A Short Introduction to Ergodic Theory of Numbers

Karma Dajani

June 13, 2013

Contents

1 Motivation and Examples **5**
- 1.1 What is Ergodic Theory? . . . 5
- 1.2 Number Theoretic Examples . . . 6

2 Measure Preserving, Ergodicity and the Ergodic Theorem **15**
- 2.1 Measure Preserving Transformations . . . 15
- 2.2 Ergodicity . . . 17
- 2.3 The Ergodic Theorem . . . 19
- 2.4 Mixing . . . 28

3 Examples Revisited **31**

4 Natural Extensions **41**
- 4.1 Natural Extensions of m-adic and β-expansions . . . 41
- 4.2 Natural Extension of Continued Fractions . . . 44
  - 4.2.1 The Doeblin-Lenstra Conjecture . . . 45
  - 4.2.2 Some Diophantine spinoff . . . 46
  - 4.2.3 A proof of the Lenstra-Doeblin Conjecture . . . 50

5 Random β-expansions **51**
- 5.1 Random β-expansions . . . 51
- 5.2 Basic Properties of Random β-transformations . . . 54
- 5.3 An ergodic measure for random β-expansions . . . 56

Chapter 1

Motivation and Examples

The aim of these short lecture notes is to show how one can use basic ideas in ergodic theory in order to understand the global behaviour of a family of series expansions of numbers in a given interval. This is done by showing that the expansions under study can be generated by iterations of an appropriate map which will be shown to be measure preserving and ergodic.

1.1 What is Ergodic Theory?

It is not easy to give a simple definition of Ergodic Theory because it uses techniques and examples from many fields such as probability theory, statistical mechanics, number theory, vector fields on manifolds, group actions of homogeneous spaces and many more.

The word *ergodic* is a mixture of two Greek words: *ergon* (work) and *odos* (path). The word was introduced by Boltzmann (in statistical mechanics) regarding his hypothesis: *for large systems of interacting particles in equilibrium, the time average along a single trajectory equals the space average*. The hypothesis as it was stated was false, and the investigation for the conditions under which these two quantities are equal lead to the birth of ergodic theory as is known nowadays.

A modern description of what ergodic theory is would be: it is the study of the long term average behavior of systems evolving in time. The collection of all states of the system form a space X, and the evolution is represented by either

– a transformation $T : X \to X$, where Tx is the state of the system at time $t = 1$, when the system (i.e., at time $t = 0$) was initially in state x. (This is analogous to the setup of discrete time stochastic processes).

– if the evolution is continuous or if the configurations have spacial structure, then we describe the evolution by looking at a group of transformations

G (like $\mathbb{Z}^2$, $\mathbb{R}$, $\mathbb{R}^2$) acting on X, i.e., every $g \in G$ is identified with a transformation $T_g : X \to X$, and $T_{gg'} = T_g \circ T_{g'}$.

The space X usually has a special structure, and we want T to preserve the basic structure on X. For example
– if X is a measure space, then T must be measurable.
– if X is a topological space, then T must be continuous.
– if X has a differentiable structure, then T is a diffeomorphism.
In these lectures our space is a probability space $(X, \mathcal{B}, \mu)$, and our time is discrete. So the evolution is described by a measurable map $T : X \to X$, so that $T^{-1}A \in \mathcal{B}$ for all $A \in \mathcal{B}$. For each $x \in X$, the orbit of x is the sequence

$$x, Tx, T^2x, \ldots.$$

If T is invertible, then one speaks of the two sided orbit

$$\ldots, T^{-1}x, x, Tx, \ldots.$$

Before we go any further with ergodic theory, let us see the connection of the above setup with a certain collection of number theoretic expansions of points in the unit interval.

1.2 Number Theoretic Examples

Example 1.2.1 (Binary Expansion) Let $X = [0, 1)$ with the Lebesgue σ-algebra $\mathcal{B}$, and Lebesgue measure λ. Define $T : X \to X$ be given by

$$Tx = 2x \bmod 1 = \begin{cases} 2x & 0 \le x < 1/2 \\ 2x - 1 & 1/2 \le x < 1. \end{cases}$$

Using T one can associate with each point in $[0, 1)$ an infinite sequence of 0's and 1's. To do so, we define the function a_1 by

$$a_1(x) = \begin{cases} 0 & \text{if } 0 \le x < 1/2 \\ 1 & \text{if } 1/2 \le x < 1, \end{cases}$$

then $Tx = 2x - a_1(x)$. Now, for $n \ge 1$ set $a_n(x) = a_1(T^{n-1}x)$. Fix $x \in X$, for simplicity, we write a_n instead of $a_n(x)$, then $Tx = 2x - a_1$. Rewriting we get $x = \frac{a_1}{2} + \frac{Tx}{2}$. Similarly, $Tx = \frac{a_2}{2} + \frac{T^2x}{2}$. Continuing in this manner, we see that for each $n \ge 1$,

$$x = \frac{a_1}{2} + \frac{a_2}{2^2} + \cdots + \frac{a_n}{2^n} + \frac{T^nx}{2^n}.$$

Since $0 < T^nx < 1$, we get

$$x - \sum_{i=1}^{n} \frac{a_i}{2^i} = \frac{T^nx}{2^n} \to 0 \text{ as } n \to \infty.$$

Thus,

$$x = \sum_{i=1}^{\infty} \frac{a_i}{2^i}.$$

We shall later see that the sequence of digits $a_1, a_2, \ldots$ forms an i.i.d. sequence of Bernoulli random variables.

Example 1.2.2 (m-ary Expansion) If we replace in the above example the transformation by $Tx = mx \bmod 1$, and $a_1(x) = k$ if $x \in [\frac{k}{m}, \frac{k+1}{m})$, $k = 0, 1, \ldots, m-1$, then for each $n \geq 1$ one has

$$x = \frac{a_1}{m} + \frac{a_2}{m^2} + \cdots + \frac{a_n}{m^n} + \frac{T^n x}{2^n},$$

and since $0 < T^n x < 1$, taking limits one gets the m-array expansion of $x = \sum_{i=1}^{\infty} \frac{a_i}{m^i}$ with digits $a_i \in \{0, 1, \cdots, m-1\}$

In the above two examples, we looked at series expansions in integer bases. In these cases, the expansion obtained by the above maps is essentially unique. The exceptional set consists of all points of the form $\frac{k}{m^n}$ which have exactly two expansions, one ending in zeros and the other ending in the digit $m-1$. In case the base is not an integer, then the situation is completely reversed, typically there are uncountably many algorithms generating expansions in noninteger base. We mention here the two extreme cases.

Example 1.2.3 (Greedy Expansions) Let $\beta > 1$ be a noninteger, define $T_\beta : [0, \lfloor\beta\rfloor/(\beta-1)) \to: [0, \lfloor\beta\rfloor/(\beta-1))$ by

$$T_\beta(x) \;=\; \begin{cases} \beta x \pmod 1, & 0 \leq x < 1, \\ \\ \beta x - \lfloor\beta\rfloor, & 1 \leq x < \lfloor\beta\rfloor/(\beta-1), \end{cases}$$

see also Figure 1.1. Similar to the above examples, we define the digits of x as

$$a_1(x) = a_1 = \begin{cases} i & \text{if } \frac{i}{\beta} \leq x < \frac{i+1}{\beta},\ i = 0, \ldots, \lfloor\beta\rfloor - 1 \\ \lfloor\beta\rfloor & \text{if } \frac{\lfloor\beta\rfloor}{\beta} \leq x < \frac{\lfloor\beta\rfloor}{\beta-1}, \end{cases}$$

and $a_n(x) = a_n = a_1(T_\beta^{n-1})$, $m \geq 2$. One easily sees that for any $n \geq 1$,

$$x = \frac{a_1}{\beta} + \frac{a_2}{\beta^2} + \cdots + \frac{a_n}{\beta^n} + \frac{T_\beta^n x}{\beta^n}.$$

Taking limits, lead to the greedy expansion of $x = \sum_{n=1}^{\infty} \frac{a_n}{\beta^n}$. If for some n one has $T_\beta^n x = 0$, then x has a finite expansion, and we do not need to take limits. It is not hard to see that for each n, the digits a_n is the largest element in $\{0, 1, \cdots, \rfloor\beta\rfloor\}$ satisfying $\sum_{i=1}^{n} \frac{a_i}{\beta^i} \leq x$.

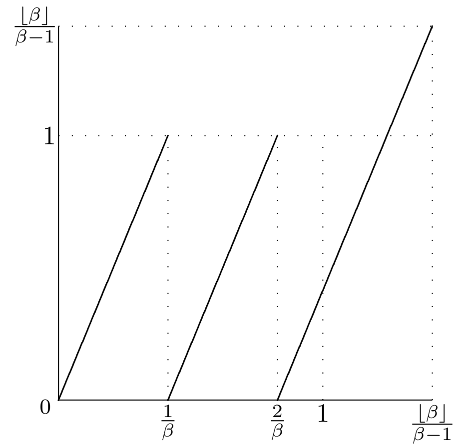

Figure 1.1: The *greedy map* T_β (here $\beta = \sqrt{2}+1$).

Example 1.2.4 (Lazy Expansions) Consider the map $S_\beta : (0, \lfloor\beta\rfloor/(\beta-1)] \to (0, \lfloor\beta\rfloor/(\beta-1)]$ by

$$S_\beta(x) \;=\; \beta x - d, \qquad \text{for } x \in \Delta(d),$$

where

$$\Delta(0) \;=\; \left(0, \frac{\lfloor\beta\rfloor}{\beta(\beta-1)}\right] \tag{1.1}$$

and

$$\Delta(d) \;=\; \left(\frac{\lfloor\beta\rfloor}{\beta(\beta-1)} + \frac{d-1}{\beta}, \frac{\lfloor\beta\rfloor}{\beta(\beta-1)} + \frac{d}{\beta}\right], \quad d \in \{1, 2, \ldots, \lfloor\beta\rfloor\}.$$

Since

$$\frac{\lfloor\beta\rfloor}{\beta(\beta-1)} \;=\; \frac{\lfloor\beta\rfloor}{\beta-1} - \frac{\lfloor\beta\rfloor}{\beta},$$

one has that

$$\Delta(d) \;=\; \left(\frac{\lfloor\beta\rfloor}{\beta-1} - \frac{\lfloor\beta\rfloor - d + 1}{\beta}, \frac{\lfloor\beta\rfloor}{\beta-1} - \frac{\lfloor\beta\rfloor - d}{\beta}\right], \quad d \in \{1, 2, \ldots, \lfloor\beta\rfloor\}. \tag{1.2}$$

Hence, to get the defining partition one starts from $\lfloor\beta\rfloor/(\beta-1)$ by taking $\lfloor\beta\rfloor$ intervals of length $1/\beta$ from right to left. The last interval with end-points 0 and $(\lfloor\beta\rfloor + 1 - \beta)/\beta(\beta-1)$, corresponding to the lazy digit 0, is longer than the rest; see also Figure 1.2. One can easily see that the interval $A_\beta = ((\lfloor\beta\rfloor + 1 - \beta)/(\beta-1), \lfloor\beta\rfloor/(\beta-1)]$ is an attractor, in the sense that for any x there exists $n \geq 0$ such that $S_\beta^m x \in A_\beta$ for all $m \geq n$. Defining now the digits of x by $a_1(x) = a_1 = d$ if $x \in \Delta(d)$, and $a_n(x) = a_n = a_1(S_\beta^{n-1}x)$ for $n \geq 2$. It is easily seen that $x = \sum_{n=1}^{\infty} \frac{a_n}{\beta^n}$, where the summation can be finite.

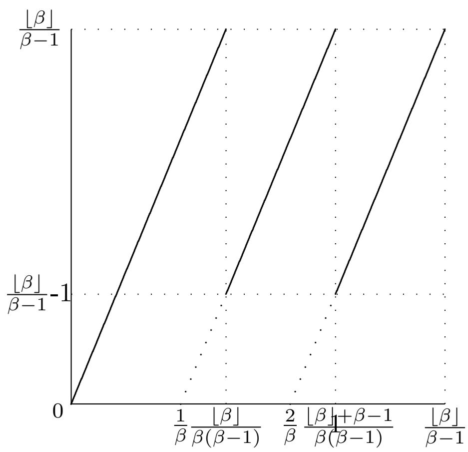

Figure 1.2: The *lazy map* S_β.

Example 1.2.5 (Lüroth Series) Another kind of series expansion, introduced by J. Lüroth [L] in 1883, motivates this approach. Several authors have studied the dynamics of such systems. Take as partition of $[0,1)$ the intervals $[\frac{1}{n+1},\frac{1}{n})$ where $n \in \mathbb{N}$. Every number $x \in [0,1)$ can be written as a finite or infinite series, the so-called Lüroth (series) expansion

$$\begin{aligned} x &= \frac{1}{a_1(x)} + \frac{1}{a_1(x)(a_1(x)-1)a_2(x)} + \cdots \\ &\quad + \frac{1}{a_1(x)(a_1(x)-1)\cdots a_{n-1}(x)(a_{n-1}(x)-1)a_n(x)} + \cdots ; \end{aligned}$$

here $a_k(x) \geq 2$ for each $k \geq 1$. How is such a series generated?

Let $T : [0,1) \to [0,1)$ be defined by

$$Tx = \begin{cases} n(n+1)x - n, & x \in [\frac{1}{n+1},\frac{1}{n}), \\ 0, & x = 0. \end{cases} \tag{1.3}$$

Let $x \neq 0$, for $k \geq 1$ and $T^{k-1}x \neq 0$ we define the digits $a_n = a_n(x)$ by

$$a_k(x) = a_1(T^{k-1}x),$$

where $a_1(x) = n$ if $x \in [\frac{1}{n},\frac{1}{n-1})$, $n \geq 2$. Now (1.3) can be written as

$$Tx = \begin{cases} a_1(x)(a_1(x)-1)x - (a_1(x)-1), & x \neq 0, \\ 0, & x = 0. \end{cases}$$

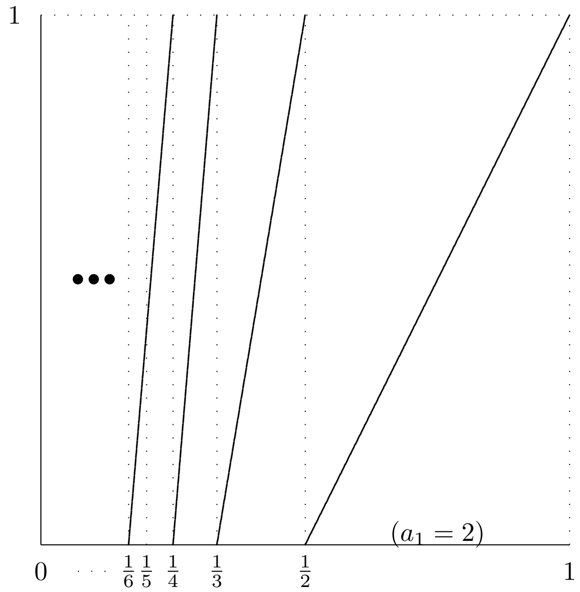

Figure 1.3: The Lüroth Series map T.

Thus[1], for any $x \in (0,1)$ such that $T^{k-1}x \neq 0$, we have

$$\begin{aligned} x = \frac{1}{a_1} + \frac{Tx}{a_1(a_1-1)} &= \frac{1}{a_1} + \frac{1}{a_1(a_1-1)}\left(\frac{1}{a_2} + \frac{T^2x}{a_2(a_2-1)}\right) \\ &= \frac{1}{a_1} + \frac{1}{a_1(a_1-1)a_2} + \frac{T^2x}{a_1(a_1-1)a_2(a_2-1)} \\ &\vdots \\ &= \frac{1}{a_1} + \cdots + \frac{1}{a_1(a_1-1)\cdots a_{k-1}(a_{k-1}-1)a_k} \\ &\quad + \frac{T^kx}{a_1(a_1-1)\cdots a_k(a_k-1)}. \end{aligned}$$

Notice that, if $T^{k-1}x = 0$ for some $k \geq 1$, and if we assume that k is the smallest positive integer with this property, then

$$x = \frac{1}{a_1} + \cdots + \frac{1}{a_1(a_1-1)\cdots a_{k-1}(a_{k-1}-1)a_k}.$$

In case $T^{k-1}x \neq 0$ for all $k \geq 1$, one gets

$$x = \frac{1}{a_1} + \frac{1}{a_1(a_1-1)a_2} + \cdots + \frac{1}{a_1(a_1-1)\cdots a_{k-1}(a_{k-1}-1)a_k} + \cdots ,$$

[1]For ease of notation we drop the argument x from the functions $a_k(x)$.

where $a_k \geq 2$ for each $k \geq 1$. Let us convince ourselves that this last infinite series indeed converges to x. Let $S_k = S_k(x)$ be the sum of the first k terms of the sum. Then

$$|x - S_k| = \left| \frac{T^k x}{a_1(a_1 - 1) \cdots a_k(a_k - 1)} \right| ;$$

since $T^k x \in [0, 1)$ and $a_k \geq 2$ for all x and all $k \geq 1$, we find

$$|x - S_k| \leq \frac{1}{2^k} \to 0 \text{ as } k \to \infty .$$

From the above we also see that if x and y have the same Lüroth expansion, then, for each $k \geq 1$,

$$|x - y| \leq \frac{1}{2^{k-1}}$$

and it follows that x equals y.

Example 1.2.6 (Generalized Lüroth Series) Consider any partition $\mathcal{I} = \{[\ell_n, r_n) : n \in \mathcal{D}\}$ of $[0, 1)$ where $\mathcal{D} \subset \mathbb{Z}^+$ is finite or countable and $\sum_{n \in \mathcal{D}} (r_n - \ell_n) = 1$. We write $L_n = r_n - \ell_n$ and $I_n = [\ell_n, r_n)$ for $n \in \mathcal{D}$. Moreover, we assume that $i, j \in \mathcal{D}$ with $i > j$ satisfy $0 < L_i \leq L_j < 1$. $\mathcal{D}$ is called the digit set; see also Figure 1.4.

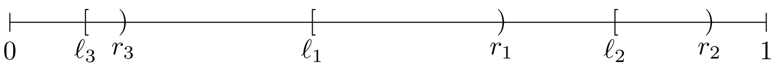

Figure 1.4: The partition $\mathcal{I}$.

We will consider the following transformation T on $[0, 1)$:

$$Tx = \begin{cases} \dfrac{1}{r_n - \ell_n} x - \dfrac{\ell_n}{r_n - \ell_n} , & x \in I_n, \ n \in \mathcal{D} , \\ 0, & x \in I_\infty = [0, 1) \setminus \bigcup_{n \in \mathcal{D}} I_n ; \end{cases} \tag{1.4}$$

see also Figure 1.5.

We want to iterate T in order to generate a series expansion of points x in $[0, 1)$, in fact of points x whose T-orbit never hits I_∞. We will show that the set of such points has measure 1.

We first need some notation. For $x \in [\ell_n, r_n)$, $n \in \mathcal{D}$, we write

$$s(x) = \frac{1}{r_n - \ell_n} \quad \text{and} \quad h(x) = \frac{\ell_n}{r_n - \ell_n} ,$$

so that $Tx = xs(x) - h(x)$. Now let

$$s_k(x) = \begin{cases} s(T^{k-1}x), & \text{if } T^{k-1}x \in \bigcup_{n \in \mathcal{D}} I_n , \\ \infty, & \text{otherwise,} \end{cases}$$

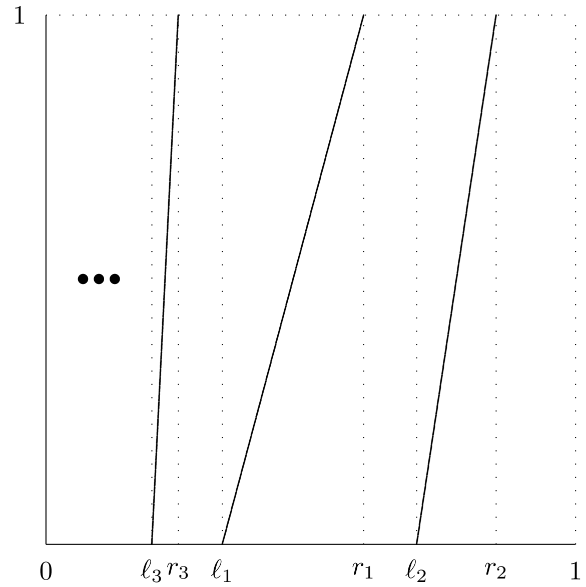

Figure 1.5: The GLS-map T.

$$h_k(x) = \begin{cases} h(T^{k-1}x), & \text{if } T^{k-1}x \in \bigcup_{n\in\mathcal{D}} I_n\,, \\ 1, & \text{otherwise} \end{cases}$$

(thus $s(x) = s_1(x)$, $h(x) = h_1(x)$). From these definitions we see that for $x \in \bigcup_{n\in\mathcal{D}} I_n \cap (0,1)$ such that $T^k x \in \bigcup_{n\in\mathcal{D}} I_n \cap (0,1)$ for all $k \geq 1$, one has

$$\begin{aligned} x &= \frac{h_1(x)}{s_1(x)} + \frac{Tx}{s_1(x)} = \frac{h_1}{s_1} + \frac{Tx}{s_1} \\ &= \frac{h_1}{s_1} + \frac{1}{s_1}\left(\frac{h_2}{s_2} + \frac{T^2x}{s_2}\right) = \frac{h_1}{s_1} + \frac{h_2}{s_1s_2} + \frac{T^2x}{s_1s_2} \\ &= \frac{h_1}{s_1} + \frac{h_2}{s_1s_2} + \cdots + \frac{h_k}{s_1s_2\cdots s_k} + \frac{T^kx}{s_1s_2\cdots s_k} \\ &= \frac{h_1}{s_1} + \frac{h_2}{s_1s_2} + \cdots + \frac{h_k}{s_1s_2\cdots s_k} + \cdots\,. \end{aligned}$$

We refer to the above expansion as the GLS($\mathcal{I}$) expansion of x with a specified digit set $\mathcal{D}$. Such an expansion converges to x. Moreover, it is unique.

To prove the first statement we define the nth GLS-convergent P_k/Q_k of x by

$$\frac{P_k}{Q_k} = \frac{h_1}{s_1} + \frac{h_2}{s_1s_2} + \cdots + \frac{h_k}{s_1s_2\cdots s_k}\,;$$

then

$$\left|x - \frac{P_k}{Q_k}\right| = x - \frac{P_k}{Q_k} = \frac{T^kx}{s_1s_2\cdots s_k}\,. \tag{1.5}$$

Notice that

$$\frac{1}{s_k} = \text{ length of the interval that } T^{k-1}x \text{ belongs to.}$$

Let $L := \max_{n\in\mathcal{D}} L_n$. Then,

$$\left| x - \frac{P_k}{Q_k} \right| \leq L^k \to 0 \quad \text{as} \quad k \to \infty \,.$$

For the proof of the second statement, use (1.5) and the triangle inequality.

Example 1.2.7 (Continued Fraction) Define a transformation $T : [0,1) \to [0,1)$ by $T0 = 0$ and for $x \neq 0$

$$Tx = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor ;$$

see Figure 1.6.

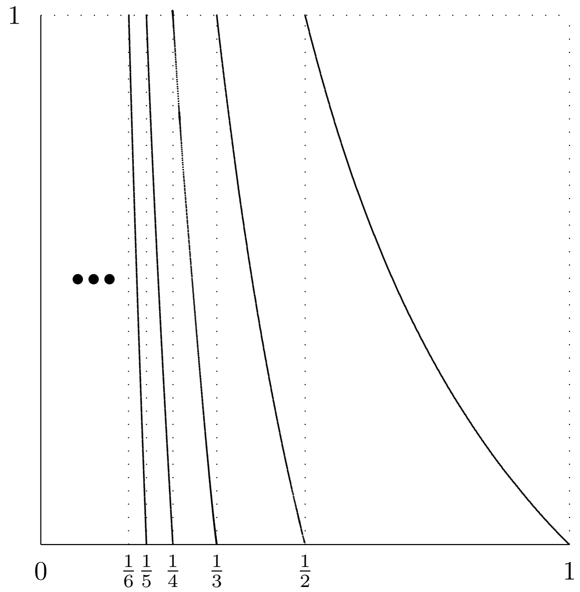

Figure 1.6: The continued fraction map T.

An interesting feature of this map is that its iterations generate the continued fraction expansion for points in $(0,1)$. For if we define

$$a_1 = a_1(x) = \begin{cases} 1 & \text{if } x \in (\frac{1}{2}, 1) \\ n & \text{if } x \in (\frac{1}{n+1}, \frac{1}{n}], n \geq 2, \end{cases}$$

then, $Tx = \frac{1}{x} - a_1$ and hence $x = \dfrac{1}{a_1 + Tx}$. For $n \geq 1$, let $a_n = a_n(x) = a_1(T^{n-1}x)$. Then, after n iterations we see that

$$x = \frac{1}{a_1 + Tx} = \cdots = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_n + T^n x}}}.$$

In fact, if $p_n, q_n \in \mathbb{Z}$, with $\gcd(p_n, q_n) = 1$, and $q_n > 0$ are such, that

$$\frac{p_n}{q_n} = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_n}}},$$

then we will show (in Section 4.2.2) that the q_n are monotonically increasing, and

$$\left| x - \frac{p_n}{q_n} \right| < \frac{1}{q_n^2} \to 0 \text{ as } n \to \infty. \tag{1.6}$$

The last statement implies that

$$x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\ddots}}}}.$$

In view of (1.6) we define for every real number x and every $n \geq 0$ the so-called *approximation coefficients* $\Theta_n(x)$ by

$$\Theta_n(x) = q_n^2 \left| x - \frac{p_n}{q_n} \right|. \tag{1.7}$$

It immediately follows from (1.6) that $\Theta_n(x) < 1$ for all irrational x and all $n \geq 0$. We will return to these approximation coefficients in Chapter 4.

Chapter 2

Measure Preserving, Ergodicity and the Ergodic Theorem

The basic setup of all the examples in the previous chapter consists of a probability space $(X, \mathcal{B}, \mu)$, where X is a set consisting of all possible outcomes, $\mathcal{B}$ is a σ-algebra, and μ is a probability measure on $\mathcal{B}$. The evolution is given by a transformation $T : X \to X$ which is measurable, i.e. $T^{-1}A = \{x \in X : Tx \in A\} \in \mathcal{B}$ for any $A \in \mathcal{B}$. We want also that the evolution is in steady state i.e. stationary. In the language of ergodic theory, we want T to be *measure preserving*.

2.1 Measure Preserving Transformations

Definition 2.1.1 *Let $(X, \mathcal{B}, \mu)$ be a probability space, and $T : X \to X$ measurable. The map T is said to be measure preserving with respect to μ if $\mu(T^{-1}A) = \mu(A)$ for all $A \in \mathcal{B}$.*

In case T is invertible, then T is measure preserving if and only if $\mu(TA) = \mu(A)$ for all $A \in \mathcal{B}$. We can generalize the definition of measure preserving to the following case. Let $T : (X_1, \mathcal{B}_1, \mu_1) \to (X_2, \mathcal{B}_2, \mu_2)$ be measurable, then T is measure preserving if $\mu_1(T^{-1}A) = \mu_2(A)$ for all $A \in \mathcal{B}_2$.

Recall that a collection $\mathcal{S}$ of subsets of X is said to be a *semi-algebra* if (i) $\emptyset \in \mathcal{S}$, (ii) $A \cap B \in \mathcal{S}$ whenever $A, B \in \mathcal{S}$, and (iii) if $A \in \mathcal{S}$, then $X \setminus A = \cup_{i=1}^{n} E_i$ is a disjoint union of elements of $\mathcal{S}$. For example if $X = [0, 1)$, and $\mathcal{S}$ is the collection of all subintervals, then $\mathcal{S}$ is a semi-algebra. Or if $X = \{0, 1\}^{\mathbb{Z}}$, then the collection of all cylinder sets $\{x : x_i = a_i, \ldots, x_j = a_j\}$ is a semi-algebra.
An *algebra* $\mathcal{A}$ is a collection of subsets of X satisfying:(i) $\emptyset \in \mathcal{A}$, (ii) if $A, B \in \mathcal{A}$, then $A \cap B \in \mathcal{A}$, and finally (iii) if $A \in \mathcal{A}$, then $X \setminus A \in \mathcal{A}$. Clearly an algebra is a semi-algebra. Furthermore, given a semi-algebra $\mathcal{S}$ one can form an algebra by taking all finite disjoint unions of elements of $\mathcal{S}$. We denote this algebra by

$\mathcal{A}(\mathcal{S})$, and we call it the *algebra generated* by $\mathcal{S}$. It is in fact the smallest algebra containing $\mathcal{S}$. Likewise, given a semi-algebra $\mathcal{S}$ (or an algebra $\mathcal{A}$), the σ-algebra generated by $\mathcal{S}$ $(\mathcal{A})$ is denoted by $\sigma(S)$ $(\sigma(A))$, and is the smallest σ-algebra containing $\mathcal{S}$ (or $\mathcal{A}$).
A *monotone class* $\mathcal{C}$ is a collection of subsets of X with the following two properties

- if $E_1 \subseteq E_2 \subseteq \ldots$ are elements of $\mathcal{C}$, then $\cup_{i=1}^{\infty} E_i \in \mathcal{C}$,
- if $F_1 \supseteq F_2 \supseteq \ldots$ are elements of $\mathcal{C}$, then $\cap_{i=1}^{\infty} F_i \in \mathcal{C}$.

The *monotone class generated* by a collection $\mathcal{S}$ of subsets of X is the smallest monotone class containing $\mathcal{S}$; for a proof see [H].

Theorem 2.1.1 *Let $\mathcal{A}$ be an algebra of X, then the σ-algebra $\sigma(A)$ generated by $\mathcal{A}$ equals the monotone class generated by $\mathcal{A}$.*

Using the above theorem, one can get an easier criterion for checking that a transformation is measure preserving.

Theorem 2.1.2 *Let $(X_i, \mathcal{B}_i, \mu_i)$ be probability spaces, $i = 1, 2$, and $T : X_1 \to X_2$ a transformation. Suppose $\mathcal{S}_2$ is a generating semi-algebra of $\mathcal{B}_2$. Then, T is measurable and measure preserving if and only if for each $A \in \mathcal{S}_2$, we have $T^{-1}A \in \mathcal{B}_1$ and $\mu_1(T^{-1}A) = \mu_2(A)$.*

Proof. Let

$$\mathcal{C} = \{B \in \mathcal{B}_2 : T^{-1}B \in \mathcal{B}_1, \text{ and } \mu_1(T^{-1}B) = \mu_2(B)\},$$

then $\mathcal{S}_2 \subseteq \mathcal{C} \subseteq \mathcal{B}_2$, and hence $\mathcal{A}(\mathcal{S}_2) \subset \mathcal{C}$. We show that $\mathcal{C}$ is a monotone class. Let $E_1 \subseteq E_2 \subseteq \ldots$ be elements of $\mathcal{C}$, and let $E = \cup_{i=1}^{\infty} E_i$. Then, $T^{-1}E = \cup_{i=1}^{\infty} T^{-1}E_i \in \mathcal{B}_1$, and

$$\begin{aligned}
\mu_1(T^{-1}E) &= \mu_1(\cup_{n=1}^{\infty} T^{-1}E_n) \\
&= \lim_{n\to\infty} \mu_1(T^{-1}E_n) \\
&= \lim_{n\to\infty} \mu_2(E_n) \\
&= \mu_2(\cup_{n=1}^{\infty} E_n) \\
&= \mu_2(E).
\end{aligned}$$

Thus, $E \in \mathcal{C}$. A similar proof shows that if $F_1 \supseteq F_2 \supseteq \ldots$ are elements of $\mathcal{C}$, then $\cap_{i=1}^{\infty} F_i \in \mathcal{C}$. Hence, $\mathcal{C}$ is a monotone class containing the algebra $\mathcal{A}(\mathcal{S}_2)$. By the monotone class theorem, $\mathcal{B}_2$ is the smallest monotone class containing $\mathcal{A}(\mathcal{S}_2)$, hence $\mathcal{B}_2 \subseteq \mathcal{C}$. This shows that $\mathcal{B}_2 = \mathcal{C}$, therefore T is measurable and measure preserving. □

For example if

– $X = [0,1)$ with the Borel σ-algebra $\mathcal{B}$, and μ a probability measure on $\mathcal{B}$. Then a transformation $T : X \to X$ is measurable and measure preserving if and only if $T^{-1}[a,b) \in \mathcal{B}$ and $\mu\left(T^{-1}[a,b)\right) = \mu([a,b))$ for any interval $[a,b)$.
– $X = \{0,1\}^{\mathbb{N}}$ with product σ-algebra and product measure μ. A transformation $T : X \to X$ is measurable and measure preserving if and only if

$$T^{-1}(\{x : x_0 = a_0, \dots, x_n = a_n\}) \in \mathcal{B},$$

and

$$\mu\left(T^{-1}\{x : x_0 = a_0, \dots, x_n = a_n\}\right) = \mu(\{x : x_0 = a_0, \dots, x_n = a_n\})$$

for any cylinder set.

Another useful lemma is the following.

Lemma 2.1.1 *Let $(X, \mathcal{B}, \mu)$ be a probability space, and $\mathcal{A}$ an algebra generating $\mathcal{B}$. Then, for any $A \in \mathcal{B}$ and any $\epsilon > 0$, there exists $C \in \mathcal{A}$ such that $\mu(A \Delta C) < \epsilon$.*

Proof. Let

$$\mathcal{D} = \{A \in \mathcal{B} : \text{ for any } \epsilon > 0, \text{ there exists } C \in \mathcal{A} \text{ such that } \mu(A\Delta C) < \epsilon\}.$$

Clearly, $\mathcal{A} \subseteq \mathcal{D} \subseteq \mathcal{B}$. By the Monotone Class Theorem (Theorem 2.1.1), we need to show that $\mathcal{D}$ is a monotone class. To this end, let $A_1 \subseteq A_2 \subseteq \cdots$ be a sequence in $\mathcal{D}$, and let $A = \bigcup_{n=1}^{\infty} A_n$, notice that $\mu(A) = \lim_{n\to\infty} \mu(A_n)$. Let $\epsilon > 0$, there exists an N such that $\mu(A\Delta A_N) = |\mu(A) - \mu(A_N)| < \epsilon/2$. Since $A_N \in \mathcal{D}$, then there exists $C \in \mathcal{A}$ such that $\mu(A_N \Delta C) < \epsilon/2$. Then,

$$\mu(A\Delta C) \leq \mu(A\Delta A_N) + \mu(A_N \Delta C) < \epsilon.$$

Hence, $A \in \mathcal{D}$. Similarly, one can show that $\mathcal{D}$ is closed under decreasing intersections so that $\mathcal{D}$ is a monotone class containg $\mathcal{A}$, hence by the Monotone Class Theorem $\mathcal{B} \subseteq \mathcal{D}$. Therefore, $\mathcal{B} = \mathcal{D}$, and the theorem is proved. □

2.2 Ergodicity

Definition 2.2.1 *Let T be a measure preserving transformation on a probability space $(X, \mathcal{F}, \mu)$. The map T is said to be ergodic if for every measurable set A satisfying $T^{-1}A = A$, we have $\mu(A) = 0$ or 1.*

Theorem 2.2.1 *Let $(X, \mathcal{F}, \mu)$ be a probability space and $T : X \to X$ measure preserving. The following are equivalent:*

(i) *T is ergodic.*

(ii) *If $B \in \mathcal{F}$ with $\mu(T^{-1}B\Delta B) = 0$, then $\mu(B) = 0$ or 1.*

(iii) *If $A \in \mathcal{F}$ with $\mu(A) > 0$, then $\mu\left(\cup_{n=1}^{\infty} T^{-n}A\right) = 1$.*

(iv) *If $A, B \in \mathcal{F}$ with $\mu(A) > 0$ and $\mu(B) > 0$, then there exists $n > 0$ such that $\mu(T^{-n}A \cap B) > 0$.*

Remark 2.2.1

1. In case T is invertible, then in the above characterization one can replace T^{-n} by T^n.

2. Note that if $\mu(B \triangle T^{-1}B) = 0$, then $\mu(B \setminus T^{-1}B) = \mu(T^{-1}B \setminus B) = 0$. Since

$$B = \left(B \setminus T^{-1}B\right) \cup \left(B \cap T^{-1}B\right),$$

and

$$T^{-1}B = \left(T^{-1}B \setminus B\right) \cup \left(B \cap T^{-1}B\right),$$

we see that after removing a set of measure 0 from B and a set of measure 0 from $T^{-1}B$, the remaining parts are equal. In this case we say that B equals $T^{-1}B$ modulo sets of measure 0.

3. In words, (iii) says that if A is a set of positive measure, almost every $x \in X$ eventually (in fact infinitely often) will visit A.

4. (iv) says that elements of B will eventually enter A.

Proof of Theorem 2.2.1.

(i) $\Rightarrow$ (ii) Let $B \in \mathcal{F}$ be such that $\mu(B \Delta T^{-1}B) = 0$. We shall define a measurable set C with $C = T^{-1}C$ and $\mu(C \Delta B) = 0$. Let

$$C = \{x \in X : T^n x \in B \text{ i.o. }\} = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} T^{-k}B.$$

Then, $T^{-1}C = C$, hence by (i) $\mu(C) = 0$ or 1. Furthermore,

$$\begin{aligned}
\mu(C \Delta B) &= \mu\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} T^{-k}B \cap B^c\right) + \mu\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} T^{-k}B^c \cap B\right) \\
&\leq \mu\left(\bigcup_{k=1}^{\infty} T^{-k}B \cap B^c\right) + \mu\left(\bigcup_{k=1}^{\infty} T^{-k}B^c \cap B\right) \\
&\leq \sum_{k=1}^{\infty} \mu\left(T^{-k}B \Delta B\right).
\end{aligned}$$

Using induction (and the fact that $\mu(E \Delta F) \leq \mu(E \Delta G) + \mu(G \Delta F)$), one can show that for each $k \geq 1$ one has $\mu\left(T^{-k}B \Delta B\right) = 0$. Hence, $\mu(C \Delta B) = 0$ which implies that $\mu(C) = \mu(B)$. Therefore, $\mu(B) = 0$ or 1.

(ii) $\Rightarrow$ (iii) Let $\mu(A) > 0$ and let $B = \bigcup_{n=1}^{\infty} T^{-n}A$. Then $T^{-1}B \subset B$. Since T is measure preserving, then $\mu(B) > 0$ and

$$\mu(T^{-1}B \Delta B) = \mu(B \setminus T^{-1}B) = \mu(B) - \mu(T^{-1}B) = 0.$$

Thus, by (ii) $\mu(B)=1$.

(iii) $\Rightarrow$ (iv) Suppose $\mu(A)\mu(B)>0$. By (iii)

$$\mu(B)=\mu\left(B\cap\bigcup_{n=1}^{\infty}T^{-n}A\right)=\mu\left(\bigcup_{n=1}^{\infty}(B\cap T^{-n}A)\right)>0.$$

Hence, there exists $k\geq 1$ such that $\mu(B\cap T^{-k}A)>0$.

(iv) $\Rightarrow$ (i) Suppose $T^{-1}A=A$ with $\mu(A)>0$. If $\mu(A^c)>0$, then by (iv) there exists $k\geq 1$ such that $\mu(A^c\cap T^{-k}A)>0$. Since $T^{-k}A=A$, it follows that $\mu(A^c\cap A)>0$, a contradiction. Hence, $\mu(A)=1$ and T is ergodic. □

The following lemma provides, in some cases, a useful tool to verify that a measure preserving transformation defined on $([0,1),\mathcal{B},\mu)$ is ergodic, where $\mathcal{B}$ is the Lebesgue σ-algebra, and μ is a probability measure equivalent to Lebesgue measure λ (i.e., $\mu(A)=0$ if and only if $\lambda(A)=0$).

Lemma 2.2.1 (Knopp's Lemma) *If B is a Lebesgue set and $\mathcal{C}$ is a class of subintervals of $[0,1)$, satisfying*

(a) *every open subinterval of $[0,1)$ is at most a countable union of disjoint elements from $\mathcal{C}$,*

(b) $\forall A\in\mathcal{C}$, $\lambda(A\cap B)\geq\gamma\lambda(A)$, *where $\gamma>0$ is independent of A,*

then $\lambda(B)=1$.

Proof. The proof is done by contradiction. Suppose $\lambda(B^c)>0$. Given $\varepsilon>0$ there exists by Lemma 2.1.1 a set E_ε that is a finite disjoint union of open intervals such that $\lambda(B^c\triangle E_\varepsilon)<\varepsilon$. Now by conditions (a) and (b) (that is, writing E_ε as a countable union of disjoint elements of $\mathcal{C}$) one gets that $\lambda(B\cap E_\varepsilon)\geq\gamma\lambda(E_\varepsilon)$.

Also from our choice of E_ε and the fact that

$$\lambda(B^c\triangle E_\varepsilon)\geq\lambda(B\cap E_\varepsilon)\geq\gamma\lambda(E_\varepsilon)\geq\gamma\lambda(B^c\cap E_\varepsilon)>\gamma(\lambda(B^c)-\varepsilon),$$

we have that

$$\gamma(\lambda(B^c)-\varepsilon)\;<\;\lambda(B^c\triangle E_\varepsilon)\;<\;\varepsilon,$$

implying that $\gamma\lambda(B^c)<\varepsilon+\gamma\varepsilon$. Since $\varepsilon>0$ is arbitrary, we get a contradiction. □

2.3 The Ergodic Theorem

The Ergodic Theorem is also known as Birkhoff's Ergodic Theorem or the Individual Ergodic Theorem (1931). This theorem is in fact a generalization of the Strong Law of Large Numbers (SLLN) which states that for a sequence

$Y_1, Y_2, \ldots$ of i.i.d. random variables on a probability space $(X, \mathcal{F}, \mu)$, with $\mathrm{E}|Y_i| < \infty$; one has

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Y_i \;=\; \mathrm{E}Y_1 \text{ (a.e.)}.$$

For example consider $X = \{0,1\}^{\mathbb{N}}$, $\mathcal{F}$ the σ-algebra generated by the cylinder sets, and μ the uniform product measure, i.e.,

$$\mu\left(\{x : x_1 = a_1, x_2 = a_2, \ldots, x_n = a_n\}\right) = 1/2^n.$$

Suppose one is interested in finding the frequency of the digit 1. More precisely, for a.e. x we would like to find

$$\lim_{n\to\infty} \frac{1}{n}\#\{1 \le i \le n : x_i = 1\}.$$

Using the Strong Law of Large Numbers one can answer this question easily. Define

$$Y_i(x) \;:=\; \begin{cases} 1, & \text{if } x_i = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Since μ is product measure, it is easy to see that $Y_1, Y_2, \ldots$ form an i.i.d. Bernoulli process, and $EY_i = E|Y_i| = 1/2$. Further, $\#\{1 \le i \le n : x_i = 1\} = \sum_{i=1}^{n} Y_i(x)$. Hence, by SLLN one has

$$\lim_{n\to\infty} \frac{1}{n}\#\{1 \le i \le n : x_i = 1\} = \frac{1}{2}.$$

Suppose now we are interested in the frequency of the block 011, i.e., we would like to find

$$\lim_{n\to\infty} \frac{1}{n}\#\{1 \le i \le n : x_i = 0, x_{i+1} = 1, x_{i+2} = 1\}.$$

We can start as above by defining random variables

$$Z_i(x) \;:=\; \begin{cases} 1, & \text{if } x_i = 0, x_{i+1} = 1, x_{i+2} = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Then,

$$\frac{1}{n}\#\{1 \le i \le n : x_i = 0, x_{i+1} = 1, x_{i+2} = 1\} = \frac{1}{n}\sum_{i=1}^{n} Z_i(x).$$

It is not hard to see that this sequence is stationary but not independent. So one cannot directly apply the strong law of large numbers. Notice that if T is the left shift on X, then $Y_n = Y_1 \circ T^{n-1}$ and $Z_n = Z_1 \circ T^{n-1}$.

In general, suppose $(X, \mathcal{F}, \mu)$ is a probability space and $T : X \to X$ a measure preserving transformation. For $f \in L^1(X, \mathcal{F}, \mu)$, we would like to know under what conditions does the limit $\lim_{n\to\infty} \dfrac{1}{n}\displaystyle\sum_{i=0}^{n-1} f(T^i x)$ exist a.e. If it does exist what is its value? This is answered by the Ergodic Theorem which was originally proved by G.D. Birkhoff in 1931. Since then, several proofs of this important theorem have been obtained; here we present a recent proof given by T. Kamae and M.S. Keane in [KK].

Theorem 2.3.1 (The Ergodic Theorem) *Let $(X, \mathcal{F}, \mu)$ be a probability space and $T : X \to X$ a measure preserving transformation. Then, for any f in $L^1(\mu)$,*

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

exists a.e., is T-invariant and $\int_X f \, \mathrm{d}\mu = \int_X f^* \, \mathrm{d}\mu$. If moreover T is ergodic, then f^* is a constant a.e. and $f^* = \int_X f \, \mathrm{d}\mu$.

For the proof of the above theorem, we need the following simple lemma.

Lemma 2.3.1 *Let $M > 0$ be an integer, and suppose $\{a_n\}_{n\geq 0}$, $\{b_n\}_{n\geq 0}$ are sequences of non-negative real numbers such that for each $n = 0, 1, 2, \ldots$ there exists an integer $1 \leq m \leq M$ with*

$$a_n + \cdots + a_{n+m-1} \geq b_n + \cdots + b_{n+m-1}.$$

Then, for each positive integer $N > M$, one has

$$a_0 + \cdots + a_{N-1} \geq b_0 + \cdots + b_{N-M-1}.$$

Proof of Lemma 2.3.1 Using the hypothesis we recursively find integers $m_0 < m_1 < \cdots < m_k < N$ with the following properties

$$m_0 \leq M,\ m_{i+1} - m_i \leq M \text{ for } i = 0, \ldots, k-1, \text{ and } N - m_k < M,$$

$$a_0 + \cdots + a_{m_0 - 1} \geq b_0 + \cdots + b_{m_0 - 1},$$
$$a_{m_0} + \cdots + a_{m_1 - 1} \geq b_{m_0} + \cdots + b_{m_1 - 1},$$
$$\vdots$$
$$a_{m_{k-1}} + \cdots + a_{m_k - 1} \geq b_{m_{k-1}} + cldots + b_{m_k - 1}.$$

Then,

$$\begin{aligned} a_0 + \cdots + a_{N-1} &\geq a_0 + \cdots + a_{m_k - 1} \\ &\geq b_0 + \cdots + b_{m_k - 1} \geq b_0 + \cdots + b_{N-M-1}. \end{aligned}$$

$\square$

Proof of Theorem 2.3.1 Assume with no loss of generality that $f \geq 0$ (otherwise we write $f = f^+ - f^-$, and we consider each part separately). Let $f_n(x) =$

$f(x)+\ldots+f(T^{n-1}x)$, $\overline{f}(x) = \limsup_{n\to\infty} \dfrac{f_n(x)}{n}$, and $\underline{f}(x) = \liminf_{n\to\infty} \dfrac{f_n(x)}{n}$. Then $\overline{f}$ and $\underline{f}$ are T-invariant, since

$$\begin{aligned} \overline{f}(Tx) &= \limsup_{n\to\infty} \frac{f_n(Tx)}{n} \\ &= \limsup_{n\to\infty} \left[\frac{f_{n+1}(x)}{n+1} \cdot \frac{n+1}{n} - \frac{f(x)}{n} \right] \\ &= \limsup_{n\to\infty} \frac{f_{n+1}(x)}{n+1} = \overline{f}(x). \end{aligned}$$

Similarly $\underline{f}$ is T-invariant. Now, to prove that f^* exists, is integrable and T-invariant, it is enough to show that

$$\int_X \underline{f}\, \mathrm{d}\mu \geq \int_X f\, \mathrm{d}\mu \geq \int_X \overline{f}\, \mathrm{d}\mu.$$

For since $\overline{f} - \underline{f} \geq 0$, this would imply that $\overline{f} = \underline{f} = f^*$. a.e.

We first prove that $\int_X \overline{f} d\mu \leq \int_X f\, \mathrm{d}\mu$. Fix any $0 < \epsilon < 1$, and let $L > 0$ be any real number. By definition of $\overline{f}$, for any $x \in X$, there exists an integer $m > 0$ such that

$$\frac{f_m(x)}{m} \geq \min(\overline{f}(x), L)(1-\epsilon).$$

Now, for any $\delta > 0$ there exists an integer $M > 0$ such that the set

$$X_0 = \{x \in X : \exists\ 1 \leq m \leq M \text{ with } f_m(x) \geq m\ \min(\overline{f}(x), L)(1-\epsilon)\}$$

has measure at least $1-\delta$. Define F on X by

$$F(x) = \begin{cases} f(x) & x \in X_0 \\ L & x \notin X_0. \end{cases}$$

Notice that $f \leq F$ (why?). For any $x \in X$, let $a_n = a_n(x) = F(T^n x)$, and $b_n = b_n(x) = \min(\overline{f}(x), L)(1-\epsilon)$ (so b_n is independent of n).We now show that $\{a_n\}$ and $\{b_n\}$ satisfy the hypothesis of Lemma 2.3.1 with $M > 0$ as above. For any $n = 0, 1, 2, \ldots$

-— if $T^n x \in X_0$, then there exists $1 \leq m \leq M$ such that

$$\begin{aligned} f_m(T^n x) &\geq m\ \min(\overline{f}(T^n x), L)(1-\epsilon) \\ &= m\ \min(\overline{f}(x), L)(1-\epsilon) \\ &= b_n + \ldots + b_{n+m-1}. \end{aligned}$$

Hence,

$$\begin{aligned} a_n + \ldots + a_{n+m-1} &= F(T^n x) + \ldots + F(T^{n+m-1}x) \\ &\geq f(T^n x) + \ldots + f(T^{n+m-1}x) = f_m(T^n x) \\ &\geq b_n + \ldots + b_{n+m-1}. \end{aligned}$$

— If $T^n x \notin X_0$, then take $m = 1$ since

$$a_n = F(T^n x) = L \geq \min(\overline{f}(x), L)(1-\epsilon) = b_n.$$

Hence by Lemma 2.3.1 for all integers $N > M$ one has

$$F(x) + \ldots + F(T^{N-1}x) \geq (N-M)\min(\overline{f}(x), L)(1-\epsilon).$$

Integrating both sides, and using the fact that T is measure preserving, one gets

$$N\int_X F(x)\,\mathrm{d}\mu(x) \geq (N-M)\int_X \min(\overline{f}(x), L)(1-\epsilon)\,\mathrm{d}\mu(x).$$

Since

$$\int_X F(x)\,\mathrm{d}\mu(x) = \int_{X_0} f(x)\,\mathrm{d}\mu(x) + L\mu(X \setminus X_0),$$

one has

$$\begin{aligned}
\int_X f(x)\,\mathrm{d}\mu(x) &\geq \int_{X_0} f(x)\,\mathrm{d}\mu(x) \\
&= \int_X F(x)\,\mathrm{d}\mu(x) - L\mu(X \setminus X_0) \\
&\geq \frac{(N-M)}{N}\int_X \min(\overline{f}(x), L)(1-\epsilon)\,\mathrm{d}\mu(x) - L\delta.
\end{aligned}$$

Now letting first $N \to \infty$, then $\delta \to 0$, then $\epsilon \to 0$, and lastly $L \to \infty$ one gets together with the monotone convergence theorem that $\overline{f}$ is integrable, and

$$\int_X f(x)\,\mathrm{d}\mu(x) \geq \int_X \overline{f}(x)\,\mathrm{d}\mu(x).$$

We now prove that

$$\int_X f(x)\,\mathrm{d}\mu(x) \leq \int_X \underline{f}(x)\,\mathrm{d}\mu(x).$$

Fix $\epsilon > 0$, and $\delta_0 > 0$. Since $f \geq 0$, there exists $\delta > 0$ such that whenever $A \in \mathcal{F}$ with $\mu(A) < \delta$, then $\int_A f d\mu < \delta_0$. Note that for any $x \in X$ there exists an integer m such that

$$\frac{f_m(x)}{m} \leq (\underline{f}(x) + \epsilon).$$

Now choose $M > 0$ such that the set

$$Y_0 = \{x \in X : \exists\, 1 \leq m \leq M \text{ with } f_m(x) \leq m\,(\underline{f}(x) + \epsilon)\}$$

has measure at least $1 - \delta$. Define G on X by

$$G(x) = \begin{cases} f(x) & x \in Y_0 \\ 0 & x \notin Y_0. \end{cases}$$

Notice that $G \le f$. Let $b_n = G(T^n x)$, and $a_n = \underline{f}(x)+\epsilon$ (so a_n is independent of n). One can easily check that the sequences $\{a_n\}$ and $\{b_n\}$ satisfy the hypothesis of Lemma 2.3.1 with $M > 0$ as above. Hence for any $M > N$, one has

$$G(x) + \cdots + G(T^{N-M-1}x) \le N(\underline{f}(x) + \epsilon).$$

Integrating both sides yields

$$(N-M)\int_X G(x)d\mu(x) \le N(\int_X \underline{f}(x)d\mu(x) + \epsilon).$$

Since $\mu(X \setminus Y_0) < \delta$, then $\nu(X \setminus Y_0) = \int_{X\setminus Y_0} f(x)d\mu(x) < \delta_0$. Hence,

$$\begin{aligned}\int_X f(x)\,\mathrm{d}\mu(x) &= \int_X G(x)\,\mathrm{d}\mu(x) + \int_{X\setminus Y_0} f(x)\,\mathrm{d}\mu(x)\\ &\le \frac{N}{N-M}\int_X (\underline{f}(x)+\epsilon)\,\mathrm{d}\mu(x) + \delta_0.\end{aligned}$$

Now, let first $N \to \infty$, then $\delta \to 0$ (and hence $\delta_0 \to 0$), and finally $\epsilon \to 0$, one gets

$$\int_X f(x)\,\mathrm{d}\mu(x) \le \int_X \underline{f}(x)\,\mathrm{d}\mu(x).$$

This shows that

$$\int_X \underline{f}\,\mathrm{d}\mu \ \ge \int_X f\,\mathrm{d}\mu \ \ge \int_X \overline{f}\,\mathrm{d}\mu,$$

hence, $\overline{f} = \underline{f} = f^*$ a.e., and f^* is T-invariant. In case T is ergodic, then the T-invariance of f^* implies that f^* is a constant a.e. Therefore,

$$f^*(x) = \int_X f^*(y)d\mu(y) = \int_X f(y)\,\mathrm{d}\mu(y).$$

□

Remark 2.3.1 (1) Let us study further the limit f^* in the case that T is not ergodic. Let $\mathcal{I}$ be the sub-σ-algebra of $\mathcal{F}$ consisting of all T-invariant subsets $A \in \mathcal{F}$. Notice that if $f \in L^1(\mu)$, then the *conditional expectation* of f given $\mathcal{I}$ (denoted by $E_\mu(f|\mathcal{I})$), is the unique a.e. $\mathcal{I}$-measurable $L^1(\mu)$ function with the property that

$$\int_A f(x)\,\mathrm{d}\mu(x) = \int_A E_\mu(f|\mathcal{I})(x)\,\mathrm{d}\mu(x)$$

for all $A \in \mathcal{I}$ i.e., $T^{-1}A = A$. We claim that $f^* = E_\mu(f|\mathcal{I})$. Since the limit function f^* is T-invariant, it follows that f^* is $\mathcal{I}$-measurable. Furthermore, for any $A \in \mathcal{I}$, by the ergodic theorem and the T-invariance of 1_A,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}(f1_A)(T^i x) = 1_A(x)\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} f(T^i x) = 1_A(x)f^*(x)\ a.e.$$

and

$$\int_X f 1_A(x)\, \mathrm{d}\mu(x) = \int_X f^* 1_A(x)\, \mathrm{d}\mu(x).$$

This shows that $f^* = E_\mu(f|\mathcal{I})$.

(2) Suppose that T is ergodic and measure preserving with respect to μ, and let ν be a probability measure which is equivalent to μ (i.e. μ and ν have the same sets of measure zero so $\mu(A) = 0$ if and only if $\nu(A) = 0$), then for every $f \in L^1(\mu)$ one has ν a.e.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) \; = \; \int_X f\, \mathrm{d}\mu$$

Corollary 2.3.1 *Let $(X, \mathcal{F}, \mu)$ be a probability space, and $T : X \to X$ a measure preserving transformation. Then, T is ergodic if and only if for all $A, B \in \mathcal{F}$, one has*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) = \mu(A)\mu(B). \tag{2.1}$$

Proof. Suppose T is ergodic, and let $A, B \in \mathcal{F}$. Since the indicator function $1_A \in L^1(X, \mathcal{F}, \mu)$, by the ergodic theorem one has

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_A(T^i x) = \int_X 1_A(x)\, \mathrm{d}\mu(x) = \mu(A) \text{ a.e.}$$

Then,

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_{T^{-i}A\cap B}(x) &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_{T^{-i}A}(x) 1_B(x) \\
&= 1_B(x) \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_A(T^i x) \\
&= 1_B(x)\mu(A) \qquad \text{a.e.}
\end{aligned}$$

Since for each n, the function $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_{T^{-i}A\cap B}$ is dominated by the constant function 1, it follows by the dominated convergence theorem that

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n} \mu(T^{-i}A \cap B) &= \int_X \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_{T^{-i}A\cap B}(x)\, \mathrm{d}\mu(x) \\
&= \int_X 1_B \mu(A)\, \mathrm{d}\mu(x) \\
&= \mu(A)\mu(B).
\end{aligned}$$

Conversely, suppose (2.1) holds for every $A, B \in \mathcal{F}$. Let $E \in \mathcal{F}$ be such that $T^{-1}E = E$ and $\mu(E) > 0$. By invariance of E, we have $\mu(T^{-i}E \cap E) = \mu(E)$, hence

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}E \cap E) = \mu(E).$$

On the other hand, by (2.1)

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}E \cap E) = \mu(E)^2.$$

Hence, $\mu(E) = \mu(E)^2$. Since $\mu(E) > 0$, this implies $\mu(E) = 1$. Therefore, T is ergodic. □

To show ergodicity one needs to verify equation (2.1) for sets A and B belonging to a generating semi-algebra only as the next proposition shows.

Proposition 2.3.1 *Let $(X, \mathcal{F}, \mu)$ be a probability space, and $\mathcal{S}$ a generating semi-algebra of $\mathcal{F}$. Let $T : X \to X$ be a measure preserving transformation. Then, T is ergodic if and only if for all $A, B \in \mathcal{S}$, one has*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) = \mu(A)\mu(B). \tag{2.2}$$

Proof. We only need to show that if (2.2) holds for all $A, B \in \mathcal{S}$, then it holds for all $A, B \in \mathcal{F}$. Let $\epsilon > 0$, and $A, B \in \mathcal{F}$. Then, by Lemma 2.1.1 (in Subsection 2.1) there exist sets A_0, B_0 each of which is a finite disjoint union of elements of $\mathcal{S}$ such that

$$\mu(A\Delta A_0) < \epsilon, \quad \text{and } \mu(B\Delta B_0) < \epsilon.$$

Since,

$$(T^{-i}A \cap B)\Delta(T^{-i}A_0 \cap B_0) \subseteq (T^{-i}A\Delta T^{-i}A_0) \cup (B\Delta B_0),$$

it follows that

$$\begin{aligned} |\mu(T^{-i}A \cap B) - \mu(T^{-i}A_0 \cap B_0)| &\leq \mu\left[(T^{-i}A \cap B)\Delta(T^{-i}A_0 \cap B_0)\right] \\ &\leq \mu(T^{-i}A\Delta T^{-i}A_0) + \mu(B\Delta B_0) \\ &< 2\epsilon. \end{aligned}$$

Further,

$$\begin{aligned} |\mu(A)\mu(B) - \mu(A_0)\mu(B_0)| &\leq \mu(A)|\mu(B) - \mu(B_0)| + \mu(B_0)|\mu(A) - \mu(A_0)| \\ &\leq |\mu(B) - \mu(B_0)| + |\mu(A) - \mu(A_0)| \\ &\leq \mu(B\Delta B_0) + \mu(A\Delta A_0) \\ &< 2\epsilon. \end{aligned}$$

Hence,

$$\left| \left(\frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) - \mu(A)\mu(B) \right) - \left(\frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A_0 \cap B_0) - \mu(A_0)\mu(B_0) \right) \right|$$

$$\begin{aligned} &\leq \frac{1}{n} \sum_{i=0}^{n-1} \left| \mu(T^{-i}A \cap B) + \mu(T^{-i}A_0 \cap B_0) \right| - |\mu(A)\mu(B) - \mu(A_0)\mu(B_0)| \\ &< 4\epsilon. \end{aligned}$$

Therefore,

$$\lim_{n\to\infty} \left[\frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) - \mu(A)\mu(B) \right] = 0.$$

□

Theorem 2.3.2 *Suppose μ_1 and μ_2 are probability measures on $(X, \mathcal{F})$, and $T : X \to X$ is measurable and measure preserving with respect to μ_1 and μ_2. Then,*

(i) *if T is ergodic with respect to μ_1, and μ_2 is absolutely continuous with respect to μ_1, then $\mu_1 = \mu_2$,*

(ii) *if T is ergodic with respect to μ_1 and μ_2, then either $\mu_1 = \mu_2$ or μ_1 and μ_2 are singular with respect to each other.*

Proof. (i) Suppose T is ergodic with respect to μ_1 and μ_2 is absolutely continuous with respect to μ_1. For any $A \in \mathcal{F}$, by the ergodic theorem for a.e. x one has

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_A(T^i x) = \mu_1(A).$$

Let

$$C_A = \{x \in X : \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_A(T^i x) = \mu_1(A)\},$$

then $\mu_1(C_A) = 1$, and by absolute continuity of μ_2 one has $\mu_2(C_A) = 1$. Since T is measure preserving with respect to μ_2, for each $n \geq 1$ one has

$$\frac{1}{n} \sum_{i=0}^{n-1} \int_X 1_A(T^i x)\, \mathrm{d}\mu_2(x) = \mu_2(A).$$

On the other hand, by the dominated convergence theorem one has

$$\lim_{n\to\infty} \int_X \frac{1}{n} \sum_{i=0}^{n-1} 1_A(T^i x) d\mu_2(x) = \int_X \mu_1(A)\, \mathrm{d}\mu_2(x).$$

This implies that $\mu_1(A) = \mu_2(A)$. Since $A \in \mathcal{F}$ is arbitrary, we have $\mu_1 = \mu_2$.

(ii) Suppose T is ergodic with respect to μ_1 and μ_2. Assume that $\mu_1 \neq \mu_2$. Then, there exists a set $A \in \mathcal{F}$ such that $\mu_1(A) \neq \mu_2(A)$. For $i = 1, 2$ let

$$C_i = \{x \in X : \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} 1_A(T^j x) = \mu_i(A)\}.$$

By the ergodic theorem $\mu_i(C_i) = 1$ for $i = 1, 2$. Since $\mu_1(A) \neq \mu_2(A)$, then $C_1 \cap C_2 = \emptyset$. Thus μ_1 and μ_2 are supported on disjoint sets, and hence μ_1 and μ_2 are mutually singular. □

2.4 Mixing

As a corollary to the ergodic theorem we found a new definition of ergodicity; namely, asymptotic average independence. Based on the same idea, we now define other notions of *weak* independence that are stronger than ergodicity.

Definition 2.4.1 *Let $(X, \mathcal{F}, \mu)$ be a probability space, and $T : X \to X$ a measure preserving transformation. Then,*

(i) *T is* weakly mixing *if for all $A, B \in \mathcal{F}$, one has*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left|\mu(T^{-i}A \cap B) - \mu(A)\mu(B)\right| = 0. \tag{2.3}$$

(ii) *T is* strongly mixing *if for all $A, B \in \mathcal{F}$, one has*

$$\lim_{n\to\infty} \mu(T^{-i}A \cap B) = \mu(A)\mu(B). \tag{2.4}$$

Notice that strongly mixing implies weakly mixing, and weakly mixing implies ergodicity. This follows from the simple fact that if $\{a_n\}$ is a sequence of real numbers such that $\lim_{n\to\infty} a_n = 0$, then $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i| = 0$, and hence $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} a_i = 0$. Furthermore, if $\{a_n\}$ is a bounded sequence, then the following are equivalent (see [W] for the proof):

(i) $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i| = 0$

(ii) $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i|^2 = 0$

(iii) there exists a subset J of the integers of density zero, i.e.

$$\lim_{n\to\infty}\frac{1}{n}\#\left(\{0,1,\ldots,n-1\}\cap J\right)=0,$$

such that $\lim_{n\to\infty, n\notin J} a_n = 0$.

Using this one can give three equivalent characterizations of weakly mixing transformations.

Proposition 2.4.1 *Let $(X,\mathcal{F},\mu)$ be a probability space, and $T : X \to X$ a measure preserving transformation. Let $\mathcal{S}$ be a generating semi-algebra of $\mathcal{F}$.*

(a) *If Equation* (2.3) *holds for all $A, B \in \mathcal{S}$, then T is weakly mixing.*

(b) *If Equation* (2.4) *holds for all $A, B \in \mathcal{S}$, then T is strongly mixing.*

Chapter 3

Examples Revisited

In this chapter, we will study the ergodic behavior of the examples given in Chapter 1. For each map we will give an invariant ergodic measure absolutely continuous with respect to Lebesgue measure. For all the examples, the invariance of the measure is verified on intervals (see Theorem 2.1.2), and the ergodicity is shown using Knopp's Lemma (Lemma 2.2.1).

Example 3.0.1 (Binary expansion revisited) Consider the map of example 1.2.1. We will show that Lebesgue measure λ is T-invariant (or that T is measure preserving with respect to λ).

For any interval $[a,b)$,

$$T^{-1}[a,b) = \left[\frac{a}{2}, \frac{b}{2}\right) \bigcup \left[\frac{a+1}{2}, \frac{b+1}{2}\right),$$

and

$$\lambda\left(T^{-1}[a,b)\right) = b - a = \lambda\left([a,b)\right).$$

hence, by Theorem 2.1.2 we see that λ is T-invariant. For ergodicity we use Knopp's Lemma. To this end, let $\mathcal{C}$ be the collection of all intervals of the form $[k/2^n,(k+1)/2^n)$ with $n \geq 1$ and $0 \leq k \leq 2^n - 1$. Notice that the the set $\{k/2^n : n \geq 1,\, 0 \leq k < 2^n - 1\}$ of dyadic rationals is dense in $[0,1)$, hence each open interval is at most a countable union of disjoint elements of $\mathcal{C}$. Hence, $\mathcal{C}$ satisfies the first hypothesis of Knopp's Lemma. Now, T^n maps each dyadic interval of the form $[k/2^n,(k+1)/2^n)$ linearly onto $[0,1)$, (we call such an interval dyadic of order n); in fact, $T^n x = 2^n x \bmod(1)$. Let $B \in \mathcal{B}$ be T-invariant, and assume $\lambda(B) > 0$. Let $A \in \mathcal{C}$, and assume that A is dyadic of order n. Then, $T^n A = [0,1)$ and $\lambda(A) = 2^{-n}$. Furthermore, for any Lebesgue measurable set C,

$$\lambda(T^{-n}C \cap A) = 2^{-n}\lambda(C).$$

Hence, by invariance of of the set B we have

$$\lambda(B \cap A) = \lambda(T^{-n}B \cap A) = 2^{-n}\lambda(B) = \lambda(B)\lambda(A),$$

Thus, the second hypothesis of Knopp's Lemma is satisfied with $\gamma = \lambda(B) > 0$. Hence, $\lambda(B) = 1$. Therefore T is ergodic.

Example 3.0.2 (m-ary expansion revisited) Consider the map T of example 1.2.2. A slight modification of the arguments used in the above example show that T is measure preserving and ergodic with respect to Lebesgue measure λ.

Example 3.0.3 (Greedy expansion revisited) Consider the transformation of example 1.2.3. It is easy to see that Lebesgue measure is not invariant. We are seeking a T_β-invariant measure μ of the form $\mu_\beta(A) = \int_A h_\beta(x) d\lambda(x)$ for any Borel set A. It is not hard to see that the interval $[0,1)$ is an attractor in the sense that for any $x \in [0, \frac{\lfloor\beta\rfloor}{1-\beta})$, there exists $n \geq 0$ such that $T_\beta^m x \in [0,1)$ for all $m \geq n$. Independently, A.O. Gel'fond (in 1959) [G] and W. Parry [P] (in 1960) showed that

$$h_\beta(x) = \begin{cases} \frac{1}{F(\beta)} \sum_{n=0}^{\infty} \frac{1}{\beta^n} \, 1_{[0,T_\beta^n(1))}(x) & x \in [0,1) \\ 0 & x \in [1, [0, \frac{\lfloor\beta\rfloor}{1-\beta}), \end{cases}$$

where $F(\beta) = \int_0^1 (\sum_{x<T_\beta^n(1)} \frac{1}{\beta^n}) \mathrm{d}x$ is a normalizing constant. The T_β-invariance of μ_β follows from the equality (proven by Parry) $\beta h_\beta(x) = \sum\limits_{y:T_\beta y = x} h_\beta(y)$.

To prove ergodicity, we need few definitions first. From now on, we will consider T_β as a map on $[0,1]$. We define fundamental intervals (of rank n) in the usual way: the intervals of rank 1 are $\Delta(i) = \{x : a_1(x) = i\} = I_i$, for $i \in \{0,1\}$, and the intervals of rank n, for $n \geq 2$ are

$$\begin{aligned} \Delta(i_1,\ldots,i_n) &= \Delta(i_1) \cap T_\beta^{-1}\Delta(i_2) \cap \cdots \cap T_\beta^{-(n-1)}\Delta(i_n) \\ &= \{x : a_1(x) = i_1, \ldots, a_n(x) = i_n\} \\ &= \left\{ x : x = \sum_{j=1}^n \frac{i_j}{\beta^j} + \frac{T_\beta^n x}{\beta^n} \right\}, \end{aligned}$$

A fundamental interval $\Delta(i_1,\ldots,i_n)$ is *full* if $T^n\Delta(i_1,\ldots,i_n) = [0,1)$, i.e. $\lambda\left(T^n(\Delta(i_1,\ldots,i_n))\right) = 1$, here λ denotes Lebesgue measure on $[0,1]$. From the above we see that if $\Delta(i_1,\ldots,i_n)$ is full, it is equal to the interval

$$[\sum_{j=1}^n \frac{i_j}{\beta^j}, \sum_{j=1}^n \frac{i_j}{\beta^j} + \frac{1}{\beta^n}).$$

Let $\mathcal{C}$ be the collection of all full fundamental intervals of all ranks. We show that $\mathcal{C}$ generates the Borel σ-algebra. To this end, let B_n be the collection of non-full intervals of rank n that are not subsets of full intervals of lower rank. Note that $\Delta(\lfloor\beta\rfloor)$ is the only member of B_1. Suppose that $\Delta(i_1,\ldots,i_n)$ is an element of B_n, then $\Delta(i_1,\ldots,i_j) \in B_j$ for $1 \leq j \leq n-1$. We claim that $\Delta(i_1,\ldots,i_n)$ contains at most one element of B_{n+1}. There are two cases:

— If $T_\beta^n 1 = \frac{k}{\beta}$ for some $k = 1, \ldots, \lfloor\beta\rfloor$, then all $(n+1)$ order fundamental intervals are full, and B_{n+1} is empty.

— If $T_\beta^n 1 = \Delta(k)^o$ (interior) for some $k = 0, 1, \ldots, \lfloor\beta\rfloor$, then $\Delta(i_1, \ldots, i_n, j)$, $j = 0, 1 \ldots, k-1$ are full, and $\Delta(i_1, \ldots, i_n, k)$ is non-full, and hence in B_{n+1}.

Since $|B_1| = 1$, it thus follows by induction from the above that $|B_n| \leq 1$ for all n. Denote by Δ_n^* the unique element of B_n (note that Δ_n^* could be empty).

We are now ready to show that the collection $\mathcal{C}$ of full intervals generates the Borel σ-algebra on $[0,1]$. Let F_n be the collection of all full intervals of rank n, and let D_n be the collection of full intervals of rank n that are not subsets of full intervals of lower rank, i.e.,

$$D_n = \{\Delta(i_1, \ldots, i_n) \in F_n : \Delta(i_1, \ldots, i_j) \notin F_j \text{ for any } 1 \leq j \leq n-1\}.$$

the union of all full intervals that are not subsets of full intervals of lower rank has full Lebesgue measure, i.e., for any $N \geq 1$,

$$\lambda\left([0,1) \setminus \bigcup_{n=1}^{N} \bigcup_{D_n} \Delta(j_1, \ldots, j_n)\right) = \lambda\left(\Delta_n^*\right)) < \frac{1}{\beta^N}.$$

Taking the limit as N tends to infinity, we get

$$\lambda\left([0,1) \setminus \bigcup_{n=1}^{\infty} \bigcup_{D_n} \Delta(j_1, \ldots, j_n)\right) = 0.$$

So applying a similar procedure to any interval, we find that any interval can be covered by a countable disjoint union of full intervals, so $\mathcal{C}$ generates, and $\mathcal{C}$ satisfies condition (i) of Knopp's Lemma. Now let B be a T_β-invariant set, and let E be a full interval of rank n. Note that $\lambda(E) = \beta^{-n}$, and for any Lebesgue measurable set C,

$$\lambda(T_\beta^{-n} C \cap E) = \beta^{-n} \lambda(C).$$

Hence, by invariance of of the set B we have

$$\lambda(B \cap E) = \lambda(T_\beta^{-n} B \cap E) = \beta^{-n} \lambda(B) = \lambda(B)\lambda(E),$$

By Knopp's Lemma with $\gamma = \lambda(B)$ we get that $\lambda(B) = 1$, and hence $\mu_\beta(B) = 1$ (since λ and μ_β are equivalent on $[0,1]$). Therefore, T_β is ergodic with respect to μ_β. □

Example 3.0.4 (Lazy expansions revisited) Consider the map S_β of Example 1.2.4. The dynamical behaviour of this map is essentially the same as that of T_β in the previous example. In the language of ergodic theory these two maps are isomorphic. To be more precise, consider the map $\psi : [0, \lfloor\beta\rfloor/(\beta-1)) \to (0, \lfloor\beta\rfloor/(\beta-1)]$ defined by

$$\psi(x) = \frac{\lfloor\beta\rfloor}{\beta - 1} - x,$$

then ψ is continuous, hence Borel measurable, and $\psi T_\beta = S_\beta \psi$. The latter equality implies that the absolutely continuous measure ρ_β defined by

$$\rho_\beta(A) = \mu_\beta(\psi^{-1}(A)),$$

(A a Borel set) is S_β- invariant. The ergodicity of S_β with respect to the measure ρ_β follows again from the commuting relation $\psi T_\beta = S_\beta \psi$. For if A is an S_β-invariant Borel set, then $\psi^{-1}A$ is a T_β invariant set. By ergodicity of T_β we have $\mu_\beta(\psi^{-1}(A))$ equals 0 or 1. Since $\rho_\beta(A) = \mu_\beta(\psi^{-1}(A))$, ergodicity of S_β follows.

Example 3.0.5 (Lüroth series revisited) Consider the map T of Example 1.2.5. We show that T is measure preserving and ergodic with respect to Lebesgue measure λ. Using the definition of T, for any interval $[a, b)$ of $[0, 1)$ one has

$$T^{-1}[a,b) = \bigcup_{k=2}^{\infty} \left[\frac{1}{k} + \frac{a}{k(k-1)}, \frac{1}{k} + \frac{b}{k(k-1)}\right),$$

Hence $\lambda(T^{-1}[a,b)) = \lambda([a,b))$, and T is measure preserving with respect to λ. Ergodicity follows again from Knopp's Lemma. The collection $\mathcal{C}$ consists in this case of all fundamental intervals of all ranks. A fundamental interval of rank n is a set of the form

$$\begin{aligned} \Delta(i_1, i_2, \ldots, i_k) &= \Delta(i_1) \cap T^{-1}\Delta(i_2) \cap \cdots \cap \Delta(i_n) \\ &= \{x : a_1(x) = i_1, a_2(x) = i_2, \ldots, a_k(x) = i_k\}. \end{aligned}$$

Notice that $\Delta(i_1, i_2, \ldots, i_k)$ is an interval with end points

$$\frac{P_k}{Q_k} \quad \text{and} \quad \frac{P_k}{Q_k} + \frac{1}{i_1(i_1-1)\cdots i_k(i_k-1)},$$

where

$$P_k/Q_k = \frac{1}{i_1} + \frac{1}{i_1(i_1-1)i_2} + \cdots + \frac{1}{i_1(i_1-1)\cdots i_{k-1}(i_{k-1}-1)i_k}.$$

Furthermore, $T^n(\Delta(i_1, i_2, \ldots, i_k)) = [0, 1)$, and T^n restricted to $\Delta(i_1, i_2, \ldots, i_k)$ has slope

$$i_1(i_1-1)\cdots i_{k-1}(i_{k-1}-1)i_k = \frac{1}{\lambda(\Delta(i_1, i_2, \ldots, i_k))}.$$

Since $\lim_{k\to\infty} \operatorname{diam}(\Delta(i_1, i_2, \ldots, i_k)) = 0$ for any sequence $i_1, i_2, \cdots$, the collection $\mathcal{C}$ generates the Borel σ-algebra. Now let A be a T-invariant Borel set of positive Lebesgue measure, and let E be any fundamental interval of rank n, then

$$\lambda(A \cap E) = \lambda(T^{-n}A \cap E) = \lambda(E)\lambda(A).$$

By Knopp's Lemma with $\gamma = \lambda(A)$ we get that $\lambda(A) = 1$; i.e. T is ergodic with respect to λ. □

Example 3.0.6 (Generalized Lüroth series revisited) Our transformation is as given in Example 1.2.6. We will show again that T is measure preserving and ergodic with respect to Lebesgue measure λ. For any interval $[a,b)$ of $[0,1)$,

$$\begin{aligned} T^{-1}[a,b) &= \left(T^{-1}[a,b) \cap \bigcup_n I_n\right) \cup \left(T^{-1}[a,b) \cap I_\infty\right) \\ &= \bigcup_n ((r_n - \ell_n)a + \ell_n, (r_n - \ell_n)b + \ell_n) \cup \left(T^{-1}(a,b) \cap I_\infty\right). \end{aligned}$$

Since $\lambda(I_\infty) = 0$, it follows that

$$\lambda\left(T^{-1}[a,b)\right) = \sum_n (r_n - \ell_n)(b-a) = b - a = \lambda([a,b)).$$

So λ is T-invariant. Before we prove ergodicity, we introduce few notations similar to those in the above examples.

A GLS expansion is identified with the partition $\mathcal{I}$ and the index (or digit) set $\mathcal{D}$. Let x have an infinite GLS($\mathcal{I}$) expansion, given by

$$x = \frac{h_1}{s_1} + \frac{h_2}{s_1 s_2} + \cdots + \frac{h_k}{s_1 s_2 \cdots s_k} + \cdots .$$

Now h_k and s_k are identified once we know in which partition element $T^{k-1}x$ lies (h_k and s_k are constants determined by partition elements). Therefore, to determine the GLS-expansion of x (for a given $\mathcal{I}$ and $\mathcal{D}$) we only need to keep track of which partition elements the orbit of x visits. For $x \in [0,1)$ we define the sequence of digits $a_n = a_n(x)$, $n \geq 1$, as follows

$$a_n = k \iff T^{n-1}x \in I_k\,,\ k \in \mathcal{D} \cup \{\infty\}.$$

Thus the values of the digits of points $x \in [0,1)$ are elements of $\mathcal{D}$; this is why $\mathcal{D}$ was called the digit set.

Notice that every GLS expansion determines a unique sequence of digits, and conversely. So

$$x = \sum_{k=1}^{\infty} \frac{h_k}{s_1 s_2 \ldots s_k} =: [\,a_1, a_2, \ldots].$$

We can now define fundamental intervals (or cylinder sets) in the usual way. Setting

$$\Delta(i) = \{x : a_1(x) = i\} \text{ if } i \in \mathcal{D} \cup \{\infty\},$$

then

$$\Delta(i) = [l_i, r_i) \text{ if } i \in \mathcal{D}, \text{ and } \Delta(\infty) = I_\infty.$$

For $i_1, i_2, \ldots, i_n \in \mathcal{D} \cup \{\infty\}$, define

$$\Delta(i_1, i_2, \ldots, i_n) = \{x : a_1(x) = i_1, a_2(x) = i_2, \ldots, a_n(x) = i_n\}.$$

Notice that, if $i_j = \infty$ for some $1 \le j \le n$, then $\Delta(i_1, i_2, \ldots, i_n)$ is a subset of a set of measure zero, namely the set consisting of all points in $(0,1)$ whose orbit hits I_∞.

Let us determine the cylinder sets $\Delta(i_1, \ldots, i_k)$, for $i_1, i_2, \ldots, i_n \in \mathcal{D}$. All points x with the same first k digits have the same first k terms in their GLS expansion. Let us call the sum of the first k terms p_k/q_k; then

$$x = \frac{p_k}{q_k} + \frac{T^k x}{s_1 \cdots s_k},$$

where $s_j = 1/L_{i_j}$ and $T^k x$ can vary freely in $[0,1)$. This implies that

$$\Delta(i_1, \ldots, i_k) = \Big[\frac{p_k}{q_k}, \frac{p_k}{q_k} + \frac{1}{s_1 \cdots s_k}\Big),$$

from which we clearly have

$$\lambda(\Delta(i_1, \ldots, i_k)) = \frac{1}{s_1 \cdots s_k},$$

where $s_1 \cdots s_k$ is the slope of the restriction of T^k to the fundamental interval $\Delta(i_1, \ldots, i_k)$. Since $L_{i_j} = 1/s_j$ for each j, we find that

$$\lambda(\Delta(i_1, \ldots, i_k)) = L_{i_1} L_{i_2} \cdots L_{i_k} = \lambda(\Delta(i_1))\lambda(\Delta(i_2)) \cdots \lambda(\Delta(i_k)).$$

Hence the digits are independent. If we let $\mathcal{C}$ be the collection of all fundamental intervals of all rank, by a similar reasoning as in the above examples, the collection $\mathcal{C}$ generates the Borel σ-algebra. let A be a T-invariant Borel set of positive Lebesgue measure, and let E be any fundamental interval of rank n, then

$$\lambda(A \cap E) = \lambda(T^{-n} A \cap E) = \lambda(E)\lambda(A).$$

By Knopp's Lemma with $\gamma = \lambda(A)$ we get that $\lambda(A) = 1$; i.e., T is ergodic with respect to λ. $\square$

Example 3.0.7 (Continued fractions revisited) Consider the map T of Example 1.2.7. One can easily see that Lebesgue measure λ is not T-invariant. However, there exists a T-invariant measure μ which is equivalent to Lebesgue measure on the interval $[0,1)$. This invariant measure was found by Gauss in 1800, and is known nowadays as the *Gauss measure* which is given by

$$\mu(A) = \frac{1}{\log 2} \int_A \frac{1}{1+x}\, \mathrm{d}x$$

for all Borel sets $A \subset [0,1)$, where log refers to the natural logarithm; see Figure 3.3.

Nobody knows how Gauss found μ, and his achievement is even more remarkable if we realize that modern probability theory and ergodic theory started almost a century later! In general, finding the invariant measure is a difficult

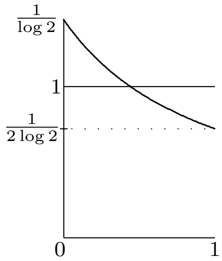

Figure 3.1: The densities of Lebesgue measure λ and Gauss measure μ.

task. The T invariance of μ can be verified on intervals of the form $[a, b)$. Easy calculations show that

$$T^{-1}[a,b) = \bigcup_{n=1}^{\infty} \left(\frac{1}{n+b}, \frac{1}{n+a} \right] ,$$

and $\mu([a, b)) = \mu(T^{-1}[a, b))$. Ergodicity is again proved by Knopp's Lemma. We first define the notion of fundamental intervals similar to the above examples. A fundamental interval of order n is a set of the form

$$\Delta(a_1, \ldots, a_n) := \{x \in [0,1) : a_1(x) = a_1, \ldots, a_n(x) = a_n\} ,$$

where $a_j \in \mathbb{N}$ for each $1 \le j \le n$. When $a_1, \ldots, a_n$ are fixed, we sometimes write Δ_n instead of $\Delta(a_1, a_2, \ldots, a_n)$. We list few properties of these sets without proofs, and we refer to [DK] for more details.

(i) $\Delta(a_1, a_2, \ldots, a_k)$ is an interval in $[0, 1)$ with endpoints

$$\frac{p_k}{q_k} \quad \text{and} \quad \frac{p_k + p_{k-1}}{q_k + q_{k-1}},$$

where $\dfrac{p_n}{q_n} = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_n}}}$.

(ii) The sequences $(p_n)_{n \ge -1}$ and $(q_n)_{n \ge -1}$ satisfy the following recurrence relations[1]

$$\begin{aligned} p_{-1} &:= 1; \quad p_0 := a_0; \quad p_n = a_n p_{n-1} + p_{n-2} , \ n \ge 1, \\ q_{-1} &:= 0; \quad q_0 := 1; \quad q_n = a_n q_{n-1} + q_{n-2} , \ n \ge 1. \end{aligned} \tag{3.1}$$

Furthermore, $p_n(x) = q_{n-1}(Tx)$ for all $n \ge 0$, and $x \in (0, 1)$.

[1] A proof of these recurrence formulas will be given in Section 4.2.2.

(iii)
$$\lambda\left(\Delta(a_1,a_2,\ldots,a_k)\right) \;=\; \frac{1}{q_k(q_k+q_{k-1})},$$

and
$$\mu(\Delta(a_k,a_{k-1},\ldots,a_1)) \;=\; \mu(\Delta(a_1,a_2,\ldots,a_k)).$$

(iv) If $0\le a<b\le 1$, that $\{x\,:\,a\le T^n x<b\}\cap\Delta_n$ equals
$$\left[\frac{p_{n-1}a+p_n}{q_{n-1}a+q_n},\,\frac{p_{n-1}b+p_n}{q_{n-1}b+q_n}\right)$$
when n is even, and equals
$$\left(\frac{p_{n-1}b+p_n}{q_{n-1}b+q_n},\,\frac{p_{n-1}a+p_n}{q_{n-1}a+q_n}\right]$$
for n odd. Here $\Delta_n \;=\; \Delta_n(a_1,\ldots,a_n)$ is a fundamental interval of rank n. This leads to
$$\lambda(T^{-n}[a,b)\cap\Delta_n)=\lambda([a,b))\lambda(\Delta_n)\frac{q_n(q_{n-1}+q_n)}{(q_{n-1}b+q_n)(q_{n-1}a+q_n)}\,.$$

Since
$$\frac{1}{2}\;<\;\frac{q_n}{q_{n-1}+q_n}\;<\;\frac{q_n(q_{n-1}+q_n)}{(q_{n-1}b+q_n)(q_{n-1}a+q_n)}\;<\;\frac{q_n(q_{n-1}+q_n)}{q_n^2}\;<\;2\,.$$

Therefore we find for every interval I, that
$$\frac{1}{2}\lambda(I)\lambda(\Delta_n)\;<\;\lambda(T^{-n}I\cap\Delta_n)\;<\;2\lambda(I)\lambda(\Delta_n)\,.$$

Let A be a finite disjoint union of such intervals I. Since Lebesgue measure is additive one has
$$\frac{1}{2}\lambda(A)\lambda(\Delta_n)\;\le\;\lambda(T^{-n}A\cap\Delta_n)\;\le\;2\lambda(A)\lambda(\Delta_n)\,. \tag{3.2}$$

The collection of finite disjoint unions of such intervals generates the Borel σ-algebra. It follows that (3.2) holds for any Borel set A.

(v) For any Borel set A one has
$$\frac{1}{2\log 2}\lambda(A)\;\le\;\mu(A)\;\le\;\frac{1}{\log 2}\lambda(A)\,, \tag{3.3}$$

hence by (3.2) and (3.3) one has
$$\mu(T^{-n}A\cap\Delta_n)\;\ge\;\frac{\log 2}{4}\mu(A)\mu(\Delta_n)\;. \tag{3.4}$$

Now let $\mathcal{C}$ be the collection of all fundamental intervals Δ_n. Since the set of all endpoints of these fundamental intervals is the set of all rationals in $[0,1)$, it follows that condition (a) of Knopp's Lemma is satisfied. Now suppose that B is invariant with respect to T and $\mu(B) > 0$. Then it follows from (3.4) that for every fundamental interval Δ_n

$$\mu(B \cap \Delta_n) \geq \frac{\log 2}{4}\mu(B)\mu(\Delta_n) .$$

So condition (b) from Knopp's Lemma is satisfied with $\gamma = \frac{\log 2}{4}\mu(B)$; thus $\mu(B) = 1$; i.e. T is ergodic. □

We now use the ergodic Theorem to give simple proofs of old and famous results of Paul Lévy; see [Le].

Proposition 3.0.2 (Paul Lévy, 1929) *For almost all $x \in [0,1)$ one has*

$$\lim_{n\to\infty} \frac{1}{n}\log q_n = \frac{\pi^2}{12\log 2}, \tag{3.5}$$

$$\lim_{n\to\infty} \frac{1}{n}\log(\lambda(\Delta_n)) = \frac{-\pi^2}{6\log 2}, \quad \textit{and} \tag{3.6}$$

$$\lim_{n\to\infty} \frac{1}{n}\log|x - \frac{p_n}{q_n}| = \frac{-\pi^2}{6\log 2}. \tag{3.7}$$

Proof. By property (ii) above, for any irrational $x \in [0,1)$ one has

$$\begin{aligned}
\frac{1}{q_n(x)} &= \frac{1}{q_n(x)}\frac{p_n(x)}{q_{n-1}(Tx)}\frac{p_{n-1}(Tx)}{q_{n-2}(T^2x)}\cdots\frac{p_2(T^{n-2}x)}{q_1(T^{n-1}x)} \\
&= \frac{p_n(x)}{q_n(x)}\frac{p_{n-1}(Tx)}{q_{n-1}(Tx)}\cdots\frac{p_1(T^{n-1}x)}{q_1(T^{n-1}x)}.
\end{aligned}$$

Taking logarithms yields

$$-\log q_n(x) = \log\frac{p_n(x)}{q_n(x)} + \log\frac{p_{n-1}(Tx)}{q_{n-1}(Tx)} + \cdots + \log\frac{p_1(T^{n-1}x)}{q_1(T^{n-1}x)}. \tag{3.8}$$

For any $k \in \mathbb{N}$, and any irrational $x \in [0,1)$, $\frac{p_k(x)}{q_k(x)}$ is a rational number close to x. Therefore we compare the right-hand side of (3.8) with

$$\log x + \log Tx + \log T^2x + \cdots + \log(T^{n-1}x).$$

We have

$$-\log q_n(x) = \log x + \log Tx + \log T^2x + \cdots + \log(T^{n-1}x) + R(n,x).$$

In order to estimate the error term $R(n,x)$, we recall from property (i) that x lies in the interval Δ_n, which has endpoints p_n/q_n and $(p_n+p_{n-1})/(q_n+q_{n-1})$. Therefore, in case n is even, one has

$$0 < \log x - \log\frac{p_n}{q_n} = (x - \frac{p_n}{q_n})\frac{1}{\xi} \leq \frac{1}{q_n(q_{n-1}+q_n)}\frac{1}{p_n/q_n} < \frac{1}{q_n},$$

where $\xi \in (p_n/q_n, x)$ is given by the mean value theorem. Let $\mathcal{F}_1, \mathcal{F}_2, \cdots$ be the sequence of Fibonacci $1, 1, 2, 3, 5, \ldots$ (these are the q_i's of the small golden ratio $g = 1/G$). It follows from the recurrence relation for the q_i's (property (ii)) that $q_n(x) \geq \mathcal{F}_n$. A similar argument shows that

$$\frac{1}{q_n} < \log x - \log \frac{p_n}{q_n},$$

in case n is odd. Thus

$$|R(n,x)| \leq \frac{1}{\mathcal{F}_n} + \frac{1}{\mathcal{F}_{n-1}} + \cdots + \frac{1}{\mathcal{F}_1},$$

and since we have

$$\mathcal{F}_n = \frac{G^n + (-1)^{n+1} g^n}{\sqrt{5}}$$

it follows that $\mathcal{F}_n \sim \frac{1}{\sqrt{5}} G^n$, $n \to \infty$. Thus $\frac{1}{\mathcal{F}_n} + \frac{1}{\mathcal{F}_{n-1}} + \cdots + \frac{1}{\mathcal{F}_1}$ is the nth partial sum of a convergent series, and therefore

$$|R(n,x)| \leq \frac{1}{\mathcal{F}_n} + \cdots + \frac{1}{\mathcal{F}_1} \leq \sum_{n=1}^{\infty} \frac{1}{\mathcal{F}_n} := \mathcal{C}.$$

Hence for each x for which

$$\lim_{n\to\infty} \frac{1}{n}(\log x + \log Tx + \log T^2 x + \cdots + \log(T^{n-1}x))$$

exists,

$$-\lim_{n\to\infty} \frac{1}{n} \log q_n(x)$$

exists too, and these limits are equal.

Now $\lim_{n\to\infty} \frac{1}{n}(\log x + \log Tx + \log T^2 x + \cdots + \log(T^{n-1}x))$ is ideally suited for the Ergodic Theorem; we only need to check that the conditions of the Ergodic Theorem are satisfied and to calculate the integral. This is left as an exercise for the reader. This proves (3.5).

It follows from Property (iii) above that

$$\lambda(\Delta_n(a_1, \ldots, a_n)) = \frac{1}{q_n(q_n + q_{n-1})};$$

thus

$$-\log 2 - 2\log q_n < \log \lambda(\Delta_n) < -2\log q_n.$$

Now apply (3.5) to obtain (3.6). Finally (3.7) follows from (3.5) and

$$\frac{1}{2q_n q_{n+1}} < \left| x - \frac{p_n}{q_n} \right| < \frac{1}{q_n q_{n+1}}, \qquad n \geq 1.$$

In Section 4.2.2 we give a proof of this last statement. □

Chapter 4

Natural Extensions

In this Chapter we will show how one constructs an invertible system associated with a given non-invertible system in such a way that all the dynamical properties of the original system are preserved. To this end, suppose $(Y, \mathcal{G}, \nu, S)$ is a *non-invertible* measure-preserving dynamical system. An invertible measure-preserving dynamical system $(X, \mathcal{F}, \mu, T)$ is called a *natural extension* of $(Y, \mathcal{G}, \nu, S)$ if there exists a measurable surjective (a.e.) map $\psi : X \to Y$ such that (i) $\psi \circ T = S \circ \psi$, (ii) $\nu = \mu \circ \psi^{-1}$, and (iii) $\vee_{m=0}^{\infty} T^m \psi^{-1} \mathcal{G} = \mathcal{F}$, where $\bigvee_{k=0}^{\infty} T^k \psi^{-1} \mathcal{G}$ is the smallest σ-algebra containing the σ-algebras $T^k \psi^{-1} \mathcal{G}$ for all $k \geq 0$.

Natural extensions were first introduced by Rohlin in the early 60's (see [Ro]). He gave a canonical way of constructing a natural extension, and he showed that his construction is unique up to isomorphism. In many examples the canonical construction may not be the easiest version to work with, especially if one is seeking an invariant measure of the original system that is absolutely continuous with respect to Lebesgue measure. In this chapter, we will construct natural extensions that are planar, easy to work with and to deduce properties of the original system.

4.1 Natural Extensions of m-adic and β-expansions

Example 4.1.1 (m-adic) For simplicity, we consider the binary map as given in Example 1.2.1, $T : [0, 1) \to [0, 1)$ given by

$$Tx = 2x \bmod 1 = \begin{cases} 2x & 0 \leq x < 1/2 \\ 2x - 1 & 1/2 \leq x < 1. \end{cases}$$

A natural extension of T is the well know Baker's transformation $\mathcal{T} : [0, 1)^2 \to [0, 1)^2$ by

$$\mathcal{T}(x, y) = \begin{cases} (2x, \frac{y}{2}) & 0 \leq x < 1/2 \\ (2x - 1, \frac{y+1}{2}) & 1/2 \leq x < 1. \end{cases}$$

It is easy to see that $\mathcal{T}$ is measurable with respect to product Lebesgue σ-algebra $\mathcal{B} \times \mathcal{B}$, and is measure preserving with respect to $\lambda \times \lambda$. Furthermore, it is straightforward to see that the map $\psi : [0,1)^2 \to [0,1)$ given by $\psi(x,y) = x$ satisfies conditions (i) and (ii) in the definition of the natural extension. It remains to verify that

$$\bigvee_{m\geq 0} \mathcal{T}^m \pi^{-1}\mathcal{B} \;=\; \bigvee_{m\geq 0} \mathcal{T}^m(\mathcal{B} \times [0,1)) \;=\; \mathcal{B} \times \mathcal{B}\,.$$

For this it suffices to show that $\bigvee_{m\geq 0} \mathcal{T}^m(\mathcal{B} \times [0,1))$ contains all the two-dimensional cylinders $\Delta(k_1, \ldots, k_n) \times \Delta(l_1, \ldots, l_m)$, where

$$\Delta(k_1, \ldots, k_n) = \{x : a_1(x) = k_1, \ldots, a_n(x)\},$$

with $a_n(x)$ the n'th binary digit of x, and $k_i \in \{0,1\}$. A closer look at the action of $\mathcal{T}$ shows that

$$\Delta(k_1, \ldots, k_n) \times \Delta(l_1, \ldots, l_m) = \mathcal{T}^m(\Delta(l_m, \ldots, l_1, k_1, \ldots, k_n) \times [0,1)) \in \mathcal{T}^m(\mathcal{B} \times [0,1))$$

□

Example 4.1.2 (Greedy β-expansions) Consider the transformation of example 1.2.3 $T_\beta x = \beta x \pmod 1$. Note that here we restrict the domain to the interval $[0,1)$ which as we saw in Example 3.0.3 is an attractor. We also saw that T_β is invariant with respect to the measure μ_β with density h_β. To build a convenient natural extension of T_β, we first look at the case β is a pseudo-golden mean (or what it commonly know as an mbonacci number, i.e. β is the positive root of the polynomial $x^m - x^{m-1} - \ldots - x - 1$.

Then,

$$1 \;=\; \frac{1}{\beta} \;+\; \frac{1}{\beta^2} \;+\; \cdots \;+\; \frac{1}{\beta^m}\;,$$

so that 1 has a finite β-expansion. Note that in the β-expansion of any $x \in [0,1)$, one can have at most $m-1$ consecutive digits equal to 1.

The underlying space of the natural extension is the set

$$X \;=\; \bigcup_{k=0}^{m-1} \left[T_\beta^{m-k}1,\, T_\beta^{m-k-1}1\right) \times \left[0,\, T_\beta^{k}1\right).$$

equipped with the Lebesgue σ-algebra $\mathcal{L}$ restricted to X, and the two dimensional Lebesgue measure $\overline{\lambda}$ restricted to X. On X we consider the transformation $\mathcal{T}$ given by

$$\mathcal{T}_\beta(x,y) \;:=\; \left(T_\beta x,\; \frac{1}{\beta}(\lfloor \beta x \rfloor + y)\right).$$

It is easy to see that the map $\mathcal{T}$ is measure preserving with respect to $\overline{\lambda}$. If one considers the map $\psi :\to [0,1)$ given by $\psi(x,y) = x$, then a proof similar to that used in the previous example shows that ψ satisfies conditions (i), (ii),

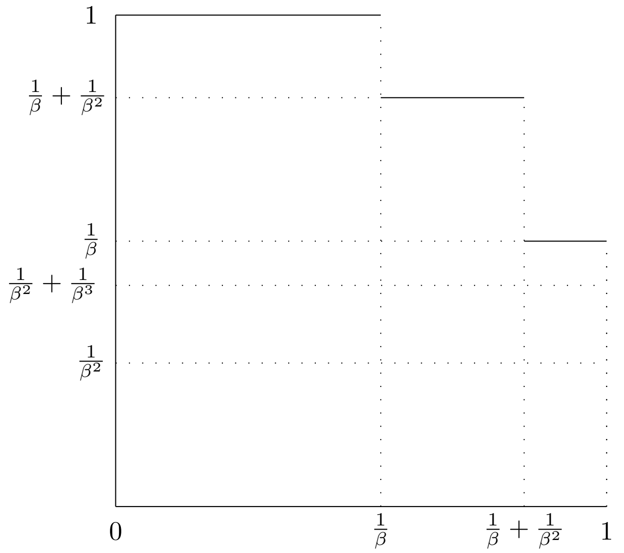

Figure 4.1: The natural extension of T_β if β is mbonacci number with $m = 3$

and (iii) in the definition of the natural extension, i.e. $(X, \mathcal{L}, \overline{\lambda}, \mathcal{T})$ is a natural extension of $([0, 1), \mathcal{B}, \mu_\beta, T_\beta)$.

The general case is a more complicated version of the pseudo-golden mean case. Our aim is to build an invertible dynamical system that captures the past as well as the future of the map T_β We will outline briefly the construction of the natural extension. Let

$$R_0 \;=\; [0,1)^2 \text{ and } R_i \;=\; [0, T_\beta^i 1) \times [0, \frac{1}{\beta^i})\,,\; i \geq 1;$$

the underlying space $\mathcal{H}_\beta$ is obtained by stacking (as pages in a book) R_{i+1} on top of R_i, for each $i \geq 0$. The index i indicates at what height one is in the stack. (In case 1 has a finite β-expansion of length n, only $n\, R_i$'s are stacked.) Let $\mathcal{B}_i$ be the collection of Borel sets of R_i, and let the σ-algebra $\mathcal{F}$ on $\mathcal{H}_\beta$ be the direct sum of the $\mathcal{L}_i$'s, i.e. $\mathcal{F} = \oplus \mathcal{B}_i$. Furthermore, the measure on $\mathcal{H}_\beta$ that is Lebesgue measure on each rectangle R_i is denoted by η, and we put $\mu = \frac{1}{\eta(\mathcal{H}_\beta)}\eta$. Finally $\mathcal{T}_\beta : \mathcal{H}_\beta \to \mathcal{H}_\beta$ is defined as follows.

Let $d^*(\beta) = .b_1 b_2 \ldots$, where $d^*(\beta)$ is the greedy expansion of 1 if it is infinite, and equals $(\overline{d_1 \cdots d_{m-1}(d_m - 1)})$ if the greedy expansion of 1 is finite and equals $(d_1 \cdots d_m)$. Let $(x, y) \in R_i$, $i \geq 0$, where $x = .d_1 d_2 \ldots$ is the β-expansionof x and $y = .\underbrace{00\ldots0}_{i-\text{times}}c_{i+1}c_{i+2}\ldots$ is the β-expansion of y (notice that $(x, y) \in R_i$ implies that $d_1 \leq b_{i+1}$). Define

$$\mathcal{T}_\beta(x,y) \;:=\; (T_\beta x, y^*) \;\in\; \begin{cases} R_0, & \text{if } d_1 < b_{i+1}\,, \\ R_{i+1}, & \text{if } d_1 = b_{i+1}\,, \end{cases} \tag{4.1}$$

where

$$y^* = \begin{cases} \dfrac{b_1}{\beta} + \cdots + \dfrac{b_i}{\beta^i} + \dfrac{d_1}{\beta^{i+1}} + \dfrac{y}{\beta} = .b_1 \cdots b_i d_1 c_{i+1} c_{i+2} \cdots , & \text{if } d_1 < b_{i+1} , \\ \qquad \dfrac{y}{\beta} = \underbrace{.000 \cdots 00}_{i+1-\text{times}} c_{i+1} c_{i+2} \ldots , & \text{if } d_1 = b_{i+1} . \end{cases}$$

Notice that in case $i = 0$ one has

$$y^* = \begin{cases} \dfrac{1}{\beta}(y + d_1) , & d_1 < b_1 , \\ \dfrac{y}{\beta} , & d_1 = b_1 . \end{cases}$$

4.2 Natural Extension of Continued Fractions

We consider the continued fraction map $T : [0,1) \to [0,1)$ as given in Example 1.2.7, i.e. $T0 = 0$ and for $x \neq 0$

$$Tx = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor .$$

We saw in Example 3.0.7 the *Gauss measure* defined by

$$\mu(A) = \frac{1}{\log 2} \int_A \frac{1}{1+x} \, \mathrm{d}x$$

is T-invariant.

A planar and a very useful version of a natural extension of the Continued fraction map was given by Ito-Nakada-Tanaka. We state it without proof.

Theorem 4.2.1 (Ito, Nakada, Tanaka, 1977; Nakada, 1981) *Let $\overline{\Omega} = [0,1) \times [0,1]$, $\overline{\mathcal{B}}$ be the collection of Borel sets of $\overline{\Omega}$. Define the two-dimensional Gauss-measure $\bar{\mu}$ on $(\overline{\Omega}, \bar{\mathcal{B}})$ by*

$$\bar{\mu}(E) = \frac{1}{\log 2} \iint_E \frac{\mathrm{d}x \, \mathrm{d}y}{(1+xy)^2} \ , \ E \in \bar{\mathcal{B}}.$$

Finally, let the two-dimensional RCF*-operator $\mathcal{T} : \overline{\Omega} \to \overline{\Omega}$ for $(x,y) \in \overline{\Omega}$ be defined by*

$$\mathcal{T}(x,y) = \left(T(x), \frac{1}{\left\lfloor \frac{1}{x} \right\rfloor + y} \right), \ x \neq 0, \quad \mathcal{T}(0,y) = (0,y). \tag{4.2}$$

Then $(\overline{\Omega}, \bar{\mathcal{B}}, \bar{\mu}, \mathcal{T})$ is the natural extension of $([0,1), \mathcal{B}, \mu, T)$. Furthermore, it is ergodic.

Clearly $\mathcal{T}$ is a bijective mapping from $\bar{\Omega}$ to $\bar{\Omega}$. For $(x, y) \in \bar{\Omega}$ let $(\xi, \eta) \in \bar{\Omega}$ be such, that $(\xi, \eta) = \mathcal{T}(x, y)$. Then

$$\xi = \frac{1}{x} - c \Leftrightarrow x = \frac{1}{c + \xi},$$

and

$$\eta = \frac{1}{c + y} \Leftrightarrow y = \frac{1}{\eta} - c.$$

Hence the above coordinate transformation has Jacobian J, which satisfies

$$J = \begin{vmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial x}{\partial \eta} \\ \frac{\partial y}{\partial \xi} & \frac{\partial y}{\partial \eta} \end{vmatrix} = \begin{vmatrix} \frac{-1}{(c+\xi)^2} & 0 \\ 0 & \frac{-1}{\eta^2} \end{vmatrix} = \frac{1}{(c+\xi)^2} \frac{1}{\eta^2},$$

and therefore we find

$$\begin{aligned} \bar{\mu}(A) &= \frac{1}{\log 2} \iint_A \frac{\mathrm{d}x\,\mathrm{d}y}{(1+xy)^2} \\ &= \frac{1}{\log 2} \iint_{\mathcal{T}A} \frac{\mathrm{d}\xi\,\mathrm{d}\eta}{(1 + \frac{1}{c+\xi}(\frac{1}{\eta} - c))^2} \frac{1}{(c+\xi)^2 \eta^2} \\ &= \frac{1}{\log 2} \iint_{\mathcal{T}A} \frac{\mathrm{d}\xi\,\mathrm{d}\eta}{(1+\xi\eta)^2} \\ &= \bar{\mu}(\mathcal{T}A). \end{aligned}$$

□

4.2.1 The Doeblin-Lenstra Conjecture

In the rest of this chapter, we show how the natural extention $\mathcal{T}$ of the continued fraction map can be used to solve a conjecture known as The Doeblin-Lenstra Conjecture. Recall first the definition approximation coefficients $\Theta_j(x)$ defined in Example 1.2.7 (see equation (1.7)). In 1981, H.W. Lenstra conjectured that for almost all x the limit

$$\lim_{n\to\infty} \frac{1}{n} \#\{j\,;\, 1 \le j \le n,\, \Theta_j(x) \le z\}, \text{ where } 0 \le z \le 1,$$

exists, and equals the distribution function $F(z)$, given by

$$F(z) \begin{cases} \dfrac{z}{\log 2} & 0 \le z \le \frac{1}{2} \\ \\ \dfrac{1}{\log 2}(1 - z + \log 2z) & \frac{1}{2} \le z \le 1, \end{cases} \tag{4.3}$$

where the $\Theta_n(x)$s are the approximation coefficients as defined in (1.7).

In other words: for almost all x the sequence $(\Theta_n(x))_{n\ge1}$ has limiting distribution F.

An immediate corollary of this conjecture is that for almost all x

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \Theta_j(x) = \frac{1}{4\log 2} = 0.360673\ldots .$$

A first attempt at Lenstra's conjecture was made by D.S. Knuth ([Knu]), who obtained the following theorem

Theorem 4.2.2 (Knuth, 1984) *Let* $K_n(z) = \{x \in [0,1) \setminus \mathbb{Q}\,;\, \Theta_n \le z\}$ *for* $0 \le z \le 1$, *then*

$$\lambda(K_n(z)) = F(z) + \mathcal{O}(g^n)\ ,$$

where F *is defined as in* (4.3).

See also [DK] for a generalization of this result.

When one tries to prove Lenstra's conjecture using the one-dimensional ergodic system $(\Omega, \mathcal{B}, \mu, T)$, one soon realizes that this approach is doomed to fail. If the continued fraction expansion of x is given by $x = [0; a_1, a_2, \ldots]$, then we will see that the variable $\Theta_n(x)$ is essentially "two-dimensional" in the sense that it depends both on the "future" $T_n = [0; a_{n+1}, \ldots]$ and the "past" $V_n = [0; a_n, \ldots, a_1]$. However, he operator T has "no memory of the past."

Lenstra's conjecture was stated earlier – in a slightly different form – by Wolfgang Doeblin, hence the name: the Lenstra-Doeblin conjecture. This conjecture was proved by W. Bosma, H. Jager and F. Wiedijk ([BJW]), using the Ito-Nakada-Tanaka natural extension of the ergodic system $(\Omega, \mathcal{B}, \mu, T)$. Before we give a proof of this result, we look at some more elementary properties.

4.2.2 Some Diophantine spinoff

It follows from the recurrence relations (3.1) that the sequence of denominators q_n is an exponentially fast growing sequence, so indeed, the approximation coefficients $\Theta_n(x)$ really give a very good idea of the quality of the approximation of x by the rational convergent p_n/q_n.

Elementary properties

Note that we actually did not give a proof of the recurrence-relations (3.1); let us fix this, and at the same time find as spinoff a number of rather strong results from Diophantine approximation. We take the long route here, diving deeply into the elementary properties of (regular) continued fractions.

Let $A \in \mathrm{SL}_2(\mathbb{Z})$, that is

$$A = \begin{bmatrix} r & p \\ s & q \end{bmatrix},$$

where $r, s, p, q \in \mathbb{Z}$ and $\det A = rq - ps \in \{\pm 1\}$. Now define the *Möbius* (or: *fractional linear*) *transformation* $A : \mathbb{C}^* \to \mathbb{C}^*$ by

$$A(z) = \begin{bmatrix} r & p \\ s & q \end{bmatrix} (z) \;=\; \frac{rz + p}{sz + q}\;.$$

Let a_1, $a_2, \ldots$ be the sequence of partial quotients of x. Put

$$A_n \;:=\; \begin{bmatrix} 0 & 1 \\ 1 & a_n \end{bmatrix}\;,\;\; n \geq 1 \tag{4.4}$$

and

$$M_n \;:=\; A_1 A_2 \cdots A_n\;,\;\; n \geq 1.$$

Writing

$$M_n \;:=\; \begin{bmatrix} r_n & p_n \\ s_n & q_n \end{bmatrix}\;,\;\; n \geq 1,$$

it follows from $M_n = M_{n-1} A_n$, $n \geq 2$, that

$$\begin{bmatrix} r_n & p_n \\ s_n & q_n \end{bmatrix} = \begin{bmatrix} r_{n-1} & p_{n-1} \\ s_{n-1} & q_{n-1} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & a_n \end{bmatrix},$$

yielding the recurrence relations (3.1).

Now

$$\omega_n \;=\; M_n(0) \;=\; \frac{p_n}{q_n}$$

and from $\det M_n \;=\; (-1)^n$ it follows, that

$$p_{n-1} q_n - p_n q_{n-1} \;=\; (-1)^n\,, \quad n \geq 1, \tag{4.5}$$

hence

$$\gcd(p_n, q_n) = 1\,,\;\; n \geq 1,$$

Setting

$$A_n^* \;:=\; \begin{bmatrix} 0 & 1 \\ 1 & a_n + T_n \end{bmatrix},$$

it follows from

$$x \;=\; M_{n-1} A_n^*(0) \;=\; [\,0;\, a_1, \ldots,\, a_{n-1},\, a_n + T_n\,],$$

that

$$x \;=\; \frac{p_n + p_{n-1} T_n}{q_n + q_{n-1} T_n}\;, \tag{4.6}$$

i.e., $x = M_n(T_n)$. From this and (4.5) one at once has that

$$x - \frac{p_n}{q_n} = \frac{(-1)^n\, T_n}{q_n(q_n + q_{n-1} T_n)}\,. \tag{4.7}$$

In fact, (4.7) yields information about the quality of approximation of the rational number $\omega_n = p_n/q_n$ to the irrational number x. Since $T_n \in [0,1)$, it at once follows that

$$\left| x - \frac{p_n}{q_n} \right| < \frac{1}{q_n^2}\ , \ n \geq 0. \tag{4.8}$$

From $1/T_n = a_{n+1} + T_{n+1}$ one even has

$$\frac{1}{2q_nq_{n+1}} < \left| x - \frac{p_n}{q_n} \right| < \frac{1}{q_nq_{n+1}}, \quad n \geq 1.$$

Notice that the recurrence relations (3.1) yield that

$$\omega_n - \omega_{n-1} = \frac{(-1)^{n+1}}{q_{n-1}q_n}, \quad n \geq 1. \tag{4.9}$$

From this and (4.7) one sees, that

$$0 = \omega_0 < \omega_2 < \omega_4 < \cdots x \cdots < \omega_3 < \omega_1 < 1\,. \tag{4.10}$$

In view of proposition 1 the following questions arise naturally: "Given a sequence of positive integers $(a_n)_{n\geq 1}$, does $\lim_{n\to\infty} \omega_n$ exist? Moreover, in case the limit exists and equals x, do we have that $x = [\,0;\, a_1, \ldots,\, a_n, \ldots]$?" We have the following proposition.

Proposition 4.2.1 *Let $(a_n)_{n\geq 1}$ be a sequence of positive integers, and let the sequence of rational numbers $(\omega_n)_{n\geq 1}$ be given by*

$$\omega_n := [\,0;\, a_1, \ldots, a_n\,],\ n \geq 1.$$

Then there exists an irrational number x for which

$$\lim_{n\to\infty} \omega_n = x$$

and we moreover have that $x = [\,0;\, a_1,\, a_2, \ldots,\, a_n, \ldots]$.

Proof. Writing $\omega_n = p_n/q_n$, $n \geq 1$, $\omega_0 := 0$, where

$$\begin{bmatrix} p_{n-1} & p_n \\ q_{n-1} & q_n \end{bmatrix} = A_1 \cdots A_n$$

and where A_i is defined as in (4.4), one has from (4.9), (4.10) and $\omega_0 := 0$ that

$$\omega_n = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{q_{k-1}q_k}\ ,$$

hence Leibniz' theorem yields that $\lim_{n\to\infty} \omega_n$ exists and equals, say, x. In order to show that the sequence of positive integers $(a_n)_{n\geq 1}$ determines a unique $x \in \mathbb{R}$

we have to show that $a_n = a_n(x)$ for $n \geq 1$, i.e. that $(a_n)_{n\geq 1}$ is the sequence of partial quotients of x. Since

$$\omega_n = [\,0;\, a_1,\, a_2, \ldots,\, a_n\,] = \frac{1}{a_1 + [\,0;\, a_2,\, a_3, \ldots,\, a_n\,]}$$

it is sufficient to show that

$$\left\lfloor \frac{1}{x} \right\rfloor = a_1 \; .$$

However,

$$\omega_n = \frac{1}{a_1 + \omega_n^*} \; ,$$

where $\omega_n^* = [\,0;\, a_2,\, a_3, \ldots,\, a_n\,]$. Hence taking limits $n \to \infty$ yields

$$x = \frac{1}{a_1 + x^*} \; ,$$

here $x^* = \lim_{n\to\infty} \omega_n^*$. From $0 < \omega_2^* < x^* < \omega_3^* < 1$, see also (4.10), and from $1/x = a_1 + x^*$ it now follows that $\lfloor 1/x \rfloor = a_1$. $\square$

Spinoff!

By definition of the approximation coefficients we have

$$\Theta_n = q_n^2 \left| x - \frac{p_n}{q_n} \right| = \frac{T^n(x)}{1 + T^n(x)\frac{q_{n-1}}{q_n}}.$$

By the recurrence-relations (3.1) we have

$$\frac{q_{n-1}}{q_n} = \frac{q_{n-1}}{a_n q_{n-1} + q_{n-2}} = \frac{1}{a_n + \frac{q_{n-2}}{q_{n-1}}},$$

so setting $V_n = q_{n-1}/q_n$ yields for $n \geq 1$ that

$$V_n = \frac{1}{a_n + V_{n-1}} = \cdots = \cfrac{1}{a_n + \cfrac{1}{a_{n-1} + \ddots + \cfrac{1}{a_1}}} = [0; a_n, a_{n-1}, \ldots, a_1].$$

We see that V_n is "the past of x at time n" (in the same way as $T_n = T^n(x)$ is the "future of x at time n"). An immediate consequence of this and (4.7) is that

$$\Theta_n = \Theta_n(x) = \frac{T_n}{1 + T_n V_n}, \quad n \geq 0. \tag{4.11}$$

Furthermore, it is an exercise to show that

$$\Theta_{n-1} = \Theta_{n-1}(x) = \frac{V_n}{1 + T_n V_n}, \quad n \geq 1. \tag{4.12}$$

4.2.3 A proof of the Lenstra-Doeblin Conjecture

In the previous subsection we saw, that for an irrational number $x \in [0,1)$, with RCF-expansion $[0; a_1, a_2 \ldots, a_n, \ldots]$ one has

$$\mathcal{T}^n(x,0) = (T_n, V_n) \quad n = 0, 1, \ldots \ .$$

Clearly one has for every $y \in [0,1]$ that

$$\lim_{n\to\infty} (\mathcal{T}^n(x,0) - \mathcal{T}^n(x,y)) \ = \ 0\,,$$

uniformly in y. An important consequence of this observation and Theorem 4.2.1 is the following result.

Theorem 4.2.3 (Jager, 1986) *For almost all $x \in [0,1)$ the two-dimensional sequence $(T_n, V_n)_{n\geq 1}$ is distributed over $\overline{\Omega}$ according to the density function d, where*

$$d(x,y) = \frac{1}{\log 2} \frac{1}{(1+xy)^2}\,.$$

Proof. Denote by E the subset of numbers $x \in \Omega$ for which the sequence $(T_n, V_n)_{n\geq 0}$ is **not** distributed according to the density function d. Since the sequence $(\mathcal{T}^n(x,0) - \mathcal{T}^n(x,y))_{n\geq 0}$ is a null-sequence, it follows that for every pair $(x,y) \in \mathcal{E}$, where $\mathcal{E} := E \times [0,1]$, the sequence $\mathcal{T}^n(x,y)_{n\geq 0}$ is **not** distributed according to the density function d. Now if E had, as a one-dimensional set, positive Lebesgue measure, so would $\mathcal{E}$ as a two-dimensional set. But this would be in conflict with Theorem 4.2.1. □

Lenstra's conjecture now follows directly from this theorem by easy calculations.

Proof of Lenstra's conjecture. Let $A_z = \{(x,y) \in \overline{\Omega} : \dfrac{x}{1+xy} \leq z\}$. Then, $\overline{\mu}(A_z) = F(z)$, and $\Theta_j(x) \leq z \Leftrightarrow \mathcal{T}^j(x,0) \in A_z$. Furthermore,

$$\frac{1}{n}\#\{j\,;\, 1 \leq j \leq n,\, \Theta_j(x) \leq z\} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{A_z}(\mathcal{T}^j(x,0)).$$

Taking limits and using the above theorem we get the required result.

Chapter 5

Random β-expansions

In this chapter we study properties of expansions to non-integer base. Typically a point has uncountably many expansions, two such expansions were discussed earlier, namely the greedy and lazy expansions. We show how one can use one map defined on a product space that generates all expansions, and we give an overview of some known results. For further results, see [DK1, DdV].

5.1 Random β-expansions

Let $\beta > 1$ be a real number, and assume β is a non-integer. A β-expansion is an expression of the form

$$x = \sum_{n=1}^{\infty} \frac{a_n}{\beta^n},$$

where $a_i \in \{0, 1, \cdots, \lfloor\beta\rfloor\}$. Note that any x that posses such an expansion must lie in the interval $[0, \frac{\lfloor\beta\rfloor}{\beta-1}]$.

We have seen earlier two maps that generate such expansions, namely the greedy map and the lazy map, see examples 1.2.3, 1.2.4 3.0.3. In Figure 5.1 we compare the two maps for the same value of β.

In Figure 5.2 we superimposed the two pictures. We see that in this case the interval $[0, \frac{\lfloor\beta\rfloor}{\beta-1}]$ is divided into types of intervals. The equality regions, those were the two maps are equal, and the switch or regions, those were the two maps differ. To be more precise, To be more precise, we partition the interval $[0, \lfloor\beta\rfloor/(\beta-1)]$ into switch regions S_ℓ and *equality regions* E_ℓ, where

$$S_\ell = \left[\frac{\ell}{\beta}, \frac{\lfloor\beta\rfloor}{\beta(\beta-1)} + \frac{\ell-1}{\beta}\right], \quad \ell = 1, \ldots, \lfloor\beta\rfloor,$$

and

$$E_\ell = \left(\frac{\lfloor\beta\rfloor}{\beta(\beta-1)} + \frac{\ell-1}{\beta}, \frac{\ell+1}{\beta}\right), \quad \ell = 1, \ldots, \lfloor\beta\rfloor - 1,$$

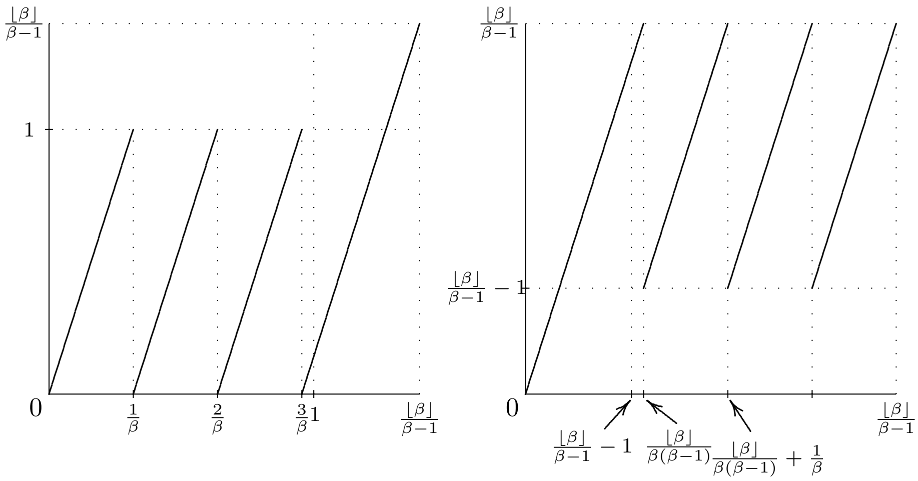

Figure 5.1: The greedy map T_β (left), and lazy map L_β (right). Here $\beta = \pi$.

$$E_0 = \left[0, \frac{1}{\beta}\right) \quad \text{and} \quad E_{\lfloor\beta\rfloor} = \left(\frac{\lfloor\beta\rfloor}{\beta(\beta-1)} + \frac{\lfloor\beta\rfloor - 1}{\beta}, \frac{\lfloor\beta\rfloor}{\beta-1}\right];$$

see Figure 5.2. On S_ℓ, the greedy map assigns the digit ℓ, while the lazy map assigns the digit $\ell-1$. Outside these switch regions both maps are identical, and hence they assign the same digits. We will now define a new random expansion in base β by randomizing the choice of the map used in the switch regions. For each switch region we assign a coin, and whenever x belongs to the ith switch region we flip the ith coin to decide which map will be applied to x, and hence which digit will be assigned.

Let

$$S = \sum_{\ell=1}^{\lfloor\beta\rfloor} S_\ell, \quad \text{and} \quad E = \sum_{\ell=0}^{\lfloor\beta\rfloor} E_\ell,$$

and consider $\Omega = \{0,1\}^{\mathbb{N}}$ with product σ-algebra. Let $\sigma : \Omega \to \Omega$ be the left shift, i.e., if $\omega = (\omega_1, \omega_2, \ldots) \in \Omega$, then $\sigma(\omega) = (\omega_2, \omega_3, \ldots)$. Define $K : \Omega \times [0, \lfloor\beta\rfloor/(\beta-1)] \to \Omega \times [0, \lfloor\beta\rfloor/(\beta-1)]$ by

$$K(\omega, x) = \begin{cases} (\omega, \beta x - \ell) & x \in E_\ell,\ \ell = 0, 1, \ldots, \lfloor\beta\rfloor, \\ (\sigma(\omega), \beta x - \ell) & x \in S_\ell \text{ and } \omega_1 = 1,\ \ell = 1, \ldots, \lfloor\beta\rfloor, \\ (\sigma(\omega), \beta x - \ell + 1) & x \in S_\ell \text{ and } \omega_1 = 0,\ \ell = 1, \ldots, \lfloor\beta\rfloor. \end{cases} \tag{5.1}$$

The elements of Ω represent the coin tosses ('heads'=1 and 'tails'=0) used every

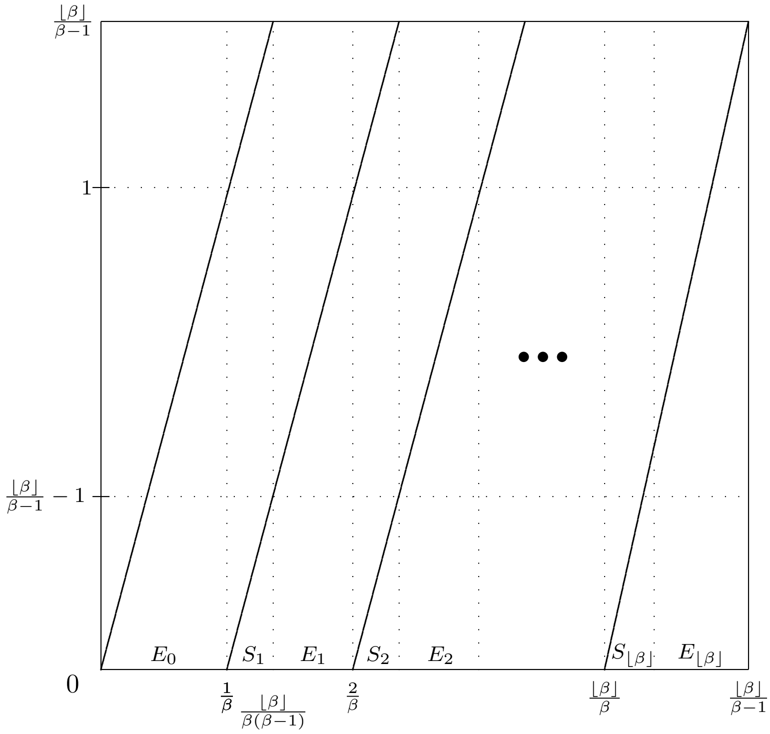

Figure 5.2: The greedy and lazy maps, and their switch regions.

time the orbit hits a switch region. Let

$$d_1 = d_1(\omega, x) = \begin{cases} \ell & \text{if } x \in E_\ell,\ \ell = 0, 1, \ldots, \lfloor\beta\rfloor, \\ & \text{or } (\omega, x) \in \{\omega_1 = 1\} \times S_\ell,\ \ell = 1, 2, \ldots, \lfloor\beta\rfloor, \\ \ell - 1 & \text{if } (\omega, x) \in \{\omega_1 = 0\} \times S_\ell,\ \ell = 1, 2, \ldots, \lfloor\beta\rfloor, \end{cases}$$

then

$$K(\omega, x) = \begin{cases} (\omega, \beta x - d_1) & \text{if } x \in E, \\ (\sigma(\omega), \beta x - d_1) & \text{if } x \in S. \end{cases}$$

Set $d_n = d_n(\omega, x) = d_1\left(K^{n-1}(\omega, x)\right)$, and let $\pi : \Omega \times [0, \lfloor\beta\rfloor/(\beta-1)] \to [0, \lfloor\beta\rfloor/(\beta-1)]$ be the canonical projection onto the second coordinate. Then

$$\pi\left(K^n(\omega, x)\right) = \beta^n x - \beta^{n-1} d_1 - \cdots - \beta d_{n-1} - d_n,$$

rewriting gives

$$x = \frac{d_1}{\beta} + \frac{d_2}{\beta^2} + \cdots + \frac{d_n}{\beta^n} + \frac{\pi\left(K^n(\omega, x)\right)}{\beta^n}.$$

Since $\pi\left(K^n(\omega,x)\right) \in [0, \lfloor\beta\rfloor/(\beta-1)]$, it follows that

$$\left| x - \sum_{i=1}^{n} \frac{d_i}{\beta^i} \right| = \frac{\pi\left(K^n(\omega,x)\right)}{\beta^n} \to 0 \qquad \text{as } n \to \infty.$$

This shows that for all $\omega \in \Omega$ and for all $x \in [0, \lfloor\beta\rfloor/(\beta-1)]$ one has that

$$x = \sum_{i=1}^{\infty} \frac{d_i}{\beta^i} = \sum_{i=1}^{\infty} \frac{d_i(\omega,x)}{\beta^i}.$$

5.2 Basic Properties of Random β-transformations

Let $<_{lex}$ and $\leq_{lex}$ denote the lexicographical ordering on both Ω, and $\{0,1,\ldots,\lfloor\beta\rfloor\}^{\mathbb{N}}$. For each $x \in [0, \frac{\lfloor\beta\rfloor}{\beta-1}]$, consider the set

$$D_x = \{(d_1(\omega,x), d_2(\omega,x), \ldots) : \omega \in \Omega\}.$$

We now show that the elements of D_x are ordered by the lexicographical ordering on Ω.

Theorem 5.2.1 *Suppose $\omega, \omega' \in \Omega$ are such that $\omega <_{lex} \omega'$, then*

$$(d_1(\omega,x), d_2(\omega,x), \ldots) \leq_{lex} (d_1(\omega',x), d_2(\omega',x), \ldots).$$

Proof. Let i be the first index where ω and ω' differ. Since $\omega <_{lex} \omega'$, then $\omega_i = 0$ and $\omega'_i = 1$. Notice that $\pi_2(K_\beta^j(\omega,x)) = \pi_2(K_\beta^j(\omega',x))$ for $j = 0, \ldots, t_i$, where $t_i \geq 0$ is the time of the i^{th} visit to the region $\Omega \times S$ of the orbit of (ω,x) under K_β. Then, $d_j(\omega,x) = d_j(\omega',x)$ for all $j \leq t_i$.

If $t_i = \infty$, then $d_j(\omega,x) = d_j(\omega',x)$ for all j.

If $t_i < \infty$, then $K_\beta^{t_i}(\omega,x) = K_\beta^{t_i}(\omega',x) \in \Omega \times S$. Since $\omega_i = 0$ and $\omega'_i = 1$, it follows that $d_{t_i+1}(\omega',x) = d_{t_i+1}(\omega,x) + 1$. Hence,

$$(d_1(\omega,x), d_2(\omega,x), \ldots) <_{lex} (d_1(\omega',x), d_2(\omega',x), \ldots).$$

□

The next theorem shows that for all $x \in [0, \lfloor\beta\rfloor/(\beta-1)]$, any representation of x of the form $x = \sum_{i=1}^{\infty} a_i/\beta^i$ with $a_i \in \{0,1,\ldots,\lfloor\beta\rfloor\}$ can be generated by means of the map K_β by choosing an appropriate $\omega \in \Omega$.

Theorem 5.2.2 *Let $x \in [0, \lfloor\beta\rfloor/(\beta-1)]$, and let $x = \sum_{i=1}^{\infty} a_i/\beta^i$ with $a_i \in \{0,1,\ldots,\lfloor\beta\rfloor\}$ be a representation of x in base β. Then there exists an $\omega \in \Omega$ such that $a_i = d_i(\omega,x)$.*

For the proof we need the following lemma.

Lemma 5.2.1 *For $x \in [0, \lfloor\beta\rfloor/(\beta-1)]$, one has*

(i) If $x \in E_j$ for some $j \in \{0, \dots, \lfloor\beta\rfloor\}$, then $a_1 = j$.

(ii) If $x \in S_j$ for some $j \in \{1, \dots, \lfloor\beta\rfloor\}$, then $a_1 \in \{j-1, j\}$.

Proof. The proof is by contradiction.

(i) Suppose $a_1 \neq j$. If $a_1 \leq j-1$, then $j \geq 1$ and $x = \sum_{i=1}^{\infty} \frac{a_i}{\beta^i} \leq \frac{j-1}{\beta} + \sum_{i=2}^{\infty} \frac{\lfloor\beta\rfloor}{\beta^i} = \frac{j-1}{\beta} + \frac{\lfloor\beta\rfloor}{\beta(\beta-1)}$. If $a_1 \geq j+1$, then $j \leq \lfloor\beta\rfloor - 1$ and $x \geq \frac{j+1}{\beta}$. In both cases $x \notin E_j$.

(ii) Suppose $a_1 \notin \{j-1, j\}$. If $a_1 \leq j-2$, then $j \geq 2$ and $x \leq \frac{j-2}{\beta} + \frac{\lfloor\beta\rfloor}{\beta(\beta-1)}$. If $a_1 \geq j+1$, then $j \leq \lfloor\beta\rfloor - 1$ and $x \geq \frac{j+1}{\beta}$. In both cases $x \notin S_j$. □

Proof 5.2.2. Define the numbers $\{x_n : n \in \mathbb{N}\}$ by $x_n = \sum_{i=1}^{\infty} \frac{a_{i+n-1}}{\beta^i}$. Notice that $x_1 = x$. Furthermore, we define a set $\{\ell_n(x) : n \in \mathbb{N}\}$ that keeps track of the number of times we flip a coin. More precisely,

$$\ell_n(x) = \sum_{i=1}^{n} \mathbf{1}_S(x_i).$$

We use induction on the number of digits already determined.

- If $x \in E_j$, then $\ell_1(x) = 0$ and by Lemma 1, $a_1 = j$. We set $\Omega_1 = \Omega$.
- If $x \in S_j$, then $\ell_1(x) = 1$ and by Lemma 1, $a_1 \in \{j-1, j\}$.
  - If $a_1 = j-1$, we set $\Omega_1 = \{\omega \in \Omega : \omega_1 = 0\}$.
  - If $a_1 = j$, we set $\Omega_1 = \{\omega \in \Omega : \omega_1 = 1\}$.

It follows that Ω_1 is a cylinder of length $\ell_1(x)$ and for all $\omega \in \Omega_1$, $d_1(\omega, x) = a_1$. By a cylinder of length 0 we mean of course the whole space Ω. Suppose we have obtained $\Omega_n \subseteq \cdots \subseteq \Omega_1$, so that Ω_n is a cylinder of length $\ell_n(x)$ and for all $\omega \in \Omega_n$, $d_1(\omega, x) = a_1, \dots, d_n(\omega, x) = a_n$. Notice that for all $\omega \in \Omega_n$, $x_{n+1} = \pi_2(K_\beta^n(\omega, x))$.

- If $x_{n+1} \in E_j$, then $\ell_{n+1}(x) = \ell_n(x)$ and for all $\omega \in \Omega_n$, $d_{n+1}(\omega, x) = d_1(K_\beta^n(\omega, x)) = j = a_{n+1}$, by Lemma 1. We set $\Omega_{n+1} = \Omega_n$.
- If $x_{n+1} \in S_j$, then $\ell_{n+1}(x) = \ell_n(x) + 1$ and $a_{n+1} \in \{j-1, j\}$ by Lemma 1.
  - If $a_{n+1} = j-1$, we set $\Omega_{n+1} = \{\omega \in \Omega_n : \omega_{\ell_{n+1}} = 0\}$. Then, for all $\omega \in \Omega_{n+1}$, $d_{n+1}(\omega, x) = d_1(K_\beta^n(\omega, x)) = j-1 = a_{n+1}$.
  - If $a_{n+1} = j$, we set $\Omega_{n+1} = \{\omega \in \Omega_n : \omega_{\ell_{n+1}} = 1\}$. Then, for all $\omega \in \Omega_{n+1}$, $d_{n+1}(\omega, x) = d_1(K_\beta^n(\omega, x)) = j = a_{n+1}$.

In all cases we see that Ω_{n+1} is a cylinder of length $\ell_{n+1}(x)$, and for all $\omega \in \Omega_{n+1}, d_1(\omega, x) = a_1, \ldots, d_{n+1}(\omega, x) = a_{n+1}$.

If the map K_β hits the switch regions infinitely many times, then $\ell_n(x) \to \infty$ and, as is well known, $\bigcap \Omega_n$ consists of a single point. If this happens only finitely many times, then the set $\{\ell_n(x) : n \in \mathbb{N}\}$ is finite and $\bigcap \Omega_n$ is exactly a cylinder set. In both cases $\bigcap \Omega_n$ is non-empty and $\omega \in \bigcap \Omega_n$ satisfies $d_j(\omega, x) = a_j, j \geq 1$. $\square$

5.3 An ergodic measure for random β-expansions

In this section we show that the map K_β on $\Omega \times [0, \frac{\lfloor\beta\rfloor}{\beta-1}]$ can be essentially identified with the left shift on $\{0, \ldots, \lfloor\beta\rfloor\}^{\mathbb{N}}$. This will enable us to define an invariant ergodic measure K_β.

Let $D = \{0, \ldots, \lfloor\beta\rfloor\}^{\mathbb{N}}$ be equipped with the product σ-algebra $\mathcal{D}$, and the uniform product measure $\mathbb{P}$. Let σ' be the left shift on D. On the set $\Omega \times [0, \lfloor\beta\rfloor/(\beta-1)]$ we consider the product σ-algebra $\mathcal{A} \times \mathcal{B}$, where $\mathcal{B}$ is the Borel σ-algebra on $[0, \lfloor\beta\rfloor/(\beta-1)]$, and $\mathcal{A}$ the product σ-algebra on Ω. Define the function $\varphi : \Omega \times [0, \frac{\lfloor\beta\rfloor}{\beta-1}] \to D$ by

$$\varphi(\omega, x) = (d_1(\omega, x), d_2(\omega, x), \ldots).$$

It is easily seen that φ is measurable, and $\varphi \circ K_\beta = \sigma' \circ \varphi$. Furthermore, Theorem 5.2.2 implies that φ is surjective. We will now show that φ restricted to an appropriate K_β-invariant subset is in fact invertible.
Let

$$Z = \{(\omega, x) \in \Omega \times [0, \frac{\lfloor\beta\rfloor}{\beta-1}] : K_\beta^n(\omega, x) \in \Omega \times S \text{ infinitely often}\},$$

and

$$D' = \{(a_1, a_2, \ldots) \in D : \sum_{i=1}^{\infty} \frac{a_{j+i-1}}{\beta^i} \in S \text{ for infinitely many } j'\text{s}\}.$$

Then, $\varphi(Z) = D'$, $K_\beta^{-1}(Z) = Z$ and $(\sigma')^{-1}(D') = D'$. Let φ' be the restriction of the map φ to Z.

Lemma 5.3.1 *The map $\varphi' : Z \to D'$ is a bimeasurable bijection.*

Proof. For any sequence $(a_1, a_2, \ldots) \in D'$, define recursively

$$\begin{aligned} r_1 &= \min\{j \geq 1 : \sum_{l=1}^{\infty} \frac{a_{j+l-1}}{\beta^l} \in S\}, \\ r_i &= \min\{j > r_{i-1} : \sum_{l=1}^{\infty} \frac{a_{j+l-1}}{\beta^l} \in S\}. \end{aligned}$$

If $\sum_{l=1}^{\infty} \frac{a_{r_i+l-1}}{\beta^l} \in S_j$ then, according to Lemma 1, $a_{r_i} \in \{j-1, j\}$. If $a_{r_i} = j-1$, let $\omega_i = 0$, otherwise let $\omega_i = 1$. Define $(\varphi')^{-1} : D' \to Z$ by

$$(\varphi')^{-1}((a_1, a_2, \ldots)) = (\omega, \sum_{i=1}^{\infty} \frac{a_i}{\beta^i}).$$

It is easily checked that $(\varphi')^{-1}$ is measurable, and is the inverse of φ'. □

Lemma 5.3.2 $\mathbb{P}(D') = 1$.

Proof. For any sequence $(a_1, a_2, \ldots) \in D$, define for $m \geq 1$,

$$x_m = \frac{1}{\beta} + \frac{a_1}{\beta^{m+1}} + \frac{a_2}{\beta^{m+2}} + \ldots.$$

Clearly $x_m \geq 1/\beta$. On the other hand,

$$x_m \leq \frac{1}{\beta} + \sum_{i=1}^{\infty} \frac{\lfloor\beta\rfloor}{\beta^{m+i}} = \frac{1}{\beta}(1 + \frac{\lfloor\beta\rfloor}{\beta^{m-1}(\beta-1)}).$$

Since $1 + \frac{\lfloor\beta\rfloor}{\beta^{m-1}(\beta-1)} \downarrow 1$ as $m \to \infty$, there exists an integer $N > 0$ such that for all $m \geq N$,

$$\frac{1}{\beta} \leq x_m \leq \frac{\lfloor\beta\rfloor}{\beta(\beta-1)},$$

i.e. $x_m \in S_1$ for all $m \geq N$. Let

$$D'' = \{(a_1, a_2, \ldots) \in D : a_j a_{j+1} \ldots a_{j+N-1} = 1 \underbrace{00\ldots0}_{(N-1)-times} \text{ for infinitely many } j\}.$$

From the above, we conclude that $D'' \subseteq D'$. Clearly $\mathbb{P}(D'') = 1$, hence $\mathbb{P}(D') = 1$. □

Now, consider the K_β-invariant measure ν_β defined on $\mathcal{A} \times \mathcal{B}$ by $\nu_\beta(A) = \mathbb{P}\left(\varphi(Z \cap A)\right)$. The following theorem is a simple consequence of Lemmas 5.3.1 and 5.3.2.

Theorem 5.3.1 *Let $\beta > 1$ be a non-integer. Then, the dynamical systems $(\Omega \times [0, \frac{\lfloor\beta\rfloor}{\beta-1}], \mathcal{A} \times \mathcal{B}, \nu_\beta, K_\beta)$ and $(D, \mathcal{D}, \mathbb{P}, \sigma')$ are measurably isomorphic.*

Corollary 5.3.1 *Under the measure ν_β the digits $(d_i)_{i \geq 1}$ form a uniform Bernoulli process.*

Bibliography

[Abr] Abramov, L.M. – *Entropy of induced automorphisms*, Dokl. Akad. Nauk SSSR **128** (1959), 647-650.

[Ada] Adams, W.W. – *On a relationship between the convergents of the nearest integer and regular continued fractions*, Math. Comp. **33** (1979), 1321-1331.

[BMcL] Bagemihl, F. and J.R. McLaughlin – *Generalization of some classical theorems concerning triples of consecutive convergents to simple continued fractions*, J. reine angew. Math. **221**, (1966), 146-149.

[Bar] Barbolosi, D. – *Sur le développement en fractions continues a quotients partiels impairs*, Monatsh. Math. **109**, (1990), 25-37.

[Bil] Billingsley, P. – *Ergodic Theory and Information*, John Wiley and Sons, 1965.

[Bor] Borel, É. – *Contribution a l'analyse arithmétiques du continu*, J. Math. Pures et Appl. (5) **9** (1903), 329-375.

[BJW] Bosma, W., H. Jager and F. Wiedijk – *Some metrical observations on the approximation by continued fractions*, ibid. **45** (1983), 281-270.

[BK] Bosma, W. and C. Kraaikamp – *Metrical theory for optimal continued fractions*, J. Number Theory **34** (1990), 251-270.

[BP] de Bruijn, N.G. and K.A. Post – *A remark on uniformly distributed sequences and Riemann integrability*, Indag. Math. **30**, (1968), 149-150.

[CFS] Cornfeld, I. P., S. V. Fomin and Ya. G. Sinai – *Ergodic Theory*, Springer-Verlag, 1981.

[DK] Dajani, K. and C. Kraaikamp – *Ergodic theory of numbers*, Carus Mathematical Monographs, 29. Mathematical Association of America, Washington, DC, 2002.

[DK1] Dajani, K., Kraaikamp, C. – *Random β-expansions*, Ergodic Theory Dynam. Systems. **23** (2003), no.2, 461-479.

[DdV] Dajani, K., de Vries, M. – *Measures of maximal entropy for random β-expansions*, J. Eur. Math. Soc. (JEMS) **7** (2005), no.1, 51-68.

[DKS] Dajani, K.; Kraaikamp, C.; Solomyak, B. – *The natural extension of the β-transformation*, Acta Math. Hungar. **73** (1996), no. 1-2, 97–109.

[Doe] Doeblin, W. – *Remarques sur la théorie métrique des fractions continues*, Comp. Math. **7**, (1940), 353-371.

[G] Gel'fond, A.O. – *A common property of number systems*, Izv. Akad. Nauk SSSR. Ser. Mat. **23** (1959) 809–814.

[H] Halmos, P.R. – *Measure Theory*, D. Van Nostrand Company, Inc., New York, N.Y., 1950.

[Hur] Hurwitz, A. – *Über eine besondere Art der Kettenbruchentwicklung reeller Grössen*, Acta Math. **12**, (1889), 367-405.

[Jag1] Jager, H. – *On the speed of convergence of the nearest integer continued fraction*, Math. Comp. **39**, (1982), 555-558.

[Jag2] Jager, H. – *The distribution of certain sequences connected with the continued fraction*, Indag. Math. **48**, (1986), 61-69.

[JK] Jager, H. and C. Kraaikamp – *On the approximation by continued fractions,* Indag. Math., **51** (1989), 289-307.

[Knu] Knuth, D.E. – *The distribution of continued fraction approximations*, J. Number Theory **19**, (1984), 443-448.

[Kr1] Kraaikamp, C. – *The distribution of some sequences connected with the nearest integer continued fraction*, Indag. Math. **49**, (1987), 177-191.

[Kr2] Kraaikamp, C. – *Statistic and ergodic properties of Minkowski's diagonal continued fraction*, Theoretical Comp. Sc. **65**, (1989), 197-212.

[Kr3] Kraaikamp, C. – *A new class of continued fractions*, Acta Arithmetica **LVII**, (1991), 1-39.

[KN] Kuipers, L. and H. Niederreiter – *Uniform Distribution of Sequences*, John Wiley and Sons, 1974.

[KK] Kamae, T., and Keane, M.S. – *A simple proof of the ratio ergodic theorem*, Osaka J. Math. **34** (1997), no. 3, 653–657.

[L] Lüroth, J. – *Ueber eine eindeutige Entwickelung von Zahlen in eine unendliche Reihe*, Math. Annalen **21** (1883), 411-423.

[Le] Lévy, P. – *Sur le loi de probabilité dont dependent les quotients complets wet incompletes d'une fraction continue*, Bull. Soc. Math. de France **57**, (1929), 178-194.

[Min] Mikowski, H. – *Über die Annäherung an eine reelle Grösse durch rationale Zahlen*, Math. Ann. **54**, (1901), 91-124.

[Nak] Nakada, H. – *Metrical theory for a class of continued fraction transformations and their natural extensions*, Tokyo J. Math. **4**, (1981), 399-426.

[Obr] Obrechkoff, N. – *Sur l'approximation des nombres irrationels par des nombres rationels*, C. R. Acad. Bulgare Sci. **3** (1), (1951), 1-4.

[P] Parry, W. – *On the β-expansion of real numbers,* Acta Math. Acad. Sci. Hungary **11** (1960), 401-416.

[Per] Perron, O. – *Die Lehre von den Kettenbrüchen*, Band I, 3. verb. u. etw. Aufl., B. G. Teubner, Stuttgart.

[Pet] Petersen, K. – *Ergodic Theory,* Cambridge Studies in Advanced Mathematics, Vol. 2, Cambridge Univ. Press, Cambridge, 1983.

[Rie1] Rieger, G. J. – *Ein Gauss-Kusmin-Lévy Satz für Kettenbrüche nach nächsten Ganzen*, Manuscripta Math. **24**, (1978), 437-448.

[Rie2] Rieger, G. J. – *Über die Länge von Kettenbrüchen mit ungeraden Teilnennern*, Abh. Braunschweig Wiss. Ges. **32**, (1981), 61-69.
– *On the metrical theory of the continued fractions with odd partial quotients*, in: *Topics in Classical Number Theory*, vol. II, Budapest 1981, 1371-1481 (Colloq. Math. Soc. János Bolyai, **34**, North-Holland 1984).

[Ro] Rohlin, V.A. – *Exact endomorphisms of a Lebesgue space,* Izv. Akad. Naik SSSR, Ser. Mat., **24** (1960); English AMS translation, Series **2**, **39** (1969), 1-36.

[RN] Ryll-Nardzewski, C. – *On the ergodic theorems (II) (Ergodic theory of continued fractions)*, Studia Math. **12**, (1951), 74-79.

[S1] Schweiger, F. – *Continued fractions with odd and even partial quotients*, Arbeitsbericht Math. Inst. Univ. Salzburg **4**, (1982), 59-70.
– *On the approximation by continued fractions with odd and even partial quotients*, Arbeitsbericht Math. Inst. Univ. Salzburg **1-2**, (1984), 105-114.

[S2] Schweiger, F. – *Ergodic Theory of Fibered Systems and Metric Number Theory*, Clarendon Press, Oxford 1995.

[Sen] Sendov, B. – *Der Vahlensche Satz über die singulären Kettenbrüche und die Kettenbrüche nach nächsten Ganzen*, Ann. Univ. Sofia, Fac. Sci. Math. Livre 1 Math. **54**, (1959/60), 251-258.

[Tie] Tietze, H. – *Über die raschesten Kettenbruchentwicklung reeller Zahlen*, Monatsh. Math. Phys. **24**, (1913), 209-241.

[V] Vahlen, K. Th. – *Über die raschesten Kettenbruchentwicklungen reeller Zahlen*, Monatsh. Math. Phys. **24**, (1913), 221-233.

[T] Tong, J. – *The conjugate property of the Borel Theorem on Diophantine Approximation*, Math. Z. **184**, (1983), 151-153.

[W] Walters, P. – *An introduction to ergodic theorey*, Springer Verlag, Graduate Textx in Mathematics **79**, 1982.