**Matching, entropy, holes and expansions**
Langeveld, N.D.S.

**Author**: Langeveld, N.D.S.
**Title:** Matching, entropy, holes and expansions
**Issue Date**: 2019-12-10

CHAPTER 1

# Introduction

In this dissertation, the main topic is representations of numbers in relation to matching, entropy and holes. In this chapter, we first briefly discuss the representations that are studied in this dissertation. This is followed by a section in which we introduce the basic definitions used in dynamical systems and ergodic theory. This is done by means of the regular continued fraction. Other representations which we study are $\beta$-expansions. These are explained in the section thereafter. In the last part of the introduction we explain the words in the title and elaborate more on what to find in which chapter.

## §1.1 Representing numbers

In general we express one number by using others. There is a family of numbers which we know how to use because we use them to count: the natural numbers. The numbers of the set $B := \{\frac{1}{n} : n \in \mathbb{N}\}$[1] are also natural to use and, since the $n$-fold of them equals 1, they are still related to counting. Now suppose we have a number $x$ between 0 and 1. We try to express $x$ using these numbers. If $x \notin B$ we can only approximate $x$ by a number $\frac{1}{n}$ such that the error is small. If we want to express $x$ with elements from $B$ without an error we should not stop here but continue! We can do two things. Either write $x = \frac{1}{n} + \varepsilon$ or $x = \frac{1}{n+\varepsilon}$ for some $\varepsilon > 0$. We can proceed with $\varepsilon$ and find an $m \in \mathbb{N}$ such that $\varepsilon$ is close to $\frac{1}{m}$ and continue in this manner. The first case corresponds to Lüroth expansions which are introduced in [78] and widely studied thereafter (see for example [6, 40, 51]). Our interest lies in the second way which leads to continued fractions. We obtain a continued fraction expansion for $x$ by using the Gauss map $T : [0,1] \to [0,1]$ which is defined by

$$T(x) = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor$$

for $x \neq 0$ and $T(0) = 0$, see Figure 1.1. Let the digits be defined as $d_1(x) = \lfloor \frac{1}{x} \rfloor$ and $d_n(x) = d_1\left(T^{n-1}(x)\right)$ for $n > 1$.

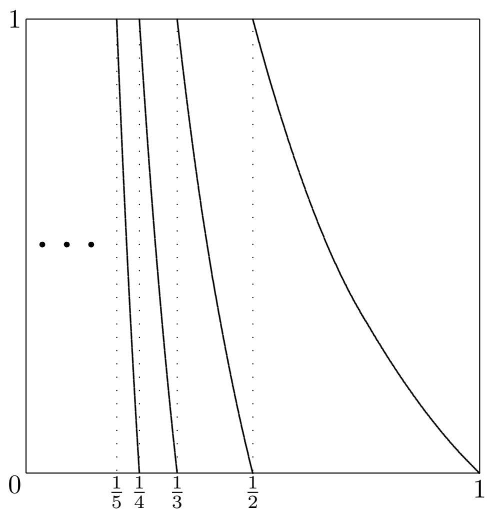


*Figure 1.1: The Gauss map.*

[1] In this dissertation 0 is not included in the set of natural numbers.

For $x \in [0,1]$ we find

$$x = \frac{1}{d_1(x) + T(x)} = \cfrac{1}{d_1(x) + \cfrac{1}{d_2(x) + T^2(x)}} = \cfrac{1}{d_1(x) + \cfrac{1}{d_2(x) + \cfrac{1}{d_3(x) + \ddots}}}$$

with $d_n \in \mathbb{N}$. We can write this in short notation as $x = [0; d_1, d_2, d_3, \ldots]$. Examples of such expansions are $e - 2 = [0; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, \ldots]$ or $\sqrt{2} - 1 = [0; 2, 2, 2, \ldots]$. Any $x \in (0,1]$ has such an expansion. For rational numbers one finds a finite continued fraction and for irrational numbers one finds an infinite continued fraction. For convergence and other basic properties of this representation see [29]. In Chapter 2, 3 and 4 variations of $T$ will be studied.

Suppose that, instead of taking the natural numbers as a given, we take a value $\frac{1}{\beta}$ with $\beta > 1$. Now we approximate numbers in $[0,1]$ by numbers of the form $\frac{m}{\beta^n}$ so that $x = \frac{m}{\beta^n} + \varepsilon$ with $n, m \in \mathbb{N}$. We do this in the following way. We first pick the smallest $n \in \mathbb{N}$ such that $\frac{1}{\beta^n} < x$ and then we take $m \in \{1, \ldots, \lfloor \beta x \rfloor\}$ maximal such that $\frac{m}{\beta^n} < x$. Then we procceed the procedure applied on $\varepsilon = x - \frac{m}{\beta^n}$. We can do this dynamically with the function $T_\beta : [0,1] \to [0,1]$ defined by $T_\beta(x) = \beta x - \lfloor \beta x \rfloor$. For an example see Figure 1.2. Now we set $d_1(x) = \lfloor \beta x \rfloor$ and $d_n = d_1\left(T^{n-1}(x)\right)$ for $n > 1$. This give us

$$x = \frac{d_1(x) + T_\beta(x)}{\beta} = \frac{d_1(x)}{\beta} + \frac{d_2(x) + T_\beta^2(x)}{\beta^2} = \frac{d_1(x)}{\beta} + \frac{d_2(x)}{\beta^2} + \frac{d_3(x)}{\beta^3} + \ldots.$$

Convergence of this representation is immediately clear since, when taking the first $n$ digits, we are at most $\frac{1}{\beta^n}$ away from $x$. The $\beta$-expansions are studied in Chapter 5. In Section 5.6.1 we see some relation to continued fractions.

## §1.1.1 Continued fractions

In this section we introduce some basic notions and results concerning continued fractions. Along the way we encounter concepts that are prominent in ergodic theory. Because of the introductory nature of this section, all the results presented in this section can also be found in [29]. Let $x \in (0,1)$ with

$$x = \cfrac{1}{d_1 + \cfrac{1}{d_2 + \cfrac{1}{d_3 + \ddots}}}.$$

We define the $n^{\text{th}}$ convergent of $x$ as

$$c_n = \frac{p_n}{q_n} = \cfrac{1}{d_1 + \cfrac{1}{d_2 + \cfrac{1}{\ddots + \cfrac{1}{d_n}}}}.$$

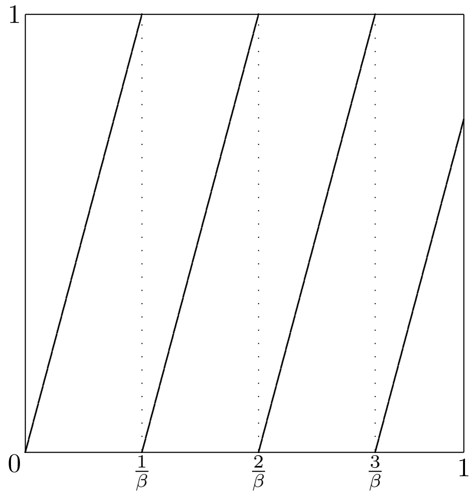


*Figure 1.2: The $\beta$-transformation with $\beta = 3.76$.*

For $p_n$ and $q_n$ we have the following recurrence relations

$$\begin{array}{llll} p_{-1} := 1; & p_0 := 0; & p_n = d_n p_{n-1} + p_{n-2}, & n \geq 1, \\ q_{-1} := 0; & q_0 := 1; & q_n = d_n q_{n-1} + q_{n-2}, & n \geq 1. \end{array}$$

The fact that $\lim_{n\to\infty} c_n = x$ follows from

$$\left| x - \frac{p_n}{q_n} \right| \leq \frac{1}{q_n^2} \tag{1.1.1}$$

and the fact that the sequence $(q_n)_{n\in\mathbb{N}}$ grows exponentially fast. A classical motivation to study continued fractions comes from approximation theory also known as Diophantine approximation. This name stems from Diophantus of Alexandria, who lived around AD 250. Let $x \in [0,1]$ and suppose that we want to find rationals $\frac{p}{q}$ such that $|x - \frac{p}{q}|$ is small. Of course for $q$ large one can probably do better. Therefore it is natural to make $|x - \frac{p}{q}|$ small relative to $q$. Hurwitz proved the following theorem in 1891.

**Theorem 1.1.1 (Hurwitz [49]).** *For every irrational number $x$ there exist infinitely many pairs of integers $p$ and $q$, such that*

$$\left| x - \frac{p}{q} \right| \leq \frac{1}{\sqrt{5}} \frac{1}{q^2}. \tag{1.1.2}$$

*The constant $\frac{1}{\sqrt{5}}$ is the best possible, i.e. for every $\varepsilon > 0$ there are $x$ such that there are only finitely many pairs of integers $p$ and $q$ such that the inequality holds when replacing $\frac{1}{\sqrt{5}}$ by $\frac{1}{\sqrt{5}} - \varepsilon$.*

To be able to find such pairs one can look at the convergents of $x$. This is displayed by a theorem of Borel from 1903.

**Theorem 1.1.2 (Borel).** *Let $n \geq 1$ and let $\frac{p_{n-1}}{q_{n-1}}, \frac{p_n}{q_n}$ and $\frac{p_{n+1}}{q_{n+1}}$ be three consecutive continued fraction convergents of the irrational number $x$. Then at least one of these convergents satisfies (1.1.2).*

There also exists a theorem that states that, when a rational approximates $x$ well, it is a convergent of $x$. This was shown by Legendre in 1798.

**Theorem 1.1.3 (Legendre [70]).** *Let $p$ and $q$ be two integers that are co-prime with $q > 0$. Furthermore, let $x \in (0,1]$ and suppose that $\left|x - \frac{p}{q}\right| \leq \frac{1}{2q^2}$. Then $\frac{p}{q}$ is a convergent of $x$.*

For a refinement of this theorem see Barbolosi, Jager 1994 [5]. Looking at the recurrence relations and (1.1.1) we can see that the higher the digits of $x$ are, the faster the continued fraction converges to $x$. For a given $x \in (0,1]$ we can simply calculate the convergents. However, we would like to make statements about *typical* points $x$ i.e. statements that hold for *almost all* $x \in (0,1]$. This is where ergodic theory comes into play. The word ergodic originates from the words ergon and odos which mean work and path respectively in Greek. Ergodic theory was used by physicists before mathematicians picked up on it in the 1930s and 1940s (from [29]). Let us first define what a dynamical system is and then give the definition of ergodicity.

**Definition 1.1.4 (Dynamical system).** *A dynamical system is a quadruple $(X, \mathcal{F}, \mu, T)$ where $X$ is a non-empty set, $\mathcal{F}$ is a $\sigma$-algebra on $X$, $\mu$ is a probability measure on $(X, \mathcal{F})$ and $T : X \to X$ is a surjective transformation such that the measure $\mu$ is $T$-invariant i.e. for all $A \in \mathcal{F}$ we have $\mu(T^{-1}(A)) = \mu(A)$. Furthermore, if $T$ is also injective we call $(X, \mathcal{F}, \mu, T)$ an invertible dynamical system.*

When dropping the condition of a probability measure and allowing the space to have an infinite measure one enters the realm of infinite ergodic theory which is studied in Chapter 2. In any case ergodic theory is characterised by ergodicity.

**Definition 1.1.5 (Ergodicity).** *Let $(X, \mathcal{F}, \mu, T)$ be a dynamical system. Then $T$ is called ergodic if for every $\mu$-measurable set $A$ satisfying $T^{-1}(A) = A$ one has $\mu(A) = 0$ or $\mu(A^c) = 0$.*

This means that, when iterating points, they will go from everywhere to everywhere and the state space $X$ cannot be divided into subsets $X_1, X_2$ with both positive measure such that $T(X_1) \subset X_1$ and $T(X_2) \subset X_2$. It is natural to wonder when two dynamical systems can be called the same. We say such maps are isomorphic.

**Definition 1.1.6 (Isomorphic).** *Two dynamical systems $(X, \mathcal{F}, \mu, T)$ and $(Y, \mathcal{C}, \nu, S)$ are isomorphic if there exists a map $\theta : X \to Y$ with the following properties.*

- *$\theta$ is bijective almost everywhere. By this we mean that, if we remove a suitable set $N_X \subset X$ with $\mu(N_X) = 0$ and a suitable set $N_Y \subset Y$ with $\nu(N_Y) = 0$ then $\theta : X \backslash N_X \to Y \backslash N_Y$ is a bijection.*
- *$\theta$ is bi-measurable, i.e. $\theta(F) \in \mathcal{C}$, for all $F \in \mathcal{F}$ and $\theta^{-1}(C) \in \mathcal{F}$, for all $C \in \mathcal{C}$.*

- *$\theta$ preserves the measures, i.e. $\nu(C) = (\nu \circ \theta^{-1})(C)$ for all $C \in \mathcal{C}$.*
- *$\theta$ preserves the dynamics, i.e. $\theta \circ T = S \circ \theta$.*

Before stating what we mean by "for almost all" we give the definition of the Lebesgue measure and the definition of an absolutely continuous measure.

**Definition 1.1.7 (Lebesgue measure).** *Let $[a, b]$ be an interval on the real line. The Borel $\sigma$-algebra is the $\sigma$-algebra generated by open intervals. The Lebesgue measure $\lambda$ is the measure such that $\lambda((c, d)) = d - c$ for all open intervals $(c, d) \subset [a, b]$.*

The Lebesgue measure is the only measure that is translation invariant. All measures studied in this dissertation are equivalent to Lebesgue. These kinds of measures are considered to be physically most relevant because they describe the statistical properties of forward orbits of a set of points with positive Lebesgue measure.

**Definition 1.1.8 (Absolutely continuous and equivalence of measures).** *Let $(X, \mathcal{F})$ be a measurable space and $\mu, \nu$ two measures on this space. The measure $\mu$ is absolutely continuous with respect to measure $\nu$ if $\nu(A) = 0$ implies $\mu(A) = 0$. Furthermore, if also $\nu$ is absolutely continuous with respect to measure $\mu$ we say that the measures are equivalent.*

Equivalence implies that if $\nu(A) = 1$ then $\mu(A) = 1$ whenever $\nu$ and $\mu$ are probability measures. With for "almost all $x \in X$" we mean with probability 1. One can see that it does not matter whether we use measure $\mu$ or $\nu$ for this statement when $\mu$ is absolutely continuous with respect to $\nu$. This way "for almost all" (or for almost every) $x$ means with respect to Lebesgue in this dissertation.

Now that we know what for almost all means we can state a theorem by Paul Lévy from 1929 that gives us the speed at which $q_n$ grows for almost all $x \in [0, 1]$.

**Theorem 1.1.9 (Lévy [72]).** *For almost all $x \in [0, 1]$ one has*

$$\lim_{n\to\infty} \frac{1}{n} \log(q_n) = \frac{\pi^2}{12 \log(2)}.$$

The fact that invariant measures are useful follows from what is the most important theorem of the field.

**Theorem 1.1.10 (The Ergodic Theorem / Birkhoff's Theorem).**
*Let $(X, \mathcal{F}, \mu)$ be a probability space and $T : X \to X$ such that $\mu$ is $T$-invariant. Then for any $f \in L^1(\mu)$,*

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

*exists almost everywhere and $\int_X f d\mu = \int_X f^* d\mu$. If moreover $T$ is ergodic, then $f^*$ is constant almost everywhere and $f^* = \int_X f d\mu$.*

This theorem is often heuristically phrased as "time average is space average" and has been proved by G.D Birkhoff in 1931 (see [8]). A question that arises is whether we can find an invariant measure for the Gauss map. The answer is yes. An invariant measure was found by Gauss in 1800. Note that this was way before most tools in ergodic theory were developed. The measure $\mu$ found by Gauss is called the Gauss measure and is given by

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

An example of what one can do with this invariant measure and Birkhoff's Theorem is to calculate frequencies of digits for typical numbers. Let $freq(i)$ be defined as

$$freq(i) := \lim_{n\to\infty} \frac{\#\text{ digits of } x \text{ equal to } i \text{ in the first } n \text{ digits}}{n}.$$

Let us also define cylinders (of order $n$)

$$\Delta(a_1, \dots a_n) = \{x \in [0,1] : d_1(x) = a_1, d_2(x) = a_2, \dots, d_n(x) = a_n\}.$$

Note that whenever $d_n(x) = i$ that $T^{n-1}(x) \in \Delta(i)$ and that $\Delta(i) = \{x \in [0,1] : d_1(x) = \lfloor \frac{1}{x} \rfloor = i\} = (\frac{1}{i+1}, \frac{1}{i}]$. Since the map $T$ is ergodic with respect to $\mu$ and $\mu$ is invariant for $T$, we can apply the Ergodic Theorem with $f = \mathbb{1}_{\Delta(i)}$ giving

$$freq(i) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n-1} f \circ T^i(x) = \frac{1}{\log(2)} \int_{\Delta(i)} \frac{1}{1+x} dx = \mu(\Delta(i)).$$

The frequencies of digits where found by Paul Lévy in 1929 (see [72]) and are given by

$$freq(i) = \mu(\Delta(i)) = \frac{1}{\log(2)} \log\left(1 + \frac{1}{i(i+2)}\right).$$

Remarkably one can find the Gauss measure through a limit of the Lebesgue measure. This is shown by the Gauss-Kuzmin-Lévy Theorem. This theorem states that the Lebesgue measure of the pre-images of a measurable set $A$ will converge to the Gauss measure i.e.

$$\lambda\left(T^{-n}(A)\right) \to \mu(A) \quad \text{as } n \to \infty. \tag{1.1.3}$$

This was stated as a hypothesis by Gauss in his mathematical diary in 1800 and proved by Kuzmin in 1928 who also obtained a bound on the speed of convergence. Independently, Lévy proved the same theorem in 1929 but found a sharper bound for the speed of convergence namely $|\lambda\,(T^{-n}(A)) - \mu(A)| = \mathcal{O}(q^n)$ with $0 < q < 1$ instead of $\mathcal{O}(q^{\sqrt{n}})$ which is the bound Kuzmin found. In [99] it is shown that (1.1.3) holds for a family of mappings $T$. In Chapter 4 we will base a numerical method on this theorem to get good estimates on invariant measures for other maps.

A very powerful tool in ergodic theory is that of a natural extension. The idea behind it is that you make a non-invertible system into an invertible one by adding dimensions. For the invertible system it can be easier to guess the invariant measure. Then one can find the invariant measure of the original system by projecting it.

**Definition 1.1.11 (Natural Extension).** *Let $(X, \mathcal{F}, \mu, T)$ be a dynamical system with $T$ a non-invertible transformation. An invertible dynamical system $(Y, \mathcal{C}, \nu, S)$ is called a natural extension of $(X, \mathcal{F}, \mu, T)$ if there exist two sets $F \in \mathcal{F}$ and $C \in \mathcal{C}$ and a function $\theta : C \to F$, such that the following properties hold.*

- $\mu(X \backslash F) = \nu(Y \backslash C) = 0,$
- $T(F) \subset F$ *and* $S(C) \subset C,$
- $\theta$ *is measurable, measure preserving and surjective,*
- $(\theta \circ S)(y) = (T \circ \theta)(y)$ *for all* $y \in C,$
- $\bigvee_{k=0}^{\infty} S^k \theta^{-1}(\mathcal{F}) = \mathcal{C}$ *where* $\bigvee_{k=0}^{\infty} S^k \theta^{-1}(\mathcal{F})$ *is the smallest $\sigma$-algebra containing all $\sigma$-algebras* $S^k \theta^{-1}(\mathcal{F})$.

Natural extensions are unique up to isomorphism and therefore we can speak of *the natural extension*. Let $\Omega = [0,1] \times [0,1]$. The natural extension of the Gauss map is given by $\mathcal{T} : \Omega \to \Omega$ with

$$\mathcal{T}(x, y) := \left( T(x), \frac{1}{d_1(x) + y} \right), \quad (x, y) \in \Omega.$$

The natural extension captures information about the future in the first dimension and of the past in the second. The following theorem gives us the invariant measure as well as ergodicity for the natural extension of the Gauss map.

**Theorem 1.1.12 (Ito, Nakada, Tanaka [82, 83]).** *Let $\bar{\mu}$ be the measure given by*

$$\bar{\mu}(A) = \frac{1}{\log(2)} \int_A \frac{1}{(1 + xy)^2} \, d\lambda(x)$$

*then $\bar{\mu}$ is an invariant probability measure for $\mathcal{T}$. Furthermore, the dynamical system $(\Omega, \mathcal{B}, \bar{\mu}, \mathcal{T})$ where $\mathcal{B}$ is the Borel $\sigma$-algebra, is an ergodic system.*

The natural extension is also used in approximation theory to get information about the quality of convergents (see [29] and the references therein). We will use the concept of natural extensions in Chapter 2 and 4.

Another notion that can be useful is that of an *induced transformation*. Let $(X, \mathcal{F}, \mu, T)$ be dynamical system and pick $A \subset X$ such that $\mu(A) > 0$. Let $n(x) := \inf\{n \geq 1 : T^n(x) \in A\}$. By the Poincare Recurrence Theorem we have that the set of $x$ for which $n(x) = \infty$ has zero measure. We remove this set from $A$ and define the induced transformation $T_A : A \to A$ as

$$T_A(x) = T^{n(x)}(x) \quad \text{for all } x \in A.$$

## §1.1.2 β-expansions

In the previous section we have seen how to write a number in its continued fraction expansion. In this section we shed light upon $\beta$-expansions. These are derived from a much simpler map $T_\beta : [0,1] \to [0,1]$ with $T_\beta(x) = \beta x - \lfloor \beta x \rfloor$ with $\beta \in (1, \infty)$. Fix $x \in [0,1]$ and set $d_1(x) = \lfloor \beta x \rfloor$ and $d_n(x) = d_1(T_\beta^{n-1}(x))$ for $n > 1$. Then for $x$ we find

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

In this case we define the convergents as $c_n = \sum_{i=1}^{n} \frac{d_i}{\beta^i}$. The convergence rate is given by $|x - c_n| \leq \frac{1}{\beta^n}$. Note that whenever $x$ has long sequences of zeros in its expansion there are $c_n$ that are fairly close to $x$ relative to $n$. On the other hand, the sequence $(d_i(x))$ is not the only sequence that will give convergence to $x$. We see that high digits result in better convergents. Since we always use the highest digit possible by taking $\lfloor \beta x \rfloor$, the map $T_\beta$ is known as the greedy $\beta$-transformation (introduced by Rényi in 1957 [95]). When instead of always taking the highest digit possible one would always take the lowest, one finds the lazy $\beta$-expansion. Fix $\beta \in (1, \infty)$ and define $S = \cup_{1 \leq i \leq \lfloor \beta \rfloor} [\frac{i}{\beta}, \frac{\lfloor \beta \rfloor}{\beta(\beta-1)} + \frac{i-1}{\beta}]$. The map used to find a lazy $\beta$-expansion is given by $L_\beta(x) = T_\beta(x)$ for $x \notin S$ and $L_\beta(x) = T_\beta(x) + 1$ for $x \in S$. The set $S$ is also called the switch region. By superimposing the two maps one can choose which of the maps to use once the orbit of $x$ falls into a switch region (see Figure 1.3). When choosing to iterate over the lazy map one will find that the digit will be one lower than when picking the greedy one. This gives us for almost every $x$ uncountably many expansions. For references on lazy $\beta$-expansions see [30, 38, 60] and for a mix of lazy and greedy [26, 31].

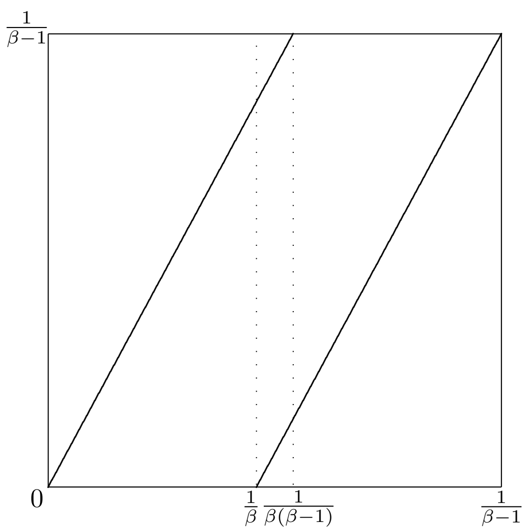


*Figure 1.3: The lazy and greedy $\beta$-transformation with $\beta = 1.85$.*

It is proven that the system $T_\beta$ is ergodic with respect to the invariant measure found by Gelfond in 1959 and Parry in 1960 independently (see [42] and [88]). The probability measure has density

$$f_\beta(x) = \frac{1}{C(\beta)} \sum_{n=0}^{\infty} \frac{1}{\beta^n} \mathbb{1}_{[0,T_\beta^n(1))}(x)$$

where $C(\beta) = \int_0^1 \sum_{n=0}^{\infty} \frac{1}{\beta^n} \mathbb{1}_{[0,T_\beta^n(1))}(x)dx$ is a normalising constant. This measure is the unique measure of maximal entropy (see Section 1.2.1 for more details). What plays a crucial role in the study of $\beta$-expansions is the quasi greedy expansion of 1. Let us first explain what a quasi greedy expansion is. Define the map $\tilde{T}_\beta(x) = T_\beta(x)$ for all $x$ with $T_\beta(x) \neq 0$ and $\tilde{T}_\beta(x) = 1$ whenever $x \neq 0$ and $T_\beta(x) = 0$. Set $\tilde{d}_1(x) = d_1(x)$ for $T_\beta(x) \neq 0$ and $\tilde{d}_1(x) = d_1(x) - 1$ whenever $x \neq 0$ and $T_\beta(x) = 0$. Furthermore, let $\tilde{d}_n(x) = \tilde{d}_1(\tilde{T}_\beta^{n-1}(x))$ for $n > 1$. The quasi greedy expansion of $x$ is given by

$$x = \sum_{i=1}^{\infty} \frac{\tilde{d}_i(x)}{\beta^i}.$$

Note that points ending up in 0 under the forward orbit of $T_\beta$ will have a finite greedy expansion. The error made by its convergent will be 0 at some point. For these points the quasi greedy expansion does a worse job in converging and there will always remain an error. For the quasi greedy expansion of 1 we write $\alpha(\beta) := \left(\tilde{d}_n(1)\right)_{n\geq 1}$. Let us now define the lexicographical ordering on sequences in $\{0, 1, \ldots, \lfloor\beta\rfloor\}^{\mathbb{N}}$. For two sequences $(x_i), (y_i) \in \{0, 1, \ldots, \lfloor\beta\rfloor\}^{\mathbb{N}}$ we write $(x_i) \prec (y_i)$ or $(y_i) \succ (x_i)$ if $x_1 < y_1$, if or there is an integer $m \geq 2$ such that $x_i = y_i$ for all $i < m$ and $x_m < y_m$. Moreover, we say $(x_i) \preccurlyeq (y_i)$ or $(y_i) \succcurlyeq (x_i)$ if $(x_i) \prec (y_i)$ or $(x_i) = (y_i)$. We can use this ordering and $\alpha(\beta)$ to prescribe which sequences are allowed in the $\beta$-expansion of any $x \in [0, 1]$. Due to Parry [88] we have that

$$\Sigma_\beta = \{(x_i) \in \{0, 1, \ldots, \lfloor\beta\rfloor\}^{\mathbb{N}} : \sigma^n((x_i)) \prec \alpha(\beta) \quad \text{for all } n \geq 0\}$$

is the set of all sequences that can occur as a $\beta$-expansion of some $x \in [0, 1]$. Here $\sigma$ denotes the shift, i.e. $\sigma((x_i)) = (x_{i+1})$. Not every sequence in $(\mathbb{N} \cup \{0\})^{\mathbb{N}}$ can occur as a quasi greedy expansion for some $\beta$. We have the following characterisation.

**Theorem 1.1.13 (Komornik and Loreti [61]).** *A sequence $(a_i) \in (\mathbb{N} \cup \{0\})^{\mathbb{N}}$ is a quasi greedy expansion of 1 for some $\beta$ if and only if*

$$0 \prec a_{n+1}a_{n+2}\ldots \preccurlyeq a_1a_2\ldots \quad \textit{for all} \quad n \geq 0.$$

In Chapter 5 there will be a constant interplay between the symbolic space $\Sigma_\beta$ and the space $[0, 1]$. We proceed by explaining the other terms in the title of this dissertation: entropy, matching and holes.

# §1.2 Explaining the terms in the title

## §1.2.1 Entropy

In this section we explain what entropy is. For any dynamical system $(\Omega, \mathcal{F}, \mu, T)$ the entropy is defined in the following way.

**Definition 1.2.1 (Entropy of a partition).** *Let $\gamma$ be a countable partition of $\Omega$, i.e a collection of pairwise disjoint ($\mu$-measurable) sets such that their union is $\Omega$ up to a $\mu$-measure 0 set. The entropy of the partition is given by*

$$h_\mu(\gamma, T) = -\sum_{\gamma_i \in \gamma} \mu(\gamma_i) \log(\gamma_i).$$

*where* $0 \log(0) = 0$.

**Definition 1.2.2 (Entropy).** *We define the entropy of $T$ by*

$$h_\mu(T) := \sup_\gamma h_\mu(\gamma, T)$$

*where we take the supremum over all countable partitions.*

Observe that different measures give different values for $h$. Often one is interested in

$$\sup_{\mu:\ \mu \text{ is invar.}} h_\mu(T)$$

and whether this value is attained for a certain measure. If so, this measure is called measure of maximal entropy. Intuitively the entropy of a system tells you something about the *amount of randomness* in a system. It is worth mentioning that entropy did not only show its importance in mathematics but also in fields like interactive particle systems [75] and information theory [25].

Unfortunately, the definition is not very helpful for applications since you have to take the supremum over all partitions. Fortunately, there are other ways to calculate the entropy. For the first method we need the notion of a generator. This generator will be a partition attaining the supremum (if it is finite). First we define

$$\gamma_1 \bigvee \gamma_2 = \{A_i \cap B_j : A_i \in \gamma_1, B_j \in \gamma_2\}$$

which allows us to define

$$\gamma_n^m = \bigvee_{k=n}^{m} T^{-k}\gamma$$

for any $n, m \in \mathbb{Z}$. Now we can define a generator.

**Definition 1.2.3 (Generator).** *Let $\sigma\left(\bigvee_{i=-\infty}^{\infty} T^{-i}\gamma\right)$ be the smallest $\sigma$-algebra containing all the partitions $\gamma_n^m$. Then $\gamma$ is called a generator w.r.t. $T$ if $\sigma\left(\bigvee_{i=-\infty}^{\infty} T^{-i}\gamma\right) = \mathcal{F}$.*

This leads us to a powerful theorem from 1959.

**Theorem 1.2.4 (Kolmogorov and Sinai [59, 101]).** *If $\gamma$ is a finite or countable generator for $T$ with $h(\gamma, T) < \infty$, then $h_\mu(T) = h_\mu(\gamma, T)$.*

We also have an existence theorem.

**Theorem 1.2.5 (Krieger [67]).** *If $T$ is an ergodic measure preserving transformation with $h_\mu(T) < \infty$, then $T$ has a finite generator.*

Note that, once we have a finite generator, we can calculate the entropy of the partition by taking a finite sum. It also gives a certificate that there are no other partitions giving a higher value. Therefore we find the entropy. A generator that works for continued fractions is the family $\{\Delta(k) = \{x : \lfloor \frac{1}{x} \rfloor = k\}\}$.

There are other ways to calculate the entropy. A theorem of Shannon, McMillan, Breiman and Chung uses any finite or countable partition. By applying it to a generator, the theorem can easily be used to find the entropy of the system. Let $A_n(x)$ be the unique element of $\bigvee_{i=0}^{n-1} T^{-i}\gamma$ such that $x \in A_n(x)$. Then we have the following theorem.

**Theorem 1.2.6 (Shannon-McMillan-Breiman-Chung).** *Let $\gamma$ be a countable partition of $X$ with $h_\mu(\gamma, T) < \infty$ then for almost every $x \in X$ we have*

$$\lim_{n\to\infty} -\frac{1}{n} \log\left(\mu(A_n(x))\right) = h_\mu(\gamma, T).$$

This theorem gives us the following insight in the setting of number expansions. If we let $\gamma$ be the collection of cylinder sets of length 1 then $A_n(x)$ is the set of $x$ starting with the same $n$ digits. Intuitively, the faster $A_n(x)$ shrinks the faster you gain information about $x$ and the higher the entropy. So if $A_n(x)$ shrinks fast we expect the convergents of $x$ to converge to $x$ fast. For continued fractions the bound on the convergence is given in terms of $q_n(x)$ so we might expect to find a formula for the entropy in terms of $q_n(x)$ as well which is indeed the case.

**Lemma 1.2.7 (Entropy formula [72]).** *Let $T$ be the Gauss map. For almost all $x \in [0, 1]$ we have*

$$h_\mu(T) = 2 \lim_{n\to\infty} \frac{1}{n} \left|\log(q_n(x))\right|. \tag{1.2.1}$$

Actually, in [72] the right-hand side is calculated and equals $\frac{\pi^2}{6}$ which later turned out to be the entropy of the Gauss map with respect to the Gauss measure. This holds in a slightly more general setting which we will prove in Chapter 3. In case the ergodic system satisfies the Rényi's condition (which is true in the case of continued fractions, $\beta$-expansions and other expansions considered in this dissertation) we can use the following formula found by Rohlin [96]:

$$h_\mu(T) = \int_\Omega \log |T'(x)| \, d\mu. \tag{1.2.2}$$

For a certain class of infinite systems this formula also holds (see [106] for example). This formula is very helpful since, once we have the density, we only need to integrate. Another way of calculating the entropy is by using Birkhoff's Theorem with a suitable $f$ which in this case gives us

$$\frac{1}{n}\sum_{i=1}^{n} \log |T'\left(T^i(x)\right)| \to \int \log |T'(x)|\, d\mu \quad \text{as } n \to \infty.$$

This formula is very useful for simulation, since we do not need the density. On the other hand, we do not obtain the density either.

The entropy of two isomorphic systems is equal (for the proof see [29]). The use of entropy is showcased in a paper of Ornstein, where he proved that Bernoulli shifts with equal entropy are isomorphic [87]. The entropy is also equal for a non-invertible system and its natural extension [10]. Furthermore, for a family of continued fractions ($\alpha$-continued fractions with a corresponding family of mappings $T_\alpha$ where $\alpha \in [0,1]$) it is shown in [65] that the product of the entropy of the dynamical system corresponding to $T_\alpha$ and the (non-normalised) measure of the domain of the natural extension is $\frac{\pi^2}{6}$.

The notion of entropy is present in every chapter of this dissertation. Since in Chapter 2 infinite ergodic theory is studied, we will introduce Krengel entropy which is calculated through (1.2.2). In Chapter 3 and 4 entropy is studied as a function of a parameter (for each different value of the parameter one has a different system). In Chapter 5 we will introduce the notion of topological entropy.

## §1.2.2 Matching

The concept of matching is relatively simple but often has big implications. Definitions vary from article to article as well as the name. The same phenomenon goes under the name of cycle property and synchronisation property. For a piecewise continuous map $T$ we define the right and left limit of a point $c$ as

$$T(c^+) := \lim_{x\downarrow c} T(x), \quad T(c^-) := \lim_{x\uparrow c} T(x).$$

We will use the definition as in [12].

**Definition 1.2.8 (Matching).** *Let $T : \Omega \to \Omega$ be a piecewise continuous map. We say that the matching condition holds for $T$ if for all discontinuity points $c$ we have that there are $N, M \in \mathbb{N}$ such that*

$$T^N(c^+) = T^M(c^-)$$

*and*

$$(T^N)'(c^+) = (T^M)'(c^-).$$

*The integers $N, M$ are called matching exponents of the discontinuity point $c$.*

Note that for piecewise linear mappings, the condition on the derivatives ensures that points in the neighbourhood of $c^+$ and $c^-$ also match, i.e. there exists $\delta > 0$ such that for all $\varepsilon \in (-\delta, \delta)$ we have $T^N(c^+ + \varepsilon) = T^M(c^- + \varepsilon)$. For continued fraction expansions it is a necessary but not sufficient condition. At first sight it is not clear whether this is useful to study. Still it showed its uses in various articles in various ways. Often a family of mappings $\{T_\alpha\}$ is studied. Matching of the discontinuities can have implications for the behaviour of an observable. For example, entropy as a function of $\alpha$ for several families of continued fractions is studied in [54, 56, 84]. Particularly, entropy is monotonic on matching intervals (see Chapter 3 for more details). In Chapter 3 we study a family for which matching holds almost everywhere. In Chapter 4 we will also encounter matching where it implies that the entropy is constant. Though, the way it is used in the proof is different from the others. It is proved that for parameters from a specific matching interval the corresponding systems are isomorphic. Also the natural extension comes into play. In Chapter 2 we study matching and natural extensions for a family of continued fraction transformations related to an infinite ergodic system. Interestingly, it does not seem to affect the entropy in the way it does for the other continued fraction systems studied.

A family related to the $\beta$-transformations is the so called generalised $\beta$-transformation which is given by $T_{\alpha,\beta} : [0,1] \to [0,1]$ with $T_{\alpha,\beta}(x) = \beta x + \alpha \bmod 1$. For a family of parameters $\beta$ it is shown that matching holds for almost every $\alpha$ (see [11]).

## §1.2.3 Holes and expansions

Up to now we looked at closed dynamical systems. That is, we had a quadruple $(\Omega, \mathcal{B}, \mu, T)$ with $T : \Omega \to \Omega$. We can make the system open by assigning a hole $H \subset \Omega$. Through this hole mass will leak out by iterating $T$. We are interested in those points that never fall into the hole.

**Definition 1.2.9 (Survivor set).** *The set*

$$S(H) := \{x \in \Omega : T^n(x) \notin H \text{ for all } n \in \mathbb{N}\}$$

*is called the survivor set of hole $H$.*

For any ergodic system we have that if $\mu(H) > 0$ then $\mu(S(H)) = 0$. In case that $\mu$ is equivalent to the Lebesgue measure $\lambda$, we find that $\lambda(S(H)) = 0$. Therefore we need a different tool to study the size of $S(H)$. To do so we use Hausdorff dimension which we denote by $\dim_H$. For the definition of Hausdorff dimension see [39]. Though, to compute the Hausdorff dimension by using the definition directly can be slightly cumbersome. What often is done to get dimensional results is to relate the set one is interested in to a set of which the dimension is already known. The following lemma helps to achieve this.

**Lemma 1.2.10 (Lipschitz [39]).** *Let $F \subset \mathbb{R}^n$. If $f : F \to \mathbb{R}^m$ is Lipschitz, then $\dim_H(f(F)) \leq \dim_H(F)$. If $f$ is also bi-Lipschitz, then $\dim_H(f(F)) = \dim_H(F)$.*

Several interesting sets pop up as survivor sets when the hole is carefully chosen. For example, let us look at the system $([0,1],\mathcal{B},\mu,T)$ where $\mathcal{B}$ is the Borel algebra, $\mu$ the Gauss measure and $T$ the Gauss map. We are interested in the set of all $x \in [0,1]$ such that the digits of $x$ are bounded, i.e. there exists an $N \in \mathbb{N}$ such that $d_i(x) \in \{1,\dots,N\}$ for all $i \in \mathbb{N}$. We find that $x$ satisfies this condition if and only if $x \in S\left((0,\frac{1}{N+1})\right)$. If we set $\mathcal{B}_N = S\left((0,\frac{1}{N+1})\right)$ then the set $BAD = \cup\mathcal{B}_N$ is known as the set of badly approximable numbers. This set is widely studied and has connections to the famous Littlewood conjecture [46]. For the size of $BAD$ we have that $\dim_H(BAD) = 1$. This is a result of Jarník from 1928 in [52]. In the same paper he also proved that

$$1 - \frac{1}{N\log(2)} \le \dim_H(\mathcal{B}_N) \le 1 - \frac{1}{8N\log(N)}$$

for $N \ge 8$. For the set $BAD$ a strong property holds which is called $\alpha$-winning (see [47]). This property implies that the set has full Hausdorff dimension and persists through taking intersections, i.e. if $A,B$ are both $\alpha$-winning then $A\cap B$ is $\alpha$-winning. Another example of an interesting set is the set of well approximable numbers. Let $K_N := \{x \in [0,1] : d_i(x) \ge N \text{ for all } i \in \mathbb{N}\} = S\left((\frac{1}{N},1)\right)$. We have good estimates for the size of $K_N$ in the case of $N \ge 20$ due to an article of Good from 1941 (see [45]):

$$\frac{1}{2} + \frac{1}{2\log(N+2)} < \dim_H(K_N) < \frac{1}{2} + \frac{\log(\log(N-1))}{2\log(N-1)}.$$

In Chapter 3 we will relate our set of interest with $K_N$.

In Chapter 5 we look at holes of the form $H_a = (0,a)$ for the greedy $\beta$-transformations. From [47] we have that $\cup H_a$ is $\alpha$-winning. Let us fix $\beta \in (1,\infty)$. For almost every $x \in [0,\frac{\lfloor\beta\rfloor}{\beta-1}]$, the greedy $\beta$-expansion is not the only representation of $x$ of the form $\sum_{i=1}^{\infty}\frac{a_i}{\beta^i}$. Holes have a connection to the set of $x$ with a unique expansion for a fixed $\beta$ (see [30]). On page 9 it is explained that for a fixed $\beta$ almost every $x$ has uncountably many expansions. Every time the orbit of the point falls into the switch region one has a choice. There are points that will never enter the switch region and therefore have a unique expansion. This is equivalent to stating that, when taking the switch region $\cup_{1\le i\le\lfloor\beta\rfloor}[\frac{i}{\beta},\frac{\lfloor\beta\rfloor}{\beta(\beta-1)}+\frac{i-1}{\beta}]$ as a hole, the survivor set will be the set of $x$ with a unique expansion. For dimensional results on such sets see [43, 62].

## §1.3 Statement of results

In Chapter 2 we study continued fractions in the domain of infinite ergodic theory. The corresponding infinite systems are obtained by flipping part of the branches of the Gauss map. We study the Krengel entropy, natural extensions and matching.

In Chapter 3 we study another family of continued fractions, namely the Ito Tanaka's $\alpha$-continued fractions. The relation between matching and entropy is shown. The focus lies on the set for which there is no matching. We characterise this set and obtain several dimensional results.

In Chapter 4 we study $N$-expansions. In the first part we also allow flips. For some cases we were able to find the natural extension and therefore the invariant density. A numerical method is used which is based on the Gauss-Kuzmin-Lévy Theorem. In the last part we study matching and entropy.

In Chapter 5 we study $\beta$-expansions with a hole around 0. We define $K_\beta(t) := \{x \in [0,1) : T_\beta^n(x) \notin (0,t) \text{ for all } n \geq 0\}$ and look at the set $E_\beta$ of all parameters $t \in [0,1)$ for which the set-valued function $t \mapsto K_\beta(t)$ is not locally constant. We show that $E_\beta$ is a Lebesgue null set of full Hausdorff dimension for all $\beta \in (1,2)$. Furthermore we characterise the topological structure of $E_\beta$ for any given $\beta \in (1,2]$.