Universiteit Utrecht

Bachelor Thesis

Ergodic Theory and Number Expansions

Author:
Lucas van Kreveld
4132777

Supervisor:
dr. K. Dajani

June 13, 2016

Contents

1 Introduction

Given a specific value of a real number, we need to be able to write it down in order to work with it. So we wish to come up with some system that attaches a unique symbolization to each real number. A part of that symbolization is formed by integers, a discrete line of values. The other part is the characterization of numbers between integers. It is the last part we will be focussing on in this thesis.
Although decimal digits are most often used for this characterization, it should not be surprising that alternatives exist. We can ask ourselves what would be a general way of defining a characterization of numbers, or how we can determine which characterization is best. The second question is hard to answer, and we will not try to cover it here. However, you can expect an extensive answer to the first one if you continue reading.

We start the thesis with a brief description of the most important measure theory. In section 3 we slowly build up the general definition of the characterization, also called an expansion, of numbers between integers, introducing four expansion types along the way. After that, we use we use ergodic theory in section 4 to prove results about digit frequency. Finally, section 5 introduces the concept of entropy and ends with a comparison theorem.

The majority of the theory written in sections 3, 4 and 5 follows the structure of [4]. However, we provide extra intuitive explanation and examples to support the theory when necessary.

2 Measure theory

This section will be devoted to a short explanation of basis measure theory and integration, as it will be needed in future sections. We will follow the approach of [8].
In particular, we will discuss the concepts of σ-algebra, measure, null sets, measurable maps and integrals. If you are already familiar with these concepts, feel free to skip this section.

2.1 Measures

Definition 2.1.1. Let X be a set. A collection $\mathcal{A}$ of subsets of X is called a *σ-algebra* if the following conditions are satisfied:
i) $X \in \mathcal{A}$.
ii) if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$.
iii) if $(A_i)_{i\in\mathbb{N}}$ is a countable selection of subsets of X with $A_i \in \mathcal{A}$ for each i, then $\bigcup\limits_{i\in\mathbb{N}} A_i \in \mathcal{A}$.
A subset C of X is called *measurable* when $C \in \mathcal{A}$. The pair $(X, \mathcal{A})$ is called a *measurable space*.

Example 2.1. It is not hard to check that $\{\emptyset, A, A^c, X\}$ is a σ-algebra for any $A \subseteq X$.

Example 2.2. Consider the set $X = [0, 1)$. It would be nice if all open subintervals of $[0, 1)$ are members of the corresponding σ-algebra. Therefore we define the *Borel σ-algebra* on $[0, 1)$ (notation: $\mathcal{B}([0, 1))$) as the smallest σ-algebra containing at least all open subintervals of $[0, 1)$. We say the open intervals form a *generator* of the Borel σ-algebra. Throughout this thesis, $\mathcal{B}([0, 1))$ is the only σ-algebra that will be used.

We would like to give a value to all measurable sets. So we define measure:

Definition 2.1.2. Let X be a set and $\mathcal{A}$ be a σ-algebra on X. A *measure* is a map $\mu : \mathcal{A} \to [0, \infty)$, satisfying:
i) $\mu(\emptyset) = 0$.
ii) if $(A_i)_{i\in\mathbb{N}} \subseteq \mathcal{A}$ are pairwise disjoint, then $\mu(\bigcup\limits_{i\in\mathbb{N}} A_i) = \sum\limits_{i\in\mathbb{N}} \mu(A_i)$.
The triple $(X, \mathcal{A}, \mu)$ we now call a *measure space*. In case $\mu(X) = 1$ we say that μ is a *probability measure* and $(X, \mathcal{A}, \mu)$ a *probability space*.

Example 2.3. When $X \subseteq \mathbb{R}$, a commonly used measure is the so-called Lebesque measure λ. For intervals $(a, b) \in \mathcal{A}$, it gives $\lambda((a, b)) = b - a$. The Lebesque measure of non-intervals then follow from the previous definition.

Of course it occurs sometimes that we wish to state a property of all points of X. However, in measure theory we often face the problem that there exist exceptional points where the property does not hold. In order to be able to say something anyway, we have the following definition.

Definition 2.1.3. Let $(X, \mathcal{A}, \mu)$ be a measure space, P some property and $N \subseteq X$ be the set of points for which property P does not hold. Then we say that P holds *almost everywhere* (or a.e) with respect to μ if $\mu(N) = 0$. In that case we call N a *null set*.

In the course of this thesis, we will consider maps T between measure spaces. We need another definition, regarding these maps.

Definition 2.1.4. Let $(X, \mathcal{A})$ be a measurable space. A map $T : X \to X$ is *measurable* if $T^{-1}(A) \in \mathcal{A}$ for all $A \in \mathcal{A}$. If X is a subset of $\mathbb{R}$, T is also called a *measurable function*.

It is often necessary to show that some map T is measurable. This definition suggests we should check the pre-image of all elements of the σ-algebra (and there can be a lot), but there is a way to make life easier, formulated in the lemma below. It will prevent us from having to check all measurable sets.

Lemma 2.1.5. *Let $(X, \mathcal{A})$ be a measurable space and let $\mathcal{G}$ be a generator of $\mathcal{A}$. Then $T : X \to X$ is measurable if and only if $T^{-1}(G) \in \mathcal{A}$ for all $G \in \mathcal{G}$.*

Proof. We have to show that $T^{-1}(A) \in \mathcal{A}$ for all $A \in \mathcal{A}$ if and only if $T^{-1}(G) \in \mathcal{A}$ for all $G \in \mathcal{G}$. Because $\mathcal{G} \subseteq \mathcal{A}$ it is clear that the first implies the second.
Now let $T^{-1}(G) \in \mathcal{A}$ for all $G \in \mathcal{G}$, and let $\mathcal{C} = \{B \in X : T^{-1}(B) \in \mathcal{A}\}$. We show that $\mathcal{C}$ is a σ-algebra. $T^{-1}(X) = X \in \mathcal{A}$, so $X \in \mathcal{C}$. If $B \in \mathcal{C}$, then $T^{-1}(B^c) = (T^{-1}(B))^c \in \mathcal{A}$, so $B^c \in \mathcal{C}$. Finally, if $(B_i)_{i\in\mathbb{N}} \subseteq \mathcal{C}$ are pairwise disjoint, $T^{-1}(\bigcup\limits_{i\in\mathbb{N}} B_i) = \bigcup\limits_{i\in\mathbb{N}} T^{-1}(B_i) \in \mathcal{A}$, so $\bigcup\limits_{i\in\mathbb{N}} B_i \in \mathcal{C}$, and $\mathcal{C}$ is a σ-algebra.
If $\sigma(\mathcal{D})$ denotes the σ-algebra generated by $\mathcal{D}$, since $\mathcal{G} \subseteq \mathcal{C}$ we have $\sigma(\mathcal{G}) \subseteq \sigma(\mathcal{C})$. Therefore $\mathcal{A} = \sigma(\mathcal{G}) \subseteq \sigma(\mathcal{C}) = \mathcal{C}$, so $\mathcal{A} \subseteq \mathcal{C}$. It follows from the definition of $\mathcal{C}$ that $T^{-1}(A) \in \mathcal{A}$ for all $A \in \mathcal{A}$, and we have completed the proof. $\square$

Apparently it suffices to check the elements of a generator of the σ-algebra to show that a map T is measurable. So with $X = [0, 1)$ and $\mathcal{A} = \mathcal{B}([0, 1))$ (the Borel σ-algebra restricted on [0,1)) we will only have to show that $T^{-1}((a, b)) \in \mathcal{B}([0, 1))$ for all intervals (a, b).

2.2 Integrals

Measure theory can be used to define a new way of integrating a function, and we will use this way of integrating in the following sections. Below we will give a brief description of the corresponding definition.

Recall that the indicator function $\mathbf{1}_A(x)$ equals 1 if $x \in A$ and 0 otherwise. We first introduce a class of functions called simple functions.

Definition 2.2.1. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $A_1, ..., A_n$ be measurable and pairwise disjoint subsets of X. A *simple function* is a function of the form $\sum\limits_{i=1}^{n} a_i \mathbf{1}_{A_i}(x)$, where $a_1, .., a_n$ are positive constants in $\mathbb{R}$.

So simple functions are constant on each A_i. A nice property of them is that each measurable function can be written as a limit of simple functions. For a proof of this property we refer to [8], theorem 8.8. The idea is now to define integrals for simple functions, and use the property to generalize the definition of integrals to arbitrary measurable functions.

Definition 2.2.2. Let $(X, \mathcal{A}, \mu)$ be a measure space and let f be a measurable function. If f is a simple function $\sum\limits_{i=1}^{n} a_i \mathbf{1}_{A_i}(x)$, then $\int\limits_X f(x)d\mu = \sum\limits_{i=1}^{n} a_i \mu(A_i)$. If f is not simple, we define

$$\int\limits_X f(x)d\mu = \sup\{\int\limits_X g(x)d\mu : g(x) \text{ simple and } g(x) \leq f(x) \text{ for all } x \in X\}.$$

We say that f is *integrable* with respect to μ if $\int\limits_X f(x)d\mu$ is finite.

We have now discussed everything we need from measure theory to analyze number expansions.

3 Various number expansions

We will now see some well and less known number expansions, introducing some notation along the way. In order to work with the notation, we will define a map which is different for each number expansion. In the remainder of this section, those maps will be analyzed as good as possible.
Since we are not interested in the integer part of a number, we can restrict our study to the interval $[0, 1)$.

Notation. An expansion of a number in the interval $[0, 1)$ can be represented by a finite or countable sequence $(a_1, a_2, ...)$, where each a_i is a non-negative integer. The a_i are called the *elements* or *digits* of a certain expansion.

Examples of such expansions will be given below.

3.1 n-ary expansions

The most common number expansion for day-to-day use is the 10-ary expansion, known as the decimal expansion. We write $0.a_1a_2a_3...$ for a number in $[0, 1)$, with $a_i \in \{0, 1, ..., 9\}$ for each $i \in \mathbb{N}$. In our general notation, the decimal expansion of this number is represented by the sequence $(a_1, a_2, a_3, ...)$. For example, it should be clear that $\frac{1}{3} = 0.333...$ gives $(3, 3, 3, ...)$ as representation of its decimal expansion.
More generally, an n-ary expansion also looks like $0.a_1a_2a_3...$, but now the possible elements are $0, 1, ..., n-1$. Consider the 2-ary or binary expansion. Here, $a_1 = 0$ if the number lies in $[0, \frac{1}{2})$, and $a_1 = 1$ if it lies in $[\frac{1}{2}, 1)$. So the first element of the binary expansion of $\frac{1}{3}$ is 0. For the second element, we look at the remaining interval $[0, \frac{1}{2})$. Note that $\frac{1}{3}$ now lies in the second half of the interval, and we have $a_2 = 1$. Continuing like this, we obtain $\frac{1}{3} = 0.01010101...$ (represented by $(0, 1, 0, 1, 0, 1, 0, 1, ...)$) in the binary expansion.
The only reason why the binary expansion of $\frac{1}{3}$ was more difficult to find than the decimal expansion, is that most people are less used to the binary expansion. So given a real number x somewhere in the interval $[0, 1)$ and a natural number n, how do we find the n-ary expansion of x? The answer is similar to how we obtained the binary expansion of $\frac{1}{3}$.
Consider the partition $\{[\frac{i}{n}, \frac{i+1}{n}) : i \in \{0, ..., n-1\}\}$ of $[0, 1)$, getting n intervals of equal length. Note that the first element of the n-ary expansion of x is i if (and only if) $x \in [\frac{i}{n}, \frac{i+1}{n})$. For the second element, divide $[\frac{i}{n}, \frac{i+1}{n})$ again into n pieces of equal length: $\{[\frac{i}{n} + \frac{j}{n^2}, \frac{i}{n} + \frac{j+1}{n^2}) : j \in \{0, ..., n-1\}\}$. Now if $x \in [\frac{i}{n} + \frac{j}{n^2}, \frac{i}{n} + \frac{j+1}{n^2})$, then $a_1 = i$ and $a_2 = j$. Of course, we can go on with this process to obtain $a_3, a_4, ...$

It is time to introduce the concept of the expansion map. This map T sends each a_i to a_{i+1} in a given expansion. Although it might not seem useful now, the map will come in handy later on.

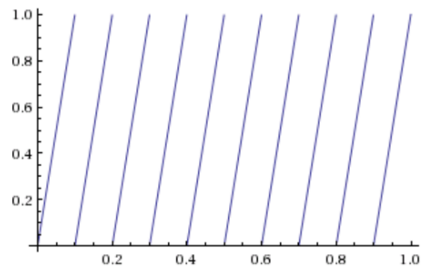
Definition 3.1.1. Let $x \in [0, 1)$ and let $(a_1, a_2, a_3, ...)$ be the expansion of x of a certain type. Then the *expansion map* is the function $T : [0, 1) \to [0, 1)$ such that if $(a_1, a_2, a_3, ...)$ is the representation of x, then $(a_2, a_3, ...)$ is the representation of $T(x)$.

Let us state some examples for the case where T represents the decimal map. We have $T(0.37841) = 0.7841$, $T(\frac{1}{3}) = T(0.3333...) = 0.333... = \frac{1}{3}$, but $T(\frac{1}{2}) = T(0.5000...) = 0.000... = 0$. It might not be entirely clear how the decimal map looks like. However, the structure of the map becomes quite clear in Figure 3.1. In fact, it turns out that the decimal map is just the function $T(x) = 10x \mod 1$. It is not hard to see that the general n-ary expansion map T_n is given by

$$T_n(x) = nx \mod 1.$$

Another concept that will be important later on is that of invariant measure.

Figure 3.1: The decimal expansion map T_{10}



Definition 3.1.2. Let $T : [0,1) \to [0,1)$ be a measurable map, and let μ be a measure on $[0,1)$. Then we say μ is invariant, or equivalently, T is measure preserving, if $\mu(T^{-1}(A)) = \mu(A)$ for every $A \in \mathcal{B}([0,1))$.

This means that T should not influence the measure of any measurable set. Unfortunately, as we have seen before, the number of measurable sets is big. We will therefor show that checking just a generator of $\mathcal{B}([0,1))$ is enough.

Lemma 3.1.3. *Suppose $T : [0,1) \to [0,1)$ is a surjective measurable map, and μ is a measure on $[0,1)$. Let $\mathcal{G} \subseteq \mathcal{B}([0,1))$ be the collection of all open intervals of $[0,1)$. Then μ is invariant if and only if $\mu(T^{-1}(G)) = \mu(G)$ for every $G \in \mathcal{G}$.*

Proof. Suppose $B \in \mathcal{B}([0,1))$. Our aim is to show that μ stays invariant when taking complements and countable unions of μ-invariant sets.
Let $G \in \mathcal{G}$, so $\mu(T^{-1}(G)) = \mu(G)$. Then, since T is surjective, $\mu(T^{-1}(G^c)) = \mu(T^{-1}(G)^c) = \mu([0,1)) - \mu(T^{-1}(G)) = \mu([0,1)) - \mu(G) = \mu(G^c)$. Moreover, note that any countable union of open intervals can be written as a countable union of disjoint open intervals, by simply merging the overlapping intervals into one.
Now, if $\{G_i\}_{i\in\mathbb{N}} \subseteq \mathcal{G}$ we write $\bigcup_{i=1}^{\infty} G_i = \bigcup_{i=1}^{\infty} G_i'$, where all G_i' are disjoint, and we have $\mu(T^{-1}(\bigcup_{i=1}^{\infty} G_i)) = \mu(T^{-1}(\bigcup_{i=1}^{\infty} G_i')) = \mu(\bigcup_{i=1}^{\infty} T^{-1}(G_i')) = \sum_{i=1}^{\infty} \mu(T^{-1}(G_i')) = \sum_{i=1}^{\infty} \mu(G_i') = \mu(\bigcup_{i=1}^{\infty} G_i') = \mu(\bigcup_{i=1}^{\infty} G_i)$.
So μ stays invariant under complements and countable unions. Because $\mathcal{G}$ is a generator of $\mathcal{B}([0,1))$, B can be written as complements and countable unions of sets in $\mathcal{G}$. We conclude that $\mu(T^{-1}(B)) = \mu(B)$ for every $B \in \mathcal{B}([0,1))$, which proves the result. $\square$

Remark. In the proceeding lemma, we chose $\mathcal{G}$ to be the set of all open intervals of $[0,1)$. However, the proof remains the same if we define $\mathcal{G}$ to be the set of all left or right half open intervals.

Just like the measurability of a map, we only have to check the elements of a (in this case particular) generator. This will be very helpful, for example in the proof of the following theorem.

Theorem 3.1.4. *The Lebesgue measure λ is invariant under the n-ary expansion map T_n.*

Proof. We have to show that T_n is measurable and, thanks to Lemma 3.1.3, only that $\lambda(T_n^{-1}((a,b))) = \lambda((a,b))$ for all $(a,b) \subseteq [0,1)$.
Note that $T_n^{-1}((a,b)) = (\frac{a}{n}, \frac{b}{n}) \cup (\frac{a+1}{n}, \frac{b+1}{n}) \cup ... \cup (\frac{a+n-1}{n}, \frac{b+n-1}{n})$, and it follows that $T_n^{-1}((a,b)) \in \mathcal{B}([0,1))$ because it is a union of intervals. We conclude that T^n is measurable.
Since the union of intervals is disjoint, we find $\lambda(T_n^{-1}((a,b))) = \lambda((\frac{a}{n}, \frac{b}{n})) + \lambda((\frac{a+1}{n}, \frac{b+1}{n})) + ... + \lambda((\frac{a+n-1}{n}, \frac{b+n-1}{n})) = n \cdot \frac{b-a}{n} = b - a = \lambda((a,b))$, which completes the proof. $\square$

3.2 GLS-expansions

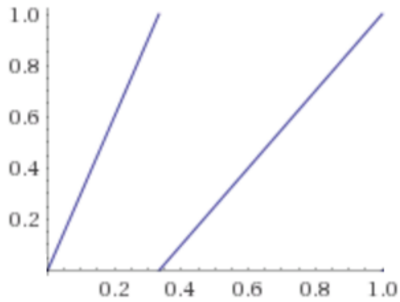
With the n-ary expansion we divided [0,1) into n intervals of equal length. Then we divided of the n new intervals again, into n^2 intervals of length $\frac{1}{n^2}$. We continued the proceeding, possibly infinitely many times, to obtain the expansion of a given number. What would happen when the subintervals do not necessarily have the same length, or when we divide into infinitely many subintervals? The resulting number expansion is a *Generalized Lüroth Series (GLS) expansion.* Below we will define this expansion step by step.
Consider a partition of $[0,1)$ of the form $\{[l_i, r_i)\}_{i \in \mathbb{N} \cup \{0\}}$, in which $l_0 = 0$ and $r_i = l_{i+1}$ for all $i \in \mathbb{N} \cup \{0\}$. This partition can be either finite or countable. In the finite case with m subintervals, we have $r_{m-1} = 1$.

Example 3.1. Such a partition could be $\{[0, \frac{1}{3}), [\frac{1}{3}, \frac{1}{2}), [\frac{1}{2}, 1)\}$. Also take a look at the infinite partition $\{[\frac{n}{n+1}, \frac{n+1}{n+2})\}_{n \in \mathbb{N} \cup \{0\}}$.

Now if x lies in the interval $[l_i, r_i)$, we define the first element of the expansion of x to be i. To find the second element, we divide $[l_i, r_i)$ in exactly the same way we divided $[0,1)$. So the j-th subinterval of $[l_i, r_i)$ will be $[l_i + l_j \cdot (r_i - l_i),\, l_i + r_j \cdot (r_i - l_i))$.
Let us examine the the expansion map $T_G(x)$ for the GLS-expansion. Without loss of generality, assume that $a_1 = i$ for the number x. So we must have $x \in [l_i, r_i)$. If $x = \frac{l_i + r_i}{2}$ is the middle of the interval, we would like a_2 to be the j such that $\frac{1}{2} \in [l_j, r_j)$. In particular, we would like $T_G(x)$ to be $\frac{1}{2}$ because T_G should send a_1 to a_2. As a consequence, T_G must grow linearly from 0 to 1 on each $[l_i, r_i)$. So we find

$$T_G(x) = \frac{1}{r_i - r_j} x - \frac{l_i}{r_i - l_i}, \qquad x \in [l_i, r_i).$$

Figure 3.2: The GLS expansion map T_G for the partition $\{[0, \frac{1}{3}), [\frac{1}{3}, 1)\}$.



See Figure 3.2 for an example of a GLS map.

We conclude the subsection with the invariant measure of the GLS expansion.

Theorem 3.2.1. *The Lebesgue measure λ is invariant under the GLS expansion map T_G.*

Proof. The proof is similar to the one of Theorem 3.1.4. Note that $T_G^{-1}((a,b)) = (l_0 + \frac{a}{r_0 - l_0}, l_0 + \frac{b}{r_0 - l_0}) \cup (l_1 + \frac{a}{r_1 - l_1}, l_1 + \frac{b}{r_1 - l_1}) \cup ... = \bigcup_{i \in \mathbb{N} \cup \{0\}} (l_i + \frac{a}{r_i - l_i}, l_i + \frac{b}{r_i - l_i})$ is a countable union of intervals, so $T_G^{-1}((a,b)) \in \mathcal{B}([0,1))$ and T_G is measurable. Furthermore, it follows from the fact that the intervals from the union are disjoint that $\lambda(T_G^{-1}((a,b))) = \sum_{i \in \mathbb{N} \cup \{0\}} \lambda((l_i + \frac{a}{r_i - l_i}, l_i + \frac{b}{r_i - l_i})) = \sum_{i \in \mathbb{N} \cup \{0\}} \frac{b-a}{r_i - l_i} = (b-a) \cdot \sum_{i \in \mathbb{N} \cup \{0\}} \frac{1}{r_i - l_i} = b - a = \lambda((a,b))$. □

3.3 β-expansions

In the n-ary expansion of Section 3.1 we made the assumption n was a natural number. Although this assumption seems reasonable and nice to work with, we ignored an interesting type of expansion: the β-expansion. Say $\beta > 1$ is a real number and consider the β-expansion. Instead of $0, 1, ..., n-1$ the possible elements become $0, 1, ..., \lfloor \beta \rfloor$, where $\lfloor \beta \rfloor$ is the largest integer smaller than or equal to β.

Similarly to the n-ary expansion, we would like an interval partition of $[0,1)$ to consist of subintervals with length $\frac{1}{\beta}$. But if β is not an integer we can obviously not have β subintervals. Instead, we take $\lfloor \beta \rfloor$ subintervals of length $\frac{1}{\beta}$, which leaves $1 - \lfloor \beta \rfloor \cdot \frac{1}{\beta} = \frac{\beta - \lfloor \beta \rfloor}{\beta}$ for the last interval. Summarizing, we obtain the partition $\{[0, \frac{1}{\beta}), [\frac{1}{\beta}, \frac{2}{\beta}), ..., [\frac{\lfloor \beta \rfloor - 1}{\beta}, \frac{\lfloor \beta \rfloor}{\beta}), [\frac{\lfloor \beta \rfloor}{\beta}, 1)\}$. Note that the last interval is smaller than the others. To see why this makes sense, take a look at the following example.

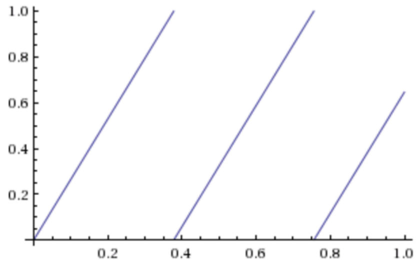
Example 3.2. Take $x \in [0,1)$ and β=9.5. The corresponding interval partition is given by $\{[0,\frac{1}{9.5}),[\frac{1}{9.5},\frac{2}{9.5}),...,[\frac{8}{9.5},\frac{9}{9.5}),[\frac{9}{9.5},1)\}$. For example, if $x \in [\frac{8}{9.5},\frac{9}{9.5})$, x must be between 0.8 and 0.9 in β-expansion notation. On the other hand, if $x \in [\frac{9}{9.5},1)$, x can only be between 0.9 and 0.95 (0.96 does not exist when $\beta = 9.5$). So the last interval should indeed be smaller.

As we have just seen, the interval partition of the β-expansion is quite different from the n-ary expansion. Even so, their corresponding expansion maps should satisfy the same properties, and thus the β-expansion map is defined by

$$T_\beta(x) = \beta x \mod 1.$$

From this map we can derive another interesting property of the β-expansion:

Figure 3.3: The β-expansion map T_β for $\beta = \sqrt{7}$.



if $a_i = \lfloor \beta \rfloor$ for some element of the β-expansion of x, not all "possible" elements $0,1,...,\lfloor \beta \rfloor$ could occur at a_{i+1}. In Example 3.2 we already remarked upon the fact that 0.96 does not exist in any β-expansion when $\beta = 9.5$. So the sequence $a_i = 9$, $a_{i+1} = 6$ is not tolerated. Another way to see this is by looking at the β-expansion map, shown in Figure 3.3.
Note that with $\beta = \sqrt{7}$, the possible elements are 0, 1 and 2. Assume that the first element of some number x is 2, or equivalently, $\frac{\lfloor\sqrt{7}\rfloor}{\sqrt{7}} \leq x < 1$. By definition of $T_{\sqrt{7}}$, $0 \leq T_{\sqrt{7}}(x) < \sqrt{7} - \lfloor\sqrt{7}\rfloor$. Since $\sqrt{7} - \lfloor\sqrt{7}\rfloor < \frac{\lfloor\sqrt{7}\rfloor}{\sqrt{7}}$, the second element of x, which is the first element of $T_{\sqrt{7}}(x)$, cannot be another 2.
Now we have defined the β-expansion map, we will focus on finding an invariant measure. Lebesgue measure might be a straightforward thought, but this measure turns out not to be invariant under T_β. Consider for example $\beta = 1.5$ and the interval $(\frac{1}{2},1)$. We find $\lambda(T_{1.5}^{-1}((\frac{1}{2},1))) = \lambda((\frac{1}{3},\frac{2}{3})) = \frac{1}{3} \neq \frac{1}{2} = \lambda((\frac{1}{2},1))$, so indeed the Lebesgue measure is not invariant.
We will soon give an explicit formula for the invariant measure of the β-expansion, but we need some notation and a lemma first. The proof of that lemma and the one of the subsequent theorem originate from [6].
Let $S(x) = \{n \in \mathbb{N} \cup \{0\} : x < T_\beta^n(1)\}$, and define $h_\beta(x) = \sum\limits_{n \in S(x)} \frac{1}{\beta^n}$.

Lemma 3.3.1 (Parry). $\sum\limits_{m=0}^{\lfloor \beta - x \rfloor} h_\beta(\frac{x+m}{\beta}) = \beta \cdot h_\beta(x)$.

Proof. Let $c_n = \begin{cases} 1 & \text{if } x < T_\beta^n(1) \\ 0 & \text{otherwise} \end{cases}$ and $c_{n,m} = \begin{cases} 1 & \text{if } \frac{x+m}{\beta} < T_\beta^n(1) \\ 0 & \text{otherwise} \end{cases}$.

Then $\sum\limits_{m=0}^{\lfloor\beta-x\rfloor} c_{n,m}$ denotes the number of times in $\{0, 1, ..., \lfloor\beta - x\rfloor\}$ that $\frac{x+m}{\beta} < T_\beta^n(1)$. Note that this is just the number of linear pieces of the graph of T_β such that $\frac{x+m}{\beta} < T_\beta^n(1)$. It follows that $\sum\limits_{m=0}^{\lfloor\beta-x\rfloor} c_{n,m} = \begin{cases} \lfloor\beta T_\beta^n(1)\rfloor + 1 & \text{if } x < T_\beta^{n+1}(1) \\ \lfloor\beta T_\beta^n(1)\rfloor & \text{otherwise} \end{cases}$.
Note also that any $y \in [0, 1)$ can be written as a sum of its expansion elements as follows: $y = \frac{a_1}{\beta} + \frac{a_2}{\beta^2} + \frac{a_3}{\beta^3} + ... = \frac{\lfloor\beta T_\beta^0(x)\rfloor}{\beta^1} + \frac{\lfloor\beta T_\beta^1(x)\rfloor}{\beta^2} + \frac{\lfloor\beta T_\beta^2(x)\rfloor}{\beta^3} + ...$ In particular, and because $T_\beta(1) = \beta - \lfloor\beta\rfloor$, we find that $\beta - \lfloor\beta\rfloor = \frac{\lfloor\beta T_\beta^0(\beta-\lfloor\beta\rfloor)\rfloor}{\beta^1} + \frac{\lfloor\beta T_\beta^1(\beta-\lfloor\beta\rfloor)\rfloor}{\beta^2} + \frac{\lfloor\beta T_\beta^2(\beta-\lfloor\beta\rfloor)\rfloor}{\beta^3} + ... = \frac{\lfloor\beta T_\beta^1(1)\rfloor}{\beta^1} + \frac{\lfloor\beta T_\beta^2(1)\rfloor}{\beta^2} + \frac{\lfloor\beta T_\beta^3(1)\rfloor}{\beta^3} + ...$ Therefore, $\beta = \frac{\lfloor\beta T_\beta^0(1)\rfloor}{\beta^0} + \frac{\lfloor\beta T_\beta^1(1)\rfloor}{\beta^1} + \frac{\lfloor\beta T_\beta^2(1)\rfloor}{\beta^2} + \frac{\lfloor\beta T_\beta^3(1)\rfloor}{\beta^3} + ... = \sum\limits_{n=0}^{\infty} \frac{\lfloor\beta T_\beta^n(1)\rfloor}{\beta^n}$.
With this knowledge we can finally move to the desired result. We obtain

$$\begin{aligned}
\sum_{m=0}^{\lfloor\beta-x\rfloor} h_\beta(\frac{x+m}{\beta}) &= \sum_{m=0}^{\lfloor\beta-x\rfloor}\sum_{n=0}^{\infty} \frac{c_{n,m}}{\beta^n} = \sum_{n=0}^{\infty}\sum_{m=0}^{\lfloor\beta-x\rfloor} \frac{c_{n,m}}{\beta^n} \\
&= \sum_{n=0}^{\infty} \frac{\lfloor\beta T_\beta^n(1)\rfloor + c_{n+1}}{\beta^n} = \sum_{n=0}^{\infty} \frac{\lfloor\beta T_\beta^n(1)\rfloor}{\beta^n} + \sum_{n=0}^{\infty} \frac{c_{n+1}}{\beta^n} \\
&= \beta + \beta \sum_{n=0}^{\infty} \frac{c_n}{\beta^n} - \beta c_0 = \beta + \beta \sum_{n\in S(x)} \frac{1}{\beta^n} - \beta \\
&= \beta \cdot h_\beta(x).
\end{aligned}$$

□

In the last equation, the sums could be interchanged because the result is finite.

Theorem 3.3.2 (Parry). *Let $\nu(E) = \int_E h_\beta(x)dx$. Ten ν is invariant under T_β.*

Proof. Let $[a, b) \subseteq [0, 1)$ be an interval such that either $a \geq \beta - \lfloor\beta\rfloor$ and $b \geq \beta - \lfloor\beta\rfloor$, or $a \leq \beta - \lfloor\beta\rfloor$ and $b \leq \beta - \lfloor\beta\rfloor$. Note that $T_\beta^{-1}([a, b)) = \bigcup\limits_{k=0}^{\lfloor\beta-b\rfloor} [\frac{a+m}{\beta}, \frac{b+m}{\beta})$.
We show that $\nu([a, b)) = \nu(T_\beta^{-1}([a, b)))$.
Using Lemma 3.3.1, we write $\nu([a, b)) = \int_a^b h_\beta(x)dx = \int_a^b \frac{1}{\beta} \sum\limits_{k=0}^{\lfloor\beta-x\rfloor} h_\beta(\frac{x+k}{\beta})dx = \frac{1}{\beta} \sum\limits_{k=0}^{\lfloor\beta-x\rfloor} \int_a^b h_\beta(\frac{x+k}{\beta})dx$. For the substitution $y = \frac{x+k}{\beta}$ we have $\frac{dx}{dy} = \beta$ and $x \in$

$[a,b)$ if and only if $y \in [\frac{a+k}{\beta}, \frac{b+k}{\beta})$. Thus,

$$\nu((a,b)) = \frac{1}{\beta} \sum_{k=0}^{\lfloor \beta - x \rfloor} \int_{\frac{a+k}{\beta}}^{\frac{b+k}{\beta}} \beta h_\beta(y) dy = \int_{\bigcup_{k=0}^{\lfloor \beta - b \rfloor} [\frac{a+k}{\beta}, \frac{b+k}{\beta})} h_\beta(y) dy$$
$$= \int_{T_\beta^{-1}([a,b))} h_\beta(y) dy = \nu(T_\beta^{-1}([a,b))).$$

If $[c,d) \subseteq [0,1)$ is an interval not satisfying our condition, note that $[c,d) = [c, \beta - \lfloor \beta \rfloor) \cup [\beta - \lfloor \beta \rfloor, d)$. This means our set of intervals $[a,b)$ generates all intervals $[c,d)$, so it also generates $\mathcal{B}([0,1))$. We conclude that T_β is measurable and ν is an invariant measure for T_β. $\square$

Remark. Unlike the invariant measures of the other expansions, ν is not (yet) a probability measure. Equivalently, $\nu([0,1)) = \int_0^1 h_\beta(x) dx$ is not necessarily 1. In order to correct this we define

$$\nu^*(E) = \frac{\int_E h_\beta(x) dx}{\int_0^1 h_\beta(x) dx}.$$

Please note that ν^* is still invariant under T_β. In the future we will only be interested in the probability measure ν^*, so from now on denote $\nu := \nu^*$.

3.4 Continued fractions expansions

Sometimes the best way to write down a number is by a continued fractions expansion. We will see why in Example 3.3. Continued fractions form the last type of expansion we will review.
The idea is as follows: let $x \in (0,1)$. Since $x < 1$, we have $\frac{1}{x} > 1$. We continue by writing $\frac{1}{x} = a_1 + r_1$, where $a_1 \in \mathbb{N}$ is the integer part of $\frac{1}{x}$ and $r_1 \in [0,1)$. If $r_1 \neq 0$, repeating this trick for r_1 yields $\frac{1}{r_1} = a_2 + r_2$, once again with $a_2 \in \mathbb{N}$ and $r_2 \in [0,1)$. We go on like this, which gives rise to the equation

$$x = \frac{1}{\frac{1}{x}} = \frac{1}{a_1 + r_1} = \frac{1}{a_1 + \frac{1}{\frac{1}{r_1}}} = \frac{1}{a_1 + \frac{1}{a_2 + r_2}} = ... = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \ddots}}}.$$

The continued fractions expansion of x then looks like $(a_1, a_2, a_3, ...)$, and is finite if and only if $r_i = 0$ for some i. Note that the possible elements of the expansion are formed by $\mathbb{N}$.

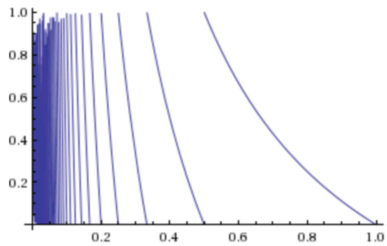
Example 3.3. Consider $\sqrt{2}-1$. Although the n-ary expansion of this number is unpredictable (for any n), its continued fraction is surprisingly elegant. From the identity $\frac{1}{\sqrt{2}-1} = \sqrt{2}+1 = 2+\sqrt{2}-1$, it follows that $\sqrt{2}-1 = \frac{1}{2+\frac{1}{2+...}} =$ $(2, 2, 2, ...)$.

Denote the continued fractions expansion map by T_c, and let $\frac{1}{a_1+\frac{1}{a_2+\frac{1}{a_3+...}}}$ be the continued fraction of x. Because $T_c(a_1, a_2, a_3, ...) = (a_2, a_3, ...)$, we must have $T_c(x) = \frac{1}{a_2+\frac{1}{a_3+...}}$. So $x = \dfrac{1}{a_1 + T_c(x)}$, and we get $T_c(x) = \frac{1}{x} - a_1$, or

$$T_c(x) = \frac{1}{x} - \lfloor \frac{1}{x} \rfloor.$$

Figure 3.4 shows how this map is different from the others. The interval

Figure 3.4: The continued fractions expansion map T_c.



partition has an infinite amount of intervals, which makes sense because there are $\mathbb{N}$ possible elements. Besides that, T_c is non-linearly decreasing (from 1 to 0) on each of the intervals.
Given that the map T_c is so different, will it also have a different invariant measure? The answer is yes.

Theorem 3.4.1. *Let μ be the measure defined by $\mu(A) = \frac{1}{\ln(2)} \int\limits_A \frac{1}{1+x} dx$. Then μ is invariant under the continued fractions map T_c.*

Proof. Let $(a, b) \subset [0, 1)$ be any interval. First of all, note that $T_c^{-1}((a, b)) = \bigcup\limits_{n=1}^{\infty} (\frac{1}{b+n}, \frac{1}{a+n})$ is a countable union of intervals, so T_c is measurable. We show that $\mu(T_c^{-1}((a, b))) = \mu((a, b))$.
Because the intervals of the union are disjoint, we find $\mu(T_c^{-1}((a, b))) = \sum\limits_{n=1}^{\infty} \mu((\frac{1}{b+n}, \frac{1}{a+n})) = \sum\limits_{n=1}^{\infty} \frac{1}{\ln(2)} \int\limits_{\frac{1}{b+n}}^{\frac{1}{a+n}} \frac{1}{1+x} dx = \frac{1}{\ln(2)} \sum\limits_{n=1}^{\infty} \ln(1+\frac{1}{a+n}) - \ln(1+\frac{1}{b+n}) = \frac{1}{\ln(2)} (\sum\limits_{n=1}^{\infty} \ln(\frac{a+n+1}{a+n}) - \sum\limits_{n=1}^{\infty} \ln(\frac{b+n+1}{b+n})) = \frac{1}{\ln(2)} (\ln(\prod\limits_{n=1}^{\infty} \frac{a+n+1}{a+n}) - \ln(\prod\limits_{n=1}^{\infty} \frac{b+n+1}{b+n})) = \frac{1}{\ln(2)} (\ln(b+1) - \ln(a+1)) = \frac{1}{\ln(2)} \int\limits_a^b \frac{1}{1+x} dx = \mu((a, b))$, which proves the theorem. □

4 Ergodic theory and dynamical systems

We have seen a few types of expansions, and the most important measure theory when it comes to analyzing them. Building on this measure theory, we move to a slightly more specific view of the expansions by looking at dynamical systems, in particular ergodic systems. The goal of this section is to discuss the Ergodic Theorem and its consequences, but first we need cover the notion of ergodicity and dynamical systems.

Definition 4.0.1. A *dynamical system* is the sequence $(X, \mathcal{A}, \mu, T)$, where:

- $(X, \mathcal{A}, \mu)$ is a probability space
- $T : X \to X$ is a surjective map
- μ is invariant under T

Example 4.1. Check that for $T(x) = 1 - x$, $((0,1), \mathcal{B}((0,1)), \lambda, T)$ is a dynamical system. To see that λ is invariant under T, note that $\lambda(T^{-1}(a,b)) = \lambda((1-b, 1-a)) = (1-a) - (1-b) = b - a = \lambda((a,b))$ is enough because of Lemma 3.1.3.

For each one of our expansions, note that $([0,1), \mathcal{B}([0,1)), \mu, T)$ is also a dynamical system. Here T should be the corresponding expansion map and μ its invariant measure, both as discussed in the previous section.

4.1 Ergodicity

Ergodicity is the last property of an expansion we need in order to make some interesting statements about it.

Definition 4.1.1. Let $(X, \mathcal{A}, \mu, T)$ be a dynamical system. Then T is called *ergodic* if for each $A = A' \cup N$ ($A' \in \mathcal{A}$ and N a subset of some null set) with $T^{-1}(A) = A$, we have $\mu(A) = 0$ or $\mu(A) = 1$. In that case, we refer to $(X, \mathcal{A}, \mu, T)$ as an *ergodic system*.

In other words, if T is ergodic, the only sets A for which $T^{-1}(A) = A$ are null sets or complements of null sets.

Example 4.2. Although the system in Example 4.1 is dynamical, it is not ergodic. For example, $\lambda((\frac{1}{4}, \frac{3}{4})) = \frac{1}{2}$ while $T^{-1}((\frac{1}{4}, \frac{3}{4})) = (1 - \frac{3}{4}, 1 - \frac{1}{4}) = (\frac{1}{4}, \frac{3}{4})$.

We will try to prove that our expansion maps are in fact ergodic. However, the definition of ergodicity forces us once again to check too many sets, namely every set $A' \cup N$ with $A' \in \mathcal{B}([0,1))$ and $N \subseteq M$ for some null set M. A lemma will be stated to prevent that.

Lemma 4.1.2 (Knopp). *Let $B = B' \cup N$ ($B' \in \mathcal{B}([0,1))$ and N a subset of some null set), and let $\mathcal{C}$ be a collection of subintervals of $[0,1)$ with the following properties:*

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

2. For all $A \in \mathcal{C}$, we have $\lambda(A \cap B) \geq \gamma\lambda(A)$, with $\gamma > 0$ independent of A.

Then $\lambda(B) = 1$.

In the proof we will make use of the measure theoretical theorem below.

Theorem 4.1.3. *Let $\mathcal{E}$ be the collection of all finite disjoint unions of subintervals of $[0, 1)$. Then for every $B \in \mathcal{B}([0, 1))$ and every $\epsilon > 0$ there exists an $E \in \mathcal{E}$ such that $\lambda(B \triangle E) < \epsilon$.* (Recall that $B \triangle E = (B \cap E) \cup (B^c \cap E^c)$.)

We will not prove this theorem here. The interested reader is referred to [5], p. 84.

Proof of Lemma 4.1.2. Assume the contrary: $\lambda(B) < 1$, or $\lambda(B^c) > 0$. This is equivalent to $\lambda(B'^c) > 0$ because $\lambda(N) = 0$.
Let $\epsilon > 0$. Now because of Theorem 4.1.3 there exists an E_ϵ which is a finite disjoint union of open intervals, such that $\lambda(B'^c \triangle E_\epsilon) < \epsilon$. Note that each of the disjoint intervals of E_ϵ can be written as an at most countable disjoint union of elements of $\mathcal{C}$. So $E_\epsilon = \bigcup\limits_i C_i$ where the C_i are pairwise disjoint elements of $\mathcal{C}$. It follows that $\lambda(B' \cap E_\epsilon) = \lambda(B' \cap \bigcup\limits_i C_i) = \lambda(\bigcup\limits_i (B' \cap C_i)) = \sum\limits_i \lambda(B' \cap C_i) \geq \sum\limits_i \gamma\lambda(C_i) = \gamma\lambda(\bigcup\limits_i C_i) = \gamma\lambda(E_\epsilon)$.
We use the last equation to write $\lambda(B'^c \triangle E_\epsilon) \geq \lambda(B' \cap E_\epsilon) \geq \gamma\lambda(E_\epsilon) \geq \gamma\lambda(B'^c \cap E_\epsilon) = \gamma\lambda(B'^c \setminus (B'^c \triangle E_\epsilon)) \geq \gamma(\lambda(B'^c) - \lambda(B'^c \triangle E_\epsilon)) > \gamma(\lambda(B'^c) - \epsilon)$.
From this we obtain $\gamma(\lambda(B'^c) - \epsilon) < \lambda(B'^c \triangle E_\epsilon) < \epsilon$, and so $\lambda(B'^c) < \frac{\epsilon + \gamma\epsilon}{\gamma}$.
Since this holds for all $\epsilon > 0$, we must have $\lambda(B'^c) = 0$, a contradiction. $\square$

Remark. In the proof, we implicitly assumed that $\lambda(B)$ and $\lambda(N)$ were defined, while it is not always true that B and N are in $\mathcal{B}([0, 1))$. However, we can define the completion of $\mathcal{B}([0, 1))$ by $\{B' \cup N : B' \in \mathcal{B}([0, 1)),\ N \subseteq M \text{ with } \lambda(M) = 0\}$. On this completion, we define a measure λ^* given by $\lambda^*(B' \cup N) = \lambda(B')$. In the probability space we just created, note that we can now measure B and N.

With Knopp's Lemma as our tool, we now prove the ergodicity of the mentioned expansions.

Notation. We would like a short notation for the interval $\{x \in [0, 1) : a_1 = i_1,\ a_2 = i_2,\ ...,\ a_n = i_n\}$. Denote this set by $\Delta(i_1, i_2, .., i_n)$, a so-called *cylinder set* of rank n. We write $\Delta_n(x) = \Delta(i_1, i_2, .., i_n)$ if $x \in \Delta(i_1, i_2, .., i_n)$.

Theorem 4.1.4. *The GLS-map T_G is ergodic.*

Proof. Let $B \in \mathcal{B}([0, 1))$ with $T_G^{-1}(B) = B$ and $\lambda(B) > 0$, and let $\mathcal{E}$ be the collection of all cylinder sets from the GLS-expansion. We see that the first condition of Knopp's Lemma has already been fulfilled.
Now note that the slope of T_G is constant on any cylinder set. So if $A_1, A_2 \subseteq E \in \mathcal{E}$, then $\frac{\lambda(A_1)}{\lambda(A_2)} = \frac{\lambda(T_G(A_1))}{\lambda(T_G(A_2))}$. Note also that

$$T_G^n(\Delta(i_1, i_2, .., i_n)) = T_G^{n-1}(\Delta(i_2, i_3, .., i_n)) = ... = T_G(\Delta(i_n)) = [0, 1)$$

results in $\lambda(T_G^n(E)) = 1$ for any cylinder set E of rank n.
Consequently, if E is a cylinder set of rank n, we obtain

$$\frac{\lambda(E \cap B)}{\lambda(E)} = \frac{\lambda(E \cap T_G^{-n}(B))}{\lambda(E)} = \frac{\lambda(T_G^n(E) \cap B)}{\lambda(T_G^n(E))} = \frac{\lambda([0, 1) \cap B)}{\lambda([0, 1))} = \lambda(B).$$

It follows that $\lambda(E \cap B) = \lambda(E)\lambda(B)$, which means the second condition holds for $\gamma = \lambda(B)$. Therefore we achieve $\lambda(B) = 1$, and T_G must be ergodic. $\square$

Keep in mind that every n-ary expansion is just a special case of a GLS-expansion. Thus ergodicity, and any other property we prove for the GLS-expansion later on, also holds for the n-ary expansion.

We proceed with the ergodicity of the β-expansion map. Note that the first part of the proof will be constructed as in [9].

Theorem 4.1.5. *The β-expansion map T_β is ergodic.*

Proof. Let $B \in \mathcal{B}([0,1))$ with $T_\beta^{-1}(B) = B$ and $\lambda(B) > 0$, and let $\mathcal{E}$ be the set of all full intervals. A full interval of rank n is a cylinder set $\Delta(i_1, i_2, ..., i_n)$ for which $\lambda(T_\beta^n(\Delta(i_1, i_2, ..., i_n))) = 1$. Since T_β has slope β, we have that $\lambda(T_\beta(\Delta(i_1, i_2, ..., i_n))) = \beta\lambda(\Delta(i_1, i_2, ..., i_n))$. It follows that

$$\lambda(T_\beta^n(\Delta(i_1, i_2, ..., i_n))) = \beta^n \lambda(\Delta(i_1, i_2, ..., i_n)) = 1,$$

and we get that $\lambda(\Delta(i_1, i_2, ..., i_n)) = \beta^{-n}$ for any full interval.
We first try to cover $[0,1)$ up to a set of (Lebesgue) measure zero with disjoint full intervals of rank n. Let $\frac{\epsilon}{3} > 0$ and $n \geq 1$ be given. Begin with filling [0,1) with all full intervals of rank n. Now fill the remainder of $[0,1)$ as far as possible with full intervals of rank $n+1$. Repeat this trick up to $n+k$. Note that the proportion of non-full intervals versus full intervals of rank $n+i$ is always $1 - \frac{\lfloor \beta \rfloor}{\beta} < \frac{1}{2}$. So we have covered $[0,1)$ up to a set of measure smaller than $(\frac{1}{2})^k < \frac{\epsilon}{3}$ if k is large enough.
Now we also covered each open subinterval I, as long as we remove those full intervals that contain the endpoints of I. For n large enough, these intervals have measure smaller than $\beta^{-n} < \frac{\epsilon}{3}$. It follows we have covered I with disjoint full intervals up to a set of measure ϵ.
In order to satisfy the first condition of Knopp's Lemma, let D_m be the cover of I we created for $\epsilon = \frac{1}{m}$. Then $I = \bigcup_{m=1}^{\infty} D_m$ is a countable disjoint union of full intervals.
We proceed in the manner of the proof of Theorem 4.1.4. If E is a full interval of rank n, we obtain $\frac{\lambda(E \cap B)}{\lambda(E)} = \frac{\lambda(E \cap T_\beta^{-n}(B))}{\lambda(E)} = \frac{\lambda(T_\beta^n(E) \cap B)}{\lambda(T_\beta^n(E))} = \frac{\lambda([0,1) \cap B)}{\lambda([0,1))} = \lambda(B)$.
It follows that $\lambda(E \cap B) = \lambda(E)\lambda(B)$, which means the second condition from Knopp's Lemma holds for $\gamma = \lambda(B)$. Therefore we achieve $\lambda(B) = 1$, and T_β must be ergodic. $\square$

Remark. It might seem odd that we use Lebesgue measure in the proof, while we have seen it is not invariant under T_β. The proof is however justified, as the invariant measure ν is of the form $\nu(A) = \int_A g(x)dx$, where $g(x)$ is positive and bounded. Note that such measures have precisely the same null sets as λ, and that is enough for showing ergodicity. Any measure with the named property is said to be *equivalent* to the Lebesgue measure.

Theorem 4.1.6. *The continued fractions map T_c is ergodic.*

Proof. Let $B \in \mathcal{B}([0,1))$ with $T_c^{-1}(B) = B$ and $\mu(B) = \frac{1}{\ln(2)} \int_B \frac{1}{1+x} dx > 0$, and let $\mathcal{E}$ be the set of all cylinders. Note that every rational number can be represented by a finite continued fractions expansion. Therefore, every rational number is a left endpoint of some cylinder, and the first condition of Knopp's Lemma is satisfied. Before we continue, we list some minor results that can be verified with number theory. Denote $\frac{p_n}{q_n}$ for the number represented by the finite expansion $(a_1, a_2, ..., a_n)$. Then:

1. $p_{n-1}q_n - p_n q_{n-1} = (-1)^n$
2. The cylinder $\Delta(a_1, a_2, ..., a_n)$ has endpoints $\frac{p_n}{q_n}$ and $\frac{p_n + p_{n-1}}{q_n + q_{n-1}}$
3. If the first n elements of the expansion of x are $a_1, a_2, ..., a_n$, then $x = \frac{p_n + T_c^n(x) p_{n-1}}{q_n + T_c^n(x) q_{n-1}}$

We now concentrate ourselves to the measure theoretic part of the proof. Let $\Delta_n = \Delta(a_1, a_2, ..., a_n)$ be a cylinder of rank n, and let $[a, b)$ be any interval. Note that if $T_c^n(x) = a$, then $x = \frac{p_n + a p_{n-1}}{q_n + a q_{n-1}}$, and we have the same with b. It follows that $T_c^{-1}([a,b)) \cap \Delta_n$ has endpoints $\frac{p_n + a p_{n-1}}{q_n + a q_{n-1}}$ and $\frac{p_n + b p_{n-1}}{q_n + b q_{n-1}}$. Since it depends on n which fraction is larger, we take the absolute value to calculate the Lebesgue measure:

$$\begin{aligned}
\lambda(T_c^{-1}([a,b)) \cap \Delta_n) &= \left| \frac{p_n + a p_{n-1}}{q_n + a q_{n-1}} - \frac{p_n + b p_{n-1}}{q_n + b q_{n-1}} \right| \\
&= \left| \frac{(p_n + a p_{n-1})(q_n + b q_{n-1}) - (p_n + b p_{n-1})(q_n + a q_{n-1})}{(q_n + a q_{n-1})(q_n + b q_{n-1})} \right| \\
&= \left| \frac{(a-b)(p_{n-1}q_n - p_n q_{n-1})}{(q_n + a q_{n-1})(q_n + b q_{n-1})} \right| \\
&= (b-a) \cdot \frac{q_n(q_{n-1} + q_n)}{q_n(q_{n-1} + q_n)} \cdot \frac{1}{(q_n + a q_{n-1})(q_n + b q_{n-1})} \\
&= \lambda([a,b)) \cdot \lambda(\Delta_n) \cdot \frac{q_n(q_{n-1} + q_n)}{(q_n + a q_{n-1})(q_n + b q_{n-1})}.
\end{aligned}$$

It is easy to check that $q_{k+1} > q_k$ for all $k \in \mathbb{N}$. As a result, $\frac{1}{2} < \frac{q_n}{q_{n-1} + q_n} < \frac{q_n(q_{n-1} + q_n)}{(q_n + a q_{n-1})(q_n + b q_{n-1})} < \frac{q_n(q_{n-1} + q_n)}{q_n^2} < 2$. From this we obtain $\frac{1}{2}\lambda([a,b))\lambda(\Delta_n) < \lambda(T_c^{-n}([a,b)) \cap \Delta_n) < 2\lambda([a,b))\lambda(\Delta_n)$ for all intervals $[a,b)$. Those intervals generate $\mathcal{B}([0,1))$, so actually

$$\frac{1}{2}\lambda(A)\lambda(\Delta_n) \leq \lambda(T_c^{-n}(A) \cap \Delta_n) \leq 2\lambda(A)\lambda(\Delta_n)$$

for all $A \in \mathcal{B}([0,1))$. Moreover, note that

$$\frac{1}{2\ln(2)}\lambda(A) \leq \mu(A) \leq \frac{1}{\ln(2)}\lambda(A)$$

because $\frac{1}{2} \leq \frac{1}{1+x} \leq 1$ when $x \in [0,1)$.

With these two inequalities, we write

$$\begin{aligned}
\mu(T_c^{-n}(A) \cap \Delta_n) &\geq \frac{1}{2\ln(2)}\lambda(T_c^{-n}(A) \cap \Delta_n) \\
&\geq \frac{1}{4\ln(2)}\lambda(A)\lambda(\Delta_n) \\
&= \frac{\ln(2)}{4}\frac{1}{\ln(2)}\lambda(A)\frac{1}{\ln(2)}\lambda(\Delta_n) \\
&\geq \frac{\ln(2)}{4}\mu(A)\mu(\Delta_n).
\end{aligned}$$

Taking $A = B$, we finally obtain $\mu(B \cap \Delta_n) = \mu(T_c^{-n}(B) \cap \Delta_n) \geq \frac{\ln(2)}{4}\mu(B)\mu(\Delta_n)$, the second condition of Knopp's Lemma with $\gamma = \frac{\ln(2)}{4}\mu(B)$. We conclude that T_c is ergodic. □

The ergodicity of the expansions will prove useful in the following subsection, where we will study the general behaviour of expansion elements.

4.2 Expansion element frequency

We are almost ready to state the Ergodic Theorem, the most important theorem in the thesis because of its many consequences. But first let us discuss the notion of normal numbers.

Definition 4.2.1. Let $x \in [0, 1)$ be a number with GLS-expansion $(a_1, a_2, ...)$, and let D be the set of all possible digits. Then x is *normal* if for any block of digits $b_1, b_2, ..., b_m$, one has $\lim_{n\to\infty} \frac{1}{n}\#\{i \in \{1, ..., n\} : a_i = b_1, a_{i+1} = b_2, ..., a_{i+m-1} = b_m\} = (r_{b_1} - l_{b_1}) \cdot (r_{b_2} - l_{b_2}) \cdot ... \cdot (r_{b_m} - l_{b_m})$.

So the frequency of every block has to be proportional to the interval length corresponding to that block. If we only allow blocks of a single digit, we can use the Strong Law of Large Numbers (SLLN) to derive an interesting property of the GLS-expansion. Recall that the SLLN is defined as follows.

Theorem 4.2.2 (Strong Law of Large Numbers)**.** *Let $X_1, X_2, ...$ be i.i.d random variables such that $E(X_1) < \infty$ Then, almost everywhere, $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = E(X_1)$.*

This theorem states that if we repeat the same random experiment often enough, the average outcome will converge to the expected value. Now let $X_i(x) = \begin{cases} 1 & \text{if } a_i(x) = j \\ 0 & \text{otherwise} \end{cases}$ Note that the X_i are independent, identically distributed and that $E(X_1) < \infty$. We find that for any $j \in D$, $\lim_{n\to\infty} \frac{1}{n}\#\{i \in \{1, ..., n\} : a_i = j\} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = E(X_1) = 1 \cdot \lambda(X_1) + 0 \cdot \lambda(X_1^c) = \lambda(\{x : a_1(x) = j\}) = r_j - l_j$ for almost all $x \in [0, 1)$.

We just proved that for almost all x, or with probability 1, the frequency of any expansion element j is equal to its interval length $r_j - l_j$. If it is not clear yet why this is interesting, note that this implies that if we pick a "random"

number in $[0,1)$, with probability 1 the numbers $0,1,...,9$ will appear equally often in the decimal expansion. The word "random" is put in quotation marks because it is hard to create such a random number.
Unfortunately, we cannot prove this property for blocks of multiple digits. Consider for example the blocks $(1,4)$ and $(4,8)$ in the decimal expansion. Note that

$$\begin{aligned}
&\lambda(X_i(1,4)\cap X_{i+1}(4,8))\\
&=\lambda(\{x:a_i(x)=1,a_{i+1}(x)=4\}\cap\{x:a_{i+1}(x)=4,a_{i+2}(x)=8\})\\
&=\lambda(\{x:a_i(x)=1,a_{i+1}(x)=4,a_{i+2}(x)=8\})\\
&=\frac{1}{10^3}\neq\frac{1}{10^4}=\frac{1}{10^2}\cdot\frac{1}{10^2}\\
&=\lambda(\{x:a_i(x)=1,a_{i+1}(x)=4\})\cdot\lambda(\{x:a_{i+1}(x)=4,a_{i+2}(x)=8\})\\
&=\lambda(X_i(1,4))\cdot\lambda(X_{i+1}(4,8)).
\end{aligned}$$

From this we conclude that the X_i are not independent anymore, and we can not use the SLLN.
A better alternative can be found in the Ergodic Theorem.

Theorem 4.2.3 (Ergodic Theorem)**.** *Let $(X, \mathcal{A}, \mu, T)$ be a dynamical system, and suppose f is an integrable function with respect to μ. Then*

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

exists a.e. and $\int\limits_X fd\mu=\int\limits_X f^*d\mu$. If T is ergodic, we also have that $f^*=\int\limits_X fd\mu$ is a constant a.e.

Remark. For our purposes we want that $f=\mathbb{1}_B=:\begin{cases}1 & \text{if } x\in B\\ 0 & \text{otherwise}\end{cases}$ for some $B\subseteq[0,1)$. Together with the ergodicity of our expansions, the Ergodic Theorem can be reformulated as

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n}\mathbb{1}_B\circ T^i(x)=\mu(B),\ \text{or}$$

$$\lim_{n\to\infty}\frac{1}{n}\#\{i\in\{0,...,n\}:T^i(x)\in B\}=\mu(B).$$

To prove Theorem 4.2.3 we need help in the form of the following lemma.

Lemma 4.2.4. *Suppose $(X,\mathcal{A},\mu)$ is a measure space and $f:X\to[0,\infty)$ is an integrable function. If $\int\limits_X fd\mu=0$, then $f=0$ a.e.*

Proof. Let $\int\limits_X fd\mu=0$.
First assume that $f=\mathbb{1}_A$ for some $A\in\mathcal{A}$. Then $\mu(A)=\int\limits_X\mathbb{1}_A d\mu=\int\limits_X fd\mu=0$, so A is a null set and $f=\mathbb{1}_A=0$ a.e.
Now assume f is a simple function $\sum\limits_{i=1}^{n}c_i\mathbb{1}_{A_i}$. We have that $\int\limits_X fd\mu=\int\limits_X\sum\limits_{i=1}^{n}c_i\mathbb{1}_{A_i}d\mu=$

$\sum_{i=1}^{n} c_i \int_X \mathbb{1}_{A_i} d\mu = \sum_{i=1}^{n} c_i\mu(A_i) = 0$. Then $c_i\mu(A_i) = 0$ for all i, and it must be that $c_i\mathbb{1}_{A_i} = 0$ a.e. It immediately follows that $f = \sum_{i=1}^{n} c_i\mathbb{1}_{A_i} = 0$ a.e.
Finally, let f be any measurable function. Note that $\int_X f d\mu = sup\{\int_X \phi d\mu :$ ϕ simple and $\phi(x) \leq f(x)$ for all $x \in X\} = 0$ implies $\int_X \phi d\mu = 0$ for all simple functions $\phi \leq f$. Since we already proved this lemma for simple functions, $\phi = 0$ a.e. holds. We derive from Theorem 8.8 of [8] that $f = 0$ a.e.

□

Proof of Theorem 4.2.3. Because it is all we need for now, we will prove only the case $f = \mathbb{1}_B$ here. Define

$$f^+(x) = \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_B \circ T^i(x)$$

and

$$f^-(x) = \liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_B \circ T^i(x).$$

Some properties of f^+ and f^- are:

- they exist, because $\sum_{i=0}^{n-1} \mathbb{1}_B \circ T^i(x) = \#\{i \in \{0, ..., n-1\} : T^i(x) \in B\} \leq n$.
- they are measurable since measurability is preserved under composition and limits.
- $0 \leq f^-(x) \leq f^+(x) \leq 1$ for all $x \in X$.
- they are invariant under T.

For the fourth property, note that $f^+(T(x)) = \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_B \circ T^{i+1}(x) = \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n} \mathbb{1}_B \circ T^i(x) - \frac{\mathbb{1}_B}{n} = \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_B \circ T^i(x) - 0 = f^+(x)$. The same argument holds for $f^-(x)$.
The main goal now is to prove $\int_X f^+ d\mu \leq \mu(B)$.

Let $\epsilon > 0$ be arbitrary. Define the partial sum $S_n(x) = \sum_{i=0}^{n-1} \mathbb{1}_B \circ T^i(x)$ and $N(x) = \min\{n \geq 1 : S_n(x) \geq n(f^+(x) - \epsilon)\}$. Clearly this set is nonempty since the inequality holds for large n. Therefore we can find an $M > 0$ such that $\mu(\{x \in X : N(x) > M\}) < \epsilon$.
Let $B' = B \cup \{x \in X : N(x) > M\}$, $S'_n(x) = \sum_{i=0}^{n-1} \mathbb{1}_{B'} \circ T^i(x)$ and $N'(x) = \begin{cases} N(x) & \text{if } N(x) \leq M \\ 1 & \text{if } N(x) > M \end{cases}$. We distinguish two cases.
If $N(x) > M$, then $x \in B'$ and so

$$S'_{N'(x)}(x) = S'_1(x) = \mathbb{1}_{B'} = 1 > f^+(x) - \epsilon = (f^+(x) - \epsilon)N'(x).$$

If on the other hand $N(x) \leq M$, then since $B \subseteq B'$ we have

$$\begin{aligned} S'_{N'(x)}(x) &= S'_{N(x)}(x) \\ &= \sum_{i=0}^{N(x)-1} \mathbb{1}_{B'} \circ T^i(x) \\ &\geq \sum_{i=0}^{N(x)-1} \mathbb{1}_{B} \circ T^i(x) \\ &= S_{N(x)}(x) \\ &\geq N(x)(f^+(x) - \epsilon) \\ &= N'(x)(f^+(x) - \epsilon). \end{aligned}$$

Thus, for all $x \in X$: $S'_{N'(x)}(x) \geq (f^+(x) - \epsilon)N'(x)$.
We continue by defining the sequence $n_0(x) = 0$, $n_{k+1}(x) = n_k(x) + N'(T^{n_k(x)}(x))$.
Also, let $n > M$ be sufficiently large and denote $l = max\{k \geq 1 : n_k(x) \leq n-1\}$.
Given that $n_{i+1}(x) - n_i(x) = N'(T^{n_i(x)}(x) \leq M$ for all i, $S'_m(y) \geq m(f^+(y) - \epsilon)$ for all $y \in X$ and the fact that $f^+(x)$ is T-invariant, we can write:

$$\begin{aligned} S'_n &= \sum_{i=0}^{n-1} \mathbb{1}_{B'} \circ T^i(x) \\ &\geq \sum_{i=0}^{n_l(x)-1} \mathbb{1}_{B'} \circ T^i(x) \\ &= \sum_{i=0}^{l} \sum_{j=n_i(x)}^{n_{i+1}(x)-1} \mathbb{1}_{B'} \circ T^j(x) \\ &= \sum_{i=0}^{l} \sum_{k=0}^{n_{i+1}(x)-n_i(x)-1} \mathbb{1}_{B'} \circ T^k(T^{n_i(x)}(x)) \\ &= \sum_{i=0}^{l} S'_{N'(T^{n_i(x)}(x))}(T^{n_i(x)}(x)) \\ &\geq \sum_{i=0}^{l} N'(T^{n_i(x)}(x))(f^+(T^{n_i}(x)) - \epsilon) \\ &= \sum_{i=0}^{l} (n_{i+1}(x) - n_i(x))(f^+(x) - \epsilon) \\ &= n_l(x)(f^+(x) - \epsilon) \\ &= (n_{l+1}(x) - (n_{l+1}(x) - n_l(x)))(f^+(x) - \epsilon) \\ &\geq (n - M)(f^+(x) - \epsilon). \end{aligned}$$

From here we obtain $\frac{1}{n} \int\limits_X S'_n d\mu \geq \frac{n-M}{n} \int\limits_X f^+ - \epsilon d\mu = \frac{n-M}{n}(\int\limits_X f^+ d\mu - \int\limits_X \epsilon d\mu) = \frac{n-M}{n}(\int\limits_X f^+ d\mu - \epsilon)$, but we also have $\frac{1}{n} \int\limits_X S'_n d\mu = \frac{1}{n} \sum\limits_{i=0}^{n-1} \int\limits_X \mathbb{1}_{B'} \circ T^i(x) d\mu =$

$\frac{1}{n}\sum_{i=0}^{n-1}\int_X \begin{cases} 1 & \text{if } T^i(x) \in B' \\ 0 & \text{otherwise} \end{cases} d\mu = \frac{1}{n}\sum_{i=0}^{n-1}\mu(T^{-n}(B')) = \frac{1}{n}\sum_{i=0}^{n-1}\mu(B') = \mu(B')$.
If $n \to \infty$, this yields $\int_X f^+ d\mu - \epsilon \leq \mu(B')$, and $\mu(B) \geq \mu(B') - \mu(\{x \in X : N(x) > M\}) \geq \mu(B') - \epsilon = \int_X f^+ d\mu - 2\epsilon$. We may now conclude that indeed $\int_X f^+ d\mu \leq \mu(B)$.
From here it is not hard to show $\mu(B) \leq \int_X f^- d\mu$. Define $g = \mathbb{1}_{B^c} = 1 - \mathbb{1}_B$ and note that $g^+(x) = \limsup_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1}(1-\mathbb{1}_B)\circ T^i(x) = \limsup_{n\to\infty}\frac{1}{n}(n - \sum_{i=0}^{n-1}\mathbb{1}_B \circ T^i(x)) = \limsup_{n\to\infty} 1 - \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{1}_B \circ T^i(x) = 1 - \liminf_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{1}_B \circ T^i(x) = 1 - f^-(x)$. We can use this and the proof of the "main goal" to check that $\mu(B) = 1 - \mu(B^c) \leq 1 - \int_X g^+ d\mu = \int_X 1 - g^+ d\mu = \int_X f^- d\mu$.
So far we have proved that $\int_X f^+ d\mu \leq \mu(B) \leq \int_X f^- d\mu$. However, $f^+ \geq f^-$ and hence $\int_X f^+ d\mu \geq \int_X f^- d\mu$. Then it must be that $\int_X f^+ d\mu = \int_X f^- d\mu = \mu(B)$. Writing $\int_X f^+ - f^- d\mu \int_X f^+ d\mu - \int_X f^- d\mu = 0$, we can use Lemma 4.2.4 to conclude $f^+ = f^-$. Denote $f^* = f^+ = f^-$ and we have completed the proof of the first part of the theorem.
What is left is to show that it T is ergodic, then f^* is constant almost everywhere. So suppose T is ergodic. Consider the collection of sets $A_r = \{x \in X : f^*(x) > r\}$ $(r \in \mathbb{R})$. We already know f^* is T-invariant, hence $T(A_r) = \{x \in X : f^*(T^{-1}(x)) > r\} = \{x \in X : f^*(x) > r\} = A_r$. Since T is ergodic, $\mu(A_r)$ equals 0 or 1 for all $r \in \mathbb{R}$. We conclude that f^* is a constant almost everywhere.

□

With the Ergodic Theorem we are able to prove what we could not with the Strong Law of Large Numbers.

Corollary 4.2.5. *Almost every $x \in [0,1)$ is normal.*

Proof. Let $x \in [0,1)$ and let $b_1, b_2, ..., b_m$ be a block of digits. Define $B = \{y \in [0,1) : a_1(y) = b_1, a_2(y) = b_2, ..., a_m(y) = b_m\}$. Then $\lim_{n\to\infty}\frac{1}{n}\#\{i \in \{1,...,n\} : a_i(x) = b_1, a_{i+1}(x) = b_2, ..., a_{i+m-1}(x) = b_m\} = \lim_{n\to\infty}\frac{1}{n}\#\{i \in \{1,...,n\} : T^i(x) \in B\} = \lambda(B) = (r_{b_1} - l_{b_1}) \cdot (r_{b_2} - l_{b_2}) \cdot ... \cdot (r_{b_m} - l_{b_m})$ □

Since non-normal numbers form a Lebesgue null set, this corollary might give the impression it's easy to construct a normal number. The opposite is true though. No rational number, for example, can be normal with respect to the n-ary expansion. To see this, note that the expansion of each rational number is eventually periodic, meaning a certain block of digits repeats itself after a while: $(a_1, ..., a_k, b_1, ..., b_m, b_1, ..., b_m, b_1, ...)$. Then consider the block $b_1, b_2, ..., b_m, b_{m+1}$, where $b_{m+1} \neq b_1$. This block never occurs in the periodic part of the expansion, so its frequency is 0. On the other hand, in the expansion of a normal number this block should have frequency $\frac{1}{n^{m+1}}$.

Example 4.3. We can construct a normal number for the n-ary expansion by listing every block of 1 digit followed by every block of 2 digits, etc. For the decimal expansion, this results in the number

$$0.012345678910111213...979899100101102....$$

Little is known about which numbers are normal. In fact, no one has even been able to prove whether $\sqrt{2}$ is a normal number. This fact makes Corollary 4.2.5 even more remarkable than it already seemed. It shows how powerful ergodic theory is.
Let us continue with a suitable example of the β-expansion. We would like a $1 < \beta < 2$ such that the block 1, 1 can only just not occur. So given that the partition is $\{[0, \frac{\lfloor\beta\rfloor}{\beta}), [\frac{\lfloor\beta\rfloor}{\beta}, 1)\}$, we want that $T_\beta(1) = \frac{\lfloor\beta\rfloor}{\beta}$. Now since $\lfloor\beta\rfloor = 1$ and by definition of T_β, we find $\beta - 1 = \frac{1}{\beta}$. It follows that β satisfies $\beta^2 - \beta - 1 = 0$, and we conclude that β is the golden mean $\frac{1+\sqrt{5}}{2}$. In order to use the Ergodic Theorem, we have to calculate the invariant measure for $\beta = \frac{1+\sqrt{5}}{2}$ explicitly.
It is clear why we chose β like this when we calculate $T^0_\beta(1) = 1$, $T^1_\beta(1) = 1 - \beta$ and $T^n_\beta(1) = 0$ for all $n \geq 2$. Now $h_\beta(x) = \sum\limits_{S(x)} \frac{1}{\beta^n} = \begin{cases} 1 + \frac{1}{\beta} & \text{if } 0 < x < \beta - 1 \\ 1 & \text{if } \beta - 1 \leq x < 1 \end{cases}$, and $\int\limits_0^1 h_\beta(x)dx = \int\limits_0^1 \sum\limits_{S(x)} \frac{1}{\beta^n}\, dx = \int\limits_0^{\beta-1} 1 + \frac{1}{\beta}\, dx + \int\limits_{\beta-1}^1 1\, dx = \frac{5-\sqrt{5}}{2}$. We finally obtain $\nu(E) = \frac{\int_E h_\beta(x)dx}{\int_0^1 h_\beta(x)dx} = \begin{cases} \int\limits_E \frac{5+3\sqrt{5}}{10} & \text{if } 0 < x < \beta - 1 \\ \int\limits_E \frac{5+\sqrt{5}}{10} & \text{if } \beta - 1 \leq x < 1 \end{cases}$. From here it is easy to find the frequency of any block with the Ergodic Theorem.

Example 4.4. We calculate the frequency of the block 0,1 with $\beta = \frac{1+\sqrt{5}}{2}$. Note first that $\{x \in [0, 1) : a_1(x) = 0, a_2(x) = 1\} = [\frac{1}{\beta^2}, \frac{1}{\beta})$. The Ergodic Theorem implies that $\lim\limits_{n\to\infty} \frac{1}{n}\#\{i \in \{1, ..., n\} : a_i(x) = 0, a_{i+1}(x) = 1\} = \lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{i=0}^{n-1} \mathbb{1}_{[\frac{1}{\beta^2}, \frac{1}{\beta})}(T^i_\beta(x)) = \nu([\frac{1}{\beta^2}, \frac{1}{\beta})) = \int\limits_{\frac{1}{\beta^2}}^{\frac{1}{\beta}} \frac{5+3\sqrt{5}}{10}\, dx = \frac{5-\sqrt{5}}{10} \approx 0.276.$

We conclude this section with a theorem about frequency of single digits in the continued fractions expansion.

Theorem 4.2.6. *Let $a \in \mathbb{N}$. Then for almost all $x \in [0, 1)$ with continued fractions expansion $(a_1, a_2, ...)$:*

$$\lim_{n\to\infty} \frac{1}{n}\#\{i \in \{1, ..., n\} : a_i(x) = a\} = \frac{1}{\ln(2)} \ln(1 + \frac{1}{a(a + 2)}).$$

Proof. The equality is a relatively simple consequence of the Ergodic Theorem. Note that $\{x \in [0, 1) : a_1(x) = a\} = (\frac{1}{a+1}, \frac{1}{a}]$. We now have $\lim\limits_{n\to\infty} \frac{1}{n}\#\{i \in \{1, ..., n\} : a_i(x) = a\} = \lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{i=0}^{n-1} \mathbb{1}_{(\frac{1}{a+1}, \frac{1}{a}]}(T^i_c(x)) = \mu((\frac{1}{a+1}, \frac{1}{a}]) = \frac{1}{\ln(2)} \int\limits_{\frac{1}{a+1}}^{\frac{1}{a}} \frac{1}{1+x} dx = \frac{1}{\ln(2)}(\ln(1+\frac{1}{a}) - \ln(1+\frac{1}{a+1})) = \frac{1}{\ln(2)} \ln(\frac{1+\frac{1}{a}}{1+\frac{1}{a+1}}) = \frac{1}{\ln(2)} \ln(\frac{\frac{a+1}{a}}{\frac{a+2}{a+1}}) = \frac{1}{\ln(2)} \ln(\frac{(a+1)^2}{a(a+2)}) = \frac{1}{\ln(2)} \ln(\frac{a(a+2)+1}{a(a+2)}) = \frac{1}{\ln(2)} \ln(1 + \frac{1}{a(a+2)})$. ∎

5 Entropy

Entropy measures the randomness of a dynamical system. It gives an indication of how predictable a particular probability space $(X, \mathcal{A}, \mu)$ is when applying some map T. From our point of view entropy gives information about the predictability of digits in expansions.
In this section we slowly develop the definition of entropy, find a way to calculate entropy of our expansions and discuss some consequences.

5.1 Definition of entropy

Consider the probability space $(X, \mathcal{A}, \mu)$ and a set $A \subseteq X$ with positive measure. Let $\delta(A)$ denote the randomness of the event A. If $\mu(A) = 1$ we would like $\delta(A) = 0$, since there is no randomness in an event that has probability 1 of occuring. If on the other hand $\mu(A)$ is small, it is hard to predict event A and so $\delta(A)$ should be large. Apart from this, there is another property we would like δ to have. It feels natural to desire that if A and B are independent events, then $\delta(A \cap B) = \delta(A) + \delta(B)$.
Summarizing, $\delta(A)$ should be nonnegative, decreasing when the measure of A increases, and $\delta(A \cap B)$ should be equal to $\delta(A) + \delta(B)$ whenever A and B are independent. Note that the following definition provides those properties.

Definition 5.1.1. Let $(X, \mathcal{A}, \mu)$ be a probability space and $A \subseteq X$ be a set of positive measure. The *randomness of the event* A is defined by $\delta(A) = -\ln(\mu(A))$.

We now move on from randomness of a single set to randomness of a partition (up to null sets).

Definition 5.1.2. Consider the same probability space as before and a finite collection of pairwise disjoint events $\alpha = \{A_1, A_2, ..., A_n\}$ of positive measure such that $\mu(\bigcup\limits_{i=1}^{n} A_i) = 1$. That is, almost every $x \in X$ is an element of some A_i. Then the *randomness of the partition* α is $H(\alpha) = \sum\limits_{i=1}^{n} \mu(A_i) \cdot \delta(A_i) = -\sum\limits_{i=1}^{n} \mu(A_i) \cdot \ln(\mu(A_i))$.

This definition makes just as much sense as the previous one. If α consists of just 1 element A, then $\mu(A) = 1$ and $H(\alpha) = 0$. Note that in that case there is indeed no randomness. However, $H(\alpha)$ should be maximal when all events have equal probability: $A_i = \frac{1}{n}$ for every i. The way to check this is by finding $H(\alpha) = \ln(n)$ for the described partition. For an arbitrary partition $\{A_1, A_2, ..., A_n\}$ we obtain with help of the Jensen's inequality: $H(\{A_1, A_2, ..., A_n\}) = n \cdot \frac{1}{n} \cdot -\sum\limits_{i=1}^{n} \mu(A_i) \cdot \ln(\mu(A_i)) \leq n \cdot -\frac{1}{n} \sum\limits_{i=1}^{n} \mu(A_i) \cdot \ln(\frac{1}{n} \sum\limits_{i=1}^{n} \mu(A_i)) = -n \cdot \frac{1}{n} \cdot \ln(\frac{1}{n}) = \ln(n)$. We have just shown $H(\alpha)$ is maximal when the events have equal probability.

We are now ready to give the full definition of entropy.

Notation. For a partition $\alpha = \{A_1, ..., A_n\}$ we denote $T^{-i}(\alpha) = \{T^{-i}(A_1), ..., T^{-i}(A_n)\}$ and $\bigvee_{i=0}^{n-1} T^{-i}(\alpha) = \{A_{k_0} \cap ... \cap A_{k_{n-1}} : A_{k_j} \in T^{-j}(\alpha)\}$. Note that both are partitions and that the last one becomes finer (i.e. has more elements) when n gets larger.

Definition 5.1.3. Let α be a finite partition up to null sets and T be measure preserving. The *entropy of T with respect to α* is given by

$$h(\alpha, T) = \lim_{n\to\infty} \frac{1}{n} \cdot H(\bigvee_{i=0}^{n-1} T^{-i}(\alpha)).$$

Finally, the *entropy* of T, which is independent of the partition, is defined as

$$h(T) = \sup_{\alpha} h(\alpha, T).$$

5.2 Entropy of expansion systems

Although Definition 5.1.3 provides us with a way to calculate entropy, it forces us to find the supremum over all partitions, of which there are many. We have seen this kind of problem before when we were defining measurability and ergodicity. Similar to those problems we have a theorem for entropy based on a property of the generator in question.

Definition 5.2.1. Suppose $(X, \mathcal{A}, \mu, T)$ is a dynamical system and α is a partition of X up to null sets. Then we say α is a *generator* (of $\mathcal{A}$) with respect to T if $\sigma(\bigvee_{i=0}^{\infty} T^{-i}(\alpha)) = \mathcal{A}$.

So if α is such that the collection $\bigvee_{i=0}^{\infty} T^{-i}(\alpha)$ forms a generator of the σ-algebra, then α itself is called a generator. Taking an expansion map T and its corresponding partition α, note that the elements of $\bigvee_{i=0}^{\infty} T^{-i}(\alpha)$ are the cylinder sets of "infinite" rank.

The following theorem makes it somewhat easier to calculate entropy. For a proof see [7].

Theorem 5.2.2 (Kolmogorov, Sinai)**.** *With the notation as before, let α be a finite or countable generator with respect to T such that $H(\alpha) < \infty$. Then $h(T) = h(\alpha, T)$.*

Lemma 5.2.3. *The partition α obtained from any GLS-, β- or continued fractions expansion generates $\mathcal{B}([0, 1))$ with respect to the corresponding map T.*

Proof. Let $n \in \mathbb{N}$. Note that every cylinder of rank n is a countable union of elements from $\bigvee_{i=0}^{\infty} T^{-i}(\alpha)$. In the proofs of Theorems 4.1.4, 4.1.5 and 4.1.6 we have already shown that any open interval (a, b) is a countable union of cylinder

sets of rank n (the first condition of Knopp's Lemma). Lastly by definition the open intervals generate $\mathcal{B}([0,1))$.
Summarizing: $\bigvee_{i=0}^{\infty} T^{-i}(\alpha)$ generates the cylinder sets of rank n, cylinder sets of rank n generate the open intervals and the open intervals generate $\mathcal{B}([0,1))$. It follows directly that $\sigma(\bigvee_{i=0}^{\infty} T^{-i}(\alpha)) = \mathcal{B}([0,1))$. $\square$

The immediate consequence of Theorem 5.2.2 and Lemma 5.2.3 is that $h(T) = h(\alpha, T)$ for α and T related to our expansions. We will use this result later, but first we state another important theorem.

Theorem 5.2.4 (Shannon-McMillan-Breiman)**.** *Let $(X, \mathcal{A}, \mu, T)$ be an ergodic system and α a partition of X up to null sets for which $H(\alpha) < \infty$. Denote the element of $\bigvee_{i=0}^{n-1} T^{-i}(\alpha)$ that contains x by $\Delta_n(x)$. Then*

$$h(\alpha, T) = \lim_{n\to\infty} -\frac{1}{n} \ln(\mu(\Delta_n(x))) \quad a.e.$$

Proof. The proof is based on the method of [2] and also uses [1], proposition 4.19. We start with a few definitions:

- $H(\alpha|\beta) = -\sum_{A\in\alpha}\sum_{B\in\beta} \ln(\frac{\mu(A\cap B)}{\mu(B)}) \cdot \mu(A\cap B)$.
- $I_\alpha(x) = -\ln(\mu(\Delta_1(x))) = -\sum_{A\in\alpha} \mathbb{1}_A(x) \cdot \ln(\mu(A))$
- $I_{\alpha|\beta}(x) = -\sum_{A\in\alpha}\sum_{B\in\beta} \mathbb{1}_{A\cap B}(x) \cdot \ln(\frac{\mu(A\cap B)}{\mu(B)})$.
- $f_n(x) = I_{\alpha|\bigvee_{i=1}^{n} T^{-i}(\alpha)}(x)$

Using this notation we have

$$\begin{aligned}
I_{\bigvee_{i=0}^{n} T^{-i}(\alpha)}(x) &= -\ln(\mu(\Delta_n(x))) \\
&= -\ln(\mu(\bigvee_{i=1}^{n} T^{-i}(\Delta_1(x)) \cap \Delta_1(x))) \\
&= -\ln(\mu(\bigvee_{i=1}^{n} T^{-i}(\Delta_1(x))) \cdot \mu(\Delta_1(x)| \bigvee_{i=1}^{n} T^{-i}(\Delta_1(x)))) \\
&= -\ln(\mu(\bigvee_{i=1}^{n} T^{-i}(\Delta_1(x)))) - \ln(\mu(\Delta_1(x)| \bigvee_{i=1}^{n} T^{-i}(\Delta_1(x)))) \\
&= I_{\bigvee_{i=1}^{n} T^{-i}(\alpha)}(x) + I_{\alpha|\bigvee_{i=1}^{n} T^{-i}(\alpha)}(x) \\
&= I_{\bigvee_{i=0}^{n-1} T^{-i}(\alpha)}(T(x)) + f_n(x).
\end{aligned}$$

Note that repeating this argument gives us $I_{\bigvee_{i=0}^{n} T^{-i}(\alpha)}(x) = \sum_{i=0}^{n} f_{n-i}(T^i(x))$. Now letting $f(x) = \lim_{n\to\infty} f_n(x)$, we can write

$$\lim_{n\to\infty} -\frac{1}{n}\ln(\mu(\Delta_n(x))) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n} f(T^i(x)) + \frac{1}{n}\sum_{i=0}^{n}(f_{n-i} - f)(T^i(x)).$$

In the remainder of the proof, our goal is to show that the first term equals $h(\alpha, T)$ and the second term equals zero.
For the first term, we follow three steps in which we encounter some results of conditional entropy.
Step 1: Suppose β and γ are any partitions of X. Then

$$\begin{aligned}
H(\beta) + H(\gamma|\beta) &= -\sum_{B\in\beta}\mu(B)\cdot\ln(\mu(B)) - \sum_{B\in\beta}\sum_{C\in\gamma}\mu(B\cap C)\cdot\ln(\frac{\mu(B\cap C)}{\mu(B)}) \\
&= -\sum_{B\in\beta}\mu(B)\cdot\ln(\mu(B)) - \sum_{\beta}\sum_{\gamma}\mu(B\cap C)\cdot\ln(\mu(B\cap C)) \\
&+ \sum_{\beta}\sum_{\gamma}\mu(B\cap C)\cdot\ln(\mu(B)) \\
&= -\sum_{B\in\beta}\mu(B)\cdot\ln(\mu(B)) + H(\beta\vee\gamma) + \sum_{B\in\beta}\mu(B)\cdot\ln(\mu(B)) \\
&= H(\beta\vee\gamma).
\end{aligned}$$

Step 2: Specifically, it follows from this equality and the fact that μ is T-invariant that $H(\alpha|\bigvee_{i=1}^{k} T^{-i}(\alpha)) = H(\bigvee_{i=0}^{k} T^{-i}(\alpha)) - H(\bigvee_{i=1}^{k} T^{-i}(\alpha)) = H(\bigvee_{i=0}^{k} T^{-i}(\alpha)) - H(\bigvee_{i=0}^{k-1} T^{-i}(\alpha))$. If we sum over k now, we get $\sum_{k=1}^{n} H(\alpha|\bigvee_{i=1}^{k} T^{-i}(\alpha)) = H(\bigvee_{i=0}^{n} T^{-i}(\alpha)) - H(\alpha)$ because the right hand side is a telescoping sum. Finally since $H(\alpha) < \infty$,

$$\begin{aligned}
\lim_{n\to\infty} H(\alpha|\bigvee_{i=1}^{n} T^{-i}(\alpha)) &= \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} H(\alpha|\bigvee_{i=1}^{k} T^{-i}(\alpha)) \\
&= \lim_{n\to\infty}\frac{1}{n}H(\bigvee_{i=0}^{n} T^{-i}(\alpha)) - 0 \\
&= h(\alpha, T).
\end{aligned}$$

Step 3: When we apply the Ergodic Theorem to f and use step 2, we obtain

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n} f(T^i(x)) &= \int_X f(x)d\mu \\
&= \int_X \lim_{n\to\infty} I_{\alpha|\bigvee_{i=1}^{n} T^{-i}(\alpha)}(x)d\mu \\
&= \int_X -\sum_{A\in\alpha}\sum_{B\in\bigvee_{i=1}^{\infty} T^{-i}(\alpha)} \mathbb{1}_{A\cap B}(x)\cdot\ln(\frac{\mu(A\cap B)}{\mu(B)})d\mu \\
&= -\sum_{A\in\alpha}\sum_{B\in\bigvee_{i=1}^{\infty} T^{-i}(\alpha)} \int_X \mathbb{1}_{A\cap B}(x)\cdot\ln(\frac{\mu(A\cap B)}{\mu(B)})d\mu \\
&= -\sum_{A\in\alpha}\sum_{B\in\bigvee_{i=1}^{\infty} T^{-i}(\alpha)} \mu(A\cap B)\cdot\ln(\frac{\mu(A\cap B)}{\mu(B)}) \\
&= H(\alpha|\bigvee_{i=1}^{\infty} T^{-i}(\alpha)) = h(\alpha, T),
\end{aligned}$$

as desired.
What is left is to show that the sequence $\{\frac{1}{n}\sum_{i=0}^{n}(f_{n-i}-f)(T^i(x))\}_{n\in\mathbb{N}}$ converges to zero. Define $F_N = \sup_{n-k\geq N}|f_{n-k}-f|$, and note that $\lim_{N\to\infty} F_N = 0$. Note also that $|f_k - f|$ is integrable for all k because f and all f_k are integrable.
Now let $\epsilon > 0$. Then there exists an N such that $F_N(x) < \epsilon$ for almost every x, and so for $n > N$,

$$\begin{aligned}
\frac{1}{n}\sum_{i=0}^{n}(f_{n-i}-f)(T^i(x)) &\leq \frac{1}{n}\sum_{i=0}^{n}|f_{n-i}-f|(T^i(x)) \\
&= \frac{1}{n}\sum_{i=0}^{n-N}|f_{n-i}-f|(T^i(x)) + \frac{1}{n}\sum_{i=n-N+1}^{n}|f_{n-i}-f|(T^i(x)) \\
&\leq \frac{1}{n}\sum_{i=0}^{n-N} F_N(T^i(x)) + \frac{1}{n}\sum_{i=0}^{N-1}|f_i - f|(T^{n-i}(x)).
\end{aligned}$$

For the left hand side, we have $\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-N} F_N(T^i(x)) = \int_X F_N d\mu < \int_X \epsilon d\mu = \epsilon$ almost everywhere. For the right hand side note that $\sum_{i=0}^{N-1}|f_i - f|$ is still integrable, and hence $\sum_{i=0}^{N-1}|f_i - f|(x)$ must be finite almost everywhere. The same then holds for $\sum_{i=0}^{N-1}|f_i - f|(T^{n-i}(x))$ since T is measure preserving. It follows that $\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{N-1}|f_i - f|(T^{n-i}(x)) = 0$ almost everywhere, which completes the proof. □

Remark. In words, the Shannon-McMillan-Breiman Theorem states that the entropy of a partition is related to the limiting size of a cylinder set created by that partition. This should not feel very strange because cylinder sets give information about the occurring events of the partition.

We just need one more lemma before we can calculate the entropy of our expansions. This lemma gives us the opportunity to use the Lebesgue measure when applying Theorem 5.2.4, a much "nicer" one than the invariant measure of the β-expansion or the continued fractions expansion.

Lemma 5.2.5. *Let θ be an invariant measure for an expansion map T_θ. If there exist a and b such that $a \cdot \theta(\Delta_n(x)) \leq \lambda(\Delta_n(x)) \leq b \cdot \theta(\Delta_n(x))$, then $\lim_{n\to\infty} \frac{\ln(\lambda(\Delta_n(x)))}{\ln(\theta(\Delta_n(x)))} = 1$.*

Proof. Taking the natural logarithm of all three sides and dividing by $\ln(\theta(\Delta_n(x)))$, we have

$$\frac{\ln(a) + \ln(\theta(\Delta_n(x)))}{\ln(\theta(\Delta_n(x)))} \leq \frac{\ln(\lambda(\Delta_n(x)))}{\ln(\theta(\Delta_n(x)))} \leq \frac{\ln(b) + \ln(\theta(\Delta_n(x)))}{\ln(\theta(\Delta_n(x)))}.$$

The result follows when $n \to \infty$ since $\lim_{n\to\infty} \ln(\theta(\Delta_n(x))) = \infty$. □

We proceed by calculating entropy of our expansion systems, using all of the previous theorems and lemmas of this subsection.

Corollary 5.2.6. *Suppose $\alpha = \{[l_i, r_i)\}_{i\in I}$ $(I \subseteq \mathbb{N} \cup \{0\})$ is a GLS partition, and denote $L_i = r_i - l_i$. Then $h(T_G) = -\sum_{i\in I} L_i \cdot \ln(L_i)$.*

Proof. From Corollary 4.2.5 it follows that when $n \to \infty$, the proportion of times that $T^{-k}(x) \in [l_i, r_i)$ (with $k \leq n$) is nL_i for almost every x. Thus,

$$\begin{aligned}\lim_{n\to\infty} \lambda(\Delta_n(x)) &= L_0^{nL_0} \cdot L_1^{nL_1} \cdot ... \\ &= e^{\ln(L_0)\cdot nL_0} \cdot e^{\ln(L_m)\cdot nL_m} \cdot ... \\ &= e^{n\sum_{i\in I} L_i \ln(L_i)},\end{aligned}$$

and $\lim_{n\to\infty} -\frac{1}{n}\ln(\lambda(\Delta_n(x))) = -\sum_{i\in I} L_i \ln(L_i)$. We can combine this equality with Theorem 5.2.2, Lemma 5.2.3 and Theorem 5.2.4 to get

$$h(T_G) = h(\alpha, T_G) = \lim_{n\to\infty} -\frac{1}{n}\ln(\lambda(\Delta_n(x))) = -\sum_{i\in I} L_i \ln(L_i)$$

. □

Example 5.1. The n-ary expansion has entropy $h(T_n) = -\sum_{i=0}^{n-1} \frac{1}{n} \cdot \ln(\frac{1}{n}) = -n \cdot \frac{1}{n} \cdot -\ln(n) = \ln(n)$.

Corollary 5.2.7. *For $\beta > 1$ and α the usual partition for a β-expansion, $h(T_\beta) = \ln(\beta)$.*

Proof. We first show that the corresponding invariant measure ν satisfies the condition of Lemma 5.2.5. Note that the density function of λ is constant, and the (unnormalized) density function of ν, $h_\beta(x) = \sum_{n:x<T_\beta^n(1)} \frac{1}{\beta^n}$, satisfies $1 < h_\beta(x) < \infty$ because $x < 1$ and $\beta > 1$. That means we can always find a and b such that $a \cdot \nu(\Delta_n(x)) \leq \lambda(\Delta_n(x)) \leq b \cdot \nu(\Delta_n(x))$, so Lemma 5.2.5 holds.

From the proof of Theorem 4.1.5, we know that every open subinterval of $[0, 1)$ can be written as a union of countably many full cylinders. Therefore full cylinders generate $\mathcal{B}([0, 1))$, and we can assume $\Delta_n(x)$ is full. Note that now $\lambda(\Delta_n(x)) = \beta^{-n}$.

Concluding, Theorems 5.2.2, 5.2.4 and Lemmas 5.2.3, 5.2.5 imply

$$\begin{aligned}
h(T_\beta) &= h(\alpha, T_\beta)\\
&= \lim_{n\to\infty} -\frac{1}{n}\ln(\nu(\Delta_n(x)))\\
&= \lim_{n\to\infty} -\frac{1}{n}\ln(\lambda(\Delta_n(x)))\\
&= \lim_{n\to\infty} -\frac{1}{n}\ln(\beta^{-n})\\
&= \lim_{n\to\infty} -\frac{n}{n}\cdot -\ln(\beta)\\
&= \ln(\beta).
\end{aligned}$$

□

Finding the entropy of the continued fraction map is slightly more complicated, as the size of its cylinder sets is not that clear. We will use the following result. A proof can be found in [4].

Theorem 5.2.8 (Lévy)**.** *Let $\Delta_n(x)$ be the cylinder set of rank n containing x with respect to the continued fractions expansion. Then*

$$\lim_{n\to\infty} \frac{1}{n}\ln(\lambda(\Delta_n(x))) = \frac{-\pi^2}{6\ln(2)} \quad a.e.$$

Corollary 5.2.9. *If α is the continued fractions partition, then $h(T_c) = \frac{\pi^2}{6\ln(2)}$.*

Proof. The inequality $\frac{1}{2\ln(2)}\lambda(A) \leq \mu(A) \leq \frac{1}{\ln(2)}\lambda(A)$ obtained from the proof of Theorem 4.1.6 yields

$$\ln(2)\mu(\Delta_n(x)) \leq \lambda(\Delta_n(x)) \leq 2\ln(2)\mu(\Delta_n(x))$$

for $A = \Delta_n(x)$. Therefore we can use Lemma 5.2.5 with μ the invariant measure of T_c.
Now because of Theorems 5.2.2, 5.2.4, 5.2.8 and Lemmas 5.2.3, 5.2.5 we have

$$h(T_c) = h(\alpha, T_c) = \lim_{n\to\infty} -\frac{1}{n}\ln(\mu(\Delta_n(x))) = \lim_{n\to\infty} -\frac{1}{n}\ln(\lambda(\Delta_n(x))) = \frac{\pi^2}{6\ln(2)}.$$

□

We end this section with a comparison between the decimal expansion and the continued fractions expansion.

Theorem 5.2.10 (Lochs)**.** *Suppose x has decimal expansion $0.d_1d_2...$ and its continued fractions expansion is represented by $(a_1, a_2, ...)$. Consider the number $x' = 0.d_1d_2...d_n$ with representation $(c_1, c_2, ..., c_k)$ of its continued fractions expansion. Define $m(n, x)$ as the largest integer such that $c_1 = a_1, ..., c_m = a_m$. Then*

$$\lim_{n\to\infty} \frac{m(n,x)}{n} = \frac{6\ln(2)\ln(10)}{\pi^2} \quad a.e.$$

Proof. Denote $\Delta_n = \Delta(a_1, ..., a_{n-1}, a_n)$ as a cylinder set of rank n with respect to the continued fractions expansion, and define $\Delta_n^+ = \Delta(a_1, ..., a_{n-1}, a_n + 1)$. A famous result in number theory is the recursion $q_n = aq_{n-1} + q_{n-2}$, where q_n is the denominator of the n-th partial continued fraction. Note that for any $y \in \Delta_n$ and $z \in \Delta_n^+$, we have $q_{n-2}(z) = q_{n-2}(y)$, $q_{n-1}(z) = q_{n-1}(y)$ and $a_n(z) = a_n(y) + 1$. Taking $q_k(y) := q_k$, we obtain

$$\begin{aligned}
\frac{\lambda(\Delta_n)}{\lambda(\Delta_n^+)} &= \frac{\frac{1}{q_n(q_n+q_{n-1})}}{\frac{1}{q_n(z)(q_n(z)+q_{n-1})}} = \frac{q_n(z)(q_n(z)+q_{n-1})}{q_n(q_n+q_{n-1})} \\
&= \frac{(a_n(z)q_{n-1}(z) + q_{n-2}(z)) \cdot (a_n(z)q_{n-1}(z) + q_{n-2}(z) + q_{n-1})}{(a_nq_{n-1} + q_{n-2}) \cdot (a_nq_{n-1} + q_{n-2} + q_{n-1})} \\
&= \frac{((a_n+2)q_{n-1} + q_{n-2}) \cdot ((a_n+1)q_{n-1} + q_{n-2})}{(a_nq_{n-1} + q_{n-2}) \cdot ((a_n+1)q_{n-1} + q_{n-2})} \\
&= 1 + \frac{2q_{n-1}}{a_nq_{n-1} + q_{n-2}} \leq 3.
\end{aligned}$$

The previous inequality will come in handy later in the proof. First, since T_c is almost everywhere decreasing, $T_c^2 = T_c(T_c)$ is increasing a.e., $T_c^3 = T_c(T_c^2)$ is decreasing a.e. and so on. Thus almost everywhere, T_c^n increases if n is even and decreases if n is odd. Consequently, the chaotic part of T_c^{n+1} (visible in Figure 3.4) happens on the left side of Δ_n if n is even and on the right side of Δ_n if n is odd. Now let I be an interval in $[0, 1)$ and Δ_n the smallest cylinder set containing I. It follows that for almost all $x \in I$, there is a $1 \leq j \leq 3$ such that either $\Delta_{n+j}(x) \subseteq I$ or $\Delta_{n+j}(x)^+ \subseteq I$. To be clear here, $x \in \Delta_{n+j}(x)$ and not $x \in \Delta_{n+j}(x)^+$.
In particular, if $D_n(x)$ is the decimal cylinder of rank n containing x, then $I = D_n(x)$ yields $\Delta_{m+j} \subseteq D_n(x) \subseteq \Delta_m(x)$ for some $1 \leq j \leq 3$, where Δ_{m+j} equals either $\Delta_{m+j}(x)$ or $\Delta_{m+j}(x)^+$. By $\lambda(\Delta_{m+j}) \leq 3\lambda(\Delta_{m+j}^+)$ we have

$$\frac{1}{3}\lambda(\Delta_{m+j}(x)) \leq \lambda(D_n(x)) \leq \lambda(\Delta_m(x)),$$

and

$$\frac{1}{m}\ln(\frac{1}{3}) + \frac{1}{m}\ln(\lambda(\Delta_{m+j}(x))) \leq \frac{n}{m}\frac{1}{n}\ln(\lambda(D_n(x))) \leq \frac{1}{m}\ln(\lambda(\Delta_m(x))).$$

From now on suppose $n \to \infty$. In that case clearly also $m \to \infty$. We can apply Theorems 5.2.2, 5.2.4 and Lemmas 5.2.3, 5.2.5 to get $h(T_c) \leq \frac{n}{m}h(T_{10}) \leq h(T_c)$. It follows that $h(T_c) = \frac{n}{m}h(T_{10})$, and we conclude from Example 5.1 and Corollary 5.2.9 that

$$\lim_{n\to\infty} \frac{m}{n} = \frac{h(T_{10})}{h(T_c)} = \frac{6\ln(2)\ln(10)}{\pi^2} \quad a.e.$$

□

Because $\frac{6\ln(2)\ln(10)}{\pi^2} \approx 0.97$, we can say that the average amount of information per digit of the continued fractions expansion is slightly higher than that of the decimal expansion. In other words, on average, 97 continued fraction digits approximately give the same information about a number as 100 decimal digits.

Without extra effort, Lochs' Theorem can be generalized for arbitrary n-ary expansions. In fact, it is possible to prove more of such comparison theorems considering many different expansions. Those theorems are beyond the scope of this thesis, but they can be found in [3].

References

[1] Barney Bramham. *Notes for the course: Ergodic Theory and Entropy*, (Bochum, 2014).

[2] K. Dajani and S. Dirksin. *A Simple Introduction to Ergodic Theory*, (2008).

[3] K. Dajani and F. Fieldsteel. *Equipartition of interval partitions and an application to number theory.* Proceedings of the American Mathematical Society, Volume 129 (2001), Number 12, 3453-3460.

[4] K. Dajani and C. Kraaikamp. *Ergodic Theory of Numbers*. The Mathematical Association of America , 2002.

[5] J. F. C. Kingman and S. J. Taylor. *Introduction to measure and probability.* Cambridge University Press, 1966.

[6] W. Parry. *On the β-expansions of Real Numbers*, (London, 1960).

[7] Karl Petersen. *Ergodic theory.* Cambridge University Press, 1989.

[8] René L. Schillings. *Measures, Integrals and Martingales.* Cambridge University Press, 2005

[9] Keith M. Wilkinson. *Ergodic Properties of a Class of Piecewise Linear Transformations.* Springer-Verlag, 1975.