Equidistribution of infinite interval substitution schemes, explicit resonances of Anosov toral maps, and the Hausdorff dimension of the Rauzy gasket

by

Benedict Adam Sewell

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy in Mathematics

Mathematics Institute

July 2021

Contents

List of Figures

Acknowledgments

First and foremost, I am indebted to my supervisor Mark Pollicott, and wish to thank him for his sustained benevolence, enthusiasm and sense of humour.

Thanks are also due to my family, particularly my mother Lise, to whom this thesis is dedicated, for their great support and encouragement over the years.

I would also like to thank David Tall, who has given a great deal of time and advice, in particular sharing many useful adages of the late great Christopher Zeeman.

Finally, I wish to thank all those people who made my time at Warwick a pleasant one. Special mention goes to Paul Colognese, Joe Manger and Marek Kurczynski.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented (including data generated) was carried out solely by the author (with the exception of cited material, i.e., results and historical background).

In particular, all figures were created by the author using TikZ-pgf with the exception of Figure 2.11, which was produced by the author using Wolfram Mathematica 11 (Student Edition).

Parts of this thesis have been submitted for publication by the author:

M Pollicott and B. Sewell. *An infinite interval version of the α-Kakutani equidistribution problem.* (2021). arXiv:2103.10235v1

Abstract

This thesis comprises three chapters:

In the first chapter, we consider an infinite analogue of the classical α-Kakutani equidistribution problem, and under mild assumptions, we prove results on uniform distribution and discrepancy, extending results of [34] and others.

In the second chapter, we provide explicit Ruelle resonances for three families of Anosov diffeomorphisms on the two-torus, following and extending results of [88].

In the third chapter, we show that the Hausdorff dimension of the Rauzy gasket is at most 1.7407, improving upon results of [10] and [47].

Chapter 1

Introduction

This thesis follows the succinct adage,

> "Three papers, a PhD",

attributed to Chistopher Zeeman [94]. More explicitly, the main body of this thesis comprises three chapters, representing expanded and expounded-upon versions of three articles.

The aim of this chapter is to present a broad overview of these three chapters. For more detail, we refer the reader to the chapters themselves.

1.1 Chapter 2: An infinite interval version of the α-Kakutani equidistribution problem

The study of uniform distribution in the unit interval has been an important area of interest for over a century. For example, it was shown by Weyl [100] that, for any irrational α, the sequence $x_n = \alpha n \pmod 1$ is uniformly distributed and Hardy and Littlewood showed that, for almost all $\lambda > 1$, the sequence $x_n = \lambda^n \pmod 1$ is uniformly distributed [54].

In chapter 2, we consider natural families of examples given by the endpoints of successively refined partitions of the interval. The historical example, due to Huzihiro Araki and Shizuo Kakutani, describes a process, for each $\alpha \in (0, 1)$ (which we will define more carefully below), in which one starts with the trivial partition and, at each stage, divides all subintervals of maximal length into two in the fixed ratio, $\alpha : 1 - \alpha$. For example, when $\alpha = \frac{1}{3}$, the first seven partitions are depicted in Figure 1.1.

In this setting, one has that the set of endpoints of the nth partition is uniformly distributed as $n \to \infty$.

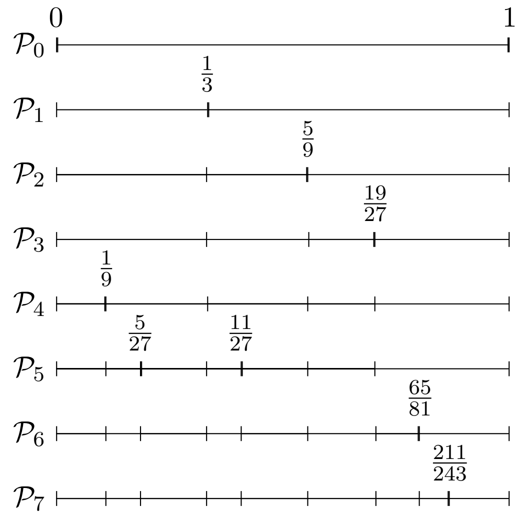


Figure 1.1: The first seven partitions $(\mathcal{P}_n)_{n=0}^{7}$ of the $\frac{1}{3}$-Kakutani interval substitution scheme.

More explicitly, in this chapter, extending a generalisation by Aliośa Volčič, we consider an infinite analogue of this interval splitting process, where one inserts infinitely many intervals at each stage. Taking a dynamical viewpoint inspired by iterated function systems, under suitable hypotheses we prove results on equidistribution and discrepancy of the endpoints of intervals which have been split. We also provide examples for which these points do not equidistribute, and begin to consider higher-dimensional analogues.

1.2 Chapter 3: Explicit examples of resonances for Anosov maps of the torus

In the study of chaotic diffeomorphisms, a natural class of examples are Anosov diffeomorphisms. In fact, it is the principle of the Cohen-Gallavotti chaotic hypothesis that chaotic behaviour can be understood through the dynamics of Anosov systems [27].

The study of Anosov dynamics is advanced by understanding various dynamical quantities, including the entropy and the resonances.

Given a map T, its resonances (assuming they are well-defined) comprise a sequence (either finite, empty, or converging to zero) of distinct complex numbers

$(\rho_n)_n$, which give all possible exponential decay rates for the correlation function

$$\int f \circ T^m g \, d\mu - \int f \, d\mu \int g \, d\mu,$$

whenever f and g are suitably smooth observables, and where μ is a particular measure (i.e., the absolutely continuous invariant probability measure in the expanding case or the SRB measure in the Anosov case).

Numerous results concerning resonances of maps of the torus exist (see [48] and references therein). However, there are very few examples of Anosov diffeomorphisms for which the resonances $(\rho_n)_n$ are actually known. For the trivial case of linear hyperbolic diffeomorphisms of the torus, such as the famous Arnol'd CAT map [5],

$$\begin{pmatrix} a \\ b \end{pmatrix} \mapsto \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \mod 1,$$

there are no (non-trivial) resonances. In the context of the two-torus,* one has only the striking work [88] of Oscar Bandtlow, Wolfram Just and Julia Slipantschuk, which provides a family of Anosov diffeomorphisms, B_λ, perturbing the above CAT map, for which the resonances $(\rho_n)_n$ are infinite and explicitly known:

$$\{\rho_n\} = \{\lambda^n, \overline{\lambda}^n \, : \, n \in \mathbb{N}_0\},$$

where λ is an arbitrary complex parameter with $|\lambda| < 1$.

The aims of chapter 3 are as follows:

- To broaden the number of examples of Anosov maps with explicit resonances.
- Moreover, to exhibit examples of resonances with more variety and structure.
- To give a simplification of the analysis of the above work, [88], by presenting a different viewpoint.

1.3 Chapter 4: A simple approach to bounding the Hausdorff dimension of the Rauzy gasket

The Rauzy gasket $\mathcal{G}$ is a self-projective fractal lying in a two-dimensional subset of $\mathbb{R}^3$, which is an important subset of parameter space arising in various settings in dynamics and topology (see [39] and references therein).

*On translation surfaces of higher genus, a complete description for resonances of linear pseudo-Anosov maps was recently given in [48].

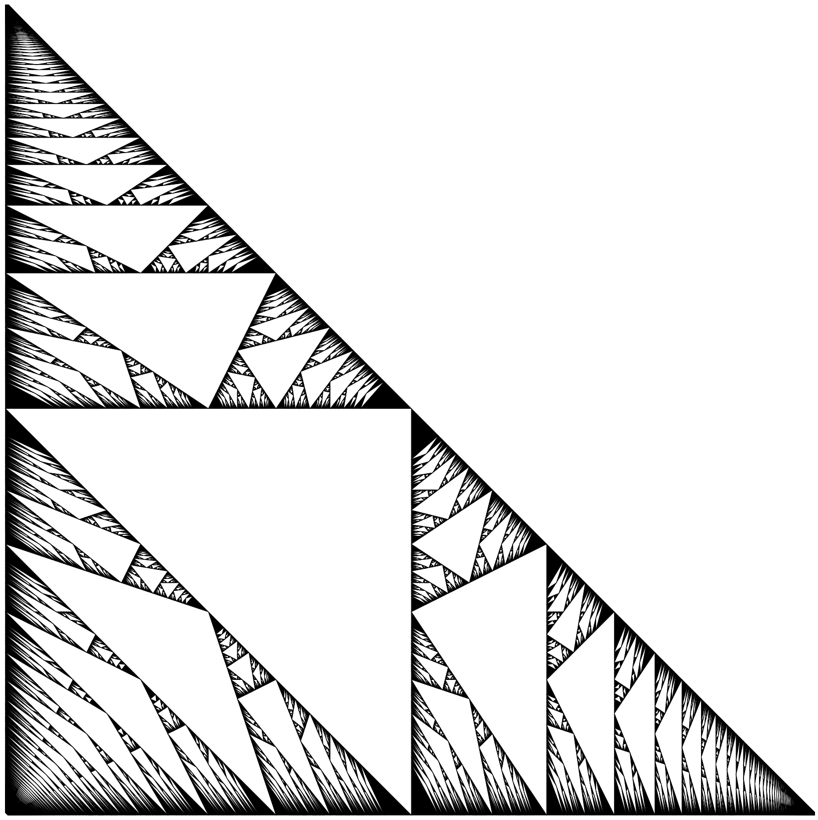


Figure 1.2: The Rauzy gasket, $\mathcal{G}$. (A right-angled realisation.)

Up to a change of variables, it can be considered as the limit set of three rational maps on the right-angled triangle, $\Delta' = \{(x,y) \in [0,1]^2 \mid x+y \leq 1\}$:

$$T_1(x,y) = \left(\frac{1}{2-x}, \frac{y}{2-x}\right), \qquad T_2(x,y) = \left(\frac{x}{2-y}, \frac{1}{2-y}\right),$$
$$T_3(x,y) = \left(\frac{x}{1+x+y}, \frac{y}{1+x+y}\right).$$

That is, $\mathcal{G}$ is the largest non-empty subset of Δ' (by inclusion) such that $\mathcal{G} = T_1(\mathcal{G}) \cup T_2(\mathcal{G}) \cup T_3(\mathcal{G})$. See Figure 1.2 for a depiction. We will later consider a more symmetric realisation: since these are bi-Lipschitz, they have the same Hausdorff dimension.

Estimating the Hausdorff dimension of a set from above is typically quite straightforward, since such estimates follow from presenting a well-chosen sequence of open covers. However, in this case, because the fixed points of the T_i (the vertices of Δ') are indifferent (i.e., the derivative is the identity matrix at the fixed point), obtaining viable upper bounds on the dimension of $\mathcal{G}$ proves to be highly non-trivial. The situation is further complicated by the system being non-conformal.

The two upper bounds known to date are those of Artur Avila, Pascal Hubert and Alexandra Skripchenko [10] ($\dim_H(\mathcal{G}) < 2$) and Charles Fougeron [47] ($\dim_H(\mathcal{G}) \leq 1.825$).

In chapter 4, using an elementary argument based on renewal theory, we give

1. a simple proof that $\dim_H(\mathcal{G}) < 2$, and
2. a proof that $\dim_H(\mathcal{G}) \leq 1.7407$.

Chapter 2

An infinite interval version of the α-Kakutani equidistribution problem

2.1 Introduction and history

Following the introduction in the previous chapter, we start with an important definition.

Definition 1 (Uniformly distributed, equidistributed)**.** We say that a sequence $(x_n)_{n=1}^{\infty}$ in the unit interval is *uniformly distributed* or *equidistributed* (as $n \to \infty$) if, for every $0 \leq a < b \leq 1$,

$$\frac{1}{N}\big|\{n = 1, 2, \ldots, N : x_n \in [a, b)\}\big| \to b - a$$

as $N \to \infty$, where $|\cdot|$ denotes the cardinality of a set. Equivalently, for all continuous functions $f : [0, 1] \to \mathbb{R}$,

$$\frac{1}{N}\sum_{n=1}^{N} f(x_n) \to \int_0^1 f(x)\,\mathrm{d}x$$

as $N \to \infty$. More generally, we say that an increasing sequence of finite subsets $(E_n)_{n=1}^{\infty}$ is uniformly distributed if

$$\frac{|E_n \cap [a, b)|}{|E_n|} \to b - a$$

as $n \to \infty$, i.e., for any continuous f,

$$\frac{1}{|E_n|} \sum_{y \in E_n} f(y) \to \int_0^1 f(x)\,\mathrm{d}x.$$

Example 1. A small selection of examples are given in the introductory chapter. In contrast, the sequence $\alpha \log(n+1) \pmod 1$ is not equidistributed, for any $\alpha \in \mathbb{R}$ [59].

As mentioned above, in this chapter we consider natural examples based on subdividing partitions of the interval. Before introducing the original motivating example, we first fix some terminology.

Definition 2 (Partition)**.** A *partition* $\mathcal{P}$ is a set of closed, positive-length intervals, which have pairwise disjoint interiors and cover $[0,1]$ up to a set of Lebesgue measure zero.

2.1.1 Equidistribution of a random interval-splitting process

In 1973, at a meeting in Oberwolfach, Huzihiro Araki posed the following problem to Shizuo Kakutani. (For more detail on this historical background, see [2].)

Conjecture 1 (Kakutani, Araki)**.** *Let x_1 be uniformly distributed in $[0,1]$. Given $(x_k)_{k=1}^n$, let x_{n+1} be uniformly distributed in the largest of the $n+1$ subintervals into which $x_1, \ldots, x_n$ subdivide $[0,1]$. Then $(x_n)_{n=1}^\infty$ is uniformly distributed almost-surely.*

A realisation of this random process is given in Figure 2.1. This conjecture was affirmed in 1978 independently by Jean-Claude Lootgier [65] and Willem Rutger van Zwet [96], using the same methods.

The theme of their proof, which is highly analogous to ours below, is that each interval appearing in the process (i.e., the interval between two adjacent x_k) is split according to the same law as the original interval, $[0,1]$.

This symmetry on multiple scales allows one to apply renewal theory to the relevant statistical quantities. The flexibility of this method subsequently lead to generalised results, where the maximal subinterval is split according to increasingly general classes of non-uniform probability laws (see [66, 67, 89, 90]).

Another interesting generalisation is to randomly choose which interval to split at each stage (then divide this interval uniformly). This was investigated in [19, 20], where it is shown that the sequence is equidistributed when the probability of choosing an interval is proportional to its length to the power $a > 0$.

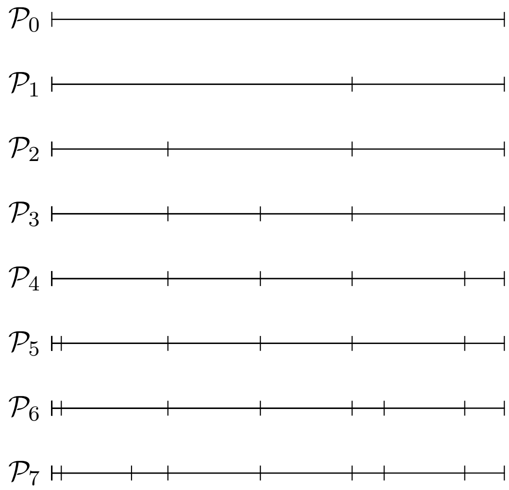


Figure 2.1: A sequence of partitions $\{\mathcal{P}_n\}_{n=0}^{\infty}$, where $\mathcal{P}_{n+1}$ is obtained by splitting the maximal subinterval in $\mathcal{P}_n$ randomly, uniformly along its length.

2.1.2 The original α-Kakutani equidistribution result

Kakutani's contribution to the above was to prove an elegant equidistribution result for a deterministic analogue of this random process. We give a description of this process, which we generalise in the next section.

Definition 3 (α-Kakutani scheme)**.** For a fixed $0 < \alpha < 1$, the *α-Kakutani scheme* is a sequence of partitions $(\mathcal{P}_n)_{n=0}^{\infty}$ defined inductively:

- $\mathcal{P}_0 = \big\{[0,1]\big\}$ is the trivial partition; and
- $\mathcal{P}_{n+1}$ is obtained from $\mathcal{P}_n$ by taking *each* interval of maximal length and subdividing it into two smaller intervals in the ratio $\alpha : 1 - \alpha$.

Example 2. Figure 1.1 in the introduction shows the first seven partitions for the choice $\alpha = 1/3$. Notice that $\mathcal{P}_5$ is obtained by splitting two maximal length intervals in $\mathcal{P}_4$ simultaneously (each of length 2/9). By contrast, the choice $\alpha = 1/2$ gives the trivial dyadic splitting.

Consider the set of endpoints at the nth stage of this process, $E_n = \bigcup_{I \in \mathcal{P}_n} \partial I$. Kakutani's result is the following.

Theorem (Kakutani, [58])**.** *For all $\alpha \in (0,1)$, the set E_n is equidistributed as $n \to \infty$.*

An alternative proof was given by Roy Adler and Leopold Flatto in [2].

The Kakutani-Fibonacci sequence

The choice of $\alpha = \phi := \frac{1}{2}(\sqrt{5}-1)$, the reciprocal of the golden ratio, is an interesting one, which gives rise to the so-called *Kakutani-Fibonacci sequence* of partitions. This particular instance of the above theorem itself received a recent, dynamical proof in [25].

Indeed, for F_n denoting the nth Fibonacci number, a simple induction shows that the nth partition in the ϕ-Kakutani scheme comprises F_n "short" intervals of length ϕ^{n+1}, and F_{n+1} "long" intervals of length ϕ^n, which, coding each short interval with an S and an L respectively, are arranged according to the nth Fibonacci word W_n: $W_0 = \text{L}$, $W_1 = \text{LS}$, and, for $n \geq 2$, W_n is the concatenation $W_n = W_{n-1}W_{n-2}$; see Figure 2.2 for a depiction.

W_n in particular comprises the first F_{n+1} symbols of the infinite Fibonacci word. This is gives an example of a Sturmian word, an infinite sequence on finitely many symbols of minimal complexity (see [46, Ch. 6] for a full definition, and the historical connection to cutting sequences on the torus). In particular, analysis of gap distributions for the nth set of endpoints corresponds to a study of subwords of W_n.

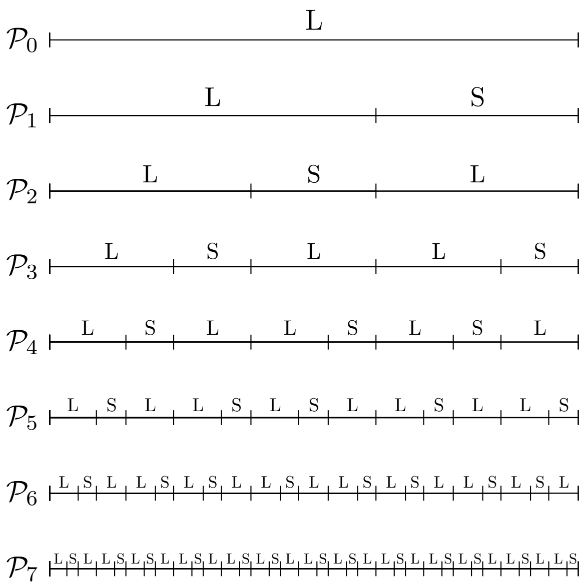


Figure 2.2: The sequence of long and short intervals in the first seven partitions, $\{\mathcal{P}\}_{n=0}^{7}$, of the ϕ-Kakutani scheme, a.k.a., the Fibonacci-Kakutani sequence.

Besides from the trivial dyadic splitting, this gives the simplest example of an LS-sequence, which we briefly introduce in the next subsection, and of *rank one* examples, which shall be defined in section 2.3.

2.1.3 Interval substitutions using multiple intervals

A natural generalisation of the α-Kakutani scheme, introduced by Aliośa Volčič in [98], is to alter the above process by splitting intervals of maximal length according to a fixed, finite partition comprising $N \geq 2$ intervals, say. That is, at each stage, one splits all intervals of maximal length into N pieces whose lengths (arranged from left to right) have a certain fixed ratio, $\alpha_1 : \alpha_2 : \cdots : \alpha_N$, where the α_i sum to 1.

Example 3. In Figure 2.3, we have the first seven partitions of the interval substitution scheme in which one splits maximal intervals according to the partition $\{[0, \frac{1}{2}], [\frac{1}{2}, \frac{2}{3}], [\frac{2}{3}, 1]\}$, i.e., with ratio $\frac{1}{2} : \frac{1}{6} : \frac{1}{3}$. By contrast, the α-Kakutani scheme corresponds to splitting according to the partition $\{[0, \alpha], [\alpha, 1]\}$.

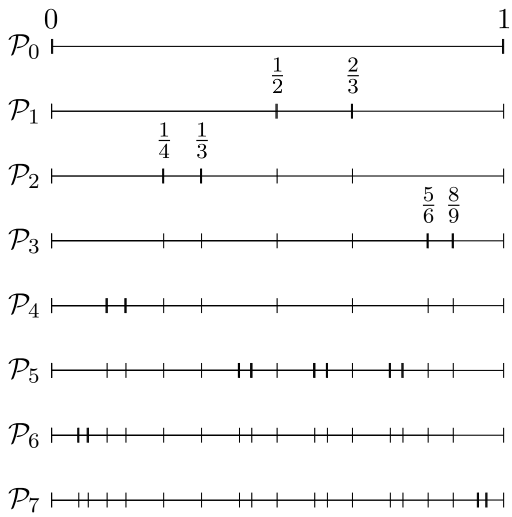


Figure 2.3: The first seven partitions $(\mathcal{P}_n)_{n=0}^{7}$ of the interval substitution scheme where one splits maximal-length intervals according to the partition $\mathcal{P}_1 = \{[0, \frac{1}{2}], [\frac{1}{2}, \frac{2}{3}], [\frac{2}{3}, 1]\}$.

A special case which we have already mentioned is that of LS-sequences, where for $L, S \in \mathbb{N}$ the partition $\mathcal{P}_1$ comprises L intervals of length x, and S intervals of length x^2, where $x > 0$ satisfies $Lx + Sx^2 = 1$.

These received a lot of recent attention [4, 21, 22, 23, 57, 99], particularly in connection to low-discrepancy sequences and β-adic van der Corput sequences in particular. (We will describe this connection in more detail at the beginning of section 2.4.)

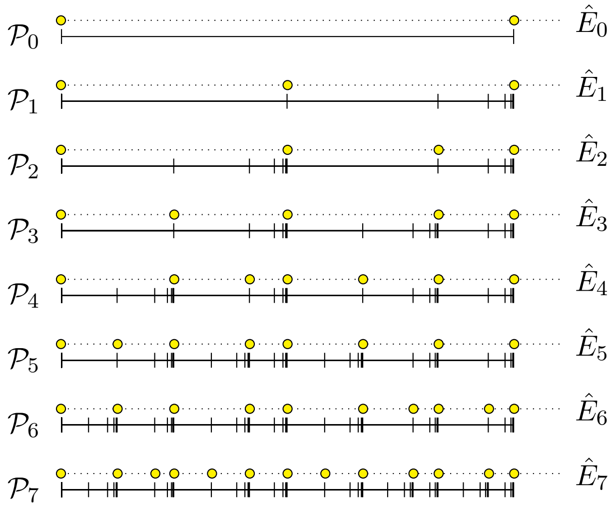


Figure 2.4: An illustration of $(\mathcal{P}_n)_{n=0}^{7}$ and $(\hat{E}_n)_{n=0}^{7}$ for the example generated by the partition $\mathcal{P} = \{[1, \frac{1}{2}]\} \cup \{[1 - \frac{1}{2} \cdot 3^{-n}, 1 - \frac{1}{6} \cdot 3^{-n}]\}_{n=0}^{\infty}$. Here the tick marks (which accumulate on certain points in the interval) denote the elements of E_n and the suspended yellow circles denote the elements of $\hat{E}_n$.

2.1.4 Interval substitutions using infinitely many intervals

In this chapter, we continue the process and ask in what way the result above still holds if at every stage we insert an *infinite* partition $\mathcal{P}$ into each maximal-length subinterval.

Consider $\hat{E}_n$, the (finite) set of endpoints of those intervals which have been split up to the $(n+1)$-st stage:

$$\hat{E}_n := \left\{ \min(I),\ \max(I)\ :\ I \in \bigcup_{i=0}^{n} \mathcal{P}_i \setminus \mathcal{P}_{n+1} \right\}.$$

Example 4. In Figure 2.4 we depict $\mathcal{P}_n$ and $\hat{E}_n$ for the infinite substitution scheme generated by $\mathcal{P} = \{[0, \frac{1}{2}]\} \cup \{[1 - \frac{1}{2} \cdot 3^{-n}, 1 - \frac{1}{6} \cdot 3^{-n}]\}_{n=0}^{\infty}$.

Example 5. One can construct many more exotic examples. For instance, one could let the intervals in $\mathcal{P}$ be the closures of the connected components of the complement of the middle-third Cantor set.

For simplicity, we restrict our attention to the set of left endpoints, which we shall denote by L_n, although we could equally well have chosen the right endpoints, midpoints, etc.

Our main result is the following generalization of Kakutani's equidistribution theorem. Throughout this chapter, let $\|I\|$ denote the length of an interval I.

Theorem 1. *Let $\mathcal{P}$ be a countable partition. Then, provided that*

$$-\sum_{I\in\mathcal{P}} \|I\| \log \|I\| < \infty,$$

the set L_n is uniformly distributed as $n \to \infty$.

Finally, we note that our interest in this problem, and the starting point for our analysis, began with the very elegant work of Yotam Smilansky [91].

Outline of the chapter

In section 2.2, we give a new dynamical viewpoint of the problem. In section 2.3, we apply renewal theory to prove Theorem 1 in two cases (rank one and higher rank). In section 2.4, we use a generating function to estimate the discrepancy in the rank one case. In section 2.5, we use methods of analytic number theory to estimate the discrepancy in the higher rank case, with a generic Diophantine-type assumption. In section 2.6, we provide examples for which the assumption in Theorem 1 is false and for which the conclusion is false and true, respectively. In section 2.7, we generalise Theorem 1 to an abstract setting, which we apply to a two-dimensional example. Finally, in section 2.8, we make some concluding remarks.

2.2 Partitions and similarities

Our approach to Theorem 1 is to express the elements of the partition $\mathcal{P}$ in terms of the images of similarities. The refinements into finer partitions, $\mathcal{P}_n$, can then be expressed in terms of words formed from the index set of $\mathcal{P}$. That is, $\mathcal{P} = \{T_i[0,1]\}_{i\in\mathcal{I}}$, where each $T_i : [0,1] \to [0,1]$ is an orientation preserving similarity with contraction ratio $\alpha_i > 0$. We see that $\mathcal{P}$ being a partition is equivalent to the following:

- $T_i[0,1) \cap T_j[0,1) = \emptyset$ for $i \neq j$; and
- $\displaystyle\sum_{i\in\mathcal{I}} \alpha_i = 1.$

For illustration, we give the following example:

Example 6. Given $0 = t_0 < t_1 < \cdots$ with $(t_n) \to 1$, the partition $\{[t_{n-1}, t_n]\}_{n\in\mathbb{N}}$ is equal to $\{T_n[0,1]\}_{n\in\mathbb{N}}$, where

$$T_n(x) = (t_n - t_{n-1})x + \sum_{k=0}^{n-1} t_n.$$

Example 7. In particular, setting $t_n = 1 - \frac{1}{6}3^{-n}$ for $n \geq 1$ gives rise to $\mathcal{P}_1$ in Figure 2.4.

We now explain how this can be used to give an explicit description of the splitting process.

Definition 4 ((T_i)-refinement)**.** Given a partition $\mathcal{P} = \{S_k[0,1]\}_k$ where the $\{S_k\}_k$ are orientation preserving similarities, and $\{T_i\}_i$ a collection of orientation preserving similarities of $[0,1]$ such that $\{T_i[0,1]\}_i$ is a partition, the (T_i)-refinement of $\mathcal{P}$ is obtained by taking all intervals of maximal length in $\mathcal{P}$ and replacing them by subintervals in the following manner: if $S[0,1] \in \mathcal{P}$ has maximal length in $\mathcal{P}$, then it is replaced by the elements of the set

$$\{S \circ T_i[0,1] \mid i \in \mathcal{I}\}.$$

Definition 5 (Interval substitution scheme)**.** The interval substitution scheme generated by $\{T_i\}_{i\in\mathcal{I}}$ is the sequence of partitions $(\mathcal{P}_n)_{n=0}^{\infty}$ defined as follows:

- $\mathcal{P}_0$ is the trivial partition, $\mathcal{P}_0 = \big\{[0,1]\big\}$; and
- $\mathcal{P}_{n+1}$ is the (T_i)-refinement of $\mathcal{P}_n$.

This gives a convenient presentation of the partitions.

Example 8. The α-Kakutani scheme is the interval substitution scheme generated by the pair $T_1 : x \mapsto \alpha x$ and $T_2 : x \mapsto (1-\alpha)x + \alpha$.

Example 9. Similarly, the interval substitution scheme generated by the triple $T_1 : x \mapsto x/2$, $T_2 : x \mapsto (x+3)/6$ and $T_3 : x \mapsto (x+2)/3$ gives the sequence of partitions depicted in Figure 2.3.

We now associate to the sequence of partitions $(\mathcal{P}_n)_{n=0}^{\infty}$ a sequence of families of left endpoints of split intervals, $(L_n)_{n=0}^{\infty}$.

Definition 6 (L_n)**.** Given an interval substitution scheme $(\mathcal{P}_n)_{n=0}^{\infty}$ generated by similarities $(T_i)_{i\in\mathcal{I}}$, we define the finite sets L_n ($n \geq 0$) to be

$$L_n = \bigcup_{k=0}^{n} \bigcup_{I\in\mathcal{P}_k\setminus\mathcal{P}_{k+1}} \min(I).$$

Remark 1. One can consider a generalisation of the above process by dropping the assumption that the $(T_i)_i$ have to be affine. We will not consider this more general setting, but note it is easy to give superficial examples where E_n and L_n are not uniformly distributed: take, for example, $T_1(x) = \sqrt{x}/2$, $T_2(x) = (x+1)/2$.

Considering the interval substitution scheme generated by $\{T_i\}_{i\in\mathcal{I}}$, it follows inductively that an interval appears at some stage in the scheme if and only if it is obtained by applying a sequence of maps from $\{T_i\}_i$ to $[0,1]$, and so each is naturally described by a finite word in $\mathcal{I}$. It is convenient to introduce the following notation.

Definition 7 ($W(\mathcal{I})$, $*$, $\alpha_{\boldsymbol{v}}$, $T_{\boldsymbol{v}}$)**.** Given a countable set $\mathcal{I}$, the word set $W(\mathcal{I})$ is the semigroup consisting of all words in $\mathcal{I}$: i.e.,

$$W(\mathcal{I}) = \{\emptyset\} \cup \bigcup_{n=1}^{\infty} \mathcal{I}^n,$$

where $\emptyset$ denotes the *empty word* (unique word of length zero), and the semigroup operation $* : W(\mathcal{I}) \times W(\mathcal{I}) \to W(\mathcal{I})$ denotes concatenation of words, for which $\emptyset$ acts as the identity:

$$(n_1, \ldots, n_k) * (m_1, \ldots, m_j) = (n_1, \ldots, n_k, m_1, \ldots, m_j); \qquad \boldsymbol{v} * \emptyset = \emptyset * \boldsymbol{v} = \boldsymbol{v}.$$

Furthermore, for ease of notation, we extend the definitions of α_i and T_i to the whole of $W(\mathcal{I})$: For the word $\boldsymbol{v} = (i_1, \ldots, i_k) \in \mathcal{I}^k$, define

$$\alpha_{\boldsymbol{v}} := \prod_{j=1}^{k} \alpha_{i_j}, \qquad T_{\boldsymbol{v}} := T_{i_1} \circ \ldots \circ T_{i_k},$$

and also define $\alpha_\emptyset = 1$ and $T_\emptyset = \mathrm{Id}_{[0,1]}$.

To paraphrase the above, any closed interval I appears in the substitution scheme if and only if $I = T_{\boldsymbol{v}}[0,1]$ for some word $\boldsymbol{v} \in W(\mathcal{I})$. It consequently has length $\alpha_{\boldsymbol{v}}$, and will be split between $\mathcal{P}_n$ and $\mathcal{P}_{n+1}$ (i.e., $I \in \mathcal{P}_n \setminus \mathcal{P}_{n-1}$) precisely when n satisfies

$$\alpha_{\boldsymbol{v}} = \max_{J\in\mathcal{P}_n} \big\{ \|J\| \big\};$$

consequently, its left endpoint $T_{\boldsymbol{v}}(0)$ will appear in L_n, if not already present in L_{n-1}.

Rather than using $\{L_n\}_{n\geq 0}$ to parametrise this process, we want to reparameterise this family to reflect the lengths of the maximal intervals, and rewrite it as

$\{X_\lambda\}$ as follows.

Definition 8 (X_λ)**.** For $\lambda > 1$, let $X_\lambda = \emptyset$, and for $\lambda \in (0, 1]$, let

$$X_\lambda := L_{n(\lambda)}, \qquad \text{where} \qquad n(\lambda) := \max\{n \geq 0 \,:\, \exists I \in \mathcal{P}_n : \; \|I\| \geq \lambda\},$$

i.e., given $\lambda \leq 1$, $P_{n(\lambda)+1}$ is the first partition in the process for which all intervals have lengths strictly smaller than λ. From the previous discussion, one obtains a convenient, dynamical formula for X_λ:

$$X_\lambda = \big\{T_{\boldsymbol{v}}(0) \,:\, \boldsymbol{v} \in W(\mathcal{I}),\, \alpha_{\boldsymbol{v}} \geq \lambda\big\}.$$

As $\lambda \to 0^+$, $n(\lambda) \to \infty$ and so the uniform distribution of L_n as $n \to \infty$ is equivalent to that of X_λ as $\lambda \to 0^+$. This is in turn equivalent to the weak-$*$ convergence of the probability measure μ_λ defined by

$$\mu_\lambda = \frac{1}{|X_\lambda|} \sum_{x \in X_\lambda} \delta_x,$$

to the Lebesgue measure (henceforth denoted as Leb in this chapter) as $\lambda \to 0^+$, where δ_x denotes the Dirac delta measure at x, and throughout, $|X|$ denotes the cardinality of a countable set X.

2.3 Proof of Theorem 1

This section is devoted to proving the main result, which we can conveniently rephrase in the following way.

Theorem 1. *Provided that* $-\sum_{I \in \mathcal{P}} \|I\| \log \|I\| < \infty$*, the measure* μ_λ *converges to* Leb *in the following sense: for any interval* $J \subset [0, 1]$*, we have* $\mu_\lambda(J) \to \|J\|$*.*

We first prove the convergence for a given interval of the form $T_{\boldsymbol{v}}[0, 1)$. For each of these elementary sets, their μ_λ-measure is intimately related to the asymptotics of $|X_\lambda|$ as $\lambda \to 0^+$.

To proceed, we can relate the cardinality of X_λ to that of the set of words,

$$A_\lambda := \{\boldsymbol{w} \in W(\mathcal{I}) \mid \alpha_{\boldsymbol{w}} \geq \lambda\},$$

and exploit a natural renewal equation involving $|A_\lambda|$.

Lemma 1. *One of the following two cases hold. Either* $T_1(0) = 0$ *for some (unique)* $1 \in \mathcal{I}$*, in which case, for all* $\lambda > 0$*,*

$$|X_\lambda| = |A_\lambda| - |A_{\lambda/\alpha_1}|; \tag{2.1}$$

or no element of $\{T_i\}_{i\in\mathcal{I}}$ *fixes 0, and* $|X_\lambda| = |A_\lambda|$ *for all* $\lambda > 0$*.*

Proof of Lemma 1. Let $\boldsymbol{w}, \boldsymbol{v} \in A_\lambda$ satisfy $T_{\boldsymbol{w}}(0) = T_{\boldsymbol{v}}(0)$. If $\boldsymbol{w}, \boldsymbol{v}$ are both non-empty, denote the first symbol by $i, j \in \mathcal{I}$ respectively (i.e., $\boldsymbol{w} = (i, \ldots)$). Then, since

$$T_i[0,1) \ni T_{\boldsymbol{w}}(0) = T_{\boldsymbol{v}}(0) \in T_j[0,1),$$

the disjointness of $\{T_i[0,1)\}_{i\in\mathcal{I}}$ tells us that $i = j$. Since T_i is injective, we have that, for $\boldsymbol{w} = i * \hat{\boldsymbol{w}}$ and $\boldsymbol{v} = i * \hat{\boldsymbol{v}}$,

$$T_{\hat{\boldsymbol{w}}}(0) = T_i^{-1} T_{\boldsymbol{w}}(0) = T_i^{-1} T_{\boldsymbol{v}}(0) = T_{\hat{\boldsymbol{v}}}(0),$$

and we can repeat the above process iteratively until $\boldsymbol{w}$ or $\boldsymbol{v}$ is empty. This implies, without loss of generality, that $\boldsymbol{w} = \boldsymbol{v} * \boldsymbol{j}$. In particular, $\boldsymbol{j} \in W(\mathcal{I})$ satisfies $T_{\boldsymbol{j}}(0) = T_{\boldsymbol{v}}^{-1} T_{\boldsymbol{w}}(0) = 0$. This gives rise to two cases.

Case 1: *There is no symbol* $i \in \mathcal{I}$ *for which* $T_i(0) = 0$. Since the T_i are orientation preserving, this is equivalent to $0 \notin \bigcup_{i\in\mathcal{I}} T_i[0,1)$. By the above logic, $\boldsymbol{j} = i * \hat{\boldsymbol{j}}$ for $i \in \mathcal{I}$ gives the contradiction $0 = T_{\boldsymbol{j}}(0) \in T_i[0,1)$. Hence $j = \emptyset$ and $\boldsymbol{w} = \boldsymbol{v}$. In sum, $\boldsymbol{v} \mapsto T_{\boldsymbol{v}}(0)$ is a bijection $A_\lambda \to X_\lambda$.

Case 2: *There exists* $1 \in \mathcal{I}$ *be such that* $T_1(0) = 0$. We note that this must occur if $\mathcal{I}$ is finite, and in any case, 1 is unique by the disjointness property. Then, if $\boldsymbol{j} \neq \emptyset$, writing $\boldsymbol{j} = i * \hat{\boldsymbol{j}}$ as before, one has $0 = T_{\boldsymbol{j}}(0) \in T_i[0,1)$, hence $i = 1$.

Reducing inductively as before, this shows that $\boldsymbol{j}$ is non-empty if and only if it is a tuple of 1's: $\boldsymbol{j} \in \{1\}^k \subset \mathcal{I}^k$ for some $k \in \mathbb{N}$.

Applying the above gives that, for each $y \in X_\lambda$, there is a shortest word $\boldsymbol{v}_0(y) \in W(\mathcal{I})$ satisfying

- $T_{\boldsymbol{v}_0(y)}(0) = y$; and
- $T_{\boldsymbol{v}}(0) = y \implies \boldsymbol{v} = \boldsymbol{v}_0(y) * \boldsymbol{j}$, for some $\boldsymbol{j} \in \{\emptyset\} \cup \{1\}^k$, $k \in \mathbb{N}$.

therefore, $T_{\boldsymbol{v}}(0) = y$ implies $\alpha_{\boldsymbol{v}} = \alpha_{\boldsymbol{v}_0(y)} \alpha_1^k$, for some $k \in \mathbb{N}_0$.

In particular, $y \in X_\lambda \iff \lambda \leq \alpha_{\boldsymbol{v}_0(y)}$, and in that case, there is exactly one element of $A_\lambda \setminus A_{\lambda/\alpha_1}$ which gets mapped onto y. Therefore, $\boldsymbol{v} \mapsto T_{\boldsymbol{v}}(0)$ this time is a bijection $A_\lambda \setminus A_{\lambda/\alpha_1} \to X_\lambda$. (2.1) follows, completing the proof. $\square$

To continue the proof of the theorem, we now use the same ideas to express the μ_λ measure of a (half-open) interval appearing in the substitution scheme in terms of a ratio involving $|X_\lambda|$.

Lemma 2. *For all $\boldsymbol{v} \in W(\mathcal{I})$ and $\lambda \in (0,1]$,*

$$|T_{\boldsymbol{v}}[0,1) \cap X_{\lambda\alpha_{\boldsymbol{v}}}| = |X_\lambda|. \tag{2.2}$$

In particular, for all $\lambda \in (0, \alpha_{\boldsymbol{v}}]$,

$$\mu_\lambda(T_{\boldsymbol{v}}[0,1)) = \frac{|X_{\lambda/\alpha_{\boldsymbol{v}}}|}{|X_\lambda|}. \tag{2.3}$$

Moreover, for $\lambda > \alpha_{\boldsymbol{v}}$,

$$T_{\boldsymbol{v}}[0,1) \cap X_\lambda = \{T_{\boldsymbol{v}}(0)\} \cap X_\lambda = \begin{cases} 1, & \text{if } \alpha_{\boldsymbol{v}_0} \geq \lambda > \alpha_{\boldsymbol{v}}; \\ 0, & \text{if } \lambda > \alpha_{\boldsymbol{v}_0}; \end{cases}$$

where $\boldsymbol{v}_0$ is the shortest word in $W(\mathcal{I})$ such that $T_{\boldsymbol{v}_0}(0) = T_{\boldsymbol{v}}(0)$. In particular, if $\boldsymbol{v} = \emptyset$, or $\boldsymbol{v} = \boldsymbol{w} * i$ with $T_i(0) \neq 0$, $\boldsymbol{v}_0 = \boldsymbol{v}$ and the previous expression is zero for all $\lambda > \alpha_{\boldsymbol{v}}$.

Remark 2. The last part of the lemma is necessary for when we later estimate the discrepancy and does not play a role in the current proof.

Proof of Lemma 2. Fix $\boldsymbol{v} \in W(\mathcal{I})$ and $\lambda \leq 1$, and consider (2.2). Suppose that $\boldsymbol{w} \in A_{\lambda\alpha_{\boldsymbol{v}}}$ and $T_{\boldsymbol{w}}(0) \in T_{\boldsymbol{v}}[0,1)$ (i.e., $T_{\boldsymbol{w}}[0,1)$ meets $T_{\boldsymbol{v}}[0,1)$). From the argument presented in the previous lemma, either $\boldsymbol{w} = \boldsymbol{v} * \boldsymbol{j}$ or $\boldsymbol{v} = \boldsymbol{w} * \boldsymbol{j}$, for some $\boldsymbol{j} \in W(\mathcal{I})$.

The second option, $\boldsymbol{v} = \boldsymbol{w} * \boldsymbol{j}$, with $\boldsymbol{j} \neq \emptyset$, gives a contradiction:

$$\alpha_{\boldsymbol{w}} = \frac{\alpha_{\boldsymbol{v}}}{\alpha_{\boldsymbol{j}}} > \alpha_{\boldsymbol{v}} \geq \lambda\alpha_{\boldsymbol{v}} \geq \alpha_{\boldsymbol{w}}.$$

Thus, the first option is necessary. It is also sufficient, since $T_{\boldsymbol{v}}T_{\boldsymbol{j}}(0) \in T_v[0,1)$ for any $\boldsymbol{j} \in W(\mathcal{I})$. This gives the following equality of discrete sets.

$$T_{\boldsymbol{v}}[0,1) \cap X_{\lambda\alpha_{\boldsymbol{v}}} = \{T_{\boldsymbol{v}}T_{\boldsymbol{j}}(0) \,:\, \boldsymbol{j} \in A_\lambda\} = T_{\boldsymbol{v}}(X_\lambda).$$

By injectivity of $T_{\boldsymbol{v}}$, this set bijectively corresponds to X_λ, giving (2.2). Thereafter, taking (2.2), with $\lambda/\alpha_{\boldsymbol{v}}$ in place of λ, dividing through by $|X_\lambda|$ yields (2.3).

To consider the last part of the lemma, now let $\lambda > 1$. This time, for $\boldsymbol{w} \in A_{\lambda\alpha_{\boldsymbol{v}}}$ and $T_{\boldsymbol{w}}(0) \in T_{\boldsymbol{v}}[0,1)$ as above, the first option, $\boldsymbol{w} = \boldsymbol{v} * \boldsymbol{j}$ gives the contradiction, for any $\boldsymbol{j} \in W(\mathcal{I})$:

$$\alpha_{\boldsymbol{w}} = \alpha_{\boldsymbol{v}}\alpha_{\boldsymbol{j}} \leq \alpha_{\boldsymbol{n}} < \lambda\alpha_{\boldsymbol{v}} \leq \alpha_{\boldsymbol{w}}.$$

Hence, $\boldsymbol{v} = \boldsymbol{w} * \boldsymbol{j}$, where $\boldsymbol{j} \neq \emptyset$. In particular, $T_{\boldsymbol{w}}(0) \in T_{\boldsymbol{w}}T_{\boldsymbol{j}}[0,1)$, so $0 \in T_{\boldsymbol{j}}[0,1)$, or equivalently, $T_{\boldsymbol{j}}(0) = 0$. As in the previous proof, this implies that there is $1 \in \mathcal{I}$ with $T_1(0) = 0$, and that $\boldsymbol{j} \in \{1\}^k$ for some k. In which case, $T_{\boldsymbol{w}}(0) = T_{\boldsymbol{w}}T_{\boldsymbol{j}}(0) = T_{\boldsymbol{v}}(0)$. Thus,

$$T_{\boldsymbol{v}}[0,1) \cap X_{\lambda\alpha_{\boldsymbol{v}}} \subset \{T_{\boldsymbol{v}}(0)\}.$$

Now, if $\boldsymbol{v} = (i_1, \ldots, i_n)$, then

$$\boldsymbol{v}_0 = (i_1, \ldots, i_k), \qquad \text{where} \qquad k := \max\{1 \leq \hat{k} \leq n : i_{\hat{k}} \neq 1\},$$

or $\emptyset$ if no such maximum exists (or if $\boldsymbol{v} = \emptyset$). Then, by the previous discussion, $\boldsymbol{v}_0$ is the shortest word such that $T_{\boldsymbol{v}_0}(0) = T_{\boldsymbol{v}}(0)$, and this point lies in X_λ precisely when $\lambda \leq \alpha_{\boldsymbol{v}_0}$. The final remark follows from the definition of $\boldsymbol{v}_0$: it equals $\boldsymbol{v}$ precisely when 1 does not exist or $i_n \neq 1$ (or $\boldsymbol{v} = \emptyset$). □

The significance of relating $|X_\lambda|$ to $|A_\lambda|$ will now become clear from the following renewal equation. As we shall see subsequently, renewal equations are a natural basis for establishing asymptotic formulae.

Lemma 3. *The following holds for all $\lambda > 0$.*

$$|A_\lambda| = \sum_{i \in \mathcal{I}} |A_{\lambda/\alpha_i}| + \chi_{\{\lambda \leq 1\}}, \tag{2.4}$$

where χ is the indicator function. Equivalently, the following renewal equation applies, for all $t \in \mathbb{R}$, where $\psi(t) := e^{-t}|A_{e^{-t}}|$:

$$\psi(t) = \sum_{i \in \mathcal{I}} \alpha_i \psi\big(t + \log(\alpha_i)\big) + e^{-t}\chi_{\{t \geq 0\}}. \tag{2.5}$$

Proof of Lemma 3. We only prove (2.4), since (2.5) follows trivially. Partitioning non-empty words in A_λ according to their first symbol gives the following disjoint union:

$$A_\lambda \setminus \{\emptyset\} = \bigcup_{i \in \mathcal{I}} \underbrace{\{i * \boldsymbol{v} \in W(\mathcal{I}) \mid v \in W(\mathcal{I}),\ \alpha_{\boldsymbol{v}} \leq \lambda/\alpha_i\}}_{\text{in bijection with } A_{\lambda/\alpha_i}}.$$

The fact that the ith factor is in bijection with A_{λ/α_i} gives rise to the sum in (2.4).

Moreover since $\alpha_\emptyset = 1$, $\emptyset \in A_\lambda$ if and only if $\lambda \leq 1$, which gives rise to the indicator term in (2.4) and completes the proof. □

To make use of this renewal equation, just as in [3, 34, 91], it is necessary to consider two cases which behave somewhat differently. These cases correspond to, for example, the α-Kakutani schemes for $\alpha = 1/3$ and $\alpha = 1/2$, as described in the introduction.

Definition 9 (Rank)**.** For $n \in \mathbb{N}$ we will say the collection $\{\alpha_i\}_{i\in\mathcal{I}}$ is *rank* n if the smallest additive subgroup of $\mathbb{R}$ containing the set $\{-\log(\alpha_i)\}_{i\in\mathcal{I}}$ is isomorphic to $\mathbb{Z}^n$. If $\{\alpha_i\}_i$ is not rank n for any $n \in \mathbb{N}$ we will say $\{\alpha_i\}_{i\in\mathcal{I}}$ is *infinite rank*. Also, whenever $\{\alpha_i\}_i$ is not rank one, we say it is *higher rank*.

Example 10. The following examples illustrate the different ranks:

1. $\{1/2^n\}_{n\in\mathbb{N}}$ is rank one;
2. $\{1/2\} \cup \{1/3^n\}_{n\in\mathbb{N}}$ is rank two;
3. $\{1/2\} \cup \{1/3\} \cup \{1/7^n\}_{n\in\mathbb{N}}$ is rank three; and
4. $\{1/n^s\}_{n=2}^\infty$ is infinite rank, where $s \approx 1.728$ satisfies $\zeta(s) = 2$ (here ζ denotes the Riemann zeta function).

2.3.1 Uniform distribution in the rank one case

In this subsection, we concentrate on the rank one case, also known as the *arithmetic*, *rationally-related* or *commensurable* case (see [3, 34, 91]). The characteristic feature of this case is that the contraction ratios are all powers of a common number, x: $\{\alpha_i\}_{i\in\mathcal{I}} \subset \{x^n\}_{n\in\mathbb{N}}$. Thereby, fixing the minimal such $x > 0$, and writing $N_k = |\{i \in \mathcal{I} \mid \alpha_i = x^k\}|$, the renewal equation for $z(n) := x^n|A_{x^n}|$ reads $z(0) = 1$ and, for all $n \in \mathbb{N}$,

$$z(n) = \sum_{k=1}^{n} N_k x^k z(n-k).$$

To this sequence, one can apply the following renewal theorem, first proved in [41]. In chapter 4, we shall describe a more general renewal theorem by William Feller [45, Theorem 1, p.330] which has a simple proof. That said, the following historical version is sufficient here.

Lemma 4 (Erdős–Feller–Pollard renewal theorem)**.** *Suppose that* $(\lambda_k)_{k=1}^\infty$ *satisfies* $\lambda_k \geq 0$, $\sum_{k=1}^\infty \lambda_k = 1$, *and that the smallest subgroup of* $\mathbb{Z}$ *containing* $\{k \in \mathbb{Z} \mid \lambda_k >$

$0\}$ *is* $\mathbb{Z}$ *itself. Then, for the sequence* $(u_n)_{n=0}^{\infty}$ *defined by* $u_0 = 1$ *and*

$$u_n = \sum_{k=1}^{n} \lambda_k u_{n-k},$$

we have the limit

$$(u_n)_{n=0}^{\infty} \to \left(\sum_{k=1}^{\infty} k\lambda_k \right)^{-1}$$

(by convention, if the sum on the right hand side is infinite, the limit is zero).

We apply this theorem to $u_n = z(n)$ to give the following corollary. To quickly check the assumptions: Firstly, if $\{k : N_k > 0\}$ was supported on a subgroup $n\mathbb{Z}$ for $n > 2$, it would follow that each α_i is a power of x^n, contradicting minimality of x. Moreover, $\sum_k k\lambda_k$ here equals

$$\sum_{k=1}^{\infty} kN_k x^k = \sum_{i\in\mathcal{I}} \alpha_i \log_x(\alpha_i) = \left(-\sum_{i\in\mathcal{I}} \alpha_i \log(\alpha_i) \right) / (-\log(x)),$$

which is finite by assumption. Thus, the lemma applies to give the first part of the following corollary.

Lemma 5. *Suppose that* $\{\alpha_i\}_{i\in\mathcal{I}}$ *is rank one, that* x *is the minimal positive number for which* $\{\alpha_i\}_{i\in\mathcal{I}} \subset \{x^n\}_{n\in\mathbb{N}}$*, and that* $H := -\sum_i \alpha_i \log(\alpha_i) < \infty$*. Then*

$$x^n |A_{x^n}| \to \frac{-\log(x)}{H} \tag{2.6}$$

as $n \to \infty$*. Consequently, for all* $\boldsymbol{v} \in W(\mathcal{I})$*,*

$$\mu_{x^n}\big(T_{\boldsymbol{v}}[0,1)\big) \to \alpha_{\boldsymbol{v}} = \|T_{\boldsymbol{v}}[0,1)\| \tag{2.7}$$

as $n \to \infty$*.*

Proof of Lemma 5. As shown in the preceding discussion, (2.6) follows from Lemma 4 applied to $\big(z(n)\big)$ above. To show how the second limit (2.7) follows as a consequence, fix $\boldsymbol{v} \in W(\mathcal{I})$ and let $\alpha_v = x^m$. Also set $C = \log(x)/H$ here for convenience. We have the following two cases, which complete the proof:

Case 1: *No* T_i *fixes* 0*, for any* $i \in \mathcal{I}$*.* Then, applying Lemma 2 followed by Lemma 1 gives, for all $n \in \mathbb{N}$,

$$\mu_{x^n}(T_{\boldsymbol{v}}[0,1)) = \frac{|X_{x^{n-m}}|}{|X_{x^n}|} = \frac{|A_{x^{n-m}}|}{|A_{x^n}|} \sim \frac{Cx^{m-n}}{Cx^{-n}} = x^m = \alpha_{\boldsymbol{v}}$$

as $n \to \infty$.

Case 2: $T_1(0) = 0$ *for some* $1 \in \mathcal{I}$. Then, for $\alpha_1 = x^j$, the same lemmas apply to give, for all $n \geq m$,

$$\begin{aligned}\mu_{x^n}\big(T_{\boldsymbol{v}}[0,1)\big) = \frac{|X_{x^{n-m}}|}{|X_{x^n}|} = \frac{|A_{x^{n-m}}| - |A_{x^{n-m-j}}|}{|A_{x^n}| - |A_{x^{n-j}}|} &= \frac{x^n|A_{x^{n-m}}| - x^n|A_{x^{n-m-j}}|}{x^n|A_{x^n}| - x^n|A_{x^{n-j}}|} \\ &\to \frac{Cx^m - Cx^{m+j}}{C - Cx^j} \\ &= x^m = \alpha_{\boldsymbol{v}}\end{aligned}$$

as $n \to \infty$. $\square$

In sum, we have shown, under the assumptions of the theorem in the rank one case, that one has the required equidistribution on intervals of the form $T_{\boldsymbol{v}}[0,1)$. It is then a simple matter to conclude equidistribution (in this case) on an arbitrary interval by packing it with subintervals of this form.

Remark 3. Considering the above proof, the same conclusion holds if

$$\frac{z(n+1)}{z(n)} \to 1$$

as $n \to \infty$, in the notation of Lemma 4. In section 2.6 below, we discuss various hypotheses which guarantee this condition, which (as we shall see) is sufficient to give equidistribution.

The following conclusion to the proof of Theorem 1 will be written in the continuous limit, $\lambda \to 0^+$, so as to align it with the higher rank case. In this case, since we naturally have

$$\{\alpha_{\boldsymbol{v}} \mid \boldsymbol{v} \in W(\mathcal{I})\} \subset \{x^n \mid x \in \mathbb{N}\},$$

$\lambda \mapsto |A_\lambda|$ is necessarily constant on intervals of the form $[x^{n+1}, x^n)$, and hence the discrete limits of the previous lemma ((2.6) and (2.7)) extend to a continuous limits, e.g.,

$$\mu_\lambda\big(T_{\boldsymbol{v}}[0,1)\big) \to \alpha_{\boldsymbol{v}}$$

as $\lambda \to 0^+$, so it makes no difference to this case.

Proof of Theorem 1. Let $I \subset [0,1]$ be an interval, and let $\alpha_* = \max_{i\in\mathcal{I}}(\alpha_i)$. Fixing $n \in \mathbb{N}$, let

$$\mathcal{U}_n := \big\{U = T_{\boldsymbol{v}}[0,1) \mid \boldsymbol{v} \in \mathcal{I}^n,\ U \subset I\big\}.$$

Since the $\{T_i[0,1)\}_{i\in\mathcal{I}}$ are pairwise disjoint, it follows inductively that the intervals in U are pairwise disjoint. Therefore,

$$\sum_{U\in\mathcal{U}_n} \mu_\lambda(U) \le \mu_\lambda(I),$$

with a similar equality for Lebesgue measure. We now show that

$$\sum_{\boldsymbol{v}\in\mathcal{U}_n} \|T_{\boldsymbol{v}}[0,1)\| \ge \|I\| - 2\alpha_*^n. \tag{2.8}$$

Let $x \in I \setminus \bigcup_{U\in\mathcal{U}_n} U$. Then, one of the following two cases hold.

Case 1: $x \in K_n := [0,1] \setminus \bigcup_{\boldsymbol{v}\in\mathcal{I}^n} T_{\boldsymbol{v}}[0,1)$. Then, since the Lebesgue measure of the complement is given by

$$\sum_{\boldsymbol{v}\in\mathcal{I}^n} \|T_{\boldsymbol{v}}[0,1)\| = \sum_{\boldsymbol{v}\in\mathcal{I}^n} \alpha_{\boldsymbol{v}} = \sum_{(i_1,\ldots,i_n)\in\mathcal{I}^n} \prod_{k=1}^{n} \alpha_{i_k} = \Big(\sum_{i\in\mathcal{I}} \alpha_i\Big)^n = 1,$$

we have that $\mathrm{Leb}(K_n) = 0$.

Case 2: $x \in T_{\boldsymbol{v}}[0,1)$ *for some* $\boldsymbol{v} \in \mathcal{I}^n$. Then, since $T_{\boldsymbol{v}}[0,1) \not\subset I$, it must meet an endpoint of I. Hence, there are at most two $\boldsymbol{v} \in \mathcal{I}^n$ (i.e., $\alpha_{\boldsymbol{v}} \le \alpha_*^n$) with this property. Hence, the union over elements of this case is at most two intervals, each with length at most α_*^n. This proves (2.8).

Proceeding with the proof, since $\mu_\lambda(U) \to \|U\|$ for each $U \in \mathcal{U}_n$, the monotone convergence theorem implies that

$$\liminf_{\lambda\to 0^+} \mu_\lambda(I) \ge \liminf_{\lambda\to 0^+} \sum_{U\in\mathcal{U}} \mu_\lambda(U) = \sum_{U\in\mathcal{U}} \|U\| \ge \|I\| - 2\alpha_*^n.$$

Repeating this argument for $[0,1] \setminus I$ (i.e., considering an analogous collection of intervals contained wholly in $[0,1] \setminus I$) gives the converse inequality,

$$\limsup_{\lambda\to 0^+} \mu_\lambda(I) = 1 - \liminf_{\lambda\to 0^+} \mu_\lambda\big([0,1] \setminus I\big) \le \|I\| + 2\alpha_{\max}^n,$$

and the proof is completed by taking $n \to \infty$. □

2.3.2 Uniform distribution in the higher rank case

In the remaining, generic case, the proof is very similar to the above, but requires a higher rank version of the renewal theorem. The first of its kind was proved by

Blackwell in [17]. His, like most proofs, is of a probabilistic nature. The following interesting two-sided version has an analytic proof in [80], attributed to Karlin. In particular, it can be considered a consequence of the Ikehara–Wiener Tauberian theorem.

Lemma 6 ([80, Theorem 9.15])**.** *Suppose that μ is a discrete Borel measure on $\mathbb{R}$ such that*

1. *μ is not supported on $k\mathbb{Z}$, for any k;*
2. *$\int_{\mathbb{R}} |x| \, \mathrm{d}\mu(x) < \infty$; and*
3. *$H = \int_{\mathbb{R}} x \, \mathrm{d}\mu(x) \neq 0$.*

Suppose also that $f : \mathbb{R} \to \mathbb{R}$ is integrable and satisfies $f(x) \to 0$ as $x \to \pm\infty$, and that $\psi : \mathbb{R} \to \mathbb{R}$ is a bounded solution to the renewal equation

$$\psi(x) = f(x) + \int_{\mathbb{R}} \psi(x - y) \, \mathrm{d}\mu(y).$$

Then the limits $\lim_{x\to\pm\infty} \psi(x)$ exist, and

$$\lim_{x\to\infty} \psi(x) - \lim_{x\to-\infty} \psi(x) = \frac{1}{H} \int_{\mathbb{R}} f(y) \, \mathrm{d}y.$$

We now apply the lemma in our particular case. In view of Lemma 3, we have $\psi(x) = e^{-x}|A_{e^{-x}}|$, $f(x) = e^{-x}\chi_{\{x\geq 0\}}$, and

$$\mu = \sum_{i\in\mathcal{I}} \alpha_i \delta_{-\log(\alpha_i)},$$

which is discrete, and evidently supported on a discrete lattice if and only if the α_i are rank one. Moreover,

$$\int_{-\infty}^{\infty} |x| \, \mathrm{d}\mu(x) = \int_{-\infty}^{\infty} x \, \mathrm{d}\mu(x) = -\sum_{i\in\mathcal{I}} \alpha_i \log(\alpha_i)$$

is finite and non-zero, and the conditions on f hold vacuously, with $\int_{\mathbb{R}} f = 1$. Finally, for boundedness of ψ, since $\psi(x) = 0$ for $x < 0$, it suffices to show that, recalling $\alpha_* = \max_i(\alpha_i)$, the sequence $(M_n)_n$ defined by

$$M_n := \max_{[0,-n\log(\alpha_*))} \psi$$

is bounded in n. This is a simple consequence of the renewal equation. Let $x \in [-n\log(\alpha_*), -(n+1)\log(\alpha_*))$. Then, for each $i \in \mathcal{I}$, $x + \log(\alpha_i) \le x + \log(\alpha_*) < -n\log(\alpha_*)$, and thus

$$\begin{aligned}\psi(x) &= \sum_{i\in\mathcal{I}} \alpha_i \psi\big(x + \log(\alpha_i)\big) + e^{-x} \\ &\le \sum_{i\in\mathcal{I}} \alpha_i M_n + e^{n\log(\alpha_*)} \\ &\le M_n + \alpha_*^n.\end{aligned}$$

I.e., $M_{n+1} \le M_n + \alpha_*^n$. Therefore, M_n is uniformly bounded:

$$M_n \le \sum_{k=0}^{n-1} \alpha_*^n < \frac{1}{1-\alpha_*} < \infty.$$

Lemma 6 thus applies to give the first limit of the following lemma. The second limit then follows from this and Lemmas 1 and 2, just like in the proof of Lemma 5.

Lemma 7. *Suppose that* $\{\alpha_i\}_i$ *is not rank one and that* $H = -\sum_i \alpha_i \log(\alpha_i)$ *is finite. Then*

$$\lambda|A_\lambda| \to \frac{1}{H}$$

as $\lambda \to 0^+$. *Consequently, for all* $\boldsymbol{v} \in W(\mathcal{I})$,

$$\mu_\lambda(T_{\boldsymbol{v}}[0,1)) \to \alpha_{\boldsymbol{v}} = \|T_{\boldsymbol{v}}[0,1)\|$$

as $\lambda \to 0^+$.

The conclusion of the proof of Theorem 1 in this case is as given previously on page 21.

2.4 Discrepancy estimates

In [98], Volčič both generalised the method used by Adler and Flatto in [2] to general finite partitions, and posed questions which inspired various other papers. Of particular interest here is the question of *discrepancy*.

Discrepancy of sequences and sets

The general study of discrepancies is of particular importance to numerical estimation of intervals, particularly in high dimensions, of bounded variation functions.

In one dimension, this can be seen from the important *Koksma inequality*: For any function $f : [0,1] \to \infty$ of bounded variation, i.e.,

$$V(f) := \sup\left\{\sum_{k=0}^{n} |x_{k+1} - x_k| \,\middle|\, n \in \mathbb{N},\ 0 = x_0 < x_1 < \cdots < x_{n+1} = 1\right\} < \infty,$$

and $X \subset [0,1]$ finite, one has that

$$\left|\frac{1}{|X|}\sum_{x\in X} f(x) - \int_0^1 f(x)\mathrm{d}x\right| \le V(f)D^*(X),$$

where $D^*(X)$ is the (star-)*discrepancy* of X,

$$D^*(X) := \max_{b\in[0,1]} \big||X \cap [0,b)| - b\big|.$$

(Remarkably, this inequality is sharp, even for smooth functions.)

In this one-dimensional setting, one has explicit lower bounds on discrepancy [74]: for any $X \subset [0,1]$, $D^*(X) \ge \frac{1}{2}|X|^{-1}$. Moreover, perhaps surprisingly, there exists a universal constant C such that, for any sequence $(x_k)_k \subset [0,1]$,

$$D^*\big((x_k)_{k=1}^n\big) \ge C\frac{\log(n)}{n}$$

for infinitely many n. Any sequence (x_n) for which the left hand side, $D^*\big((x_k)_{k=1}^n\big)$, decays at this optimal rate is known as a *low-discrepancy sequence.* Classical examples are given by van der Corput sequences [62, p.127].

Example 11 (van der Corput)**.** Given $b \in \mathbb{N}_{\ge 2}$, let

$$x_n^{(b)} := \sum_{k=0}^{K} d_k(n)b^{-k-1},$$

where $d_k(n) \in \{0,\dots,b-1\}$ and we have the unique representation

$$n = \sum_{k=0}^{K} d_k(n)b^k.$$

For example,

$$x_n^{(2)} = \left(\frac{1}{2},\frac{1}{4},\frac{3}{4},\frac{1}{8},\frac{5}{8},\frac{3}{8},\frac{7}{8},\frac{1}{16},\frac{9}{16},\frac{5}{16},\frac{13}{16},\dots\right).$$

We note that there is an analogue, the β-adic van der Corput sequences, where $b = \beta > 1$ need not be integer (see, e.g., [23] and references therein).

Interval substitution schemes naturally provide increasing sequences of sets E_n (of endpoints) which equidistribute. For the sake of numerical integration, equidistributing sequences of points are a more natural object of study, and so one might consider constructing such a sequence (x_n) from the (E_k) by enumerating the points in $E_{k+1} \setminus E_k$ for each k. However, such enumerations do not always lead to equidistributing sequences.* For example, the naive left-to-right enumeration of new points (i.e., those in $E_{n+1} \setminus E_n$) fails to be uniformly distributed, even for the trivial dyadic splitting, as we now show explicitly:

Example 12. Let $(x_n)_{n=1}^{\infty} = (\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{3}{8}, \frac{5}{8} \ldots)$ be given by

$$x_n := \frac{2(n-2^k)+1}{2^{k+1}}, \qquad \text{if } 2^k \leq n < 2^{k+1},\ k \in \mathbb{N}_0.$$

Consider the proportion of x_n which lie in $[0, \frac{1}{2}]$. Since $x_n \leq \frac{1}{2}$ whenever $k \geq 1$ and

$$2^k \leq n \leq 2^k + 2^{k-1} = 3 \cdot 2^{k-1},$$

and $x_n > \frac{1}{2}$ if $3 \cdot 2^{k-1} \leq n < 2^k$; one has, for each $N \in \mathbb{N}$,

$$\begin{aligned} \frac{\left|\left\{1 \leq n \leq 3 \cdot 2^{N-1} : x_n \leq \frac{1}{2}\right\}\right|}{3 \cdot 2^{N-1}} &= \frac{1}{3 \cdot 2^{N-1}} \left|\left\{1 \leq n < 2^{N+1} : x_n \leq \frac{1}{2}\right\}\right| \\ &= \frac{1}{3 \cdot 2^{N-1}} \left(1 + \sum_{k=1}^{N} \left|\left\{2^k \leq n < 2^{k+1} : x_n \leq \frac{1}{2}\right\}\right|\right) \\ &= \frac{1}{3 \cdot 2^{N-1}} \left(1 + \sum_{k=1}^{N} 2^{k-1}\right) \\ &= \frac{2^N}{3 \cdot 2^{N-1}} = \frac{2}{3} \not\to \frac{1}{2}. \end{aligned}$$

This shows that (x_n) is not equidistributed.

On the other hand, there are positive results which indicate that the situation is generically more favourable. Although it is not directly applicable, we cannot help but mention a classical result of von Neumann, which says that any dense set subset of the interval can be enumerated so that the resultant sequence is uniformly distributed. In [98], Volčič proves the following advancement of this result:

Proposition 1 (Volčič)**.** *Given any increasing sequence of finite subsets of the interval $(E_n)_{n=1}^{\infty}$ which is uniformly distributed, the sequence $(x_n)_{n=1}^{\infty}$ obtained by ran-*

*Indeed, this is why one insists that maximal intervals are all split simultaneously in the definition of interval substitution schemes.

domly ordering the points in $E_k \setminus E_{k-1}$ is uniformly distributed as $n \to \infty$, almost-surely.

Not only is this true, but in the very specific context of LS-sequences, Ingrid Carbone in [22] shows that a van der Corput–style enumeration gives rise to low-discrepancy sequences in certain cases, as per the following result.

Proposition 2 (Carbone)**.** *Given $L, S \in \mathbb{N}_0$, there exists an enumeration of the endpoints appearing in the corresponding LS-sequence, $(x_n)_n$, with the following discrepancies:*

- *If $S \leq L$, there exists C such that $D^*((x_n)_{n=1}^N) \leq C \log(N)/N$ for all $N \in \mathbb{N}$.*
- *If $S = L + 1$, there exists C such that $D^*((x_n)_{n=1}^N) \leq C(\log(N))^2/N$ for all $N \in \mathbb{N}$.*
- *If $S \geq L + 2$, there exists C such that $D^*((x_n)_{n=1}^N) \leq C \log(N) N^{-\log_x(S)-1}$ for all $N \in \mathbb{N}$, where $x \in (0, 1)$ satisfies $Sx^2 + Lx = 1$.*

Beyond the very explicit structure exhibited by LS-sequences, Maria Rita Iaco and Volker Ziegler in [57] provide more general classes of low-discrepancy sequences arising from (finite) rank one examples.

Naturally, our extension to the infinite setting is not necessarily applicable to the applications-focused search for computable low-discrepancy sequences, and we henceforth focus on the discrepancy of (X_λ) as a function of λ.

To this end, the most general results to date are due to Michael Drmota and Maria Infusino in [34], extending the work of Carbone in [21], which we paraphrase with the following:

Proposition 3 (Drmota–Infusino)**.** *Suppose that $\mathcal{P}$ is a finite partition, corresponding to the set of lengths $\{\alpha_k\}_{k=1}^n$. Then, we have the following two cases:*

If $\{\alpha_k\}_k$ is rank one, there exist $C, \beta, d > 0$ such that, for $x > 0$ the maximal value such that $\{\alpha_1, \dots \alpha_n\} \subset \{x^k : k \in \mathbb{N}\}$,

$$D^*(X_{x^n}) \leq \begin{cases} Cn^d x^{\beta n}, & \text{if } \eta \leq 1; \\ Cx^n, & \text{if } \eta > 1; \end{cases}$$

Moreover, these estimates are (generically) optimal: if $\eta \leq 1$ there exists a constant C' such that $D^*(X_{x^n}) > C' n^d x^{\beta n}$ for infinitely many n.

If $n = 2$ and $\{\alpha_1, \alpha_2\}$ is higher rank and moreover if

$$\gamma = \frac{\log(\alpha_1)}{\log(\alpha_2)}$$

is Diophantine (a.k.a., 0-badly approximable, as defined prior to Theorem 3 below), there exists $C > 0$ such that

$$D^*(X_\lambda) \leq C \left(\frac{-\log(\lambda)}{\lambda} \right)^{\frac{1}{4}}$$

for all $\lambda > 0$. Alternatively, if γ is algebraic, one has an inequality like the previous, with the exponent replaced with some positive, computable constant.

In this section, we in turn extend this result to the context of infinite interval substitution schemes. As above the results obtained here are different, depending on whether we are in the rank one case or the higher rank case.

Remark 4. An interesting connection between this chapter and chapter 4 is that the 7th volume of Uniform Distribution Theory (which we cite with [34]) is dedicated to the life of Gérard Rauzy, after whom the Rauzy gasket is named, and contains an article on his contribution to that field [64].

2.4.1 Discrepancy estimates in the rank one case

In this section, we extend the analysis of the rank one case to estimate the discrepancy between the measure μ_λ and the Lebesgue measure. More precisely, we have the following result.

Theorem 2. *Suppose that*

1. *$\{\alpha_i\}_{i\in\mathcal{I}}$ is rank one,*
2. *$x > 0$ is the smallest number for which $\{\alpha_i\}_{i\in\mathcal{I}} \subset \{x^n\}_{n\in\mathbb{N}}$, and*
3. *there is some $\varepsilon > 0$ for which $\sum_{i\in\mathcal{I}} \alpha_i^{1-\varepsilon} < \infty$.*

Then there is an R^* (made explicit in Lemma 8 below) such that, for all $\rho \in \big(x/R^*, 1\big) \subset (x^{-\varepsilon}, 1)$, there is a constant $C > 0$ such that, for all $n \in \mathbb{N}$ and all intervals $I \subset [0, 1]$,

$$\big|\, \mu_{x^n}(I) - \|I\| \,\big| \leq C\rho^n.$$

I.e., $D^*(X_{x^n}) \leq C\rho^n$ for all $n \in \mathbb{N}$.

The proof of Theorem 2 begins with the following lemma, which in particular defines R^* in terms of a generating function for $|A_{x^n}|$.

Lemma 8. *Given $\{\alpha_i\}_{i\in\mathcal{I}}$ as in Theorem 2, the function formally defined by*

$$g(z) = (z - x) \sum_{n=0}^{\infty} |A_{x^n}| z^n$$

entends to a holomorphic function on the open disk of radius R^* about 0, where, for $\alpha_j = x^{n_j}$,

$$R^* := \max\left\{ R \in \big(x, \min(x^{1-\varepsilon}, 1)\big] \;\middle|\; |z| = R \implies \sum_{j\in\mathcal{I}} z^{n_j} \neq 1 \right\}.$$

Therefore, writing

$$g(z) = \sum_{n=0}^{\infty} b_n z^n,$$

for any $R < R^*$, there exists $C > 0$ such that, for all $n \in \mathbb{N}$, $b_n \leq CR^{-n}$.

Proof of Lemma 8. From the renewal equation of Lemma 3, one has, for $|z| \leq x^{1-\varepsilon}$ and $z \neq x$,

$$\begin{aligned}
\frac{g(z)}{z-x} = \sum_{n=0}^{\infty} |A_{x^n}| z^n &= \sum_{n=0}^{\infty} \left(\sum_{j\in\mathcal{I}} |A_{x^{n-n_j}}| + 1 \right) z^n \\
&= \sum_{n=0}^{\infty} \sum_{j\in\mathcal{I}} |A_{x^{n-n_i}}| z^n + \frac{1}{1-z} \\
&= \sum_{j\in\mathcal{I}} z^{n_j} \sum_{n=0}^{\infty} |A_{x^{n-n_j}}| z^{n-n_i} + \frac{1}{1-z} \\
&= \sum_{j\in\mathcal{I}} z^{n_j} \frac{g(z)}{z-x} + \frac{1}{1-z},
\end{aligned}$$

which rearranges to

$$g(z) = \frac{z-x}{(z-1)(\sum_{j\in\mathcal{I}} z^{n_j} - 1)}.$$

Therefore, g has a meromorphic expansion on the disk of convergence of

$$f(z) = \sum_{j\in\mathcal{I}} z^{n_j}, \tag{2.9}$$

which has radius at least $x^{1-\varepsilon}$, since, by assumption

$$\sum_{j\in\mathcal{I}} x^{(1-\varepsilon)n_j} = \sum_{j\in\mathcal{I}} \alpha_j^{1-\varepsilon} < \infty.$$

We now consider the zeros of $f - 1$ (i.e., the poles of g not equal to one).

Firstly, if $|z| < 1$, then

$$\left| \sum_{j \in \mathcal{I}} z^{n_j} \right| \le \sum_{j \in \mathcal{I}} |z|^{n_j} < \sum_{j \in \mathcal{I}} x^{n_j} = \sum_{j \in \mathcal{I}} \alpha_j = 1,$$

so that no z with $|z| < x$ satisfies $f(z) = 1$. In this inequality, we also see that $f(x) = 1$. Since this is a simple zero, i.e.,

$$f'(1) = \sum_{j \in \mathcal{I}} n_j x^{n_j} = \sum_{j \in \mathcal{I}} \alpha_j \log_x(\alpha_j) \neq 0,$$

it cancels with the simple zero in the numerator of g. Furthermore, suppose for contradiction that there exists $\theta \in (0, 1)$ with $f(xe^{2\pi i\theta}) = 1$. Writing this as

$$\sum_{j \in \mathcal{I}} \alpha_j = \sum_{j \in \mathcal{I}} x^{n_j} e^{2\pi n_j \theta} = \sum_{j \in \mathcal{I}} \alpha_j e^{2\pi n_j \theta},$$

this implies that $e^{2\pi\theta n_j} = 1$ for each j, i.e., $\theta n_j \in \mathbb{N}$, which in particular gives

$$\alpha_j = (x^{1/\theta})^{\theta n_j},$$

contradicting the minimality of $x > x^{1/\theta}$ assumed in the theorem.

In summary, g has no poles on the closed disk of radius x, and therefore is holomorphic on the slightly larger disk of radius $R^* > x$, where R^* is the absolute value of the next smallest root of (2.9), or $x^{1-\varepsilon}$ if no such root exists, which yields the lemma. □

The final stage of the proof of Theorem 2 is similar to that of Theorem 1, but the method of decomposing a general interval requires a little more care. We use the following concept of *itinerary.*

Definition 10 (Itinerary)**.** Given $b \in [0, 1)$, we say $\boldsymbol{v} = \mathrm{It}_n(b)$ if $\boldsymbol{v} \in \mathcal{I}^n$ and $b \in T_{\boldsymbol{v}}[0, 1)$ (this is unique by the disjointness of $T_i[0, 1)$), and we let $\mathrm{It}_0(b) = \emptyset$.

If $\mathrm{It}_n(b)$ exists for all $n \in \mathbb{N}$, we say b has *infinite itinerary.* In which case (by disjointness again), there exists $(i_k)_{k=1}^{\infty} \in \mathcal{I}^{\mathbb{N}}$ such that

$$\mathrm{It}_k(b) = (i_1, i_2, \ldots, i_k) \tag{2.10}$$

for all $k \in \mathbb{N}$. Otherwise, we say b has *finite itinerary*, and for $n = \max\{k : \mathrm{It}_k(b) \text{ exists}\}$ and $\mathrm{It}_n(b) = (i_1, \ldots, i_n)$, (2.10) holds for $0 < k \le n$.

Finally, for simplicity in later sums, we adopt the convention that $\alpha_{\mathrm{It}_k(b)} = 0$

whenever $\mathrm{It}_k(b)$ does not exist.

We are now ready to prove Theorem 2.

Proof of Theorem 2. Let $\{\alpha_i\}$, x and $\rho \in (x/R^*, 1)$ be as in the statement of the theorem. Fix $b \in (0,1)$ and consider $I = [0,b)$ (i.e, $\|I\| = b$).

We consider a natural "greedy algorithm" for filling I with intervals of the form $T_{\boldsymbol{v}}[0,1)$. First, we fill I with as many intervals as possible from the "first generation",

$$\{T_i[0,1) \mid i \in \mathcal{I}\},$$

then, in the remaining gap, as many from the "second generation",

$$\{T_{i_1}T_{i_2}[0,1) \mid (i_1,i_2) \in \mathcal{I}^2\},$$

and so on. Accordingly, let $V_1 := \{i \in \mathcal{I} \mid T_i[0,1) \subset I\}$ and for $k \geq 2$,

$$V_k := \big\{\boldsymbol{w} * i \in \mathcal{I}^k \mid (\boldsymbol{w}, i) \in \mathcal{I}^{k-1} \times \mathcal{I},\ \underbrace{T_{\boldsymbol{w}}T_i[0,1) \subset I}_{①},\ \underbrace{T_{\boldsymbol{w}}[0,1) \not\subset I}_{②}\big\}.$$

Notice that conditions ① and ② imply that $T_{\boldsymbol{w}}[0,1)$ is an interval which meets both I and $[0,1]\backslash I$. It thus contains b, so $\boldsymbol{w} = \mathrm{It}_{k-1}(b)$.

In this way, for $k \in \mathbb{N}$, $V_k = \emptyset$ if $\mathrm{It}_{k-1}(b)$ does not exist, and otherwise,

$$V_k \subset \{\mathrm{It}_{k-1}(b)\} \times \mathcal{I}.$$

In any case, for $V := \bigcup_{k\in\mathbb{N}} V_k$, the union

$$\mathcal{V} = \bigcup_{\boldsymbol{v}\in V} T_{\boldsymbol{v}}[0,1)$$

is disjoint and wholly contained in I. To show that they have the same Lebesgue measure, we note this construction by definition contains the construction from the proof of Theorem 1 on page 21. That is, for $\mathcal{U}_n := \{\boldsymbol{v} \in \mathcal{I}^n \mid T_{\boldsymbol{v}}[0,1) \subset I\}$, and arbitrary $n \in \mathbb{N}$,

$$\mathcal{V} \supset \bigcup_{k=1}^{n} \bigcup_{\boldsymbol{v}\in V_k} T_{\boldsymbol{v}}[0,1) \supset \bigcup_{U\in\mathcal{U}_n} U.$$

Therefore, (2.8) applies to give

$$b \geq \sum_{\boldsymbol{v} \in V} \alpha_{\boldsymbol{v}} = \mathrm{Leb}(\mathcal{V}) \geq \mathrm{Leb}\left(\bigcup_{U \in \mathcal{U}_n} U\right) \geq b - 2\alpha_*^n,$$

for $\alpha_* = \max_i(\alpha_i)$. Hence, taking $n \to \infty$,

$$\sum_{\boldsymbol{v} \in V} \alpha_v = b.$$

Unfortunately, the respective equality for μ_λ does not necessarily apply (specifically, in the case that no $i \in \mathcal{I}$ fixes 0, i.e., $0 \notin \bigcup_i T_i[0,1)$), but nevertheless the difference decays in a favourable manner. Let

$$K := \{T_{\boldsymbol{v}}(0) \in I \setminus \mathcal{V} \mid \boldsymbol{v} \in W(\mathcal{I})\},$$

then $\mu_\lambda(I) = \mu_\lambda(\mathcal{V}) + \mu_\lambda(K)$ for all $\lambda > 0$. Suppose that $\boldsymbol{v} \in \mathcal{I}^n$ has $T_{\boldsymbol{v}}(0) \in I \setminus \mathcal{V}$. Then $T_{\boldsymbol{v}}[0,1) \not\subset I$, since otherwise we have the middle inclusion in

$$T_{\boldsymbol{v}}(0) \in T_{\boldsymbol{v}}[0,1) \subset \bigcup_{U \in \mathcal{U}_n} U \subset \mathcal{V}.$$

Thus, $T_{\boldsymbol{v}}[0,1) \not\subset I$ which, by the above reasoning, implies that $b \in T_{\boldsymbol{v}}[0,1)$, i.e., $\boldsymbol{v} = \mathrm{It}_n(b)$. Hence, by Lemma 2,

$$\begin{aligned}
\mu_{x^n}(K) &\leq \mu_{x^n}\{T_{\mathrm{It}_k(b)}(0) \mid k \in \mathbb{N}_0,\ \mathrm{It}_k(b) \text{ exists}\} \\
&= \frac{1}{|A_{x^n}|}|\{k \in \mathbb{N}_0 \mid \alpha_{\mathrm{It}_k(b)} \geq x^n\}| \\
&\leq \frac{1}{|A_{x^n}|}|\{k \in \mathbb{N}_0 \mid \alpha_*^k \geq x^n\}| \\
&= \frac{1}{|A_{x^n}|}\left(\left\lfloor \frac{n}{\log_x(\alpha_*)} \right\rfloor + 1\right), \qquad (2.11)
\end{aligned}$$

which is plainly $\mathcal{O}(y^n)$ as $n \to \infty$ for every $y > x$, since $|A_{x^n}|^{-1} \sim Hx^n/\log(x)$.

Moving forwards, we write (by disjointness)

$$\begin{aligned}
\mu_{x^n}(I) - b &= \mu_{x^n}(K) + \mu_{x^n}(\mathcal{V}) - \sum_{v \in V} \alpha_v \\
&= \mu_{x^n}(K) + \underbrace{\sum_{\boldsymbol{v} \in V} \mu_{x^n}\big(T_{\boldsymbol{v}}[0,1)\big) - \alpha_{\boldsymbol{v}}}_{\Lambda}.
\end{aligned}$$

We now consider the decay of Λ in two cases.

Case 1: *No* $i \in \mathcal{I}$ *fixes* 0. In this case, by Lemma 2, $\mu_{x^n}(T_{\boldsymbol{v}}[0,1)) = 0$ whenever $\alpha_{\boldsymbol{v}} > x^n$, and therefore

$$\Lambda = \sum_{\substack{\boldsymbol{v}\in V\\ \alpha_{\boldsymbol{v}}\geq x^n}} \frac{|X_{x^n/\alpha_{\boldsymbol{v}}}| - \alpha_{\boldsymbol{v}}|X_{x^n}|}{|X_{x^n}|} \quad - \sum_{\substack{\boldsymbol{v}\in V\\ \alpha_{\boldsymbol{v}}< x^n}} \alpha_{\boldsymbol{v}}.$$

The second sum is simple to bound, using the fact that $V_k \subset \{\mathrm{It}_{k-1}(b)\} \times \mathcal{I}$ if $\mathrm{It}_{k-1}(b)$ exists, and is empty otherwise. This gives the inequality below, recalling the convention that $\alpha_{\mathrm{It}_n(b)} = 0$ if $\mathrm{It}_n(b)$ doesn't exist (i.e., if b has finite itinerary):

$$\begin{aligned}
\sum_{\substack{\boldsymbol{v}\in V\\ \alpha_{\boldsymbol{v}}< x^n}} \alpha_{\boldsymbol{v}} &\leq x^{\varepsilon n} \sum_{\boldsymbol{v}\in V} \alpha_{\boldsymbol{v}}^{1-\varepsilon}\\
&= x^{\varepsilon n} \sum_{k=1}^{\infty} \sum_{\boldsymbol{v}\in V_k} \alpha_{\boldsymbol{v}}^{1-\varepsilon}\\
&\leq x^{\varepsilon n} \sum_{k=1}^{\infty} \sum_{i\in\mathcal{I}} \alpha_{\mathrm{It}_{k-1}(b)}^{1-\varepsilon} \alpha_i^{1-\varepsilon}\\
&\leq x^{\varepsilon n} \sum_{k=1}^{\infty} \sum_{i\in\mathcal{I}} \alpha_*^{(1-\varepsilon)k} \alpha_i^{1-\varepsilon}\\
&= \frac{\alpha_* x^{n\varepsilon}}{1-\alpha_*} \sum_{i\in\mathcal{I}} \alpha_i^{1-\varepsilon},
\end{aligned}$$

which is a bounded multiple of $x^{n\varepsilon} \leq (x/R^*)^n < \rho^n$.

To bound the remaining first sum, we consider the generating function, for $m \in \mathbb{N}$, of $|A_{x^{n-m}}| - x^m|A_{x^n}|$, which relates to g from Lemma 8 in the following manner:

$$\begin{aligned}
\sum_{n=0}^{\infty} \big(|A_{x^{n-m}}| - x^m|A_{x^n}|\big) z^n &= \big(z^m - x^m\big) \sum_{n=0}^{\infty} |A_{x^n}| z^n\\
&\leq \left(\frac{z^m - x^m}{z - x}\right) g(z)\\
&= \big(z^{m-1} + xz^{m-2} + \cdots + x^{m-1}\big) g(z).
\end{aligned}$$

Writing $g(z) = \sum_{n\in\mathbb{Z}} b_n z^n$ (setting $b_n = 0$ for $n < 0$), equating coefficients gives, for all $n \in \mathbb{N}$,

$$|A_{x^{n-m}}| - x^m|A_{x^n}| = b_{n-m+1} + xb_{n-m+2} + \cdots + x^{m-1}b_n.$$

By Lemma 8, since $x/\rho < R^*$, there exists $C_1 > 0$ such that $|b_n| < C_1(\rho/x)^n$ for all $n \in \mathbb{N}_0$, which trivially applies for negative n also. Thus, we have

$$\begin{aligned}\big||A_{x^{n-m}}| - x^m|A_{x^n}|\big| &\le C_1 \frac{\rho^{n-m-1} + \rho^{n-m-2} + \cdots + \rho^n}{x^{n-m+1}} \\ &= C_1 \left(\frac{\rho}{x}\right)^n \frac{(x/\rho)^m - x^m}{x(1-\rho)} \\ &\le C_2 \left(\frac{\rho}{x}\right)^n \left(\left(\frac{x}{\rho}\right)^m - x^m\right),\end{aligned}$$

where $C_1 = C_2 x(1-\rho)$. In particular, if $x^m = \alpha_{\boldsymbol{v}}$, then $(x/\rho)^m \le (R^*)^m < x^{m(1-\varepsilon)} \le \alpha_{\boldsymbol{v}}^{1-\varepsilon}$, and, since $x^n|A^n|$ is bounded away from zero (since it is positive and convergent to a positive limit), the previous inequality gives, for $x^n \le \alpha_{\boldsymbol{v}}$,

$$\left|\frac{|A_{x^n/\boldsymbol{v}}| - \alpha_{\boldsymbol{v}}|A_{x^n}|}{|A_{x^n}|}\right| \le \frac{C_2}{x^n|A_{x^n}|}\,\rho^n\left(\alpha_{\boldsymbol{v}}^{1-\varepsilon} - \alpha_{\boldsymbol{v}}\right) \le C_3\,\rho^n\left(\alpha_{\boldsymbol{v}}^{1-\varepsilon} - \alpha_{\boldsymbol{v}}\right),$$

for some constant $C_3 > 0$. Summing over $\boldsymbol{v}$ thus gives

$$\sum_{\substack{\boldsymbol{v}\in V\\ \alpha_{\boldsymbol{v}} \ge x^n}} \left|\frac{|A_{x^n/\alpha_{\boldsymbol{v}}}| - \alpha_{\boldsymbol{v}}|A_{x^n}|}{|A_{x^n}|}\right| \le C_3\rho^n \sum_{\boldsymbol{v}\in V}\big(\alpha_{\boldsymbol{v}}^{1-\varepsilon} - \alpha_{\boldsymbol{v}}\big).$$

The right hand side is a bounded multiple of ρ^n, completing the proof in this case.

Case 2: *There exists* $1 \in \mathcal{I}$ *such that* $T_1(0) = 0$. The proof here is similar to the previous case, but with some necessary modifications. Firstly, since it is now possible that $\mu_{x^n}(T_{\boldsymbol{v}}[0,1)) > 0$ if $x^n > \alpha_{\boldsymbol{v}}$, we bound

$$\sum_{\substack{\boldsymbol{v}\in V\\ \alpha_{\boldsymbol{v}} < \lambda}} \mu_\lambda\big(T_{\boldsymbol{v}}[0,1)\big)$$

as follows. As in the proof of Lemma 2, if $\mu_\lambda\big(T_{\boldsymbol{v}}[0,1)\big) > 0$ for $\lambda > \alpha_{\boldsymbol{v}}$, there exists $\boldsymbol{w} \in W(\mathcal{I})$ and $\boldsymbol{j} \in \{1\}^k$ such that

$$T_{\boldsymbol{v}}[0,1) = T_{\boldsymbol{w}}T_{\boldsymbol{j}}[0,1)$$

and $\mu_\lambda\big(T_{\boldsymbol{v}}[0,1)\big) = \mu_\lambda\{T_{\boldsymbol{v}}(0)\} = \mu_\lambda\{T_{\boldsymbol{w}}(0)\}$. Moreover, by construction, $\boldsymbol{v} \in V$ again implies that $T_{\boldsymbol{w}}[0,1) \not\subset I$ and hence that $\boldsymbol{w} = \mathrm{It}_n(b)$ for some $n \in \mathbb{N}_0$ (i.e.,

$T_{\boldsymbol{v}}(0) = T_{\mathrm{It}_n(b)}(0))$. Therefore, similarly to $\mu_\lambda(K)$ considered above, one has

$$\sum_{\substack{\boldsymbol{v}\in V\\ \alpha_{\boldsymbol{v}}<\lambda}} \mu_{x^n}\big(T_{\boldsymbol{v}}[0,1)\big) \le \mu_{x^n}\{T_{\mathrm{It}_n(b)}(0) \mid n\in\mathbb{N}\} = \mathcal{O}(y^n)$$

for all $y > x$, as $n \to \infty$. Therefore, up to a quantity of this order, we may write $\mu_{x^n}(I) - b$ again as

$$\Lambda = \sum_{\substack{\boldsymbol{v}\in V\\ \alpha_{\boldsymbol{v}}\ge x^n}} \frac{|X_{x^n/\alpha_{\boldsymbol{v}}}| - \alpha_{\boldsymbol{v}}|X_{x^n}|}{|X_{x^n}|} - \sum_{\substack{\boldsymbol{v}\in V\\ \alpha_{\boldsymbol{v}}>x^n}} \alpha_{\boldsymbol{v}}.$$

The second sum is $\mathcal{O}(x^{n\varepsilon})$, by the argument in the first case. The first sum yields to a similar treatment as before. For $\alpha_{\boldsymbol{v}} = x^m$ and $\alpha_1 = x^j$, the numerator reads

$$|X_{x^{n-m}}| - x^m|X_{x^n}| = |A_{x^{n-m}}| - x^m|A_{x^n}| - |A_{x^{n-m-j}}| - x^m|A_{x^{n-j}}|,$$

so that

$$\sum_{k=0}^{m} |X_{x^{n-m}}| - x^m|X_{x^n}| = \big(z^m - x^m - z^{m+j} + x^m z^j\big)\frac{g(z)}{z-x}.$$

Equating coefficients as before and applying the inequalities at the end of the last case (in particular that $x/\rho \le x^{1-\varepsilon}$) then gives

$$\begin{aligned}\big||X_{x^{n-m}}| - x^m|X_{x^n}|\big| &\le \big|\big(b_{n-m+1} + b_{n-m+2}x + \cdots + x^{m-1}b_n\big)\\ &\qquad + \big(b_{n-m-j+1} + b_{n-m-j+2}x + \cdots + x^{m-1}b_{n-j}\big)\big|\\ &\le C_2\left(\frac{\rho}{x}\right)^n\left(\left(\frac{x}{\rho}\right)^m - x^m + \left(\frac{x}{\rho}\right)^{m-j} - x^{m-j}\right)\\ &\le C_2\left(\frac{\rho}{x}\right)^n\left(\alpha_{\boldsymbol{v}}^{1-\varepsilon} - \alpha_{\boldsymbol{v}} + \left(\frac{\alpha_{\boldsymbol{v}}}{\alpha_1}\right)^{1-\varepsilon} - \frac{\alpha_{\boldsymbol{v}}}{\alpha_1}\right).\end{aligned}$$

Finally, as in the last case, dividing through by $|X_{x^n}| \sim (1-\alpha_1^{-1})Hx^{-n}/\log(x)$, and summing over $\boldsymbol{v} \in V$, gives a bounded multiple of ρ^n, completing the proof in this case. $\square$

2.4.2 Discrepancy estimates in the higher rank case

When the collection $\{\alpha_j\}_{j\in\mathcal{I}}$ is not rank one, we require not only a strict decay property on the $\{\alpha_j\}$, as assumed by Theorem 2, but also a kind of Diophantine condition, which is based on the following definition.

Definition 11 (r-badly approximable)**.** For $r \in [0, \infty)$ we say a number $\gamma \in \mathbb{R}$ is r-badly approximable if there exists $d > 0$ such that

$$\forall\, (l, k) \in \mathbb{Z}^2 \ s.t.\ l \neq 0, \qquad \left| \gamma - \frac{k}{l} \right| > \frac{d}{|l|^{2+r}}.$$

Remark 5. Larger values of r correspond to more easily approximable numbers:

- For $r = 0$, the property is equivalent to γ having bounded continued fraction coefficients. (Such γ comprise a set of measure 0 containing all quadratic algebraic numbers.)
- For $r > 0$, the property holds Lebesgue almost-everywhere: the Hausdorff dimension of the complementary set is $2/(2+r) < 1$, by Jarnik's theorem [42, Theorem 10.3].
- Certain transcendental numbers (e.g., *Liouville numbers*) do not satisfy the property for any r whatsoever.

We have the following positive result, using techniques of algebraic number theory.

Theorem 3. *Suppose that*

- $\{\alpha_i\}_{i \in \mathcal{I}}$ *is higher rank;*
- $\sum_i \alpha_i^{1-\varepsilon} < \infty$ *for some* $\varepsilon > 0$*; and*
- *there is a pair* $\alpha, \beta \in \{\alpha_i\}_{i \in \mathcal{I}}$ *with* $\alpha < \beta$*, such that* $\log(\alpha)/\log(\beta)$ *is* r*-badly approximable for some* $r \in [0, 1/2)$*.*

Then, for all positive P *strictly less than*

$$P^* = \frac{1 - 2r}{8(1 + r)},$$

there exists $C > 0$ *such that, for all intervals* $I \subset [0, 1]$ *and* $\lambda \in (0, 1)$*,*

$$\big|\, \mu_\lambda(I) - \|I\| \,\big| \leq C \big(- \log(\lambda) \big)^{-P}.$$

The proof of Theorem 3 requires us to consider the Mellin transform of $|A_{1/t}|$,

$$g(z) = \int_0^\infty t^{-z-1}|A_{1/t}| \,\mathrm{d}t.$$

The following lemma gives this an explicit form, as a consequence of the renewal equation for $|A_\lambda|$. In the following, we let $\Re(z)$, $\Im(z)$ denote the real and imaginary parts of a complex number z, respectively.

Lemma 9. *For all $\Re(z) > 1$, the Mellin transform g takes the form*

$$g(z) = \frac{1}{z}\left(1 - \sum_{j\in\mathcal{I}} \alpha_j^{z}\right)^{-1}.$$

In particular, if $\varepsilon > 0$ and $\sum_j \alpha_j^{1-\varepsilon} < \infty$, g has a meromorphic extension to the half-plane $\{\Re(z) > 1-\varepsilon\}$.

Proof of Lemma 9. We compute directly. Since $|A_{1/t}| \sim t/H$, the dominated convergence theorem implies that, for $\Re(z) > 1$,

$$\begin{aligned}
g(z) &:= \int_0^\infty t^{-z-1}|A_{1/t}| \,\mathrm{d}t \\
&= \int_0^\infty t^{-z-1} \sum_{\boldsymbol{v}\in W(\mathcal{I})} \chi_{\{t\alpha_{\boldsymbol{v}}\geq 1\}} \,\mathrm{d}t \\
&= \sum_{\boldsymbol{v}\in W(\mathcal{I})} \int_{\alpha_{\boldsymbol{v}}^{-1}}^\infty t^{-z-1} \,\mathrm{d}t \\
&= \sum_{\boldsymbol{v}\in W(\mathcal{I})} \int_{\alpha_{\boldsymbol{v}}^{-1}}^\infty t^{-z-1} \,\mathrm{d}t \\
&= \frac{1}{z} \sum_{\boldsymbol{v}\in W(\mathcal{I})} \alpha_{\boldsymbol{v}}^z \\
&= \frac{1}{z}\left(1 + \sum_{k=1}^\infty \sum_{(i_1,\dots i_k)\in\mathcal{I}^k} \alpha_{i_1}^z\alpha_{i_2}^z\cdots\alpha_{i_k}^z\right) \\
&= \frac{1}{z}\sum_{k=0}^\infty \left(\sum_{j\in\mathcal{I}}\alpha_j^z\right)^k \\
&= \frac{1}{z}\left(1 - \sum_{j\in\mathcal{I}}\alpha_j^z\right)^{-1},
\end{aligned}$$

the geometric series converging since

$$\left|\sum_{j\in\mathcal{I}}\alpha_j^z\right| \leq \sum_{j\in\mathcal{I}}\alpha_j^{\Re(z)} < \sum_{j\in\mathcal{I}}\alpha_j = 1. \tag{2.12}$$

□

As is well-known, one can obtain asymptotic information about a function from the distribution of poles of its Mellin transform. In this case, these are given by the zeros of the almost-periodic function

$$f(z) := \sum_{j\in\mathcal{I}}\alpha_j^z - 1.$$

There is a lot one can say straight away. By (2.12), $f(z) = 0$ implies that $\Re(z) \leq 1$. Indeed, $f(1) = 0$ and $f'(1) = -1/H \neq 0$, so f has a simple zero at 1. This is the only zero of f on the line $\{\Re(z) = 1\}$ since, if $\theta \neq 0$ satisfies $f(1+i\theta) = 1$, i.e.,

$$\sum_{j\in\mathcal{I}}\alpha_j^{1+i\theta} = \sum_{j\in\mathcal{I}}\alpha_j e^{i\theta\log(\alpha_j)} = \sum_{j\in\mathcal{I}}\alpha_j,$$

we must have $\alpha_j^{1+\theta} = \alpha_j$, i.e., $e^{i\theta\log(\alpha_j)} = 1$, so that $\log(\alpha_j) \in \frac{2\pi}{\theta}\mathbb{Z}$, which contradicts the assumption that $\{\alpha_j\}_j$ is higher rank.

The role of the poles of the Mellin transform is illustrated in the proof of the following negative result, which in particular shows that these poles must accumulate towards the ends of $\{\Re(z) = 1\}$.

Proposition 4. *For any given collection of positive numbers $\{\alpha_j\}_{j\in\mathcal{I}}$ which sum to 1 such that $H = -\sum_j \alpha_j\log(\alpha_j) < \infty$, there is no $\varepsilon > 0$ for which*

$$|A_\lambda| = \frac{1}{H\lambda} + \mathcal{O}(\lambda^{\varepsilon-1}) \tag{2.13}$$

as $\lambda \to 0^+$.

Proof of Proposition 4. Fix $\{\alpha_j\}_{j\in\mathcal{I}}$ and assume for contradiction that there is an $\varepsilon > 0$ for which (2.13) holds. Taking the Mellin transform of (2.13) yields that

$$\frac{-1}{zf(z)} = g(z) = \frac{1}{H(z-1)} + \int_1^\infty \mathcal{O}\big(t^{-z-\varepsilon}\big)\ \mathrm{d}t.$$

Since the integral converges absolutely for all z with $\Re(z) > 1-\varepsilon$, the left hand side has a meromorphic extension to the half plane $\mathcal{H} = \{z \in \mathbb{C} \mid \Re(z) > 1-\varepsilon\}$ and $\mathcal{H}$ contains exactly one pole at $z = 1$ (i.e., with only one zero of f). This leads to a contradiction, via the theory of almost-periodic functions, which we now recall.

Definition 12 (Almost-periodic)**.** For f, a complex function defined on a half plane, we say f is *periodic* if there exists $L > 0$ satisfying $f(z) = f(z + iL)$ for all z. More generally, we say that f is *almost-periodic* if, for any $\delta > 0$, there exist infinitely many $L > 0$ satisfying

$$|f(z + iL) - f(z)| < \delta$$

for all z.

From this point, we follow the proof of the corollary to [29, Theorem 3.6]. As the limit of a sequence of periodic functions, f is almost-periodic (see, e.g., the corollary to [29, Theorem 3.12]). Moreover, f is bounded on the half-plane $\mathcal{H}' = \{\Re(z) \geq 1 - \varepsilon/2\}$ since there one has

$$|f(z)| \leq 1 + \sum_{i \in \mathcal{I}} \alpha^{1-\varepsilon/2} < \infty.$$

Since $f(1) = 0$, the definition of almost-periodicity provides an increasing sequence of positive numbers $(y_n)_{n=1}^{\infty}$, with $f(1 + iy_n) \to 0$ as $n \to \infty$. In other words, the holomorphic functions

$$f_n(z) = f(z + iy_n)$$

are each bounded on $\mathcal{H}'$, and $f_n(1) \to 0$.

Furthermore, fixing $j \in \mathcal{I}$, we see that every interval of length $L := -2\pi/\log(\alpha_j)$ contains a v such that $\alpha_j^{1+iv} = -\alpha_j$, and hence

$$|f\,(1 + iv)| \geq 1 + \alpha_j - \sum_{k \in \mathcal{I} \setminus \{j\}} \alpha_k = 2\alpha_j > 0.$$

Thus, $\sup_{|v| \leq L} |f_n(1 + iv)|$ is bounded away from zero uniformly in n.

An application of Montel's theorem on the rectangle $(1 - \varepsilon/2, 1 + \varepsilon/2) + i(-L, L)$ shows that (f_n) uniformly converges (passing to a subsequence if necessary) to an analytic function f_∞. From the previous two considerations, $f_\infty(1) = 0$ and f_∞ is not identically zero.

Now, taking a circle about 1 in this rectangle small enough that f_∞ has no zeros on it, Hurwitz's theorem applies to show that f_n has a zero inside this circle for all n sufficiently large (which converge to 1 as $n \to \infty$).

This gives a sequence of zeros of f, $(z_n)_n$, such that $\Re(z_n) \to 1$ and $\Im(z_n) \to \infty$. Hence, there are infinitely many zeros of f in $\mathcal{H}$, contradicting the assumption that there is only one. $\square$

Considering now the proof of Theorem 3, the following lemma is where we use the r-badly approximable hypothesis, after [34], to manage the rate at which the zeros of f converge upon $\{\Re(z) = 1\}$.

Lemma 10. *Suppose that $(\alpha_j)_{j\in\mathcal{I}}$ satisfies the conditions of Theorem 3. Then there exists $C > 0$ such that, whenever $z = 1 - u + iv \in \mathbb{C} \setminus \{1\}$ satisfies both $f(z) = 0$ and $u < C$, then $u > 0$ and*

$$|v|^{2+2r} \geq \frac{C}{u}.$$

The basic idea of the proof is the following. If $f(z) = 0$, and $\Re(z) \approx 1$, then α^z and β^z need to be aligned, in the sense that $\alpha^z \approx \alpha$ and $\beta^z \approx \beta$ (i.e., their polar arguments lie close to the lattice $2\pi\mathbb{Z}$). This leads to a rational approximation of $\log(\alpha)/\log(\beta)$. The badly-approximable hypothesis then forces the denominator, which is roughly proportional to $|v|$, to be relatively large, compared to the efficacy of the approximation.

Proof of Lemma 10. The fact that $u > 0$ follows from the discussion preceding the previous proposition. We first show, assuming that $f(z) = 0$, the argument of α^z is $\mathcal{O}(\sqrt{u})$ as $u \to 0^+$. That is, for $\eta_\alpha \in (-\pi, \pi]$ satisfying

$$e^{i\eta_\alpha} = \frac{\alpha^z}{|\alpha^z|} = \frac{\alpha^z}{\alpha^{1-u}},$$

there exists a constant C such that $|\eta_\alpha| < C\sqrt{u}$, for all u sufficiently small. This uses the triangle inequality and some basic trigonometry, as we now detail. Firstly, we have that

$$|1 - \alpha^z| = \left|\sum_{j\in\mathcal{I}} \alpha^z - \alpha^z\right| \leq \sum_{j\in\mathcal{I}} \alpha^{1-u} - \alpha^{1-u} = 1 - \alpha + H(u),$$

where

$$H(u) := \sum_{n\in\mathcal{I}} \alpha_n^{1-u} - \alpha^{1-u} - 1 + \alpha.$$

In particular, since H is differentiable at 0 and $H(0) = 0$, we have $H(u) = \mathcal{O}(u)$ as $u \to 0^+$. Therefore, for u sufficiently small, $H(u) < \alpha$, which gives rise to the picture in Figure 2.5(i). Consequently, $|\eta_\alpha| < \theta$, where $\theta = \theta(u)$ is depicted in Figure 2.5(ii). By the cosine rule,

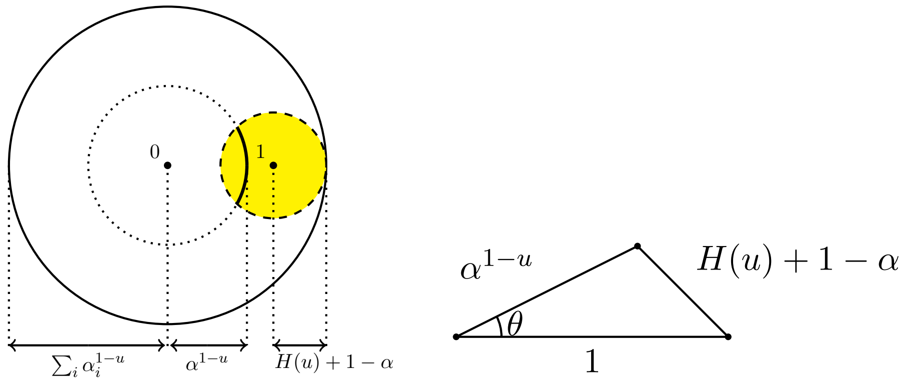


Figure 2.5: (i) The region in which α^z must lie to have $f(z) = 0$. (ii) The triangle defining θ, the maximum possible value of $|\eta_\alpha|$.

$$\begin{aligned}\cos(\theta) &= \frac{1+\alpha^{2-2u} - \big(1-\alpha+H(u)\big)^2}{2\alpha^{1-u}} \\ &= \alpha^u + \frac{1}{2}\left(\alpha^{1-u} - \alpha^{1+u}\right) - \frac{1}{2}\alpha^{1-u}H(u)\big(1-\alpha+H(u)\big) \\ &= 1 - \mathcal{O}(u),\end{aligned}$$

as $u \to \infty$. Using, for example, that

$$\lim_{y\to 0^+} \frac{\arccos(1-y)}{\sqrt{y}} = \sqrt{2}, \qquad \arccos : [-1,1] \to [0,\pi],$$

one has that $|\eta_\alpha| \leq \theta = \mathcal{O}(\sqrt{u})$ as $u \to 0^+$.

The above applies with α in place of β to show that $\eta_\beta = \mathcal{O}(\sqrt{u})$ as $u \to 0^+$, where $\eta_\beta \in (-\pi, \pi]$ is the argument of β^z.

We now apply the badly approximable assumption. To set this up, first write

$$v\log(\alpha) = 2\pi k + \eta_\alpha, \qquad \text{and} \qquad v\log(\beta) = 2\pi l + \eta_\beta,$$

for $l, k \in \mathbb{Z}$, and assume that $-\log(\beta)|v| \geq 2\pi$, so that l and k are non-zero and

$$-|v|\log(\alpha) \geq 2\pi|l| - \pi \geq 2\pi|l| + \frac{\log(\beta)}{2}|v| \geq 2\pi|l| + \frac{\log(\alpha)}{2}|v|,$$

i.e., $|l| \leq -\frac{3}{2}\log(\alpha)|v|$. Similarly, $|k| \leq -\frac{3}{2}\log(\beta)|v|$. The mean value theorem provides λ, such that $0 < |\lambda| < |\eta_\beta|$ and

$$\frac{2\pi k}{2\pi l + \eta_\beta} - \frac{k}{l} = -\eta_\beta \frac{2\pi k}{\big(2\pi l + \lambda\big)^2}.$$

Now, applying this in the definition of $\log(\alpha)/\log(\beta)$ being r-badly approximable, and subsequently using that $|\lambda|, |\eta_\alpha|, |\eta_\beta| < \pi$, yields that

$$\begin{aligned}\frac{d}{|l|^{2+r}} \leq \left|\frac{\log(\alpha)}{\log(\beta)} - \frac{k}{l}\right| = \left|\frac{2\pi k + \eta_\alpha}{2\pi l + \eta_\beta} - \frac{k}{l}\right| &= \left|\frac{\eta_\alpha}{2\pi l + \eta_\beta} + \left(\frac{2\pi k}{2\pi l + \eta_\beta} - \frac{k}{l}\right)\right| \\ &\leq \frac{|\eta_\alpha|}{|2\pi l + \eta_\beta|} + \frac{2\pi |k||\eta_\beta|}{(2\pi l + \lambda)^2} \\ &= \frac{|\eta_\alpha|}{2\pi |l|}\left|1 + \frac{\eta_\beta}{2\pi l}\right|^{-1} + \frac{|k||\eta_\beta|}{2\pi |l|^2}\left(1 + \frac{\lambda}{2\pi l}\right)^{-2} \\ &\leq \frac{|\eta_\alpha|}{\pi |l|} + \frac{2|k||\eta_\beta|}{\pi |l|^2}.\end{aligned}$$

Therefore, the previous asymptotics yield that there exists $C > 0$ such that, whenever u is sufficiently small and $-\log(\beta)|v| \geq 2\pi$,

$$0 < d \leq \frac{1}{\pi}\left(|\eta_\alpha||l|^{1+r} + 2|\eta_\beta||k||l|^r\right) \leq C u^{1/2} |v|^{1+r}.$$

This yields the required inequality, since one can manipulate C to cater for the finitely-many zeros of f corresponding to $0 < -\log(\beta)|v| < 2\pi$ and $u < \varepsilon/2$, say. □

The next lemma is a variant on the last and allows us to estimate decay of the Mellin inverse integral inside the zero-free region just constructed.

Lemma 11. *Suppose* $(\alpha_j)_{j \in \mathcal{I}}$ *is as given in Theorem 3. Then there exists* $C > 0$ *such that, whenever* $z = 1 - u + iv \in \mathbb{C}$ *with* $u \geq 0$ *and* $\sigma > 0$ *sufficiently small, the inequality*

$$|f(z)| < \sigma$$

implies that either

$$-\log(\beta)|v| \leq 2\pi,$$

or

$$|v|^{2+2r} > \frac{C}{\max(u, \sigma)}.$$

Proof of Lemma 11. The proof is an adaptation of that for Lemma 10. This time, the distance of α^z to 1 is at most $H(u) + 1 - \alpha + \sigma$, which, for u and σ sufficiently small, gives rise to a similar triangle to Figure 2.5(ii). Correspondingly, for u, σ sufficiently small,

$$\cos(\eta_\alpha) \geq \frac{1}{2}\frac{1 + \alpha^{2-2u} - (1 - \alpha + H(u) + \sigma)^2}{\alpha^{1-u}} = 1 - \mathcal{O}(\max(u, \sigma))$$

as $(u,\sigma) \to 0$, which gives, as $\max(u,\sigma) \to (0,0)$,

$$|\eta_\alpha| \leq \frac{\pi}{2}\sqrt{1-\cos(\theta)} = \mathcal{O}(\sqrt{\max(u,\sigma)}).$$

An analogous asymptotic formula holds for η_β. The diophantine approximation part proceeds verbatim, with the following ending: for $-\log(\beta)|v| \geq 2\pi$ and (u,σ) sufficiently small, there exists $C > 0$ such that

$$0 < d \leq \frac{1}{\pi}\Big(|\eta_\alpha||l|^{1+r} + 2|\eta_\beta||k||l|^r\Big) \leq C\max(u,\sigma)^{1/2}|v|^{1+r},$$

which rearranges to an inequality of the required form. □

The next crucial lemma is the analogue of Lemma 8 in the higher rank case.

Lemma 12. *Under the assumptions of Theorem 3,*

$$|A_\lambda| = \frac{1}{H\lambda} + \mathcal{O}\big(\lambda^{-1}(-\log(\lambda))^{-P}\big) \qquad \lambda \in (0,1),$$

for any $P \in (0, P^*)$, where P^* is given in Theorem 3.

Proof of Lemma 12. The proof uses simple complex analysis to estimate the integral

$$F(t) := \frac{-1}{2\pi i}\int_{2-i\infty}^{2+i\infty} \frac{t^{z+3}}{z(z+1)(z+2)(z+3)f(z)}\mathrm{d}z \qquad (t>0), \tag{2.14}$$

which we relate to $|A_{1/t}|$ in the following manner. On the line $\{\Re(z) = 2\}$, f is uniformly bounded away from zero, since

$$|f(z)| \geq 1 - \sum_{j\in\mathcal{I}} \alpha_j^2 > 0,$$

so $F(t)$ absolutely converges for all $t > 0$. Therefore, by the Mellin inversion theorem [69, Theorem 8.11], the Mellin transform of $t \mapsto F(t)/t^3$ equals

$$\int_0^\infty t^{-z-4}F(t)\ \mathrm{d}t = \frac{-1}{z(z+1)(z+2)(z+3)f(z)} = \frac{g(z)}{(z+1)(z+2)(z+3)}.$$

That is,

$$\int_0^\infty t^{z-1}|A_{1/t}|\ \mathrm{d}t = (z+1)(z+2)(z+3)\int_0^\infty t^{z-4}F(t)\ \mathrm{d}t = -\int_0^\infty \frac{\mathrm{d}^3}{\mathrm{d}t^3}\big(t^{z-1}\big)F(t)\ \mathrm{d}t,$$

which, using integration by parts and the uniqueness of the Mellin transform, shows

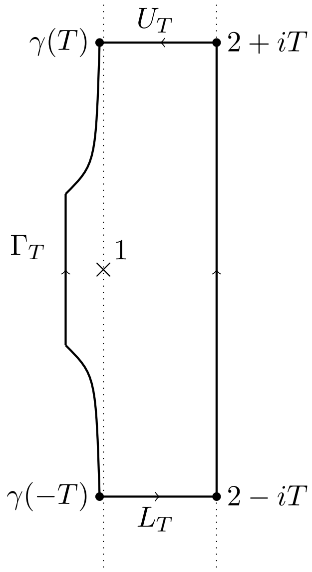


Figure 2.6: The contour Γ_T used in the proof of Lemma 10.

that, for all $x > 0$,

$$F''(x) = \int_0^x |A_{1/t}| \, \mathrm{d}t.$$

We now relate the integral in (2.14) to that over the contour Γ, parametrised by

$$\gamma : \mathbb{R} \to \mathbb{C}, \qquad \gamma(v) = 1 + iv - C \min(1, |v|^{-2-r})$$

(see Figure 2.6). Here, $C > 0$ is chosen sufficiently small so that the previous two lemmas apply as follows: firstly, the only zero of f which lies on or to the right of Γ is at 1, and secondly, there exists $D > 0$ such that whenever z lies on or to the right of Γ, $\Re(z) \leq 1$ and $|\Im(z)| \geq -2\pi(\log(\beta))^{-1}$, one has

$$|f(z)| \geq D|\Im(z)|^{-2-2r}. \tag{2.15}$$

Consider, for $T \geq 1$, the contour Γ_T depicted in Figure 2.6. Cauchy's residue theorem gives, for $F^*(z)$ denoting the integrand of $F(t)$,

$$\int_{2-iT}^{2+iT} F^*(z) \, \mathrm{d}z = 2\pi i \frac{t^4}{24H} + \int_{\Gamma_T} F^*(z) \, \mathrm{d}z + \int_{U_T} F^*(z) \, \mathrm{d}z + \int_{L_T} F^*(z) \, \mathrm{d}z,$$

where H is recalled below.

Note that U_T and L_T have bounded lengths. Thus, to show that

$$\left|\int_{U_T} F^*(z)\,\mathrm{d}z\right| = \left|\int_{L_T} F^*(z)\,\mathrm{d}z\right| \to 0$$

as $T \to \infty$, since trivially

$$\int_{U_T} |F^*(z)|\,|\mathrm{d}z| \le T^{-3} \int_{U_T} |f(z)|^{-1}\,|\mathrm{d}z|,$$

it suffices, for $\kappa(v) := \min\big(1, |v|^{-2-r}\big)$, to show that

$$\sup_{1-\kappa(T)\le t\le 2} |f(t+iT)|^{-1} = o(T^3)$$

as $T \to \infty$. This we do in three cases. Firstly, for $1-\kappa(T) \le t \le 1$, (2.15) applies to give

$$|f(t+iT)| \ge DT^{2+2r}.$$

Secondly, since $H = -\sum_j \alpha_j \log(\alpha_j) = \max_{1\le\Re(z)\le 2} |f'(z)|$, one has that, whenever $1 < t \le \frac{2}{DH}T^{-2-2r}+1$,

$$|f(t+iT)| \ge \frac{D}{2}T^{2+2r}.$$

Finally, for all $t > \frac{2}{DH}T^{-2-2r}+1$,

$$\begin{aligned}
|f(t+iT)| &\ge 1 - \sum_{j\in\mathcal{I}} \alpha_j^t \\
&\ge 1 - \sum_{j\in\mathcal{I}} \alpha_j^{\frac{2}{DH}T^{-2-2r}+1} \\
&\sim H\left(\frac{2}{DH}T^{-2-2r}\right) = \frac{2}{D}T^{-2-2r}
\end{aligned}$$

as $T \to \infty$, using the limit definition of $f'(1) = -H$. Therefore in summary, there is a constant $C > 0$ such that for all T sufficiently large,

$$\sup_{1-\kappa(T)\le t\le 2} |f(t+iT)|^{-1} \le C\big(T^{2+2r}\big),$$

which yields the required decay of the integrals over U_T and L_T as $T \to \infty$. Thus, in this limit, we have

$$F(t) = \frac{t^4}{24H} + \frac{1}{2\pi i}\int_\Gamma F^*(z)t^{-z-3}\,\mathrm{d}z. \tag{2.16}$$

From this point, the proof follows along the lines of [40, Theorem 4.6]. Since $t \mapsto |A_{1/t}|$ is non-decreasing, we have the following, for any positive t and h:

$$\begin{aligned}
\frac{F(t+3h) - 3F(t+2h) + 3F(t+h) - F(t)}{h^3} &= \frac{1}{h^3}\int_t^{t+h} F'(\hat{t}+2h) - 2F'(\hat{t}+h) + F'(\hat{t}\,)\,\mathrm{d}\hat{t} \\
&= \frac{1}{h^3}\int_t^{t+h}\int_{\hat{t}}^{\hat{t}+h} F''(\hat{\hat{t}}+h) - F''(\hat{\hat{t}}\,)\,\mathrm{d}\hat{\hat{t}}\,\mathrm{d}\hat{t} \\
&= \frac{1}{h^3}\int_t^{t+h}\int_{\hat{t}}^{\hat{t}+h}\int_{\hat{\hat{t}}}^{\hat{\hat{t}}+h} |A_{\hat{\hat{\hat{t}}}^{-1}}|\,\mathrm{d}\hat{\hat{\hat{t}}}\,\mathrm{d}\hat{\hat{t}}\,\mathrm{d}\hat{t} \\
&\geq \frac{1}{h^2}\int_t^{t+h}\int_{\hat{t}}^{\hat{t}+h} |A_{\hat{\hat{t}}^{-1}}|\,\mathrm{d}\hat{\hat{t}}\,\mathrm{d}\hat{t} \\
&\geq \frac{1}{h}\int_t^{t+h} |A_{\hat{t}^{-1}}|\,\mathrm{d}\hat{t} \\
&\geq |A_{1/t}|.
\end{aligned}$$

Similarly, for any $t > 0$ and $h \in (0, t/3)$, we have that

$$\frac{F(t-3h) - 3F(t-2h) + 3F(t-h) - F(t)}{-h^3} \leq |A_{1/t}|.$$

Substituting (2.16) into the left hand side of these expressions yields

$$\frac{F(t\pm 3h) - 3F(t\pm 2h) + 3F(t\pm h) - F(t)}{\pm h^3} =$$
$$\frac{t}{H} \pm \frac{3}{2H}h - \frac{1}{2\pi i}\int_\Gamma \frac{(t\pm 3h)^{z+3} - 3(t\pm 2h)^{z+3} + 3(t\pm h)^{z+3} - t^{z+3}}{\pm h^3} F^*(z)\,\mathrm{d}z,$$

which uses the marvellous equality,

$$\frac{(t+3h)^4 - 3(t+2h)^4 + 3(t+h)^4 - t^4}{24} = t + \frac{3}{2}h$$

(which applies for all $h \in \mathbb{R}$). Thus,

$$\begin{aligned}
|A_{1/t}| = {} & \frac{t}{H} + \mathcal{O}(h) + \\
& \mathcal{O}\left(\int_\Gamma \left|\frac{(t\pm 3h)^{z+3} - 3(t\pm 2h)^{z+3} + 3(t\pm h)^{z+3} - t^{z+3}}{\pm h^3(z+1)(z+2)(z+3)}\right| \left|\frac{1}{zf(z)}\right| |\mathrm{d}z|\right).
\end{aligned}$$

From now on, let $h = h(t) \in (0, t/3)$ be a function of t to be determined later. To begin to estimate the integral, consider $|\Delta_\pm(t,h,z)|$ for $t>1$ and $z\in\Gamma$, where

$$\Delta_\pm(t,h,z) := \frac{(t\pm 3h)^z - 3(t\pm 2h)^z + 3(t\pm h)^z - t^z}{\pm h^3(z+1)(z+2)(z+3)}.$$

We can bound $\Delta_\pm$ in two different ways. First, we express it as a series of nested integrals (like with the fraction involving F) and use that $0<\Re(z)<1$ for $z\in\Gamma$, i.e., $t\mapsto t^{\Re(z)}$ is increasing in t. That is,

$$\begin{aligned}
|\Delta_\pm(t,h,z)| &= \left| \frac{1}{\pm h}\int_t^{t\pm h} \frac{1}{\pm h}\int_{\hat{t}}^{\hat{t}\pm h} \frac{1}{\pm h}\int_{\hat{\hat{t}}}^{\hat{\hat{t}}\pm h} \hat{\hat{\hat{t}}}^z \,\mathrm{d}\hat{\hat{\hat{t}}}\,\mathrm{d}\hat{\hat{t}}\,\mathrm{d}\hat{t} \right| \\
&\le \frac{1}{h}\int_t^{t+h} \frac{1}{h}\int_{\hat{t}}^{\hat{t}\pm h} \frac{1}{h}\int_{\hat{\hat{t}}}^{\hat{\hat{t}}+h} \hat{\hat{\hat{t}}}^{\Re(z)} \,\mathrm{d}\hat{\hat{\hat{t}}}\,\mathrm{d}\hat{\hat{t}}\,\mathrm{d}\hat{t} \\
&\le (t+3h)^{\Re(z)} \\
&\le 2t^{\Re(z)}. \qquad (2.17)
\end{aligned}$$

An alternative bound is given by the triangle inequality. Namely,

$$\begin{aligned}
|\Delta_\pm(t,h,z)| &\le \frac{(2t)^{\Re(z)+3} + 3(2t)^{\Re(z)+3} + 3(2t)^{\Re(z)} + t^{\Re(z)+3}}{h^3|z+1||z+2||z+3|} \\
&= \frac{113t^{\Re(z)+3}h^{-3}}{|z+1||z+2||z+3|}. \qquad (2.18)
\end{aligned}$$

Moreover, since $\Re(\gamma(v))>0$ for all v (i.e., $C<\varepsilon\le 1$ necessarily), for all $k\in\mathbb{N}$, one has that

$$|\gamma(v)+k| \ge \left(|v|^2+k^2\right)^{1/2} \ge \left(|v|^2+1\right)^{1/2} \ge \frac{1+|v|}{\sqrt{2}},$$

where the last inequality is simply a rearrangement of $(|v|-1)^2\ge 0$. Applying this in (2.18) and combining it with (2.17) thus yields

$$|\Delta_\pm\big(t,h,\gamma(v)\big)| \le t^{\Re(\gamma(v))}\min\left(2, \frac{226\sqrt{2}\,t^3}{h^3\,(1+|v|)^3}\right).$$

From (2.15), one can easily deduce that $\left|\gamma(v)f\big(\gamma(v)\big)\right|^{-1} = \mathcal{O}\big((1+|v|)^{1+2r}\big)$, for all $v\in\mathbb{R}$. Combining the previous three inequalities and using the boundedness of $|\gamma'(v)|$ gives the following.

$$|A_{1/t}| = \frac{t}{H} + \mathcal{O}(h) + \mathcal{O}\left(\int_{-\infty}^{\infty} (1+|v|)^{1+2r} t^{\Re(\gamma(v))} \min\left(1, \frac{t^3}{(1+|v|)^3}\right)\mathrm{d}v\right).$$

Because the integral is symmetric in v, it suffices to estimate the integral from 0 to ∞. Recalling $\kappa(v) = D\min(1, |v|^{-2-2r})$, the previous equation simplifies to the following.

$$\begin{aligned}\frac{|A_{1/t}|}{t} - \frac{1}{H} &= \mathcal{O}\left(\frac{h}{t}\right) + \mathcal{O}\left(\int_0^\infty (1+v)^{1+2r} t^{-\kappa(v)} \min\left(1, \frac{(t/h)^3}{(1+v)^3}\right) \mathrm{d}v\right) \\ &= \mathcal{O}\left(\frac{h}{t}\right) + \mathcal{O}\left(\int_0^\infty (1+v)^{-2+2r} t^{-\kappa(v)} \min\left((1+v)^3, \left(\frac{t}{h}\right)^3\right) \mathrm{d}v\right) \\ &= \mathcal{O}\left(\frac{h}{t}\right) + \mathcal{O}\left(\int_1^\infty v^{-2+2r} t^{-\kappa(v-1)} \min\left(v^3, \left(\frac{t}{h}\right)^3\right) \mathrm{d}v\right).\end{aligned}$$

Now let $\delta \in (0, 1-2r)$. For $v, t \geq 1$, we have, since κ is decreasing on $[0, \infty)$,

$$\begin{aligned}v^{-2+2r} t^{-\kappa(v-1)} &\leq v^{-2+2r} t^{-\kappa(v)} \\ &= v^{-2+2r} \exp(-\kappa(v)\log(t)) \\ &= v^{-2+2r+\delta} \exp(-\kappa(v)\log(t) - \delta\log(v)) \\ &= v^{-2+2r+\delta} \exp\left(-Dv^{-2-2r}\log(t) - \delta\log(v)\right) \\ &= v^{-2+2r+\delta} \exp(-\xi_\delta(t)),\end{aligned}$$

where

$$\begin{aligned}\xi_\delta(t) &:= \inf_{v\geq 1}\left(Dv^{-2-2r}\log(t) + \delta\log(v)\right) \\ &= \frac{\delta}{2+2r}\left(1 + \log\left(\frac{D(2+2r)\log(t)}{\delta}\right)\right).\end{aligned}$$

This last equality holds for all t sufficiently large, by elementary calculus. Hence

$$\frac{|A_{1/t}|}{t} - \frac{1}{H} = \mathcal{O}\left(\frac{h}{t}\right) + \mathcal{O}\left(e^{-\xi_\delta(t)} \int_1^\infty v^{-2+2r+\delta} \min\left(v^3, \left(\frac{t}{h}\right)^3\right) \mathrm{d}v\right).$$

Now, writing $\omega = vh/t$ and substituting, the integral becomes

$$\left(\frac{t}{h}\right)^{2+2r+\delta} \int_{h/t}^\infty \omega^{-2+2r+\delta} \min(\omega^3, 1)\, \mathrm{d}\omega.$$

This last integral can be split into two parts,

$$\int_{h/t}^1 \omega^{1+2r+\delta}\, \mathrm{d}\omega + \int_1^\infty \omega^{-2+2r+\delta}\, \mathrm{d}\omega,$$

both of which are finite, since $2r + \delta \in (0, 1)$. Therefore, for all $t \geq 1$,

$$\frac{|A_{1/t}|}{t} - \frac{1}{H} = \mathcal{O}\left(\frac{h}{t}\right) + \mathcal{O}\left(e^{-\xi_\delta(t)}\left(\frac{t}{h}\right)^{2+2r+\delta}\right).$$

Finally, choosing $h(t) = \frac{t}{3}\exp(\frac{-\xi_\delta(t)}{3+2r+\delta})$ ensures both terms have the same order of magnitude. More explicitly, this gives the required expression:

$$\begin{aligned}|A_{1/t}| &= \frac{t}{H} + \mathcal{O}\left(e^{-\xi_\delta(t)/(3+2r+\delta)}\right) \\ &= \frac{t}{H} + \mathcal{O}\left(t \log(t)^{-P(r,\delta)}\right),\end{aligned}$$

where

$$0 < P(r, \delta) = \frac{\delta}{(2+2r)(3+2r+\delta)} < P^*,$$

where, on account of $P(r, \delta)$ increasing in δ,

$$P^* := \sup_{0<\delta<1-2r} P(r, \delta) = P(r, 1-2r) = \frac{1-2r}{8(1+r)}.$$

$\square$

Remark 6. The above lemma is the only part of the proof of Theorem 3 that uses that $r < 1/2$. Indeed, one can use the same method to prove an analogous asymptotic formula for larger values of r. In this case, to ensure decay of the contour integrals over U_T and L_T, and to guarantee the integrals at the end of the proof converge, one should consider the following adaptation of the integrand of $F(t)$:

$$F^*(z) = \frac{-t^{z+n}}{z(z+1)(z+2)\cdots(z+n)f(z)},$$

where $n = \lceil 2 + 2r \rceil$.

The concluding stages of the proof of Theorem 3 are similar to those of Theorem 2, but with some differences.

Proof of Theorem 3. Let $\lambda \in (0, 1)$ and $I = [0, b)$, and recall $V = \bigcup_{k\geq 1} V_k$ and $\mathcal{V} \subset I$ from the proof of Theorem 2 (see page 29 onwards). Also recall that

$$\sum_{\boldsymbol{v}\in V} \alpha_{\boldsymbol{v}} = \mathrm{Leb}(\mathcal{V}) = b$$

and $\mu_\lambda(\mathcal{V}) \geq \mu_\lambda(I) - \mu_\lambda(K)$, where

$$K = \{T_{\mathrm{It}_n(b)}(0) \mid n \in \mathbb{N}_0,\ \mathrm{It}_n(b) \text{ exists}\}.$$

In particular, the proof of (2.11) from Theorem 2 shows that

$$\mu_\lambda(K) \leq \frac{1}{|X_\lambda|}\left(\left\lfloor \frac{\log(\lambda)}{\log(\alpha_1)} \right\rfloor + 1\right),$$

i.e., $\mu_\lambda(K) = \mathcal{O}(-\lambda \log(\lambda))$ as $\lambda \to 0^+$, so decays much faster than required. We now consider the familiar two cases.

Case 1: *No $i \in \mathcal{I}$ fixes 0.* Recalling $\alpha_* = \max_i(\alpha_i)$, we have the following decomposition, accounting for the divergence of $\log(t)^{-1}$ at $t = 1$:

$$\begin{aligned} \mu_\lambda(I) - \|I\| - \mu_\lambda(K) &= \sum_{\boldsymbol{v}\in V} \mu_\lambda(T_{\boldsymbol{v}}[0,1)) - \alpha_{\boldsymbol{v}} \\ = \sum_{\substack{\boldsymbol{v}\in V \\ \lambda \leq \alpha_{\boldsymbol{v}} < \lambda/\alpha_*}} \frac{|X_{\lambda/\alpha_{\boldsymbol{v}}}|}{|X_\lambda|} &+ \sum_{\substack{\boldsymbol{v}\in V \\ \alpha_{\boldsymbol{v}} \geq \lambda/\alpha_*}} \left(\frac{|X_{\lambda/\alpha_{\boldsymbol{v}}}|}{|X_\lambda|} - \alpha_{\boldsymbol{v}}\right) - \sum_{\substack{\boldsymbol{v}\in V \\ \alpha_{\boldsymbol{v}} < \lambda/\alpha_*}} \alpha_{\boldsymbol{v}}. \end{aligned} \tag{2.19}$$

We estimate each sum separately. The third sum decays favourably, as in the proof of Theorem 2:

$$\begin{aligned} \sum_{\substack{\boldsymbol{v}\in V \\ \alpha_{\boldsymbol{v}} \leq /\alpha_*}} \alpha_v &\leq \left(\frac{\lambda}{\alpha_{\boldsymbol{v}}}\right)^{\varepsilon} \sum_{\boldsymbol{v}\in V} \alpha_{\boldsymbol{v}}^{1-\varepsilon} \\ &\leq \lambda^\varepsilon \frac{\alpha_*^{-\varepsilon}}{1-\alpha_*} \sum_{i\in\mathcal{I}} \alpha_i^{1-\varepsilon} = \mathcal{O}(\lambda^\varepsilon) \end{aligned}$$

as $\lambda \to 0^+$.

Now, assuming that b has an infinite itinerary, recalling that $V_n \subset \{\mathrm{It}_{n-1}(b)\} \times \mathcal{I}$ for each n, and exploiting the fact that $|X_\lambda| = 1$ for $\alpha_* < \lambda \leq 1$ gives that

$$\begin{aligned} \sum_{\substack{v\in V \\ \lambda \leq \alpha_{\boldsymbol{v}} < \lambda/\alpha_*}} |X_{\lambda/\alpha_{\boldsymbol{v}}}| = \quad & \left|\{\boldsymbol{v} \in V : \lambda \leq \alpha_{\boldsymbol{v}} < \lambda/\alpha_*\}\right| \\ &= \sum_{n\in\mathbb{N}} \left|\{\boldsymbol{v} \in V_n : \lambda \leq \alpha_{\boldsymbol{v}} < \lambda/\alpha_*\}\right| \\ &\leq \sum_{n\in\mathbb{N}} \left|\{i \in \mathcal{I} : \lambda \leq \alpha_{\mathrm{It}_{n-1}(b)}\alpha_i < \lambda/\alpha_*\}\right| \end{aligned}$$

(continued on next page)

(continued from previous page)

$$\begin{aligned}
&= \sum_{n\in\mathbb{N}} \big|\{i \in \mathcal{I} : \lambda/\alpha_{\mathrm{It}_{n-1}(b)} \le \alpha_i < \lambda/(\alpha_*\alpha_{\mathrm{It}_{n-1}(b)})\}\big| \\
&\le \sum_{n\in\mathbb{N}} \big|\{i \in \mathcal{I} : \lambda/\alpha_{\mathrm{It}_{n-1}(b)} \le \alpha_i < \lambda/\alpha_{\mathrm{It}_{n}(b)}\}\big| \\
&= \qquad \big|\{i \in \mathcal{I} : \lambda \le \alpha_i\}\big| \\
&\le \lambda^{\varepsilon-1} \sum_{i\in\mathcal{I}} \alpha_i^{1-\varepsilon}.
\end{aligned}$$

The finite itinerary case follows similarly, except that, after the first inequality, the sum on the right hand side is finite, and hence leads to the same inequality. Therefore, dividing both sides by $|X_\lambda|$, we see that the first sum is also $\mathcal{O}(\lambda^\varepsilon)$.

It thus remains to bound the first sum of (2.19), using the asymptotics for $|X_\lambda| = |A_\lambda|$ provided by the previous lemma. More explicitly, given P as in the statement of the theorem, there exists C such that, for any $\boldsymbol{v} \in V$ with $\alpha_{\boldsymbol{v}} \ge \lambda/\alpha_*$, the following holds (in particular, the leading order terms cancel):

$$\begin{aligned}
\big||X_{\lambda/\alpha_{\boldsymbol{v}}}| - \alpha_{\boldsymbol{v}}|X_\lambda|\big| &\le \frac{C\alpha_{\boldsymbol{v}}}{\lambda}\left(\big(-\log(\lambda/\alpha_{\boldsymbol{v}})\big)^{-P} + \big(-\log(\lambda)\big)^{-P}\right) \\
&\le C\alpha_{\boldsymbol{v}}\big(-\log(\lambda)\big)^{-P}\left(1 + \max_{\lambda\le\alpha_*\alpha_{\boldsymbol{v}}}\left(\frac{\log(\lambda)}{\log(\lambda/\alpha_{\boldsymbol{v}})}\right)^P\right).
\end{aligned}$$

In fact, for $\lambda \le \alpha_*\alpha_{\boldsymbol{v}}$ (since $\alpha_* \ge \alpha_{\boldsymbol{v}}$),

$$\frac{\log(\lambda)}{\log(\lambda/\alpha_{\boldsymbol{v}})} = 1 + \frac{\log(\alpha_{\boldsymbol{v}})}{\log(\lambda) - \log(\alpha_{\boldsymbol{v}})} \le 1 - \frac{\log(\alpha_{\boldsymbol{v}})}{\log(\alpha_*)} \le 2.$$

Therefore, summing the previous inequality over $\boldsymbol{v} \in V$, one bounds the second term above by an expression of the order of $(-\log(\lambda))^{-P}$. This completes the proof in this case.

Case 2: *There exists* $1 \in \mathcal{I}$ *with* $T_1(0) = 0$. As in the corresponding case in the proof of Theorem 2, this case is similar to the above, with some modifications. As in the previous proof,

$$\begin{aligned}
\sum_{\substack{\boldsymbol{v}\in V\\ \alpha_{\boldsymbol{v}}>\lambda}} \mu_\lambda(T_{\boldsymbol{v}}[0,1)) &\le \mu_\lambda\{T_{\mathrm{It}_n(b)}(0) \mid n \in \mathbb{N}_0,\ \mathrm{It}_n(b) \text{ exists}\} \\
&= \mu_\lambda(K) = \mathcal{O}(\lambda\log(\lambda)),
\end{aligned}$$

where K is recalled from the start of the proof. Thus, up to a expression of the order λ^ε, we may write, similarly to the previous case,

$$\begin{aligned}\mu_\lambda(I) - b &= \sum_{\substack{v\in V\\ \alpha_{\boldsymbol{v}}\alpha_1\alpha_*<\lambda\le\alpha_{\boldsymbol{v}}}} \frac{|X_{\lambda/\alpha_{\boldsymbol{v}}}|}{|X_\lambda|} + \sum_{\substack{v\in V\\ \lambda\le\alpha_{\boldsymbol{v}}\alpha_1\alpha_*}} \frac{|X_{\lambda/\alpha_{\boldsymbol{v}}}| - \alpha_{\boldsymbol{v}}|X_\lambda|}{|X_\lambda|} - \sum_{\substack{v\in V\\ \lambda>\alpha_{\boldsymbol{v}}\alpha_1\alpha_*}} \alpha_{\boldsymbol{v}} \\ &= \mathcal{O}(\lambda)\Bigg[\sum_{\substack{v\in V\\ \alpha_{\boldsymbol{v}}\alpha_1\alpha_*<\lambda\le\alpha_{\boldsymbol{v}}}} |X_{\lambda/\alpha_{\boldsymbol{v}}}| + \sum_{\substack{v\in V\\ \lambda\le\alpha_{\boldsymbol{v}}\alpha_1\alpha_*}} |X_{\lambda/\alpha_{\boldsymbol{v}}}| - \alpha_{\boldsymbol{v}}|X_\lambda|\Bigg] - \sum_{\substack{v\in V\\ \lambda>\alpha_{\boldsymbol{v}}\alpha_1\alpha_*}} \alpha_{\boldsymbol{v}}.\end{aligned} \tag{2.20}$$

By an analogy with the previous case, we have

$$\sum_{\substack{v\in V\\ \lambda>\alpha_{\boldsymbol{v}}\alpha_1\alpha_*}} \alpha_{\boldsymbol{v}} = \mathcal{O}(\lambda^\varepsilon),$$

and also that, for $m \in \mathbb{N}$ such that $\alpha_*^{m-1} \le \alpha_1$, assuming that b has an infinite itinerary,

$$\begin{aligned}\sum_{\substack{v\in V\\ \alpha_{\boldsymbol{v}}\alpha_1\alpha_*\le\lambda}} |X_{\lambda/\alpha_{\boldsymbol{v}}}| &\le \sum_{\substack{v\in V\\ \alpha_{\boldsymbol{v}}\alpha_1\alpha_*<\lambda\le\alpha_{\boldsymbol{v}}}} |X_{(\alpha_1\alpha_{\boldsymbol{v}})^{-1}}| \\ &= |X_{(\alpha_1\alpha_{\boldsymbol{v}})^{-1}}| \sum_{k=1}^{\infty} |\{v\in V_k \mid \alpha_{\boldsymbol{v}}\alpha_1\alpha_* < \lambda \le \alpha_{\boldsymbol{v}}\}| \\ &\le |X_{(\alpha_1\alpha_{\boldsymbol{v}})^{-1}}| \sum_{k=1}^{\infty} |\{v\in V_k \mid \alpha_{\boldsymbol{v}}\alpha_*^m < \lambda \le \alpha_{\boldsymbol{v}}\}| \\ &\le |X_{(\alpha_1\alpha_{\boldsymbol{v}})^{-1}}| \sum_{k=0}^{\infty} |\{i\in \mathcal{I} \mid \alpha_{\mathrm{It}_k(b)}\alpha_*^m\alpha_i < \lambda \le \alpha_{\mathrm{It}_k(b)}\alpha_i\}| \\ &\le |X_{(\alpha_1\alpha_{\boldsymbol{v}})^{-1}}| \sum_{k=1}^{\infty} |\{i\in \mathcal{I} \mid \alpha_{\mathrm{It}_{k+m}(b)}\alpha_i < \lambda \le \alpha_{\mathrm{It}_k(b)}\alpha_i\}| \\ &\le |X_{(\alpha_1\alpha_{\boldsymbol{v}})^{-1}}| \sum_{k=1}^{\infty} \min(k,m)|\{i\in \mathcal{I} \mid \alpha_{\mathrm{It}_{k+1}(b)}\alpha_i < \lambda \le \alpha_{\mathrm{It}_k(b)}\alpha_i\}| \\ &\le m|X_{(\alpha_1\alpha_{\boldsymbol{v}})^{-1}}|\,|\{i\in\mathcal{I} \mid \alpha_i \ge \lambda\}| \\ &\le \left(m|X_{(\alpha_1\alpha_{\boldsymbol{v}})^{-1}}| \sum_{i\in\mathcal{I}} \alpha_i^{1-\varepsilon}\right)\lambda^{\varepsilon-1}.\end{aligned}$$

the finite itinerary case is again similar (with a finite sum) and yields the same result.

Therefore, the first and third sums in (2.20) are $\mathcal{O}(\lambda^\varepsilon)$ and it remains to estimate the first using the asymptotics of Lemma 12. Considering a typical summand, we have, by Lemmas 1 and 12, given P as in the latter, there exists $C > 0$ such that, for all $\lambda \leq \alpha_{\boldsymbol{v}}\alpha_1\alpha_*$,

$$\begin{aligned}
|X_{\lambda/\alpha_{\boldsymbol{v}}}| - \alpha_{\boldsymbol{v}}|X_\lambda| &= |A_{\lambda/\alpha_{\boldsymbol{v}}}| - |A_{\lambda/(\alpha_{\boldsymbol{v}}\alpha_1)}| - \alpha_{\boldsymbol{v}}|A_\lambda| + \alpha_{\boldsymbol{v}}|X_{\lambda/\alpha_1}| \\
&= \frac{\alpha_{\boldsymbol{v}}}{\lambda H} + \mathcal{O}\left(\frac{\alpha_{\boldsymbol{v}}}{\lambda}(-\log(\lambda/\alpha_{\boldsymbol{v}}))^{-P}\right) - \frac{\alpha_{\boldsymbol{v}}\alpha_1}{\lambda H} + \mathcal{O}\left(\frac{\alpha_{\boldsymbol{v}}\alpha_1}{\lambda}(-\log(\lambda/(\alpha_{\boldsymbol{v}}\alpha_1)))^{-P}\right) \\
&\quad - \frac{\alpha_{\boldsymbol{v}}}{\lambda H} + \mathcal{O}\left(\frac{\alpha_{\boldsymbol{v}}}{\lambda}(-\log(\lambda))^{-P}\right) + \frac{\alpha_{\boldsymbol{v}}\alpha_1}{\lambda H} + \mathcal{O}\left(\frac{\alpha_{\boldsymbol{v}}\alpha_1}{\lambda}(-\log(\lambda/\alpha_1))^{-P}\right) \\
&\leq \frac{C\alpha_{\boldsymbol{v}}}{\lambda}\Big((-\log(\lambda/\alpha_{\boldsymbol{v}}))^{-P} + (-\log(\lambda/(\alpha_v\alpha_1)))^{-P} \\
&\quad + (-\log(\lambda))^{-P} + (-\log(\lambda/\alpha_1))^{-P}\Big) \\
&\leq \frac{7C\alpha_{\boldsymbol{v}}}{\lambda}(-\log(\lambda))^{-P},
\end{aligned}$$

the last inequality following by a similar argument to that in the previous case. Hence, summing over $\boldsymbol{v}$ gives that the second term of (2.20) is $\mathcal{O}((-\log(\lambda))^{-P})$ for $\lambda \in (0,1)$, completing the proof of this case. □

2.5 Examples and non-examples in the infinite mean case

The previous theorems all explicitly depend on summability conditions on $\{\alpha_i\}_{i\in\mathcal{I}}$ in their proofs, i.e.,

$$-\sum_{i\in\mathcal{I}} \alpha_i \log(\alpha_i) < \infty \qquad \text{or} \qquad \sum_{i\in\mathcal{I}} \alpha_i^{1-\varepsilon}$$

for some $\varepsilon > 0$, suggesting some connection between the discrepancy, or more generally equidistribution, and these summability properties of the α_i.

In this section, we begin to investigate this connection by going in the opposite direction, by considering examples for which

$$-\sum_{i\in\mathcal{I}} \alpha_i \log(\alpha_i) = \infty$$

in the rank one case. Since the left hand side corresponds to the mean of the measure appearing in the renewal equation (2.5) for $e^{-t}|A_{e^{-t}}|$, or the measure-theoretic entropy of the original partition $\mathcal{P} = \mathcal{P}_1$ with respect to the Lebesgue measure, it is natural to refer to this setting as the *infinite mean* or *infinite entropy* case.

Firstly, following a construction of Erdős and De Bruijn [31, §4] gives rise to the following negative result. We expect that a more conscientious adaptation should provide examples for which the discrepancy decays arbitrarily slowly.

Theorem 4. *There exist partitions $\mathcal{P}$ such that the corresponding set of lengths $\{\alpha_i\}_{i\in\mathcal{I}}$ is rank one and X_λ is not equidistributed as $\lambda \to 0^+$.*

Proof of Theorem 4. The construction above hinges on the following two observations, given $(\lambda_k)_{k\in\mathbb{N}}$ and $u_0 := 1$, $u_n := \sum_{k=1}^{n} \lambda_k u_{n-k}$ (i.e., as in Lemma 4):

1. The value of u_n depends only on $\lambda_1, \dots, \lambda_n$, for each $n \in \mathbb{N}$.

2. If $\lambda_k \geq 0$ and $\sum_{k=1}^{\infty} \lambda_k < 1$, then $(u_n) \to 0$.

The first point is obvious, and the second follows from a basic inequality: Since $\lambda_k \geq 0$ implies $u_k \geq 0$, we have

$$\sum_{k=0}^{m} u_k = 1 + \sum_{k=1}^{m}\sum_{j=1}^{k} \lambda_j u_{k-j} = 1 + \sum_{j=1}^{m} \lambda_j \sum_{k=0}^{m-j} u_k \leq 1 + \sum_{j=1}^{m} \lambda_j \sum_{k=0}^{m} u_k,$$

which rearranges to

$$\sum_{k=0}^{m} u_k \leq \left(1 - \sum_{k=1}^{m} \lambda_j\right)^{-1},$$

i.e., taking $m \to \infty$, the series on the left is bounded, hence convergent.

Applying these two points repeatedly, we construct $(\lambda_k)_{k=1}^{\infty}$ such that

- $\lambda_k 3^k \in \mathbb{N}$, for each $k \in \mathbb{N}$;
- $\sum_{k=1}^{\infty} \lambda = 1$; and
- $\big(u_{n+1}/u_n\big)_{n=0}^{\infty}$ is unbounded.

To construct this sequence, we define $\lambda_k \geq 3^{-k}$ inductively (i.e., one index at a time), with the occasional boost when u_k becomes small.

First, iteratively defining λ_k (and thereby u_k) by $\lambda_k = 3^{-k}$, by 2, there exists $K_1 > 1$ such that $u_{K_1} < \frac{1}{9}$, and we set $\lambda_{K_1+1} = 3^{-K_1-1} + \frac{1}{3}$.

More generally, given $(K_j)_{j=1}^{n}$ and $(\lambda_k)_{k=0}^{K_n+1}$ satisfying

- $K_{j+1} > K_j$ for all $j \leq n$ (in particular, $K_j \geq j$);
- For all $k \leq K_n + 1$ and $j \leq k$,

$$\lambda_k = \begin{cases} 3^{-k} + 3^{-j}, & k = K_j + 1; \\ 3^{-k}, & \text{otherwise}; \end{cases} \tag{2.21}$$

- $u_{K_j} < 9^{-j}$ for each $j \leq n$;

we continue to set $\lambda_k = 3^{-k}$ for $k > K_n + 1$, up to the first index $K_{n+1} > K_n$ satisfying $u_{K_{n+1}} < 9^{-n-1}$. This index exists, because if one extends this definition indefinitely, ② applies to show that $(u_k) \to 0$. One puts $\lambda_{K_{n+1}+1} = 3^{-k} + 3^{-n-1}$.

By induction, this gives rise to the sequences $(\lambda_k)_k$ and $(K_n)_n$ satisfying these last three bullet points for arbitrarily large n. From these, we derive the desired properties of (λ_k) mentioned above, as follows:

Firstly, since (K_n) is strictly increasing, $K_n \geq n$ and hence, looking at (2.21), $\lambda_k 3^k \in \mathbb{N}$ for each k. Secondly, it should be clear from (2.21) that $\sum_{k=1}^{\infty} \lambda_k = 2\sum_{j=1}^{\infty} 3^{-j} = 1$. Finally, since $u_{K_n} \leq 9^{-n}$ and $\lambda_{K_n+1} > 3^{-n}$ for each n,

$$\frac{u_{K_n+1}}{u_{K_n}} \geq \frac{\lambda_{K_n+1}}{u_{K_n}} > \frac{3^{-n}}{9^{-n}} = 3^n,$$

and hence (u_{n+1}/u_n) is unbounded.

Now to derive the result. Let $\mathcal{P}$ be any partition comprising $\lambda_n 3^n$ intervals of length 3^{-n} for each n, and assume for simplicity that these intervals accumulate at 0, so that none of the similarities corresponding to $\mathcal{P}$ fix 0. In the earlier notation, we have $\{\alpha_i\}_{i\in\mathcal{I}} \subset \{3^{-n}\}_{n\in\mathbb{N}}$ (i.e., $\{\alpha_i\}$ is rank one), such that

$$\left|\{i \in \mathcal{I} : \alpha_i = 3^{-n}\}\right| = \lambda_n 3^n.$$

In particular, the renewal equation (2.4) for $z(n) = 3^{-n}|A_{3^{-n}}|$ reads $z(n) = 1$ and

$$z(n) = \sum_{i\in\mathcal{I}} \alpha_i z\big(n + \log_3(\alpha_i)\big) = \sum_{k=1}^{n} (\lambda_k 3^k) 3^{-k} z(n-k) = \sum_{k=1}^{n} \lambda_k z(n-k),$$

i.e., $3^{-n}|A_{3^{-n}}| = u_n$ above.

Let $I = [a, b)$ denote a half-open interval in $\mathcal{P}$ of length $\frac{1}{3}$ (which exists since $\lambda_1 > 0$). Then, by Lemmas 1 and 2, its measure is given by

$$\mu_{3^{-n}}(I) = \frac{|A_{3^{1-n}}|}{|A_{3^{-n}}|} = \frac{1}{3}\frac{u_{n-1}}{u_n},$$

if $n \geq 2$. This, by construction, has a subsequence converging to 0. Hence $\mu_{3^{-n}}(I) \not\to \frac{1}{3} = \|I\|$ and thus, L_n (i.e., X_λ) does not equidistribute, as required. $\square$

Contrary to the previous result, as mentioned in a prior remark, numerous simple criteria exist which guarantee, in the general language above, that $(u_{n+1}/u_n) \to 1$. Whenever this is satisfied for $u_n = x^n|A_{x^n}|$, we have that $\mu_{x^n}(I) \to$

$\|I\|$ for all intervals I appearing in the substitution scheme, which extends to equidistribution on the whole interval (see the following proof for details).

The simplest criterion, given by Garsia, Orey and Rodemich [52], is

$$\limsup_{k\to\infty} \frac{\lambda_{k+1}}{\lambda_k} \leq 1, \tag{2.22}$$

or more generally, for some $N \in \mathbb{N}$,

$$\limsup_{k\to\infty} \frac{\lambda_{k+1} + \cdots + \lambda_{k+N+1}}{\lambda_k + \cdots + \lambda_{k+N}} \leq 1.$$

Other, weighted versions of this are also given in [50] and alternative (albeit unwieldy) necessary and sufficient conditions are discussed in [31].

On a different note, the Garsia–Lamperti renewal theorem [51] yields that, if $D > 0$ and $\kappa \in (\frac{1}{2}, 1)$ exist such that

$$\sum_{k\geq n} \lambda_k \sim Dn^{-\kappa},$$

then

$$u_n \sim \frac{n^{\alpha-1}}{D} \frac{\sin(\pi\kappa)}{\sin(\pi)},$$

which would also lead to the required equidistribution.

We now use the first and simplest of these criteria, (2.22), to provide families of equidistributing examples in the infinite mean case.

Proposition 5. *There exist partitions $\mathcal{P}$ with $\{\alpha_i\}_{i\in\mathcal{I}}$ rank one, such that*

$$-\sum_{i\in\mathcal{I}} \alpha_i \log(\alpha_i) = \infty,$$

and for which X_λ is uniformly distributed as $\lambda \to 0^+$.

Proof of Proposition 5. Fix $x > 0$ and $a \in (1, 2)$, and let $D^{-1} = \sum_{n\in\mathbb{N}} n^{-a}$. For $k \in \mathbb{N}$, we define $N_k \in \mathbb{Z}$ as follows:

$$N_1 := \left\lfloor \frac{D}{x} \right\rfloor,$$

and given $N_1, \ldots, N_k$, set

$$N_{k+1} := \left\lfloor \frac{D\sum_{j=1}^{k+1} j^{-a} - \sum_{j=1}^{k} N_j x^j}{x^{k+1}} \right\rfloor.$$

That is, $\sum_{j=1}^{k} N_j x^j$ approximates $D\sum_{j=1}^{k} j^{-a}$, up to a difference of at most x^k. More precisely, by induction, for all $k \in \mathbb{N}$,

$$D\sum_{j=1}^{k} j^{-a} - \sum_{j=1}^{k} N_j x^j \in [0, x^k),$$

which has the following easy consequences:

- $N_k \geq 0$ for all $k \in \mathbb{N}$.
- $\displaystyle\sum_{k=1}^{\infty} N_k x^k = D\sum_{k=1}^{\infty} j^{-a} = 1.$
- $|N_k x^k - Dk^{-a}| \leq x^k + x^{k-1}$ and thus, $N_j x^k \sim Dk^{-a}$.

In view of the second bullet point, take any partition $\mathcal{P}$ comprising N_k intervals of length x^k. In which case, one has that

$$-\sum_{i\in\mathcal{I}} \alpha_i \log(\alpha_i) = -\log(x)\sum_{k=1}^{\infty} kN_k x^k,$$

and since $kN_k x^k \sim Dk^{1-a}$, this series diverges to $+\infty$.

The renewal equation for $z(n) = x^{-n}|A_{x^n}|$ from Lemma 3 reads $z(0) = 1$ and

$$z(n) = \sum_{k=1}^{n} N_k x^k z(n-k).$$

By the third bullet point above,

$$\frac{N_{k+1}x^{k+1}}{N_k x^k} \sim \frac{k^a}{(k+1)^a} \to 1,$$

and hence the criterion (2.22) applies to give $\big(z(n+1)/z(n)\big) \to 1$. Thus, by the product rule, for any $k \in \mathbb{N}$,

$$\frac{|A_{x^{n-k}}|}{|A_{x^n}|} = x^k \frac{z(n-k)}{z(n)} \to x^k.$$

Therefore, by Lemma 2.1, either for all $n \in \mathbb{N}_0$, one has

$$\frac{|X_{x^{n-k}}|}{|X_{x^n}|} = \frac{|A_{x^{n-k}}|}{|A_{x^n}|} \to x^k,$$

or for $\alpha_1 = x^j$ (corresponding to the interval $T_1[0,1) \ni 0$), and for all $n \geq j+k$,

$$\frac{|X_{x^{n-k}}|}{|X_{x^n}|} = \frac{|A_{x^{n-k}}| - |A_{x^{n-k-j}}|}{|A_{x^n}| - |A_{x^{n-j}}|} = \frac{|A_{x^{n-k}}|/|A_{x^n}| - |A_{x^{n-k-j}}|/|A_{x^n}|}{1 - |A_{x^{n-j}}|/|A_{x^n}|} \to \frac{x^k - x^{k+j}}{1-x^j} = x^k.$$

Thus, by Lemma 2, whenever I appears as an interval in the substitution scheme, $\mu_\lambda(I) \to \|I\|$. The proof of equidistribution of X_λ as $\lambda \to 0^+$ then follows precisely along the lines of Theorem 1, from page 21. $\square$

2.6 Higher-dimensional considerations

2.6.1 Extensions of the α-Kakutani scheme into higher dimensions

In comparison to the one-dimensional setting, higher-dimensional analogues of the α-Kakutani scheme are relatively undeveloped, perhaps because there is no obviously canonical way to extend the original definition. On the other hand, this is somewhat surprising, particularly in regard to discrepancy theory, since quasi–Monte Carlo methods (i.e., numerical integration using low-discrepancy sequences, see [35] and [74]) are all the more important in higher dimensions.

We briefly recount the work done by so far, which is limited to three papers:

Volčič and Carbone consider in [24] an analogue of the α-Kakutani scheme on the hypercube $[0,1]^m$. In brief, at each stage, starting with $\{[0,1]^m\}$, one splits each cuboid of maximal volume into two, along its longest side in the fixed ratio $\alpha : 1-\alpha$. If there are multiple such edges, the one with minimal index is chosen. See Figure 2.7 for a depiction in two dimensions with $\alpha = \phi$, the reciprocal of the golden ratio. In this setting, the "left-endpoints" (i.e., those closest to zero) are proven to be uniformly distributed.

Christoph Aistleitner, Markus Hofer and Volker Ziegler consider in [4] another multidimensional generalisation, this time of LS-sequences of points (i.e., particular enumerations of the sequence of endpoints of the LS-sequences, which are low-discrepancy), on the m-dimensional hypercube. Their motivation is the Halton sequences: the Halton sequence corresponding to $b \in \mathbb{N}^m_{\geq 2}$, is simply given by

$$(x_n^{(b_1)}, x_n^{(b_2)}, \ldots, x_n^{(b_m)})_n \subset [0,1]^m,$$

recalling the definition of the van der Corput sequences $(x_n^{(b_k)})_n \subset [0,1]$. These sequences equidistribute if and only if the b_k are coprime (and indeed, if they do, they exhibit the [conjectured] optimal decay rate of discrepancy). The authors of [4] consider a similar coupling of (enumerated) LS-sequences, and obtain some preliminary negative results analogous to the non-coprime case for Halton sequences.

Since the rectangles for the example depicted in Figure 2.7 admit only finitely many base-to-height ratios, and the manner of which these rectangles are split is dictated precisely by this ratio, this instance of the Carbone–Volčič construction is a particular instance of a *prototile substitution scheme*, as considered by Smilansky [91], which we now describe in more generality.

Broadly speaking, given

- a finite collection of subsets of $\mathbb{R}^m$ (typically polygonal, but can be more exotic) known as *prototiles*, $\{P_1, \ldots, P_k\}$; and
- a set of *substitution rules*, which describe a tiling of each P_i $(i = 1, \ldots, k)$ by scale copies of the k prototiles;

this defines a *tile substitution scheme*, a sequence $(\mathcal{P})_{n=0}^{\infty}$ of tilings of P_1 as follows:

- $\mathcal{P}_0$ is the tiling of P_1 by itself.
- Given $\mathcal{P}_n$, a tiling of P_1 by tiles which are scale copies of the P_i, $\mathcal{P}_{n+1}$ is obtained by subdividing all tiles of maximal volume in $\mathcal{P}_n$ into smaller tiles according to the corresponding substitution rule.

A very simple example of a substitution rule on two prototiles is shown in Figure 2.8, and the subsequent tile substitution scheme is depicted in Figure 2.9.

The main result of [91] states that, for any tile substitution scheme satisfying a natural irreducibility assumption, any sequence of sets $(E_n)_{n=1}^{\infty}$, such that, for each $n \in \mathbb{N}$ and tile $T \in \mathcal{P}_n$, $|T \cap E_n| = 1$; the set E_n is equidistributed as $n \to \infty$.

2.6.2 An abstraction of Theorem 1

Our first contribution to the higher dimensional setting is to simply generalise the proof of Theorem 1 to the following result.

Theorem 5. *Suppose that ν is a probability measure on the set $\mathbb{X}$, and there is a collection $\{T_i, \alpha_i\}_{i\in\mathcal{I}}$ such that*

a) *for all $i \in \mathcal{I}$, $T_i : \mathbb{X} \to \mathbb{X}$ is injective;*

b) *for all $i, j \in \mathcal{I}$ distinct, $T_i(\mathbb{X}) \cap T_j(\mathbb{X}) = \emptyset$;*

c) *for all $\boldsymbol{n} \in W(\mathcal{I})$, $T_{\boldsymbol{n}}(\mathbb{X})$ is ν-measurable and $\nu(T_{\boldsymbol{n}}(\mathbb{X})) = \alpha_{\boldsymbol{n}}$;*

d) *$\alpha_i > 0$ for all $i \in \mathcal{I}$ and $\sum_{i\in\mathcal{I}} \alpha_i = 1$; and*

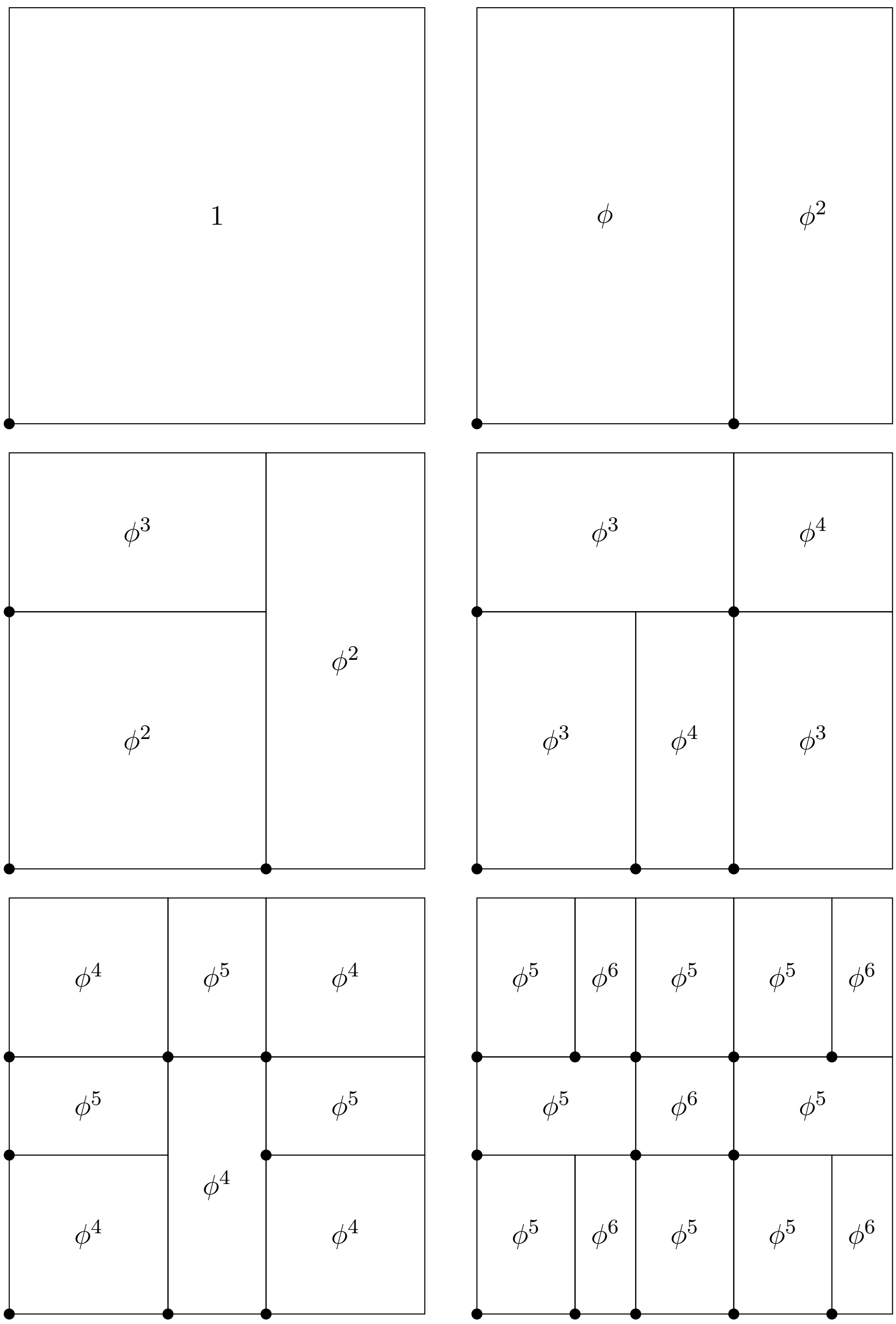


Figure 2.7: The first 6 partitions of given by Carbone–Volčič's generalisation of the α-Kakutani scheme, when the dimension is two and $\alpha = \phi$ is the reciprocal of the golden ratio. Here, the inset numbers denote areas and the circled vertices are uniformly distributed as $n \to \infty$.

Figure 2.8: A simple example of a tile substitution rule on two prototiles.

e) $-\sum_{i\in\mathcal{I}} \alpha_i \log(\alpha_i) < \infty$.

Then, for any $x \in \mathbb{X} \setminus \bigcup_{i\in\mathcal{I}} T_i(\mathbb{X})$, *and for any* ν*-measurable set* S *expressible as a disjoint union of sets from* $\{T_{\boldsymbol{n}}(\mathbb{X}) \mid \boldsymbol{n} \in \mathbb{X}\}$ *or, more generally, satisfying*

$$\nu\Bigg(\bigcup_{\substack{\boldsymbol{v}\in\mathcal{I}^n:\\ T_{\boldsymbol{v}}(\mathbb{X})\not\subset S, \mathbb{X}\setminus S}} T_{\boldsymbol{v}}(\mathbb{X}) \Bigg) \to 0 \tag{2.23}$$

as $n \to \infty$, *then*

$$\mu_\lambda(S) := \frac{|S \cap X_\lambda(x)|}{|X_\lambda(x)|} \to \nu(S),$$

as $\lambda \to 0^+$, *where* $X_\lambda(x) = \{T_{\boldsymbol{n}}(x) \,:\, \boldsymbol{n} \in W(\mathcal{I}),\ \alpha_{\boldsymbol{n}} \geq \lambda\}$.

Remark 7. The limitations of the theorem as given stem from two of the hypotheses. Firstly, assumption c), which constrains either ν or the T_i, and secondly, the requirement that x lies outside the images of the T_i, so that no combination of the T_i fix x. If ν corresponds to the Lebesgue measure, the former naturally forces the T_i to be affine.

Before we illustrate the theorem with an example, we sketch its proof.

Proof of Theorem 5. The proof precisely mirrors that of Theorem 1. Firstly, our assumption on the basepoint $x \notin \bigcup_i T_i(\mathbb{X})$ together with the injectivity of T_i and the disjointness of $T_i(\mathbb{X})$ allow us to conclude that, for all $\boldsymbol{m}, \boldsymbol{n} \in W(\mathcal{I})$,

- $T_{\boldsymbol{m}}(x) = T_{\boldsymbol{n}}(x) \iff \boldsymbol{m} = \boldsymbol{n}$; and
- $T_{\boldsymbol{m}}(x) \in T_{\boldsymbol{n}}(\mathbb{X}) \iff \boldsymbol{m} = \boldsymbol{n} * \boldsymbol{j}$, for some $\boldsymbol{j} \in W(\mathcal{I})$.

Thereby, one obtains analogues of Lemmas 1 and 2 in their simplest form: i.e., for all $\lambda > 0$ and $\boldsymbol{n} \in W(\mathcal{I})$, we have

$$|X_\lambda(x)| = |A_\lambda|, \quad \text{and} \quad |T_{\boldsymbol{n}}(\mathbb{X}) \cap X_{\lambda\alpha_{\boldsymbol{n}}}(x)| = |X_\lambda(x)|,$$

where, as before, $A_\lambda = \{\boldsymbol{n} \in W(\mathcal{I}) \mid \alpha_{\boldsymbol{n}} \leq \lambda\}$. $|A_\lambda|$ satisfies the renewal equation in Lemma 3, and the renewal theory of that section applies verbatim to show that

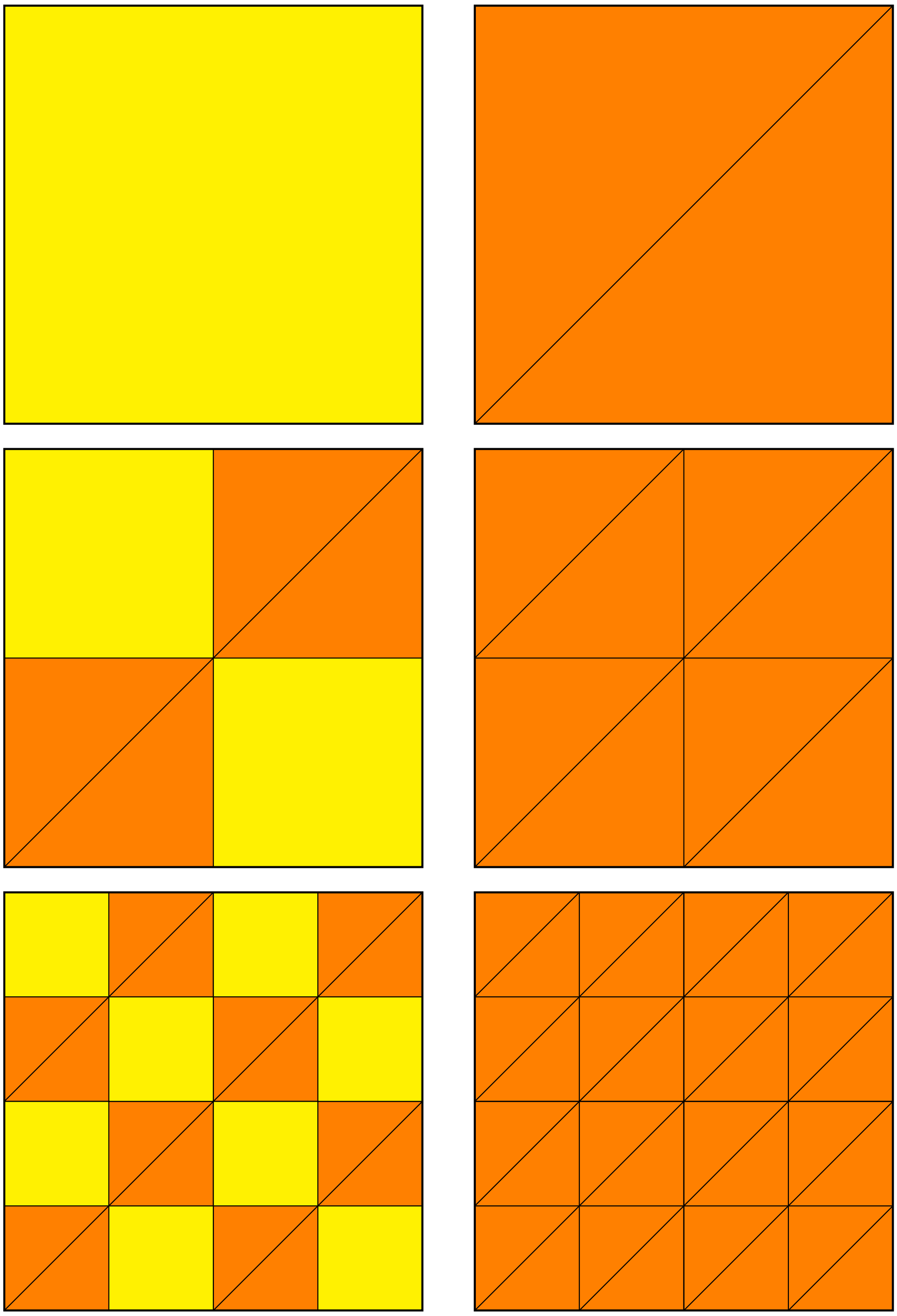


Figure 2.9: The first five tilings, $\{\mathcal{P}_n\}_{n=0}^{5}$, in the tile substitution scheme given in Figure 2.8.

$\lambda|A_\lambda|$ converges to a positive limit as $\lambda \to 0^+$. Hence, for each $\boldsymbol{n} \in W(\mathcal{I})$,

$$\mu_\lambda\big(T_{\boldsymbol{n}}(\mathbb{X})\big) = \frac{|X_{\lambda/\alpha_{\boldsymbol{n}}}(x)|}{|X_\lambda(x)|} = \frac{|A_{\lambda/\alpha_{\boldsymbol{n}}}|}{|A_\lambda|} \to \alpha_{\boldsymbol{n}} = \mu(T_{\boldsymbol{n}}(\mathbb{X})).$$

The extension to any $S \subset \mathbb{X}$ satisfying (2.23) follows as in proof of the original theorem. That is, considering unions over

$$\{T_{\boldsymbol{n}}(\mathbb{X}) \mid \boldsymbol{n} \in \mathcal{I}^n,\ T_{\boldsymbol{n}}(\mathbb{X}) \subset S\}, \qquad \{T_{\boldsymbol{n}}(\mathbb{X}) \mid \boldsymbol{n} \in \mathcal{I}^n,\ T_{\boldsymbol{n}}(\mathbb{X}) \subset \mathbb{X} \setminus S\}$$

gives, for each $n \in \mathbb{N}$,

$$\limsup_{\lambda \to 0^+} \mu_\lambda(S) - \nu(S) \le \nu\Bigg(\bigcup_{\substack{\boldsymbol{v} \in \mathcal{I}^n: \\ T_{\boldsymbol{v}}(\mathbb{X}) \not\subset S, \mathbb{X} \setminus S}} T_{\boldsymbol{v}}(\mathbb{X}) \Bigg),$$

and

$$\nu(S) - \liminf_{\lambda \to 0^+} \mu_\lambda(S) \le \nu\Bigg(\bigcup_{\substack{\boldsymbol{v} \in \mathcal{I}^n: \\ T_{\boldsymbol{v}}(\mathbb{X}) \not\subset S, \mathbb{X} \setminus S}} T_{\boldsymbol{v}}(\mathbb{X}) \Bigg).$$

Taking $n \to \infty$ thus yields the required convergence, $\mu_\lambda(S) \to S$. □

To demonstrate Theorem 5, we apply it to a novel self-affine example in the plane, inspired by a geometric progression with similar triangles (note that the corresponding maps will not be similarities, unless $h = 1$).

Example 13 (The Shark's Teeth)**.** Let $h > 0$, and consider the triangle with one side removed: $\mathbb{X} = \{(x, y) \in \mathbb{R}^2 \,|\, 0 \le x < 1,\ 0 \le y \le hx\}$, equipped with normalised area, $\nu = (2/h)\,\text{Leb}$. For $n \in \mathbb{N}$, we define $T_n : \mathbb{X} \to \mathbb{X}$ to be the unique, orientation-

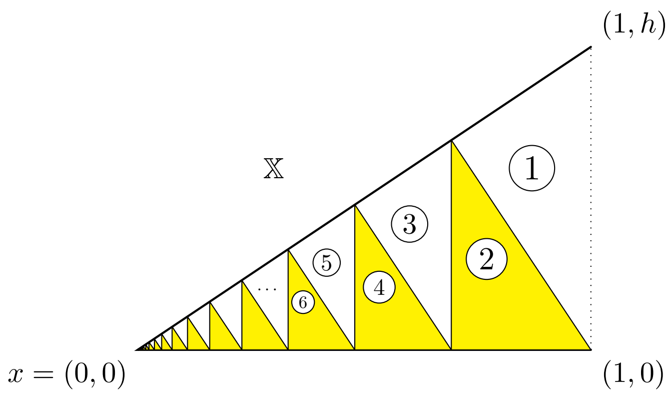


Figure 2.10: Depiction of the triangle $\mathbb{X}$, and the images $T_n(\mathbb{X})$ for $n = 1, \ldots, 6$.

preserving affine map taking $\mathbb{X}$ onto the triangle labelled n, and taking the line

$\{(1,t) \mid 0 \le t \le h\}$ onto the hypotenuse of this triangle. From these requirements, the maps can be calculated explicitly:

$$T_n(x,y) = \begin{cases} (1+h^2)^{-n/2}\left(\begin{pmatrix} h^2 & -h \\ 0 & 1 \end{pmatrix} \cdot (x,y) + (1,0) \right), & \text{if } n \text{ is even;} \\ (1+h^2)^{-(n+1)/2}\left(\begin{pmatrix} h^2 & 0 \\ -h & 1+h^2 \end{pmatrix} \cdot (x,y) + (1,h) \right), & \text{if } n \text{ is odd;} \end{cases} \tag{2.24}$$

where we denote $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \cdot (x,y) = (ax+by, cx+dy)$.

By construction, the triangles $T_n(\mathbb{X})$ are pairwise disjoint and $\bigcup_{n\in\mathbb{N}} T_n(\mathbb{X}) = \mathbb{X}\backslash\{(0,0)\}$. Furthermore from (2.24), for each $n \in \mathbb{N}$, T_n is an affine map which scales area by a factor of

$$\alpha_n = h^2(1+h^2)^{-n},$$

from which it is easy to check that assumptions a)–e) of Theorem 5 hold. The theorem thus applies to show, for $x = (0,0)$, that the measure μ_λ converges to ν on every Borel set S satisfying (2.23). (For an illustration of $X_\lambda(x)$ when $h = 1$, see Figure 2.11.)

We now claim that (2.23) is satisfied whenever S is Borel and its topological boundary, ∂S, has zero measure. Consider, for each $k \in \mathbb{N}$,

$$E_k := \{\boldsymbol{m} \in \mathcal{I}^k \mid T_{\boldsymbol{m}}(\mathbb{X}) \text{ meets both } B \text{ and } \mathbb{X}\backslash B\},$$

let $\mathcal{E}_k = \bigcup_{\boldsymbol{m}\in E_k} T_{\boldsymbol{m}}(\mathbb{X})$, so that (2.23) can be written equivalently as $\lim_{k\to\infty} \nu(\mathcal{E}_k) = 0$. Since $(\mathcal{E}_k)_{k=1}^\infty$ is a decreasing sequence of sets, this limit exists:

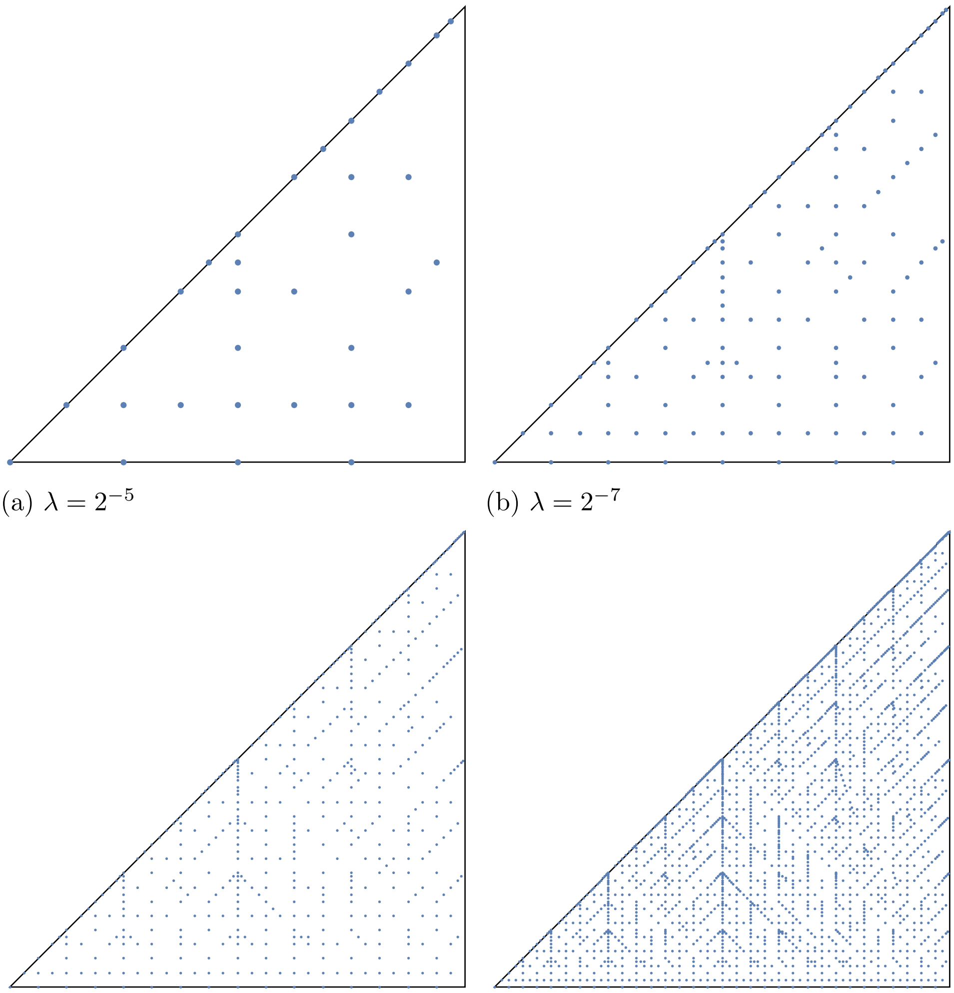
$$\lim_{k\to\infty} \nu(\mathcal{E}_k) = \nu\left(\bigcap_{k=1}^{\infty} \mathcal{E}_k \right).$$

We now show that the intersection on the right hand side lies in ∂S and thus has zero ν-measure, as required. This uses that every $p \in \mathbb{X}$ either has finite itinerary, or

$$\operatorname{diam}\left(T_{\mathrm{It}_n(p)}(\mathbb{X})\right) \to 0 \tag{2.25}$$

as $n \to \infty$ (this last property holding ν-almost everywhere would also suffice). If p lies in $\bigcap_{k=1}^\infty \mathcal{E}_k$, it must have infinite itinerary, and in particular, $\mathrm{It}_k(p) \in E_k$ for all $k \in \mathbb{N}$. That is, $T_{\mathrm{It}_k(p)}(\mathbb{X})$ meets both S and $\mathbb{X} \setminus S$ for each k, and thus (2.25) implies that $p \in \partial S$.

Finally, to prove (2.25), observe that, considering the matrices above, T_n is



(a) $\lambda = 2^{-5}$ (b) $\lambda = 2^{-7}$

(c) $\lambda = 2^{-9}$ (d) $\lambda = 2^{-11}$

Figure 2.11: $X_\lambda(x)$, for $x = (0, 0)$ and various values of λ, in the case $h = 1$.

a strict contraction for $n \geq 2$, and T_1 does not increase distances. Thus, given $p \in \mathbb{X}$ with infinite itinerary, it suffices to show that, for the sequence $(i_n)_{n\in\mathbb{N}}$ such that $\mathrm{It}_n(p) = (i_1, \ldots, i_n)$ for each n, that $i_n \neq 1$ infinitely often. Suppose for contradiction that this is not true, i.e., there exists $n \in \mathbb{N}$ such that $i_k = 1$ for all $k \geq n$. Then,

$$p' = (T_{\mathrm{It}_n(p)})^{-1}(p) \in \mathbb{X}$$

is a point with itinerary $(1, 1, \ldots)$, i.e., $p' \in T_1^k(\mathbb{X})$ for every $k \in \mathbb{N}$. But, as shown by the first coordinate of the formula

$$T_1(1 + x, y) = \left(1 + \frac{h^2 x}{h^2 + 1}, y - \frac{hx}{h^2 + 1}\right),$$

T_1 contracts points towards the line $L = \{1\} \times \mathbb{R}$. I.e., for all $(x, y) \in \mathbb{R}^2$,

$$d\big(T_1(x, y), L\big) \leq \frac{h^2}{1 + h^2}\, d\big(T_1(x, y), L\big),$$

where $d\big((x, y), L\big) = |x - 1|$ is the distance from (x, y) to L. Hence,

$$\bigcap_{k=1}^{\infty} T_1^k(\mathbb{X}) \subset L,$$

which is disjoint from $\mathbb{X}$. Thus, p' cannot exist, and one thus obtains (2.25). In summary, we have shown that $\mu_\lambda(S) \to \nu(S)$ whenever S is Borel and its topological boundary has zero area.

2.7 Conclusions and further work

There are plenty of potential avenues left to explore. The most pertinent of these is to expand Theorem 5 to a more comprehensive theory in multiple dimensions. Our viewpoint seems particularly applicable to equidistribution on fractal examples, namely iterated function schemes (such as the examples presented later in section 4.1.2).

One a more superficial level, it would be interesting to estimate discrepancy of the higher dimensional analogues of the Kakutani-Fibonacci sequence (as depicted in Figure 2.7 in two dimensions), for example.

Returning to one dimension, one possible generalisation would be the natural extension of the following result of Aistleitner and Hofer [3], which appears to follow from a simple consideration of the proof of Theorem 1.

Proposition 6. *Suppose that one has a sequence of intervals $(\mathcal{P}_k)_{k=0}^{\infty}$, for which $\mathcal{P}_{k+1}$ is the $(T_i)_{i=1}^{n}$ refinement of $\mathcal{P}_k$. Then, letting E_n denote the endpoints of $\mathcal{P}_n$, we have the following, for $\alpha_i > 0$ corresponding to T_i:*

- *If $\{\alpha_i\}_{i=1}^{n}$ are higher rank, then E_n is equidistributed as $n \to \infty$.*
- *If $\{\alpha_i\}_{i=1}^{n}$ are rank one, then E_n is equidistributed if and only if, for $x > 0$ the minimum number such that $\{\alpha_i\}_{i=1}^{n} \in \{x^k \mid k \in \mathbb{N}_0\}$, we also have*

$$\{\|I\| : I \in \mathcal{P}_0\} \subset \{x^k \mid k \in \mathbb{N}_0\}.$$

More significantly, in all of the above proofs it is clear that it is the α_i, and not the T_i, that determine equidistribution. Therefore, one should be able to conclude equidistribution for generalised interval substitution schemes, in which a maximal interval is split according to an *arbitrary* partition (one is tempted to say "random partition"), obtained by permuting the intervals in $\mathcal{P}$. This should be quite simple; whether such a generalisation affects the proofs of discrepancy is much less clear.

Other natural questions exist, such as "What happens if you alter the refinement at each stage?" (for example, if one alternates between splitting according to two different partitions).

There is no shortage of questions one can ask. This is perhaps one of the main advantages of considering low-dimensional problems which have a strong, visual component.

Chapter 3

Explicit examples of resonances for Anosov maps of the torus

3.1 Introduction to resonances

The general study of resonances is thought to originate with the article [81] of David Ruelle in 1976, who considered the zeros of a dynamical zeta functions corresponding to expanding maps on manifolds, although the term "resonances" first appears ten years later, in another article of his [82].* For our purposes, it is instructive to consider the particular case of expanding maps on the circle.

3.1.1 Resonances of expanding maps on the circle

The simplest example of an expanding map on the circle is the doubling map $T_2 : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$:

$$T_2 : x \mapsto 2x \mod 1.$$

More generally, a continuously differentiable map T of the circle is *expanding* if there exists $\lambda > 1$ such that, for all $x \in \mathbb{R}/\mathbb{Z}$,

$$|T'(x)| \geq \lambda > 1.$$

Assuming that an expanding map T is C^r, for some $r > 1$ (here, fractional r indicates that the $\lfloor r \rfloor$-th derivative of T exists and is $(r - \lfloor r \rfloor)$-Hölder), there exists a unique invariant ergodic measure, μ say, which is absolutely continuous with respect to the Lebesgue measure (*acim*). The acim μ satisfies a host of properties. Our focus

*Resonances are often referred to as *Ruelle resonances* for this reason.

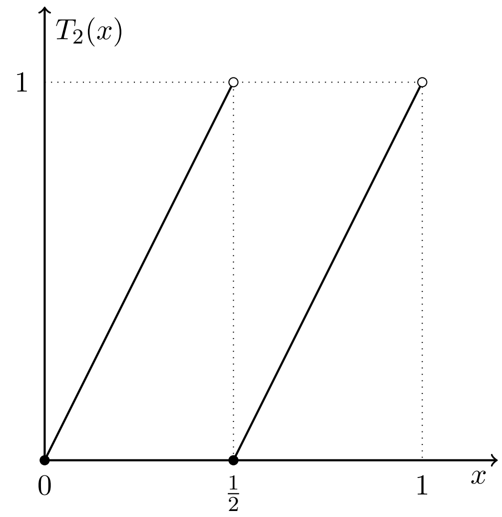


Figure 3.1: Graph of the doubling map, T_2 as a map on $[0, 1)$.

is the following result, which is stated in more generality (i.e., for expanding maps on manifolds) as [12, Corollary 2.6].

Proposition 7. *Given a C^r expanding circle map T with acim μ, such that $|T'| \geq \lambda > 1$; then, for any $0 < \beta < r - 1$, and any $\theta > \lambda^{-\beta}$, there exists $N \in \mathbb{N}_0$ and a collection of N distinct complex numbers $\{\rho_n\}_{n=1}^N$, non-increasing in modulus, such that $\theta < |\rho_n| < 1$ for each k, and, for every pair f, g of C^β functions on the circle, there exist (explicit) polynomials $(c_n)_{n=1}^N$, for which we have the following asymptotic formula for the correlation function:*

$$\int f \circ T^m g \, d\mu = \int f \, d\mu \int g \, d\mu + \sum_{n=1}^{N} c_n(m) \rho_n^m + \mathcal{O}(\theta^m) \tag{3.1}$$

as $m \to \infty$.

Remark 8. This result is very strong, in particular implying exponential decay of correlations, i.e.,

$$\int f \circ T^m g \, d\mu - \int f \, d\mu \int g \, d\mu = \mathcal{O}(\eta^m)$$

as $m \to \infty$, for all $\eta > |\rho_1|$. This in turn implies the famous Birkhoff ergodic theorem, and a host of statistical properties pertaining to μ. For example, for each $f : S^1 \to \mathbb{R}$ which is C^β for β as in the proposition, we have the following central limit theorem: for the random variable x distributed with law μ, the sequence of random variables

$$\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f(T^n(x)) - \sqrt{n}\int f \, \mathrm{d}\mu$$

converges in law to a centred normal distribution as $n \to \infty$ (see the introduction to [12] and references therein for more details).

The values ρ_n (which we henceforth consider to include 1 and 0), are called the *resonances*. Note that, if T is smooth, there can be infinitely many such resonances, which necessarily converge to 0, and the size of θ in (3.1) is limited only by the smoothness of the observables f and g.

Relationship to composition and transfer operators

In principle, the resonances are calculated as the eigenvalues of the *composition operator*, $\mathcal{C}_T(f) := f \circ T$ or its dual, the *transfer operator*,

$$\mathcal{L}f(x) := \sum_{T(y)=x} \frac{f(y)}{|T'(y)|},$$

acting on a suitable Banach space of functions, $\mathcal{B}$ say.† More explicitly, if $\mathcal{L}$ is *quasicompact* on $\mathcal{B}$, in the sense that there is some $\theta < 1$ such that $\mathcal{L}$ is the sum of a finite rank operator and one of operator norm at most θ, then the generating function of the correlation function (3.1) for each fixed $f, g \in \mathcal{B}$,

$$G_{f,g}(z) = \sum_{m=1}^{\infty} \left(\int f \circ T^m g \; d\mu \right) z^m,$$

is meromorphic on the disk of radius θ^{-1}, and within this disk, poles can only occur at the reciprocals of eigenvalues of $\mathcal{L}$, with the degree of the pole not exceeding the size of the largest Jordan block of the corresponding eigenvalue (see [12, Ch. 2] for more details).

More explicitly, writing these eigenvalues as $\{\rho_1, \ldots, \rho_N\}$, and writing i_n as the size of the largest Jordan block of $\mathcal{L}$ corresponding to ρ_n, there exist constants $a_{n,k}$ and b_m such that

$$G_{f,g}(z) = \sum_{n=1}^{N}\sum_{k=1}^{i_n} \frac{a_{n,k}}{(1-\rho_n z)^k} + \sum_{m=1}^{\infty} b_m z^m$$

†In this expanding case, $\mathcal{B}$ can simply be the completion of the set of C^β functions (with β as above) with respect to a suitable norm, e.g. the Sobolev norm.

and $b_m = \mathcal{O}(\theta^m)$ as $m \to \infty$. Applying Newton's formula, namely

$$\frac{1}{(1-\rho_n z)^k} = \sum_{m=1}^{\infty} \binom{m}{k} \rho_n^k z^k,$$

to the first sum, we recover (3.1) with the ρ_n given by the above eigenvalues of $\mathcal{L}$, and

$$c_n(m) = \sum_{k=1}^{i_n} a_{n,k} \binom{m}{k}$$

In particular, if ρ_n is a semi-simple eigenvalue of $\mathcal{L}$ (i.e., if $i_n = 1$), then c_n is constant.

A lack of explicit examples

Unfortunately, there are very few examples of expanding maps for which the resonances are known. The doubling map (or more generally, the m-tupling map) has only the trivial resonances, 0 and 1.

In contrast, Frédéric Naud proved in [72] that generic real analytic expanding maps have infinitely many non-trivial resonances, and gave lower bounds on the rate that these resonances converge to zero. More explicitly, given such a generic map, the resonances $(\rho_n)_{n=1}^{\infty}$ satisfy

$$\liminf_{n\to\infty} |\rho_n| e^{n(1-\varepsilon)} \geq 1$$

for each $\varepsilon > 0$. Indeed, this complements a general upper bound for real analytic expanding maps given by Oscar Bandtlow and Oliver Jenkinson in [14]: namely, there exists $a > 0$ such that

$$\sup_{n\in\mathbb{N}} |\rho_n| e^{an} < \infty.$$

Moreover, Bandtlow and Naud in [16] show that, for a dense subset of the collection of expanding maps analytic on an annular neighbourhood of the circle,

$$\limsup_{n\to\infty} |\rho_n| e^{n^{1+\varepsilon}} > 0$$

holds for every $\varepsilon > 0$. (The proof in fact uses the result of [87, 15] below.)

Critically, in spite of the great deal of work in this area, the only explicitly known examples of resonances remained those of the trivial case above, until the breakthrough articles [87, 15] of Oscar Bandtlow, Wolfram Just and Julia Slipantschuk, which provide families of expanding maps on the circle for which the resonances are explicit:

$$\{\rho_n\} = \{0, 1\} \cup \{\lambda^m, \overline{\lambda}^m \,:\, m \in \mathbb{N}\}, \tag{3.2}$$

where λ is an arbitrary complex parameter with $|\lambda| < 1$. Their examples are given by families of *Blaschke products* on the circle, which we now describe.

Blaschke products on the circle

Historically, Blaschke products play an important role in one-dimensional complex dynamics. They are known to be the only holomorphic maps on the open unit disk which extend to a continuous map on the unit circle, and map this circle into itself [26]. Moreover, it follows as a simple consequence of the Riemann mapping theorem that any meromorphic map whose Julia set bounds a simply connected invariant set is conjugate to a Blaschke product which is expanding on the unit circle [78].

We first make some simple definitions.

Definition 13. In this chapter, we consider the circle and the torus to be embedded in complex one- or two-space respectively. Let

$$\mathbb{T}^1 = \big\{z \in \mathbb{C} : |z| = 1\big\}, \qquad \mathbb{T}^2 = \mathbb{T}^1 \times \mathbb{T}^1 \subset \mathbb{C}^2,$$

and also let

$$\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}, \qquad \mathbb{D}_\infty = \{z \in \mathbb{C} : |z| > 1\} \cup \{\infty\},$$

considered as subsets of the Riemann sphere, $\mathbb{C} \cup \{\infty\}$.

Remark 9. Regarding changes of coordinates: whenever we need to translate between $\mathbb{T}^2$ as above and $\mathbb{R}^2/\mathbb{Z}^2$ (e.g., to show hyperbolicity), we use the natural conjugation

$$(e^{2\pi i\theta_1}, e^{2\pi i\theta_2}) \mapsto (\theta_1, \theta_2) \mod 1.$$

We define Blaschke products as maps on the Riemann sphere.

Definition 14 (Blaschke product)**.** A Blaschke product is a holomorphic map $B : \mathbb{C} \cup \{\infty\} \to \mathbb{C} \cup \{\infty\}$ of the following form, for $\lambda_1, \ldots, \lambda_m \in \mathbb{D}$ and $\theta \in \mathbb{T}^1$:

$$B : z \mapsto \theta \prod_{k=1}^{m} \frac{z + \lambda_k}{1 + \overline{\lambda_k} z}, \tag{3.3}$$

where $\overline{\lambda_k}$ denotes the complex conjugate of λ_k.

For any a Blaschke product B, as noted above, $B(\mathbb{D}) = \mathbb{D}$ and $B(\mathbb{T}^1) = \mathbb{T}^1$. (Also, $B(\mathbb{D}_\infty) = \mathbb{D}_\infty$.) In particular, B restricts to a degree m map of the circle.

Resonances of expanding Blaschke products

Following [30] and [15], we now present an informal description of the resonances of B, under the assumption that B is a Blaschke product which restricts to an expanding map on $\mathbb{T}^1$. More explicitly, we show that this restriction has resonances as given in (3.2), where λ is the multiplier of the unique fixed point of B in $\mathbb{D}$. This will inform our calculations in the Anosov case.

So as to not overly complicate the analysis here, we assume that $\mathcal{C}_B$ acts compactly on a suitable Hilbert space $\mathcal{H}$ of functions which has the set of monomials $\{e_n : z \mapsto z^n\}_{n\in\mathbb{Z}}$ as an orthogonal basis (a family of such spaces is noted below). Acting on such a space, the spectrum of $\mathcal{C}_B$ can be derived quite simply, by considering its action on the subspaces spanned by $\{e_n : n \geq 0\}$ and $\{e_n : n < 0\}$ respectively.

We first apply the following result of [95].

Lemma 13. *A Blaschke product B restricts to an expanding map on $\mathbb{T}^1$ if and only if $m \geq 2$ and $z \mapsto \alpha B(z)$ has a fixed point in $\mathbb{D}$ for every $\alpha \in \mathbb{T}^1$.*

Remark 10. An example of a non-expanding Blaschke product for $m \geq 2$ is given by

$$B(z) = \frac{3z^2+1}{z^2+3} = \frac{\left(z - \frac{i}{\sqrt{3}}\right)}{\left(1 + \frac{i}{\sqrt{3}}z\right)} \frac{\left(z + \frac{i}{\sqrt{3}}\right)}{\left(1 - \frac{i}{\sqrt{3}}z\right)},$$

which has an indifferent fixed point at $1 \in \mathbb{T}$, i.e., $B(1) = 1$ and $B'(1) = 1$.

By the lemma, if B is expanding then, up to conjugation by a Möbius map, we can assume that it has a fixed point at 0: i.e., $B(0) = 0$. That is, without loss of generality,

$$B(z) = \theta z \prod_{k=1}^{m-1} \frac{z + \lambda_k}{1 + \overline{\lambda_k} z}.$$

To consider the action of $\mathcal{C}_B$ on the subspace with basis $\{e_n : n \geq 0\}$, one formally takes a Taylor expansion about 0, exploiting the assumption that $B(0) = 0$:

$$\big(B(z)\big)^n = \big(B'(0)z + \mathcal{O}(z^2)\big)^n = \big(B'(0)\big)^n z^n + \sum_{k=1}^{\infty} a_{n,k} z^{n+k},$$

for some coefficients $a_{n,k} \in \mathbb{C}$. That is, we can formally write

$$\mathcal{C}_B(e_n) = \big(B'(0)\big)^n e_n + \sum_{k=1}^{\infty} a_{n,k} e_{n+k}. \tag{3.4}$$

This leads to the following infinite lower-triangular matrix representation for $\mathcal{C}_B$:

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ 0 & B'(0) & 0 & 0 & \cdots \\ 0 & a_{1,1} & \big(B'(0)\big)^2 & 0 & \cdots \\ 0 & a_{1,2} & a_{2,1} & \big(B'(0)\big)^3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

In particular, $\mathcal{C}_B^*$, the adjoint of $\mathcal{C}_B$, is represented in block form by

$$A^* = \begin{pmatrix} (A_n)^* & C_n \\ 0 & D_n \end{pmatrix},$$

where A_n is the $n \times n$ truncation of A and $\|D_n\|_{\mathrm{op}} \to 0$ as $n \to \infty$, by the compactness of $\mathcal{C}_B$, for $\|\cdot\|_{\mathrm{op}}$ the appropriate operator norm.‡

Now, given a non-zero eigenvalue λ of $\mathcal{C}_B^*$ (i.e., $\mathcal{C}_B$), assume that n is large enough that $\|D_n\|_{\mathrm{op}} < |\lambda|$. Then, the eigenvalue equation for $v = (v_1, v_2)$ reads

$$\lambda v_1 = (A_n)^* v_1 + C_n v_2 \qquad \text{and} \qquad \lambda v_2 = D_n v_2.$$

However, since $\|D_n\|_{\mathrm{op}} < |\lambda|$, the second equation implies $v_2 = 0$, and thus $\lambda v_1 = (A_n)^* v_1$. That is, λ is an eigenvalue of A_n, for all sufficiently large n. Since A_n is a triangular matrix, these eigenvalues are given by the entries on the diagonal, i.e., the non-negative powers of $B'(0)$. Moreover, since these values are distinct (i.e., $|B'(0)| < 1$, by the Schwarz–Pick theorem [26, p.13]), these eigenvalues are all simple.

As for the action of $\mathcal{C}_B$ on the complementary subspace spanned by $\{e_{-n} : n \in \mathbb{N}\}$, one can exploit the following symmetry of B:

$$B(z^{-1})^{-1} = \overline{\big(B(\overline{z})\big)},$$

‡That is, in this case, the operator norm on the codimension-n subspace $\mathcal{H}_n$ with basis $\{e_k \mid k \geq n\}$ (see the proof of Lemma 18 for more details).

or, expressed more explicitly (and using (3.4)),

$$\big(B(z)\big)^{-n} = \left(\theta^{-1}z^{-1}\prod_{k=1}^{m-1}\left(\frac{z+\lambda_k}{1+\overline{\lambda_k}z}\right)^{-1}\right)^n = \left(\overline{\theta}z^{-1}\prod_{k=1}^{m-1}\frac{z^{-1}+\overline{\lambda_k}}{1+\lambda_k z^{-1}}\right)^n$$
$$= \big(\overline{B'(0)}\big)z^{-n} + \sum_{k=1}^{\infty}\overline{a_{n,k}}z^{-n-k}.$$

Therefore, $\mathcal{C}_B(e_{-n}) = \big(\overline{B'(0)}\big)^n e_{-n} + \sum_{k=1}^{\infty}\overline{a_{n,k}}e_{-n-k}$, and the action of $\mathcal{C}_B$ on the subspace with basis $\{e_{-n}\}_{n\in\mathbb{N}}$ is represented by the matrix

$$\begin{pmatrix} \overline{B'(0)} & 0 & 0 & \cdots \\ \overline{a_{1,1}} & \big(\overline{B'(0)}\big)^2 & 0 & \cdots \\ \overline{a_{1,2}} & \overline{a_{2,1}} & \big(\overline{B'(0)}\big)^3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Since $\mathcal{C}_B$ is compact and this matrix is lower triangular, by the above reasoning, the restriction of the operator to this space has simple eigenvalues consisting of positive integer powers of $\overline{B'(0)}$.

Summary: Whenever there exists a Hilbert space of functions $\mathcal{H}$ which has $\{e_n\}_{n\in\mathbb{Z}}$ as an orthogonal basis and on which $\mathcal{C}_B$ acts compactly, its spectrum is given by

$$\{0,1\}\cup\big\{\big(B'(0)\big)^n, \big(\overline{B'(0)}\big)^n : n\in\mathbb{N}\big\}.$$

Moreover, each non-zero eigenvalue is simple, unless $B'(0)$ has a rational argument. In the latter case, coincidences in value occur when $\big(B'(0)\big)^n = \big(\overline{B'(0)}\big)^n$ for some n, and this eigenvalue has algebraic and geometric multiplicity 2.

Remark 11. Note that we more generally call any eigenvalue which has the same algebraic and geometric multiplicity *semi-simple*.

Examples of such a space, for each $\phi > 1$, are given by

$$\mathcal{H} = \left\{ f : z \mapsto \sum_{n=-\infty}^{\infty} a_n z^n \;\middle|\; \sum_{n=-\infty}^{\infty} |a_n|^2\phi^{-|n|} < \infty \right\},$$

equipped with the natural inner product

$$\left\langle \sum_n a_n z^n, \sum_n b_n z^n \right\rangle = \sum_n a_n \overline{b_n}\phi^{-|n|}.$$

(That $\mathcal{C}_B$ acts compactly on $\mathcal{H}$ is shown in detail in [15], and can also be derived using the elementary arguments of [84, p.23].)[§] Then, the transfer operator $\mathcal{L}$ (i.e., the dual of $\mathcal{C}_B : \mathcal{H} \to \mathcal{H}$, which has the same spectrum) acts compactly on the dual space

$$\mathcal{H}^* = \left\{ f : z \mapsto \sum_{n=-\infty}^{\infty} a_n z^n \;\middle|\; \sum_{n=-\infty}^{\infty} |a_n|^2 \phi^{|n|} < \infty \right\},$$

which contains all functions analytic on a neighbourhood of the annulus

$$\left\{ z \in \mathbb{C} : \frac{1}{\sqrt{\phi}} \le |z| \le \sqrt{\phi} \right\}.$$

Therefore, given any pair of functions f, g on T^1 which extend analytically to some neighbourhood of the torus, ϕ can be chosen sufficiently close to one so that $f, g \in \mathcal{H}^*$. By the discussion in the preceding subsections, this shows that there exist constants c_n, d_n such that, for any $N \in \mathbb{N}$, and μ the acim corresponding to $B|_{\mathbb{T}^1}$,

$$\int_{\mathbb{T}^1} f \circ B^m g \, d\mu = \int_{\mathbb{T}^1} f \, d\mu \int_{\mathbb{T}^1} g(z) \, d\mu + \sum_{n=1}^{N} c_n \lambda_n^m + d_n \overline{\lambda}^m + \mathcal{O}(|\lambda|^{N+1}),$$

i.e., the non-zero eigenvalues in the spectrum of $\mathcal{C}$ are all resonances of B. By the density of analytic functions (i.e., of polynomials), there can be no further resonances.

As a final note, since, for any $\lambda \in \mathbb{D}$,

$$B(z) = z \frac{z + \lambda}{1 + \overline{\lambda} z}$$

has $B'(0) = \lambda$, $B'(0)$ can take any value in $\mathbb{D}$. In particular, we can choose the "exponential rate of mixing" $|\rho_1| = |\lambda|$ to be arbitrarily close to the unit circle, i.e., arbitrarily slow.

3.1.2 Resonances of Anosov diffeomorphisms

The situation governing resonances of Anosov diffeomorphisms is highly analogous to that of expanding maps.

We first give the definition of an Anosov diffeomorphism, which makes concrete the informal description that such a map is characterised by "locally contracting in some directions and expanding in others".

[§] In brief: decomposing this space into two, corresponding to negative and non-negative n, each corresponds to the Hardy-Hilbert space of one of two disks (one based at zero, and one based at infinity). Then, since B can be shown to map these disks into their interiors, it follows that the composition operator acts compactly on each.

Definition 15 (Anosov)**.** Given a compact connected manifold M, we say that the diffeomorphism $T : M \to M$ is *Anosov* if it is differentiable and there exists a continuous splitting of the tangent bundle, $\mathcal{T}_x(M) = E^s_x \oplus E^u_x$, and constants $C > 0$ and $\lambda \in (0, 1)$ such that the following hold for all $x \in M$:

- $D_x T(E^s_x) = E^s_{T(x)}$, and $D_x T(E^u_x) = E^u_{T(x)}$.
- $\|D_x T^n v\| \leq C\lambda^n \|v\|$ for all $v \in E^s_x$.
- $\|D_x T^{-n} v\| \leq C\lambda^n \|v\|$ for all $v \in E^u_x$.

We also say that a diffeomorphism $T : \mathbb{T}^2 \to \mathbb{T}^2$ is *area-preserving* if it leaves the normalised Lebesgue measure (i.e., unit volume measure) invariant: that is,

$$\int_M f \circ T \, d\mu = \int_M f \, d\mu$$

for all Borel functions f on M, and μ the unit volume measure on M.

One can show, by an easy proof, that any continuous map which has a fully supported, ergodic measure is transitive. Thus, in view of the following result, the resonances for T are well-defined. Note that, if T is area-preserving, the measure μ below is simply the unit volume measure on M. (For a gentle introduction to SRB measures in general, see [101].)

Proposition 8 ([12, Theorem 7.11])**.** *For $r > 1$, suppose that T is a C^r transitive Anosov diffeomorphism on a compact connected manifold M. Then there exists a unique probability measure μ (the SRB measure corresponding to T) such that, for λ as in Definition 15, for any $0 < \beta < r - 1$, and any $\theta > \lambda^{-\beta}$, there exists $N \in \mathbb{N}_0$ and a collection of N complex numbers $\{\rho_n\}_{n=1}^N$ with $\theta < |\rho_n| < 1$ for each k, such that, for every pair f, g of C^β functions on M, there exist (explicit) polynomials $(c_n)_{n=1}^\infty$ such that*

$$\int f \circ T^m g \, d\mu = \int f \, d\mu \int g \, d\mu + \sum_{n=1}^N c_n(m)\rho_n^m + \mathcal{O}(\theta^m) \tag{3.5}$$

as $m \to \infty$.

Remark 12. Note that, since we need not reference the measure μ in what follows, we abuse notation and later use μ also as a complex parameter.

As in the expanding case, there are very few examples of Anosov diffeomorphisms for which the resonances are known. Until recently, the only known examples were given by the trivial examples of linear hyperbolic diffeomorphisms, which represent all hyperbolic diffeomorphisms up to isotopy (see [43] for a proof of this fact):

Definition 16. Given any matrix $M \in \mathbb{Z}^{n,n}$ such that $\det(M) = \pm 1$, we say that

$$\mathbb{R}^n/\mathbb{Z}^n \to \mathbb{R}^n/\mathbb{Z}^n, \qquad v \mapsto M \cdot v \mod 1$$

is a *linear* (toral) *diffeomorphism*. Note that if M is a hyperbolic matrix (i.e., it has no eigenvalues on $\mathbb{T}^1$), then the map is *hyperbolic*.

These examples, including the Arnol'd CAT map of [5] mentioned above, have only 0 and 1 as resonances, similarly to the doubling map above.

More generally, Alexander Adam [1] proved that generic small perturbations of linear diffeomorphisms yield at least one non-trivial resonance (i.e., not equal to 0 or 1). However, the only explicit non-trivial examples known to date are those provided by the authors of [87] and [88]. In particular, given B_λ as defined below, it has the following resonances:

$$\{\rho_n\} = \{0, 1\} \cup \{\lambda^m, \overline{\lambda}^m \,:\, m \in \mathbb{N}_0\}$$

(where $\lambda \in \mathbb{D}$ is again an arbitrary parameter). These B_λ are given by so-called two-dimensional Blaschke products, introduced in more generality by [78] (see also [77]). More explicitly, they are given the following definition in [88, p.2669]:

Definition 17 (B_λ)**.** For $\lambda \in \mathbb{D}$, let $B_\lambda : \mathbb{T}^2 \to \mathbb{T}^2$ be given by

$$B_\lambda : (z, w) \mapsto \left(\left(\frac{z + \lambda}{1 + \overline{\lambda} z} \right) zw, \left(\frac{z + \lambda}{1 + \overline{\lambda} z} \right) w \right).$$

This family of maps smoothly perturbs the Arnol'd CAT map, represented (on $\mathbb{T}^2$ and $\mathbb{R}^2/\mathbb{Z}^2$ respectively) by

$$B_0 : (z, w) \mapsto (z^2 w, zw) \qquad \text{or} \qquad \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \mod 1.$$

In comparison to the previous setting of expanding maps, results such as those alluded to above pose much more of a technical challenge. This is partly due to the fact that constructing a space of functions (or more generally distributions) on which the composition operator (or transfer operator) of a hyperbolic map is quasicompact is a non-trivial matter.

Such Banach spaces exist specifically for this purpose, known as *anisotropic spaces*. In each case, these have to be tailored to the diffeomorphism itself, to exploit the fact that, locally, it is expanding in some particular directions and contracting in others.

There is no canonical way to construct these spaces. A rough classification of the three main varieties of anisotropic spaces seen in the literature are given a light overview in [36] and are given a more thorough account in the survey [13] (see also references therein).

In considering families of maps such as the above, by analogy with the expanding case, it is natural to consider the composition operator acting by formal Taylor expansion on spaces comprising sums of monomials $\{e_{m,n} : \mathbb{T}^2 \to \mathbb{T}\}_{(m,n)\in\mathbb{Z}^2}$, defined by

$$e_{m,n}(z, w) = z^m w^n.$$

More explicitly, it is natural to consider the composition operator acting on anisotropic Hilbert spaces which have $\{e_{m,n}\}_{m,n}$ as an orthogonal basis. As we shall describe later in more detail, such a space is completely determined by the norms given to the individual $e_{m,n}$. The particular anisotropic space used in [88], following a construction in [44], consists in weighting the norms of the $e_{m,n}$ according to the eigenvectors of the CAT map (we will describe this in more detail in the next section).

In this chapter, following [88], we give a new account of the resonances of B_λ. In particular, we introduce new families of Hilbert spaces which allow us to simplify the analysis substantially. These spaces weight the norms of the $e_{m,n}$ according to a *degree function* (broadly speaking, a signed l^1 norm).

This approach also allows us to prove new results on the resonances of two other families, (T_λ) and $(T_\lambda \circ T_\mu)$. In particular, the resonances of the latter are studied empirically in [87, 2683–2685], and we give a rigorous proof.

Remark 13. In a related account, the resonances of all linear pseudo-Anosov maps acting on translation surfaces of genus at least 2 were recently made explicit in [48].

Contents of this chapter

In summary, to prove their result on the above resonances of B_λ, the authors of [88] take the following, important steps (as illustrated in the expanding case above):

- Providing a Hilbert space of distributions (with $\{e_{m,n}\}$ as a a basis) on which the composition operator of B_λ acts compactly.

- Providing an ordering of the basis such that the matrix of the composition operator is lower-triangular, allowing the eigenvalues to be simply read off the main diagonal.
- Showing that this space can be chosen to contain any given pair of functions analytic on a neighbourhood of the torus.

In each of the sections 3.2, 3.3 and 3.4, we follow this general strategy for three different families of maps ((B_λ), (T_λ) and $(T_\lambda \circ T_\mu)$ respectively):

- We first prove that the maps considered are Anosov and area-preserving, so that the resonances are well-defined.
- We then exhibit a corresponding family of anisotropic Hilbert spaces, and show that these can be chosen to contain any pair of functions analytic on a neighbourhood of the torus.
- We also show that the composition operator acts compactly on this space (so that the eigenvalues of this operator give the resonances of the map).
- Finally, we calculate the spectrum of the operator using a convenient, block-triangular matrix form.

Finally, in section 3.5, we make some closing remarks.

3.2 The spectrum of $\mathcal{C}_{B_\lambda}$

In this section, we prove the following result on the resonances of B_λ. As mentioned above, this result and its proof are entirely analogous to the main result of [88], and we provide a new, simplified perspective.

Theorem 6. *There exists a Hilbert space, $\mathcal{H}_a$, consisting of distributions on the torus, on which, for each $\lambda \in \mathbb{D}$, the composition operator $\mathcal{C}_{B_\lambda} : f \mapsto f \circ B_\lambda$ acts compactly and has spectrum given by*

$$\{0, 1\} \cup \{\lambda^m, \overline{\lambda}^m \, : \, m \in \mathbb{N}\}.$$

Moreover, each non-zero value is a simple eigenvalue, up to coincidences in value.

Remark 14. Note that, by “up to coincidences in value”, we mean that λ^n and $\overline{\lambda}^n$ are distinct for each $n \in \mathbb{N}$, i.e., the argument of λ is an irrational multiple of π. Otherwise, whenever $\lambda^n = \overline{\lambda}^n$, we will see that it is semi-simple, of multiplicity 2.

Remark 15. The methods of this section naturally extend to the following families of diffeomorphisms, indexed by $K \in \mathbb{N}$ and $\lambda \in \mathbb{D}$:

$$B_{\lambda,K} : (z, w) \mapsto \left(\left(\frac{z+\lambda}{1+\overline{\lambda}z} \right)^{K^2+1} w^K, \left(\frac{z+\lambda}{1+\overline{\lambda}z} \right)^K w \right),$$

considered, for each K, as a perturbation of the hyperbolic linear toral automorphism

$$B_{0,K} : (z, w) \mapsto (z^{K^2} z w^K, z^K w) \qquad \longleftrightarrow \qquad \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} K^2 & K \\ K & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \mod 1.$$

This being said, since the resonances of $B_{\lambda,K}$ equal the resonances of B_{λ^K}, these families contribute nothing new to the variety of spectra presented here. We note however, that the same space, $\mathcal{H}_a$, can be used successfully for each family, without needing to know anything additional about the eigenvectors of each matrix (compare with the construction (3.7) below).

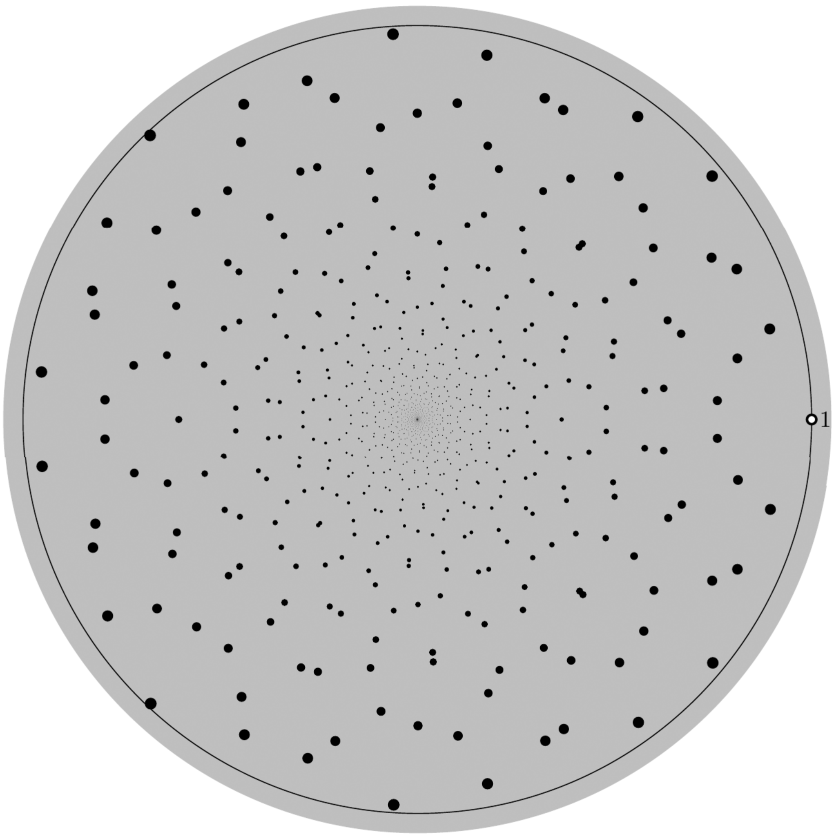


Figure 3.2: The spectrum of $\mathcal{C}_{B_\lambda}$, for $\lambda = 0.99e^{37i\pi/50}$.

3.2.1 Hyperbolicity of B_λ

Before we prove Theorem 6, we first show, for each $\lambda \in \mathbb{D}$, that B_λ is an Anosov, area-preserving map of the torus. Since B_λ is smooth, by Proposition 8, this shows that the resonances of B_λ are well-defined.

We prove the Anosov property in a standard way, using expanding and co-expanding cone families, as per the following definition (which we have made specific to the torus, following the terminology of [73]).

Definition 18 (Cone family, (co-)expanding)**.** We say that $C \subset \mathbb{R}^2$, is a *cone* if it is proper, non-empty and can be written as

$$C = \{v \in \mathbb{R}^2 \mid Q(v) \geq 0\}$$

for some real quadratic form Q. In this case, we call $Q^{-1}(-\infty, 0]$ the *complementary cone* (to C). We call a collection $\{C_x\}_{x\in M}$, where C_x is a cone in the tangent space of $x \in \mathbb{T}^2$ (which we naturally identify with $\mathbb{R}^2$) for each $x \in \mathbb{T}^2$, a *cone field*, and we say it is *expanding* with respect to the diffeomorphism $f : \mathbb{T}^2 \to \mathbb{T}^2$ if

$$D_x f(C_x) \subset \operatorname{int}(C_{f(x)}) \cup \{0\};$$

we say it is *co-expanding* with respect to f if the cone family $\{C'_x\}_{x\in M}$, where C'_x is complementary to C_x, is expanding with respect to f^{-1}.

The standard proof of the Anosov property alluded to above simply uses the following fact, after [73, Theorem 1.4].

Fact 1. A diffeomorphism f of $\mathbb{T}^2$ is Anosov if and only if there exists a cone field which is expanding and co-expanding with respect to a power of f.

This leads naturally to the following.

Proposition 9. *For each $\lambda \in \mathbb{D}$, B_λ is an area-preserving Anosov diffeomorphism.*

Proof of Proposition 9. This is a simple consideration of the tangent map $D_x B_\lambda$ in standard coordinates. In these coordinates, we have that $B_\lambda(x, y) = (x', y')$, where

$$\begin{aligned}(e^{2\pi i x'}, e^{2\pi i y'}) &= B_\lambda(e^{2\pi i x}, e^{2\pi i y}) \\ &= \left(e^{2\pi i(2x+y)}\left(e^{-2\pi i x}\frac{e^{2\pi i x} + \lambda}{1 + \overline{\lambda} e^{2\pi i x}}\right), e^{2\pi i(x+y)}\left(e^{-2\pi i x}\frac{e^{2\pi i x} + \lambda}{1 + \overline{\lambda} e^{2\pi i x}}\right)\right).\end{aligned}$$

To focus on the inner brackets: writing $\lambda = |\lambda| e^{i\theta}$, let $z \in \mathbb{R}$ satisfy

$$e^{iz} = e^{-ix} \frac{e^{ix} + \lambda}{1 + \overline{\lambda} e^{ix}} = \frac{1 + |\lambda| e^{i(\theta - x)}}{1 + |\lambda| e^{i(x-\theta)}} = \frac{\left(1 + |\lambda| e^{i(\theta-x)}\right)^2}{\left|1 + |\lambda| e^{i(x-\theta)}\right|^2}.$$

The polar argument of e^{iz} is therefore twice that of

$$1 + |\lambda| e^{i(\theta - x)}.$$

Dividing imaginary by real components, this gives

$$\tan\left(\frac{z}{2}\right) = \frac{2|\lambda| \sin(\theta - x)}{1 + |\lambda| \cos(\theta - x)}.$$

Adapting this argument (i.e., replacing x with $2\pi x$), we obtain

$$B_\lambda(x, y) = \big(2x + y + \psi(x),\ x + y + \psi(x)\big) \mod 1,$$

where $\psi : \mathbb{R}\backslash\mathbb{Z} \to \mathbb{R}\backslash\mathbb{Z}$ is given by

$$\tan(\pi\psi(x)) := -\frac{2|\lambda| \sin(2\pi x - \theta)}{1 + |\lambda| \cos(2\pi x - \theta)}. \tag{3.6}$$

Differentiating both sides of this expression gives, for all $x \in \mathbb{R}$,

$$\psi'(x) = -\frac{2|\lambda|\big(\cos(2\pi x - \theta) + |\lambda|\big)}{|\lambda|^2 + 2|\lambda| \cos(2\pi x - \theta) + 1} \geq -\frac{2\,|\lambda|}{|\lambda| + 1} > -1.$$

Thus, the tangent map takes the form

$$D_{(x,y)} B_\lambda = \begin{pmatrix} 1 + \varepsilon(x) & 1 \\ \varepsilon(x) & 1 \end{pmatrix},$$

where $\varepsilon(x) := \psi'(x) + 1 \geq \frac{1-|\lambda|}{1+|\lambda|} > 0$.

Obviously, $\det(D_{(x,y)} B_\lambda) = 1$ for all $(x, y) \in \mathbb{T}^2$, which shows that B_λ is area-preserving (i.e., via the integration by substitution formula).

To prove that B_λ is Anosov, one can simply show that the cone given by the positive quadrants

$$C^+ = \{(u, v) \in \mathbb{R}^2 \mid uv \geq 0\},$$

is mapped strictly inside itself under $D_x B_\lambda$, and that the complementary cone

$$C^- = \{(u, v) \in \mathbb{R}^2 \mid uv \leq 0\}$$

is mapped strictly inside itself by $(D_{(x,y)}B_\lambda)^{-1}$, for each $(x, y) \in \mathbb{T}^2$. The first is clear from the fact that $D_{(x,y)}B_\lambda$ has all entries positive for all $x \in \mathbb{T}^2$, and the second follows since, for $Q(u, v) := uv$, we have that the inequality

$$Q \circ D_{(x,y)}B_\lambda(u, v) = (\varepsilon(x)u + v)\big((1 + \varepsilon(x))u + v\big) \leq 0$$

prescribes a cone contained strictly within C^-, since $\varepsilon(x) > 0$. Therefore, by Fact 1, B_λ is Anosov as required. $\square$

3.2.2 The Hilbert space $\mathcal{H}_a$

As mentioned in the introduction, all of the Hilbert spaces discussed in this chapter can all be described with the following general construction:

Suppose we want to construct a complex Hilbert space $\mathcal{H}$ which has $\{e_{m,n}\}_{(m,n)\in\mathbb{Z}^2}$ as an orthogonal basis, recalling the notation

$$e_{m,n} : (z, w) \mapsto z^m w^n.$$

Given such a space $\mathcal{H}$, for $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ denoting the inner product and norm on $\mathcal{H}$ respectively, we can write

$$\left\langle \sum_{(m,n)\in\mathbb{Z}^2} b_{m,n}e_{m,n}, \sum_{(m,n)\in\mathbb{Z}^2} c_{m,n}e_{m,n} \right\rangle = \sum_{(m,n)\in\mathbb{Z}^2} b_{m,n}\overline{c_{m,n}}\,\|e_{m,n}\|^2$$

and

$$\Big\| \sum_{(m,n)\in\mathbb{Z}^2} b_{m,n}e_{m,n} \Big\|^2 = \sum_{(m,n)\in\mathbb{Z}^2} |b_{m,n}|^2\|e_{m,n}\|^2.$$

Moreover, we can define $\mathcal{H}$ set-wise to comprise series which have finite $\|\cdot\|_a$ norm:

$$\mathcal{H} = \left\{ \sum_{(m,n)\in\mathbb{Z}^2} b_{m,n}e_{m,n} \;\middle|\; b_{m,n} \in \mathbb{C}, \sum_{(m,n)\in\mathbb{Z}^2} |b_{m,n}|^2\|e_{m,n}\|^2 < \infty \right\}.$$

In this way, the nature of the space $\mathcal{H}$ is completely characterised by the values of $\|e_{m,n}\|$, which we call the *weights*.

Remark 16. For any $a > 0$, classical examples of such spaces include the Sobolev space of a-times weakly differentiable functions, with weak derivatives in $l^2(\mathbb{T}^2)$ [93, p.42], which can be defined by

$$\|e_{m,n}\| = \min\left(\big(|m| + |n|\big)^a, 1\right);$$

or the Hardy-Hilbert space of functions analytic on the polydisk $e^{-a}(\mathbb{D}\times\mathbb{D})$ (see, e.g., [85]), where

$$\|e_{m,n}\| = \begin{cases} e^{-a(m+n)}, & \text{if } m,n\geq 0; \\ \infty, & \text{otherwise.} \end{cases}$$

To obtain quasi-compactness (or more strongly, compactness, as is required here) for the composition operator acting on $\mathcal{H}$, one is required to define the weights in an anisotropic manner. In particular, taking limits along rays based at the origin, the weights decay to zero in some directions and diverge to infinity in others.

For example, in [88], the authors base these weights on the eigenvectors of the CAT map B_0: i.e., for $a>0$,

$$\|e_{m,n}\| = \exp\left(-a\left|\frac{\sqrt{5}+1}{2}m+n\right| + a\left|\frac{1-\sqrt{5}}{2}m+n\right|\right). \tag{3.7}$$

These are a particular instance of the anisotropic spaces introduced in greater generality by Frédéric Faure and Nicholas Roy in [44] and also used by Alexander Adam [1] to show his result on generic resonances.

We stress that the usefulness of the thus-obtained Hilbert space in [88] is principally that the composition operator $\mathcal{C}_{B_\lambda}$ acts compactly on this space, and that $a>0$ can be chosen so that the space contains any given pair of functions analytic on a neighbourhood of the torus.

Contrastingly, assuming it acts compactly, the spectrum of the composition operator acting on $\mathcal{H}$ as above is independent of the weights used (provided that they are all finite). We therefore propose simple alternative weightings, yielding new families of anisotropic Hilbert spaces. These spaces will be particularly simple for (B_λ) below; but will need a small adjustment when we later come to consider (T_λ).

The definition of the spaces $\mathcal{H}_a$, pertaining to (B_λ), make use of the signed *degree function* $\deg_1$, which we now present:

Definition 19 ($\deg_1$, $\|\cdot\|_a$, $\mathcal{H}_a$)**.** Let $\deg_1:\mathbb{Z}^2\to\mathbb{Z}$ be given by

$$\deg_1 : (m,n)\mapsto \operatorname{sign}(mn)\big(|m|+|n|\big),$$

where

$$\operatorname{sign}(k) = \begin{cases} 1, & \text{if } k\geq 0; \\ -1, & \text{if } k<0. \end{cases}$$

(Figure 3.3 shows some level sets of $\deg_1$.) We define, for $a > 0$,

$$\|e_{m,n}\|_a := e^{-a\deg(m,n)}.$$

Extending this norm in the manner above, we let $\mathcal{H}_a$ be the vector space of series in $e_{m,n}$ with finite $\|\cdot\|_a$ norm, i.e.,

$$\mathcal{H}_a = \left\{ f = \sum_{(m,n)\in\mathbb{Z}^2} b_{m,n}e_{m,n} \;\middle|\; b_{m,n} \in \mathbb{C},\ \|f\|_a < \infty \right\},$$

where, more explicitly, the $\|\cdot\|_a$ norm to

$$\left\| \sum_{(m,n)\in\mathbb{Z}^2} b_{m,n}e_{m,n} \right\|_a^2 := \sum_{(m,n)\in\mathbb{Z}^2} |b_{m,n}|^2 e^{-2a\deg_1(m,n)}.$$

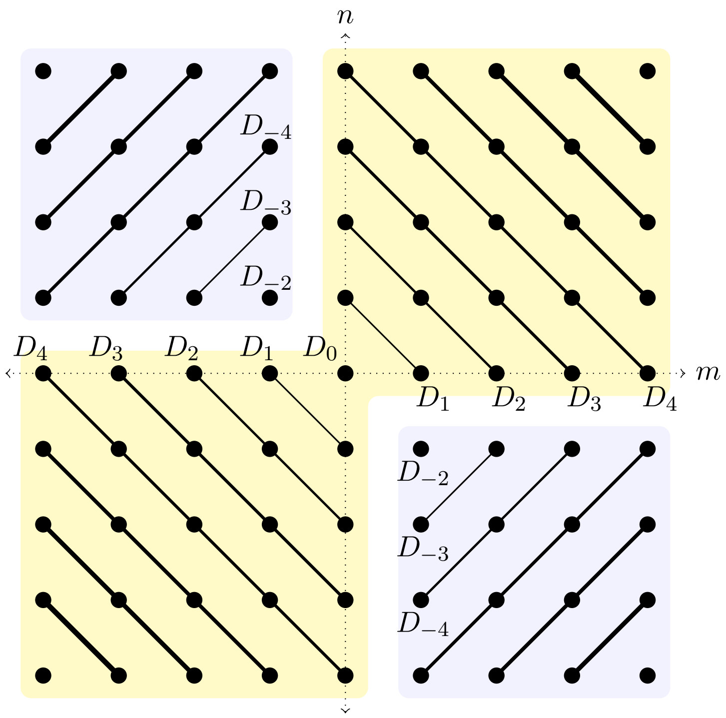


Figure 3.3: The level sets of $\deg_1 : \mathbb{Z}^2 \to \mathbb{Z}$. Here, for each $n \in \{-4, \ldots, 4\}$, D_n denotes ${\deg_1}^{-1}(n)$.

The benefits of using $\mathcal{H}_a$ (and later $\mathcal{H}_{a,\phi}$) over the family of anisotropic spaces defined by (3.7) are as follows:

- The proofs for compactness and the inclusion of analytic functions are simpler and more direct.

- The construction permits more flexibility (considering, e.g., $B_{\lambda,K}$ mentioned above).
- There is a clearer link between the structure of the space and the simple form for the matrix of the operator (block-triangular).

The following result shows that any pair of analytic functions on a neighbourhood of the torus will be contained in some $\mathcal{H}_a$, allowing us to equate the resonances of T_λ with the spectrum described in Theorem 6.

Proposition 10. *Let $a > 0$ and suppose that f is an analytic function on a neighbourhood of the poly-annulus*

$$P_a := \left\{(z, w) \in \mathbb{C}^2 \;\middle|\; e^{-a} \leq |z| \leq e^a,\ e^{-a} \leq |w| \leq e^a\right\}.$$

Then $f \in \mathcal{H}_a$. In particular, every function analytic on a neighbourhood of $\mathbb{T}^2$ is contained in $\mathcal{H}_a$ for all sufficiently small a.

Proof of Proposition 10. Fix $a > 0$ and f analytic on a neighbourhood of P_a. By construction, the Laurent series for f converges absolutely on P_a. In particular, writing this expansion as

$$f(z, w) = \sum_{(m,n)\in\mathbb{Z}^2} b_{m,n} z^m w^n, \tag{3.8}$$

we have, by definition of $\|\cdot\|_a$,

$$\|f\|_a^2 = \sum_{(m,n)\in\mathbb{Z}^2} |b_{m,n}|^2 \, e^{-2a \deg_1(m,n)} \leq \sum_{(m,n)\in\mathbb{Z}^2} |b_{m,n}|^2 \, e^{2a(|m|+|n|)},$$

which we want to show is finite. Note that

$$\sum_{(m,n)\in\mathbb{Z}^2} |b_{m,n}| e^{a(|m|+|n|)} \leq \sum_{(m,n)\in\mathbb{Z}^2} |b_{m,n}| \left(e^{a(m+n)} + e^{a(m-n)} + e^{a(n-m)} + e^{-a(m+n)}\right)$$

is finite, since (3.8) converges absolutely for all $(z, w) \in P_a$, i.e., the sums

$$\sum_{m,n} |b_{m,n}| e^{a(m+n)}, \qquad \sum_{m,n} |b_{m,n}| e^{a(m-n)},$$
$$\sum_{m,n} |b_{m,n}| e^{a(n-m)}, \qquad \sum_{m,n} |b_{m,n}| e^{-a(m+n)}$$

are each finite, since $(e^{\pm a}, e^{\pm a}) \in P_a$. In particular, the left hand side is square-summable, and hence $f \in \mathcal{H}_a$, as required. $\square$

3.2.3 $\mathcal{C}_{B_\lambda}$ is Hilbert-Schmidt

To prove compactness of the composition operator acting on $\mathcal{H}_a$, we need to estimate the Taylor coefficients of arbitrary powers of a given Möbius function.

Estimates on the Taylor coefficients of a simple Blaschke product

In view of the definition of B_λ, the action of $\mathcal{C}_{B_\lambda}$ on $\mathcal{H}_a$ involves formally expanding functions of the form

$$\left(\frac{z+\lambda}{1+\overline{\lambda}z}\right)^m \tag{3.9}$$

as Taylor or Laurent series, depending on the sign of $m \in \mathbb{Z}$. To this end, we introduce the coefficients $\alpha_{m,k} = \alpha_{m,k}(\lambda)$, which we use throughout this chapter.

Definition 20 ($\alpha_{m,k}$)**.** For $m \in \mathbb{N}_0$, since (3.9) is analytic on $\mathbb{C} \setminus \{-\overline{\lambda}^{-1}\}$, the following expansion is valid for $|z| < |\lambda|^{-1}$ (and converges uniformly on every closed disk of radius strictly less than $|\lambda|^{-1}$):

$$\left(\frac{z+\lambda}{1+\overline{\lambda}z}\right)^m = \sum_{k=0}^{\infty} \alpha_{m,k} z^k, \tag{3.10}$$

where the complex coefficients $\alpha_{m,k}$ are given by the Cauchy integral equation

$$\alpha_{m,k} = \frac{1}{2\pi i} \int_{|s|=1} s^{-(k+1)} \left(\frac{s+\lambda}{1+\overline{\lambda}s}\right)^m \, \mathrm{d}s,$$

or alternatively, via Newton's identity,

$$\alpha_{m,k} = \sum_{j=0}^{\min(m,k)} \binom{m}{j} \binom{m+k-j-1}{k-j} (-\overline{\lambda})^{k-j} \lambda^{m-j}.$$

Explicitly, we have $\alpha_{m,0} = \lambda^m$ for all $m \in \mathbb{N}_0$, and $\alpha_{0,k} = 0$ for all $k \in \mathbb{N}$.

Reciprocating the above Möbius function, one also obtains a related Taylor expansion about ∞: if $m \leq -1$,

$$\left(\frac{z+\lambda}{1+\overline{\lambda}z}\right)^m = \left(\frac{z^{-1}+\overline{\lambda}}{1+\lambda z^{-1}}\right)^{-m} = \sum_{k=0}^{\infty} \overline{\alpha_{-m,k}} z^{-k}. \tag{3.11}$$

For simplicity, we adopt the notation that $\alpha_{-m,k} = \overline{\alpha_{m,k}}$ for all m and k.

The proof of compactness of $\mathcal{C}_{B_\lambda}$ (here and in [88]) reduces to estimating sums of the form

$$\sum_{k=0}^{\infty} |\alpha_{m,k}|^2 e^{-2ak}$$

for each $m \in \mathbb{Z}$, where $a > 0$ is fixed, and the $\alpha_{m,k}$ are given by (3.10) and (3.11). To this end, the authors of [88] decompose the sum into two and apply the following estimates, derived from the Cauchy integral formula for $\alpha_{m,k}$.

Lemma 14 ([88, Lemma 2.3])**.** *For all $\lambda \in \mathbb{D}$,*

1. $|\alpha_{m,k}| \le 1$ *for all $m, k \in \mathbb{N}_0$.*
2. *For all*
$$\beta \in \left(0, \frac{1-|\lambda|}{1+|\lambda|}\right),$$
there exists $\gamma > 0$ such that for all $m, k \in \mathbb{N}_0$ satisfying $k < \beta m$,
$$|\alpha_{m,k}| \le e^{\gamma(k+1-\beta m)}. \tag{3.12}$$

We now present an alternative result, which has the advantages of being direct, simple and more explicit. There is also clear scope for improvement if needed for later applications. This emerged by considering the proof of the following neat formula, which implies $|\alpha_{m,k}| \le 1$.

Fact 2. For any $\lambda \in \mathbb{D}$ and $m \in \mathbb{Z}$,

$$\sum_{k=0}^{\infty} |\alpha_{m,k}|^2 = 1.$$

Proof of Fact 2. Since $|\alpha_{m,k}| = |\alpha_{-m,k}|$, we may assume that $m \ge 0$. Recalling the following basic cancellation fact for Lebesgue integrals,

$$\frac{1}{2\pi}\int_{\mathbb{T}^1} z^{k-j}\,|\mathrm{d}z| := \frac{1}{2\pi}\int_0^{2\pi} e^{i(k-j)\theta}\,\mathrm{d}\theta = \chi_{\{k=j\}}$$

(where χ again denotes the indicator function), one writes

$$\sum_{k=0}^{\infty} |\alpha_{m,k}|^2 = \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \alpha_{m,k}\overline{\alpha_{m,j}}\,\chi_{\{k=j\}}$$

(continued on next page)

(continued from previous page)

$$
\begin{aligned}
&= \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \alpha_{m,k}\overline{\alpha_{m,j}}\,\frac{1}{2\pi}\int_{\mathbb{T}^1} z^{k-j}\;|\mathrm{d}z| \\
&= \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \frac{1}{2\pi}\int_{\mathbb{T}^1} \alpha_{m,k}\overline{\alpha_{m,j}} z^{k-j}\;|\mathrm{d}z| \\
&= \frac{1}{2\pi}\int_{\mathbb{T}^1} \sum_{k=0}^{\infty} \alpha_{m,k} z^k \sum_{j=0}^{\infty} \overline{\alpha_{m,j}} z^{-j}\;|\mathrm{d}z| \\
&= \frac{1}{2\pi}\int_{\mathbb{T}^1} \left(\frac{z+\lambda}{1+\overline{\lambda}z}\right)^m \left(\frac{z+\lambda}{1+\overline{\lambda}z}\right)^{-m}\;|\mathrm{d}z| \\
&= \frac{1}{2\pi}\int_{\mathbb{T}^1} 1\;|\mathrm{d}z| \\
&= 1.
\end{aligned} \tag{3.13}
$$

In particular, both Taylor expansions on the fourth line converge uniformly on $\mathbb{T}^1$, which allows for the interchange of sum and integral. □

The method of proof generalises to give (the second part of) the following.

Lemma 15. *For all $\lambda \in \mathbb{D}$ and $a > 0$,*

$$M_{a,\lambda} := \max_{|z|=e^{-2a}} \left|\frac{z+\lambda}{1+\overline{\lambda}z}\right| < 1. \tag{3.14}$$

Moreover, $M_{a,\lambda}$ satisfies, for all $m \in \mathbb{Z}$,

$$\sum_{k=0}^{\infty} |\alpha_{m,k}|^2 e^{-2ak} \leq M_{a,\lambda}^{|m|}.$$

Proof of Lemma 15. We again assume that $m \geq 0$. Repeating the argument (3.13) above, this time also using that (3.10) converges uniformly on $e^{-2a}\mathbb{T}^1 \subset \mathbb{D}$, gives the following:

$$
\begin{aligned}
\sum_{k=0}^{\infty} |\alpha_{m,k}|^2 e^{-2ak} &= \frac{1}{2\pi}\sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \alpha_{m,k}\overline{\alpha_{m,j}} \int_{\mathbb{T}^1} (ze^{-2a})^k z^{-j}\;|\mathrm{d}z| \\
&= \frac{1}{2\pi}\int_{\mathbb{T}^1} \left(\frac{ze^{-2a}+\lambda}{1+\overline{\lambda}e^{-2a}z}\right)^m \left(\frac{z+\lambda}{1+\overline{\lambda}z}\right)^{-m}\;|\mathrm{d}z|.
\end{aligned}
$$

In particular, a uniform estimate on this integral gives

$$\sum_{k=0}^{\infty} |\alpha_{m,k}|^2 e^{-2ak} \leq \max_{z\in\mathbb{T}^1} \left|\frac{ze^{-2a}+\lambda}{1+\overline{\lambda}e^{-2a}z}\right|^m \underbrace{\left|\frac{z+\lambda}{1+\overline{\lambda}z}\right|^{-m}}_{=1}$$

$$= \max_{|ze^{2a}|=1} \left|\frac{z+\lambda}{1+\overline{\lambda}z}\right|^m = M_{a,\lambda}^m,$$

which proves (3.14). Finally, given $a > 0$, the Möbius function

$$z \mapsto \frac{z+\lambda}{1+\overline{\lambda}z}$$

takes $\mathbb{D}$ into itself, and hence maps the circle $e^{-2a}\mathbb{T}^1$ to one strictly contained in $\mathbb{D}$, which is thus bounded away from $\mathbb{T}^1$. Hence, the maximum modulus obtained on this circle is strictly less than 1, giving $M_{a,\lambda} < 1$. □

Application to $\mathcal{C}_{B_\lambda}$

The above estimates suffice to prove the following property for $\mathcal{C}_{B_\lambda}$.

Definition 21 (Hilbert-Schmidt, $\|\cdot\|_{\mathrm{HS}}$)**.** The Hilbert-Schmidt norm of an operator $\mathcal{C} : \mathcal{H} \to \mathcal{H}$ acting on a separable Hilbert space $\mathcal{H}$ takes the following form, for a given orthonormal basis $\{\hat{e}_i\}_{i\in\mathcal{I}}$ of $\mathcal{H}$:

$$\|\mathcal{C}\|_{\mathrm{HS}}^2 = \sum_{i\in\mathcal{I}} \|\mathcal{C}(\hat{e}_i)\|^2.$$

We say that $\mathcal{C}$ is *Hilbert-Schmidt* if it has finite Hilbert-Schmidt norm.

Well-known properties of Hilbert-Schmidt operators (see [28, p.267]) include that they are compact and that their singular values are square-summable (respecting multiplicity). Every trace-class operator is Hilbert-Schmidt, and an operator is trace-class if and only if it is the composition of two Hilbert Schmidt operators. (We say that a compact operator is *trace-class* if its singular values are summable.)

The Hilbert-Schmidt property is relatively easy to verify. In particular, the norm equivalently reads as

$$\|\mathcal{C}\|_{\mathrm{HS}}^2 = \sum_{i\in\mathcal{I}} \left(\frac{\|\mathcal{C}(e_i)\|}{\|e_i\|}\right)^2$$

for any *orthogonal* basis $\{e_i\}_{i\in\mathcal{I}}$ of $\mathcal{H}$, and it is therefore natural to prove the Hilbert-Schmidt property by estimating these summands. This is indeed how we prove the following lemma.

Lemma 16. *For all $\lambda \in \mathbb{D}$ and $a > 0$, the composition operator $\mathcal{C}_{B_\lambda} : \mathcal{H}_a \to \mathcal{H}_a$ is Hilbert-Schmidt.*

Remark 17. Since $B_\lambda = T_0 \circ T_\lambda$, we will later see that $\mathcal{C}_{B_\lambda}$ is in fact trace-class (which we recall is a stronger property than being Hilbert-Schmidt) as an operator on the related space $\mathcal{H}_{a,\phi}$.

The proof of Lemma 16 uses the following simple lemma.

Lemma 17. *For all $(m, n) \in \mathbb{Z}^2$ with $n \neq 0$,*

$$\deg_1(m + \operatorname{sign}(n), n) \geq \deg_1(m, n) + 1, \tag{3.15}$$

and similarly, $\deg_1(m, \operatorname{sign}(m) + n) \geq \deg_1(m, n) + 1$ *if $m \neq 0$.*

Although the statement of the lemma is somewhat terse, an intuitive description may suffice to convince the reader of its correctness in advance of the formal proof: Broadly speaking (considering $\deg_1$ as a signed l^1-norm) the lemma states that, starting from a lattice point in the positive quadrants, $\{(m, n) \in \mathbb{Z}^2 \mid mn \geq 0\}$, moving to an adjacent point further from an axis increases the value of $\deg_1$ (since it increases the l^1 norm), whereas starting from a lattice point in the negative quadrants, $\{(m, n) \mid mn < 0\}$ and moving to an adjacent point closer to the axis increases the value of $\deg_1$ (since it decreases the l^1 norm). The previous depiction of level sets in Figure 3.3 may be helpful for the reader's intuition. Also, in Figure 3.4 below, (3.18) corresponds to the fact that all horizontal arrows point right for $n > 0$ and left for $n < 0$.

We now give an algebraic proof, whose three cases correspond to the shaded regions in Figure 3.4.

Proof of Lemma 17. We only prove the first equality, since the second follows by symmetry. Considering the two factors of

$$\deg_1(m + \operatorname{sign}(n), n) = \operatorname{sign}\big(m + \operatorname{sign}(n), n\big)\big(|m + \operatorname{sign}(n)| + |n|\big),$$

we first observe that

$$\big(m + \operatorname{sign}(n)\big)n = mn + |n| \geq 0 \iff \begin{cases} \text{either} & mn \geq 0, \\ \text{or} & mn < 0 \text{ and } |m| = 1. \end{cases}$$

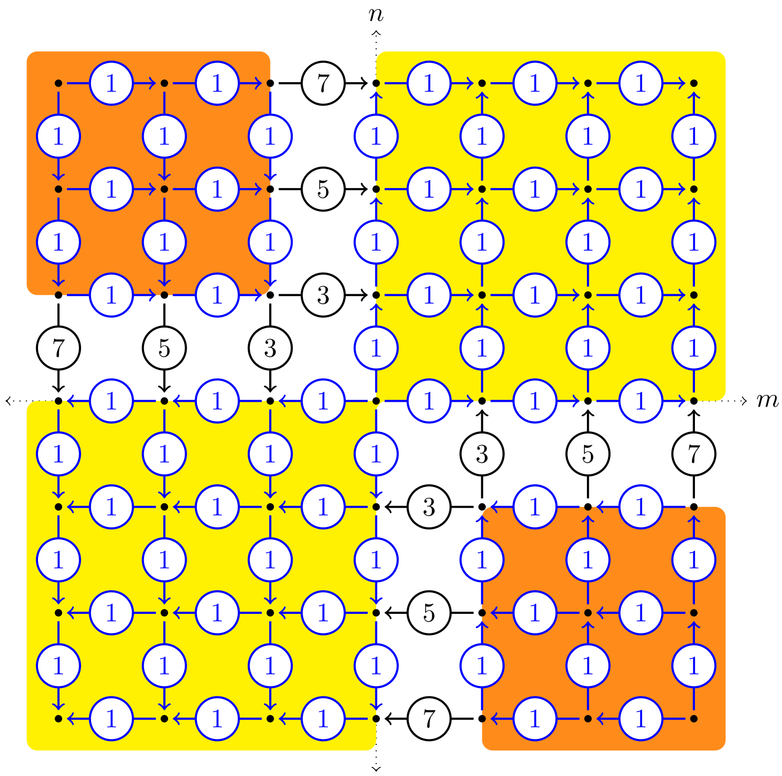


Figure 3.4: Increments in value for $\deg_1$:

- The arrows indicate the directions in which $\deg_1$ is increasing.
- The numbers show the differences in value between adjacent points.

Moreover, since $n \neq 0$, trivially

$$|m + \operatorname{sign}(n)| = \begin{cases} |m| + 1, & \text{if } mn \geq 0; \\ |m| - 1, & \text{if } mn < 0. \end{cases}$$

Applying these two equations in order gives three cases:

$$\deg_1(m + \operatorname{sign}(n), n) = \begin{cases} |m + \operatorname{sign}(n)| + |n|, & \text{if } mn \geq 0; \\ |m + \operatorname{sign}(n)| + |n|, & \text{if } mn < 0 \text{ and } |m| = 1; \\ -|m + \operatorname{sign}(n)| - |n|, & \text{if } mn < 0 \text{ and } |m| > 1; \end{cases}$$

(continued on next page)

(continued from previous page)

$$
\begin{aligned}
&= \begin{cases} |m|+|n|+1, & \text{if } mn \geq 0; \\ |m|+|n|-1 = |n|, & \text{if } mn < 0 \text{ and } |m| = 1; \\ 1-|m|-|n|, & \text{if } mn < 0 \text{ and } |m| > 1; \end{cases} \\
&= \begin{cases} \deg_1(m,n)+1, & \text{if } mn \geq 0; \\ \deg_1(m,n)+1+2|n|, & \text{if } mn < 0 \text{ and } |m| = 1; \\ \deg_1(m,n)+1, & \text{if } mn < 0 \text{ and } |m| > 1; \end{cases} \\
&\geq \deg_1(m,n)+1,
\end{aligned}
$$

which proves (3.15). In particular, the third equality for the middle case follows since, for $mn < 0$ and $|m| = 1$, $\deg_1(m,n) = -1-|n|$. $\square$

We now apply this lemma to prove that $\mathcal{C}_{B_\lambda}$ is Hilbert-Schmidt.

Proof of Lemma 16. Fix $\lambda \in \mathbb{D}$ and $a > 0$ above, and consider $\mathcal{C}_{B_\lambda}(e_{m,n})$. Using the earlier Taylor expansions of (3.10) and (3.11) gives

$$
\begin{aligned}
e_{m,n}\big(B_\lambda(z,w)\big) &= \left(\left(\frac{z+\lambda}{1+\bar{\lambda}z}\right)zw\right)^m \left(\left(\frac{z+\lambda}{1+\bar{\lambda}z}\right)w\right)^n \\
&= \left(\frac{z+\lambda}{1+\bar{\lambda}z}\right)^{m+n} z^m w^{m+n} \\
&= \begin{cases} \displaystyle\sum_{k=0}^{\infty} \alpha_{m+n,k} z^{m+k} w^{m+n}, & \text{if } m+n > 0; \\ z^m w^{m+n}, & \text{if } m+n = 0; \\ \displaystyle\sum_{k=0}^{\infty} \alpha_{m+n,k} z^{m-k} w^{m+n}, & \text{if } m+n < 0. \end{cases}
\end{aligned}
$$

That is,

$$
\mathcal{C}_{B_\lambda}(e_{m,n}) = \begin{cases} \displaystyle\sum_{k=0}^{\infty} \alpha_{m+n,k}\, e_{m+k,m+n}, & \text{if } m+n > 0; \\ e_{m,m+n} = e_{m,0}, & \text{if } m+n = 0; \\ \displaystyle\sum_{k=0}^{\infty} \alpha_{m+n,k}\, e_{m-k,m+n}, & \text{if } m+n < 0. \end{cases} \tag{3.16}
$$

Consider first the case that $m+n \neq 0$, and let $\sigma = \operatorname{sign}(m+n)$. To estimate

$$\left(\frac{\|\mathcal{C}_{B_\lambda} e_{m,n}\|_a}{\|e_{m,n}\|_a}\right)^2 = \sum_{k=0}^{\infty} |\alpha_{|m+n|,k}|^2 \left(\frac{\|e_{m+\sigma k,m+n}\|_a}{\|e_{m,n}\|_a}\right)^2, \tag{3.17}$$

we first bound the ratio

$$\frac{\|e_{m+\sigma k,m+n}\|_a}{\|e_{m,n}\|_a} = \exp\left[-a\Big(\deg_1(m+\sigma k, m+n) - \deg_1(m,n)\Big)\right], \tag{3.18}$$

for each $k \in \mathbb{N}_0$. To this end, we apply Lemma 17 in two different ways. First, since $m+n \neq 0$, applying the lemma k times gives

$$\deg_1(m+\sigma k, m+n) = \deg_1\Big(m + \operatorname{sign}(m+n)k, m+n\Big) \geq \deg_1(m, m+n) + k.$$

Second, applying the lemma $|m|$ times to the term on the right hand side gives

$$\deg_1(m, m+n) = \deg_1\Big(m, |m|\operatorname{sign}(m) + n\Big) \geq \deg_1(m,n) + |m|$$

(if $m=0$, the inequality is trivial). That is,

$$\deg_1(m+\sigma k, m+n) \geq \deg_1(m,n) + |m| + k. \tag{3.19}$$

Thus, by (3.18),

$$\frac{\|e_{m+\sigma k,m+n}\|_a}{\|e_{m,n}\|_a} \leq e^{-a\big(|m|+k\big)}.$$

We can now bound (3.17) with a straightforward application of Lemma 15:

$$\begin{aligned}
\left(\frac{\|\mathcal{C}_{B_\lambda} e_{m,n}\|_a}{\|e_{m,n}\|_a}\right)^2 = & \sum_{k=0}^{\infty} |\alpha_{m+n,k}|^2 \left(\frac{\|e_{m+\sigma k,m+n}\|_a}{\|e_{m,n}\|_a}\right)^2 \\
& \leq e^{-2a|m|} \sum_{k=0}^{\infty} |\alpha_{m+n,k}|^2 \, e^{-2ak} \\
& \leq e^{-2a|m|} M_{a,\lambda}^{|m+n|} \\
& \leq e^{-2\delta\big(|m+n|+|m|\big)} \\
& \leq e^{-\delta\big(|m|+|n|\big)},
\end{aligned} \tag{3.20}$$

where $\delta = \min(-\frac{1}{2}\log M_{a,\lambda}, a) > 0$. The last inequality in particular follows from the reverse triangle inequality:

$$|m+n|+|n| \geq \begin{Bmatrix} |m|, & \text{if } |m| \geq |n| \\ 2|n|-|m|, & \text{if } |n| \geq |m| \end{Bmatrix} \geq \frac{1}{2}(|m|+|n|).$$

(3.20) also extends to the complementary case of $m+n=0$. That is, using that $\deg_1(m,-m) = -2|m|$ and $\deg_1(m,0) = |m|$,

$$\begin{aligned} \left(\frac{\|\mathcal{C}_{B_\lambda} e_{m,-m}\|_a}{\|e_{m,-m}\|_a}\right)^2 = \frac{\|e_{m,0}\|_a^2}{\|e_{m,-m}\|_a^2} = e^{2a\deg_1(m,0)-2a\deg(m,-m)} &= e^{-6a|m|} \\ &= e^{-3a\left(|m|+|-m|\right)} \\ &\leq e^{-\delta\left(|m|+|-m|\right)}. \end{aligned}$$

This inequality is sufficient to finish the proof, recalling the expression for the Hilbert-Schmidt norm in terms of an orthogonal basis:

$$\|\mathcal{C}_{B_\lambda}\|_{\mathrm{HS}}^2 = \sum_{(m,n)\in\mathbb{Z}^2} \left(\frac{\|\mathcal{C}_{B_\lambda} e_{m,n}\|_a}{\|e_{m,n}\|_a}\right)^2 \leq \sum_{(m,n)\in\mathbb{Z}^2} e^{-\delta(|m|+|n|)} < \infty.$$

Therefore $\mathcal{C}_{B_\lambda}$ is Hilbert-Schmidt, as required. □

3.2.4 The spectrum of $\mathcal{C}_{B_\lambda}$

As mentioned above, the calculation of the eigenvalues of the composition operator is independent of the weights $\|e_{m,n}\|_a$, provided that they are all finite.

It is useful to think of $\mathcal{C}_{B_\lambda}$ as bi-infinite matrix. In view of this consideration, we present the following notion of a block-triangular form, which generalises that of a block-triangular matrix, i.e., a matrix of the form

$$A = \begin{pmatrix} A_1 & 0 & 0 & \cdots & 0 \\ * & A_2 & 0 & \cdots & 0 \\ * & * & A_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ * & * & * & \cdots & A_n \end{pmatrix},$$

where the A_k are square matrices.

The results of this subsection in particular generalise the basic fact that the eigenvalues of A are simply those of the "blocks" A_k, whose algebraic multiplicity is obtained by summing the respective multiplicities over each block.

Remark 18. This generality, although it is not required for the family B_λ, is convenient for when we later consider the family (T_λ), and is particularly so when we extend the analysis to $(T_\lambda \circ T_\mu)$.

Block–lower triangular form for compact operators

We now give a useful definition.

Definition 22. [Block-triangular form] We say that a linear operator $\mathcal{C}$, acting on a Hilbert space $\mathcal{H}$ with orthogonal basis $\mathcal{B} = \{e_i\}_{i\in\mathcal{I}}$, has a *block-triangular form* (with respect to $\mathcal{B}$) if one has

$$\mathcal{H} = \bigoplus_{k\in\mathbb{Z}} D_k$$

such that, for each $k \in \mathbb{Z}$,

- D_k has a basis consisting of a finite (non-empty) subset of $\mathcal{B}$, and
- $\mathcal{C}(D_k) \subset \bigoplus_{j=k}^{\infty} D_j$.

We now state the following result which reduces eigenvalue computations of block-triangular operators to those of their finite-dimensional blocks.

Lemma 18. *Suppose $\mathcal{C}$ and D_k are as in Definition 22, and suppose further that $\mathcal{C}$ is Hilbert-Schmidt. Then its non-zero eigenvalues are precisely the union of the eigenvalues for each finite rank operator $\mathcal{C}_k$ ($k \in \mathbb{Z}$):*

$$\mathcal{C}_k = \Pi_{D_k} \circ \mathcal{C} \circ \Pi_{D_k},$$

where Π_D denotes the orthogonal projection onto the subspace D.

Moreover, if a given non-zero eigenvalue of $\mathcal{C}$ is an eigenvalue of only one $\mathcal{C}_k$, then its algebraic and geometric multiplicities for these two operators coincide.

Because this result is quite intuitive (in view of the finite dimensional case) and admits an elementary but long proof, we defer this proof to appendix A.

Now, to apply this result, each of the composition operators in this chapter will be block-triangular with respect to $(e_{m,n})_{m,n}$, with the subspaces D_k given by

$$D_k = \mathrm{Span}\{e_{m,n} \mid \deg_1(m,n) = k\}. \tag{3.21}$$

Since $|\deg_1(m,n)| = k \iff |m| + |n| = k$, each D_k is finite dimensional, and the above lemma applies to any Hilbert-Schmidt operator that *increases* $\deg_1$, in the following sense.

Definition 23 (Increase)**.** If $\mathcal{H}$ is a Hilbert space which has $(e_{m,n})_{(m,n)\in\mathbb{Z}^2}$ as an orthogonal basis, we say the endomorphism $\mathcal{C} : \mathcal{H} \to \mathcal{H}$ *increases* $\deg_1$ if, for each $(m,n) \in \mathbb{Z}^2$, $\mathcal{C}(e_{m,n})$ lies in the closure of

$$\mathrm{Span}\{e_{m',n'} \mid \deg_1(m',n') \geq \deg_1(m,n)\}.$$

I.e., $\mathcal{C}(D_k) \subset \oplus_{j\geq k} D_k$ for each $k \in \mathbb{Z}$, where the D_k are given in (3.21).

Application to the spectrum of $\mathcal{C}_{B_\lambda}$

We apply the above machinery to obtain the following important result, completing the proof of Theorem 6.

Lemma 19. *For all $a > 0$, $\mathcal{C}_{B_\lambda} : \mathcal{H}_a \to \mathcal{H}_a$ has spectrum*

$$\{0,1\} \cup \{\lambda^k,\ \overline{\lambda}^k \mid k \in \mathbb{N}\}.$$

where each non-zero eigenvalue has algebraic and geometric multiplicity equal to the frequency with which it appears in the above (in particular, they are all semi-simple).

The proof of this result is a straightforward application of Lemma 18.

Proof of Lemma 19. We first show that $\mathcal{C}_{B_\lambda}$ increases $\deg_1$. Recalling the expansion

$$\mathcal{C}_{B_\lambda}(e_{m,n}) = \begin{cases} \displaystyle\sum_{k=0}^{\infty} \alpha_{m+n,k}\, e_{m+k,m+n}, & \text{if } m+n > 0; \\ \quad e_{m,0}, & \text{if } m+n = 0; \\ \displaystyle\sum_{k=0}^{\infty} \alpha_{m+n,k}\, e_{m-k,m+n}, & \text{if } m+n < 0; \end{cases} \tag{3.22}$$

consider first the case that $m+n \neq 0$. Then $\mathcal{C}_{B_\lambda}(e_{m,n})$ is a linear combination of

$$\{e_{m+\sigma k,m+n} \mid k \in \mathbb{N}_0\},$$

where $\sigma = \mathrm{sign}(m+n)$. Recall (3.19), which shows that the index of each of these terms in (3.22) take a higher value of $\deg_1$ than (m,n):

$$\deg_1(m+\sigma k, m+n) \geq \deg_1(m,n) + |m| + k. \tag{3.23}$$

Similarly, in the case that $m+n = 0$, one has, from the definition of $\deg_1$,

$$\deg_1(m,0) = \deg_1(m,-m) + 3|m|.$$

Thus, $\mathcal{C}_{B_\lambda}$ increases $\deg_1$. Furthermore, recalling the notation used in Lemma 18, for each $j \in \mathbb{Z}$, the form for $\left(\mathcal{C}_{B_\lambda}\right)_j : D_j \to D_j$,

$$\left(\mathcal{C}_{B_\lambda}\right)_j = \Pi_{D_j} \circ \mathcal{C}_{B_\lambda} \circ \Pi_{D_j},$$

can be obtained quite simply from (3.22). For $e_{m,n} \in D_j$, i.e., $\deg_1(m,n) = j$, $\left(\mathcal{C}_{B_\lambda}\right)_j(e_{m,n})$ is obtained by removing all terms from the right hand side except for those for which the index of the basis vector (i.e., $e_{m+\sigma k,m+n}$) has the same value of $\deg_1$ as (m,n). In view of (3.23), one sees that the only possible term that can remain in the $m+n \neq 0$ case is the one corresponding to $k=0$, and this remains only if $m=0$. Similarly, in the $m+n=0$ case, the (single) term survives only if $m=0$.

Indeed, setting $m=0$, the zeroth term of $\mathcal{C}_{B_\lambda}(e_{0,n})$ in (3.22) is a multiple of $e_{0,n}$. More explicitly,

$$\left(\mathcal{C}_{B_\lambda}\right)_{|n|} e_{0,n} = \alpha_{n,0}\, e_{0,n} = \begin{cases} \lambda^n e_{0,n}, & \text{if } n \geq 0; \\ \overline{\lambda}^n e_{0,n}, & \text{if } n < 0. \end{cases}$$

In other words, for $k<0$, $(\mathcal{C}_{B_\lambda})_k$ is the zero map, and for $k \geq 0$, it is the diagonal operator

$$\left(\mathcal{C}_{B_\lambda}\right)_k (e_{m,n}) = \begin{cases} \lambda^k e_{m,n}, & (m,n) = (0, \quad k); \\ \overline{\lambda}^k e_{m,n}, & (m,n) = (0, -k); \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, if $k>0$, $(\mathcal{C}_{B_\lambda})_k$ contributes two non-zero eigenvalues, λ^k and $\overline{\lambda}^k$, and $(\mathcal{C}_{B_\lambda})_0$ contributes the eigenvalue 1.

Finally, since $|\lambda| < 1$, these eigenvalues are distinct, except for when $\lambda^k = \overline{\lambda}^k$, i.e., when λ^k is real. In any case, it is clear that these eigenvalues are semi-simple (because they appear on the diagonal of $(\mathcal{C}_{B_\lambda})_k$, which has all other entries equal to zero). $\square$

This completes the proof of Theorem 6.

3.3 The spectrum of $\mathcal{C}_{T_\lambda}$

In this section, we consider a family of Anosov maps which give a richer, more varied resonances. This time, they will be perturbations of the orientation-reversing square root of the CAT map, T_0:

$$T_0 : (z, w) \mapsto (zw, z) \qquad \longleftrightarrow \qquad \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{mod } 1.$$

Definition 24. For $\lambda \in \mathbb{D}$, consider

$$T_\lambda : (z, w) \mapsto \left(\left(\frac{z + \lambda}{1 + \bar{\lambda} z} \right) w, z \right),$$

There are two minor differences between the argument presented in this section and the previous:

1. It is necessary to use a slightly more complicated family of Hilbert spaces, $\mathcal{H}_{a,\phi}$, depending on a generalisation of $\deg_1$.
2. The cone families used to show that T_λ is Anosov must vary as $|\lambda| \to 1$ (but the cones will still be independent of the basepoint).

The main result of this section is the following, which gives resonances for each T_λ.

Theorem 7. *For each $\lambda \in \mathbb{D}$, there exists a Hilbert space $\mathcal{H}_{a,\phi}$ of distributions on $\mathbb{T}^2$, such that the composition operator $\mathcal{C}_{T_\lambda} : \mathcal{H}_{a,\phi} \to \mathcal{H}_{a,\phi}$ given by $f \mapsto f \circ T_\lambda$ is compact and has spectrum as follows: for λ_1 a square root of λ,*

$$\{0, 1\} \cup \{\omega \lambda_1^m \overline{\lambda_1}^n \mid m, n \in \mathbb{N}_0,\, m + n \geq 1,\, \omega = \pm 1\}. \tag{3.24}$$

All non-zero eigenvalues have algebraic multiplicities as given in Lemma 21. Moreover, if the argument of λ is not a rational multiple of π, then all non-zero eigenvalues are semi-simple (i.e., their algebraic and geometric multiplicities coincide).

Remark 19. In this section, one could again extend the analysis to related families of examples: i.e., for $K \in \mathbb{N}$ and $\lambda \in \mathbb{D}$,

$$T_{\lambda,K} : (z, w) \mapsto \left(\left(\frac{z + \lambda}{1 + \bar{\lambda} z} \right)^K w, z \right),$$

perturbing, for each K, the hyperbolic linear automorphism

$$T_{0,K} : (z, w) \mapsto (z^K w, z) \qquad \longleftrightarrow \qquad \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} K & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

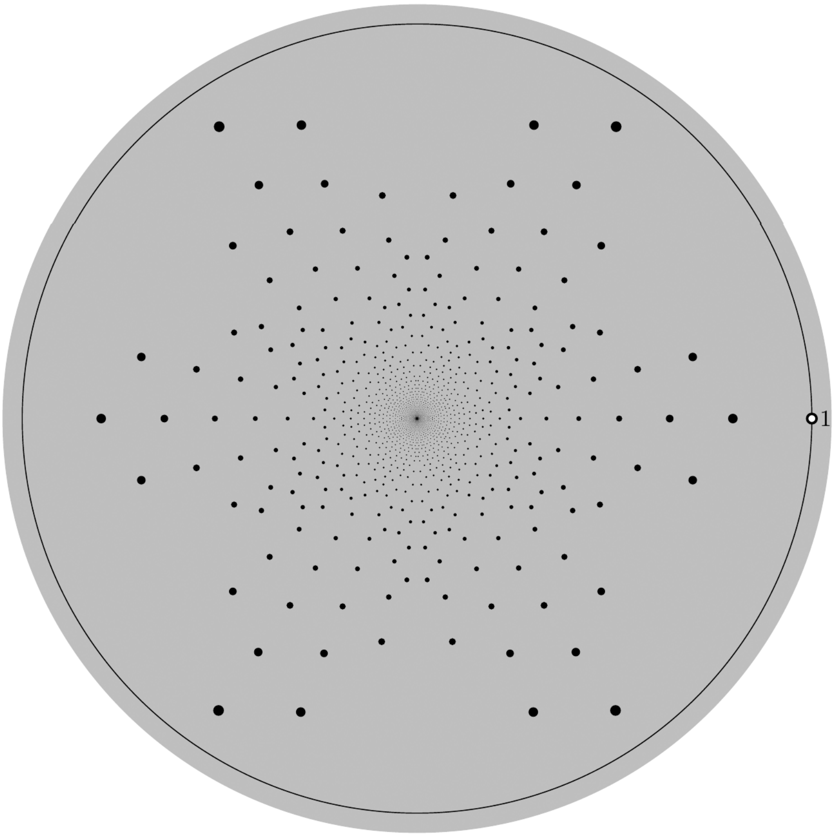


Figure 3.5: A plot of the spectrum of $\mathcal{C}_{T_\lambda}$, for $\lambda = 0.8e^{31i\pi/50}$.

the orientation-reversing square root of $B_{0,K}$. However again, we would find that that the spectrum of $T_{\lambda,K}$ equals that of T_{λ^K}, so these families again contribute nothing extra in variety.

3.3.1 Hyperbolicity of T_λ

As before, we first show that T_λ is area-preserving and Anosov.

Proposition 11. *For all $\lambda \in \mathbb{D}$, T_λ is an area-preserving Anosov diffeomorphism of $\mathbb{T}^2$.*

Proof of Proposition 11. Fix $\lambda \in \mathbb{D}$. Considering the proof of Proposition 9, one has the following form for T_λ in standard coordinates:

$$T_\lambda(x, y) = (x + y + \psi(x), x) \mod 1,$$

where $\psi : \mathbb{R}/\mathbb{Z} \to \mathbb{R}$ is defined by (3.6) on page 83. This gives

$$D_{(x,y)}(T_\lambda) = \begin{pmatrix} \varepsilon(x) & 1 \\ 1 & 0 \end{pmatrix},$$

where $\varepsilon(x) = 1 + \psi'(x) \geq \frac{1-|\lambda|}{1+|\lambda|} > 0$ is as before. Because this matrix has determinant minus one, T_λ is an area-preserving (and orientation-reversing) diffeomorphism.

To show that T_λ is Anosov, it again suffices to provide a cone family which is expanding and co-expanding with respect to this map. To this end, given $\lambda \in \mathbb{D}$, fix κ such that

$$\kappa > \frac{1+|\lambda|}{1-|\lambda|} = \max_{x \in \mathbb{T}} \left(\big(\varepsilon(x)\big)^{-1} \right) > 0, \tag{3.25}$$

and consider the cone C^κ defined by $Q_\kappa(u,v) := v(\kappa u - v)$:

$$C^\kappa_{(x,y)} := \{(u,v) \in \mathbb{R}^2 \in \mathcal{T}_{(x,y)}(\mathbb{T}^2) \ : \ Q_\kappa(u,v) \geq 0\}.$$

(See Figure 3.6.) Considering C^κ as a subset of $\mathcal{T}_x(\mathbb{T}^2)$ by the standard coordinates, we show first that $\{C^\kappa\}_x$ is expanding with respect to T_λ, one considers the action of $D_{(x,y)}(T_\lambda)$ and its inverse on the vector $(1,\eta)$, where $\eta \in [0,\kappa]$:

$$D_{(x,y)}(T_\lambda) \begin{pmatrix} 1 \\ \eta \end{pmatrix} = \begin{pmatrix} \varepsilon(x) + \eta \\ 1 \end{pmatrix}$$

lies on the line with gradient $0 < (\varepsilon(x) + \eta)^{-1} \leq \big(\varepsilon(x)\big)^{-1} < \kappa$. By linearity, one sees that each line in C^κ is mapped to a line in the interior of C^κ, i.e., $\{C^\kappa\}_x$ is expanding.

For co-expansiveness of $\{C^\kappa\}_x$, it suffices to consider the action on the boundary of $C^\kappa_{(x,y)}$, i.e., the lines spanned by $(1,0)$ and $(1,\kappa)$:

$$\big(D_{(x,y)}(T_\lambda)\big)^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -\varepsilon(x) \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

and $\big(D_{(x,y)}(T_\lambda)\big)^{-1}(1,\kappa)^T = \big(\kappa, 1 - \kappa\varepsilon(x)\big)^T$. Since both of these vectors lie outside of the closure of C^κ, the situation is as depicted in Figure 3.6.

Therefore the cone family $\{C^\kappa\}_{x \in \mathbb{T}}$ is expanding and co-expanding with respect to T_λ, and hence T_λ is Anosov by Fact 1, as required. $\square$

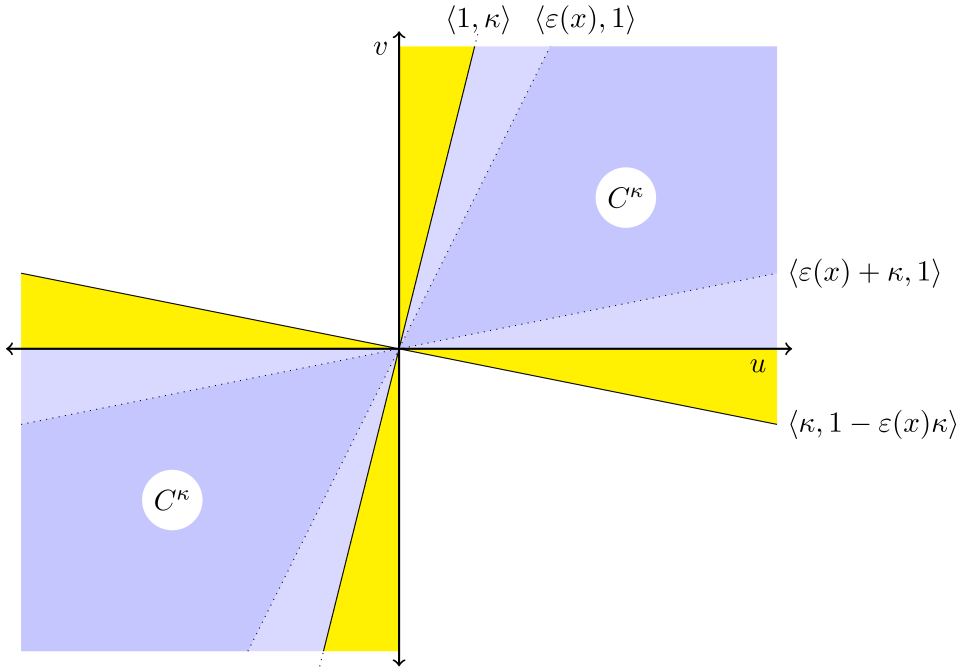


Figure 3.6: The cone C^κ (light blue) and its images under $D_{(x,y)}(T_\lambda)$ and its inverse (deeper shade of blue and yellow, respectively), showing that C^κ is expanding and co-expanding with respect to T_λ. To have these inclusions, the constant κ is chosen to satisfy (3.25).

3.3.2 The Hilbert space $\mathcal{H}_{a,\phi}$

The space $\mathcal{H}_{a,\phi}$ is defined analogously to $\mathcal{H}_a$. The weights here, $\|e_{m,n}\|_{a,\phi}$, depend on the following convenient generalisation, $\deg_\phi$, of $\deg_1$.

Definition 25 ($\deg_\phi$, $\|\cdot\|_{a,\phi}$, $\mathcal{H}_{a,\phi}$)**.** For $\phi > 1$, let

$$\deg_\phi(m,n) := \deg_1(m, \phi^{-\operatorname{sign}(m,n)} n) = \begin{cases} \;\;|m| + \phi^{-1}|n| & \text{if } mn \geq 0; \\ -|m| - \phi \;\;\; |n| & \text{if } mn < 0. \end{cases}$$

For $a > 0$, one has

$$\|e_{m,n}\|_{a,\phi} := e^{-a \deg_\phi(m,n)}.$$

This norm extends to arbitrary linear combinations of the $e_{m,n}$ as before:

$$\Big\| \sum_{m,n} b_{m,n} e_{m,n} \Big\|_{a,\phi}^2 = \sum_{m,n} |b_{m,n}|^2 \|e_{m,n}\|_{a,\phi}^2,$$

which gives rise to the Hilbert space (with an implicit inner product)

$$\mathcal{H}_{a,\phi} = \left\{ f = \sum_{(m,n)\in\mathbb{Z}^2} b_{m,n} \;\middle|\; b_{m,n} \in \mathbb{C},\ \|f\|_{a,\phi} < \infty \right\}.$$

The following result shows that, like $\mathcal{H}_a$, $\mathcal{H}_{a,\phi}$ can be chosen to contain any pair of functions analytic functions on a neighbourhood of the torus.

Proposition 12. *For $a > 0$ and $\phi > 1$, suppose that f is an analytic function on a neighbourhood of the poly-annulus*

$$P_{a,\phi} := \{(z, w) \in \mathbb{C}^2 \mid e^{-a} \le |z| \le e^a,\ e^{-a\phi} \le |w| \le e^{a\phi}\}.$$

Then $f \in \mathcal{H}_{a,\phi}$. In particular, every function analytic on a neighbourhood of $\mathbb{T}^2$ is contained in $\mathcal{H}_{a,\phi}$, for all (a, ϕ) such that $a\phi$ is sufficiently small.

The proof of the proposition is very similar to that of Proposition 10.

Proof of Proposition 12. Fix a, ϕ and f as above. By construction, the expansion

$$f(z, w) = \sum_{(m,n)\in\mathbb{Z}^2} b_{m,n}\, z^m w^n \tag{3.26}$$

converges absolutely for all $(z, w) \in P_{a,\phi}$. Also, one has the following bound from the definition of $\|f\|_{a,\phi}$, using that $-\deg_\phi(m, n) \le |m| + \phi|n|$:

$$\|f\|_{a,\phi}^2 := \sum_{(m,n)\in\mathbb{Z}^2} |b_{m,n}|^2\, e^{-2a\deg_\phi(m,n)} \le \sum_{(m,n)\in\mathbb{Z}^2} |b_{m,n}|^2\, e^{2a(|m|+\phi|n|)}. \tag{3.27}$$

Considering the right hand side, one bounds a related sum

$$\begin{aligned}\sum_{(m,n)\in\mathbb{Z}^2)} |b_{m,n}|\, e^{a(|m|+\phi|n|)} \le & \sum_{(m,n)\in\mathbb{Z}^2)} |b_{m,n}| e^{a(m+\phi n)} + \sum_{(m,n)\in\mathbb{Z}^2)} |b_{m,n}| e^{a(m-\phi n)} \\ & + \sum_{(m,n)\in\mathbb{Z}^2)} |b_{m,n}| e^{a(-m-\phi n)} + \sum_{(m,n)\in\mathbb{Z}^2)} |b_{m,n}| e^{a(-m+\phi n)},\end{aligned}$$

each of which is convergent by the absolute convergence of (3.26) for all $(z, w) \in \{(e^{\pm a}, e^{\pm a\phi})\} \subset P_{a,\phi}$. In particular, the sum on the left is square-summable, i.e., the sum on the right hand side of (3.27) is finite. Thus, $f \in \mathcal{H}_{a,\phi}$ as required. □

3.3.3 $\mathcal{C}_{T_\lambda}$ is Hilbert-Schmidt

To motivate the use of $\mathcal{H}_{a,\phi}$, we now provide the following negative result, which in particular shows that $\mathcal{C}_{T_\lambda}$ does not act compactly on either $\mathcal{H}_a$ or the anisotropic space used in [88] (i.e., defined by (3.7)), for any non-zero λ.

Proposition 13. *Suppose that $\mathcal{H}$ is a Hilbert space which has $\{e_{m,n}\}_{m,n}$ as an orthogonal basis, and satisfies, for all $(m,n) \in \mathbb{Z}^2$,*

$$\|e_{m,n}\| = \|e_{n,m}\|.$$

Then, $\mathcal{C}_{T_\lambda}$ is not compact on $\mathcal{H}$, for any $\lambda \in \mathbb{D} \setminus \{0\}$ (and may even be unbounded).

Proof of Proposition 13. Fix $m \in \mathbb{N}$ and $\lambda \neq 0$. Then, one writes

$$e_{m,n}\big(T_\lambda(z,w)\big) = w^m \left(\frac{z+\lambda}{1+\bar{\lambda}z}\right)^m z^n = \sum_{k=0}^{\infty} \alpha_{m,k} z^{n+k} w^m,$$

where the $\alpha_{m,k}$ are the Taylor coefficients defined in §3.2.3. That is,

$$\mathcal{C}_{T_\lambda}(e_{m,n}) = \sum_{k=0}^{\infty} \alpha_{m,k} e_{n+k,m}.$$

In particular, by orthogonality,

$$\|\mathcal{C}_{T_\lambda}(e_{m,n})\|^2 \geq |\alpha_{m,0}|^2 \|e_{n,m}\|^2 = |\lambda|^{2m} \|e_{m,n}\|^2. \tag{3.28}$$

Now assume for contradiction that $\mathcal{C}_{T_\lambda}$ is compact. Let $\hat{e}_{m,n} = e_{m,n}/\|e_{m,n}\|$ for each n. Then $\big(\mathcal{C}_{T_\lambda}(\hat{e}_{m,n})\big)_{n=1}^{\infty}$ must have a convergent subsequence. The limit of this subsequence must be zero since, for each $y \in \mathcal{H}$ and for $\langle \cdot, \cdot \rangle$ denoting the inner product on $\mathcal{H}$,

$$\langle y, \mathcal{C}_{T_\lambda}(\hat{e}_n) \rangle = \langle (\mathcal{C}_{T_\lambda})^* (y), \hat{e}_n \rangle \to 0,$$

by Bessel's inequality [28, p.15]. But by (3.28), $\mathcal{C}_{T_\lambda}(\hat{e}_{m,n}) \geq |\lambda|^{2m}$ for each n, a contradiction. Therefore $\mathcal{C}_{T_\lambda}$ is non-compact on $\mathcal{H}$, as required. $\square$

To begin the proof of Theorem 7, we now give the appropriate positive result. Note that, fixing a and λ below, the hypothesis of this lemma is be satisfied for all ϕ sufficiently close to 1.

Lemma 20. *Given $\lambda \in \mathbb{D}$, $a > 0$ and $\phi > 1$, if*

$$2a(\phi - 1) < -\log M_{a,\lambda},$$

the composition operator $\mathcal{C}_{T_\lambda} : \mathcal{H}_{a,\phi} \to \mathcal{H}_{a,\phi}$ is Hilbert-Schmidt.

The proof of this result is analogous to that of Lemma 16.

Proof of Proposition 20. Formally expanding

$$e_{m,n}\big(T_\lambda(z,w)\big) = w^m \left(\frac{z+w}{1+\overline{w}z}\right)^m z^n,$$

we have the following:

$$\mathcal{C}_{T_\lambda}(e_{m,n}) = \begin{cases} \displaystyle\sum_{k=0}^{\infty} \alpha_{m,k}\, e_{n+k,m}, & m > 0; \\ \qquad e_{n,m}, & m = 0; \\ \displaystyle\sum_{k=0}^{\infty} \alpha_{m,k}\, e_{n-k,m}, & m < 0. \end{cases} \tag{3.29}$$

First consider the case that $m \neq 0$. For $\sigma = \mathrm{sign}(m)$, one has

$$\begin{aligned} \left(\frac{\|\mathcal{C}_{T_\lambda}(e_{m,n})\|_{a,\phi}}{\|e_{m,n}\|_{a,\phi}}\right)^2 &= \sum_{k=0}^{\infty} |\alpha_{m,k}|^2 \left(\frac{\|e_{n+\sigma k,m}\|_{a,\phi}}{\|e_{m,n}\|_{a,\phi}}\right)^2 \\ &= \sum_{k=0}^{\infty} |\alpha_{m,k}|^2 e^{2a\left(\deg_\phi(m,n) - \deg_\phi(n+\sigma k,m)\right)} \\ &= e^{2a\left(\deg_\phi(m,n) - \deg_\phi(n,m)\right)} \sum_{k=0}^{\infty} |\alpha_{m,k}|^2 e^{2a\left(\deg_\phi(n,m) - \deg_\phi(n+\sigma k,m)\right)}. \end{aligned}$$

First considering the prefactor, for all $(m,n) \in \mathbb{Z}^2$, one writes

$$\begin{aligned} I(m,n) := \deg_\phi(m,n) - \deg_\phi(n,m) &= \begin{cases} |m| + \phi^{-1}|n| - \big(|n| + \phi^{-1}|m|\big), & \text{if } mn \geq 0; \\ -|m| - \phi|n| - \big(-|n| - \phi|m|\big), & \text{if } mn < 0; \end{cases} \\ &= \begin{cases} \phi^{-1}(\phi-1)\big(|m| - |n|\big), & \text{if } mn \geq 0; \\ (\phi-1)\big(|m| - |n|\big), & \text{if } mn < 0. \end{cases} \end{aligned}$$

Also, as in the proof of Lemma 17, one has three cases for $\deg_\phi(n+\sigma,m)-\deg_\phi(n,m)$:

$$\begin{aligned}
\deg_\phi(n+\sigma,m)-\deg_\phi(n,m) &= \begin{cases} |n+\sigma|+\phi^{-1}|m|-|n|-\phi^{-1}|m|, & \text{if } mn\geq 0;\\ |n+\sigma|+\phi^{-1}|m|+|n|+\phi|m|, & \text{if } mn<0,\ |m|=1;\\ -|n+\sigma|-\phi|m|+|n|+\phi|m|, & \text{if } mn<0,\ |m|>1;\end{cases}\\
&= \begin{cases} |n|+1-|n|, & \text{if } mn\geq 0;\\ |n|-1+(\phi^{-1}+\phi)|m|+|n|, & \text{if } mn<0,\ |m|=1;\\ 1-|n|+|n|, & \text{if } mn<0,\ |m|>1;\end{cases}\\
&\geq 1.
\end{aligned}$$

Therefore by induction, for all $k\in\mathbb{N}$, $\deg_\phi(n+\sigma k,m)-\deg_\phi(n,m)\geq k$. Thus, for all $(m,n)\in(\mathbb{Z}\backslash\{0\})\times\mathbb{Z}$ (applying Lemma 15),

$$\begin{aligned}
\left(\frac{\|\mathcal{C}_{T_\lambda}(e_{m,n})\|_{a,\phi}}{\|e_{m,n}\|_{a,\phi}}\right)^2 &\leq e^{2aI(m,n)}\sum_{k=0}^{\infty}|\alpha_{m,k}|^2e^{-2ak}.\\
&\leq e^{2aI(m,n)}M_{a,\lambda}^{|m|}\\
&= \begin{cases} e^{2a\phi^{-1}(\phi-1)\big(|m|-|n|\big)}M_{a,\lambda}^{|m|}, & \text{if } mn\geq 0;\\ e^{2a(\phi-1)\big(|m|-|n|\big)}M_{a,\lambda}^{|m|}, & \text{if } mn<0.\end{cases}
\end{aligned}$$

Considering the exponents on the right hand side, if

$$2a(\phi-1)=2a\max\big(\phi-1,\phi^{-1}(\phi-1)\big)<-\log(M_{a,\lambda}),$$

then $\delta:=\min\big(2a\phi^{-1}(\phi-1),2a(1-\phi)-\log(M_{a,\lambda})\big)$ is positive and satisfies

$$\left(\frac{\|\mathcal{C}_{T_\lambda}(e_{m,n})\|_{a,\phi}}{\|e_{m,n}\|_{a,\phi}}\right)^2\leq e^{-\delta\big(|m|+|n|\big)}$$

whenever $m\neq 0$. This inequality also applies in the $m=0$ case:

$$\left(\frac{\|\mathcal{C}_{T_\lambda}(e_{0,n})\|_{a,\phi}}{\|e_{0,n}\|_{a,\phi}}\right)^2=\left(\frac{\|e_{n,0}\|_{a,\phi}}{\|e_{0,n}\|_{a,\phi}}\right)^2=e^{-2aI(0,n)}\leq e^{-2a\phi^{-1}(\phi-1)|n|}\leq e^{-\delta|n|}.$$

Thus, one has

$$\|\mathcal{C}_{T_\lambda}\|_{\mathrm{HS}}^2=\sum_{(m,n)\in\mathbb{Z}^2}\left(\frac{\|\mathcal{C}_{T_\lambda}(e_{m,n})\|_{a,\phi}}{\|e_{m,n}\|_{a,\phi}}\right)^2\leq\sum_{(m,n)\in\mathbb{Z}^2}e^{-\delta(|m|+|n|)}<\infty,$$

i.e., $\mathcal{C}_{T_\lambda}$ is Hilbert-Schmidt, as required. $\square$

Remark 20. In the vein of Proposition 13, the smallness of $a(\phi-1)$ is necessary for the boundedness of $\mathcal{C}_{T_\lambda}$ on $\mathcal{H}_{a,\phi}$: Briefly speaking, let $m \geq 0$ and fix $n < 0$. Then, considering the first term of the expansion in this case gives

$$\frac{\|\mathcal{C}_{T_\lambda}(e_{m,n})\|_{a,\phi}}{\|e_{m,n}\|_{a,\phi}} \geq |\lambda|^m \frac{\|e_{n,m}\|_{a,\phi}}{\|e_{m,n}\|_{a,\phi}} = |\lambda|^m e^{aI(m,n)} = |\lambda|^m e^{a(\phi-1)(m+n)}.$$

Thus, if $-\log|\lambda| < a(\phi-1)$, this expression is unbounded in m.

3.3.4 The spectrum of $\mathcal{C}_{T_\lambda}$

We now give the following conclusion to Theorem 7.

Remark 21. Note that by "up to coincidences in value", we mean under the assumption that $(m,n,\omega) \mapsto \omega\lambda_1^m\overline{\lambda}_1^n$ is injective (i.e., if the argument of λ is not a rational multiple of π). Whenever this is not the case, one simply sums the respective multiplicities.

Lemma 21. *For λ, a and ϕ as in Lemma 20, the spectrum of $\mathcal{C}_{T_\lambda} : \mathcal{H}_{a,\phi} \to \mathcal{H}_{a,\phi}$ is as follows, where λ_1 is a square root of λ:*

$$\{0,1\} \cup \left\{\omega\lambda_1^m\overline{\lambda}_1^n \mid \omega = \pm 1,\ (m,n) \in \mathbb{N}_0^2 \setminus \{(0,0)\}\right\}.$$

Up to coincidences in value, the eigenvalues $\omega\lambda_1^k$, $\omega\overline{\lambda}_1^k$ have algebraic multiplicity

$$N(k,\omega) = \begin{cases} \left\lfloor \frac{k}{2} \right\rfloor + 1, & \text{if } \omega = 1; \\ \left\lfloor \frac{k+1}{2} \right\rfloor, & \text{if } \omega = -1; \end{cases}$$

and all other non-zero eigenvalues are simple. Moreover, if the argument of λ is not a rational multiple of π, the above eigenvalues are all semi-simple.

The proof of this lemma is analogous to the proof of Lemma 19.

Proof of Lemma 21. Recalling that $\alpha_{m,0} = \lambda^m$ for $m \in \mathbb{N}$, the expansion (3.29) for $\mathcal{C}_{T_\lambda}(e_{m,n})$ reads

$$\mathcal{C}_{T_\lambda}(e_{m,n}) = \begin{cases} \lambda^m \ e_{n,m} + \sum_{k=1}^{\infty} \alpha_{m,k}\, e_{n+k,m}, & \text{if } m > 0; \\ \qquad e_{n,m}, & \text{if } m = 0; \\ \overline{\lambda}^{|m|} e_{n,m} + \sum_{k=1}^{\infty} \alpha_{m,k}\, e_{n-k,m}, & \text{if } m < 0. \end{cases}$$

Recalling Lemma 17, one has that, for any $m \neq 0$ and $k \in \mathbb{N}$,

$$\deg_1(m,n) = \deg_1(n,m) < \deg_1\big(n + \operatorname{sign}(m)k, m\big).$$

Since the first equality applies for $m = 0$ also, it follows immediately, considering the above expansion, that $\mathcal{C}_{T_\lambda}$ increases $\deg_1$, and moreover that $\big(\mathcal{C}_{T_\lambda}\big)_k(e_{m,n})$ is obtained by eliminating the sums from the right hand side, for $k = \deg_1(m,n)$. I.e.,

$$\big(\mathcal{C}_{T_\lambda}\big)_k = \Pi_{D_k} \circ \mathcal{C}_{T_\lambda} \circ \Pi_{D_k} : e_{m,n} \mapsto \begin{cases} \lambda^m e_{n,m}, & \text{if } m \geq 0,\ \deg_1(m,n) = k; \\ \overline{\lambda}^{|m|} e_{n,m}, & \text{if } m < 0,\ \deg_1(m,n) = k; \\ 0, & \text{otherwise}; \end{cases}$$

where $D_k := \operatorname{Span}\{e_{m,n} \mid \deg_1(m,n) = k\}$ as before. Thus, pairing up the $e_{m,n}$ and $e_{n,m}$ for $m \neq n$, one recovers the following block-diagonal matrix representation of $\big(\mathcal{C}_{T_\lambda}\big)_k$, depending on the value of k:

$$(\mathcal{C}_{T_\lambda})_k \cong \begin{cases} \big(1\big), & k = 0; \\ \displaystyle\bigoplus_{n=0}^{(k-2)/2} \begin{pmatrix} 0 & \lambda^n \\ \lambda^{k-n} & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & \overline{\lambda}^n \\ \overline{\lambda}^{k-n} & 0 \end{pmatrix} \oplus \big(\lambda^{k/2}\big) \oplus \big(\overline{\lambda}^{k/2}\big), & k \in 2\mathbb{N}; \\ \displaystyle\bigoplus_{n=1}^{(k-1)/2} \begin{pmatrix} 0 & \lambda^n \\ \lambda^{k-n} & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & \overline{\lambda}^n \\ \overline{\lambda}^{k-n} & 0 \end{pmatrix}, & k \in 2\mathbb{N} - 1; \\ \displaystyle\bigoplus_{n=1}^{k-1} \begin{pmatrix} 0 & \lambda^n \\ \overline{\lambda}^{k-n} & 0 \end{pmatrix}, & k < 0. \end{cases}$$

By Lemma 18, the non-zero eigenvalues of $\mathcal{C}_{T_\lambda}$ correspond precisely to those of the above matrices.

More explicitly, every non-zero eigenvalue is a square root of $\lambda^m \overline{\lambda}^n$ for some $(m,n) \in (\mathbb{N}_0)^2$. Counting the algebraic multiplicity, the square roots of $\lambda^m \overline{\lambda}^n$ are simple eigenvalues unless $mn = 0$; otherwise, for $\lambda_1^2 = \lambda$, the multiplicity of $\omega\lambda_1^k$ and $\omega\overline{\lambda}_1^k$ equals $N(k,\omega)$, where $N(2k,1) = k+1$, $N(2k,-1) = k$ and $N(2k-1,1) = N(2k-1,-1) = k$. This agrees with the formula given in the statement of the lemma:

$$N(k,\omega) = \begin{cases} \left\lfloor \frac{k}{2} \right\rfloor + 1, & \text{if } \omega = 1; \\ \left\lfloor \frac{k+1}{2} \right\rfloor, & \text{if } \omega = -1. \end{cases}$$

Finally, since the $\big(\mathcal{C}_{T_\lambda}\big)_k$ are diagonalisable, if λ is non-zero and its argument is an irrational multiple of π, the $\left(\mathcal{C}_{T_\lambda}\right)_k$ do not share any eigenvalues, which implies that every non-zero eigenvalue is semi-simple, completing the proof. □

3.4 The spectrum of $\mathcal{C}_{T_\lambda \circ T_\mu}$

Having established the machinery for T_λ, the following result for $T_\lambda \circ T_\mu$ $(\lambda, \mu \in \mathbb{D})$ will be very easy to prove. Again, we note that this family of examples appears in an appendix of [88], where their resonances are announced and numerically studied. We here provide a rigorous argument, as per the following result.

Theorem 8. *For $\lambda, \mu \in \mathbb{D}$ and $\mathcal{H}_{a,\phi}$ defined as above, if $a > 0$ and $\phi > 1$ satisfy*

$$2a(\phi - 1) < -\log\big(\max(M_{a,\lambda}, M_{a,\mu})\big), \tag{3.30}$$

then $\mathcal{C}_{T_\lambda \circ T_\mu} = \mathcal{C}_{T_\mu} \circ \mathcal{C}_{T_\lambda}$ acts compactly on $\mathcal{H}_{a,\phi}$ and has spectrum

$$\{0, 1\} \cup \big\{\lambda^m \mu^n,\ \lambda^m \overline{\mu}^n,\ \overline{\lambda}^m \overline{\mu}^n,\ \overline{\lambda}^m \mu^n \mid (m, n) \in \mathbb{N}_0^2 \setminus \{(0, 0)\}\big\}.$$

Moreover, all non-zero eigenvalues are simple, up to coincidences in value.

3.4.1 Hyperbolicity of $T_\lambda \circ T_\mu$

As before, we relate the spectrum above to the resonances by showing that the underlying map is Anosov and area-preserving.

Proposition 14. *For all $(\lambda, \mu) \in \mathbb{D} \times \mathbb{D}$, $T_\lambda \circ T_\mu$ is an area-preserving Anosov map of the torus.*

Proof of Proposition 14. This is a simple consequence of the statement and proof of Proposition 11. Firstly, the composition of two area-preserving maps is obviously area-preserving. Secondly, recalling the cone family $\{C^\kappa\}_{x \in \mathbb{T}}$ from the earlier proof, we recall it is expanding and co-expanding with respect to T_λ, T_μ respectively if

$$\kappa > \frac{1 + |\lambda|}{1 - |\lambda|}, \qquad \kappa > \frac{1 + |\mu|}{1 - |\mu|}.$$

That is, one can easily take κ large enough that $\{C^\kappa\}_{x \in \mathbb{T}}$ is expanding and co-expanding simultaneously with respect to both maps. It then follows trivially that it is expanding and co-expanding with respect to $T_\lambda \circ T_\mu$, and hence by Fact 1, this map is Anosov. □

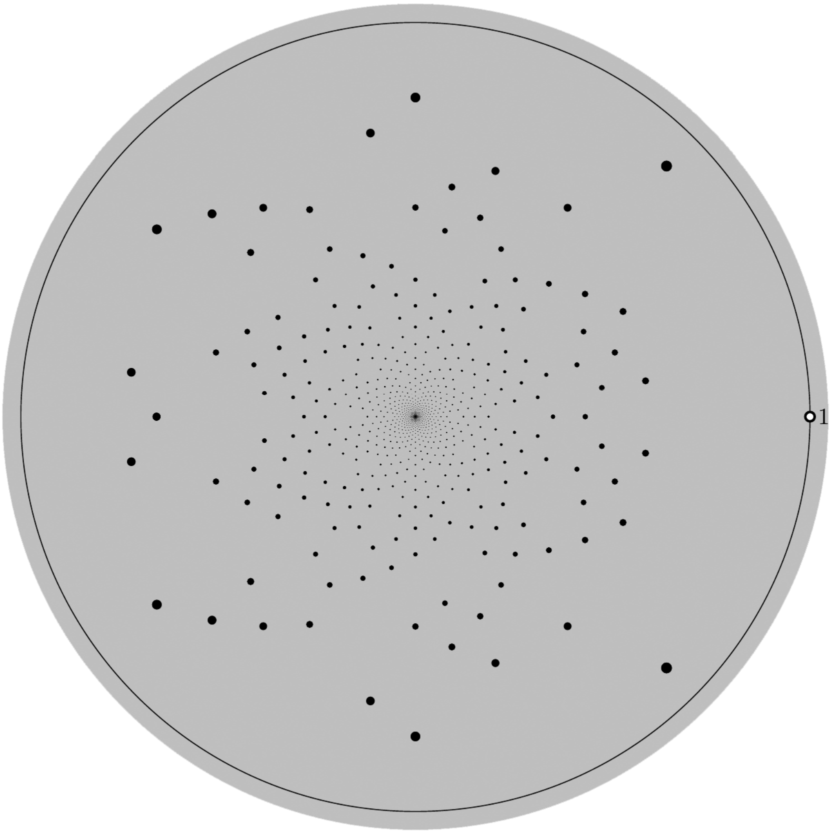


Figure 3.7: The spectrum of $\mathcal{C}_{T_\lambda \circ T_\mu}$, for $\lambda = 0.9e^{i\pi/4}$, $\mu = 0.65e^{6i\pi/5}$.

3.4.2 $\mathcal{C}_{T_\lambda \circ T_\mu}$ is trace-class

To begin the proof of Theorem 8, one has the following simple corollary of Lemma 20.

We first recall from [28, p.267] that a compact operator is trace-class if its singular values are summable, that being trace-class is a stronger property than being Hilbert-Schmidt, and that an operator is trace class if and only if it is the composition of two Hilbert-Schmidt operators.

Lemma 22. *For λ, μ, a, ϕ as in Theorem 8, $\mathcal{C}_{T_\lambda \circ T_\mu} : \mathcal{H}_{a,\phi} \to \mathcal{H}_{a,\phi}$ is trace-class.*

Remark 22. Again, since $B_\lambda = T_0 \circ T_\lambda$ for all λ, this shows that $\mathcal{C}_{B_\lambda}$ is trace-class as an operator on $\mathcal{H}_{a,\phi}$.

Proof of Lemma 22. By the hypothesis (3.30), Lemma 20 applies twice to show that $\mathcal{C}_{T_\lambda}$ and $\mathcal{C}_{T_\mu}$ are both Hilbert-Schmidt on $\mathcal{H}_{a,\phi}$. Thus $\mathcal{C}_{T_\lambda \circ T_\mu} = \mathcal{C}_{T_\mu} \circ \mathcal{C}_{T_\lambda}$ is the composition of two Hilbert-Schmidt operators, hence trace-class as required. $\square$

3.4.3 The spectrum of $\mathcal{C}_{T_\lambda} \circ \mathcal{C}_{T_\mu}$

The calculation of the spectrum likewise follows simply from the corresponding calculation for $\mathcal{C}_{T_\lambda}$. This uses the following lemma, which is again an extension of a corresponding intuitive result in finite dimensions:

$$\begin{pmatrix} A_1 & 0 & 0 & \cdots & 0 \\ * & A_2 & 0 & \cdots & 0 \\ * & * & A_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ * & * & * & \cdots & A_n \end{pmatrix} \begin{pmatrix} B_1 & 0 & 0 & \cdots & 0 \\ * & B_2 & 0 & \cdots & 0 \\ * & * & B_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ * & * & * & \cdots & B_n \end{pmatrix} = \begin{pmatrix} A_1B_1 & 0 & 0 & \cdots & 0 \\ * & A_2B_2 & 0 & \cdots & 0 \\ * & * & A_3B_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ * & * & * & \cdots & A_nB_n \end{pmatrix},$$

where A_k and B_k are square matrices with the same dimensions, for each k.

Lemma 23. *Let $\mathcal{H}$ be a Hilbert space such that $\{e_{m,n}\}_{(m,n)\in\mathbb{Z}^2}$ is an orthogonal basis, and let $\mathcal{C}_1, \mathcal{C}_2 : \mathcal{H} \to \mathcal{H}$ increase* $\deg_1$. *Then $\mathcal{C}_1 \circ \mathcal{C}_2$ increases* $\deg_1$ *and satisfies, for each k,*

$$\big(\mathcal{C}_1 \circ \mathcal{C}_2\big)_k = \big(\mathcal{C}_1\big)_k \circ \big(\mathcal{C}_2\big)_k. \tag{3.31}$$

Proof of Lemma 23. Recall that an equivalent way to write that $\mathcal{C}_i$ $(i = 1, 2)$ increases $\deg_1$ is the following, for $D_k = \operatorname{Span}\{e_{m,n} \mid \deg_1(m,n) = k\}$:

$$\mathcal{C}_i(D_k) \subset \bigoplus_{j\geq k} D_j. \tag{3.32}$$

It is then immediate that $\mathcal{C}_1 \circ \mathcal{C}_2$ increases deg:

$$\mathcal{C}_1 \circ \mathcal{C}_2(D_k) \subset \mathcal{C}_1\left(\bigoplus_{j\geq k} D_j\right) \subset \bigoplus_{j\geq k} D_j.$$

To prove (3.31), let $v \in D_k$. Also let $\mathcal{C}_2(v) = v_1 + v_2$, where $v_1 = \Pi_{D_k}\mathcal{C}_2(v)$. In particular, $v_2 \in \bigoplus_{j>k} D_j$ by (3.32), and hence

$$\mathcal{C}_1(v_2) \in \bigoplus_{j>k} D_k,$$

also by (3.32). Therefore, $\Pi_{D_k} \circ \mathcal{C}_1(v_1) = 0$ and

$$\Pi_{D_k} \circ \mathcal{C}_1 \circ \mathcal{C}_2(v) = \Pi_{D_k} \circ \mathcal{C}_1(v_1) + \Pi_{D_k} \circ \mathcal{C}_1(v_2) = \Pi_{D_k} \circ \mathcal{C}_1(v_1) = \Pi_{D_k} \circ \mathcal{C}_1 \circ \Pi_{D_k} \circ \mathcal{C}_2(v).$$

That is,

$$\big(\mathcal{C}_1 \circ \mathcal{C}_2\big)_k(v) = \big(\mathcal{C}_1\big)_k \circ \big(\mathcal{C}_2\big)_k(v).$$

The arbitrariness of $v \in D_k$ completes the proof. $\square$

We now apply this lemma to give the resonances of $T_\lambda \circ T_\mu$, following the earlier proofs.

Lemma 24. *For each* $(\lambda, \mu) \in \mathbb{D}^2$*, the spectrum of* $\mathcal{C}_{T_\lambda \circ T_\mu}$ *is given by*

$$\{0,1\} \cup \big\{\lambda^m \mu^n,\ \lambda^m \overline{\mu}^n,\ \overline{\lambda}^m \overline{\mu}^n,\ \overline{\lambda}^m \mu^n \mid (m,n) \in \mathbb{N}_0^2 \setminus \{(0,0)\}\big\}. \tag{3.33}$$

Moreover, each non-zero eigenvalue has algebraic multiplicity equal to the frequency with which it appears in (3.33)*.*

Proof of Lemma 24. Applying Lemmas 23 and 18 reduces the proof to a consideration of the eigenvalues of $\big(\mathcal{C}_{T_\lambda \circ T_\mu}\big)_k = \big(\mathcal{C}_{T_\lambda}\big)_k \circ \big(\mathcal{C}_{T_\mu}\big)_k$. We recall from the proof of Lemma 21 that, for $k = \deg_1(m,n)$,

$$\big(\mathcal{C}_{T_\mu}\big)_k(e_{m,n}) = \Pi_{D_k} \circ \mathcal{C}_{T_\mu}(e_{m,n}) = \begin{cases} \mu^m\ e_{n,m}, & \text{if } m \geq 0; \\ \overline{\mu}^{|m|} e_{n,m}, & \text{if } m < 0. \end{cases}$$

Thus, for $k = \deg_1(m,n) = \deg_1(n,m)$,

$$\big(\mathcal{C}_{T_\lambda \circ T_\mu}\big)_k(e_{m,n}) = \begin{cases} \mu^m\ \lambda^n\ e_{m,n}, & \text{if } m \geq 0,\ n \geq 0; \\ \overline{\mu}^{|m|} \lambda^n\ e_{m,n}, & \text{if } m < 0,\ n \geq 0; \\ \mu^m\ \overline{\lambda}^{|n|} e_{m,n}, & \text{if } m \geq 0,\ n < 0; \\ \overline{\mu}^{|m|} \overline{\lambda}^{|n|} e_{m,n}, & \text{if } m < 0,\ n < 0. \end{cases}$$

That is, each $\big(\mathcal{C}_{T_\lambda \circ T_\mu}\big)_k$ is diagonal. Since the prefactor of $e_{m,n}$ is unique (up to coincidences in value), this shows that the spectrum is given by

$$\{0,1\} \cup \big\{\lambda^m \mu^n,\ \lambda^m \overline{\mu}^n,\ \overline{\lambda}^m \overline{\mu}^n,\ \overline{\lambda}^m \mu^n \mid (m,n) \in \mathbb{N}_0^2 \setminus \{(0,0)\}\big\},$$

and that the non-zero eigenvalues are simple, up to coincidences in value (e.g. if λ, μ and μ/λ are non-zero and have arguments which are irrational multiples of π). $\square$

This completes the proof of Theorem 8.

3.5 Conclusions and further work

The arguments used in the preceding section leave many questions for the reader. For example:

- Does the method apply to give other families of examples?
- In particular, can one produce examples in higher dimensions or with more intricate structure?
- How representative are these resonances of the general picture?

To begin to answer these questions, we propose the following generalisation of B_λ and T_λ to maps on the n-torus,

$$\mathbb{T}^n := \{z = (z_1, \dots, z_n) \in \mathbb{C}^n : |z_k| = 1 \text{ for all } k\}.$$

Definition 26. Given the matrix $A = (A_{j,k})_{j,k} \in \mathbb{Z}^{n,n}$, and n-tuples $b, r \in \mathbb{Z}^n$, let $S_\lambda : \mathbb{T}^n \to \mathbb{T}^n$ be given by

$$S_\lambda(z) = \left(z^{A_{\cdot,k}} \left(\frac{z^b + \lambda}{1 + \overline{\lambda} z^b} \right)^{r_k} \right)_{k=1}^{n},$$

where we denote

$$z^b := \prod_{j=1}^{k} z_j^{b_j} \qquad \text{and} \qquad z^{A_{\cdot,k}} := \prod_{j=1}^{k} z_j^{A_{j,k}}.$$

Denoting $B \in \mathbb{Z}^{n,n}$ as the rank one matrix given by $B_{j,k} = b_j r_k$, S_λ is a perturbation of

$$S_0 : z \mapsto (z^{A_{\cdot,k} + B_{\cdot,k}})_{k=1}^{n},$$

which in standard coordinates is the linear toral map given by $A + B$, i.e., $v \mapsto (A + B)v \pmod 1$.

Example 14. As mentioned, this family generalises (B_λ) and (T_λ) above:

- B_λ corresponds to $A = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, $b = (1, 0)$ and $r = (1, 1)$.
- T_λ corresponds to $A = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, $b = (1, 0)$ and $r = (1, 0)$.

In view of this definition, we conclude this chapter by alluding to future work regarding S_λ. This we do with two remarks. The first remark considers obstacles to the method presented above, in the context of attaining new examples.

Remark 23. By assuming hypotheses on S_λ, its composition operator $\mathcal{C}_{S_\lambda}$ and the spectrum of $\mathcal{C}_{S_\lambda}$,[¶] we obtain the following obstructions to the above method, expressed as linear algebraic conditions on A, B, b and r:

[¶]Namely, that a) S_0 is hyperbolic and area-preserving, b) the composition operator is compact on a suitable space and admits a block-triangular form with respect to an ordering of $\{e_v : z \mapsto z^v\}_{v \in \mathbb{Z}^n}$, and c) the spectrum is non-trivial and computable from the first term of the formal Taylor expansion

- $\det(A + B)$ is hyperbolic matrix which has determinant ± 1.
- A has an eigenvalue which is a root of unity.
- There exists $\deg : \mathbb{Z}^n \to \mathbb{Z}$, such that $|\deg^{-1}(m)| < \infty$ for each $m \in \mathbb{Z}$, and for each $v \in \mathbb{Z}^n$,
$$\deg(A^T v) \geq \deg(v)$$
and, whenever $r \cdot v \neq 0$, for all $k \geq \mathbb{N}$,
$$\deg(A^T v + (-1)^{r \cdot v} kb) > \deg(v).$$

The second remark is more positive: assuming the method applies, we can make certain structural statements about the spectrum.

Remark 24. Assuming the composition operator $\mathcal{C}_{S_\lambda}$ acts compactly on a suitable Hilbert space with orthogonal basis $\{e_v : z \mapsto z^v\}_{v \in \mathbb{Z}^n}$ and satisfies the conditions of the previous remark, we have the following:

- Each non-zero eigenvector of $\mathcal{C}_{S_\lambda}$ can be written as a multiple of non-negative powers of the mth roots of λ and $\overline{\lambda}$, where m is the maximum period of a periodic orbit of the linear action of A on $\mathbb{Z}^n$.
- If α is an eigenvalue of $\mathcal{C}_{S_\lambda}$, then so too is $\overline{\alpha}$ and α^k, for every $k \in \mathbb{N}$.

It is our expectation that if there are more examples to be derived from the method of this chapter, the family S_λ is a natural place to begin to search. However, the conditions in the first remark have so far proven to be quite exclusive. Simultaneously, proving that candidate examples do not satisfy these criteria, mainly showing the non-existence of the function deg above, has proven to be quite involved, especially when one leaves the comfort of two dimensions. To simplify these conditions would make an interesting technical challenge in its own right.

Chapter 4

A simple approach to bounding the Hausdorff dimension of the Rauzy gasket

4.1 Introduction

4.1.1 A brief history

The Rauzy gasket (after Gérard Rauzy), named by Pierre Arnoux and Štěphán Starosta in [7],* has a long and varied history, appearing multiple times in different contexts in dynamical systems, topology and combinatorics on words.

In each case, the gasket, denoted $\mathcal{G}$, represents an important subset of two-dimensional parameter space Δ, the standard two-simplex, corresponding to exotic and rare behaviour (for more details, we refer the reader to [39]):

- In 1993, Gilbert Levitt in [63] considers a pseudogroup of partially defined rotations of the circle, indexed by Δ, where minimal elements (i.e., those having no non-empty, proper and closed invariant subset) are those indexed by $\mathcal{G}$, which is shown to have zero two-dimensional Lebesgue measure, attributed to Jean-Christophe Yoccoz.
- Arnoux and Rauzy in [6], generalising a construction by Arnoux–Yoccoz in [8] of a minimal interval exchange transformation,† produce a family, indexed by

*It is also referred to in [32] as the Levitt–Yoccoz gasket, in honour of those who first described the gasket's properties.

†Their construction (and its generalisations) permute intervals which have rationally dependent lengths, so their minimality is not covered by the classical results of M. Keane [60].

Δ of interval exchange transformations (on six or seven intervals), which are minimal precisely when the index lies in $\mathcal{G}$.

- In connection to Novikov's problem, regarding the connectivity of intersections with triply periodic surfaces in $\mathbb{R}^3$ with families of planes orthogonal to some v, Roberto De Leo and Ivan Dynnikov in [33] investigate the case of a particular piecewise linear surface, $\{4, 6 \mid 4\}$. For this surface, the $v \in \Delta$ which admit so-called chaotic behaviour are precisely those which lie in $\mathcal{G}$. The authors give an independent proof that $\mathcal{G}$ has Lebesgue measure zero, and also empirically estimate the box-counting dimension of $\mathcal{G}$ (see section 4.1.3 below).

- The above work of Arnoux–Rauzy in fact focuses on a natural class of Sturmian words (sequences of minimal complexity on finitely many symbols, again see [46] for a definition and the historical connection to cutting sequences), known as *episturmian words* or *Arnoux–Rauzy words*. Arnoux and Štěpán Starosta in [7] show that all such words on three symbols are indexed by $\mathcal{G}$. Also, by relating $\mathcal{G}$ to the set of vectors in Δ for which the fully subtractive algorithm (a particular generalisation of the continued fraction algorithm in two or more dimensions) converges, the authors give yet another proof that the gasket has zero Lebesgue measure, using a result originally due to Meester–Nowicki [71].

- Finally, Pascal Hubert and Olga Paris-Romaskevich in [56] recently proved, in the context of triangular tiling billiards, that the triangles exhibiting "chaotic behaviour" (which is analogous to the Novikov case above) are those whose angles, after a simple transformation, lie in $\mathcal{G}$.

Before mentioning the results pertaining to the Hausdorff dimension of the gasket, we compare $\mathcal{G}$ to two related (and better-known) gaskets.

We note here that the term "gasket", in the context of engine design, refers to seals which fill the space between distinct (non-mating) parts of an engine, which typically have holes of many different sizes. Indeed, the fractals considered below can be naturally thought of as the limit of a process of iteratively excising triangular or circular holes of decreasing sizes from a region in the plane.

4.1.2 Three related fractals

To illustrate the difficulty in considering the Rauzy gasket, we consider two related fractals, which are mutually homeomorphic [7].

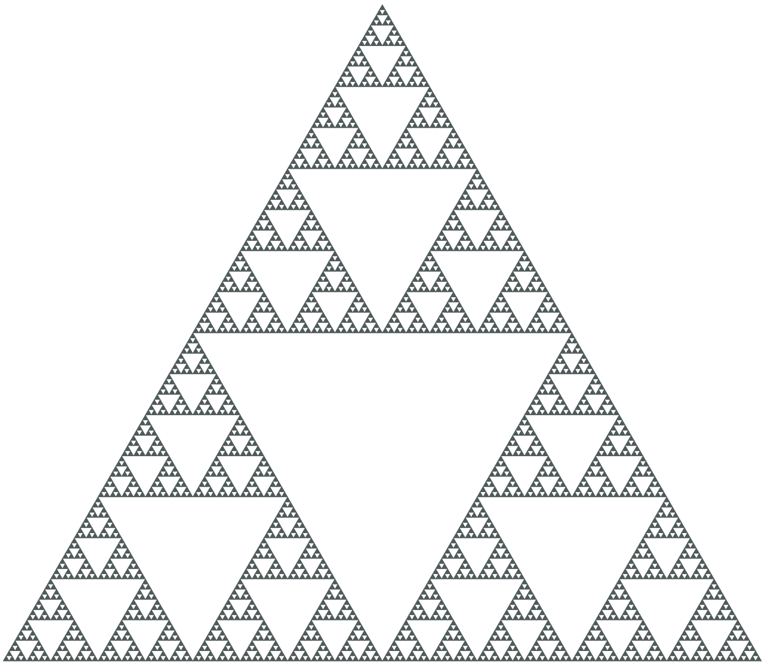


Figure 4.1: The Sierpiński gasket, $\mathcal{S}$.

The Sierpiński gasket

The Sierpiński gasket, $\mathcal{S}$, introduced by Wacław Sierpiński in a 1915 paper [86], is the familiar fractal example seen in first courses in fractal geometry. It is the limit set of the three similarities $x \mapsto \frac{1}{2}(x + e_k)$, where $(e_k)_{k=1}^3$ denote the vertices of a triangle in the plane. See Figure 4.1 for an equilateral realisation.

Owing to the rigidity of these maps, it is simple to deduce from the definitions that this gasket has Hausdorff dimension (and indeed box-counting and other dimensions) precisely equal to

$$\dim_H(S) = \frac{\log(3)}{\log(2)} = 1.5849\ldots.$$

Finally, we note that although quite a stock example, there are interesting connections to, for example, cellular automata and Pascal's triangle (see the Wikipedia page for more details).

The Apollonian gasket

For more details on the following history, see [76]. The Apollonian gasket is named after the Greek geometer Apollonius of Perga (ca. 200BC), who first studied the geometric problem of constructing circles tangent to three given circles. In particular, he proved that, given any three mutually tangent circles, there exist two circles, known as Apollonian circles, which are tangent to all three (see Figure 4.2). Observe

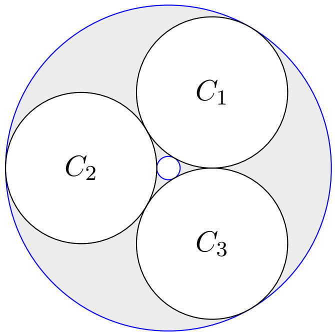


Figure 4.2: The two Apollonian circles (blue) which are tangent to C_1, C_2 and C_3.

in Figure 4.2 that one may continue this process by adding an Apollonian circle inside each of the six grey curvilinear triangles, creating six new curvilinear triangles, and so on indefinitely. For each choice of C_1 up to C_3, this defines a packing of the larger blue circle by infinitely many Apollonian circles, known generally as an *Apollonian circle packing*.

Returning again to Figure 4.2, suppose that C_4 is an Apollonian circle tangent to C_1, C_2 and C_3 (which are mutually tangent), and denote the radii of C_k by r_k, with the convention that the sign of r_k is reversed if C_k contains the other three circles (i.e., in this case the radius of C_k is $-r_k$). Then the following famous formula relates these radii (a simple proof is given in [83]):

$$2\left(\frac{1}{r_1^2}+\frac{1}{r_2^2}+\frac{1}{r_3^2}+\frac{1}{r_4^2}\right)=\left(\frac{1}{r_1}+\frac{1}{r_2}+\frac{1}{r_3}+\frac{1}{r_4}\right)^2.$$

This formula was originally proclaimed by the philosopher Rene Descartes in a 1643 letter to Princess Elizabeth of Bohemia, who gave an independent proof. The Nobel laureate Frederick Soddy, upon rediscovering this formula, published it in the poem, “A Kiss Precise”, in a 1936 edition of Nature [92]. For this reason, the circles in the Apollonian packing are sometimes known as “Soddy circles”.

Enumerating the radii of the Apollonian/Soddy circles by $(r_k)_{k=1}^{\infty}$, this sequence satisfies some very interesting number-theoretic properties (see e.g., [49]). For example, if the reciprocals r_1^{-1}, r_2^{-1} and r_3^{-1} (corresponding to the circles C_1, C_2 and C_3 above) are integer, then r_k^{-1} is integer for all $k \in \mathbb{N}$.

One may also consider the rate at which the radii converge to zero. The following is a result of Alex Kontorovich and Hee Oh [61]:

Proposition 15. *For $T > 0$, let $N(T) = |\{k \in \mathbb{N} : r_k \geq T^{-1}\}|$ denote the number of circles in a given Apollonian circle packing which have radii at least $1/T$. Then there exists C, $d > 0$ such that*

$$\frac{N(T)}{T^d} \to C$$

as $T \to \infty$, where d is independent of the radii of the original three circles.

This value d is known as the *packing exponent*. As we shall see, it is strongly related to the Hausdorff dimension of $\mathcal{A}$, the Apollonian gasket.

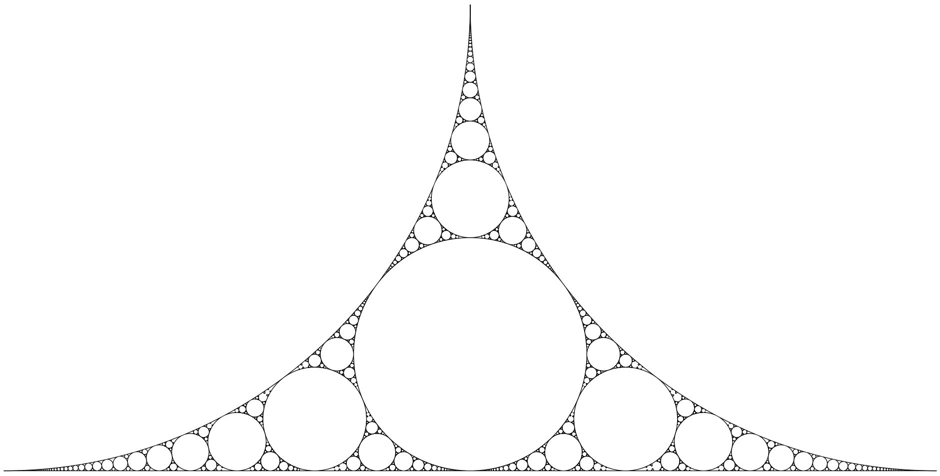


Figure 4.3: The Apollonian gasket, $\mathcal{A}$

Up to a conformal change of coordinates, the Apollonian gasket $\mathcal{A}$ is the restriction of the Apollonian circle packing to a single curvilinear triangle. It also admits a convenient description as the limit set of three Möbius maps, $(f_k)_{k=1}^3$. More explicitly, considering its vertices to be at ± 1 and i, as depicted in Figure 4.3, these maps are given by

$$f_1(z) = \frac{z-1}{z+3}, \qquad f_2(z) = \frac{z+1}{3-z}, \qquad \text{and} \qquad f_3(z) = \frac{1}{z-2i}.$$

Unlike the case of the Sierpiński gasket, these maps are not contractions, in the sense that they each have Lipschitz constant equal to one. More precisely, the fixed points are indifferent: each f_k fixes a corner of $\mathcal{A}$, z_k say, and $f'(z_k) = 1$.

Visually, this fact corresponds to the "slow decay" seen in the size of the circles approaching one vertex: For example, starting with the central circle in Figure 4.3 and considering the chain of circles proceeding towards minus 1, the nth circle encountered (i.e., f_1 applied $n-1$ times to this central circle) has radius equal to $(n+2)^{-2}$. In contrast, the corresponding triangles of the Sierpiński gasket have diameter proportional to 3^{-n}.

That the f_k are not contracting makes estimating the dimension of $\mathcal{A}$ much more difficult than the Sierpiński gasket. The redeeming feature of these maps is their conformality, which corresponds visually to the fact that the Soddy circles are all circular, irrespective of their size. In particular, this leads to the following precise formula relating the Hausdorff dimension to the packing exponent d, due to David Boyd [18].

Proposition 16.

$$\dim_H(\mathcal{A}) = \inf\left\{t > 0 : \sum_{k=1}^{\infty} r_k^t < \infty\right\} = d,$$

where $(r_k)_k$ and d are as above.

Despite this convenient formulation for the Hausdorff dimension, its exact value is not known. We have rigorous bounds due again to Boyd, $1.300197 < \dim_H(\mathcal{A}) < 1.314534$ [18], and an estimate of Curtis McMullen in [70] gives $\dim_H(\mathcal{A}) \approx 1.30568\ldots$, which is in accordance with more recent empirical estimates for the dimension (see [11] and references therein).

The Rauzy gasket

The Rauzy gasket in the fractal geometry sense represents the worst of all worlds, in the sense that its three attracting maps are neither conformal nor strict contractions. Visually, not only do the triangles in Figure 1.2 decay in size polynomially as they approach the corners, but also show signs of being highly distorted or sheared (e.g., close to the bottom edge in the figure).

These phenomena make a study of the Hausdorff dimension of the gasket even more difficult than that of the Apollonian gasket, particularly in regard to lower bounds. We now present what is known so far in the literature.

4.1.3 Results on the Hausdorff dimension of the Rauzy gasket

Since, in the contexts in which it appears, the Rauzy gasket parametrises exotic behaviour that goes "unseen" (particularly in relation to Novikov's problem, which corresponds to real-world experiments), it is a natural expectation that the gasket has zero two-dimensional Lebesgue measure. A stronger conjecture (now affirmed) is the following, a particular case of an open conjecture of Novikov and Maltsev.

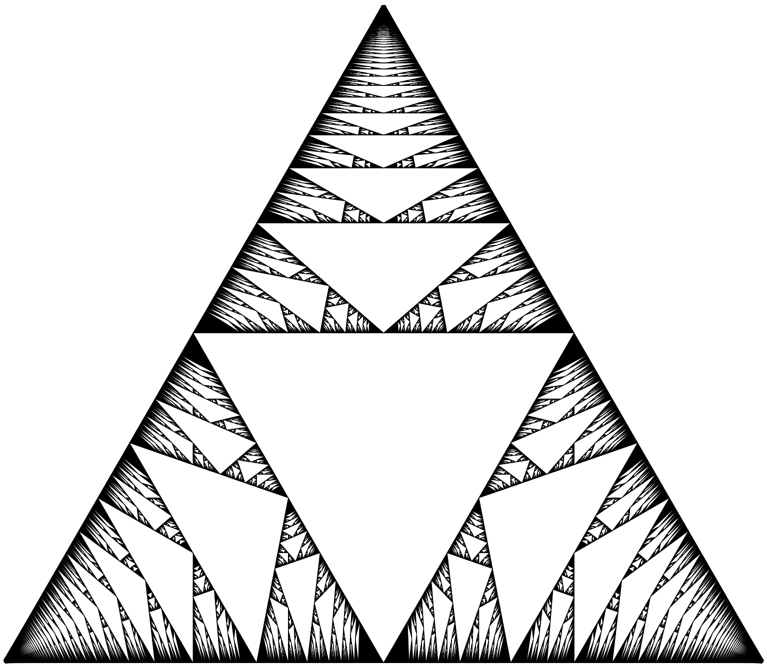


Figure 4.4: The Rauzy gasket, $\mathcal{G}$.

Conjecture 2 (Novikov–Maltsev, [75])**.** *The Rauzy gasket has Hausdorff dimension strictly between 1 and 2.*

Perhaps motivated by this conjecture, there have been several recent attempts to study the Hausdorff dimension of $\mathcal{G}$ (including a contribution from the Fields Medallist Artur Avila):

- As mentioned above, De Leo and Dynnikov in [33] also provide non-rigorous estimates for the box-counting dimension, giving $\dim_B(\mathcal{G}) \approx 1.72$, which suggests this figure as an upper bound for $\dim_H(\mathcal{G})$.[‡]

- Artur Avila, Pascal Hubert and Alexandra Skripchenko in [10] rigorously showed that $\dim_H(\mathcal{G}) < 2$. Although their method does yield a precise bound, they suggest that it would not be worth the effort to calculate it.

- With respect to lower bounds on the dimension, a recent publication of Rodolfo Gutérrio-Romo and Carlos Matheus [53] showed $\dim_H(\mathcal{G}) \geq 1.19$, completing the proof of Conjecture 2.

[‡]The authors state that these estimates suggest $\dim_H(\mathcal{G})$ to lie between 1.7 and 1.8, but it is not clear why.

- In [32], De Leo advances a conjecture, supported by numerical evidence, which would imply that $\dim_H(\mathcal{G}) \geq 1.63$.
- Finally, and most recently, a preprint [47] of Charles Fougeron shows that $\dim_H(\mathcal{G}) \leq 1.825$, as a consequence of a more general theory involving suspension flows and thermodynamic formalism.

It is worth emphasising, with the two works, [10] and [47], which rigorously bound the dimension above, that the methods used therein are long and complicated.

As mentioned in the introduction, the aim of the present chapter is to provide a simple proof that $\dim_H(\mathcal{G}) < 2$, and to prove stronger bounds for $\dim_H(\mathcal{G})$. In summary, our best bound is the following, rounded upwards to four decimal places.

Theorem 9. *$\dim_H(\mathcal{G}) \leq 1.7407$.*

This bound can be improved upon with greater computational power, subject to limiting returns. The above figure was confirmed in 180 seconds on Mathematica 11 on the author's laptop (we give more details in subsubsection 4.6.4 below), with memory seemingly being the limiting factor.

The method of proof in both cases is quite elementary, relying on ideas from Markov theory and the renewal theorem.

Contents of the chapter

On the way to proving Theorem 9, we prove four results which each give upper bounds on the dimension, of increasing complexity and efficacy; this allows us to build up the method in a gradual, systematic way. These results, corresponding to the contents of sections 4.3–4.5 and 4.7 below, are based on a key lemma (Lemma 25). In more detail, we have the following sections:

- In section 4.2, after giving some necessary definitions and prerequisites, we prove Lemma 25 which, by providing a sequence of open covers, allows us to bound the Hausdorff dimension in terms of parameter values for which the sequence (Y_n), whose definition involves areas and diameters of "level n triangles", converges to zero.
- In section 4.3, by eschewing the diameter factors in (Y_n), we obtain a related sequence (X_n), with $X_n \geq Y_n$, and we give it an explicit formula.
- In section 4.4, we give a simple proof that $\dim_H(\mathcal{G}) < 2$, using the renewal theorem to give parameter values for which $(X_n) \to 0$.

- In section 4.5, we refine the method of section 4.4, to give a decreasing sequence of upper bounds for the dimension, showing that $\dim_H(\mathcal{G}) \leq 1.8203$.
- In section 4.6, taking a break from upper bounds, we apply the methods of the preceding two sections to give lower bounds for the limit inferior of the above sequence of upper bounds.
- Finally, in section 4.7, we apply the methods of sections 3–5 to the original sequence, (Y_n), to obtain a sequence of upper bounds which are more effective still; in particular, proving Theorem 9.

4.2 Prerequisites and a preliminary result

4.2.1 Definition of the Rauzy gasket

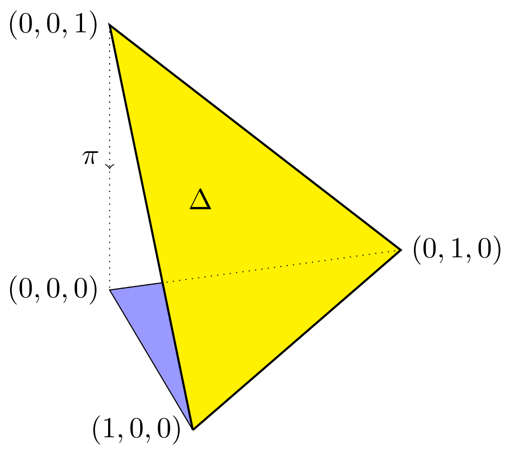


Figure 4.5: The two-simplex Δ and the projection $\pi : (x,y,z) \mapsto (x,y)$, used in section 4.3.

We begin in earnest by defining the Rauzy gasket $\mathcal{G}$ as the attractor of three maps on the standard two-simplex, Δ (depicted in Figure 4.5)

$$\Delta := \big\{(x,y,z) \in [0,1]^3 \,:\, x+y+z=1\big\},$$

as follows.

Definition 27 ($\mathcal{G}$)**.** The Rauzy gasket $\mathcal{G}$ is the limit set of $(\phi_i)_{i=1}^3$:

$$\phi_i : \Delta \to \Delta, \qquad \phi_i(x) = \frac{N_i \cdot x}{\|N_i \cdot x\|}$$

(i.e., $\mathcal{G} \subset \Delta$ is the largest set [by inclusion] such that $\mathcal{G} = \phi_1(\mathcal{G}) \cup \phi_2(\mathcal{G}) \cup \phi_3(\mathcal{G})$, where $\|x\| := (1,1,1) \cdot x$ denotes the l^1-norm of x, and the N_i are as follows:

$$N_1 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; N_2 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \; N_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix},$$

which we consider to act by left multiplication: i.e.,

$$N_1 \cdot (x,y,z) = (x+y+z, y, z), \qquad N_2 \cdot (x,y,z) = (x, x+y+z, z),$$
$$N_3 \cdot (x,y,z) = (x, y, x+y+z).$$

Remark 25. An explicit formula (see, e.g., [42]) is the following:

$$\mathcal{G} = \bigcap_{n \in \mathbb{N}} \bigcup_{i \in \{1,2,3\}^n} \phi_{i_1} \circ \phi_{i_2} \circ \cdots \circ \phi_{i_n}(\Delta). \tag{4.1}$$

Note that, in view of this, we will informally call the elements of

$$\{\phi_{i_1} \circ \phi_{i_2} \circ \cdots \circ \phi_{i_n}(\Delta) \mid i \in \{1,2,3\}^n\}$$

level n triangles for each $n \in \mathbb{N}$.

Definitions of Hausdorff and box-counting dimension

First, recall that the diameter of a subset A of a metric space (M,d) is given by

$$\operatorname{diam}(A) := \sup\{d(x,y) \mid x, y \in A\},$$

and that a collection $\mathcal{U}$ comprising open subsets of M is an open cover of A if

$$A \subset \bigcup_{U \in \mathcal{U}} U.$$

This leads to the following standard definition for the Hausdorff dimension.

Definition 28 ($\dim_H$)**.** Given a closed subset A of a metric space M, its *Hausdorff dimension* is given by

$$\dim_H(A) = \inf\{t > 0 \mid \mathcal{H}_t(A) = 0\},$$

where

$$\mathcal{H}_t(A) = \lim_{\varepsilon \searrow 0} \inf \left\{ \sum_{U \in \mathcal{U}} \operatorname{diam}(U)^t \,\middle|\, \mathcal{U} \text{ is an open cover of } A,\ \sup_{U \in \mathcal{U}} \big(\operatorname{diam}(U)\big) < \varepsilon \right\}.$$

Since the estimates of [33] are in terms of the box-counting dimension, we define it for completeness.

Definition 29 ($\dim_B$)**.** Given a closed subset A of a metric space M, its (upper) *box-counting dimension* (also known as Minkowski dimension) is given by

$$\dim_B(A) = \limsup_{\varepsilon \to 0} \left(\frac{-\log\big(N(A,\varepsilon)\big)}{\log(\varepsilon)} \right),$$

where $N(A,\varepsilon)$ is defined by

$$N(A,\varepsilon) := \min\{|\mathcal{U}| : \mathcal{U} \text{ is an open cover of } A \text{ by balls of radius } \varepsilon\},$$

where $|\mathcal{U}|$ denotes the cardinality of $\mathcal{U}$.

It is well-known that $\dim_H(A) \le \dim_B(A)$ for any A, which can be deduced from the definitions. To show that the difference can be quite pronounced, consider the following example: if $\alpha > 1$, then

$$A = \{0\} \cup \{n^{-\alpha}\}_{n=1}^{\infty}$$

satisfies $\dim_H(A) = 0$ and $\dim_B(A) = \alpha^{-1}$. (These equalities can be deduced simply from the definitions.)

For other definitions of dimension (e.g., topological dimension and Assuad dimension, satisfying similar inequalities), and proofs of the above statements, see [42].

4.2.2 Upper bounds on the Hausdorff dimension via a sequence of covers

Throughout this chapter, we adopt an analogue of the convenient notation for the indices of compositions and products used in Chapter 2.

Definition 30. Given $i = (i_1, i_2, \ldots, i_n) \in \{1,2,3\}^n$, we denote

- $N_i = N_{i_1} N_{i_2} \cdots N_{i_n}$.
- $\phi_i = \phi_{i_1} \circ \phi_{i_2} \circ \cdots \circ \phi_{i_n} : x \mapsto \dfrac{N_i \cdot x}{\|N_i \cdot x\|}$,

- $\Delta_i := \phi_i(\Delta) = \phi_{i_1} \circ \phi_{i_2} \circ \cdots \circ \phi_{i_n}(\Delta)$.

As mentioned in the introduction, our upper bounds for $\dim_H(\mathcal{G})$ are all based on the following simple lemma.

Lemma 25. *Let $\delta > 0$. Assume the sequence*

$$Y_n = \sum_{i \in \{1,2,3\}^n} \text{area}\left(\Delta_i\right)^{\delta} \text{diam}\left(\Delta_i\right)^{1-\delta} \to 0 \tag{4.2}$$

as $n \to \infty$. Then $\dim_H(\mathcal{G}) \leq 1 + \delta$.

Proof of Lemma 25. This is based on some simple bounds. Fix $\delta > 0$ satisfying (4.2). Referring to the definition of $\dim_H$ above, it suffices to provide a suitable sequence of open covers $\mathcal{U}_n$ of $\mathcal{G}$ such that, as $n \to \infty$,

$$\sup_{U \in \mathcal{U}_n} \{\text{diam}(U)\} \to 0 \qquad \text{and} \qquad \sum_{U \in \mathcal{U}_n} \text{diam}(U)^{1+\delta} \to 0$$

(note that the second implies the first). Here, we exploit the following fact, viewing a triangle in $\mathbb{R}^3$ as the convex hull of its three vertices.

Fact 3. For each i, the image $\Delta_i = \phi_i(\Delta)$ is a triangle with vertices

$$\phi_i(1,0,0),\ \phi_i(0,1,0) \text{ and } \phi_i(0,0,1).$$

Proof of Fact 3. The fact follows inductively from the assertion that each ϕ_j ($j = 1$, 2 or 3) maps triangles onto triangles (and vertices onto vertices). For this purpose, considering the definition of ϕ_j, it is sufficient to consider how the projection map $v \mapsto v/\|v\|$, with $\|v\| = (1,1,1) \cdot v$, acts on triangles, since the linear action of N_j obviously maps triangles in Δ onto triangles in $[0,\infty)^3 \setminus \{0\}^3$.

Consider the projection of a linear parametrisation: for $w, v \in (0,\infty)^3 \setminus \{0\}^3$ and $t \in [0,1]$, one directly computes that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{tw + (1-t)v}{\|tw + (1-t)v\|}\right) = \frac{\tilde{v}}{\|tw + (1-t)v\|^2},$$

where $\tilde{v} \in \mathbb{R}^3$ is constant and non-zero.

As the derivative is a multiple of a fixed vector, the projected parametrisation prescribes a line segment between $\phi_j(w)$ and $\phi_j(v)$; moreover, because the multiple is strictly positive for all t, the parametrisation is bijective (i.e., monotonic). In other words, for any $t \in [0,1]$, there is an $s \in [0,1]$ (with s strictly increasing in t) such that

$$\frac{tw + (1-t)v}{\|tw + (1-t)v\|} = s\frac{w}{\|w\|} + (1-s)\frac{v}{\|v\|}.$$

I.e., if x is a convex combination of w, v in $\mathbb{R}^3$, then $x/\|x\|$ is a convex combination of $w/\|w\|$ and $v/\|v\|$. An analogous statement follows for convex combinations of 3 (or more) points in $\mathbb{R}^3 \setminus \{0\}^3$, which precisely shows that the projection, and hence ϕ_j, maps triangles onto triangles (and vertices onto vertices), as required. □

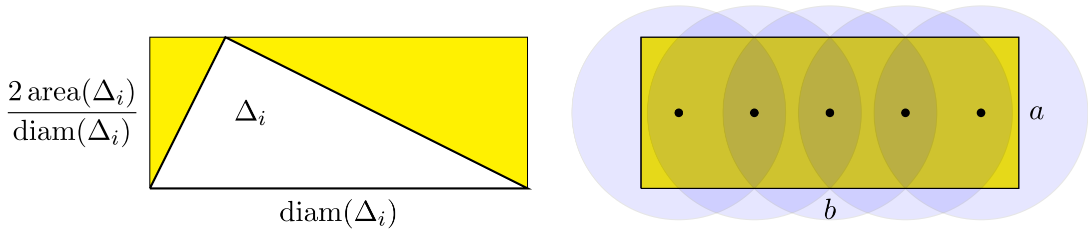


Figure 4.6: Left: covering Δ_i by a rectangle. Right: covering a rectangle by open disks.

Returning to the proof of the lemma: Given $i \in \{1,2,3\}^n$, since Δ_i is triangular, it is wholly contained in a rectangle as depicted in Figure 4.6. Moreover, any rectangle with base b and height $a \le b$ can be covered naively by $\lfloor 2b/a \rfloor$ disks of diameter $2a$, with centres spaced evenly at distance $a/2$ along the line at height $a/2$ and parallel to the b-length sides, as depicted in Figure 4.6. Combining these two facts, with $a = 2\,\mathrm{area}(\Delta_i)/\,\mathrm{diam}(\Delta_i)$ and $b = \mathrm{diam}(\Delta_i)$, shows that Δ_i can be covered by (the floor of)

$$\frac{2b}{a} = \frac{\mathrm{diam}(\Delta_i)^2}{\mathrm{area}(\Delta_i)}$$

open disks of diameter

$$2a = \frac{4\,\mathrm{area}(\Delta_i)}{\mathrm{diam}(\Delta_i)}.$$

Taking the union of the disks for each $i \in \{1,2,3\}^n$ to get a collection $\mathcal{U}_n$, we see that $\mathcal{U}_n$ is an open cover of the union

$$\bigcup_{i \in \{1,2,3\}^n} \Delta_i,$$

which contains $\mathcal{G}$ by (4.1). Moreover, $\mathcal{U}_n$ satisfies, for any $\delta > 0$,

$$\begin{aligned}\sum_{u\in\mathcal{U}_n} \operatorname{diam}(U)^{1+\delta} &\leq 4^{1+\delta} \sum_{i\in\{1,2,3\}^n} \frac{\operatorname{diam}(\Delta_i)^2}{\operatorname{area}(\Delta_i)} \left(\frac{\operatorname{area}(\Delta_i)}{\operatorname{diam}(\Delta_i)}\right)^{1+\delta} \\ &= 4^{1+\delta} \sum_{i\in\{1,2,3\}^n} \operatorname{area}(\Delta_i)^{\delta} \operatorname{diam}(\Delta_i)^{1-\delta}. \\ &=: 4^{1+\delta}\, Y_n\end{aligned}$$

which, by assumption, converges to zero as $n \to \infty$, as required. □

4.3 Proof that $\dim_H(\mathcal{G}) < 2$: an explicit form for X_n

We have just related bounds on the dimension to the convergence of (Y_n). For the time being, we restrict our attention to a simpler sequence, (X_n), where X_n, presented below, is obtained by dropping the $\operatorname{diam}(\Delta_i)^{1-\delta}$ factor in the summands of Y_n.

Definition 31 $(X_n, \hat{\delta})$**.** Let

$$X_n = X_n(\delta) := \sum_{i\in\{1,2,3\}^n} \operatorname{area}\left(\Delta_i\right)^{\delta}$$

and let

$$\hat{\delta} := \inf\left\{\delta \in [0,1] \mid \left(X_n\right)_{n=1}^{\infty} \to 0\right\},$$

Since $\operatorname{diam}(\Delta_i) \leq \operatorname{diam}(\Delta) = \sqrt{2}$ for each i, we have that $X_n \geq \sqrt{2}^{1-\delta} Y_n$ if $\delta \leq 1$. In particular, if $\delta \leq 1$ is such that $(X_n) \to 0$, then $\dim_H(\mathcal{G}) \leq 1+\delta$. Thus, assuming that $\hat{\delta} < 1$ (as we shall prove below), $\dim_H(\mathcal{G}) \leq 1 + \hat{\delta}$.

Why should one consider this simplified sequence (X_n)? Principally, it simplifies the following proof that $\dim_H(\mathcal{G}) < 2$, and also allows us to build up our methodology in a gradual fashion.

Furthermore, (X_n) is arguably an interesting object in its own right, and since it is easier to be precise about the areas of the Δ_i (indeed, we present a simple formula in terms of the N_i in the next subsection) than the diameters, the value of $\hat{\delta}$ can be estimated with much more accuracy.

In fact, estimates on $\hat{\delta}$ alone already offer a mild improvement on the best upper bound for $\dim_H(\mathcal{G})$ known previously (1.825, by Fougeron). This is seen by the following result, which is the culmination of sections 5 and 6.

Theorem 10. $0.8095 \leq \hat{\delta} \leq 0.8203$. *Consequently,* $\dim_H(\mathcal{G}) \leq 1.8203$.

In these sections we in fact obtain a sequence of upper and lower bounds for $\hat{\delta}$ ($\delta^{(m)}$ and δ_m), which appear to be monotonic towards $\hat{\delta}$ (see Figure 4.14 on page 168). Assuming these sequences converge to $\hat{\delta}$, sequence acceleration (i.e., numerically estimating the limit) gives a heuristic estimate of $\hat{\delta} \approx 0.8135$ (i.e., $\dim_H(\mathcal{G}) \leq \approx 1.8135$).

The discrepancy between the values in Theorem 10 and the estimated value of 1.72 from [33] perhaps unsurprisingly shows the importance of the diameter factors that we have eschewed. Indeed, the sequence of bounds we give in section 4.7, incorporating a naive estimate of these factors, quickly surpass the values above.

Our first step in considering (X_n) is to give it the following explicit form.

Lemma 26. *For all $n \in \mathbb{N}$, up to a constant factor,*

$$X_n = \sum_{i \in \{1,2,3\}^n} \|N_i \cdot e_1\|^{-\delta} \|N_i \cdot e_2\|^{-\delta} \|N_i \cdot e_3\|^{-\delta},$$

where, for $k = 1, 2, 3$, e_k denotes the kth standard basis vector of $\mathbb{R}^3$, i.e., $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, etc., and $\| \cdot \|$ again denotes the l^1 norm.

Remark 26. Note that each e_k is a vertex of Δ, and the unique fixed point of ϕ_k. Also, note that the coordinate permutations, interchanging the e_k, are symmetries of Δ.

Proof of Lemma 26. For a given triangle $A \subset \Delta$, let $\text{area}'(A) = \text{area}(\pi(A))$, where π is the simple projection map

$$\pi : \Delta \to \Delta' = \{(x, y) \in [0, 1]^2 \mid x + y \leq 1\}, \qquad \pi(x, y, z) = (x, y),$$

depicted in Figure 4.5 on page 124. Since π is linear and bijective, there exists $C > 0$ such that

$$\text{area}(A) = C\, \text{area}'(A)$$

for all $A \subset \Delta$. In fact, putting $A = \Delta$ shows that $C = \sqrt{3}$. To consider $\text{area}'(\Delta_i)$ for $i \in \{1, 2, 3\}^n$, integrating by substitution we have (in the abridged notation introduced as the beginning of this section)

$$\text{area}'(\Delta_i) = \int_{\Delta'} \text{Jac}_{T_i}(x, y)\, \mathrm{d}(x, y), \tag{4.3}$$

where $T_i := \pi \circ \phi_i \circ \pi^{-1}$ and

$$\text{Jac}_{T_i}(x, y) := \left| \det D_{(x,y)} T_i \right|,$$

where $D_{(x,y)}T_i$ is the Fréchet derivative of T_i at (x, y). More explicitly, for $i = 1, 2, 3$, the T_i are as in chapter 1:

$$T_1(x,y) = \left(\frac{1}{2-x}, \frac{y}{2-x}\right), \qquad T_2(x,y) = \left(\frac{y}{2-y}, \frac{1}{2-y}\right),$$
$$T_3(x,y) = \left(\frac{x}{1+x+y}, \frac{1}{1+x+y}\right).$$

In regards to the Jacobeans,

$$\mathrm{Jac}_{T_1}(x,y) = \det\begin{pmatrix} \frac{1}{(2-x)^2} & 0 \\ \frac{y}{(2-x)^2} & \frac{1}{2-x} \end{pmatrix} = (2-x)^{-3},$$

similarly, $\mathrm{Jac}_{T_2}(x,y) = (2-y)^{-3}$, and

$$\mathrm{Jac}_{T_3}(x,y) = \det\begin{pmatrix} \frac{y+1}{(x+y+1)^2} & \frac{-x}{(x+y+1)^2} \\ \frac{-y}{(x+y+1)^2} & \frac{x+1}{(x+y+1)^2} \end{pmatrix} = (1+x+y)^{-3} = \big(2-(1-x-y)\big)^{-3}.$$

In other words, for all $v \in \Delta$, $\mathrm{Jac}_{T_k}\big(\pi(v)\big) = (2-v_k)^{-3} = \|N_k \cdot v\|^{-3}$. This last equality follows from the definition of N_k: for example, if $v = (x, y, z)$,

$$\|N_1 \cdot (x,y,z)\| = \|(x+y+z, y, z)\| = x + 2y + 2z = 2 - x.$$

The formula relating N_k to Jac_{T_k} in fact extends to arbitrary combinations of the T_k, by the chain rule. That is, for any $v \in \Delta$ and $i \in \{1,2,3\}^n$, we have the telescoping product

$$\begin{aligned}
\mathrm{Jac}_{T_i}\big(\pi(v)\big) &= \prod_{k=1}^{n} \mathrm{Jac}_{T_{i_k}}\big(T_{i_{k+1}} \cdots T_{i_n}(\pi(v))\big) \\
&= \prod_{k=1}^{n} \mathrm{Jac}_{T_{i_k}}\big(\pi\big(\phi_{i_{k+1}} \cdots \phi_{i_n}(v)\big)\big) \\
&= \prod_{k=1}^{n} \|N_{i_k} \cdot \big(\phi_{i_{k+1}} \cdots \phi_{i_n}(v)\big)\|^{-3} \\
&= \prod_{k=1}^{n} \left\| N_{i_k} \cdot \left(\frac{N_{i_{k+1}} \cdots N_{i_n} \cdot v}{\|N_{i_{k+1}} \cdots N_{i_n} \cdot v\|}\right)\right\|^{-3} \\
&= \prod_{k=1}^{n} \frac{\big\|N_{i_k} N_{i_{k+1}} \cdots N_{i_n} \cdot v\big\|^{-3}}{\big\|N_{i_{k+1}} \cdots N_{i_n} \cdot v\big\|^{-3}} \\
&= \|N_{i_1} N_{i_2} \cdots N_{i_n} \cdot v\|^{-3} \qquad = \|N_i \cdot v\|^{-3}.
\end{aligned}$$

More explicitly, writing $v = (x, y, z)$,

$$\begin{aligned}\mathrm{Jac}_{T_i}(x, y) &= \big((1,1,1)\cdot N_i \cdot (x,y,z)\big)^{-3} \\ &= \big(x(1,1,1)\cdot N_i \cdot e_2 + y(1,1,1)\cdot N_i\cdot e_2 + z(1,1,1)\cdot N_i \cdot e_k)\big)^{-3} \\ &= (\lambda_1 x + \lambda_2 y + \lambda_3 z)^{-3},\end{aligned}$$

where $\lambda_k = \lambda_k(i)$ is the sum over the kth column of N_i:

$$\lambda_k = (1,1,1)\cdot N_i \cdot e_k = \|N_i \cdot e_k\|.$$

Inserting this into (4.3) yields a simple formula for $\mathrm{area}'(\Delta_i)$:

$$\begin{aligned}\mathrm{area}'\left(\Delta_i\right) &= \int_{\Delta'} \mathrm{Jac}_{T_i}(x, y, 1-x-z)\,\mathrm{d}(x,y) \\ &= \int_0^1 \int_0^{1-y} \big(\lambda_1 x + \lambda_2 y + \lambda_3(1-x-y)\big)^{-3} \mathrm{d}x\,\mathrm{d}y \\ &= \frac{1}{2}\frac{1}{\lambda_1\lambda_2\lambda_3}.\end{aligned}$$

Raising to the power δ and summing over $i \in \{1,2,3\}^n$ with the above values for λ_k then gives the required expression for X_n, up to a constant factor:

$$X_n = \sum_{i\in\{1,2,3\}^n} \mathrm{area}(\Delta_i)^\delta = \left(\frac{\sqrt{3}}{2}\right)^\delta \sum_{i\in\{1,2,3\}^n} \prod_{k=1}^{3} \|N_i \cdot e_k\|^{-\delta}.$$

□

Definition 32. For economy of space and simplicity, we will often write $F(i)$ for the ith summand of X_n:

$$F(i) := \|N_i \cdot e_1\|^{-\delta}\|N_i \cdot e_2\|^{-\delta}\|N_i \cdot e_3\|^{-\delta}.$$

4.4 Proof that $\dim_H(\mathcal{G}) < 2$: a renewal method to show $(X_n) \to 0$

Continuing with the simple proof that $\dim_H(\mathcal{G}) < 2$, we now provide a (partial) decomposition of X_n as follows. The following definition is combinatorial, but its dynamical relevance will soon be made clear.

Definition 33 ($A_{n,k}$, $X_{n,k}$)**.** Given $n, k \in \mathbb{N}$ such that $1 \le k < n$, let $A_{n,k}$ denote the elements of $\{1,2,3\}^n$, for which the first entry is repeated exactly k times: i.e.,

$$A_{n,k} := \{i \in \{1,2,3\}^n \,|\, i_1 = i_2 = \cdots = i_k \neq i_{k+1}\}\,.$$

Accordingly, let

$$X_{n,k} := \sum_{i \in A_{n,k}} F(i) = \sum_{i \in A_{n,k}} \|N_i \cdot e_1\|^{-\delta} \|N_i \cdot e_2\|^{-\delta} \|N_i \cdot e_3\|^{-\delta}.$$

It should be clear from the definition of the $A_{n,k}$ that we have the following disjoint union, for each $n \geq 2$:

$$\{1,2,3\}^n = \bigcup_{k=1}^{n-1} A_{n,k} \cup \{1\}^n \cup \{2\}^n \cup \{3\}^n, \tag{4.4}$$

i.e., the decomposition of X_n into the $X_{n,k}$ leaves out three terms. Since they play a tangential role, it is convenient to define the following terminology.

Definition 34 (Constant)**.** For each $n \in \mathbb{N}$, call the three sequences in $\{1\}^n \cup \{2\}^n \cup \{3\}^n$ (which don't lie in any $A_{n,k}$) *constant.*

The following simple but important lemma relates the symbolic partition in (4.4) to a geometric partition of Δ.

Lemma 27. *For all $n \geq 2$ and $n < k$, given $i \in \{1,2,3\}^n$, the following are equivalent:*

- $i \in A_{n,k}$.
- $\Delta_i \subset R_k$.
- $\phi_i(e_k) \in R_k$ *for each* $k \in \mathbb{N}$.

where, for $k \in \mathbb{N}$,

$$R_k := \mathrm{cl}\left(\bigcup_{j=1}^{3} \phi_j^k(\Delta) \setminus \phi_j^{k+1}(\Delta)\right),$$

where cl *denotes the topological closure. This gives the partition of Δ depicted in Figure 4.7, which partitions each Δ_j into "strips".*

Remark 27. The fact that no analogue of the above lemma holds for constant sequences (more explicitly, for each j, $\phi_j^n(\Delta)$ meets every R_k with $k \geq n$) is the precise reason for not defining $A_{n,n}$, to maintain a distinction.

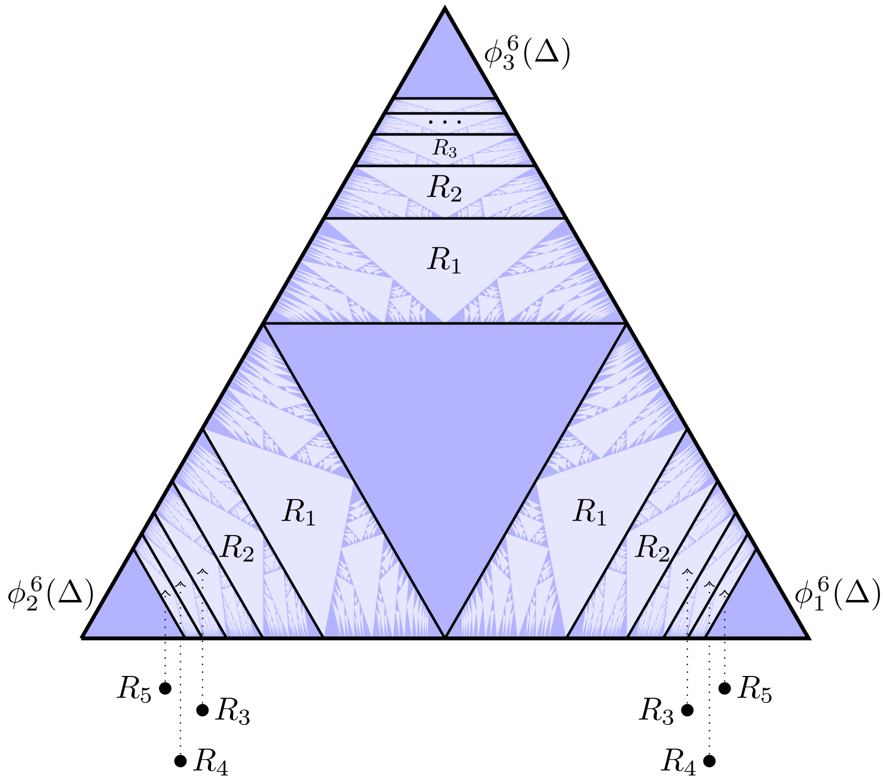


Figure 4.7: The sets R_k, for $k = 1, \ldots, 5$.

Proof of Lemma 27. The equivalence of the second and third bullet points is evident from the fact that Δ is the convex hull of $\{e_k\}_{k=1}^3$.

Consider the equivalence of the first and second bullet points. If $i \in \{1, 2, 3\}^n$ is not constant, then $i \in A_{n,k}$, where

$$i_1 = i_2 = \cdots = i_k \neq i_{k+1}.$$

In particular,

$$\Delta_i = \phi_{i_1}^k \phi_{i_{k+1}}(\Delta_{(i_{k+2},\cdots,i_n)}) \subset \phi_{i_1}^k(\Delta_{i_{k+1}}).$$

Since Δ_1, Δ_2 and Δ_3 have pairwise disjoint interiors, $\phi_{i_1}^{k+1}(\Delta) = \phi_{i_1}^k(\Delta_{i_1})$ and $\phi_{i_1}^k(\Delta_{i_{k+1}})$ also have disjoint interiors, and thus

$$\phi_{i_1}^k(\Delta_{i_{k+1}}) \subset \mathrm{cl}\left(\phi_{i_1}^k(\Delta) \setminus \phi_{i_1}^k(\Delta_{i_1})\right) \subset R_k.$$

Now consider the converse implication. Supposing that $i \in \{1, 2, 3\}^n$ is not constant and does not lie in $A_{n,k}$, then by (4.4) it lies in $A_{n,k'}$ for some $k' \neq k$. Thus, the above proof shows that $\Delta_i \subset R_{k'}$, and hence $\Delta_i \not\subset R_k$ (since Δ_i has positive area and R_k, $R_{k'}$ can only meet at their boundaries). $\square$

Using the above partition, we bound $\{X_{n+1,k}\}_{k \leq n}$ in terms of a linear combination on $\{X_{n,k}\}_{k<n}$, as per the following lemma. To express its result in matrix form, extending $X_{n,k} = 0$ for $k \geq n$ to make an infinite vector, we have, for the functions a_k, b_k, and c_n of δ defined below,

$$\begin{pmatrix} X_{n+1,1} \\ X_{n+1,2} \\ X_{n+2,3} \\ X_{n+1,4} \\ X_{n+1,5} \\ \vdots \end{pmatrix} \leq \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & \cdots \\ b_1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & b_2 & 0 & 0 & 0 & \cdots \\ 0 & 0 & b_3 & 0 & 0 & \cdots \\ 0 & 0 & 0 & b_4 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} X_{n,1} \\ X_{n,2} \\ X_{n,3} \\ X_{n,4} \\ X_{n,5} \\ \vdots \end{pmatrix} + \begin{pmatrix} c_n \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix} \tag{4.5}$$

for any $n \geq 2$, where

$$\begin{pmatrix} x_1 \\ \vdots \\ x_k \\ \vdots \end{pmatrix} \leq \begin{pmatrix} y_1 \\ \vdots \\ y_k \\ \vdots \end{pmatrix}$$

here means that $x_k \leq y_k$ for each $k \geq 1$.

Remark 28. For visual intuition: if the c_n were all equal to zero, this in particular would bound $X_{n,k}$ in terms of the sum, over all edge paths from node 1 to node k of length $n-1$ in the graph depicted in Figure 4.8, of the product of the edge weights along that path. Practically, the c_n will have a negligible effect, so this is a reasonable consideration. In particular, the fact that all closed loops go through the node 1 is a precursor to the renewal equation that will be a natural consequence of the lemma.

We now state the anticipated lemma.

Lemma 28. *For each $k, n \in \mathbb{N}$, we define the following maxima on $R_k \cap \Delta_1 = \phi_1^k(\Delta) \setminus \phi_1^{k+1}(\Delta)$:*

$$\begin{aligned} a_k &:= \max_{v \in R_k \cap \Delta_1} \|N_2 \cdot v\|^{-3\delta} + \|N_3 \cdot v\|^{-3\delta} \\ &= \max_{v \in R_k \cap \Delta_1} (2 - v_2)^{-3\delta} + (2 - v_3)^{-3\delta}, \end{aligned}$$

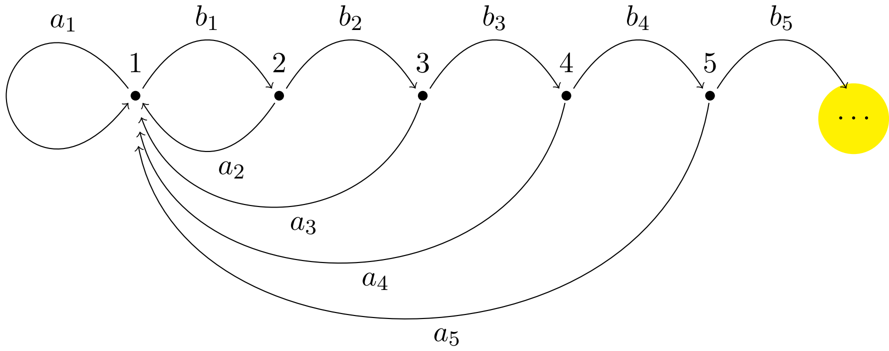


Figure 4.8: The edge-weighted graph corresponding to (4.5). Here, the edge from r to s indicates that there is a contribution from $X_{n,r}$ in the bound of $X_{n+1,s}$, and the edge labels show the proportions.

$$\begin{aligned} b_k &:= \max_{v \in R_k \cap \Delta_1} \|N_1 \cdot v\|^{-3\delta} \\ &= \max_{v \in R_k \cap \Delta_1} (2 - v_1)^{-3\delta}, \end{aligned}$$

and

$$c_n := 6 \cdot 4^{-\delta}(n+1)^{-\delta}(2n+1)^{-\delta}.$$

then, For all $n \geq 2$ and $k < n$, we have the following inequalities:

$$X_{n+1,1} \leq \sum_{j=1}^{n-1} a_j X_{n,j} + c_n, \tag{4.6}$$

$$X_{n+1,k+1} \leq b_k X_{n,k}. \tag{4.7}$$

Remark 29. The definitions of a_k and b_k (and our subsequent proofs) exploit the symmetries between the different N_j, i.e., that these definitions are still valid if one permutes the subscripts 1, 2 and 3 therein.

Remark 30. In our proofs, we will only need to know that $c_n > 0$ for each n and that $\sum_{n=1}^{\infty} c_n < \infty$ if and only if $\delta > \frac{1}{2}$, which are both clear from the definition. Its exact value can be forgotten.

Before proving the lemma, we define some useful terminology.

Definition 35 (Successor)**.** We say, for $j \in \{1,2,3\}$, that $(j;i) := (j, i_1 \ldots, i_n) \in \{1,2,3\}^{n+1}$ is a *successor* of $i = (i_1, \ldots, i_n)$, and more precisely, the *j-successor* of i. Moreover, if $j = i_1$, we say it is the *principal successor.*

Proof of Lemma 28. We first note, from the definition of $A_{n,k}$, that if $i \in A_{n,k}$ for some k,

- the principal successor of i lies in $A_{n+1,k+1}$, and
- the other two successors lie in $A_{n+1,1}$.

Similarly, if $i \in \{1\}^n \cup \{2\}^n \cup \{3\}^n$ is constant, its principal successor is also constant, and its other two successors lie in $A_{n+1,1}$.

We now prove (4.7). From the above bullet points, $i' \in A_{n+1,k+1}$ if and only if i' is the principal successor of a unique $i \in A_{n,k}$. That is, for each $i' \in A_{n+1,k+1}$, there exists a unique $i \in A_{n,k}$ such that $i' = (i_1; i) = (i_1, i_1, \ldots, i_n)$. Consequently,

$$X_{n+1,k+1} = \sum_{i' \in A_{n+1,k+1}} F(i') = \sum_{i \in A_{n,k}} F(i_1; i). \tag{4.8}$$

Applying Lemma 27, we now bound the ratio $F(i_1; i)/F(i)$:

$$\begin{aligned} \frac{F(i_1;i)}{F(i)} &= \prod_{j=1}^{3} \frac{\|N_{i_1} N_i \cdot e_j\|^{-\delta}}{\|N_i \cdot e_j\|^{-\delta}} \\ &= \prod_{j=1}^{3} \|N_{i_1} \cdot \frac{N_i \cdot e_j}{\|N_i \cdot e_j\|}\|^{-\delta} \\ &= \prod_{j=1}^{3} \|N_{i_1} \cdot \phi_i(e_j)\|^{-\delta}. \\ &\le \prod_{j=1}^{3} \max_{v \in R_k \cap \Delta_{i_1}} \|N_{i_1} \cdot v\|^{-\delta} \\ &= \prod_{j=1}^{3} \max_{v \in R_k \cap \Delta_1} \|N_1 \cdot v\|^{-\delta} \\ &= \max_{v \in R_k \cap \Delta_1} \|N_1 \cdot v\|^{-3\delta} \\ &=: b_k, \end{aligned} \tag{4.9}$$

where the last two equalities use symmetry of the N_j under coordinate permutations, and the fact that $v \mapsto \|N_1 \cdot v\|$ is linear, respectively. Applying this estimate to (4.8) gives the required inequality:

$$X_{n+1,k+1} \le \sum_{i \in A_{n,k}} b_k \, F(i) = b_k X_{n,k}.$$

The proof of (4.6) is similar but slightly more nuanced. From our first consideration, we see that $A_{n+1,1}$ comprises all non-principal successors of elements in $\{1,2,3\}^n$. That is, we can write

$$X_{n+1,1} = \sum_{i\in\{1,2,3\}^n} \sum_{\substack{1\le\omega\le 3:\\ \omega\ne i_1}} F(\omega;i) \tag{4.10}$$

Recalling (4.4), i.e.,

$$\{1,2,3\}^n = \bigcup_{k=1}^{n-1} A_{n,k} \cup \{1\}^n \cup \{2\}^n \cup \{3\}^n,$$

we bound $\sum_{\omega\ne i} F(\omega;i)/F(i)$ in two cases:

Case 1: $i \in A_{n,k}$ *for some* $k \in \{1,2,\dots,n-1\}$. In this case, we have a similar bound to (4.9), but this time first using the AM-GM inequality:

$$\begin{aligned}
\frac{\sum_{\omega\ne i_1} F(\omega;i)}{F(i)} &= \sum_{\omega\ne i_1}\prod_{j=1}^{3} \frac{\|N_\omega N_i\cdot e_j\|^{-\delta}}{\|N_i\cdot e_j\|^{-\delta}} \\
&= \sum_{\omega\ne i_1}\prod_{j=1}^{3} \|N_\omega\cdot\phi_i(e_j)\|^{-\delta} \\
&\le \frac{1}{3}\sum_{\omega\ne i_1}\sum_{j=1}^{3} \|N_\omega\cdot\phi_i(e_j)\|^{-3\delta} \\
&= \frac{1}{3}\sum_{j=1}^{3}\sum_{\omega\ne i_1} \|N_\omega\cdot\phi_i(e_j)\|^{-3\delta} \\
&\le \max_{v\in R_k\cap\Delta_{i_1}} \sum_{\omega\ne i_1} \|N_\omega\cdot v\|^{-3\delta} \\
&= \max_{v\in R_k\cap\Delta_1} \|N_2\cdot v\|^{-3\delta} + \|N_2\cdot v\|^{-3\delta} \\
&=: a_k,
\end{aligned} \tag{4.11}$$

where we are again using symmetry of the N_k under permutation. Summing over the $i \in A_{n,k}$ hence gives the kth term of the sum in (4.6):

$$\sum_{i\in A_{n,k}}\sum_{\omega\ne i_1} F(\omega;i) \le \sum_{i\in A_{n,k}} a_k F(i) = a_k X_{n,k}.$$

Case 2: $i \in \{1\}^n \cup \{2\}^n \cup \{3\}^n$ *is constant.* By symmetry, the six terms of this case are all equal, so it suffices to calculate just one. Considering $(1,2,\dots,2)$ for example, one has

$$(1,1,1)N_1N_2^n \cdot e_j = (1,1,1)\begin{pmatrix} n+1 & 1 & n+1 \\ n & 1 & n \\ 0 & 0 & 1 \end{pmatrix} \cdot e_j = (2n+1, 2, 2n+2) \cdot e_j$$

and therefore $F(1,2,2,\ldots,2) = 4^{-\delta}(n+1)^{-\delta}(2n+1)^{-\delta} = c_n/6$. Summing hence gives c_n in (4.6), completing the formula. $\square$

Remark 31. We note that, if one were to apply this program to the Sierpiński gasket, say, one would obtain an analogue of Lemma 28 with the values $a_k = b_k = \frac{1}{3}$ and $c_n = 2 \cdot 3^{-n}$. In this way, in view of the next lemma, the values of a_k and b_k below distinguish the Rauzy gasket from other attractors on three maps (at least in the context presented here). It would be interesting to compute the corresponding values for the Apollonian gasket.

The next lemma provides values for the a_k, b_k defined in the previous lemma, as well as corresponding minima for section 4.6, where we estimate $\hat{\delta}$ from below.

Lemma 29. *For all $v \in \mathcal{P}_k := R_k \cap \Delta_1$, we have the following tight bounds.*

$$\left(\frac{k+1}{k+2}\right)^{3\delta} \leq (2-v_1)^{-3\delta} \leq \left(\frac{k+2}{k+3}\right)^{3\delta}, \tag{4.12}$$

$$2^{3\delta+1}\left(\frac{k+2}{4k+7}\right)^{3\delta} \leq (2-v_2)^{-3\delta} + (2-v_3)^{-3\delta} \leq \left(\frac{k+1}{2k+1}\right)^{3\delta} + 2^{-3\delta}. \tag{4.13}$$

In particular, the maxima from Lemma 28 take the following values:

$$a_k = \left(\frac{k+1}{2k+1}\right)^{3\delta} + 2^{-3\delta}, \qquad b_k = \left(\frac{k+2}{k+3}\right)^{3\delta}.$$

For continuity, we defer the proof of this lemma to appendix B. The final ingredient in the proof is the following renewal theorem of Feller [45, p.330, Theorem 1], generalising the Erdős–Feller–Pollard renewal theorem (Lemma 4) from chapter 2.

Lemma 30 (Feller renewal theorem)**.** *Suppose we have non-negative sequences $(u_n)_{n=0}^{\infty}$, $(\lambda_k)_{n=1}^{\infty}$ and $(\nu_n)_{n=0}^{\infty}$ such that, for all $n \in \mathbb{N}$,*

$$u_n = \nu_n + \sum_{k=1}^{n} \lambda_k u_{n-k},$$

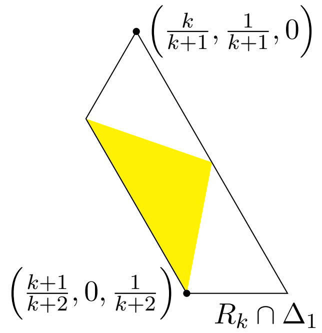


Figure 4.9: Two maximisers giving the values of a_k and b_k in Lemma 29.

and

$$\lambda = \sum_{k=1}^{\infty} \lambda_k < \infty, \qquad \nu = \sum_{n=1}^{\infty} \nu_n < \infty,$$

and that the smallest additive subgroup of $\mathbb{Z}$ *containing* $\{n \in \mathbb{N} \mid \lambda_n > 0\}$ *is* $\mathbb{Z}$*. Then we have the following cases:*

- *If* $\lambda < 1$*, then*

$$\sum_{n=0}^{\infty} u_n = \frac{\nu}{1-\lambda}.$$

  In particular, $(u_n) \to 0$*.*

- *If* $\lambda = 1$*, then*

$$(u_n) \to \frac{b}{\mu}^{-1},$$

  where $\mu = \sum_{n=1}^{\infty} n u_n$ *(if* $\mu = \infty$*, the limit equals zero).*

- *If* $\lambda > 1$*, then there exists a unique* $x \in (0,1)$ *such that*

$$\sum_{k=1}^{\infty} \lambda_k x^k = 1$$

  and $(x^n u_n)$ *converges to a positive limit. In particular,* $(u_n) \to \infty$*.*

This comprehensive renewal theorem in particular allows us to accommodate the remainder terms c_n (which will play the role of the ν_n, above), which is a necessary improvement on Lemma 4.

Applying the first case of the previous lemma, we are ready to give our first bound on the dimension, in particular giving an elementary proof that $\dim_H(\mathcal{G}) < 2$.

Proposition 17. *Let* $\delta^* = 0.893368\ldots$ *be the unique positive value of* δ *satisfying*

$$\sum_{k=1}^{\infty} a_k \prod_{i=1}^{k-1} b_i = 1,$$

i.e.,

$$3^{3\delta^*} \sum_{k=1}^{\infty} \left(\frac{k+1}{(k+2)(2k+1)} \right)^{3\delta^*} + \left(\frac{3}{2} \right)^{3\delta^*} \sum_{k=1}^{\infty} (k+2)^{-3\delta^*} = 1.$$

Then $\big(X_n(\delta)\big)_n \to 0$ *for all* $\delta > \delta^*$. *Consequently,* $\dim_H(\mathcal{G}) \leq 1 + \delta^* = 1.893368\ldots$.

The proof uses the previous three lemmas in a simple and direct way.

Proof of Proposition 17. We first show that $(X_n) \to 0$ if $(X_{n,1}) \to 0$. For any $i \in \{1,2,3\}^n$, the tuple $(1,2;i) := (1,2,i_1,i_2,\ldots,i_n)$ lies in $A_{n+2,1}$. Therefore,

$$X_{n+2,1} \geq \sum_{i \in \{1,2,3\}^n} F(1,2;i). \tag{4.14}$$

As in the proof of Lemma 28, we estimate the ratio $F(1,2;i)/F(i)$ from the definition:

$$\begin{aligned} \frac{F(1,2;i)}{F(i)} = \prod_{j=1}^{3} \frac{\|N_1 N_2 N_i \cdot e_j\|^{-\delta}}{\|N_i \cdot e_j\|^{-\delta}} &= \prod_{j=1}^{3} \left\| N_1 N_2 \cdot \frac{N_i(e_j)}{\|N_i(e_j)\|} \right\|^{-\delta} \\ &= \prod_{j=1}^{3} \|N_1 N_2 \cdot \phi_i(e_j)\|^{-\delta} \\ &\geq \min_{v \in \Delta} \|N_1 N_2 \cdot v\|^{-3\delta} \\ &= \min_{(x,y,z) \in \Delta} (2x + 3y + 4z)^{-3\delta} \\ &= 4^{-3\delta}, \end{aligned} \tag{4.15}$$

using that

$$N_1 N_2 \cdot (x,y,z) = \begin{pmatrix} 2 & 1 & 2 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \cdot (x,y,z) = (2x + y + 2z, x + y + z, z),$$

i.e., $\|N_1 N_2 \cdot (x,y,z)\| = 2x + 3y + 4z$, which takes its maximum on Δ when $(x,y,z) =$

$(0, 0, 1)$. Therefore, inserting (4.15) into (4.14) gives

$$X_{n+2,1} \geq 4^{-3\delta} \sum_{i \in \{1,2,3\}^n} F(i) = 4^{-3\delta} X_n,$$

and thus, $(X_n) \to 0$ if $(X_{n,1}) \to 0$, as claimed.

We next combine the two inequalities of Lemma 28, (4.6) and (4.7), to give a renewal-style inequality for $(X_{n+2,1})_{n\geq 0}$:

$$X_{n+2,1} \leq \sum_{k=1}^{n} a_k X_{n+1,k} + c_{n+1} \leq \sum_{k=1}^{n} a_k \prod_{i=1}^{k-1} b_i\, X_{n+2-k,1} + c_{n+1}.$$

Thereby, taking $(\hat{X}_n)_{n=0}^{\infty}$ such that $\hat{X}_0 := X_{2,1}$ and, for each $n \geq 1$,

$$\hat{X}_n = \sum_{k=1}^{n} a_k \prod_{i=1}^{k-1} b_i \hat{X}_{n-k} + c_{n+1},$$

we have that $\hat{X}_n \geq X_{n+2,1}$ for all $n \in \mathbb{N}$, and by the renewal theorem in Lemma 30, $(\hat{X}_n) \to 0$ if

$$\sum_{k=1}^{\infty} a_k \prod_{i=1}^{k-1} b_i < 1 \qquad \text{and} \qquad \sum_{n=2}^{\infty} c_n < \infty.$$

If $\delta > \delta^*$, then $\delta > \frac{1}{2}$, which implies the latter condition (as noted earlier). Also, since a_k, b_k are strictly decreasing functions in δ for all k (which follows either from the definitions or their explicit values given in Lemma 29), $\delta > \delta^*$ also implies the former condition.

Finally, the explicit formula defining δ^* is easy to show, using the convenient identity

$$\prod_{i=1}^{k-1} b_i = \prod_{i=1}^{k-1} \left(\frac{i+2}{i+3}\right)^{3\delta} = 3^{3\delta}(k+2)^{-3\delta}.$$

□

4.5 Refined upper bounds for $\hat{\delta}$: the sequence $(\delta^{(m)})$

We now give our first refinement of the proof of Proposition 17, or rather, sequence of refinements.

In brief: for each fixed $m \geq 2$, we consider a refinement of the decomposition $\{X_{n,k}\}$ of X_n, which corresponds to a refinement of the geometric partition $\{R_k\}_k$

above.[§] More explicitly,

- if $k \geq m$, we leave $X_{n,k}$ alone; and
- if $k < m$, we decompose $X_{n,k} = \sum_{v \in \mathcal{V}_k} X_{n,v}$, where $|\mathcal{V}_k| = 3^{m-k}$ is defined in subsection 4.5.2.

In particular, for $v \in \mathcal{V}_k$, we will have

$$X_{n,v} = \sum_{i \in A_{n,v}} F(i)$$

where $A_{n,v} \subset A_{n,k} \subset \{1,2,3\}^n$ satisfies, analogously to Lemma 27,

$$i \in A_{n,v} \iff \Delta_i \subset R_v,$$

where R_v is the union of six triangles in R_k, each the images of each other under the symmetries of Δ. (See the next subsection for an illustration for the case that $m = 2$).

An upper bound on the dimension emerges from this decomposition as follows: first, an analogue of Lemma 28 holds, i.e., one can bound each $X_{n+1,v}$ (for $v \in \bigcup_{k=1}^{m-1} \mathcal{V}_k \cup \{m, \ldots, n\}$) above by linear combinations of the $X_{n,v}$. (Again, see the next subsection for the case of $m = 2$.)

In particular, there will be a distinguished index, $\circledast \in \mathcal{V}_1$ (whose definition depends on m), such that $X_{n,\circledast}$ plays a similar role to $X_{n,1}$ from Lemma 28. This leads to a renewal-style inequality for $(X_{n,\circledast})_{n \geq m+1}$ which, by the renewal theorem, provides a threshold $\delta^{(m)}$, such that $X_{n,\circledast}$ (and hence X_n) converges to zero whenever $\delta > \delta^{(m)}$.

4.5.1 The example of $m = 2$

For clarity, we first briefly consider the case of $m = 2$. Here, for all $n \geq 3$, one retains $X_{n,k}$ for $k > 1$, decomposes $X_{n,1}$ into three parts,

$$X_{n,1} = X_{n,121} + X_{n,122} + X_{n,123},$$

and sets $\circledast = 122$. The corresponding regions in the geometric partition which decompose R_1 (R_{121}, R_{122} and R_{123}) are depicted in Figure 4.10. More specifically, for each $j = 1, 2, 3$, R_{12j} is the orbit of the level three triangle $\Delta_{(1,2,j)} = \phi_1 \phi_2 \phi_j(\Delta)$ under the symmetries of Δ.

[§] Heuristically, taking the regions smaller allows us to make more precise estimates, leading to

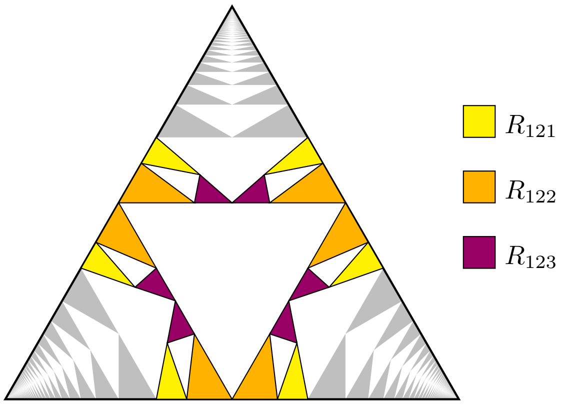


Figure 4.10: The refinement of R_1 into regions R_{121}, R_{122} and R_{123}, corresponding to $m = 2$.

Then, with some careful combinatorics (again considering successors), we can show that $X_{n,121}$, $X_{n,\circledast}$ and $X_{n,123}$ contribute to $X_{n+1,121}$, $X_{n+1,123}$ and $X_{n+1,2}$ in a linear fashion, and $\sum_{k\geq 2} X_{n,k}$ into $X_{n,\circledast}$. More precisely, bounding ratios of successive terms gives the following matrix inequality, for all $n \geq 3$:

$$\begin{pmatrix} X_{n+1,\circledast} \\ X_{n+1,121} \\ X_{n+1,123} \\ X_{n+1,2} \\ X_{n+1,3} \\ X_{n+1,4} \\ \vdots \end{pmatrix} \leq \begin{pmatrix} 0 & 0 & 0 & a_2 & a_3 & a_4 & \cdots \\ \left(\frac{6}{11}\right)^{3\delta} & \left(\frac{4}{7}\right)^{3\delta} & \left(\frac{7}{13}\right)^{3\delta} & 0 & 0 & 0 & \cdots \\ \left(\frac{2}{3}\right)^{3\delta} & \left(\frac{3}{5}\right)^{3\delta} & \left(\frac{5}{8}\right)^{3\delta} & 0 & 0 & 0 & \cdots \\ \left(\frac{5}{7}\right)^{3\delta} & \left(\frac{7}{10}\right)^{3\delta} & \left(\frac{3}{4}\right)^{3\delta} & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & b_2 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & b_3 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} X_{n,\circledast} \\ X_{n,121} \\ X_{n,123} \\ X_{n,2} \\ X_{n,3} \\ X_{n,4} \\ \vdots \end{pmatrix} + \begin{pmatrix} c_n \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix}. \tag{4.16}$$

Visually, the square matrix in (4.16) corresponds to the graph in Figure 4.11 (c.f. Figure 4.8 above). From this matrix, we define $B = B(2, \delta)$ as the 4×4 concatenation, omitting the a_2 entry:

$$B = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \left(\frac{6}{11}\right)^{3\delta} & \left(\frac{4}{7}\right)^{3\delta} & \left(\frac{7}{13}\right)^{3\delta} & 0 \\ \left(\frac{2}{3}\right)^{3\delta} & \left(\frac{3}{5}\right)^{3\delta} & \left(\frac{5}{8}\right)^{3\delta} & 0 \\ \left(\frac{5}{7}\right)^{3\delta} & \left(\frac{7}{10}\right)^{3\delta} & \left(\frac{3}{4}\right)^{3\delta} & 0 \end{pmatrix}.$$

B in particular corresponds to the subgraph on nodes $\{122, 123, 121, m\}$ with unla-

tighter bounds.

belled edges in Figure 4.11: i.e., the unlabelled edge from node y to node v should have label $B_{v,y}$ (for visual clarity, these labels have been omitted).

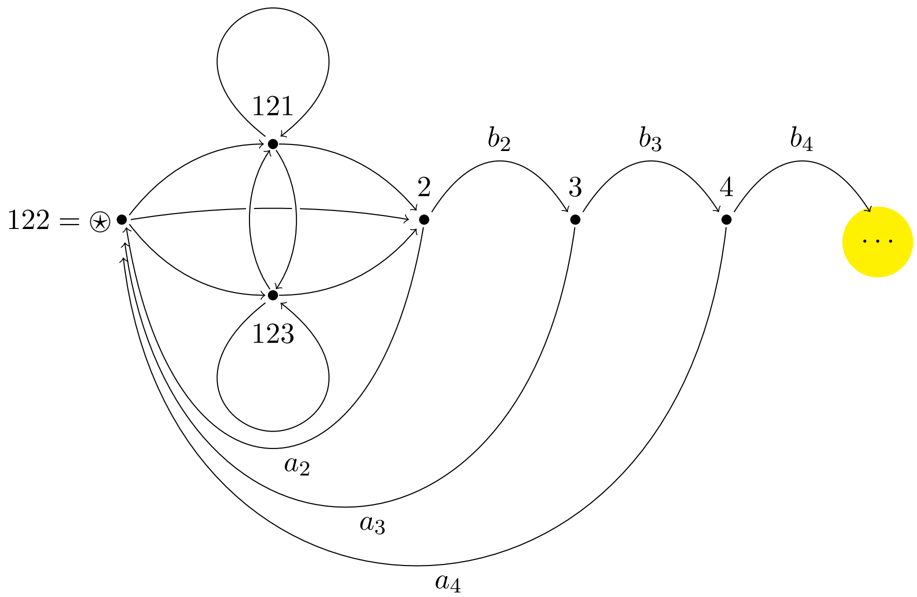


Figure 4.11: The graph for $m = 2$, corresponding to (4.16). The unlabelled edges correspond to the entries of B; see the discussion following (4.16).

(4.16) leads to a renewal-style inequality for $(X_{n+3,\circledast})_{n=1}^{\infty}$, with a more complicated remainder term, and coefficients in terms of B, a_k and b_k. That is, we obtain

$$X_{n+3,\circledast} \leq \sum_{k=1}^{n} \lambda_k X_{n+3-k,\circledast} + (\text{remainder term}),$$

where, taking the graph viewpoint, each λ_k is the sum over every loop (i.e., closed edge path, visiting the node $\circledast$ exactly once) of length k in the graph in Figure 4.11, of the product of the edge weights along that loop.

Correspondingly, the left hand side of the renewal condition $\sum_{k=1}^{\infty} \lambda_k < 1$, which we use to show that $(X_{n,\circledast}) \to 0$ (and hence $(X_n) \to 0$), corresponds to the sum over *all* loops in the graph (that visit $\circledast$ exactly once). This sum splits into a product of two, since each loop decomposes into two parts:

- A path (not necessarily a simple path) from $\circledast$ to node m in the unlabelled subgraph (the "finite-but-complicated part").
- A simple path from node m to $\circledast$ on the subgraph labelled by a's and b's (the "infinite-but-simple" part).

More explicitly, this leads to Theorem 11 with $m = 2$, which states that $\delta^{(2)} \geq \hat{\delta} \geq \dim_H(\mathcal{G}) - 1$, where $\delta^{(2)}$ is the value of δ which satisfies the equality

$$\sum_{k=1}^{n} \lambda_k = \sum_{k=0}^{\infty} \left(B^k\right)_{4,1} \sum_{k=2}^{\infty} a_k \prod_{i=2}^{k-1} b_i = 1, \tag{4.17}$$

where $\left(B^k\right)_{4,1}$ is the entry in B^k corresponding to m and $\circledast$, respectively.

Now consider this number, $\delta^{(2)}$. Since a_k, b_k and the entries of B are all non-negative, decreasing functions of δ, the left hand side of (4.17) is decreasing in δ, which shows there is at most one value $\delta^{(2)}$ satisfying (4.17). However, since the first sum, involving powers of B, is liable to diverge, a priori it requires a small leap of faith to assume that $\delta^{(2)}$ exists at all. To this end, let us sketch a general argument for the existence of $\delta^{(2)}$ (which generalises to show that $\delta^{(m)}$, δ_m and ε_m, defined below, exist for each $m \geq 2$).

First, B clearly converges to the zero matrix as $\delta \to \infty$. Thus, for $\rho(B)$ denoting the spectral radius of B, the infimum

$$\hat{x} = \inf\{x \in \mathbb{R} \mid \rho(B) < 1 \text{ for all } \delta > x\}$$

exists. That is, for all $\delta > \hat{x}$, $\rho(B) < 1$ and we have (applying Cramer's formula for inverses [79])

$$\sum_{k=0}^{\infty} (B^k)_{4,1} = ((I - B)^{-1})_{4,1} = -\frac{\det\left(\mathrm{Minor}_{1,4}(I - B)\right)}{\det(I - B)}, \tag{4.18}$$

where $\mathrm{Minor}_{1,4}(I - B)$ is the submatrix of $(I - B)$ obtained by removing the first row and fourth column from $I - B$ (i.e., those corresponding to zeros in B). (See (4.20) below for an explicit formula.)

Since the numerator and denominator of (4.18) are polynomial in the entries of B, this quotient extends to a meromorphic function of the plane, with poles located at (some of the) values of $\delta \in \mathbb{C}$ for which B has 1 as an eigenvalue. But then, since B is a continuous function of δ, $\rho(B)$ is also continuous in δ, implying that $\rho(B|_{\delta=\hat{x}}) = 1$. An analogue of the Perron–Frobenius theorem (see the proof of Lemma 32) then implies that 1 is an eigenvalue of $B_{\delta=\hat{x}}$, and more particularly that $(B^k)_{4,1}|_{\delta=\hat{x}}$ converges to a positive constant as $k \to \infty$, i.e., (4.18) has a pole at $\hat{x}$. Since, for all $\delta > x$, (4.18) is finite and positive (i.e., since $\rho(B) < 1$ and B is non-negative), the divergence of (4.18) to $+\infty$ as $\delta \to \hat{x}^+$ implies that the image of (4.18) on $(\hat{x}, \infty)$ is $(0, \infty)$, by the intermediate value theorem. The existence of $\delta^{(2)}$

follows.

For $m = 2$, we can be more explicit about the quantities above. The spectral radius of B is given by its maximal eigenvalue,

$$\rho(B) = \frac{1}{2}\left(\left(\frac{4}{7}\right)^{3\delta} + \left(\frac{5}{8}\right)^{3\delta} + \left(\left(\left(\frac{4}{7}\right)^{3\delta} + \left(\frac{5}{8}\right)^{3\delta}\right)^2 + 4\left(\left(\frac{21}{65}\right)^{3\delta} - \left(\frac{5}{14}\right)^{3\delta}\right)\right)^{\frac{1}{2}}\right),$$

which is less than one for $\delta > 0.429\ldots$ (this function is decreasing by [97, Theorem 2.1]). For these values, we have the following, explicit form for (4.18):

$$\det\begin{pmatrix} \left(\frac{6}{11}\right)^{3\delta} & \left(\frac{4}{7}\right)^{3\delta} - 1 & \left(\frac{7}{13}\right)^{3\delta} \\ \left(\frac{2}{3}\right)^{3\delta} & \left(\frac{3}{5}\right)^{3\delta} & \left(\frac{5}{8}\right)^{3\delta} - 1 \\ \left(\frac{5}{7}\right)^{3\delta} & \left(\frac{7}{10}\right)^{3\delta} & \left(\frac{3}{4}\right)^{3\delta} \end{pmatrix} / \det(I - B)$$

$$= \left(2^{-3\delta} + \left(\frac{5}{7}\right)^{3\delta} + \left(\frac{21}{55}\right)^{3\delta} + \left(\frac{25}{98}\right)^{3\delta} + \left(\frac{27}{110}\right)^{3\delta} + \left(\frac{49}{195}\right)^{3\delta} - \left(\frac{2}{7}\right)^{3\delta} - \left(\frac{3}{13}\right)^{3\delta}\right. \tag{4.19}$$

$$\left. - \left(\frac{20}{49}\right)^{3\delta} - \left(\frac{25}{56}\right)^{3\delta} - \left(\frac{21}{88}\right)^{3\delta}\right) \Big/ \left(1 + \left(\frac{5}{14}\right)^{3\delta} - \left(\frac{4}{7}\right)^{3\delta} - \left(\frac{5}{8}\right)^{3\delta} - \left(\frac{21}{65}\right)^{3\delta}\right). \tag{4.20}$$

Also using an explicit form for $\sum_{k\geq 2} a_k \prod_{i=2}^{k-1} b_i$, one calculates that $\delta^{(2)} = 0.8798\ldots$, by standard root-finding techniques. This value gives a modest improvement on the bound $\delta^* = 0.8933\ldots$ given in Proposition 17.

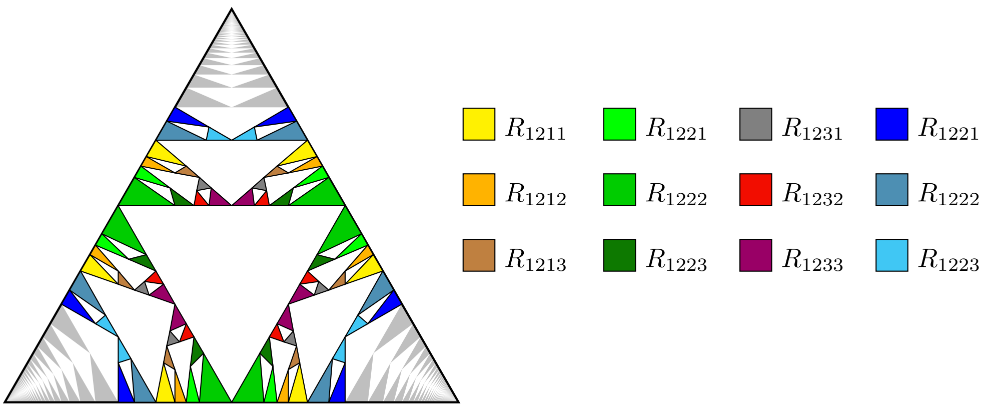


Figure 4.12: The refinement of $R_1 \cup R_2$ into 12 regions, when $m = 3$.

4.5.2 The general case

Deferring the general definitions of $B = B(m, \delta)$, a matrix defined on the index set $\bigcup_{k=1}^{m-1} \mathcal{V}_k \cup \{m\}$, and the distinguished index $\circledast \in \mathcal{V}_1$ until later, we have the following theorem, which defines the upper bound $\delta^{(m)}$ of $\hat{\delta}$ (hence of $\dim_H(\mathcal{G}) - 1$). The existence of $\delta^{(m)}$ is established by a similar argument to the above.

Theorem 11. *For each $m \geq 2$, recalling a_n, b_n from Lemma 28; whenever $\delta > \frac{1}{2}$ satisfies*

$$\sum_{k=0}^{\infty} \left(B^k\right)_{m,\circledast} \sum_{k=m}^{\infty} a_k \prod_{i=m}^{k-1} b_i < 1, \tag{4.21}$$

then $\left(X_n(\delta)\right) \to 0$.

In particular, for the value, $\delta = \delta^{(m)}$, for which the left hand side of (4.21) *equals 1, then*

$$\dim_H(\mathcal{G}) \leq 1 + \hat{\delta} \leq 1 + \max\left(\delta^{(m)}, \frac{1}{2}\right).$$

Remark 32. The results of the next section imply that $(X_n) \to 0$ for values of δ greater than $\frac{1}{2}$. Thus in hindsight, one can remove all references to $\frac{1}{2}$ above.

Remark 33. Naturally, we can consider Proposition 17 as representing the $m = 1$ case (as we do from now on), and we define $\delta^{(1)} := \delta^* = 0.8933\ldots$.

For values of $\delta^{(m)}$ up to $m = 9$, we refer the reader to the table in Figure 4.14 and the corresponding subsection, 4.6.4.

We now give a proof of Theorem 11 which will follow along the lines of Proposition 17, as outlined in the introduction to this section.

In what follows, we fix $m \in \{2, 3, 4, \ldots\}$ and suppress it from the subsequent notation. We also henceforth assume, unless otherwise stated, that $n \geq m + 1$ (a fact we will sometimes emphasise).

Partitioning $\{1, 2, 3\}^{m+1}$ via symmetries of Δ

We now lay down some simple prerequisites necessary for defining the (mth) decomposition of X_n, and thereby the matrix B (this matrix resembles $B(2, \delta)$ above, but is increasingly sparse as one increases m).

The definitions below follow from our attempt to exploit the following symmetries of Δ.

Definition 36 (Sym_3)**.** Let Sym_3 denote the group of permutations on $\{1, 2, 3\}$. We define two related actions of this group as follows:

- For $v = (v_1, v_2, v_3) \in \mathbb{R}^3$, set $\sigma^*(v) = (v_{\sigma^{-1}(1)}, v_{\sigma^{-1}(2)}, v_{\sigma^{-1}(3)})$, i.e., so that $\sigma^*(e_k) = e_{\sigma(k)}$ for each k.
- For $i \in \{1,2,3\}^n$, let $\sigma \cdot i = (\sigma(i_1), \sigma(i_2), \ldots, \sigma(i_n))$.

The connection between these two actions is the following conjugacy, which is clear from the definition of the N_j. Namely, for all $i \in \{1,2,3\}$ and $v \in \Delta$,

$$\sigma^*(N_i \cdot v) = N_{\sigma \cdot i} \cdot \sigma^*(v) \tag{4.22}$$

and in particular,

$$\sigma^* \circ \phi_i = \phi_{\sigma \cdot i} \circ \sigma^*. \tag{4.23}$$

By induction, (4.22) and (4.23) extend to all $i \in \{1,2,3\}^n$ and $n \in \mathbb{N}$.

As hinted in the introduction, our decomposition of $(X_{n,k})$ consists of partitioning the elements of $(A_{n,k})_{k=1}^{m-1}$ according to equivalence classes given by the second action above. More explicitly, we have the following definition.

Definition 37 ($[\cdot]$, matching)**.** For $v \in \{1,2,3\}^{m+1}$, let

$$[v] := \{\sigma \cdot v \mid \sigma \in \mathrm{Sym}_3\}$$

denote the equivalence class (i.e., orbit) of v in $\{1,2,3\}^{m+1}$. For $i \in \{1,2,3\}^n$, we say that i *matches* v if

$$(i_1, i_2, \ldots, i_{m+1}) \in [v],$$

i.e., there exists $\sigma \in \mathrm{Sym}_3$ such that $(i_1, \ldots, i_{m+1}) = \sigma \cdot v$, which implies the important inclusion

$$\Delta_i = \phi_{i_1} \cdots \phi_{i_{m+1}}(\Delta_{(i_{m+2},\ldots,i_n)}) = \phi_{\sigma \cdot v}(\Delta_{(i_{m+2},\ldots,i_n)}) \subset \Delta_{\sigma \cdot v}. \tag{4.24}$$

Remark 34. Note that we do not need to refer explicitly to the regions $R_v \subset \Delta$ in the subsequent proof (these regions were included for the sake of exposition in the introduction), but these can be defined by taking the union of six triangles, as follows. For non-constant $v \in \{1,2,3\}^{m+1}$, let

$$R_v = \bigcup_{i \in [v]} \Delta_i.$$

These R_v are as depicted in Figures 4.10 and Figures 4.12 for the cases $m = 2$ and $m = 3$, respectively.

We now define the decomposition using these classes.

Definition 38 ($A_{n,v}$, $X_{n,v}$)**.** For any $v \in \{1,2,3\}^{m+1}$ such that v is not constant, let

$$A_{n,v} := \{i \in \{1,2,3\}^n \mid i \text{ matches } v\},$$

and let

$$X_{n,v} := \sum_{i \in A_{n,v}} F(i).$$

The above definition of $A_{n,v}$ obviously introduces duplicates. In particular, $i \in A_{n,(1,1,\ldots,1,2)}$ is equivalent to $i_1 = \cdots = i_m \neq i_{m+1}$, i.e., $i \in A_{n,m}$. Considering the remaining classes on $\{1,2,3\}^{m+1} \setminus \{1\}^n \cup \{2\}^n \cup \{3\}^n$, they admit the following "standard" representatives (whose corresponding triangles all lie in one-sixth of Δ).

Remark 35. Here and below, the use of v, $\mathcal{V}$ etc. connotes "vertex", from the graph viewpoint exhibited above.

Definition 39 ($\mathcal{V}_k$, $\mathcal{V}$)**.** For each $k = 1, \ldots m-1$, let

$$\mathcal{V}_k := \{1\}^k \times \{2\} \times \{1,2,3\}^{m-k}.$$

and let

$$\mathcal{V} := \bigcup_{k=1}^{m-1} \mathcal{V}_k.$$

Of the elements of $\mathcal{V}_1$, we denote the distinguished element

$$\circledast = (1,2,\ldots,2) \in \{1\} \times \{2\}^m.$$

Some consideration shows that these $v \in \mathcal{V}$ uniquely represent every class except $[(1,1,\ldots,1,2)]$ and $[(1,1,\ldots,1)]$. More explicitly, given $i \in \{1,2,3\}^{m+1}$ with $i_1 = \cdots = i_k \neq i_{k+1}$ for k between 1 and $m-1$, we see that $\sigma \cdot i$, where $\sigma(i_1) = 1$ and $\sigma(i_{k+1}) = 2$ lies in $\mathcal{V}_k$ (and is unique).

More generally, this shows that, for each $k \leq m-1$, if $i \in A_{n,k}$, then $i \in A_{n,v}$ for a unique $v \in \mathcal{V}_k$. That is, we have the disjoint union

$$A_{n,k} = \bigcup_{v \in \mathcal{V}_k} A_{n,k}.$$

This in turn implies the disjoint union

$$\{1,2,3\}^n \setminus \left(\{1\}^n \cup \{2\}^n \cup \{3\}^n\right) = \bigcup_{v \in \mathcal{V}} A_{n,v} \cup \bigcup_{k=m}^{n-1} A_{n,k}.$$

Having established this decomposition, we now aim to recover an analogue of Lemma 28 (i.e., pertaining to the matrix inequalities above). This requires the definition of the matrix B, which in turn requires the following definitions of S_j for $j = 1, 2, 3$. These describe adjacency in the graph picture above (or more concretely, which entries in B are non-zero), and are defined by successors of the standard representatives, $v \in \mathcal{V}$.

Definition 40 (S_j)**.** Let $S_1 : \mathcal{V} \to \mathcal{V} \cup \{m\}$ and $S_2, S_3 : \mathcal{V} \to \mathcal{V}$ be defined by the following equation, for each $j \in \{1, 2, 3\}$ and $v \in \mathcal{V}$:

$$(j; v) = (j, v_1, \ldots, v_m) \in A_{m+2,S_j(v)},$$

i.e., if $v \in \mathcal{V}_{m-1}$, then $S_1(v) = m$:

$$(1, v_1, \ldots, v_m) = (1, 1, \ldots, 1, 2) \in A_{m+1,m},$$

and otherwise there exists $\sigma \in \mathrm{Sym}_3$ such that

$$\sigma \cdot (j, v_1, \ldots, v_m) = S_j(v).$$

An important consequence of this definition, which we use in the proof of Lemma 33 below, is that, if $i \in A_{n,v}$ for any $v \in \mathcal{V}$ and $\sigma \cdot (i_1, \ldots, i_{m+1}) = v$, then

$$(\sigma^{-1}(j), i_1, \ldots, i_{m+1}) = \sigma^{-1} \cdot (j; v) \in A_{m+2,S_j(v)}, \tag{4.25}$$

i.e., $(\sigma^{-1}(j); i) \in A_{n,S_j(v)}$.

We now define B as a weighted adjacency matrix on $\mathcal{V} \cup \{m\}$, where adjacency is defined by the S_j, and the weights are defined analogously to a_k, b_k above.

Definition 41 (B)**.** let $B \in \mathbb{R}^{\mathcal{V}\cup\{m\}\times\mathcal{V}\cup\{m\}}$ be the non-negative matrix defined by

$$B_{v,y} = \begin{cases} \max_{x\in\Delta_y} \|N_j \cdot x\|^{-3\delta} = \max_{x\in\Delta_y} (2 - x_j)^{-3\delta}, & \text{if } v = S_j(y) \text{ for some } j \in \{1, 2, 3\}; \\ 0, & \text{otherwise.} \end{cases}$$

In particular, $B_{v,m} = 0$ for all $v \in \mathcal{V} \cup m$.

We now state two lemmas, the proofs of which we defer to appendix B to focus on the main part of the proof. The first gives an explicit form for the S_j (which is useful for computation), and in a final remark, proves two statements that show first that B is well-defined, and one that implies, with the exception of one zero row (corresponding to $\circledast$) and column (corresponding to m), that B^{m+1} contains only positive entries. The latter allows us to apply the Perron–Frobenius theorem to prove the second lemma, which is needed to control the remainder term in the important renewal-style inequality for $X_{n,\circledast}$ that we derive later.

Lemma 31. *The S_j $(j = 1, 2, 3)$ are given by the following formulae: Firstly,*

$$S_1(v) = \begin{cases} m, & \text{if } v \in \mathcal{V}_k; \\ (1, v_1, \ldots v_m), & \text{otherwise.} \end{cases}$$

Secondly,

$$S_2(v) = \tau \cdot (2, v_1, \ldots, v_m),$$

where

$$\tau := \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

is the transposition interchanging 1 and 2, e.g., $\tau \cdot (2, 1, 2, 3) = (1, 2, 1, 3)$. Similarly,

$$S_3(v) = \kappa \cdot (3, v_1, \ldots, v_3),$$

where

$$\kappa := \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$$

is the cycle taking 3 to 1 and 1 to 2, e.g., $\kappa \cdot (3, 1, 2, 3) = (1, 2, 3, 1)$.

Moreover, $j \mapsto S_j(v)$ is injective for any $v \in \mathcal{V}$, and for each $v, y \in \mathcal{V} \setminus \{\circledast\}$, there exists a tuple $j \in \{1, 2, 3\}^{m+1}$ such that

$$v = S_{j_1} S_{j_2} \cdots S_{j_{m+1}}(y).$$

Lemma 32. *There exists a non-negative matrix, $D \in \mathbb{R}^{\mathcal{V} \cup \{m\} \times \mathcal{V} \cup \{m\}}$ such that, for each $v, y \in \mathcal{V} \cup \{m\}$,*

$$\left(B^n\right)_{v,y} \rho(B)^{-n} \to D_{v,y}$$

as $n \to \infty$. Moreover, $D_{v,y} > 0$ if and only if $v \neq \circledast$ and $y \neq m$.

Returning to the main course of the proof, we have the following analogue of Lemma 28.

Lemma 33. *For all $n > k > m$ and $v \in \mathcal{V} \cup \{m\} \setminus \{\circledast\}$, we have the following, recalling a_k,b_k and c_k from Lemma 28:*

$$X_{n+1,k+1} \leq b_k X_{n,k}, \tag{4.26}$$

$$X_{n+1,\circledast} \leq c_n + \sum_{k=m}^{n-1} a_k X_{n,k}, \tag{4.27}$$

$$X_{n+1,v} \leq \sum_{y \in \mathcal{V}} B_{v,y} X_{n,y}. \tag{4.28}$$

Remark 36. Note that (4.27) in particular implies that there does not exist $v \in \mathcal{V}$ and $j = 1, 2, 3$ such that $S_j(v) = \circledast$. In particular, $B_{\circledast,v} = 0$ for all $v \in \mathcal{V} \cup \{m\}$, giving a zero row in B, and hence $(B^n)_{\circledast,v} = 0$ for all such v and $n \in \mathbb{N}$. (A similar statement holds for $(B^n)_{v,m}$ from the definition.)

The proof is similar to that of Lemma 28, again using the notion of successor.

Proof of Lemma 33. We prove the inequalities in order:

The inequality (4.26) is the same as (4.7) proven in Lemma 28, and is merely stated here for completeness. We note for later that this case accounts for all principal successors of all $i \in \bigcup_{k \geq m} A_{n,k}$.

Regarding the proof of (4.27), by the definition of $\circledast$, the j-successor of $i \in \{1, 2, 3\}^n$ matches $\circledast$ if and only if

$$j \neq i_1 = i_2 = \cdots = i_m.$$

That is, $j \neq i_1$ and i is either constant or lies in $i \in A_{n,k}$ for some $k \geq m$. This gives the following analogue of (4.10) from Lemma 28:

$$X_{n+1,\circledast} = c_n + \sum_{k=m}^{n-1} \sum_{j \neq i_1} F(j; i) \leq c_n + \sum_{k=m}^{n-1} \max_{i \in A_{n,k}} \left(\sum_{j \neq i_1} \frac{F(j; i)}{F(i)} \right) X_{n,k}, \tag{4.29}$$

where $c_n = 6\, F(1, 2, 2, \ldots, 2)$ is, as before, the contribution from elements of $\bigcup_{k \neq j} \{k\} \times \{j\}^n$. Recalling (4.11), i.e.,

$$\sum_{j \neq i_1} \frac{F(j; i)}{F(i)} \leq a_k$$

for any $i \in A_{n,k}$, and applying it in (4.29) thus yields (4.27).

For the proof of (4.28), note that every successor of each $i \in A_{n,k}$ for $k \geq m$ is accounted for in the previous two cases. Therefore, fixing $v \in \mathcal{V} \cup \{m\} \setminus \{\circledast\}$, if $i \in \{1, 2, 3\}^n$ has a successor in $A_{n+1,v}$, then $i \in A_{n,y}$ for some $y \in \mathcal{V}$.

In particular, if $S_j(y) = v$ for some j, then it is unique by Lemma 31. Writing σ_i for the unique permutation such that

$$y = \sigma_i \cdot (i_1, \ldots, i_{m+1})$$

(this is well-defined for each i, since y and j are unique), one has, from (4.25),

$$X_{n+1,v} = \sum_{j=1}^{3} \sum_{y \in S_j^{-1}(v)} \sum_{i \in A_{n,y}} F\big(\sigma_i^{-1}(j); i\big). \tag{4.30}$$

We now bound the ratio $F(\sigma_i^{-1}(j); i)/F(i)$ in the same way as before. Firstly, using the symmetries (4.22) and (4.23) (and that $\|\sigma^*(x)\| = \|x\|$ for any x), we have that, for any $\sigma \in \mathrm{Sym}_3$,

$$\begin{aligned} \|N_{\sigma^{-1}(j)} \cdot \phi_i(e_k)\| &= \big\|\sigma^*\big(N_{\sigma(j)} \cdot \phi_i(e_k)\big)\big\| \\ &= \|N_j \cdot \big(\phi_{\sigma \cdot i} \circ \sigma^*(e_k)\big)\| \\ &= \|N_j \cdot \phi_{\sigma \cdot i}(e_{\sigma(k)})\|. \end{aligned} \tag{4.31}$$

Then, if $i \in A_{n,y}$ such that $v = S_j(y)$, applying (4.31) with $\sigma = \sigma_i$, and using that

$$\Delta_{\sigma_i \cdot i} = \phi_{\sigma_i \cdot (i_1, \ldots, i_{m+1})}(\Delta_{i_{m+2}, \ldots, i_n}) = \phi_y(\Delta_{i_{m+2}, \ldots, i_n}) \subset \Delta_y,$$

we have the following:

$$\begin{aligned} \frac{F\big(\sigma_i^{-1}(j); i\big)}{F(i)} &= \big\|N_{\sigma_i^{-1}(j)} \cdot \phi_i(e_1)\big\|^{-\delta} \big\|N_{\sigma_i^{-1}(j)} \cdot \phi_i(e_2)\big\|^{-\delta} \big\|N_{\sigma_i^{-1}(j)} \cdot \phi_i(e_3)\big\|^{-\delta} \\ &= \big\|N_j \cdot \phi_{\sigma_i \cdot i}(e_{\sigma_i(1)})\big\|^{-\delta} \big\|N_j \cdot \phi_{\sigma_i \cdot i}(e_{\sigma_i(2)})\big\|^{-\delta} \big\|N_j \cdot \phi_{\sigma_i \cdot i}(e_{\sigma_i(3)})\big\|^{-\delta} \\ &\le \max_{x \in \Delta_y} \|N_j \cdot x\|^{-3\delta} \\ &=: B_{v,y}. \end{aligned} \tag{4.32}$$

Consequently,

$$\begin{aligned} X_{n+1,v} &\le \sum_{j=1}^{3} \sum_{y \in S_j^{-1}(v)} \sum_{i \in A_{n,y}} B_{v,y} F(i) \\ &= \sum_{j=1}^{3} \sum_{y \in S_j^{-1}(v)} B_{v,y} X_{n,y} = \sum_{y \in \mathcal{V}} B_{v,y} X_{n,y}, \end{aligned}$$

proving (4.28). $\square$

We are now in a position to prove the theorem. The remainder of the proof follows that of Proposition 17, but is naturally more involved.

Proof of Theorem 11. We start by relating (X_n) and $(X_{n,\circledast})$ asymptotically. Observe that $(1, 2^m; i) := (1, \underbrace{2, \ldots, 2}_{m}, i_1, i_2, \ldots, i_n) \in A_{n+m+1,\circledast}$ for any $i \in \{1,2,3\}^n$. Thus,

$$X_{n+m+1;\circledast} \geq \sum_{i\in\{1,2,3\}^n} F(1, 2^m; i). \tag{4.33}$$

Following the proof of (4.15), we estimate:

$$\begin{aligned}\frac{F(1,2^m;i)}{F(i)} &= \prod_{j=1}^{3} \|N_1 N_2^m \cdot \phi_i(e_j)\|^{-\delta} \\ &\geq \min_{v\in\Delta} \quad \|N_1 N_2^m \cdot v\|^{-3\delta} \\ &= \min_{(x,y,z)\in\Delta} \big((2m+1)x + 2y + (2m+2)z\big)^{-3\delta} \\ &= (2m+2)^{-3\delta}. \end{aligned} \tag{4.34}$$

Inserting this into (4.33), we thus obtain

$$(2m+2)^{-3\delta} X_{n,k} \leq X_{n+m+1,\circledast}.$$

Therefore, $(X_n) \to 0$ if $(X_{n,\circledast}) \to 0$.

We now deduce a renewal-style inequality for $(X_{n,\circledast})_{n=m}^{\infty}$ as follows. First, applying $X_{n+1,k+1} \leq b_k X_{n,k}$ repeatedly in (4.27) gives, for all $n \geq m+1$,

$$X_{n+1,\circledast} \leq c_n + \sum_{k=m}^{n-1} a_k \prod_{i=m}^{k} b_i X_{n+m-k,m}. \tag{4.35}$$

Moreover, for all $\hat{n} \geq m+1$ and $v \in \mathcal{V} \cup \{m\} \setminus \{\circledast\}$, applying (4.28) $\hat{n}+1-m$ times yields

$$X_{\hat{n},v} \leq \sum_{k=1}^{\hat{n}-m-1} \left(B^k\right)_{v,\circledast} X_{\hat{n}-k,\circledast} + \sum_{\substack{y\in\mathcal{V}\\ y\neq\circledast}} \left(B^{\hat{n}-m-1}\right)_{v,y} X_{m+1,y}, \tag{4.36}$$

as we now show inductively. The base case of $\hat{n} = m+1$ is trivial. Inductively assuming (4.36), and using that $(B^j)_{\circledast,v} = 0$ and $(B^j)_{v,m} = 0$ for all $j \in \mathbb{N}$ and

$v \in \mathcal{V} \cup \{m\}$ (see the remark following Lemma 33), we have the following:

$$
\begin{aligned}
X_{\hat{n}+1,v} &\leq \sum_{v\in\mathcal{V}} B_{v,y} X_{\hat{n},y} \\
&= B_{v,\circledast} X_{\hat{n},\circledast} + \sum_{\substack{y\in\mathcal{V}\\ y\neq\circledast}} B_{v,y} X_{\hat{n},y} \\
&\leq B_{v,\circledast} X_{\hat{n},\circledast} + \sum_{j=1}^{\hat{n}-m-1} \sum_{\substack{y\in\mathcal{V}\\ y\neq\circledast}} B_{v,y} (B^j)_{y,\circledast} X_{m+1,\circledast} + \sum_{\substack{w,y\in\mathcal{V}\\ w,y\neq\circledast}} B_{v,y} (B^{\hat{n}-m-1})_{y,w} X_{m+1,w} \\
&= B_{v,\circledast} X_{\hat{n},\circledast} + \sum_{j=1}^{\hat{n}-m-1} (B^{j+1})_{y,\circledast} X_{\hat{n}-j,\circledast} + \sum_{\substack{w\in\mathcal{V}\\ w\neq\circledast}} (B^{\hat{n}-m})_{y,w} X_{m+1,w} \\
&= \sum_{j=1}^{\hat{n}-m} (B^j)_{y,\circledast} X_{\hat{n}+1-j,\circledast} + \sum_{\substack{w\in\mathcal{V}\\ w\neq\circledast}} (B^{\hat{n}-m})_{y,w} X_{m+1,w}.
\end{aligned}
\tag{4.37}
$$

This completes the induction for (4.36). Applying it now with $m = y$ and $\hat{n} = n + m - k$ in the summands of (4.35) gives

$$
\begin{aligned}
X_{n+1,\circledast} &\leq c_n + \sum_{k=m}^{n-1} a_k \prod_{i=m}^{k-1} b_i X_{n+m-k,m}, \\
&\leq c_n + \sum_{k=m}^{n-1} a_k \prod_{i=m}^{k-1} b_i \left(\sum_{j=1}^{n-k-1} \left(B^j\right)_{m,\circledast} X_{n+m-k-j,\circledast} + \sum_{\substack{y\in\mathcal{V}\\ y\neq\circledast}} \left(B^{n-k-1}\right)_{m,y} X_{m+1,y} \right) \\
&= \mathcal{E}_n + \sum_{k=m}^{n-1} a_k \prod_{i=m}^{k-1} b_i \sum_{j=1}^{n-k-1} \left(B^j\right)_{m,\circledast} X_{n+m-k-j,\circledast} \\
&= \mathcal{E}_n + \sum_{l=0}^{n-m-1} \sum_{\substack{k+j=l\\ j\geq 0, k\geq 1}} a_{m+j} \prod_{i=m}^{m+j-1} b_i \left(B^k\right)_{m,\circledast} X_{n-l,\circledast} \\
&= \mathcal{E}_n + \sum_{k=1}^{n-m} \lambda_k X_{n+1-k,\circledast},
\end{aligned}
\tag{4.38}
$$

where we have defined

$$
\mathcal{E}_n := c_n + \sum_{k=m}^{n-1} a_k \prod_{i=m}^{k-1} b_i \sum_{\substack{y\in\mathcal{V}\\ y\neq\circledast}} \left(B^{n-k-1}\right)_{m,y} X_{m+1,y}
$$

and

$$\lambda_k := \sum_{\substack{i+j=k\\ i,j\geq 0}} \left(B^i\right)_{m,\circledast} a_{m+j} \prod_{l=m}^{m+j-1} b_l$$

(this definition uses that $(B^0)_{m,\circledast} = 0$, i.e., $m \neq \circledast$).

Remark 37. We again remark that, from the graph viewpoint (see Figure 4.11 for the case of $m = 2$), λ_k corresponds to the sum of products over loops of length k which meet $\circledast$ once, and $\mathcal{E}_n$ to edge paths of length $n-m-1$ which start at vertices in $\mathcal{V} \setminus \{\circledast\}$, end at $\circledast$ and do not visit $\circledast$ as an intermediate vertex.

In view of the renewal-style inequality above, since $\lambda_k > 0$ for all large enough k, Lemma 30 implies that $(X_{n,\circledast}) \to 0$ whenever

1. $\sum_{n=1}^{\infty} \lambda_n < 1$, and
2. $\sum_{n=1}^{\infty} \mathcal{E}_n < \infty$.

We now show that these two numbered points are implied by the conditions of Theorem 11. To consider the second, note that, if (4.21) holds, then

$$\sum_{k=0}^{\infty} (B^k)_{m,\circledast}$$

converges. Since $(B^k)_{m,\circledast} \sim D_{m,\circledast}\rho(B)^k$ (by Lemma 32), this implies that $\rho(B) < 1$. Furthermore, by direct calculation,

$$a_k \prod_{i=m}^{k-1} b_i = \left(\frac{m+2}{k+2}\right)^{3\delta} \left(\left(\frac{k+1}{2k+1}\right)^{-3\delta} + 2^{-3\delta}\right) = \mathcal{O}\big((k+2)^{-3\delta}\big) \qquad (k \in \mathbb{N}). \tag{4.39}$$

Combining this with another application of Lemma 32, there exist $C, C' > 0$ such that

$$\begin{aligned}\sum_{n=1}^{\infty} \mathcal{E}_n - c_n &= \sum_{n=1}^{\infty} \sum_{k=m}^{n-1} a_k \prod_{i=m}^{k-1} b_i \sum_{\substack{y\in\mathcal{V}\\ y\neq\circledast}} \left(B^{n-k}\right)_{m,y} X_{m,y} \\ &\leq C \sum_{n=1}^{\infty} \sum_{k=m}^{n-1} a_k \prod_{i=m}^{k-1} b_i \; \rho(B)^{n-k}\end{aligned}$$

(continued on next page)

(continued from previous page)

$$\leq C' \sum_{n=1}^{\infty} \sum_{k=m}^{n-1} (k+2)^{-3\delta} \, \rho(B)^{n-k}$$
$$= C' \sum_{k=m}^{\infty} \sum_{n=k+1}^{\infty} (k+2)^{-3\delta} \, \rho(B)^{n-k}$$
$$= \frac{C'\rho(B)}{1-\rho(B)} \sum_{k=m}^{\infty} (k+2)^{-3\delta},$$

which is finite if $\delta > \frac{1}{3}$. Recalling that $\sum_{n=1}^{\infty} c_n$ converges if and only if $\delta > \frac{1}{2}$, we see that $\sum_{n=1}^{\infty} \mathcal{E}_n$ converges if and only if $\delta > \frac{1}{2}$.

Expressed more simply, the first condition is equivalent to (4.21):¶

$$1 > \sum_{k=1}^{\infty} \lambda_k = \sum_{k=1}^{\infty} \sum_{\substack{i+j=k \\ i,j \geq 0}} \left(B^i\right)_{\circledast,m} a_{m+j} \prod_{l=m}^{m+j-1} b_l = \sum_{k=1}^{\infty} \left(B^k\right)_{\circledast,m} \sum_{k=m}^{\infty} a_k \prod_{i=m}^{k-1} b_i.$$

Since the a_k and b_k and all the entries of B are all decreasing functions in δ, this last condition is implied by $\delta > \delta^{(m)}$, completing the proof. $\square$

4.6 Lower bounds for $\hat{\delta}$

In this section, we apply the above methodology to obtain lower bounds for $\hat{\delta}$. Although these bounds do not contribute to bounding the dimension, they show the limitations of deriving upper bounds for the dimension established by considering (X_n).

Moreover, as we shall see, the methods to prove lower bounds carry through with fewer complications, and bounds can also be obtained by other techniques.

4.6.1 Naive lower bounds

Before applying the more complex methods using renewal theory, we derive lower bounds for $\hat{\delta}$ via two distinct, elementary methods.

For example, we use Jensen's inequality to show the following result.

Proposition 18. *$(X_n) \to \infty$ for*

$$\delta > \frac{\log(3)}{3\log(5/3)} = 0.716887\ldots.$$

¶We recall that this condition corresponds, in the graph picture, in the graph picture, to the sum over all loops which meet $\circledast$ once (of the product over their edge weights).

Proof of Proposition 18. Noting that $N_1 \cdot e_1$ and $N_1 \cdot e_j = e_1 + e_j$ if $j \neq 1$, we have, by symmetry,

$$\begin{aligned}
X_{n+1} &= 3 \sum_{i \in \{1,2,3\}^n} F(i_1, \ldots, i_n, 1) \\
&= 3 \sum_{i \in \{1,2,3\}^n} \|N_i N_1 \cdot e_1\|^{-\delta} \|N_i N_1 \cdot e_2\|^{-\delta} \|N_i N_1 \cdot e_3\|^{-\delta} \\
&= 3 \sum_{i \in \{1,2,3\}^n} \|N_i \cdot e_1\|^{-\delta} \|N_i \cdot e_1 + N_i \cdot e_2\|^{-\delta} \|N_i \cdot e_1 + N_i \cdot e_3\|^{-\delta} \\
&\geq 3 \cdot 4^{\delta-1} \sum_{i \in \{1,2,3\}^n} \|N_i \cdot e_1\|^{-\delta} \big(\|N_i \cdot e_1\|^{-\delta} + \|N_i \cdot e_2\|^{-\delta}\big)\big(\|N_i \cdot e_1\|^{-\delta} + \|N_i \cdot e_3\|^{-\delta}\big) \\
&\geq 3 \cdot 4^{\delta-1} \sum_{i \in \{1,2,3\}^n} \|N_i \cdot e_1\|^{-3\delta},
\end{aligned}$$

where the first inequality uses the convexity of $t \mapsto t^{-\delta}$ and the triangle inequality.

Expressing the right hand side as an expectation involving Z_n, a random matrix distributed uniformly in $\big\{N_i \mid i \in \{1,2,3\}^n\big\}$, Jensen's inequality [37, Theorem 1.6.2] applies to give

$$\frac{X_{n+1}}{3^{n+1}} \geq 4^{\delta-1}\, \mathbb{E}\big(\|Z_n \cdot e_1\|^{-\delta}\big) \geq 4^{\delta-1}\, \mathbb{E}\big(\|Z_n \cdot e_1\|\big)^{-\delta}.$$

This last expectation has a simple form: writing $\mathbb{P}_n$ as shorthand for the probability $\mathbb{P}\big(Z_n \cdot e_1 = (a,b,c)\big)$, we have

$$\begin{aligned}
\mathbb{E}\big(\|Z_{n+1} \cdot e_1\|\big) &= \frac{1}{3} \sum_{(a,b,c) \in \mathbb{N}_0} \sum_{j=1}^{3} \|N_j \cdot (a,b,c)\|\, \mathbb{P}_n \\
&= \frac{1}{3} \sum_{(a,b,c) \in \mathbb{N}_0} \big(\|(a+b+c,b,c)\| + \|(a,a+b+c,c)\| + \|(a,b,a+b+c)\|\big)\mathbb{P}_n \\
&= \frac{1}{3} \sum_{(a,b,c) \in \mathbb{N}_0} \big((a+2b+2c) + (2a+b+2c) + (2a+2b+c)\big)\, \mathbb{P}_n \\
&= \frac{1}{3} \sum_{(a,b,c) \in \mathbb{N}_0} 5(a+b+c)\mathbb{P}_n \\
&= \frac{5}{3} \sum_{(a,b,c) \in \mathbb{N}_0} \|(a,b,c)\|\, \mathbb{P}_n
\end{aligned}$$

(continued on next page)

(continued from previous page)

$$= \frac{5}{3}\mathbb{E}\big(\|Z_n \cdot e_1\|\big),$$

and hence $\mathbb{E}(\|Z_n \cdot e_1\|) = \left(\frac{5}{3}\right)^n \|e_1\| = \left(\frac{5}{3}\right)^n$. Thus, X_n is bounded below by a multiple of

$$3^n \left(\frac{5}{3}\right)^{-3\delta\, n},$$

which diverges to infinity when

$$\delta > \frac{\log(3)}{3\log(5/3)},$$

as required. □

We now present a simpler method, the ancestor of the renewal-theoretic methods used in this chapter, which gives a stronger lower bound. More specifically, in the proof, one simply bounds from below each ratio $\sum_{\omega=1}^{3} F(\omega;i)/F(i)$ uniformly over $i \in \{1,2,3\}$.

Proposition 19. *For all $n \in \mathbb{N}$,*

$$\frac{X_{n+1}}{X_n} \geq \min_{v \in \Delta_1 \cup \Delta_2 \cup \Delta_3} \Big((2-v_1)^{-3\delta} + (2-v_2)^{-3\delta} + (2-v_3)^{-3\delta} \Big) \tag{4.40}$$

$$= \left(\frac{2}{3}\right)^{3\delta} + 2\left(\frac{4}{7}\right)^{3\delta}. \tag{4.41}$$

Therefore, $(X_n) \to \infty$ whenever the right hand side is greater than 1*, i.e.,* $\delta > 0.729\ldots$.

Proof of Proposition 19. Using that

$$v \mapsto -\log(\|N_j \cdot v\|) = -\log(2 - v_j)$$

is convex on Δ for each $j \in \{1,2,3\}$ (i.e., as a convex function with a linear argument), we first have

$$\log\left(\prod_{k=1}^{3}\|N_j\cdot\phi_i(e_k)\|^{-\delta}\right) = -\frac{3\delta}{3}\log\left(\prod_{k=1}^{3}\|N_j\cdot\phi_i(e_k)\|\right)$$
$$= 3\delta\cdot\frac{1}{3}\sum_{k=1}^{3} -\log\left(\|N_j\cdot\phi_i(e_k)\|\right)$$
$$\geq -3\delta\log\left(\left\|N_j\cdot\frac{1}{3}\sum_{k=1}^{3}\phi_i(e_k)\right\|\right)$$
$$= \log\left(\left\|N_j\cdot\frac{1}{3}\sum_{k=1}^{3}\phi_i(e_k)\right\|^{-3\delta}\right). \tag{4.42}$$

To apply this last estimate, note that $\frac{1}{3}\sum_{k=1}^{3}\phi_i(e_k) \in \Delta_i \subset \Delta_1\cup\Delta_2\cup\Delta_3$, and recall, for each $i\in\{1,2,3\}^n$ and $j = 1,2,3$,

$$\frac{F(j;i)}{F(i)} = \prod_{k=1}^{3}\|N_j\cdot\phi_i(e_k)\|^{-\delta}$$

(see, e.g., (4.9)–(4.11)). This altogether gives the following:

$$\begin{aligned}
X_{n+1} &= \sum_{i\in\{1,2,3\}^n}\sum_{j=1}^{3}F(j;i)\\
&= \sum_{i\in\{1,2,3\}^n}\sum_{j=1}^{3}\prod_{k=1}^{3}\|N_j\cdot\phi_i(e_k)\|^{-\delta}F(i)\\
&\geq \sum_{i\in\{1,2,3\}^n}\sum_{j=1}^{3}\left\|N_j\cdot\frac{1}{3}\sum_{k=1}^{3}\phi_i(e_k)\right\|^{-3\delta}F(i)\\
&\geq \min_{v\in\Delta_1\cup\Delta_2\cup\Delta_3}\left(\sum_{j=1}^{3}\|N_j\cdot v\|^{-3\delta}\right)\sum_{i\in\{1,2,3\}^n}F(i)\\
&= \min_{v\in\Delta_1\cup\Delta_2\cup\Delta_3}\left(\sum_{j=1}^{3}(2-v_j)^{-3\delta}\right)X_n,
\end{aligned}$$

which proves (4.40).

The minimum can be calculated using the fact that, for any $v\in\Delta_1\cup\Delta_2\cup\Delta_3$, the convex hull of $\{\sigma\cdot v \mid \sigma\in\mathrm{Sym}_3\{1,2,3\}\}$ contains the point $(\frac{1}{2},\frac{1}{4},\frac{1}{4})$, as sketched in Figure 4.13. Then, as in the proof of Lemma 29, convexity shows that the value taken at this interior point is smaller than the maximum of the values taken on the vertices $\{\sigma \cdot v \mid \sigma \in \mathrm{Sym}_3\{1,2,3\}\}$, which by symmetry is the value taken at v. Since v is arbitrary, this implies that $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ is a minimiser, leading to (4.41). $\square$

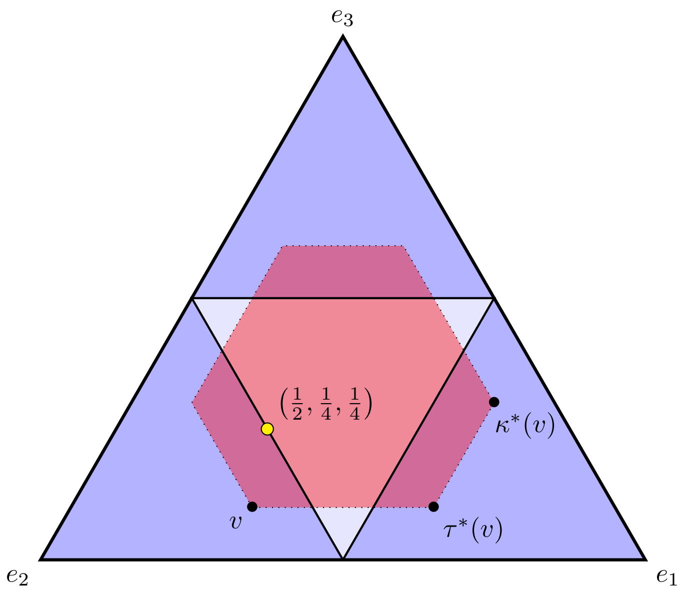


Figure 4.13: Sketch for the proof of (4.41). The hexagon is the convex hull of the images of $v \in \Delta_2$ under the symmetries of Δ, which contains $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$. ($\kappa, \tau \in \mathrm{Sym}_3$ are defined in Lemma 31.)

4.6.2 Renewal method: the value δ_1

We now follow the method of the previous sections to bound $\hat{\delta}$ from below. For clarity, in this subsection we consider the basic case corresponding to Proposition 17 (and $m = 1$), before considering the more general case ($m \geq 2$) in the next subsection. The only difference in the proofs here as opposed to the previous sections are that we now estimate, e.g.,

$$\frac{F(i_1; i)}{F(i)} \qquad \text{and} \qquad \sum_{\omega \neq i_1} \frac{F(\omega; i)}{F(i)}$$

from below rather than from above.

To begin, one has the following lemma, which uses the values for the lower bounds provided by Lemma 29.

Lemma 34. *For*

$$\tilde{a}_k := \min_{v \in R_k \cap \Delta_1} (2-v_2)^{-3\delta} + (2-v_3)^{-3\delta} = 2^{3\delta+1}\left(\frac{k+2}{4k+7}\right)^{3\delta}$$

and

$$\tilde{b}_k := \min_{v \in R_k \cap \Delta_1} (2-v_2)^{-3\delta} + (2-v_3)^{-3\delta} = \left(\frac{k+1}{k+2}\right)^{3\delta},$$

(i.e., as given in Lemma 29), we have that, for all $n > k \geq 1$,

$$X_{n+1,1} \geq \sum_{j=1}^{n-1} \tilde{a}_j X_{n,j}, \tag{4.43}$$

$$X_{n+1,k+1} \geq \tilde{b}_k X_{n,k}. \tag{4.44}$$

The proof of this lemma is a natural adaptation of the proof of Lemma 28.

Proof of Lemma 34. Recalling (4.8) gives

$$\begin{aligned}
X_{n+1,k+1} &= \sum_{i \in A_{n,k}} F(i_1; i) \\
&\geq \min_{i \in A_{n,k}} \left(\frac{F(i_1;i)}{F(i)}\right) X_{n,k} \\
&= \min_{i \in A_{n,k}} \left(\prod_{j=1}^{3} \|N_{i_1} \cdot \phi_i(e_k)\|^{-\delta}\right) X_{n,k} \\
&\geq \min_{x \in R_k \cap \Delta_{i_1}} \|N_{i_1} \cdot x\|^{-3\delta} X_{n,k} \\
&= \min_{x \in R_k \cap \Delta_1} (2-x_1)^{-3\delta} X_{n,k}. \\
&=: \tilde{b}_k X_{n,k},
\end{aligned}$$

which proves (4.44).

The proof of (4.43) differs from that of its predecessor, (4.6), since it uses convexity (as in the proof of Proposition 18) rather than the AM-GM inequality.

To start with, recalling (4.10), and using that $c_n \geq 0$, we have that

$$\begin{aligned} X_{n+1,1} \geq X_{n+1,1} - c_n &= \sum_{k=1}^{n-1} \sum_{i \in A_{n,k}} \sum_{\omega \neq i_1} F(\omega; i) \\ &\geq \sum_{k=1}^{n-1} X_{n,k} \min_{i \in A_{n,k}} \left(\sum_{\omega \neq i_1} \frac{F(\omega; i)}{F(i)} \right) \\ &\geq \sum_{k=1}^{n-1} X_{n,k} \min_{i \in A_{n,k}} \left(\sum_{\omega \neq i_1} \prod_{j=1}^{3} \|N_\omega \cdot \phi_i(e_j)\|^{-\delta} \right). \end{aligned} \tag{4.45}$$

Now, recalling (4.42) from the proof of Proposition 19:

$$\prod_{j=1}^{3} \|N_\omega \cdot e_j\|^{-\delta} \geq \left\| N_\omega \cdot \frac{1}{3} \sum_{j=1}^{3} e_j \right\|^{-3\delta},$$

we have

$$\begin{aligned} \sum_{\omega \neq i_1} \frac{F(\omega; i)}{F(i)} &= \sum_{\omega \neq i_1} \prod_{j=1}^{3} \|N_\omega \cdot \phi_i(e_j)\|^{-\delta} \\ &\geq \sum_{\omega \neq i_1} \left\| N_\omega \cdot \frac{1}{3} \sum_{j=1}^{3} e_j \right\|^{-3\delta} \\ &\geq \min_{x \in R_k \cap \Delta_{i_1}} \left(\sum_{\omega \neq i_1} \|N_\omega \cdot x\|^{-3\delta} \right) \\ &= \min_{x \in R_k \cap \Delta_1} (2 - x_2)^{-3\delta} + (2 - x_3)^{-3\delta} \\ &=: \tilde{a}_k. \end{aligned} \tag{4.46}$$

Applying this estimate in the summands of (4.45) yields (4.43), as required. □

We now apply this lemma to give the best lower bound for $\hat{\delta}$ so far, as per the following result.

Proposition 20. *$(X_n) \to \infty$ if*

$$\sum_{k=1}^{\infty} \tilde{a}_k \prod_{i=1}^{k-1} \tilde{b}_i = 2^{6\delta+1} \sum_{k=1}^{\infty} \left(\frac{k+2}{(k+1)(4k+7)} \right)^{3\delta} > 1.$$

Therefore, $\hat{\delta} \geq \delta_1$, where $\delta_1 = 0.7681\ldots$ is the value of δ for which the left hand side equals one.

Proof of Proposition 20. Since $X_{n,1} \le X_n$ for all $n \ge 2$, $(X_{n,1}) \to \infty$ implies $(X_n) \to \infty$. From the previous lemma (i.e., applying (4.44) in the summands of (4.43)), we obtain the following renewal-style inequality for $(X_{n,1})_{n\ge 2}$:

$$X_{n+1,1} \ge \sum_{k=1}^{n-1} \tilde{a}_k \prod_{i=1}^{k-1} \tilde{b}_i X_{n-k,1}.$$

Hence, by the renewal theorem (Lemma 30), $(X_{n,1}) \to \infty$ if

$$\sum_{k=1}^{\infty} \tilde{a}_k \prod_{i=1}^{k-1} \tilde{b}_i = 2^{6\delta+1} \sum_{k=1}^{\infty} \left(\frac{k+2}{(k+1)(4k+7)} \right)^{3\delta} > 1,$$

as required. In particular, this last condition is implied by $\delta > \delta_1$ (since $\tilde{a}_k$ and $\tilde{b}_k$ are decreasing in δ), which yields the final remark. □

Remark 38. Notice how the absence of the remainder term in the above proof makes the argument slightly simpler than that of Proposition 17. The difference will be even more pronounced in the proof of the next result.

4.6.3 Refined renewal method: the sequence (δ_m)

We now advance the renewal method, to give the following analogue of Theorem 11.

Proposition 21. *Given $m \ge 2$, with $\tilde{a}_k$ and $\tilde{b}_k$ as defined in Lemma 34, and $\mathcal{V}$ as defined on page 150, let $\tilde{B} = \tilde{B}(m, \delta) \in \mathbb{R}^{\mathcal{V}\cup\{m\}\times\mathcal{V}\cup\{m\}}$ be given by*

$$\tilde{B}_{v,y} := \begin{cases} \min\limits_{x\in\Delta_y} (2 - x_j)^{-3\delta}, & v = S_j(y); \\ 0, & \textit{otherwise.} \end{cases}$$

Then $(X_n) \to \infty$ whenever

$$\sum_{k=0}^{\infty} \left(\tilde{B}^k\right)_{m,\circledast} \sum_{k=m}^{\infty} \tilde{a}_k \prod_{i=m}^{k-1} \tilde{b}_i > 1. \tag{4.47}$$

Consequently, for the value $\delta = \delta_m$ for which the left hand side of (4.47) *equals 1, $\hat{\delta} \ge \delta_m$.*

For the remainder of this subsection, we again fix m and suppress it from the notation. We use the following lemma, which follows verbatim to Lemma 33, with minima replacing maxima.

Lemma 35. *For all $n \geq m+1$,*

$$X_{n+1,\circledast} \geq \sum_{k=m}^{n-1} \tilde{a}_k X_{n,k}, \tag{4.48}$$

and for each $v \in \mathcal{V} \cup \{m\} \setminus \{\circledast\}$,

$$X_{n+1,v} \geq \sum_{y \in \mathcal{V}} \tilde{B}_{v,y} X_{n,y}. \tag{4.49}$$

Proof of Lemma 35. Recalling (4.29), we have

$$X_{n+1,\circledast} \geq X_{n+1,\circledast} - c_n = \sum_{k=m}^{n-1} \sum_{i \in A_{n,k}} \sum_{j \neq i_1} F(j;i) \geq \sum_{k=m}^{n-1} \min_{i \in A_{n,k}} \left(\frac{F(j;i)}{F(i)} \right).$$

Then, for each $i \in A_{n,k}$, applying (4.46), i.e.,

$$\frac{F(j;i)}{F(i)} \geq \tilde{a}_k,$$

in the summands above yields (4.48).

To prove (4.49), recall (4.30):

$$X_{n+1,v} = \sum_{j=1}^{3} \sum_{y \in S_j^{-1}(v)} \sum_{i \in A_{n,y}} F\big(\sigma_i^{-1}(j); i\big).$$

Then, following the argument leading up to (4.32) on page 154, we have that, for each $i \in A_{n,y}$ such that $S_j(y) = v$,

$$\frac{F\big(\sigma_i^{-1}(j); i\big)}{F(i)} \geq \min_{x \in \Delta_y} \|S_j \cdot x\|^{-3\delta} =: \tilde{B}_{v,y}.$$

Consequently,

$$\begin{aligned} X_{n+1,v} &\geq \sum_{j=1}^{3} \sum_{y \in S_j^{-1}(v)} \sum_{i \in A_{n,y}} \tilde{B}_{v,y} F(i) \\ &= \sum_{j=1}^{3} \sum_{y \in S_j^{-1}(v)} \tilde{B}_{v,y} X_{n,y} = \sum_{y \in \mathcal{V}} \tilde{B}_{v,y} X_{n,y}, \end{aligned}$$

proving (4.49), as required. □

We proceed directly with the proof of the proposition, which follows along the lines of Theorem 11.

Proof of Proposition 21. Let $\delta > \delta_m$. Since $X_{n,\circledast} \le X_n$ for all $n \ge m+1$, it suffices to show $(X_{n,\circledast}) \to \infty$.

To obtain a renewal-style inequality for $(X_{n,\circledast})_{n\ge m+1}$: repeating the inductive proof of (4.36) (see (4.37) on page 156) gives, for all $\hat{n} \ge m+1$,

$$X_{\hat{n},m} \ge \sum_{k=1}^{\hat{n}-m-1} \left(\tilde{B}^k\right)_{m,\circledast} X_{\hat{n}-k,\circledast} + \sum_{v\in\mathcal{V}\setminus\{\circledast\}} \left(\tilde{B}^{n-m-1}\right)_{m,v} X_{m+1,\circledast} \ge \sum_{k=1}^{\hat{n}-m-1} \left(\tilde{B}^k\right)_{v,\circledast} X_{\hat{n}-k,\circledast}.$$

Hence,

$$\begin{aligned} X_{n,\circledast} &\ge \sum_{k=m}^{n-1} \tilde{a}_k \prod_{i=1}^{k-1} \tilde{b}_i \, X_{n+m-k,m} \\ &\ge \sum_{k=m}^{n-1} \tilde{a}_k \prod_{i=m}^{k-1} \tilde{b}_i \sum_{j=1}^{n-k-1} \left(\tilde{B}^j\right)_{m,\circledast} X_{n+m-k-j,\circledast} \\ &= \sum_{k=1}^{n-m} \left(\sum_{\substack{i+j=k \\ i,j\ge 0}} \left(\tilde{B}^i\right)_{m,\circledast} \tilde{a}_{m+j} \prod_{l=m}^{m+j-1} \tilde{b}_l \right) X_{n+1-k}. \end{aligned}$$

Hence, Lemma 30 applies to show that $(X_{n,\circledast}) \to \infty$ if

$$1 > \sum_{k=1}^{\infty} \left(\sum_{\substack{i+j=k \\ i,j\ge 0}} \left(\tilde{B}^i\right)_{m,\circledast} \tilde{a}_{m+j} \prod_{l=m}^{m+j-1} \tilde{b}_l \right) = \sum_{i=1}^{\infty} \left(\tilde{B}^i\right)_{m,\circledast} \sum_{j=m}^{\infty} \tilde{a}_j \prod_{l=m}^{j-1} \tilde{b}_l,$$

as required. As before, the final remark (concerning δ_m) simply follows from the fact that the entries of $\tilde{B}$, $\tilde{a}_k$ and $\tilde{b}_k$ are decreasing in δ. □

4.6.4 The values of δ_m and $\delta^{(m)}$

We now compare the computed values for the upper and lower bounds for $\hat{\delta}$ obtained by the renewal methods of the previous two sections. Rounded values for $\delta^{(m)}$ and δ_m for $m \le 9$ are given in Figure 4.14.

These bounds imply Theorem 10, i.e., that $\hat{\delta} \in [0.8095, 0.8204]$. In particular, $\delta^{(8)}$ and $\delta^{(9)}$ lead to slightly stronger bounds for $\dim_H(\mathcal{G})$ than that given by Fougeron (1.825). The lower bounds show that one cannot use X_n to prove, e.g., that $\dim_H(G) \le 1.8$.

From these data, we naturally conjecture that δ_m and $\delta^{(m)}$ each form monotonic sequences converging to $\hat{\delta}$. In particular, this would show that $\hat{\delta}$ is a threshold in the sense that $(X_n) \to \infty$ whenever $\delta < \hat{\delta}$. As mentioned above, assuming this convergence, standard sequence acceleration methods (i.e., numerical limit estimation, see [9]) yield a non-rigorous estimate of $\hat{\delta} \approx 1.8135$, which is depicted in the graph below.

m	δ_m	$\delta^{(m)}$
1	0.7681	0.8934
2	0.7766	0.8799
3	0.7862	0.8592
4	0.7939	0.8447
5	0.7994	0.8353
6	0.8034	0.8291
7	0.8061	0.8250
8	0.8080	0.8223
9	0.8095	0.8204

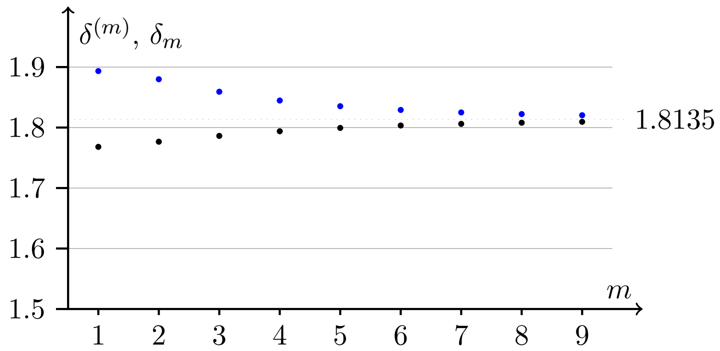


Figure 4.14: Left: values of δ_m and $\delta^{(m)}$ for $m \leq 9$, rounded (down and up respectively) to four decimal places, calculated using Wolfram Mathematica on a Lenovo ThinkPad X240 laptop (with an Intel core i5-4300 2.49GHz processor and 4GB of RAM). The ninth terms each took less than 3 minutes to verify. Right: a graphical depiction.

4.7 Enhanced upper bounds using contraction ratios: the sequence (ε_m)

Having established the sequence of upper bounds $(\delta^{(m)})$ for $\hat{\delta}$, we are now ready to introduce the next improvement to the method for bounding the dimension, which is to consider the original sequence from Lemma 25,

$$Y_n = \sum_{i \in \{1,2,3\}^n} \operatorname{area}(\Delta_i)^{\delta} \operatorname{diam}(\Delta_i)^{1-\delta}.$$

In particular, recall that $\dim_H(\mathcal{G}) \leq 1 + \delta$ whenever $(Y_n) \to 0$. Let

$$G(i) := \operatorname{area}(\Delta_i)^{\delta} \operatorname{diam}(\Delta_i)^{1-\delta} = F(i) \operatorname{diam}(\Delta_i)^{-\delta}$$

(where the last equality is up to some fixed multiple). Whereas previously, the method we used bounded ratios such as

$$\frac{F(j;i)}{F(i)}, \qquad \sum_{j\neq i_1} \frac{F(j;i)}{F(i)},$$

according to i lying in some $A_{n,v}$ or $A_{n,k}$, now we bound the corresponding quantities with F replaced by G.

This is in fact quite simple to do, at least in a naive way. In the next subsection, we state a lemma to bound the ratios of successive diameters, i.e., expressions of the form

$$\frac{\operatorname{diam}\big(\phi_j(\Delta_i)\big)}{\operatorname{diam}(\Delta_i)},$$

by using a local Lipschitz constant for ϕ_j. In the next subsection, we consider the basic ($m = 1$) case, before moving onto the more general ($m \geq 2$) setting in the final subsection, which proves the main theorem of this chapter.

4.7.1 A lemma on local contraction

The following lemma compares the diameters of $\Delta_{(j;i)}$ to Δ_i, for each $i \in \{1,2,3\}^n$ and $j = 1,2,3$. Because the proof of this result is computer-assisted and somewhat unenlightening, it is deferred to appendix B.

Lemma 36. *Treating Δ as an immersed manifold in $\mathbb{R}^3$, the tangent map $D_v\phi_j : \mathcal{T}_v\Delta \to \mathcal{T}_{\phi_j(v)}\Delta$ has maximal singular value satisfying*

$$\|D_v\phi_j\|_{\mathrm{op}} \leq (2 - v_j)^{-\lambda}, \tag{4.50}$$

where $\lambda := \frac{3}{2} - \frac{1}{\sqrt{3}} = 0.9226\ldots$. Consequently, for all $i \in \{1,2,3\}^n$,

$$2^{\lambda-3} \leq \left(\frac{\operatorname{diam}\big(\phi_j(\Delta_i)\big)}{\operatorname{diam}(\Delta_i)}\right) \leq \max_{v\in\Delta_i}(2 - v_j)^{-\lambda}. \tag{4.51}$$

4.7.2 The basic case, $m = 1$

From the previous lemma and Lemma 28 we can deduce the following, which we need to prove the first main result of this section, an analogue of Proposition 17.

Lemma 37. *For $i \in \{1,2,3\}^n$, let*

$$Y_{n,k} := \sum_{i \in A_{n,k}} G(i).$$

Then, for the values

$$\hat{b}_k := \left(\frac{k+2}{k+3}\right)^{\lambda(1-\delta)} \quad b_k = \left(\frac{k+2}{k+3}\right)^{3\delta+\lambda(1-\delta)}$$

and

$$\hat{a}_k := \left(\frac{k+1}{2k+1}\right)^{\lambda(1-\delta)} \quad a_k = \left(\frac{k+1}{2k+1}\right)^{3\delta+\lambda(1-\delta)} + 2^{-3\delta}\left(\frac{k+1}{2k+1}\right)^{\lambda(1-\delta)},$$

there exist positive constants $(\hat{c}_n)_{n=1}^{\infty}$ such that, for all $n > k \geq 1$ and $\delta \in [0,1]$,

$$Y_{n+1,1} \leq \hat{c}_n + \sum_{j=1}^{n-1} \hat{a}_j Y_{n,j} \tag{4.52}$$

and

$$Y_{n+1,k+1} \leq \hat{b}_j Y_{n,k}. \tag{4.53}$$

Moreover, there exists $C > 0$ such that

$$\hat{c}_n \sim C n^{-3\delta} \tag{4.54}$$

as $n \to \infty$ (i.e., $\sum_{n=1}^{\infty} \hat{c}_n < \infty$ if and only if $\delta > \frac{1}{3}$).

The proof of this lemma combines the inequalities from the proof of Lemma 28 with the upper bound of (4.51) from Lemma 36.

Proof of Lemma 37. From the proof of Lemma 28, repeating the proof of (4.8) with G in place of F gives

$$Y_{n+1,k+1} = \sum_{i \in A_{n,k}} G(i_1; i) \leq \max_{i \in A_{n,k}} \left(\frac{G(i_1; i)}{G(i)}\right) Y_{n,k}. \tag{4.55}$$

Recall (4.9): for all $i \in A_{n,k}$,

$$\frac{F(i_1, i)}{F(i)} \leq b_k.$$

Applying this inequality and then (4.51) gives the following:

$$\begin{aligned}
\frac{G(i_1,i)}{G(i)} &= \frac{F(i_1,i)}{F(i)}\frac{\operatorname{diam}(\phi_{i_1}\Delta_i)^{1-\delta}}{\operatorname{diam}(\Delta_i)^{1-\delta}} \\
&\le b_k \frac{\operatorname{diam}(\phi_{i_1}\Delta_i)^{1-\delta}}{\operatorname{diam}(\Delta_i)^{1-\delta}} \\
&\le b_k \max_{x\in\Delta_i}(2-x_{i_1})^{-\lambda(1-\delta)} \\
&\le b_k \max_{x\in R_k\cap\Delta_{i_1}}(2-x_{i_1})^{-\lambda(1-\delta)} \\
&= b_k \max_{x\in R_k\cap\Delta_1}(2-x_1)^{-\lambda(1-\delta)} \\
&= \big(2-\max_{x\in R_k\cap\Delta_1}(x_1)\big)^{-\lambda(1-\delta)} \\
&= b_k\left(\frac{k+2}{k+3}\right)^{\lambda(1-\delta)} \\
&=: \hat{b}_k,
\end{aligned}$$

the sixth line following from the fact that

$$R_k \cap \Delta_1 = \left\{x\in\Delta \;\middle|\; \frac{k}{k+1} \le x_1 \le \frac{k+1}{k+2}\right\} \tag{4.56}$$

(see the proof of Lemma 29 in the appendix). Combining this estimate with (4.55) thus proves (4.53).

Similarly, from the proof of (4.10) in Lemma 28,

$$\begin{aligned}
Y_{n+1,1} &= \hat{c}_n + \sum_{k=1}^{n-1}\sum_{i\in A_{n,k}}\sum_{j\ne i_1} G(j;i) \\
&\le \hat{c}_n + \sum_{k=1}^{n-1}\max_{i\in A_{n,k}}\left(\sum_{j\ne i_1}\frac{G(j;i)}{G(i)}\right)Y_{n,k},
\end{aligned} \tag{4.57}$$

where (up to a fixed multiple)

$$\hat{c}_n := 6\operatorname{area}(\phi_1\phi_2^n(\Delta))^{\delta}\operatorname{diam}(\phi_1\phi_2^n(\Delta))^{1-\delta} = c_n\operatorname{diam}(\phi_1^n\phi_2^n(\Delta))^{1-\delta} \tag{4.58}$$

is the contribution from the non-principal successors of $i\in\{1\}^n\cup\{2\}^n\cup\{3\}^n$.

More explicitly considering (4.58), recall

$$N_1 N_2^n = \begin{pmatrix} n+1 & 1 & n+1 \\ n & 1 & n \\ 0 & 0 & 1 \end{pmatrix}.$$

From the columns of this matrix, one sees that the triangle $\phi_1\phi_2^n(\Delta)$ has vertices

$$\phi_1\phi_2^n(e_1) = \left(\frac{n+1}{2n+1}, \frac{n}{2n+1}, 0\right), \qquad \phi_1\phi_2^n(e_2) = \left(\frac{1}{2}, \frac{1}{2}, 0\right),$$
$$\phi_1\phi_2^n(e_3) = \left(\frac{n+1}{2n+2}, \frac{n}{2n+2}, \frac{1}{2n+2}\right),$$

and thereby (by a routine calculation) has diameter

$$\max\left(\frac{\sqrt{6n^2+6n+2}}{2(n+1)(2n+1)}, \frac{1}{\sqrt{2}(2n+1)}, \frac{1}{\sqrt{2}(n+1)}\right).$$

Since each of these last expressions is asymptotic to a positive multiple of n^{-1} as $n \to \infty$, and since c_n is asymptotic to a positive multiple of n^{-2} (see, e.g., the definition in Lemma 28), (4.58) shows that $\hat{c}_n$ is asymptotic to a positive multiple of $n^{-3\delta}$, proving (4.54).

Considering now the summands of (4.57), let $i \in A_{n,k}$. Recalling (4.11), i.e.,

$$\sum_{j \neq i_1} \frac{F(j;i)}{F(i)} \leq a_k,$$

and again applying (4.51) from Lemma 36, we have that

$$\begin{aligned}
\sum_{j\neq i_1} \frac{G(j;i)}{G(i)} &= \sum_{j\neq i_1} \frac{F(j;i)}{F(i)} \frac{\operatorname{diam}(\phi_j(\Delta_i))^{1-\delta}}{\operatorname{diam}(\Delta_i)^{1-\delta}} \\
&\leq a_k \max_{j\neq i_1} \left(\frac{\operatorname{diam}(\phi_j(\Delta_i))^{1-\delta}}{\operatorname{diam}(\Delta_i)^{1-\delta}} \right) \\
&\leq a_k \max_{\substack{j\neq i_1 \\ x\in\Delta_i}} (2-x_j)^{\lambda(1-\delta)} \\
&\leq a_k \max_{\substack{j\neq i_1 \\ x\in R_k\cap\Delta_{i_1}}} (2-x_j)^{\lambda(1-\delta)} \\
&= a_k \max_{\substack{j=2,3 \\ x\in R_k\cap\Delta_1}} (2-x_j)^{\lambda(1-\delta)} \\
&= a_k \left(\frac{k+1}{2k+1} \right)^{\lambda(1-\delta)} \\
&=: \hat{a}_k.
\end{aligned} \tag{4.59}$$

Inserting this last estimate into (4.57) thus yields (4.52), completing the proof. □

We now consider the following marked improvement on Proposition 17. In fact, $\varepsilon_1 < \delta^{(3)}$, as can be seen in Figure 4.15.

Proposition 22. *Let $\varepsilon_1 = 0.851101\ldots$ be the value of $\delta > 0$ for which*

$$\sum_{k=1}^{\infty} \hat{a}_k \prod_{i=1}^{k-1} \hat{b}_i = 1,$$

i.e., more explicitly,

$$3^{3\varepsilon_1} \sum_{k=1}^{\infty} \left(\frac{k+1}{(k+2)(2k+1)} \right)^{3\varepsilon_1+\lambda(1-\varepsilon_1)} + \left(\frac{3}{2} \right)^{3\varepsilon_1} \sum_{k=1}^{\infty} \left(\frac{k+1}{k+2} \right)^{\lambda(1-\varepsilon_1)} (k+2)^{-3\varepsilon_1+\lambda(1-\varepsilon_1)} = 1.$$

Then $(Y_n) \to 0$ for all $\delta > \varepsilon_1$. In particular, $\dim_H(\mathcal{G}) \leq 1.851101\ldots$.

Proof of Proposition 22. Assuming that ε_1 exists and takes the above value, fix $\delta > \varepsilon_1$. Then $\delta > \frac{1}{3}$ (and therefore $\sum_n \hat{c}_n < \infty$), and since $\hat{a}_k$, $\hat{b}_k$ are decreasing in δ for each k,

$$\sum_{k=1}^{\infty} \hat{a}_k \prod_{i=1}^{k-1} \hat{b}_i < 1.$$

From Lemma 37, applying (4.53) iteratively in the summands of (4.52) as before gives the following renewal-style inequality for $(Y_{n,1})_{n\geq 2}$:

$$Y_{n+1,1} \leq \hat{c}_n + \sum_{k=1}^{n-1} \hat{a}_k \prod_{i=1}^{k-1} \hat{b}_i Y_{n+1-k,1}.$$

Thus, Lemma 30 (together with the second sentence above) implies that $(Y_{n,1}) \to 0$.

To finish the proof, we relate Y_n and $Y_{n,1}$ as in the proof of Proposition 17:

$$Y_{n+2,1} \geq \sum_{i\in\{1,2,3\}^n} G(1,2;i) \geq \min_{i\in\{1,2,3\}^n} \left(\frac{G(1,2;i)}{G(i)}\right) Y_n.$$

Let $i \in \{1,2,3\}^n$. Then, recalling (4.15):

$$\left(\frac{\operatorname{area}\big(\phi_1\phi_2(\Delta_i)\big)}{\operatorname{area}(\Delta_i)}\right)^{\delta} \geq 4^{-3\delta},$$

and applying this with the lower bound of (4.51) twice, one has

$$\begin{aligned}
\frac{G(1,2;i)}{G(i)} &\geq \left(\frac{\operatorname{area}\big(\phi_1\phi_2(\Delta_i)\big)}{\operatorname{area}(\Delta_i)}\right)^{\delta} \left(\frac{\operatorname{diam}\big(\phi_1\phi_2(\Delta_i)\big)}{\operatorname{diam}(\Delta_i)}\right)^{1-\delta} \\
&\geq 4^{-3\delta} \left(\frac{\operatorname{diam}\big(\phi_1\phi_2(\Delta_i)\big)}{\operatorname{diam}\big(\phi_2(\Delta_i)\big)} \frac{\operatorname{diam}\big(\phi_2(\Delta_i)\big)}{\operatorname{diam}(\Delta_i)}\right)^{1-\delta} \\
&\geq 4^{-3\delta} \left(2^{\lambda-3}\right)^{2(1-\delta)}.
\end{aligned}$$

That is, for all $n \in \mathbb{N}$,

$$Y_n \leq 4^{\lambda-3-\lambda\delta} Y_{n+2,1} \to 0$$

as $n \to \infty$, as required. □

4.7.3 The advanced case, $m \geq 2$

Proceeding as before, we now prove the the following, main result of this chapter which, for $m = 9$, gives Theorem 9.

Theorem 12. *Fix $m \geq 2$, and let $\hat{B} = \hat{B}(m,\delta) := B\big(m, \delta + \frac{1}{3}\lambda(1-\delta)\big)$, i.e.,*

$$(\hat{B}_m)_{v,y} := \begin{cases} \max_{x\in\Delta_y} (2-x_j)^{-3\delta+\lambda(1-\delta)}, & v = S_j(y); \\ 0, & \textit{otherwise.} \end{cases}$$

Then $(Y_n) \to 0$ *for all* $\delta > \max\left(\varepsilon_m, \frac{1}{3}\right)$*, where* ε_m *is the value of* δ *for which*

$$\sum_{k=0}^{\infty} \left(\hat{B}^k\right)_{m,\circledast} \sum_{k=m}^{\infty} \hat{a}_k \prod_{i=m}^{k-1} \hat{b}_i = 1.$$

Consequently, $\dim_H(\mathcal{G}) \leq 1 + \max\left(\varepsilon_m, \frac{1}{3}\right)$.

Values for the ε_m up to $m = 9$ are written and plotted below in Figure 4.15, where they are compared to $\delta^{(m)}$. Assuming that (ε_m) converges, we numerically estimate its limit to be ≈1.7368, via Aitken acceleration [9, p.83]. Note that we expect this (conjectured) limit to be strictly greater than the real value of the dimension (and indeed, the infimal value for which $(Y_n) \to 0$), owing to the looseness of the upper bound in Lemma 36, which is also suggested by the difference between this value and the numerical upper bound for the box-counting dimension, ≈1.72, of [33].

m	$\delta^{(m)}$	ε_m
1	0.8934	0.8512
2	0.8799	0.8285
3	0.8592	0.7978
4	0.8447	0.7764
5	0.8353	0.7624
6	0.8291	0.7534
7	0.8250	0.7475
8	0.8223	0.7435
9	0.8204	0.7407

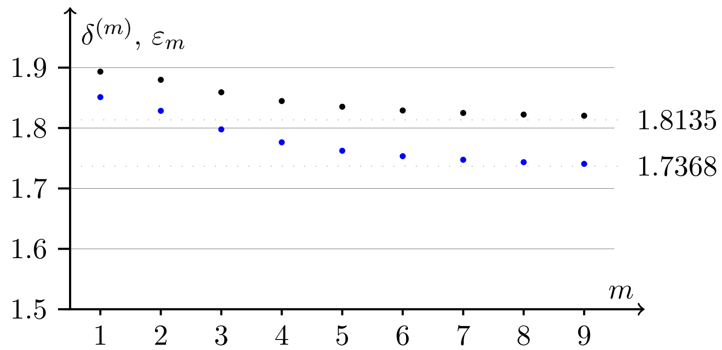


Figure 4.15: Left: values of $\delta^{(m)}$ and ε_m for $m \leq 9$, rounded up to four decimal places; calculated using Wolfram Mathematica on the author's laptop (see Figure 4.14 for details). Right: a graphical depiction, with numerically estimated limits.

The proof of the theorem again follows that of Theorem 11. As in the statement of the theorem, fix $m \geq 2$. The following is an enhancement of Lemma 33, again incorporating (4.51) from Lemma 36.

Lemma 38. *For any* $v \in \mathcal{V}$*, write*

$$Y_{n,v} = \sum_{i \in A_{n,y}} G(i).$$

Then, for all $n \geq m + 1$ *(recalling* $\mathcal{V}$*,* $\circledast = (1, 2, \ldots, 2)$ *and* $\hat{a}_n$*,* $\hat{b}_n$ *and* $\hat{c}_n$ *from* *above),*

$$Y_{n+1,\circledast} \le \hat{c}_n + \sum_{k=m}^{n-1} \hat{a}_k Y_{n,k}, \tag{4.60}$$

and for any $v \in \mathcal{V} \cup \{m\} \setminus \{\circledast\}$*,*

$$Y_{n+1,v} \le \sum_{y \in \mathcal{V}} \hat{B}_{v,y} Y_{n,y}. \tag{4.61}$$

Proof of Lemma 38. The proof of (4.60) follows similarly to the proof of (4.27). First, recalling the proof of (4.29), we have

$$\begin{aligned} Y_{n+1,\circledast} &= \hat{c}_n + \sum_{k=m}^{n-1} \sum_{i \in A_{n,k}} \sum_{j \neq i_1} G(j;i) \\ &\le \hat{c}_n + \sum_{k=m}^{n-1} \max_{i \in A_{n,k}} \left(\sum_{j \neq i_1} \frac{G(j;i)}{G(i)} \right) Y_{n,k} \\ &\le \hat{c}_n + \sum_{k=m}^{n-1} \hat{a}_k Y_{n,k}, \end{aligned}$$

where $\hat{c}_n \ge 0$ is, as before, the contribution from non-principal successors of constant $i \in \{1,2,3\}^n$, and the last inequality simply uses (4.59):

$$\max_{i \in A_{n,k}} \left(\sum_{j \neq i_1} \frac{G(j;i)}{G(i)} \right) \le \hat{a}_k.$$

Similarly, to prove (4.61), by analogy with the proof of (4.28), we have, for any $v \in \mathcal{V} \cup \{m\} \setminus \{\circledast\}$ and $n \ge m+1$,

$$Y_{n+1,v} = \sum_{j=1}^{3} \sum_{y \in S_j^{-1}(v)} \sum_{i \in A_{n,y}} G(\sigma_i^{-1}(j); i).$$

Moreover, for each $j = 1,2,3$ and $i \in A_{n,y}$ such that $S_j(y) = v$, first applying (4.32), i.e.,

$$\frac{F\big(\sigma_i^{-1}(j); i\big)}{F(i)} \le B_{v,y},$$

and then (4.51) from Lemma 36 in turn gives the following:

$$\begin{aligned}\frac{G\big(\sigma_i^{-1}(j);i\big)}{G(i)} &= \frac{F\big(\sigma_i^{-1}(j);i\big)}{F(i)}\frac{\operatorname{diam}\big(\phi_{\sigma_i^{-1}(j)}\Delta_i\big)^{1-\delta}}{\operatorname{diam}(\Delta_i)^{1-\delta}}\\ &\leq B_{v,y}\frac{\operatorname{diam}(\phi_{\sigma_i^{-1}(j)}\Delta_i)^{1-\delta}}{\operatorname{diam}(\Delta_i)^{1-\delta}}\\ &\leq B_{v,y}\max_{x\in\Delta_i}\big(2-x_{\sigma_i^{-1}(j)}\big)^{\lambda(1-\delta)}\\ &\leq B_{v,y}\max_{x\in\Delta_y}(2-x_j)^{\lambda(1-\delta)}\\ &=: \hat{B}_{v,y}.\end{aligned}$$

(4.61) then follows as in the proof of (4.28):

$$\begin{aligned}Y_{n+1,v} &\leq \sum_{j=1}^{3}\sum_{y\in S_j^{-1}(v)}\hat{B}_{v,y}\sum_{i\in A_{n,y}}G(i)\\ &= \sum_{j=1}^{3}\sum_{y\in S_j^{-1}(v)}\hat{B}_{v,y}Y_{n,y} = \sum_{y\in\mathcal{V}}\hat{B}_{v,y}Y_{n,y},\end{aligned}$$

as required. □

Finally, we now conclude the proof of the main theorem of this chapter.

Proof of Theorem 12. Following verbatim the proof of Theorem 11 on page 155 (with a generous application of circumflex accents), the results of Lemmas 37 and 38 apply to give the following renewal-style inequality for $(Y_{n,\circledast})_{n\geq m+1}$: for all $n \geq m+1$,

$$Y_{n+1,\circledast} \leq \sum_{k=1}^{n-m}\hat{\lambda}_k Y_{n+1-k,\circledast} + \hat{\mathcal{E}}_n,$$

where, similarly to before,

$$\hat{\lambda}_k := \sum_{\substack{i+j=k\\ i,j\geq 0}}\big(\hat{B}^i\big)_{m,\circledast}\hat{a}_{m+j}\prod_{l=m}^{m+j-1}\hat{b}_l$$

and

$$\hat{\mathcal{E}}_n := \hat{c}_n + \sum_{k=m}^{n-1}\hat{a}_k\prod_{i=m}^{k-1}\hat{b}_i\sum_{\substack{y\in\mathcal{V}\\ y\neq\circledast}}\big(\hat{B}^{n-k-1}\big)_{m,y}Y_{m+1,y}.$$

Therefore, if $\delta > \varepsilon_m$ and $\delta > \frac{1}{3}$, we have both that $\sum_{n=1}^{\infty} \hat{c}_n < \infty$ and

$$\sum_{k=1}^{\infty} \hat{\lambda}_k = \sum_{k=0}^{\infty} \big(\hat{B}^k\big)_{m,\circledast} \sum_{k=m}^{\infty} \hat{a}_k \prod_{i=m}^{k-1} \hat{b}_i < 1 \tag{4.62}$$

(since the functions on the left hand side of the inequality are all decreasing in δ).

Consider the summability of the remainder terms, $\hat{\mathcal{E}}_n$. Since $\hat{B}_{v,y} = 0 \iff B_{v,y} = 0$, the proof of Lemma 32 applies to show that $\big(\hat{B}^k\big)_{m,\circledast}$ is asymptotic to a positive multiple of $\rho(\hat{B})^n$, and hence the convergence of (4.62) implies $\rho(\hat{B}) < 1$. Furthermore, each component of B^k is bounded above by a multiple of $\rho(\hat{B})^k$. Using this, noting that $\hat{a}_k \le a_k$ and $\hat{b}_k \le b_k$ for any $\delta \le 1$ and $k \ge 1$, and recalling (4.39), i.e.,

$$a_k \prod_{i=m}^{k-1} b_i = \mathcal{O}\big((k+2)^{-3\delta}\big)$$

as $k \to \infty$, there exist $C, C' > 0$ such that

$$\begin{aligned}
\sum_{n=m}^{\infty} \hat{\mathcal{E}}_n - \hat{c}_n &= \sum_{n=m}^{\infty} \sum_{k=m}^{n-1} \hat{a}_k \prod_{i=m}^{k-1} \hat{b}_i \sum_{\substack{y \in \mathcal{V} \\ y \ne \circledast}} \big(\hat{B}^{n-k-1}\big)_{m,y} Y_{m+1,y} \\
&\le C \sum_{n=m}^{\infty} \sum_{k=m}^{n-1} a_k \prod_{i=m}^{k-1} b_i \; \rho(\hat{B})^{n-k} \\
&\le C' \sum_{k=m}^{n-1} (k+2)^{-3\delta} \rho(\hat{B})^{n-k} \\
&\le \frac{C' \rho(\hat{B})}{1-\rho(\hat{B})} \sum_{k=m}^{\infty} (k+2)^{-3\delta},
\end{aligned}$$

which is finite since $\delta > \frac{1}{3}$. Hence, $\sum_{n=1}^{\infty} \hat{\mathcal{E}}_n < \infty$, since $\sum \hat{c}_n < \infty$ for $\delta > \frac{1}{3}$ (recalling that c_n is asymptotic to a multiple of $n^{-3\delta}$ as $n \to \infty$, from Lemma 37). Thus the renewal theorem of Lemma 30 gives $Y_{n,\circledast} \to 0$.

Finally, recalling $(1, 2^m; i) = (1, 2, \cdots, 2, i_1, ..., i_n) \in A_{n+m+1,\circledast}$, repeating the proof of (4.33) on page 155 gives

$$Y_{n+2,\circledast} \ge \sum_{i \in \{1,2,3\}^n} G(1, 2^m; i) \ge \min_{i \in \{1,2,3\}^n} \left(\frac{G(1, 2^m; i)}{G(i)} \right) Y_n.$$

Considering this minimum, recalling (4.34):

$$\min_{i\in\{1,2,3\}^n}\left(\frac{\operatorname{area}\left(\phi_1\phi_2^m(\Delta_i)\right)^\delta}{\operatorname{area}(\Delta_i)^\delta}\right) \geq (2m+2)^{-3\delta},$$

and applying the lower bound of (4.51) from Lemma 36 $m+1$ times, we have

$$\begin{aligned}\frac{G(1,2^m;i)}{G(i)} &= \left(\frac{\operatorname{area}\left(\phi_1\phi_2^m(\Delta_i)\right)}{\operatorname{area}(\Delta_i)}\right)^\delta \left(\frac{\operatorname{diam}\left(\phi_1\phi_2^m(\Delta_i)\right)}{\operatorname{diam}(\Delta_i)}\right)^{1-\delta}\\ &\geq (2m+2)^{-3\delta}\left(\frac{\operatorname{diam}\left(\phi_1\phi_2^m(\Delta_i)\right)}{\operatorname{diam}(\Delta_i)}\right)^{1-\delta}\\ &\geq (2m+2)^{-3\delta}\left(2^{\lambda-3}\right)^{(1-\delta)(m+1)}.\end{aligned}$$

That is, there exists a constant $C > 0$ such that, for each $n \in \mathbb{N}$,

$$Y_n \leq C\,Y_{n+m+1,\circledast} \to 0$$

as $n \to \infty$, as required. □

4.8 Conclusions and further work

Of all of the above chapters, it is probably this chapter which offers the most promise for future work.

Firstly, one might consider on how to improve the method in this context (particularly Lemma 36). Indeed, simple plots of $(Y_n)_{n=1}^{11}$ suggest that the infimal value for its convergence lies somewhere between 0.68 and 0.72, and so there is plenty of scope for improvement.

On the other hand, the main benefit of the method presented above seems to be its generality, which suggests potential applications to other fractal examples:

For example, it would be interesting to attempt to apply this method to the Apollonian gasket. In view of the result of Boyd, this method should yield both upper and lower bounds for the dimension, and the conformality of the attracting maps also means that the method should be more accurate, since the relationship between areas and diameters of level n triangles (or Soddy circles) is much more rigid (i.e., one does not have to rely on a loose result like Lemma 36). It would be interesting, in particular, to compare results that can be obtained by this method with the historical bounds. It has been noted that the basic case ($m = 1$) in the above seems broadly equivalent to the technique of *inducing* (see, e.g., [68] for an account).

Another natural candidate of study also presents itself in three dimensions. A collaboration was recently announced, between Ivan Dynnikov, Pascal Hubert, Paul Mercat, Olga Paris-Romaskevich and Alexandra Skripchenko [38]. They consider a self-projective fractal in the three-simplex, which they name the Novikov gasket, which they claim to show that has Hausdorff dimension strictly less than 3, following the method of Charles Fougeron in [47]. This gasket, which apparently arises in the context of Novikov's problem (see [75]), has not yet been publicly presented.

A more challenging task by far is to bound the dimension of the Rauzy gasket from below. It is presently unclear to the author if a renewal-style method can be used to provide bounds in this non-conformal setting.

Finally, as a closing remark on the thesis as a whole, we recall another (paraphrased) aphorism of Christopher Zeeman:

> "You earn a PhD by teaching your supervisor something new".

Appendix A

The proof of Lemma 18

We here prove the result on the eigenvalues of Hilbert-Schmidt operators which admit block triangular-form with respect to an orthogonal basis. For clarity, we recall the following definition.

Definition 22. [Block-triangular form] We say that a linear operator $\mathcal{C}$, acting on a Hilbert space $\mathcal{H}$ with orthogonal basis $\mathcal{B} = \{e_i\}_{i\in\mathcal{I}}$, has a *block-triangular form* (with respect to $\mathcal{B}$) if one has

$$\mathcal{H} = \bigoplus_{k\in\mathbb{Z}} D_k$$

such that, for each $k \in \mathbb{Z}$,

- D_k has a basis consisting of a finite (non-empty) subset of $\mathcal{B}$, and
- $\mathcal{C}(D_k) \subset \bigoplus_{j=k}^{\infty} D_j$.

We now prove the following lemma.

Lemma 18. *Suppose $\mathcal{C}$ and D_k are as in Definition 22, and suppose further that $\mathcal{C}$ is Hilbert-Schmidt. Then its non-zero eigenvalues are precisely the union of the eigenvalues for each finite rank operator $\mathcal{C}_k$ ($k \in \mathbb{Z}$):*

$$\mathcal{C}_k = \Pi_{D_k} \circ \mathcal{C} \circ \Pi_{D_k},$$

where Π_D denotes the orthogonal projection onto the subspace D.

Moreover, if a given non-zero eigenvalue of $\mathcal{C}$ is an eigenvalue of only one $\mathcal{C}_k$, then its algebraic and geometric multiplicities for these two operators coincide.

Proof of Lemma 18. In the notation of the lemma, for all $-\infty \leq N \leq M \leq \infty$, let

$$\Pi_N^M = \sum_{k=N}^{M} \Pi_{D_k}$$

denote the orthogonal projection onto the subspace $\bigoplus_{k=N}^M D_k$, and let $\mathcal{C}_N^M = \Pi_N^M \circ \mathcal{C} \circ \Pi_N^M$. Also, as before, denote the inner product and norm on $\mathcal{H}$ by $\langle \cdot, \cdot \rangle$, $\| \cdot \|$, respectively.

We now show, using that $\mathcal{C}$ is Hilbert-Schmidt, that the "lower-right concatenation", $\mathcal{C}_M^\infty$, converges to zero in operator norm, $\| \cdot \|_{\mathrm{op}}$, as $M \to \infty$. In particular, this uses the following basic fact ([28, p.267]): for any linear map $L : \mathcal{H} \to \mathcal{H}$,

$$\|L\|_{\mathrm{op}} \leq \|L\|_{\mathrm{HS}}. \tag{A.1}$$

By definition, letting $\hat{e}_i = e_i/\|e_i\|$ be the normalisation of e_i for convenience,

$$\mathcal{C}_M^\infty(\hat{e}_i) = \Pi_M^\infty \circ \mathcal{C} \circ \Pi_M^\infty(\hat{e}_i) = \begin{cases} 0, & \text{if } i \in D_n,\ n < M; \\ \sum_{k=M}^\infty \sum_{j \in \mathcal{I} : e_j \in D_k} \langle \mathcal{C}(\hat{e}_i), \hat{e}_j \rangle \, \hat{e}_j, & \text{if } i \in D_n,\ n \geq M. \end{cases}$$

Therefore, by Parseval's identity [28, Theorem 4.13],

$$\left\| \mathcal{C}_M^\infty(\hat{e}_i) \right\|^2 = \begin{cases} 0, & \text{if } i \in D_n,\ n \leq M; \\ \sum_{k=M}^\infty \sum_{j \in \mathcal{I} : e_j \in D_k} |\langle \mathcal{C}(\hat{e}_i), \hat{e}_j \rangle|^2, & \text{if } i \in D_n,\ n \geq M. \end{cases}$$

For each $i \in \mathcal{I}$, this is non-negative, bounded above by $\|\mathcal{C}(\hat{e}_i)\|^2$, and convergent to zero, by Bessel's inequality [28, p.15]. Since, by assumption,

$$\|\mathcal{C}\|_{\mathrm{HS}}^2 = \sum_{i \in \mathcal{I}} \|\mathcal{C}(\hat{e}_i)\|^2 < \infty,$$

the dominated convergence theorem implies that

$$\left\| \mathcal{C}_M^\infty \right\|_{\mathrm{op}}^2 \leq \left\| \mathcal{C}_M^\infty \right\|_{\mathrm{HS}}^2 = \sum_{i \in \mathcal{I}} \left\| \mathcal{C}_M^\infty(\hat{e}_i) \right\|^2 \to 0$$

as $M \to \infty$. A similar argument shows that $\|\mathcal{C}_{-\infty}^N\|_{\mathrm{op}} \to 0$ as $N \to -\infty$.

We now repeat the argument used earlier to obtain the spectrum of an expanding Blaschke product on the circle in section 3.1.1, following [30]. This time, since the corresponding matrix is bi-infinite, we need to apply the method twice. Let $\lambda \neq 0$ be such an eigenvalue, and take $N < M$ such that $\|\mathcal{C}_{-\infty}^N\|_{\mathrm{op}} < |\lambda|$

and $\|\mathcal{C}_M^\infty\|_{\mathrm{op}} < |\lambda|$. Then, for v an eigenvector corresponding to λ, decomposing $v = v_1 + v_2$, where

$$v_1 = \Pi_{-\infty}^N(v), \qquad v_2 = \Pi_{N+1}^\infty(v),$$

the eigenvector equation for v reads

$$\begin{pmatrix} \mathcal{C}_{-\infty}^N & \Pi_{-\infty}^N \circ \mathcal{C} \circ \Pi_{N+1}^\infty \\ \Pi_{N+1}^\infty \circ \mathcal{C} \circ \Pi_{-\infty}^N & \mathcal{C}_{N+1}^\infty \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \lambda \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}. \tag{A.2}$$

By the definition of block-triangularity, the upper-right operator in the matrix is zero:

$$\Pi_{-\infty}^N \circ \mathcal{C} \circ \Pi_{N+1}^\infty(\mathcal{H}) = \Pi_{-\infty}^N \circ \mathcal{C}\left(\bigoplus_{k=N+1}^\infty D_k\right) \subset \Pi_{-\infty}^N\left(\bigoplus_{k=N+1}^\infty D_k\right) = \{0\}. \tag{A.3}$$

Hence, the first component of (A.2) simply reads

$$\mathcal{C}_{-\infty}^N(v_1) = \lambda v_1.$$

Since $v_1 \neq 0$ would imply that $\|\mathcal{C}_{-\infty}^N\|_{\mathrm{op}} \geq |\lambda|$, a contradiction, we must have $v_1 = 0$, so that the second component of (A.2) reads

$$\mathcal{C}_{N+1}^\infty(v_2) = \lambda v_2,$$

i.e., λ is an eigenvalue of $\mathcal{C}_{N+1}^\infty$. Regarding the geometric multiplicity of λ, considering the generalised eigenvector equation,

$$(\mathcal{C} - \lambda I)^n v = \begin{pmatrix} \left(\mathcal{C}_{-\infty}^N - I\right)^n & 0 \\ * & \left(\mathcal{C}_{N+1}^\infty - I\right)^n \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0,$$

the first component, if $v_1 \neq 0$, implies that λ is a (generalised) eigenvector of $\mathcal{C}_{-\infty}^N$, which again contradicts the assumption that its operator norm is strictly less than $|\lambda|$. The second component then gives that v_2 is a generalised eigenvector for $\mathcal{C}_{N+1}^\infty$, thus showing that λ has the same multiplicity (algebraic and geometric) in the spectrum of $\mathcal{C}$ as in the spectrum of $\mathcal{C}_{N+1}^\infty$.

For the second iteration of this method, consider λ now as an eigenvalue of the adjoint of $\mathcal{C}_{N+1}^\infty$, $(\mathcal{C}_{N+1}^\infty)^*$, treated as an operator on $\bigoplus_{k=N+1}^\infty D_k$. Now, decomposing a corresponding eigenvector as $w = w_1 + w_2$, where

$$w_1 = \Pi_{N+1}^{M-1}(w), \qquad w_2 = \Pi_M^\infty(w),$$

and using block-triangularity as in (A.3) to show that $\Pi_{N+1}^{M-1} \circ \mathcal{C} \circ \Pi_M^{\infty} = 0$, the eigenvector equation for w reads

$$\begin{pmatrix} (\mathcal{C}_{N+1}^{M-1})^* & (\Pi_M^{\infty} \circ \mathcal{C} \circ \Pi_{N+1}^{M-1})^* \\ 0 & (\mathcal{C}_M^{\infty})^* \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \lambda \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}. \tag{A.4}$$

The second component of this equality again shows, if $w_2 \neq 0$, that λ is an eigenvalue of $(\mathcal{C}_M^{\infty})^*$, and hence that $|\lambda| > \|\mathcal{C}_M^{\infty}\|_{\mathrm{op}} = \|(\mathcal{C}_M^{\infty})^*\|_{\mathrm{op}} \geq |\lambda|$. Therefore $w_2 = 0$ and, from the first component of (A.3), λ is an eigenvalue of $(\mathcal{C}_{N+1}^{M-1})^*$, i.e., of $\mathcal{C}_{N+1}^{M-1}$. The extended argument applied in the first iteration also applies to show that λ has the same algebraic and geometric multiplicity in the spectra of $\mathcal{C}_{N+1}^{M-1}$ as in that of $\mathcal{C}_{N+1}^{\infty}$ and hence of $\mathcal{C}$.

To now consider this latter, finite concatenation $\mathcal{C}_{N+1}^{M-1}$, the block-triangularity of $\mathcal{C}$ gives this operator an according finite block-triangular form:

$$\mathcal{C}_{N+1}^{M-1} = \begin{pmatrix} \mathcal{C}_{N+1} & 0 & 0 & \cdots & 0 \\ * & \mathcal{C}_{N+2} & 0 & \cdots & 0 \\ * & * & \mathcal{C}_{N+3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ * & * & * & \cdots & \mathcal{C}_{M-1} \end{pmatrix},$$

where we recall the operators $\mathcal{C}_k = \Pi_{D_k} \circ \mathcal{C} \circ \Pi_{D_k}$ from the statement of the lemma, now considered as operators on D_k. Elementary linear algebra shows that λ is an eigenvalue of $\mathcal{C}_{N+1}^{M-1}$ if and only if it is an eigenvalue of one of the "blocks" $\mathcal{C}_k$; moreover, the algebraic multiplicity of λ in the spectrum of $\mathcal{C}_{N+1}^{M-1}$ (i.e., $\mathcal{C}$) is obtained by summing over the multiplicities in these blocks, and furthermore, if it is the eigenvalue of $\mathcal{C}_k$ for some unique k, then its geometric multiplicity is exactly that of $\mathcal{C}_k$. $\square$

Appendix B

Proofs of auxiliary results from chapter 4

B.1 The proof of Lemma 29

We now prove the following lemma on the values of the constants a_k and b_k (as well as $\tilde{a}_k$ and $\tilde{b}_k$), upon which all of the upper bounds for $\dim_H(\mathcal{G})$ in chapter 4 depend sensitively.

Lemma 29. *For all $v \in \mathcal{P}_k := R_k \cap \Delta_1$, we have the following tight bounds.*

$$\left(\frac{k+1}{k+2}\right)^{3\delta} \leq (2-v_1)^{-3\delta} \leq \left(\frac{k+2}{k+3}\right)^{3\delta}, \tag{4.12}$$

$$2^{3\delta+1}\left(\frac{k+2}{4k+7}\right)^{3\delta} \leq (2-v_2)^{-3\delta} + (2-v_3)^{-3\delta} \leq \left(\frac{k+1}{2k+1}\right)^{3\delta} + 2^{-3\delta}. \tag{4.13}$$

In particular, the maxima from Lemma 28 take the following values:

$$a_k = \left(\frac{k+1}{2k+1}\right)^{3\delta} + 2^{-3\delta}, \qquad b_k = \left(\frac{k+2}{k+3}\right)^{3\delta}.$$

Proof of Lemma 29. The first inequality (4.12) follows swiftly from a consideration of the vertices of $\mathcal{P}_k$. Considering the coordinates of these vertices,

$$\phi_1^k(e_2) = \left(\frac{k}{k+1}, \frac{1}{k+1}, 0\right), \qquad \phi_1^{k+1}(e_2) = \left(\frac{k+1}{k+2}, \frac{1}{k+2}, 0\right),$$
$$\phi_1^k(e_3) = \left(\frac{k}{k+1}, 0, \frac{1}{k+1}\right), \qquad \phi_1^{k+1}(e_3) = \left(\frac{k+1}{k+2}, 0, \frac{1}{k+2}\right),$$

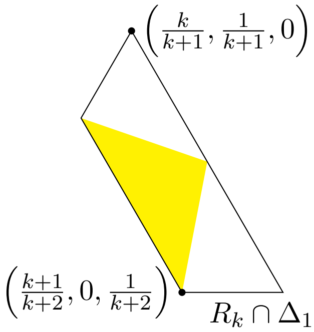


Figure B.1: Two maximisers giving the values of a_k and b_k in Lemma 29.

one obtains a convenient definition for $\mathcal{P}_k$:

$$\mathcal{P}_k = R_k \cap \Delta_1 = \left\{ x \in \Delta \;\middle|\; \frac{k+1}{k+2} \leq x_1 \leq \frac{k+2}{k+3} \right\}.$$

Hence,

$$\left(\frac{k+1}{k+2}\right)^{3\delta} = \left(2 - \frac{k}{k+1}\right)^{-3\delta} \leq (2 - v_1)^{-3\delta} \leq \left(2 - \frac{k}{k+1}\right)^{-3\delta} = \left(\frac{k+1}{k+2}\right)^{3\delta}.$$

The second inequality (4.13) can be deduced from properties of the function, $f : \Delta \to \mathbb{R}$, which we are maximising:

$$f(v) = (2 - v_2)^{-3\delta} + (2 - v_3)^{-3\delta}.$$

It is clear from the formula that f takes its global minimum at e_1 and that f is symmetric in the second and third coordinates. We can also see that f is convex in the sense that, for any $x, y \in \Delta$ and $t \in [0, 1]$,

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

This follows since its two terms are convex (as inverse powers of linear maps), and the sum of any two convex functions is convex.

We apply the convexity of f in two different ways to obtain each bound of (4.13). For the upper bound, note that if $v \in \mathcal{P}_k$, then in particular, v lies in the triangle with vertices e_1, $\left(\frac{k}{k+1}, \frac{1}{k+1}, 0\right)$ and $\left(\frac{k}{k+1}, 0, \frac{1}{k+1}\right)$. That is, there exists $\lambda \in [0, 1]^3$ such that $\lambda_1 + \lambda_2 + \lambda_3 = 1$ and

$$v = \lambda_1 e_1 + \lambda_2 \left(\frac{k}{k+1}, \frac{1}{k+1}, 0\right) + \lambda_3 \left(\frac{k}{k+1}, 0, \frac{1}{k+1}\right).$$

Applying the definition of convexity twice to this expression gives

$$
\begin{aligned}
f(v) &\le \lambda_1 f(e_1) + \lambda_2 f\left(\frac{k}{k+1}, \frac{1}{k+1}, 0\right) + \lambda_3 f\left(\frac{k}{k+1}, 0, \frac{1}{k+1}\right) \\
&\le \max\left(f(e_1), f\left(\frac{k}{k+1}, \frac{1}{k+1}, 0\right), f\left(\frac{k}{k+1}, 0, \frac{1}{k+1}\right)\right)
\end{aligned}
$$

But then, since $f(e_1) \le f\left(\frac{k}{k+1}, \frac{1}{k+1}, 0\right) = f\left(\frac{k}{k+1}, 0, \frac{1}{k+1}\right)$, this simplifies to the required upper bound:

$$
\begin{aligned}
f(x) \le f\left(\frac{k}{k+1}, \frac{1}{k+1}, 0\right) &= \left(2 - \frac{1}{k+1}\right)^{-3\delta} + 2^{-3\delta} \\
&= \left(\frac{k+1}{2k+1}\right)^{3\delta} + 2^{-3\delta}.
\end{aligned}
$$

To prove the lower bound of (4.13), note that, as illustrated in Figure B.2, for each $v \in \mathcal{P}_k$, the convex hull of v, $v' = (v_1, v_3, v_2)$ and e_1 contains the point in $\mathcal{P}_k$ closest to e_1,

$$
\left(\frac{k+1}{k+2}, \frac{1}{2(k+2)}, \frac{1}{2(k+2)}\right),
$$

and convexity thus gives that f takes its maximal value on one of the vertices. Since $f(v) = f(v') \ge f(e_1)$, this gives the required lower bound:

$$
f(v) \ge f\left(\frac{k+1}{k+2}, \frac{1}{2(k+2)}, \frac{1}{2(k+2)}\right) = 2\left(2 - \frac{1}{2k+4}\right)^{-3\delta} = 2^{3\delta+1}\left(\frac{k+1}{4k+7}\right)^{3\delta}.
$$

□

B.2 The proof of Lemma 31

In view of the definitions of S_j and $\mathcal{V}$, $\mathcal{V}_k$ on pages 39–40, we prove the following basic lemma, which considers some simple properties of the S_j.

Lemma 31. *The S_j ($j = 1, 2, 3$) are given by the following formulae: Firstly,*

$$
S_1(v) = \begin{cases} m, & \text{if } v \in \mathcal{V}_k; \\ (1, v_1, \ldots v_m), & \text{otherwise.} \end{cases}
$$

Secondly,

$$
S_2(v) = \tau \cdot (2, v_1, \ldots, v_m),
$$

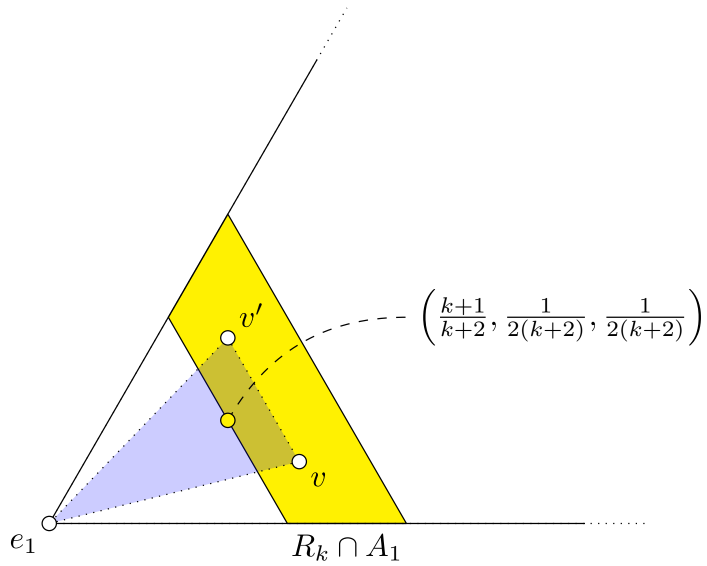


Figure B.2: Sketch of the proof of the lower bound for (4.13).

where

$$\tau := \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

is the transposition interchanging 1 and 2, e.g., $\tau \cdot (2,1,2,3) = (1,2,1,3)$. *Similarly,*

$$S_3(v) = \kappa \cdot (3, v_1, \ldots, v_3),$$

where

$$\kappa := \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$$

is the cycle taking 3 *to* 1 *and* 1 *to* 2*, e.g.,* $\kappa \cdot (3,1,2,3) = (1,2,3,1)$.

Moreover, $j \mapsto S_j(v)$ *is injective for any* $v \in \mathcal{V}$*, and for each* $v, y \in \mathcal{V} \setminus \{\circledast\}$*, there exists a tuple* $j \in \{1,2,3\}^{m+1}$ *such that*

$$v = S_{j_1} S_{j_2} \cdots S_{j_{m+1}}(y).$$

Proof of Lemma 31. We show the properties in order. The proof is elementary and direct.

To prove the explicit formulae for S_j, fix $v \in \mathcal{V}$ and write $v = (v_1, \ldots, v_{m+1})$, where

$$v_1 = v_2 = \cdots = v_k = 1, \qquad v_{k+1} = 2$$

for some $k \in \{1, 2, 3, \ldots, m-1\}$ (i.e., $v \in \mathcal{V}_k$). Also recall the definition of $S_j(v)$:

$$(j; v) \in A_{m+2, S_j(v)} \qquad \Longleftrightarrow \qquad (j, v_1, \ldots, v_m) \in [S_j(v)]. \tag{B.1}$$

First consider $j = 1$. As noted in the definition of the S_j, if $v \in \mathcal{V}_{m-1}$, then $(1; v) \in A_{m+2,m}$, i.e., $S_1(v) = m$. Otherwise (if $k < m-1$), $(1; v)$ trivially matches

$$(1, v_1, \ldots, v_m) \in \mathcal{V}_{k+1}.$$

For $j = 2$, 2 appears before 1, and 1 before 3 in $(2; v) = (2, 1, \ldots)$, and hence $\tau \cdot (2; v) = (1, 2, \ldots, v_{m+1})$ matches

$$\tau \cdot (2, v_1, \ldots, v_m) = (1, 2, \ldots, \tau(v_m)) \in \mathcal{V}_1.$$

Similarly for $j = 3$, 3 appears before 1 before 2 in $(3; v) = (3, 1, \ldots, 1, 2, \ldots, v_{m+1})$, and thus matches

$$\kappa \cdot (3, v_1, \cdots, v_m) = (1, 2, \ldots, 2, 3, \ldots, \kappa(v_m)) \in \mathcal{V}_1.$$

In view of (B.1), this proves the required formulae.

To prove the injectivity claim, again consider $v \in \mathcal{V}_k$ above, so that $v_{k+1} = 2$. Then, from the formulae above,

- $S_1(v) = m$ or $\big(S_1(v)\big)_{k+2} = v_{k+1} = 2$,
- $\big(S_2(v)\big)_{k+2} = \tau \cdot v_{k+1} = 1$, and
- $\big(S_3(v)\big)_{k+2} = \kappa \cdot v_{k+1} = 3$.

This difference in the $(k+1)$-st digit is sufficient to establish that $\{S_\omega(v)\}_{\omega=1}^3$ are distinct, as required.

Regarding the final remark in the lemma, fix $v, y \in \mathcal{V} \setminus \{\circledast\}$. Let $j \in [v]$ be such that $j_{m+1} \neq y_1$. Then, since $v \neq \circledast$, j has at most $m-1$ consecutively equal entries. This implies that, since $j_{m+1} \neq y_1$, the $(2m+2-k)$-tuple

$$(j_k, j_{k+1}, \ldots, j_{m+1}, y_1, y_2, \ldots, y_{m+1})$$

has at most the first $\min(m+2-k, m-1)$ entries all equal to j_k, and hence lies in

$\bigcup_{v\in\mathcal{V}} A_{2m+2-k,v}$, for each $1 \leq k \leq m+1$. Since this tuple matches v when $k=1$, this gives that there is some $j' \in \{1,2,3\}^{m+1}$ such that

$$v = S_{j'_1} S_{j'_2} \cdots S_{j'_{m+1}}(y),$$

as required. $\square$

B.3 The proof of Lemma 32

Applying the previous lemma, the following result is necessary to control the remainder terms in in the proof of Theorem 11. The statement and proof also apply verbatim to $\hat{B}$ as in the statement of Theorem 12.

Lemma 32. *There exists a non-negative matrix, $D \in \mathbb{R}^{\mathcal{V}\cup\{m\}\times\mathcal{V}\cup\{m\}}$ such that, for each $v, y \in \mathcal{V}\cup\{m\}$,*

$$\left(B^n\right)_{v,y} \rho(B)^{-n} \to D_{v,y}$$

as $n \to \infty$. Moreover, $D_{v,y} > 0$ if and only if $v \neq \circledast$ and $y \neq m$.

Proof of Lemma 32. Consider B as above. By definition, the column $(B_{v,m})_{v\in\mathcal{V}\cup\{m\}}$ is the zero vector. Moreover, the row $(B_{\circledast,v})_{v\in\mathcal{V}\cup\{m\}}$ is also the zero vector (i.e., there is no $v \in \mathcal{V}$ such that $S_j(v) = \circledast$: see the proof of (4.27) in Lemma 33). Therefore, for all $n \in \mathbb{N}$ and $v \in \mathcal{V}\cup\{m\}$,

$$(B^n)_{\circledast,v} = (B^n)_{v,m} = 0.$$

More explicitly, writing $\breve{B}$ as the submatrix of B on $\mathcal{V}\setminus\circledast \times \mathcal{V}\setminus\circledast$ (i.e., removing two rows and two columns from B), we can write B in the following block form, taking $\circledast$ and m to be the first and last indices of $\mathcal{V}\cup\{m\}$, respectively:

$$B = \left(\begin{array}{c|ccc|c} 0 & 0 & \cdots & 0 & 0 \\ \hline & & & & 0 \\ B_{\bullet,\circledast} & & \breve{B} & & \vdots \\ & & & & 0 \\ \hline B_{m,\circledast} & & B_{m,\bullet} & & 0 \end{array}\right),$$

where we have defined the row vector $B_{m,\bullet} := (B_{m,v})_{v\in\mathcal{V}\setminus\{\circledast\}}$ and the column vector

$B_{\bullet,\circledast} := (B_{v,\circledast})_{v\in\mathcal{V}\setminus\{\circledast\}}$. By induction, for each $n \in \mathbb{N}$,

$$B^n \quad = \quad \left(\begin{array}{c|ccc|c} 0 & 0 & \cdots & 0 & 0 \\ \hline & & & & 0 \\ \breve{B}^{n-1} \cdot B_{\bullet,\circledast} & & \breve{B}^n & & \vdots \\ & & & & 0 \\ \hline (B^n)_{m,\circledast} & & B_{m,\bullet} \cdot \breve{B}^{n-1} & & 0 \end{array}\right) \tag{B.2}$$

(the dot denoting matrix multiplication). In particular, $(B^n)_{v,y} = (\breve{B}^n)_{v,y}$ for all $v, y \in \mathcal{V} \setminus \{\circledast\}$ and $n \in \mathbb{N}$.

We now show that $\breve{B}^{m+1}$ is a positive matrix, i.e., $(B^{m+1})_{v,y} > 0$ for all $v, y \in \mathcal{V} \setminus \{\circledast\}$. Fixing such a pair of indices, Lemma 31 provides a tuple $j \in \{1, 2, 3\}^{m+1}$ such that

$$v = S_{j_1} S_{j_2} \cdots S_{j_{m+1}}(y)$$

(in particular, $S_{j_k} \cdots S_{j_{m+1}}(y) \in \mathcal{V} \setminus \{\circledast\}$ for each $1 \leq k \leq m+1$). Therefore,

$$(\breve{B}^{m+1})_{v,y} \geq (\breve{B}^{m+1})_{v,y} \geq \prod_{n=1}^{m-1} B_{S_{j_k} S_{j_{k+1}} \cdots S_{j_{m+1}}(y), S_{j_{k+1}} \cdots S_{j_{m+1}}(y)} > 0,$$

so $\breve{B}^{m+1}$ is positive, as claimed. The Perron–Frobenius theorem [97, Theorem 2.7] then states that the spectral radius $\rho(\breve{B})$ of $\breve{B}$ is a positive eigenvalue of $\breve{B}$, that this eigenvalue is simple, and that it is the only eigenvalue of $\breve{B}$ with this absolute value.

In particular, there exists a positive, rank-one matrix $D \in \mathbb{R}^{\mathcal{V}\setminus\circledast \times \mathcal{V}\setminus\circledast}$ such that, as $n \to \infty$,

$$\left(\frac{\breve{B}^n}{\rho(\breve{B})^n}\right) \to D,$$

i.e., for each $v, y \in \mathcal{V} \setminus \{\circledast\}$,

$$(B^n)_{v,y} \rho(\breve{B})^{-n} = (\breve{B}^n)_{v,y} \rho(\breve{B})^{-n} \to D_{v,y} > 0 \tag{B.3}$$

as $n \to \infty$.

We now simply extend this formula. First, for any $v \in \mathcal{V} \setminus \{\circledast\}$ and $n \in \mathbb{N}$, in view of (B.2) we have, by linearity,

$$\rho(\breve{B})^{-n} (B^{n+1})_{v,\circledast} = \rho(\breve{B})^{-n} (\breve{B}^n \cdot B_{\bullet,\circledast})_v \to \big(D \cdot B_{\bullet,\circledast}\big)_v,$$

as $n \to \infty$, and similarly,

$$\rho(\breve{B})^{-n}(B^{n+1})_{v,\circledast} = \rho(\breve{B})^{-n}(B_{m,\bullet} \cdot \breve{B}^n)_v \to \big(B_{m,\bullet} \cdot D\big)_v.$$

Finally, for each $n \in \mathbb{N}$,

$$\rho(\breve{B})^{-n}(B^{n+2})_{m,\circledast} = \rho(\breve{B})^{-n}\left(B_{m,\bullet} \cdot B^n \cdot B_{\bullet,\circledast}\right) \to B_{m,\bullet} \cdot D \cdot B_{\bullet,\circledast}$$

as $n \to \infty$. We thus extend the definition of D by the following, for $v \in \mathcal{V} \setminus \circledast$:

- $D_{v,\circledast} := \rho(\breve{B})^{-1}\big(D \cdot B_{\bullet,\circledast}\big)_{v,\circledast}$.
- $D_{m,v} := \rho(\breve{B})^{-1}\big(D \cdot B_{m,\bullet}\big)_{v,\circledast}$.
- $D_{m,\circledast} := \rho(\breve{B})^{-2}\big(B_{m,\bullet} \cdot D \cdot B_{\bullet,\circledast}\big)_{m,\circledast}$.
- $D_{v,m} = D_{\circledast,v} = D_{\circledast,m} = 0$.

From our previous considerations, it is clear that (B.3) extends to all $v, y \in \mathcal{V} \cup \{m\}$. Moreover, that these first three bullet points are positive follows from the fact that D is positive matrix, and that $B_{\bullet,\circledast}$ and $B_{m,\bullet}$ are non-zero and non-negative vectors (i.e., by the definition of B, $\{B_{m,v}\}_{v \in \mathcal{V} \cup \{m\}}$ and $\{B_{v,\circledast}\}_{v \in \mathcal{V} \cup \{m\}}$ each have three positive entries, and at most one of these is shared).

Finally, since each entry of B^n is asymptotic to a multiple of $\rho(\breve{B})^n$, it follows that $\rho(B) = \rho(\breve{B})$, completing the proof. □

B.4 The proof of Lemma 36

In the final part of this appendix, we prove the following mildly technical lemma which is needed to augment the proofs of Proposition 17 and Theorem 11 in proving Proposition 22 and Theorem 12.

Lemma 36. *Treating Δ as an immersed manifold in $\mathbb{R}^3$, the tangent map $D_v\phi_j : \mathcal{T}_v\Delta \to \mathcal{T}_{\phi_j(v)}\Delta$ has maximal singular value satisfying*

$$\|D_v\phi_j\|_{\mathrm{op}} \leq (2 - v_j)^{-\lambda}, \tag{4.50}$$

where $\lambda := \frac{3}{2} - \frac{1}{\sqrt{3}} = 0.9226\ldots$. Consequently, for all $i \in \{1,2,3\}^n$,

$$2^{\lambda-3} \leq \left(\frac{\mathrm{diam}\left(\phi_j(\Delta_i)\right)}{\mathrm{diam}(\Delta_i)}\right) \leq \max_{v \in \Delta_i}(2 - v_j)^{-\lambda}. \tag{4.51}$$

Proof of Lemma 36. Since the metric on Δ is invariant under coordinate permutations, it suffices to prove (4.51) for $j = 1$, which we do using a computer-aided calculation. We extend the definition

$$\phi := \phi_1 : (x, y, z) \mapsto \frac{1}{2-x}(1, y, z)$$

to a map $\hat{\phi} : (0, \infty)^3 \to (0, \infty)^3$. Then, since the orthogonal projection

$$v \mapsto v \cdot \left(0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)\left(0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) + v \cdot \left(\sqrt{\frac{2}{3}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right)\left(\sqrt{\frac{2}{3}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right)$$

gives an isometric embedding from Δ into $\{(x, y, z) \mid x + y + z = 0\}$, the map

$$(x, y, z) \mapsto \begin{pmatrix} 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\sqrt{\frac{2}{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix} \cdot (x, y, z),$$

is an isometric embedding of Δ into $\mathbb{R}^2$. Therefore, up to congruency, we have the following matrix representation for $D_{(x,y,z)}\phi$, which respects the metric on Δ:

$$D_{(x,y,z)}\phi = \begin{pmatrix} 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\sqrt{\frac{2}{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix} D_{(x,y,z)}\hat{\phi} \begin{pmatrix} 0 & -\sqrt{\frac{2}{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{pmatrix}.$$

By a direct computation,

$$D_{(x,y,z)}\hat{\phi} = \frac{1}{(2-x)^2} \begin{pmatrix} 1 & 0 & 0 \\ y & 2-x & 0 \\ z & 0 & 2-x \end{pmatrix},$$

and hence

$$D_{(x,y,z)}\phi = \frac{1}{(2-x)^2} \begin{pmatrix} 2-x & \frac{1}{\sqrt{3}}(y-z) \\ 0 & 1 \end{pmatrix}.$$

Thus,

$$(D_{(x,y,z)}\phi)(D_{(x,y,z)}\phi)^T = \frac{1}{(2-x)^4} \begin{pmatrix} (2-x)^2 + \frac{1}{3}(y-z)^2 & \frac{y-z}{\sqrt{3}} \\ \frac{y-z}{\sqrt{3}} & 1 \end{pmatrix},$$

and therefore the maximal singular value of $D_{(x,y,z)}\phi$, $\|D_{(x,y,z)}\phi\|_{\mathrm{op}}$, is the square root of the largest eigenvalue of this last matrix, which simplifies to

$$\frac{(2-x)^{-2}}{\sqrt{3}}\Bigg(8-7x-2y+2x^2+2xy+2y^2 \\ +2\sqrt{\left(7-5x-y+x^2+xy+y^2\right)\left(1-2x-y+x^2+xy+y^2\right)}\Bigg)^{1/2}.$$

From this formula, with the assistance of a computer, one can verify that

$$\max_{(x,y,z)\in\Delta}\frac{\|D_{(x,y,z)}\phi\|_{\mathrm{op}}}{(2-x)^{-\lambda}}=1,$$

which is attained at $(1,0,0)$. This proves (4.50), and shows that it is tight.

We now deduce (4.51) using the mean value theorem: Let $\|\cdot\|_2$ denote the Euclidean norm. Then, given $v,w\in\Delta$, there exists y on the line segment $[v,w]$ joining v to w, such that

$$\phi_j(w)-\phi_j(v)=D_y\phi_j\left(\frac{(w-v)}{\|w-v\|_2}\right)\cdot(w-v).$$

Therefore,

$$\|\phi_j(w)-\phi_j(v)\|_2\leq\max_{y\in[v,w]}\|D_y\phi_j\|_{\mathrm{op}}\|w-v\|_2. \tag{B.4}$$

Similarly, we have a lower bound in terms of the minimum singular value of $D_y\phi_j$:

$$\mathrm{msv}(D_y):=\min_{\|z\|_2=1}\frac{\|D_y(z)\|_2}{\|z\|_2}$$

as follows:

$$\|\phi_j(w)-\phi_j(v)\|_2\geq\min_{y\in[v,w]}\left(\mathrm{msv}(D_y\phi_j)\right)\|w-v\|_2. \tag{B.5}$$

To give the minimum singular value a more explicit form: since the singular eigenvalues of a matrix multiply to give its determinant, we have that

$$\mathrm{msv}(D_y\phi_j)\|D_y\phi_j\|_{\mathrm{op}}=\mathrm{Jac}_{\phi_j}(y),$$

i.e.,

$$\mathrm{msv}(D_y\phi_j)\geq\frac{\mathrm{Jac}_{\phi_j}(y)}{\|D_y\phi_j\|_{\mathrm{op}}}\geq\frac{(2-y_j)^{-3}}{(2-y_j)^{-\lambda}}\geq(2-y_j)^{\lambda-3}\geq 2^{\lambda-3}. \tag{B.6}$$

Now let $i \in \{1,2,3\}^n$. Applying (B.4) yields the upper bound of (4.51) as follows:

$$\begin{aligned}
\operatorname{diam}\big(\phi_j(\Delta_i)\big) &= \sup\big\{\|\phi_j(v)-\phi_j(w)\|_2 : v,w \in \Delta_i\big\}\\
&\le \sup\big\{\max_{y\in[v,w]} \|D_y\phi_j\|_{\mathrm{op}}\|v-w\|_2 : v,w\in\Delta_i\big\}\\
&\le \max_{y\in\Delta_i}\|D_y\phi_j\|_{\mathrm{op}}\sup\big\{\|v-w\|_2 : v,w\in\Delta_i\big\}\\
&\le \max_{y\in\Delta_i}(2-y_j)^{-\lambda}\operatorname{diam}(\Delta_i).
\end{aligned}$$

Similarly for the lower bound, applying (B.5) and (B.6) gives the following:

$$\begin{aligned}
\operatorname{diam}\big(\phi_j(\Delta_i)\big) &= \sup\big\{\|\phi_j(v)-\phi_j(w)\|_2 : v,w \in \Delta_i\big\}\\
&\ge \sup\big\{\min_{y\in[v,w]}\big(\operatorname{msv}(D_y\phi_j)\big)\|v-w\|_2 : v,w\in\Delta_i\big\}\\
&\ge 2^{\lambda-3}\sup\big\{\|v-w\|_2 : v,w\in\Delta_i\big\}\\
&\ge 2^{\lambda-3}\operatorname{diam}(\Delta_i).
\end{aligned}$$

This completes the proof of (4.51), and hence of Lemma 36. $\square$

Bibliography

[1] A. Adam. Generic non-trivial resonances for Anosov diffeomorphisms. *Nonlinearity* **30** (2017), no. 3, 1146–64. doi:10.1088/1361-6544/aa59a9

[2] R. L. Adler and L. Flatto. Uniform distribution of Kakutani's interval splitting procedure. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **38** (1977), no. 4, 253–259. doi:10.1007/BF00533157

[3] C. Aistleitner and M. Hofer. Uniform distribution of generalized Kakutani's sequences of partitions. *Ann. Mat. Pura Appl. (4)* **192** (2013), no. 4, 529–538. doi:10.1007/s10231-011-0235-9

[4] C. Aistleitner, M. Hofer and V. Ziegler. On the uniform distribution modulo 1 of multidimensional LS-sequences. *Ann. Mat. Pura Appl. (4)* **193**, no. 5 (2014), 1329–1344. doi:10.1007/s10231-013-0331-0

[5] V. I. Arnold and A. Avez. *Problèmes ergodiques de la mécanique classique.* Gauthier-Villars: Paris, 1967.

[6] P. Arnoux, and G. Rauzy. Représentation géométrique de suites de complexité 2n+1. *Bull. Soc. Math. France* **119** (1991), no. 2, 199–215. doi:10.24033/bsmf.2164

[7] P. Arnoux and Š. Starosta. *The Rauzy gasket.* In: Barral, J., Seuret, S. (eds) Further developments in fractals and related fields, pp.1–23. Birkhäuser/Springer: New York, 2013. doi:10.1007/978-0-8176-8400-6

[8] P. Arnoux, and J.-C. Yoccoz. Construction de difféomorphismes pseudo-Anosov. *C. R. Acad. Sci. Paris Sér. I Math.* **292** (1981), no. 1, 75–78.

[9] K. E. Atkinson. *An introduction to numerical analysis.* Second edition. John Wiley & Sons: New York, 1989.

[10] A. Avila, P. Hubert and A. Skripchenko. On the Hausdorff dimension of the Rauzy gasket. *Bull. Soc. Math. France* **144** (2016), no. 3, 539–568. doi:10.24033/bsmf.2722

[11] Z.-Q. Bai and S. R. Finch. Precise calculation of Hausdorff dimension of Apollonian gasket. *Fractals* **26** (2018), no. 4, 1850050–1850059. doi:10.1142/S0218348X18500500

[12] V. Baladi. *Dynamical zeta functions and dynamical determinants for hyperbolic maps: A functional approach.* A Series of Modern Surveys in Mathematics, 3rd Series, **68**. Springer: Cham, 2018. ISBN:978-3-319-77660-6

[13] V. Baladi. The quest for the ultimate anisotropic Banach space. *J. Stat. Phys.* **166** (2017), no. 3–4, 525–557. doi:10.1007/s10955-016-1663-0

[14] O. Bandtlow and O. Jenkinson. On the Ruelle eigenvalue sequence. *Ergodic Theory and Dynam. Systems* **28** (2008), no. 6, 1701–1711. doi:10.1017/S0143385708000059

[15] O. Bandtlow, W. Just and J. Slipantschuk. Spectral structure of transfer operators for expanding circle maps. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **34** (2017), no. 1, 31–43. doi:10.1016/j.anihpc.2015.08.004

[16] O. Bandtlow and F. Naud. Lower bounds for the Ruelle spectrum of analytic expanding circle maps. *Ergodic Theory Dynam. Systems* **39** (2019), no. 2, 289–310. doi:10.1017/etds.2017.29

[17] D. Blackwell. A renewal theorem. *Duke Math. J.* **15** (1948), 145–150. doi:10.1215/S0012-7094-48-01517-8

[18] D. W. Boyd. The residual set dimension of the Apollonian packing. *Mathematika* **20** (1973), 170–174.

[19] M. D. Brennan and R. Durrett. Splitting intervals. *Ann. Probab.* **14** (1986), no. 3, 1024–1036.

[20] M. D. Brennan and R. Durrett. Splitting intervals II: Limit laws for lengths. *Probab. Theory Related Fields* **75** (1987), no. 1, 109–127.

[21] I. Carbone. Discrepancy of LS-sequences of partitions and points. *Ann. Mat. Pura Appl. (4)* **191** (2012), no. 4, 819–844. doi:10.1007/s10231-011-0208-z

[22] I. Carbone. Extension of van der Corput algorithm to LS-sequences. *Appl. Math. Comput.* **255** (2015), 207–213. doi:10.1016/j.amc.2014.08.063

[23] I. Carbone. *Comparison between LS-sequences and β-adic van der Corput sequences.* In: Cools, R., Nuyens, D. (eds) Monte Carlo and quasi-Monte Carlo methods, 261–270. Springer Proc. Math. Stat., **163**. Springer: 2016. doi:10.1007/978-3-319-33507-0_11

[24] I. Carbone, M. R. Iacò and A. Volčič. Kakutani's splitting procedure in higher dimension. *Rend. Istit. Mat. Univ. Trieste* **39** (2007), 119–126.

[25] I. Carbone, M. R. Iacò and A. Volčič. A dynamical system approach to the Kakutani-Fibonacci sequence. *Ergodic Theory Dynam. Systems* **34** (2014), no. 6, 1794–1806. doi:10.1017/etds.2013.20

[26] L. Carleson and T. W. Gamelin. *Complex Analysis.* Universitext: Tracts in Math. Springer-Verlag: New York, 1993. doi:10.1007/978-1-4612-4364-9

[27] E. G. D. Cohen and G. Gallavotti. Dynamical Ensembles in Nonequilibrium Statistical Mechanics. *Phys. Rev. Lett.* **74** (1995), 2694–2697. doi:10.1103/PhysRevLett.74.2694

[28] J. B. Conway. *A course in functional analysis.* Second edition. Graduate Texts in Math. **96**. Springer-Verlag: New York, 1990. doi:10.1007/978-1-4757-4383-8

[29] C. Corduneanu. *Almost Periodic Functions.* Interscience Tracts in Pure and App. Math. **22**. Wiley Interscience: London, 1961.

[30] C. C. Cowen. *Composition Operators on Spaces of Analytic Functions II.* Lecture notes: Spring School of Functional Analysis, Rabat, 19–21 May 2009. math.iupui.edu/∼ccowen/Talks/CompOp0905slidesB.pdf

[31] N. G. de Bruijn and P. Erdős. Some linear and some quadratic recursion formulas. *I. Nederl. Akad. Wetensch. Proc. Ser. A.* ***54*** *= Indagationes Math.* **13** (1951), 374–382.

[32] R. De Leo. A conjecture on the Hausdorff dimension of attractors of real self-projective iterated function systems. *Exp. Math.* **24** (2015), no. 3, 270–288. doi:10.1080/10586458.2014.987884

[33] R. De Leo and I. A. Dynnikov. Geometry of plane sections of the infinite regular skew polyhedron $\{4, 6 \mid 4\}$. *Geom. Dedicata* **138** (2009), 51–67. doi:10.1007/s10711-008-9298-1

[34] M. Drmota and M. Infusino. On the discrepancy of some generalized Kakutani's sequences of partitions. *Unif. Distrib. Theory* **7** (2012), no. 1, 75–104. doi:10.1515/udt-2012-0014

[35] M. Drmota and R. Tichy. *Sequences, Discrepancies and Applications.* Lecture notes in math. **1651**. Springer: Berlin, 1997. doi:10.1007/BFb0093404

[36] M. F. Demers. A gentle introduction to anisotropic Banach spaces. *Chaos Solitons Fractals* **116** (2018), 29–42. doi:10.1016/j.chaos.2018.08.028

[37] R. Durrett. *Probability: theory and examples.* Fifth edition. Cambridge Series in Statistical and Probabilistic Mathematics **49**. Cambridge University Press: Cambridge, 2019. doi:10.1017/9781108591034

[38] I. Dynnikov, P. Hubert, P. Mercat, O. Paris-Romaskevich and A. Skripchenko. *Novikov gasket has zero Lebesgue measure.* Ongoing project, announced on romaskevich.carrd.co/#research

[39] I. A. Dynnikov, P. Hubert and A. Skripchenko. *Dynamical systems around the Rauzy gasket and their ergodic properties* (2020). Preprint: arXiv:2011.15043

[40] F. Ellison and W. Ellison. *Prime Numbers.* Wiley Interscience: London, 1985. ISBN:0-471-82653-7

[41] P. Erdős, W. Feller and H. Pollard. A property of power series with positive coefficients. *Bull. Amer. Math. Soc.* **55** (1949), 201–204. doi:10.1090/S0002-9904-1949-09203-0

[42] K. Falconer. *Fractal geometry: Mathematical foundations and applications.* Third edition. John Wiley & Sons: Chichester, 2014. ISBN:1-119-94239-X

[43] B. Farb and D. Margalit. *A primer on mapping class groups.* Princeton Mathematical Series, **49**. Princeton University Press: Princeton, 2012. ISBN:978-0-691-14794-9

[44] F. Faure and N. Roy. Ruelle-Pollicott resonances for real analytic hyperbolic maps. *Nonlinearity* **19** (2006), no. 6, 1233–1252. doi:10.1088/0951-7715/19/6/002

[45] W. Feller. *An introduction to probability theory and its applications vol. I.* Third edition. John Wiley & Sons, Ltd.: London 1968.

[46] N. P. Fogg. *Substitutions in dynamics, arithmetics and combinatorics.* Lecture Notes in Mathematics, **1794**. Springer-Verlag: Berlin, 2002. doi:10.1007/b13861

[47] C. Fougeron. *Dynamical properties of simplicial systems and continued fraction algorithms* (2020). Preprint: arXiv:2001.01367

[48] F. Faure, S. Gouëzel and E. Lanneau. Ruelle spectrum of linear pseudo-Anosov maps. *J. Éc. polytech. Math.* **6** (2019), 811–877. doi:10.5802/jep.107

[49] E. Fuchs. Counting problems in Apollonian packings. *Bull. Amer. Math. Soc.* **50** (2013), no. 2, 229–266. doi:10.1090/S0273-0979-2013-01401-0

[50] A. M. Garsia. Some Tauberian theorems and the asymptotic behavior of probabilities of recurrent events. *J. Math. Anal. Appl.* **7** (1963), 146–162. doi:10.1016/0022-247X(63)90084-2

[51] A. M. Garsia and J. Lamperti. A discrete renewal theorem with infinite mean. *Comment. Math. Helv.* **37** (1962), 221–234. doi:10.1007/BF02566974

[52] A. M. Garsia, S. Orey and E. Rodemich. Asymptotic behavior of successive coefficients of some power series. *Illinois J. Math.* **6** (1962), 620–629. doi:10.1215/ijm/1255632709

[53] R. Gutiérrez-Romo and C. Matheus. Lower bounds on the dimension of the Rauzy gasket. *Bull. Soc. Math. France* **148** (2020), no. 2, 321–327. doi:10.24033/bsmf.2807

[54] G. H. Hardy and J. E. Littlewood. Some problems of diophantine approximation II. *Acta Math.* **37** (1914), no. 1, 193–239. doi:10.1007/BF02401833

[55] M. Hofer, M. R. Iacò and R., Tichy. Ergodic properties of β-adic Halton sequences. *Ergodic Theory Dynam. Systems* **35** (2015), no. 3, 895–909. doi:10.1017/etds.2013.70

[56] P. Hubert and O. Paris-Romaskevich. Triangle tiling billiards and the exceptional family of their escaping trajectories: circumcenters and Rauzy gasket. *Exp. Math.* (2019). doi:10.1080/10586458.2019.1583615

[57] M. R. Iacò and V. Ziegler. Discrepancy of generalized LS-sequences. *Unif. Distrib. Theory* **12** (2017), no. 2, 37–63. doi:10.1515/udt-2017-0014

[58] S. Kakutani. A problem of equidistribution on the unit interval $[0, 1]$. *Measure theory (Proc. Conf., Oberwolfach, 1975)*. Lecture Notes in Math. **541**. Springer: Berlin 1976. 369–375.

[59] A. Kar. Weyl's Equidistribution theorem. *Resonance* **8** (2003), no. 5, 30–37.

[60] M. Keane Interval exchange transformations. *Math. Z.* **141** (1975), 25–31. doi:10.1007/BF01236981

[61] A. Kontorovich and H. Oh. Apollonian circle packings and closed horospheres on hyperbolic 3-manifolds. *J. Am. Math. Soc.* **24** (2011), 603–648. doi:10.1090/S0894-0347-2011-00691-7

[62] L. Kuipers and H. Niederreiter. *Uniform distribution of sequences.* Pure and Applied Mathematics. Wiley Interscience: New York-London-Sydney, 1974.

[63] G. Levitt. La dynamique des pseudogroupes de rotations. *Invent. Math.* **113** (1993), no. 3, 633–670. doi:10.1007/BF01244321

[64] P. Liardet. Gérard Rauzy (1938–2010). *Unif. Distrib. Theory* **7** (2012), no. 1, 1–9.

[65] J.-C. Lootgieter. Sur la répartition des suites de Kakutani I. *Ann. Inst. H. Poincaré Sect. B* **13** (1977), no. 4, 385–410.

[66] J.-C. Lootgieter. Sur la répartition des suites de Kakutani II. *Ann. Inst. H. Poincaré Sect. B* **14** (1978), no. 3, 279–302.

[67] J.-C. Lootgieter. Processus généraux de branchements et procédure de Kakutani généralisée. *C. R. Acad. Sci. Paris Sér. A-B* **288** (1979), no. 7, A427–A430.

[68] R. D. Mauldin and M. Urbański. Parabolic iterated function systems. *Ergodic Theory Dynam. Systems* **20** (2000), no. 5, 1423–1447.

[69] N. W. McLachlan. *Complex variable theory and transform calculus with technical applications.* Second edition. Cambridge University Press: New York, 1963.

[70] C. T. McMullen. Hausdorff Dimension and Conformal Dynamics, III: Computation of Dimension. *Am. J. Math.* **120** (1998), 691–721. doi:10.1353/ajm.1998.0031

[71] R. Meester and Th. Nowicki. Infinite clusters and critical values in two-dimensional circle percolation. *Israel J. Math.* **68** (1989), no. 1, 63–81. doi:10.1007/BF02764969

[72] F. Naud. The Ruelle spectrum of generic transfer operators. *Discrete Contin. Dynam. Systems* **32** (2012), no. 7, 2521–2531. doi:10.3934/dcds.2012.32.2521

[73] S. Newhouse. *Cone-fields, domination, and hyperbolicity.* In: Brin, M., Hasselblatt, B., Pesin, Ya. (eds) Modern dynamical systems and applications. pp.419–432. Cambridge Univ. Press: Cambridge, 2004.

[74] H. Niederreiter. Random number generation and quasi-Monte Carlo methods. *CBMS-NSF Regional Conference Series in Applied Mathematics,* **63**. SIAM: Philadelphia, 1992. ISBN:0-89871-295-5

[75] S. P. Novikov and A. Ya. Maltsev. Dynamical systems, topology and conductivity in normal metals. *J. of Stat. Physics* **115** (2004), no. 1–2, 31–46.

[76] M. Pollicott. *Apollonian Circle Packings.* In: Bandt, C., Falconer, K., Zähle, M. (eds) Fractal Geometry and Stochastics V, pp.121–142. Birkhäuser: Cham, 2015. doi:10.1007/978-3-319-18660-3_8

[77] E. R. Pujals and R. K. W. Roeder. Two-dimensional Blaschke products: degree growth and ergodic consequences. *Indiana Univ. Math. J.* **59** (2010), no. 1, 301–325.

[78] E. R. Pujals and M. Shub. Dynamics of two-dimensional Blaschke products. *Ergodic Theory Dynam. Systems* **28** (2008), no. 2, 575–85. doi:10.1017/S0143385707000752

[79] S. M. Robinson. A Short Proof of Cramer's Rule. *Math. Magazine* **43** (1970), no. 2. 94–95.

[80] W. Rudin. *Functional analysis.* Second edition. International Series in Pure and Applied Mathematics. McGraw-Hill: New York, 1991. ISBN:0-07-054236-8

[81] D. Ruelle. Zeta-functions for expanding maps and Anosov flows. *Invent. Math.* **34** (1976), 231-242 doi:10.1007/BF01403069

[82] D. Ruelle. Resonances of chaotic dynamical systems. *Phys. Rev. Lett.* **56** (1986), no. 5, 405–407. doi:10.1103/PhysRevLett.56.405

[83] P. Sarnak. Integral Apollonian packings. *Am. Math. Mon.* **118** (2011), 291–306.

[84] J. H. Shapiro. *Composition operators and classical function theory.* Universitext: Tracts in Mathematics. Springer-Verlag: New York, 1993. ISBN:0-387-94067-7

[85] K. R. Shrestha. Hardy spaces on the polydisk. *Eur. J. Pure Appl. Math.* **9** (2016), no. 3, 292–304.

[86] W. Sierpiński, Sur une courbe dont tout point est un point de ramification. *C. R. Acad. Sci. Paris* **160** (1915), 302–305.

[87] J. Slipantschuk, O. Bandtlow and W. Just. Analytic expanding circle maps with explicit spectra. *Nonlinearity* **26** (2013), no. 12, 3231–45. doi:10.1088/0951-7715/26/12/3231

[88] J. Slipantschuk, O. Bandtlow and W. Just. Complete spectral data for analytic Anosov maps of the torus. *Nonlinearity* **30** (2017), no. 7, 2667–86. doi:10.1088/1361-6544/aa700f

[89] E. Slud. Entropy and maximal spacing for random partitions. *Z. Wahrscheinlichkeitsthorie verw. Gebiete* **41** (1978), 341–352.

[90] E. Slud. Correction to "Entropy and maximal spacing for random partitions". *Z. Wahrscheinlichkeitsthorie verw. Gebiete* **60** (1982), 139–141.

[91] Y. Smilansky. Uniform Distribution of Kakutani Partitions Generated By Substitution Schemes. *Israel J. Math.* **240** (2020), 667–710. doi:10.1007/s11856-020-2075-z

[92] F. Soddy. The Kiss Precise. *Nature* **137** (1936), 1021.

[93] T. J. Sullivan. *Introduction to uncertainty quantification.* Texts in Applied Mathematics **63**. Springer: Cham, 2015. doi:10.1007/978-3-319-23395-6

[94] D. Tall. Private conversation. February 2018.

[95] D. Tischler. Blaschke products and expanding maps of the circle. *Proc. Amer. Math. Soc.* **128** (2000), no. 2, 621–622.

[96] W. R. van Zwet. A proof of Kakutani's conjecture on random subdivision of longest intervals. *Ann. Probability* **6** (1978), no. 1, 133–137.

[97] R. S. Varga. *Matrix Iterative Analysis.* Second edition. Springer Series of Computational Mathematics **27**. Springer-Verlag: Berlin, 2002. doi:10.1007/978-3-642-05156-2

[98] A. Volčič. A generalization of Kakutani's splitting procedure. *Ann. Mat. Pura Appl. (4)* **190** (2011), no. 1, 45–54. doi:10.1007/s10231-010-0136-3

[99] C. Weiss. On the classification of LS-sequences. *Unif. Distrib. Theory* **13** (2018), no. 2, 83–92.

[100] H. Weyl. Über die Gleichverteilung von Zahlen mod. Eins. *Math. Ann.* **77** (1916), no. 3, 313–352. doi:10.1007/BF01475864

[101] L.-S. Young. What are SRB measures, and which dynamical systems have them? *J. Stat. Phys.* **108** (2002), no. 5-6, 733–754. doi:10.1023/A:1019762724717