Gaps, Frequencies and Spacial Limits of Continued Fraction Expansions

de Jonge, Jaap

DOI
10.4233/uuid:e0b37188-c8b6-4c96-9d04-93ac1f6899e3

Publication date
2020

Document Version
Final published version

Citation (APA)
de Jonge, J. (2020). *Gaps, Frequencies and Spacial Limits of Continued Fraction Expansions*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:e0b37188-c8b6-4c96-9d04-93ac1f6899e3

GAPS, FREQUENCIES AND SPACIAL LIMITS OF CONTINUED FRACTION EXPANSIONS

GAPS, FREQUENCIES AND SPACIAL LIMITS OF CONTINUED FRACTION EXPANSIONS

Proefschrift

ter verkrijging van de graad van doctor

aan de Technische Universiteit Delft,

op gezag van de Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen,

voorzitter van het College voor Promoties,

in het openbaar te verdedigen op vrijdag 24 januari 2020 om 12:30 uur

door

Cornelis Jacobus DE JONGE

Doctorandus in de Wiskunde
Universiteit van Amsterdam, Nederland,

geboren te Apeldoorn, Nederland.

Dit proefschrift is goedgekeurd door de

promotor: dr. C. Kraaikamp
promotor: prof. dr. F. H. J. Redig

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Dr. C. Kraaikamp,	Technische Universiteit Delft
Prof. dr. F. H. J. Redig,	Technische Universiteit Delft
Onafhankelijke leden:	
Dr. K. Dajani,	Universiteit Utrecht
Prof. dr. ir. G. Jongbloed,	Technische Universiteit Delft
Prof. dr. S. Marmi,	Scuola Normale Superiore di Pisa, Italië
Prof. dr. H. Nakada,	Keio University, Japan
Overig lid:	
Prof. dr. H. Jager,	Universiteit van Amsterdam
Reservelid:	
Prof. dr. ir. M. C. Veraar,	Technische Universiteit Delft

TU Delft Delft University of Technology
NWO Nederlandse Organisatie voor Wetenschappelijk Onderzoek

Printed by: GVO drukkers & vormgevers B. V.

ISBN 978-94-6384-087-3

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

CONTENTS

Summary

In this thesis continued fractions are studied in three directions: semi-regular continued fractions, Nakada's α-expansions and N-expansions. Whereas the first two had been studied quite thoroughly already, the third was still young and hardly explored.

In Chapter 1 the general concept of a continued fraction is given, involving an operator that yields the partial quotients or digits of a continued fraction expansion. The approximation coefficients $\vartheta_n(x) := q_n^2|x - p_n/q_n|$ are introduced, where p_n/q_n, $n = 0, 1, 2, \ldots$ are the convergents of the continued fraction. Some well-known results on semi-regular continued fractions are given. Finally, the concept of 'natural extension' is explained.

Chapter 2 is about orders (called patterns) of triplets of three consecutive approximation coefficients $\vartheta_{n-1}(x)$, $\vartheta_n(x)$ and $\vartheta_{n+1}(x)$. The asymptotic frequency of pattern $\mathscr{X}(n)$ is defined by

$$AF(\mathscr{X}(n)) := \lim_{N\to\infty} \frac{1}{N} \#\{n \in \mathbb{N} \,|\, 2 \le n \le N, \mathscr{X}(n)\}.$$

Starting with the regular continued fraction (RCF), it is shown that, for instance, the asymptotic frequency as $n \to \infty$ of the pattern $\vartheta_{n-1}(x) < \vartheta_n(x) < \vartheta_{n+1}(x)$ is smaller than the asymptotic frequency of the pattern $\vartheta_n(x) < \vartheta_{n+1}(x) < \vartheta_{n-1}(x)$. The asymptotic frequencies in the case of the RCF are explicitly given: two of them are $0.1210\cdots$, the others are $0.1894\cdots$. After this, these patterns are studied of two other semi-regular continued fractions: the optimal continued fraction (OCF) and the nearest integer continued fraction (NICF). The asymptotic frequencies of the OCF prove to be more equally distributed: the two less frequent patterns of the RCF now have the asymptotic frequency $0.1603\cdots$, where this is $0.1698\cdots$ for the other patterns. The asymptotic frequencies of the NICF prove to be different for all six patterns. However, summation of specific pairs yield once $2\cdot 0.1603\cdots$ and two times $2\cdot 0.1698\cdots$, thus showing a great correspondence with the OCF.

Chapter 3 is dedicated to the natural extension of Nakada's α-expansions. By means of singularisations and insertions in these continued fraction expansions, involving the removal or addition of partial quotients 1 in exchange with partial quotients with a minus sign, the interval on which the natural extension of Nakada's continued fraction map T_α is given is extended from $[\sqrt{2}-1, 1)$ to $[\sqrt{10}-2)/3, 1)$. From our construction it follows that Ω_α, the domain of the natural extension of T_α, is metrically isomorphic to Ω_g for $\alpha \in [g^2, g)$, where g is the small golden mean. Finally, although Ω_α proves to be very intricate and unmanageable for $\alpha \in [g^2, (\sqrt{10}-2)/3)$, the α-Legendre constant $L(\alpha)$ on this interval is explicitly given.

In Chapter 4 N-expansions are introduced for natural numbers N larger than 1. These expansions, like semi-regular continued fraction expansions, are also sequences of partial quotients, called orbits, existing in the interval $I_\alpha = [\alpha, \alpha+1]$ for some $\alpha \in (0, \sqrt{N}-1]$. Depending on N and α, there is a finite number of consecutive digits that occur as partial quotient. It appears that there are conditions (that is, combinations of N and α) such that these orbits eventually do not land in certain parts of the interval I_α, called gaps. It is proved that if the number of digits is at least five, no gaps exist. If the number of digits is four, there do not exist gaps for most N, but in the cases that there are α such that I_α contains a gap, there is only one and it covers the lion's part of I_α. When the number of digits is two or three, the number of gaps varies, but it is possible to give very clear conditions under which there are no gaps.

Samenvatting

In dit proefschrift worden de resultaten gepresenteerd van onderzoek op het gebied van kettingbreuken in drie richtingen: de klassieke semi-reguliere kettingbreuken, Nakada's α-ontwikkelingen en N-ontwikkelingen. De eerste twee gebieden hadden vóór dit onderzoek al een lange geschiedenis, terwijl de N-ontwikkelingen nog in de kinderschoenen staan.

In Hoofdstuk 1 worden algemene eigenschappen van kettingbreuken genoemd. Elk type heeft een eigen afbeelding die de wijzergetallen van de kettingbreuk voortbrengt. Ook de benaderingscoëfficiënten $\vartheta_n(x) := q_n^2|x - p_n/q_n|$ worden geïntroduceerd, waarin $p_n/q_n, n = 0,1,2,\ldots$ de convergenten van de bijbehorende kettingbreuk zijn. Verder worden een paar klassieke resultaten op het gebied van kettingbreuken genoemd. Ten slotte wordt het begrip 'natuurlijke uitbreiding' uitgelegd.

Hoofdstuk 2 gaat over volgorden (die 'patronen' worden genoemd) van drietallen opeenvolgende benaderingscoëfficiënten $\vartheta_{n-1}(x)$, $\vartheta_n(x)$ en $\vartheta_{n+1}(x)$. De asymptotische frequentie van patroon $\mathscr{X}(n)$ wordt gedefinieerd als

$$AF(\mathscr{X}(n)) := \lim_{N\to\infty} \frac{1}{N} \#\{n \in \mathbb{N} \mid 2 \le n \le N, \mathscr{X}(n)\}.$$

Eerst wordt in het geval van de reguliere kettingbreuk (RCF) aangetoond dat, bijvoorbeeld, de asymptotische frequentie voor $n \to \infty$ van het patroon $\vartheta_{n-1}(x) < \vartheta_n(x) < \vartheta_{n+1}(x)$ kleiner is dan die van het patroon $\vartheta_n(x) < \vartheta_{n+1}(x) < \vartheta_{n-1}(x)$. De asymptotische frequenties in het geval van de RCF worden expliciet gegeven: twee ervan zijn $0.1210\cdots$ en de andere zijn $0.1894\cdots$. Daarna worden deze patronen van twee andere semi-reguliere kettingbreuken onderzocht: de optimale kettingbreuk (OCF) en de kettingbreuk van het dichtstbijzijnde gehele getal. De asymptotische frequenties van de OCF blijken meer gelijkelijk te zijn verdeeld: de twee minder vaak voorkomende patronen van de RCF hebben asymptotische frequentie $0.1603\cdots$, terwijl die $0.1698\cdots$ is van de andere patronen. De asymptotische frequenties van de NICF zijn voor alle patronen verschillend, maar sommatie van specifieke paren geven een keer de somfrequentie $2\cdot 0.1603\cdots$ en twee keer $2\cdot 0.1698\cdots$, wat precies de waarden van de OCF zijn.

In hoofdstuk 3 wordt de natuurlijke uitbreiding van Nakada's α-ontwikkelingen behandeld. Een belangrijke rol is daarbij weggelegd voor singularisaties en invoegingen in deze kettingbreuken, waarbij wijzergetallen 1 worden verwijderd of juist toegevoegd in ruil voor wijzergetallen met een minteken. Daarmee wordt het interval waarop de natuurlijke uitbreiding van Nakada's kettingbreukafbeelding T_α bekend is verlengd van $[\sqrt{2}-1, 1)$ tot $[(\sqrt{10}-2)/3, 1)$. Uit de manier waarop dit gebeurt blijkt dat Ω_α, het domein van de natuurlijke uitbreiding van T_α, metrisch isomorf is met Ω_g voor $\alpha \in [g^2, g)$, waar g de kleine gulden snede is. Hoewel Ω_α uiterst ingewikkeld blijkt op $\alpha \in [g^2, (\sqrt{10}-2)/3)$ en niet meer goed te beschrijven, wordt de α-Legendre constante $L(\alpha)$ op dat interval expliciet gegeven.

In hoofdstuk 4 worden de N-ontwikkelingen geïntroduceerd, waarbij N een natuurlijk getal groter dan 1 is. Net als bij de reguliere kettingbreuken hebben we hier te maken met rijen wijzergetallen, die we 'banen' noemen. Deze banen bestaan in het interval $I_\alpha = [\alpha, \alpha+1]$, waarbij $\alpha \in (0, \sqrt{N}-1]$. Afhankelijk van N en α is er een eindig aantal opeenvolgende getallen dat als wijzergetal in een baan kan voorkomen. Het blijkt dat er omstandigheden (combinations van N en α) zijn waaronder banen vanaf een bepaald moment niet meer in bepaalde delen van I_α terechtkomen, die we 'gaten' noemen. Zulke gaten blijken niet voor te komen als het aantal verschillende wijzergetallen ten minste vijf is. Als dat aantal vier is, zijn er meestal geen gaten, maar als dat toch het geval is, dan betreft het precies één groot gat, dat het grootste deel van I_α beslaat. Als het aantal verschillende wijzergetallen twee of drie is, dan varieert het aantal gaten, maar het is mogelijk zeer duidelijke voorwaarden te geven waaronder geen gaten bestaan.

PREFACE

This thesis is the conclusion of my research as a PhD-student while being a professional teacher of mathematics in secondary education. This research was supported by the Netherlands Organisation for Scientific Research (NWO) under project number: 023.003.036.

Chapter 2 is a slightly adapted combination of two separately published papers: "On the approximation by three consecutive continued fraction convergents", published in Indagationes Mathematicae in 2014, and "Three consecutive approximation coefficients: asymptotic frequencies in semi-regular cases", published in Tohoku Mathematical Journal in 2018. Chapter 3, only superficially adapted, was published with the title "Natural Extensions for Nakada's alpha-expansions: descending from 1 to g^2" in Journal of Number Theory in 2018. Chapter 4 is not yet published, but will be the first of two papers on gaps in orbits of N-expansions, both to be submitted for publication soon.

Cornelis Jacobus de Jonge
Delft, September 2019

1

INTRODUCTION

1.1. GENERAL INTRODUCTION

This thesis is about continued fractions. The first two chapters in which I present the results of my research are on *semi-regular continued fractions*; the last, third chapter is on *N-continued fractions*. Although both types of continued fractions are about approximating irrational numbers by (rational) fractions, the main difference is that the semi-regular ones involve an infinite alphabet of *partial quotients* or *digits*, while N-expansions make use of a finite set of digits.

Generally speaking, a semi-regular continued fraction expansion of an irrational number $x \in [\alpha, \alpha+1]$, with $\alpha \in [-1, 0]$, is associated with an operator

$$K(x) = \left|\frac{1}{x}\right| - c,$$

where digit c is a natural number such that $K(x) \in [\alpha, \alpha+1] \setminus \mathbb{Q}$. By means of such an operator *any* irrational number can be expanded as a continued fraction if, in the case of numbers with an absolute value larger than 1, the subtracting with a natural number is done first.

In the case of N-expansions, the operator involved is

$$L(x) = \frac{N}{x} - d,$$

where $x \in [\alpha, \alpha+1]$ for some $\alpha \in (0, \sqrt{N}-1]$ and digit d is an integer such that $L(x) \in [\alpha, \alpha+1)$.

Whereas the semi-regular continued fractions had already been profoundly researched when I started my own research, the N-expansions have a very recent origin. Therefore, Chapters 2 and 3 are quite specific, built on an extensive literature on semi-regular continued fractions, while Chapter 4 is a true exploration of fundamental properties of N-expansions, more specifically of *orbits* of N-expansions.

In the next section I will give an introduction to semi-regular continued fractions. In the subsequent section I will introduce the *natural extension* of an invertible measure-preserving dynamical system, which is important for both Chapters 2 and 3. The introduction to N-continued fractions will be given in the related Chapter 4 itself.

1.2. SEMI-REGULAR CONTINUED FRACTIONS

A *continued fraction* of a real number x is defined as a finite or infinite fraction

$$a_0 + \cfrac{\varepsilon_1}{a_1 + \cfrac{\varepsilon_2}{a_2 + \cfrac{\varepsilon_3}{a_3 + \ddots}}} \tag{1.1}$$

In this expression one has $\varepsilon_n = \pm 1, n \geq 1$, $a_0 \in \mathbb{Z}$ and $a_n \in \mathbb{N}, n \geq 1$. In the following the more convenient notation $[a_0; \varepsilon_1 a_1, \varepsilon_2 a_2, \varepsilon_3 a_3, \dots]$ will be used for a continued fraction.

A finite or infinite continued fraction is called a *semi-regular continued fraction* (SCRF) when $a_n \geq 1, n \geq 1$; $\varepsilon_{n+1} + a_n \geq 1$, $n \geq 1$, and, in the infinite case, $\varepsilon_{n+1} + a_n \geq 2$ infinitely often; see for instance [P] or [K]. In this thesis solely infinite continued fractions are investigated.

The SRCFs have been studied extensively (e.g. [B, K]), as have their *approximation coefficients*, defined by

$$\vartheta_n(x) := q_n^2 \left| x - \frac{p_n}{q_n} \right|, \quad n = 0, 1, 2, \dots,$$

where x is a real irrational with continued fraction expansion (1.1) and p_n/q_n, with $n = 0, 1, 2, \dots,$ is the corresponding sequence of convergents, obtained by truncation of the infinite continued fraction (1.1). These approximation coefficients, giving an indication of the quality of the approximation of x by p_n/q_n, have been studied intensively; in Chapter 2 some important results will be presented. For convenience, in the rest of this thesis the suffix '(x)' behind ϑ_n will often be omitted.

The sequence $(p_n, q_n)_{n \geq -1}$ has also been studied much. Many well-known properties can be found in, for instance, [DK], among which $\gcd(p_n, q_n) = 1$, $n \geq -1$, and

$$\begin{aligned} p_{-1} &:= 1, & p_0 &:= a_0, & p_n &= a_n p_{n-1} + \varepsilon_n p_{n-2}, \ n \geq 1; \\ q_{-1} &:= 0, & q_0 &:= 1, & q_n &= a_n q_{n-1} + \varepsilon_n q_{n-2}, \ n \geq 1. \end{aligned} \tag{1.2}$$

In the study of the sequence ϑ_n the 'future' and 'past' of the continued fraction expansion $[a_0; \varepsilon_1 a_1, \varepsilon_2 a_2, \varepsilon_3 a_3, \dots]$ of an irrational number x play an important role, defined as

$$t_n := [0; \varepsilon_{n+1} a_{n+1}, \varepsilon_{n+2} a_{n+2}, \dots],\ n \geq 0$$

respectively

$$v_0 := 0 \text{ and } v_n := \frac{q_{n-1}}{q_n} = [0; a_n, \varepsilon_n a_{n-1}, \dots, \varepsilon_2 a_1],\ n \geq 1.$$

The following relations, which can be found in, for instance, [JK, p. 303], are essential for Chapter 2:

$$\vartheta_{n-1} = \frac{v_n}{1 + t_n v_n},\ n \geq 1, \tag{1.3}$$

$$\vartheta_n = \frac{\varepsilon_{n+1} t_n}{1 + t_n v_n},\ n \geq 1, \tag{1.4}$$

and

$$\vartheta_{n+1} = \varepsilon_{n+2}(\varepsilon_{n+1}\vartheta_{n-1} + a_{n+1}\sqrt{1 - 4\varepsilon_{n+1}\vartheta_{n-1}\vartheta_n} - a_{n+1}^2 \vartheta_n),\ n \geq 1. \tag{1.5}$$

1.3. THE NATURAL EXTENSION

In Chapters 2 and 3 the continued fraction operators are embedded in dynamical systems, defined as follows (see [DK], page 16):

Definition 1. A dynamical system is a quadruple $(X, \mathscr{F}, \varrho, T)$, where X is a non-empty set, $\mathscr{F}$ is a σ-algebra on X, ϱ is a probability measure on $(X, \mathscr{F})$ and $T : X \to X$ is a surjective ϱ-measure preserving transformation.

In particular, these chapters are about invertible dynamical systems, requiring the notion of *natural extension*, for which the definition of a dynamical system as a *factor* is needed; both definitions are taken from [DK] (pages 98 and 99) as well:

Definition 2. Let $(X, \mathscr{F}, \mu, T)$ and $(Y, \mathscr{C}, \nu, S)$ be two dynamical systems. Then $(Y, \mathscr{C}, \nu, S)$ is said to be a *factor* of $(X, \mathscr{F}, \mu, T)$ if there exists a measurable and surjective map $\Psi : X \to Y$ such that

(*i*) $\Psi^{-1}\mathscr{C} \subset \mathscr{F}$ (Ψ preserves the measure structure);

(*ii*) $\Psi T = S \Psi$ (Ψ preserves the dynamics);

(*iii*) $\mu(\Psi^{-1}E) = \nu(E), \forall E \in \mathscr{C}$ (Ψ preserves the measure).

The dynamical system $(X, \mathscr{F}, \mu, T)$ is called an *extension* of $(Y, \mathscr{C}, \nu, S)$ and Ψ is called a *factor map*. In case $(Y, \mathscr{C}, \nu, S)$ is a *non-invertible* measure-preserving dynamical system, the measure-preserving dynamical system $(X, \mathscr{F}, \mu, T)$ is called a *natural extension* of $(Y, \mathscr{C}, \nu, S)$ if Y is a factor of X and the factor map Ψ satisfies $\bigvee_{m=0}^{\infty} T^m \Psi^{-1}\mathscr{C} = \mathscr{F}$. For convenience it is common to speak of the natural extension of an operator S instead of the natural extension of the full dynamical system $(Y, \mathscr{C}, \nu, S)$.

In Chapter 2 the natural extension is used to calculate the asymptotic frequency of patterns such as $\vartheta_{n-1} < \vartheta_n < \vartheta_{n+1}$; in Chapter 3 it is investigated itself, in the sense that a natural extension is constructed.

References

[B] W. Bosma, *Optimal continued fractions*, Nederl. Akad. Wetensch. Indag. Math. 49 (1987), no. 4, 353–379.

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

[JK] H. Jager and C. Kraaikamp, *On the approximation by continued fractions*, Nederl. Akad. Wetensch. Indag. Math. 51 (1989), no. 3, 289–307.

[K] C. Kraaikamp, *A new class of continued fraction expansions*, Arith. 57 (1991), no. 1, 1–39.

[P] O. Perron, *Die Lehre von den Kettenbrüchen*, Chelsea, New York, 1929.

2

On the approximation by three consecutive approximation coefficients of continued fraction expansions

2.1. The regular continued fraction

Let $x \in \Omega := [0, 1] \setminus \mathbb{Q}$ have the regular continued fraction (RCF) expansion

$$x = [0; a_1, a_2, \ldots]$$

and let p_n/q_n, $n = 1, 2, 3, \ldots$, be the corresponding sequence of convergents. Let the operator $T : \Omega \to \Omega$ be defined by

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

Since for $x = [0; a_1, a_2, a_3, \ldots]$ we have $T(x) = [0; a_2, a_3, \ldots]$, T is called the *one-sided shift operator* connected with the continued fraction, also known as the *Gauss operator*.

In 1798, Legendre ([L]) showed that if $p, q \in \mathbb{Z}$, $q > 0$, $\gcd(p, q) = 1$, and

$$\left| x - \frac{p}{q} \right| < \frac{1}{2} \frac{1}{q^2},$$

then there exists an $n \geq 1$ for which $p_n/q_n = p/q$, with p_n/q_n the nth RCF-convergent of x. In 1895, Vahlen ([V]) showed that for all irrational x and all $n \geq 2$,

$$\min\{\vartheta_{n-1}, \vartheta_n\} < \frac{1}{2},$$

while Borel ([Bor]) showed in 1905 that

$$\min\{\vartheta_{n-1}, \vartheta_n, \vartheta_{n+1}\} < \frac{1}{\sqrt{5}}.$$

In the course of the 20th century several authors sharpened Borel's result:

$$\min\{\vartheta_{n-1}, \vartheta_n, \vartheta_{n+1}\} < \frac{1}{\sqrt{a_{n+1}^2 + 4}};$$

see e.g. [BM]. In fact, J. Tong ([T1]) showed in 1983 that one also has the *converse property*:

$$\max\{\vartheta_{n-1}, \vartheta_n, \vartheta_{n+1}\} > \frac{1}{\sqrt{a_{n+1}^2 + 4}}.$$

For the *optimal continued fraction* (OCF) expansion, which will be discussed in Section 2.4, one has even more impressive Diophantine properties:

$$\min\{\vartheta_{n-1}, \vartheta_n\} < \frac{1}{\sqrt{5}}.$$

Unfortunately, this is not the case for the *nearest integer continued fraction* (NICF) expansion, which will be discussed in Section 2.5.

In 1995, J. Tong showed in [T2] that for all irrational x, for all $n \geq 2$ and for all $k \geq 1$ one has that:

$$\min\{\vartheta_{n-1}, \vartheta_n, \ldots, \vartheta_{n+k}\} < \frac{1 + \left(\frac{3-\sqrt{5}}{2}\right)^{2k+3}}{\sqrt{5}}.$$

Note that

$$\lim_{k\to\infty} \frac{1 + \left(\frac{3-\sqrt{5}}{2}\right)^{2k+3}}{\sqrt{5}} = \frac{1}{\sqrt{5}}.$$

In various papers the distribution for almost all x of the sequences (ϑ_n) and $(\vartheta_n, \vartheta_{n+1})$ for $n \geq 1$ has been studied for the RCF, OCF, NICF, and several other continued fraction algorithms; see e.g. [BJW, BK1]. In this chapter, we will focus on the *asymptotic frequency* of triplets $(\vartheta_{n-1}, \vartheta_n, \vartheta_{n+1})$; see Section 2.2 for a definition.

Now let $\bar{\Omega} := \Omega \times [0, 1]$ and $x = [0; a_1, a_2, \ldots]$. The natural extension of T is the *two-sided shift operator* $\mathscr{T} : \bar{\Omega} \to \bar{\Omega}$, defined by

$$\mathscr{T}(x, y) := \left(T(x), \frac{1}{y + a_1}\right) = \left(\frac{1}{x} - a_1, \frac{1}{y + a_1}\right). \tag{2.1}$$

In particular, $\mathscr{T}$ is measure-preserving with regard to the measure m with density function μ, where

$$\mu(x, y) := \frac{1}{\log 2} \cdot \frac{1}{(1 + xy)^2}. \tag{2.2}$$

We remark that

$$\mathscr{T}^n(x,0) = (t_n, v_n), n \geq 1,$$

with the 'future' t_n and the 'past' v_n as defined in Chapter 1. In this section we will prove various arithmetical properties of $\mathscr{T}$, from which we will later deduce some metrical results on triplets of three consecutive approximation coefficients.

To make the description of these properties easier, we introduce a slightly different operator, $\mathscr{S} : \bar{\Omega} \to \bar{\Omega}$, defined by

$$\mathscr{S} := \mathscr{R}\mathscr{T},$$

$\mathscr{R}$ being the reflection

$$\mathscr{R}(t,v) = (v,t).$$

One easily sees that $\mathscr{S}$ is an involution. Moreover, $\mathscr{S}$ is measure-preserving with respect to the measure m.

We will study the effect of $\mathscr{S}$ on vertical strips

$$R_a := \left(\frac{1}{a+1}, \frac{1}{a}\right) \times [0,1], \quad a = 1,2,\ldots.$$

Observe that $(t_n, v_n) \in R_a$ if and only if $a_{n+1} = a$. It follows at once that $\mathscr{S}$ is a one-to-one mapping of R_a onto itself. Furthermore, $\mathscr{S}$ maps horizontal line segments in R_a onto vertical segments in R_a and vice versa.

Now consider the curve given by $t/(1+tv) = c$ (see for instance [BJW]). Provided that $c < 1/2$, $c \notin 1/n$, $n = 2,3,\ldots$, this curve intersects both R_a and R_{a-1}, $a = 2,3,\ldots$. Under $\mathscr{S}$, both curve segments are flipped, each in its own vertical strip, resulting in an overall invariance of the curve under $\mathscr{S}$.

While the operator $\mathscr{T}$ has but one fixed point per strip R_a, i.e.

$$G_a = (g_a, g_a) := ([0;\bar{a}],[0;\bar{a}]) = \left(\frac{1}{2}(-a+\sqrt{a^2+4}), \frac{1}{2}(-a+\sqrt{a^2+4})\right),$$

the fixed points of $\mathscr{S}$ on R_a are the points of the form

$$\left(t, \frac{1}{t} - a\right),$$

i.e. all the points of the graph of T. Figure 2.1 shows the strips R_2 and R_3 with the associated curves $t/(1+tv) = c$ and the fixed points of $\mathscr{S}$.

Since the equations

$$v = \frac{1}{t} - a \quad \text{and} \quad v = (1-at)(a+v)$$

are equivalent, we derive from (1.3) and (1.5) that the graph of the Gauss operator T divides R_a into two regions: one, to the right, corresponding with

$$\vartheta_{n+1} < \vartheta_{n-1},$$

2

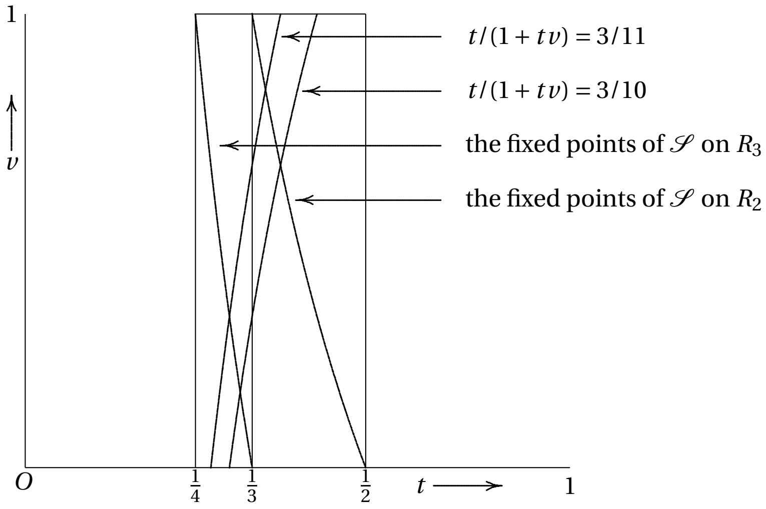


Figure 2.1: The fixed points of $\mathscr{S}$ as well as two curves $t/(1+tv) = c$ on R_2 and R_3

the other one, to the left, with

$$\vartheta_{n+1} > \vartheta_{n-1}.$$

These two regions are mapped by $\mathscr{S}$ onto each other, so they have the same measure m.

This can be generalised in the following way. On R_a, it follows from (1.3) and (1.5) that instead of

$$\vartheta_{n+1} - \vartheta_{n-1} = \lambda$$

we can write (omitting indices)

$$\frac{(1-at)(a+v)}{1+tv} - \frac{v}{1+tv} = \lambda, \tag{2.3}$$

which is equivalent with

$$v = \frac{a-\lambda}{a+\lambda}\frac{1}{t} - \frac{a^2}{a+\lambda}. \tag{2.4}$$

The corresponding hyperbola, h_λ, has a non-empty intersection with R_a if and only if

$$-\frac{a}{a+1} < \lambda < \frac{a}{a+1}.$$

Again we see from (1.3) and (1.5) that h_λ divides R_a into two regions, the one to the right of h_λ corresponding with

$$\vartheta_{n+1} - \vartheta_{n-1} < \lambda,$$

and the one to the left with

$$\vartheta_{n+1} - \vartheta_{n-1} > \lambda.$$

A simple calculation shows that

$$\mathscr{S} h_\lambda = h_{-\lambda}, \quad -\frac{a}{a+1} < \lambda < \frac{a}{a+1}. \tag{2.5}$$

In the next section we will use this to determine the mean value, for almost all x, of the sequence $|\vartheta_{n+1} - \vartheta_{n-1}|$, $n = 2, 3, \ldots$.

2

2.2. THE MEAN VALUE OF $|\vartheta_{n+1} - \vartheta_{n-1}|$

The sequence $|\vartheta_n - \vartheta_{n-1}|$, $n = 2, 3, \ldots$, is for almost all x distributed over the unit interval with density function f, where

$$f(\lambda) = \frac{1}{\log 2}\left(\frac{\pi}{2} - 2\arctan\lambda\right) \tag{2.6}$$

and as a consequence, for almost all x,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=2}^{N} |\vartheta_n - \vartheta_{n-1}| = \frac{4-\pi}{4\log 2} = 0.3096\cdots, \tag{2.7}$$

see ([J1]). A similar result for the sequence $|\vartheta_{n+1} - \vartheta_{n-1}|$, $n = 2, 3, \ldots$, is:

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=2}^{N} |\vartheta_{n+1} - \vartheta_{n-1}| = \frac{2\gamma + 1 - \log(2\pi)}{2\log 2} = 0.2283\cdots, \tag{2.8}$$

for almost all x, where γ is Euler's constant; see ([J2]). In this section we will show how to use the operator $\mathscr{S}$ on the strip R_a to deduce a crucial result for the proof of (2.8), in a much simpler way than in ([J2]). We need the following theorem; see [DK], Lemma 5.3.11 for a proof:

Theorem 1. *The two-dimensional sequence (t_n, v_n), $n = 1, 2, \ldots$, is for almost all irrational x distributed over the unit square in the (t, v)-plane according to the density function μ defined in (2.2).*

Now let $0 \le \lambda < \dfrac{a}{a+1}$ and let $L_{a,\lambda}$ be the region where $\vartheta_{n+1} - \vartheta_{n-1} > \lambda$ and $L_{a,-\lambda}$ be the one where $\vartheta_{n+1} - \vartheta_{n-1} < -\lambda$. From (2.5) it follows that

$$m(L_{a,\lambda}) = m(L_{a,-\lambda}), \quad 0 \le \lambda < \frac{a}{a+1},$$

an important ingredient in the proof of (2.8).

Straightforward calculation shows that the boundaries of $L_{a,\lambda}$ are the line segment given by $v = 0$, $t \in \left(\dfrac{1}{a+1}, \dfrac{a-\lambda}{a^2}\right)$, the line segment given by $t = \dfrac{1}{a+1}$, $v \in \left(0, \dfrac{a - a\lambda - \lambda}{a+\lambda}\right)$ and the part of h_λ that lies in R_a; see Figure 2.2.

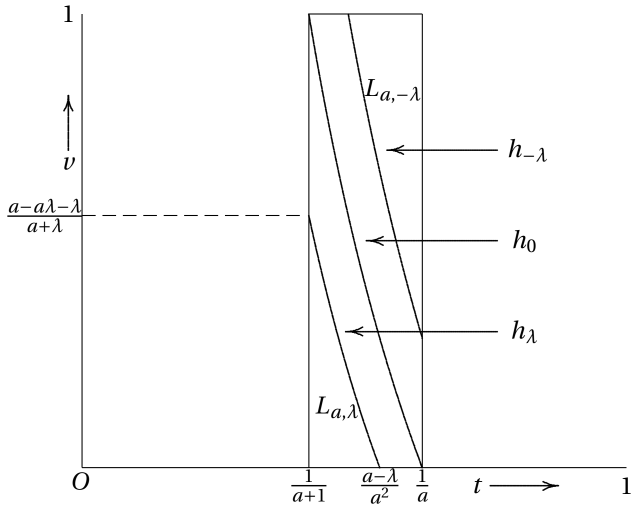


Figure 2.2: The regions $L_{a,\lambda}$ and $L_{a,-\lambda}$ in R_a

We define the asymptotic frequency of $\vartheta_{n+1} - \vartheta_{n-1} > \lambda$ as

$$P(\vartheta_{n+1} - \vartheta_{n-1} > \lambda) = \lim_{N\to\infty} \frac{1}{N} \#\{n; 2 \le n \le N, \vartheta_{n+1} - \vartheta_{n-1} > \lambda\} \tag{2.9}$$

and that of $\vartheta_{n+1} - \vartheta_{n-1} > \lambda$ under the condition $a_{n+1} = a$ as

$$P_a(\vartheta_{n+1} - \vartheta_{n-1} > \lambda) = \lim_{N\to\infty} \frac{1}{N} \#\{n; 2 \le n \le N, \vartheta_{n+1} - \vartheta_{n-1} > \lambda, a_{n+1} = a\}.$$

These limits exist for almost all x and

$$P(\vartheta_{n+1} - \vartheta_{n-1} > \lambda) = \sum_{a=1}^{\infty} P_a(\vartheta_{n+1} - \vartheta_{n-1} > \lambda).$$

From Theorem 1 it follows that

$$P_a(\vartheta_{n+1} - \vartheta_{n-1} > \lambda) = \frac{1}{\log 2} \int\int_{L_{a,\lambda}} \frac{\mathrm{d}t\,\mathrm{d}v}{(1+tv)^2}.$$

For our computation it is convenient to rewrite (2.4) as $t = \dfrac{a-\lambda}{a^2 + (a+\lambda)v}$.

We find

$$
\begin{aligned}
P_a(\vartheta_{n+1}-\vartheta_{n-1}>\lambda) &= \frac{1}{\log 2}\int_0^{\frac{a-a\lambda-\lambda}{a+\lambda}}\int_{\frac{1}{a+1}}^{\frac{a-\lambda}{a^2+(a+\lambda)v}} \frac{\mathrm{d}t\,\mathrm{d}v}{(1+tv)^2} \\
&= \frac{1}{\log 2}\int_0^{\frac{a-a\lambda-\lambda}{a+\lambda}} \left(\frac{t}{1+tv}\Bigg|_{\frac{1}{a+1}}^{\frac{a-\lambda}{a^2+(a+\lambda)v}} \right)\mathrm{d}v \qquad (2.10)\\
&= \frac{1}{2\log 2}\left(\log\frac{(a+1)^2}{a(a+2)}+\frac{\lambda}{a}\log\frac{a}{a+2}+\log\left(1-\frac{\lambda^2}{a^2}\right)+\frac{\lambda}{a}\log\frac{a+\lambda}{a-\lambda}\right).
\end{aligned}
$$

This result, crucial for the proof of (2.8), is also obtained in ([J2]), but with considerably more and inconvenient calculations. The remainder of the proof of (2.8) runs similar as the one in ([J2]).

2.3. Six patterns

Now suppose that, for instance, $\vartheta_{n-1}<\vartheta_n$. Then, regarding (2.7) and (2.8), $\vartheta_n<\vartheta_{n+1}$ seems less probable than $\vartheta_{n+1}<\vartheta_n$. In the latter case one might ask whether $\vartheta_{n+1}<\vartheta_{n-1}$ is more likely than $\vartheta_{n-1}<\vartheta_{n+1}<\vartheta_n$. In this section we determine, for almost all x, the asymptotic frequencies of the six possible patterns

$$
\begin{array}{lll}
\mathscr{A}:\vartheta_{n-1}<\vartheta_n<\vartheta_{n+1}, & \mathscr{B}:\vartheta_{n-1}<\vartheta_{n+1}<\vartheta_n, & \mathscr{C}:\vartheta_n<\vartheta_{n-1}<\vartheta_{n+1},\\
\mathscr{D}:\vartheta_n<\vartheta_{n+1}<\vartheta_{n-1}, & \mathscr{E}:\vartheta_{n+1}<\vartheta_{n-1}<\vartheta_n, & \mathscr{F}:\vartheta_{n+1}<\vartheta_n<\vartheta_{n-1}
\end{array} \qquad (2.11)
$$

where the asymptotic frequency of a pattern is defined in a similar way as in Section 2.2. As remarked, we may not expect the six patterns to all have the same asymptotic frequency, which would be $0.1666\cdots$.

Our investigation of the patterns is based on the division into six corresponding regions of the strip R_a. From the equations (1.3), (1.4) and (1.5) we deduce that g, the curve dividing the region where $\vartheta_{n-1}<\vartheta_n$ from the region where $\vartheta_n<\vartheta_{n-1}$, is given by $v=t$; that h, the curve dividing the region where $\vartheta_{n-1}<\vartheta_{n+1}$ from the region where $\vartheta_{n+1}<\vartheta_{n-1}$, is given by the graph of the Gauss operator T, $v=\frac{1}{t}-a$ on R_a. Further k, the curve dividing the region where $\vartheta_n<\vartheta_{n+1}$ from the region where $\vartheta_{n+1}<\vartheta_n$, is given by $v=\frac{1}{\frac{1}{t}-a}-a$, which is the graph of T^2 on the fundamental interval $\Delta_2(0;a,a)$, i.e.the set of numbers with a continued fraction expansion of the form $[0;a,a,\ldots]$.

It is easily verified that the three curves all intersect in G_a, the fixed point of $\mathscr{T}$ in R_a, and that the six regions actually correspond with the six patterns; see Figure 2.3.

2

For the frequencies of the six patterns that we are investigating, we define

$$Q_i := \bigcup_{a=1}^{\infty} Q_{i,a},\ i = 1,\ldots,6,$$

where the meaning of $Q_{i,a}$ can be read from Figure 2.3. As observed earlier, $\mathscr{S}h = h$ (by taking $\lambda = 0$ in (2.5)). In addition, straightforward computation shows that g and k are mapped onto each other by $\mathscr{S}$. Even stronger, if in R_a we divide both g and k in a left part and a right part (with regard to the fixed point), we have $\mathscr{S}g_l = k_r$ and $\mathscr{S}k_l = g_r$ (and vice versa). Also, under $\mathscr{S}$ the horizontal edge of $Q_{3,a}$ and the vertical edge of $Q_{2,a}$ are mapped onto each other, as holds similarly for the horizontal edge of $Q_{6,a}$ and the vertical edge of $Q_{5,a}$; for the horizontal edge of $Q_{3,a} \cup Q_{4,a}$ and the vertical edge of $Q_{1,a} \cup Q_{2,a}$; and for the horizontal edge of $Q_{1,a} \cup Q_{6,a}$ and the vertical edge of $Q_{4,a} \cup Q_{5,a}$.

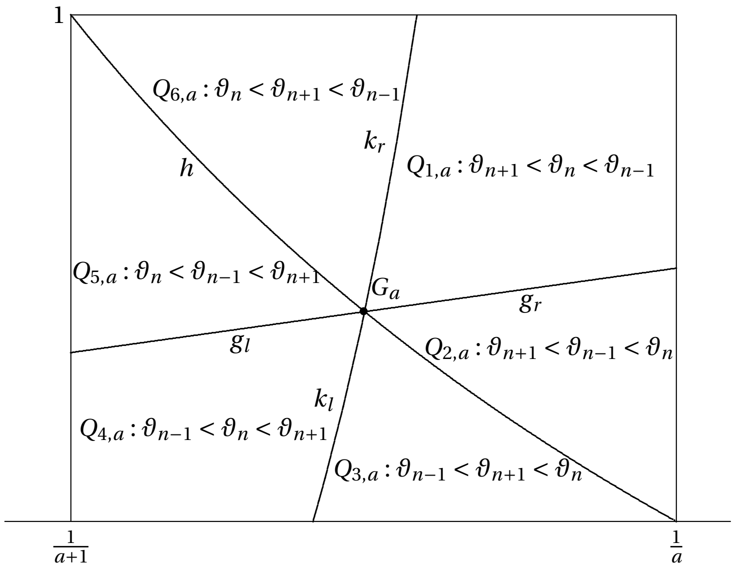


Figure 2.3: The six regions of R_a

From this, applying Theorem 1, we derive

$$m(Q_1) = m(Q_4),\ m(Q_2) = m(Q_3) \text{ and } m(Q_5) = m(Q_6).$$

Clearly

$$m(Q_1) + m(Q_5) + m(Q_6) = m(Q_2) + m(Q_3) + m(Q_4) = \frac{1}{2}.$$

and hence it follows that

$$m(Q_2) = m(Q_3) = m(Q_5) = m(Q_6) \quad \text{and} \quad m(Q_1) = m(Q_4).$$

Since the six asymptotic frequencies add up to 1, we only have to compute one. It is convenient to choose $Q_{2,a}$, and we find:

$$\begin{aligned} m(Q_{2,a}) &= \frac{1}{\log 2}\int_{g_a}^{\frac{1}{a}}\int_{\frac{1}{t}-a}^{t}\frac{\mathrm{d}v\,\mathrm{d}t}{(1+tv)^2} \\ &= \frac{1}{\log 2}\int_{g_a}^{\frac{1}{a}}\left(\frac{v}{1+tv}\bigg|_{\frac{1}{t}-a}^{t}\right)\mathrm{d}t \\ &= \frac{1}{2\log 2}\left(\log\left(1+\frac{1}{a^2}\right)-\log\left(\frac{1}{2}+\frac{1}{2}\sqrt{1+\frac{4}{a^2}}\right)\right). \end{aligned}$$

Summing over a, we come across two infinite products, the first of which is the well-known

$$\prod_{a=1}^{\infty}\left(1+\frac{1}{a^2}\right)=\frac{\sinh\pi}{\pi}=3.6760\cdots.$$

The factors of the second product are $1+a^{-2}-a^{-4}+O(a^{-6})$, $a\to\infty$. We find [1]

$$\rho := \prod_{a=1}^{\infty}\left(\frac{1}{2}+\frac{1}{2}\sqrt{1+\frac{4}{a^2}}\right)=2.8269\cdots.$$

Thus

$$m(Q_2)=\frac{1}{2\log 2}\cdot\log\frac{\sinh\pi}{\pi\rho}=0.1894\cdots.$$

Considering our earlier remarks, we have now proved

Theorem 2. *Define the constant ρ by*

$$\rho := \prod_{a=1}^{\infty}\left(\frac{1}{2}+\frac{1}{2}\sqrt{1+\frac{4}{a^2}}\right)=2.8269\cdots.$$

Then, with the notation introduced in (2.9) and (2.11), one has for almost all x:

$$\begin{cases} P(\mathscr{B})=P(\mathscr{C})=P(\mathscr{D})=P(\mathscr{E})=\dfrac{1}{2\log 2}\cdot\log\dfrac{\sinh\pi}{\pi\rho}=0.1894\cdots \\ P(\mathscr{A})=P(\mathscr{F})=\dfrac{1}{2}-\dfrac{1}{\log 2}\cdot\log\dfrac{\sinh\pi}{\pi\rho}=0.1210\cdots. \end{cases}\tag{2.12}$$

2.4. THE OPTIMAL CONTINUED FRACTION

In this section we will investigate the six asymptotic frequencies of the previous section for the *optimal continued fraction* (OCF). In general, we take the same approach as we

[1] using www.wolframalpha.com/examples/Sums.html

used for the regular RCF, based on the division of the (t, v)-plane in vertical strips in each of which the measures of all areas corresponding to the six patterns are computed, by applying Theorem 1. In this section we will show how to adopt this approach to the OCF, where the situation is more complicated. For convenience we will identify a pattern $\mathscr{P} \in \{\mathscr{A}, \ldots, \mathscr{F}\}$, with the region corresponding to this pattern. Throughout this thesis, we will use $g := \frac{1}{2}\sqrt{5} - \frac{1}{2} = 0.6180\cdots$ and $G := \frac{1}{2}\sqrt{5} + \frac{1}{2} = 1.6180\cdots$ as abbreviations of the two golden means. Note that $G = g + 1$ and that $g = \frac{1}{G}$.

As remarked above, we obtain the convergents p_n/q_n, $n = 1, 2, 3, \ldots$, by truncating the infinite continued fraction (1.1) expansion of a real irrational number x, so as to obtain good approximations of x. The approximation coefficients ϑ_n, $n = 1, 2, 3, \ldots$, provide a way of measuring the quality of the approximants. In [B], Wieb Bosma introduced the optimal continued fraction as a continued fraction that is both *fastest* (i.e. having an expansion for which the growth rate of the denominators is *maximal*) and *closest* (i.e. having expansions for which $\sup\{\vartheta_k : \vartheta_k = q_k|q_k x - p_k|\}$ is *minimal*).

Optimal as this fraction may be as to its approximating qualities, in [B, BK1] it is shown that both the subset of $\mathbb{R}^2$, which we denote by Υ_O, and the two-sided shift operator $\mathscr{T}_O : \Upsilon_O \to \Upsilon_O$ of the ergodic system underlying the OCF are less accessible than those of the RCF:

$$\Upsilon_O = \left\{(t, v) \in (-1, 1) \times (-1, 1) : v \leq \min\left(\frac{2t+1}{t+1}, \frac{t+1}{t+2}\right) \text{and } v \geq \max\left(0, \frac{2t-1}{1-t}\right)\right\},$$

see also Figure 2.4, and

$$\mathscr{T}_O(t, v) := \left(\left|\frac{1}{t}\right| - a(t, v), \frac{1}{a(t, v) + \operatorname{sign}(t)v}\right)$$

where

$$a(t, v) := \left\lfloor \left|\frac{1}{t}\right| + \frac{\left\lfloor\left|\frac{1}{t}\right|\right\rfloor + \operatorname{sign}(t)v}{2\left(\left\lfloor\left|\frac{1}{t}\right|\right\rfloor + \operatorname{sign}(t)v\right) + 1} \right\rfloor. \tag{2.13}$$

It is not hard to see that $\mathscr{T}_O$ works on Υ_O in way similar to $\mathscr{T}$ on $\bar{\Omega}$ in the RCF case:

$$\mathscr{T}_O(t, v) = \left(\frac{\varepsilon_1}{t} - a_1, \frac{1}{\varepsilon_1 v + a_1}\right), \text{ and } (t_n, v_n) = \mathscr{T}_O^n(t, 0),\ n \geq 0.$$

In [BK1] it is shown that $(\Upsilon_O, \mathscr{B}_{\Upsilon_O}, \bar{\mu}_{\Upsilon_O}, \mathscr{T}_O)$ forms an ergodic system, with $\mathscr{B}_{\Upsilon_O}$ the collection of Borel subsets of Υ_O and $\bar{\mu}_{\Upsilon_O}$ the measure with density function

$$d_O(t, v) := \frac{1}{\log G} \cdot \frac{1}{(1+tv)^2}, \text{ for } (t, v) \in \Upsilon_O. \tag{2.14}$$

In particular, we have that (apart from sets with Lebesgue measure 0) $\mathscr{T}_O : \Upsilon_O \to \Upsilon_O$ is bijective and that $\bar{\mu}_{\Upsilon_O}$ is invariant under $\mathscr{T}_O$. In [BK1] the following version of Theorem 1 was obtained:

Theorem 3. *The two-dimensional sequence* (t_n, v_n), $n = 1,2,\ldots$, *is for almost all irrational* x *distributed over* Υ_O *according to the density function* d_O *in* (2.14).

For more detailed information about the metric properties of the OCF, see for instance [BK1].

From Theorem 3 we derive that for every Borel measurable set $A \subseteq \Upsilon_O$

$$AF_O(A) = \iint_A \frac{1}{\log G} \cdot \frac{1}{(1+tv)^2}\,\mathrm{d}t\,\mathrm{d}v. \tag{2.15}$$

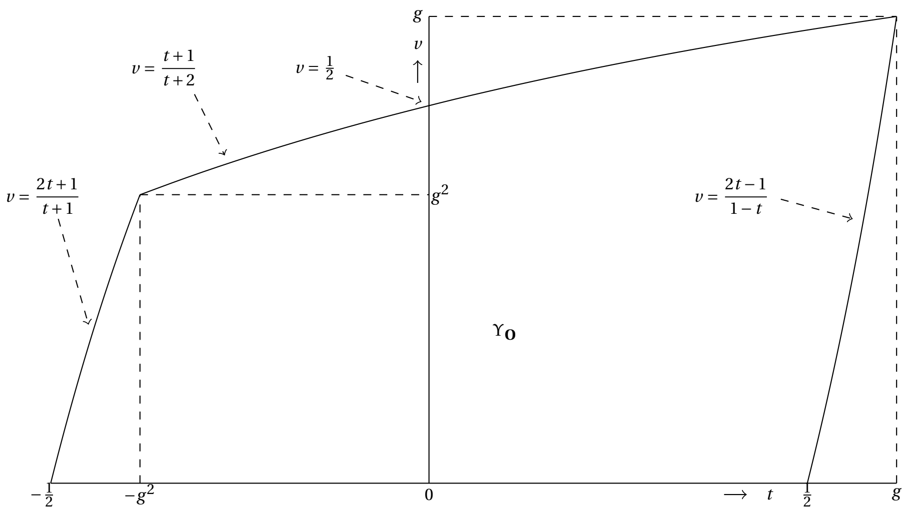


Figure 2.4: The domain of the OCF

The most important obstacle to following the approach taken in the case of the RCF is that a 'stripwise' computation (with $1/a_{n+1} < t_n < 1/a_n$, $n = 1,2,\ldots$) is not possible, due to the curved boundary of the domain of the natural extension of the OCF; see Figure 2.4. In view of (2.13), an obvious solution of this problem is regarding curved regions for every $a_{n+1} = 2,3,\ldots$. For $(t,v) \in \Upsilon_O$, the sign of t is obvious. However, it is not easy to find the regions of Υ_O where the values of the digit $a(t,v)$ is fixed using (2.13). We first will show that, apart from sets of Lebesgue measure 0, for every $(t,v) \in \Upsilon_O$ a unique integer $a \geq 2$ exists for which

$$\mathscr{T}_O(t,v) = \left(\frac{\varepsilon}{t} - a, \frac{1}{a+\varepsilon v}\right) =: (T,V) \in \Upsilon_O,$$

where $\varepsilon = \operatorname{sign}(t)$. We first consider the points (t,v) that are sent under $\mathscr{T}_O$ to the boundary of Υ_O. Let $(T,V) \in \partial(\Upsilon_O)$, then we have the following three cases:

1. (T,V) satisfies $V = \frac{2T+1}{1+T}$. In this case we obviously have that

$$(\alpha,\beta) := \left(\frac{\varepsilon}{t} - a{-}\mathbf{1}, \frac{1}{a{+}\mathbf{1} + \varepsilon v}\right) \notin \Upsilon_O,$$

since $T{-}\mathbf{1} = \frac{\varepsilon}{t} - a{-}\mathbf{1} < -\frac{1}{2}$. Now consider the point (α,β), given by

$$(\alpha,\beta) := \left(\frac{\varepsilon}{t} - a{+}\mathbf{1}, \frac{1}{a{-}\mathbf{1} + \varepsilon v}\right).$$

In this case we have

$$\beta = \frac{1}{a+\varepsilon v - 1} = \frac{1}{\frac{1}{V} - 1} = \frac{2\alpha - 1}{1-\alpha};$$

i.e. (α,β) is on one of the other boundary curves of Υ_O. We conclude that $(t,v) \in \Upsilon_O$ was on the boundary of the regions where the digit is either equal to a or to $a-1$.

2. (T,V) satisfies $V = \frac{2T-1}{1-T}$. In this case we obviously have that

$$(\alpha,\beta) := \left(\frac{\varepsilon}{t} - a{+}\mathbf{1}, \frac{1}{a{-}\mathbf{1} + \varepsilon v}\right) \notin \Upsilon_O,$$

since $T{+}\mathbf{1} = \frac{\varepsilon}{t} - a{+}\mathbf{1} > \frac{3}{2} > g$. Now consider the point (α,β), given by

$$(\alpha,\beta) := \left(\frac{\varepsilon}{t} - a{-}\mathbf{1}, \frac{1}{a{+}\mathbf{1} + \varepsilon v}\right).$$

In this case we have

$$\beta = \frac{1}{a+1+\varepsilon v} = \frac{1}{\frac{1}{V}+1} = \frac{2T-1}{T} = \frac{2\alpha+1}{1+\alpha};$$

i.e. (α,β) is on one of the other boundary curves of Υ_O. Again we conclude that $(t,v) \in \Upsilon_O$ was on the boundary of the regions where the digit is either equal to a or to $a+1$.

3. (T,V) satisfies $V = \frac{1+T}{2+T}$. In this case we obviously have that

$$(\alpha,\beta) := \left(\frac{\varepsilon}{t} - a{+}\mathbf{1}, \frac{1}{a{-}\mathbf{1} + \varepsilon v}\right) \notin \Upsilon_O,$$

since $T+1 \geq g$. Now consider the point (α,β), given by

$$(\alpha,\beta) := \left(\frac{\varepsilon}{t} - a{-}\mathbf{1}, \frac{1}{a{+}\mathbf{1} + \varepsilon v}\right).$$

In this case we have

$$\beta = \frac{1}{a+1+\varepsilon v} = \frac{1}{\frac{1}{V}+1} = \frac{2+T}{3+2T} = \frac{\alpha+3}{2\alpha+5} \notin \Upsilon_O.$$

In this case the digit a was unique.

Now let $(t,v) \in \Upsilon_O$ be such that $\mathcal{T}_O(t,v) \in \mathrm{Int}(\Upsilon_O)$ (here $\mathrm{Int}(S)$ denotes the interior of the set S). Then from the above it follows that we must have that

$$(\alpha,\beta) := \left(\frac{\varepsilon}{t} - a \pm 1, \frac{1}{a \mp 1 + \varepsilon v}\right) \notin \Upsilon_O,$$

so we must have that $a = a(t,v)$, i.e.

$$a = \left\lfloor \left|\frac{1}{t}\right| + \frac{\left\lfloor\left|\frac{1}{t}\right|\right\rfloor + \mathrm{sign}(t)v}{2\left(\left\lfloor\left|\frac{1}{t}\right|\right\rfloor + \mathrm{sign}(t)v\right) + 1} \right\rfloor.$$

In the regular case, the value of a_{n+1} depends on t_n only, but in the optimal case it depends on both t_n and v_n. We want to know how to determine the curves between which a_{n+1} is constant, given t_n and v_n. For convenience, we will generally omit the indices n for t and v and $n+1$ for a in what follows. We start in the leftmost corner of Υ_O, where $a = 2$, $\varepsilon_{n+1} = -1$ and $\varepsilon_{n+2} = +1$. So

$$\mathcal{T}_O(t,v) = \left(\frac{-1}{t} - 2, \frac{1}{2-v}\right) \qquad (a = 2, \varepsilon_{n+1} = -1, \varepsilon_{n+2} = +1). \tag{2.16}$$

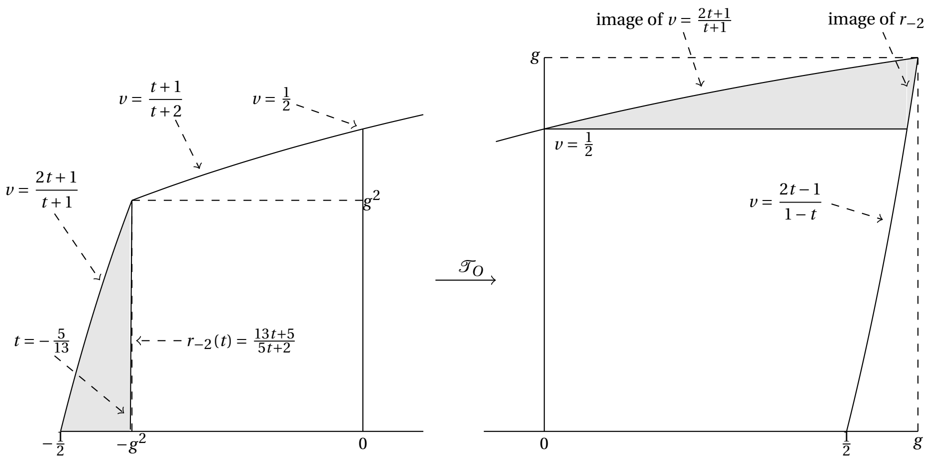


Figure 2.5: The map of the leftmost corner applying $\mathcal{T}_O$

The left boundary is given by $(t,(2t+1)/(t+1))$, for t between $-1/2$ and $-g^2$, which $\mathcal{T}_O$ maps to the curve $(T,V) = (-1/t-2, 1/(2-(2t+1)/(t+1))) = (-1/t-2, t+1)$, which we can write as $(T,(T+1)/(T+2))$, for T between 0 and g. The horizontal line segment with v-coordinate 0 is mapped to the horizontal line segment with V-coordinate $1/2$. We now determine the right boundary, denoted by $r_{-2} = r_{-2}(t)$, such that r_{-2} is mapped to the upper right boundary of Υ_O. Applying (2.16), we want to write $(-1/t-2, 1/(2-r_{-2}(t)))$ as $(T, (2T-1)/(1-T))$. A straightforward calculation yields $r_{-2}(t) = (13t+5)/(5t+2)$ (see Figure 2.5).

This procedure is easily copied to the rightmost side of Υ_O. This time we have $a = 2$, $\varepsilon_{n+1} = \varepsilon_{n+2} = +1$. Now

$$\mathscr{T}_O(t, v) = \left(\frac{1}{t} - 2, \frac{1}{2+v}\right) \qquad (a = 2, \varepsilon_{n+1} = \varepsilon_{n+2} = 1). \tag{2.17}$$

The right boundary is given by $(t, (2t-1)/(1-t))$, for t between $1/2$ and g, which $\mathscr{T}_O$ maps to the curve $(T, (T+1)/(T+2))$, for T between $-g^2$ and 0. The upper boundary is part of $(t, (t+1)/(t+2))$, its rightmost point being (g, g), which is mapped to the curve $(T, (2T+5)/(5T+13))$, with leftmost point on $T = -g^2$. We now determine the left boundary, denoted by $l_2 = l_2(t)$, such that l_2 is mapped to $(T, (2T-1)/(1-T))$. Applying (2.17), we want to write $(1/t-2, 1/(2+l_2(t)))$ as $(T, (2T-1)/(1-T))$. We find $l_2(t) = (13t-5)/(2-5t)$; see Figure 2.6.

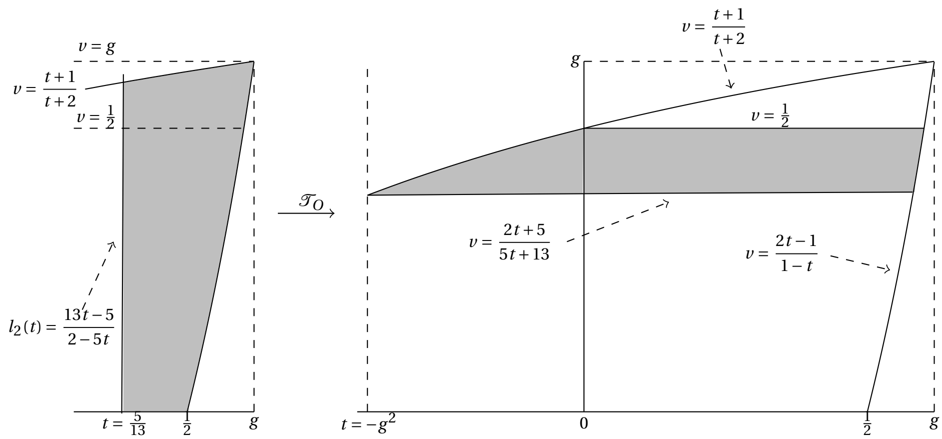


Figure 2.6: The map of the rightmost strip applying $\mathscr{T}_O$

Proceeding similarly, we establish formulas for all combinations of $a = 3, 4, \ldots$; ε_{n+1} and ε_{n+2}. We remark that the boundary between two regions with equal a and ε_{n+1} are separated by the line $t = \frac{\varepsilon_{n+1}}{a}$, where

$$r_{-a}(t) = \frac{(2a^2+2a+1)t+2a+1}{(2a+1)t+2}$$

and

$$l_a(t) = \frac{-(2a^2+2a+1)t+2a+1}{(2a+1)t-2}.$$

We conclude that $\mathscr{T}_O$ maps vertical regions from the left and the right side of Υ_O alternately to horizontal regions from the top of Υ_O downwards; see Figure 2.7.

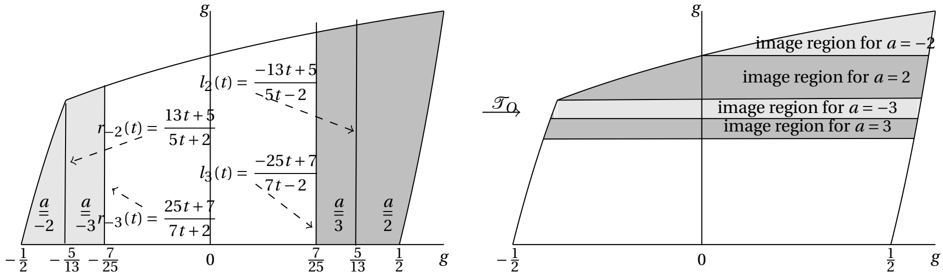


Figure 2.7: The alternating character of the map $\mathcal{T}_O$

In Figure 2.7 we have not yet processed the value of ε_{n+2}, which is indispensable for determining the six patterns. In Figure 2.8, confining ourselves to the leftmost part of Υ_O, we show how the six patterns are spread out over Υ_O, for $a = 2, 3, \ldots$. We have filled the regions with different shades of grey, such that pattern $\mathscr{A}$ has the lightest shade and $\mathscr{F}$ has the darkest.

$(\varepsilon_{n+1}, \varepsilon_{n+2})$	(-1,-1)	(-1,1)	(1,-1)	(1,1)
$\vartheta_{\mathbf{n-1}} = \vartheta_{\mathbf{n}}$	$v = -t$	$v = -t$	$v = t$	$v = t$
$\vartheta_{\mathbf{n-1}} = \vartheta_{\mathbf{n+1}}$	$v = a + \frac{1}{t}$	$v = \frac{a^2t+a}{at+2}$	$v = \frac{a^2t-a}{-at+2}$	$v = -a + \frac{1}{t}$
$\vartheta_{\mathbf{n}} = \vartheta_{\mathbf{n+1}}$	$v = \frac{(a^2-1)t+a}{at+1}$	$v = \frac{(a^2+1)t+a}{at+1}$	$v = \frac{(-a^2+1)t+a}{at-1}$	$v = \frac{(a^2+1)t-a}{-at+1}$
$\vartheta_{\mathbf{n-1}} = \vartheta_{\mathbf{n}} = \vartheta_{\mathbf{n+1}}$	$t = \frac{-a+\sqrt{a^2-4}}{2}$	$t = \frac{-a^2-2+\sqrt{a^4+4}}{2a}$	$t = \frac{2-a^2+\sqrt{a^4+4}}{2a}$	$t = \frac{-a+\sqrt{a^2+4}}{2}$

Table 2.1: The curves and their intersection per sign tuple.

Now that we have established a way of dividing Υ_O in regions where a, ε_{n+1} and ε_{n+2} are constant, we will show how we compute the measure of all regions. From the relations (1.3), (1.4) and (1.5) we derive for each of the four possible ordered sign tuples $(\varepsilon_{n+1}, \varepsilon_{n+2})$ the three curves that establish the six patterns. In each strip, that is, for every $a \geq 2$, we will now draw the curves that divide the strip in regions that correspond with the patterns $\mathscr{A}$ through $\mathscr{F}$, for which we will use Table 2.1. Recall that for convenience we use $t := t_n$, $v := v_n$ and $a := a_{n+1}$.

Finally, in Figure 2.9 we have a generic situation for the patterns: we know the values of ε_{n+1} and ε_{n+2} (which in Figure 2.9 is -1 for both of them) and all patterns actually occur, which is not the case in the leftmost and the rightmost regions. In Figure 2.9 we have indicated the formulas belonging to the curves drawn and some noteworthy values

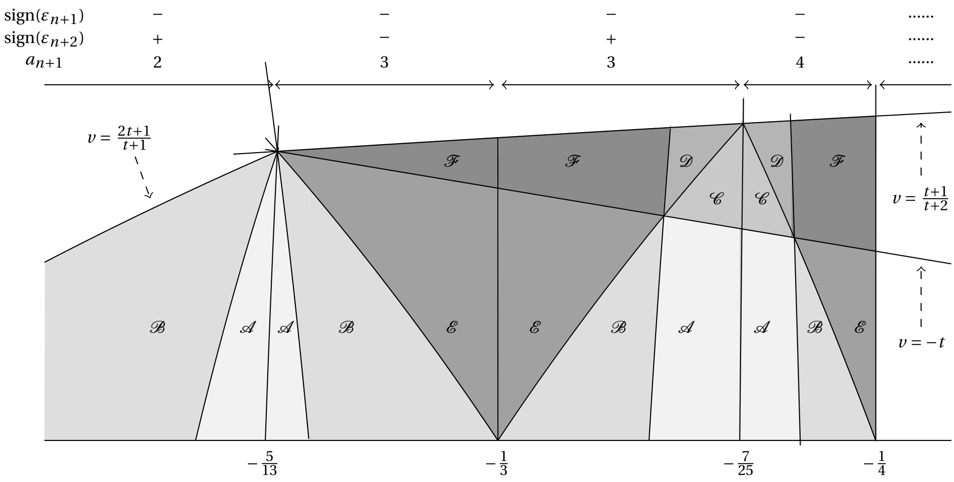


Figure 2.8: The six patterns for $\varepsilon_{n+1} = -1$

of t.

For convenience, we have omitted the coordinates of most intersection points, which are a bit lenghty in some cases. For instance, the t-coordinate of the intersection of $v = \frac{(2a^2-2a+1)t+2a-1}{(2a-1)t+2}$ (which is actually r_{a-1}) and $v = -t$ is

$$\frac{\sqrt{4a^4-8a^3+4a+5}-(2a^2-2a+3)}{4a-2}. \tag{2.18}$$

The calculation of the measure of areas such as $\mathscr{C}_a$ involves computing the sum of two double integrals, the limits of which are expressions such as (2.18). Computing the measures of all pattern regions for all four cases would obviously be very tedious and demanding, and therefore it is convenient that several areas prove to have the same measure. As in [JJ], we use a composed operator, which in the case of $\varepsilon_{n+1} = -1$ is

$$\mathscr{S}_O^- := \mathscr{R}^- \mathscr{T}_O,$$

$\mathscr{R}^-$ being the reflection

$$\mathscr{R}^-(t,v) = (-v,-t).$$

This operator $\mathscr{S}_O^-$ is an involution that is measure-preserving with respect to the measure m that was introduced on page 6. We will show how $\mathscr{S}_O^-$ works on the regions shown in Figure 2.9, where $\varepsilon_{n+2} = -1$ holds as well. We have (leaving the computations to the reader)

$$\begin{cases} \mathscr{S}_O^-\{(t,v) : v = \frac{1}{t} + a\} = \{(t,v) : v = \frac{1}{t} + a\}; \\ \mathscr{S}_O^- r_{-(a-1)} = \{(t,v) : v = \frac{t+1}{t+2}\}; \\ \mathscr{S}_O^-\{(t,v) : v = 0\} = \{(t,v) : t = -\frac{1}{a}\}; \\ \mathscr{S}_O^-\{(t,v) : v = -t\} = \{(t,v) : t = \frac{(a^2-1)t+a}{at+1}\}. \end{cases} \tag{2.19}$$

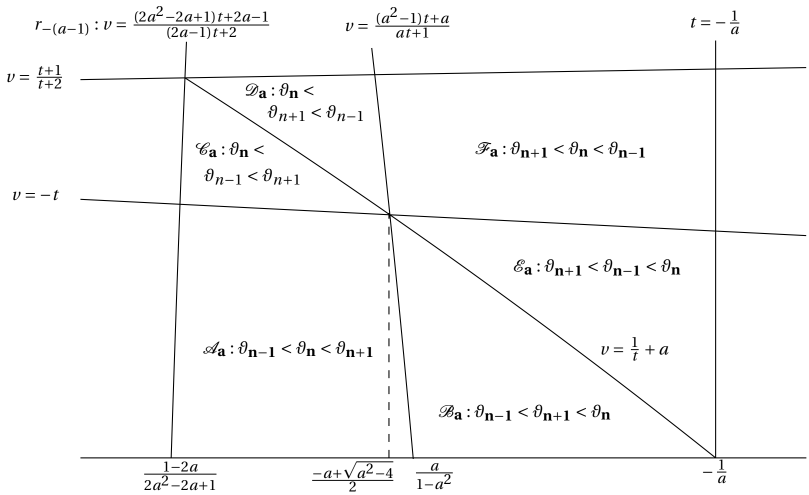


Figure 2.9: The six regions for $\varepsilon_{n+1} = \varepsilon_{n+2} = -1$

For $\mathscr{X} \in \{\mathscr{A}, \dots, \mathscr{F}\}$, we set $\mathscr{X}_a^{\varepsilon_1/\varepsilon_2} = \{(x, y) \in \mathscr{X} \mid a_1(x) = a, \varepsilon_1(x) = \varepsilon_1, \varepsilon_2(x) = \varepsilon_2\}$. Now, using (2.19), we easily derive the following (while $\varepsilon_{n+1} = \varepsilon_{n+2} = -1$):

$$\begin{cases} m(\mathscr{A}_a^{-/-}) = m(\mathscr{F}_a^{-/-}); \\ m(\mathscr{B}_a^{-/-}) = m(\mathscr{E}_a^{-/-}); \\ m(\mathscr{C}_a^{-/-}) = m(\mathscr{D}_a^{-/-}); \\ m(\mathscr{A}_a^{-/-} \cup \mathscr{B}_a^{-/-} \cup \mathscr{C}_a^{-/-}) = m(\mathscr{D}_a^{-/-} \cup \mathscr{E}_a^{-/-} \cup \mathscr{F}_a^{-/-}). \end{cases} \tag{2.20}$$

This is exactly what was found in the case of the RCF. We note, however, that at this place we are only dealing with the situation $\varepsilon_{n+1} = \varepsilon_{n+2} = -1$, while in the case of the RCF one always has $\varepsilon_{n+1} = \varepsilon_{n+2} = 1$. Still, we can confine ourselves to computing three relatively easy measures, say $m(\mathscr{C}_a^{-/-})$, $m(\mathscr{E}_a^{-/-})$ and $m(\mathscr{D}_a^{-/-} \cup \mathscr{E}_a^{-/-} \cup \mathscr{F}_a^{-/-})$.

We have (in the case $\varepsilon_{n+1} = \varepsilon_{n+2} = -1$, $a \geq 4$):

$$m(\mathscr{C}_a^{-/-}) = \int_{\frac{\sqrt{4a^4-8a^3+4a+5}-(2a^2-2a+3)}{4a-2}}^{\frac{-a+\sqrt{a^2-2a+2}}{a-1}} \int_{-t}^{\frac{(2a^2-2a+1)t+2a-1}{(2a-1)t+2}} \frac{\mathrm{d}t\,\mathrm{d}v}{(1+tv)^2} + \int_{\frac{-a+\sqrt{a^2-2a+2}}{a-1}}^{\frac{-a+\sqrt{a^2-4}}{2}} \int_{-t}^{\frac{1}{t}+a} \frac{\mathrm{d}t\,\mathrm{d}v}{(1+tv)^2},$$

which is

$$\frac{1}{2}\left(\log\frac{\sqrt{4a^4-8a^3+4a+5}+2a^2-2a-1}{2} + \log\frac{a-\sqrt{a^2-4}}{2} + \log(\sqrt{a^2-2a+2}-a+1)\right)$$

and can be written as

$$\frac{1}{2}\log\frac{\sqrt{(2a^2-2a-1)^2+4}+2a^2-2a-1}{(a+\sqrt{a^2-4})(\sqrt{(a-1)^2+1}+a-1)}.$$

Then,

$$m(\mathscr{E}_a^{-/-}) = \int_{\frac{-a+\sqrt{a^2-4}}{2}}^{\frac{-1}{a}}\int_{\frac{1}{t}+a}^{-t}\frac{\mathrm{d}t\,\mathrm{d}v}{(1+tv)^2} = \frac{1}{2}\log\frac{a-\sqrt{a^2-4}}{2}+\frac{1}{2}\log\frac{a^2-1}{a}.$$

Finally,

$$\begin{aligned} m(\mathscr{D}_a^{-/-}\cup\mathscr{E}_a^{-/-}\cup\mathscr{F}_a^{-/-}) &= \int_{\frac{-a+\sqrt{a^2-2a+2}}{a-1}}^{\frac{-1}{a}}\int_{\frac{1}{t}+a}^{\frac{t+1}{t+2}}\frac{\mathrm{d}t\,\mathrm{d}v}{(1+tv)^2} \\ &= \frac{1}{2}\log\frac{2a^2-2a+1}{a}+\frac{1}{2}\log(\sqrt{(a-1)^2+1}-(a-1)). \end{aligned}$$

Applying (2.20), we find for $\varepsilon_{n+1}=\varepsilon_{n+2}=-1$ and $a\geq 4$:

$$\begin{cases} m(\mathscr{A}_a^{-/-}) = m(\mathscr{F}_a^{-/-}) = \frac{1}{2}\log\frac{(2a^2-2a+1)(a\sqrt{a^2-4}+a^2-2)}{(a^2-1)(\sqrt{(2a^2-2a-1)^2+4}+(2a^2-2a-1))}; \\ m(\mathscr{B}_a^{-/-}) = m(\mathscr{E}_a^{-/-}) = \frac{1}{2}\log\frac{a-\sqrt{a^2-4}}{2}+\frac{1}{2}\log\frac{a^2-1}{a}; \\ m(\mathscr{C}_a^{-/-}) = m(\mathscr{D}_a^{-/-}) = \frac{1}{2}\log\frac{\sqrt{(2a^2-2a-1)^2+4}+2a^2-2a-1}{(a+\sqrt{a^2-4})(\sqrt{(a-1)^2+1}+a-1)}. \end{cases} \tag{2.21}$$

We remark that although for $a=3$ patterns $\mathscr{C}_a^{-/-}$ and $\mathscr{D}_a^{-/-}$ do not occur, the formula in (2.21) still holds, for it gives $m(\mathscr{C}_a^{-/-})=m(\mathscr{D}_a^{-/-})=0$.

In the case that $\varepsilon_{n+1}=\varepsilon_{n+2}=1$, the approach is completely analogous, including the use of

$$\mathscr{S}_O^+ := \mathscr{R}^+\mathscr{T}_O,$$

$\mathscr{R}^+$ being the reflection

$$\mathscr{R}^+(t,v)=(v,t),$$

instead of $\mathscr{S}_O^-$. In this case we find, for $a\geq 3$,

$$\begin{cases} m(\mathscr{A}_a^{+/+}) = m(\mathscr{F}_a^{+/+}) = \frac{1}{2}\log\frac{(a^2+2+a\sqrt{a^2+4})(2a^2+2a+1)}{(a^2+1)(2a^2+2a+3+\sqrt{(2a^2+2a+3)^2-4})}; \\ m(\mathscr{B}_a^{+/+}) = m(\mathscr{E}_a^{+/+}) = \frac{1}{2}\log\frac{\sqrt{a^2+4}-a}{2}+\frac{1}{2}\log\frac{a^2+1}{a}; \\ m(\mathscr{C}_a^{+/+}) = m(\mathscr{D}_a^{+/+}) = \frac{1}{2}\log\frac{2a^2+2a+3+\sqrt{(2a^2+2a+3)^2-4}}{(\sqrt{a^2+4}+a)(\sqrt{(a+1)^2+1}+(a+1))}. \end{cases} \tag{2.22}$$

In the cases where $\varepsilon_{n+1} \cdot \varepsilon_{n+2} = -1$, we get hold of the six patterns with a mixture of $\mathscr{S}_O^-$ and $\mathscr{S}_O^+$:

$$\begin{cases} \mathscr{S}_O^+(\mathscr{A}_a^{-/+}) = \mathscr{F}_a^{+/-}; & \mathscr{S}_O^-(\mathscr{F}_a^{+/-}) = \mathscr{A}_a^{-/+}; \\ \mathscr{S}_O^+(\mathscr{B}_a^{-/+}) = \mathscr{E}_a^{+/-}; & \mathscr{S}_O^-(\mathscr{E}_a^{+/-}) = \mathscr{B}_a^{-/+}; \\ \mathscr{S}_O^+(\mathscr{C}_a^{-/+}) = \mathscr{D}_a^{+/-}; & \mathscr{S}_O^-(\mathscr{D}_a^{+/-}) = \mathscr{C}_a^{-/+}; \\ \mathscr{S}_O^+(\mathscr{D}_a^{-/+}) = \mathscr{C}_a^{+/-}; & \mathscr{S}_O^-(\mathscr{C}_a^{+/-}) = \mathscr{D}_a^{-/+}; \\ \mathscr{S}_O^+(\mathscr{E}_a^{-/+}) = \mathscr{B}_a^{+/-}; & \mathscr{S}_O^-(\mathscr{B}_a^{+/-}) = \mathscr{E}_a^{-/+}; \\ \mathscr{S}_O^+(\mathscr{F}_a^{-/+}) = \mathscr{A}_a^{+/-}; & \mathscr{S}_O^-(\mathscr{A}_a^{+/-}) = \mathscr{F}_a^{-/+}, \end{cases} \tag{2.23}$$

and we find, for $a \geq 3$,

$$\begin{cases} m(\mathscr{A}_a^{-/+}) = m(\mathscr{F}_a^{+/-}) = \frac{1}{2}\log \frac{(\sqrt{a^4+4}-a^2)(a^2+1)}{(2a^2+2a+1)(\sqrt{(2a^2+2a-1)^2+4}-(2a^2+2a-1)}; \\ m(\mathscr{B}_a^{-/+}) = m(\mathscr{E}_a^{+/-}) = \frac{1}{2}\log \frac{a^2+2+\sqrt{a^4+4}}{2(a^2+1)}; \\ m(\mathscr{C}_a^{-/+}) = m(\mathscr{D}_a^{+/-}) = \frac{1}{2}\log \frac{(a^2-2+\sqrt{a^4+4})(a^2+a+1+(a+1)\sqrt{a^2+1})}{a^2(\sqrt{(2a^2+2a-1)^2+4}+(2a^2+2a-1))}; \\ m(\mathscr{D}_a^{-/+}) = m(\mathscr{C}_a^{+/-}) = \frac{1}{2}\log \frac{(a^2+2+\sqrt{a^4+4})(a^2-a+1+(a-1)\sqrt{a^2+1})}{a^2(2a^2-2a+3+\sqrt{(2a^2-2a+3)^2-4})}; \\ m(\mathscr{E}_a^{-/+}) = m(\mathscr{B}_a^{+/-}) = \frac{1}{2}\log \frac{a^2-2+\sqrt{a^4+4}}{2(a^2-1)}; \\ m(\mathscr{F}_a^{-/+}) = m(\mathscr{A}_a^{+/-}) = \frac{1}{2}\log \frac{(\sqrt{a^4+4}-a^2)(a^2-1)}{(2a^2-2a+1)((2a^2-2a+3-\sqrt{(2a^2-2a+3)^2-4})}. \end{cases} \tag{2.24}$$

We are almost able to give the total sum measure of all the six patterns. To actually do so, we have yet to compute the measures of the regions in the leftmost and the rightmost part of Υ_O, where $a = 2$. In the case $\varepsilon_{n+1} = \varepsilon_{n+2} = 1$, we can apply the formulas of (2.22). In the case $\varepsilon_{n+1} \cdot \varepsilon_{n+2} = -1$, the patterns $\mathscr{C}$ and $\mathscr{D}$ do not occur. In fact, on the left side we only have $\mathscr{A}$ and $\mathscr{B}$ and on the right side we only have $\mathscr{E}$ and $\mathscr{F}$, which can be mutually mapped onto each other as in (2.23); see Figure 2.10[2]. To compute their measures, we apply the formulas in (2.24).

[2]For visual purposes, we used different scaling for the left part and the right of Υ_O, as a result of which not everything seems to fit.

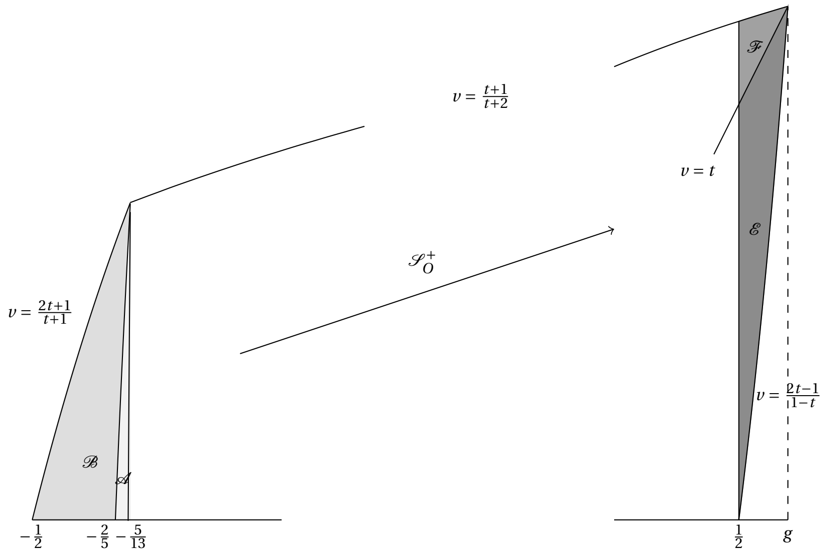


Figure 2.10: The four deviant regions, where $a = 2$ and $\varepsilon_{n+1} \cdot \varepsilon_{n+2} = -1$

Now we can compute the total sum measures of all regions:

Patterns $\mathscr{A}$ and $\mathscr{F}$:

$$m(\mathscr{A}) = m(\mathscr{F}) = m(\mathscr{A}_2^{-/+}) + m(\mathscr{A}_2^{+/+}) + \sum_{a=3}^{\infty} \left(m(\mathscr{A}_a^{-/-}) + m(\mathscr{A}_a^{-/+}) + m(\mathscr{A}_a^{+/-}) + m(\mathscr{A}_a^{+/+}) \right)$$

$$= \tfrac{1}{2}\log\tfrac{3+\sqrt{5}}{2} \ + \ \tfrac{1}{2}\log\tfrac{5}{13} \ + \ \tfrac{1}{2}\log\tfrac{13(6+4\sqrt{2})}{5(15+\sqrt{221})} \ + \ \sum_{a=3}^{\infty}\tfrac{1}{2}\log\tfrac{2a^2+2a-1+\sqrt{4a^4+8a^3-4a+5}}{2a^2-2a-1+\sqrt{4a^4-8a^3+4a+5}} \ +$$

$$\sum_{a=3}^{\infty}\tfrac{1}{2}\log\frac{\left(2a^2-2a+3+\sqrt{4a^4-8a^3+16a^2-12a+5}\right)\left(a\sqrt{a^2-4}+a^2-2\right)\left(a\sqrt{a^2+4}+a^2+2\right)}{2\left(2a^2+2a+3+\sqrt{4a^4+8a^3+16a^2+12a+5}\right)\left(a^4+2+a^2\sqrt{a^4+4}\right)}.$$

Applying the principle of telescoping series, we can reduce this to

$$\tfrac{1}{2}\log\left(3+2\sqrt{2}\right) + \tfrac{1}{2}\log\left(\sqrt{5}-2\right) + \sum_{a=3}^{\infty}\frac{1}{2}\log\frac{\left(a\sqrt{a^2-4}+a^2-2\right)\left(a\sqrt{a^2+4}+a^2+2\right)}{2\left(a^4+2+a^2\sqrt{a^4+4}\right)},$$

and finally to

$$\log\left(\sqrt{2}+1\right) - \tfrac{1}{2}\log G^3 + \sum_{a=3}^{\infty}\log\frac{\left(a+\sqrt{a^2-4}\right)\left(a+\sqrt{a^2+4}\right)}{2\left(a^2+\sqrt{a^4+4}\right)},$$

which can be simplified further to

$$\tfrac{3}{2}\log G+\sum_{a=2}^{\infty}\log\frac{\left(a+\sqrt{a^2-4}\right)\left(a+\sqrt{a^2+4}\right)}{2\left(a^2+\sqrt{a^4+4}\right)}.$$

In order to facilitate numerical computations, we write this last expression as

$$\frac{3}{2}\log G+\sum_{a=2}^{\infty}\log\frac{\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)}{2\left(1+\sqrt{1+\frac{4}{a^4}}\right)}.$$

Patterns $\mathscr{B}$ and $\mathscr{E}$:

$$m(\mathscr{B})=m(\mathscr{E})=m(\mathscr{B}_2^{-/+})+m(\mathscr{B}_2^{+/+})+\sum_{a=3}^{\infty}m(\mathscr{B}_a^{-/-})+m(\mathscr{B}_a^{-/+})+m(\mathscr{B}_a^{+/-})+m(\mathscr{B}_a^{+/+})$$

$=\tfrac{1}{2}\log\left(\frac{3+\sqrt{5}}{5}\right)\ +\ \tfrac{1}{2}\log(\sqrt{2}-1)+\tfrac{1}{2}\log\tfrac{5}{2}\ +\ \sum_{a=3}^{\infty}\tfrac{1}{2}\log\frac{2\left(a^2+\sqrt{a^4+4}\right)}{\left(a+\sqrt{a^2-4}\right)\left((a+\sqrt{a^2+4}\right)}$, shortly

$\tfrac{1}{2}\log G^2+\tfrac{1}{2}\log(\sqrt{2}-1)+\sum_{a=3}^{\infty}\tfrac{1}{2}\log\frac{2\left(a^2+\sqrt{a^4+4}\right)}{\left(a+\sqrt{a^2-4}\right)\left((a+\sqrt{a^2+4}\right)}$, which can be simplified further to

$-\tfrac{1}{2}\log G+\sum_{a=2}^{\infty}\tfrac{1}{2}\log\frac{2\left(a^2+\sqrt{a^4+4}\right)}{\left(a+\sqrt{a^2-4}\right)\left((a+\sqrt{a^2+4}\right)}$ or

$$-\frac{1}{2}\log G+\sum_{a=2}^{\infty}\frac{1}{2}\log\frac{2\left(1+\sqrt{1+\frac{4}{a^4}}\right)}{\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)}.$$

Patterns $\mathscr{C}$ and $\mathscr{D}$:

$$m(\mathscr{C})=m(\mathscr{D})=m(\mathscr{C}_2^{+/+})+\sum_{a=3}^{\infty}m(\mathscr{C}_a^{-/-})+m(\mathscr{C}_a^{-/+})+m(\mathscr{C}_a^{+/-})+m(\mathscr{C}_a^{+/+})$$

$$\begin{aligned}=\sum_{a=3}^{\infty}\Bigg(&\frac{1}{2}\log\frac{\left(2a^2-2a-1+\sqrt{4a^4-8a^3+4a+5}\right)\left(2a^2+2a+3+\sqrt{4a^4+8a^3+16a^2+12a+5}\right)}{\left(2a^2+2a-1+\sqrt{4a^4+8a^3-4a+5}\right)\left(2a^2-2a+3+\sqrt{4a^4-8a^3+16a^2-12a+5}\right)}\\ &+\frac{1}{2}\log\frac{2\left(a^2+\sqrt{a^4+4}\right)\left(2a^2+1+2a\sqrt{a^2+1}\right)}{\left(a+\sqrt{a^2-4}\right)\left(a+\sqrt{a^2+4}\right)\left(a+1+\sqrt{a^2+2a+2}\right)\left(a-1+\sqrt{a^2-2a+2}\right)}\Bigg)\\ &+\ \tfrac{1}{2}\log\frac{15+\sqrt{221}}{2}+\tfrac{1}{2}\log\left(\sqrt{10}-3\right)+\tfrac{1}{2}\log\left(\sqrt{2}-1\right).\end{aligned}$$

Again applying the principle of telescoping series, we can reduce this to

$\frac{1}{2}\log G^2 + \frac{1}{2}\log\left(\sqrt{2}-1\right) + \sum_{a=3}^{\infty} \frac{1}{2}\log \frac{2\left(a^2+\sqrt{a^4+4}\right)}{\left(a+\sqrt{a^2-4}\right)\left(a+\sqrt{a^2+4}\right)}$, which can be simplified further to

$-\frac{1}{2}\log G + \sum_{a=2}^{\infty} \frac{1}{2}\log \frac{2\left(a^2+\sqrt{a^4+4}\right)}{\left(a+\sqrt{a^2-4}\right)\left((a+\sqrt{a^2+4}\right)}$ or

$$-\frac{1}{2}\log G + \sum_{a=2}^{\infty} \frac{1}{2}\log \frac{2\left(1+\sqrt{1+\frac{4}{a^4}}\right)}{\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)},$$

being the same as for patterns $\mathscr{B}$ and $\mathscr{E}$.

All we have to do to find the asymptotic frequencies for the six patterns $\mathscr{A}$ through $\mathscr{F}$ is dividing these expressions by the normalising constant $\log G$ from (2.15) and rendering these numerically[3]. We have now proved the following theorem:

Theorem 4. *For the optimal continued fraction, the asymptotic frequencies of the six patterns of three consecutive approximation constants are given by*

$$AF_O(\mathscr{A}) = AF_O(\mathscr{F}) = \frac{3}{2} + \frac{1}{\log G}\left(\sum_{a=2}^{\infty} \log \frac{\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)}{2\left(1+\sqrt{1+\frac{1}{a^4}}\right)}\right)$$
$$\approx 0.1603\cdots;$$

$$AF_O(\mathscr{B}) = AF_O(\mathscr{C}) = AF_O(\mathscr{D}) = AF_O(\mathscr{E})$$
$$= -\frac{1}{2} + \frac{1}{2\log G}\left(\sum_{a=2}^{\infty} \log \frac{2\left(1+\sqrt{1+\frac{4}{a^4}}\right)}{\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)}\right)$$
$$\approx 0.1698\cdots.$$

We conclude that, similar to the case of the RCF, the OCF has only two different values for the asymptotic frequencies of the six patterns, similarly divided over these patterns, but mutually differing considerably less than in the case of the RCF (where these values are $0.1210\cdots$ and $0.1894\cdots$).

2.5. THE NEAREST INTEGER CONTINUED FRACTION

Like the OCF, the NICF is an example of a continued fraction with better approximation properties than those of the regular one. Although the NICF is merely fastest (and not closest), it is a continued fraction that is much studied; see for instance [WB] and [W]. As with the OCF, the convergents of the NICF and the OCF form a subsequence of the

[3] For obtaining numerical values we used Mathematica from WolframAlpha.

sequence of the RCF-convergents. It can be obtained from the RCF by a singularisation process concerning all partial quotients with value 1 ([K1]), yielding a continued fraction $[a_0; \varepsilon_1 a_1, \varepsilon_2 a_2, \ldots]$ that – in the case of the NICF – satisfies

$$\varepsilon_n \in \{-1, 1\}, n \geq 1; \quad a_0 \in \mathbb{Z}; \quad a_n \geq 2, n \geq 1; \quad \varepsilon_{n+1} + a_n \geq 2, n \geq 1.$$

The NICF of an $x \in \mathbb{R} \setminus \mathbb{Q}$ can also be obtained directly, by an algorithm explaining the name of this continued fraction. We remark that for $\alpha \in (-1/2, 1/2) \setminus \{0\}$ the expression $\lfloor 1/\alpha + 1/2 \rfloor$ is the rounding off of $1/\alpha + 1/2$ to its *nearest integer*, the absolute value of which is at least 2. It is also this expression in the NICF operator $\tau : (-1/2, 1/2) \to (-1/2, 1/2)$ that yields the partial quotients a_n of the NICF, where

$$\tau(t) := \frac{\varepsilon}{t} - \left\lfloor \frac{\varepsilon}{t} + \frac{1}{2} \right\rfloor, \; t \neq 0; \quad \tau(0) := 0,$$

with ε being the sign of t.

The values of ε_n and a_n, $n \geq 1$, are determined by repeated application of this operator:

$$\varepsilon_n = \operatorname{sgn}\left(\tau^{n-1}(t)\right) \quad \text{and} \quad a_n = \left\lfloor \frac{\varepsilon_n}{\tau^{n-1}(t)} + \frac{1}{2} \right\rfloor,$$

provided $\tau^{n-1}(t) \neq 0$ – which is always true in the case of $t \in \mathbb{R} \setminus \mathbb{Q}$.

Now put $\Omega_N := [-1/2, 1/2] \setminus \mathbb{Q}$ and let $[0; \varepsilon_1 a_1, \varepsilon_2 a_2, \ldots]$ be the NICF expansion of $t \in \Omega_N$.

We define

$$\rho(t) := \begin{cases} \dfrac{1}{\log G} \cdot \dfrac{1}{G + t}, & 0 \leq t \leq \frac{1}{2} \\[2ex] \dfrac{1}{\log G} \cdot \dfrac{1}{G + t + 1}, & -\frac{1}{2} \leq t < 0 \end{cases}.$$

Let μ be the measure with density function ρ. Then (Ω_N, μ, τ) forms an ergodic system, as was proved by G.J. Rieger ([Ri]) and A.M. Rockett ([Ro]).

For our investigation we use the natural extension of τ, which is the same as the one of the OCF, but defined on a different domain:

$$\mathcal{T}_N(t, v) := \left(\tau(t), \frac{1}{\varepsilon_1 v + a_1}\right) = \left(\frac{\varepsilon_1}{t} - a_1, \frac{1}{\varepsilon_1 v + a_1}\right),$$

where

$$(t, v) \in \Upsilon_N := [-\tfrac{1}{2}, 0] \setminus \mathbb{Q} \times [0, g^2] \cup [0, \tfrac{1}{2}] \setminus \mathbb{Q} \times [0, g];$$

see Figure 2.11. This natural extension domain was first obtained by H. Nakada ([N]), who showed that $(\Upsilon_N, \mathcal{B}_{\Upsilon_N}, \bar{\mu}_{\Upsilon_N}, \mathcal{T}_N)$ forms an ergodic system, where the $\mathcal{T}_N$-invariant probability measure $\bar{\mu}_{\Upsilon_N}$ has density function

$$d_N(t, v) := \frac{1}{\log G} \cdot \frac{1}{(1 + tv)^2} 1_{\Upsilon_N}(t, v).$$

Note that projecting this measure on the first coordinate axis yields a τ-invariant probability measure with Rieger's density function ρ; see also [K1]. As in the case of the OCF, we have to deal with ε_{n+1} and ε_{n+2}, yielding four different cases. At first, it seems convenient that we can take the approach from the RCF by regarding vertical strips in the (t,v)-plane. These strips $R_a^{\varepsilon_{n+1}/\varepsilon_{n+2}}$, defined below, are determined by the values of ε_{n+1} and ε_{n+2} and a_{n+1}, about which we remark that

$$\mathscr{T}_N^n(t,0) \in R_a^{\varepsilon_{n+1}/\varepsilon_{n+2}} \Leftrightarrow a_{n+1} = a,\ n \geq 0.$$

We define

$$\begin{aligned}
R_a^{-/-} &:= \left(-\frac{2}{2a-1}, -\frac{2}{2a}\right) \times [0, g^2], a = 3,4,5,\ldots \\
R_a^{-/+} &:= \left(-\frac{2}{2a}, -\frac{2}{2a+1}\right) \times [0, g^2], a = 2,3,4,\ldots \\
R_a^{+/-} &:= \left(\frac{2}{2a}, \frac{2}{2a-1}\right) \times [0, g], a = 3,4,5,\ldots \\
R_a^{+/+} &:= \left(\frac{2}{2a+1}, \frac{2}{2a}\right) \times [0, g], a = 2,3,4,\ldots
\end{aligned}$$

In Figure 2.11 we have drawn these strips in the $(t,v)-$square. Note that $g \approx 0.618$ and $g^2 \approx 0.382$. Also, $\varepsilon_{n+1} = -1$ implies $t < 0$ and $a_n \geq 3$, from which follows $v < g^2$. Secondly, $\varepsilon_{n+2} = -1$ implies $a = a_{n+1} \geq 3$, and therefore $|t| < 2/5$.

An important difference with the regular case is that the measure of each region is not given by one formula for every a for which the patterns exist. As we can see in Figure 2.12, in both $R_2^{-/+}$ and $R_3^{-/-}$ not all patterns are present. In Figures 2.13 and 2.14, we have filled all regions according to the same pattern with the same shade of grey, where – similar to the case of the OCF – darker shades correspond with ascending alphabetical order from $\mathscr{A}$ to $\mathscr{F}$.

In the case of the patterns $\mathscr{C}$ and $\mathscr{D}$ it is only from $a = 6$ on that the same formula holds for every value of a. This absence of one formula for every a is connected with the inutility of measure-preserving maps within the strips, from which the clear distribution of asymptotic frequencies was derived, in case of the RCF in [JJ] and in the case of the OCF in the previous sections. Actually, if both regions would have $0 \leq v \leq 1/2$ instead of $0 \leq v \leq g^2$ (in the case $\varepsilon_{n+1} = -1$) and $0 \leq v \leq g$ (in the case $\varepsilon_{n+1} = 1$), we could have applied the same maps for $R_a^{-/-}$ and $R_a^{+/+}$ as in the regular and the optimal case. But even then we would still have to deal with the less 'convenient' two other strips. There is not much we can do but calculate the measure for each region in the most exterior strips and then find general formulas for the regions in all other strips.

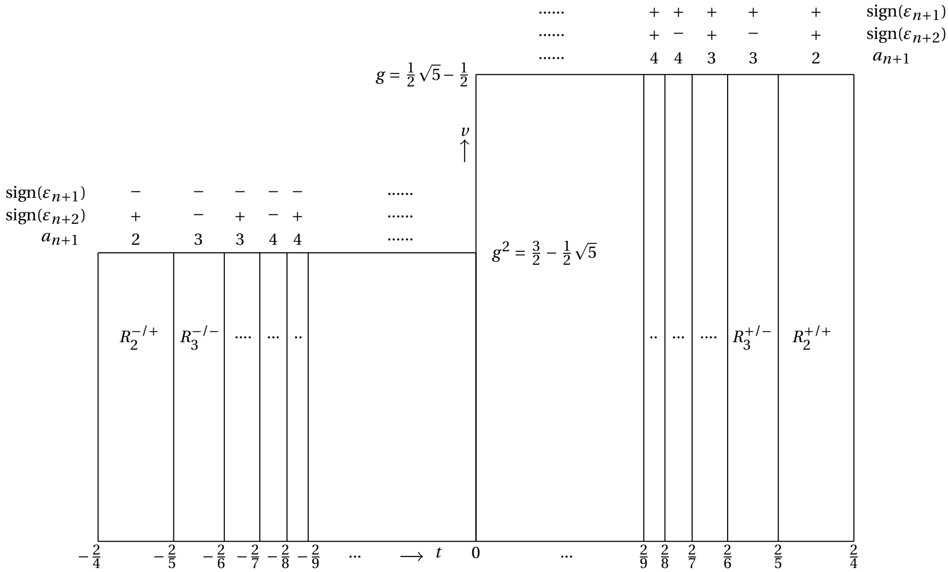


Figure 2.11: The strips $R_a^{*/*}$

2.6. THE MEASURES OF THE SIX PATTERNS OF THE NICF

For the frequencies of the patterns $\mathscr{A}$ through $\mathscr{F}$ that we are investigating, we define

$$\mathscr{X} := \bigcup_{a=2}^{\infty} \mathscr{X}_a, \mathscr{X} = \mathscr{A}, \ldots, \mathscr{F}.$$

A major complication in trying to determine a clear expression for the sum of the measures of all regions belonging to each pattern, is that quite a lot of different smaller expressions are involved, among which three different square root arguments, as we have seen in Table 1. In this section we wil evaluate these measures for each pattern separately. The calculations involve laborious double integrals that we will mostly omit. In establishing the formulas below, however, some terrifying expressions can be reduced using basic calculus. We come across expressions such as (in case $\mathscr{A}$)

$$\tfrac{1}{2}\log\left(a^6 + a^4 + 4a^2 + 4 - (a^4 + a^2 + 2)\sqrt{a^4+4}\right) - \tfrac{1}{2}\log\left(-a^4 - 4 + (a^2+2)\sqrt{a^4+4}\right),$$

which at first seem hard to handle. Applying long division and some other basic techniques, though, we can reduce this to

$$\tfrac{1}{2}\log\tfrac{-a^2+\sqrt{a^4+4}}{2}.$$

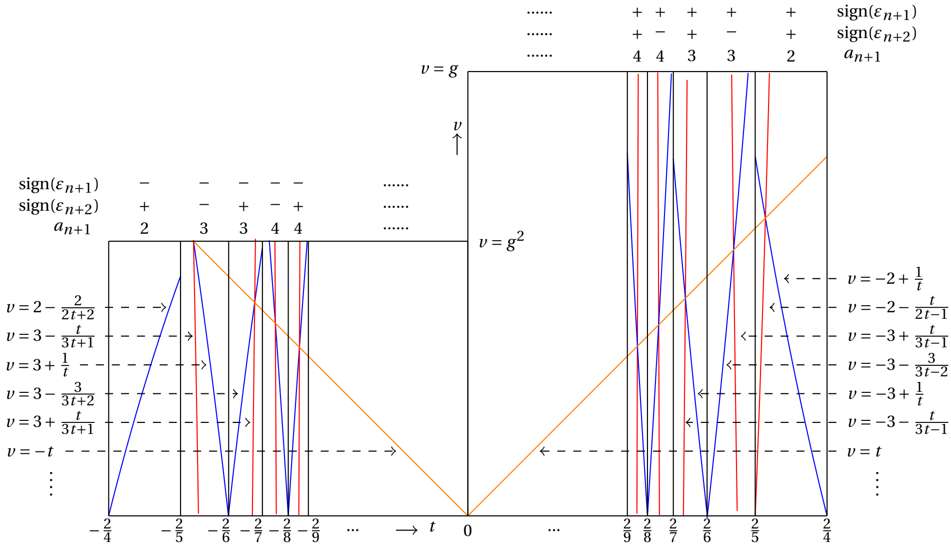


Figure 2.12: The curves dividing the regions of the patterns $\mathscr{A}$ through $\mathscr{F}$

It is in fact possible to evaluate the sum measures for four patterns as more or less well-arranged expressions of a definite form; for the patterns $\mathscr{C}$ and $\mathscr{D}$ a bit too many terms are involved. To find the asymptotic frequency of each pattern, we merely have to divide these measures by the normalizing constant, that is by

$$m(\Upsilon_N) = \int_{-\frac{1}{2}}^{0} \int_{0}^{g^2} \frac{\mathrm{d}t\,\mathrm{d}v}{(1+tv)^2} + \int_{0}^{\frac{1}{2}} \int_{0}^{g} \frac{\mathrm{d}t\,\mathrm{d}v}{(1+tv)^2} = \log G.$$

Unfortunately but not unexpectedly, the attractive conciseness of the regular case will not be reached, and we shall see that the asymptotic frequencies are different in all six cases.

2.6.1. PATTERN $\mathscr{A}$

There are two regions of pattern $\mathscr{A}$, in the strips $R_2^{+/+}$ and $R_3^{-/-}$, with deviant measures, that is, not computable with the formulas for all other strips stated below. These measures are $\log(\sqrt{2}+1) + \frac{1}{2}\log\frac{5}{29}$ and $\frac{1}{2}\log(7\sqrt{5}-15) + \log\frac{5}{4}$, respectively. The measures of all other $\mathscr{A}$-regions can be expressed as functions of a_{n+1} or shortly a, for $a \geq 3$, as in the

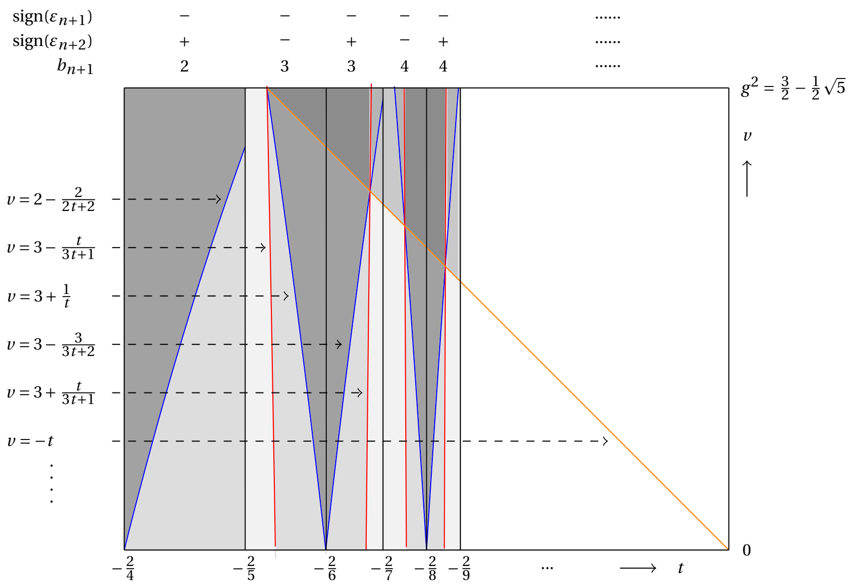


Figure 2.13: The patterns $\mathscr{A}$ through $\mathscr{F}$ for $\varepsilon_{n+1} < 0$

previous section. These are

$$
\begin{array}{lll}
R_a^{-/-}, a \geq 4: & \frac{1}{2}\log\frac{(2a-1)^2}{((2a-1)^2-4)(a^2-1)} & -\log\frac{a-\sqrt{a^2-4}}{2}; \\
R_a^{-/+}, a \geq 3: & \frac{1}{2}\log\frac{((2a+1)^2-4)(a^2+1)}{(2a+1)^2} & +\frac{1}{2}\log\frac{-a^2+\sqrt{a^4+4}}{2}; \\
R_a^{+/-}, a \geq 3: & \frac{1}{2}\log\frac{((2a-1)^2+4)(a^2-1)}{(2a-1)^2} & +\frac{1}{2}\log\frac{-a^2+\sqrt{a^4+4}}{2}; \\
R_a^{+/+}, a \geq 3: & \underbrace{\frac{1}{2}\log\frac{(2a+1)^2}{(2a+1)^2+4)(a^2+1)}}_{\text{part I}} & \underbrace{-\log\frac{a-\sqrt{a^2+4}}{2}.}_{\text{part II}}
\end{array}
$$

A close inspection of the factors in the arguments of the logarithms in part I unfolds that all terms in part I are mutually canceled, except for the one of $R_3^{+/-}$, the canceling term of which would be found partly in $R_2^{+/+}$ and partly in $R_3^{-/-}$. This yields a measure of $\frac{1}{2}\log\frac{29\cdot 8}{25}$. If we now sum all terms of both parts from $a = 3$, we mistakenly add a 'virtual' $-\frac{1}{2}\log\frac{3^2-2-3\sqrt{3^2-4}}{2}$ from $R_3^{-/-}$ that we have to cancel by adding $\frac{1}{2}\log\frac{7-3\sqrt{5}}{2}$. Finally, summing up to $a = n$, we should add the value $\frac{1}{2}\log(n^2+4n-3) - \frac{1}{2}\log(n^2+4n+5)$ from $R_n^{-/+}$ and $R_n^{+/+}$ that has by then not yet been canceled by the corresponding terms in $R_{n+1}^{-/-}$ and $R_{n+1}^{+/-}$. However, $\lim_{n\to\infty}\frac{1}{2}\log(n^2+4n-3) - \frac{1}{2}\log(n^2+4n+5) = 0$, so for the infinite summation this makes no difference. So far, having summed the aforementioned constants,

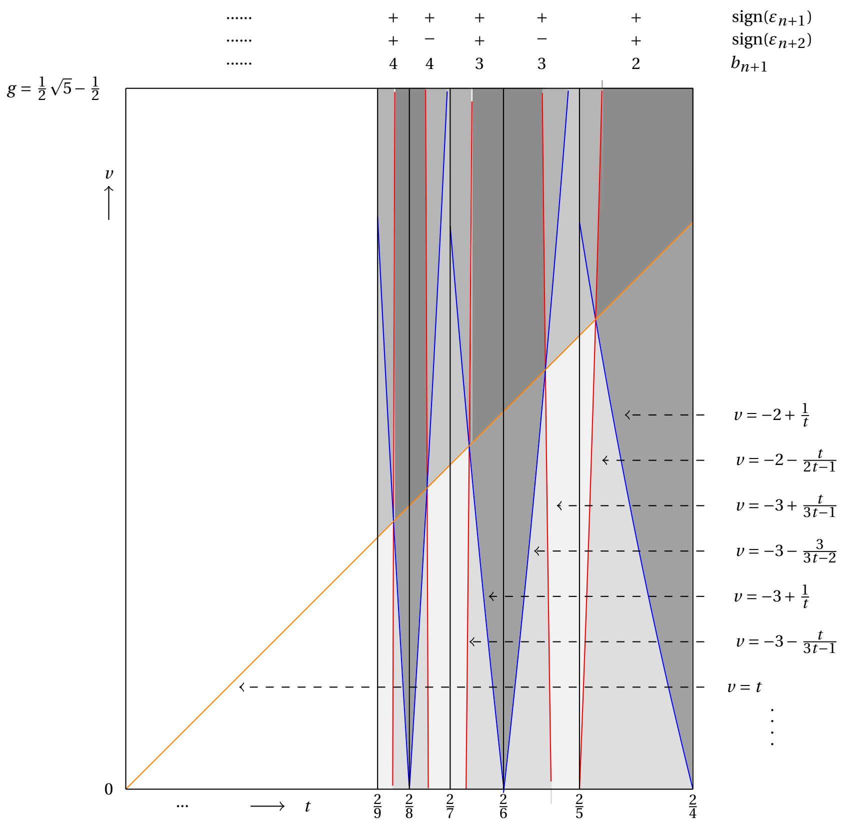


Figure 2.14: The patterns $\mathscr{A}$ through $\mathscr{F}$ for $\varepsilon_{n+1} > 0$

we have found the sum measure of the regions $\mathscr{A}$ to be:

$$\frac{1}{2}\log\frac{235\sqrt{5}-525}{2}+\log(\sqrt{2}+1)+\sum_{a=3}^{\infty}(\text{expressions in part II}). \tag{2.25}$$

The final step in getting hold of the sum measure is reducing the sum of the four expressions in part II, which turns out to be $\log\frac{\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)}{2\left(1+\sqrt{1+\frac{4}{a^4}}\right)}$. This means that we can write (2.25) as

$$\frac{1}{2}\log\frac{235\sqrt{5}-525}{2}+\log\left((\sqrt{2}+1)\prod_{a=3}^{\infty}\frac{\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)}{2\left(1+\sqrt{1+\frac{4}{a^4}}\right)}\right). \tag{2.26}$$

Unfortunately, we cannot but give a numerical approximation for this expression, which is $0.08122410\cdots$. Dividing this by the normalizing constant, $\log G$, we find

$$AF_N(\mathscr{A}) = 0.168790\cdots.$$

2.6.2. PATTERN $\mathscr{B}$

There is one region of pattern $\mathscr{B}$, in the strip $R_2^{-/+}$, with a deviant (see 2.6.1) measure, which is $\frac{1}{2}\log\frac{26}{25}$. The measures of all other $\mathscr{B}$-regions can be expressed as functions of a. These are

2

$$\begin{array}{lll}
R_a^{-/-}, a\geq 3: & \frac{1}{2}\log\left(a-\frac{1}{a}\right) & +\frac{1}{2}\log\frac{a-\sqrt{a^2-4}}{2}; \\
R_a^{-/+}, a\geq 3: & -\frac{1}{2}\log\left(a+\frac{1}{a}\right) & +\frac{1}{2}\log\frac{a^2+2+\sqrt{a^4+4}}{2a}; \\
R_a^{+/-}, a\geq 3: & -\frac{1}{2}\log\left(a-\frac{1}{a}\right) & +\frac{1}{2}\log\frac{a^2-2+\sqrt{a^4+4}}{2a}; \\
R_a^{+/+}, a\geq 2: & \underbrace{\frac{1}{2}\log\left(a+\frac{1}{a}\right)}_{\text{part I}} & \underbrace{+\frac{1}{2}\log\frac{-a+\sqrt{a^2+4}}{2}}_{\text{part II}}.
\end{array}$$

It appears that in part I all terms are mutually canceled, except for $R_2^{+/+}$, where we also have to take the corresponding value in part II into account, yielding the sum value $\frac{1}{2}\log 5-\frac{1}{2}\log 2+\frac{1}{2}\log(\sqrt{2}-1)=\frac{1}{2}\log\frac{5\sqrt{2}-5}{2}$. Adding this number to the aforementioned $\frac{1}{2}\log\frac{26}{25}$, we find the sum measure of the regions $\mathscr{B}$ to be:

$$\frac{1}{2}\log\frac{13\sqrt{2}-13}{5}+\sum_{a=3}^{\infty}(\text{expressions in part II}). \tag{2.27}$$

We can write the sum of the four expressions in part II as $\frac{1}{2}\log\frac{2\left(1+\sqrt{1+\frac{4}{a^4}}\right)}{\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)}$. This means that we can write (2.27) as

$$\frac{1}{2}\log\left(\frac{13\sqrt{2}-13}{5}\prod_{a=3}^{\infty}\frac{2\left(1+\sqrt{1+\frac{4}{a^4}}\right)}{\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)}\right). \tag{2.28}$$

A numerical approximation for this expression is $0.07825923\cdots$. Dividing this by the normalizing constant, we find

$$AF_N(\mathscr{B})=0.162629\cdots.$$

2.6.3. PATTERN $\mathscr{C}$

There is one region of pattern $\mathscr{C}$, in the strip $R_3^{-/+}$, with deviant (see 2.6.1) measure, which is $\frac{1}{2}\log(7+\sqrt{85})+\frac{1}{2}\log\frac{5}{81}$. The measures of all other $\mathscr{C}$-regions can be expressed as generic functions only from $a \geq 6$. These are

$$
\begin{aligned}
&R_a^{-/-}, a \geq 4: && \frac{1}{2}\log\frac{2a-1)^2-4}{(2a+\sqrt{5}-4)^2} + \frac{\log(a+\sqrt{5}-3)+\log\frac{a-\sqrt{a^2-4}}{2}}{2};\\
&R_a^{-/+}, a \geq 4: && \frac{1}{2}\log\frac{(2a+\sqrt{5}-2)^2}{(2a+1)^2-4} - \frac{\log(a^2+(\sqrt{5}-3)a+7-3\sqrt{5})-\log\frac{a^2-2+\sqrt{a^4+4}}{2}}{2};\\
&R_a^{+/-}, a \geq 6: && -\frac{1}{2}\log((2a-1)^2+4) + \frac{\log\left(4-\frac{4}{a}+\frac{2}{a^2}\right)+\log\frac{a^2+2+\sqrt{a^4+4}}{2}}{2};\\
&R_a^{+/+}, a \geq 2: && \underbrace{\frac{1}{2}\log((2a+1)^2+4)}_{\text{part I}} \underbrace{-\frac{\log(4a+4)-\log\frac{-a+\sqrt{a^2+4}}{2}}{2}.}_{\text{part II}}
\end{aligned}
$$

Unfortunately, we also have to deal with

$$
R_a^{+/-}, a \in \{3,4,5\}: \quad \underbrace{-\frac{1}{2}\log((2a-1)^2+4)}_{\text{part I}} \underbrace{+\frac{\log\frac{(2a+\sqrt{5}-2)^2}{a^2+(\sqrt{5}-1)a+3-\sqrt{5}}+\log\frac{a^2+2+\sqrt{a^4+4}}{2}}{2}.}_{\text{part II}}
$$

Both parts make the computations for pattern $\mathscr{C}$ more troublesome than for $\mathscr{A}$ and $\mathscr{B}$. Luckily, in part I all terms, including those for $a \in \{3,4,5\}$, are mutually canceled, except for $\frac{1}{2}\log\frac{45}{21+8\sqrt{5}}$ in $R_4^{-/-}$ when we sum the terms of both parts from $a = 4$. This summation seems the most convenient in the sense of restricting the number of loose constants to be summed separately. Still, there are too many of those to write them sensibly as one logarithm. They are:

$$
\begin{aligned}
&\text{from } R_a^{+/-}: && -\tfrac{1}{2}\log(2\sqrt{5}+9), -\tfrac{1}{2}\log(3\sqrt{5}+15), -\tfrac{1}{2}\log(4\sqrt{5}+23), \log(4+\sqrt{5}),\\
& && \log(6+\sqrt{5}), \log(8+\sqrt{5}), \tfrac{1}{2}\log\tfrac{11+\sqrt{85}}{2}, \tfrac{1}{2}\log\tfrac{16}{50}, \tfrac{1}{2}\log\tfrac{25}{82};\\
&\text{from } R_a^{+/+}: && -\tfrac{1}{2}\log 12, \tfrac{1}{2}\log(\sqrt{2}-1), -\tfrac{1}{2}\log 16, \tfrac{1}{2}\log\tfrac{\sqrt{13}-3}{2}.
\end{aligned}
$$

Summing all constants aforementioned yields a value of $-1.90648232\cdots$. Summing all terms of both parts up to $a = n$, we should add the value

$$
-\tfrac{1}{2}\log(n^2+4n-3)+\log(2n+\sqrt{5}-2)+\tfrac{1}{2}\log(n^2+4n+5)
$$

from $R_n^{-/+}$ and $R_n^{+/+}$ that has by then not yet been canceled by the corresponding terms in $R_{n+1}^{-/-}$ and $R_{n+1}^{+/-}$. Note that $\lim_{n\to\infty}\frac{1}{2}\log(n^2+4n-3)-\frac{1}{2}\log(n^2+4n+5)=0$ (as used in the case of pattern $\mathscr{A}$), so we have to compute the sum of the terms of part II from $a = 4$ up to n and $\log(2n+\sqrt{5}-2)$ as $n \to \infty$. It seems like we cannot write this as anything shorter than

$$
\begin{aligned}
\lim_{n\to\infty}\Bigg(\sum_{a=4}^{n}\tfrac{1}{2}\log\frac{\left(1+\frac{\sqrt{5}-3}{a}\right)\left(1+\sqrt{1+\frac{4}{a^4}}\right)\left(2-\frac{2}{a}+\frac{1}{a^2}\right)}{\left(1+\frac{\sqrt{5}-3}{a}+\frac{7-3\sqrt{5}}{a^2}\right)\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)\left(1+\frac{1}{a}\right)}\\
+\log(2n+\sqrt{5}-2)\Bigg), \qquad (2.29)
\end{aligned}
$$

which is 1.98689899⋯. So we can approximate the sum measure of the $\mathscr{C}$-regions by 1.98689899⋯ − 1.90648232⋯ ≈ 0.08041667. Dividing this by $\log G$, we find

$$AF_N(\mathscr{C}) = 0.167112\cdots.$$

2.6.4. PATTERN $\mathscr{D}$

Pattern $\mathscr{D}$ demands even more bothersome computations than pattern $\mathscr{C}$. There is also one region of pattern $\mathscr{D}$, in the strip $R_3^{-/+}$, with deviant (see 2.6.1) measure

$$\frac{1}{2}\log\frac{(3-\sqrt{5})(21+8\sqrt{5})(11+\sqrt{85})}{600},$$

which is 0.00057163⋯. The measures of all other $\mathscr{D}$-regions can be expressed as generic expressions only from $a \geq 6$. These are

$$\begin{aligned}
R_a^{-/-}:&\quad \frac{1}{2}\log(a^2+(\sqrt{5}-3)a+\tfrac{5}{2}-\tfrac{3}{2}\sqrt{5}) - \frac{\log(a+\sqrt{5}-3)-\log\frac{a-\sqrt{a^2-4}}{2}}{2};\\
R_a^{-/+}:&\quad -\frac{1}{2}\log(a^2+(\sqrt{5}-3)a+\tfrac{9}{2}-\tfrac{3}{2}\sqrt{5}) + \frac{\log(a^2+(\sqrt{5}-3)a+7-3\sqrt{5})-\log\frac{a^2+2-\sqrt{a^4+4}}{2}}{2};\\
R_a^{+/-}:&\quad \frac{1}{2}\log\frac{(2a+\sqrt{5}-2)^2}{a^2+(\sqrt{5}-1)a+\frac{1}{2}-\frac{1}{2}\sqrt{5}} - \frac{\log\left(4-\frac{4}{a}+\frac{2}{a^2}\right)+\log\frac{-a^2+2+\sqrt{a^4+4}}{2}}{2};\\
R_a^{+/+}:&\quad \underbrace{\frac{1}{2}\log\frac{a^2+(\sqrt{5}-1)a+\frac{5}{2}-\frac{1}{2}\sqrt{5}}{(2a+\sqrt{5})^2}}_{\text{part I}} + \underbrace{\frac{\log(4a+4)+\log\frac{-a+\sqrt{a^2+4}}{2}}{2},}_{\text{part II}}
\end{aligned}$$

for $a \geq 4$ in the cases of $R_a^{-/-}$ and $R_a^{-/+}$; $a \geq 6$ in the case $R_a^{+/-}$ and $a \geq 2$ in the case $R_a^{+/+}$. As in the case of pattern $\mathscr{C}$, there are some deviant expressions, viz.

$$R_a^{+/-}, a \in \{3,4,5\}: \tfrac{1}{2}\log\frac{a^2+(\sqrt{5}-1)a+3-\sqrt{5}}{a^2+(\sqrt{5}-1)a+\frac{1}{2}-\frac{1}{2}\sqrt{5}} - \tfrac{1}{2}\log\frac{-a^2+2+\sqrt{a^4+4}}{2}.$$

Again, both parts make the computations more troublesome than for $\mathscr{A}$ and $\mathscr{B}$. Luckily, in part I almost all terms, including those for $a \in \{3,4,5\}$, are mutually canceled when we sum the terms of both parts from $a = 4$. This summation seems the most convenient, similar to the case of pattern $\mathscr{A}$. The number of loose constants to be summed separately are:

from $R_a^{+/-}$: $\tfrac{1}{2}\log(2\sqrt{5}+9)$, $\tfrac{1}{2}\log(3\sqrt{5}+15)$, $\tfrac{1}{2}\log(4\sqrt{5}+23)$, $-\tfrac{1}{2}\log(-\tfrac{7}{2}+\tfrac{1}{2}\sqrt{85})$, $\tfrac{1}{2}\log\tfrac{50}{16}$, $\tfrac{1}{2}\log\tfrac{82}{25}$;

from $R_a^{+/+}$: $-\log(4+\sqrt{5})$, $-\log(6+\sqrt{5})$, $-\log(8+\sqrt{5})$, $\tfrac{1}{2}\log 12$, $\tfrac{1}{2}\log 16$, $\tfrac{1}{2}\log(\sqrt{2}-1)$, $\tfrac{1}{2}\log(\tfrac{9}{2}+\tfrac{3}{2}\sqrt{5})$, $\tfrac{1}{2}\log(\tfrac{1}{2}\sqrt{13}-\tfrac{3}{2})$.

2

Summing all constants aforementioned yields a value of $2.03969322\cdots$. Summing all terms of both parts up to $a = n$, we should add the value

$$-\tfrac{1}{2}\log(n^2+(\sqrt{5}-1)n+\tfrac{1}{2}-\tfrac{1}{2}\sqrt{5})-\log(2n+\sqrt{5})+\tfrac{1}{2}\log(n^2+(\sqrt{5}-1)n+\tfrac{5}{2}-\tfrac{1}{2}\sqrt{5})$$

from $R_n^{-/+}$ and $R_n^{+/+}$ that has by then not yet been canceled by the corresponding terms in $R_{n+1}^{-/-}$, $R_{n+1}^{-/+}$ and $R_{n+1}^{+/-}$. Note that $\lim_{n\to\infty} -\frac{1}{2}\log(n^2+(\sqrt{5}-1)n+\frac{1}{2}-\frac{1}{2}\sqrt{5})+\frac{1}{2}\log(n^2+(\sqrt{5}-1)n+\frac{5}{2}-\frac{1}{2}\sqrt{5}) = 0$, so we have to compute the sum of the terms of part II from $a = 4$ up to n and $-\log(2n+\sqrt{5})$ as $n\to\infty$. It seems like we cannot write this as anything shorter than

$$\lim_{n\to\infty}\left(\sum_{a=4}^{n}\tfrac{1}{2}\log\frac{\left(1+\frac{\sqrt{5}-3}{a}+\frac{7-3\sqrt{5}}{a^2}\right)\left(1+\sqrt{1+\frac{4}{a^4}}\right)(4+\frac{4}{a})}{\left(1+\frac{\sqrt{5}-3}{a}\right)\left(2-\frac{2}{a}+\frac{1}{a^2}\right)\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)}-\log(2n+\sqrt{5})\right), \tag{2.30}$$

which is $-1.95667971\cdots$. We approximate the sum measure of the $\mathscr{D}$-regions by $2.03969322\cdots - 1.95667971 = 0.08301351\cdots$. Dividing this by $\log G$, we find

$$AF_N(\mathscr{D}) = 0.172509\cdots.$$

2.6.5. PATTERN $\mathscr{E}$

There is one region of pattern $\mathscr{E}$, in the strip $R_2^{-/+}$, with a deviant (see 2.6.1) measure, which is $\log(3+\sqrt{5})-\frac{1}{2}\log 26$. The measures of all other $\mathscr{E}$-regions can be expressed as functions of a. These are

$$\begin{array}{lll}
R_a^{-/-}, a\geq 3: & \frac{1}{2}\log\left(a^2-1\right) & +\frac{1}{2}\log\frac{a-\sqrt{a^2-4}}{2a};\\
R_a^{-/+}, a\geq 3: & -\frac{1}{2}\log\left(a^2-1\right) & +\frac{1}{2}\log\frac{a^2-2+\sqrt{a^4+4}}{2};\\
R_a^{+/-}, a\geq 3: & -\frac{1}{2}\log\left(a^2+1\right) & +\frac{1}{2}\log\frac{a^2+2+\sqrt{a^4+4}}{2};\\
R_a^{+/+}, a\geq 2: & \underbrace{\frac{1}{2}\log\left(a^2+1\right)}_{\text{part I}} & \underbrace{+\frac{1}{2}\log\frac{-a+\sqrt{a^2+4}}{2a}.}_{\text{part II}}
\end{array}$$

It appears that in part I all terms are mutually canceled, except for $R_2^{+/+}$, where we also have to take the corresponding value in part II into account, with the sum value $\frac{1}{2}\log 5 + \frac{1}{2}\log(-\frac{1}{2}+\frac{1}{2}\sqrt{2})$. Adding this number to the aforementioned two constants, we find the sum measure of the regions $\mathscr{E}$ to be:

$$\frac{1}{2}\log\frac{35+15\sqrt{5}}{26(\sqrt{2}+1)}+\sum_{a=3}^{\infty}(\text{expressions in part II}). \tag{2.31}$$

Attentive inspection of the four expressions in part II reveals that they have the same sum value as part II of pattern $\mathscr{B}$. This means that we can write (2.31) as

$$\frac{1}{2}\log\left(\frac{35+15\sqrt{5}}{26(\sqrt{2}+1)}\prod_{a=3}^{\infty}\frac{2\left(1+\sqrt{1+\frac{4}{a^4}}\right)}{\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)}\right). \tag{2.32}$$

A numerical approximation for this expression, which is $0.08517144\cdots$. Dividing this by $\log G$, we find

$$AF_N(\mathscr{E}) = 0.176993\cdots.$$

2.6.6. PATTERN $\mathscr{F}$

There is one region of pattern $\mathscr{F}$, in the strip $R_3^{-/-}$, with deviant (see 2.6.1) measures, which is $\frac{1}{2}\log(9\sqrt{5}+20)-\frac{1}{2}\log 40$.
The measures of all other $\mathscr{F}$-regions can be expressed as functions of a. These are

$$\begin{array}{lll}
R_a^{-/-}, a\geq 4: & \frac{1}{2}\log\frac{(2a+\sqrt{5}-3)^2}{(2a^2-2)(2a^2+2(\sqrt{5}-3)a+5-3\sqrt{5})} & -\log\frac{a-\sqrt{a^2-4}}{2}; \\
R_a^{-/+}, a\geq 3: & \frac{1}{2}\log\frac{(2a^2-2)(2a^2+2(\sqrt{5}-3)a+9-3\sqrt{5})}{(2a+\sqrt{5}-3)^2} & +\frac{1}{2}\log\frac{-a^2+\sqrt{a^4+4}}{2}; \\
R_a^{+/-}, a\geq 3: & \frac{1}{2}\log\frac{(2a^2-2)(2a^2+2(\sqrt{5}-1)a+1-\sqrt{5})}{(2a+\sqrt{5}-1)^2} & +\frac{1}{2}\log\frac{-a^2+\sqrt{a^4+4}}{2}; \\
R_a^{+/+}, a\geq 2: & \underbrace{\frac{1}{2}\log\frac{(2a+\sqrt{5}-1)^2}{(2a^2-2)(2a^2+2(\sqrt{5}-1)a+5-\sqrt{5})}}_{\text{part I}} & \underbrace{-\log\frac{a-\sqrt{a^2+4}}{2}}_{\text{part II}}.
\end{array}$$

All terms in part I are mutually canceled. If we sum all terms of both parts from $a=3$, we have to add a corrective $\frac{1}{2}\log(\frac{7}{2}-\frac{3}{2}\sqrt{5})$ in $R_3^{-/-}$, $\frac{1}{2}\log 8$ in $R_3^{-/+}$ and of $-\frac{1}{2}\log(3-2\sqrt{2})-\frac{1}{2}\log 5$ in $R_2^{+/+}$. Finally, summing up to $a=n$, we should add the value $\frac{1}{2}(\log(2n^2+2(\sqrt{5}-1)n+1-\sqrt{5})-\log(2n^2+2(\sqrt{5}-1)n+5-\sqrt{5}))$ from $R_n^{+/-}$ and $R_n^{+/+}$ that has by then not yet been canceled by the corresponding terms in $R_{n+1}^{-/-}$ and $R_{n+1}^{-/+}$. However, as $\lim_{n\to\infty}$, this difference approaches 0, so for the infinite summation this makes no difference. So far, having summed the aforementioned constants, we have found the sum measure of the regions $\mathscr{F}$ to be:

$$\frac{1}{2}\log\frac{(5+3\sqrt{5})(3+2\sqrt{2})}{50}+\sum_{a=3}^{\infty}(\text{expressions in part II}). \tag{2.33}$$

It appears that part II is completely the same as part II of pattern $\mathscr{A}$, and we can write (2.33) as

$$\frac{1}{2}\log\frac{5+3\sqrt{5}}{50}+\log\left((\sqrt{2}+1)\prod_{a=3}^{\infty}\frac{\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)}{2\left(1+\sqrt{1+\frac{4}{a^4}}\right)}\right). \tag{2.34}$$

2

A numerical approximation for this expression is $0.07312636\cdots$. Dividing this by $\log G$, we find

$$AF_N(\mathscr{F}) = 0.151962\cdots.$$

Summarizing, we have obtained the following result.

Theorem 5. *For the nearest integer continued fraction, the asymptotic frequencies of the six patterns of three consecutive approximation constants are given by*

$$AF_N(\mathscr{A}) = 0.168790\cdots \qquad AF_N(\mathscr{B}) = 0.162629\cdots \qquad AF_N(\mathscr{C}) = 0.167112\cdots,$$
$$AF_N(\mathscr{D}) = 0.172509\cdots \qquad AF_N(\mathscr{E}) = 0.176993\cdots \qquad AF_N(\mathscr{F}) = 0.151962\cdots.$$

At first sight, these numbers seem quite unremarkable; they are all different, and that's it. But something interesting happens when we add the expressions for patterns $\mathscr{A}$ and $\mathscr{F}$, for patterns $\mathscr{B}$ and $\mathscr{E}$, and for patterns $\mathscr{C}$ and $\mathscr{D}$. First, adding (2.26) and (2.34), belonging to patterns $\mathscr{A}$ and $\mathscr{F}$, we get

$$\frac{1}{2}\log\frac{235\sqrt{5}-525}{2}+\log\left((\sqrt{2}+1)\prod_{a=3}^{\infty}\frac{\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)}{2\left(1+\sqrt{1+\frac{4}{a^4}}\right)}\right)$$

$$+\frac{1}{2}\log\frac{5+3\sqrt{5}}{50}+\log\left((\sqrt{2}+1)\prod_{a=3}^{\infty}\frac{\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)}{2\left(1+\sqrt{1+\frac{4}{a^4}}\right)}\right),$$

which equals

$$\log\left(g^3\left((\sqrt{2}+1)\prod_{a=3}^{\infty}\frac{\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)}{2\left(1+\sqrt{1+\frac{4}{a^4}}\right)}\right)^2\right).$$

A close inspection of this last expression reveals that it equals

$$3\log G+2\cdot\sum_{a=2}^{\infty}\log\frac{\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)}{2\left(1+\sqrt{1+\frac{4}{a^4}}\right)}, \tag{2.35}$$

which is exactly the same as the sum of the expressions for patterns $\mathscr{A}$ and $\mathscr{F}$ in the OCF case.

Adding (2.28) and (2.32), belonging to patterns $\mathscr{B}$ and $\mathscr{E}$, yields

$$\frac{1}{2}\log\left(\frac{13\sqrt{2}-13}{5}\prod_{a=3}^{\infty}\frac{2\left(1+\sqrt{1+\frac{4}{a^4}}\right)}{\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)}\right)$$

$$+\frac{1}{2}\log\left(\frac{35+15\sqrt{5}}{26(\sqrt{2}+1)}\prod_{a=3}^{\infty}\frac{2\left(1+\sqrt{1+\frac{4}{a^4}}\right)}{\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)}\right),$$

which equals

$$\log\left(G^2(\sqrt{2}-1)\prod_{a=3}^{\infty}\frac{2\left(1+\sqrt{1+\frac{4}{a^4}}\right)}{\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)}\right). \tag{2.36}$$

Completely similarly to the case of patterns $\mathscr{A}$ and $\mathscr{F}$, we can rewrite (2.36) so as to find

$$-\log G+\sum_{a=2}^{\infty}\log\frac{2\left(1+\sqrt{1+\frac{4}{a^4}}\right)}{\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)}, \tag{2.37}$$

which is exactly the same as the sum of the expressions for patterns $\mathscr{B}$ and $\mathscr{E}$ in the OCF case.
Since the sum of all expressions for patterns $\mathscr{A}$ through $\mathscr{F}$ equals $\log G$, the sum of the patterns $\mathscr{C}$ and $\mathscr{D}$ equals $\log G$ - (2.35) - (2.37), equalling

$$-\log G+\sum_{a=2}^{\infty}\log\frac{2\left(1+\sqrt{1+\frac{4}{a^4}}\right)}{\left(1+\sqrt{1-\frac{4}{a^2}}\right)\left(1+\sqrt{1+\frac{4}{a^2}}\right)}, \tag{2.38}$$

which is exactly the same as the sum of the expressions for patterns $\mathscr{C}$ and $\mathscr{D}$ in the OCF case.

2.7. DISCUSSION - CONNECTING THE RESULTS WITH OUR UNDERSTANDING OF THE CONTINUED FRACTIONS INVOLVED

In the previous sections we showed that in the case of both the OCF and the NICF the distribution of asymptotic frequencies (as defined above) is more even than in the case of the RCF. Moreover, we showed that the two different values occurring in the case of the OCF were distributed over the six patterns in a way completely similar to the RCF case. Finally, the six different values in the case of the NICF proved to be remarkably connected with the values belonging to the OCF.

Although some adjustments were needed and the calculations were more troublesome, the approach we took in computing the asymptotic frequencies in the RCF case proved to be applicable to the other two continued fractions. In fact, it is applicable to all continued fractions of which we know the natural extension, such as the Nakada α-expansions (for $\sqrt{2}-1\leq\alpha\leq 1$), all Rosen fractions, the odd and the even continued fraction expansion et cetera.

The fact that in the case of the RCF we have

$$\mathrm{AF}(\mathscr{A}) = \mathrm{AF}(\mathscr{F}), \quad \mathrm{AF}(\mathscr{B}) = \mathrm{AF}(\mathscr{E}), \quad \text{and } \mathrm{AF}(\mathscr{C}) = \mathrm{AF}(\mathscr{D}), \tag{2.39}$$

follows from the properties of the natural extension of the RCF, see also [N], with natural extension map $\mathscr{T} : \Omega = [0,1) \times [0,1] \to \Omega$ given by

$$\mathscr{T}(x,y) = \left(T(x) = \tfrac{1}{x} - a, \frac{1}{a+y} \right), \quad \text{for } (x,y) \in \Omega, \text{ with } \tfrac{1}{a+1} < x \le \tfrac{1}{a}, 0 \le y \le 1,$$

and $\mathscr{T}(0,y) = (0,y)$, for $0 \le y \le 1$; see also (2.1) on page 6.

As a natural extension we have that $\mathscr{T} : \Omega \to \Omega$ is bijective almost surely. So apart from a set of Lebesgue measure zero, $\mathscr{T}^{-1} : \Omega \to \Omega$ exists, and is also a bijection (almost surely). Note that

$$\mathscr{T}^{-1}(x,y) = \left(\frac{1}{a+x}, \tfrac{1}{y} - a \right), \quad \text{for } (x,y) \in \Omega, \text{ with } 0 \le x \le 1, \tfrac{1}{a+1} < y \le \tfrac{1}{a},$$

so essentially we again find the natural extension of the Gauss map T. We can understand this as follows: if $(x,y) \in \Omega$, where $x = [0; a_1, a_2, \dots]$ and $y = [0; a_0, a_{-1}, a_{-2}, \dots]$ are the RCF-expansions of x resp. y, then - apart from a set of measure zero - we can write (x,y) symbolically also as a bi-infinite sequence

$$[\dots, a_{-3}, a_{-2}, a_{-1}, a_0; a_1, a_2, a_3, \dots],$$

where for each $m \in \mathbb{Z}$ the expression $[0; a_m, a_{m+1}, \dots]$ is the RCF-expansion of some point $x_m \in [0,1]$.

Now $\mathscr{T}$ acts on $[\dots, a_{-3}, a_{-2}, a_{-1}, a_0; a_1, a_2, a_3, \dots]$ as the left-shift τ, i.e. we can write $\mathscr{T}(x,y)$ as

$$\tau\left([\dots, a_{-3}, a_{-2}, a_{-1}, a_0; a_1, a_2, a_3, \dots]\right) = [\dots, a_{-3}, a_{-2}, a_{-1}, a_0, a_1; a_2, a_3, \dots].$$

Note that for each $m \in \mathbb{Z}$ the expression $[0; a_m, a_{m-1}, \dots]$ is the RCF-expansion of some point $y_m \in [0,1]$. Since in $\mathscr{T}^{-1}$ time runs "backwards," we can view $\mathscr{T}^{-1}$ also as the left-shift on

$$[\dots, a_3, a_2, a_1; a_0, a_{-1}, a_{-2}, a_{-3}, \dots].$$

Since $\mathscr{T}$ and $\mathscr{T}^{-1}$ are essentially the same algorithm (but with "time n running forward resp. backward") we find the equalities (2.39).

Due to the way the OCF and also Minkowski's Diagonal continued fraction (DCF) expansion can be described as S-expansions, it can be shown that these continued fraction algorithms have the same property: the second coordinate map of $\mathscr{T}^{-1}$ of the corresponding natural extension map $\mathscr{T}$ behaves essentially the same as the first coordinate map of $\mathscr{T}$; see [K1, BK1, BK2, K2]. This is essentially due to symmetry in the singularisations yielding these continued fraction expansions. Consequently, (2.39) also holds for these two continued fraction algorithms.

In the previous section we found that

$$\mathrm{AF}_N(\mathscr{A}) \neq \mathrm{AF}_N(\mathscr{F}), \quad \mathrm{AF}_N(\mathscr{B}) \neq \mathrm{AF}_N(\mathscr{E}), \quad \mathrm{AF}_N(\mathscr{C}) \neq \mathrm{AF}_N(\mathscr{D}), \tag{2.40}$$

and also that

$$\frac{\mathrm{AF}_N(\mathscr{A}) + \mathrm{AF}_N(\mathscr{F}))}{2} = \mathrm{AF}_O(\mathscr{A}), \tag{2.41}$$

and similarly for the other two means.

Although we do not have an explanation for these phenomena, we think that they are due to the fact that the NICF and *Hurwitz' singular continued fraction* (see [P], §44) are closely related; again see [K1], Section (2.11). In the latter case, the continued fraction map $T_H : [g-1, g) \to [g-1, g)$ defined by

$$T_H(x) = \left|\frac{1}{x}\right| - \left\lfloor \left|\frac{1}{x}\right| + 1 - g \right\rfloor, \quad \text{for } x \in [g-1, g) \setminus \{0\},$$

and $T_H(0) = 0$, is used. Nakada showed in [N] that the natural extension Ω_H of this continued fraction algorithm can be obtained by reflecting the natural extension of the NICF in the line $Y = X$ for those points in Ω_N for which $x \geq 0$, and by reflecting (x, y) in $Y = -X$ if $x < 0$; so the natural extension region of T_H is given by

$$\Omega_H = [g-1, g) \times [0, \tfrac{1}{2}].$$

One could say that Hurwitz' singular continued fraction 'suffers reflectedly' from what we noted at the end of Section 3 about the NICF: "Actually, if both regions would have $0 \leq v \leq \frac{1}{2}$ instead of $0 \leq v \leq g^2$ (in the case $\varepsilon_{n+1} = -1$) and $0 \leq v \leq g$ (in the case $\varepsilon_{n+1} = 1$), we could have applied the same maps for $R_a^{-/-}$ and $R_a^{+/+}$ as in the regular and the optimal case." It is not hard to show that the second coordinate-map of $\mathscr{T}_N$ is essentially the map T_H, **not** the map T_N, which explains (2.40). "Running back the time" now **only** yields that

$$\mathrm{AF}_N(\mathscr{A}) = \mathrm{AF}_H(\mathscr{F}), \quad \mathrm{AF}_N(\mathscr{B}) = \mathrm{AF}_H(\mathscr{E}), \quad \mathrm{AF}_N(\mathscr{C}) = \mathrm{AF}_H(\mathscr{D})$$

and that

$$\mathrm{AF}_N(\mathscr{F}) = \mathrm{AF}_H(\mathscr{A}), \quad \mathrm{AF}_N(\mathscr{E}) = \mathrm{AF}_H(\mathscr{B}), \quad \mathrm{AF}_N(\mathscr{D}) = \mathrm{AF}_H(\mathscr{C}),$$

something which is immediately clear if we view these two continued fraction algorithms as S-expansions; see the examples in [K1], Section (2.11). However, we think that an explanation of (2.41) should follow from understanding how the singularisation areas of the OCF, NICF and Hurwitz' singular continued fraction are related.

ACKNOWLEDGMENTS

We would like to thank Dion Gijswijt from Delft University of Technology and Frits Beukers from Utrecht University for their helpful suggestions in tackling the infinite log sums in this chapter using Mathematica. Moreover, we would like to thank our referee for helping us to improve our presentation and for asking us to go deeper into understanding our results from similarities and differences between the continued fractions involved, which we did in the final section.

2

REFERENCES

[B] W. Bosma, *Optimal continued fractions*, Nederl. Akad. Wetensch. Indag. Math. 49 (1987), no. 4, 353–379.

[BK1] W. Bosma and C. Kraaikamp, *Metrical Theory for Optimal Continued Fractions*, Journal of Number Theory 34 (1990), no. 3, 251–270.

[BK2] W. Bosma and C. Kraaikamp, *Optimal approximation by continued fractions*, J. Austral. Math. Soc. Ser. A50 (1991), no. 3, 481–504.

[BJW] W. Bosma, H. Jager and F. Wiedijk, *Some metrical observations on the approximation by continued fractions* , Nederl. Akad. Wetensch. Indag. Math. 45 (1983), no. 3, 281–299.

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

[Bor] É. Borel, *Contribution à l'analyse arithmétique du continu*, Journal de Mathématiques Pures et Appliquées, 5e série, **9** (1903), 329–375.

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

[J1] H. Jager, *The distribution of certain sequences connected with the continued fraction*, Indag. Math. Series A **89**, 1986, 61-69.

[J2] H. Jager, *A metrical result on the appoximation by continued fractions*, Mathematics of Computation, vol. 81, no. 280, 2012, p. 2377-2382.

[JJ] H. Jager and J. de Jonge, *On the approximation by three consecutive continued fraction convergents*, Indag. Math. (N.S.) 25 (2014), no. 4, 816–824.

[JK] H. Jager and C. Kraaikamp, *On the approximation by continued fractions*, Nederl. Akad. Wetensch. Indag. Math. 51 (1989), no. 3, 289–307.

[K] J. F. Koksma, *Diophantische Approximationen*, Verlag Julius Springer, Berlin 1937.

[K1] C. Kraaikamp, *A new class of continued fraction expansions*, Arith. 57 (1991), no. 1, 1–39.

[K2] C. Kraaikamp, *Statistic and ergodic properties of Minkowski's diagonal continued fraction*, Theoret. Comput. Sci. 65 (1989), no. 2, 197–212.

[L] A.M. Legendre, *Essay sur la théorie des nombres*, Duprat, Paris, An VI, 1798.

[N] H. Nakada, *Metrical theory for a class of continued fraction transformations and their natural extensions*, Tokyo J. Math. 4 (1981), no. 2, 399–426.

[P] O. Perron, *Die Lehre von den Kettenbrüchen*, Chelsea, New York, 1929.

[Ri] G.J. Rieger, *Mischung und Ergodizität bei Kettenbrüchen nach nächsten Ganzen*, J. Reine Angew. Math. 310 (1979), 171–181.

[Ro] A.M. Rockett, *The metrical theory of continued fraction to the nearest integer*, Acta Arith. 38 (1980), no. 2, 97–103.

[T1] J.C. Tong, *The conjugate property of the Borel theorem on Diophantine approximation*, Math. Z. 184 (1983), no. 2, 151–153.

[T2] J.C. Tong, *Approximation by nearest integer continued fractions, II*, Math. Scand. 74 (1994), no. 1, 17–18.

[V] K.Th. Vahlen, *Ueber Näherungswerte und Kettenbrüche*, J. Reine Angew. Math. 115 (1895), 221–233.

[W] H.C. Williams, *On mid-period criteria for the nearest integer continued fraction expansion of* $\sqrt{D}$, Utilitas Math. 27 (1985), 169–185.

[WB] H.C. Williams and P.A. Buhr, *Calculation of the regulator of* $Q(\sqrt{D})$ *by use of the nearest integer continued fraction algorithm*, Math. Comp. 33 (1979), no. 145, 369–381.

2

3

THE NATURAL EXTENSION OF NAKADA'S α-EXPANSIONS

3.1. NAKADA'S α-EXPANSIONS

In 1981, Hitoshi Nakada introduced in [9] a family of continued fraction maps

$$T_\alpha : [\alpha-1,\alpha] \to [\alpha-1,\alpha), \text{ with } \alpha \in [0,1],$$

and obtained and studied their natural extensions for $\alpha \in [1/2, 1]$; see [9], [4] and [6]. Nakada defined these maps by

$$T_\alpha(x) := \frac{\varepsilon(x)}{x} - a(x), \quad x \neq 0; \quad T_\alpha(0) := 0, \tag{3.1}$$

where $\varepsilon(x) = \text{sign}(x)$ and $a : (-1,1) \to \mathbb{N} \cup \{\infty\}$ is defined by

$$a(x) := \left\lfloor \left|\frac{1}{x}\right| + 1 - \alpha \right\rfloor, \quad x \neq 0; \quad a(0) := \infty.$$

In case $\alpha = 1$, we have the *regular* (or: *simple*) *continued fraction expansion* (RCF), while the case $\alpha = \frac{1}{2}$ is the *nearest integer continued fraction expansion* (NICF); see Chapter 2. In the case $\alpha = 0$, we have the *by-access continued fraction expansion*; see for instance [14].

For $\alpha \in [0,1]$, if $T_\alpha^{n-1}(x) \neq 0$ for $n \in \mathbb{N}$, we define

$$\varepsilon_n = \varepsilon_n(x) := \varepsilon\left(T_\alpha^{n-1}(x)\right), \quad \text{and} \quad a_n = a_n(x) := \left\lfloor \left|\frac{1}{T_\alpha^{n-1}(x)}\right| + 1 - \alpha \right\rfloor.$$

Applying (3.1), we obtain

$$x = T_\alpha^0(x) = \cfrac{\varepsilon_1}{a_1 + T_\alpha(x)} = \cfrac{\varepsilon_1}{a_1 + \cfrac{\varepsilon_2}{a_2 + T_\alpha^2(x)}} = \cdots = \cfrac{\varepsilon_1}{a_1 + \cfrac{\varepsilon_2}{a_2 + \cfrac{\varepsilon_3}{a_3 + \ddots}}}, \tag{3.2}$$

which we will write as $x = [0; \varepsilon_1 a_1, \varepsilon_2 a_2, \varepsilon_3 a_3, \ldots]$, like we did in the previous chapters.

Let $\Omega_\alpha \in \mathbb{R}^2$ be the domain for the natural extension of T_α. We define $\mathcal{T}_\alpha : \Omega_\alpha \to \Omega_\alpha$ by

$$\mathcal{T}_\alpha(x, y) := \left(T_\alpha(x), \frac{1}{a(x) + \varepsilon(x) y} \right).$$

Without giving Ω_α yet (the determination of which is the main topic of this chapter), we note that – apart from a set of (Lebesgue) measure 0 – the map $\mathcal{T}_\alpha$ is bijective almost everywhere on Ω_α, with inverse map

$$\mathcal{T}_\alpha^{-1}(x, y) := \left(\frac{\varepsilon(x)}{x + a(x)}, \frac{1 - a(x) y}{\varepsilon(x) y} \right) = \left(\frac{\varepsilon(x)}{x + a(x)}, \frac{\varepsilon(x)}{y} - \varepsilon(x) a(x) \right);$$

note how the coordinates giving information about 'pasts' and 'futures' of the continued fraction expansion traded places with those of $\mathcal{T}_\alpha$.

A very helpful tool in the arithmetic of t_n and v_n is the well-known *Möbius transformation*:
Let $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, with $a, b, c, d \in \mathbb{Z}, ad - bc = \pm 1$. The Möbius transformation $M : \mathbb{R} \cup \{\infty\} \to \mathbb{R} \cup \{\infty\}$ associated with M[1] is defined by

$$M(x) := \frac{ax + b}{cx + d}, \quad x \neq \frac{-d}{c}, \infty \quad \text{and} \quad M\left(\frac{-d}{c}\right) = \infty; \quad M(\infty) = \frac{a}{c}.$$

So, applying Möbius transformations, we can rephrase (4.2) as

$$x = \begin{pmatrix} 0 & \varepsilon_1 \\ 1 & a_1 \end{pmatrix}(t_1) = \begin{pmatrix} 0 & \varepsilon_1 \\ 1 & a_1 \end{pmatrix}\begin{pmatrix} 0 & \varepsilon_2 \\ 1 & a_2 \end{pmatrix}(t_2) = \begin{pmatrix} 0 & \varepsilon_1 \\ 1 & a_1 \end{pmatrix}\begin{pmatrix} 0 & \varepsilon_2 \\ 1 & a_2 \end{pmatrix}\begin{pmatrix} 0 & \varepsilon_3 \\ 1 & a_3 \end{pmatrix}\cdots.$$

Writing $M_n := \begin{pmatrix} 0 & \varepsilon_n \\ 1 & a_n \end{pmatrix}$, we also see that

$$t_n = M_{n+1}(t_{n+1}) \quad (\text{so } t_{n+1} = M_{n+1}^{-1}(t_n))$$

and generally,

$$t_n = M_{n+1} \cdots M_{n+k}(t_{n+k}) \quad (\text{and } t_{n+k} = M_{n+k}^{-1} \cdots M_{n+1}^{-1}(t_n)).$$

[1] Throughout this chapter we will omit the distinction between a matrix and its associated Möbius transformation.

Likewise,

$$v_{n+1} = M_{n+1}^T(v_n)$$

and generally

$$v_{n+k} = M_{n+k}^T \cdots M_{n+1}^T(v_n),$$

from which we derive the very useful fact that

$$t_{n+k} = M(t_n) \quad \text{if and only if} \quad v_{n+k} = (M^T)^{-1}(v_n), \tag{3.3}$$

with M a Möbius transformation.

3

The main topic of this chapter is the construction of all domains Ω_α with $\alpha \in [g^2, 1)$. We will show how to obtain these by transforming Ω_α into $\Omega_{\alpha'}$, with $\alpha' < \alpha$, starting with $\alpha = 1$. We will also show that for $\alpha \le (\sqrt{10}-2)/3$, Ω_α becomes rapidly extremely intricate, finally making it senseless if not impossible to give the associated set of pairs (t, v) explicitly when α approaches g^2. Our approach is based on *singularisations* and *insertions* in the continued fraction expansion of α; see [1] and also [6], where singularisations are used to define a suitable induced transformation. For $A, B \in \mathbb{N}$, $B \ge 2$, $\xi \in \mathbb{R}$ and $\varepsilon \in \{-1, 1\}$, the first of these operations is based on the equation

$$A + \cfrac{1}{1 + \cfrac{\varepsilon}{B + \xi}} = A + 1 - \frac{\varepsilon}{B + \varepsilon + \xi} \quad \text{(partial quotient 1 between } A \text{ and } B \text{ is \textit{singularised});} \tag{3.4}$$

and the second one on

$$A + \frac{\varepsilon}{B + \xi} = A + \varepsilon - \cfrac{\varepsilon}{1 + \cfrac{1}{B - 1 + \xi}} \quad \text{(partial quotient 1 is \textit{inserted} between } A \text{ and } B). \tag{3.5}$$

In terms of Möbius transformations, (3.4) is similar to

$$\begin{pmatrix} 0 & \varepsilon_n \\ 1 & a_n \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & \varepsilon_{n+2} \\ 1 & a_{n+2} \end{pmatrix} = \begin{pmatrix} 0 & \varepsilon_n \\ 1 & a_n + 1 \end{pmatrix} \begin{pmatrix} 0 & -\varepsilon_{n+2} \\ 1 & a_{n+2} + \varepsilon_{n+2} \end{pmatrix}$$

and (3.5) to

$$\begin{pmatrix} 0 & \varepsilon_n \\ 1 & a_n \end{pmatrix} \begin{pmatrix} 0 & \varepsilon_{n+1} \\ 1 & a_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & \varepsilon_n \\ 1 & a_n + \varepsilon_{n+1} \end{pmatrix} \begin{pmatrix} 0 & -\varepsilon_{n+1} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & \varepsilon_{n+1} \\ 1 & a_{n+1} - \varepsilon_{n+1} \end{pmatrix}.$$

Singularisations *shorten* the (regular) continued fraction expansion of numbers $x \in \mathbb{R}$. Considering (4.2), this implies a *loss* of convergents (as will be illustrated in the next section). Insertions obviously *lengthen* a continued fraction expansion, yielding so-called *mediants*, that are usually of lesser approximative quality compared to convergents. It be remarked, however, that inserting partial quotients 1 in a given continued fraction expansion involves the possibility of creating points (t, v) lying in a part of Ω_1 that had been removed in a previous stage of constructing. As long as we can compensate each insertion with a singularisation, we can avoid this. Applying this *compensated insertion* is actually the base of our construction of Ω_α, from $\alpha = g$ downwards, that we will gradually unfold in the next sections.

It is easy to see that compensation is only possible as long as an insertion involves the introduction of an extra 1, that is, if $B-1$ in (3.5) equals 1, implying $B=2$. In this case we deal with

$$A+\cfrac{\varepsilon_1}{2+\cfrac{\varepsilon_2}{C+\xi}} = A+\varepsilon_1-\cfrac{\varepsilon_1}{1+\cfrac{1}{1+\cfrac{\varepsilon_2}{C+\xi}}} = A+\varepsilon_1-\cfrac{\varepsilon_1}{2-\cfrac{\varepsilon_2}{C+\varepsilon_2+\xi}}.$$

Of course, similar equations could be given in terms of Möbius transformations, but at this point we want to stress that either notation has its pros and cons as to readibility and illuminating power. Depending on the context, for our computations we will use either continued fractions or Möbius transformations.

In the next sections we will step-by-step transform the domain $\Omega_1 := [0,1)\times[0,1]$ for the natural extension of the regular continued fraction to the domain Ω_α of the natural extension of any α-expansion for $\alpha\in[g^2,1)$. Note that $t_n=1$ only occurs in the case $x=1$ and $n=0$, associated with the single point $(1,0)$, that is sent to $(0,1)$ under T_1. So, in view of (3.1), one should actually define $\Omega_1 := [0,1)\times[0,1]\cup\{(1,0)\}$. We will suppress subtleties like these in similar cases throughout this chapter.

We will often use some well-known equations involving g and G, such as $g^2 = 1-g$ and $1/g = G = g+1$. Although we let α decrease continuously from 1 downwards, we speak of 'step-by-step', as the generic form of Ω_α varies over several intervals, according to feasability for singularisation (possible in Section 3.3 only) or compensated insertion. In Section 3.2 we will explain how to determine these intervals. We remark that the results for $\alpha\ge\sqrt{2}-1$ are not new (see for instance [6] or [9]). However, our approach is best introduced by starting at $\alpha=1$ and has some illuminative qualities. For one thing, we will show how our construction of Ω_α calls for the term *quilting*, as introduced in [7]. More importantly, some interesting results on α-expansions come with the determination of Ω_α. Since the construction of Ω_α will prove to be quite intricate, we will present these results mostly separately at the end of this chapter. In Section 3.8 we will go into ergodic properties of $\mathcal{T}_\alpha$. Representing the collection of subsets of Ω_α by $\mathcal{B}$ and defining μ_α as the probability measure with density

$$\frac{1}{N_\alpha}\cdot\frac{1}{(1+tv)^2}$$

on $(\Omega_\alpha,\mathcal{B})$ (with N_α a normalising constant), we will also show that $(\Omega_\alpha,\mu_\alpha,\mathcal{T}_\alpha)$ is an ergodic dynamical system.

Finally, in Section 3.9 we will extend a result of Rie Natsui ([12]), based on the well-known theorem of Legendre in the theory of regular continued fractions; see for instance [2]:

Theorem 6. *Let $x\in\mathbb{R}\backslash\mathbb{Q}$ and $p/q\in\mathbb{Q}$, with $\gcd(p,q)=1$ and $q>0$, such that $q^2|x-p/q|<1/2$. Then there exists a non-negative integer n such that $p/q=p_n/q_n$, in words: p/q is a regular convergent of x. This constant $1/2$ is best possible in the sense that for every $\varepsilon>0$*

an irrational x and a rational number p/q, with $\gcd(p,q)=1$ *and* $q>0$*, exist such that* p/q *is not a regular convergent and* $q^2|x-p/q|<1/2+\varepsilon$*.*

In [12], Natsui proves the existence of the α*-Legendre constant* for α-continued fractions, defined by

$$L(\alpha) := \sup\{c>0 : q^2\left|x-\tfrac{p}{q}\right| < c, \gcd(p,q)=1 \Rightarrow \tfrac{p}{q}=\tfrac{p_n}{q_n}, n\geq 0\}$$

for $0<\alpha\leq 1$; here p_n/q_n is an α-convergent of x. She expands the result given by Ito ([5]; see also [6]) for values $1/2\leq\alpha\leq 1$ to giving explicit values of $L(\alpha)$ for $\sqrt{2}-1\leq\alpha<1/2$. In Section 3.9 we will augment this result with the interval $[g^2,\sqrt{2}-1)$ and the numbers $\alpha=1/n$, for $n\in\mathbb{N}$, $n\geq 3$.

3

We will first concentrate on the construction of Ω_α for various values of α.

3.2. APPLYING SINGULARISATIONS AND INSERTIONS

One way to look at the domains Ω_α, $\alpha<1$, is that in Ω_α no points (t,v) exist for which $\alpha\leq t\leq 1$. In combination with the definition of (t_n,v_n) as $\mathcal{T}_\alpha^n(x,0)$, $n\in\mathbb{N}\cup\{0\}$, for any $x\in[\alpha-1,\alpha]$ (see Chapter 2), this forms the basis of our approach for constructing the Ω_α with $\alpha\in[g^2,1)$. We start with Ω_1 and fix α such that each time $\alpha\leq t_n\leq 1$, we replace the point (t_n,v_n) by a point (t_n^*,v_n^*) belonging to the continued fraction expansion of x after singularising. When fixing α, we use the continued fraction map T_1 of the associated ergodic system. Then we follow the same procedure, this time starting with Ω_α, fixing an $\alpha'<\alpha$ and applying T_α in order to construct $\Omega_{\alpha'}$. From Section 3.4 on, we will show how to make use of compensated insertion when $\alpha\leq g$.

We remark that under T_α the continued fraction expansion of numbers can vary considerably, according to the value of α. Denoting with x_α the α-expansion of $x\in(0,1)$, we have, for instance, $g_1=[0;\overline{1}]$ and $g_g=[0;2,\overline{-3}]$, the bar indicating infinite repetition. Constructing $\Omega_{\alpha'}$ from Ω_α, the way points are replaced depends on the way x_α changes under singularisation or (compensated) insertion. As a start, in Section 3.3 we will remove all points (t,v) for which $t=[0;1,\ldots]$, to which end mere singularisation suffices. In Sections 3.4 and 3.5 we will remove all points (t,v) for which $t=[0;2,\ldots]$ by means of a single compensated insertion. In Section 3.6 we will show that removing points (t,v) for which $t=[0;3,\ldots]$ (the ultimate one being $t=g^2$) is possible only in special cases, applying compensated insertion twice.

Not only the continued fraction expansion of a number x depends on which continued fraction map we use, the expansion of the associated numbers t_n does also. In order to determine the interval $[\alpha',\alpha]$, $\alpha'<\alpha\leq 1$, such that $t_n\in(\alpha',\alpha)$ implies $t_n=[0;1,\ldots]$, we observe that on $(0,1]$ both $\frac{1}{\alpha}$ and $1-\alpha$ are decreasing functions of α. So, for $\alpha\in(\alpha',1]$,

$$1=\left\lfloor\frac{1}{1}+1-1\right\rfloor \leq \left\lfloor\frac{1}{\alpha}+1-\alpha\right\rfloor \leq a_1(\alpha'_\alpha) = \left\lfloor\frac{1}{\alpha'}+1-\alpha\right\rfloor \leq \left\lfloor\frac{1}{\alpha'}+1-\alpha'\right\rfloor, \qquad (3.6)$$

from which we derive $a_1(\alpha'_\alpha) = 1$ if and only if $1/\alpha' + 1 - \alpha' < 2$, which is the case if and only if $g < \alpha' < 1$. This is why the next section is about the case $\alpha \in (g, 1]$. There (and more so in the sections following it) we will show that although replacing points by other points is in fact only part of the construction of the Ω_α, the heart of the construction is still the procedure sketched above, which is actually a way to determine *α-fundamental intervals* $\Delta_{n,\alpha} = \Delta_\alpha(i_1, i_2, \ldots, i_n)$ of rank n, that we define as follows:

$$\Delta_\alpha(i_1, i_2, \ldots, i_n) := \{\alpha \in [0,1] : a_1(\alpha_\alpha) = i_1, a_2(\alpha_\alpha) = i_2, \ldots, a_n(\alpha_\alpha) = i_n\},$$

where $i_j \in \mathbb{Z} \setminus \{0\}$ for each $1 \le j \le n$. In the previous paragraph, for instance, we found $\Delta_\alpha(1) = \{\alpha \in [0,1] : a_1(\alpha_\alpha) = 1\} = (g, 1]$.

3.3. THE CASE $\alpha \in (g, 1]$

We start with the natural extension domain for the regular continued fraction, the square Ω_1, and the classic *Gauss map* T_1, given by Nakada in [9]. Let $x \in [0,1]$ and $n \ge 0$ be the smallest integer for which $t_n \in [\alpha, 1)$, i.e., for which $(t_n, v_n) \in R_\alpha := [\alpha, 1) \times [0,1]$. Then the RCF of x satisfies

$$x = [0; a_1, \ldots, a_n, 1, a_{n+2}, a_{n+3}, \ldots];$$

observe that in Ω_1 we have $\varepsilon_k = 1$, $k \ge 1$. Singularising $a_{n+1} = 1$, we obtain

$$x = [0; a_1, \ldots, (a_n + 1), -(a_{n+2} + 1), a_{n+3}, \ldots].$$

Let the sequence $(t_n^*, v_n^*)_{n \ge 0}$ denote the futures and pasts of the singularised continued fraction expansion of x. Then

$$(t_n^*, v_n^*) = ([0; -(a_{n+2} + 1), a_{n+3}, \ldots], [0; a_n + 1, a_{n-1}, \ldots, a_1]),$$

while

$$(t_n, v_n) = ([0; 1, a_{n+2}, a_{n+3}, \ldots], [0; a_n, a_{n-1}, \ldots, a_1]).$$

Note that $v_n = 1/(a_n + v_{n-1})$, which is equivalent to $a_n + v_{n-1} = 1/v_n$, and so

$$v_n^* = \frac{1}{a_n + 1 + v_{n-1}} = \frac{1}{1 + \frac{1}{v_n}} = \frac{v_n}{1 + v_n}.$$

Applying (3.5) with $A = 0$, $\varepsilon = -1$, $B = a_n + 1$ and $\xi = t_{n+2}$, we get

$$t_n^* = \frac{-1}{a_n + 1 + t_{n+2}} = -1 + \frac{1}{1 + \frac{1}{a_n + t_{n+2}}} = t_n - 1.$$

Since $(t_n, v_n) \in R_\alpha$, we find that $(t^*, v^*) \in A_\alpha := [\alpha - 1, 0) \times [0, 1/2]$.

Note that the map $\mathscr{M} : R_\alpha \to A_\alpha$, defined by $\mathscr{M}(t, v) := (t - 1, v/(1 + v))$ is a bijection. In Section 3.8 we will use this to obtain ergodicity of the dynamical system $(\Omega_\alpha, \mathscr{B}, \mu_\alpha, \mathscr{T}_\alpha)$, $g \le \alpha \le 1$.

It is obvious that $t^*_{n+1} = t_{n+2}$, and since

$$v^*_{n+1} = \frac{1}{a_{n+2}+1-v^*_n} = \frac{1}{a_{n+2}+1-\frac{v_n}{1+v_n}} = \frac{1}{a_{n+2}+\frac{1}{1+v_n}} = \frac{1}{a_{n+2}+v_{n+1}} = v_{n+2},$$

we conclude that

$$(t^*_{n+1}, v^*_{n+1}) = (t_{n+2}, v_{n+2}). \tag{3.7}$$

Having singularised the first partial quotient 1 in the continued fraction expansion of x and so obtained an alternative expansion, we repeat the procedure ad infinitum. When doing so, we actually not only remove R_α from Ω_1, but $\mathscr{T}_1(R_\alpha) := (0,(1-\alpha)/\alpha] \times [1/2,1]$ as well, since $(t_{n+1}, v_{n+1}) = ((1-t_n)/t_n, 1/(1+v_n))$. Equation (3.7) implies $\mathscr{T}_1(A_\alpha) = \mathscr{T}_1^2(R_\alpha)$, so the loss of points is confined to R_α and $\mathscr{T}_1(R_\alpha)$. Removing from Ω_1 the strips R_α and $\mathscr{T}_1(R_\alpha)$ and adding A_α yields Ω_α; see Figures 3.1 and 3.2. Here we see a first glimpse of the similarity with the process of quilting as described in [7]. Expanding this metaphor, our construction in the next sections may be described as a form of recursive quilting.

We remark that Ω_α can be constructed from Ω_1 immediately, using $\mathscr{T}_1$, or from some $\Omega_{\alpha'}$, $\alpha < \alpha' < 1$, using $\mathscr{T}_{\alpha'}$. Indeed, in these cases we have something similar to (3.6):

$$\left\lfloor \frac{1}{\alpha} + 1 - \alpha' \right\rfloor < \left\lfloor \frac{1}{g} + 1 - g \right\rfloor = 2,$$

implying that changing "$t_n \in [\alpha, 1)$" in the beginning of this section into "$t_n \in [\alpha, \alpha')$" would leave the construction of Ω_α unaffected, save for the initial domain.

Rewriting $x = [0; a_1, \dots, a_n, 1, a_{n+2}, \dots]$ as $x = [0; a_1, \dots, (a_n+1), -(a_{n+2}+1), \dots]$ does *not* leave the sequence of convergents unaffected: since $p^*_k = p_k$ and $q^*_k = q_k$, $k < n$, we find, applying (1.2),

$$\begin{aligned} p^*_n &= (a_n+1)p^*_{n-1} + p^*_{n-2} = a_n p_{n-1} + p_{n-2} + p_{n-1} = p_n + p_{n-1} = p_{n+1}; \\ p^*_{n+1} &= 2p^*_n - p^*_{n-1} = a_{n+2}p_{n+1} + p_{n+1} - p_{n-1} = a_{n+2}p_{n+1} + p_n = p_{n+2}. \end{aligned}$$

Similarly, we find $q^*_n = q_{n+1}$ and $q^*_{n+1} = q_{n+2}$. We conclude that in the transformation from Ω_1 to Ω_α, $\alpha \in (g,1)$, a convergent p_n/q_n is lost every time that $(t_n, v_n) \in R_\alpha$.

Note that $\lim_{\alpha \downarrow g}(1-\alpha)/\alpha = (1-g)/g = g = \lim_{\alpha \downarrow g} \alpha$, which is mirrored in the fact that for $\alpha = g$ the left boundary of R_α coincides with the right boundary of $\mathscr{T}_1(R_\alpha)$. We now have obtained the following result:

Theorem 7. *Let* $\alpha \in (g,1)$*. Then* $\Omega_\alpha = [\alpha-1, 0) \times [0, \frac{1}{2}] \cup [0, \frac{1-\alpha}{\alpha}] \times [0, \frac{1}{2}) \cup (\frac{1-\alpha}{\alpha}, \alpha) \times [0,1]$.

This theorem was already obtained by Nakada in [9]; see also [6]. Figure 3.2 shows an example of the generic form of Ω_α, $\alpha \in (g,1)$. In Figure 3.3 we have written $-g^2$ for the equivalent $g-1$.

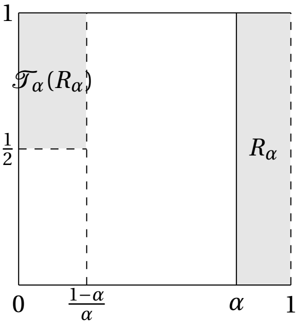


Figure 3.1: Ω_1, $\alpha = 0.8$

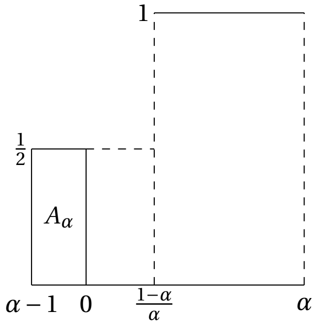


Figure 3.2: Ω_α, $\alpha = 0.8$

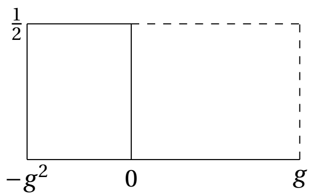


Figure 3.3: Ω_g

3.4. THE CASE $\alpha \in (\frac{1}{2}, g]$

An important implication of our construction of Ω_g in the previous section is that in continued fraction expansions associated with Ω_α, $\alpha \le g$, the partial quotient 1 is non-existent. In the current and the following section we will similarly remove all partial quotients 2 (with plus sign). We determine the interval $(\alpha', \alpha]$, $\alpha' < \alpha \le g$, such that $t_n \in (\alpha', \alpha]$ implies $t_n = [0; 2, \dots]$, in a way similar to the one at the end of Section (3.2): if $x \in [0, g)$ and $\alpha \in (x, g]$, then

$$2 = \left\lfloor \frac{1}{g} + 1 - g \right\rfloor \le \left\lfloor \frac{1}{\alpha} + 1 - \alpha \right\rfloor \le a_1(\alpha'_\alpha) = \left\lfloor \frac{1}{\alpha'} + 1 - \alpha \right\rfloor \le \left\lfloor \frac{1}{\alpha'} + 1 - \alpha' \right\rfloor,$$

from which we derive $a_1(\alpha'_\alpha) = 2$ if and only if $1/\alpha' + 1 - \alpha' < 3$, which (given that $\alpha' \in [0, g)$) is the case if and only if $\sqrt{2} - 1 < \alpha' < g$; indeed $\sqrt{2} - 1_{\sqrt{2}-1} = [0; 3, \overline{-2, -4}]$. We conclude that $\Delta_\alpha(2) = (\sqrt{2} - 1, g]$, and therefore we will now investigate the case $\alpha \in (\sqrt{2} - 1, g]$.

In the current section we will confine ourselves to the case $\alpha \in (1/2, g]$. To make a cut at $\alpha = 1/2$ is because obviously $t \in (1/2, g]$ if and only if $t = [0; 2, -a, \dots]$, with $a \in \mathbb{N}_{\ge 2}$, while $t \in (\sqrt{2} - 1, 1/2]$ if and only if $t = [0; 2, b, \dots]$, with $b \in \mathbb{N}_{\ge 2}$. We will see how the distinction shows in different ways of transforming Ω_α.

Let $x \in [g - 1, g)$ and let $n \ge 0$ be the smallest integer for which $t_n \in [\alpha, g]$, i.e., for which $(t_n, v_n) \in R_\alpha := [\alpha, g] \times [0, 1/2)$. We already know that $x = [0; \varepsilon_1 a_1, \dots, \varepsilon_n a_n, 2, -a_{n+2}, \dots]$ in this case. Since the continued fraction map of the related dynamical system is T_g, for $\alpha \in (1/2, g]$ we have

$$a_2(\alpha) = \left\lfloor \left| \frac{1}{T_g(\alpha)} \right| + 1 - g \right\rfloor = \left\lfloor \left| \frac{1}{\frac{1}{\alpha} - 2} \right| + 1 - g \right\rfloor \ge \left\lfloor \left| \frac{1}{\frac{1}{g} - 2} \right| + 1 - g \right\rfloor = G + 1 + 1 - g = 3,$$

so in this case

$$x = [0; \varepsilon_1 a_1, \dots, \varepsilon_n a_n, 2, -a_{n+2}, \dots], \quad a_{n+2} \ge 3.$$

To remove the partial quotient 2, we first insert -1, so as to write

$$x = [0; \varepsilon_1 a_1, \ldots, \varepsilon_n(a_n+1), -1, 1, -a_{n+2}, \ldots].$$

Now we singularise 1 in this expansion and get

$$x = [0; \varepsilon_1 a_1, \ldots, \varepsilon_n(a_n+1), -2, (a_{n+2}-1), \ldots].$$

Having removed the first partial quotient 2 in the continued fraction expansion of x and so obtained the alternative expansion, we repeat the procedure ad infinitum. Note that the next $+2$ could be $a_{n+2}-1$ in the rewritten expansion, since $a_{n+2} \geq 3$. In the current case we have $R_\alpha = [\alpha, g] \times [0, 1/2)$ and, considering that $(t_{n+1}, v_{n+1}) = ((1-2t_n)/t_n, 1/(2+v_n))$, we have $\mathscr{T}_g(R_\alpha) = [-g^2, (1-2\alpha)/\alpha] \times (2/5, 1/2]$, knowing that $(1-2g)/g = -g^2$; see Figure 3.4, where we have taken $\alpha = 0.52$ for illustrative purposes. Having determined what to remove (in the sense of the previous section), we use similar calculations as before to find what to add. Using the same notation as in the previous case, we find

$$\begin{aligned} (t_n^*, v_n^*) &= (t_n - 1, \tfrac{v_n}{1+v_n}) \quad \text{(as in the previous case)},\\ (t_{n+1}^*, v_{n+1}^*) &= (-\tfrac{t_{n+1}}{1+t_{n+1}}, 1 - v_{n+1}) = (\tfrac{1-2t_n}{t_n-1}, \tfrac{v_n+1}{v_n+2}) \text{ and}\\ (t_{n+2}^*, v_{n+2}^*) &= (t_{n+2}, v_{n+2}). \end{aligned}$$

So there are two regions to be added: one is $A_\alpha = [\alpha - 1, -g^2] \times [0, 1/3)$, the other is $\mathscr{T}_g(A_\alpha) = [(1-2\alpha)/(\alpha-1), (1-2g)/(g-1)] \times [1/2, 3/5)$. With regard to the last region, we note that $(1-2g)/(g-1) = g$. Because of the latter, we have not yet established to construct Ω_α, since $\mathscr{T}_g(A_\alpha)$ contains points (t, v) that do not exist in Ω_α, i.e. those for which $t \in (\alpha, g]$; see the greyed region in Figure 3.5. We will call the newly constructed region $\Omega_{\alpha,1}^+$, defined by

$$\Omega_{\alpha,1}^+ := \Omega_g \cup A_\alpha \cup \mathscr{T}_g(A_\alpha) \setminus (R_\alpha \cup \mathscr{T}_g(R_\alpha)).$$

Putting $R_{\alpha,1} := \mathscr{T}_g(A_\alpha) \cap [\alpha, g] \times [0, 3/5)$ as the part yet to be removed, we define

$$\Omega_{\alpha,1} := \Omega_{\alpha,1}^+ \setminus R_{\alpha,1}.$$

We will follow the same procedure of removal and addition regarding $R_{\alpha,1}$ as in the case of R_α. Note that this renders the same t-coordinates as in the creation of $\Omega_{\alpha,1}$. To efficiently calculate the v-coordinates, it is convenient to use Möbius transformations. We saw that in the current construction of Ω_α the equation $v_n^* = v_n/(v_n+1)$ holds, that we now write as $v_n^* = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}(v_n)$; similarly we write $v_{n+1}^* = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}(v_n)$. It follows that the v-coordinate of points in $A_{\alpha,1}$ is given by

$$v_{n,1}^* = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}(v_n), \tag{3.8}$$

with v_n the v-coordinate of points in R_α; for the v-coordinate of points in $\mathscr{T}_g(A_{\alpha,1})$ we have

$$v_{n+1,1}^* = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}(v_n). \tag{3.9}$$

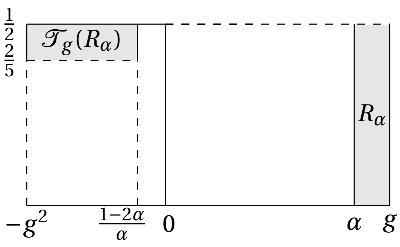


Figure 3.4: Ω_g with R_α and $\mathcal{T}_g(R_\alpha)$, $\alpha = 0.52$

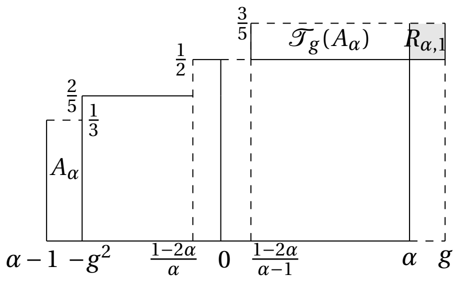


Figure 3.5: $\Omega^+_{\alpha,1}$, $\alpha = 0.52$

For the construction of $\Omega_{\alpha,2}$ we also need to calculate $\mathcal{T}_g(R_{\alpha,1})$. Since $v_{n+1} = \frac{1}{v_n+2} = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}(v_n)$, we find that the v-coordinates of $\mathcal{T}_g(R_{\alpha,1})$ are given by

$$v_{n+1,1} = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}(v_n), \tag{3.10}$$

v_n still being the v-coordinate of points in R_α. Since $v \in [0, 1/2)$ for points in R_α, calculating the boundaries of $\Omega_{\alpha,2}$ is now straightforward. We have written the relevant values of t and v in Figure 3.6 and write

$$\Omega^+_{\alpha,2} := \Omega^+_{\alpha,1} \cup A_{\alpha,1} \cup \mathcal{T}_g(A_{\alpha,1}) \setminus (R_{\alpha,1} \cup \mathcal{T}_g(R_{\alpha,1}))$$

Defining $R_{\alpha,2} := \mathcal{T}_g(A_{\alpha,1}) \cap [\alpha, g] \times [3/5, 8/13)$ as the part yet to be removed, we define

$$\Omega_{\alpha,2} := \Omega^+_{\alpha,2} \setminus R_{\alpha,2}.$$

Proceeding this way, we construct the sequence $\Omega_{\alpha,k}$, $k = 1, 2, \ldots$, and bring forth Ω_α as being $\lim_{k\to\infty} \Omega_{\alpha,k}$. To do so, we need to calculate $\lim_{k\to\infty} v^*_{n,k}$, $\lim_{k\to\infty} v^*_{n+1,k}$ and $\lim_{k\to\infty} v_{n+1,k}$, using formulas associated with (3.8) through (3.10):

$$\begin{cases} v^*_{n,k} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}^k (v_n), \; k \geq 1, \\ v^*_{n+1,k} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}^k (v_n), \; k \geq 1, \quad \text{and} \\ v_{n+1,k} = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}^k (v_n), \; k \geq 1. \end{cases}$$

We note that $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ is the square of $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$, the matrix that maps a pair of consecutive Fibonacci numbers (F_{i-1}, F_i) to (F_i, F_{i+1}); we put $F_0 = 0$ and $F_1 = 1$. It follows that

$$\lim_{k\to\infty} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}^k (v_n) = \lim_{k\to\infty} \frac{F_{2k-1}v_n + F_{2k}}{F_{2k}v_n + F_{2k+1}}.$$

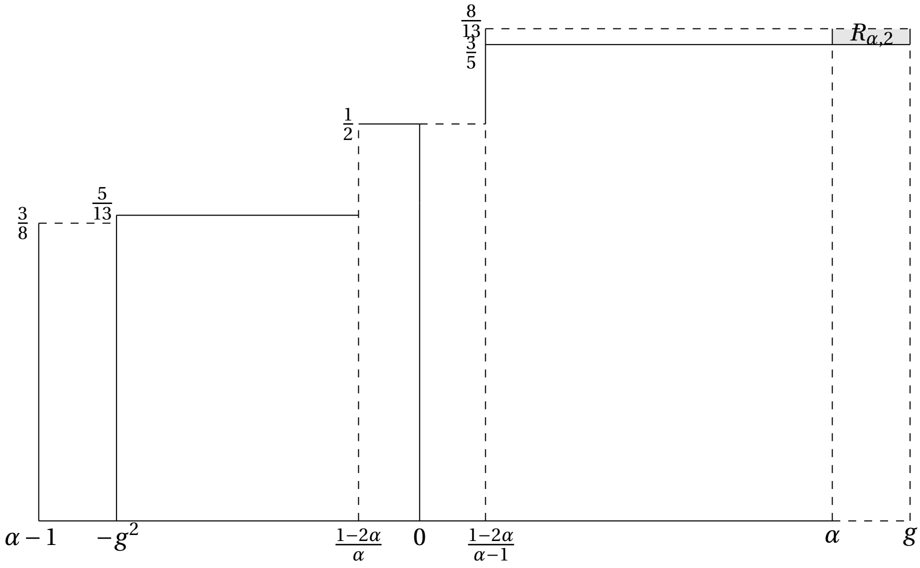


Figure 3.6: $\Omega_{\alpha,2}$, $\alpha = 0.52$

Now suppose $\lim_{k\to\infty} \frac{F_{2k-1}v_n+F_{2k}}{F_{2k}v_n+F_{2k+1}} = L$. Then

$$L = \lim_{k\to\infty} \frac{1}{\frac{F_{2k}v_n+F_{2k+1}}{F_{2k-1}v_n+F_{2k}}} = \lim_{k\to\infty} \frac{1}{1+\frac{F_{2k-2}v_n+F_{2k-1}}{F_{2k-1}v_n+F_{2k}}} = \frac{1}{1+L},$$

from which we derive

$$\lim_{k\to\infty} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}^k (v_n) = g,$$

whence

$$\begin{cases} \lim_{k\to\infty} v^*_{n,k} = \frac{g}{g+1} = g^2, \\ \lim_{k\to\infty} v^*_{n+1,k} = \frac{g+1}{g+2} = g \quad \text{and} \\ \lim_{k\to\infty} v_{n+1,k} = \frac{1}{g+2} = g^2. \end{cases}$$

We conclude that

$$\Omega_\alpha = \Omega_g \setminus \left([\alpha, g] \times [0, \tfrac{1}{2}) \bigcup_{k=1}^{\infty} [-g^2, \tfrac{1-2\alpha}{\alpha}] \times \left(\frac{F_{2k+1}}{F_{2k+3}}, \frac{F_{2k-1}}{F_{2k+1}} \right) \right)$$
$$\bigcup_{k=1}^{\infty} \left([\alpha-1, -g^2] \times \left[\frac{F_{2(k-1)}}{F_{2k}}, \frac{F_{2k}}{F_{2(k+1)}} \right) \cup [\tfrac{1-2\alpha}{\alpha-1}, \alpha) \times \left[\frac{F_{2k}}{F_{2k+1}}, \frac{F_{2k+2}}{F_{2k+3}} \right) \right),$$

from which we derive the following results, again obtained by Nakada in [9] (cf. [6]), illustrated by Figures 3.7 and 3.8.

Theorem 8. *Let* $\alpha \in (\frac{1}{2}, g]$. *Then*

$$\begin{aligned}\Omega_\alpha = &[\alpha-1, -g^2) \times [0, g^2) \cup [-g^2, \tfrac{1-2\alpha}{\alpha}] \times [0, g^2] \\ &\cup (\tfrac{1-2\alpha}{\alpha}, 0) \times [0, \tfrac{1}{2}] \cup [0, \tfrac{1-2\alpha}{\alpha-1}] \times [0, \tfrac{1}{2}) \cup (\tfrac{1-2\alpha}{\alpha-1}, \alpha) \times [0, g).\end{aligned}$$

Moreover, we have

$$\Omega_{1/2} = [-\tfrac{1}{2}, 0] \times [0, g^2] \cup (0, \tfrac{1}{2}) \times [0, g).$$

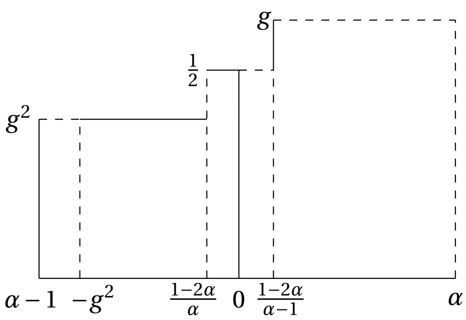


Figure 3.7: Ω_α, $\alpha = 0.52$

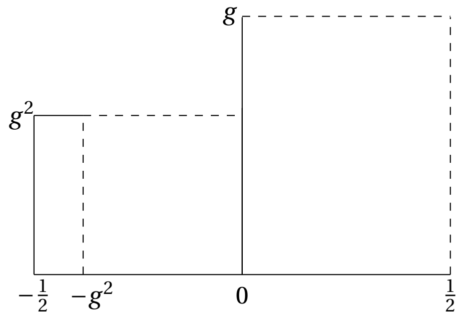


Figure 3.8: $\Omega_{\frac{1}{2}}$

In order to find the effect of the current transformation on the sequence of convergents, we recall that we rewrote

$$\begin{aligned}x &= [0; \varepsilon_1 a_1, \ldots, \varepsilon_n a_n, 2, -a_{n+2}, \ldots] \quad \text{as} \\ x &= [0; \varepsilon_1 a_1, \ldots, \varepsilon_n (a_n+1), -2, (a_{n+2}-1), \ldots].\end{aligned}$$

Since $p_k^* = p_k$ and $q_k^* = q_k$, $k < n$, we find, applying (1.2) and omitting the straightforward calculations,

$$\begin{cases} p_n^* = p_n + p_{n-1}; \\ p_{n+1}^* = p_{n+1}; \\ p_{n+2}^* = p_{n+2}. \end{cases}$$

Similarly, we find $q_n^* = q_n + q_{n-1}$, $q_{n+1}^* = q_{n+1}$ and $q_{n+2}^* = q_{n+2}$. It would seem that in the transformation from Ω_g to Ω_α, $\alpha \in [1/2, g)$, the convergent p_n/q_n is replaced by the so-called *mediant* $(p_n + p_{n-1})/(q_n + q_{n-1})$. We emphasise, however, that this is only a mediant with respect to a small interval with upper bound g; actually, this '*α-mediant*' is a convergent of the regular continued fraction. This comes as no surprise: from [6] we know that regular convergents only disappear and are not replaced by mediants when moving from the regular expansion to the α-expansion for $1/2 \le \alpha \le g$. The regular convergent that replaces another one in the transformation from Ω_g to Ω_α, $\alpha \in [1/2, g)$ had previously been 'singularised away' in the transformation from Ω_1 to Ω_g. We will illustrate this with an example.

Let $x = (\sqrt{17}-3)/4 = 1/(3+(\sqrt{17}-3)/2) = 0.28077\ldots$; in the second, more intricate expression for x, the number $(\sqrt{17}-3)/2$ is explicitly displayed for reasons that will show instantly. We distinguish

$$\begin{cases} x_\alpha = [0;\overline{3,1,1}], & \alpha \in (g,1]; \\ x_\alpha = [0;3,\overline{2,-4}], & \alpha \in ((\sqrt{17}-3)/2, g]; \\ x_\alpha = [0;\overline{4,-2}], & \alpha \in (1/2, (\sqrt{17}-3)/2]. \end{cases}$$

If we focus at the first partial quotients, we have

$$\begin{cases} \text{if } x_\alpha = [0;3,1,1,3,1,1\ldots], & \text{then } (\frac{p_0}{q_0},\ldots,\frac{p_6}{q_6}) = (\frac{0}{1},\frac{1}{3},\frac{1}{4},\frac{2}{7},\frac{7}{25},\frac{9}{32},\frac{16}{57}); \\ \text{if } x_\alpha = [0;3,2,-4,2,\ldots], & \text{then } (\frac{p_0}{q_0},\ldots,\frac{p_4}{q_4}) = (\frac{0}{1},\frac{1}{3},\frac{2}{7},\frac{7}{25},\frac{16}{57}); \\ \text{if } x_\alpha = [0;4,-2,4,-2,\ldots], & \text{then } (\frac{p_0}{q_0},\ldots,\frac{p_4}{q_4}) = (\frac{0}{1},\frac{1}{4},\frac{2}{7},\frac{9}{32},\frac{16}{57}). \end{cases}$$

What happens is that we create $[0;3,2,-4,2,\ldots]$ from $[0;3,1,1,3,1,1\ldots]$ by singularising the second 1 from each pair $(1,1)$, in which process some of the (regular) convergents get lost, in accordance with what we saw in Section 3.3. When creating $[0;4,-2,4,-2,\ldots]$ from $[0;3,2,-4,2,\ldots]$, we first insert a 1 between each 2 followed by -4, so as to retrieve $[0;3,1,1,3,1,1]$ (and its convergents), and subsequently singularise the first 1 from each pair $(1,1)$, so as to lose other convergents of the regular case. Generally, we conclude that in the transformation from Ω_g to Ω_α, $\alpha \in [1/2, g)$, some regular convergents p_n/q_n are replaced by others.

Although direct singularisation, as used in for instance [6], would be more straightforward, it is limited to $\alpha \ge 1/2$, while our approach enables a decrease beyond $\alpha = 1/2$, as we will show in the next sections. Moreover, it shows how to transform a domain Ω_α into another domain $\Omega_{\alpha'}$ in a uniform manner, where α' is smaller than but close to α, for any $\alpha \in (g^2, 1]$. On top of that, in Section 3.8 we will go into the isomorphism between Ω_α and Ω_g for $g^2 \le \alpha < g$.

3.5. THE CASE $\alpha \in (\sqrt{2}-1, \frac{1}{2}]$

In this section we will show how to derive Ω_α from $\Omega_{1/2}$, with $\sqrt{2}-1 \le \alpha < 1/2$. Let $x \in [-1/2, 1/2)$ and let $n \ge 0$ be the smallest integer for which $t_n \in [\alpha, 1/2]$, i.e., for which $(t_n, v_n) \in R_\alpha := [\alpha, 1/2] \times [0, g)$. We already know that $x = [0;\varepsilon_1 a_1, \ldots, \varepsilon_n a_n, 2, a_{n+2}, \ldots]$ in this case. Since the continued fraction map of the related dynamical system is $T_{1/2}$, for

$\alpha \in (\sqrt{2}-1, 1/2]$ we have

$$a_2(\alpha) = \left\lfloor \frac{1}{T_{1/2}(\alpha)} + 1 - \frac{1}{2} \right\rfloor = \left\lfloor \frac{1}{\frac{1}{\alpha}-2} + \frac{1}{2} \right\rfloor \geq \left\lfloor \frac{1}{\frac{1}{\sqrt{2}-1}-2} + \frac{1}{2} \right\rfloor = \left\lfloor \sqrt{2} + \frac{3}{2} \right\rfloor = 2,$$

so in this case

$$x = [0; \varepsilon_1 a_1, \ldots, \varepsilon_n a_n, 2, a_{n+2}, \ldots],\ a_{n+2} \geq 2.$$

The difference with the case in the previous section is merely the sign of a_{n+2}, which shows in the great resemblance of equations that we need to construct Ω_α. Applying insertion and singularisation in a similar way, we find

$$x = [0; \varepsilon_1 a_1, \ldots, \varepsilon_n(a_n+1), -2, -(a_{n+2}+1), \ldots]$$

this time. The other equations are exactly the same as in the previous section:

$$\begin{aligned}
(t_{n+1}, v_{n+1}) &= \left(\frac{1-2t_n}{t_n}, \frac{1}{2+v_n}\right) \\
(t_n^*, v_n^*) &= \left(t_n - 1, \frac{v_n}{1+v_n}\right) \\
(t_{n+1}^*, v_{n+1}^*) &= \left(\frac{1-2t_n}{t_n-1}, \frac{v_n+1}{v_n+2}\right) \\
(t_{n+2}^*, v_{n+2}^*) &= (t_{n+2}, v_{n+2}).
\end{aligned}$$

Using similar calculations as in the previous section, we find that $R_\alpha = [\alpha, 1/2) \times [0, g)$ implies $\mathcal{T}_{1/2}(R_\alpha) = (0, (1-2\alpha)/\alpha) \times (g^2, 1/2]$; see Figure 3.9, where we have taken $\alpha = 0.43$ for illustrative purposes. The first region to be added is: $A_\alpha = [\alpha-1, -\frac{1}{2}] \times [0, g^2)$; the second is $\mathcal{T}_{1/2}(A_\alpha) = [(1-2\alpha)/(\alpha-1), 0] \times [1/2, g)$. An important difference with the previous case is that $(R_\alpha \cup \mathcal{T}_{1/2}(R_\alpha)) \cap (A_\alpha \cup \mathcal{T}_{1/2}(A_\alpha)) = \emptyset$, so all we have to do is remove R_α and $\mathcal{T}_{1/2}(R_\alpha)$ from $\Omega_{1/2}$ and add A_α and $\mathcal{T}_{1/2}(A_\alpha)$ to it. Doing so, we find

Theorem 9. *Let $\alpha \in (\sqrt{2}-1, 1/2]$. Then*

$$\Omega_\alpha = [\alpha-1, \alpha) \times [0, g^2) \cup [\tfrac{1-2\alpha}{\alpha}, \alpha) \times [g^2, \tfrac{1}{2}) \cup [\tfrac{1-2\alpha}{\alpha-1}, \alpha) \times [\tfrac{1}{2}, g).$$

Since $(1-2\alpha)/\alpha$ equals α for $\alpha = \sqrt{2}-1$, at $\alpha = \sqrt{2}-2$ the domain Ω_α is split in two parts:

$$\Omega_{\sqrt{2}-1} = [\sqrt{2}-2, \sqrt{2}-1) \times [0, g^2) \cup [\tfrac{1}{2}\sqrt{2}-1, \sqrt{2}-1) \times [\tfrac{1}{2}, g);$$

see Figure 3.10.

With a great similarity of calculations as in the previous section, we find that in this case also

$$\begin{cases} p_n^* = p_n + p_{n-1}; \\ p_{n+1}^* = p_{n+1}; \\ p_{n+2}^* = p_{n+2}. \end{cases}$$

The difference with the case $\alpha \in (1/2, g]$ is that now the convergent p_n/q_n is really replaced by the mediant $(p_n + p_{n-1})/(q_n + q_{n-1})$ in the sense of the regular case.

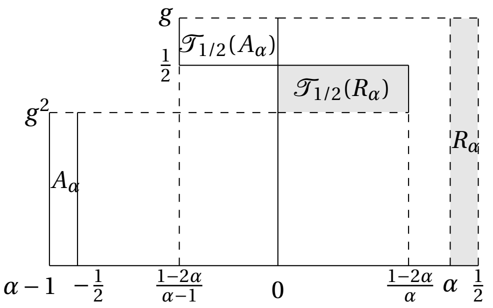


Figure 3.9: $\Omega_{\frac{1}{2}}, \alpha = 0.43$

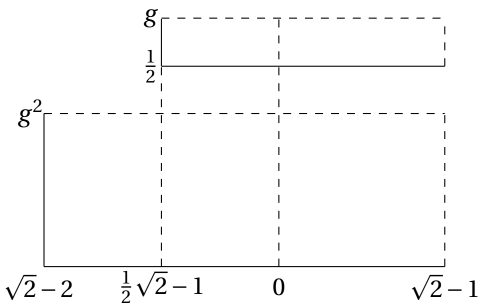


Figure 3.10: $\Omega_{\sqrt{2}-1}$

3.6. THE CASE $\alpha \in (\frac{\sqrt{10}-2}{3}, \sqrt{2}-1]$

In Section 3.4 we noticed that $\sqrt{2}-1_{\sqrt{2}-1} = [0;3,\overline{-2,-4}]$; actually, $\sqrt{2}-1$ is the largest $\alpha \in [0,1]$ such that no partial quotients 1 or 2 (with plus sign) occur in α_α. In this section we want to remove points (t_n, v_n) with $t_n = [0;3,\varepsilon_{n+2}a_{n+2},\dots]$, that is, points associated with numbers $x = [0;\varepsilon_1 a_1,\dots,\varepsilon_n a_n, 3, \varepsilon_{n+2}a_{n+2},\dots]$. Similar to the cases in Sections 3.4 and 3.5, the removal of a 3 requires compensated insertion, for which we need a partial quotient 2, as explained in Section 3.1. Since we removed all partial quotients 2 with a plus sign, we are dealing with the following numbers:

$$x = [0;\varepsilon_1 a_1,\dots,\varepsilon_n a_n, 3, -2, \varepsilon_{n+3}a_{n+3},\dots].$$

Note that uncompensated insertion would leave us with partial quotients 1 or +2, involving points outside Ω_α. Because of the necessity of compensated insertion, the next interval to explore is not simply $\Delta_\alpha(3) = ((\sqrt{13}-3)/2, \sqrt{2}-1]$ but the much smaller $\Delta_\alpha(3,-2)$. To find the boundaries of this interval, we determine $(\alpha',\alpha]$, $\alpha' < \alpha \le \sqrt{2}-1$, such that $t_n \in (\alpha',\alpha]$ implies $t_n = [0;3,-2,\dots]$. If $\alpha' \in [0,\sqrt{2}-1)$ and $\alpha \in (\alpha',\sqrt{2}-1]$, then

$$\begin{aligned}
2 &= \left\lfloor \frac{1}{3-\frac{1}{\sqrt{2}-1}} + 2 - \sqrt{2} \right\rfloor \le \left\lfloor \frac{1}{3-\frac{1}{\alpha}} + 1 - \alpha \right\rfloor \le a_2(\alpha'_\alpha) \\
&= \left\lfloor \frac{1}{3-\frac{1}{\alpha'}} + 1 - \alpha \right\rfloor \le \left\lfloor \frac{1}{3-\frac{1}{\alpha'}} + 1 - \alpha' \right\rfloor,
\end{aligned}$$

from which we derive $a_2(\alpha'_\alpha) = 2$ if and only if $1/(3-1/\alpha') + 1 - \alpha' < 3$. Given that $\alpha' \in [0,\sqrt{2}-1))$, this is the case if and only if $(\sqrt{10}-2)/3 < \alpha' < \sqrt{2}-1$; indeed,

$$(\sqrt{10}-2)/3_{(\sqrt{10}-2)/3} = [0;3,\overline{-3,-2,-3,-4}].$$

So $\Delta_\alpha(3,-2) = ((\sqrt{10}-2)/3, \sqrt{2}-1]$. Since

$$\sqrt{2}-1 = [0;3,\overline{-2,-4}] \text{ and } \lim_{a\to\infty}[0;3,-2,-a] = \frac{2}{5},$$

we will first show how to construct Ω_α from $\Omega_{\sqrt{2}-1}$, with $2/5 \le \alpha < \sqrt{2}-1$, which will prove to be more complicated than the work we have done so far. After that, the step from $\Omega_{2/5}$ to $\Omega_{(\sqrt{10}-2)/3}$ will prove to be fairly small.

Let $\alpha \in (2/5, \sqrt{2}-1]$, $x \in [\sqrt{2}-2, \sqrt{2}-1)$ and $n \ge 0$ be the smallest integer for which $t_n \in [\alpha, \sqrt{2}-1)$, i.e., for which

$$(t_n, v_n) \in R_\alpha := [\alpha, \sqrt{2}-1] \times [0, g^2) \cup [\alpha, \sqrt{2}-1] \times [\tfrac{1}{2}, g). \tag{3.11}$$

Considering the fact that $\sqrt{2}-1_{\sqrt{2}-1} = [0; 3, \overline{-2, -4}]$, it is not hard to see that in this case

$$x = [0; \varepsilon_1 a_1, \dots, \varepsilon_n a_n, 3, -2, -a_{n+3}, \dots], \ a_{n+3} \ge 4.$$

First we insert 1 between 3 and -2, yielding

$$x = [0; \varepsilon_1 a_1, \dots, \varepsilon_n a_n, 2, 1, 1, -a_{n+3}, \dots], \ a_{n+3} \ge 4.$$

Next we insert -1 between a_n and 2, yielding

$$x = [0; \varepsilon_1 a_1, \dots, \varepsilon_n (a_n+1), -1, 1, 1, \mathbf{1}, -a_{n+3}, \dots], \ a_{n+3} \ge 4.$$

Then we singularise the last $+1$, yielding

$$x = [0; \varepsilon_1 a_1, \dots, \varepsilon_n (a_n+1), -1, \mathbf{1}, 2, -2, a_{n+3}-1, \dots], \ a_{n+3} \ge 4.$$

Finally we singularise the remaining $+1$, so as to get

$$x = [0; \varepsilon_1 a_1, \dots, \varepsilon_n (a_n+1), -2, -3, a_{n+3}-1, \dots], \ a_{n+3} \ge 4.$$

For the first time we are confronted with the singularised expansion differing at four places from the original expansion, necessitating one more stage of calculating successive (t, v)-pairs (in the sense of the previous sections). Omitting the straightforward calculations, we find

$$\begin{aligned} (t_{n+1}, v_{n+1}) &= (\tfrac{1-3t_n}{t_n}, \tfrac{1}{3+v_n}), & (t_{n+2}, v_{n+2}) &= (\tfrac{2-5t_n}{3t_n-1}, \tfrac{v_n+3}{2v_n+5}), \\ (t_n^*, v_n^*) &= (t_n - 1, \tfrac{v_n}{1+v_n}), & (t_{n+1}^*, v_{n+1}^*) &= (\tfrac{1-2t_n}{t_n-1}, \tfrac{v_n+1}{v_n+2}), \\ (t_{n+2}^*, v_{n+2}^*) &= (\tfrac{2-5t_n}{2t_n-1}, \tfrac{v_n+2}{2v_n+5}), & (t_{n+3}^*, v_{n+3}^*) &= (t_{n+3}, v_{n+3}). \end{aligned}$$

We see that in the current situation (i.e. for the current values of α), constructing Ω_α is much more complicated than in the previous sections. There is one more step of removing and adding regions, and these regions are all split in two disjoint parts. In particular, $R_{\alpha,1} := \mathcal{T}^2_{\sqrt{2}-1}(A_\alpha) \cap R_\alpha \ne \emptyset$, similar to Section 3.4. Since (3.11) holds, we can draw $\Omega_{\alpha,1}$, defined similarly as in Section 3.4. We will first sketch roughly what is removed from $\Omega_{\sqrt{2}-1}$ (see Figure 3.11; the removed parts are in grey), consisting of

$$\begin{aligned} R_\alpha &= [\alpha, \sqrt{2}-1] \times [0, g^2) \cup [\alpha, \sqrt{2}-1] \times [\tfrac{1}{2}, g), \\ \mathcal{T}_{\sqrt{2}-1}(R_\alpha) &= [\sqrt{2}-2, \tfrac{1-3\alpha}{\alpha}] \times (\tfrac{1}{3+g^2}, \tfrac{1}{3}] \cup [\sqrt{2}-2, \tfrac{1-3\alpha}{\alpha}] \times (\tfrac{1}{3+g}, \tfrac{2}{7}] \quad \text{and} \\ \mathcal{T}^2_{\sqrt{2}-1}(R_\alpha) &= [\tfrac{1}{2}\sqrt{2}-1, \tfrac{2-5\alpha}{3\alpha-1}] \times (\tfrac{3+g^2}{5+2g^2}, \tfrac{3}{5}] \cup [\tfrac{1}{2}\sqrt{2}-1, \tfrac{1-3\alpha}{\alpha}] \times (\tfrac{3+g}{5+2g}, \tfrac{7}{12}]. \end{aligned}$$

The removal of R_α does not show in the resulting figure (it is 'cut off' from $\Omega_{\sqrt{2}-1}$), but the four parts of which $\mathcal{T}_{\sqrt{2}-1}(R_\alpha)$ and $\mathcal{T}^2_{\sqrt{2}-1}(R_\alpha)$ consist, show as coves. In Figure 3.12 we see Figure 3.11 with the added regions; $R_{\alpha,1}$ is in grey:

$$A_\alpha = [\alpha-1, \sqrt{2}-2] \times [0, \tfrac{g^2}{1+g^2}) \cup [\alpha-1, \sqrt{2}-2] \times [\tfrac{1}{3}, g^2),$$

$$\mathcal{T}_{\sqrt{2}-1}(A_\alpha) = [\tfrac{1-2\alpha}{\alpha-1}, \tfrac{1}{2}\sqrt{2}-1] \times [\tfrac{1}{2}, \tfrac{1+g^2}{2+g^2}) \cup [\tfrac{1-2\alpha}{\alpha-1}, \tfrac{1}{2}\sqrt{2}-1] \times [\tfrac{3}{5}, g) \text{ and}$$

$$\mathcal{T}^2_{\sqrt{2}-1}(A_\alpha) = [\tfrac{2-5\alpha}{2\alpha-1}, \sqrt{2}-1] \times [\tfrac{2}{5}, \tfrac{2+g^2}{5+2g^2}) \cup [\tfrac{2-5\alpha}{2\alpha-1}, \sqrt{2}-1] \times [\tfrac{5}{12}, \tfrac{2+g}{5+2g}).$$

Similar to the construction in Section 3.4, we are now left with $R_{\alpha,1}$ to remove, consisting of the protuberant parts of two small rectangles between the parts that we already removed. Figure 3.13 shows a detailed example of $\Omega_{\alpha,1}$, $2/5 < \alpha < \sqrt{2}-1$. This first stage shows how Ω_α is not only characterised by more and more protuberant parts and coves as α decreases, but is also splitting in more and more disjoint regions. Nevertheless, the construction of Ω_α in the current case is very similar to the construction of Ω_α in Section 3.4: likewise, we construct the sequence $\Omega_{\alpha,k}$, $k = 1,2,\ldots$, and bring forth Ω_α as being $\lim_{k\to\infty} \Omega_{\alpha,k}$.

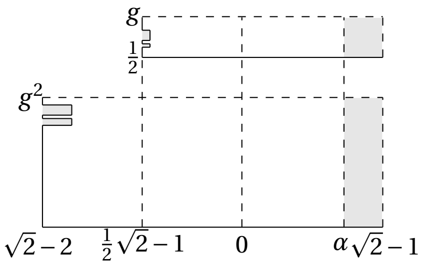


Figure 3.11: $\Omega_{\alpha,1}$, removed parts, rough sketch

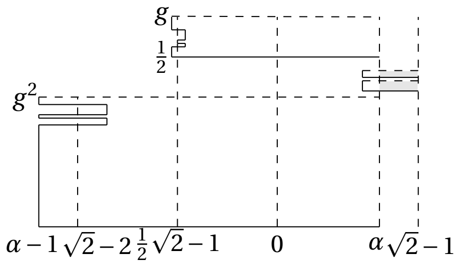


Figure 3.12: $\Omega_{\alpha,1}$, removed and added parts, rough sketch

We remark that fractions involving g or G can be represented in many ways. In this particular case, substituting g for v_n in $1/(3+v_n)$ (which is associated with the v-bounds of $\mathcal{T}_{\sqrt{2}-1}(R_\alpha)$) yields $1/(3+g)$, which is equal to $g^2/(1+g^2)$; this is the fraction we find when we substitute g^2 for v_n in $v_n/(1+v_n)$, which is associated with the v-bounds of A_α; see the lower corner on the left side of Figure 3.13.

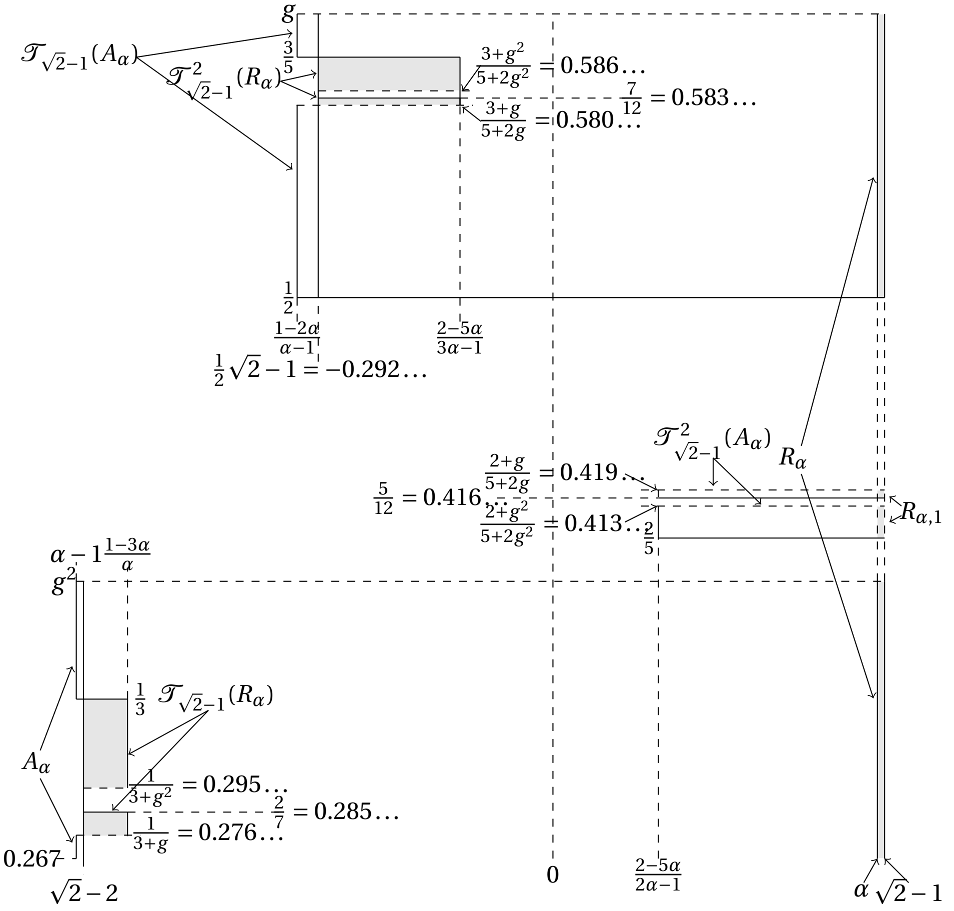


Figure 3.13: $\Omega_{\alpha,1}$, $\alpha = 0.405$. The lower rectangle $[\alpha-1, \sqrt{2}-1] \times [0, .267)$ has been omitted.

To obtain Ω_α as a 'limit' of the sequence $\Omega_{\alpha,k}$, we need to infinitely remove and add regions. Similar to the case $1/2 < \alpha \leq g$, only new v-values need to be calculated, to which end we use the Möbius transformations associated with v_n^*, v_{n+1}^*, v_{n+2}^*, v_{n+1} and v_{n+2}. The matrices involved are immediately derived from the way v_n^*, v_{n+1}^* et cetera are expressed in v_n. In the recurrent process of removing and adding regions, the matrix $\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}$ - the one belonging to v_{n+2}^* - plays a central role, because it determines the domain of v-values in each new stage of constructing $\Omega_{\alpha,k}$, $k \geq 1$, similar to $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ in

Section 3.4. Using similar notations for new v-values also, we find

$$\begin{cases} v^*_{n,k} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}^k (v_n),\ k \geq 1, \\ v^*_{n+1,k} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}^k (v_n),\ k \geq 1, \\ v^*_{n+2,k} = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}^k (v_n),\ k \geq 1 \end{cases}$$

and

$$\begin{cases} v_{n+1,k} = \begin{pmatrix} 0 & 1 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}^k (v_n),\ k \geq 1, \quad \text{and} \\ v_{n+2,k} = \begin{pmatrix} 1 & 3 \\ 2 & 5 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}^k (v_n),\ k \geq 1. \end{cases}$$

It be remarked that graphic representations of $\Omega_{\alpha,k}, k \geq 2$ would show merely very small changes compared to $\Omega_{\alpha,1}$. The six regions that are removed as well as the six regions that are added in each new stage are very small. We have, for example:

$$A_{\alpha,1} = [\alpha-1, \sqrt{2}-2] \times [\tfrac{2}{7}, \tfrac{2+g^2}{7+3g^2}) \cup [\alpha-1, \sqrt{2}-2] \times [\tfrac{5}{17}, \tfrac{2+g}{7+3g}),$$

while $\frac{2}{7} = 0.285714\ldots$; $\frac{2+g^2}{7+3g^2} = 0.292412\ldots$; $\frac{5}{17} = 0.294117\ldots$ and $\frac{2+g}{7+3g} = 0.295686\ldots$;

$$\mathscr{T}_{\sqrt{2}-1}(A_{\alpha,1}) = [\tfrac{1-2\alpha}{\alpha-1}, \tfrac{1}{2}\sqrt{2}-1] \times [\tfrac{7}{12}, \tfrac{7+3g^2}{12+5g^2}) \cup [\tfrac{1-2\alpha}{\alpha-1}, \tfrac{1}{2}\sqrt{2}-1] \times [\tfrac{17}{29}, \tfrac{7+3g}{12+5g}),$$

while $\frac{7}{12} = 0.583333\ldots$; $\frac{7+3g^2}{12+5g^2} = 0.585621\ldots$; $\frac{17}{29} = 0.586206\ldots$ and $\frac{7+3g}{12+5g} = 0.586746\ldots$;

$$\mathscr{T}^2_{\sqrt{2}-1}(A_{\alpha,1}) = [\tfrac{2-5\alpha}{2\alpha-1}, \sqrt{2}-1] \times [\tfrac{12}{29}, \tfrac{12+5g^2}{29+12g^2}) \cup [\tfrac{2-5\alpha}{2\alpha-1}, \sqrt{2}-1] \times [\tfrac{29}{70}, \tfrac{12+5g}{29+12g}),$$

while $\frac{12}{29} = 0.413793\ldots$; $\frac{12+5g^2}{29+12g^2} = 0.414185\ldots$; $\frac{29}{70} = 0.414285\ldots$ and $\frac{12+5g}{29+12g} = 0.414378\ldots$;

$$\mathscr{T}_{\sqrt{2}-1}(R_{\alpha,1}) = [\sqrt{2}-2, \tfrac{1-3\alpha}{\alpha}] \times (\tfrac{5+2g^2}{17+7g^2}, \tfrac{5}{17}] \cup [\sqrt{2}-2, \tfrac{1-3\alpha}{\alpha}] \times (\tfrac{5+2g}{17+7g}, \tfrac{12}{41}],$$

while $\frac{5}{17} = 0.294117\ldots$; $\frac{5+2g^2}{17+7g^2} = 0.292975\ldots$; $\frac{12}{41} = 0.292682\ldots$ and $\frac{5+2g}{17+7g} = 0.292412\ldots$ and

$$\mathscr{T}^2_{\sqrt{2}-1}(R_{\alpha,1}) = [\tfrac{1}{2}\sqrt{2}-1, \tfrac{2-5\alpha}{3\alpha-1}] \times (\tfrac{17+7g^2}{29+12g^2}, \tfrac{17}{29}] \cup [\tfrac{1}{2}\sqrt{2}-1, \tfrac{1-3\alpha}{\alpha}] \times (\tfrac{17+7g}{29+12g}, \tfrac{41}{70}],$$

while $\frac{17}{29} = 0.586206\ldots$; $\frac{17+7g^2}{29+12g^2} = 0.585814\ldots$; $\frac{41}{70} = 0.585714\ldots$ and $\frac{17+7g}{29+12g} = 0.585621\ldots$. In Figure 3.14 the leftmost corner of $\Omega_{\alpha,2}$ is shown. Similar to Figure 3.13, we have omitted the lower rectangle (where $v < 2/7$), in order not to waste too much blank space.

Even more than Figure 3.14, Figure 3.15 illustrates the rapid decrease of the surface of the rectangles, the union of which finally builds Ω_α. In particular,

$$R_{\alpha,2} = [\alpha, \sqrt{2}-1] \times [\tfrac{12}{29}, \tfrac{12+5g^2}{29+12g^2}) \cup [\alpha, \sqrt{2}-1] \times [\tfrac{29}{70}, \tfrac{12+5g}{29+12g}),$$

the union of two very small rectangles between the two rectangles of $R_{\alpha,1}$. In the next step, we would find $R_{\alpha,3}$ consisting of two very small rectangles between the two rectangles of $R_{\alpha,2}$ - and so on.

Since $\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}^2$, we see that at the base of our current construction lies a matrix that is almost equal to the one in Section 3.4 that renders consecutive Fibonacci numbers. In this case, we have a matrix that renders the recurrent sequence E_n given by $E_{n+1} = 2E_n + E_{n-1}, n \geq 0$, with $E_{-1} = 1$ and $E_0 = 0$. Now suppose $\lim_{k\to\infty} \frac{E_{2k-1}v_n + E_{2k}}{E_{2k}v_n + E_{2k+1}} = L$. Then

$$L = \lim_{k\to\infty} \frac{1}{\frac{E_{2k}v_n+E_{2k+1}}{E_{2k-1}v_n+E_{2k}}} = \lim_{k\to\infty} \frac{1}{2 + \frac{E_{2k-2}v_n+E_{2k-1}}{E_{2k-1}v_n+E_{2k}}} = \frac{1}{2+L},$$

from which we derive

$$\lim_{k\to\infty} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}^k (v_n) = \sqrt{2}-1, \quad \text{whence}$$

$$\begin{cases} \lim_{k\to\infty} v^*_{n,k} = \frac{\sqrt{2}-1}{\sqrt{2}} = 1 - \frac{1}{2}\sqrt{2} = 0.292893\ldots, \\ \lim_{k\to\infty} v^*_{n+1,k} = \frac{\sqrt{2}}{\sqrt{2}+1} = 2 - \sqrt{2} = 0.585786\ldots, \\ \lim_{k\to\infty} v^*_{n+2,k} = \frac{\sqrt{2}+1}{2\sqrt{2}+3} = \sqrt{2}-1 = 0.414213\ldots, \\ \lim_{k\to\infty} v_{n+1,k} = \frac{1}{\sqrt{2}+2} = 1 - \frac{1}{2}\sqrt{2} \quad \text{and} \\ \lim_{k\to\infty} v_{n+2,k} = \frac{\sqrt{2}+2}{2\sqrt{2}+3} = 2 - \sqrt{2}. \end{cases}$$

Omitting straightforward calculations, we conclude that

$$\Omega_\alpha = \Omega_{\sqrt{2}-1} \setminus \left(\bigcup_{k=0}^{\infty} R_{\alpha,k} \cup \mathcal{T}_{\sqrt{2}-1}(R_{\alpha,k}) \cup \mathcal{T}^2_{\sqrt{2}-1}(R_{\alpha,k}) \right)$$
$$\bigcup_{k=0}^{\infty} \left(A_{\alpha,k} \cup \mathcal{T}_{\sqrt{2}-1}(A_{\alpha,k}) \cup \mathcal{T}^2_{\sqrt{2}-1}(A_{\alpha,k})^* \right),$$

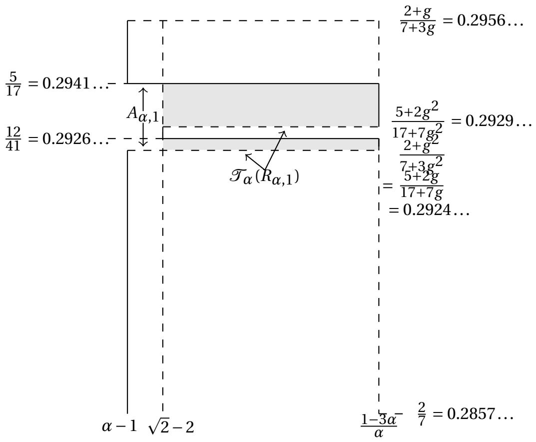


Figure 3.14: The leftmost corner of $\Omega_{\alpha,2}, \alpha = 0.405$.

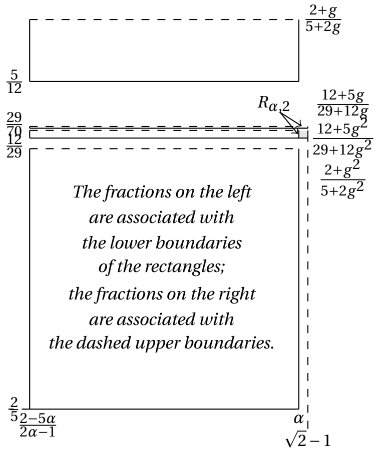


Figure 3.15: The rectangles between the upper and lower block of $\Omega_{\alpha,2}, \alpha = 0.405$.

with

$$
\begin{aligned}
R_{\alpha,k} &= [\alpha, \sqrt{2}-1] \times [\tfrac{E_{2k}}{E_{2k+1}}, \tfrac{E_{2k}+g^2E_{2k-1}}{E_{2k+1}+g^2E_{2k}}) \cup [\alpha, \sqrt{2}-1] \times [\tfrac{E_{2k+1}}{E_{2k+2}}, \tfrac{E_{2k}+gE_{2k-1}}{E_{2k+1}+gE_{2k}});\\
\mathcal{T}_{\sqrt{2}-1}(R_{\alpha,k}) &= [\sqrt{2}-2, \tfrac{1-3\alpha}{\alpha}] \times (\tfrac{E_{2k+1}+g^2E_{2k}}{E_{2k+2}+E_{2k+1}+g^2(E_{2k+1}+E_{2k})}, \tfrac{E_{2k+1}}{E_{2k+2}+E_{2k+1}}]\\
&\quad \cup [\sqrt{2}-2, \tfrac{1-3\alpha}{\alpha}] \times (\tfrac{E_{2k+1}+gE_{2k}}{E_{2k+2}+E_{2k+1}+g(E_{2k+1}+E_{2k})}, \tfrac{E_{2k+2}}{E_{2k+3}+E_{2k+2}}];\\
\mathcal{T}^2_{\sqrt{2}-1}(R_{\alpha,k}) &= [\tfrac{1}{2}\sqrt{2}-1, \tfrac{2-5\alpha}{3\alpha-1}] \times (\tfrac{E_{2k+2}+E_{2k+1}+g^2(E_{2k+1}+E_{2k})}{E_{2k+3}+g^2E_{2k+2}}, \tfrac{E_{2k+2}+E_{2k+1}}{E_{2k+3}}]\\
&\quad \cup [\tfrac{1}{2}\sqrt{2}-1, \tfrac{2-5\alpha}{3\alpha-1}] \times (\tfrac{E_{2k+2}+E_{2k+1}+g(E_{2k+1}+E_{2k})}{E_{2k+3}+gE_{2k+2}}, \tfrac{E_{2k+3}+E_{2k+2}}{E_{2k+4}}];\\
A_{\alpha,k} &= [\alpha-1, \sqrt{2}-2] \times [\tfrac{E_{2k}}{E_{2k+1}+E_{2k}}, \tfrac{E_{2k}+g^2E_{2k-1}}{E_{2k+1}+E_{2k}+g^2(E_{2k}+E_{2k-1})})\\
&\quad \cup [\alpha-1, \sqrt{2}-2] \times [\tfrac{E_{2k+1}}{E_{2k+2}+E_{2k+1}}, \tfrac{E_{2k}+gE_{2k-1}}{E_{2k+1}+E_{2k}+g(E_{2k}+E_{2k-1})});\\
\mathcal{T}_{\sqrt{2}-1}(A_{\alpha,k}) &= [\tfrac{1-2\alpha}{\alpha-1}, \tfrac{1}{2}\sqrt{2}-1] \times [\tfrac{E_{2k+1}+E_{2k}}{E_{2k+2}}, \tfrac{E_{2k+1}+E_{2k}+g^2(E_{2k}+E_{2k-1})}{E_{2k+2}+g^2E_{2k+1}})\\
&\quad \cup [\tfrac{1-2\alpha}{\alpha-1}, \tfrac{1}{2}\sqrt{2}-1] \times [\tfrac{E_{2k+2}+E_{2k+1}}{E_{2k+3}}, \tfrac{E_{2k+1}+E_{2k}+g(E_{2k}+E_{2k-1})}{E_{2k+2}+gE_{2k+1}});\\
\mathcal{T}^2_{\sqrt{2}-1}(R_{\alpha,k})^* &= [\tfrac{2-5\alpha}{2\alpha-1}, \alpha) \times [\tfrac{E_{2k+2}}{E_{2k+3}}, \tfrac{E_{2k+2}+g^2E_{2k+1}}{E_{2k+3}+g^2E_{2k+2}}) \cup [\tfrac{2-5\alpha}{2\alpha-1}, \alpha) \times [\tfrac{E_{2k+3}}{E_{2k+4}}, \tfrac{E_{2k+2}+gE_{2k+1}}{E_{2k+3}+gE_{2k+2}}),
\end{aligned}
$$

where $\mathcal{T}^2_{\sqrt{2}-1}(R_{\alpha,k})^* := \mathcal{T}^2_{\sqrt{2}-1}(R_{\alpha,k}) \setminus R_{\alpha,k+1}$, $R_{\alpha,0} := R_\alpha$ and $A_{\alpha,0} := A_\alpha$.
In order to find the effect of the current transformation on the sequence of convergents, we recall that we rewrote

$$
\begin{aligned}
x &= [0; \varepsilon_1 a_1, \ldots, \varepsilon_n a_n, 3, -2, -a_{n+3}, \ldots] \quad \text{as}\\
x &= [0; \varepsilon_1 a_1, \ldots, \varepsilon_n (a_n+1), -2, -3, a_{n+3}-1, \ldots].
\end{aligned}
$$

Since $p_k^* = p_k$ and $q_k^* = q_k$, $k < n$, we find, applying (1.2) and this time omitting the straightforward calculations,

$$\begin{cases} p_n^* = p_n + p_{n-1}; \\ p_{n+1}^* = p_{n+1} - p_n; \\ p_{n+2}^* = p_{n+2}; \\ p_{n+3}^* = p_{n+3}. \end{cases}$$

Of course, we find similar relations for q_n^* through q_{n+3}^*. We conclude that in the transformation from $\Omega_{\sqrt{2}-1}$ to Ω_α, $\alpha \in [2/5, \sqrt{2}-1)$, the convergent p_n/q_n is replaced by the mediant $(p_n + p_{n-1})/(q_n + q_{n-1})$ and that p_{n+1}/q_{n+1} is replaced by the mediant $(p_{n+1} - p_n)/(q_{n+1} - q_n)$.

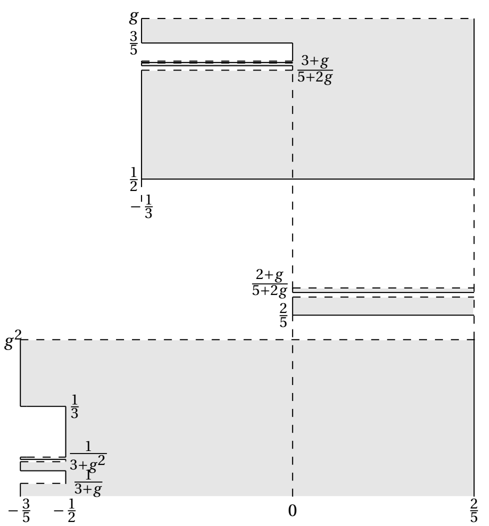


Figure 3.16: $\Omega_{\frac{2}{5}}$, 'rough sketch'.

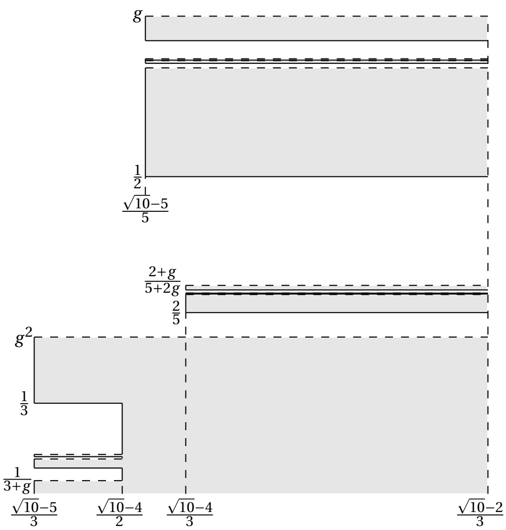


Figure 3.17: $\Omega_{\frac{\sqrt{10}-2}{3}}$, 'rough sketch'

We will now show how to construct Ω_α from $\Omega_{2/5}$, with $\sqrt{10}-2)/3 < \alpha < 2/5$. Let $\alpha \in ((\sqrt{10}-2)/3, 2/5)$, $x \in [-3/5, 2/5)$ and $n \geq 0$ be the smallest integer for which $t_n \in [\alpha, 2/5]$, i.e., for which

$$(t_n, v_n) \in \bigcup_{k=0}^{\infty} \left([\alpha, \tfrac{2}{5}] \times \left([\tfrac{E_{2k}}{E_{2k+1}}, \tfrac{E_{2k}+g^2 E_{2k-1}}{E_{2k+1}+g^2 E_{2k}}) \cup [\tfrac{E_{2k+1}}{E_{2k+2}}, \tfrac{E_{2k}+g E_{2k-1}}{E_{2k+1}+g E_{2k}}) \right) \right).$$

We already know that in this case

$$x = [0; \varepsilon_1 a_1, \dots, \varepsilon_n a_n, 3, -2, a_{n+3}, \dots],$$

that we rewrite as

$$x = [0; \varepsilon_1 a_1, \ldots, \varepsilon_n(a_n+1), -2, -3, -(a_{n+3}+1), \ldots];$$

it is not hard to find that $a_{n+3} \geq 3$. Remember that in Section 3.5 we saw how much the case $\alpha \in (\sqrt{2}-1, 1/2]$ resembled the case $\alpha \in (1/2, g]$. Here we find a similar resemblance between $\alpha \in ((\sqrt{10}-2)/3, 2/5)$ and $\alpha \in [2/5, \sqrt{2}-1]$. Of course, in the current case the continued fraction map is $T_{2/5}$ instead of $T_{\sqrt{2}-1}$, but straightforward calculations show that exactly the same equations for calculating the (t, v)-pairs for the regions R_α through $\mathscr{T}^2(A_\alpha)$ hold, which proves to be very convenient for the construction of Ω_α, $\alpha \in ((\sqrt{10}-2)/3, 2/5)$. It appears that the transformation consists solely of expanding the regions already added and removed, in other words: both the protuberant parts as the coves are extended. Of special interest are the coves in the upper block: as α decreases, they scoop out the upper block until it splits in infinitely many rectangles at $t = (\sqrt{10}-2)/3$; see Figure 3.17. Finally, extending our research from $[2/5, \sqrt{2}-1]$ to $((\sqrt{10}-2)/3, \sqrt{2}-1]$ makes no difference for the way convergents are replaced by mediants as described earlier in this section.
We have now proved the following theorem:

Theorem 10. *Let $\alpha \in ((\sqrt{10}-2)/3, \sqrt{2}-1]$ and let the sequence E_n, $n \geq -1$, be defined by $E_{n+1} := 2E_n + E_{n-1}$, $n \geq 0$, with $E_{-1} := 1$ and $E_0 := 0$. Define*

$$V_{1,k} := [\tfrac{E_{2k+2}}{E_{2k+3}}, \tfrac{E_{2k+2}+g^2E_{2k+1}}{E_{2k+3}+g^2E_{2k+2}}) \cup [\tfrac{E_{2k+3}}{E_{2k+4}}, \tfrac{E_{2k+2}+gE_{2k+1}}{E_{2k+3}+gE_{2k+2}});$$
$$V_{2,k} := [\tfrac{E_{2k+1}+E_{2k}}{E_{2k+2}}, \tfrac{E_{2k+1}+E_{2k}+g^2(E_{2k}+E_{2k-1})}{E_{2k+2}+g^2E_{2k+1}}) \cup [\tfrac{E_{2k+2}+E_{2k+1}}{E_{2k+3}}, \tfrac{E_{2k+1}+E_{2k}+g(E_{2k}+E_{2k-1})}{E_{2k+2}+gE_{2k+1}});$$
$$V_{3,k} := [\tfrac{E_{2k}}{E_{2k+1}+E_{2k}}, \tfrac{E_{2k}+g^2E_{2k-1}}{E_{2k+1}+E_{2k}+g^2(E_{2k}+E_{2k-1})}) \cup [\tfrac{E_{2k+1}}{E_{2k+2}+E_{2k+1}}, \tfrac{E_{2k}+gE_{2k-1}}{E_{2k+1}+E_{2k}+g(E_{2k}+E_{2k-1})}).$$

Then

$$\Omega_\alpha = [\tfrac{1-3\alpha}{\alpha}, \alpha) \times [0, g^2) \cup [\tfrac{2-5\alpha}{3\alpha-1}, \alpha) \times [\tfrac{1}{2}, g)$$
$$\bigcup_{k=0}^{\infty} \left([\tfrac{2-5\alpha}{2\alpha-1}, \alpha) \times V_{1,k} \cup [\tfrac{1-2\alpha}{\alpha-1}, \tfrac{2-5\alpha}{3\alpha-1}) \times V_{2,k} \cup [\alpha-1, \tfrac{1-3\alpha}{\alpha}) \times V_{3,k} \right).$$

3.7. THE LONG WAY DOWN TO g^2

In the previous section we rewrote

$$x = [0; \varepsilon_1 a_1, \ldots, \varepsilon_n a_n, 3, -2, \varepsilon_{n+3} a_{n+3}, \ldots]$$

as

$$x = [0; \varepsilon_1 a_1, \ldots, \varepsilon_n(a_n+1), -2, -3, -\varepsilon_{n+3}(a_{n+3}+\varepsilon_{n+3}), \ldots].$$

3

In terms of Möbius transformations this is

$$\begin{aligned}
&\begin{pmatrix} 0 & \varepsilon_n \\ 1 & a_n \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 0 & \varepsilon_{n+3} \\ 1 & a_{n+3} \end{pmatrix} \\
=&\begin{pmatrix} 0 & \varepsilon_n \\ 1 & a_n+1 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 0 & -\varepsilon_{n+3} \\ 1 & a_{n+3}+\varepsilon_{n+3} \end{pmatrix} \qquad (3.12)\\
=&\begin{pmatrix} 5\varepsilon_n & 5\varepsilon_n a_{n+3}+3\varepsilon_n\varepsilon_{n+3} \\ 5a_n+2 & 5a_n a_{n+3}+3\varepsilon_{n+3}a_n \\ & +2a_{n+3}+\varepsilon_{n+3} \end{pmatrix};
\end{aligned}$$

We can extend and then generalise this: just as

$$\begin{aligned}
&\begin{pmatrix} 0 & \varepsilon_n \\ 1 & a_n \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 3 \end{pmatrix}\underbrace{\begin{pmatrix} 0 & -1 \\ 1 & 3 \end{pmatrix}\cdots\begin{pmatrix} 0 & -1 \\ 1 & 3 \end{pmatrix}}_{k-1 \text{ times}}\begin{pmatrix} 0 & -1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 0 & \varepsilon_{n+k+2} \\ 1 & a_{n+k+2} \end{pmatrix} \\
=&\begin{pmatrix} 0 & \varepsilon_n \\ 1 & a_n+1 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 2 \end{pmatrix}\underbrace{\begin{pmatrix} 0 & -1 \\ 1 & 3 \end{pmatrix}\cdots\begin{pmatrix} 0 & -1 \\ 1 & 3 \end{pmatrix}}_{k \text{ times}}\begin{pmatrix} 0 & -\varepsilon_{n+k+2} \\ 1 & a_{n+k+2}+\varepsilon_{n+k+2} \end{pmatrix} \qquad (3.13)\\
=&\begin{pmatrix} F_{2k+5}\varepsilon_n & F_{2k+5}\varepsilon_n a_{n+k+2}+F_{2k+4}\varepsilon_n\varepsilon_{n+k+2} \\ F_{2k+5}a_n+F_{2k+4} & F_{2k+5}a_n a_{n+k+2}+F_{2k+4}\varepsilon_{n+k+2}a_n \\ & +F_{2k+3}a_{n+k+2}+F_{2k+2}\varepsilon_{n+k+2} \end{pmatrix},
\end{aligned}$$

which can be derived from (3.12) by inserting $\begin{pmatrix} 0 & -1 \\ 1 & 3 \end{pmatrix}$-matrices at the right place, we have

$$\begin{aligned}
&[0;\varepsilon_1 a_1,\dots,\varepsilon_n a_n, 3, \underbrace{-3,\dots,-3}_{k-1 \text{ times}}, -2, \varepsilon_{n+k+2}a_{n+k+2},\dots] \\
=&[0;\varepsilon_1 a_1,\dots,\varepsilon_n(a_n+1), -2, \underbrace{-3,\dots,-3}_{k \text{ times}}, -\varepsilon_{n+k+2}(a_{n+k+2}+\varepsilon_{n+k+2}),\dots].
\end{aligned}$$

We observe that

$$\lim_{k\to\infty}[0;3,\underbrace{-3,\dots,-3}_{k-1 \text{ times}}] = g^2.$$

The good part of this is that it appears to be possible to apply compensated insertion beyond $(\sqrt{10}-2)/3$, by means of extending the $(3,-2)$-insertion by adding partial quotients -3 between 3 and -2. The bad part is that this possibility vanishes once we reach g^2. Taking the approach we took so far, the thing to do is letting α decrease from $(\sqrt{10}-2)/3$ to g^2 via the sequence $[0;3,\underbrace{-3,\dots,-3}_{k-1 \text{ times}},-2,\varepsilon_{k+2}a_{k+2},\dots]$, starting with the case $k=2$, which is $[0;3,-3,-2,\varepsilon_4 a_4,\dots]$.

In the previous sections we have seen how our approach of singularisation and insertion gradually involved more details to process. We will show how things rapidly become even more complex, making it impossible to continue this approach much longer. So far, we have been able to construct Ω_α for all $\alpha \in [(\sqrt{10}-2)/3, 1)$. The next set of numbers to tackle would seem to consist of numbers $x = [0; \varepsilon_1 a_1, \ldots, \varepsilon_n a_n, 3, -3, -2, \varepsilon_{n+4} a_{n+4}, \ldots]$ and $t_n = [0; 3, -3, -2, \varepsilon_{n+4} a_{n+4}, \ldots]$, the largest of which is

$$(\sqrt{10}-2)/3 = [0; 3, \overline{-3, -2, -3, -4}].$$

Let us first determine the interval (α', α), $\alpha' < \alpha \le (\sqrt{10}-2)/3$, such that $t_n \in (\alpha', \alpha)$ implies $t_n = [0; 3, -3, -2, \ldots]$. We remark that

$$f_2(x) := \cfrac{1}{3 - \cfrac{1}{3 - \frac{1}{x}}},$$

where the index 2 indicates the number of successive 3s, is a decreasing function of x on $\mathbb{R} \setminus \{0\}$, so if $\alpha' \in [0, (\sqrt{10}-2)/3)$ and $\alpha \in (\alpha', (\sqrt{10}-2)/3]$, then

$$2 = \left\lfloor f\left(\frac{\sqrt{10}-2}{3}\right) + 1 - \frac{\sqrt{10}-2}{3} \right\rfloor \le a_3(\alpha'_\alpha) = \lfloor f(\alpha') + 1 - \alpha \rfloor \le \lfloor f(\alpha') + 1 - \alpha' \rfloor, \qquad (3.14)$$

from which we derive $a_3(\alpha'_\alpha) = 2$ if and only if $f(\alpha') + 1 - \alpha' < 3$. Since $\alpha' \in [0, (\sqrt{10}-2)/3))$, this is the case if and only if $(\sqrt{65}-5)/8 < \alpha' < (\sqrt{10}-2)/3$; indeed

$$(\sqrt{65}-5)/8_{(\sqrt{65}-5)/8} = [0; 3, \overline{-3, -3, -2, -3, -3, -4}].$$

So the next step of our investigation would be the case

$$\alpha \in \Delta_\alpha(3, -3, -2) = ((\sqrt{65}-5)/8, (\sqrt{10}-2)/3].$$

Proceeding in a similar way, a sequence of successive intervals to be investigated would come into view, the boundaries of which are the positive roots R_k of

$$f_k(x) := 1/\underbrace{(3 - 1/(3 - \ldots 1/(3}_{k \text{ times a } 3} - 1/x)\ldots)) - x - 2.$$

Straightforward calculation shows that

$$R_k = \frac{\sqrt{F_{2k+1} F_{2k+3}} - F_{2k+1}}{F_{2k+2}}$$

and

$$\lim_{k\to\infty} R_k = \lim_{k\to\infty} \left(\sqrt{\frac{F_{2k+1} F_{2k+3}}{F_{2k+2}^2}} - \frac{F_{2k+1}}{F_{2k+2}} \right) = \sqrt{g \cdot G} - g = 1 - g = g^2,$$

as was to be expected. For our investigations we would also need that

$$R_{k R_k} = [0; 3, \overline{\underbrace{-3, \ldots, -3}_{k \text{ times}}, -2, \underbrace{-3, \ldots, -3}_{k \text{ times}}, -4}].$$

Although this perspective would already promise a lot of intricacies, when we turn back to the case $\alpha \in ((\sqrt{65}-5)/8, (\sqrt{10}-2)/3]$, we are confronted with a problem we had not encountered yet: if we apply compensated insertion to numbers equal to or only slightly smaller than $(\sqrt{10}-2)/3$, we get

$$\alpha_\alpha = [0;3,-3,-2,-3,-a,\ldots] = [1;-2,-3,-3,2,-a,\ldots],\ \ a \geq 4.$$

Of course, the occurrence of the partial quotient 2 is not allowed, which calls for one more compensated insertion, so as to get

$$\alpha_\alpha = [0;3,-3,-2,-3,-a,\ldots] = [1;-2,-3,-3,2,-a,\ldots] = [1;-2,-3,-4,-2,a-1,\ldots],$$

with $a \geq 4$. To determine the interval where these numbers occur, we have to solve

$$\cfrac{1}{2-\cfrac{1}{3-\cfrac{1}{3-\cfrac{1}{\alpha}}}} + 1 - \alpha = 4.$$

The positive solution of this equation is

$$(5\sqrt{13}-13)/13_{(5\sqrt{13}-13)/13} = [0;3,\overline{-3,-2,-4,-2,-3,-4,-2,-4}] = 0.386750\ldots.$$

So the next case to investigate is actually $\alpha \in \Delta_\alpha(3,-3,-2,-3) = ((5\sqrt{13}-13)/13, (\sqrt{10}-2)/3]$. More precisely, let $x \in [(\sqrt{10}-5)/3, (\sqrt{10}-2)/3)$ and $n \geq 0$ be the smallest integer for which $t_n \in [\alpha, (\sqrt{10}-2)/3]$, i.e., for which

$$(t_n, v_n) \in \bigcup_{k=0}^{\infty} \left\{ [\alpha, (\sqrt{10}-2)/3] \times \left([\tfrac{E_{2k}}{E_{2k+1}}, \tfrac{E_{2k}+g^2E_{2k-1}}{E_{2k+1}+g^2E_{2k}}) \cup [\tfrac{E_{2k+1}}{E_{2k+2}}, \tfrac{E_{2k}+gE_{2k-1}}{E_{2k+1}+gE_{2k}}) \right) \right\}.$$

Above we saw that in this case

$$\begin{aligned} x &= [0;\varepsilon_1 a_1,\ldots,\varepsilon_n a_n, 3,-3,-2,-3,-a_{n+5},\ldots]\ , a_{n+5} \geq 4,\ \text{that we rewrite as} \\ x &= [0;\varepsilon_1 a_1,\ldots,\varepsilon_n (a_n+1),-2,-3,-4,-2,a_{n+5}-1,\ldots]\ , a_{n+5} \geq 4. \end{aligned} \tag{3.15}$$

We now have $(t_n^*, v_n^*) = (t_n - 1, \frac{v_n}{v_n+1})$ and

$$\begin{aligned}
(t_{n+1}, v_{n+1}) &= (\tfrac{1-3t_n}{t_n}, \tfrac{1}{v_n+3}) & (t_{n+1}^*, v_{n+1}^*) &= (\tfrac{1-2t_n}{t_n-1}, \tfrac{v_n+1}{v_n+2}) \\
(t_{n+2}, v_{n+2}) &= (\tfrac{3-8t_n}{3t_n-1}, \tfrac{v_n+3}{3v_n+8}) & (t_{n+2}^*, v_{n+2}^*) &= (\tfrac{2-5t_n}{2t_n-1}, \tfrac{v_n+2}{2v_n+5}) \\
(t_{n+3}, v_{n+3}) &= (\tfrac{5-13t_n}{8t_n-3}, \tfrac{3v_n+8}{5v_n+13}) & (t_{n+3}^*, v_{n+3}^*) &= (\tfrac{7-18t_n}{5t_n-2}, \tfrac{2v_n+5}{7v_n+18}) \\
(t_{n+4}, v_{n+4}) &= (\tfrac{12-31t_n}{13t_n-5}, \tfrac{5v_n+13}{12v_n+31}) & (t_{n+4}^*, v_{n+4}^*) &= (\tfrac{12-31t_n}{18t_n-7}, \tfrac{7v_n+18}{12v_n+31}), \\
(t_{n+5}, v_{n+5}) &= (t_{n+5}^*, v_{n+5}^*)
\end{aligned}$$

Although it would still be possible to explicitly give Ω_α for $\alpha \in ((5\sqrt{13}-13)/13, (\sqrt{10}-2)/3]$, it may be clear that it becomes quite unmanageable. Determining the effect of the current transformation on the sequence of convergents is still easy, though. Since

$p_k^* = p_k$ and $q_k^* = q_k$, $k < n$, we find, applying (1.2) and referring to (3.15) (while omitting the straightforward calculations),

$$\begin{cases} p_n^* = p_n + p_{n-1}; \\ p_{n+1}^* = p_{n+1} - p_n; \\ p_{n+2}^* = p_{n+2} - p_{n+1}; \\ p_{n+3}^* = p_{n+4} - p_{n+3}; \\ p_{n+4}^* = p_{n+4}; \\ p_{n+5}^* = p_{n+5}. \end{cases}$$

Of course, we find similar relations for q_n^* through q_{n+5}^*. We conclude that in the transformation from $\Omega_{(\sqrt{10}-2)/3}$ to Ω_α, $\alpha \in [(5\sqrt{13}-13)/13, (\sqrt{10}-2)/3)$, the convergent p_n/q_n is replaced by the mediant $(p_n+p_{n-1})/(q_n+q_{n-1})$, that p_{n+1}/q_{n+1} is replaced by the mediant $(p_{n+1}-p_n)/(q_{n+1}-q_n)$, p_{n+2}/q_{n+2} is replaced by the mediant $(p_{n+2}-p_{n+1})/(q_{n+2}-q_{n+1})$ and that p_{n+3}/q_{n+3} is replaced by the mediant $(p_{n+4}-p_{n+3})/(q_{n+4}-q_{n+3})$.

A substantial part of our construction of Ω_α, $(\sqrt{10}-2)/3 \le \alpha < 1$, consisted in infinitely repeating the removal and addition of rectangles in $[(\sqrt{10}-5)/3, 1] \times [0,1]$, in the cases that part of the added regions overlapped part of the regions to be removed. In Section 3.4, for instance, we worked at the case $x = [0; \varepsilon_1 a_1, \ldots, \varepsilon_n a_n, 2, -a_{n+2}, \ldots]$, $a_{n+2} \ge 3$, and rewrote it as

$$x = [0; \varepsilon_1 a_1, \ldots, \varepsilon_n (a_n+1), -2, (a_{n+2}-1), \ldots]. \tag{3.16}$$

The overlapping stems from the possibility that (3.16) can be written as

$$x = [0; \varepsilon_1 a_1, \ldots, \varepsilon_n (a_n+1), -2, 2, \varepsilon_{n+3} a_{n+3}, \ldots],\ a_{n+3} \ge 3, \tag{3.17}$$

i.e. when $a_{n+2} = 3$. One could ask whether the work of infinite removal and addition could have possibly been avoided by immediately performing one more compensated insertion and so rewriting (3.17) as

$$x = [0; \varepsilon_1 a_1, \ldots, \varepsilon_n (a_n+1), -3, -2, -\varepsilon_{n+3}(a_{n+3}+\varepsilon_{n+3}), \ldots].$$

This, however, confronts us with some intricacies, among which $t_{n+2}^* \ne t_{n+2}$, that appear to at least cancel the benefits of this approach. Earlier in this section we showed that for $\alpha \le (\sqrt{10}-2)/3$, this repeated compensated insertion is unavoidable, as is a further complication of the construction of Ω_α. One obvious problem is the growth of the number of regions to be removed and added in each step of construction, due to the lengthening of the continued fraction expansions of the numbers involved on the way down to

3

$g^2 = 0.381966\ldots$, the first of which are (in decreasing order):

$$
\begin{aligned}
(5\sqrt{13}-13)/13_{(5\sqrt{13}-13)/13} &= [0;3,\overline{-3,-2,-4,-2,-3,-4,-2,-4}] \\
&= 0.3867504\ldots, \\
5/13_{5/13} &= [0;3,-3,-2] \\
&= 0.3846153\ldots, \\
(\sqrt{65}-5)/8_{(\sqrt{65}-5)/8} &= [0;3,\overline{-3,-3,-2,-3,-3,-4}] \\
&= 0.3827822\ldots, \\
(\sqrt{14401}-89)/81_{(\sqrt{14401}-89)/81} &= [0;3,-3,-3,-2,-3,-4,-2,-3,-3,-4,-3,-2,\ldots] \\
&= 0.3827674\ldots, \\
(\sqrt{2210}-34)/34_{(\sqrt{2210}-34)/34} &= [0;3,\overline{-3,-3,-2,-4,-2,-3,-3,-4,-2,-4}] \\
&= 0.3826657\ldots, \\
13/34_{13/34} &= [0;3,-3,-3,-2] \\
&= 0.3823529\ldots, \\
(\sqrt{442}-13)/21_{(\sqrt{442}-13)/21} &= [0;3,\overline{-3,-3,-3,-2,-3,-3,-3,-4}] \\
&= 0.3820855\ldots, \\
(\sqrt{670762}-610)/547_{(\sqrt{670762}-610)/547} &= [0;3,-3,-3,-3,-2,-3,-3,-4,-2,-3,-3,-3\ldots] \\
&= 0.3820852\ldots
\end{aligned}
$$

The red thread in this list is the first appearance of the partial quotient 2 with minus sign in the continued fraction expansion α_α of the related decreasing numbers α, always precedented by the sequence $3,\underbrace{-3,\ldots,-3}_{k \text{ times}},\ k>0$. With k increasing, the number of partial quotients 3 with minus sign following this 2 before the first appearance of the partial quotient 4 with minus sign may be from 0 to k, each case calling for an α-fundamental interval to be investigated separately. On our way down from

$$(\sqrt{65}-5)/8_{(\sqrt{65}-5)/8} = [0;3,\overline{-3,-3,-2,-3,-3,-4}]$$

to

$$(\sqrt{442}-13)/21_{(\sqrt{442}-13)/21} = [0;3,\overline{-3,-3,-3,-2,-3,-3,-3,-4}],$$

for instance, we have to distinguish between the cases associated with the sequences starting with $3,-3,-3,-2,-3,-3,-4$, then $3,-3,-3,-2,-3,-4$, then $3,-3,-3,-2,-4$ and $3,-3,-3,-2$ before arriving at $3,-3,-3,-3,\ldots$.

In sections 3.3, 3.5 and 3.6 we saw how (compensated) insertion in most cases involves loss of convergents, although for $\alpha \le 1/2$ each convergent is replaced by a mediant of the form $(p_k \pm p_{k-1})/(q_k \pm q_{k-1})$ for some $k \in \mathbb{N}$. As α decreases beyond $(5\sqrt{13}-13)/13$, this will remain the case, which we can best illustrate by again using Möbius transformations. If p_{n-1} and p_n are two consecutive denominators of a convergent of some number

$x = [0; \varepsilon_1 a_1, \dots, \varepsilon_{n-1} a_{n-1}, \varepsilon_n a_n, \dots]$, then

$$\begin{pmatrix} p_{n+1} \\ p_n \end{pmatrix} = \begin{pmatrix} a_{n+1} & \varepsilon_{n+1} \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p_n \\ p_{n-1} \end{pmatrix}.$$

The denominators p^*_{n+k}, $k \in \{0, \dots, 5\}$ that we calculated in the previous section, can easily be derived by comparing the partial products of

$$\begin{pmatrix} p^*_{n+5} \\ p^*_{n+4} \end{pmatrix} = \begin{pmatrix} a_{n+5}-1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 4 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_n+1 & \varepsilon_n \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p_{n-1} \\ p_{n-2} \end{pmatrix}$$

with those of

$$\begin{pmatrix} p_{n+5} \\ p_{n+4} \end{pmatrix} = \begin{pmatrix} a_{n+5} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 3 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_n & \varepsilon_n \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p_{n-1} \\ p_{n-2} \end{pmatrix},$$

associated with the sequence $\varepsilon_n(a_n+1), -2, -3, -4, -2, a_{n+5}-1$ as derived from the sequence $\varepsilon_n a_n, 3, -3, -2, -3, -a_{n+5}$ by applying compensated insertion.
On the interval $((\sqrt{14401}-89)/81, (\sqrt{65}-5)/8]$ we could similarly compare the partial products of

$$\begin{pmatrix} p^*_{n+7} \\ p^*_{n+6} \end{pmatrix} = \begin{pmatrix} a_{n+7}-1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 4 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 3 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_n+1 & \varepsilon_n \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p_{n-1} \\ p_{n-2} \end{pmatrix}$$

with those of

$$\begin{pmatrix} p_{n+7} \\ p_{n+6} \end{pmatrix} = \begin{pmatrix} a_{n+7} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 3 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_n & \varepsilon_n \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p_{n-1} \\ p_{n-2} \end{pmatrix},$$

associated with the sequence $\varepsilon_n(a_n+1), -2, -3, -3, -4, -3, -2, a_{n+7}-1$ as derived from the sequence $\varepsilon_n a_n, 3, -3, -3, -2, -3, -3, -a_{n+7}$ by applying compensated insertion. It is not hard to see that as α decreases to g^2, it is just the number of lost convergents that increases, but not the way in which they are replaced by mediants as described above; the difference is merely the number of matrices $\begin{pmatrix} 3 & -1 \\ 1 & 0 \end{pmatrix}$ in both products, similar to the difference made by the number of matrices $\begin{pmatrix} 0 & -1 \\ 1 & 3 \end{pmatrix}$ in (3.13).

3.8. THE ERGODIC SYSTEMS $(\Omega_\alpha, \mathscr{B}, \mu_\alpha, \mathscr{T}_\alpha)$

Starting with Ω_1, we constructed domains $\Omega_{\alpha'}$ by removing sets of points from a given Ω_α, with $\alpha' < \alpha$, and adding other sets to it. Each stage of construction – corresponding with successive sections of this chapter – had its own characteristic set of singularisations, such as simply singularising a partial quotient 1 for $\alpha' \in (g, 1]$. The first step of constructing consisted of fixing a subset $R_{\alpha'}$ of Ω_α, consisting of all points $(t, v) \in \Omega_\alpha$ for which $\alpha' \le t < \alpha$.

Now let the collection of subsets of Ω_α be denoted by $\mathscr{B}$ and μ_α be defined as the probability measure with density

$$\frac{1}{N_\alpha} \cdot \frac{1}{(1+tv)^2}$$

on $(\Omega_\alpha, \mathscr{B})$, where N_α is a normalising constant. In [9], Nakada showed that for $\alpha \in [1/2, 1]$ the dynamical systems $(\Omega_\alpha, \mathscr{B}, \mu_\alpha, \mathscr{T}_\alpha)$ are ergodic (he actually obtained stronger mixing properties); see also [6]. Here we will use this result for $\alpha = 1$ only, which was the starting point in [6] as well. Note that for $g^2 < \alpha < g$ our construction actually is a bijection between $\Omega_{\alpha'}$ and Ω_α, $\alpha' < \alpha$. Clearly this is not the case for $g < \alpha' < \alpha \le 1$, where one region is added to Ω_α while two are removed from it when constructing $\Omega_{\alpha'}$.

We will now show that μ_α is an invariant measure on Ω_α. Let $x_\alpha = [0; \varepsilon_1 a_1, \dots, \varepsilon_n a_n, \dots]$ and $D := [t_1, t_2] \times [v_1, v_2] \subset R_{\alpha'}$. Then

$$m(D) := \iint_D \frac{1}{(1+tv)^2}\, dt\, dv = \log \frac{(1+t_2 v_2)(1+t_1 v_1)}{(1+t_2 v_1)(1+t_1 v_2)}.$$

Let $D_k := \mathscr{T}_\alpha^k(D)$ and M_n be the Möbius transformation associated with the matrix

$$\begin{pmatrix} 0 & \varepsilon_n \\ 1 & a_n \end{pmatrix}.$$

We define $\widetilde{M}_k := M_{n+1} \cdots M_{n+k}$. From (3.3) we derive that

$$D_k = \left[\widetilde{M}_k^{-1}(t_1), \widetilde{M}_k^{-1}(t_2)\right] \times \left[\widetilde{M}_k^{T}(v_1), \widetilde{M}_k^{T}(v_2)\right],$$

from which it follows that

$$\begin{aligned} m(D_k) &= \iint_{D_k} \frac{1}{(1+tv)^2}\, dt\, dv \\ &= \log \frac{(1+\widetilde{M}_k^{-1}(t_2)\widetilde{M}_k^{T}(v_2))(1+\widetilde{M}_k^{-1}(t_1)\widetilde{M}_k^{T}(v_1))}{(1+\widetilde{M}_k^{-1}(t_2)\widetilde{M}_k^{T}(v_1))(1+\widetilde{M}_k^{-1}(t_1)\widetilde{M}_k^{T}(v_2))} \\ &= \log \frac{|\widetilde{M}_k|(t_2 v_2 + 1)|\widetilde{M}_k|(t_1 v_1 + 1)}{|\widetilde{M}_k|(t_2 v_1 + 1)|\widetilde{M}_k|(t_1 v_2 + 1)} \\ &= \log \frac{(1+t_2 v_2)(1+t_1 v_1)}{(1+t_2 v_1)(1+t_1 v_2)} = m(D). \qquad (*) \end{aligned}$$

Rewriting

$$x_\alpha = [0; \varepsilon_1 a_1, \varepsilon_2 a_2, \ldots, \varepsilon_n a_n, a_{n+1}, \varepsilon_{n+2} a_{n+2}, \ldots] \text{ as}$$
$$x_\alpha = [0; \varepsilon_1 a_1, \varepsilon_2 a_2, \ldots, \varepsilon_n (a_n + 1), -a^*_{n+1}, \varepsilon^*_{n+2} a^*_{n+2}, \ldots],$$

as we have done throughout this chapter, we define

$$D^* := [t_1 - 1, t_2 - 1] \times [v_1/(1+v_1), v_2/(1+v_2)] \in A_{\alpha'}.$$

With a straightforward calculation one finds that $m(D^*) = m(D)$. Similarly to what we have written above, we define $D^*_k := \mathscr{T}^k_\alpha(D^*)$ and let M^*_n be the Möbius transformation associated with the matrix $\begin{pmatrix} 0 & \varepsilon^*_n \\ 1 & a^*_n \end{pmatrix}$. We define $\widetilde{M}^*_k := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} M^*_{n+1} \cdots M^*_{n+k}$, where $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is associated with mapping $R_{\alpha'}$ to $A_{\alpha'}$. Again applying (3.3), we find that

$$D^*_k = \left[(\widetilde{M}^*_k)^{-1}(t_1), (\widetilde{M}^*_k)^{-1}(t_2)\right] \times \left[(\widetilde{M}^*_k)^T(v_1), (\widetilde{M}^*_k)^T(v_2)\right],$$

from which it follows that

$$\begin{aligned} m(D^*_k) &= \iint_{D_k} \frac{1}{(1+tv)^2}\, dt\, dv \\ &= \log \frac{(1 + (\widetilde{M}^*_k)^{-1}(t_2)(\widetilde{M}^*_k)^T(v_2))(1 + (\widetilde{M}^*_k)^{-1}(t_1)(\widetilde{M}^*_k)^T(v_1))}{(1 + (\widetilde{M}^*_k)^{-1}(t_2)(\widetilde{M}^*_k)^T(v_1))(1 + (\widetilde{M}^*_k)^{-1}(t_1)(\widetilde{M}^*_k)^T(v_2))} \\ &= \log \frac{|\widetilde{M}^*_k|(t_2 v_2 + 1)|\widetilde{M}^*_k|(t_1 v_1 + 1)}{|\widetilde{M}^*_k|(t_2 v_1 + 1)|\widetilde{M}^*_k|(t_1 v_2 + 1)} \\ &= \log \frac{(1 + t_2 v_2)(1 + t_1 v_1)}{(1 + t_2 v_1)(1 + t_1 v_2)} = m(D). \end{aligned}$$

From this and (*) it follows that $\mu(D^*_k) = \mu(D)$ for any pair of α-measurable sets D_k and D. Since the dynamical system $(\Omega_1, \mathscr{B}, \mu_1, T_1)$ is ergodic, from our construction it easily follows that $(\Omega_\alpha, \mathscr{B}, \mu_\alpha, \mathscr{T}_\alpha)$ forms an ergodic system as well. Indeed, from our construction and regarding the above observations on m, it follows that for $\alpha \in [g^2, g)$, all Ω_α are isomorphic to Ω_g, so all dynamical systems $(\Omega_\alpha, \mathscr{B}, \mu_\alpha, \mathscr{T}_\alpha)$, $\alpha \in [g^2, 1)$, will 'inherit' ergodicity from $(\Omega_g, \mathscr{B}, \mu_g, \mathscr{T}_g)$. Obviously, for $\alpha \in [g, 1)$ the dynamical systems $(\Omega_\alpha, \mathscr{B}, \mu_\alpha, \mathscr{T}_\alpha)$ are *not* isomorphic to Ω_1, but from [6] we know they are isomorphic to an induced transformation of Ω_1, which is also clear from our construction; for $g < \alpha < 1$, let $I_\alpha \subseteq \Omega_\alpha$ be invariant under $\mathscr{T}_\alpha$. We define $I_{\alpha,1} := I_\alpha \cap A_\alpha$ and $I_{\alpha,2} := I_\alpha \setminus I_{\alpha,1}$. Then $I_{\alpha,2} \subseteq \Omega_1$. Furthermore, putting $I_{1,1} := \mathscr{M}^{-1}(I_{\alpha,1})$ (with $\mathscr{M}$ the map defined in Section 3.3), $I_{1,2} := \mathscr{T}_1(I_{1,1})$, and $I_1 := I_{\alpha,2} \cup I_{1,1} \cup I_{1,2}$, then $I_1 \subseteq \Omega_1$ is by construction $\mathscr{T}_1$-invariant, and therefore I_1 has μ_1-measure 0 or 1. Since $m(I_{\alpha,1}) = m(I_{1,1}) = m(I_{1,2})$, it follows that $m(I_\alpha) = m(I_1) - m(I_{1,1})$. If $\mu_1(I_1) = 0$, we are done. If $\mu_1(I_1) = 1$, we see that $m(I_\alpha) = \log 2(\mu_1(I_1) - \mu_1(I_{1,1})) = \log 2(1 - \mu_1(R_\alpha)) = \log 2 - m(R_\alpha) = N_\alpha$, and we see that

$$\mu_\alpha(I_\alpha) = \frac{1}{N_\alpha} m(I_\alpha) = 1;$$

so if $I_\alpha \subseteq \Omega_\alpha$ is a measurable $\mathscr{T}_\alpha$-invariant set, we find that $\mu_\alpha(I_\alpha) \in \{0,1\}$, and we conclude that $\mathscr{T}_\alpha$ is ergodic.

Writing $h(\mathscr{T}_\alpha)$ for the *entropy* of $\mathscr{T}_\alpha$, in [9] Nakada also proved that

$$h(\mathscr{T}_\alpha) = \begin{cases} \frac{\pi^2}{6\log(1+\alpha)}, & \alpha \in (g,1]; \\ \frac{\pi^2}{6\log(G)}, & \alpha \in [\frac{1}{2}, g]. \end{cases}$$

For $g < \alpha < 1$ this result is essentially due to the fact that Ω_α is isomorphic to an appropriate induced transformation of $(\Omega_1, \mathscr{B}, \mu_1, T_1)$ and Abramov's formula; see [6] for explicit details. This explains the increase of entropy as α decreases from 1 to g. In the previous sections, we have shown that for $\alpha \in (g^2, g]$, the loss of R_α is completely compensated by the addition of A_α, from which it follows that the systems $(\Omega_\alpha, \mathscr{B}, \mu_\alpha, \mathscr{T}_\alpha)$ are metrically isomorphic to $(\Omega_g, \mathscr{B}, \mu_g, \mathscr{T}_g)$. We conclude that $h(\mathscr{T}_\alpha) = \pi^2/(6\log G), \alpha \in (g^2, g]$, a result previously obtained also in [7]. For α smaller than g^2, the entropy map $h(\alpha)$ behaves behaves quite irregularly; see [8, 11] and Giulio Tiozzo's thesis [13] for further details.

3.9. THE α-LEGENDRE CONSTANT

In ([10]), Nakada obtains a very strong connection between two constants that play an important role in the study of continued fractions: the *Legendre constant* and the *Lenstra constant*. The first is associated with the theorem of Legendre, mentioned on page 48, where we also introduced the α-Legendre constant

$$L(\alpha) := \sup\{c > 0 : q^2 \left| x - \frac{p}{q} \right| < c, \gcd(p,q) = 1 \Rightarrow \frac{p}{q} = \frac{p_n}{q_n}, n \geq 0\}.$$

The Lenstra constant is associated with the following conjecture by W. Doeblin and Hendrik Lenstra on regular continued fractions, proved by Wieb Bosma, Hendrik Jager and Freek Wiedijk in [3]:

Theorem 11. *For almost all x and all $t \in [0,1]$, the limit*

$$\lim_{n\to\infty} \frac{1}{n} \#\{j; 1 \leq j \leq n : \vartheta_j \leq t\}$$

exists, and is equal to the distribution function

$$F(t) = \begin{cases} \frac{t}{\log 2}, & 0 \leq t \leq 1/2, \\ \frac{1}{\log 2}(1 - t + \log 2t), & 1/2 \leq t \leq 1. \end{cases}$$

The number 1/2, the upper bound of the interval where $F(t)$ is linear, is called the *Lenstra constant of regular continued fractions*; for α-expansions the Lenstra constant can be defined similarly.

Let $\tilde{g} := \dfrac{g-2+\sqrt{g^2+4}}{2g} = 0.57549\ldots$. In 1988, Shunji Ito ([5]; see also [6]) proved that

$$L(\alpha) = \begin{cases} \dfrac{\alpha}{\alpha+1}, & g \le \alpha \le 1; \\ 1-\alpha, & \tilde{g} \le \alpha \le g; \\ \dfrac{\alpha}{1+g\alpha}, & \frac{1}{2} \le \alpha \le \tilde{g}. \end{cases}$$

In 2010, Nakada proved in [10] that for a continued fraction algorithm the Legendre constant equals the Lenstra constant whenever the Legendre constant exists; one year later Natsui ([12]) proved the existence of the α-Legendre constant for $0 < \alpha \le 1$. She also extends Ito's result to

$$L(\alpha) = \frac{\alpha}{1+g\alpha}, \quad \sqrt{2}-1 \le \alpha \le \tilde{g}.$$

In this section we will in turn extend Natsui's result to $g^2 \le \alpha < \tilde{g}$. Moreover, we will show how we can use a result of Laura Luzzi and Stefano Marmi ([8]) to make some observations on $L(\alpha)$ for $\alpha < g^2$.

From Section 3.7 it is clear that it is both very hard and unmanageable to give Ω_α explicitly for $\alpha \in [g^2, (\sqrt{10}-2)/3]$. We can, however, give some useful characteristics for this case with regard to $L(\alpha)$. The first thing we note is that $0 \le v < g$ for $(t, v) \in \Omega_\alpha$, where $\alpha \in [g^2, 1/2]$. This follows from the fact that in our method of constructing Ω_α from $\Omega_{\alpha'}$, with $\alpha < \alpha'$, any -2 that is introduced by insertion and singularisation is precedented by either a 3 with minus sign or at least a 4. Any second coordinate v in Ω_α will therefore be smaller than $1/(2-g^2) = g$. From this it follows that the bottom side of the lowermost cove in Figure 3.18 is on the line $v = 1/(3+g)$; see our calculations in Section 3.6. Figure 3.18 is an example of Ω_α with α near $(\sqrt{10}-2)/3$, and is more or less the same as Figures 3.16 and 3.17. The difference between Figure 3.18 and the other two figures is the addition of curves associated with $\vartheta_{n-1} = \alpha/(1+g\alpha)$ and $\vartheta_{n-1} > \alpha/(1+g\alpha)$, the choice of which we will explain now.

Recall that in the introduction we defined $\vartheta_n(x) := q_n^2|x - p_n/q_n|$ and mentioned the equations $\vartheta_{n-1} = v_n/(1+t_n v_n)$ and $\vartheta_n = (\varepsilon_n t_n)/(1+t_n v_n)$ (omitting the suffix '(x)'). In [6], $L(\alpha)$ is found by determining the largest C for which the curve $\vartheta_n = C$ contains no points in what is called a singularisation area: indeed, this area is associated with convergents that are replaced (when $1/2 < \alpha \le g$) or lost (when $g < \alpha < 1$ or $\alpha \le 1/2$), as we have shown in Sections 3.3 through 3.7. Note that this C is also the Lenstra constant for the α-expansion under consideration. In the current case ($g^2 \le \alpha < \sqrt{2}-1$), we find it convenient to consider curves $\vartheta_{n-1} = C$ rather than $\vartheta_n = C$, if only for the numerator of $\vartheta_n = \varepsilon_n t_n/(1+t_n v_n)$, yielding two branches that we would have to take into account separately. Since curves associated with the equation $\vartheta_{n-1} = C$, C small enough, will be in the lower part of Ω_α, a first guess would be that $L(\alpha)$ is determined by either the points $M_\alpha := ((1-3\alpha)/\alpha, 1/(3+g))$ or $N_\alpha := (\alpha, g^2)$. No matter how intricate Ω_α becomes for $\alpha \le \sqrt{2}-1$, from our earlier observations it follows that these v-coordinates $1/(3+g)$ and g^2 do not change as α decreases to g^2. In Figure 3.18 we see how curves associated with $\vartheta_{n-1} > \alpha/(1+g\alpha)$ contain points in the cove left and above M_α. We find

$\vartheta_{n-1}(M_\alpha) = \alpha/(1+g\alpha)$, while $\vartheta_{n-1}(N_\alpha) = 1/(G^2+\alpha)$. Since

$$\min\{\alpha/(1+g\alpha), 1/(G^2+\alpha)\} = \alpha/(1+g\alpha)$$

on $[g^2, \sqrt{2}-1)$, we conclude that

$$L(\alpha) = \frac{\alpha}{1+g\alpha}, \quad g^2 \leq \alpha < \sqrt{2}-1,$$

hence extending Natsui's result.

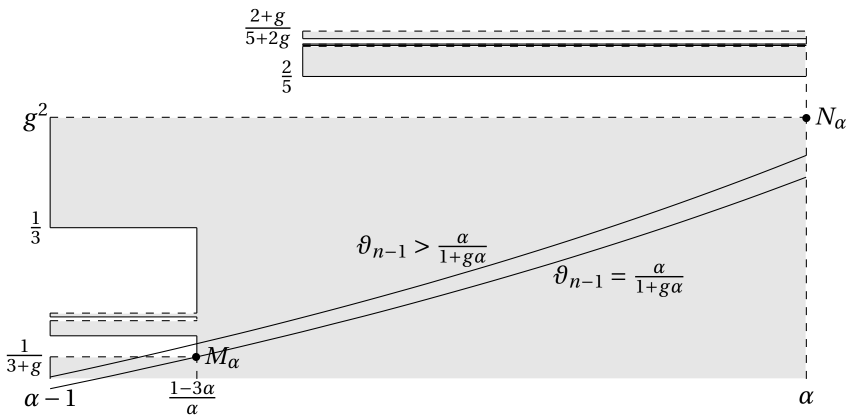


Figure 3.18: The lower part of Ω_α, with α near $\frac{\sqrt{10}-2}{3}$

For various reasons, this approach cannot be extended to values of $\alpha < g^2$. For one thing, g will cease to be the upper bound for the values of v, since insertions and singularisations in the case $\alpha < g^2$ involve strings of more than one -2, yielding values of v larger than g. This implies that points have to be removed with v-values under $1/(3+g)$. Unfortunately, we cannot say much more about Ω_α for general α smaller than g^2. Still, a result of Luzzi and Marmi (see [8], pages 27 and 28) enables us to at least find how $L(\alpha)$ evolves for these rational values of α. In [8], Luzzi and Marmi use the aforementioned strings of partial quotients 2 with minus sign to give a description of Ω_α for $\alpha = 1/r$, $r \in \mathbb{N}_{\geq 3}$. Although their description is not very explicit, it is not hard to find that it yields that $\Omega_{1/r}$ consists of the bottom part

$$[-1+1/r, 0] \times [0, (r+1-\sqrt{(r+1)^2-4})/2] \cup [0, 1/r] \times [0, (\sqrt{(r+1)^2-4}-r+1)/(2r-2)]$$

and a myriad of vertically disjunct rectangles above it; see Figure 3.19 for the bottom part. Taking the approach of considering curves $\vartheta_{n-1} = C$ as we did above, we see that the Lenstra constant for $\alpha = 1/r$ is determined by the vertex $(0, (r+1-\sqrt{(r+1)^2-4})/2)$ (which is the point $(0, (5-\sqrt{21})/2)$ in Figure 3.19). Thanks to Nakada's result in [10] it follows that

$$L\left(\tfrac{1}{r}\right) = \frac{r+1-\sqrt{(r+1)^2-4}}{2}. \tag{3.18}$$

Although this formula is different from the one we found for $g^2 \le \alpha \le 1/2$, we remark that $L(1/2) = g^2$ fits both formulas. What's more, we suspect that with (3.18) a sharpening is possible of Natsui's result on the boundaries of $L(\alpha)$ in [12], which is:

Theorem 12. *Let $r \in \mathbb{N}$. If $\frac{1}{r+1} < \alpha < \frac{1}{r}$, then $\frac{1}{r+2} < L(\alpha) < \frac{1}{r}$.*

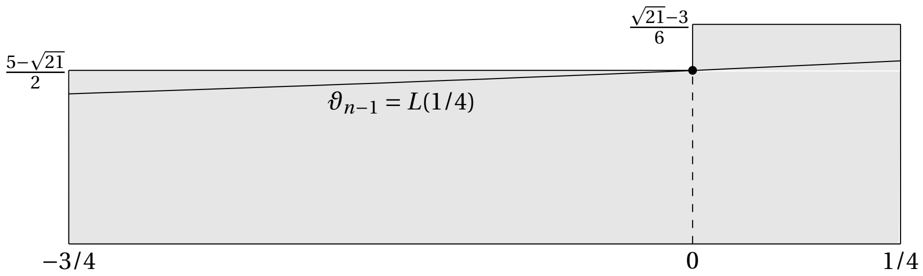


Figure 3.19: The bottom part of $\Omega_{1/4}$.

Since (3.18) yields

$$L(\tfrac{1}{r}) = \frac{r+1-\sqrt{(r+1)^2-4}}{2} = \frac{1}{r+1-L(\frac{1}{r})}, \qquad (3.19)$$

we suspect that in fact (if $1/(r+1) < \alpha < 1/r$)

$$\frac{1}{r+2} < L(\alpha) < \frac{1}{r+1-L(\frac{1}{r})}.$$

REFERENCES

[1] W. Bosma, *Optimal continued fractions*, Indag. Math. 49 (1987) 353-379.

[2] D. Barbolosi and H. Jager, *On a theorem of Legendre in the theory of continued fractions*, Journal de Théorie des Nombres de Bordeaux, tome 6, No. 1 (1994), 81 – 94.

[3] W. Bosma, H. Jager and F. Wiedijk, *Some metrical observations on the approximation by continued fractions*, Nederl. Akad. Wetensch. Indag. Math. 45 (1983), no. 3, 281–299.

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

[5] Sh. Ito, *On Legendre's theorem related to Diophantine approximation*, Séminaire de Théorie des Nombres, 1987 – 1988 (Talence, 1987 – 1988), pages Exp. No.44, 19 Univ. Bordeaux I, Talence, 1988.

3

[6] C. Kraaikamp, *A new class of continued fraction expansions*, Acta Arith. 57 (1991), no. 1, 1–39.

[7] C. Kraaikamp, T.A. Schmidt and I. Smeets, *Quilting natural extensions for α-Rosen Fractions*, Journal of the Mathematical Society of Japan Vol. 62 (2010) No. 2 P 649-671.

[8] L. Luzzi and S. Marmi, *On the entropy of Japanese continued fractions*, Discrete Contin. Dyn. Syst. **20** (2008), no. 3, 673–711.

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

[10] H. Nakada, *On the Lenstra constant associated to the Rosen continued fractions*, J. Eur. Math. Soc. 12 (2010), 55 – 70.

[11] H. Nakada and R. Natsui, *The non-monotonicity of the entropy of α-continued fraction transformations*, Nonlinearity **21** (2008), no. 6, 1207–1225.

[12] R. Natsui, *On the Legendre constant of α-continued fractions*, Journal of Number Theory 131 (2011), 487-507.

[13] G. Tiozzo, *Entropy, dimension and combinatorial moduli for one-dimensional dynamical systems.* Thesis (Ph.D.), Harvard University (2013).

[14] D.B. Zagier, *Zetafunktionen und quadratische Körper*, Springer-Verlag, Berlin Heidelberg, 1981.

4

ORBITS OF N-EXPANSIONS

4.1. INTRODUCTION

In various fields of mathematics transformations of the unit interval onto itself have been studied. When it comes to measure theory and ergodic theory, the work of Rényi on the β-transformation $T\omega = \beta\omega$ (mod one), $\beta > 1$, in the fifties of the twentieth century, preceded and inspired many others, such as Rohlin, Parry and Shiokawa. Then, in the seventies, Keith M. Wilkinson introduced a class of piecewise linear transformations of the unit interval onto itself, all having an invariant measure equivalent to the Lebesgue measure and being weak Bernoulli; see [14]. A few years later Marion R. Palmer studied the linear mod one transformations $T_{\beta\alpha}(x) = \beta x + \alpha$ (mod 1) of the unit interval I for $1 < \beta < 2$ and $0 \leq \alpha < 1$. In the same period, a slightly more applied approach was taken by Ito, Nakada and Tanaka in their study of unimodular linear transformations with respect to chaotic phenomena as observed in nature. In the mid-eighties Saito took the study of one-dimensional, piecewise linear mappings to the field of electrical and electronic engineering.

Whereas the work mentioned so far allows for the existence of gaps, there is no mixing. In this paper we will study N-expansions associated with a transformation of $[\alpha, \alpha + 1] \to [\alpha, \alpha + 1)$, with α real and positive, allowing for gaps but also being mixing and as such connected with more recent work of, for instance, Haas and Molnar in 2003 or Ma and Nair in 2017. In this paper we will study only the 'basic properties' of N-expansions. More specifically, we will focus on conditions for gaplessness.

The fact that this paper is almost entirely dedicated to gaplessness is indicative for the complexity of the subject.The interdependency of the variables α, N and the number of digits makes it very hard to prove general statements, while the occurrence of gaps is subject to the smallest of differences in these variables, which may be overlooked in computer simulations such as Figure 4.5 on page 85.

In this section we will introduce the basic definitions and concepts of N-expansions and gaps. We will finish it with Theorem 13, on a remarkable relationship between N, α and the largest digit $d(\alpha)$ under certain conditions. In Section 4.2 (cf. Theorems 18 and 19 in Subsection 4.2.2) we will prove that when $N \geq 8$, having at least six digits is a sufficient condition for gaplessness of I_α. We also present sufficient conditions for gaplessness in the case I_α has only two or three digits (cf. Theorems 20 in Subsection 4.2.3 and 27 in Subsection 4.2.4); later, in Section 4.4, we will prove that these are actually sufficient (cf. Remark 12). In Section 4.3 we will show that when I_α has four digits gaps exist only in rare cases (cf. Theorem 31) and that for $N \geq 2$ no gaps exist when I_α consists of five digits or more (cf. Theorem 34). In Section 4.4 we will give a preview on gaps in two- and three-cylinder cases. In a subsequent paper we will give an inventory of these cases. Now we will start with the introduction of N-expansions.

4

For $N \in \mathbb{N}_{\geq 2}$ and $\alpha \in \mathbb{R}$ such that $0 < \alpha \leq \sqrt{N} - 1$, let $I_\alpha := [\alpha, \alpha+1]$ and $I_\alpha^- := [\alpha, \alpha+1)$. We define the N-expansion map $T_\alpha : I_\alpha \to I_\alpha^-$ as

$$T_\alpha(x) := \frac{N}{x} - d(x), \tag{4.1}$$

where $d : I_\alpha \to \mathbb{N}$ is defined by

$$d(x) := \left\lfloor \frac{N}{x} - \alpha \right\rfloor, \quad \text{if either } x \in (\alpha, \alpha+1] \text{ or both } x = \alpha \text{ and } N/\alpha - \alpha \notin \mathbb{Z}$$

and [1]

$$d(\alpha) = \left\lfloor \frac{N}{\alpha} - \alpha \right\rfloor - 1, \quad \text{if } N/\alpha - \alpha \in \mathbb{Z}.$$

For a fixed $\alpha \in (0, \sqrt{N} - 1]$ and $x \in I_\alpha$ we define for $n \in \mathbb{N}$

$$d_n = d_n(x) := d(T_\alpha^{n-1}(x)).$$

Note that for $\alpha \in (0, \sqrt{N} - 1]$ fixed, there are only *finitely* many possibilities for each d_n.

Applying (4.1), we obtain for every $x \in I_\alpha$ a continued fraction expansion of the form

$$x = T_\alpha^0(x) = \frac{N}{d_1 + T_\alpha(x)} = \frac{N}{d_1 + \frac{N}{d_2 + T_\alpha^2(x)}} = \cdots = \frac{N}{d_1 + \frac{N}{d_2 + \frac{N}{d_3 + \ddots}}}, \tag{4.2}$$

which we will throughout this paper write as $x = [d_1, d_2, d_3, \ldots]_{N,\alpha}$ (note that this expansion is infinite for every $x \in I_\alpha$, since $0 \notin I_\alpha$); we will call the numbers d_i, $i \geq 1$, the *partial quotients* or *digits* of this *N-continued fraction expansion* of x; see [2] and [5], where these continued fractions (also with a finite set of digits) were introduced and elementary properties were studied (such as the convergence in [2]).

[1] Note that if $N/\alpha - \alpha \in \mathbb{Z}$, we have that $T_\alpha(\alpha) = \alpha + 1$.

Remark 1. While the limited set of digits of N-expansions has its attractions, there is an obvious weak spot: fast approximations are out of the question. The N-continued fraction expansions may even seem nothing but curious, when we approximate 7 by $[7,6,6,7,6,6,6,6,7,6,5]_{100,6.9} = 7.0144\cdots$, for instance. Not surprisingly, our focus is not on approximating properties, but on N-expansions as a source of interesting dynamical systems.

The map T_α obviously has one *fixed point* f_i in each *cylinder set* $\Delta_i := \{x \in I_\alpha; d(x) = i\}$ of rank 1, with $d_{\min} \le i \le d_{\max}$, where $d_{max} := d(\alpha)$ is the largest partial quotient, and $d_{min} := d(\alpha+1)$ the smallest one given N and α. In the sequel we will write simply 'cylinder set' for 'cylinder set of rank 1'.

It is easy to see that

$$f_i = \frac{\sqrt{4N+i^2}-i}{2}, \text{ for } d_{\min} \le i \le d_{\max}. \tag{4.3}$$

Note that $N/\alpha - \alpha \in \mathbb{Z}$ if and only if for some $d \ge 2$ we have that $\alpha = f_{d+1}$. In that case we have that $d(\alpha) = d$.

Given $N \in \mathbb{N}_{\ge 2}$, we let $\alpha_{max} = \sqrt{N} - 1$ be the largest value of α we consider. The reason for this is that for larger values of α we would have 0 as a partial quotient as well. Since $T'_\alpha(x) = -N/x^2$ and because $0 < \alpha \le \sqrt{N} - 1$, we have $|T'_\alpha(x)| > 1$ on I^-_α. From this it follows that the fixed points act as *repellers* and that the maps T_α are *expanding*.

Each pair of consecutive cylinders sets (Δ_i, Δ_{i-1}) is divided by a *discontinuity point* $p_i(N,\alpha)$ of[2] T_α, satisfying $N/p_i - i = \alpha$, so $p_i = N/(\alpha + i)$. A cylinder set Δ_i is called *full* if $T_\alpha(\Delta_i) = I^-_\alpha$. When a cylinder set is not full, it contains either α (in which case $T_\alpha(\alpha) < \alpha+1$) or $\alpha+1$ (in which case $T_\alpha(\alpha+1) > \alpha$), and is called *incomplete*. On account of our definition of T_α, cylinder sets will always be an interval, *never* consist of one single point.

The main object of this paper is the sequence $T^n_\alpha(x)$, $n = 0,1,2,\ldots$, for $x \in I_\alpha$, which is called the *orbit of* x *under* T_α. As an example we consider the case $N = 50, \alpha = 6$, taking $x = 6.8$. In this case the map is $T_6 : [6,7] \to [6,7)$, defined by

$$T_6(x) := \begin{cases} \dfrac{50}{x} - 1 \text{ for } x \in (\frac{25}{4}, 7]; \\ \dfrac{50}{x} - 2 \text{ for } x \in [6, \frac{25}{4}]. \end{cases} \tag{4.4}$$

We have $I_6 = \Delta_2 \cup \Delta_1$, where the cylinder sets Δ_2 and Δ_1 are separated by the discontinuity point $p_2 = 25/4$. The fixed points are $f_1 = (\sqrt{201} - 1)/2 = 6.5887\cdots$ and $f_2 = \sqrt{51} - 1 = 6.1414\cdots$; see Figure 4.1. In Figure 4.2 we have drawn not only $x(= 6.8)$ but also $T_6(x)$ and $T^2_6(x)$, illustrating the repelling character of f_1, that causes the orbit to move to Δ_2 at some point; see Figure 4.3, where we have added eight more points of the orbit of $x = 6.8$, writing x_i as an abbreviation of $T^i_6(x)$; note the position of x_9 and x_{10}, on the outermost parts of Δ_1.

[2]We will usually omit the suffix '(N,α)'.

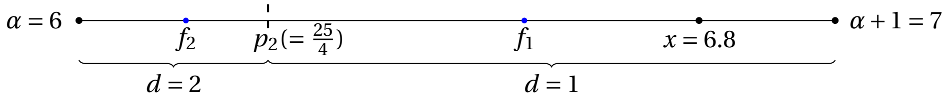


Figure 4.1: $N = 50$, $\alpha = 6$, $x = 6.8$

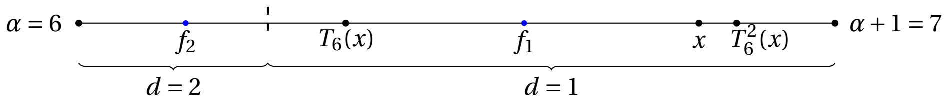


Figure 4.2: $N = 50$, $\alpha = 6$, $x = 6.8$

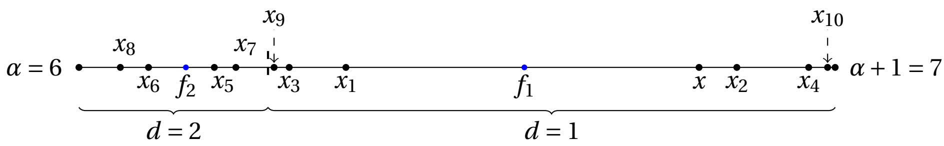


Figure 4.3: $N = 50$, $\alpha = 6$, $x = 6.8$

Figure 4.4 gives a more general impression of the orbits under T_6. We see that, due to its expansiveness, cylinder set Δ_2 is mapped onto Δ_2 plus a small subinterval $L_1 = (p_2, T_6(6)]$ on the left in Δ_1. This subinterval, which for the present values of N and α does not contain f_1, is sent under T_6 to a second interval $R_1 = T_6(L_1) = [T_6^2(6), 7)$ on the right in Δ_1 (also not containing f_1). The interval R_1 in turn is mapped to the interval $T_6(R_1) = (T_6(7), T_6^3(6)]$, which contains p_2, since for the present values of N and α we have $p_2 < T_6^3(6)$. Now let $R_2 = T_6(R_1) \cap \Delta_2 = (T_6(7), p_2]$. Then $f_2 \notin R_2$ for our present values of N and α (where $f_2 = 6.1414\cdots$ and $T_6(7) = 6.1428\cdots$). Finally, set $L_2 = T_6(R_2)$.

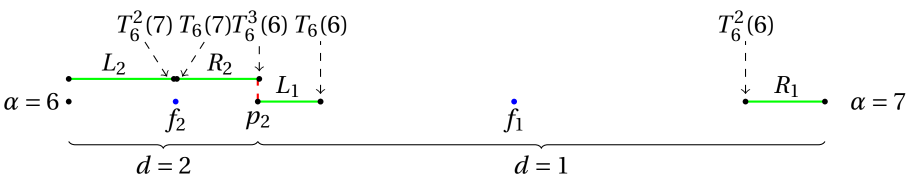


Figure 4.4: $N = 50$, $\alpha = 6$

The intervals L_1, L_2, R_1, R_2 determine where the orbit of any point x different from the fixed points or their pre-images will eventually exist. Any such orbit will either start in Δ_2 or will land there coming from Δ_1, due to the repelling action of f_1. After finitely many times it will be impossible to land in either $G_1 = \Delta_1 \setminus (L_1 \cup R_1) = (T_6(6), T_6^2(6))$ or $G_2 = \Delta_2 \setminus (L_2 \cup R_2) = (T_6^2(7), T_6(7))$.

In a moment we will define the intervals G_2 around f_2 and G_1 around f_1 as "gaps". Clearly such gaps will never be empty; for every $n \in \mathbb{N}$ there exist points x either close to f_1 or f_2 for which $T_\alpha^n(x)$ will still be in the gaps around f_1 or f_2. However, there is an $m > n$ for which $T_\alpha^m(x)$ exits for the first time the cylinder Δ_i that contains x. Then for all $k \geq m$ we have that $T_\alpha^k(x)$ will **never** return to the "gap" around f_i (and also will stay outside the gap around the other fixed point). Moreover, the boundaries of these two gaps are given by the first orbit points of the endpoints of I_α.

In [5] computer simulations were used to get a general impression of the orbits of N-expansions. For a lot of values of α, with $0 < \alpha \leq \alpha_{\max}$, the orbits of many points were calculated, yielding intervals I_α containing 'gaps' in the sense explained above. By stacking these intervals, for $0 < \alpha \leq \alpha_{\max}$, figures such as Figure 4.5 were obtained, with the values of α on the vertical axis and at each height the corresponding interval I_α drawn. In the same paper plots (such as Figure 4.5) are given in which gaps are visible for $N = 9, 20, 36$ and 100. Since the plots in [5] are based on computer simulations, they do not actually show very small gaps (smaller than pixel size) nor clarify much the connection between the gaps for each N. Still, the suggestion is strong that for α sufficiently small there are no gaps. The calculated orbits are all sufficiently long to suggest that indeed, for α large enough, gaps occur in I_α; apparently, the orbits (that can start anywhere on I_α) will eventually not return to certain intervals of I_α. What is more, these gaps may be quite large, even larger than half of the interval. It also seems that for α large enough several disjoint gaps may occur. In Figure 4.5 we see this for α near $\alpha_{\max} = \sqrt{50} - 1$.

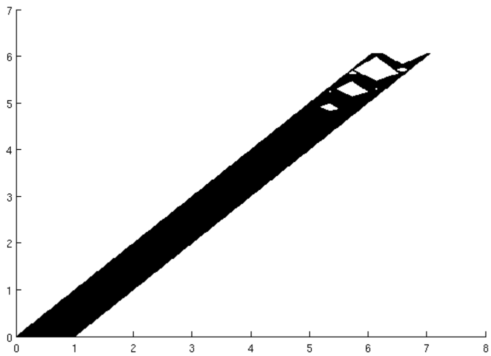


Figure 4.5: A simulation of intervals I_α with gaps if existent, for $0 < \alpha \leq \sqrt{50} - 1$ and $N = 50$

In order to get a better understanding of the orbits of N-expansions, it is useful to consider the graphs of T_α, which are drawn in the square $\Upsilon_{N,\alpha} := I_\alpha \times I_\alpha^-$. This square is divided in rectangular sets of points $\Delta_i := \{(x, y) \in \Upsilon_\alpha : d(x) = i\}$. By reusing this notation, we identify these two-dimensional *fundamental regions* Δ_i with the one-dimensional cylinder sets we already use. We will call these regions shortly *cylinders*. It is obvious that T_α has one fixed point $F_i := (f_i, f_i)$ in each Δ_d. We will denote the dividing line between Δ_i and Δ_{i-1} by l_i, which is the set $\{p_i\} \times [\alpha, \alpha + 1)$, with p_i the discontinuity point between Δ_i and Δ_{i-1}. In case $T_\alpha(\Delta_i) = I_\alpha^-$, we will call the cylinder set Δ_i *full* and the branch of the graph of T_α in Δ_α *complete*; if a cylinder set is not full, we will call it and its associated branch of T_α *incomplete*. We will call the collection of Υ_α and its associated branches, fixed points and dividing lines an *arrangement* of Υ_α. When Υ_α is a union of full cylinder sets, we will call the associated arrangement also full.

Figure 4.6 is an example of such an arrangement, in which a part of the *cobweb* is drawn associated with the orbit we investigated previously. The discontinuity point $p_2 = 25/4$ is now visible as a dividing line between Δ_1 and Δ_2.

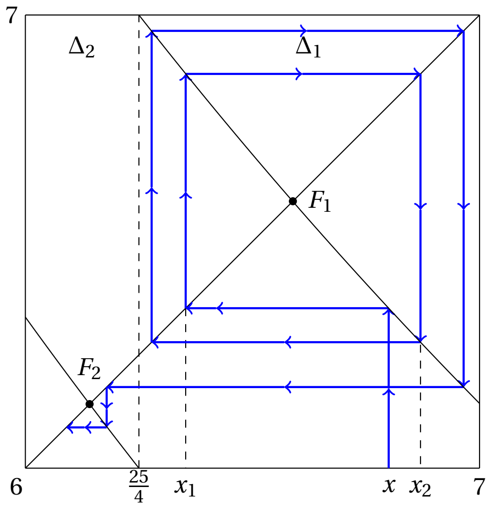


Figure 4.6: $N = 50$, $\alpha = 6$, $x = 6.8$

In [2] and [5] the arrangement for $N = 4$ and $\alpha = 1$ is studied, consisting of two full cylinders Δ_1 and Δ_2 and not showing any gaps. On the other hand, the demonstration of the interval $(5/2, 13/5)$ being a gap of $[2,3]$ in the case $(N,\alpha) = (9,2)$ in [5] is done without referring to such an arrangement. In this paper we will show that arrangements may considerably support the insight of the way gaps exist around fixed points in orbits of N-expansions, which is the main purpose of this paper. More specifically, we will study the conditions under which the course of any orbit is free, in the sense that no gaps in I_α exist. Defining a gap is slightly delicate, since $T_\alpha(I_\alpha) = I_\alpha^-$. We have the following definition:

Definition 3. An open interval $(a,b) \subset I_\alpha$ is called a *gap* of I_α[3] if for almost every[4] $x \in I_\alpha$ there is an $n_0 \in \mathbb{N}$ for which $T_\alpha^n(x) \notin (a,b)$ for all $n \geq n_0$.

When I_α consists of full cylinder sets only, we obviously have no gaps. In this situation the mutual relations between N, α and $d(\alpha)$ show a great coherence, as expressed

[3] We will usually omit the addition 'of I_α'.

[4] Here we use 'for almost all x' (and not 'for all x') because we want to exclude fixed points and pre-images of fixed points, i.e. points that are mapped under T_α to a fixed point, which may never leave an interval (a,b). All 'for all' statements in this paper are with respect to Lebesgue measure.

in the following theorem:

Theorem 13. *The interval I_α consists of m full cylinders sets, with $m \geq 2$ an integer, if and only if there is a positive integer k such that*

$$\begin{cases} \alpha = k \\ N = mk(k+1) \\ d(\alpha) = (m-1)(k+1) \end{cases} \tag{4.5}$$

Proof of Theorem 13: Writing $d := d(\alpha)$, the interval I_α is the union of m full cylinder sets if and only if

$$\begin{cases} \frac{N}{\alpha} - d = \alpha + 1 \\ \frac{N}{\alpha+1} - (d - m + 1) = \alpha \end{cases} \tag{4.6}$$

Note that the first equation in (4.6) can be written as

$$N = \alpha^2 + (d+1)\alpha,$$

while the second equation in (4.6) equals

$$N = \alpha^2 + (d+2-m)\alpha + d + 1 - m,$$

Subtracting the first of these equations from the last we find

$$(1-m)\alpha + d + 1 - m = 0,$$

yielding that

$$\alpha = \frac{d+1-m}{m-1}. \tag{4.7}$$

Since $d \geq m \geq 2$ are positive integers, we find that α is a positive rational number. We *claim* that α is a positive integer. To see this, let $\alpha = k/\ell$, where $k, \ell \in \mathbb{N}$, and $\gcd(k, \ell) = 1$. Now assume that $\ell \geq 2$. From the first equation in (4.6) we then find, that

$$\frac{N\ell}{k} = \frac{k}{\ell} + d + 1.$$

In case $k \big| N$ (e.g. when $k = 1$), we find that

$$\frac{k}{\ell} = \frac{N\ell}{k} - d - 1 \in \mathbb{Z},$$

which implies that $\ell \big| k$, which is in contradiction with the assumptions that $\ell \geq 2$ and $\gcd(k, \ell) = 1$. It follows that k is **not** a divisor of N. Since the first equation in (4.6) yields that the rational numbers

$$\frac{N\ell}{k} \quad \text{and} \quad \frac{k + (d+1)\ell}{\ell}$$

must be equal, we are led to a contradiction, because $\gcd(k, \ell) = 1$ and $\ell \geq 2$. We find that $\ell = 1$, so $\alpha = k$ is a positive integer. From the equations in (4.6) it follows immediately that both $\alpha = k$ and $\alpha+1 = k+1$ are divisors of N. In fact, applying Lemma 17 yields $m = N/(\alpha(\alpha+1))$, i.e. $N = mk(k+1)$. It immediately follows from (4.7) that $d = (m-1)(k+1)$.

Conversely, let k be a positive integer such that the relations of (4.5) hold. Then $N/\alpha - \alpha \in \mathbb{Z}$, implying that the left cylinder set is full. Moreover, Lemma 17 yields $b(N,\alpha) = mk(k+1)/(k(k+1)) = m$. It follows that the right cylinder set is full too and that I_α consists of m full cylinder sets. □

4.2. SUFFICIENT CONDITIONS FOR GAPLESSNESS

4.2.1. PRELIMINARY RESULTS

4

Before we will present our results on sufficient conditions for gaplessness, we introduce some concepts and preliminary results. The first thing to mention is that, depending on N and α, the value of the largest partial quotient $d(\alpha)$ may vary a lot:

Lemma 14. *Given N and α, let $d := d(\alpha)$ be the largest possible digit. Then*

$$d \geq N-1 \textit{ if and only if } \alpha < 1.$$

The proof of this lemma is left to the reader.

When $\alpha = \alpha_{\max} = \sqrt{N} - 1$, we have

$$\begin{cases} d = \left\lfloor \frac{2}{\sqrt{2}-1} - (\sqrt{2}-1) \right\rfloor = 4 & \text{for } N = 2; \\ d = \left\lfloor \frac{3}{\sqrt{3}-1} - (\sqrt{3}-1) \right\rfloor = 3 & \text{for } N = 3; \\ d = \left\lfloor \frac{4}{\sqrt{4}-1} - (\sqrt{4}-1) \right\rfloor - 1 = 2 & \text{for } N = 4; \\ d = \left\lfloor \frac{N}{\sqrt{N}-1} - (\sqrt{N}-1) \right\rfloor = \left\lfloor 2 + \frac{\sqrt{N}+1}{N-1} \right\rfloor = 2 & \text{for } N \geq 5. \end{cases}$$

On the other hand we have, for $N \in \mathbb{N}$, $N \geq 2$ fixed:

$$\lim_{\alpha \downarrow 0} d(\alpha) = \lim_{\alpha \downarrow 0} \left\lfloor \frac{N}{\alpha} - \alpha \right\rfloor = \infty.$$

The following lemma provides for a lower bound for the rate of increase of d compared with the rate of decrease of α.

Lemma 15. *Let $N \geq 2$ be fixed and $d := d(\alpha)$. Then d is constant for $\alpha \in [f_{d+1}, f_d)$, and d increases overall more than twice as fast as α decreases.*

Proof of Lemma 15: Starting from $\alpha_{\max}$, d increases by 1 each time α decreases beyond a fixed point, i.e. when $N/\alpha - \alpha \in \mathbb{N}$. For the difference between two successive fixed points f_{d-1} and f_d we have

$$f_{d-1} - f_d = \frac{\sqrt{4N+(d-1)^2}-(d-1)}{2} - \frac{\sqrt{4N+d^2}-d}{2} = \frac{\sqrt{4N+(d-1)^2}-\sqrt{4N+d^2}+1}{2} < \frac{1}{2}.$$

This finishes the proof. □

Closely related to the previous lemma is the following one, the proof of which is left to the reader.

Lemma 16. *Let $d \geq 2$ be fixed and let $f_d(N)$ be defined by the equation $N/f_d(N) - d = f_d(N)$ (so $f_d(N)$ is the fixed point of the map $x \mapsto N/x - d$ for $x \in (0, N/d)$). Then*

$$f_{d-1}(N+1) - f_d(N+1) > f_{d-1}(N) - f_d(N).$$

So, for d fixed, the distance between two consecutive fixed points increases when N increases. We have, in fact, for d fixed:

$$\lim_{N\to\infty} \left(f_{d-1}(N) - f_d(N)\right) = \tfrac{1}{2};$$

cf. the proof of Lemma 15. For N fixed, on the other hand, we have:

$$\lim_{d\to\infty} f_d(N) = 0.$$

While $d(\alpha)$ is a monotonously non-increasing function of α, the number of cylinder sets is not. The reason is obvious: starting from $\alpha = \alpha_{\max}$, the number of cylinder sets changes every time either α or $\alpha+1$ decreases beyond the value of a fixed point; in the first case, the number increases by 1, and in the second case, it decreases by 1. Since $T'_\alpha(x) = -N/x^2 < 0$ and $T''_\alpha(x) = 2N/x^3 > 0$ on I_α, $T_\alpha(x)$ is decreasing and convex on I_α, implying that the per saldo increase of the number of cylinder sets is larger than the decrease. Still, for N and α large enough, it may take a long time of α decreasing from $\alpha_{\max}$ before the amount of cylinder sets stops alternating between two successive numbers $k \geq 2$ and $k+1$, and starts to alternate between the numbers $k+1$ and $k+2$. As an example, we take $N = 100$. When α decreases from $\alpha_{\max}$, the interval I_α consists of two cylinder sets until α decreases beyond f_3 and cylinder set Δ_3 emerges; then, when $\alpha+1$ decreases beyond f_1, cylinder set Δ_1 disappears and so on, until α decreases beyond f_8 and Δ_9 emerges while Δ_6 has not yet disappeared.

In order to get a grip on counting the number of cylinder sets, the following arithmetic will be useful: a full cylinder set counts for 1, an incomplete left one counts for $N/\alpha - d_{\max} - \alpha$, and an incomplete right one for $\alpha + 1 - (N/(\alpha+1) - d_{\min})$, giving rise to the following definition:

Definition 4. Let $N \in \mathbb{N}_{\geq 2}$ and $\alpha \in \mathbb{R}$ such that $0 < \alpha \leq \sqrt{N} - 1$ and T_α the N-continued fraction map. The branch number[5] $b(N,\alpha)$ is defined as

$$\begin{aligned} b(N,\alpha) := {} & d_{\max} - d_{\min} - 1 \text{ (the number of full cylinder sets save for the outermost ones)} \\ & + \frac{N}{\alpha} - d_{\max} - \alpha \text{ (the length of the image of the leftmost cylinder set)} \\ & + \alpha + 1 - \left(\frac{N}{\alpha+1} - d_{\min}\right) \text{ (the length of the image of the rightmost cylinder set),} \end{aligned}$$

[5]The word ‘branch’ refers to the part of the graph of T_α on the concerning cylinder set.

From this the next lemma follows immediately:

Lemma 17. *For $N \in \mathbb{N}$, $N \geq 2$ and $0 < \alpha \leq \sqrt{N} - 1$ we have*

$$b(N,\alpha) = \frac{N}{\alpha} - \frac{N}{\alpha+1} = \frac{N}{\alpha(\alpha+1)}.$$

It follows that for fixed N the branch number $b(N,\alpha)$ is a strictly decreasing function of α.

Remark 2. Applying Lemma 17, we find

$$b(N,\alpha_{\max}) = \frac{N}{(\sqrt{N}-1)\sqrt{N}} = 1 + \frac{1}{\sqrt{N}-1}. \tag{4.8}$$

It follows that $b(N,\alpha) > 1$ for all $N \geq 2$, so the number of cylinders is always at least 2. On the other hand, from Lemma 17 it follows that the number of cylinders increases to infinity as α decreases from α_{max} to 0. Actually, we have infinitely many digits *only* when $\alpha = 0$. In this case the corresponding N-expansion is the *greedy* N-expansion, studied in [1] and [2].

The relation $N/(\alpha(\alpha+1)) = b$ yields

$$\alpha = \frac{\sqrt{\frac{4N}{b}+1}-1}{2}, \tag{4.9}$$

from which we derive that $d(\alpha) = d$ (or $d(\alpha) = d-1$ in case $N/\alpha - \alpha \in \mathbb{Z}$), where d is given by

$$d = \left\lfloor \frac{(b-1)\sqrt{\frac{4N}{b}+1}+b+1}{2} \right\rfloor. \tag{4.10}$$

4.2.2. GAPLESSNESS WHEN THE BRANCH NUMBER IS LARGE ENOUGH

So far, we merely discussed the way I_α is divided in cylinder sets, depending on the values of $N, \alpha, d(\alpha)$ and the branch number b. We are now ready to prove some first results on sufficient conditions for gaplessness.

Theorem 18. *Let $N \in \mathbb{N}$, $N \geq 3$, and let $0 < \alpha \leq \sqrt{N} - 1$. Let $|T'_\alpha(x)| > 2$ for all $x \in I_\alpha$. Then I_α contains no gaps.*

Proof of Theorem 18: In case $N = 2$, the condition $|T'_\alpha(x)| > 2$ for all $x \in I_\alpha$ is not satisfied, so we assume $N \geq 3$. The condition implies $N/(\alpha+1)^2 > 2$, yielding $\alpha < \sqrt{N/2} - 1$. From Lemma 17 it follows that

$$b(N,\alpha) > \frac{2\sqrt{2N}}{\sqrt{2N}-2},$$

which is larger than 2 for all $N \in \mathbb{N}_{\geq 3}$. So I_α consists of at least three cylinder sets. Since $|T'_\alpha(x)| > 2$ for all $x \in I_\alpha$, there exists an $\varepsilon > 0$ such that for any open interval J_0 that is contained in a cylinder set of T_α we have

$$|T_\alpha(J_0)| \geq (2+\varepsilon)|J_0|,$$

where $|J|$ denotes the length (i.e. Lebesgue measure) of an interval J.

If $T_\alpha(J_0)$ contains two consecutive discontinuity points p_{i+1}, p_i of T_α, then

$$(p_{i+1}, p_i) \subset T_\alpha(J_0),$$

and we immediately have that

$$I_\alpha^O := (\alpha, \alpha+1) = T_\alpha(p_{i+1}, p_i) \subset T_\alpha^2(J_0).$$

4

If $T_\alpha(J_0)$ contains only one discontinuity point p of T_α, then $T_\alpha(J_0)$ is the disjoint union of two subintervals located in two consecutive cylinders:

$$T_\alpha(J_0) = J'_1 \cup J'_2.$$

Obviously,

$$|T_\alpha(J_0)| = |J'_1| + |J'_2|.$$

Now select the larger of these two intervals J'_1, J'_2, and call this interval J_1. Then

$$|J_1| \geq (1+\tfrac{\varepsilon}{2})|J_0|.$$

In case $T_\alpha(J_0)$ does not contain any discontinuity point of T_α, we set $J_1 = T_\alpha(J_0)$. Induction yields that there exists an $\ell \geq 1$ such that

$$|J_\ell| \geq \left(1+\frac{\varepsilon}{2}\right)^\ell |J_0|,$$

whenever $J_{\ell-1}$ includes no more than one discontinuity point of T_α. But then there must exist an integer $k \geq 1$ such that J_k contains two (or more) consecutive discontinuity points of T_α, and we find that $T_\alpha(J_k) = I_\alpha^O$. □

The next theorem, a corollary of the previous one, gives an even more explicit condition for gaplessness.

Theorem 19. *Let $N \geq 8$ and $b(N,\alpha) \geq 4$. Then I_α contains no gaps.*

Proof of Theorem 19: When $b(N,\alpha) \geq 4$, equation (4.9) yields that $\alpha \leq (\sqrt{N+1}-1)/2$. On I_α the value of $|T'_\alpha(x)| = N/x^2$ is smallest for $x = \alpha+1 \leq (\sqrt{N+1}+1)/2$ and this minimal value equals at least $4(N+2-2\sqrt{N+1})/N$. This last expression is an increasing function of N and larger than 2 for $N \geq 9$. Applying Theorem 18 completes the proof for all $N \geq 9$. When $N = 8$, the expression is also larger than 2 on the whole interval I_α save for the endpoint $\alpha+1$, where we have $|T'_\alpha(\alpha)| = 2$. From the proof of Theorem 18 it follows that this does not interfere with the condition for gaplessness. □

When $N \geq 8$, Theorem 19 implies that arrangements with at least six cylinders are gapless. In the following we will go into conditions involving gaps for arrangements consisting of less than six cylinders. We will start with two cylinders and will use the results for arrangements with three, four and five cylinders.

4.2.3. GAPLESSNESS IN TWO-CYLINDER CASES

4

In general, when the branch number is not much larger than 1 (which is when α is not much smaller than $\alpha_{\max}$), the overall expanding power of T_α, determined by T'_α (or $|T'_\alpha|$, which we will often use), is not enough to exclude the existence of gaps; we shall elaborate this in a subsequent article; see also Section 4.4. However, in the two-cylinder case $I_\alpha = \Delta_d \cup \Delta_{d-1}$, there is a very clear condition under which this power suffices:

Theorem 20. *Let $I_\alpha = \Delta_d \cup \Delta_{d-1}$, with $d := d(\alpha)$. If $T_\alpha(\alpha) \geq f_{d-1}$* **and** *$T_\alpha(\alpha+1) \leq f_d$, then I_α is gapless.*

For the proof of Theorem 20 we need some propositions and lemmas that we will prove first. Then, immediately following Remark 6 on page 98, we will prove Theorem 20 itself.

Since arrangements under the condition of Theorem 20 play an important role in this section, we introduce the following notation:

Definition 5. Let $N \in \mathbb{N}_{\geq 4}$ be fixed. For $d \in \mathbb{N}_{\geq 2}$, we define $\mathscr{F}(d)$ as the family of all arrangements $\Upsilon_{N,\alpha}$ such that $I_\alpha = \Delta_d \cup \Delta_{d-1}$, $T_\alpha(\alpha) \geq f_{d-1}$ and $T_\alpha(\alpha+1) \leq f_d$. We will write $\mathscr{F}^*(d)$ in case we have even $T_\alpha(\alpha) = f_{d-1}$.

Remark 3. Note that for each N and d there is at most one α such that $\mathscr{F}^*(d)$ is not void.

If the expanding power of T_α is large enough to exclude the existence of gaps for *the largest* α for which an arrangement in $\mathscr{F}(d)$ exists, there will not be gaps in any arrangement in $\mathscr{F}(d)$. We will now first show how to find these largest α, which takes some effort. When we have finished that, we will go into the expanding power of $|T'_\alpha|$ in these arrangements with largest α.

Note that in case $N \in \{2,3\}$ the condition $I_\alpha = \Delta_d \cup \Delta_{d-1}$ is never satisfied. For $4 \leq N \leq 8$, with $d = 2$, the largest α is $\alpha_{\max}$, since then $T_\alpha(\alpha_{\max}) > f_1$, while $T_\alpha(\alpha_{\max}+1) = \alpha_{\max} < f_2$. For $N \geq 9$ we have $T_\alpha(\alpha_{\max}) < f_1$. When $d = 2$ we can find α such that $\Upsilon_{N,\alpha}$ in $\mathscr{F}^*(2)$ for each $9 \leq N \leq 17$; see Figure 4.7, where ten arrangements in various $\mathscr{F}(d)$ are drawn. Underneath each arrangement we have mentioned an approximation of $\sigma(\alpha) := |T'_\alpha(\alpha+1)|$, which we will later return to. This σ is important, because it is the expanding power on the rightmost cylinder that may be too weak to exclude gaps.

When $d = 2$ and $N \geq 18$, the condition $T_\alpha(\alpha) = f_{d-1}$ yields $T_\alpha(\alpha+1) > f_d$, and d has to *increase by* 1 so as to find an arrangement in $\mathscr{F}(3)$. When $d = 3$, for $18 \leq N \leq 24$ we find that the largest α is $f_{d-2} - 1$, in which case $T_\alpha(\alpha+1) = \alpha < f_d$ and $T_\alpha(\alpha) > f_{d-1}$ (so in this case the arrangement with the largest α is in $\mathscr{F}(3)$ but not in $\mathscr{F}^*(3)$); for $25 \leq N \leq 49$, the largest α is such that $T_\alpha(\alpha) = f_2$. When $N \geq 50$, the family $\mathscr{F}(3)$ is empty and d has to increase further; see Figure 4.7 once more. In the proof of Lemma 25 this approach (of exhausting $\mathscr{F}(d)$ for successive values of N and going to $\mathscr{F}(d+1)$ for larger values of N) will be formalised into a proof by induction.

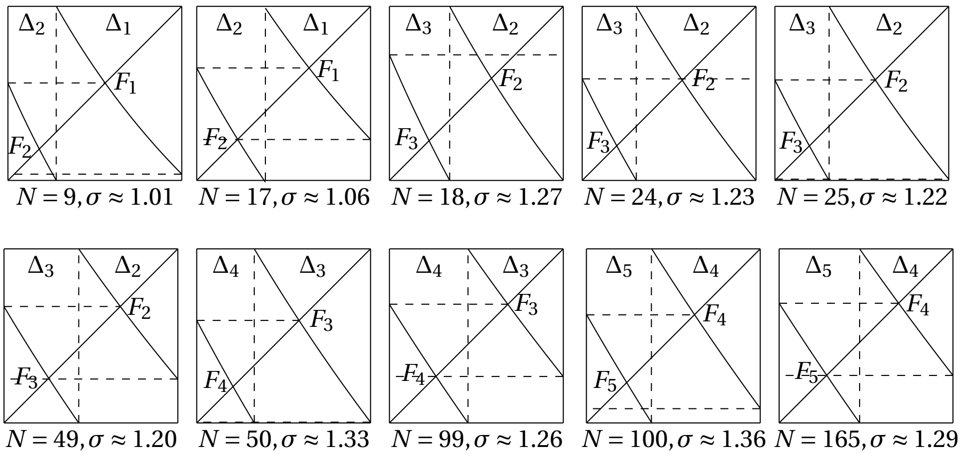


Figure 4.7: Arrangements in $\mathscr{F}(d)$, $d \in \{2,3,4,5\}$, where α is maximal

Note that this inductive approach works since for each d only *finitely* many N exist such that there are α with $\Upsilon_{N,\alpha} \in \mathscr{F}(d)$. To see why this claim holds, note that for fixed N and d, the smallest α for which $d = d(\alpha) = d_{\max}$ is α_d, given by

$$\alpha_d = f_{d+1} = \frac{\sqrt{4N + (d+1)^2} - (d+1)}{2};$$

cf. (4.3). For this α it is not necessarily so, that $I_{\alpha_d} = \Delta_d \cup \Delta_{d-1}$, i.e. that I_{α_d} consists of 2 cylinders. According to Lemma 17, the branch number $b(N, \alpha_d)$ satisfies

$$b(N, \alpha_d) = \frac{N}{f_{d+1}(f_{d+1}+1)} = \frac{4N}{4N + (d+1)^2 - 2d\sqrt{4N + (d+1)^2} + d^2 - 1}.$$

Keeping d fixed and letting $N \to \infty$, we find

$$\lim_{N\to\infty} b(N, \alpha_d) = 1.$$

In view of this and Lemma 16 (and the results mentioned directly thereafter), we choose N sufficiently large, such that for $\alpha \geq \alpha_d$ we have $b(N, \alpha) < 5/4$ and $f_{d-1} - f_d > 1/4$.

Now suppose that for such a sufficiently large value of N there exists an $\alpha \ge \alpha_d$, such that $\alpha \in \mathscr{F}(d)$. Then by Definition 4 of branch number and the assumption that $\alpha \in \mathscr{F}(d)$, we have that

$$b(N,\alpha) \ge 1 + f_{d-1} - f_d > 1\tfrac{1}{4},$$

which is *impossible* since for N sufficiently large, d fixed and $\alpha \ge \alpha_d$ we have

$$b(N,\alpha) \le b(N,\alpha_d) < 1\tfrac{1}{4}.$$

It follows that for N sufficiently large, $\mathscr{F}(d)$ is void.

We will prove (in Lemma 25) that when $N \ge 25$ there exists a minimal $d \ge 3$ such that the arrangement in $\mathscr{F}(d)$ with α maximal lies in $\mathscr{F}^*(d)$. Before we will prove this, we will explain the relation between d and N for arrangements in $\mathscr{F}^*(d)$.

In Figure 4.7 we see that for $N \in \{49, 99, 165\}$ the arrangements in $\mathscr{F}^*$ are very similar, and that the arrangement for $N = 100$ is more similar to these than the arrangement for $N = 50$. Moreover, the last three arrangements look hardly curved. The following lemmas, the proofs of which we leave to the reader, show why (cf. 4.9).

Lemma 21. *Let $b(N,\alpha) = b$ be fixed. Then*

$$|T'_\alpha(\alpha)| = \frac{b(\sqrt{4bN+b^2}+2N+b)}{2N}.$$

Lemma 21 implies that for fixed b we have that $|T'_\alpha(\alpha)|$ is a decreasing function of N, approaching b as $N \to \infty$ and that for a fixed branch number the steepest slope at $T_\alpha(\alpha)$ is given by the smallest N for which the branch number b exists.

Lemma 22. *Let $b(N,\alpha) = b$ be fixed. Then*

$$|T'_\alpha(\alpha)| - |T'_\alpha(\alpha+1)| = \frac{b\sqrt{4bN+b^2}}{N}.$$

Since $(b\sqrt{4bN+b^2})/N$ is a decreasing function of N, approaching 0 from above as $N \to \infty$, Lemma 22 implies that for a fixed branch number b the branches become less curved as N increases; i.e., the curves approach linearity as $N \to \infty$ and b is fixed. Although in $\mathscr{F}^*(d)$ the branch number is not so much fixed as bounded between 1 and 2, we have a similar decrease of curviness as N increases. The arrangements for $N \in \{49, 99, 165\}$ in Figure 4.7 suggest that (assuming $T_\alpha(\alpha) = f_{d-1}$) when $N \to \infty$ (and $d \to \infty$ and $\alpha \to \infty$ accordingly), the difference $f_d - T_\alpha(\alpha+1)$ tends to 0, yielding a 'limit graph' of T_α that consists of two parallel line segments (the straightened branch curves of T_α); see Figure 4.8, obtained by translating the graph over $(-\alpha, -\alpha)$. In this situation we have both $a := T_\alpha(\alpha) \pmod{\alpha} = f_{d-1} \pmod{\alpha}$ and $T_\alpha(\alpha+1) = f_d$ (also $\pmod{\alpha}$ in Figure 4.8). We also have that $(0, a+1)$ lies on the prolonged right line segment, from which we derive that the line segments have slope $-1/a$. The line with equation $y = -x/a + a + 1$ intersects the line $y = 1$ at $(a^2, 1)$ (so the dividing line is $x = a^2$) and intersects the line

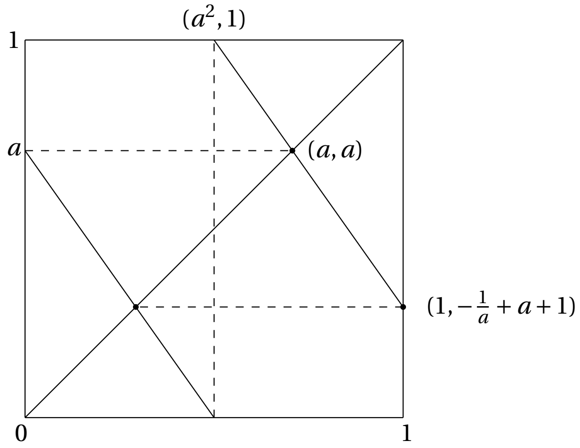


Figure 4.8: The 'limit graph' of T_α, translated over $(-\alpha, -\alpha)$, under the conditions $I_\alpha = \Delta_d \cup \Delta_{d-1}$ and $N/\alpha - d = f_{d-1}$ for $N \to \infty$ (and $\alpha, d \to \infty$ accordingly)

$x = 1$ in $(1, -1/a + a + 1)$, yielding the point $(-1/a + a + 1, -1/a + a + 1)$ on the line through $(0, a)$ with equation $y = -x/a + a$ (since $T_\alpha(\alpha + 1) = f_d$). From this we derive $2a^2 = 1$, so $a = \sqrt{1/2}$.

From Figure 4.8 we almost immediately find that the branch number for the limit case is $\sqrt{2}$ and that the dividing line is at $1/2$. We use this heuristic to find a formula describing the relation between N and d for arrangements in $\mathscr{F}^*(d)$ very precisely. Note that for arrangements similar to the limit graph we have

$$1 + f_{d-1} - f_d = \frac{\sqrt{4N + (d-1)^2} - \sqrt{4N + d^2} + 1}{2} + 1 \approx b(N, \alpha) \approx \sqrt{2},$$

from which we derive

$$N \approx (4 + 3\sqrt{2})(d^2 - d) + 2 \quad \text{or} \quad d \approx \frac{1}{2}\left(1 + \sqrt{(6\sqrt{2} - 8)(N-2) + 1}\right). \tag{4.11}$$

If, for d fixed, we determine arrangements in $\mathscr{F}^*(d)$ such that the difference $f_d - T_\alpha(\alpha + 1)$ is positive and as small as possible according to our heuristic, the best function seems to be $N = (4 + 3\sqrt{2})(d^2 - d)$, yielding the right N (after rounding off to the nearest integer) for $d \in \{3, \ldots, 500\} \setminus \{9, 50, 52, 68, 69, 80, 97, 129, 167, 185, 210, 231, 289, 330, 416, 440, 444, 479, 485\}$, in all of which cases the rounding off should have been up in stead of down. For $d = 2$ we find $N = \lceil 2(4 + 3\sqrt{2}) \rceil = 17$, for $d = 3$ we find $N = \lfloor 6(4 + 3\sqrt{2}) \rfloor = 49$, for $d = 4$ we find $N = \lceil 12(4 + 3\sqrt{2}) \rceil = 99$ and or $d = 5$ we find $N = \lceil 20(4 + 3\sqrt{2}) \rceil = 165$; see Figure 4.7 once more.

Although we do not know generally when rounding off to the nearest integer yields the right N, with (4.11) we can find a very good overall indication of the relation between

d and N for arrangements in $\mathscr{F}^*(d)$ by looking at the difference between the image of $\alpha(N)+1$ and f_d. Here $\alpha(N)$ is defined as the (positive) root of $N/\alpha - d = f_{d-1}$. With some straightforward calculations[6] we find that $N/\alpha - d = f_{d-1}$ yields

$$\alpha(N) = \frac{N\left(\sqrt{4N+(d-1)^2}-(d+1)\right)}{2(N-d)} \tag{4.12}$$

Applying (4.12), we write $f_d - (N/(\alpha(N)+1)-(d-1))$ as

$$j_d(N) := \frac{(N^2+dN+d)\sqrt{4N+d^2} - N^2\sqrt{4N+(d-1)^2} - (N^2-d(d-4)N-d(d-2))}{2(N^2+dN+d)}$$

and, more generally, define

$$j_d(x) := \frac{(x^2+dx+d)\sqrt{4x+d^2} - x^2\sqrt{4x+(d-1)^2} - (x^2-d(d-4)x-d(d-2))}{2(x^2+dx+d)} \tag{4.13}$$

for $x \in [25,\infty)$.

With this we can prove the following lemma, which is illustrated by the arrangements for $N \in \{49, 99, 165\}$ in Figure 4.7.

Lemma 23. *Let $d \in \mathbb{N}_{\geq 2}$ and $j_d(x)$ be defined as in (4.13). Then the largest N for which there are α such that $I_\alpha = \Delta_d \cup \Delta_{d-1}$, $T_\alpha(\alpha) = f_{d-1}$ and $T_\alpha(\alpha+1) \leq f_d$ is either $\lfloor (4+3\sqrt{2})(d^2-d) \rfloor$ or $\lceil (4+3\sqrt{2})(d^2-d) \rceil$.*

Proof of Lemma 23: We remark that we can confine ourselves to showing that

$$\begin{cases} j_d((4+3\sqrt{2})(d^2-d)) > 0; \\ j_d((4+3\sqrt{2})(d^2-d)+1) < 0, \end{cases} \tag{4.14}$$

since the first equation implies that $j_d(\lfloor (4+3\sqrt{2})(d^2-d) \rfloor) > 0$, while the second implies that $j_d(\lceil (4+3\sqrt{2})(d^2-d) \rceil + 1) < 0$. The work to be done is straightforward and is left to the reader. □

Before we will show that for $N \geq 25$ there are a $d \geq 3$ and an α such that $\Upsilon_{N,\alpha} \in \mathscr{F}^*(d)$, we note that $I_\alpha = \Delta_d \cup \Delta_{d-1}$ implies $N/(\alpha+1)-(d-1) \geq \alpha$. In case $\alpha = \alpha(N)$, we have

$$\frac{N}{\alpha(N)+1} - (d-1) = \frac{N^2\sqrt{4N+(d-1)^2} - \left((d-1)N^2 + 2d(d-2)N + 2d(d-1)\right)}{2(N^2+dN+d)}. \tag{4.15}$$

Applying (4.15), for the difference $h_d(N) := N/(\alpha(N)+1)-(d-1)-\alpha(N)$ we write

$$h_d(N) := \frac{2N^3 + 4dN^2 + (2d^3-5d^2+3d)N + 2d^2(d-1) - dN(2N+1)\sqrt{4N+(d-1)^2}}{2(N-d)(N^2+dN+d)}, \tag{4.16}$$

With this, we can prove the following lemma:

[6]We have throughout this paper frequently used *(Wolfram) Mathematica* for making intricate calculations, all of which are nonetheless algebraically basic. We leave it to the reader to guess where Mathematica may have been used.

Lemma 24. *Let $d \in \mathbb{N}_{\geq 2}$. Let $N \in \mathbb{N}_{\geq 25}$ be such that $I_{\alpha(M)} = \Delta_d \cup \Delta_{d-1}$ and $T_{\alpha(M)}(\alpha(M)) = f_{d-1}$ for $M \in \{N, N+1\}$. Then*

$$T_{\alpha(N+1)}(\alpha(N+1)+1) - \alpha(N+1) > T_{\alpha(N)}(\alpha(N)+1) - \alpha(N),$$

i.e. $h_d(N+1) > h_d(N)$.

Proof of Lemma 24: We want to show that $h_d(N)$ from (4.16) is an increasing sequence, and do so by calculating the derivative of

$$h_d(x) := \frac{2x^3 + 4dx^2 + (2d^3 - 5d^2 + 3d)x + 2d^2(d-1) - dx(2x+1)\sqrt{4x + (d-1)^2}}{2(x-d)(x^2 + dx + d)}, \quad (4.17)$$

with $x \in [25, \infty)$, and then showing that $h_d'(x) > 0$ on $[25, \infty)$. Although a little bit intricate, the work is straightforward and and is left to the reader. □

4

Now we can prove the following lemma:

Lemma 25. *Let $N \in \{9, \ldots, 17, 25, 26, \ldots\}$. Then there are $d \in \mathbb{N}_{\geq 2}$ and $\alpha \in (0, \sqrt{N} - 1)$ such that $I_\alpha = \Delta_d \cup \Delta_{d-1}$, $T_\alpha(\alpha) = f_{d-1}$ and $T_\alpha(\alpha+1) \leq f_d$.*

Proof of Lemma 25: We will use induction on d. For $N \in \{9, \ldots, 17, 25, 26, \ldots, 99\}$ and $d \in \{2, 3, 4\}$ we refer to Figure 4.7 and leave the calculations to the reader. Specifically, when $d = 4$, we have for $50 \leq N \leq 99$ that there is an α such that $\Upsilon_{N,\alpha} \in \mathscr{F}^*(d)$. It is easily seen that there is also an α_{99} such that $\Upsilon_{99,\alpha_{99}} \in \mathscr{F}^*(5)$. Due to Lemma 23, there is an $N_5 > 99$ for which there is an α_{N_5} such that $\Upsilon_{N_5,\alpha_{N_5}} \in \mathscr{F}^*(5)$. Applying Lemma 4.15, we see that for all $N \in \{99, \ldots, N_5\}$ there are α such that $\Upsilon_{N,\alpha} \in \mathscr{F}^*(5)$. For the induction step, let $d \geq 5$ be such that there is an α for which $\Upsilon_{N_d,\alpha} \in \mathscr{F}^*(d)$, where N_d is the largest such N possible, regarding Lemma 23. If we can show that for this N_d there is an α' such that $\Upsilon_{N_d,\alpha'} \in \mathscr{F}^*(d+1)$, we are finished. This can be done by showing that

$$h_{d+1}((4+3\sqrt{2})(d^2-d)-1) > 0, \quad (4.18)$$

for this implies $h_{d+1}(N_d) > 0$, in which case α' is such that $N_d/\alpha' - (d+1) = f_d$. Although intricate, the calculations are straightforward and are left to the reader. □

Remark 4. Although Lemma 25 is about N in the first place, our approach is actually based on increasing d and then determining all N such that arrangements $\Upsilon_{N,\alpha} \in \mathscr{F}^*(d)$ exist. The proof of Lemma 25 yields the arrangements with the *smallest* d (and therefore the largest α) for which $\Upsilon_{N,\alpha} \in \mathscr{F}^*(d)$, as illustrated by the last five arrangements of Figure 4.7.

Example 1. For $d = 4$ we have $N = \lfloor (4+3\sqrt{2})(d^2-d) \rfloor = 98$. Then $\Upsilon_{N,\alpha(N_0)} \in \mathscr{F}^*(d)$ and $\Upsilon_{N+1,\alpha(N+1)} \in \mathscr{F}^*(d)$, while $\Upsilon_{N_0+2,\alpha(N+2)} \in \mathscr{F}^*(d+1)$; see Figure 4.7. It follows immediately from our construction of $\alpha(N+2)$ that this is the largest α such that $\Upsilon_{N+2,\alpha} \in \mathscr{F}(d+1)$.

With manual calculations we can quickly calculate the expanding power of T_α in $\alpha+1$ for arrangements in $\mathscr{F}$ and $\mathscr{F}^*$ and N not too large, say $N \le 49$, where the smallest values are found where α is as large as possible. The next proposition gives a lower bound for $|T'_\alpha(\alpha+1)|$ for such arrangements for most N.

Proposition 26. *Let $N \in \{18\} \cup \{50, 51, \ldots\} \setminus \{95, \ldots, 99\}$ and $\alpha \in (0, \sqrt{N}-1]$ such that $I_\alpha = \Delta_d \cup \Delta_{d-1}$ for some $d \in \mathbb{N}$, $d \ge 2$. Furthermore, suppose that $T_\alpha(\alpha) \ge f_{d-1}$ and $T_\alpha(\alpha+1) \le f_d$. Then $|T'_\alpha(\alpha+1)| > \sqrt[3]{2}$.*

Proof of Proposition 26: Considering Lemma 25, we can confine ourselves to arrangements in $\mathscr{F}^*$ with α as large as possible. For $\alpha = \alpha(N)$ (cf. (4.12)) we can write $|T'_\alpha(\alpha+1)| = N/(\alpha+1)^2$ as

4

$$k_d(N) = \frac{2N^4 + (d-1)^2N^3 + 2d(d-1)N^2 + 2d^2N + ((d-1)N^3 + 2dN^2)\sqrt{4N + (d-1)^2}}{2(N^4 + 2dN^3 + d(d+2)N^2 + 2d^2N + d^2)}. \tag{4.19}$$

It is not hard to find that, for d fixed, k_d is a decreasing sequence, with $\lim_{N\to\infty} k_d(N) = 1$; we leave this to the reader. However, from (4.11) it follows that if $N \to \infty$ we have that also $d \to \infty$ in a precise manner. Due to the previous lemmas, for each d we can confine ourselves to considering only $N/(\alpha+1)^2$ for the largest N and α such that $\Upsilon_{N,\alpha} \in \mathscr{F}^*(d)$. Applying Lemma 23, an easy way to check if indeed $|T'_{\alpha(N)}(\alpha(N)+1)| > \sqrt[3]{2}$ is considering $k_d(x)$, with $x \in [100, \infty)$, and then calculating $k_d((4+3\sqrt{2})(d^2-d)+1)$ for $d \ge 5$, which is amply larger than $\sqrt[3]{2} = 1.2599\cdots$. For $d = 3$ and $N = 18$ and for $d = 4$ and $N \in \{50, 51, \ldots, 94\}$ it is easily checked manually that indeed $|T'_{\alpha(N)}(\alpha(N)+1)| > \sqrt[3]{2}$. □

Remark 5. Considering our previous remarks concerning arrangements in $\mathscr{F}^*$, it may be clear that $\lim_{N\to\infty} N/(\alpha(N)+1)^2 = \sqrt{2}$.

Remark 6. For our purposes the value $\sqrt[3]{2}$ in Proposition 26 can be replaced by 1.26, the third power of which is 2.000376. The choice for $\sqrt[3]{2}$ relates to the proof of Theorem 20 and also to the proofs of Proposition 28 and Theorem 18, where the numbers $\sqrt{2}$ and 2 have a similar importance. In fact, we could replace $\sqrt[3]{2}$ by

$$\frac{94}{(\alpha(94)+1)^2} = \frac{20480015 + 320305\sqrt{385}}{21233664} = 1.2604\cdots.$$

Finally we are ready to prove Theorem 20, stating that $I_\alpha = \Delta_d \cup \Delta_{d-1}$, with $d := d(\alpha)$, is gapless if $T_\alpha(\alpha) \ge f_{d-1}$ **and** $T_\alpha(\alpha+1) \le f_d$.

Proof of Theorem 20: First we note that the conditions imply $N \in \mathbb{N}_{\ge 4}$. Now let $\Upsilon_{N,\alpha} \in \mathscr{F}(d)$ and let $K \subset I_\alpha$ be any open interval. Since K expands under T_α, there is an $n \ge 0$ such that $T_\alpha^n(K)$ contains for the first time a fixed point or the discontinuity point p_d, in the former case of which we are finished. So we assume that $T_\alpha^n(K) \cap \Delta_d = (b, p_d] =: L$, with $f_d < b < p_d$. Note that $T_\alpha(L) = [\alpha, T_\alpha(b)) \subset [\alpha, f_d)$. Similarly, we may assume that $T_\alpha^2(L) = (T_\alpha^2(b), T_\alpha(\alpha)]$, with $f_{d-1} < T_\alpha^2(b) < \alpha+1$ (since otherwise $f_{d-1} \in T_\alpha^2(L)$, and again we are done).

Now suppose that $T_\alpha^3(L)$ contains p_d, excluding $f_d \in T_\alpha^3(L)$. Then $T_\alpha^3(L) = L_1 \cup M_1$, with $L_1 = [T_\alpha^2(\alpha), p_d]$ and $M_1 = (p_d, T_\alpha^3(b))$. First we confine ourselves to $N \in \{18\} \cup \{50,\ldots\} \setminus \{95,\ldots,99\}$. Since then $|T_\alpha^3(L)| > 2.000376|L|$ (cf. Remark 6), we have certainly $|L_1| > 1.001|L|$ or $|M_1| > 1.001|L|$. If we consider the images of L_1 and M_1 under T_α, T_α^2 and T_α^3 similarly as we did with the images of L, we find that due to expansiveness (see the proof of Theorem 18) there *must* be an m such that $f_d \in T_\alpha^{3m}(L_1)$ or $f_{d-1} \in T_\alpha^{3m}(M_1)$ and we are finished. If $T_\alpha^3(L)$ does *not* contain p_d, the expansion of L will only go on longer, yielding even larger L_1' and M_1' and the reasoning would only be stronger that no gaps can exist.

For $N \in \{4,\ldots,17,19,20,\ldots,49,95,96,\ldots,99\}$ a similar approach can be taken, but there is no useful general lower bound for $|T_\alpha'(x)|$ on I_α. For these cases, however, the moderate expanding power in Δ_{d-1} is easily made up for by a relatively strong expanding power in Δ_d, and the gaplessness is easily checked (cf. Examples 2 and 3 below). This finishes the proof of Theorem 20. □

4

Example 2. In case $N=7$, there exist $\alpha \in (0,\sqrt{7}-1]$ for which $I_\alpha = \Delta_2 \cup \Delta_1$. The largest α for which $\Upsilon_{7,\alpha} \in \mathscr{F}(2)$ is $\alpha_{\max} = \sqrt{7}-1$, in which case $|T_\alpha'(\alpha+1)| = 1$. However, $|T_\alpha'(f_2)| = 2.0938\cdots > 2$, and the approach taken above works if only for the expanding power of T_α on $[\alpha, f_2)$.

Example 3. In case $N = 99$, we have $I_\alpha = \Delta_4 \cup \Delta_3$, and $\Upsilon_{99,\alpha} \in \mathscr{F}^*$ for $\alpha = 99(\sqrt{405}-5)/190 = 7.8807\cdots$. Then $|T_\alpha'(\alpha+1)| = 1.2552\cdots$, $|T_\alpha'(f_3)| = 1.3503\cdots$ and $|T_\alpha'(f_4)| = 1.4908\cdots$. So for an interval (p_4, x), with $x \in (p_d, f_3)$, assuming that $f_4 \notin T_\alpha^3(p_4,x)$, we have $|T_\alpha^3(p_4,x)| > 1.3503\cdots \times 1.2552\cdots \times 1.4908\cdots \times |(p_4,x)| > 2|(p_4,x)|$, implying enough expanding power for T_α^3 to exclude the existence of gaps.

Remark 7. We can also prove that $|T_\alpha'(x)| > \sqrt{2}$ on Δ_d for all arrangements under the assumptions of Theorem 20, but we cannot do without knowledge about the slope in Δ_{d-1}.

Next we will make preparations for formulating a sufficient condition for gaplessness in case I_α consists of three cylinder sets. Proving it involves more subtleties on the one hand, but will have a lot of similarities with the two-cylinder case on the other. Once we have finished that, not much work remains to be done for gaplessness in four- and five-cylinder cases.

4.2.4. A SUFFICIENT CONDITION FOR GAPLESSNESS WHEN I_α CONSISTS OF THREE CYLINDERS OR MORE

When $I_\alpha = \Delta_d \cup \ldots \cup \Delta_{d-m}$, there is a sufficient condition for gaplessness that resembles the condition for gaplessness in the two-cylinder case a lot:

Theorem 27. *Let $I_\alpha = \Delta_d \cup \ldots \cup \Delta_{d-m}$, with $d := d(\alpha)$ and $m \geq 2$. Then I_α is gapless if*

$$T_\alpha(\alpha) \geq f_{d-1} \textbf{ or } T_\alpha(\alpha+1) \leq f_{d-m+1}.$$

4

We will prove this theorem in parts. In the subsection named *Gaplessness in three-cylinder cases*, following shortly, we will prove Theorem 27 for $m = 2$; in the subsection named *A sufficient condition for gaplessness in case I_α consists of more than three cylinders*, on page 106, we will extend the result of Subsection *Gaplessness in three-cylinder cases* to $m \geq 2$.

GAPLESSNESS IN THREE-CYLINDER CASES

Since we have $m = 2$, the condition $T_\alpha(\alpha) \geq f_{d-1}$ can be split in

$$\begin{cases} 1.\ T_\alpha(\alpha) \geq f_{d-1} \geq T_\alpha(\alpha+1); \\ 2.\ T_\alpha(\alpha+1) \geq T_\alpha(\alpha) \geq f_{d-1}; \\ 3.\ T_\alpha(\alpha) \geq T_\alpha(\alpha+1) \geq f_{d-1}; \end{cases}$$

of course the condition $T_\alpha(\alpha+1) \leq f_{d-1}$ can be split in a similar way. We will prove Theorem 27 by proving gaplessness according to this distinction in three cases, associated with Lemma 28, 29 and 30 respectively. The first of these is not very hard to prove:

Lemma 28. *Let $I_\alpha = \Delta_d \cup \Delta_{d-1} \cup \Delta_{d-2}$, $d := d(\alpha)$. If $T_\alpha(\alpha) \geq f_{d-1}$* **and** *$T_\alpha(\alpha+1) \leq f_{d-1}$, then I_α is gapless.*

Proof of Lemma 28: The assumptions imply that $b(N,\alpha) > 2$, yielding $\sigma(\alpha) = |T'_\alpha(\alpha+1)| > \sqrt{2}$ for $N \geq 17$. If $N \geq 17$, we let $K \subset I_\alpha$ be any open interval. Since K expands under T_α, there is an $n \geq 0)$ such that $T^n_\alpha(K)$ contains a fixed point or a discontinuity point p_{d-i} (with $i \in \{0,1\}$) , in the former case of which we are finished. So we assume that $T^n_\alpha(K) \supset L$, where $L = (b, p_{d-i}]$, with $f_{d-i} < b < p_{d-i}, with \in \{0,1\}$. If $T_\alpha(L)$ contains a fixed point, we are finished. If $T_\alpha(L)$ does not contain a fixed point, then it cannot contain a discontinuity point, and we have that $|T^2_\alpha(L)| > 2|L|$, implying enough expanding power of T_α to ensure gaplessness of at least one cylinder set (which might be non-full). Since both $T_\alpha(\alpha) \geq f_{d-1}$ and $T_\alpha(\alpha+1) \leq f_{d-1}$, it follows that I_α is gapless. For $2 \leq N \leq 16$ the slopes on I_α may differ considerably: for some N, such as $N = 7$ and $N = 16$ we also have $\sigma > \sqrt{2}$, but when this is not he case, the steepness left of f_{d-2} is amply larger then $\sqrt{2}$; see Figure 4.9 for some examples where α is as large as possible. This finishes the proof of Lemma 28. □

If $I_\alpha = \Delta_d \cup \Delta_{d-1} \cup \Delta_{d-2}$ under the condition $T_\alpha(\alpha) \geq f_{d-1}$ and $T_\alpha(\alpha+1) > f_{d-1}$ or under the condition $T_\alpha(\alpha) < f_{d-1}$ and $T_\alpha(\alpha+1) \leq f_{d-1}$, I_α is gapless as well, but this is much harder to prove. The following definition will be convenient:

Definition 6. Let $I_\alpha = \Delta_d \cup \ldots \cup \Delta_{d-m}$, with $d := d(\alpha)$ and $1 \leq m \leq d-1$. If $T_\alpha(\alpha) \leq f_{d-1}$ or $T_\alpha(\alpha+1) \geq f_{d-m+1}$, the cylinder set Δ_d respectively Δ_{d-m} is called *small.*

Taking a similar approach as in the proof of Theorem 20, one can show that the map T_α has enough expansive power to ensure that for any open interval $K \subset I_\alpha$ there exists a non-negative integer n such that $T^n_\alpha(K)$ contains a fixed point. If this fixed point is in a

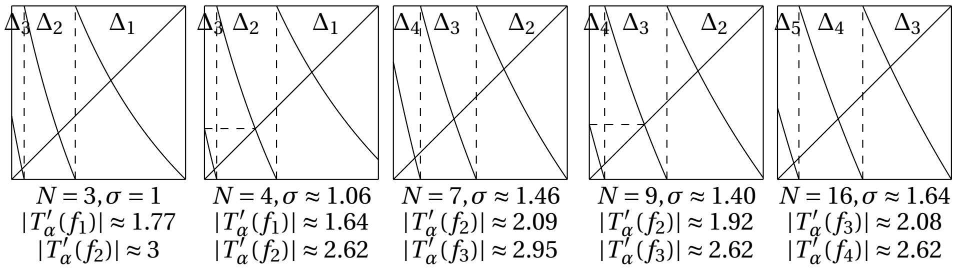


Figure 4.9: Arrangements with largest α such that there is a d with $I_\alpha = \Delta_d \cup \Delta_{d-1} \cup \Delta_{d-2}$ under the condition $T_\alpha(\alpha) \geq f_{d-1}$ *and* $T_\alpha(\alpha+1) \leq f_{d-1}$

non-small or even full cylinder, we are done (as in the proofs of Theorem 20 and Lemma 28). However, if this fixed point is from the *small* cylinder, then it only follows that every point of the small cylinder is in the orbit under T_α of some point in K. Note this implies that the *small* cylinder is *gapless*. So we assume that the small cylinder set is gapless. First we assume that the left cylinder set is small. We define $L := T_\alpha(\Delta_d) \setminus \Delta_d$. Since Δ_d is gapless, we have $L = (p_d, T(\alpha)] \subset (p_d, f_{d-1})$, so $T_\alpha(L) = [T_\alpha^2(\alpha), \alpha+1)$. If $T_\alpha^2(\alpha) \leq f_{d-2}$, we are finished, so we assume that $T_\alpha^2(\alpha) > f_{d-2}$. We then have $T_\alpha^2(L) = (T_\alpha(\alpha+1), T_\alpha^3(\alpha)]$. If $T_\alpha^3(\alpha) \geq f_{d-1}$ we are finished, since then $f_{d-1} \in T_\alpha^2(L)$. The question arises whether it is possible to keep avoiding fixed points if we go on with letting T_α work on L and its images. We will prove that this is not possible in the two most plausible cases for gaps to exist, involving the least expansion.

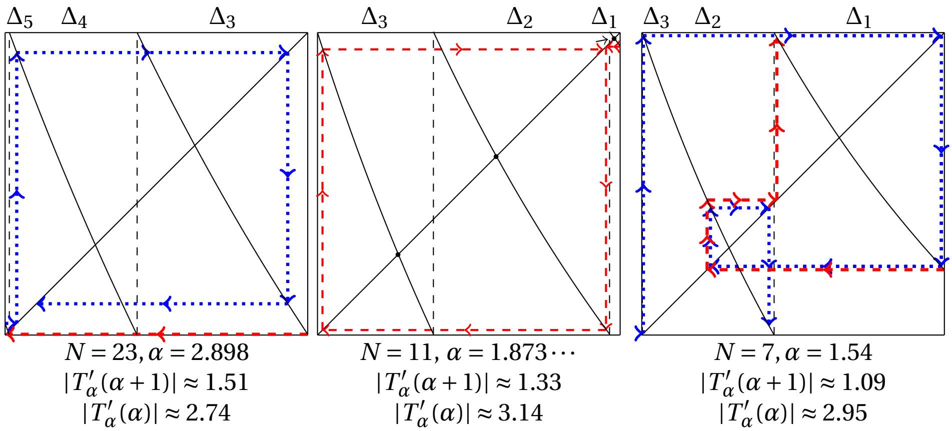


Figure 4.10: Arrangements with one very small cylinder

The first case is illustrated with two arrangements in Figure 4.10, one where $N = 23$ and one where $N = 11$. In both arrangements one outer cylinder is very small while the other one is full or almost full. In the arrangement where $N = 23$, we see that L is a very narrow strip between p_5 and $T_\alpha(\alpha)$, $T_\alpha^2(L)$ is not so narrow anymore, and $T_\alpha^3(L)$ is definitely wide enough to make clear that avoiding fixed points f_4 and f_3 is not possible. The middle arrangement, where $N = 11$, is an example of the case where Δ_{d-2} is small and Δ_d is actually full. Here we have that $R := T_\alpha(\Delta_1) \setminus \Delta_1$ is a very narrow strip between $T_\alpha(\alpha+1)$ and p_2 and that $T_\alpha^2(R)$ is only slightly larger than $T_\alpha(\Delta_1)$, whence eventually there will be an $n \in \mathbb{N}$ such that $f_3 \in T_\alpha^n(R)$ or $f_2 \in T_\alpha^n(R)$.

The rightmost arrangement in Figure 4.10 is an is an illustration of the second plausible case for the existence of gaps: here Δ_3 is small, while Δ_1 is incomplete. This arrangement illustrates the role p_d might play in avoiding fixed points: in this case, taking $L := T_\alpha(\Delta_3) \setminus \Delta_3$, we have $T_\alpha^3(L) = M_1 \cup M_2$, with $M_1 = [T_\alpha^4(\alpha), p_2]$ and $M_2 = (p_2, T_\alpha^2(\alpha+1)]$. Since $T_\alpha^3(L)$ contains a discontinuity point, the expansion under T_α is interrupted. If $T_\alpha(M_1)$ would be a subset of $T_\alpha(\Delta_3)$ and $T_\alpha(M_2)$ would be a subset of $T_\alpha(L)$, the expansion would be finished and we would have three gaps: $(T_\alpha(\alpha), T_\alpha(\alpha+1))$, $(T_\alpha^3(\alpha), T_\alpha^4(\alpha))$ and $(T_\alpha^2(\alpha+1), T_\alpha^2(\alpha))$ – but this is not the case, as we will shortly prove.

Of course arrangements exist such that one of the outer cylinder sets is small, fixed points are avoided (in the sense we used above) for a long time and it takes more of T_α working on L or R before one of the discontinuity points is captured. But in these cases the interruption of the expansion is even weaker than in the cases above. We will first show that arrangements such as the rightmost one of Figure 4.10 exclude the existence of gaps (cf. Lemma 29) and will then consider cases such as the first two arrangements of Figure 4.10 (cf. Lemma 30).

Lemma 29. *Let $I_\alpha = \Delta_d \cup \Delta_{d-1} \cup \Delta_{d-2}$, with $d := d(\alpha)$. Then I_α is gapless if*

$$f_{d-1} \le T_\alpha(\alpha) \le T_\alpha(\alpha+1) \text{ or } T_\alpha(\alpha) \le T_\alpha(\alpha+1) \le f_{d-1}.$$

Proof of Lemma 29: We will confine ourselves to the first case of this lemma, that is when $f_{d-1} \le T_\alpha(\alpha) \le T_\alpha(\alpha+1)$, since the second one is proved similarly. Regarding our observations above, we may assume that the small cylinder set is gapless (cf. the remarks after Definition 6). We will show that this implies the gaplessness of the other cylinder sets as well. We define $R := T_\alpha(\Delta_{d-2}) \setminus \Delta_{d-2}$ and try to determine α such that $p_d \in T_\alpha^3(R)$ (see the remark immediately preceding this lemma). Necessary conditions for this are $T_\alpha^2(\alpha) < p_d < T_\alpha^4(\alpha+1)$ (assuming that $f_d \notin T_\alpha(R)$ and $f_{d-1} \notin T_\alpha^2(R)$, since in either case we would be done). If these conditions are satisfied, we write $T_\alpha^3(R) = N_1 \cup N_2$, with $N_1 = [T_\alpha^2(\alpha), p_d]$ and $N_2 = (p_d, T_\alpha^4(\alpha+1)]$. We will show that we cannot have both $T_\alpha(N_1) \subset T_\alpha(R)$ and $T_\alpha(N_2) \subset T_\alpha(\Delta_{d-2})$, which is necessary for limiting the expansion of R under T_α and so not eventually capturing f_d and f_{d-1}; see Figure 4.11.

We take an approach that is similar to the proof of Theorem 20, for which several lemmas and propositions where used, partially concerning a relation between N and d in the arrangements involved, partially concerning the slope in $\alpha+1$. In this proof we will not explicitly formulate similar statements as lemmas or propositions, nor do we prove them, since they require similar basic but very intricate calculations that we prefer to omit.

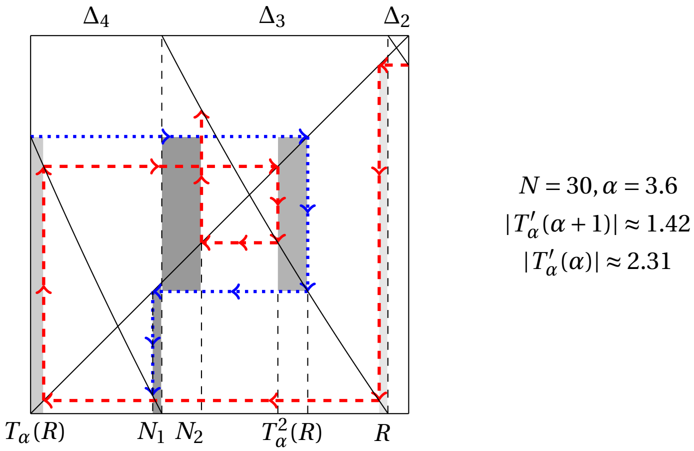


Figure 4.11: Arrangement illustrating Proposition 29

In order to find the relationship between N and d for arrangements with the conditions $T^2_\alpha(\alpha) < p_d$ and $T^4_\alpha(\alpha+1) > p_d$ mentioned above, we refer to some more relevant arrangements, as shown in Figure 4.12.

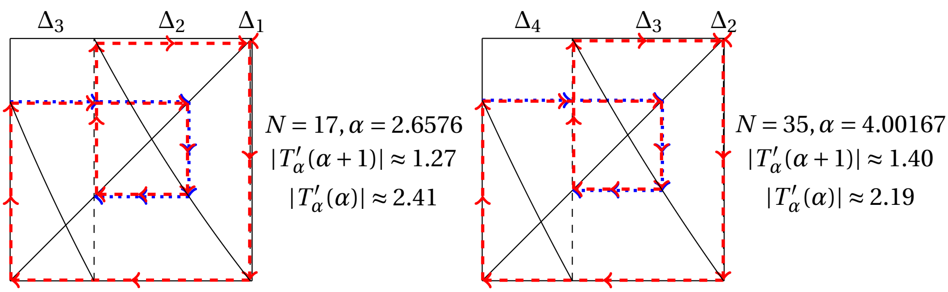


Figure 4.12: Two arrangements in which almost $T^2_\alpha(\alpha) < p_d < T^4_\alpha(\alpha+1)$

In both cases in Figure 4.12, α is such that $T^2_\alpha(\alpha) = p_d$, which is a value of α that is only a little larger than the values for which $T^2_\alpha(\alpha) < p_d$ and $T^4_\alpha(\alpha+1) > p_d$. A 'limit arrangement' (where the third, rightmost cylinder is infinitely small), similar to the 'limit arrangement' used in the proof of Theorem 20, is shown in Figure 4.13. The assumptions yield $a^3 + a^2 - 1 = 0$, with real root $a = 0.75487\cdots =: \gamma$.

Similar to the proof of Theorem 20 we then find that for arrangements as in Figure 4.12 we have

$$N \approx \frac{(d-1)(d-1+\gamma)(1+\gamma)}{\gamma^2}.$$

Using this relationship, we can take a similar approach as in the proof of Proposition

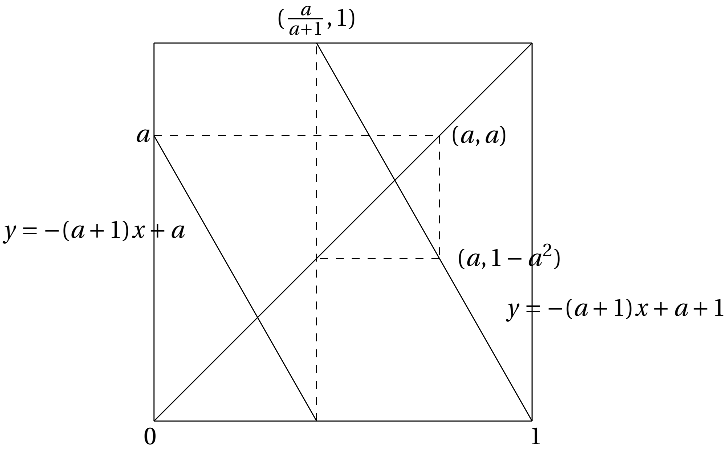


Figure 4.13: The 'limit graph' of T_α, translated over $(-\alpha, -\alpha)$, under the conditions $I_\alpha = \Delta_d \cup \Delta_{d-1}$ and $N/(N/\alpha - d) - (d-1) = p_d$ for $N \to \infty$ (and $\alpha, d \to \infty$ accordingly). This 'arrangement' can be seen as one with three cylinders, where Δ_{d-2} (mod α), the one on the right, is infinitely small; see also the arrangements in Figure 4.12.

26. We leave out the tedious steps and confine ourselves to observing that the slope of the line segments in Figure 4.13 is $-(\gamma+1) = -1.75487\cdots$ and that in arrangements where $T^2_\alpha(\alpha) < p_d$ and $T^4_\alpha(\alpha+1) > p_d$, the slope $T'_\alpha(\alpha+1)$ approaches $-(\gamma+1)$ as N tends to infinity. However, for our proof the inequality $|T'_\alpha(\alpha+1)| > 1/2(\sqrt{5}+1) = 1.61803\cdots =: G$ suffices, which holds for $N \geq 273$. We will use this to show that for $N \geq 273$ we have $|T^3_\alpha(R)| > |T_\alpha(\Delta_{d-2})| + |T_\alpha(R)|$. Since $|T'_\alpha(x)|$ is a decreasing function on I_α, and writing $\beta := |\Delta_{d-2}|$, we have

$$|T_\alpha(\Delta_{d-2})| > |T'_\alpha(\alpha+1)| \cdot \beta, \text{ so } |R| > (|T'_\alpha(\alpha+1)| - 1)\beta.$$

It follows that

$$|T_\alpha(R)| > (|T'_\alpha(\alpha+1)| - 1) \cdot |T'_\alpha(p_{d-1})|\beta,$$

that

$$|T^2_\alpha(R)| > (|T'_\alpha(\alpha+1)| - 1) \cdot |T'_\alpha(p_{d-1})| \cdot |T'(f_d)|\beta,$$

and finally that

$$|T^3_\alpha(R)| > (|T'_\alpha(\alpha+1)| - 1) \cdot |T'_\alpha(p_{d-1})|^2 \cdot |T'(f_d)|\beta.$$

We also have $|T_\alpha(\Delta_{d-2})| < |T'_\alpha(p_{d-1})|\beta$, so

$$|R| < (|T'_\alpha(p_{d-1})| - 1)\beta \text{ and } |T_\alpha(R)| < |T'_\alpha(f_{d-1})| \cdot (|T'_\alpha(p_{d-1})| - 1)\beta.$$

It follows that

$$\begin{aligned} |T_\alpha(\Delta_{d-2})| + |T_\alpha(R)| &< (|T'_\alpha(p_{d-1})| + |T'_\alpha(f_{d-1})| \cdot (|T'_\alpha(p_{d-1})| - 1))\beta \\ &= (|T'_\alpha(p_{d-1})| - |T'_\alpha(f_{d-1})| + |T'_\alpha(f_{d-1})| \cdot |T'_\alpha(p_{d-1})|)\beta \\ &< |T'_\alpha(f_{d-1})| \cdot |T'_\alpha(p_{d-1})|\beta. \end{aligned}$$

We conclude that $|T_\alpha^3(R)| > |T_\alpha(\Delta_{d-2})| + |T_\alpha(R)|$ if

$$|T'_\alpha(f_{d-1})| \cdot |T'_\alpha(p_{d-1})| < (|T'_\alpha(\alpha+1)| - 1) \cdot |T'_\alpha(p_{d-1})|^2 \cdot |T'_\alpha(f_d)|,$$

that is, if

$$1 < (|T'_\alpha(\alpha+1)| - 1) \cdot |T'_\alpha(p_{d-1})| \cdot \frac{|T'_\alpha(f_d)|}{|T'_\alpha(f_{d-1})|}. \tag{4.20}$$

Since

$$(|T'_\alpha(\alpha+1)|-1)\cdot|T'_\alpha(p_{d-1})|\cdot\frac{|T'_\alpha(f_d)|}{|T'_\alpha(f_{d-1})|} > (|T'_\alpha(\alpha+1)|-1)\cdot|T'_\alpha(p_{d-1})| > (|T'_\alpha(\alpha+1)|-1)\cdot|T'_\alpha(\alpha+1)|,$$

4

we know that (4.20) holds for $|T'_\alpha(\alpha+1)| > G$, that is for $N \geq 273$. We remark that this value is quite a wide upper bound, since we did a rough approximation. Still, checking that we cannot have both $T_\alpha(N_1) \subset T_\alpha(R)$ and $T_\alpha(N_2) \subset T_\alpha(\Delta_{d-2})$ for smaller N is not that hard and is left to the reader. This finishes the proof of Lemma 29. □

Lemma 29 implies that in case $I_\alpha = \Delta_d \cup \Delta_{d-1} \cup \Delta_{d-2}$ and $f_{d-1} \leq T_\alpha(\alpha) \leq T_\alpha(\alpha+1)$ or $T_\alpha(\alpha) \leq T_\alpha(\alpha+1) \leq f_{d-1}$ the division of an interval containing p_d in two smaller ones cannot prevent an overall expansion that excludes any gaps. The other plausible three-cylinder case in which gaps might exist is when one outer cylinder is very small, while the other one is full or almost full, such that either $T_\alpha^3(\alpha+1) \geq T_\alpha(\alpha+1)$ (when Δ_{d-2} is the small cylinder set) or $T_\alpha^3(\alpha) \leq T_\alpha(\alpha+1)$ (when Δ_d is the small cylinder set). We will show that this is not possible either:

Lemma 30. *Let $I_\alpha = \Delta_d \cup \Delta_{d-1} \cup \Delta_{d-2}$, with $d := d(\alpha)$. Then I_α is gapless if*

$$f_{d-1} \leq T_\alpha(\alpha+1) \leq T_\alpha(\alpha) \text{ or } T_\alpha(\alpha+1) \leq T_\alpha(\alpha) \leq f_{d-1}.$$

Proof of Lemma 30: Taking into account our observations immediately following Definition 6 and the arrangements of Figure 4.10 for $N = 23$ and $N = 11$, we only have to prove that there are *no* α such that $T_\alpha^3(\alpha) < T_\alpha(\alpha)$ is possible when $T_\alpha(\alpha+1) \leq T_\alpha(\alpha) \leq f_{d-1}$ or such that $T_\alpha^3(\alpha+1) > T_\alpha(\alpha+1)$ is possible when $f_{d-1} \leq T_\alpha(\alpha+1) \leq T_\alpha(\alpha)$. Note that the conditions $T_\alpha^3(\alpha) < T_\alpha(\alpha)$ and $T_\alpha^3(\alpha+1) > T_\alpha(\alpha+1)$ imply that the branch number is slightly larger than 2. Now remember that I_α consists of m full cylinder sets if and only if $\alpha = k$, $N = mk(k+1)$ and $d = (m-1)(k+1)$ for some $k \in \mathbb{N}$, cf. Theorem 13. Figure 4.14 shows for increasing values of N a sequence of arrangements where the branch number b is 2, from one full arrangement (here for $N = 4$) with two cylinders to the next one (here for $N = 12$). Since $|T'_\alpha(\alpha)| > |T'_\alpha(\alpha+1)|$, Figure 4.14 suggests that in case $b = 2$, the most favourable arrangement for $T_\alpha^3(\alpha+1) = T_\alpha(\alpha+1)$ to have real roots is when $N = 2k^2 + 2k - 1$, where $k \geq 2$, while for $T_\alpha^3(\alpha) = T_\alpha(\alpha)$ to have real roots is when $N = 2k^2 + 2k + 1$, where $k \geq 1$. We will confine ourselves to investigating only the possibility of $T_\alpha^3(\alpha+1) = T_\alpha(\alpha+1)$; the calculations for the other case are similar.

So we will try and find out if for $N = 2k^2 + 2k - 1$, $d = k+1$, with $k \geq 2$, the positive

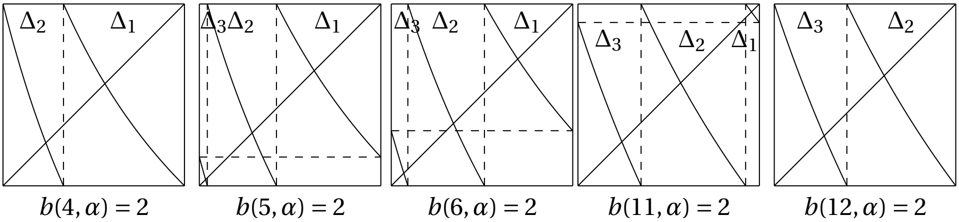


Figure 4.14: Arrangements of $\Upsilon_{N,\alpha}$ when the branch number b is 2

root of $T_\alpha^3(\alpha+1) = T_\alpha(\alpha+1)$ lies in I_α. To do this, we solve

$$\cfrac{2k^2+2k-1}{\cfrac{2k^2+2k-1}{\cfrac{2k^2+2k-1}{\alpha+1}-(k-1)}-k}-(k+1) = \frac{2k^2+2k-1}{\alpha+1}-(k-1),$$

which is reducible to

$$(2k^3+6k^2-k-1)\alpha^2+(2k^4+5k^2+k-2)\alpha-(4k^5+6k^4+2k^3-3k^2-k+1)=0,$$

yielding

$$\alpha = \frac{\sqrt{36k^8+144k^7+164k^6-12k^5-95k^4-2k^3+21k^2-4k}-(2k^4+5k^2+k-2)}{2(2k^3+6k^2-k-1)}. \quad (4.21)$$

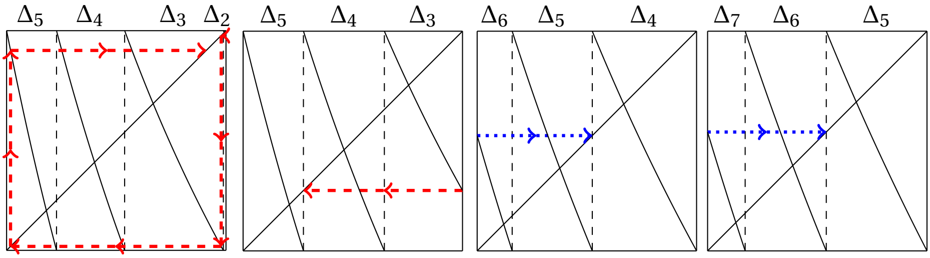
A straightforward computation shows that this last expression is smaller than f_{k+2}, meaning that the root (4.21) lies outside I_α when $I_\alpha = \Delta_{k+1}\cup\Delta_k\cup\Delta_{k-1}$. Since $N = 2k^2+2k-1$ was the most favourable option for investigation, this finishes our proof. □

Remark 8. The arrangement for $N = 11$ in Figure 4.10 illustrates that the difference between $T_\alpha^3(\alpha+1)$ and $T_\alpha(\alpha+1)$ may be very small.

A SUFFICIENT CONDITION FOR GAPLESSNESS IN CASE I_α CONSISTS OF MORE THAN THREE CYLINDERS

When I_α consists of four cylinders, the analogons of Lemmas 28 and 30 are easy to prove, since they involves a branch number larger than 3, in which case $|T'_\alpha(\alpha+1)| > 2$ when $N \geq 18$ (and Theorem 18 yields the desired result). The cases $2 \leq N \leq 17$ can be checked manually and are left to the reader; in Figure 4.15 the arrangement for $N = 11$, associated with Lemma 30 , illustrates that gaps are out of the question. The arrangements for $N = 15$, $N = 24$ and $N = 35$ in Figure 4.15 are interesting illustrations of the analogon of Lemma 29 in the case of two full cylinders instead of one.

We will confine ourselves to the arrangement for $N = 15$; the other ones have similar properties. The arrangement for $N = 15$ is the boundary case with four cylinders where



$N = 11, \alpha = f_6$ | $N = 15, \alpha = f_6$ and $T_\alpha(\alpha+1) = p_5$ | $N = 24, \alpha = f_3 - 1$ and $T_\alpha(\alpha) = p_5$ | $N = 35, \alpha = f_4 - 1$ and $T_\alpha(\alpha) = p_6$

Figure 4.15: Four arrangements with two full cylinders

the left one (that would be Δ_6 in this example) is extremely small and the right one is such that almost $p_5 \in T_\alpha^2(T_\alpha(\Delta_6) \setminus \Delta_6)$. The interesting thing is that this option would imply a quick interruption of the expansion of $T_\alpha(\Delta_6) \setminus \Delta_6$, involving two large gaps. But it is not really an option: the arrangement for $N = 15$ in Figure 4.15 is exceptional among relatively small N (as well as the arrangements for $N = 24$ and $N = 35$ are), while for $N > 36$ we have $|T'_\alpha(\alpha+1)| > 2$ when $I_\alpha = \Delta_d \cup \Delta_{d-1}\Delta_{d-2}$ and $N/(\alpha+1) - (d-2) = p_d$ or $N/\alpha - d = p_d$. We derived this in a similar way as in the proof of Lemma 29 (see Figure 4.13) or the preparations for Theorem 20 (see Figure 4.8). Figure 4.16 shows the associated 'limit graph', from which it is easily found that $a = 1/2(3-\sqrt{5})$, yielding branch number $2 + g$, with $g = 1/G$ the small golden section.

When I_α consists of five cylinders or more, there are only a few N such that α exist for which $|T'_\alpha(\alpha+1)| < 2$. However, in these cases the expanding power of T_α in the left part of the arrangements excludes any existence of gaps.

With this, we conclude the proof of Theorem 27. □

So far, we have proved a theorem on gaplessness in the case of more than five cylinder sets and one on gaplessness under a specific condition. In the next section we will prove that in the case of five cylinder sets I_α is always gapless and that in the case of four cylinders sets gaps exist only in very rare cases and if they do, that they are very large.

4.3. GAPLESSNESS IN CASE I_α CONTAINS TWO OR THREE FULL CYLINDERS

In this section we will proof that arrangements of four cylinders generally do not contain a gap, save for special values of N. Moreover, we will prove that arrangements of five cylinders cannot contain a gap. The core of these proofs are the equations

$$T_\alpha(\alpha) = T_\alpha^3(\alpha) \text{ and } T_\alpha(\alpha+1) = T_\alpha^3(\alpha+1).$$

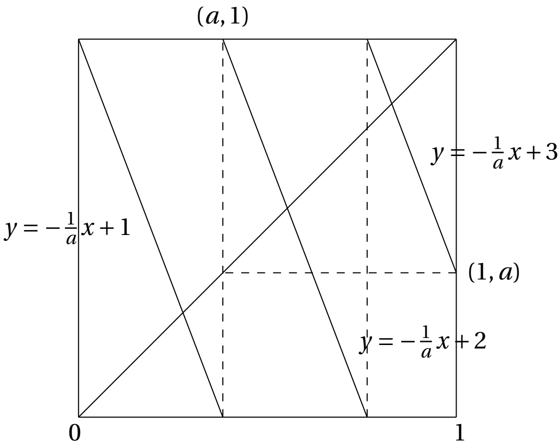


Figure 4.16: The 'limit graph' of T_α, translated over $(-\alpha,-\alpha)$, under the conditions $I_\alpha = \Delta_d \cup \Delta_{d-1}\Delta_{d-2}$ and $N/(\alpha+1)-(d-2) = p_d$ for $N\to\infty$ (and $\alpha, d\to\infty$ accordingly)

The central theorem of this section is the following:

Theorem 31. *Let $N\in\mathbb{N}_{\geq 2}$. There are $\alpha\in(0,\sqrt{N}-1)$ such that*

$$\begin{cases} I_\alpha = \Delta_d\cup\Delta_{d-1}\cup\Delta_{d-2}\cup\Delta_{d-3}, \text{ with } d := d(\alpha);\\ \text{there is a gap in } I_\alpha;\\ \text{the gap contains } f_{d-1} \text{ and } f_{d-2};\\ \Delta_d \text{ and } \Delta_{d-3} \text{ are gapless}\end{cases}$$

if and only if $N = 2k^2+2k-i$, with $k>1$ and $i\in\{1,2,3\}$.

Proof of Theorem 31: Suppose that there is a gap containing f_{d-1} and f_{d-2} in I_α and that Δ_d and Δ_{d-3} are gapless. Then the interval (f_{d-1}, f_{d-2}) is a gap. Since $f_{d-1} < f_{d-2}$, $N/(f_{d-1}+d-1) = f_{d-1}$ and $N/(f_{d-2}+d-2) = f_{d-2}$, we know that

$$(f_{d-1}, f_{d-2}) \subsetneq \left(\frac{N}{f_{d-2}+d-1}, \frac{N}{f_{d-1}+d-2}\right),$$

where the larger open interval is a gap as well. What is more, the infinite sequence of intervals

$$(f_{d-1}, f_{d-2}) \subsetneq \left(\frac{N}{f_{d-2}+d-1}, \frac{N}{f_{d-1}+d-2}\right) \subsetneq \left(\frac{N}{\frac{N}{f_{d-1}+d-2}+d-1}, \frac{N}{\frac{N}{f_{d-2}+d-1}+d-2}\right) \subsetneq \dots$$

consists of the union of (f_{d-1}, f_{d-2}) with pre-images of (f_{d-1}, f_{d-2}) in Δ_{d-1} and Δ_{d-2} respectively and therefore of gaps containing f_{d-1} and f_{d-2}. It is contained in the closed interval $[q,r]$, with

$$q = [\overline{d-1, d-2}]_{N,\alpha} \in \Delta_{d-1} \quad \text{and} \quad r = [\overline{d-2, d-1}]_{N,\alpha} \in \Delta_{d-2},$$

yielding

$$T_\alpha^2(q) = q,\ T_\alpha(q) = r,\ T_\alpha(r) = q \text{ and } T_\alpha^2(r) = r. \tag{4.22}$$

Since Δ_d and Δ_{d-3} are gapless, $T_\alpha(\alpha)$ and $T_\alpha(\alpha+1)$ lie outside the interval (q, r), which is to say

$$p_d < T_\alpha(\alpha) \le q \text{ and } r \le T_\alpha(\alpha+1) < p_{d-2}.$$

For the images of α under T_α this means that either $T_\alpha^2(\alpha) \in \Delta_{d-3}$ or $T_\alpha^2(\alpha) \in \Delta_{d-2}$, in the latter case of which we have, due to the expansiveness of T_α and the equalities of (4.22),

$$|T_\alpha(\alpha) - q| \le |T_\alpha^2(\alpha) - r| \le |T_\alpha^3(\alpha) - q|,$$

with equalities only in the case $T_\alpha(\alpha) = q$. From this we derive that

$$\text{either } T_\alpha^2(\alpha) \in \Delta_{d-3} \text{ or } T_\alpha^2(\alpha) \in \Delta_{d-2} \wedge T_\alpha^3(\alpha) \le T_\alpha(\alpha) \tag{4.23}$$

and, similarly, that

$$\text{either } T_\alpha^2(\alpha+1) \in \Delta_d \text{ or } T_\alpha^2(\alpha+1) \in \Delta_{d-1} \wedge T_\alpha^3(\alpha+1) \ge T_\alpha(\alpha+1). \tag{4.24}$$

In the following we will write $\alpha_u(N, m)$ (u for 'upper') for the positive root of the equation $T_\alpha^3(\alpha+1) = T_\alpha(\alpha+1)$ and $\alpha_l(N, m)$ (l for 'lower') for the positive root of the equation $T_\alpha^3(\alpha) = T_\alpha(\alpha)$, with m the number of full cylinder sets; in the current case we have $m = 2$. We remark that we have to do with arrangements very similar to the ones related to the proof of Lemma 30, albeit now two full cylinders in between *two* very narrow ones; see Figures 4.17 and 4.18. We recall that I_α consists of m full cylinder sets if and only if $\alpha = k$, $N = mk(k+1)$ and $d = (m-1)(k+1)$ for some $k \in \mathbb{N}$, cf. Theorem 13. Note that in the arrangements of this theorem (with $m = 2$) we have $d = k+2$. Since Δ_{d-3} decreases and Δ_d increases as α decreases, we see that the assumption that there is a gap containing f_{d-1} and f_{d-2} in I_α implies $\alpha_u(N,2) \ge \alpha_l(N,2)$. We will shortly show that the only values of N for which $\alpha_u(N,2) \ge \alpha_l(N,2)$ are $N = 2k^2 + 2k - i$, with $k > 1$ and $i \in \{1,2,3\}$; in all cases $d = k+2$. Although we could keep i as a variable in our calculations, we can limit ourselves to the case $i = 3$, since $i = 3$ is the least favourable value of i allowing for a gap, as is suggested in Figures 4.14 and 4.17 through 4.20. We will show that for $i = 3$ indeed $\alpha_u(N,2) \ge \alpha_l(N,2)$. Subsequently we will show that for $4 \le i \le 4k$ no gaps exist; the upper bound is $4k$, since $2k^2 + 2k - 4k = 2(k-1)^2 + 2(k-1)$, so as to confine the calculations to the group of arrangements where $d = k+2$.

So, let $N = 2k^2 + 2k - 3$ and $d = k+2$. Then (omitting straightforward calculations)

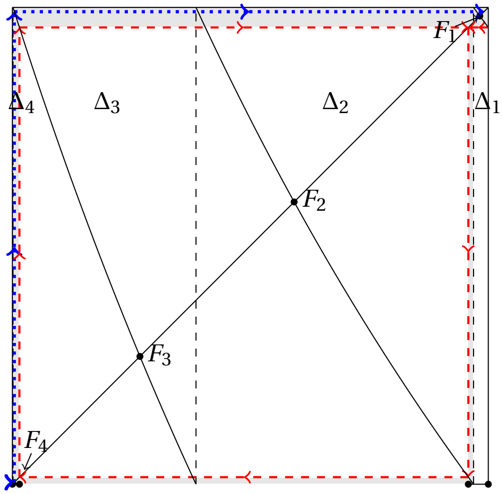
$$\alpha_u(2k^2+2k-3, 2) = \frac{\sqrt{D} - (2k^4 + 3k^2 + 3k - 6)}{4k^3 + 12k^2 - 6k - 6}$$

and

$$\alpha_l(2k^2+2k-3, 2) = \frac{\sqrt{D} - (2k^4 + 12k^3 + 15k^2 - 7k - 12)}{4k^3 - 18k - 8},$$

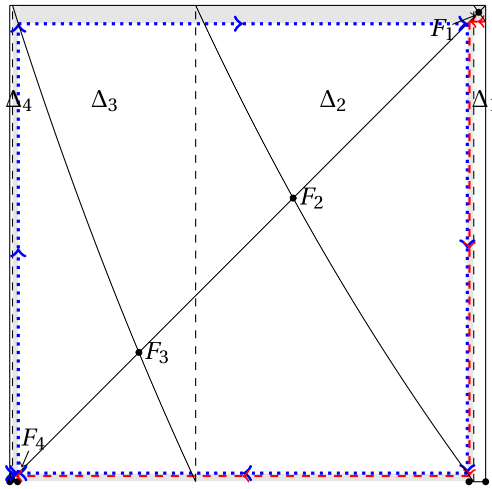
with

$$D = 36k^8 + 144k^7 + 60k^6 - 324k^5 - 207k^4 + 294k^3 + 117k^2 - 108k.$$



$N = 11, \alpha = 1.8719\cdots$ $(k = 2,\ i = 1)$
$T_\alpha^3(\alpha+1) = T_\alpha(\alpha+1)$

Figure 4.17



$N = 11, \alpha = 1.8687\cdots$ $(k = 2,\ i = 1)$
$T_\alpha^3(\alpha) = T_\alpha(\alpha)$

Figure 4.18

Since we assume that there is a gap containing f_{d-1} and f_{d-2}, we have $\alpha_u \geq \alpha_l$, which inequality is equivalent to

$$(6k^2+6k+1)\sqrt{D} \leq 36k^6+108k^5+36k^4-108k^3-57k^2+15k+12.$$

Some elementary calculations show that

$$\begin{aligned}(6k^2+6k+1)\sqrt{D} &< (6k^2+6k+1)\left(6k^4+12k^3-7k^2-13k+\frac{14}{3}+\frac{10}{9k^2}-\frac{10}{9k^3}+\frac{46}{27k^4}-\frac{61}{27k^5}\right)\\ &= 36k^6+108k^5+36k^4-108k^3-57k^2+15k+\frac{34}{3}+\frac{14}{3k^2}-\frac{40}{9k^3}-\frac{320}{27k^4}-\frac{61}{27k^5}\\ &< 36k^6+108k^5+36k^4-108k^3-57k^2+15k+12,\end{aligned}$$

where the last inequality holds for $k \geq 4$. In the case $k = 3$ (and so $N = 21$), we have

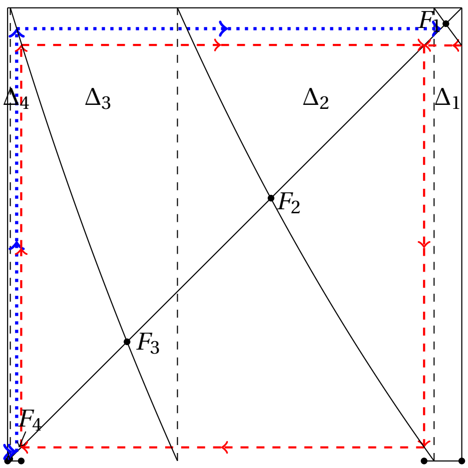
$$\alpha_u(2k^2{+}2k{-}3,2) = \frac{\sqrt{508032}-192}{192} = 2.7123\cdots > 2.7122\cdots = \frac{\sqrt{508032}-588}{46} = \alpha_l(2k^2{+}2k{-}3,2);$$

see Figure 4.22.

Although we have not explicitly calculated the cases $N = 2k^2+2k-1$ and $N = 2k^2+2k-2$, these cases allow for larger intervals $[\alpha_l, \alpha_u]$ where large gaps exist; see the next examples.

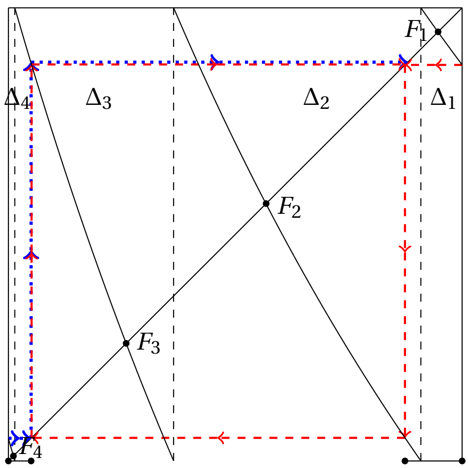
$$\alpha_u(11,2) = \frac{\sqrt{9075}-26}{37} = 1.8719\cdots \quad \text{and} \quad \alpha_l(11,2) = \frac{99-\sqrt{9075}}{2} = 1.8686\cdots$$
$$\alpha_u(10,2) = \frac{\sqrt{1725}-12}{17} = 1.7372\cdots \quad \text{and} \quad \alpha_l(10,2) = \frac{45-\sqrt{1725}}{2} = 1.7334\cdots$$
$$\alpha_u(9,2) = \frac{\sqrt{5103}-22}{31} = 1.5946\cdots \quad \text{and} \quad \alpha_l(9,2) = \frac{27-\sqrt{567}}{2} = 1.5941\cdots$$
$$\alpha_u(8,2) = \frac{\sqrt{228}-5}{7} = 1.4428\cdots \quad \text{and} \quad \alpha_l(8,2) = 9-\sqrt{57} = 1.4501\cdots$$

We see that the intervals $\alpha_u - \alpha_l$ decrease as N decreases, until (for $N = 8$) the 'interval' would have negative length, hence does not exist.



$N = 10, \alpha = 1.7372\cdots$ $(k = 2,\ i = 2)$
$T^3_\alpha(\alpha+1) = T_\alpha(\alpha+1)$

Figure 4.19



$N = 9, \alpha = 1.5946\cdots$ $(k = 2,\ i = 3)$
$T^3_\alpha(\alpha+1) = T_\alpha(\alpha+1)$

Figure 4.20

Now suppose $N = 2k^2 + 2k - 4$ (note that voor $k = 2$ we have $N = 8$). Then

$$\alpha_u(2k^2+2k-4,2) = \frac{(k+2)\sqrt{D}-(k^3+k^2+2k+4)}{2(k^2+4k+2)}$$

and

$$\alpha_l(2k^2+2k-4,2) = \frac{(k-1)\sqrt{D}-(k^3+4k^2-k-4)}{2(k^2-2k-1)},$$

with

$$D = 9k^4 + 18k^3 - 7k^2 - 16k.$$

There are no gaps provided $\alpha_l(2k^2+2k-4,2)-\alpha_u(2k^2+2k-4,2)>0$, which inequality is equivalent with $(3k^2+3k)\sqrt{D}-(9k^4+18k^3+k^2-8k-4)>0$. Since

$$D=9k^4+18k^3-7k^2-16k>9k^4+18k^3-7k^2-16k-\frac{4}{45}-\frac{36}{5k}+\frac{32}{5k^2}+\frac{36}{25k^4}=(3k^2+3k-\tfrac{8}{3}-\tfrac{6}{5k^2})^2,$$

we find

$$\begin{aligned}&(3k^2+3k)\sqrt{D}-(9k^4+18k^3+k^2-8k-4)\\ >&(3k^2+3k)(3k^2+3k-\tfrac{8}{3}-\tfrac{6}{5k^2})-(9k^4+18k^3+k^2-8k-4)\\ =&9k^4+18k^3+k^2-8k-\tfrac{18}{5k}-\tfrac{18}{5}-(9k^4+18k^3+k^2-8k-4)=\tfrac{2}{5}-\tfrac{18}{5k}>0,\end{aligned}$$

for $k>9$. A manual inspection of the cases $k\in\{3,\ldots,9\}$ shows that then $\alpha_u<\alpha_l$ as well. And since we had already explicitly calculated the case $k=2$ (and so $N=8$), we conclude that there are no gaps in case $N=2k^2+2k-4$. When we replace the number 4 in $N=2k^2+2k-4$ by larger integers (if possible), there will not be any gaps either: the length of the 'interval' $[\alpha_l,\alpha_u]$ would only become more negative. This concludes the proof that if $I_\alpha=\Delta_d\cup\Delta_{d-1}\cup\Delta_{d-2}\cup\Delta_{d-3}$, with $d:=d(\alpha)$ and there is a gap containing f_{d-1} and f_{d-2} in I_α, then $N=2k^2+2k-i$, with $k>1$ and $i\in\{1,2,3\}$.

For the converse statement, we assume that $N=2k^2+2k-i$, with $k\in\mathbb{N}$ and $i\in\{1,2,3\}$. Earlier in this proof we showed that only then $\alpha_l(N,\alpha)\le\alpha_u(N,\alpha)$. We will show that for α such that $\alpha_l(N,\alpha)\le\alpha\le\alpha_u(N,\alpha)$, implying both (4.23) and (4.24), there is a gap in I_α containing both f_{d-1} and f_{d-2} and that Δ_d and Δ_{d-3} are gapless. To do so, we will apply the following two lemmas:

Lemma 32. *Let $N=2k^2+2k-i$, where $k>1$ and $i\in\{1,2,3\}$. Let $\alpha\in[\alpha_l,\alpha_u]$ be such that $I_\alpha=\Delta_d\cup\Delta_{d-1}\cup\Delta_{d-2}\cup\Delta_{d-3}$, with $d:=d(\alpha)$. Suppose that $T_\alpha^2(\alpha)>f_{d-3}$. Then $T_\alpha^3(\alpha)>p_{d-2}$.*

Lemma 33. *Let $N=2k^2+2k-i$, where $k>1$ and $i\in\{1,2,3\}$. Let $\alpha\in[\alpha_l,\alpha_u]$ be such that $I_\alpha=\Delta_d\cup\Delta_{d-1}\cup\Delta_{d-2}\cup\Delta_{d-3}$, with $d:=d(\alpha)$. Let α_1 be the positive root of $T_\alpha^2(\alpha)=f_{d-3}$ and α_2 be the positive root of $T_\alpha^3(\alpha+1)=p_{d-2}$ Then $\alpha_1>\alpha_2$.*

Proof of Lemma 32: We assume that $T_\alpha^2(\alpha)>f_{d-3}$, so

$$T_\alpha^2(\alpha)\in\Delta_{d-3}\text{ and }T_\alpha(\alpha+1)<T_\alpha^3(\alpha)<f_{d-3}.\qquad(4.25)$$

For reasons we explained in the proofs of the previous lemmas, we can limit ourselves to taking $i=1$. Even stronger, we can take $\alpha=\alpha_u$, for this value of α renders the largest possible $T_\alpha^2(\alpha)$ and the smallest $T_\alpha^3(\alpha)$ for which (4.25) holds. So we need to prove that

$$T_{\alpha_u}^3(\alpha_u)>p_{d-2}=\frac{2k^2+2k-1}{\alpha_u+k},\qquad(4.26)$$

where α_u equals the right part of (4.21) on page 106. Note that $\alpha_u\in\Delta_{k+2}$. We can write (4.26) as

$$(5k^3+8k^2-k-3)\alpha_u^2+(5k^4+14k^3+11k^2-3)\alpha_u-(10k^5+22k^4+5k^3-10k^2-k+1)<0.\quad(4.27)$$

Substituting the right part of (4.21) for α_u, we can write (4.27) as $s\sqrt{D}-t<0$, with

$$\begin{cases} s = 42k^6+78k^5+8k^4-24k^3+3k^2+4k-3; \\ D = 36k^8+144k^7+164k^6-12k^5-95k^4-2k^3+21k^2-4k \text{ (see (4.21))}; \\ t = 252k^{10}+972k^9+1058k^8-76k^7-558k^6-90k^5+15k^4+43k^3+13k^2-11k+2. \end{cases}$$

One can tackle this inequality by noting that for $k\geq 1$ we have

$$\begin{aligned} D &< 36k^8+144k^7+164k^6-12k^5-\tfrac{851}{9}k^4-\tfrac{10}{9}k^3+\tfrac{557}{27}k^2-\tfrac{130}{27}k+\tfrac{25}{81} \\ &= (6k^4+12k^3+\tfrac{5}{3}k^2-\tfrac{13}{3}k+\tfrac{5}{9})^2, \end{aligned}$$

yielding

$$s\sqrt{D}-t < -4k^8-24k^7-\tfrac{40}{3}k^6+\tfrac{356}{3}k^5+\tfrac{1156}{9}k^4-\tfrac{296}{3}k^3-\tfrac{101}{3}k^2+\tfrac{236}{9}k-\tfrac{11}{3},$$

which is smaller than 0 for $k\geq 3$. Since $k>1$ and we already discussed the case $k=2$, this finishes the proof of Lemma 32. □

Proof of Lemma 33: Again, we can limit ourselves to $i=1$. With some elementary calculations we find that

$$\alpha_1 = \frac{(4k^4+8k^3-4k+1)\sqrt{9k^2+6k-3}-(4k^5+4k^4-24k^3-40k^2-7k+9)}{8k^4+24k^3+32k^2+30k+14}$$

and

$$\alpha_2 = \frac{\sqrt{225k^8+390k^7-201k^6-384k^5+183k^4+102k^3-75k^2+12k}-(5k^4+5k^3-3k^2-k)}{2(5k^3+4k^2-3k)}.$$

Proving $\alpha_1>\alpha_2$ for $k\geq 2$ is elementary, despite the large fractions, and is left to the reader. This finishes the proof of Lemma 33. □

Proceeding with the proof of Theorem 31, we assume that $T_\alpha^2(\alpha+1)\geq T_\alpha(\alpha)$; the case $T_\alpha^2(\alpha)\leq T_\alpha(\alpha+1)$ is similar. We may assume that at least one of the small cylinders Δ_d and Δ_{d-3} is gapless (see for instance the proof of Theorem 20). First suppose that Δ_{d-3} is gapless. Then so are $T_\alpha(\Delta_{d-3})=[T_\alpha(\alpha+1),\alpha+1)$ and $T_\alpha^2(\Delta_{d-3})=[T_\alpha(\alpha+1),\alpha+1)\cup[\alpha,T_\alpha^2(\alpha+1)\supset\Delta_d$. Now suppose that Δ_d is gapless. Then so are $T_\alpha(\Delta_d)=[\alpha,T_\alpha(\alpha)]$ and $T_\alpha^2(\Delta_d)=[\alpha,T_\alpha(\alpha)]\cup[T_\alpha^2(\alpha),\alpha+1)$. If $T_\alpha^2(\alpha)\leq f_{d-3}$, then $T_\alpha^3(\Delta_d)\supset\Delta_{d-3}$ and Δ_{d-3} is gapless as well. If $T_\alpha^2(\alpha)>f_{d-3}$, we apply Lemma 32, yielding that $T_\alpha^4(\Delta_d)\supset(T_\alpha(\alpha+1),p_{d-2})$. It follows that $T_\alpha^5(\Delta_d)\supset(p_d,T_\alpha^2(\alpha+1))$ and $T_\alpha^6(\Delta_d)\supset(T_\alpha^3(\alpha+1),\alpha+1)$. Applying Lemma 33, we conclude that $f_{d-3}\in T_\alpha^6(\Delta_d)$, from which the gaplessness of Δ_{d-3} follows.

The conclusion so far, still assuming $T_\alpha^2(\alpha+1)\geq T_\alpha(\alpha)$, is that orbits once having passed through Δ_d or Δ_{d-3} will never enter $(T_\alpha^2(\alpha+1),T_\alpha(\alpha+1))$. We will show that this interval is actually a gap.

Clearly, for any $x\in A=I_\alpha\setminus(T_\alpha^2(\alpha+1),T_\alpha(\alpha+1))$, we have that $T_\alpha^n(x)\in A$ for all $n\in\mathbb{N}$. So we only need to show that – apart of a set of measure zero – for every $x\in(T_\alpha^2(\alpha+$

$1), T_\alpha(\alpha+1))$ there exists an $n_o \geq 1$ for which $T_\alpha^{n_0}(x) \in A$. To show this, consider the map $T : I_\alpha \to I_\alpha$, defined by

$$T(x) = \begin{cases} \frac{-x}{p_d-\alpha} + \frac{(\alpha+1)p_d-\alpha^2}{p_d-\alpha}, & \text{if } x \in \Delta_d; \\ \frac{N}{x} - (d-1), & \text{if } x \in \Delta_{d-1}; \\ \frac{N}{x} - (d-2), & \text{if } x \in \Delta_{d-2}; \\ \frac{-x}{\alpha+1-p_{d-2}} + \frac{(\alpha+1)^2-\alpha p_{d-2}}{\alpha+1-p_{d-2}}, & \text{if } x \in \Delta_{d-3}. \end{cases} \tag{4.28}$$

So on Δ_d and on Δ_{d-3} we have that T is a straight line segment with negative slope, through $(\alpha, \alpha+1)$ and (p_d, α) on Δ_d, resp. through $(p_{d-2}, \alpha+1)$ and $(\alpha+1, \alpha)$ on Δ_{d-3}. For $x \in \Delta_{d-1} \cup \Delta_{d-2}$ we have that $T(x) = T_\alpha(x)$.

4

According to Definition 1.1.1 from Fritz Schweiger's monograph [12], we have that (I_α, T) is a *fibred system*, i.e.

(a) there is a finite or countable set $\mathscr{I}$ (called the digit set; here $\mathscr{I} = \{d-3, d-2, d-1, d\}$);

(b) there is a map $k : I_\alpha \to \mathscr{I}$. Then the sets

$$B(i) = k^{-1}\{i\} = \{x \in B : k(x) = i\}$$

form a partition of I_α (here $B(i) = \Delta_i$, for $i \in \mathscr{I}$);

(c) the restriction of T to any $B(i)$ is an injective map.

We now use results from [12] to show that the map T from (4.28) is ergodic and has a unique T-invariant measure μ which is absolutely continuous with respect to Lebesgue measure λ on I_α. Once this has been established, almost every $x \in G = I_\alpha \setminus A$ enters A with positive frequency. Since $G \subset \Delta_{d-1} \cup \Delta_{d-2}$, this and the definition of T implies that G is a gap for the map T_α.

In order to show that T is ergodic, one only needs to show that T satisfies the conditions of Adler's *Folklore Theorem*; see Theorem 15.2.1 in [12]. Once the conditions of this theorem are satisfied (which is easy for our map T), we know that the so-called *Rényi Condition* holds, which is condition (c) of Rényi's Theorem 9.5.3 in [12]. Now applying Corollary 9.5.4 (from [12]) immediately yields that T is ergodic.

Furthermore, since the conditions of Rényi's Theorem 9.5.3 in [12] are satisfied, Rényi's Theorem 15.1.2 in [12] now yields that there exists a unique invariant probability measure μ on $[0,1)$ such that for some positive constant C

$$C^{-1}\lambda(E) \leq \mu(E) \leq C\lambda(E),$$

where E is any Borel measurable subset of $[0,1)$. This finishes the proof of Theorem 31. □

We stress that the in case of $N = 2k^2 + 2k - 3$ the intervals $[\alpha_l, \alpha_u]$ on which gaps exist may be very small; see Figure 4.22. On the other hand, in case $N = 2k^2 + 2k - 4$, the gaplessness may be a very close call; see Figure 4.21. Table 4.1 illustrates how fast these differences between α_l and α_u decrease as N increases:

	$\alpha_l(N,2)$	$\alpha_u(N,2)$
$N = 9$	$1.594119\cdots$	$1.594686\cdots$
$N = 21$	$2.712252\cdots$	$2.712310\cdots$
$N = 37$	$3.776839\cdots$	$3.776851\cdots$
$N = 57$	$4.817672\cdots$	$4.817675\cdots$
$N = 8$	$1.450165\cdots$	$1.442809\cdots$
$N = 20$	$2.613247\cdots$	$2.611575\cdots$
$N = 36$	$3.700989\cdots$	$3.700407\cdots$
$N = 56$	$4.756087\cdots$	$4.755832\cdots$

Table 4.1: The thin thread between having a gap or not

Remark 9. While a fixed point f_i is repellent for points within Δ_i, the fixed points in two adjacent cylinder sets are in a way attracting points from each other. As a consequence, it may take quite some time before the orbit of points in the full cylinders of gap arrangements with four cylinders leave these full cylinders for the first time. As an example we take the gap arrangement for $k = 50$ (according to the notations used above). Then $N = 2\cdot 50^2 + 2\cdot 50 - 3 = 5097$, $d = 52$ and $\alpha \approx \alpha_u \approx \alpha_l \approx 49.98019737$. Table 4.2 shows for ten values of x between α and $\alpha + 1$ the smallest n such that $T_\alpha^n(x) \notin \Delta_{51} \cup \Delta_{50}$.

x	50	50.1	50.2	50.3	50.4	50.5	50.6	50.7	50.8	50.9
n	5417	2090	3568	1123	4776	185	5816	16231	5646	7604

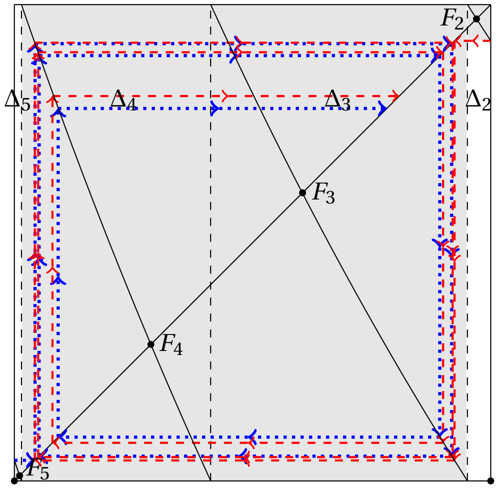
Table 4.2: The difficulty of leaving the gap: with $N = 5097$, $\alpha = 49.98019737$, for each of ten values of $x \in [\alpha, \alpha+1]$ the smallest n is given such that $T_\alpha^n(x) \notin \Delta_{51} \cup \Delta_{50}$.

When I_α consists of five cylinder sets, there are no special cases for N allowing gaps. But more than that, we remark the following: when $N = 18$ and $\alpha = 2$, we have a full arrangement with three cylinders, so the branch number is 3, while $|T'_\alpha(\alpha + 1)| = 2$. Any arrangement with five cylinders has a branch number larger than 3 and therefore has $|T'_\alpha(\alpha + 1)| > 2$. We leave it to the reader to manually check the cases $2 \le N \le 17$, and conclude:

Theorem 34. *Let I_α consist of five cylinders or more. Then I_α has no gaps.*

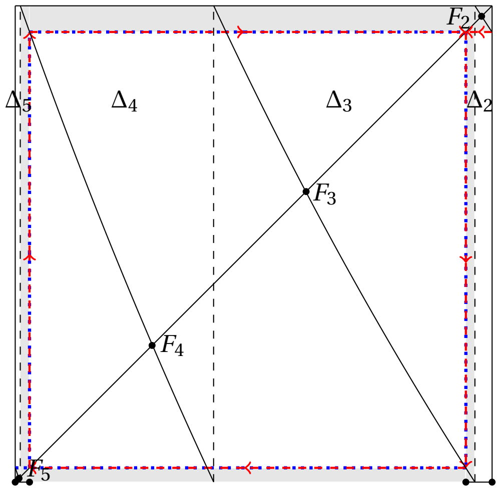
Now that we have proved some theorems for determining gaplessness and found the unique way in which a gap can occur in four-cylinder cases, in the next section we will give an introduction to the occurrence of gaps in case I_α consists of two or three cylinder sets.

4



$N = 20, \alpha = 2.6124$
$\alpha_l = 2.6132\cdots$ and $\alpha_u = 2.6115\cdots$

Figure 4.21



$N = 21, \alpha = 2.7123$
$\alpha_l = 2.7122\cdots$ and $\alpha_u = 2.7123\cdots$

Figure 4.22

4.4. A WORLD OF GAPS

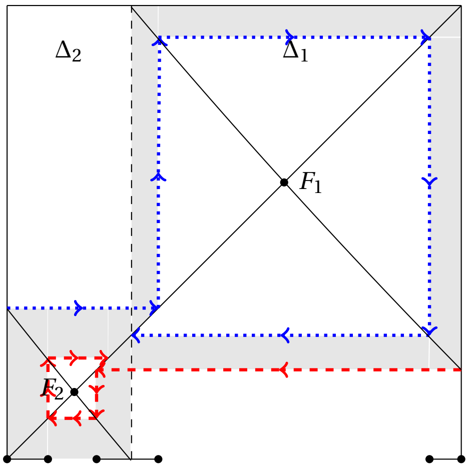
In this section we will give some results on sufficient conditions for gaps in two- and three-cylinder cases, but these results hardly unveil the complex and diverse occurrence of gaps in these cases. At the end of this section we will give two examples (without further explanation) of arrangements with many gaps, one containing two cylinders, one containing three cylinders. In a subsequent paper we will give a complete inventory of the existence of gaps in two- and three-cylinder cases.

When I_α exists of two cylinder sets only, it is not hard to find sufficient conditions for the existence of gaps. We distinguish two cases. The first one is the most obvious (see also the example $N = 50$ and $\alpha = 6$ from Section 1):

Proposition 35. *Let $I_\alpha = \Delta_d \cup \Delta_{d-1}$. If $T_\alpha(\alpha) < f_{d-1}$, there is a gap containing f_{d-1}; if $T_\alpha(\alpha+1) > f_d$, there is a gap containing f_d.*

Proof of Proposition 35: Let $I_\alpha = \Delta_d \cup \Delta_{d-1}$ and suppose $T_\alpha(\alpha) < f_{d-1}$. Since T_α is expanding and $T_\alpha(\alpha) < f_{d-1}$, we find that $f_{d-1} < T_\alpha^2(\alpha)$. Again due to the fact that T_α is expanding we have that $(T_\alpha(\alpha+1), T_\alpha^3(\alpha)) \subset [\alpha, T_\alpha(\alpha)]$. Consequently $T_\alpha^n(\Delta_d) \subset I_\alpha \setminus (T_\alpha(\alpha), T_\alpha^2(\alpha))$ for $n \geq 2$, with $f_{d-1} \in (T_\alpha(\alpha), T_\alpha^2(\alpha))$. Now let $x \in I_\alpha \setminus \{f_{d-1}\}$. Then there is a smallest integer n_0 (possibly 0) such that $T_\alpha^{n_0}(x) \in \Delta_d$. Since $T_\alpha^n(\Delta_d) \subset I_\alpha \setminus (T_\alpha(\alpha), T_\alpha^2(\alpha))$ for $n \geq 2$, it follows that $T_\alpha^n(x) \notin (T_\alpha(\alpha), T_\alpha^2(\alpha))$ for $n \geq n_0$. We conclude that $(T_\alpha(\alpha), T_\alpha^2(\alpha))$ is a gap containing f_{d-1}. The reasoning is completely similar in case $T_\alpha(\alpha+1) > f_d$. □

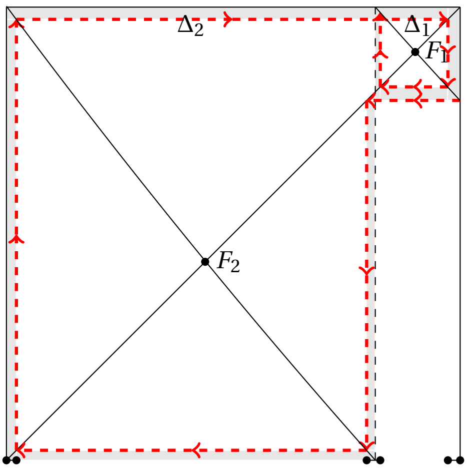
Remark 10. Proposition 35 allows for the situation in which both $T_\alpha(\alpha) < f_{d-1}$ and $T_\alpha(\alpha+1) > f_d$, in which case there are gaps around both fixed points; see Figure 4.23.



$N = 100, \alpha = \frac{\sqrt{2311076}-176}{151} = 8.9021\cdots$

$T_\alpha^3(\alpha) = p_2$

Figure 4.23



$N = 100, \alpha = \frac{\sqrt{409}-3}{2} = 8.61187\cdots$

$\alpha = f_3$

Figure 4.24

The second case in which gaps occur when I_α exists of two cylinder sets only is given by the following proposition, the proof of which is very similar to the one of Proposition 35.

Proposition 36. *Let $I_\alpha = \Delta_d \cup \Delta_{d-1}$. If $T_\alpha(\alpha) < f_{d-1}$ and $T_\alpha^3(\alpha) < f_d$, or if $T_\alpha(\alpha+1) > f_d$ and $T_\alpha^3(\alpha+1) > f_{d-1}$, then both f_{d-1} and f_d are contained in a gap.*

Proof of Proposition 36: Let $I_\alpha = \Delta_d \cup \Delta_{d-1}$ and suppose that $T_\alpha(\alpha) < f_{d-1}$ and $T_\alpha^3(\alpha) < f_d$. Due to the previous result we have that $(T_\alpha(\alpha), T_\alpha^2(\alpha))$ is a gap in Δ_{d-1} containing f_{d-1}. Furthermore, $T_\alpha((p_d, T_\alpha(\alpha)]) = [T_\alpha^2(\alpha), \alpha+1)$ and $T_\alpha^2((p_d, T_\alpha(\alpha)]) \subset (\alpha, f_d)$. From this it follows that $T_\alpha^n((p_d, T_\alpha(\alpha)]) \subset I_\alpha \setminus \{(T_\alpha^3(\alpha), T_\alpha^4(\alpha)) \cup (T_\alpha(\alpha), T_\alpha^2(\alpha))\}$ for $n \geq 3$, with $f_{d-1} \in (T_\alpha(\alpha), T_\alpha^2(\alpha))$ and $f_d \in (T_\alpha^3(\alpha), T_\alpha^4(\alpha))$ (*). Now let $x \in I_\alpha \setminus \{f_{d-1}\}$ be such that no integer m exists for which $T_\alpha^m(x) = f_d$. Then there is a smallest integer n_0 (possibly 0) such that $T_\alpha^{n_0}(x) \in (p_d, T_\alpha(\alpha)]$. Since (*) holds, it follows that $T_\alpha^n(x) \notin (T_\alpha(\alpha), T_\alpha^2(\alpha)) \cup (T_\alpha^3(\alpha), T_\alpha^4(\alpha))$ for $n \geq n_0$. We conclude that $(T_\alpha(\alpha), T_\alpha^2(\alpha))$ is a gap containing f_{d-1} and $(T_\alpha^3(\alpha), T_\alpha^4(\alpha))$ is a gap containing f_d. The reasoning is completely similar if we suppose $T_\alpha(\alpha+1) > f_d$ and $T_\alpha^3(\alpha+1) > f_{d-1}$. □

Remark 11. The conditions of Proposition 36 imply that one of the cylinder sets is almost or completely full, while the other is very small; see Figure 4.24. One might even wonder whether gaps exist that contain α or $\alpha+1$. This is not the case, as stated in the following proposition.

Proposition 37. *Let $I_\alpha = \Delta_d \cup \Delta_{d-1} \cup \ldots \cup \Delta_{d-k}$, with $d := d(\alpha)$ and $1 \le k \le d-1$. Then for every open interval $K \subset I_\alpha$ there exist an $n \in \mathbb{N}$ and $s \in \Delta_d$, $s > \alpha$, and $t \in \Delta_{d-k}$, $t < \alpha + 1$ such that both $[\alpha, s)$ and $(t, \alpha+1)$ are contained in $T_\alpha^n(K)$.*

Proof of Proposition 37: Suppose that $K \subset I_\alpha$ is an open interval and suppose that K contains a discontinuity point p_i. Then $T_\alpha(K) \supset [\alpha, v) \cup (w, \alpha+1)$ for some $\alpha < v < f_d$ and $f_{d-k} < w < \alpha + 1$ and we are done. Suppose then that K does not contain a discontinuity point. Note that for $x \in [\alpha, f_{d-k})$ there exists an $\varepsilon > 0$ such that $|T'_\alpha(x)| > 1 + \varepsilon$. Now suppose that for all $m \ge 0$ and all discontinuity points p_i we have that $p_i \notin T_\alpha^m(K)$. Since $T_\alpha(f_{d-k}, \alpha+1)) \subset [\alpha, f_{d-k})$, it follows that – assuming that the interval $K \ne \emptyset$ – it would follow that

$$|T_\alpha^{2m}(K)| \ge (1+\varepsilon)^m |K|,$$

which is impossible since $(1+\varepsilon)^m \to \infty$ as $m \to \infty$. □

Now suppose I_α consists of three cylinder sets. Then there is at least one gap when both conditions for gaplessness of Theorem 27 are unfulfilled:

Proposition 38. *Let $I_\alpha = \Delta_d \cup \Delta_{d-1} \cup \Delta_{d-2}$, and suppose that $T_\alpha(\alpha) < f_{d-1} < T_\alpha(\alpha+1)$. Let $a = \max\{T_\alpha(\alpha), T_\alpha^2(\alpha+1)\}$ and $b = \min\{T_\alpha^2(\alpha), T_\alpha(\alpha+1)\}$. Then (a, b) is a gap containing f_{d-1}.*

Proof of Proposition 38: Let $I_\alpha = \Delta_d \cup \Delta_{d-1} \cup \Delta_{d-2}$, and suppose that $T_\alpha(\alpha) < f_{d-1} < T_\alpha(\alpha+1)$. Then $T_\alpha((p_d, T_\alpha(\alpha)]) = [T_\alpha^2(\alpha), \alpha+1)$ and $T_\alpha([T_\alpha(\alpha+1), p_{d-1})) = (\alpha, T_\alpha^2(\alpha+1)]$ (*). Now let $x \in I_\alpha \setminus \{f_{d-1}\}$ be such that *no* non-negative integer m exists for which $T_\alpha^m(x) \in \{f_d, f_{d-2}\}$. Then there are smallest integers n_0 and n_1 (possibly 0) such that $T_\alpha^{n_0}(x) \in (p_d, T_\alpha(\alpha)]$ and $T_\alpha^{n_1}(x) \in [T_\alpha(\alpha+1), p_{d-1})$. Let $n = \max\{n_0, n_1\}$ and let $a = \max\{T_\alpha(\alpha), T_\alpha^2(\alpha+1)\}$ and $b = \min\{T_\alpha^2(\alpha), T_\alpha(\alpha+1)\}$. Since (*) holds, it follows that $T_\alpha^{n'+1}(x) \notin (a, b)$ for $n' \ge n$. We conclude that (a, b) is a gap containing f_{d-1}. □

Remark 12. When we combine Proposition 35 and Theorem 20, we conclude that $I_\alpha = \Delta_d \cup \Delta_{d-1}$ is gapless *if and only if* $T_\alpha(\alpha) \ge f_{d-1}$ **and** $T_\alpha(\alpha+1) \le f_d$; similarly, when we combine Proposition 38 and Theorem 27, we conclude that $I_\alpha = \Delta_d \cup \Delta_{d-1} \cup \Delta_{d-2}$ is gapless *if and only if* $T_\alpha(\alpha) \ge f_{d-1}$ **or** $T_\alpha(\alpha+1) \le f_{d-1}$.

Since the branch number is a decreasing function of α (cf. Lemma 17), there will be a maximal α such that the condition $T_\alpha(\alpha) < f_{d-1} < T_\alpha(\alpha+1)$ of Proposition 38 can be fulfilled in case I_α consists of three cylinder sets. Due to Theorem 27 this is the smallest α for which gaps exist in case I_α consists of less than four cylinder sets. We remark that $T_\alpha(\alpha) < f_{d-1} < T_\alpha(\alpha+1)$ implies that $b(N, \alpha) < 2$ and leave it to the reader to prove the following:

Corollary 39. *Let $I_\alpha = \Delta_d \cup \Delta_{d-1} \cup \Delta_{d-2}$. Then I_α is gapless when*

$$\alpha \le \min_\alpha \{0 < \alpha \le \sqrt{N} - 1 : T_\alpha(\alpha) = f_{d-1} \text{ and } b(N, \alpha) < 2\}.$$

For $N = 9, 20, 36, 100$ this yields $\alpha = 9/(\sqrt{10}+2) = 1.7434\cdots, 40/(\sqrt{89}+5) = 2.7712\cdots$, $36/(\sqrt{40}+3) = 3.8607\cdots, 200/(\sqrt{449}+9) = 6.6247\cdots$ respectively; these numbers are supported by the simulations in [5] save for the case $N = 100$, the explanation of which is probably the smallness of the associated last gap; straightforward calculations show that this gap exists for $6.6247\cdots = 200/(\sqrt{449}+9) < \alpha < (25\sqrt{449} - 178)/53 = 6.6366\cdots$. This gap is largest when $T_\alpha(\alpha) = T_\alpha^2(\alpha+1)$, with root $\alpha = (\sqrt{1041209} - 397)/94 = 6.6318\cdots$, yielding a maximum width of $0.0242\cdots$.

Before, in the case of I_α existing of two cylinder sets, we already saw how Propositions 35 and 36 yielded two types of arrangements having two gaps. However, arrangements with two cylinders exist with much more gaps; see Figure 4.28. In the case of three cylinders the variety of possible arrangements is even larger, although this does not necessarily relates to larger numbers of gaps. Before we will finish this section with two arrangements (see Figures 4.28 and 4.27) illustrating the complexity of gaps in two- and three-cylinder cases, we will make some observations regarding N-continued fraction expansions in case I_α has gaps.

4

By definition, gaps are open intervals in I_α that are eventually excluded by orbits of any point of I_α, giving rise to the following definition:

Definition 7. Let $N \in \mathbb{N}_{\geq 7}$ and $0 < \alpha \leq \sqrt{N} - 1$. Suppose that I_α has at least one gap and let $\{I_1, \ldots, I_m\}$, $m \in \mathbb{N}$, be all gaps of I_α containing a fixed point. Let $n_0 \in \mathbb{N}$ be the smallest integer such that $T_\alpha^{n_0}(x) \cap (I_1 \cup \ldots \cup I_m) = \emptyset$ for almost every $x \in I_\alpha$. The *characteristic part of the orbits in* I_α is defined as the orbit $(T_\alpha^{n_0}(x), T_\alpha^{n_0+1}(x), T_\alpha^{n_0+2}(x), \ldots)$.

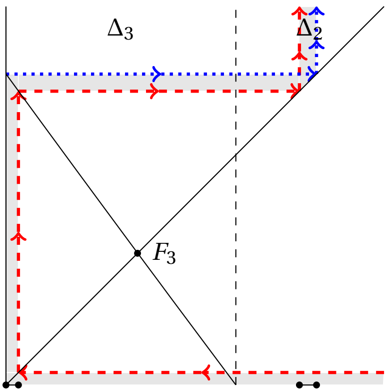
In the characteristic part digits associated with gaps will only occur a limited number of times in succession as partial quotients of the related N-continued fraction expansion. This is specifically interesting in case all fixed points are contained in a gap, as the following example will illustrate.

Again, we take $N = 100$; for α we take the root of $T_\alpha^2(\alpha) = p_2$, yielding $\alpha = (\sqrt{281684} - 78)/53 = 8.5422\cdots$. Figures 4.25 and 4.26, each with a different scale, show the bottom left part and the upper right part of the associated arrangement.

Figures 4.25 and 4.26 immediately show that in the characteristic part of the orbits digit 3 only occurs in isolation, in between two digits 2, while the sequence 2, 3, 2 only occurs in between two digits 1. The figures also show that digit 1 occurs only as an even multiple. What is does not show is the maximum number of successive ones in the characteristic part of the orbits. However, since

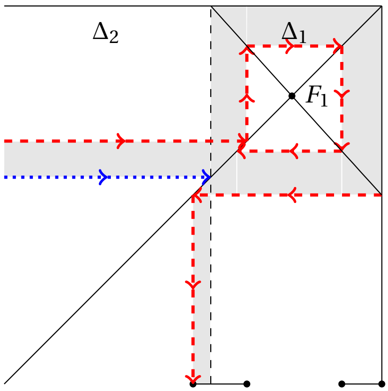
$$p_2 < T_\alpha^8(\alpha+1) < T_\alpha^4(\alpha+1) \quad \text{while} \quad T_\alpha(\alpha+1) < T_\alpha^{10}(\alpha+1) < p_2,$$

we conclude that digit 1 does not occur more than six times in succession. The following example illustrates the great similarity of the characteristic part of the orbits of two random numbers in case $N = 100$ and $\alpha = (\sqrt{281684} - 78)/53$.



$N = 100, \alpha = 8.5422\cdots$
$T_\alpha^2(\alpha) = p_2;\ b(N,\alpha) \approx 1.23$

Figure 4.25



$N = 100, \alpha = 8.5422\cdots$
$T_\alpha^2(\alpha) = p_2;\ b(N,\alpha) \approx 1.23$

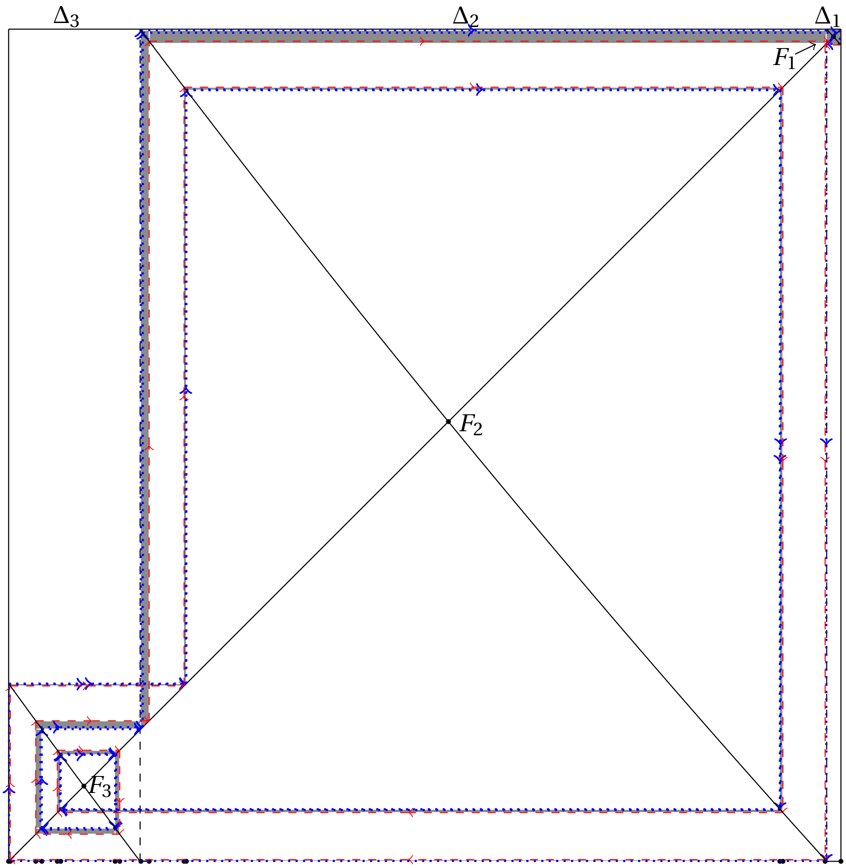
Figure 4.26

Example 4.

$$8.8 = [2,2,2,\underbrace{1,\dots,1}_{13\text{ times }1},2,3,2,1,1,1,1,1,1,2,3,2,1,1,1,1,2,3,2,1,1,1,1,2,3,2,1,1,2,3,2,1,1,\\ 1,1,1,1,2,3,2,1,1,2,3,2,1,1,1,1,2,3,2,\dots]_{100,(\sqrt{281684}-78)/53};$$

$$9.5 = [2,2,2,\underbrace{1,\dots,1}_{8\text{ times }1},2,3,2,1,1,2,3,2,1,1,1,1,2,3,2,1,1,2,3,2,1,1,2,3,2,1,1,1,1,1,1,2,3,2,\\ 1,1,1,1,1,1,2,3,2,1,1,2,3,2,1,1,2,3,2,\dots]_{100,(\sqrt{281684}-78)/53}$$

Whereas the previous arrangement is an interesting example of the way the endpoints of I_α determine the gaps and the characteristic parts of the orbits, the next arrangement shows the possible complexity of the occurrence of gaps. In Figure 4.27 we have $N = 100$ and $\alpha = 8.5215$. There is a very small, gapless cylinder set Δ_1, as opposed to cylinder sets Δ_2 and Δ_3, where gaps leave almost no room for the characteristic parts of the orbits. Note that this abundance of gaps goes with a relatively high steepness on the left of $\Upsilon_{100,8.5215}$, with $|T'_\alpha(8.5215)| = 1.3771\cdots$.



$N = 100, \alpha = 8.5215$

Figure 4.27

Finally, Figure 4.28 illustrates that two-cylinder cases may come with a large number of gaps. For convenience, we use the notation α_i for $T_\alpha^i(\alpha)$. In this case, the expanding power on I_α is very small: we have $|T'_\alpha(\alpha)| = 1.0203\cdots$ and $|T'_\alpha(\alpha+1)| = 1.0000\cdots$.

4

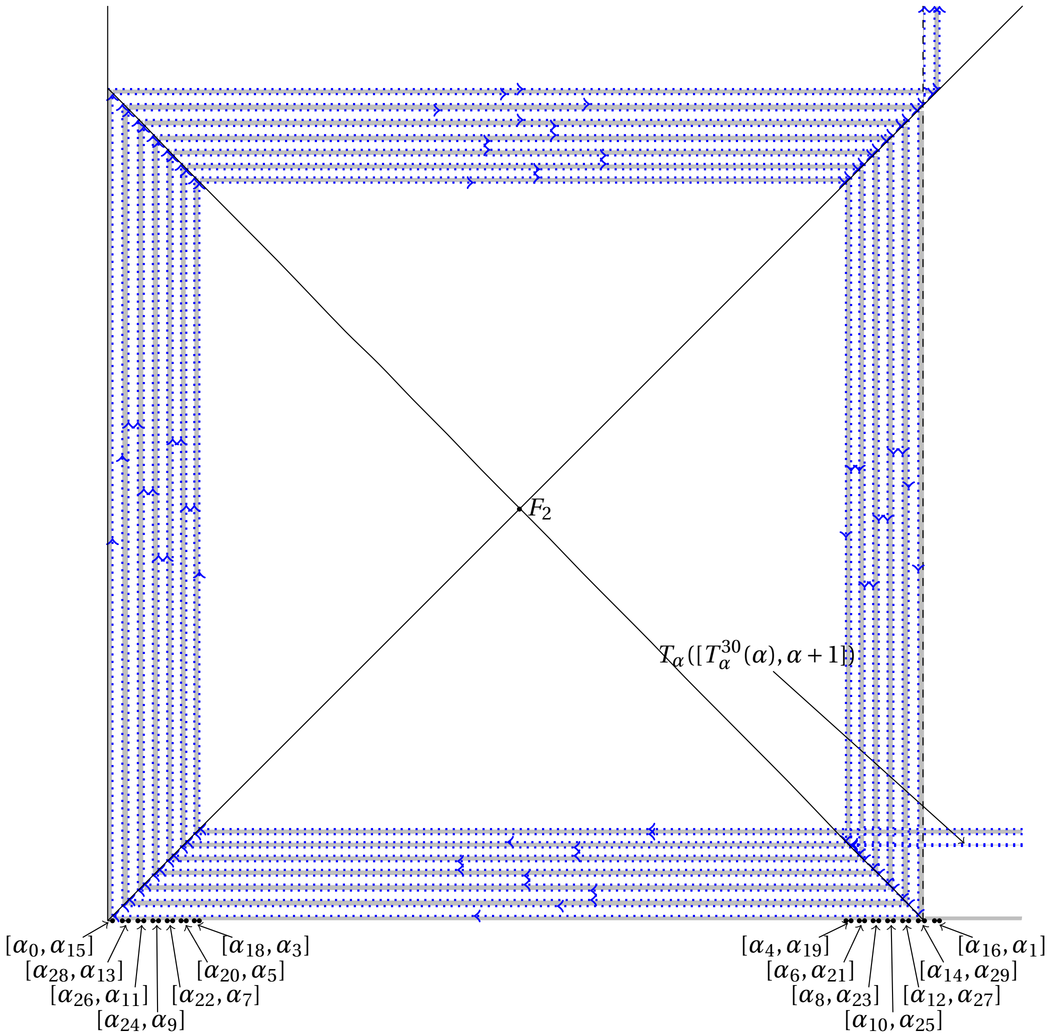


Figure 4.28: $N = 10,000; \alpha = 98.9995$ bottom left. The notation α_i is short for $T_\alpha^i(\alpha)$. The small gaps and the large gap in Δ_2 are visible as well as the leftmost, small gap of Δ_1 and part of the large gap containing f_1. Not visible is the one gap remaining, the one between $T_\alpha^{17}(\alpha)$ and $T_\alpha^{30}(\alpha)$.

ACKNOWLEDGEMENTS

The first author of this paper is very thankful to the Dutch organisation for scientific research NWO, which funded his research for this paper with a so-called *Promotiebeurs voor Leraren*, with grant number 023.003.036. We also want to thank Julian Lyczak, then working at Leiden University, for his help with some programming in the preparatory phase of Section 4.3. We also thank Pjotr Buys of the University of Amsterdam for his support in using Mathematica for Section 4.4. Finally, we thank Niels Langeveld for providing us with Figure 4.5.

REFERENCES

[1] M. Anselm, S.H. Weintraub, *A Generalization of continued fractions*, Journal of Number Theory 131 (12) (2011), 2442 – 2460.

[2] K. Dajani, C. Kraaikamp, N. van der Wekken, *Ergodicity of N-continued fraction expansions*, Journal of Number Theory 133 (9) (2013), 3183 – 3204.

[3] A. Haas, D. Molnar, *Metrical Diophantine approximation for continued fraction like maps of the interval*, Transactions of the American Mathematical Society, Vol. 356 (7) (2003), 2851 – 2870.

[4] S. Ito, S. Tanaka, H. Nakada, *On unimodal linear transformations and chaos I and II*, Tokyo J. Math. Vol. 2 (1979), 221 – 259.

[5] C. Kraaikamp, N. Langeveld, *Invariant measures for continued fraction algorithms with finitely many digits*, Journal of Mathematical Analysis and Applications 454 (1) (2017), 106 –126.

[6] L. Ma, R. Nair, *Haas-Molnar continued fractions and metric diophantine approximation*, Proceedings of the Steklov Institute of Mathematics, Vol. 299 (2017), 157 – 177.

[7] M.R. Palmer, *On the classification of measure preserving transformations of Lebesgue spaces*, thesis submitted for the degree of Doctor of Philosophy, Dept. of Mathematics, University of Warwick (1979).

[8] W. Parry, *Representations for real numbers*, Acta Math. Acad. Sci. Hungar. 15 (1964), 95 – 105.

[9] A. Rényi, *Representations for real numbers and their ergodic properties*, Acta Math. Acad. Sci. Hungar. 8 (1957), 477 – 493.

[10] V.A. Rohlin, *Exact endomorphisms of a Lebesgue space*, Izv. Akad. Nauk SSSR Ser. Mat. 25 (1961), 499 – 530 = Amer. Math. Soc. Transl. (2) 39 (1964) 1 – 36.

[11] T. Saito, *A chaos generator based on a quasi-harmonic oscillator*, IEEE Transactions on Circuits and Systems, Vol. Cas-32, 4 (1985), 320 – 331.

[12] S. Schweiger, *Ergodic theory of fibred systems and metric number theory*, Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1995.

[13] I. Shiokawa, *Ergodic properties of piecewise linear transformations*, Proc. Japan Acad. 46 (1970), 1122 – 1125.

[14] K.M. Wilkinson, *Ergodic Properties of a Class of Piecewise Linear Transformations*, Zeitschrift für Wahrscheinlichkeitstheorie verw. Gebiete 31 (1975), 303 – 328.

5

CONCLUSION

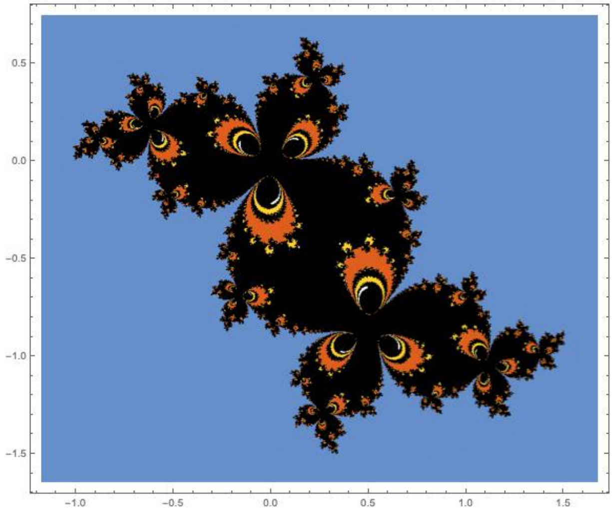
In the previous chapters, in which the results of my research are presented, no attention is given to any implications for society, i.e. to any technical applications or connections with other sciences. Still, the research findings are significant for society as well, because they extend the theory of a subject with many well known applications. Before going into this extension, I will mention some of these applications.

Probably the best known example is Huygens' planetarium. In the seventeenth century, the Dutch scientist Christiaan Huygens tried to build a planetarium. He wanted to approximate the ratios between the revolving times of planets around the sun in our solar system, so as to build gears with a limited amount of teeth that would still give a good impression of the planets revolving around the sun. For this, he used convergents (introduced in Chapter 1) of continued fractions of these ratios with strict, practical limits for the numerators and denominators.

Another example is electrical resistance in networks. Where a network of a series of two resistances R and S has resistance $R+S$, the resistance in a parallel network with resistances T and U is $1/(1/T+1/U)$. In some networks where series of resistances and parallel resistances exist, the network resistance can be expressed by using a regular continued fraction in which the partial quotients (also introduced in Chapter 1) are the individual resistances; see for instance [S].

The theory of continued fractions has also applications within mathematics, such as complex analysis. As an example, germs of holomorphic functions with linear part $e^{2\pi i\alpha}$ are linearisable if α is a *Brjuno number*; see https://en.wikipedia.org/wiki/Brjuno_number. The nice figure of a Julia Set illustrating this was composed by Han Peters from the University of Amsterdam.

A last example to be mentioned here is the theory of knots and tangles, see for in-



5

Figure 5.1: Julia Set

stance [C], where Conway describes tangles using continued fractions. He even makes use of singularisations, introduced in Chapter 2, to demonstrate the equivalence of tangles. This theory is used in, for instance, biomedical research on DNA; see [KL].

In Chapter 2 the theory on approximation coefficients is extended, which are useful for indicating the quality of approximation by convergents of continued fractions. The research sprang from the observation that there appeared to be differences in the asymptotic frequencies of orders of three successive approximation coefficients. In Chapter 2 it is shown how these frequencies can be calculated, not only in the case of the regular continued fraction, but also several other semi-regular cases. The conclusion is that in all cases there is a strong connection, sometimes even symmetrical, between the asymptotic frequencies of orders and their reverses, and that they are never evenly distributed. The results can be partly explained by properties of the natural extensions, partly by considering the continued fraction algorithms as S-expansions. More research on singularisation areas of continued fractions might contribute to understanding better the differences between the various distributions of the asymptotic frequencies of these continued fractions.

In Chapter 3 the interval for which the natural extension of Nakada's α-expansions is known is extended. This contributes to a better understanding of these expansions as a dynamical system, since it can be seen as the smallest invertible dynamical system containing T_α, the operator for α-expansions; see [DK].

In Chapter 4 a lot of results are presented on N-expansions. Whereas the approximating qualities of these expansions are poor, there is the appeal is that only finitely many different digits are involved. More than that, N-expansions give rise to very interesting dynamical systems. Orbits of N-expansions having only two, three or four different digits may have gaps that yield clear patterns regarding the order of the digits. Especially when all fixed points of the related operator are contained in gaps, these patterns are extremely similar for infinitely many different numbers.

REFERENCES

[C] J. Conway, *An enumeration of knots and links and some of their algebraic properties*. In J. Leech (ed.), Computational Problems in Abstract Algebra, Oxford, England: Pergamon Press, 329 – 358

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

[KL] L.H. Kauffmann, and S. Lambropoulou, *From Tangle Fractions to DNA*, Chelsea, New York, 2003.

[S] C. Sanna, *Continued Fractions: Introduction and Applications*, Proc. of the Roman Number Theory Association 2 (2017), no. 1, 61–81.

ACKNOWLEDGEMENTS

One motive for writing this thesis was justifying my presence in a world in which I met so many great and inspiring people: Sven Aerts, Sjoerd Beentjes, Frits Beukers, Wim Berkelmans, Raymond van Bommel, Tom van den Bosch, Alex van den Brandhof, Sam van den Brink, Pjotr Buys, Jan van de Craats, Karma Dajani, Birgit van Dalen, Jan Donkers, Dion Gijswijt, Helmuth Götz, Anne de Haan, KP Hart, Charlene Kalle, Geert Kapteijns, Larissa Krawczyk, Arno Kret, Niels Langeveld, Niek Lamoree, Andries Lenstra, Hendrik Lenstra, Julian Lyczak, Milan Lopuhaä, Hitoshi Nakada, Gerben Oling, Han Peters, Quintijn Puite, Frank Redig, Guus Regts, Huub Rutjes, Karandeep Singh, Melanie Steentjes, Peter Stevenhagen, Marco Swaen, Dirk Temme, Kees Temme, Samuel Tiersma, Allard Veldman, Yde Venema, Tom Verhoeff, Frank Welling, Jan Wiegerinck, Chris Zaal and Bart Zevenhek. They all taught, helped and inspired me, often beyond the realm of mathematics.

I did not yet mention Annemieke Schimmel, a math teacher just like me, whom I met at my beloved Gymnasium Felisenum, and who is now evolving as a great school leader. But I do not really thank her; I simply love her.

With regard to the writing of this thesis, two men stand out.

The first is of course Cor Kraaikamp, the perfect promotor: always patient, enthusiastic and encouraging about my endlessly repeated first steps, sharp as a razor when it came to dots, commas and other mathematics. Although I never thought it was worth much what I was doing, I was glad he noticed and praised my efforts. And wow - such good company he is! Thank you so much, Cor! Thanks a lot for so smoothly guiding me to the finish.

The second is Hendrik Jager, under whom I graduated almost thirty years ago, some years after he had been the supervisor of Cor's doctoral thesis. He was by far the best teacher I have ever had (and there have been many good ones). Had I not written a paper together with him, so long ago now, I would not even have considered writing this thesis. I think it is wonderful that more than twenty years after that first joint paper we wrote another one so as to obtain the 'NWO promotiebeurs voor leraren'.

Finally: thank you, KP, for helping me with actually composing this thesis according to the wishes of the TU Delft.

Curriculum Vitæ

Cornelis Jacobus de Jonge

25-09-1966	Born in Apeldoorn, The Netherlands.

Education

1984	Gymnasium-B, Pius X College, Beverwijk
1990	MSc Mathematics, University of Amsterdam
1991	MA Philosophy, University of Amsterdam
2006	MA Educational Management, Dutch School for Educational Management, Amsterdam

Work

1991	Teacher of Mathematics, Dr. Ariëns MAVO, Krommenie
1991 – 1993	Lecturer's assistant, University of Amsterdam
1993 – 2018	Teacher of Mathematics, Gymnasium Felisenum, Velsen-Zuid
2007 –	Lecturer, University of Amsterdam
2018 –	Teacher of Mathematics, Cygnus Gymnasium, Amsterdam

List of Publications

8. **J. de Jonge**, *Nederlands succes bij de IMC geen toevalstreffer*, Nieuw Archief voor Wiskunde (2018).

7. **J. de Jonge, C. Kraaikamp**, *Natural extensions for Nakada's α-expansions: Descending from 1 to g^2*, Journal of Number Theory, Volume 183, February 2018, Pages 172-212.

6. **J. de Jonge, C. Kraaikamp**, *Three consecutive approximation coefficients: asymptotic frequencies in semi-regular cases*, Tohoku Mathematical Journal, Volume 70, Number 2 (2018), 285-317.

5. **H. Jager, J. de Jonge**, *On the approximation by three consecutive continued fraction convergents*, Indagationes Mathematicae, Volume 25, Issue 4, 27 June 2014, Pages 816-824.

4. **J. de Jonge**, *Op zoek naar 2013*, Pythagoras (2014).

3. **J. de Jonge**, *Opkomst en ondergang van de rekentoets*, Euclides, jaargang 88 (2013).

2. **J. de Jonge**, *De boze buitenwereld*, Nieuw Archief voor Wiskunde (2000).

1. **H. Jager, J. de Jonge**, *The Circular Dispersion Spectrum*, Journal of Number Theory, Volume 49, Issue 3, December 1994, Pages 360-384.