMARK. If we take  $\mathcal{A}$  to be the trivial sigma-algebra  $\{X, \emptyset\}$  then in the Rohlin lemma  $X_1$  is trivial,  $X_2 = X$  and  $T \times_{\mathcal{A}} T = T \times T$ . We see by Proposition 11.8 that  $T$  being weak-mixing relative to  $\{X, \emptyset\}$  is equivalent to  $T : X \rightarrow X$  being weak-mixing.

The next result shows that the (UMR) property is preserved under weak-mixing extensions.

PROPOSITION 16.4. *Let  $T : (X, \mathcal{B}, \mu) \rightarrow (X, \mathcal{B}, \mu)$  be ergodic. Let  $T : (X, \mathcal{B}) \rightarrow (X, \mathcal{B})$  be a weak-mixing extension of  $T : (X, \mathcal{A}) \rightarrow (X, \mathcal{A})$  which satisfies (UMR); then  $T : (X, \mathcal{B}) \rightarrow (X, \mathcal{B})$  also satisfies (UMR).*

Notice that when  $\mathcal{A} = \{X, \emptyset\}$  then Proposition 16.4 reduces to Proposition 16.3.

We now introduce a property that complements that of weak mixing.

DEFINITION. The system  $(X, \mathcal{B})$  is *compact* if for every  $f \in L^2(X, \mu)$  the closure in  $L^2(X, \mu)$  of the orbit  $\{f \circ T^n\}_{n \geq 0}$  is compact.

EXAMPLE. Let  $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  be a rotation  $T(x) = x + \alpha \pmod{1}$  ( $\alpha \in \mathbb{R}$ ); then this is a compact system. (The same applies for any rotation on a compact group.)

The next proposition complements Proposition 16.3 by giving a second special case in which (UMR) can be readily established.

PROPOSITION 16.5. *If the system  $T : (X, \mathcal{B}) \rightarrow (X, \mathcal{B})$  is compact then property (UMR) holds.*

**16.3.2: The non-weak-mixing case.** In view of Proposition 16.3 we have to concentrate on the alternative where  $T : (X, \mathcal{B}) \rightarrow (X, \mathcal{B})$  is not weak-mixing.

PROPOSITION 16.6. *If the system  $T : (X, \mathcal{B}) \rightarrow (X, \mathcal{B})$  is not weak-mixing then there exists a non-trivial compact factor  $T : (X, \mathcal{B}_1) \rightarrow (X, \mathcal{B}_1)$  which satisfies property (UMR).*

DEFINITION. We denote by  $\mathcal{F}$  the family of factors  $\mathcal{A} \subset \mathcal{B}$  for which  $T : (X, \mathcal{A}) \rightarrow (X, \mathcal{A})$  satisfies property (UMR).

By Proposition 16.6,  $\mathcal{F}$  is non-empty (and non-trivial). The main result on  $\mathcal{F}$  is the following.

PROPOSITION 16.7. *There exists a maximal factor in  $\mathcal{F}$  (i.e.  $\exists \mathcal{B}_\infty \in \mathcal{F}$  such that  $\mathcal{B}_1 \subset \mathcal{B}_\infty, \forall \mathcal{B}_1 \in \mathcal{F}$ ).*

**16.3.3: (UMR) for compact extensions.** Given  $\mathcal{A} \subset \mathcal{B}$ , a  $T$ -invariant sigma-algebra, and thus by Rohlin's lemma a quadruple  $(X_1, \mathcal{B}_1, \mu_1, S_1)$  identified with  $(X, \mathcal{A}, \mu, T)$ , and  $x_1 \in X$  we write  $\mu_{x_1}(B) = \mathbb{E}(\chi_B | \mathcal{A})(x_1)$ . We need the following definition.

DEFINITION. A function  $f \in L^2(X, \mathcal{B}, \mu)$  is *almost periodic* (AP) relative to the factor  $\mathcal{A} \subset \mathcal{B}$  if for every  $\delta > 0$  there exist functions  $F_1, \dots, F_n \in L^2(X, \mathcal{A}, \mu)$  such that for every  $j \in \mathbb{N}$  we have  $\inf_{1 \leq s \leq n} \|f \circ T^j - F_s\|_{L^2(\mu_y)} < \delta$  for almost all  $x_1 \in X$ .

We let  $\mathcal{P}(\mathcal{A})$  denote the set of all almost periodic functions in  $L^2(X, \mu)$ , relative to  $\mathcal{A}$ .

DEFINITION. If  $\mathcal{P}(\mathcal{A}) \subset L^2(X, \mathcal{B}, \mu)$  is dense then  $T : (X, \mathcal{B}) \rightarrow (X, \mathcal{B})$  is a *compact extension* of  $T : (X, \mathcal{A}) \rightarrow (X, \mathcal{A})$

The following result complements Proposition 16.4 by showing that (UMR) is also preserved under compact extensions.

PROPOSITION 16.8. *Let  $T : (X, \mathcal{C}) \rightarrow (X, \mathcal{C})$  be a compact extension of  $T : (X, \mathcal{A}) \rightarrow (X, \mathcal{A})$  which satisfies (UMR); then  $T : (X, \mathcal{C}) \rightarrow (X, \mathcal{C})$  also satisfies (UMR).*

**16.3.4: The last step.** The final ingredient in the proof of Theorem 16.2 is the following.

PROPOSITION 16.9. *Let  $T : (X, \mathcal{B}) \rightarrow (X, \mathcal{B})$  be a not relatively weak-mixing extension of  $T : (X, \mathcal{A}) \rightarrow (X, \mathcal{A})$ ; then there exists an intermediate factor  $T : (X, \mathcal{C}) \rightarrow (X, \mathcal{C})$  which is a non-trivial compact extension of  $T : (X, \mathcal{A}) \rightarrow (X, \mathcal{A})$ .*

To finish the proof of Theorem 16.2 we proceed as follows. If  $T : (X, \mathcal{B}) \rightarrow (X, \mathcal{B})$  is weak-mixing, then we are done (by Proposition 16.3). Alternatively, if  $T : (X, \mathcal{B}) \rightarrow (X, \mathcal{B})$  is not weak-mixing then there are a further two

possibilities: either the maximal sigma-algebra  $\mathcal{B}_\infty$  (given by Proposition 16.7) is equal to  $\mathcal{B}$  or it is a non-trivial sub-sigma-algebra of  $\mathcal{B}$ . In the first case the proof is complete because  $\mathcal{B} = \mathcal{B}_\infty \in \mathcal{F}$  satisfies (UMR) by definition. In the second case, we may assume that the maximal factor  $\mathcal{B}_\infty \subset \mathcal{B}$  is not relatively weak-mixing since otherwise we immediately have that  $T : (X, \mathcal{B}) \rightarrow (X, \mathcal{B})$  satisfies (UMR), by Proposition 16.4. Under this assumption there exists (by Proposition 16.9) an intermediate non-trivial extension  $T : (X, \mathcal{C}) \rightarrow (X, \mathcal{C})$  (i.e.  $\mathcal{B}_\infty \subset \mathcal{C} \subset \mathcal{B}$ ) which is relatively compact (see diagram).

$$\begin{array}{c} \text{compact} \\ \underbrace{\mathcal{B}_\infty \subset \mathcal{C} \subset \mathcal{B}} \\ \text{not relatively weak-mixing} \end{array}$$

But then (by Proposition 16.8) we know that  $T : (X, \mathcal{C}) \rightarrow (X, \mathcal{C})$  satisfies property (UMR). Since by construction  $\mathcal{C} \neq \mathcal{B}_\infty$  this gives a contradiction to  $\mathcal{B}_\infty$  being maximal, and so the second case cannot occur. Thus the proof of Theorem 16.2 is complete. ■

### 16.4 Appendix to section 16.3.

In this rather lengthy appendix, we shall give the omitted proofs of propositions from section 16.3.

**16.4.1 The proofs of Propositions 16.3 and 16.4.** Proposition 16.3 is a special case of Proposition 16.4. To simplify notation, we shall write  $E(f|\mathcal{A})$  for  $E(f|\mathcal{A})\psi^{-1}$ .

LEMMA 16.10. *Let  $T : (X, \mathcal{B}) \rightarrow (X, \mathcal{B})$  be a relatively weak-mixing extension of  $T : (X, \mathcal{A}) \rightarrow (X, \mathcal{A})$  and  $f, g \in L^\infty(\mu)$ . Then*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int \{E(f(gT^n)|\mathcal{A}) - E(f|\mathcal{A})E(gT^n|\mathcal{A})\}^2 d\mu_1 = 0.$$

*In particular, for the characteristic functions  $\chi_A$  and  $\chi_B$  we have*

$$\int \frac{1}{N} \sum_{n=1}^N |\mu_{x_1}(A \cap T^{-n}B) - \mu_{x_1}(A)\mu_{x_1}(B)| d\mu_1(x_1) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

PROOF. We can assume that  $E(f|\mathcal{A}) = 0$  (otherwise we simply replace  $f$  by  $f - E(f|\mathcal{A})$ ).

Define  $f \otimes_{\mathcal{A}} f : X_1 \times (X_2 \times X_2) \rightarrow X_1 \times (X_2 \times X_2)$  by  $f \otimes_{\mathcal{A}} f(x_1, x_2, y_2) = f(x_1, x_2)f(x_1, y_2)$ ; then

$$\begin{aligned} & \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=1}^N \int \{E(f \cdot g \circ T^n | \mathcal{A})\}^2 d\mu_1 \\ &= \lim_{N \rightarrow +\infty} \int (f \otimes_{\mathcal{A}} f) \cdot \left( \frac{1}{N} \sum_{n=1}^N g \otimes_{\mathcal{A}} g(T \times_{\mathcal{A}} T)^n \right) d(\mu_1 \times \mu_2 \times \mu_2). \end{aligned} \quad (16.1)$$

By assumption  $T \times_{\mathcal{A}} T$  is ergodic and so by the Birkhoff ergodic theorem

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{N} \sum_{n=1}^N g \otimes_{\mathcal{A}} g(T \times_{\mathcal{A}} T)^n &= \int g \otimes_{\mathcal{A}} g d(\mu_1 \times \mu_2 \times \mu_2) \\ &= \int \{E(g | \mathcal{A})\}^2 d\mu_1 \end{aligned}$$

(where the last line follows from the definitions of  $g \otimes_{\mathcal{A}} g$ ).

Hence the limit in (16.1) is

$$\int (f \otimes_{\mathcal{A}} f) d(\mu_1 \times \mu_2 \times \mu_2) \cdot \left\{ \int \{E(g | \mathcal{A})\}^2 d\mu_1 \right\}^2 = 0$$

since  $\int (f \otimes_{\mathcal{A}} f) d(\mu_1 \times \mu_2 \times \mu_2) = \int \{E(f | \mathcal{A})\}^2 d\mu_1 = 0$  (because we assumed  $E(f | \mathcal{A}) = 0$ ). ■

**LEMMA 16.11.** *Let  $T : (X, \beta) \rightarrow (X, \beta)$  be a relatively weak-mixing extension of  $\mathcal{A}$ . Then  $T \times_{\mathcal{A}} T$  is also relatively weak-mixing extension of  $\mathcal{A}$ .*

**PROOF.** This is an immediate consequence of Lemma 16.10 and the equality  $E(gf | \mathcal{A}) = gE(f | \mathcal{A})$ , whenever  $g \in L^1(X, \mathcal{A})$ . ■

**LEMMA 16.12.** *Let  $T : (X, C) \rightarrow (X, C)$  be a relatively weak-mixing extension of  $T : (X, \mathcal{A}) \rightarrow (X, \mathcal{A})$ . Then if  $f_l \in L^\infty(\mu)$ ,  $l = 1, \dots, k$ , we have,*

$$\begin{aligned} (1)_k \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int \left\{ E\left(\prod_{l=1}^k f_l T^{ln} | \mathcal{A}\right) - \prod_{l=1}^k E(f_l | \mathcal{A}) \right\}^2 d\mu_1 &= 0, \\ (2)_k \lim_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=1}^N \left( \prod_{l=1}^k f_l T^{ln} - \prod_{l=1}^k E(f_l | \mathcal{A}) T^{ln} \right) \right\|_{L^2(\mu)} &= 0. \end{aligned}$$

**PROOF.** First we show (2)<sub>k</sub> by induction. For  $k = 1$  the limit (2)<sub>1</sub> follows from the ergodic theorem since

$$\begin{cases} \frac{1}{N} \sum_{n=1}^N f_1(T^n x) \rightarrow \int f_1 d\mu & \text{(in } L^2 \text{ norm),} \\ \frac{1}{N} \sum_{n=1}^N E(f_1 | \mathcal{A})(T^n x) \rightarrow \int E(f_1 | \mathcal{A}) d\mu = \int f_1 d\mu & \text{(in } L^2 \text{ norm).} \end{cases}$$

Assume for the inductive step that the result has been established for  $k-1$  functions (i.e.  $(2)_{k-1}$  is valid).

By the simple identity

$$\prod_{l=1}^k a_l - \prod_{l=1}^k b_l = \sum_{j=1}^k \left( \prod_{l=1}^{j-1} a_l \right) (a_j - b_j) \left( \prod_{l=j+1}^k b_l \right),$$

with  $a_1, \dots, a_k, b_1, \dots, b_k \in \mathbb{R}$  we can write

$$\begin{aligned} & \left( \int \left| \frac{1}{N} \sum_{n=1}^N (f_1(T^n x) f_2(T^{2n} x) \dots f_k(T^{kn} x)) \right. \right. \\ & \quad \left. \left. - E(f_1 | \mathcal{A})(T^n x) E(f_2 | \mathcal{A})(T^{2n} x) \dots E(f_k | \mathcal{A})(T^{kn} x) \right|^2 d\mu(x) \right)^{\frac{1}{2}} \\ & \leq \sum_{j=1}^k \left( \int \left| \frac{1}{N} \sum_{n=1}^N (E(f_1 | \mathcal{A})(T^n x) \dots E(f_{j-1} | \mathcal{A})(T^{(j-1)n} x)) \right. \right. \\ & \quad \left. \left. \times f_j(T^{jn} x) f_{j+1}(T^{(j+1)n} x) \dots f_k(T^{kn} x) \right|^2 d\mu(x) \right)^{\frac{1}{2}}. \end{aligned}$$

We fix a choice of  $1 \leq j \leq k$  and we may assume without loss of generality that  $E(f_j | \mathcal{A}) = 0$  (otherwise we need only replace  $f_j$  by  $f_j - E(f_j | \mathcal{A})$ ). Thus it suffices to show that

$$\alpha_j(N) := \int \left| \frac{1}{N} \sum_{n=1}^N \beta_j(n) \right|^2 d\mu(x) \rightarrow 0$$

as  $N \rightarrow +\infty$ , where

$$\beta_j(n) = E(f_1 | \mathcal{A})(T^n x) \dots E(f_{j-1} | \mathcal{A})(T^{(j-1)n} x) f_j(T^{jn} x) \dots f_k(T^{kn} x).$$

For any  $1 \leq m \leq N$  we can now bound

$$\begin{aligned} & \alpha_j(N) \\ & \leq \int \left| \frac{1}{N} \sum_{n=1}^N \left( \frac{1}{m} \sum_{i=0}^{m-1} \beta_j(n+i) \right) \right|^2 d\mu(x) + \frac{2m |f_1|_\infty \dots |f_k|_\infty}{N} \\ & \leq \frac{1}{N} \sum_{n=1}^N \left( \int \left| \frac{1}{m} \sum_{i=0}^{m-1} \beta_j(n+i) \right|^2 d\mu(x) \right) + \frac{2m |f_1|_\infty \dots |f_k|_\infty}{N} \\ & = \frac{1}{N} \sum_{n=1}^N \left( \frac{1}{m^2} \sum_{i, i'=0}^{m-1} \int \beta_j(n+i) \beta_j(n+i') d\mu(x) \right) + \frac{2m |f_1|_\infty \dots |f_k|_\infty}{N}. \end{aligned}$$

Collecting together similar terms we get that

$$\begin{aligned} \alpha_j(N) &\leq \frac{1}{N} \sum_{n=1}^N \sum_{r=-m}^m \frac{(m-|r|)}{m^2} \int \prod_{s=1}^{j-1} (E(f_s|\mathcal{A}) \cdot E(f_s|\mathcal{A}) \circ T^{sr}) (T^{(s-1)n}x) \\ &\quad \times \prod_{t=j}^k (f_t \cdot f_t \circ T^{tr}) (T^{(t-1)n}x) d\mu(x) \\ &\quad + \frac{2m|f_1|_\infty \cdots |f_k|_\infty}{N}. \end{aligned}$$

We can now write that

$$\begin{aligned} \limsup_{N \rightarrow +\infty} |\alpha_j(N)|^2 &\leq \sum_{r=-m}^m \frac{m-|r|}{m^2} \|E(f_1|\mathcal{A})\|_\infty^2 \cdots \|E(f_{j-1}|\mathcal{A})\|_\infty^2 \\ &\quad \times \|E(f_{j+1}|\mathcal{A})\|_\infty^2 \cdots \|E(f_k|\mathcal{A})\|_\infty^2 \\ &\quad \times \left| \int E(f_j \circ T^{jr} \cdot f_j|\mathcal{A}) d\mu \right|. \end{aligned}$$

since in the limit we can replace the terms  $\prod_{t=j+1}^k (f_t \cdot f_t \circ T^{tr}) (T^{(t-1)n}x)$  by the terms  $\prod_{t=j+1}^k (E(f_t|\mathcal{A}) \cdot E(f_t|\mathcal{A}) \circ T^{tr}) (T^{(t-1)n}x)$  using the inductive hypothesis.

Finally, we know that the averages over terms  $\int E(f_j|\mathcal{A}) \circ T^{jr} E(f_j|\mathcal{A}) d\mu$ , for  $-m \leq r \leq m$ , are small for large  $m$  (since using Lemma 16.10 we can show that the terms tend to 0 for a sub-sequence of density one). This completes the proof of  $(2)_k$ .

To prove  $(1)_k$  we proceed as follows. By  $(2)_k$  we know that for  $f \in L^\infty(\mu)$ ,

$$\begin{aligned} &\lim_{N \rightarrow +\infty} \left\{ \int f_0 \left( \frac{1}{N} \sum_{n=1}^N \prod_{l=1}^k f_l \circ T^{ln} \right) d\mu \right\} \\ &= \lim_{N \rightarrow +\infty} \left\{ \int f_0 \left( \frac{1}{N} \sum_{n=1}^N \prod_{l=1}^k E(f_l|\mathcal{A}) \circ T^{ln} \right) d\mu \right\}. \end{aligned}$$

Since  $T \times_{\mathcal{A}} T$  is also relatively weak-mixing (by Lemma 16.11) we can replace  $f_l$  by  $f_l \otimes_{\mathcal{A}} f_l$  and  $T$  by  $T \times_{\mathcal{A}} T$  so that we obtain

$$\begin{aligned} &\lim_{N \rightarrow +\infty} \left\{ \int f_0 \otimes_{\mathcal{A}} f_0 \left( \frac{1}{N} \sum_{n=1}^N \prod_{l=1}^k (f_l \otimes_{\mathcal{A}} f_l) \circ (T \times_{\mathcal{A}} T)^{ln} \right) d(\mu_1 \times \mu_2 \times \mu_2) \right\} \\ &= \lim_{N \rightarrow +\infty} \left\{ \int f_0 \otimes_{\mathcal{A}} f_0 \left( \frac{1}{N} \sum_{n=1}^N \prod_{l=1}^k E(f_l \otimes_{\mathcal{A}} f_l|\mathcal{A}) \circ (T \times_{\mathcal{A}} T)^{ln} \right) d(\mu_1 \times \mu_2 \right. \\ &\quad \left. \times \mu_2) \right\}. \end{aligned} \tag{16.2}$$



A similar argument to the proof of Lemma 16.10 allows us to see that the right hand side of (16.2) is equal to

$$\int \left\{ E \left( \prod_{l=1}^k f_l | \mathcal{A} \right) \right\}^2 d\mu_1 \cdot \int \{ E(f_0 | \mathcal{A}) \}^2 d\mu_1. \quad (16.3)$$

Let us assume for the present that  $E(f_0 | \mathcal{A}) = 0$ , then we see that (16.3) (and thus the right hand side of (16.2)) is identically zero. On the other hand, the left hand side of (16.2) is equal to

$$\begin{aligned} & \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=1}^N \int \left\{ E \left( f_0 \times \prod_{l=1}^k f_l \circ T^{ln} | \mathcal{A} \right) \right\}^2 d\mu_1 \\ &= \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=1}^N \int \left\{ E \left( \prod_{l=0}^k f_l \circ T^{ln} | \mathcal{A} \right) \right\}^2 d\mu_1 \end{aligned}$$

showing  $(1)_k$ .

It remains to consider the case that  $E(f_0 | \mathcal{A}) \neq 0$ . We then write  $f_0 = (f_0 - E(f_0 | \mathcal{A})) + E(f_0 | \mathcal{A})$ . The proof of  $(1)_k$  for  $(f_0 - E(f_0 | \mathcal{A}))$  is as above and for the proof of  $(1)_k$  for  $E(f_0 | \mathcal{A})$  we may write

$$\begin{aligned} & \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int \left\{ E \left( E(f_0 | \mathcal{A}) \prod_{l=1}^k f_l \circ T^{ln} | \mathcal{A} \right) - \prod_{l=0}^k E(f_l \circ T^{ln} | \mathcal{A}) \right\}^2 d\mu_1 \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int E(f_0 | \mathcal{A})^2 \left\{ E \left( \prod_{l=1}^k f_l \circ T^{ln} | \mathcal{A} \right) - \prod_{l=1}^k E(f_l \circ T^{ln} | \mathcal{A}) \right\}^2 d\mu_1 \\ &\leq \|E(f_0 | \mathcal{A})^2\|_\infty \\ &\quad \times \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int \left\{ E \left( \prod_{l=1}^k f_l \circ T^{ln} | \mathcal{A} \right) - \prod_{l=1}^k E(f_l \circ T^{ln} | \mathcal{A}) \right\}^2 d\mu_1 = 0 \end{aligned}$$

using  $(1)_{k-1}$ . ■

To complete the proof of Proposition 16.4 we proceed as follows. Let  $A \in \mathcal{B}$  with  $\mu(A) > 0$ . Let  $\epsilon > 0$  be a small number so that for  $A_1 = \{x : E(\chi_A | \mathcal{B}_1) \geq \epsilon\}$  we have  $\mu_1(A_1) > 0$ . It follows from Lemma 16.12 and  $E(\chi_A | \mathcal{B}_1) \geq \epsilon \chi_{A_1}$  that we have the following:

$$\frac{1}{N} \sum_{j=1}^N \mu(\cap_{l=0}^k T^{-lj} A) > \frac{1}{2} \epsilon^{k+1} \frac{1}{N} \sum_{j=1}^N \mu_1(\cap_{l=0}^k S_1^{-lj} A_1),$$

for all  $k > 0$ . ■

**16.4.2 The proof of Proposition 16.5.** Let  $f$  satisfy  $0 \leq f \leq 1$ . We first note that if  $F_0, \dots, F_k$  are measurable functions with  $0 \leq F_i \leq 1$  satisfying  $\|f - F_i\|_2 < \epsilon$  ( $i = 0, \dots, k$ ), then

$$\begin{aligned} \left| \int \prod_{l=0}^k F_l d\mu - \int f^{k+1} d\mu \right| &\leq \sum_{l=0}^k \int \prod_{j=0}^{l-1} F_j |F_l - f| f^{k-l} d\mu \\ &\leq \epsilon \sum_{l=0}^k \int \prod_{j=0}^{l-1} F_j f^{k-l} d\mu \\ &\leq (k+1)\epsilon. \end{aligned}$$

So if we put  $a = \int f^{k+1} d\mu > 0$  and choose  $\epsilon < a/(k+1)$ , we have  $\int \prod_{l=0}^k F_l d\mu \geq (k+1)\epsilon - a > 0$ .

LEMMA 16.13. *For a set of  $n$  of positive lower density,  $\|f \circ T^n \cdot f - f\|_2 < \frac{\epsilon}{k}$  for  $l = 1, \dots, k$ .*

PROOF. Since  $\text{cl}\{f \circ T^n\} \subset L^2(\mu)$  is compact we can find a finite set  $\{f \circ T^{m_1}, f \circ T^{m_2}, \dots, f \circ T^{m_r}\}$  which is  $\frac{\epsilon}{k}$ -separated, i.e.

$$\left( \int |f \circ T^{m_i} - f \circ T^{m_j}|^2 d\mu \right)^{\frac{1}{2}} \geq \frac{\epsilon}{k}, \quad \text{for } 1 \leq i < j \leq r.$$

We can assume that  $r$  is the maximal cardinality of all such subsets.

For all  $n \geq 0$  we have that  $\{f \circ T^{n+m_1}, f \circ T^{n+m_2}, \dots, f \circ T^{n+m_r}\}$  is again  $\frac{\epsilon}{k}$ -separated. For each  $n$  there exists  $1 \leq i(n) \leq r$  such that  $\left( \int |f - f \circ T^{n+m_{i(n)}}| d\mu \right)^{\frac{1}{2}} \geq \frac{\epsilon}{k}$ . In particular, the sequence  $\{n + m_{i(n)}\}_{n=0}^{\infty}$  is a sequence of positive lower density with the required property. ■

### 16.4.3 The proof of Proposition 16.6.

DEFINITION. A function  $f \in L^2(\mu)$  is *almost periodic* if its orbit closure is compact.

In particular,  $T : (X, \mathcal{B}) \rightarrow (X, \mathcal{B})$  is compact if every  $f \in L^2(X, \mathcal{B})$  is almost periodic.

LEMMA 16.14. *If  $T \times T : (X \times X, \mathcal{B} \times \mathcal{B}) \rightarrow (X \times X, \mathcal{B} \times \mathcal{B})$  is not ergodic, then there exists a non-constant function  $f \in L^2(\mu)$  which is almost periodic.*

PROOF. Since  $T \times T$  is not ergodic we can choose  $g(x, y)$  to be a non-constant  $(T \times T)$ -invariant function in  $L^2(\mu \times \mu)$ . In the case that  $T$  is ergodic, the existence of a non-constant almost periodic function is trivial, so we suppose that  $T$  is not ergodic. Define a metric  $d$  on  $X$  by

$$d(x, y) = \int |g(x, z) - g(y, z)| d\mu(z).$$

If we identify points that are 0 distance, then we have an invariant sub-sigma-algebra (i.e., factor) so that  $T$  acts on this factor space as an isometry. Furthermore if we can show that the metric space is totally bounded (i.e. for any  $\epsilon > 0$  there exist a finite number  $N(\epsilon)$  of points which are  $\epsilon$ -separated) then we can apply the well-known result: *for any function  $f$  defined on a totally bounded metric space the orbit closure of  $\{f \circ T^n\}$  is compact when  $T$  is an isometry.* Without loss of generality, we can assume that the  $T$ -invariant function  $\int g(x, y)d\mu(x)$ , which is constant by ergodicity, vanishes (else we subtract the constant from  $g$ ). Since  $g$  is not identically 0 there exists a function  $h \in L^2(\mu)$  satisfying  $\int g(x, y)h(y)d\mu(y) \neq 0$  on a set of positive measure. Now define a function  $H(x) = \int g(x, y)h(y)d\mu(y)$  which is non-constant. Then we see that the function  $H$  is the desired almost periodic function. In fact,  $H \circ T^n(x) = \int g(x, y)h \circ T^n(y)d\mu(y)$  (because of the  $T$ -invariance of  $\mu$ ) and the integral operator

$$\begin{cases} \mathcal{G} : L^2(\mu) \rightarrow L^2(\mu), \\ \mathcal{G}\phi(x) = \int g(x, y)\phi(y)d\mu(y) \end{cases}$$

is a compact operator. Since the closure of  $\{H \circ T^n\}_{n>0}$  coincides with the closure of  $\{\mathcal{G}(h \circ T^n)\}_{n>0}$  and the norms of  $h \circ T^n$  are constant, we have the desired result. ■

Now we suppose that  $T : (X, \mathcal{B}) \rightarrow (X, \mathcal{B})$  is not weak-mixing. By lemma 16.14 we have a non-constant function  $f \in L^2(X, \mathcal{B})$  which is almost periodic. Recall that a subset of a complete metric space has compact closure if and only if for any  $\epsilon > 0$  there are finitely many balls of radius less than  $\epsilon > 0$  which are cover the subset. It follows from this fact that the set of  $f \in L^2(X, \mathcal{B})$  which are almost periodic is a closed linear subspace of  $L^2(X, \mathcal{B})$ . In fact, we can see that the set of almost periodic functions is closed under the (lattice) operations  $(h_1, h_2) \mapsto \max(h_1, h_2)$  and  $(h_1, h_2) \mapsto \min(h_1, h_2)$ . Let  $\mathcal{B}_0$  be the smallest sigma-algebra of sets with respect to which  $f$  is measurable. In particular, every characteristic function  $\chi_A$  ( $A \in \mathcal{B}_0$ ) is also almost periodic. Let  $\mathcal{B}_1$  be the smallest sigma-algebra of sets with respect to which  $f, f \circ T, \dots, f \circ T^n, \dots$  are measurable. Since  $f$  is almost periodic if and only if  $f \circ T$  is almost periodic we know that each  $\chi_A$ ,  $A \in \mathcal{B}_1$ , is almost periodic. If for every  $A \in \mathcal{B}_1$  we know  $\chi_A \in \mathcal{P}(\mathcal{B}_1)$  then any  $f' \in L^2(X, \mathcal{B}_1)$  must be almost periodic because the set of almost periodic functions is a closed linear subspace of  $L^2(X, \mathcal{B})$  as we have mentioned above. Finally, we have a non-trivial  $T$ -invariant sigma-algebra  $\mathcal{B}_1$  so that the factor  $T : (X, \mathcal{B}_1) \rightarrow (X, \mathcal{B}_1)$  is compact.

Using Proposition 16.5 the proof of Proposition 16.6 is complete. ■

**16.4.4 The proof of Proposition 16.7.** We wish to apply Zorn's lemma. Thus it suffices to show that every chain contains a maximal element, i.e. if

$\{A_\alpha\} \subset \mathcal{F}$  is a totally ordered chain then  $\cup_\alpha A_\alpha \in \mathcal{F}$  (where this represents the sigma-algebra generated by the union).

Let  $A \in \cup_\alpha A_\alpha$  with  $\mu(A) > 0$  and fix  $k > 0$ . Take  $\rho = \frac{1}{2(k+1)}$ ; then for  $\alpha_0$  sufficiently large we can choose  $A'_0 \in \mathcal{A}_{\alpha_0}$  with

$$\mu(A \Delta A'_0) < \frac{1}{4} \rho \mu(A). \quad (16.4)$$

By assumption  $T : (X, \mathcal{A}_{\alpha_0}) \rightarrow (X, \mathcal{A}_{\alpha_0})$  satisfies (UMR). We can use Rohlin's lemma to find a system  $S_1 : (X_1, \mathcal{B}_1, \mu_1) \rightarrow (X_1, \mathcal{B}_1, \mu_1)$  and a map  $\pi : X \rightarrow X_1$  so that  $A_{\alpha_0} = \pi^{-1}(\mathcal{B}_1)$ . Let  $A''_0 = \pi(A'_0)$ ; then by (16.3) we have  $\mu(A'_0) \geq \mu(A) - \frac{1}{4} \rho \mu(A)$ . This implies that  $\mu_1(x_1 \in A''_0 : \mu_{x_1}(A) < 1 - \eta) < \frac{1}{4} \mu(A)$ .

Write  $A_0 = \{x_1 \in A''_0 : \mu_{x_1}(A) > 1 - \eta\}$ . Then  $A_0 \in \mathcal{B}_1$  and

$$\mu_1(A_0) > \mu_1(A''_0) - \frac{1}{4} \mu(A) = \mu(A'_0) - \frac{1}{4} \mu(A) > \frac{1}{2} \mu(A).$$

By hypothesis  $T : (X, \mathcal{A}_{\alpha_0}) \rightarrow (X, \mathcal{A}_{\alpha_0})$  satisfies (UMR). We claim that for every  $j > 0$

$$\frac{1}{2} \mu_1(A_0 \cap S_1^{-j} A_0 \cap \dots \cap S_1^{-kj} A_0) < \mu(A \cap T^{-j} A \cap \dots \cap T^{-kj} A) \quad (16.5)$$

To see this it suffices to show that for  $x_1 \in A_0 \cap S_1^{-j} A_0 \cap \dots \cap S_1^{-kj} A_0$

$$\mu_{x_1}(A \cap T^{-j} A \cap \dots \cap T^{-kj} A) > \frac{1}{2} \quad (16.6)$$

because (16.5) follows from (16.6) by integration. But if  $x_1 \in S_1^{-lj} A_0$ ,  $l = 0, \dots, k$ , then by definition of  $A_0$  we see that  $\mu_{x_1}(S_1^{-lj} A_0) > 1 - \eta$  and (16.6) follows easily.

Averaging over  $j = 1, \dots, N$  and letting  $N \rightarrow +\infty$  we get

$$\liminf_{N \rightarrow +\infty} \frac{1}{N} \sum_{j=1}^N \mu(A \cap T^{-j} A \cap \dots \cap T^{-kj} A) > 0.$$

■

**16.4.5 Proof of Proposition 16.8.** We shall show that for  $A \in \mathcal{C}$  with  $\mu(A) > 0$  and for  $k > 0$  we have

$$\liminf_{N \rightarrow +\infty} \frac{1}{N} \sum_{j=1}^N \mu(\cap_{l=1}^k T^{-jl} A) > 0. \quad (16.7)$$

Since the above inequality follows for any subset of  $\mathcal{A}$ , we can assume (without loss of generality) that  $\mu_{x_1}(A) \geq \frac{1}{2} \mu(A)$  for  $x_1 \in A_1$  with  $\mu_1(A_1) > \frac{1}{2} \mu(A)$  and  $\mu_{x_1}(A) = 0$  for  $x_1 \notin A_1$  (by removing from  $A$  those fibres projecting to  $x_1$  satisfying  $\mu_{x_1}(A) \leq \frac{1}{2} \mu(A)$ ). The claim follows.

LEMMA 16.15.  $\chi_A$  is almost periodic

PROOF. Compactness of  $(X, \mathcal{C})$  implies that for every  $\epsilon > 0$  there exists an almost periodic function  $f'$  such that  $(\int |f - f'|^2 d\mu)^{\frac{1}{2}} < \epsilon^2$ . This means that the set  $E_\epsilon = \{x_1 \in X_1 : \int |f - f'|^2 d\mu_{x_1}\} \geq \epsilon^2\}$  satisfies  $\mu_1(E_\epsilon) < \epsilon^2$ . (If this were not the case we would have a contradiction since

$$\begin{aligned} \epsilon^4 &> \int \int |f - f'| d\mu_{x_1} d\mu_1(x_1) \\ &\geq \int_{E_\epsilon} \int |f - f'|^2 d\mu_{x_1} d\mu_1(x_1) \\ &\geq \epsilon^2 \mu_1(E_\epsilon) \geq \epsilon^4. \end{aligned}$$

Let  $A_\epsilon = A \cap E_\epsilon^c$  then on every fibre and for every  $j > 0$  either

- (1)  $\int |\chi_{A_\epsilon} \circ T^j - f' \circ T^j|^2 d\mu_{x_1} < \epsilon^2$ , or
- (2)  $\int |\chi_{A_\epsilon} \circ T^j| d\mu_{x_1} = 0$ .

Since  $f'$  is almost periodic we have from the definition that for each  $\delta > 0$  there exist  $F_1, \dots, F_n \in L^2(\mu)$  such that  $\inf_{0 \leq s \leq n} (\int |\chi_{A_\epsilon} \circ T^j - F_s|^2 d\mu_{x_1})^{\frac{1}{2}} < \delta + \epsilon$ , for almost all  $x_1$  and  $j > 0$ .

We can replace  $A_\epsilon$  by the intersection in  $\mathcal{A}$ ; then providing we choose a sequence  $\{\epsilon_j\}_{j>0}$  with  $\sum_{j>0} \epsilon_j^2 < +\infty$  the above procedure gives a sequence of sets  $A_{\epsilon_j}$ . The set  $A_\infty = \bigcap_{j>0} A_{\epsilon_j}$  has non-zero measure and the characteristic function  $\chi_A$  is almost periodic. ■

By the above lemma  $f = \chi_A$  is almost periodic. Given  $x_1$  let  $\oplus_{i=0}^k L^2(\mu_{x_1})$  have the norm

$$\|(f_1, \dots, f_k)\|_{x_1} = \max_{1 \leq i \leq k} \left( \int |f_i|^2 d\mu_{x_1} \right)^{\frac{1}{2}}.$$

Let  $V(k, f, x_1)$  be the set of vectors of the form  $(f, f \circ T^n, \dots, f \circ T^{kn})_y$ ,  $n \in \mathbb{N}$ . This has compact closure in  $L^2(\mu_{x_1})$ .

Denote by  $V^*(k, f, x_1)$  the subset of  $V(k, f, x_1)$  consisting of elements all of whose components are non-zero ( $\| \cdot \| \geq \frac{1}{2} (\mu(A))^{\frac{1}{2}}$ ).

Given  $x_1 \in A_1$  and  $\epsilon > 0$  we denote by  $M(\epsilon, x_1)$  the maximum cardinality of an  $\epsilon$ -separated set in  $V^*(k, f, x_1)$ . Since  $V^*(k, f, x_1)$  has compact closure,  $M(\epsilon, x_1)$  is bounded for  $x_1 \in A_1$ .

We choose  $\frac{\mu(A)}{10k} > \epsilon_0 > 0$ ,  $\eta > 0$  and  $A_2 \subset A_1$  with  $\mu_1(A_2) > 0$  such that  $M(\epsilon, x_1) = M$ ,  $\forall \epsilon_0 - \eta \leq \epsilon \leq \epsilon_0$ ,  $\forall x_1 \in A_2$ . Since  $M(\epsilon, x_1)$  is an integer valued monotone decreasing function of  $\epsilon$  it is locally constant (except at a countable set of  $\epsilon$ ) and is a measurable function.

Choose  $x_1^0 \in A_2$  and  $m_1, \dots, m_M$  for  $x_1^0$  such that  $\{(f, f \circ T^{m_j}, \dots, f \circ T^{km_j})_{x_1^0} : j = 1, \dots, M\}$  is a maximal  $\epsilon_0$ -separated set in  $V^*(k, f, x_1^0)$ .

As a function on the factor space  $X_1$

$$x_1 \mapsto \left( \int |f \circ T^{lm_i} - f \circ T^{lm_j}| d\mu_{x_1} \right)^{\frac{1}{2}}$$

is measurable (for  $1 \leq i < j \leq M$  and for  $l = 0, \dots, k$ ).

Without loss of generality we can assume that every neighbourhood of the image of  $x_1^0$  has positive  $\mu$ -measure in  $A_2$ . Now we let  $A_3 \subset A_2$ ,  $\mu(A_3) > 0$ , be the set such that

$$\left( \int |f \circ T^{lm_i} - f \circ T^{lm_j}| d\mu_{x_1^0} \right)^{\frac{1}{2}} - \left( \int |f \circ T^{lm_i} - f \circ T^{lm_j}| d\mu_{x_1} \right)^{\frac{1}{2}} < \eta$$

for  $y \in A_3$ .

Assume that  $T : (X, \mathcal{A}) \rightarrow (X, \mathcal{A})$  satisfies (UMR). Let  $n > 0$  satisfy  $\mu_1 \left( \bigcap_{l=0}^k S_1^{-nl} A_3 \right) > 0$  and fix  $x_1 \in \bigcap_{l=0}^k S_1^{-nl} A_3$ . It follows from the definition of  $A_3$  and  $V^*(k, f, x_1)$  that

$$A_3 \subset \bigcap_{l=0}^k \left( S_1^{-lm_j} A_1 \cap S^{-ln} A_1 \right)$$

for  $j = 1, \dots, M$ . Since the vectors  $\{(f, f \circ T^{n+m_j}, \dots, f \circ T^{k(n+m_j)})_{x_1}\}_{j=1}^M$  are  $(\epsilon_0 - \eta)$ -separated in  $V^*(k, f, x_1)$ , these form a maximal set which is  $(\epsilon_0 - \eta)$ -dense in  $V^*(k, f, x_1)$ .

Applying the argument in the proof of Lemma 16.13 to  $(f, \dots, f) \in V^*(k, f, x_1)$  we can choose  $j > 0$  such that  $(f, f \circ T^{n+m_j}, \dots, f \circ T^{k(n+m_j)})_{x_1}$  is  $\epsilon_0$ -close to it. Then

$$\mu_y \left( \bigcap_{l=0}^k T^{-l(n+m_j)} A \right) = \int \prod_{l=0}^k f \circ T^{n+m_j} d\mu_y \geq \int \chi_A d\mu_{x_1} - k\epsilon_0 > C\mu(A)$$

for some  $C > 0$ . Summing over  $j = 1, \dots, N$ , integrating over  $x_1 \in \bigcap_{l=0}^k S_1^{-ln} A_3$ ; then averaging over  $1 \leq n \leq N$  and letting  $N \rightarrow +\infty$  finally gives (16.2). ■

**16.4.6 Proof of Proposition 16.9.** We can assume  $T : (X, \mathcal{B}) \rightarrow (X, \mathcal{B})$  is ergodic and use Rohlin's lemma so that we identify  $(X, \mathcal{B})$  and  $(X_1 \times X_2, \mathcal{B}_1 \times \mathcal{B}_2)$ .

Assume that  $T \times_{\mathcal{A}} T$  is not ergodic; then there exists a bounded function  $g(x, y)$  on  $X \times_{\mathcal{A}} X := X_1 \times X_2 \times X_2$  ( $= \bigcup_{x_1 \in X_1} \pi^{-1}(x_1) \times \pi^{-1}(x_1)$ ) which is invariant under  $T \times_{\mathcal{A}} T$  but is not purely a function of  $x \in \pi^{-1}(x_1)$ , nor purely a function of  $y \in \pi^{-1}(x_1)$ . We write  $g(x, y) = g(x_1, x_2, x'_2)$  and define an integral operator

$$\begin{cases} \mathcal{G} : L^2(X, \mathcal{B}) \rightarrow L^2(X, \mathcal{B}), \\ \mathcal{G}\phi(x_1, x_2) := \int g(x_1, x_2, x'_2)\phi(x_1, x'_2)d\mu_2(x'_2) \end{cases}$$

which is a compact operator.

By analogy with the proof of Lemma 16.14 we have the following. There exists a function  $h \in L^2(X, \mathcal{B}, \mu)$  such that

$$\mathcal{G}h(x_1, x_2) := \int g(x_1, x_2, x'_2)h(x_1, x'_2)d\mu_2(x'_2)$$

is not a function of  $x_1$  alone and  $\text{cl}\{(\mathcal{G}h)T^j : j \geq 0\} = \text{cl}\{\mathcal{G}(h \circ T^j) : j \geq 0\} \subset L^2(X, \mathcal{B})$ . For  $\delta > 0$  there exists  $M = M(x_1, \delta) > 0$  such that  $\{(\mathcal{G}h) \circ T^j\}_{j=-M}^M$  is  $\epsilon$ -dense in  $\{(\mathcal{G}h) \circ T^j\}_{j \in \mathbb{N}}$  (in the  $L^2(\mu_{x_1})$ -norm).

For every  $\epsilon > 0$  we choose a sufficiently large  $M_{\epsilon, \delta}$  and a set  $E(\delta, \epsilon) \subset X_1$  with  $\mu_1(E(\delta, \epsilon)) < \epsilon$  such that  $M(x_1, \delta) < M_{\epsilon, \delta}$  for all  $x_1 \in E(\delta, \epsilon)^c$ . For a positive sequence  $\{\delta_j\}_{j \in \mathbb{N}}$  (with  $\delta_j \rightarrow 0$  as  $j \rightarrow +\infty$ ) and a positive sequence  $\{\epsilon_j\}_{j \in \mathbb{N}}$  (with  $\sum_{j=1}^{\infty} \epsilon_j$  sufficiently small) we define  $f : L^2(X, \mathcal{B}) \rightarrow L^2(X, \mathcal{B})$ :

$$f(x_1, x_2) = \begin{cases} 0 & \text{if } x_1 \in \cup_{j \in \mathbb{N}} E(\delta_j, \epsilon_j), \\ \mathcal{G}h & \text{otherwise.} \end{cases}$$

For each  $x_1$  the integral operator is compact and so  $\|f - \mathcal{G}h\|_2 \leq \|g\|_{L^\infty} \cdot \|h\|_{L^2(X)} \sum_{j \in \mathbb{N}} \epsilon_j$ . Furthermore, for every  $\delta > 0$  and large  $M$  the family  $\{0\} \cup \{(\mathcal{G}h) \circ T^j\}_{j=-M}^M$  is  $\delta$ -dense in  $\{fT^j\}_{j \in \mathbb{N}}$  (in the  $L^2(\mu_{x_1})$ -norm for every  $x_1$ ).

Let  $\mathcal{F}$  be the algebra spanned by  $\{\mathcal{G}h : g \in L^\infty(X \times_{\mathcal{A}} X), g(T \times_{\mathcal{A}} T) = g, h \in L^\infty(X)\}$ ; then  $\mathcal{F}$  is  $T$ -invariant and the almost periodic functions in  $\mathcal{F}$  are dense in  $\mathcal{F}$ .

Let  $\mathcal{B}^* \subset \mathcal{B}$  denote the smallest sigma algebra with respect to all of the elements of  $\mathcal{F}$  are measurable.  $\mathcal{B}^*$  is  $T$ -invariant and  $\mathcal{B}_1 \subset \mathcal{B}^*$ . Moreover,  $\mathcal{F} \subset L^2(X, \mathcal{B}^*, \mu)$  is dense and so the set of almost periodic functions is dense in  $L^2(X, \mathcal{B}^*, \mu)$ .

Finally, the desired compact factor corresponds to  $T : (X, \mathcal{B}^*) \rightarrow (X, \mathcal{B}^*)$ . ■

## 16.5 Comments and references

Most of the details are taken from the article of Furstenberg, Katznelson and Ornstein [2]. Alternative accounts of the proof appear in [1] and [5] and [3] contains an overview of the proof.

### References

1. H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, P.U.P., Princeton, N.J., 1981.
2. H. Furstenberg, I. Katznelson, and D. Ornstein, *The ergodic theoretical proof of Szemerédi's theorem*, Bull. Amer. Math. Soc. **7** (1982), 527-552.
3. K. Petersen, *Ergodic Theory*, C.U.P., Cambridge, 1983.
4. V. Rohlin, *Selected topics from the metric theory of dynamical systems*, Transl. Amer. Math. Soc. (2) **49** (1966), 171-209.
5. J.-P. Thouvenot, *La démonstration de Furstenberg du théorème de Szemerédi sur les progressions arithmétiques*, Séminaire Bourbaki, 30 (1977-78), 518.

- Abramov's lemma, 79
- Abramov's theorem
  - for measure theoretic entropy, 148
  - for topological entropy, 30
- adding machine, 68
- almost periodic functions, 171
- aperiodic interval transformation, 128
- aperiodic matrix, 23
- arithmetic progression, 11
  
- Baire's theorem, xi
- Bernoulli shifts, 88
- Birkhoff ergodic theorem
  - for ergodic theorems, 102
  - for invariant measures, 111
- Birkhoff recurrence theorem, 7
- Birkhoff-Poincaré theorem, 141
- Bochner-Herglotz theorem, 119
- Brouwer's plane theorem, 141
  
- $A$ -cartesian product, 164
- $\epsilon$ -chain, 139
- chain recurrent,
  - homeomorphism, 139
  - point, 139
- conditional expectation  $E(f|A)$ , 73
- conditional information  $I(f|A)$ , 74
- conjugacy, 30
  - and topological entropy, 31
- compact
  - extensions, 165
  - systems, 164
- continued fraction expansion, 108
- continued fraction transformation, 106
  
- density, upper (for sequences), 161
- Denjoy's theorem, 61
- dominated convergence theorem, xiii
  
- doubling map, 1
  - (measure theoretic) entropy, 79
  - invariant Haar measure, 68
  - topological entropy, 29
  - transitivity, 2
  
- entropy
  - of  $\alpha$  relative to  $T$   $h(T, \alpha)$ , 77
  - conditional entropy  $H(\alpha|A)$ , 76
  - (measure theoretic) entropy  $h_\mu(T)$ , 79
  - of a partition  $H(\alpha)$ , 76
- ergodic measures, 91
  - existence, 93
- exact endomorphism, 125
  
- first return time, 92
- fixed points, 31
- Franks' theorem, 141
  - (measure theoretic) entropy  $h_\mu(T)$ , 79
- Furstenberg's conjecture, 153
  
- Gauss measure, 106
- Generator
  - for a homeomorphism, 21
  - for a probability space, 79
  - strong: for a homeomorphism, 21; for a probability space, 79
- geodesic flow, 70
- geodesic preserving diffeomorphisms, 109
- horocycle flow, 70
- hyperbolic toral automorphisms, 47
  - periodic points, 48
  - semiconjugacy, 54
  - topological entropy, 50
  
- induced transformation, 95
- information function  $I(\alpha)$ , 73
- interval maps, 33



- Markov property, 39
- periodic points, 34
- piecewise expanding, 39
- topological entropy, 37
- intervals of monotonicity, 37
- invariant measures
  - for flows, 69
  - for transformations, 65; existence, 65
  - for  $Z^k$ -actions, 71
  - for interval maps, 127
- irreducible matrix, 5
- isomorphism, 88
  - and entropy, 88
- Kac's theorem, 92
- Kolmogorov extension theorem, xii
- law of large numbers, 99
- linear toral automorphism, 47
- Markov measures, 68
  - entropy, 81
  - strong mixing, 122
- martingale theorem, increasing, 83
- minimality, 6
  - equivalent definitions, 7
- multiple recurrence, 162
- natural extensions, 96
- von Neumann ergodic theorem, 99
  - generalization for weak-mixing  $T$ , 116
- normal numbers, 106
- partitions (for a probability space), 73
- periodic points, 31
  - for interval maps, 34
  - for Markov interval maps, 43
  - for toral automorphisms, 48
- Perron-Frobenius operator, 126
- Perron-Frobenius theorem, 23
- Poincaré recurrence theorem, 91
- positive definite sequences, 118
- pseudo-orbit, 52
- Radon-Nikodym derivative, xii
- refinement
  - of open covers, 19
  - of measurable partitions, 74
- Riemann-Lebesgue lemma, 119
- Riesz representation theorem, xii
- rotations (of the circle), 1
  - measure theoretic entropy, 80
  - minimality (for irrational rotations), 6
  - topological entropy, 28
  - transitivity (for irrational rotations), 3
- Rohlin tower, 96
- Rohlin's entropy formula, 133
- Rohlin's lemma (for sigma-algebras), 163
- rotations (on tori), 67
  - invariant measures, 67
- rotation numbers  $\rho(T)$ , 57
- Rudolph's theorem, 153
- semi-conjugacy, 31
  - for hyperbolic toral automorphisms, 54
  - for Markov interval maps, 42
  - and topological entropy, 31
- separated sets, 26
- sequential compactness, xii
- shadowing property, 52
- Shannon-McMillan-Breiman Theorem, 134
- Sharkovski's theorem, 36
- shift map, 1
- Sinai's lemma (for entropy), 79
- skew products, 95

- spanning sets, 26
- spectral measure, 119
- spectrum
  - continuous, 118
  - pure point, 118
- strong mixing, 122
- subadditive sequences, xiii
- subcover, 19
- subshift of finite type, 5
  - topological entropy, 25
  - transitivity, 5
- suspended flow, 70
- Szemerédi's theorem, 161
  - Furstenberg's proof, 162
  
- topological entropy
  - of an open cover  $H(\alpha)$ , 19
  - of  $T$  relative to  $\alpha$   $h(T, \alpha)$ , 20
  - of a transformation  $h(T)$ , 21
- transitive, 2
  - equivalent definitions, 4
- transitive, forwardly, 3
- transitive point, 3
  
- uniformly close homeomorphisms, 54
  
- variational principle, 147
- van der Waerden's theorem, 11
- weak mixing, 113
  - and density one sub-sequences, 114
  - equivalent definitions, 120
- weak-mixing extension, 164
  
- Zorn's lemma, xii

# ERRATA FOR "DYNAMICAL SYSTEMS AND ERGODIC THEORY"

## PRELIMINARIES

Theorem G should read:

**Theorem G (Radon-Nikodym).** *If  $\mu$  is absolutely continuous with respect to  $\nu$  then there exists a (unique) function  $f \in L^1(X, \mathcal{B}, d\nu)$  such that for any  $A \in \mathcal{B}$  we can write  $\mu(A) = \int_A f d\nu$ .*

## CHAPTER I

**page 1, Example 1:** It should read  $0 < \theta < 1$  (Tex error).

**page 2, line 2:**  $y = (y_n)_{n \in \mathbb{Z}}$  instead of  $x = (x_n)_{n \in \mathbb{Z}}$ , again.

**page 2, Proof of Lemma 1.1:** The superscript *ks* should become *ms*, say (Too many *ks*)

**page 3, Example 1:** It should read " $T^k x = \sum_{n=0}^{+\infty} \frac{x_{n+k-1}}{2^{n+1}} \pmod{1}$ ."

**page 7, Theorem 1.8:** The sets in  $\mathcal{E}$  should also be closed.

## CHAPTER 2

**page 12, Proof of Theorem 1.8:** We should define  $T_i := T^i = (\sigma \circ \dots \circ \sigma)$

**page 13, line 2:** It should read  $(T^b T_i^a x)_0 = x_{b+ia} = z_a \in \{1, \dots, k\}$

**page 13, bottom:** It should read "In particular,  $d_{\mathcal{X}_N}(S^{n_i} \underline{z}', z) \rightarrow 0$  as  $n_i \rightarrow +\infty$  where  $\underline{z} = (z, \dots, z), \underline{z}' = (T_N^{-n_i} z, \dots, T_N^{-n_i} z) \in \mathcal{D}_N$ ."

**page 14-15, Sublemma 2.2.4 (and proof):** Each  $\mathcal{D}$  is really  $\mathcal{D}_N$  (subscripts omitted)

**page 15, mid-page:** It should read "The proof of Theorem 2.3 is finished".

**page 16, after (2.3):** It should read:  $d_{\mathcal{D}_N}(S^n z, z') < \delta$  ( $S^n$  for  $T^n$ ).

**page 16, displayed equation (2.3):** It should read  $d_{\mathcal{X}_N}(\hat{T}^{n_{1j}} \circ \dots \circ \hat{T}^{n_{Nj}} z, \hat{T}^{n_{1j}} \circ \dots \circ \hat{T}^{n_{Nj}} z') < \frac{\epsilon}{4}$ . (no negative powers.)

**page 16, displayed equation (2.4):**  $d_{\mathcal{X}_N}(S^n(\hat{T}^{n_{1j}} \circ \dots \circ \hat{T}^{n_{Nj}} z), \hat{T}^{n_{1j}} \circ \dots \circ \hat{T}^{n_{Nj}} z') < \frac{\epsilon}{4}$ . ( $S^n$  for  $T^n$ )

**page 16, last displayed equation :**

$$\begin{aligned} d_{\mathcal{X}_N}(S^n y, x) &\leq d_{\mathcal{X}_N}(S^n y, \hat{T}^{n_{1j}} \circ \dots \circ \hat{T}^{n_{Nj}} z') + d_{\mathcal{X}_N}(\hat{T}^{n_{1j}} \circ \dots \circ \hat{T}^{n_{Nj}} z', x) \\ &\quad + d_{\mathcal{X}_N}(\hat{T}^{n_{1j}} \circ \dots \circ \hat{T}^{n_{Nj}} z, \hat{T}^{n_{1j}} \circ \dots \circ \hat{T}^{n_{Nj}} z') \\ &\quad + d_{\mathcal{X}_N}(\hat{T}^{n_{1j}} \circ \dots \circ \hat{T}^{n_{Nj}} z, x) \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{4} + \frac{\epsilon}{4} = \epsilon. \end{aligned}$$

(term missing)

**page 17, after (2.5):** Another  $\mathcal{D}$  which should be a  $\mathcal{D}_N$  (subscript omitted).

**Comments and references:** It should read "An account also appears in [5]."

### CHAPTER 3

**page 19, last displayed equation:** There is a  $\cap$  missing.

**page 22, first paragraph of proof:** It should read : This means that  $N(\bigvee_{i=0}^{k-1} T^{-i}(\bigvee_{n=0}^{N-1} T^{-n}\alpha)) \geq N(\bigvee_{i=0}^{k-1} T^{-i}\beta)$ , for  $k \geq 1$ .

**page 22, Definition:** It should read The matrix  $A$  is called *aperiodic* if  $\exists N > 0$ ,  $\forall 1 \leq i, j \leq k$ ,  $A^N(i, j) \geq 1$ . (Power of  $N$  missing)

**page 22, mid-page.** The displayed equation should read

$$\bigvee_{n=-N}^N \sigma^{-n}\alpha = \{[i_{-N}, \dots, i_0, \dots, i_N]_{-N}^N : i_{-N}, \dots, i_0, \dots, i_N \in \{1, \dots, k\}\}$$

and several  $\sigma^{-i}\alpha$  should change to  $\sigma^{-n}\alpha$ .

**page 23, mid-page.** It should read  $A(v_1, \dots, v_k)^T = (v'_1, \dots, v'_k)^T = (\sum_{i=1}^k A(i, 1)v_i, \dots, \sum_{i=1}^k A(i, k)v_i)$  (the transpose  $\cdot^T$  is omitted)

**page 24, end of proof.** It should read: However, elementary calculus shows that the supremum is realised where  $z = (\frac{bd}{ac})^{1/2}$  and thus

$$C < \frac{\nu^{1/2} - \nu^{-1/2}}{\nu^{1/2} + \nu^{-1/2}} < 1,$$

where  $\nu := \frac{ad}{bc} > 1$ . **page 24, bottom of page.** It should read "a contraction with respect to this metric". (The error seems to appear in the source article: G. Birkhoff, *Extensions of Jentzsch's theorem*, Trans. Amer. Math. Soc., 85 (1957) 219-227)

**page 25, second paragraph.** It should read " $C < 1$  is independent of the choices  $x, y$ ."

**page 25, last displayed equation.** It should read

$$H(\bigvee_{n=0}^{N-1} \sigma^{-n}\alpha) = \log \text{Card}(\bigvee_{n=0}^{N-1} \sigma^{-n}\alpha) = \dots$$

**page 26, near top.** It should read "... Jordan block matrices  $B_2, \dots, B_l, \dots$ "

**page 26, Lemma 3.6 (i)** It should read:  $s(n, \epsilon') \geq s(n, \epsilon)$  and  $r(n, \epsilon') \geq r(n, \epsilon)$ .

**page 27, near top** For  $r(n, \epsilon) \geq s(n, \epsilon)$  read  $s(n, \epsilon) \geq r(n, \epsilon)$ .

**page 28, near the bottom** It should read:  $\limsup_{n \rightarrow +\infty} \frac{1}{n} \log N(\bigvee_{i=0}^{n-1} T^{-i}\beta)$  (by Lemma 3.7 (H instead of N))

**page 30, near the top** It should read:  $r_{T^m}(n, \epsilon) \leq r_T(nm, \epsilon)$  ( $r$  missing).

**page 30, middle of page.** It should read  $R$  is also an  $(n, \epsilon)$ -spanning set for  $T^m$  ( $T^m$  not  $T^n$ ).

### CHAPTER 4

**page 33, Proof of Lemma 4.3** It should read "Let  $J_2 = [a, b]$ "

**page 35, Sublemma 4.4.1** It should read:

(ii)  $T^{n-1}(I_{n-1}) = J'$ ; and

(iii)  $T^n(I_{n-1}) \supset J''$ .

**page 36, second paragraph.** The two  $J'$  should be  $J''$ .

**page 37, Proof of Lemma 4.5.** The intervals of monotonicity for  $S_1 \circ S_2 : I \rightarrow I$  take the form  $J_j \cap S_2^{-1}(I_i)$ . Thus  $\mathcal{N}(S_1 \circ S_2) \leq \text{Card}\{(i, j) : J_j \cap S_2^{-1}(I_i) \neq \emptyset\} \leq N(S_1) \cdot N(S_2)$ .

**page 37, Remark.** The chain rule says:  $(T^n)'(x) = \prod_{i=0}^{n-1} T'(T^i x)$ .

**page 37, Proof of Theorem 4.6.** It should read:  $0 \leq r_i \leq n-1$  and  $0 \leq r_i \leq n-1$ .

**page 37, penultimate paragraph.** It is better to write:  $\{x_{i_0} < x_{i_1} < \dots < x_{i_{m-1}}\} \subset E_n$

**page 37, last paragraph.** It should read: there are at least  $\frac{m\epsilon}{C}$  intervals of monotonicity for  $T^r$ , i.e.  $\mathcal{N}(T^r) \geq \frac{m\epsilon}{C} \geq \frac{N}{Cn}\epsilon$ . Thus

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \log \mathcal{N}(T^n) \geq \lim_{n \rightarrow +\infty} \frac{1}{n} (\log s(n, \epsilon) -$$

**page 38, after first displayed equation.**

It should read.

$$\begin{aligned} m \left( \lim_{n \rightarrow +\infty} \frac{1}{n} \log \mathcal{N}(T^n) \right) &= m \left( \limsup_{n \geq 1} \frac{1}{n} \log \mathcal{N}(T^n) \right) \\ &\geq \limsup_{k \geq 1} \frac{1}{k} \log \mathcal{N}(S^k) \\ &= \lim_{k \rightarrow \infty} \frac{1}{k} \log \mathcal{N}(S^k) \end{aligned}$$

(since  $\left(\frac{m \log \mathcal{N}(S^k)}{k}\right)_{k \in \mathbb{N}}$  is a sub-sequence of  $\left(\frac{\log \mathcal{N}(T^n)}{n}\right)_{n \in \mathbb{N}}$ . Moreover, we can estimate

$$\begin{aligned} &m \left( \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathcal{N}(T^n) \right) \\ &\leq \lim_{n \rightarrow \infty} \frac{m}{n} \log \left( \mathcal{N}(S^{\lfloor \frac{n}{m} \rfloor}) \mathcal{N}(T^{n - \lfloor \frac{n}{m} \rfloor m}) \right) \quad (\text{by Lemma 4.5}) \\ &\leq \lim_{n \rightarrow \infty} \frac{m}{n} \log \left( \mathcal{N}(S^{\lfloor \frac{n}{m} \rfloor}) + \max_{0 \leq i \leq m-1} \{\log \mathcal{N}(T^i)\} \right) \\ &= \lim_{k \rightarrow \infty} \frac{1}{k} \log \mathcal{N}(S^k). \end{aligned}$$

**page 39, first displayed equation.** It should read

$$\geq \limsup_{k \rightarrow +\infty} \frac{1}{k} \log (\mathcal{N}(S^k)/3^k) \quad (\text{since } \mathcal{N}(S^k) \leq 3^k N(\sqrt[k]{S^{-r} \alpha}))$$

**page 42, Proof of (i).** It should read for every point  $x$  in the complement of the dense set  $\cup_{n \in \mathbb{Z}^+} T^{-n} \{x_0, \dots, x_k\}$

**page 42, Proof of (ii).** It should read  $\pi(w) = z$  (twice) instead of  $\pi(w) = x$ .

**page 43, Proof of Theorem 4.9.** We actually use Proposition 3.11.

**page 43, Proof of sublemma 4.9.1.** We actually use Proposition 3.5.

**page 44, Proof of sublemma 4.10.1.** For  $\lambda_0$  read  $\lambda_1$ .

## CHAPTER 5

**page 47** We should use  $GL(2, \mathbb{Z})$  instead of  $SL(2, \mathbb{Z})$ .

**page 47** To give the broadest definition of hyperbolicity we should use

Let  $A \in GL(2, \mathbb{Z})$  have eigenvalues  $\lambda_1, \lambda_2$  then if  $|\lambda_1| > 1 > |\lambda_2| (= \frac{1}{|\lambda_1|})$  we call the matrix  $A$  *hyperbolic*

**page 49, below figure 5.2** It should read: (where  $(x_1, x_2) \in [0, 1) \times [0, 1)$ ).

**page 50, near the top** Delete the upper bound on  $k$  (For the "boxes" below to cover we should require  $k \leq \frac{4}{\epsilon^2 \sin \theta}$ , where  $\theta$  is the angle between the eigenvectors.)

**page 50, Next paragraph add** "We can choose boxes such that  $\text{Box}(x_1^i, x_2^i)$ ,  $i = 1, \dots, n$  cover the torus."

**page 50, Proof of Sublemma 5.3.1** Change the first line(s) to: "For any point  $(z_1, z_2) \in \mathbb{R}^2/\mathbb{Z}^2$  we can choose some  $i = 1, \dots, k$  such that  $(z_1, z_2) \in \text{Box}(x_1^i, x_2^i)$  and so ..."

**page 51, displayed equation (5.2)** Change to:  $\lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow +\infty} \frac{1}{n} \log r(n, \epsilon) \leq \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow +\infty} \frac{1}{n} \log (k(2[|\lambda_1|^n] + 1))$  ■

**page 51, sublemma 5.3.2.** The statement should read: " $S$  is an  $(n, 2\epsilon/|\lambda_1|)$ -separated".

## CHAPTER 6

**page 59, proof of part (i).** In fact, it is the argument in part (ii).

**page 59, statement of Lemma 6.3.** We are assuming  $\rho$  is irrational.

**page 60, first line.** Assume  $n_1 > n_2$

**page 60, first paragraph.** There  $T$ s should be  $\hat{T}$  (i.e. hats missing).

**page 61, lower part of page.** It should read: "It is easy to see that if  $T: \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  is  $C^2$  with  $|T''|/|T'|$  is bounded then  $\text{Var}(\log |T'|)$  is finite."

**page 62, Statement of Sublemma 6.5.1.** The displayed equation should be  $|(T^{q_n})'(x)| \cdot |(T^{-q_n})'(x)| \geq C$ .

**page 62, Displayed equation, mid-page.** It should read  $|I_{q_n}| + |I_{-q_n}|$  (misplaced minus sign).

**page 62, last line.** It should read  $|T^{q_n} x - x| = \delta_n$ .

**page 63, displayed equation (6.4).** The second line should read  $\geq$  instead of  $=$ .

## CHAPTER 7

**page 67, first displayed equation .** In the first line the upper limit in the sum should be  $N_n^{(n)} - 1$ ; the second line should be  $f_k(T^{N_n^{(n)}}x) + \sum_{n=0}^{N_n^{(n)}-1} f_k(T^n x) - f_k(x)$  (reverse signs).

**page 68, item (2).** It should read "for every non-empty open set  $U \subset X$ "

**page 68, mid-page.** It should read  $\mu([z_0, \dots, z_n]) = \left(\frac{1}{2}\right)^{n+1}$  ( $n+1$  instead of  $n$ ).

**page 69, second line.** For  $(1, \dots, 1)$  read  $(1, \dots, 1)^T$  (i.e. the transpose).

## CHAPTER 8

**page 74, Lemma 8.1 .** Part (1) should read  $I(\alpha|\{\emptyset, X\})(x) = I(\alpha)(x)$ ;

Part(2) should read  $I(\alpha|\mathcal{A})(Tx) = I(T^{-1}\alpha|T^{-1}\mathcal{A})(x)$ .

**page 74, line before definition.** Delete the erroneous comment "(in particular,  $\hat{\alpha}$  is countable and consists of all unions of elements from  $\alpha$ )."

**page 75, last displayed equation.** It should read  $I(\alpha \vee \beta|\hat{\gamma})(x)$  (hat missing in  $\hat{\gamma}$ ).

**page 76, proof of lemma 8.3.** It should be corrected to:

"For part (4) we have

$$\begin{aligned} H(\alpha|\gamma) &= - \sum_{A \in \alpha, C \in \gamma} \mu(A \cap C) \log \left( \frac{\mu(A \cap C)}{\mu(C)} \right) \\ &= - \sum_{C \in \gamma} \mu(C) \left[ \sum_{A \in \alpha} \frac{\mu(A \cap C)}{\mu(C)} \log \left( \frac{\mu(A \cap C)}{\mu(C)} \right) \right] \\ &\leq - \sum_{A \in \alpha} \left[ \sum_{C \in \gamma} \mu(A \cap C) \right] \log \left[ \sum_{C \in \gamma} \mu(A \cap C) \right] \\ &\leq - \sum_{A \in \alpha} \mu(A) \log \mu(A) = H(\alpha) \end{aligned}$$

since for fixed  $A \in \alpha$  we can bound

$$- \sum_{C \in \gamma} \frac{\mu(A \cap C)}{\mu(C)} \log \left( \frac{\mu(A \cap C)}{\mu(C)} \right) \leq - \left[ \sum_{C \in \gamma} \mu(A \cap C) \right] \log \left[ \sum_{C \in \gamma} \mu(A \cap C) \right]$$

using concavity of  $t \mapsto -t \log t$ ."

## CHAPTER 9

**page 95, line 3 :** Replace  $X \times \mathbb{R}/\mathbb{Z}$  by  $X \times G$ .

**page 95, line 25 from below :**  $\mu(A) > 0$  should be  $0 < \mu(A) < \infty$ .

**page 96, line 5 from below :**  $\hat{T} : X \rightarrow X$  should be  $\hat{T} : \hat{X} \rightarrow \hat{X}$ .

**page 96, line 1 :** *natural extension* of  $X$  should be *natural extension* of  $T$ .

**page 96, line 2 :**  $\pi((x_n)_{n \in \mathbb{Z}^+}) = X$  should be  $\pi((x_n)_{n \in \mathbb{Z}^+}) = x_0$

## CHAPTER 10

**page 100, line 2 from below :**  $\int f(x)g(x)d\mu(x)$  should be removed.

**page 101, line 6 :**  $\epsilon \|g\|_2$  should be  $\sqrt{\epsilon} \|g\|_2$ .

**page 101, lines 11,13 :** Replace  $(\int f d\mu \int g d\mu - \epsilon, \int f d\mu \int g d\mu + \epsilon)$  should be  $(\int f_1 g d\mu - \epsilon, \int f_1 g d\mu + \epsilon)$ . **Mark ?** Hooever..??? Or are you assuming ergodicity ?

**page 102, line 5 from below :** We assume  $\min(f_+, f_-) = 0$ .

**page 105, line 9 :** We assume ergodicity.

**page 106, lines 17-18 :** Replace  $2^k, r_k$  by  $2^n, r_n$  respectively.

**page 107 :** (ii) should be corrected as follows. (ii) there exists  $D > 0$  such that  $\sup_{x,y,z \in I_k} \frac{|T''(x)|}{|T'(y)T'(z)|} \leq D, \forall 0 < x < 1$ . Then  $|\frac{\psi'_{k_0 \dots k_n}(x)}{\psi'_{k_0 \dots k_n}(y)}|$  is bounded from above by

$$\prod_{i=0}^n \left( 1 + \frac{D}{4^{[(n-i-1)/2]}} |T^n x - T^n y| \right) \leq C.$$

We assume in (a) that  $l(E \cap I_{k_0} \cap T^{-1}I_{k_1} \cap \dots \cap T^{-n}I_{k_n}) > 0$ .

**page 109, line 18 :** Replace " Absolutely continuous " by " nonsingular ".

**page 110, line 2 : Mark ? :** We should put a reference for the ergodicity of the geodesic flow....

**page 111, line 19 from below :** Replace "sublemmas 10.5.1 and 10.5.2" by "sublemmas 10.2.1 and 10.2.2."

## CHAPTER 11

**page 115 : Lemma 11.4 :** " a bounded sequence of real numbers  $\{a_n\}$  " should read " a bounded sequence of positive numbers  $\{a_n\}$ . "

**page 115, line 3 from below :**  $N_k, N_{k+1}$  should be  $n_k, n_{k+1}$  respectively.

**page 117, line 6 from below :** " (11-1), (11.2) and (11.3) " should be " (11-2), (11.3) and (11.4) "

**page 118, line 3 from below :** It should read  $n \in \mathbb{Z}$ .

**page 119, line 1 :** It should read  $n \in \mathbb{Z}$ .

**page 120** All  $C^\perp$  should be  $\mathbb{C}^\perp$ .

**120, Proof of Proposition 11.8 :**  $\bar{\mu}$  denotes the spectral measure.

**page 121,(2)  $\implies$  (3):** All  $n_i$  should be  $n_k$ .

**page 121, line 8 from below :**  $\mu(T^{-n_i}B_i \cap D_j)$  should be removed.

**page 123, line 9 :** It should read " (by writing  $P$  in terms of Jordan forms) "

**page 123, line 7 from below :** It should read  $\mu(A \cap T^{-(n+l)}B) \rightarrow \mu(A)\mu(B)$  as  $n \rightarrow +\infty$ .

## CHAPTER 12

**page 125, line 8 from below :** It should read "  $\{k_i \circ T^n\}_{i=0}^{j_n}$  for  $L^2(T^{-n}\mathcal{B}) \ominus L^2(T^{-(n+1)}\mathcal{B})$ . It follows that  $\{k_i \circ T^n\}_{i=0}^{j_n} \stackrel{\infty}{n=0}$ . "

**page 126, line 4 :** Example 1 should be one-sided aperiodic Markov shifts(aperiodicity is missing).

**page 127, line 4 :**  $\int_A f(x)dx$  should be  $\int_{T^{-1}A} f(x)dx$ .

**page 128, line 6 :**  $K \exp\left(|x-y| \frac{D'}{1-\frac{1}{\beta}}\right)$  should be  $K\left(|x-y| \frac{D'}{1-\frac{1}{\beta}}\right)$

**page 129, line 15 :** It should read " we can choose a finite disjoint set of cylinders  $\{I_{\underline{j}} : \underline{j} = (j_1, \dots, j_l)\}$ , with  $\mu\left(\left(\cup_{\underline{j}} I_{\underline{j}}\right) \Delta A\right) < \epsilon$ ."

**page 129 line 2 from below :** It should read

$$\sup_{x,y \in T^n I_{\underline{i}}} \frac{|\psi'_{\underline{i}}(x)|}{|\psi'_{\underline{i}}(y)|} \leq C. \quad (12.2)$$

**page 130, Sublemma 12.5.1.** The statement should read "There exist  $S > 0$  and a subset  $I'$  of  $T^l(I_{\underline{j}})$  which is a finite disjoint union of elements of  $\bigvee_{i=0}^{S-1} T^{-i}\{I_1 \dots I_k\}$ "



and satisfies  $T^S(I') = I$ . " In the proof,  $T^{s_i}U_i \supset T^{s_i}I_{m_1, \dots, m_{s_j}}^{(i,j)}$  should be  $T^{s_j}U_i \supset T^{s_j}I_{m_1, \dots, m_{s_j}}^{(i,j)}$ .

**page 131, Proof of Proposition 12.6 :** All  $I_{tm_1 \dots m_{s_i}}^{(i)}$  should be  $I_{m_1 \dots m_{s_i}}^{(i)}$ .  
Relace  $\psi_{tm_1 \dots m_{s_i} h_1 \dots h_t}(x)$  by  $\psi_{m_1 \dots m_{s_i}}(x)$ . "Here we take  $t = S - s_i$ " on the line 17 should be put before " Let  $l = l(\epsilon)$ " on the line 15.

**page 132, lines 5-6 :** It should read "  $\mathcal{L}^*(\mu) = \mu$ , i.e. the dual operator  $\hat{\mathcal{L}}^*$  acting on measures (defined by  $(\mathcal{L}^*\mu)(A) = \int \mathcal{L}\chi_A d\mu$ ) fixes  $\mu$ . "

**page 124, line 7 :** it should read " Channon-McMillan- Breiman theorem ".

**page 134, line 14 :** It should read  $\alpha_n = \bigvee_{i=0}^{n-1} T^{-i}\alpha$ .

**page 135 line 1 from below :**  $I(\alpha | \bigvee_{j=1}^{n-i} T^{-j}\alpha)T^i$  should be replacded by  $I(\alpha | \bigvee_{j=1}^{n-1-i} T^{-j}\alpha)T^i$ .

**page 135-136 : Mark ? How to correct.....**

Define

$$F_m(x) =: |I(\alpha | \bigvee_{j=1}^m T^{-j}\alpha)(x) - I(\alpha | \bigvee_{j=1}^{\infty} T^{-j}\alpha)(x)|.$$

Then we see that

$$\frac{1}{n} \left| I(\bigvee_{i=0}^{n-1} T^{-i}\alpha) - \sum_{i=0}^{n-1} fT^i \right| \leq \frac{1}{n} \sum_{j=0}^{n-1} F_{n-j}T^j(x).$$

It follows from Corollary 1.2 in page 96 in [Mane's book] that if  $F_n$  is a sequence that converges to 0 almost everywhere and in  $L^1$  then  $\frac{1}{n} \sum_{j=0}^{n-1} F_{n-j}T^j(x)$  converges to 0 almost everywhere and in  $L^1$ . It is easy (??) to see pointwise convergence and  $L^1$ -convergence of  $F_n$  to 0 (!)

## CHAPTER 14

**page 147, line 12 :** It should read " a compact metric space. "

**page 147, line 12 from below :** It should read " a finite Borel measurable partition "

**page 148, line 7 :** Sub-lemma 14.1 should be Sub-lemma 14.7.

**page 148, line 2 from below :** It should read

$$" h_{\mu}(T^k, \bigvee_{i=0}^{k-1} T^{-i}\alpha) = \lim_{n \rightarrow +\infty} \frac{1}{n} H_{\mu}(\bigvee_{i=0}^{n-1} T^{-ik}(\bigvee_{j=0}^{k-1} T^{-j}\alpha)) "$$

**page 149, line 8 :** It should read

$$" h_{\mu}(T^k) \geq h_{\mu}(T^k, \bigvee_{i=0}^{k-1} T^{-i}\alpha) \geq kh_{\mu}(T, \alpha) \geq kh_{\mu}(T) - k\epsilon. "$$

**page 151, line 2 :** It should read "  $C \in T^{-l}\alpha^{(N)}$  with  $(T^l)^*\nu_{n_i}(D) := \nu_{n_i}(T^{-l}D) = \nu_{n_i}(C)$  and  $C = T^{-l}D$ . "

**page 151, line 5 :** It should read :

$$" N \log(s(n_i, \epsilon)) \leq \sum_{l=0}^{n_i-1} \left( - \sum_{D \in \alpha^{(N)}} (T^l)^*\nu_{n_i}(D) \log((T^l)^*\nu_{n_i}(D)) \right) + 2N^2 \log k "$$

**page 151, line 9 from below :** It should read

$$" \leq -\frac{1}{N} \sum_{C \in \alpha^{(N)}} \mu_{n_i}(C) \log \mu_{n_i}(C) + \frac{2N \log k}{n_i},$$

**page 151, line 5 to page 152, line 2 :** All  $H_\nu$  should be  $H_\mu$ .

## CHAPTER 15

**page 154, (3) :** It should read

$$" \int f(x+a) d\mu = \lim_{n \rightarrow +\infty} \int E(f(x+a) | S^{-n} \mathcal{B}) d\mu = \lim_{n \rightarrow +\infty} \int E(f(x) | S^{-n} \mathcal{B}) d\mu = \int f(x) d\mu$$

and thus we know that  $\mu$  is the Haar-Lebesgue measure. "

**page 154, (4) :** It should read

$$" T'(Sx) \cdot S'(x) = (TS)'(x) = (ST)'(x) = S'(Tx) \cdot T'(x). "$$

**page 155, line 6 :** it should read " $E(\cdot | T^{-1} \mathcal{B})(x) : L^2(X, \mathcal{B}, \mu) \rightarrow L^2(X, \mathcal{B}, \mu)$  is an orthogonal projection which is a contraction "

**page 155, line 7 from below :** It should read " $S^{n'}$  and  $S'$  are  $\mathcal{A}_n$ -measurable "

**Proof 157 :** It should read : " Thus by subadditivity the limit  $h(T|\mathcal{A}) := \lim_{n \rightarrow +\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i} \gamma | \mathcal{A})$  exists.

By the basic equalities for entropy we see that

$$\begin{aligned} h(T|\mathcal{A}) &= \lim_{n \rightarrow +\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i} \gamma | \mathcal{A}) \\ &= \lim_{n \rightarrow +\infty} \frac{1}{n} \left( H(\bigvee_{i=0}^{n-2} T^{-i} \gamma | \mathcal{A}) + H(T^{-(n-1)} \gamma | \mathcal{A} \vee (\bigvee_{i=0}^{n-2} T^{-i} \gamma)) \right) \\ &= h(T) \end{aligned}$$

since

$$h(T) = \lim_{n \rightarrow +\infty} \frac{1}{n} \left( H(\bigvee_{i=0}^{n-2} T^{-i} \gamma) \right)$$

and

$$H\left(T^{-(n-1)} \gamma | \mathcal{A} \vee (\bigvee_{i=0}^{n-2} T^{-i} \gamma)\right) \leq H\left(T^{-(n-1)} \gamma | \mathcal{A}\right) = H(\gamma | \mathcal{A}) < +\infty.$$

It follows from Sub-lemma 15.1.3 that  $h(T|\mathcal{A}) = \frac{\log 3}{\log 2} h(S|\mathcal{A})$  and so we have that the following limit exists

$$h(S|\mathcal{A}) := \lim_{n \rightarrow +\infty} \frac{1}{n} H(\bigvee_{i=0}^{s_n-1} T^{-i} \gamma | \mathcal{A}).$$

Observe that if we replace  $\mathcal{A}$  by the trivial sigma-algebra then the same argument gives that  $h(T) = \frac{\log 3}{\log 2} h(S)$ . Comparing these identities we see that  $h(S) = h(S|\mathcal{A})$ . ■

"

**page 158 :** In Sub-lemma 15.1.5 and in the proof, all  $\mathcal{B}, \beta$  should be  $\gamma$ .

## MARKOV EXTENSIONS FOR INTERVAL MAPS

In section 4.3 we introduced piecewise expanding Markov interval maps and showed they can be effectively modelled by a (one sided) subshift of finite type. This was particularly useful in estimating numbers of periodic points. In this chapter we shall describe a more general construction which works for non-Markov piecewise expanding maps.

## 13.1 BASIC CONSTRUCTIONS

Assume that  $T : I \rightarrow I$  is a interval map which is expanding on each of the intervals  $I = I_1 \cup \dots \cup I_k$ . We shall denote their endpoints by  $\{x_0, x_1, \dots, x_k\}$  (i.e.  $I_1 = [x_0, x_1], \dots, I_k = [x_{k-1}, x_k]$ ).

The following simple lemma will be used several times.

**Lemma 13.1.** *Let  $J$  be a sub interval of one of the intervals  $I_1, \dots, I_k$ . The image  $T(J)$  will be a union either of the form:*

- (a)  $T(J) = D_J \subset I_i$ ;
- (b)  $T(J) = \cup_{j=s}^t I_j$ ;
- (c)  $T(J) = D_J^- \cup (\cup_{j=s}^t I_j)$  or  $T(J) = (\cup_{j=s}^t I_j) \cup D_J^+$
- (d)  $T(J) = D_J^- \cup (\cup_{j=s}^t I_j) \cup D_J^+$

where  $D_J, D_J^- \subset I_{s-1}$  and  $D_J^+ \subset I_{t+1}$  are each closed non-degenerate (strict) sub-intervals.

fifth.eps

FIGURE 13.1. The image  $T(J)$  of  $J \in \mathcal{E}$  gives rise to two more intervals  $D_J^-$  and  $D_J^+$  in  $\mathcal{E}$

*Proof.* Since  $T(J)$  is a closed interval in  $I$  it is clear that this exhausts all possibilities.

*Definition.* We call  $\mathcal{E}_0 = \{I_1, \dots, I_k\}$  the *zeroth generation of intervals*

We can apply Lemma 13.1 to each  $J = I_i$  ( $i = 1, \dots, n$ ) to get new sub-intervals. For example,  $D_{I_i}$  (when  $T(I_i)$  is as in case (a) or (c)) and  $D_{I_i}^-, D_{I_i}^+$  (when  $T(I_i)$  is as in case (d)). If  $T(I_i)$  corresponds to case (b) then no new intervals are created.

*Definition.* We call  $\mathcal{E}_1 = \{D_{I_i}^-, D_{I_i}^+ \text{ or } D_{I_i} : i = 1, \dots, k\}$  the *first generation of intervals*.

We define collections of intervals  $\mathcal{E}_n$ ,  $n \geq 1$ , (each a closed sub-interval of one of the original intervals  $I_1, \dots, I_k$ ) inductively. Given a collection of intervals  $\mathcal{E}_{n-1}$  (each a closed sub-interval of one of the original intervals  $I_1, \dots, I_k$ ) then we can

apply Lemma 13.1 to each interval  $J \in \mathcal{E}_{n-1}$ . We define the  $n$ th generation of intervals to be the collection of closed intervals

$$\mathcal{E}_n = \{D_J^-, D_J^+ \text{ or } D_J : J \in \mathcal{E}_{n-1}\}$$

We write  $\mathcal{E} = \cup_{n=0}^{\infty} \mathcal{E}_n$ .

If  $J \in \mathcal{E}_n$ , say, then each of the sets  $J'$  in the appropriate union for  $T(J)$  (in Lemma 13.1) is called a *successor* to  $J$ . We write  $J \rightarrow J'$ .

The essence of this construction is to compensate for the absence of the Markov property by introducing more and more intervals so that the image of any  $J \in \mathcal{E}$  is again the union of other intervals from  $\mathcal{E}$ . The complications over and above those in the Markov case are that the larger family  $\mathcal{E}$  will not in general be finite, and intervals in  $\mathcal{E}$  will not in general have disjoint interiors.

The following simple example illustrates this construction.

*Example ( $\beta$ -transformation).* Let  $I = [0, 1]$  and choose  $2 > \beta > 1$ . We can define an interval map  $T : I \rightarrow I$  by

$$Tx = \begin{cases} \beta x & \text{if } 0 \leq x \leq \frac{1}{\beta} \\ \beta x - 1 & \text{if } \frac{1}{\beta} < x \leq 1 \end{cases}$$

In this case, we have that  $\mathcal{E}_0 = \{[0, \frac{1}{\beta}], [\frac{1}{\beta}, 1]\}$ . Since  $\beta < 2$  we have that  $T[0, \frac{1}{\beta}] = [0, \frac{1}{\beta}] \cup [\frac{1}{\beta}, 1]$  and  $T[\frac{1}{\beta}, 1] = [0, \frac{1}{\beta}] \cup [\frac{1}{\beta}, \beta - 1]$ . Thus  $\mathcal{E}_1 = \{[\frac{1}{\beta}, 1 - \beta]\}$ . If  $\beta < \frac{5^{1/2}-1}{2}$ , say, then  $T[\frac{1}{\beta}, \beta - 1] = [0, \beta^2 - \beta - 1] \subset [0, \frac{1}{\beta}]$  and so  $\mathcal{E}_2 = \{[0, \beta^2 - \beta - 1]\}$ .

For the purposes of exposition we want to make the following simplifying assumption (which, for example, holds for the  $\beta$ -transformation when  $\beta > 1$  is not an algebraic number.)

*Simplifying Assumption.* There are no choices  $1 \leq i \leq k$  and  $n, m \geq 1$  such that  $T^n(T^m x_i) = T^m x_i$  (i.e. none of the endpoints is pre-periodic).

The next lemma shows that there are some restrictions on the possible successors to a given interval in  $\mathcal{E}_n$ .

**Lemma 13.2 (Restrictions on successors).**

(1) If  $J \in \mathcal{E}_n$  ( $n \geq 1$ ) then there are two possibilities. Either  $J' = T(J)$  or  $T(J) = J'' \cup (\cup_{i=r}^{s-1} I_i) \cup J'$  where

- (i)  $J'' \in \mathcal{E}_j$ , for some  $0 \leq j \leq n$  (perhaps with  $J'' = \emptyset$ ),
- (ii)  $I_r, \dots, I_{s-1}$  are original intervals (perhaps with  $\cup_{i=r}^{s-1} I_i = \emptyset$ ), and
- (ii)  $J' \in \mathcal{E}_{n+1}$ ,

(2) Each element  $J \in \mathcal{E}_n$  must be of one of the two types:

- (i)  $J = T^n(I_i)$ , for some  $1 \leq i \leq k$ ,
- (ii)  $J$  has endpoints  $T^m x_i, T^n x_j$  with  $n - 1 \geq m \geq 0$  (for some  $1 \leq i, j \leq k$ ) (i.e.  $J = [T^m x_i, T^n x_j]$  or  $J = [T^m x_i, T^n x_j]$ )

sixth.eps

FIGURE 13.2

*Proof.* We prove the lemma by induction on  $n$ .

For  $n = 1$  both parts of the lemma are immediate from Lemma 4.11 and Assumption (2). Assume that both part (1) and part (2) of the lemma are known for  $J \in \mathcal{E}_n$ .

If we first assume that  $J = T^n I_r = [T^n x_s, T^n x_{s+1}]$ , say, then either

- (a)  $T(J) = T^{n+1} I_r$ ; or
- (b)  $x_{i-1} \leq T^{n+1} x_s < x_i < \dots < x_j < T^{n+1} x_{s+1} \leq x_{j+1}$ , say (perhaps with the order reversed)

In case (a) we immediately have  $T(J) \in \mathcal{E}_{n+1}$  (i.e. the “either” case in part (1) of the statement and case (i) in part (2)). In case (b) we have by Lemma 13.1 that the successors to  $J$  are  $[T^{n+1} x_s, x_i], I_{i+1}, \dots, I_j, [x_j, T^{n+1} x_{s+1}]$ , (i.e. the “or” case in part (1) of the statement). Moreover, we see that  $[T^{n+1} x_s, x_i] \in \mathcal{E}_{n+1}$  is in the form of case (ii) of part (2).

Alternatively, we can assume  $J = [a, b] = [T^m(x_j), T^n(x_k)] \in \mathcal{E}_n$ , say, with  $0 \leq m \leq n - 1$  (the case with the endpoints reversed being similar).

There are now two possibilities.

- (c)  $x_{r-1} < T(a) < x_r < \dots < x_s < T(b) < x_{s+1}$ , say.
- (d)  $x_r < T(a) < T(b) < x_{r+1}$

(where in each case  $T(a)$  and  $T(b)$  might be in the reverse order).

In case (c) we can take  $J' = [T(a), x_r] \in \mathcal{E}_{n+1}$  and  $J'' = [x_s, T(x_j)] \in \mathcal{E}_{m+1}$ . In case (d) we take  $J'' = [T(a), T(b)] \in \mathcal{E}_{n+1}$ . (The case  $x_r < T(b) < T(a) < x_{r+1}$  being similar). These both correspond to the “Or” case of part (1) in the statement. Moreover, in both cases (c) and (d)  $J'' \in \mathcal{E}_{n+1}$

This completes the proof of the Lemma

*Remarks.*

(a) If we don't assume the simplifying Assumption then there are yet more possibilities with  $J' \in D_i$ ,  $1 \leq i \leq n$ .

(b) Since the map  $T : I \rightarrow I$  is clearly piecewise expanding, there are at most finitely many successors of the form  $T^n(I_i)$ .

(c) Assume that  $x \in J_0 = [a, b] \in \mathcal{E}_n$ . It is useful to enumerate the various possibilities  $J_0 \rightarrow J_1$  with  $T(x) \in J_1$  (cf. Figure 13.3):

- (1) If  $Ta < x_i < Tx < Tb < x_{i+1}$  then  $J_1 = [x_i, Tb]$ ;
- (2) If  $Ta < x_i < Tx < x_{i+1} < Tb$  then  $J_1 = [x_i, x_{i+1}]$ ;
- (3) If  $x_i < Ta < Tx < x_{i+1} < Tb$  then  $J_1 = [Ta, x_{i+1}]$ ;
- (4) If  $x_i < Ta < Tx < Tb < x_{i+1}$  then  $J_1 = [Ta, Tb]$

and four more cases if  $T|J$  reverses orientation.

What characterises these cases is:

- (i) In case (2) *none* of the endpoints of  $J_1$  are images of  $J_0$ ;
- (ii) in cases (1) and (3) one of the end points of  $J_1$  is the image of an endpoint of  $J_0$ ; and,
- (iii) in case (4) *both* of the endpoints of  $J_1$  are images of endpoints of  $J_0$ .

seventh.eps

FIGURE 13.3

Let  $(J_n)_{n \in \mathbb{Z}^+}$  be a sequence of successor intervals (i.e.  $J_0 \rightarrow J_1 \rightarrow J_2 \rightarrow \dots$ ) then we can associate a unique element  $\pi((J_n)_{n \in \mathbb{Z}^+}) \in I$  by

$$\pi((J_n)_{n \in \mathbb{Z}^+}) = \bigcap_{n \in \mathbb{Z}^+} T^{-n} J_n$$

(in complete analogy with the Markov case).

Although  $x \in I$  will not in general have a unique sequence  $(J_n)_{n \in \mathbb{Z}^+}$  such that  $\pi((J_n)_{n \in \mathbb{Z}^+}) = x$ , the next lemma shows the remarkable fact that (generically) the “tail” of such a sequence is uniquely determined.

**Lemma 13.3.** *Let  $(J_n)_{n \in \mathbb{Z}^+}$  and  $(J'_n)_{n \in \mathbb{Z}^+}$  be two sequences of successors such that  $\pi((J_n)_{n \in \mathbb{Z}^+}) = \pi((J'_n)_{n \in \mathbb{Z}^+}) = x \in I$ . If  $\forall m \geq 0, T^m(x) \notin \bigcup_{n \geq 0} T^n \{x_1, \dots, x_k\}$  then  $\exists N \geq 0$  with  $J_n = J'_n$  for  $n \geq N$ .*

*Proof.*

Let us denote  $J_0 = [a, b]$  and  $J'_0 = [c, d]$ . The proof is a little involved, depending on analysing a number of similar cases. Here is a typical case (the others being variations on the same theme). The approach is to analyse the endpoints of the intervals  $J_n$  and  $J'_n$  and to show that for sufficiently large  $n$  they must agree.

**Step One (Comparing  $J_0$  and  $J'_0$ ):** Assume (for definiteness) that we begin with  $a < c < x < b < d$ . By monotonicity of  $T$  on  $I_i$  the images of these points must have the same order (or be reversed). Assume that the order of the images is preserved and that  $Ta < x_j < Tc < Tx < Tb < x_{j+1} < Td$ , say. The successors to  $J_0, J'_0$  containing  $x$  are then  $J_1 = [x_j, T(b)]$  and  $J'_1 = [T(c), x_{j+1}]$  (by cases (1) and (3) above, respectively).

**Step Two ( $J_N$  and  $J'_N$  with a common endpoint):** Let  $N$  be the smallest positive integer such that  $T^N(c, b) \cap \{x_0, \dots, x_k\} \neq \emptyset$ . For  $1 \leq n \leq N - 1$  we see (by repeated application of cases (1), (3) and (4) above) that the successors  $J_n$  each have one of their endpoints being equal to  $T^n(b)$ . Similarly,  $J'_n$  will have an endpoint equal to  $T^n(c)$ .

Assume that we arrive at  $T^N(c) < x_i < T^N(x) < T^N(b) < x_{i+1}$ , say, then  $x_i$  now usurps of  $T^N(c)$  as the (left) endpoint of  $J'_N$ , by case (1) above. Furthermore,  $x_i$  is also the (left) endpoint of  $J_N$ .

eighth.eps

FIGURE 4.8

**Step Three ( $J_n$  and  $J'_n$  ( $n \geq N$ ) with a common endpoint):** We need the following easy sublemma.

**Sublemma 13.3.1.** *If  $J_N \subset J'_N$  with one (or both) of their endpoints the same, then the same is true for  $J_n, J'_n, n \geq N$ .*

*Proof of sublemma 13.3.1.* This is something which one sees from the four cases (1) to (4) above.

**Four ( $J_M$  and  $J'_M$  ( $M \geq N$ ) with two common endpoints):** Assume that  $M > N$  is the least integer for which  $T^M(x, b) \cap \{x_0, \dots, x_k\} \neq \emptyset$ . We see from

sublemma 4.13.1 that for  $N \leq n < M$  the point  $T^{n-N}x_i$  is an endpoint for  $J_n, J'_n$ . Assume that  $T^M(x) < x_j < T^M(b)$ , say, then  $x_j$  usurps  $T^M(b)$  as an endpoint for  $J_M$  (by cases (1) or (2) above). Moreover,  $x_j$  is also an end point for  $J'_M$ . We conclude that  $J_M = J'_M$  and so  $J_n = J'_n$  for  $n \geq M$  (by sublemma 4.13.1 now applied to both endpoints).

The other cases are similar. This completes the proof ♣

If  $T : I \rightarrow I$  is *not* Markov then  $\mathcal{E} = \cup_{n=0}^{\infty} \mathcal{E}_n$  is infinite (by the Simplifying Assumption). We can define an *infinite* matrix  $A$  with entries 0 or 1, whose rows and columns are indexed by  $\mathcal{E}$ , in the following way

*Definition.* For  $J_1, J_2 \in \mathcal{E}$

$$A(J_1, J_2) = \begin{cases} 1 & \text{if } J_2 \text{ succeeds } J_1 \\ 0 & \text{otherwise} \end{cases}$$

The next Proposition gives us (amongst other things) a nice characterization of periodic orbits for  $T : I \rightarrow I$ .

**Proposition 13.4.**

- (i) *There exists  $C > 0$  such that for all  $n \geq 1$   $\text{Card}\mathcal{E}_n \leq C$*
- (ii) *There is a bijection between periodic points  $T^n x = x$  and strings  $(i_0, i_1, \dots, i_{n-1})$  such that  $A(i_0, i_1) = A(i_1, i_2) = \dots = A(i_{n-1}, i_0) = 1$  (except possibly for endpoints of the original intervals  $I_i$ )* ■
- (iii) *There exists  $N > 0$  such that every  $J \in \mathcal{E}$  has at most  $N$  successors.*

*Proof.*

(i) From Lemma 13.1 we see that each  $I_i$  ( $i = 1, \dots, k$ ) gives rise to either two, one or no elements in  $\mathcal{E}_1$ . Thus  $\text{Card}\mathcal{E}_1 \leq 2k$ . For  $n \geq 1$  we see from Lemma 4.12 (i) that each  $J \in \mathcal{E}_n$  gives rise to at most one element  $J' \in \mathcal{E}_{n+1}$ . Thus  $\text{Card}\mathcal{E}_{n+1} \leq \text{Card}\mathcal{E}_n \leq \dots \leq \text{Card}\mathcal{E}_N$ . Thus part (i) is proved with  $C = 2k$ .

(ii) Let  $(J_n)_{n=0}^{\infty}$  be a sequence of successors which is periodic (i.e.  $\exists n \geq 1$  with  $J_{m+n} = J_m$ , for  $k \geq 0$ ). Then if  $x \in I$  is the corresponding point (i.e.  $\{x\} = \cap_{m \geq 0} T^{-m} J_m$ ) then  $T^n(x)$  is the corresponding point for  $(J_{n+m})_{n=0}^{\infty} = (J_n)_{n=0}^{\infty}$  (i.e.  $\{T^n(x)\} = \cap_{m \geq 0} T^{-m} J_{n+m}$ ). Thus  $T^n(x) = x$

Assume that  $T^n(x) = x$  and choose a corresponding sequence of successors  $(J_n)_{n \in \mathbb{Z}^+}$ . We know by Lemma 13.3 that provided  $x \notin \cup_{n \geq 0} T^{-n} (\cup_{m \geq 0} T^m \{x_1, \dots, x_k\})$  then the two sequences of successors (for  $x = T^n x$ )  $(J_n)_{n \in \mathbb{Z}^+}$  and  $(J_{n+m})_{n \in \mathbb{Z}^+}$  agree for  $n \geq N$ , say. If we choose  $nm \geq N$  then the periodic sequence  $(J'_n)_{n \in \mathbb{Z}^+} = (J_{n+mN})_{n \in \mathbb{Z}^+}$  corresponds to  $x$  since  $\{x = T^{mN} x\} = \cap_{n \in \mathbb{Z}^+} T^{-n} J'_n$ .

Assume that  $T^n(x) = x$  and  $T^m(x) = T^q(x_i)$ , for some  $1 \leq i \leq k$  and  $q, m \geq 0$ . In particular, if we choose  $(r-1)n \leq m < rn$  then we see  $x = T^{rn}(x) = T^{q+(rn-m)}(x_i)$ . (i.e.  $x$  is characterised as the (unique) periodic orbit  $\{x, Tx, \dots, T^{n-1}x\}$  into which the orbit of  $x_i$  eventually falls). Since there are only  $k$  boundary points  $x_1, \dots, x_k$  this means the bijection fails on at most  $k$  periodic orbits.

- (iii) This is an easy consequence of the definitions.

**Corollary 13.4.1.** *The number of periodic points  $T^n(x) = x$  of period  $n$  is given by*

$$\text{CardFix}(T^n) := \sum_{J \in \mathcal{E}} A^n(J, J)$$

### 13.2 THE GROWTH OF PERIODIC POINTS

In this section we want apply the analysis from the previous section to study the number of periodic points  $\text{CardFix}(T^n)$  via the matrix  $\mathcal{A}$ . Rather than try to analyse infinite matrices  $\mathcal{A}$  it is simpler to fix some large  $k$  and consider the truncated (finite) matrix  $\mathcal{A}_k(J, J') = \mathcal{A}(J, J')$  if  $J, J' \in \cup_{n=0}^k \mathcal{E}_n$

Proposition 4.?? we know that every string of successors corresponding to a fixed point for  $T^n : I \rightarrow I$  must be of length  $n$ . Moreover, since we can always choose a string starting in  $\mathcal{E}_0$  then the uniqueness conclusion of Proposition 13.(ii) tell us that the string contains an interval from  $\mathcal{E}_0$ . Finally we see by 13.2 that all the elements of the string must therefore be in  $\cup_{i=1}^n \mathcal{E}_i$ . Thus by Corollary 13.4.1 we see that

$$\text{CardFix}(T^n) = \sum_{J \in \mathcal{E}} A^n(J, J) = \sum_{J \in \mathcal{E}} \mathcal{A}_k^n(J, J) = \text{trace}(\mathcal{A}_k^n)$$

whenever  $k \geq n$ .

Since  $\mathcal{A}_k$  can be thought of as a finite matrix, we have the standard matrix identity

$$\text{Det}(I - z\mathcal{A}_k) = \exp\left(\sum_{m=1}^{\infty} \frac{z^m}{m} \text{trace}(\mathcal{A}_k^m)\right)$$

The right hand-side carries information on the traces  $\text{trace}(\mathcal{A}_k^m)$ . To study the left hand side, it helps to fix  $1 \leq N \leq k$  and write  $\mathcal{A}_k = \begin{pmatrix} A_N & B_N \\ C_N & D_N \end{pmatrix}$  where  $A_N$  is the square submatrix with indices  $\cup_{i=0}^N \mathcal{E}_i$ , and  $D_N$  is the square sub-matrix with indices  $\cup_{i=N+1}^k \mathcal{E}_i$ , etc.

**Lemma 13.5.** *For any  $k \geq N \geq 1$ , the sub-matrix  $D_N$  has all of its eigenvalues with modulus at most  $N^{1/N}$*

We shall postpone the proof until later and first consider the implications of this lemma.

Since  $D_N - zI$  is invertable whenever  $|z| \geq N^{1/N}$  (by Lemma 13.5), we have the matrix identity

$$\begin{pmatrix} A_N - z & B_N \\ C_N & D_N - z \end{pmatrix} = \begin{pmatrix} (A_N - z) - B_N(D_N - z)^{-1}C & B(D_N - z)^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ C & D - z \end{pmatrix} \blacksquare$$

and taking determinants gives that

$$\text{Det}(\mathcal{A}_k - zI) = \text{Det}\left(\left((A_N - z) - B_N(D_N - z)^{-1}CB(D_N - z)^{-1}\right) \text{Det}(D_N - z)\right).$$

#### The argument needs completing

The following theorem shows that the number of periodic points has a very distinctive pattern.



**Theorem 4.16.** For any  $\beta > 1$  there exist constants

- (i)  $\lambda_1, \dots, \lambda_M$ , with  $|\lambda_i| > \beta$  ( $i = 1, \dots, M$ ), and
- (ii)  $C_1, \dots, C_M \in \mathbb{C}$

such that

$$N(n) = \sum_{i=1}^M C_i \lambda_i^n + E(n)$$

where the error term  $E(n)$  satisfies  $|E(n)| \leq \text{Constant} \beta^n$

*Proof.* The above identity shows that provided  $N^{1/n} < \beta$  the eigenvalues of  $\mathcal{A}_k$  (for any  $k \geq N$ ) are determined by the determinant of a  $N \times N$  matrix (by Lemma 4.15).

#### The argument needs completing

Finally, with some additional bounds we can let  $k \rightarrow +\infty$  to see that the power series  $\sum_{m=1}^{\infty} \frac{z^m}{m} \text{trace} \mathcal{A}_k^m$ .

It only remains to supply the proof of Lemma 4.15

*Proof of Lemma 4.15.* We can define a norm on the matrix  $D_k$  (and its powers) by  $\|D_N\| = \sup \left\{ \sum_{D' \in \cup_{i=N+1}^k \mathcal{E}_i} D_N(D, D') : D \in \cup_{i=N+1}^k \mathcal{E}_i \right\}$ . By the spectral radius theorem it suffices to show that  $\limsup_{n \rightarrow +\infty} \|D_N^n\|^{\frac{1}{n}} \leq N^{\frac{1}{N}}$

For  $p \geq 0$  we let  $\mathcal{N}(p)$  denote the number of possible strings of successors

$$J_1 \rightarrow J_2 \rightarrow \dots \rightarrow J_{p-1} \rightarrow J_p$$

with  $J_i \in \cup_{m=k+1}^{\infty} \mathcal{E}_m$  for  $1 \leq i \leq p$ . If we set

$$T_r = \begin{cases} 0 & \text{if } r \leq 0 \\ 2k & \text{if } r = 1 \end{cases}$$

and then define inductively  $T_{r+1} = T_r + T_{r-k}$  ( $r \geq 1$ ) then we claim that  $\mathcal{N}(p) \leq T_p$ .

For  $p = 1$  this is true by definition. We proceed by induction. and assume that  $\mathcal{N}(r) \leq T_r$  for  $r = 1, \dots, p$ . Consider a typical string of successors

$$J_1 \rightarrow J_2 \rightarrow \dots \rightarrow J_p \rightarrow J_{p+1}$$

(of length  $p + 1$ ) with  $J_i \in \cup_{m=k+1}^{\infty} \mathcal{E}_m$  for  $1 \leq i \leq p + 1$ .

By Lemma 4.12 we see that there are three possible types of successor for  $J_1 \in \mathcal{E}_n$  ( $n \geq k$ ). There are three possibilities

- (i) If  $J_1 = [T^n x_i, T^l x_j]$  (with  $0 \leq l \leq n - 1$ ), say, then assume that  $J_2 = [x_k, T^l x_j]$ , say. To satisfy  $J_i \in \cup_{m=k+1}^{\infty} \mathcal{E}_m$  for  $2 \leq i \leq k + 3$  the only possibility is that  $J_3 = [T x_q, T^{l+1} x_j]$ ,  $J_4 = [T^2 x_q, T^{l+2} x_j]$ ,  $\dots$ ,  $J_{i+2} = [T^i x_q, T^{l+i} x_j]$ ,  $\dots$ ,  $J_{k+3} = [T^{k+1} x_q, T^{l+k+1} x_j]$  (or with the endpoints reversed). Thus only the remaining intervals  $J_{k+4}, \dots, J_{p+1}$  are not determined, and by the inductive hypothesis the number of possible choices is bounded by  $\mathcal{N}(p - k) \leq T_{n-k}$ .
- (ii) Assume that  $J_2$  is a unique successor with  $J_2 \in \mathcal{E}_{n+1}$ . Then by the inductive hypothesis there are at most  $\mathcal{N}(n) \leq T_n$  choices for the remaining successors  $J_i$ ,  $2 \leq i \leq n + 1$
- (iii) Finally, the possibility  $J_2 \in \mathcal{E}_0$  is eliminated by our requirement that  $J_2 \in \cup_{m=k+1}^{\infty} \mathcal{E}_m$ .

Thus we see that  $T_{p+1} \leq T_p + T_{p-k}$ , completing the inductive step.

We now see that

$$\begin{aligned} T_p &\leq T_{p-1} + T_{p-k-1} \\ &\leq (T_{p-2} + T_{p-k-2}) + T_{p-k-1} \\ &\leq \dots \leq T_{p-k} + (T_{p-2k} + \dots + T_{p-k-2}) + T_{p-k-1} \\ &\leq kT_{p-k} \end{aligned}$$

and so  $T_{mk} \leq k^m T_0$ , for  $m \geq 1$ . In particular, we see that  $\|D_N^n\| \leq N(n)$  for  $n \geq 1$ .

Since  $\|(\mathcal{A} - \mathcal{A}_k)^n\|$  is sub-additive in  $n$  we see that

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \|(\mathcal{A} - \mathcal{A}_k)^n\|_1^{1/n} &= \limsup_{m \rightarrow +\infty} \|(\mathcal{A} - \mathcal{A}_k)^{mk}\|_1^{\frac{1}{mk}} \\ &\leq \limsup_{m \rightarrow +\infty} N(mk)^{\frac{1}{mk}} \\ &\leq \limsup_{m \rightarrow +\infty} (T_{mk})^{\frac{1}{mk}} \leq k^{\frac{1}{k}} \end{aligned}$$

This completes the proof of the sublemma.

### 13.3 COMMENTS AND REFERENCES

The theory of Markov extensions was originated by Hofbauer [H]. The main application we give to periodic points is based on work of Hofbauer and Keller.

#### REFERENCES

1. F. Hofbauer, *Israel J. Math* **34** ??? (19??), ???-???
2. F. Hofbauer and G. Keller, *????? ??? (??)*, ???-???