



HAL
open science

Études probabilistes en théorie des nombres et combinatoire des mots : exemples d'analyse dynamique

Pablo Rotondo

► **To cite this version:**

Pablo Rotondo. Études probabilistes en théorie des nombres et combinatoire des mots : exemples d'analyse dynamique. Number Theory [math.NT]. Université Sorbonne Paris Cité; Universidad de la República (Montevideo), 2018. English. NNT : 2018USPCC213 . tel-02463068

HAL Id: tel-02463068

<https://theses.hal.science/tel-02463068>

Submitted on 31 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**THÈSE DE DOCTORAT DE
UNIVERSIDAD DE LA REPÚBLICA ET DE
L'UNIVERSITÉ SORBONNE PARIS CITÉ
PRÉPARÉE À L'UNIVERSITÉ PARIS DIDEROT
ÉCOLE DOCTORALE DE SCIENCES MATHÉMATIQUES DE
PARIS CENTRE ED 386
INSTITUT DE RECHERCHE EN INFORMATIQUE FONDAMENTALE**

Probabilistic studies in Number Theory and Word Combinatorics: instances of dynamical analysis

Par Pablo Rotondo

Thèse de doctorat d'Informatique

Codirigée par Valérie Berthé, Brigitte Vallée et Alfredo Viola

Présentée et soutenue publiquement à l'Université Paris Diderot le 27 septembre 2018

Président du jury	Cyril NICAUD	Université Paris-Est
Rapporteurs	Bruno SALVY Pierre ARNOUX	INRIA Institut de Mathématique de Luminy
Examineurs	Franco ROBLEDO Thomas STOLL Brigitte VALLÉE	Universidad de la República Université de Lorraine CNRS, Université de Caen
Directeurs de thèse	Valérie BERTHÉ Alfredo VIOLA	CNRS, Université Paris Diderot Universidad de la República

Abstract. Dynamical Analysis incorporates tools from dynamical systems, namely the Transfer Operator, into the framework of Analytic Combinatorics, permitting the analysis of numerous algorithms and objects naturally associated with an underlying dynamical system. This dissertation presents, in the integrated framework of Dynamical Analysis, the probabilistic analysis of seemingly distinct problems in a unified way: the probabilistic study of the recurrence function of Sturmian words, and the probabilistic study of the Continued Logarithm algorithm.

Sturmian words are a fundamental family of words in Word Combinatorics. They are in a precise sense the simplest infinite words that are not eventually periodic. Sturmian words have been well studied over the years, notably by Morse and Hedlund (1940) who demonstrated that they present a notable number theoretical characterization as discrete codings of lines with irrational slope, relating them naturally to dynamical systems, in particular the Euclidean dynamical system. These words have never been studied from a probabilistic perspective. Here, we quantify the recurrence properties of a “random” Sturmian word, which are dictated by the so-called “recurrence function”; we perform a complete asymptotic probabilistic study of this function, quantifying its mean and describing its distribution under two different probabilistic models, which present different virtues: one is a naturally choice from an algorithmic point of view (but is innovative from the point of view of dynamical analysis), while the other allows a natural quantification of the worst-case growth of the recurrence function. We discuss the relation between these two distinct models and their respective techniques, explaining also how the two seemingly different techniques employed could be linked through the use of the Mellin transform. In this dissertation we also discuss our ongoing work regarding two special families of Sturmian words: those associated with a quadratic irrational slope, and those with a rational slope (not properly Sturmian). Our work seems to show the possibility of a unified study.

The Continued Logarithm Algorithm, introduced by Gosper in Hakmem (1978) as a mutation of classical continued fractions, computes the greatest common divisor of two natural numbers by performing division-like steps involving only binary shifts and subtractions. Its worst-case performance was studied recently by Shallit (2016), who showed a precise upper-bound for the number of steps and gave a family of inputs attaining this bound. In this dissertation we employ dynamical analysis to study the average running time of the algorithm, giving precise mathematical constants for the asymptotics, as well as other parameters of interest. The underlying dynamical system is akin to the Euclidean one, and was first studied by Chan (around 2005) from an ergodic point of view, but the presence of powers of 2 in the quotients ingrains into the central parameters a dyadic flavour that cannot be grasped solely by studying this system. We thus introduce a dyadic component and deal with a two-component system. With this new mixed system at hand, we then provide a complete average-case analysis of the algorithm by Dynamical Analysis.

Key words. *Dynamical Analysis, dynamical systems, Word Combinatorics, Sturmian words, recurrence functions, greatest common divisor, continued fractions, continued logarithm expansion, transfer operator, Riemann sums, Dirichlet series, Tauberian theorem.*

Résumé. L'analyse dynamique intègre des outils propres aux systèmes dynamiques (comme l'opérateur de transfert) au cadre de la combinatoire analytique, et permet ainsi l'analyse d'un grand nombre d'algorithmes et objets qu'on peut associer naturellement à un système dynamique. Dans ce manuscrit de thèse, nous présentons, dans la perspective de l'analyse dynamique, l'étude probabiliste de plusieurs problèmes qui semblent à priori bien différents : l'analyse probabiliste de la fonction de récurrence des mots de Sturm, et l'étude probabiliste de l'algorithme du "logarithme continu".

Les mots de Sturm constituent une famille omniprésente en combinatoire des mots. Ce sont, dans un sens précis, les mots les plus simples qui ne sont pas ultimement périodiques. Les mots de Sturm ont déjà été beaucoup étudiés, notamment par Morse et Hedlund (1940) qui en ont exhibé une caractérisation fondamentale comme des codages discrets de droites à pente irrationnelle. Ce résultat relie ainsi les mots de Sturm au système dynamique d'Euclide. Les mots de Sturm n'avaient jamais été étudiés d'un point de vue probabiliste. Ici nous introduisons deux modèles probabilistes naturels (et bien complémentaires) et y analysons le comportement probabiliste (et asymptotique) de la "fonction de récurrence" ; nous quantifions sa valeur moyenne et décrivons sa distribution sous chacun de ces deux modèles : l'un est naturel du point de vue algorithmique (mais original du point de vue de l'analyse dynamique), et l'autre permet naturellement de quantifier des classes de plus mauvais cas. Nous discutons la relation entre ces deux modèles et leurs méthodes respectives, en exhibant un lien potentiel qui utilise la transformée de Mellin. Nous avons aussi considéré (et c'est un travail en cours qui vise à unifier les approches) les mots associés à deux familles particulières de pentes : les pentes irrationnelles quadratiques, et les pentes rationnelles (qui donnent lieu aux mots de Christoffel).

L'algorithme du logarithme continu est introduit par Gosper dans Hakmem (1978) comme une mutation de l'algorithme classique des fractions continues. Il calcule le plus grand commun diviseur de deux nombres naturels en utilisant uniquement des shifts binaires et des soustractions. Le pire des cas a été étudié récemment par Shallit (2016), qui a donné des bornes précises pour le nombre d'étapes et a exhibé une famille d'entrées sur laquelle l'algorithme atteint cette borne. Dans cete thèse, nous étudions le nombre moyen d'étapes, tout comme d'autres paramètres importants de l'algorithme. Grâce à des méthodes d'analyse dynamique, nous exhibons des constantes mathématiques précises. Le système dynamique ressemble à première vue à celui d'Euclide, et a été étudié d'abord par Chan (2005) avec des méthodes ergodiques. Cependant, la présence des puissances de 2 dans les quotients change la nature de l'algorithme et donne une nature dyadique aux principaux paramètres de l'algorithme, qui ne peuvent donc pas être simplement caractérisés dans le monde réel. C'est pourquoi nous introduisons un nouveau système dynamique, avec une nouvelle composante dyadique, et travaillons dans ce système à deux composantes, l'une réelle, et l'autre dyadique. Grâce à ce nouveau système mixte, nous obtenons l'analyse en moyenne de l'algorithme.

Mots clés. *Analyse dynamique, systèmes dynamiques, combinatoire des mots, mots de Sturm, fonction de récurrence, plus grand commun diviseur, fractions continues, logarithme continu, opérateur de transfert, sommes de Riemann, series de Dirichlet, théorème tauberien, probabilités, modèle probabiliste.*

RÉSUMÉ LONG

Contexte général

Dans ce manuscrit de thèse, nous présentons l'étude probabiliste de plusieurs objets provenant de disciplines qui semblent a priori bien distinctes : d'abord, une famille très importante et classique de *mots*, appelés *mots de Sturm*, qui jouent un rôle fondamental en combinatoire des mots; deuxièmement, un algorithme de pgcd (l'algorithme du "logarithme continu" – *CL* pour ses initiales en anglais), propre à la théorie des nombres. Ces objets ont été déjà beaucoup étudiés, en particulier, les "ordres de croissance maximaux" de certains de leurs paramètres caractéristiques sont bien connus. Nous adoptons ici un point de vue différent et nous en faisons une étude probabiliste. Au lieu d'être motivés par la question de savoir "quel est le meilleur/pire cas?", nous considérons des questions comme "comment décrire un mot (de Sturm) aléatoire? comment décrire une exécution aléatoire de l'algorithme de pgcd?"

Même si nos objets d'étude (mots, algorithmes de pgcd) proviennent de disciplines qui semblent éloignées, ils peuvent être décrits dans un cadre commun de théorie des nombres. Ce cadre inclut les fractions continues (classiques pour le cas des mots de Sturm, et une famille qui n'est pas si classiques pour le cas du *CL* où les puissances de 2 jouent un rôle central), et leurs systèmes dynamiques respectifs.

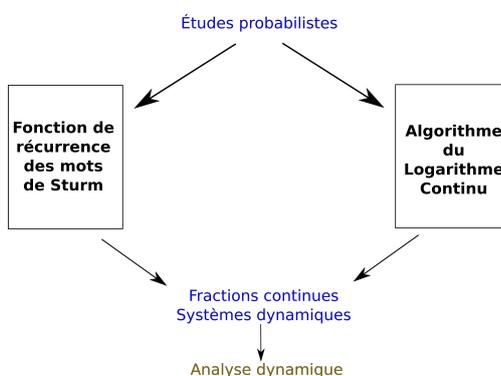


Figure 1: Le schéma de la thèse.

Ici nous utilisons des outils propres aux systèmes dynamiques (comme l'opérateur de transfert) et au cadre de la combinatoire analytique [FS09]. La combinatoire analytique a pour objet fondamental les fonctions génératrices (ici de type "Dirichlet", typiques de la théorie des nombres [Ten15]), avec des coefficients qui comptent des objets combinatoires (ou de théorie des nombres), et relie leur comportement analytique (les

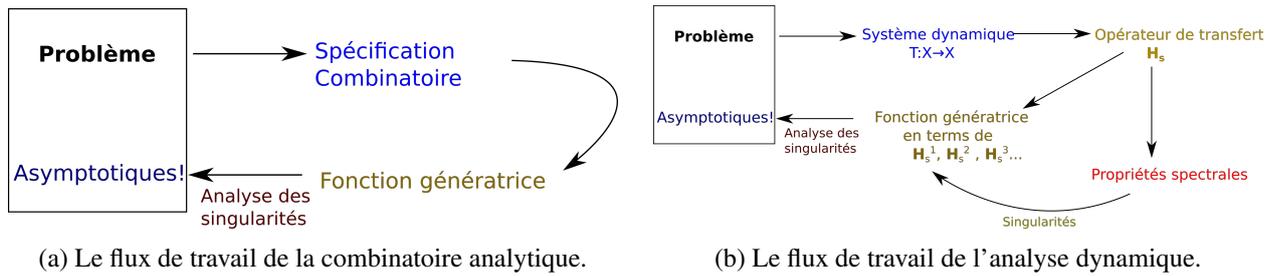


Figure 2: La combinatoire analytique et son prolongement à l'analyse dynamique.

singularités) aux asymptotiques de leur coefficients (grâce aux théorèmes tauberiens). La situation est illustrée dans la figure 2a. La combinatoire analytique est largement utilisée dans l'étude d'algorithmes, de structures de données, ou même en combinatoire per se, probabilistiquement. Quand les objets étudiés sont engendrés par un système dynamique, la combinatoire analytique peut (et doit) être complétée par une autre classe de méthodes, donnant lieu à ce qui s'appelle l'analyse dynamique [FV98], introduite par Baladi, Flajolet, Vallée et d'autres. L'outil clé de l'analyse dynamique est l'opérateur de transfert du système dynamique sous-jacent, qui étend l'opérateur transformateur de densité du système, et suit naturellement l'évolution de nos paramètres pendant l'itération du système. Nous pouvons alors utiliser cet opérateur pour engendrer des fonctions génératrices. Quand l'opérateur agit sur un espace fonctionnel approprié, il présente une valeur propre dominante, qui joue le rôle de la singularité dominante en combinatoire analytique.

L'analyse dynamique est illustrée dans la figure 2b. Nous expliquons brièvement comment on y arrive. La puissance k -ième (par composition) de l'opérateur de transfert H_s décrit la situation après k itérations du système dynamique (sous-jacent à l'algorithme ou processus). Nous cherchons alors des expressions pour notre fonction génératrice en termes de puissances de H_s , souvent avec toutes les puissances apparaissant en même temps $(\mathbf{I} - H_s)^{-1} = \mathbf{I} + H_s + H_s^2 + \dots$, quand nous considérons toutes les exécutions possibles d'un algorithme. Trouver une telle expression peut parfois être impossible avec la combinatoire analytique classique, car l'utilisation des outils nécessite une certaine indépendance entre les différentes étapes de l'algorithme. Une fois que nous avons trouvé ces expressions pour les fonctions génératrices, nous utilisons les propriétés spectrales de l'opérateur; si l'opérateur a de bonnes propriétés (dans un espace fonctionnel approprié), l'action de la puissance H_s^k est déterminée par la *valeur propre dominante* et la *projection* sur l'espace propre associé. Par bonnes propriétés, nous entendons que l'opérateur présente un *saut spectral* [BV03] : la valeur propre dominante est unique et simple, et est séparée du reste du spectre. Nous remarquons que cette situation est similaire à celle du théorème de Perron-Frobenius pour des matrices; le comportement cherché est analogue mais dans un espace de dimension infinie. Le choix de l'espace fonctionnel est délicat car il y a un compromis à trouver : l'espace doit être suffisamment grand pour contenir des fonctions utiles, mais suffisamment petit pour avoir un saut spectral. Une fois établi que les puissances de l'opérateur de transfert sont déterminées par les puissances de la valeur propre dominante, nous pouvons finir l'analyse et déterminer les singularités principales des fonctions génératrices.

Étude probabiliste des mots de Sturmian

Les *mots de Sturm* constituent une famille omniprésente en combinatoire des mots (voir e.g., [Fog02] et [Lot02]). Ce sont précisément les mots les plus simples qui ne sont pas ultimement périodiques, au sens qu'ils ont le plus petit nombre possible de facteurs de chaque longueur n , c'est-à-dire $n+1$. Les mots de Sturm apparaissent naturellement en relation avec la géométrie digitale et les quasicristaux, et, par conséquence, ont été beaucoup étudiés.

Sur l'alphabet binaire $\{0, 1\}$, Morse et Hedlund [MH40] fournissent une description arithmétique des mots de Sturm fondamentale, qui montre un lien profond avec les fractions continues. Plus précisément, ils

ont démontré qu'à chaque mot de Sturm correspond est associé un nombre irrationnel α de l'intervalle unité (appelé sa pente), qui décrit la fréquence des 1 dans le mot : chaque mot de Sturm peut être écrit systématiquement en fonction de sa pente comme S_α . Cette représentation correspond, en fait, à un codage discret de droite comme décrit dans la figure Figure 3.

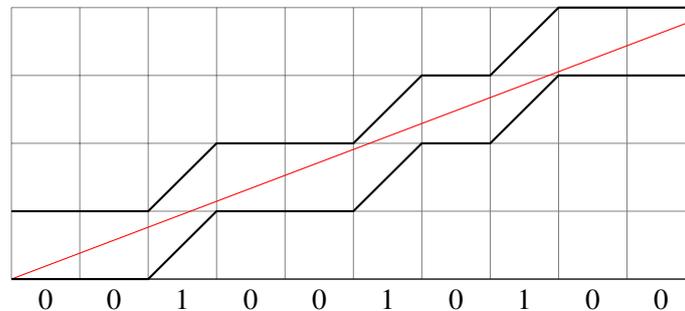


Figure 3: Morse et Hedlund ont relié les mots de Sturm aux codages discrets de droites avec une pente irrationnelle α .

La fonction de récurrence du mot de Sturm S_α décrit la façon dont les facteurs finis de longueur n arrivent dans le mot infini S_α . En particulier, $R_\alpha(n)$ indique le “temps d’attente” maximum qu’il faut pour découvrir tous les facteurs de S_α de longueur n . La fonction $R_\alpha(n)$ dépend, d’une façon assez élégante, du couple (α, n) . Plus précisément, Morse et Hedlund [MH40] ont relié la fonction de récurrence $R_\alpha(n)$ au développement en fraction continue de α , plus particulièrement à ses continuants $q_k(\alpha)$, les dénominateurs des convergents de la fraction continue de α .

Morse and Hedlund ont démontré que, quand n appartient à l’intervalle $[q_{k-1}(\alpha), q_k(\alpha)[$ entre deux continuants consécutifs $q_{k-1}(\alpha)$ et $q_k(\alpha)$ de α , la fonction de récurrence $R_\alpha(n)$ admet une expression simple. De plus, les auteurs exhibent un comportement en “ $n \log n$ ” pour le pire cas de $R_\alpha(n)$, dont l’apparition dépend fortement du choix spécifique de (α, n) , mais qui se produit presque sûrement pour une infinité de n .

La question que nous posons est : *Si le couple (α, n) est tiré au hasard de façon systématique, quel est le comportement probabiliste de la fonction de récurrence $R_\alpha(n)$?* Dans cette thèse, nous proposons deux modèles probabilistes différents pour répondre à la question.

Premier modèle probabiliste

Notre premier modèle, décrit dans le chapitre 5, a été publié en [BCR⁺15] pour la conférence MFCS 2015. C’est une première approche pour le pire cas d’un point de vue probabiliste obtenue en conditionnant avec des intervalles de plus en plus petits qui contiennent les pires cas. Dans ce modèle, nous tirons la pente α “uniformément” dans l’intervalle unité puis nous prenons de sous-suites particulières $k \mapsto n_k$ qui fixent la position barycentrique μ de n relative à l’intervalle $[q_{k-1}(\alpha), q_k(\alpha)[$. Nos résultats [BCR⁺15], obtenus par des méthodes d’analyse dynamique, quantifient l’incidence de la position μ sur le comportement du pire cas de la fonction de récurrence, et montrent effectivement un comportement du type “ $n \log n$ en moyenne” sur certaines sous-suites $k \mapsto n_k$.

Nous utilisons des méthodes d’analyse dynamique, ici avec le système dynamique classique d’Euclide (avec l’application de Gauss) et son opérateur transformateur de densités. Nous démontrons que l’espérance associé à la k -ième étape n_k s’écrit en termes de la puissance k -ième de l’opérateur transformateur de densités. Sur l’espace fonctionnel des fonctions de variation bornée, cet opérateur présente un saut spectral dont on a besoin, et les puissances de l’opérateur sont alors approchées par les (véritables) puissances de la valeur propre dominante. Dans ce cas nous avons eu besoin aussi de résultats sur le terme de reste, qui sont assez connus.

Deuxième modèle probabiliste

Le modèle précédent est très utile quand il s’agit de décrire le pire cas de la fonction de récurrence et l’incidence de la position relative μ . Dans notre deuxième modèle nous considérons, à nouveau, une pente aléatoire α , mais la taille d’entrée n est fixée et indépendante de α . L’analyse de la fonction de récurrence dans ce modèle “ $n \rightarrow \infty$ fixé” est décrite au chapitre 4, et a été publiée dans les actes de ANALCO 2017 [RV17]. Ce modèle peut être utilisé aussi afin d’étudier d’autres fonctions, comme la position relative μ , qui jouent un rôle dans l’analyse des mots de Sturm, ainsi que les fractions continues elles-mêmes.

Nous obtenons trois résultats principaux dans [RV17]; nous considérons les variables aléatoires $\alpha \mapsto (1/n)R_\alpha(n)$ et les étudions pour n grand. Nous exhibons leur distribution limite, et nous démontrons l’existence de la densité limite. Nous étudions aussi l’espérance conditionnelle du quotient de récurrence $(1/n)R_\alpha(n)$, quand nous excluons la possibilité que n se trouve trop proche du bord gauche de l’intervalle $[q_{k-1}(\alpha), q_k(\alpha)]$. Enfin, nous exhibons une classe d’événements pour lesquels l’ordre de cette espérance conditionnelle est exactement $\log n$. Ce dernier résultat peut être considéré comme une extension probabiliste du résultat classique de Morse et Hedlund.

Nos preuves utilisent des méthodes élémentaires : elles reposent sur une comparaison précise entre une intégrale et sa somme de Riemann; cependant, l’intégrale est impropre (mais convergente) et la somme de Riemann implique une condition supplémentaire de primalité, ce que nous avons appelé une “somme de Riemann première”. Les sommes de Riemann premières apparaissent aussi dans [BCZ03], où les auteurs étudient la suite de Farey, mais travaillent dans des domaines bornés. Ici nous adaptons les méthodes pour des domaines non bornés, avec des termes d’erreur précis pour la convergence vers les intégrales.

Comportement probabiliste de familles “particulières” de mots de Sturm

Il y a deux familles particulières de mots de Sturm qui sont importantes en tant que telles. Leur étude constitue un travail un cours qui, en fait, vise à unifier les approches pour les deux familles et le cas générique du deuxième modèle. Nos résultats actuels (non encore publiés) montrent que le comportement de ces familles est semblable au cas générique. Les preuves utilisent aussi d’autres outils utiles tels que les séries de Dirichlet et les théorèmes taubériens.

Christoffel words Quand la pente α est rationnelle, le mot S_α est périodique, ce que l’on appelle un mot de Christoffel [BLRS08]. Dans ce cas, la question est : *Est-ce vrai que, quand la longueur du développement en fraction continue d’ α tend vers l’infini, la fonction de récurrence a un comportement en moyenne semblable à celui d’un mot de Sturm générique?*

Mots de Sturm engendrés par morphisms Il y a une deuxième famille importante, à savoir, les mots de Sturm qui sont engendrés par de morphisms [Al198]. Ils sont associés à des pentes α irrationnelles quadratiques, qui ont des développements en fractions continues périodiques. Ici nous posons une question similaire : *Est-ce vrai que, quand la période de l’irrationnel quadratique α tend vers l’infini, le comportement de la fonction de récurrence, en moyenne, est semblable à celui du cas générique?* L’analyse dans ce cadre est plus compliquée car elle dépend de (n, α, ℓ) où ℓ dénote la quantité de tours de la période.

Analysis of the Continued Logarithm Algorithm

L’algorithme du logarithme continu –CL pour ses initiales en anglais– est introduite par Gosper dans Hakmem [Gos78] comme une “mutation” de l’algorithme classique des fractions continues. Cet algorithme

calcule le plus grand commun diviseur de deux nombres entiers en utilisant uniquement des shifts binaires et des soustractions. Le pire cas a été étudié récemment par Shallit [Sha16], qui a donné des bornes précises pour le nombre d'étapes et a exhibé une famille d'entrées sur laquelle l'algorithme atteint cette borne. Dans cete thèse, nous étudions le nombre moyen d'étapes, tout comme d'autres paramètres importants de l'algorithme. Grâce à des méthodes d'analyse dynamique, nous exhibons des constantes mathématiques précises.

Plus précisément, nous considérons les couples (p, q) , avec $1 \leq p \leq q \leq N$, avec la probabilité uniforme, et nous étudions les valeurs moyennes du nombre d'étapes de pseudo-divisons et de shifts binaires, quand $N \rightarrow \infty$. Dans notre résultat principal, le théorème 7.2, nous démontrons que ces valeurs moyennes sont asymptotiquement linéaires par rapport à la taille $\log N$, et nous décrivons précisément leur comportement asymptotique quand $N \rightarrow \infty$.

Le système dynamique sous-jacent ressemble à première vue à celui d'Euclide, et a été étudié d'abord par Chan [Cha05] et Borwein et al [BCLM17], avec des méthodes ergodiques, mais la présence des puissances de 2 dans les quotients change la nature de l'algorithme et donne une nature dyadique aux principaux paramètres de l'algorithme, qui ne peuvent donc pas être simplement caractérisés dans le monde réel. C'est pourquoi nous introduisons un nouveau système dynamique, avec une nouvelle composante dyadique, et travaillons dans ce système à deux composantes, l'une réelle, et l'autre dyadique. Grâce à ce nouveau système mixte, nous obtenons l'analyse en moyenne de l'algorithme.

Acknowledgements. I begin by thanking my thesis directors. First Alfredo Viola, who has unconditionally supported me since the times of the “Low-correlation sequences” seminar and introduced me to my other thesis directors Valérie Berthé, Brigitte Vallée, as well as the community in general. I keep very fond memories of our time together and hope we will continue to have seminars or work discussions for a long time. I am profoundly grateful to Brigitte, who has helped me and taught me a lot, through her seemingly unbounded energy. We have worked a lot together, and I think that, even though we have had our share of divergence and communication problems, overall it has been very positive and led to interesting and unexpected ideas (in part due to this divergence of thought process!). I would like to thank Valérie, who has always supported me, both academically and administratively, and directed me towards interesting topics and problems which resulted in this thesis. I hope we can continue to work on the pending topics and unify it with the rest of the work.

I would like to express my gratitude to the referees, Bruno Salvy and Pierre Arnoux, for having accepted to review my thesis through the summer, and for all of their corrections and suggestions. They have certainly helped improve the quality of this dissertation. I also extend my thanks to all of the members of the jury for having accepted to be here for my defence, I feel honoured.

I thank the whole IRIF, the PhD students, the ATERs and everyone that has made me feel welcome. Heartfelt thanks go to the people from the AMACC team in Caen, Ali, Julien, Loïck, as well as Eda Cesaratto from Universidad Nacional de General Sarmiento, for the useful discussions, work and seminars enjoyed together. I thank the secretaries from GREYC in Caen, who have helped me with various “missions”.

My sincere thanks to Alberto Pardo and Franco Robledo, from Universidad de la República, as well as the secretary María Inés, who have made the cotutelle possible.

My research would have been impossible without the funding from the ANR DynA3S.

Finally, I want to thank my family, my friends, both from Uruguay, from the Maison de l’Italie and IRIF, for their constant support.

CONTENTS

Contents	11
Introduction	15
I Presentation of the general context	23
1 Continued Fractions and the Gauss map	25
1.1 The numeration process	25
1.1.1 The Gauss map	26
1.1.2 Continued fractions and the Euclidean Algorithm	27
1.1.3 Basic properties of continuants	28
1.1.4 Inverse branches of the Gauss map	30
1.1.5 Fundamental Intervals	31
1.2 Dynamical systems and the Perron-Frobenius operator	32
1.2.1 A general definition of a dynamical system	33
1.2.2 Dynamical systems of interest	34
1.2.3 Pushforward measure – Invariant measure	35
1.2.4 The Perron Frobenius operator	36
1.2.5 The case of the Gauss map; the Gauss density	37
1.2.6 The case of the CL map	38
1.2.7 Entropy and dynamical systems	38
1.3 Almost everywhere properties	39
1.3.1 Introduction	39
1.3.2 Birkhoff’s Ergodic Theorem	40
1.3.3 Ergodicity of the Gauss map and CL map	40
1.3.4 Consequences of ergodicity: frequency of digits	42
1.3.5 Large digits in the expansion: the Borel-Bernstein Theorem	46
1.4 Real probabilistic framework	51
1.4.1 Concepts from functional analysis	52
1.4.2 The spectral radius of the Perron Frobenius operator	53
1.4.3 Quasi-compactness	55
1.4.4 Spectral gap	58
1.4.5 Eigenvalues and the ergodic properties of the Perron Frobenius operator	59

1.4.6	An application of the spectral decomposition to the values of the digits	61
1.4.7	The transfer operator	62
1.5	Continued fractions: Rationals and quadratics irrationals	64
1.5.1	Probabilistic model for rational numbers	65
1.5.2	Asymptotic probabilistic properties of digits for rational numbers	65
1.5.3	Probabilistic model for quadratics irrationals	66
2	Concepts from Analytic Combinatorics	69
2.1	Generating functions	70
2.1.1	Ordinary generating functions	70
2.1.2	Exponential generating functions	72
2.1.3	Power series and singularities: the second step	73
2.1.4	Dirichlet generating functions	78
2.2	Analytic Combinatorics for Dirichlet generating functions	80
2.2.1	Tauberian theorems	81
2.2.2	An example of Analytic Combinatorics in arithmetics	81
2.2.3	What is Dynamical Analysis? Why is it useful?	85
II	Studies in Word Combinatorics	87
3	Sturmian words	89
3.1	Concepts from Combinatorics on Words	89
3.1.1	The complexity function of an infinite word	90
3.1.2	Definition of Sturmian words	91
3.1.3	Basic properties of Sturmian words	91
3.2	The arithmetic nature of Sturmian words	93
3.2.1	The slope of a Sturmian word	95
3.2.2	The language of a Sturmian word	98
3.2.3	End of the proof of the characterization of Morse-Hedlund Theorem 3.1	100
3.3	A second concept from Word Combinatorics: recurrence	101
3.3.1	Definitions and basic properties	101
3.3.2	The frequencies of the factors of a Sturmian word	102
3.3.3	The Morse-Hedlund formula for recurrence of Sturmian words	103
3.4	The growth of the recurrence function of Sturmian words	105
3.4.1	Classical results: worst-case analysis	106
3.4.2	Position parameters	107
3.4.3	Our framework: probabilistic analyses	108
3.4.4	Two probabilistic models	109
4	A first probabilistic study of Sturmian words and Q-functions	113
4.1	Introduction	113
4.2	Framework and results.	114
4.2.1	Position parameters.	114
4.2.2	Q -functions.	115
4.2.3	Probabilistic setting.	116
4.2.4	Distributions.	116
4.2.5	Results - densities.	117
4.2.6	Conditional expectations.	119
4.3	Proofs: distributions and densities.	119
4.3.1	Continued fractions, fundamental intervals and continuants.	119

4.3.2	Distributions. Strategy of the proof.	120
4.3.3	Distributions. Proof of Theorem 4.1.	124
4.3.4	Proof of Theorem 4.2. First step.	125
4.3.5	Proof of Theorem 4.2. Second step.	125
4.4	Proofs: conditional expectations.	128
4.4.1	Limit expectation of bounded \mathcal{LQ} - functions.	128
4.4.2	Case of the recurrence quotient.	128
4.4.3	General conditional expectations.	129
4.4.4	Conditional expectation of the recurrence quotient. Proof of Theorem 4.3.	130
4.5	Other applications and extensions	130
4.5.1	Number of continuants in an interval.	131
4.5.2	The smallest distance and \mathcal{Q} \int -functions	131
4.5.3	Independence from the initial distribution	138
4.6	Conclusions	142
5	The recurrence function and the relative position	145
5.1	Introduction	145
5.2	The recurrence function of Sturmian words	145
5.3	Probabilistic model and main results	147
5.3.1	Position parameter μ	147
5.3.2	Probabilistic model	148
5.3.3	Results for a constant position μ	149
5.3.4	Results when the sequence $\mu_k \rightarrow 0$	150
5.4	Strategy for the proofs.	150
5.4.1	Smooth sequences	151
5.4.2	The dynamical system and the Perron-Frobenius operator	151
5.4.3	Smooth random variables and Perron-Frobenius operator	153
5.4.4	Asymptotic study of smooth variables	153
5.4.5	Third Step of the proof of Theorem 5.1.	154
5.4.6	Comparison between densities $s^{(\mu)}$ and $s^{(0)}$	156
5.4.7	End of the proof of Theorem 5.2	156
6	Comparison between the models and special families of slopes	157
6.1	Relation between the two models.	157
6.2	Relationship between the techniques employed	158
6.2.1	From Riemann sums to the Transfer Operator	158
6.2.2	Riemann sums resembling the Perron-Frobenius operator	160
6.3	Special families of slopes	161
6.3.1	Eventually periodic words: α rational	161
6.3.2	Sturmian words and morphisms: α quadratic irrational	164
6.3.3	The size of a quadratic irrational and the model	164
6.3.4	The generating functions	165
6.3.5	The number of complete cycles ℓ	166
6.3.6	Generating function for the first cycle: $\ell = 0$	166
6.3.7	Results for $\ell = 0$	168
6.3.8	The case $\ell \rightarrow \infty$ and future work	169
III	Studies in Arithmetics	171
7	The Continued Logarithm Algorithm	173

7.1	Introduction	173
7.1.1	The continued logarithm expansion	175
7.1.2	The continued logarithm algorithm	176
7.2	The CL dynamical system	178
7.2.1	The continuants of the CL expansion	178
7.2.2	The Perron Frobenius operator	181
7.3	Costs and model for the algorithm	182
7.3.1	Generating functions for our main costs	182
7.3.2	Probabilistic model for the rational case	183
7.4	The extended dynamical system.	184
7.4.1	Extension of the dynamical system	184
7.4.2	Measures on \mathbb{Q}_2	185
7.4.3	Density transformer and transfer operator	187
7.4.4	Alternative expressions of the Dirichlet generating functions	188
7.5	Functional Analysis	188
7.5.1	Lasota-Yorke for the classical CL dynamical system	188
7.5.2	Functional space	189
7.5.3	Action of the operator $\mathbf{H}_{t,u,v,z}$ on \mathcal{F}	190
7.5.4	Dominant spectral properties of the operator	193
7.6	Final result for the analysis of the CL algorithm	195
	Conclusions	197
	Bibliography	199

INTRODUCTION

General context

In this dissertation, we study objects coming from a priori diverse fields: first, a very important family of *words*, the so-called *Sturmian words*, fundamental in Combinatorics on Words; second, a gcd algorithm (the Continued Logarithm Algorithm – *CL* for short), stemming from Number Theory. These objects have been studied extensively, and the “extreme orders” of some of their important characteristics are now well-understood. We adopt a different point of view; we wish to study them from a probabilistic perspective. Rather than being motivated the question “what is the worst/best case scenario?”, we consider questions such as “what does a random (Sturmian) word look like? what does a random gcd algorithm execution look like ?

Even though the objects we consider (words, gcd algorithms) come from seemingly distant fields, they can both be described within a common Number Theoretic framework. This framework includes continued fraction expansions (the usual one for Sturmian words, and a less classical one for the *CL* algorithm, where powers of two play a central role), and their corresponding underlying dynamical systems.

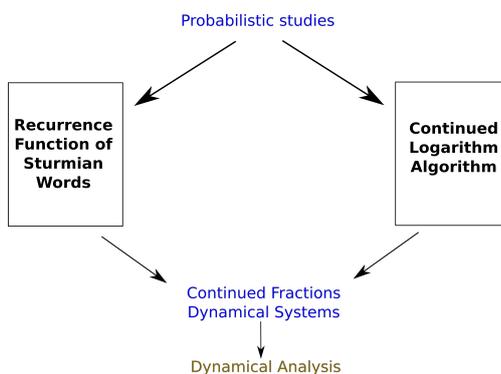


Figure 4: The diagram of this thesis.

Here we apply techniques originating from Analytic Combinatorics and Dynamical Systems. Analytic Combinatorics [FS09] deals with generating functions (here of Dirichlet type, stemming from Number Theory [Ten15]), with coefficients counting combinatorial (number theoretical) objects, and relates their analytic behavior (notably their singularities) to the asymptotics of their coefficients (here via Tauberian Theorems). The situation is illustrated in Figure 5a. Analytic Combinatorics is widely used to study algorithms, data structures, or combinatorial objects per se, probabilistically. When the objects of interest are generated by a

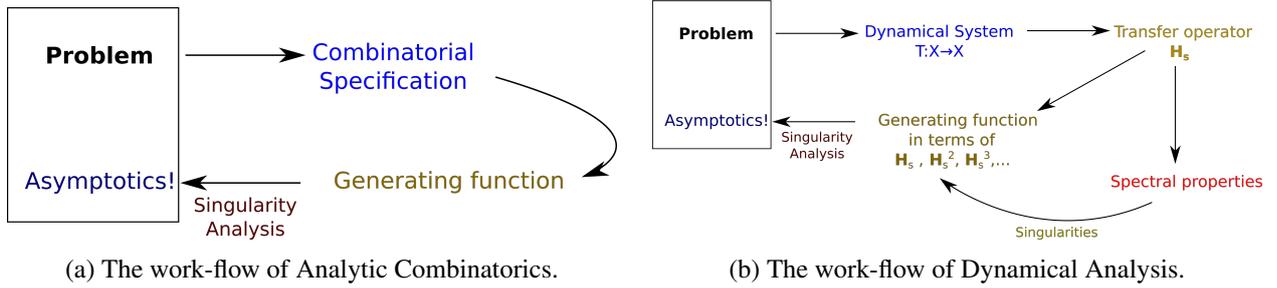


Figure 5: Analytic Combinatorics and the extension to Dynamical Analysis.

dynamical system, Analytic Combinatorics can be (and must be) complemented by another class of methods, giving rise to the so-called Dynamical Analysis [FV98], pioneered by Baladi, Flajolet, Vallée and others. The key object in Dynamical Analysis is the transfer operator of the underlying dynamical system, which extends the density transformer operator of the system, and naturally tracks the evolution of our parameters of interest through its iterates. It thus can be viewed as a generating operator that generates itself the generating functions of interest. When it acts on a convenient functional space, this operator has a dominant eigenvalue, which plays the same role as the dominant singularity in classical Analytic Combinatorics.

The extension to Dynamical Analysis is illustrated in Figure 5b. We briefly explain how this is achieved. The k -th power (by composition) of the transfer operator \mathbf{H}_s reflects the situation after k iterations of the dynamical system (underlying the algorithm or process). Thus we must first exploit this to give expressions for our target generating functions in terms of the powers of \mathbf{H}_s , often with all powers at the same time $(\mathbf{I} - \mathbf{H}_s)^{-1} = \mathbf{I} + \mathbf{H}_s + \mathbf{H}_s^2 + \dots$ when considering every possible execution of an algorithm. Finding such an expression is sometimes impossible in classical Analytic Combinatorics, as the usual techniques require some sort of independence between distinct steps of the algorithm. Given expressions for our generating functions in terms of the powers of \mathbf{H}_s , we exploit the spectral properties of the operator; if the operator is well-behaved (in an appropriately chosen functional space), the action of the power \mathbf{H}_s^k is determined by the *dominant eigenvalue* and the *projection* over its eigenspace. By well-behaved we mean it presents a *spectral gap* [BV03]: the dominant eigenvalue is unique and simple, and separated from the rest of the spectrum. The reader may be familiar with the Perron-Frobenius Theorem for matrices; the target behavior is analog but in a space of infinite dimension. The choice of the functional space is then a delicate one as there must be a balance: it must be big enough to contain useful input functions, but small enough to give us our desired spectral gap. Then, once we have established that the powers of the operator follow the powers of the dominant eigenvalue, we may complete the analysis and determine the main singularities of the generating functions.

This dissertation is structured into 3 parts, following the bottom of the diamond in Figure 4:

- Part I is concerned with the common background for the whole thesis, which corresponds to the bottom part of Figure 4. We introduce the background in continued fractions and dynamical systems in Chapter 1. Particular attention has been paid to introducing continued fractions, dynamical systems and the functional properties of the transfer operator. The concepts we require from Analytic Combinatorics are introduced in Chapter 2, in particular Dirichlet Generating Functions and the Tauberian Theorem.
- Part II deals with the problems coming from Combinatorics on Words. Therein we define all the necessary notions concerning Sturmian words and our problematic (in Chapter 3), to then proceed to our probabilistic study of Sturmian words (in chapters 4 and 5).
- Finally, Part III closes the thesis with a probabilistic study of the CL algorithm.

Now we give further details regarding the contents of Part II and Part III.

Probabilistic study of Sturmian words

Sturmian words are central objects in Word Combinatorics (see e.g., [Fog02] and [Lot02]). These are precisely the simplest infinite words that are not eventually periodic, in the sense that they have the absolutely smallest number of factors of each length n , that is $n + 1$. Sturmian words turn up naturally in relation to digital geometry and quasicrystals, and have been widely studied.

On the binary alphabet $\{0, 1\}$, Morse and Hedlund [MH40] provide a powerful arithmetic description of Sturmian words and relate them to continued fraction expansions. Specifically, they show that each Sturmian word is in strong correspondence with an irrational number α of the unit interval (called its slope), describing the frequency of ones in the word: each Sturmian word may be written as S_α for some irrational α in the unit interval. This representation is related to the discrete coding of lines such as the one in Figure 6.

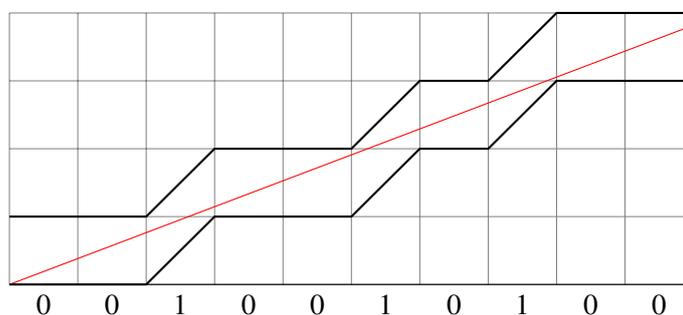


Figure 6: Morse and Hedlund connected Sturmian words to discretized lines of irrational slope α .

The recurrence function of the Sturmian word S_α describes how the finite factors of length n occur inside the infinite word S_α . In particular $R_\alpha(n)$ denotes the maximum “waiting time” that is needed to discover all the factors of S_α of length n . The function $R_\alpha(n)$ depends nicely upon the pair (α, n) . More precisely, Morse and Hedlund [MH40] relate the recurrence function $R_\alpha(n)$ to the continued fraction expansion of α , more particularly to its continuants $q_k(\alpha)$, the denominators resulting from the convergents of the continued fraction for α .

Morse and Hedlund showed that when n belongs to the interval $[q_{k-1}(\alpha), q_k(\alpha))$ between two consecutive continuants $q_{k-1}(\alpha)$ and $q_k(\alpha)$ of α , the recurrence function $R_\alpha(n)$ admits a simple expression. Further, the authors exhibited a “ $n \log n$ ” worst-case behavior for $R_\alpha(n)$, whose occurrence depends strongly on the specific choices of (α, n) , but occurs almost surely for infinitely many n .

We present the general background regarding Sturmian words as well as Morse and Hedlund’s results, rewritten in our notation, in Chapter 3. We deemed, in particular, that the link between the recurrence of Sturmian words and continuants had to be explained thoroughly as we make extensive use of it. Finally, at the end of the chapter, in Section 3.4, we discuss Morse and Hedlund’s classical results concerning the growth of the recurrence function and we present briefly our context, questions, and results.

The question we pose is: *If the pair (α, n) is chosen randomly in some systematic way. What is the probabilistic behavior of the recurrence function $R_\alpha(n)$?* In this dissertation, we consider two distinct probabilistic models to answer this question. Roughly speaking, our first model [BCR⁺15] considers sequences of n , appropriately chosen for each α , in turn chosen uniformly at random, while our second model [RV17] leaves n fixed and large ($n \rightarrow \infty$) and chooses α at random. The latter is described and studied in Chapter 4 while the former in Chapter 5, and the relation between the two models is studied in Chapter 6. For each model, we answer the preceding question by giving precise *limit distributions* and *densities*, and studying how these relate to the worst-case “ $n \log n$ ” behavior, found by Morse and Hedlund, through the study of appropriate *conditional probabilities*.

First probabilistic model

Our first model, described in Chapter 5, was published in [BCR⁺15] for MFCS 2015. It was a first attempt to approach this worst-case scenario from a probabilistic setting by actually conditioning to smaller and smaller sets which encapsulate the worst cases. In this setting, one picks the slope α “uniformly at random” from the unit interval and then considers particular subsequences $k \mapsto n_k$ that fix the barycentric position μ of n within the interval $[q_{k-1}(\alpha), q_k(\alpha)]$. Our results in [BCR⁺15], following a Dynamical Analysis, quantify the incidence of the position μ on the worst-case behavior of the recurrence function, and exhibit a “ $n \log n$ average behavior” over certain sequences $k \mapsto n_k$.

More precisely, we exhibit the asymptotic value, as $k \rightarrow \infty$, of the distribution of $(1/n_k)R_\alpha(n_k)$. As this analysis may be performed even for varying relative position μ_k , we show a kind of Morse-Hedlund result “on average”: the expectation is of order $n \log n$ when $\mu_k \rightarrow 0$ at a prescribed exponential rate.

Considering the dynamical system underlying the problem (here the classical Euclidean one, with the Gauss map), we use methods from Dynamical Analysis, in this case the plain density transformer operator of the system. We express the expectations relative to the k -th step in terms of the k -th power of the density transformer. Over an appropriate functional space (here the functions of bounded variation), the operator presents our desired spectral gap, and the powers of the operator are approximated by the (true) k -th power of the dominant eigenvalue. In this case we also require specific knowledge regarding the remainder term, which has been studied extensively for the Euclidean system.

Main results for the first probabilistic model. The main results obtained are summarized here:

▷ **Theorem 5.1.** For a fixed relative (barycentric) position μ , we characterize the limit expectations and the limit density of $(1/n_k)R_\alpha(n_k)$ as $k \rightarrow \infty$.

▷ **Theorem 5.2.** For a *varying* relative (barycentric) position $\mu_k \rightarrow 0$, we characterize the the limit density of $(1/n_k)R_\alpha(n_k)$ as $k \rightarrow \infty$, demonstrating the rate of convergence to the case of fixed $\mu = 0$ from Theorem 5.1. Moreover, we demonstrate that the expectations of $(1/n_k)R_\alpha(n_k)$ do have a $\log n_k$ behavior when μ_k tends to zero exponentially (not too fast).

Second probabilistic model

The previous model for Sturmian words is very useful when it comes to describing the worst-case behavior of the recurrence function and the incidence of the relative position μ . Our second model considers, again, a random slope α , but somewhat orthogonally to the previous model, we take an input size n independent from α . The analysis of the recurrence function within this “fixed $n \rightarrow \infty$ ” model is described in Chapter 4, and was published in the proceedings of ANALCO 2017 [RV17]. This model can be employed to study other functions, such as the relative position μ , playing a role in the analysis of Sturmian words, or continued fractions.

We obtain three main results in [RV17]; we consider the random variables $\alpha \mapsto (1/n)R_\alpha(n)$ and study them for large n . We exhibit a limit for their distribution, and prove the existence of a limit density. We also study the conditional expectation of the recurrence quotient $(1/n)R_\alpha(n)$, when we exclude the possibility that n be too close to the left-end of the interval $[q_{k-1}(\alpha), q_k(\alpha)]$. Finally, we describe a class of events for which the order of this conditional mean value is exactly of order $\log n$. This can be viewed as a probabilistic extension of the Morse and Hedlund result.

Our proofs use elementary methods: they are based on a precise comparison between an integral and its Riemann sum; however, the integral is improper (but convergent) and the Riemann sum is constrained by a coprimality condition, what we call a “coprime Riemann sum”. Coprime Riemann sums appear in [BCZ03],

where the authors study the Farey sequence, but limited to bounded domains. Here we adapt the methods of proof to unbounded domains, getting tight error bounds for the convergence towards the integrals.

We introduce a general family of functions, called continuant-functions or \mathcal{Q} -functions, which are defined via the sequence of continuants $k \mapsto q_k(\alpha)$. Ustinov in [Ust09] had already considered similar functions to answer a question from Sinai and Ulcigrai [SU08]; here we generalize the class of functions to \mathcal{Q} -functions, demonstrating its ubiquity in continued fraction problems. The recurrence quotient $(1/n)R_\alpha(n)$ is an instance of this family, and the other “geometric” parameters of interest provide natural examples of such a notion. The class of \mathcal{Q} -functions lead naturally to coprime Riemann sums. Thus the paper describes a framework for studying the more general \mathcal{Q} -functions, giving special attention to the recurrence function.

Main results for the second probabilistic model. The main results obtained are summarized here:

▷ **Theorem 4.1.** We show that for a wide family of \mathcal{Q} -functions, which we call \mathcal{LQ} -functions, the random variables $(1/n)R_\alpha(n)$ have a limiting distribution as $n \rightarrow \infty$. This distribution is expressed in terms of an analog $\psi(x, y)$ of the Gauss measure. Moreover, we show explicit bounds for the remainder term.

▷ **Theorem 4.2.** This result makes precise when the histograms do converge to the derivative of the distribution, thus making sense of it as a density. Since the distributions of \mathcal{Q} -functions, such as $(1/n)R_\alpha(n)$, are discrete, we have to be careful when speaking of the limit density. Here we characterize this limit density of \mathcal{LQ} -functions completely, giving also remainder terms.

▷ **Theorem 4.3.** This result demonstrates that if we condition to an event such as $\mu \geq \epsilon(n)$, which prevents that n be too close to the left-end of the interval $[q_{k-1}(\alpha), q_k(\alpha))$, the expected value of $(1/n)R_\alpha(n)$ involves a $\log \epsilon(n)$. This is the counter-part of the results by Morse and Hedlund for this probabilistic model. The proof exploits strongly the knowledge of the remainder term from Theorem 4.1.

We highlight also that the convergence in distribution for \mathcal{Q} -functions still holds for more general conditions, but without any guarantee for the remainder term however, see Theorem 4.10.

Probabilistic behavior of “particular” Sturmian words

There are two kinds of special Sturmian words, both interesting in their own right.

Christoffel words When the parameter α is rational, the word S_α is periodic and is called a Christoffel word [BLRS08]. It is still interesting to study such words probabilistically, particularly how the word evolves (when the length $p(\alpha)$ of its continued fraction becomes large) towards a Sturmian word, notably from the point of view of its recurrence function. Our main question is: *Is it true that, when the length of the continued fraction of α becomes large, the behavior of the recurrence function becomes close to the recurrence function of a generic Sturmian word?*

We have carried out this study, yielding results in the same vein as the ANALCO paper. This results are not yet published, and we explain them in subsection 6.3.1. We consider the set Ω_N of rational numbers from the unit interval with a denominator at most N , endowed with the uniform probability. We introduce generating functions of Dirichlet kind, in order to sieve the right rationals for our asymptotics (through a Tauberian Theorem). We show that when the bound for the denominator N tends to infinity, the analogous distributions for the recurrence quotient are given by the same coprime Riemann-sum as before.

Sturmian words generated by morphisms There is a second important family of Sturmian words, namely Sturmian words that are generated by word morphisms [All98]. They are associated to quadratic irrationals α , that give rise to periodic continued fraction expansions. Our main question is: *Is it true that, when the*

period $\pi(\alpha)$ of the quadratic irrational α becomes large, the behavior of the recurrence function resembles that of a generic Sturmian word? The analysis is more difficult because it now depends on the triple (n, α, ℓ) where the integer ℓ describes the number of times the period is needed.

We have already obtained interesting results for the case $\ell = 1$ (not yet published), where we get the same distributions from our second probabilistic model but through substantially different methods. There seems to be a stationary behavior as $\ell \rightarrow \infty$, and this is work still in progress. The results obtained thus far are discussed in subsection 6.3.2.

Our proofs mix methods from dynamical analysis and elementary methods like those in the second model: as for rational numbers, we introduce generating functions of Dirichlet type to manage the quadratic irrationals, which are endowed with their usual notion of size, here closely related to the fundamental unit of the associated quadratic field. The associated generating functions are (again) expressed in terms of the transfer operator of the Euclid dynamical system, but now via their traces. This leads to a more involved study.

A specific interesting result This study also leads us to a specific interesting result on finite continued fractions. Such a continued fraction is defined by a finite sequence (m_1, m_2, \dots, m_k) of partial quotients and represents a rational p/q . The “mirror” continued fraction defined by the mirror sequence $(m_k, m_{k-1}, \dots, m_1)$ represents another reduced rational p'/q that has the same denominator as the previous one (but not the same numerator). Very often, the two expansions occur together in our studies (and in many other studies), and the two associated rational numbers seem *a priori* to be correlated in a strong way. We adapt a result of Shparlinski [Shp12, Theorem 13] and we prove that, when one draws a rational p/q uniformly at random, the two rational numbers p/q and p'/q asymptotically behave in an independent way, as the denominator q becomes large. This is explained in Section 4.5.3.

Analysis of the Continued Logarithm Algorithm

The Continued Logarithm Algorithm –CL for short– introduced by Gosper in 1978 [Gos78], computes the gcd of two integers; it employs efficient operations, as it only performs (binary) shifts and subtractions. Shallit [Sha16] has studied its worst-case complexity in 2016 and showed it to be linear, and he proposed the problem of determining the average-case analysis of the algorithm to us. We answer his question in the publication [RVV18], accepted in LATIN 2018 and described in Chapter 7: we study its main parameters (number of iterations K , total number of shifts S) and obtain precise asymptotics for their mean values.

More precisely, we consider the set Ω_N gathering all integer pairs (p, q) with $1 \leq p \leq q \leq N$, endowed with the uniform probability, and we study the mean values of K and S as $N \rightarrow \infty$. In our main result, Theorem 7.2, we prove that these mean values are asymptotically linear in the size $\log N$, and describe precisely their asymptotic behavior as $N \rightarrow \infty$. This result is to be expected intuitively, since the algorithm resembles the Euclidean one where both the worst and average case are linear [Val06]. The study leading up to this average case asymptotics, however, presents several interesting (and non-trivial) aspects.

The dynamical analysis involves the dynamical system underlying the algorithm, which produces continued fraction expansions whose quotients are powers of 2. Even though this CL system has already been studied by Chan [Cha05] and Borwein et al [BCLM17], the presence of powers of 2 in the quotients ingrains into the central parameters a dyadic flavor that cannot be grasped solely by studying the CL system. Indeed, even if the input of the CL algorithm is a pair of coprime integers, the algorithm builds a sequence q_k of remainders, for which the pair (q_{k-1}, q_k) is no longer coprime. The successive $\gcd(q_{k-1}, q_k)$ are now powers of 2, and it appears experimentally that $(1/k) \log_2 \gcd(q_{k-1}, q_k)$ gets close to $1/2$ as k becomes large.

In order to take into account this involved dyadic phenomenon, central to our analysis, we add a second dyadic component to the (usual) CL dynamical system, and work with a two-component system, which

allows us to keep track of all the interesting parameters under study. This new dynamical system with its two components, is not classical at all, but we succeed in finding a convenient space where its transfer operator acts nicely, with a single dominant eigenvalue. With this new mixed system at hand, we provide a complete average-case analysis of the *CL* algorithm, with explicit constants.

The extended dynamical system and its properties are presented in Section 7.4. In particular, in Section 7.4.2 we discuss to a certain extent the appropriate probability measures on the dyadics, while in Section 7.4.3 we provide the properties of the transfer operator that are needed to complete the analysis. This becomes significant because the use of dyadics in Dynamical Analysis performed here is not commonplace. Even though there exist works studying other gcd algorithms that employ dyadics, namely the so-called “Turtle and the Hare” algorithm [DMDV05], and the Binary algorithm [Val98a], the *CL* algorithm evolves and uses the dyadics in a novel way.

Main results for the *CL* algorithm. The main result obtained is summarized here:

▷ **Theorem 7.2.** The mean value of the total number of iterations K and the total number of shifts S are asymptotically linear in the size $\log N$ as $N \rightarrow \infty$, and we provide explicit constants.

We remark that we are working on the so-called “real case” for the Continued Logarithm expansion. We consider a random real in the unit interval and, a given number k , we wish to describe the evolution of the main parameters associated with the expansion truncated after k steps, when the depth k tends to ∞ . This is a work still in progress.

Part I

Presentation of the general context

CHAPTER 1

CONTINUED FRACTIONS AND THE GAUSS MAP

We kick off by introducing several useful concepts that are recurrent in our studies and will prove fundamental for both the solution and conception of our problems.

1.1 The numeration process

Continued fractions can be introduced in several ways, and arise in contexts as seemingly diverse as diophantine approximation and Pell's equation. The most classical text describing extensively the elementary properties of continued fractions is definitely "Continued Fractions" by Khinchin [Khi97].

A continued fraction is a "formal" expression of the form

$$\frac{1}{m_1 + \frac{1}{m_2 + \frac{1}{m_3 + \ddots}}},$$

which we denote by $[m_1, m_2, \dots]$, where the coefficients m_1, m_2, \dots , known as quotients or partial quotients, are positive integers.

We can make sense of this as a limit of the so-called *convergents*

$$[m_1, \dots, m_k] := \frac{1}{m_1 + \frac{1}{\ddots + \frac{1}{m_k}}}, \quad (1.1)$$

that is, the truncated expansion considering only the first k coefficients. This can be realized as a finite continued fraction, or by filling in with 0s as follows $[m_1, \dots, m_k, 0, 0, 0 \dots]$, thus explaining the notation.

More precisely define

$$[m_1, m_2, \dots] := \lim [m_1, \dots, m_k],$$

and this limit is well-defined (see e.g., [Khi97]), for any choice of quotients $m_1, m_2, \dots \geq 1$. We will now describe how the continuants evolve, and in fact derive the existence of the limits too, as well as other interesting properties.

The finite continued fraction $[m_1, \dots, m_k]$ represents a rational number

$$\frac{p_k}{q_k} = \frac{1}{m_1 + \frac{1}{\dots + \frac{1}{m_k}}}, \quad \gcd(p_k, q_k) = 1, \quad (1.2)$$

where we enforce that p_k and q_k to be coprime to make the choice unique.

Note p_k and q_k defined in (1.2) depend only on the vector $(m_1, \dots, m_k) \in \mathbb{Z}_{\geq 1}^k$ and hence we will often write $p(m_1, \dots, m_k)$ and $q(m_1, \dots, m_k)$ to underline this dependence, or even $p_k(\mathbf{m})$ and $q_k(\mathbf{m})$ when the whole sequence $\mathbf{m} = (m_1, \dots, m_k, \dots)$ is fixed beforehand, thus emphasizing the “truncation” aspect. We write simply p_k and q_k as above when there is no danger of confusion.

The limit $[m_1, m_2, \dots]$ in (1.1) actually exists for any choice of coefficients $(m_k)_{k=1}^{\infty}$ (this is proved e.g., in [Khi97, Theorem 10, p.10]) and represents a real number $\alpha \in [0, 1]$. Conversely, every real number $\alpha \in (0, 1]$ has a continued fraction expansion

$$\alpha = \frac{1}{m_1 + \frac{1}{m_2 + \dots}}, \quad (1.3)$$

and we write $m_i = m_i(\alpha)$ when there is danger of confusion. This expansion is unique when α is irrational and, of course, necessarily infinite. For rationals we have two expansions, both finite (we will point out why later on), by means of the equality

$$[m_1, \dots, m_{k-1}, m_k] = [m_1, \dots, m_{k-1}, m_k - 1, 1] \quad (1.4)$$

which holds for $m_k \geq 2$.

Thus rationals have two finite expansions; one of them ending with the digit 1. It is direct to see that this is the only possible “redundancy” in the continued fraction expansion (this can be seen by direct comparisons), somewhat analogously to the case of the binary base representation, where the redundancies come from the cases of the form $0.a_1 \dots a_k 100 \dots = 0.a_1 \dots a_k 011 \dots$. This is to say, if two continued fraction expansions represent the same real number, then we are necessarily in the case (1.4) described above.

When considering the quotients $m_1(\alpha), m_2(\alpha), \dots$ coming from the expansion of α , we will also write $p_k(\alpha)$ and $q_k(\alpha)$ to denote the numerators and denominators of the convergents. The sequence $(q_k(\alpha))_{k=1}^{\infty}$ of denominators is known as the sequence of *continuants*, and plays a fundamental role in our studies.

Given $\alpha \in (0, 1)$ it will be useful to explain how its expansion is computed. We first note that if equality (1.3) is to hold, then

$$\frac{1}{\alpha} = m_1 + [m_2, m_3, \dots]$$

which implies $m_1(\alpha) = \lfloor \frac{1}{\alpha} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the floor function. Then the continued fraction $[m_2, m_3, \dots]$ corresponds, in its turn, to the rational part $\{1/\alpha\}$ and the procedure is iterated.

Thus we may think of this procedure as a dynamical system producing the digits $m_1(\alpha), m_2(\alpha), \dots$

1.1.1 The Gauss map

The process of computing a continued fraction expansion, its successive digits, can be described somewhat more.

Definition 1.1 (Gauss map). Let $\mathcal{I} := (0, 1)$, we consider the shift map T_g , called the *Gauss map*, defined by

$$T_g: \mathcal{I} \rightarrow \mathcal{I}, \quad \alpha \mapsto \left\{ \frac{1}{\alpha} \right\}, \quad (1.5)$$

where $\{\cdot\}$ denotes the fractional part $\{\alpha\} := \alpha - \lfloor \alpha \rfloor$. We further define the digit function m by

$$m: \mathcal{I} \rightarrow \mathbb{N}, \quad \alpha \mapsto \left\lfloor \frac{1}{\alpha} \right\rfloor. \quad (1.6)$$

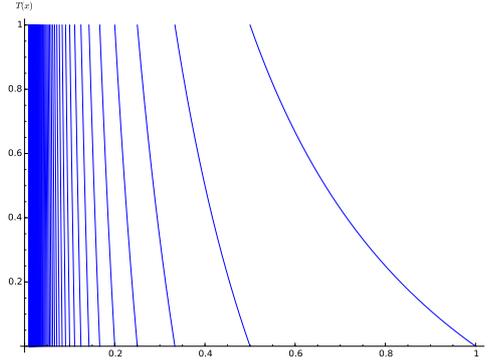


Figure 1.1: The Gauss map

We remark then that the digits $m_i(\alpha)$ are retrieved from the equality $m_i(\alpha) = m(T_g^{i-1}\alpha)$, thus making the continued fraction expansion a coding of the trajectory $\{\alpha, T_g\alpha, T_g^2\alpha, \dots\}$ by m . Note that this is well defined when $\alpha \in \mathcal{I} \setminus \mathbb{Q}$, while the trajectory will be finite when α is rational, as we will soon explain.

The evolution of the orbits through the Gauss map T_g from Definition 1.1 constitutes a fundamental example in dynamical systems. We extend the notions given for the Gauss map to more general dynamical systems in Section 1.2, where we define complete interval dynamical systems in Definition 1.3.

Observation 1.1. It is important to remark that the Gauss map T_g and the digit function m are defined so that

$$\alpha = \frac{1}{m_1(\alpha) + T_g(\alpha)}. \quad (1.7)$$

This equation may be iterated giving

$$\alpha = \frac{1}{m_1(\alpha) + \frac{1}{m_2(\alpha) + \frac{1}{\ddots + \frac{1}{m_k(\alpha) + T_g^k \alpha}}}}, \quad (1.8)$$

which we denote by $[m_1(\alpha), m_2(\alpha), \dots, m_{k-1}(\alpha), m_k(\alpha) + T_g^k \alpha]$ in a little abuse of notation (as we are introducing non-integer coefficients).

1.1.2 Continued fractions and the Euclidean Algorithm

The Euclidean Algorithm computes the greatest common divisor (gcd) of a pair of positive integers $a \leq b$ by exploiting the equality $\gcd(a, b) = \gcd(r, a)$ where $r = b \bmod a$ is the remainder of the division of b by a , and then proceeding likewise with (r, a) until we get to a pair $(0, g)$ for which it is clear that $\gcd(0, g) = g$.

This algorithm is very efficient, having linear complexity (on the bit-length) of a and b . Here we explain this fact briefly, by showing that there is a fundamental connection between the Euclidean Algorithm and continued fractions of rational numbers.

Continued fractions are naturally equivalent to the execution of the Euclidean algorithm. Indeed, let us pick a rational number a/b . One step of the expansion gives:

$$\frac{a}{b} = \frac{1}{b/a} = \frac{1}{[b/a] + \{b/a\}},$$

and here $\{b/a\} = \frac{b \bmod a}{a}$ so that our original pair (a, b) has become $(b \bmod a, a)$ after one iteration, exactly like in the Euclidean Algorithm. As we know, the Euclidean algorithm certainly terminates, as the first entry always decreases strictly until becoming 0 (clearly $b \bmod a < a$), thus giving a finite representation

$$\frac{a}{b} = \frac{1}{m_1 + \frac{1}{\ddots + \frac{1}{m_k}}},$$

where m_1, m_2, \dots are the quotients of the divisions in the Euclidean algorithm applied to (a, b) !

In particular, as (1.4) is the only redundancy in the representation, rational numbers have only finite expansions (exactly two of them). Second, this means that one may study the Euclidean algorithm by studying the continued fraction expansion or viceversa.

1.1.3 Basic properties of continuants

We now develop important properties regarding the convergents p_k/q_k , providing information regarding the growth of the sequence of continuants q_k . These will prove useful in proving that $\alpha = [m_1(\alpha), m_2(\alpha), \dots]$ indeed holds as expected, as the convergence rate of $[m_1(\alpha), m_2(\alpha), \dots, m_k(\alpha)]$ towards α will be dictated by the size of q_k as we explain in Proposition 1.4. Along the way we will introduce several properties which are fundamental to our studies.

We start off by studying the recurrence equation satisfied by the sequences $(p_k)_k$ and $(q_k)_k$, providing its matricial form too.

Proposition 1.1. *Let $(m_k)_{k=1}^\infty$ be a sequence of positive integers. The sequences $(p_k)_{k=1}^\infty \subset \mathbb{N}$ and $(q_k)_{k=1}^\infty \subset \mathbb{N}$ of successive numerators and denominators of the continued fraction expansion, defined by*

$$\frac{p_k}{q_k} = [m_1, \dots, m_k], \quad \gcd(p_k, q_k) = 1,$$

satisfy the recurrences

$$p_{k+1} = m_{k+1}p_k + p_{k-1}, \quad q_{k+1} = m_{k+1}q_k + q_{k-1}, \quad (1.9)$$

for all $k \geq 0$, where we consider $p_0 = 0, p_{-1} = 1$ and $q_0 = 1, q_{-1} = 0$.

This can be written in matricial form as

$$\begin{pmatrix} q_{k+1} & p_{k+1} \\ q_k & p_k \end{pmatrix} = \begin{pmatrix} m_{k+1} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_k & p_k \\ q_{k-1} & p_{k-1} \end{pmatrix}, \quad (1.10)$$

along with

$$\begin{pmatrix} q_0 & p_0 \\ q_{-1} & p_{-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (1.11)$$

Proof. For convenience will prove the proposition in matricial form. Observe that the result is equivalent to

$$\begin{pmatrix} q(m_1, \dots, m_k) & p(m_1, \dots, m_k) \\ q(m_1, \dots, m_{k-1}) & p(m_1, \dots, m_{k-1}) \end{pmatrix} = \begin{pmatrix} m_k & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} m_1 & 1 \\ 1 & 0 \end{pmatrix}, \quad (1.12)$$

for all $k \geq 0$ and $(m_1, \dots, m_k) \in \mathbb{N}^k$. This expression is particularly useful as what we effectively do in the inductive step is add the coefficient m_1 to the beginning of $(m_2, \dots, m_k) \in \mathbb{N}^k$, for which we assume the result to hold by induction.

The proof proceeds by strong induction over k . It is clear that the base case $p_0 = 0, p_{-1} = 1$ and $q_0 = 1, q_{-1} = 0$ holds for any choice of the quotients. Now assume that the recurrence holds for j up to k , for any choice of quotients m_1, \dots, m_k (strong induction). Let us consider now a concrete m_1, \dots, m_k, m_{k+1} and show (1.12) for tuple of length $k + 1$.

We have

$$\frac{p(m_1, \dots, m_{k+1})}{q(m_1, \dots, m_{k+1})} = \frac{1}{m_1 + \frac{p(m_2, \dots, m_{k+1})}{q(m_2, \dots, m_{k+1})}} = \frac{q(m_2, \dots, m_{k+1})}{m_1 q(m_2, \dots, m_{k+1}) + p(m_2, \dots, m_{k+1})},$$

and therefore $p(m_1, \dots, m_{k+1}) = q(m_2, \dots, m_{k+1})$ as well as $q(m_1, \dots, m_{k+1}) = m_1 q(m_2, \dots, m_{k+1}) + p(m_2, \dots, m_{k+1})$ because their gcd equals $\gcd(q(m_2, \dots, m_{k+1}), p(m_2, \dots, m_{k+1})) = 1$.

As the previous equalities will also hold when we substitute $k \mapsto k - 1$, we have

$$\begin{pmatrix} q(m_1, \dots, m_{k+1}) & p(m_1, \dots, m_{k+1}) \\ q(m_1, \dots, m_k) & p(m_1, \dots, m_k) \end{pmatrix} = \begin{pmatrix} q(m_2, \dots, m_{k+1}) & p(m_2, \dots, m_{k+1}) \\ q(m_2, \dots, m_k) & p(m_2, \dots, m_k) \end{pmatrix} \begin{pmatrix} m_1 & 1 \\ 1 & 0 \end{pmatrix}.$$

It follows from the inductive hypothesis that

$$\begin{pmatrix} q(m_2, \dots, m_{k+1}) & p(m_2, \dots, m_{k+1}) \\ q(m_2, \dots, m_k) & p(m_2, \dots, m_k) \end{pmatrix} = \begin{pmatrix} m_{k+1} & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} m_2 & 1 \\ 1 & 0 \end{pmatrix},$$

thus

$$\begin{pmatrix} q(m_1, \dots, m_{k+1}) & p(m_1, \dots, m_{k+1}) \\ q(m_1, \dots, m_k) & p(m_1, \dots, m_k) \end{pmatrix} = \begin{pmatrix} m_{k+1} & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} m_2 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} m_1 & 1 \\ 1 & 0 \end{pmatrix},$$

which proves the result for $k + 1$. ■

An immediate corollary of the recurrence is that the sequence of continuants grows at least exponentially.

Corollary 1.1. *Let $(m_k)_{k=1}^{\infty}$ be a sequence of positive integers. The sequence $(q_k)_{k=1}^{\infty} \subset \mathbb{N}$ of successive continuants of the continued fraction expansion satisfies*

$$q_k \geq 2^{(k-1)/2}.$$

Proof. Observe that $q_{k+1} \geq 2q_{k-1}$ by the recurrence. ■

Observation 1.2 (Precise bound). This last inequality can be made somewhat more precise by considering that $q_{k+1} \geq q_k + q_{k-1}$. Then by induction we conclude that $q_k \geq f_k$ where f_k is the k -th Fibonacci number, defined from $f_0 = 0, f_1 = 1$ and $f_{j+1} = f_j + f_{j-1}$ for $j \geq 1$. Equality can only hold for each k if $\alpha = [1, 1, 1, \dots]$ which then satisfies $\alpha = 1/(1 + \alpha)$ so that $\alpha = (\sqrt{5} - 1)/2$.

Finally, we recall that $f_k = \lfloor \Phi^k / \sqrt{5} \rfloor$ where $\lfloor \cdot \rfloor$ is the ‘‘round to the nearest integer’’ function and $\Phi = \frac{\sqrt{5}+1}{2}$ is the Golden ratio. This means that $q_k(\alpha) \geq \Phi^{k-2}$ for all $k \geq 1$. ◇

Notice that the previous corollary gives a bound for the depth $K(a, b)$ of the continued fraction expansion of a reduced rational number a/b , as then $b = q_{K(a,b)} \geq 2^{(K(a,b)-1)/2}$ and therefore $K(a, b) \leq 2 \log_2 b + 1$.

By definition we have that $\gcd(p_k, q_k) = 1$, and the recurrence in Proposition 1.1 implies also $\gcd(q_{k-1}, q_k) = 1$ and $\gcd(p_{k-1}, p_k) = 1$. All of these greatest common divisors can be deduced at once too from the following ‘‘determinant’’ property of the convergents.

Corollary 1.2 (Determinant). *Let $(m_k)_{k=1}^\infty$ be a sequence of positive integers. The sequences $(p_k)_{k=1}^\infty \subset \mathbb{N}$ and $(q_k)_{k=1}^\infty \subset \mathbb{N}$ of successive numerators and denominators of the continued fraction expansion satisfy*

$$q_k p_{k-1} - p_k q_{k-1} = (-1)^k, \quad (1.13)$$

for all $k \geq 0$.

Proof. Take determinants in (1.10). ■

The determinant property (1.13) gives a non-trivial relation between the sequences $(p_k)_k$ and $(q_k)_k$ of numerators and denominators. For instance, notice that

$$\frac{p_{k-1}}{q_k} = \frac{p_k}{q_k} \frac{q_{k-1}}{q_k} + \frac{(-1)^k}{q_k^2} = \frac{p_k}{q_k} \frac{q_{k-1}}{q_k} + O(2^{-k}).$$

Another property that will play a fundamental role in our studies is the so-called ‘‘mirror property’’ (described for instance in [AA07]) which tells us what happens when we consider the convergents of the mirror sequence $(m_1, \dots, m_k) \mapsto (m_k, \dots, m_1)$: the numerator p_k becomes the continuant q_{k-1} .

Corollary 1.3 (Mirror property). *Let $(m_k)_{k=1}^\infty$ be a sequence of positive integers*

$$q(m_k, m_{k-1}, \dots, m_1) = q(m_1, \dots, m_k), \quad p(m_k, m_{k-1}, \dots, m_1) = q(m_1, \dots, m_{k-1}), \quad (1.14)$$

for all $k \geq 0$.

Proof. Transpose the matrices. ■

This tells us at once that $\gcd(q_{k-1}, q_k) = 1$ is no surprise: actually q_{k-1} and q_k give the reduced convergent for the mirror sequence!

1.1.4 Inverse branches of the Gauss map

To actually get to the matters of convergence, it is important to relate the partial expansions $[m_1, \dots, m_k]$ to the complete expansion $[m_1, m_2, \dots]$ (which equals our number $\alpha \in \mathcal{I}$).

We recall that $\alpha \in \mathcal{I}$ itself is of the form $\alpha = [m_1(\alpha), \dots, m_{k-1}(\alpha), m_k(\alpha) + z]$ for $z = T_g^k \alpha \in [0, 1]$ (see Equation 1.8), making significant the function $z \mapsto [m_1(\alpha), \dots, m_{k-1}(\alpha), m_k(\alpha) + z]$.

Definition 1.2 (Inverse branches). The *inverse branches* of the system (T, \mathcal{I}) defined in Definition 1.1 are given by

$$h_m(x) := \frac{1}{m+x}, \quad \mathcal{H} := \{h_m : m \in \mathbb{N}\}. \quad (1.15)$$

While the depth k inverse branches, equivalently, the inverse branches of T^k , are given by

$$h_{m_1, \dots, m_k}(x) := h_{m_1} \circ \dots \circ h_{m_k}(x), \quad \mathcal{H}^k := \{h_m : (m_1, \dots, m_k) \in \mathbb{N}^k\}. \quad (1.16)$$

Later on in Section 1.2 we shall define the concept of inverse branches for more general dynamical systems having a countable number of complete branches.

Notice that by definition $h_{m_1, \dots, m_k}(x) = [m_1, \dots, m_{k-1}, m_k + x]$.

Proposition 1.2. *Let $(m_k)_{k=1}^\infty$ be a sequence of positive integers. Consider the sequences $(p_k)_{k=1}^\infty \subset \mathbb{N}$ and $(q_k)_{k=1}^\infty \subset \mathbb{N}$ of successive numerators and denominators of the continued fraction expansion associated with the quotients $(m_k)_{k=1}^\infty$, then*

$$h_{m_1, \dots, m_k}(z) = \frac{p_k + z p_{k-1}}{q_k + z q_{k-1}}. \quad (1.17)$$

Proof. We proceed by induction. The result is clear for $k = 0$ as $q_0 = p_{-1} = 1$ and $p_0 = q_{-1} = 0$ by definition. Assume the result to hold for k , we will prove it for $k + 1$. Indeed

$$\begin{aligned} [m_1, \dots, m_k, m_{k+1} + z] &= \left[m_1, \dots, m_k + \frac{1}{m_{k+1} + z} \right] \\ &= \frac{p_k + \frac{p_{k-1}}{m_{k+1} + z}}{q_k + \frac{q_{k-1}}{m_{k+1} + z}} \\ &= \frac{m_{k+1}p_k + p_{k-1} + zp_k}{m_{k+1}q_k + q_{k-1} + zq_k}, \end{aligned}$$

and the result follows from Proposition 1.1. ■

Corollary 1.4. *Let m_1, m_2, \dots be the quotients in the continued fraction expansion of $\alpha \in \mathcal{I}$, then for every $k \geq 0$ we have*

$$\alpha = h_{m_1, \dots, m_k}(T^k \alpha) = \frac{p_k + (T_g^k \alpha)p_{k-1}}{q_k + (T_g^k \alpha)q_{k-1}}. \quad (1.18)$$

We remark that the inverse branch $h(z) = [m_1, \dots, m_k + z]$ satisfies

$$h'(z) = \frac{(-1)^k}{(q_k + zq_{k-1})^2},$$

thanks to (1.13). Note then that $|h'(x)| \leq 1/q_k^2$ for all $x \in \mathcal{I}$.

It is important to underline the simplicity of the formula for $h'(z)$, which will play a crucial role in the results to come. In particular, we draw the attention to the equality $|h'(0)|^{-1/2} = q_k$, which yields an analytic formula for the denominators purely in terms of the inverse branches $h \in \mathcal{H}^*$.

We also note that the sign of the derivative $h'(z)$ is not that surprising. Indeed, the inverse branch $h_a: x \mapsto \frac{1}{a+x}$ is decreasing for any a , hence $h_{m_1} \circ h_{m_2}$ is increasing, $h_{m_1} \circ h_{m_2} \circ h_{m_3}$ decreasing, and so on. A simple, but very informational consequence of this property is

$$\frac{p_0}{q_0} < \frac{p_2}{q_2} < \dots < \alpha < \dots < \frac{p_3}{q_3} < \frac{p_1}{q_1}, \quad (1.19)$$

namely that the even convergents of α form an increasing subsequence, while the odd convergents constitute a decreasing subsequence. Observe in particular that this means that

$$|\alpha q_k - p_k| = (-1)^k (\alpha q_k - p_k). \quad (1.20)$$

1.1.5 Fundamental Intervals

The monotonicity of the inverse branches $h = h_{m_1, \dots, m_k} \in \mathcal{H}^k$ implies that the cylinder $\mathcal{I}_{m_1, \dots, m_k} := h(\mathcal{I})$, which will consist of all reals from \mathcal{I} having a continued fraction expansion starting with the digits m_1, \dots, m_k , is an interval with endpoints $h(0) = \frac{p_k}{q_k}$ and $h(1) = \frac{p_k + p_{k-1}}{q_k + q_{k-1}}$, which has length

$$|\mathcal{I}_{m_1, \dots, m_k}| = |h(1) - h(0)| = \frac{1}{q_k(q_k + q_{k-1})}. \quad (1.21)$$

The intervals $\mathcal{I}_{m_1, \dots, m_k} = h_{m_1, \dots, m_k}(\mathcal{I})$ are known as the *fundamental intervals* of the dynamical system.

We summarize the previous discussion.

Proposition 1.3 (Fundamental intervals). *The inverse branches $h = h_{m_1, \dots, m_k} \in \mathcal{H}^k$ of the Euclidean dynamical system are monotonic. The fundamental interval $\mathcal{I}_{m_1, \dots, m_k} := h_{m_1, \dots, m_k}(\mathcal{I})$, which is the set of all reals from \mathcal{I} having a continued fraction expansion starting with the digits m_1, \dots, m_k , has length*

$$|\mathcal{I}_{m_1, \dots, m_k}| = |h(1) - h(0)| = \frac{1}{q_k(q_k + q_{k-1})}.$$

More generally, the monotonicity implies $|h([x, y])| = |h(y) - h(x)|$.

In particular, from (1.21) we deduce

$$\left| \alpha - \frac{p_k}{q_k} \right| \leq \frac{1}{q_k^2} \leq \frac{2}{2^k},$$

but we can be even more precise

Proposition 1.4. *Let $\alpha \in \mathcal{I}$ be a real number having a continued fraction expansion starting with the digits $m_1, \dots, m_k \in \mathbb{N}$, then we have the inequalities*

$$\frac{1}{q_k(q_k + q_{k+1})} \leq \left| \alpha - \frac{p_k}{q_k} \right| \leq \frac{1}{q_k q_{k+1}}.$$

Proof. Apply (1.13) and (1.18) substituting $k \mapsto k + 1$. ■

As a consequence our convergents $[m_1(\alpha), m_2(\alpha), \dots, m_k(\alpha)]$ constitute very good approximations to α , the rate of convergence being exponential. In any case something much stronger holds. Continued fractions in fact correspond to optimal approximations of reals by rational numbers. This is explained in detail e.g., in [Khi97, Theorems 16 and 17] which we cite here

Proposition 1.5. *Let $\alpha \in \mathcal{I} = (0, 1)$, $\alpha \neq 1/2$. Then the convergents of α satisfy (if the index k does not exceed the depth of α , if the number α were rational)*

$$|\alpha p_k(\alpha) - q_k(\alpha)| < |\alpha p - q|,$$

for all $(p, q) \in \mathbb{N}$, with $p/q \neq p_k/q_k$ and $0 < q \leq q_k(\alpha)$.

We remark that, for $\alpha = 1/2$, we have $|1 \cdot \alpha - 1| = |1 \cdot \alpha - 0|$, and here $1/1$ is not a convergent by definition, but this is, in fact, the only exceptional case.

It is therefore said that the convergents correspond to the best approximants of the “second-kind” p/q , minimizing $|\alpha p - q|$ when one bounds q . Proposition 1.5 then tells us that continued fractions do occur naturally when we attempt to approximate reals by rationals.

Observation 1.3 (Dirichlet’s Theorem). The inequalities in Proposition 1.4 yield the existence of good approximants to $\alpha \in (0, 1)$ in the following sense: for any $Q \geq 1$, there are $p \in \mathbb{Z}$ and $q \in \mathbb{N}$ such that

$$\begin{cases} |q\alpha - p| & \leq 1/Q \\ q & \leq Q. \end{cases}$$

Indeed, pick $(p, q) = (p_k, q_k)$ where k is the index such that $q_k \leq Q < q_{k+1}$. This result is known as “Dirichlet’s Theorem”, and may also be proved as a direct application of the “pigeon-hole principle” (see for instance in [HW08]).

1.2 Dynamical systems and the Perron-Frobenius operator

In (discrete-time) dynamical systems we are interested in the evolution of the discrete orbit of a map $f: X \rightarrow X$ on a set X , more precisely, the orbit $\{x, f(x), f^2(x), \dots\}$ for an initial point $x \in X$. When the set X

is given some extra structure, be it topological, or a weight (measure), we would like to know how these evolve in time, along the orbits, for given initial points $x \in X$.

In general we will be concerned with a concrete (and widely applicable) kind of dynamical systems called the *interval dynamical systems with complete branches*. These systems are naturally associated with the process of numeration, they give a coding of the orbits which characterizes the initial point x as is the case of continued fractions or the decimal expansion.

In Section 1.3 we will consider properties related to the so called *measure theoretic dynamical systems*. A measure theoretic dynamical system is a tuple (Ω, T, Σ, μ) consisting of a compact set Ω , a σ -algebra Σ , a measurable map $T: (\Omega, \Sigma) \rightarrow (\Omega, \Sigma)$ and a probability measure μ on Σ . Such systems are quite general and give the context for Ergodic theory.

1.2.1 A general definition of a dynamical system

In this section we introduce the general notion of a dynamical system associated with a numeration process. Of course, numeration involves “digits” (a.k.a., letters) from a countable set \mathcal{A} , known as the alphabet. This alphabet is \mathbb{N} for the case of continued fractions. Each letter of the alphabet codes information with regard to the position of the orbit at a given moment of time.

Definition 1.3 (Interval dynamical systems). An interval dynamical system of class \mathcal{C}^k is defined from the following elements

- a countable set \mathcal{A} , known as the *alphabet*.
- disjoint open intervals I_a for $a \in \mathcal{A}$, such that $\mathcal{I} = \bigcup_{a \in \mathcal{A}} \overline{I_a}$ (a *topological partition*).
- a *coding map* $\sigma: \mathcal{I} \rightarrow \mathcal{A}$ satisfying $\sigma(x) = a$ for $x \in I_a$.
- a map $T: \mathcal{I} \rightarrow \mathcal{I}$, called the *shift map*, which satisfies that $T|_{I_a}: I_a \rightarrow J_a := T(I_a)$ is bijective for each $a \in \mathcal{A}$ and of class $\mathcal{C}^k(I_a)$. The inverses $h_a := T|_{J_a}^{-1}: J_a \rightarrow I_a$ are known as the *inverse branches* of T and we denote $\mathcal{H} := \{h_a : a \in \mathcal{A}\}$.

Observe that the inverse branches $h \in \mathcal{H}$ are necessarily monotone, as they are bijections. This means that $|I_a| = |h(B) - h(A)|$ if $J_a = (A, B)$.

Such a dynamical system is called *complete* if $J_a = \mathcal{I}$ for each $a \in \mathcal{A}$, and is called *markovian* when for each $a \in \mathcal{A}$, the images J_a are made out of the elements from the topological partition: $J_a = \bigcup_{a' \in S_a} I_{a'}$ for some set $S_a \subset \mathcal{A}$. That is, an interval system is markovian when $I_a \cap J_b \neq \emptyset$ implies $I_a \subset J_b$.

Observe that for a markovian system we have that $\sigma(x) = a$ implies $\sigma(Tx) \in S_a$, hence the name.

Given an initial point $x \in \mathcal{I}$ we say that $\mathcal{T}(x) := (x, Tx, T^2x, \dots)$ is the orbit of x and that $M(x) := (\sigma x, \sigma(Tx), \sigma(T^2x) \dots)$ is the infinite word (coding) associated to x .

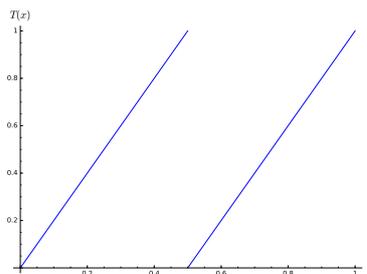


Figure 1.2: The shift map of the binary dynamical system

In what follow we will only consider *complete* interval dynamical systems of class at least \mathcal{C}^1 .

Example 1.1. To give a first example of such a complete dynamical system let us consider the binary

expansion:

- Consider the alphabet $\mathcal{A} = \{0, 1\}$.
- The partition $I_0 = (0, 1/2)$ and $I_1 = (1/2, 1)$.
- The map $T(x) = 2x \pmod{1}$.

Then $M(x)$ is the base two representation of x and the inverse branches are given by

$$h_0(x) = x/2, \quad h_1(x) = (x + 1)/2.$$

We remark that numbers from $h_1(\mathcal{I})$ begin their binary expansion by 1 while those in $h_0(\mathcal{I})$ begin their binary expansion by 0. Similarly, for $d_1, \dots, d_k \in \mathcal{A}$ the sets

$$h_{d_1} \circ \dots \circ h_{d_k}(\mathcal{I}),$$

correspond to the $x \in \mathcal{I}$ having $M(x)$ starting with d_1, \dots, d_k , i.e.,

$$h_{d_1} \circ \dots \circ h_{d_k}(\mathcal{I}) = \{x \in \mathcal{I} : (\sigma(x), \sigma(Tx), \dots, \sigma(T^{k-1}x)) = (d_1, \dots, d_k)\}$$

It is easy to see that this simple remark applies to all (complete) interval dynamical systems. \diamond

Definition 1.4 (Fundamental intervals). Consider a complete interval dynamical system of class \mathcal{C}^k with alphabet \mathcal{A} and inverse branches $\mathcal{H} = \{h_a : a \in \mathcal{A}\}$.

The fundamental interval $\mathcal{I}_{a_1, \dots, a_k}$ associated with the digits a_1, \dots, a_k is defined by

$$\mathcal{I}_{a_1, \dots, a_k} := h_{a_1} \circ \dots \circ h_{a_k}(\mathcal{I}). \quad (1.22)$$

This is indeed an interval as the inverse branches are continuous. We also note that $\mathcal{I}_{a_1, \dots, a_k}$ corresponds exactly to the reals $x \in \mathcal{I}$ that have a coding beginning with a_1, \dots, a_k .

Definition 1.5 (Expanding maps). A map $T: \mathcal{I} \rightarrow \mathcal{I}$ is said to be expanding if and only if there exists a topological partition of \mathcal{I} into countable disjoint open intervals $(I_a)_{a \in \mathcal{A}}$ such that

- the restriction $T_a|_{I_a}: I_a \rightarrow \mathcal{I}$, $T_a := T|_{I_a}$ is monotone and $C^1(I_a)$ for each $a \in \mathcal{A}$.
- $|(T^n)'| \geq \delta > 1$ for some $\delta > 0$ and $n \in \mathbb{N}$.

Proposition 1.6. Consider a complete interval dynamical system of class \mathcal{C}^j , $j \geq 1$ associated with the map T . If T is expanding, then the lengths of the fundamental intervals satisfy $|\mathcal{I}_{a_1, \dots, a_k}| \rightarrow_k 0$ uniformly.

A consequence of this proposition is that the fundamental intervals generate the so-called *Borel* σ -algebra of \mathcal{I} . The Borel σ -algebra of \mathcal{I} is the one generated by the open intervals of \mathcal{I} . Such property is key for several proofs because, when this is the case, it is enough to prove many things just over the fundamental intervals. We will come back to this property later on in Section 1.3.

1.2.2 Dynamical systems of interest

In this thesis there are three fundamental dynamical systems: the euclidean dynamical system (associated with the Gauss map), the rotation dynamical system and the Continued Logarithm dynamical system. We have already discussed the Euclidean system, here we introduce the other two.

Example 1.2 (Circle rotations). We consider a unit length circle, more precisely $\mathbb{T}^1 = \mathbb{R}/\mathbb{Z}$, which corresponds to the interval $[0, 1]$ when we identify the points 0 and 1, and a rotation angle $\alpha \in [0, 1)$.

The associated rotation is given by $\mathcal{R}_\alpha(x) := (x + \alpha) \pmod{1}$. This means that $\mathcal{R}_\alpha(x) = x + \alpha$ for $x \in [0, 1 - \alpha)$ and $\mathcal{R}_\alpha(x) = x + \alpha - 1$ otherwise (in an abuse of notation identifying $[0, 1)$ with \mathbb{T}^1).

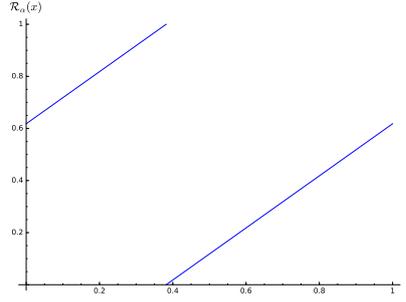


Figure 1.3: The rotation $\mathcal{R}_\alpha(x)$ of angle $\alpha = (\sqrt{5} - 1)/2$.

For reasons that will be made clear in Section 3.2, we introduce the alphabet $\mathcal{A} = \{0, 1\}$ and the topological partition given by $I_0 = [0, 1 - \alpha)$ (or $(0, 1 - \alpha]$) and $I_1 = [1 - \alpha, 1)$ (respectively $(1 - \alpha, 1]$ taking $0 \equiv 1$).

We note that the shift map \mathcal{R}_α is clearly invertible as $\mathcal{R}_\alpha^{-1} = \mathcal{R}_{-\alpha}$, but this system is not complete (note $\mathcal{R}_\alpha(I_a) \neq \mathcal{I}$) and cannot be made even markovian when $\alpha \notin \mathbb{Q}$.

Example 1.3 (Continued logarithm). The continued logarithm expansion is a mutation of the classical continued fractions, introduced by Gosper [Sha16] in Hakmem.

Formally, the expansion of a number $\alpha \in \mathcal{I}$ is of the form

$$\alpha = \frac{2^{-a_1}}{1 + \frac{2^{-a_2}}{\ddots}}, \quad a_1, a_2, \dots \geq 0,$$

and, accordingly, reversing the process, the associated shift T_c is given by

$$T_c(x) = \frac{2^{-a}}{x} - 1, \quad x \in I_a := (2^{-a-1}, 2^{-a}),$$

where the symbols a belong to the alphabet $\mathcal{A} = \{0, 1, 2, \dots\}$. This system is complete and of class \mathcal{C}^∞ , with inverse branches $h_a: \mathcal{I} \rightarrow I_a$ given by

$$h_a(x) = \frac{1}{2^a(1+x)}.$$

We will come back to this system in Chapter 7, when we will study its associated gcd algorithm.

1.2.3 Pushforward measure – Invariant measure

Almost every dynamical system of interest is what is informally called “chaotic”. In such systems it is difficult to predict anything significant regarding the position of $T^k(x)$, for large k , without knowing x itself. This is why it is more natural to consider the effect of dynamical systems when an initial distribution or density is given, and study how these evolve as we iterate with the shift map T . The pushforward measure describes precisely how a (probability) measure μ evolves after one iteration of a measurable map T .

Definition 1.6 (Pushforward measure, Measure preserving transformation, invariant measure). Given measurable spaces (Ω_1, Σ_1) and (Ω_2, Σ_2) , i.e., Σ_i is a σ -algebra over Ω_i , a measurable map $T: (\Omega_1, \Sigma_1) \rightarrow (\Omega_2, \Sigma_2)$ and a measure $\mu: \Sigma_1 \rightarrow [0, \infty]$, we define the pushforward measure $T_*\mu$ by

$$(T_*\mu)(B) := \mu(T^{-1}B), \quad (1.23)$$

for $B \in \Sigma_2$.

If the spaces coincide, i.e., $(\Omega_1, \Sigma_1) = (\Omega_2, \Sigma_2)$, the map T is said to *preserve the measure* μ if $T_*\mu = \mu$. When a probability space $(\Omega_1, \Sigma_1, \mu)$ is fixed before-hand, we will simply say that T is *measure preserving* when $T_*\mu = \mu$. Equivalently, it is said that the measure μ is invariant with respect to T .

The importance of finding invariant measures is realized intuitively from the stationary behavior for $T^k x$ for large k . If there were a kind of stationary distribution for $T^k x$ then we must have, loosely speaking $T_*(T_*^k \mu) \approx T_*^k \mu$ so that $T_*^k \mu$ approaches an invariant (our “stationary probability”) as $k \rightarrow \infty$.

Since (1.23) can be seen as $\int_{\Omega} \mathbf{1}_B(x) d(T_* \mu(x)) = \int_{\Omega} \mathbf{1}_{B \circ T}(x) d\mu(x)$ for all $B \in \Sigma$, it follows by measure-theoretic induction that

Proposition 1.7 (Integration with respect to a pushforward measure). *Consider a measure space (Ω, Σ, μ) and a measurable map $T: \Omega \rightarrow \Omega$. A measurable map $g: \Omega \rightarrow \mathbb{R}$ is integrable with respect to the pushforward measure $T_* \mu$ if and only if $g \circ T$ is integrable with respect to μ , and then we have the equality*

$$\int_{\Omega} g(x) d(T_* \mu(x)) = \int_{\Omega} g \circ T(x) d\mu(x).$$

Example 1.4. We return to the example of circle rotations on the unit circle $\mathbb{T}^1 = \mathbb{R}/\mathbb{Z}$ modulo 1.

For a rotation $\mathcal{R}_{\alpha}(x) := (x + \alpha) \bmod .1$. It is clear that the uniform (Lebesgue) measure on \mathbb{T}^1 is invariant.

When α rational, we may build other invariant measures. For instance, if $\alpha = p/q$ with $\gcd(p, q) = 1$, consider the measure μ given by

$$\mu(A) = \#\{a/q \in A : a \in \mathbb{Z}\}/q,$$

so that the measure is actually uniform on the points $0, 1/q, \dots, (q-1)/q$. Our rotation \mathcal{R}_{α} just permutes these points, hence the reason the measure μ is invariant.

Of course, the invariant measures we just constructed for rational α have “atoms”, i.e., measurable sets A such that any strict subset (measurable) of it has 0 measure. For irrationals, as the sequence $\alpha, 2\alpha \bmod 1, 3\alpha \bmod 1, \dots$ is dense on \mathcal{I} (we will prove this in the next section), we will have a very different behavior: the Lebesgue measure is the only (unique) invariant Borel measure. Intuitively speaking, we may approach any possible translation $I + a$ of an interval I by rotating I enough times. Hence I and $I + a$ should have the same measure, for any invariant measure, and this characterizes the Lebesgue measure. \diamond

1.2.4 The Perron Frobenius operator

We have described how the map T defining a (interval) dynamical system acts on a measure μ . For the cases of interval dynamical systems, we may speak more concretely about how it acts on a density with respect to the Lebesgue measure λ_{Leb} on \mathcal{I} (the uniform probability).

The way the densities evolve as we apply a map T is given by what is called the *Perron-Frobenius operator*, known as the density transformer.

Definition 1.7 (Perron-Frobenius operator). Consider an interval dynamical system of class \mathcal{C}^1 with inverse branches \mathcal{H} . Then the Perron-Frobenius operator $\mathbf{H}: C^1(\mathcal{I}) \rightarrow C^1(\mathcal{I})$ of the system is defined by

$$\mathbf{H}[g](x) = \sum_{h \in \mathcal{H}} |h'(x)| g(h(x)). \quad (1.24)$$

Proposition 1.8 (Density transformer). *Consider an interval dynamical system of class \mathcal{C}^1 with map T . If X is drawn from \mathcal{I} with density (w.r.t., the Lebesgue measure λ_{Leb}) g then $T(X)$ has density $\mathbf{H}[g]$ given by the Perron Frobenius operator.*

Proof. The proof is direct and is illustrated in Figure 1.4. \blacksquare

It is then clear that an invariant probability measure having a density g must then satisfy $\mathbf{H}[g] \equiv g$ and conversely, if $g \in L^1(\mathcal{I})$ satisfies $\mathbf{H}[g] \equiv g$, then the measure associated with this density is invariant.

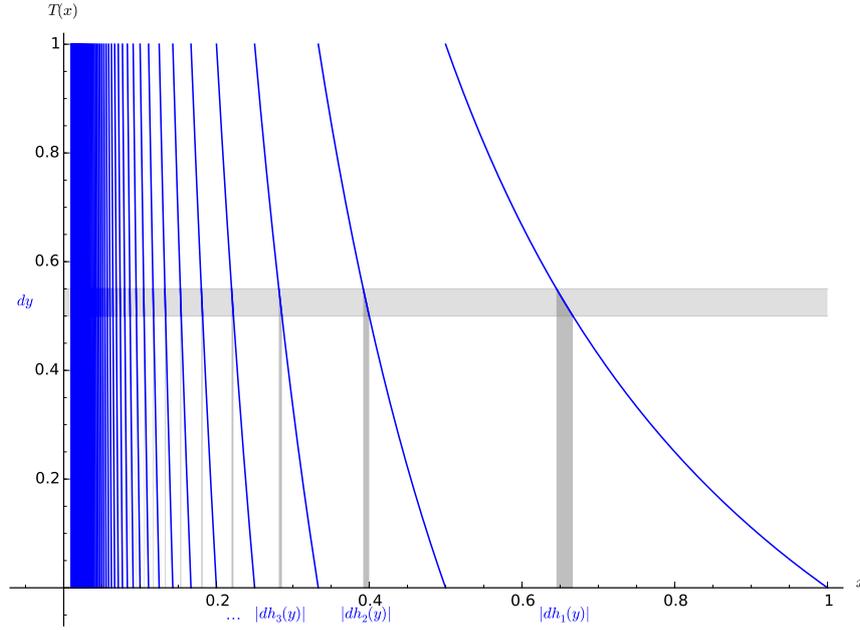


Figure 1.4: Illustration of the derivation of the density of $T_g x$ from that of x . The probability of $T_g x$ falling around y in a small interval of length “ dy ” is the sum of the probabilities of x falling within an interval of length “ $|dh_m(y)|$ ” from the respective points $h_m(y)$.

1.2.5 The case of the Gauss map; the Gauss density

Let us describe the case of the dynamical system associated to the Gauss map. As we know, this system has inverse branches

$$h_m(x) = \frac{1}{m+x}, \quad |h'_m(x)| = \frac{1}{(m+x)^2},$$

so that its associated Perron Frobenius operator is

$$\mathbf{H}[g](x) = \sum_{m=1}^{\infty} \frac{1}{(m+x)^2} g\left(\frac{1}{m+x}\right).$$

In the case of the Gauss map, as well as other interval dynamical systems, the iterates of the Perron Frobenius operator will play a big role, generating expressions of interest. We give here the formula for the Gauss map.

Proposition 1.9. *For the Euclidean dynamical system (\mathcal{I}, T_g) the iterates of the Perron-Frobenius operator take on the form*

$$\mathbf{H}^k[g](x) := \sum_{(m_1, \dots, m_k) \in \mathbb{N}^k} \frac{1}{(q_k(\mathbf{m}) + x q_{k-1}(\mathbf{m}))^2} g\left(\frac{p_k(\mathbf{m}) + x p_{k-1}(\mathbf{m})}{q_k(\mathbf{m}) + x q_{k-1}(\mathbf{m})}\right). \quad (1.25)$$

Proof. Recall Proposition 1.2. ■

To motivate the ensuing limiting density, known as the Gauss density, we observe that formally when $x \rightarrow \infty$ (of course, we do actually have the constraint $x \in \mathcal{I}$)

$$\mathbf{H}[g](x) \approx g(0) \sum_{m=1}^{\infty} \frac{1}{(m+x)^2} \approx g(0) \int_1^{\infty} \frac{dt}{(t+x)^2} = \frac{g(0)}{1+x},$$

which is indeed almost the right guess. Indeed

$$\mathbf{H}\left[t \mapsto \frac{1}{1+t}\right](x) = \sum_{m=1}^{\infty} \frac{1}{(m+x)^2} \frac{1}{1+\frac{1}{m+x}} = \sum_{m=1}^{\infty} \frac{1}{(m+x)(m+x+1)},$$

which is a telescoping sum because $\frac{1}{(m+x)(m+x+1)} = \frac{1}{m+x} - \frac{1}{m+x+1}$. Thus we get

$$\mathbf{H}\left[t \mapsto \frac{1}{1+t}\right](x) = \frac{1}{1+x}.$$

This density (with a normalization factor) is known as the Gauss density, and is key to the study of continued fractions as will be made clear in the rest of the chapter. In a letter to Laplace in 1812, Gauss noted [AS17] that the map T_g preserves the derived probability measure μ_g .

Definition 1.8 (Gauss density – Gauss measure). The Gauss density is given by

$$\psi(x) := \frac{1}{\log 2} \frac{1}{1+x}.$$

This defines a probability measure over the Borel σ -algebra $\mathcal{B}_{\mathcal{I}}$ on \mathcal{I} , known as the Gauss measure

$$\mu_g(A) = \frac{1}{\log 2} \int_A \frac{dx}{1+x},$$

for $A \in \mathcal{B}_{\mathcal{I}}$.

1.2.6 The case of the CL map

Let us now describe the case of the dynamical system associated to the CL map. As we know, this system has inverse branches

$$h_a(x) = \frac{2^{-a}}{1+x}, \quad |h'_a(x)| = \frac{2^{-a}}{(1+x)^2},$$

so its associated Perron Frobenius operator is given by

$$\mathbf{H}[g](x) = \frac{1}{(1+x)^2} \sum_{a=0}^{\infty} 2^{-a} g\left(\frac{2^{-a}}{1+x}\right).$$

Here the right invariant density ψ_c , due to [Cha05], is slightly more difficult to guess. The *CL density* is given by

$$\psi_c(x) := \frac{1}{\log(4/3)} \frac{1}{(x+1)(x+2)}. \quad (1.26)$$

1.2.7 Entropy and dynamical systems

The entropy of a dynamical system is a fundamental quantity that describes the way the system evolves.

Entropy in Information Theory. The reader may be familiar with the concept of entropy in *Information Theory* [CT06]. For a discrete random variable X taking values with probabilities p_1, p_2, \dots the entropy $H(X)$ is defined by

$$H(X) = - \sum_i p_i \log p_i, \quad (1.27)$$

where it is understood that $0 \log 0$ is formally 0.

Given a stochastic process $(X_i)_{i=1}^{\infty}$ (i.e., a series of random variables), its entropy, if it exists, is defined by

$$H(\mathcal{X}) = \lim_{n \rightarrow \infty} \frac{1}{n} H(X_1, \dots, X_n), \quad (1.28)$$

where $H(X_1, \dots, X_n)$ is the entropy of the random vector (X_1, \dots, X_n) .

Entropy in Dynamical Systems. Consider now a complete interval dynamical system with shift map T , alphabet \mathcal{A} , inverse branches $h_{a_1, \dots, a_k} \in \mathcal{H}^k$ and fundamental intervals $\mathcal{I}_{a_1, \dots, a_k} = h_{a_1, \dots, a_k}(\mathcal{I})$.

The digits $a_1(X), a_2(X), \dots$, when we consider a distribution over $X \in \mathcal{I}$ with measure μ (i.e. $\mathbb{P}(X \in A) = \mu(A)$), constitute random variables which give rise to a stochastic process. The entropy of this stochastic process H_μ is said to be the entropy of the dynamical system with respect to the measure μ and satisfies

$$H_\mu(T) = - \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{(a_1, \dots, a_k) \in \mathcal{A}^k} \mu(\mathcal{I}_{a_1, \dots, a_k}) \log \mu(\mathcal{I}_{a_1, \dots, a_k}). \quad (1.29)$$

Observation 1.4. Notice that in (1.29), the entropy can be rewritten as

$$H_\mu(T) = - \lim_{k \rightarrow \infty} \mathbb{E} \left[\frac{1}{k} \log \mu(\mathcal{J}_k(X)) \right], \quad (1.30)$$

where $\mathcal{J}_k(x)$ is the fundamental interval of depth k containing x , and X is distributed according to μ .

The entropy is, then, strongly related to what we call the “real probabilistic framework”, and in Section 1.4 we shall point out how the transfer operator relates to the entropy.

Notation 1.1. When we speak about “the entropy” without reference to a measure, the Lebesgue measure is assumed. From Equation 1.30 it is actually seen that this does not change the entropy when μ and λ_{Leb} are equivalent measures in the sense that there exist constants $A, B > 0$ with $A\mu \leq \lambda_{\text{Leb}} \leq B\mu$.

In Example 1.5 we derive the entropy of the Euclidean system, which equals $\frac{\pi^2}{6 \log 2}$, from a classical result of Lévy (Proposition 1.10) and the Dominated Convergence Theorem. We could also derive it with the techniques from Section 1.4 and we point this out in Section 1.4.6.

1.3 Almost everywhere properties

1.3.1 Introduction

In Section 1.1 we described some classical properties of continued fractions, however, the great interest of continued fractions lies also in their analytic properties! We now get into some classical, yet very important statistics regarding the continued fraction expansion of a real number from \mathcal{I} .

We study the properties of the Gauss map $T_g: x \mapsto \{1/x\}$ in more detail, in particular its interplay with the Lebesgue measure λ_{Leb} , demonstrating a sort of stationary behavior of the dynamical system, namely its ergodicity.

First, it is important to remark that the σ -algebra generated by the *fundamental intervals* $\mathcal{I}_{m_1, \dots, m_k} := h_{m_1} \circ \dots \circ h_{m_k}(\mathcal{I})$, with $k \geq 0$ and $m_1, \dots, m_k \geq 1$, is the Borel σ -algebra $\mathcal{B}_{\mathcal{I}}$.

It is clear that the σ -algebra generated by the fundamental intervals is a subset of $\mathcal{B}_{\mathcal{I}}$, but the fact that they generate the whole of $\mathcal{B}_{\mathcal{I}}$ will come in handy later on: it will be enough to prove most things over fundamental intervals, on which T_g behaves nicely.

To see why the fundamental intervals $\mathcal{I}_{m_1, \dots, m_k}$ generate the Borel σ -algebra, notice that the lengths $|\mathcal{I}_{m_1, \dots, m_k}|$ tend to 0 uniformly as $k \rightarrow \infty$ (we have stated a more general result in Proposition 1.6), hence any interval $(a, b) \subset \mathcal{I}$ may actually be produced as a countable union of fundamental intervals by considering the continued fraction expansions of the borders a and b of the interval.

The other important concept, Ergodicity, implies a “stationary behavior” (Theorem 1.1). It can be thought of as a sort of “escape property”, points from anywhere in the dynamical system will always escape, as we apply our map T , from any set smaller than the whole space (in terms of measure).

Definition 1.9 (Ergodic transformation – ergodic measure). Let $(\Omega, \mathcal{B}, \mu)$ be a probability space and let $T: \Omega \rightarrow \Omega$ be a measurable transformation. We say that T is ergodic (or μ is ergodic with respect to T) if and only if the implication

$$T^{-1}(B) = B \implies \mu(B) = 0 \text{ or } \mu(B) = 1, \quad (1.31)$$

holds for every measurable set B .

We note that [Bil65] actually adds the condition that T be a measure-preserving transformation to the definition of ergodicity, thus not adding the condition to the Ergodic Theorem below (Theorem 1.1). On the other hand, [PY98] or [EW11] define an ergodic transformation the way we do, and add explicitly the condition that the transformation be measure-preserving to the Ergodic Theorem.

1.3.2 Birkhoff’s Ergodic Theorem

The fundamental consequence of the “ergodicity” is Birkhoff’s Ergodic Theorem (Theorem 1.1 below), which roughly tells us that the “time averages” approach the “space average”: the proportion of time spent by the orbit $x, S(x), S^2(x), \dots$ on some part B of the space Ω is given $\mu(B)$.

Theorem 1.1 (Birkhoff’s Ergodic Theorem). *Let $(\Omega, \mathcal{B}, \mu)$ be a probability space and let $T: \Omega \rightarrow \Omega$ be an ergodic measure-preserving transformation. Then the time and space averages coincide, more precisely*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) = \mathbb{E}_\mu[f],$$

for almost every $x \in \Omega$ and $f \in L^1(\Omega, \mathcal{B}, \mu)$.

Birkhoff’s Ergodic Theorem may also be thought of as a generalization of the law of large numbers. Indeed, the law of large numbers corresponds to the case in which $x, T(x), T^2(x), \dots$ are independent and identically distributed ^a.

We remark that this is not the most general form of Theorem 1.1. The hypothesis that T be an ergodic transformation may be dropped (it remains measure-preserving), with the cost that the resulting limit becomes a random variable given by a conditional expectation. See [EW11] for more details.

1.3.3 Ergodicity of the Gauss map and CL map

We now prove that the system $(\mathcal{I}, T_g, \mathcal{B}_{\mathcal{I}}, \mu_g)$ associated with the Gauss Map $T_g(x) = \frac{1}{x} \bmod 1$ is ergodic.

Theorem 1.2. *The measure preserving dynamical system $(\mathcal{I}, T_g, \mathcal{B}_{\mathcal{I}}, \mu_g)$ is ergodic.*

Proof. The plan is as follows: we will prove that if $T_g^{-1}A = A$ with A measurable, then

$$\mu_g(A)\mu_g(B) \leq C\mu_g(A \cap B) \quad (1.32)$$

holds for a certain constant $C > 0$ and all $B \in \mathcal{B}_{\mathcal{I}}$. Of course, taking $B = A^c$, this implies that either $\mu_g(A) = 0$ or $\mu_g(A) = 1$, the desired conclusion.

^aMore precisely this happens on a product space with T being $T(X_1, X_2, \dots) = (X_2, X_3, \dots)$.

Since the fundamental intervals $(\mathcal{I}_{m_1, \dots, m_k})$ generate the σ -algebra $\mathcal{B}_{\mathcal{I}}$, it is enough to prove (1.32) for B being an arbitrary fundamental interval $B = \mathcal{I}_{m_1, \dots, m_k}$. Let A be an invariant set, that is $T_g^{-1}A = A$. Then we want to show that $\mu_g(A)$ is either 0 or 1.

Since $T_g^{-1}A = A$, we observe that (always bearing we chose $B = \mathcal{I}_{m_1, \dots, m_k}$)

$$\mu_g(A \cap \mathcal{I}_{m_1, \dots, m_k}) = \mu_g\left(\left(T_g^{-k}A\right) \cap \mathcal{I}_{m_1, \dots, m_k}\right),$$

thus, if it is true that for a certain constant $C > 0$ and all $A \in \mathcal{B}_{\mathcal{I}}$ we have

$$\mu_g(A)\mu_g(\mathcal{I}_{m_1, \dots, m_k}) \leq C\mu_g\left(\left(T_g^{-k}A\right) \cap \mathcal{I}_{m_1, \dots, m_k}\right) \quad (1.33)$$

for all $A \in \mathcal{B}_{\mathcal{I}}$ and $m_1, \dots, m_k \geq 1$, the result will follow.

Clearly, it is enough to prove this result changing the Gauss measure μ_g for the Lebesgue measure λ_{Leb} because they are equivalent measures as we have (bounding the integrand)

$$\frac{1}{2 \log(2)} \lambda_{\text{Leb}}(D) \leq \mu_g(D) \leq \frac{1}{\log(2)} \lambda_{\text{Leb}}(D), \forall D \in \mathcal{B}_{\mathcal{I}}. \quad (1.34)$$

Therefore, it suffices to prove that for a certain constant $\tilde{C} > 0$

$$\lambda_{\text{Leb}}(A) \lambda_{\text{Leb}}(\mathcal{I}_{m_1, \dots, m_k}) \leq \tilde{C} \lambda_{\text{Leb}}\left(\left(T_g^{-k}A\right) \cap \mathcal{I}_{m_1, \dots, m_k}\right)$$

holds for all $k \geq 0$, $m_1, \dots, m_k \geq 1$ and $A \in \mathcal{B}_{\mathcal{I}}$.

Since we want to prove the latter inequality for all $A \in \mathcal{B}_{\mathcal{I}}$, it is enough to prove it for an arbitrary interval $A = [x, y]$ with $x < y$, as these generate the σ -algebra of Borel sets $\mathcal{B}_{\mathcal{I}}$.

Now observe (key!) that $(T_g^{-k}A) \cap \mathcal{I}_{m_1, \dots, m_k} = h_{m_1, \dots, m_k}([x, y])$ so that

$$\lambda_{\text{Leb}}\left(\left(T_g^{-k}A\right) \cap \mathcal{I}_{m_1, \dots, m_k}\right) = \lambda_{\text{Leb}}(h_{m_1, \dots, m_k}([x, y])) = |h_{m_1, \dots, m_k}(y) - h_{m_1, \dots, m_k}(x)|$$

which reduces, by the determinant property (1.13), to

$$|h_{m_1, \dots, m_k}(y) - h_{m_1, \dots, m_k}(x)| = \left| \frac{p_k + yp_{k-1}}{q_k + yq_{k-1}} - \frac{p_k + xp_{k-1}}{q_k + xq_{k-1}} \right| = \frac{|y - x|}{(q_k + yq_{k-1})(q_k + xq_{k-1})}.$$

Now $\lambda_{\text{Leb}}(\mathcal{I}_{m_1, \dots, m_k}) = \lambda_{\text{Leb}}(h_{m_1, \dots, m_k}([0, 1])) = \frac{1}{q_k(q_k + q_{k-1})}$ and of course we have $q_{k-1} \leq q_k$, hence

$$\frac{1}{q_k(q_k + q_{k-1})} \leq 2 \times \frac{1}{(q_k + q_{k-1})^2} \leq 2 \times \frac{1}{(q_k + yq_{k-1})(q_k + xq_{k-1})},$$

thus proving that

$$\lambda_{\text{Leb}}(A) \lambda_{\text{Leb}}(\mathcal{I}_{m_1, \dots, m_k}) \leq 2 \lambda_{\text{Leb}}\left(\left(T_g^{-k}A\right) \cap \mathcal{I}_{m_1, \dots, m_k}\right),$$

and the result follows. ■

Observation 1.5. We note that the previous proof comes down to the following key points:

- (i) the fundamental intervals generate the Borel σ -algebra $\mathcal{B}_{\mathcal{I}}$.
- (ii) the given invariant measure μ is equivalent to the Lebesgue measure.
- (iii) there is a constant C such that $|h'(y)|/|h'(x)| \leq C$ for all $x, y \in \mathcal{I}$ and inverse branches $h \in \mathcal{H}^*$.

For an expanding map T , we have immediately (i) by the same argument given for the Gauss map, while for (iii) we note that the inverse branches are monotone and therefore h' does not change sign, hence (iii) may be restated in terms of having $|(\log |h'(x)|)'| = |h''(x)/h'(x)| \leq C_1$ for some C_1 .

All of these properties are immediate for the CL dynamical system, while (ii) is also immediate as the invariant density $\psi_c(x)$ satisfies

$$\frac{1}{6 \log(4/3)} \leq \psi_c(x) = \frac{1}{\log(4/3)} \frac{1}{(x+1)(x+2)} \leq \frac{1}{2 \log(4/3)}.$$

1.3.4 Consequences of ergodicity: frequency of digits

We now explain the classical estimates for continued fractions that derive from this Theorem.

Corollary 1.5. *For almost every real number $x = [m_1, m_2, \dots] \in \mathcal{I}$, the frequency of the digit $j \geq 1$ equals*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \#\{i : 1 \leq i \leq N, m_i(x) = j\} = \frac{1}{\log 2} \log \left(1 + \frac{1}{j(j+2)} \right). \quad (1.35)$$

Proof. Note that $m(x) = j$ if and only if $x \in \mathcal{I}_j = \left(\frac{1}{j+1}, \frac{1}{j} \right]$. Thus, by the Birkhoff Ergodic Theorem applied to the characteristic function $\mathbf{1}_{\mathcal{I}_j}$ we get

$$\lim_{N \rightarrow \infty} \frac{1}{N} \#\{i : 1 \leq i \leq N, m_i(x) = j\} = \mu_g(\mathcal{I}_j) = \int_{1/(j+1)}^{1/j} \psi(x) dx,$$

which yields the result. ■

Observation 1.6. The above distribution given by (1.35) is known as the Gauss-Kuzmin distribution. Similar kinds of stationary distributions can be derived for any number of consecutive partial quotients

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \#\{i : 1 \leq i \leq N, m_{i+1}(x) = j_1, \dots, m_{i+k}(x) = j_k\} &= \mu_g(\mathcal{I}_{j_1, \dots, j_k}) \\ &= \frac{1}{\log 2} \frac{1}{1 + \frac{p_k(\mathbf{j})}{q_k(\mathbf{j})}} |\mathcal{I}_{j_1, \dots, j_k}| + O\left(|\mathcal{I}_{j_1, \dots, j_k}|^2\right). \end{aligned}$$

This can be restated as follows

$$\lim_{N \rightarrow \infty} \frac{1}{N} \#\{i : 1 \leq i \leq N, T^i x \in [y, y + \delta]\} = \frac{1}{\log 2} \frac{1}{1 + y} \delta + O(\delta^2), \quad \delta \rightarrow 0.$$

Corollary 1.6. *For almost every real number $x = [m_1, m_2, \dots] \in \mathcal{I}$, the arithmetic and geometric means of the quotients satisfy*

$$\lim_{N \rightarrow \infty} (m_1 \cdots m_N)^{1/N} = \prod_{j=1}^{\infty} \left(1 + \frac{1}{j(j+2)} \right)^{\log j / \log 2}. \quad (\text{Kuzmin constant})$$

and

$$\lim_{N \rightarrow \infty} \frac{m_1 + \cdots + m_N}{N} = \infty. \quad (1.36)$$

Proposition 1.10 (Lévy's constant). *For almost every real number $x = [m_1, m_2, \dots] \in \mathcal{I}$, the continuants $(q_k(x))_{k=1}^{\infty}$ satisfy*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log q_N(x) = \frac{\pi^2}{12 \log 2}, \quad (1.37)$$

and the constant appearing on the right-hand side is known as the Lévy constant.

Proof. In this case applying the Ergodic Theorem is not immediate. To take the expression $\lim_{N \rightarrow \infty} \frac{1}{N} \log q_N(x)$ to the form of the Ergodic Theorem, we note that $p_k(x) = q_{k-1}(T_g x)$ so that

$$\frac{1}{q_N(x)} = \frac{p_N(x)}{q_N(x)} \frac{p_{N-1}(T_g x)}{q_{N-1}(T_g x)} \cdots \frac{p_1(T_g^{N-1} x)}{q_1(T_g^{N-1} x)},$$

because of the telescoping taking place, indeed $p_{N-j}(T_g^j x) = q_{N-j-1}(T_g^{j+1} x)$. Thus we get

$$\log \frac{1}{q_N(x)} = \sum_{k=0}^{N-1} \log \frac{p_{N-k}(T_g^k x)}{q_{N-k}(T_g^k x)},$$

which is almost of the form needed to apply the Ergodic Theorem, were it not for the fact that the function $x \mapsto \frac{p_{N-k}(x)}{q_{N-k}(x)}$ depends on k and N . To get around this, we note that $x \mapsto \frac{p_{N-k}(x)}{q_{N-k}(x)}$ is in general a good approximation of the identity function. Thus we must find a bound for

$$\Delta_N(x) := \left| \sum_{k=0}^{N-1} \left(\log \frac{p_{N-k}(T_g^k x)}{q_{N-k}(T_g^k x)} - \log T_g^k x \right) \right|,$$

which makes $\frac{1}{N} \Delta_N(x) \rightarrow 0$, in order to conclude

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \frac{1}{q_N(x)} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} \log T_g^k x = \frac{1}{\log 2} \int_0^1 \frac{\log x}{1+x} dx.$$

Showing that the integral gives the desired constant $\int_0^1 \frac{\log x}{1+x} dx = -\frac{\pi^2}{12}$ follows simply from term-wise integration of the sum $\frac{1}{1+x} = 1 - x + x^2 - \cdots + (-1)^m x^m + (-1)^{m+1} \frac{x^{m+1}}{1+x}$ and $\sum (-1)^k / k^2 = -\pi^2 / 12$.

Fix k , let us look at how big the difference

$$\left| \log \frac{p_{N-k}(T_g^k x)}{q_{N-k}(T_g^k x)} - \log T_g^k x \right| = \left| \log \left((T_g^k x) \times \frac{q_{N-k}(T_g^k x)}{p_{N-k}(T_g^k x)} \right) \right|,$$

actually is. Of course, to study $\log u$ with u close to 1, we look at the corresponding $u - 1$, for us

$$(T_g^k x) \times \frac{q_{N-k}(T_g^k x)}{p_{N-k}(T_g^k x)} - 1 = \left(T_g^k x - \frac{p_{N-k}(T_g^k x)}{q_{N-k}(T_g^k x)} \right) \times \frac{q_{N-k}(T_g^k x)}{p_{N-k}(T_g^k x)},$$

and we recall $\left| T_g^k x - \frac{p_{N-k}(T_g^k x)}{q_{N-k}(T_g^k x)} \right| \leq (q_{N-k}(T_g^k x))^{-2}$ hence for $N - k \geq 2$

$$\left| (T_g^k x) \times \frac{q_{N-k}(T_g^k x)}{p_{N-k}(T_g^k x)} - 1 \right| \leq \frac{1}{p_{N-k}(T_g^k x) q_{N-k}(T_g^k x)} \leq \frac{1}{2}.$$

Since $|\log u| \leq 2|u - 1|$ for $u \in [\frac{1}{2}, \frac{3}{2}]$, which applies to our difference whenever $k \leq N - 2$

$$\begin{aligned}
\Delta_N(x) &\leq \sum_{k=0}^{N-1} \left| \log \frac{p_{N-k}(T_g^k x)}{q_{N-k}(T_g^k x)} - \log T_g^k x \right| = \sum_{k=0}^{N-1} \left| \log \left((T_g^k x) \times \frac{q_{N-k}(T_g^k x)}{p_{N-k}(T_g^k x)} \right) \right| \\
&\leq 2 \sum_{k=0}^{N-1} \left| (T_g^k x) \times \frac{q_{N-k}(T_g^k x)}{p_{N-k}(T_g^k x)} - 1 \right| \\
&\leq 2 \sum_{k=0}^{N-1} \frac{1}{p_{N-k}(T_g^k x) q_{N-k}(T_g^k x)} \leq 2 \sum_{k=0}^{N-1} 2^{-(N-k-1)/2} \\
&\leq 2 \sum_{k=0}^{\infty} 2^{-k/2} = \frac{2}{1 - \sqrt{2}/2}.
\end{aligned}$$

Thus $\frac{1}{N} \Delta_N(x)$ tends to 0, yielding the result. \blacksquare

Example 1.5 (Entropy of the Euclidean system). We now compute the entropy of the Euclidean dynamical system with respect to the Lebesgue measure, recall the definition of entropy from Section 1.2.7. By definition

$$H(T_g) = - \lim_{k \rightarrow \infty} \mathbb{E} \left[\frac{1}{k} \log \lambda_{\text{Leb}}(\mathcal{J}_k(X)) \right],$$

where $J_k(x)$ is the fundamental interval of depth k containing x and X is drawn according to the uniform distribution. We recall that for the Euclidean system, the interval of depth k containing X is the one delimited by $\frac{p_k(X)}{q_k(X)}$ and $\frac{p_k(X)+p_{k-1}(X)}{q_k(X)+q_{k-1}(X)}$ as follows from Section 1.1.5. Thus we derive

$$H(T_g) = - \lim_{k \rightarrow \infty} \mathbb{E} \left[\frac{1}{k} \log \frac{1}{q_k(X)(q_k(X) + q_{k-1}(X))} \right].$$

Since $0 < q_{k-1}(X) \leq q_k(X)$ we deduce that

$$H(T_g) = 2 \lim_{k \rightarrow \infty} \mathbb{E} \left[\frac{1}{k} \log q_k(X) \right].$$

From Proposition 1.10 we would expect the entropy H to equal $\frac{\pi^2}{6 \log 2}$. This is not immediate, as the random variable $\frac{1}{k} \log q_k(X)$ is not bounded in general (build X from a sequence m_1, m_2, \dots that grows very fast). However, $\frac{1}{k} \log q_k(X)$ is bounded almost everywhere by a fixed constant $B > 0$ (see [Khi97, Theorem 31, pp.65-68]), whence we derive the entropy by the Dominated Convergence Theorem (see [Fol99]).

In all

$$H(T_g) = \frac{\pi^2}{6 \log 2}. \quad (1.38)$$

Example 1.6 (Continued logarithm). We continue with the example of the Continued Logarithm from Example 1.3. We noted already in Observation 1.5 that the map T_c is Ergodic with respect to $(\mathcal{I}, \mathcal{B}_{\mathcal{I}}, \mu_c)$, where μ_c is the invariant measure given by $d\mu_c(x) = \psi_c(x) dx$, a fact that was already noted by Chan in [Cha05].

It follows that we may apply Birkhoff's Ergodic Theorem, from which we derive the following result.

Proposition 1.11. *For almost every $x \in \mathcal{I}$, the digits $(a_i(x))$ of the CL expansion of x from Example 1.3 satisfy*

$$\lim_{N \rightarrow \infty} \frac{a_1(x) + \dots + a_N(x)}{N} = \frac{\log(3/2)}{\log(4/3)}. \quad (1.39)$$

We notice immediately that the digits of the CL expansion are much smaller than those of classical continued fractions (see Corollary 1.6). This is to be expected from the fact that we are coding the *exponents*. In

fact, it is similarly true that the averages of $\log m_i(x)$ converge almost everywhere (see the first part of Corollary 1.6).

Proof. Since for almost every x we have $a_1(x) = \lfloor \log_2(1/x) \rfloor$ and $a_k(x) = a_1(T_c^{k-1}x)$, it follows at once from the Ergodic Theorem that

$$\lim_{N \rightarrow \infty} \frac{a_1(x) + \dots + a_N(x)}{N} = \frac{1}{\log(4/3)} \int_0^1 \lfloor \log_2(1/z) \rfloor \frac{dz}{(z+1)(z+2)}.$$

We now compute this integral

$$\int_0^1 \lfloor \log_2(1/z) \rfloor \frac{dz}{(z+1)(z+2)} = \sum_{k=1}^{\infty} k \int_{2^{-k-1}}^{2^{-k}} \frac{dz}{(z+1)(z+2)},$$

and we remark that $\frac{1}{(z+1)(z+2)} = \frac{1}{z+1} - \frac{1}{z+2}$, thus

$$\int_{2^{-k-1}}^{2^{-k}} \frac{dz}{(z+1)(z+2)} = \int_{2^{-k-1}}^{2^{-k}} \frac{dz}{z+1} - \int_{2^{-k-1}}^{2^{-k}} \frac{dz}{z+2} = \int_{2^{-k-1}}^{2^{-k}} \frac{dz}{z+1} - \int_{2^{-k-2}}^{2^{-k-1}} \frac{dz}{z+1}$$

by performing a change of variables in the second integral. Thus if $I_k := \int_{2^{-k-1}}^{2^{-k}} \frac{dz}{z+1}$ we get

$$\int_0^1 \lfloor \log_2(1/z) \rfloor \frac{dz}{(z+1)(z+2)} = - \sum_{k=1}^{\infty} k \Delta I_k = I_1 + \sum_{k=1}^{\infty} I_{k+1} = \int_0^{1/2} \frac{dz}{z+1},$$

were we applied summation by parts, and this equals $\log(1/2 + 1) = \log(3/2)$. ■

Example 1.7. We show that, for an irrational $\alpha \notin \mathbb{Q}$, the circle rotation $\mathcal{R}_\alpha: x \mapsto (x + \alpha) \bmod 1$ is ergodic with respect to the Lebesgue measure. For the proof we follow [EW11].

Two important remarks

- Observe that if B is an invariant measurable set, meaning that $\mathcal{R}_\alpha^{-1}B = B$, then $\mathbf{1}_B = \mathbf{1}_B \circ \mathcal{R}_\alpha$ almost everywhere.
- Now the key observation. Since α is irrational, the points $0, \alpha, 2\alpha, 3\alpha, \dots$ taken in \mathbb{T}^1 (i.e., modulo 1) are dense in the circle \mathbb{T}^1 . To see why, note that $n \cdot \alpha + m \cdot 1 = \delta$ can be arbitrary small (this follows from Observation 1.3) and then consider for $s \in (0, 1)$ consider $\lfloor \frac{s}{\delta} \rfloor n \cdot \alpha$ modulo 1, which satisfies $|\lfloor \frac{s}{\delta} \rfloor n \cdot \alpha + \lfloor \frac{s}{\delta} \rfloor m - s| \leq |\delta|$. Thus the distance $d_{\mathbb{T}^1}(s, \lfloor \frac{s}{\delta} \rfloor n \cdot \alpha \bmod 1)$ on \mathbb{T}^1 is at most δ .

Suppose then that B were an invariant measurable set. Let $f \in C(\mathbb{T}^1)$ be such that $\|f - \mathbf{1}_B\|_1 < \epsilon$.

Then we have $\|f \circ \mathcal{R}_\alpha^k - f\|_1 < 2\epsilon$, as follows from the first remark and the fact that for any integrable g we have $\int g = \int g \circ \mathcal{R}_\alpha$, because the Lebesgue measure is translation invariant. Since $k\alpha$ can be made arbitrarily close to any $s \in \mathbb{T}^1$ (by density) and f is continuous, we deduce that the inequality $\|f \circ \mathcal{R}_s - f\|_1 < 2\epsilon$ holds too.

Thus

$$\left\| f - \int f(s) ds \right\|_1 = \int \left| \int (f(x) - f(x+s)) ds \right| dx \leq \int \|f - f \circ \mathcal{R}_s\|_1 ds \leq 2\epsilon,$$

by applying the triangle inequality and Tonelli's Theorem [Fol99].

Thus we deduce that

$$\|\mathbf{1}_B - \lambda_{\text{Leb}}(B)\|_1 \leq \|\mathbf{1}_B - f\|_1 + \left\| f - \int f(s) ds \right\|_1 + \left\| \int f(s) ds - \lambda_{\text{Leb}}(B) \right\|_1 < 4\epsilon.$$

Since $\epsilon > 0$ was arbitrary, we conclude that actually $\|\mathbf{1}_B - \lambda_{\text{Leb}}(B)\|_1 = 0$ and so $\mathbf{1}_B(x) = \lambda_{\text{Leb}}(B)$ for almost every $x \in \mathbb{T}^1$, implying that either $\lambda_{\text{Leb}}(B)$ is 0 or 1. ◇

As a consequence of Definition 3.4 and Theorem 1.1, for every irrational angle $\alpha \in \mathcal{I} \setminus \mathbb{Q}$ we have that for almost every x the “time average” equals the space average

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} f \circ \mathcal{R}_\alpha^k(x) = \int_0^1 f(s) ds,$$

for any $f \in L^1(\mathcal{I}, \mathcal{B}_\mathcal{I}, \lambda_{\text{Leb}})$.

In particular this tells us that for almost every initial point $x \in \mathbb{S}^1$ and interval $I \subset \mathbb{S}^1$

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} \mathbf{1}_I(\mathcal{R}_\alpha^k(x)) = |I|,$$

the frequency of time spent by the rotation on an interval I equals its length (as we have a unit length circle). However there is still this somewhat annoying “almost every x ” condition regarding the initial point. The convergence actually holds for *every* initial point $x \in \mathbb{S}^1$ and $f \in C(\mathbb{S}^1)$, as the system is what is called “uniquely ergodic”. We will not get into much detail regarding unique ergodicity, in any case it is fair to say that a dynamical system is uniquely ergodic if and only if it has a unique invariant measure.

The fact that it holds for continued functions will imply (by bounding from above and below) the result for intervals that we will later need.

Proposition 1.12. *Let α be an irrational. Given an interval $I \subset \mathbb{T}^1$ and x is any point $x \in \mathbb{T}^1$, the orbit of the irrational rotation \mathcal{R}_α of angle α , starting from x , spends a fraction of time proportional to the length $|I|$ on the interval I , namely*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} \mathbf{1}_I(\mathcal{R}_\alpha^k x) = |I|. \quad (1.40)$$

The above proposition could have been derived as a consequence of the so-called “uniform distribution modulo 1”, u.d.mod1 for short. Informally, a sequence x_1, x_2, \dots is uniformly distributed modulo 1 if and only if the fractional parts $\{x_k\}$ spend $|J|$ proportion of the time on the interval $J \subset \mathcal{I}$. There is a classical criterion (the Weil criterion) which gives a necessary and sufficient condition for u.d.mod1, which is quite direct to verify in the case of the irrational rotations. A book of reference on the subject is [KN74].

1.3.5 Large digits in the expansion: the Borel-Bernstein Theorem

There is finally one other important property regarding the quotients m_k of the continued fraction expansions that will be of great interest to us. We know so far that $m_k = s$ occurs with frequency $\sim \frac{1}{\log 2} \frac{1}{s^2}$ almost surely (see Corollary 1.5), and we know that the arithmetic mean $(m_1 + \dots + m_N)/N$ tends to infinity almost surely (see Corollary 1.6). Can we give asymptotic bounds for the “large quotients” m_k in terms of the index k ? The answer to this question is given by the so-called Borel-Bernstein Theorem [KMS16] which we prove in two parts.

The principle of the proof is not reduced only to continued fractions, and we will underline the corresponding results for the continued logarithm. These, as far as the author is aware, do not appear in the literature.

As we intend to prove an almost everywhere result, which concerns an infinite number of quotients m_N , it comes as no surprise that the Borel-Cantelli Lemma will be involved. The first Borel-Cantelli Lemma gives a criterion to decide when a series of events on a measurable space are “transient”, meaning that almost every point of the space will be in only finitely many events.

Thus it is important to highlight the following definition which comes in handy when stating the Borel-Cantelli Lemma

Definition 1.10. Let (X, \mathcal{M}, μ) be a measure space and suppose A_1, A_2, \dots are measurable sets. We define the set *i.o.* A_n , which reads infinitely often A_n , by

$$\text{i.o. } A_n := \limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k. \quad (1.41)$$

In other words, $x \in X \setminus (\text{i.o. } A_n)$ iff from some $n_0(x)$ we have $x \notin A_k$ for all $k \geq n_0(x)$.

Theorem 1.3 (Borel-Cantelli Lemma). *Let (X, \mathcal{M}, μ) be a measure space and suppose A_1, A_2, \dots are measurable sets such that $\sum \mu(A_k) < \infty$.*

Then the set $\text{i.o. } A_n = \limsup A_n$ has measure 0. In other words, for almost every $x \in X$ there is $n_0(x)$ such that $x \notin A_k$ for all $k \geq n_0(x)$.

Proof. Let $B_n := \bigcup_{k \geq n} A_k$ and observe that

$$\mu(B_n) = \mu\left(\bigcup_{k \geq n} A_k\right) \leq \sum_{k \geq n} \mu(A_k) < \infty.$$

This is the tail of a convergent series, therefore $\sum_{k \geq n} \mu(A_k) \rightarrow 0$ as $n \rightarrow \infty$, hence $\mu(B_n) \rightarrow 0$.

Clearly B_n is a decreasing sequence of sets, and the above equation proves that $\mu(B_1) < \infty$, thus

$$\mu(\text{i.o. } A_n) = \mu\left(\bigcap_{n \geq 1} B_n\right) = \lim_{n \rightarrow \infty} \mu(B_n),$$

and this number equals 0. ■

Now we come to the digits (or quotients) $m_n(x)$ of the continued fraction expansion of x . The corresponding application of the previous Borel-Cantelli Lemma gives

Theorem 1.4 (Borel-Bernstein). *Let $f: \mathbb{N} \rightarrow \mathbb{R}_{>0}$ be such that $\sum_n 1/f(n) < \infty$. Then for almost every $x \in \mathcal{I}$ the inequality*

$$m_n(x) < f(n) \quad (1.42)$$

holds for all large enough n .

Proof. Let us consider $A_k := \{x \in \mathcal{I} : m_k(x) \geq f(k)\}$. Then we note that

$$\lambda_{\text{Leb}}(A_k) \leq \sum_{m_1, \dots, m_{k-1} \geq 1} \left| h_{m_1, \dots, m_{k-1}} \left(\left(0, \frac{1}{f(k)}\right] \right) \right|,$$

simply because $m_k(x) \geq f(k)$ implies $1/T^{k-1}x \geq f(k)$, i.e., $x \in h_{m_1, \dots, m_{k-1}} \left(\left(0, \frac{1}{f(k)}\right] \right)$ for some coefficients $m_1, \dots, m_{k-1} \geq 1$.

Here by the monotonicity of h and the mean-value theorem we have

$$\left| h_{m_1, \dots, m_{k-1}} \left(\left(0, \frac{1}{f(k)}\right] \right) \right| = \left| h_{m_1, \dots, m_{k-1}}(0) - h_{m_1, \dots, m_{k-1}}\left(\frac{1}{f(k)}\right) \right| \leq \frac{|h'_{m_1, \dots, m_{k-1}}(0)|}{f(k)},$$

and since $|h'_{m_1, \dots, m_{k-1}}(0)| = 1/q_{k-1}^2 \leq |\mathcal{I}_{m_1, \dots, m_{k-1}}|$ we obtain, from $\sum_{m_1, \dots, m_k \geq 1} |\mathcal{I}_{m_1, \dots, m_{k-1}}| = 1$, that

$$\lambda_{\text{Leb}}(A_k) \leq \frac{1}{f(k)}.$$

Thus the result follows from the Borel-Cantelli Lemma Theorem 1.3. ■

We see that Theorem 1.4 tells at once that almost surely

$$m_n(x) < n(\log n)^{1+\epsilon},$$

for all large enough n , when $\epsilon > 0$. We will see conversely that $n \log n$ is, in a way, a kind of threshold for $m_n(x)$ as $m_n(x) > n \log n$ will hold almost surely for infinitely many indices n .

Observation 1.7. The same proof reads almost ad verbum for the continued logarithm and many other dynamical systems. Indeed, for a complete interval dynamical system having digits in an infinite alphabet $\mathcal{A} \subset \mathbb{N}$ we require just the following key steps and hypothesis:

- Define

$$A_k = \bigcup_{a_1, \dots, a_{k-1} \in \mathcal{A}} h_{a_1, \dots, a_{k-1}} \left(\bigcup_{j \in \mathcal{A}: j \geq f(k)} I_j \right), \quad I_j = h_j(\mathcal{I}).$$

- There is a constant $C > 0$ such that $|h'(y)/h'(x)| \leq C$ for all $x, y \in \mathcal{I}$ and $h \in \mathcal{H}^*$.
- The function $f(k)$ is strictly increasing for large enough k , hence invertible for large enough k . It is key to note then that $\#\{k : a \geq f(k)\} = f^{-1}(a) + O(1)$.

And we get the following result for the digit sequence $a_1(x), a_2(x), \dots$, for almost every $x \in \mathcal{I}$.

Proposition 1.13. Consider a complete interval dynamical system of class \mathcal{C}^1 with associated shift map $T: \mathcal{I} \rightarrow \mathcal{I}$ and inverse branches h_a labeled by $\mathcal{A} \subset \mathbb{N}$. Suppose further that, for some fixed constant $C > 0$ independent from the choice of the inverse branch, every inverse branch $h \in \mathcal{H}^*$ of the system satisfies $|h'(y)/h'(x)| \leq C$ for all $x, y \in \mathcal{I}$. Let f be a strictly increasing function such that the condition

$$\sum_{a \in \mathcal{A}} f^{-1}(a) \times \lambda_{\text{Leb}}(h_a(\mathcal{I})) < \infty \quad (1.43)$$

is satisfied. Then for almost every $x \in \mathcal{I}$, the digit sequence $a_1(x), a_2(x), \dots$ satisfies

$$a_n(x) < f(n),$$

for all large enough n .

Example 1.8 (Continued fractions). The difference in the statements of Theorem 1.4 and Proposition 1.13 is simply due to the fact that during the proof of Theorem 1.4 we exploited the particular ordering of the fundamental intervals, which gives the equality $\{x : m(x) \geq f(k)\} = (0, \frac{1}{f(k)}]$, while for the other, we cannot suppose such an equality and we are forced to reverse a double sum and count.

Nevertheless, Proposition 1.13 does yield that the digits of classical continued fractions satisfy $m_n(x) < n(\log n)^{1+\epsilon}$ for large enough n .

Indeed, we show how to reverse $f(n) = n(\log n)^{1+\epsilon}$ asymptotically. We have

$$\begin{aligned} \log n &= \log f(n) - (1 + \epsilon) \log \log n \\ &= \log f(n) - (1 + \epsilon) \log(\log f(n) - (1 + \epsilon) \log \log n) \\ &= \log f(n) - (1 + \epsilon) \log \log f(n) + \log \left(1 - (1 + \epsilon) \frac{\log \log n}{\log f(n)} \right), \end{aligned}$$

and here $\log n < \log f(n)$. Thus

$$f^{-1}(a) = \frac{a}{(\log a)^{1+\epsilon}} \times \left(1 - (1 + \epsilon) \frac{\log \log f^{-1}(a)}{\log a} \right),$$

and $\log f^{-1}(a) < \log a$ implies $\frac{\log \log f^{-1}(a)}{\log a} = O\left(\frac{\log \log a}{\log a}\right)$ as $a \rightarrow \infty$. Given this asymptotic formula and $\lambda_{\text{Leb}}(h_a(\mathcal{I})) = \frac{1}{a(a+1)}$, Proposition 1.13 yields that for a.e., x , $m_a(x) < a(\log a)^{1+\epsilon}$ for large enough a .

Example 1.9 (Continued logarithms). Here $\lambda_{\text{Leb}}(h_a(\mathcal{I})) = 1/2^{a+1}$ for $a \in \mathcal{A} = \{0, 1, 2, \dots\}$, thus picking $\epsilon > 0$ and $f^{-1}(a) = (2 - \epsilon)^a$ we get that for almost every $x \in \mathcal{I}$

$$a_n(x) < \frac{\log n}{\log(2 - \epsilon)} \text{ holds for all large enough } n.$$

A more precise analysis yields that for almost every $x \in \mathcal{I}$

$$a_n(x) < \log_2 n + \log_2 \log_2 n + (1 + \epsilon) \log_2 \log_2 \log_2 n, \text{ holds for all large enough } n.$$

Theorem 1.5 (Borel-Bernstein). *Let $f: \mathbb{N} \rightarrow \mathbb{R}_{>0}$ be such that $\sum_n 1/f(n) = \infty$. Then the inequality $m_n(x) \geq f(n)$ holds for infinitely many indices n .*

If the variables $m_n(x)$ for $n \geq 1$ were independent, this result would follow from the classical ‘‘converse’’ of the Borel-Cantelli Lemma (on a probability space), which tells us that if $\sum \mathbb{P}(A_n) = \infty$ and (A_i) are mutually independent, then $\mathbb{P}(i.o.A_n) = 1$. This is not our case however, but the same proof still holds if an inequality such as

$$\mathbb{P}(A_{n+1} = s | A_1 = s_1, \dots, A_n = s_n) \geq C \mathbb{P}(A_{n+1} = s),$$

holds for some $C > 0$.

Thus, in order to prove this result, we rely on a lemma which tells us that the probability of finding $m_{k+1}(x) = s$, given the knowledge of m_1, \dots, m_k , is always of order $\Theta(1/s^2)$.

Lemma 1.1. *Consider integers $m_1, \dots, m_k \geq 1$ and $s \geq 1$. Then, the conditional probability, with respect to the Lebesgue measure on \mathcal{I} , of having $m_{k+1}(x) = s$ given that $m_1(x) = m_1, \dots, m_k(x) = m_k$, which is given by $|\mathcal{I}_{m_1, \dots, m_k, s}| / |\mathcal{I}_{m_1, \dots, m_k}|$, satisfies the inequalities*

$$\frac{1}{3} \frac{1}{s^2} \leq \frac{|\mathcal{I}_{m_1, \dots, m_k, s}|}{|\mathcal{I}_{m_1, \dots, m_k}|} \leq 2 \frac{1}{s^2}. \quad (1.44)$$

Proof. We remark that by from Proposition 1.3

$$|\mathcal{I}_{m_1, \dots, m_k, s}| = \left| h_{m_1, \dots, m_k} \left(\frac{1}{s+1} \right) - h_{m_1, \dots, m_k} \left(\frac{1}{s} \right) \right| = \left| \frac{p_k(s+1) + p_{k-1}}{q_k(s+1) + q_{k-1}} - \frac{p_k s + p_{k-1}}{q_k s + q_{k-1}} \right|$$

which by (1.13) equals

$$|\mathcal{I}_{m_1, \dots, m_k, s}| = \frac{1}{(sq_k + q_{k-1})((s+1)q_k + q_{k-1})}.$$

Therefore

$$\frac{|\mathcal{I}_{m_1, \dots, m_k, s}|}{|\mathcal{I}_{m_1, \dots, m_k}|} = \frac{1}{s^2} \frac{1 + \frac{q_{k-1}}{q_k}}{\left(1 + \frac{q_{k-1}}{sq_k}\right) \left(1 + \frac{1}{s} + \frac{q_{k-1}}{sq_k}\right)},$$

hence we find

$$\frac{1}{3} \frac{1}{s^2} \leq \frac{|\mathcal{I}_{m_1, \dots, m_k, s}|}{|\mathcal{I}_{m_1, \dots, m_k}|} \leq 2 \frac{1}{s^2}. \quad \blacksquare$$

Proof. [Theorem 1.5] If we had $m_n(x) < f(n)$ for all n from N to $N + M$, an event which we call $E_{N,M}$, we note that conditioning

$$\lambda_{\text{Leb}}(E_{N,M+1}) = \sum_{s < f(n)} \lambda_{\text{Leb}}(m_{N+M+1}(x) = s | E_{N,M}) \lambda_{\text{Leb}}(E_{N,M}).$$

Here we are going to apply our bounds on the conditional probability, first we write the sum in a more convenient form

$$\sum_{s < f(n)} \lambda_{\text{Leb}}(m_{N+M+1} = s | E_{N,M}) = 1 - \sum_{s \geq f(n)} \lambda_{\text{Leb}}(m_{N+M+1} = s | E_{N,M}),$$

and therefore, from the bound we get

$$\sum_{s < f(n)} \lambda_{\text{Leb}}(m_{N+M+1} = s | E_{N,M}) \leq 1 - \frac{1}{3} \sum_{s \geq f(n)} \frac{1}{s^2} \leq 1 - \frac{1}{3} \frac{1}{f(n)},$$

where we have applied $\sum_{s \geq t} 1/s^2 \geq \int_t^\infty ds/s^2 = 1/t$. Thus we derive

$$\lambda_{\text{Leb}}(E_{N,M+1}) \leq \left(1 - \frac{1}{3} \frac{1}{f(n)}\right) \lambda_{\text{Leb}}(E_{N,M}) \leq \dots \leq \prod_{k=0}^M \left(1 - \frac{1}{3f(n)}\right),$$

which converges to 0 as $M \rightarrow \infty$ by the divergence of $\sum 1/f(n)$. Indeed, from the classical inequality $e^x \geq 1 + x$ for all $x \in \mathbb{R}$ we derive $1 - \frac{1}{3f(n)} \leq \exp(-\frac{1}{3f(n)})$.

Thus we have proved, as $\lambda_{\text{Leb}}(\mathcal{I}) = 1 < \infty$ and $E_{N,M+1} \subset E_{N,M}$ decreases in M , that we have $\lambda_{\text{Leb}}(E_N) = 0$ where $E_N := \bigcap_{M \geq N} E_{N,M}$. Now we are done, because the probability that there exists N such that $m_n(x) < f(n)$ for all $n \geq N$ is $\lambda_{\text{Leb}}(\bigcup_{N=1}^\infty E_N) = 0$. ■

Thus, the previous Theorems tell us that $m_N(x)$ is roughly $N \log N$ in the “worst-case”.

Corollary 1.7. *Let $\epsilon > 0$. For almost every $x \in \mathcal{I}$ the partial quotients $m_N(x)$ of the continued fraction expansion of x satisfy the limits*

$$\lim_{N \rightarrow \infty} \frac{m_N(x)}{N(\log N)^{1+\epsilon}} = 0,$$

and

$$\limsup_{N \rightarrow \infty} \frac{m_N(x)}{N \log N} = \infty,$$

meaning that there is an increasing sequence N_k of positive integers (depending on x) such that

$$\lim_{k \rightarrow \infty} \frac{m_{N_k}(x)}{N_k \log N_k} = \infty.$$

The Borel-Bernstein Theorem (divided into Theorem 1.4 and Theorem 1.5) as well as Corollary 1.7 apply in the context of the recurrence function of Sturmian words, namely Theorem 3.4, first proved by Morse and Hedlund in [MH40], which gives information regarding the “worst-case” for the recurrence function of Sturmian words.

Observation 1.8. For the continued logarithms we do have a similar result, which we point out here. The proof employs Corollary 7.1 from Chapter 7, which explains how to compute the lengths of the fundamental intervals. The digits $a_1(x), a_2(x), \dots$ of the continued logarithms satisfy, for almost every $x \in \mathcal{I}$, that

$$a_n(x) \geq \log_2 n + \log_2 \log n$$

infinitely often. As mentioned, the proof remains basically the same because $\frac{|\mathcal{I}_{m_1, \dots, m_k, s}|}{|\mathcal{I}_{m_1, \dots, m_k}|}$ is again of order $\Theta(|\mathcal{I}_s|)$, for the fundamental intervals of this system.

Comparing with the case of continued fractions, we conclude that the digits $a_1(x), a_2(x), \dots$ of the continued logarithm expansion of a real $x \in \mathcal{I}$ are almost surely $a_n(x) \sim \log_2 n + \log_2 \log n$ in the “worst-case”. More formally, for almost every x

$$\limsup_{n \rightarrow \infty} \frac{a_n(x) - \log_2 n}{\log_2 \log n} = 1. \quad (1.45)$$

1.4 Real probabilistic framework

In Section 1.3 we studied the continued fraction expansion of almost every real number $x \in \mathcal{I}$, namely up to a set of measure (or probability) zero. It is worth noting that the results from Section 1.3, even though the limits are explicit, all involve convergence rates that may vary widely according to the real number $x \in \mathcal{I}$ that is considered (fixed). In this section we introduce what we call the “Real probabilistic framework”: here we do not study a fixed $x \in \mathcal{I}$ anymore, but the behavior of larger sets on *average* or *distribution*.

Given a random $X \in I$ with continued fraction expansion $m_1(X), m_2(X), \dots$, we stop the expansion at a given position k and study the average properties of the resulting rational $[m_1(X), m_2(X), \dots, m_k(X)]$, which will be our random variable of interest.

Similarly, we may ask analogous questions for the continued logarithm expansion and other interval dynamical systems. The main tool to perform such studies is in fact the Perron Frobenius operator, as well as a generalization of it: the so-called *transfer operator*.

We recall Definition 1.7: the Perron-Frobenius operator \mathbf{H} for an interval dynamical system of class C^1 is defined formally by the expression

$$\mathbf{H}[g](x) := \sum_{h \in \mathcal{H}} |h'(x)| g(h(x)),$$

for an input function g that must belong to an “appropriate” functional space.

The Perron Frobenius operator describes how the densities evolve after one iteration of the corresponding map of the system. Thus, if g is the density of the random variable X , then $\mathbf{H}[g]$ is the density of $T(X)$. In particular, we remark that g is an invariant density if and only if it satisfies $\mathbf{H}[g] \equiv g$.

When we iterate the Perron Frobenius operator k times we get

$$\mathbf{H}^k[g](x) := \sum_{h \in \mathcal{H}^k} |h'(x)| g(h(x)),$$

involving the depth k inverse branches \mathcal{H}^k . Thus the powers of the Perron Frobenius operator follow the evolution of the densities of the system.

Often, the Perron-Frobenius operator \mathbf{H} has a behavior which resembles that of matrices; the iterate \mathbf{H}^k is determined by the so-called spectrum of \mathbf{H} , with the dominant eigenvalues being the most determinant.

Our target properties for \mathbf{H} are the following:

(P1) There is a unique and simple dominant eigenvalue $\lambda = 1$ for \mathbf{H} .

(P2) There is a so-called spectral gap: the dominant eigenvalue is separated from the rest of the spectrum.

Then the iterates of \mathbf{H} , which describe the evolution of the process, are asymptotically determined by the corresponding projection onto the eigenspace of $\lambda = 1$ and the corresponding eigenvector ϕ , which gives the invariant density after normalization.

Indeed, then [BV03] one has an appropriate decomposition $\mathbf{H} = \mathbf{P} + \mathbf{N}$, where \mathbf{P} is the projector over the dominant eigenspace (given by $\mathbf{P}[f](t) = \phi(t) \int_{\mathcal{I}} f(x) dx$), and \mathbf{N} is an operator with the same spectrum except for $\lambda = 1$. Iterating gives $\mathbf{H}^k = \mathbf{P} + \mathbf{N}^k$ and the spectral radius of \mathbf{N}^k tends to 0 as $k \rightarrow \infty$.

In subsection 1.4.1 we give several useful concepts and results concerning the spectrum of an operator. Then subsection 1.4.3 introduces the concept of *quasi-compactness*, a property strongly related to (P2), in particular we give sufficient conditions for the quasi-compactness. In subsection 1.4.4 we complete the set of properties (P1) and (P2) by explaining how the spectral gap is obtained from the uniqueness of the dominant eigenvalue and the quasi-compactness.

In 1.4.7 we describe the transfer operator, which generalizes the Perron Frobenius operator.

In this section we follow mainly [Bal00, pp. 28–36] as well as [Kat95], and the notes in [Sar12] and [BV03].

1.4.1 Concepts from functional analysis

Let $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ be a Banach space, i.e., a vector space \mathcal{B} along with a norm $\|\cdot\|_{\mathcal{B}}$ which makes it complete. We recall that an operator $A: (\mathcal{B}, \|\cdot\|_{\mathcal{B}}) \rightarrow (\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ is continuous if and only if it is bounded, i.e., the set $\{A(v) : \|v\|_{\mathcal{B}} \leq 1\}$ is bounded, or equivalently, there is constant $C > 0$ such that $\|A(v)\|_{\mathcal{B}} \leq C\|v\|_{\mathcal{B}}$. The minimal such constant C is the *norm* of the operator $\|A\|$, and the operator norm makes the space of bounded operators into a normed space which further satisfies $\|AB\| \leq \|A\|\|B\|$.

For a bounded operator $A: (\mathcal{B}, \|\cdot\|_{\mathcal{B}}) \rightarrow (\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ we say, equivalently, that A *acts* on the Banach space.

Consider a bounded linear operator $A: (\mathcal{B}, \|\cdot\|_{\mathcal{B}}) \rightarrow (\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ acting on our Banach space.

Definition 1.11 (Spectrum and eigenvalues). The *spectrum* $\text{SP}(A)$ of an operator A acting on a Banach space \mathcal{F} is the set of all complex numbers λ such that $(A - \lambda\mathbf{I})$ does not have a bounded (norm) inverse, where \mathbf{I} is the identity operator.

The *spectral radius* of A is defined by

$$R(A) = \sup\{|z| : z \in \text{SP}(A)\}. \quad (1.46)$$

We say $\lambda \in \text{SP}(A)$ is an eigenvalue of A if and only if $(A - \lambda\mathbf{I})$ is not injective. Equivalently, this means that there is a non zero $v \in \mathcal{B}$ such that $A(v) = \lambda v$. We then say that $v \neq 0$ is an eigenvector associated with the eigenvalue λ . The multiplicity of an eigenvalue λ is the dimension of the eigenspace $\{v \in \mathcal{B} : A(v) = \lambda v\}$.

Observation 1.9 (Perron Frobenius operator). In this example we do a formal computation with regard to the Perron Frobenius operator. We note that if λ were an eigenvalue associated with the eigenvector g in L^1 , then

$$\mathbf{H}[g](x) = \lambda g(x).$$

Integrating and noting that $\int_{\mathcal{I}} \mathbf{H}[g](x)dx = \int_{\mathcal{I}} g(x)dx$, we derive that either $\int_{\mathcal{I}} g(x)dx = 0$ or $\lambda = 1$. \diamond

Proposition 1.14. Let $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ be a Banach space and let A be a bounded operator acting on \mathcal{B} . If $\|A\| < 1$ then $\mathbf{I} - A$ is invertible and

$$(\mathbf{I} - A)^{-1} = \mathbf{I} + A + A^2 + \dots, \quad (1.47)$$

where the series actually converges to a bounded operator with norm at most $1/(1 - \|A\|)$.

Proof. Simply note that $F := \mathbf{I} + A + A^2 + \dots$ defines a linear operator, due the completeness of \mathcal{B} , which is bounded because $\|A^k\| \leq \|A\|^k$ implies $\|F\| \leq 1/(1 - \|A\|)$.

Due to the continuity of A we have, point-wise

$$(I - A)F = F - AF = F - (A + A^2 + A^3 + \dots) = \mathbf{I},$$

and similarly from the other side. \blacksquare

Corollary 1.8. Let $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ be a Banach space and let A be a bounded operator acting on \mathcal{B} .

Then the following inequality holds for the spectral radius

$$R(A) \leq \|A\|. \quad (1.48)$$

Proof. Let λ be a complex number with $|\lambda| > \|A\|$. Then we note that

$$\|A - \lambda\mathbf{I}\| = \lambda \left\| \mathbf{I} - \frac{A}{\lambda} \right\|,$$

and here $\|\frac{A}{\lambda}\| < 1$, thus it follows from Proposition 1.14 that $A - \lambda\mathbf{I}$ is invertible.

Since this holds for arbitrary $\lambda \in \mathbb{C}$ with $|\lambda| > \|A\|$, we conclude that $R(A) \leq \|A\|$. \blacksquare

Corollary 1.9. *Let A be a bounded linear operator acting on a Banach space \mathcal{B} . The spectrum $\text{SP}(A)$ of A is compact.*

Proof. From Proposition 1.14 we see that $\mathbb{C} \setminus \text{SP}(A)$ is open, as the proposition actually tells us that the set of invertible bounded operators is open (with the operator norm). From Corollary 1.8 it is bounded. ■

There is a general formula connecting the spectral radius and the norm, see [Yos95, pp.209–212].

Theorem 1.6 (Spectral radius formula). *Let $(\mathcal{B}, \|\cdot\|)$ be a Banach space and let $A: \mathcal{B} \rightarrow \mathcal{B}$ be a linear bounded operator acting on it. Then we have the following spectral radius formula*

$$R(A) = \lim \|A^n\|^{1/n}. \quad (1.49)$$

Notice, then, that the spectral radius tells us straight away the order of growth of $\|A^n\|$, this is related to the radius of convergence of power series we shall see in Proposition 2.3. In the case of matrices this is usually known as Gelfand's formula and, in that case, it is independent from the choice of norm.

Notation 1.2. Throughout this text we will make use of several classical norms which constitute the basis of more complex ones. To start with, the $\|\cdot\|_1$ (or $\|\cdot\|_{L^1}$) norm from $L^1(X, \mu)$, given by the integral of the absolute value

$$\|f\|_1 := \int_X |f(x)| d\mu(x).$$

We recall that this norm makes L^1 into a Banach space.

Second, the *supremum norm* (or uniform convergence norm) $\|\cdot\|_\infty$ for the continuous functions on $C^0(X, \mathbb{C})$, given by

$$\|f\|_\infty := \sup_{x \in X} |f(x)|.$$

Again, this makes $C^0(X, \mathbb{C})$ into a Banach space. ◇

1.4.2 The spectral radius of the Perron Frobenius operator

Coming back to the Perron Frobenius operator defined in Definition 1.7, it is clear that it acts on the space $L^1(\mathcal{I})$ of integrable function on \mathcal{I} . The space L^1 is complete under the norm $\|g\|_1 = \int_{\mathcal{I}} g(x) dx$. Observe

$$\int_{\mathcal{I}} |\mathbf{H}[g](x)| dx \leq \int_{\mathcal{I}} |g(x)| dx, .$$

This follows from an application of the triangle inequality, which yields $|\mathbf{H}[g](x)| \leq \mathbf{H}[|g|](x)$, as well as the fact that \mathbf{H} preserves the integrals $\int \mathbf{H}[|g|] = \int |g|$. It follows that $\|\mathbf{H}\| \leq 1$ and $R(\mathbf{H}) \leq 1$ over L^1 .

For the Euclidean dynamical system, we already know that $x \mapsto 1/(1+x)$ is an eigenvector associated with the eigenvalue $\lambda = 1$, hence $R(\mathbf{H}) = 1$ in this case. We explain a generalization from [BDV02].

Throughout the text we will encounter several spaces other than $L^1(\mathcal{I})$. Concretely, we will consider $\text{BV}(\mathcal{I})$, the subspace of functions of bounded variation (see [Fol99]), and the even smaller space $C^1(\mathcal{I})$ of continuously differentiable functions.

Definition 1.12 (Bounded variation). The *total variation* of a function $f: [a, b] \rightarrow \mathbb{C}$ from $L^1([a, b])$ is defined by

$$V_a^b(f) = \sup_{P \in \mathcal{P}} \sum_{i=0}^{n_P-1} |f(x_{i+1}) - f(x_i)|, \quad (1.50)$$

where the supremum is taken over, \mathcal{P} , the of all finite partitions $P = (x_0, x_1, \dots, x_{n_P})$ of $[a, b]$ with

$$a = x_0 < x_1 < \dots < x_{n_P-1} < x_{n_P} = b.$$

The function f is said to be of *bounded variation*, BV for short, on $[a, b]$ if and only if $V_a^b(f) < \infty$. The set of all functions of bounded variation on $[a, b]$ is denoted by $BV([a, b])$.

The total variation satisfies various useful (and simple) properties

Proposition 1.15. *Let f, g belong to $BV([a, b])$ then*

1. *If $f \in C^1([a, b])$ we have $V_a^b(f) = \int_a^b |f'(x)| dx$.*
2. *Triangle inequality: $V_a^b(f + g) \leq V_a^b(f) + V_a^b(g)$.*
3. *Additivity: $V_a^b(f) = V_a^c(f) + V_c^b(f)$ if $c \in (a, b)$.*
4. *Product rule: $V_a^b(fg) \leq \|f\|_\infty V_a^b(g) + \|g\|_\infty V_a^b(f)$. Also $V_a^b(fg) \leq \int_a^b |f(x)g'(x)| dx + \|g\|_\infty V_a^b(f)$ when $g \in C^1([a, b])$.*
5. *Boundedness: $\|f\|_\infty \leq V_a^b(f) + \frac{1}{b-a} \|f\|_1$.*
6. *Composition: $V_c^d(f \circ j) = V_a^b(f)$ if $j: [c, d] \rightarrow [a, b]$ is a monotone bijection.*

Observation 1.10. Consider a complete interval dynamical system of class \mathcal{C}^2 with $|h'| \leq \Delta_1$ and $|h''| \leq C|h'|$ for all $h \in \mathcal{H}$. Then associated Perron Frobenius operator \mathbf{H} acts on $BV(\mathcal{I})$ under the norm

$$\|f\|_{BV} := \|f\|_1 + V_0^1(f). \quad (1.51)$$

The proof of this fact is simple enough. Indeed, if $f \in BV(\mathcal{I})$, we have $V_0^1(\mathbf{H}[f]) \leq \sum_{h \in \mathcal{H}} V_0^1(|h'|f \circ h)$ by Proposition 1.15 and for each particular $h \in \mathcal{H}$ we note

$$V_0^1(|h'|f \circ h) \leq \|h'\|_\infty V_{\mathcal{I}_h}(f) + \int_0^1 |h''(x)f(h(x))| dx \leq \|h'\|_\infty V_{\mathcal{I}_h}(f) + C \int_0^1 |h'(x)||f(h(x))| dx,$$

where $\mathcal{I}_h = h(\mathcal{I})$ is the fundamental interval associated with h . Thus it follows, summing up over all h , that

$$V_0^1(\mathbf{H}[f]) \leq \Delta_1 V_0^1(f) + C \|f\|_1.$$

Finally, since $\|\mathbf{H}[f]\|_1 \leq \|f\|_1$ we get

$$\|\mathbf{H}[f]\|_{BV} \leq \Delta_1 V_0^1(f) + (1 + C) \|f\|_1, \quad (1.52)$$

in particular $\mathbf{H}[f]$ is of bounded variation. \diamond

Here is the main theorem of this subsection, which is a simplified version of [BDV02, Proposition 4].

Theorem 1.7. *Consider a complete interval dynamical system of class \mathcal{C}^2 (see Definition 1.3) with inverse branches $h \in \mathcal{H}$. Assume further that $\Delta_n := \sup\{\|h'\|_\infty : h \in \mathcal{H}^n\}$ satisfies*

1. *(Weak expansion) The quantity Δ_1 satisfies $\Delta_1 \leq 1$.*
2. *(Strong expansion) There exists an integer m and a real constant $\gamma < 1$ such that $\Delta_m \leq \gamma$.*
3. *(Bounded distortion) There exists a constant $C > 0$ such that $|h''(x)| \leq C|h'(x)|$ for all $h \in \mathcal{H}$ and $x \in \mathcal{I}$.*

Then the system admits an invariant density of bounded variation, in particular the spectral radius $R(\mathbf{H})$ of the Perron Frobenius operator is indeed $R(\mathbf{H}) = 1$.

Proof. From (1.52) applied to the iteration of the system m times, so with \mathbf{H}^m rather than \mathbf{H} , we get

$$\|\mathbf{H}^m[f]\|_{BV} \leq \gamma \|f\|_{BV} + C_m \|f\|_1,$$

for some finite C_m . Note that we have a bounded distortion for $h \in \mathcal{H}^m$ too by induction (use $\Delta_1 < \infty$).

Now iterating and recalling $\|\mathbf{H}\|_1 \leq 1$ we get

$$\|\mathbf{H}^{m \times k}[f]\|_{\text{BV}} \leq \gamma^k \|f\|_{\text{BV}} + \frac{C_m}{1 - \gamma} \|f\|_1. \quad (1.53)$$

This means that the set

$$\mathcal{F} := \left\{ f_n = \frac{1}{n} \sum_{j=0}^{n-1} \mathbf{H}^{m \times j}[\mathbf{1}] : n \geq 1 \right\}, \quad \mathbf{1}(x) := x,$$

is a bounded set for the $\|\cdot\|_{\text{BV}}$ norm. Thus Helly's selection Theorem implies that a certain subsequence of f_n converges to a function f_∞ of $\text{BV}(\mathcal{I})$. Since the elements f_n are densities, so is f_∞ .

Note that $\mathbf{H}^m[f_\infty] = f_\infty$, and from this we derive an invariant density (note it is positive)

$$g := \frac{1}{m} \sum_{j=0}^{m-1} \mathbf{H}^j[f_\infty]$$

for our original system. ■

1.4.3 Quasi-compactness

To be able to analyze the long term behavior of \mathbf{H}^k as k gets larger we need to characterize the dominant eigenvalues as well as the corresponding eigenvectors. If these are separated by a gap from the rest of the spectrum, we will get asymptotics for \mathbf{H}^k .

Definition 1.13 (Essential spectral radius). The essential spectral radius $R_e(A)$ of a bounded linear operator A acting on the Banach space $(\mathcal{B}, \|\cdot\|)$ is the infimum of the numbers $R \geq 0$ such that for all $\lambda \in \text{SP}(A)$ with $|\lambda| > R$ we have that λ is an isolated *eigenvalue* of finite multiplicity.

When $R_e(A)$ is strictly smaller than $R(A)$, we can work with the ‘‘dominant’’ eigenvalues (those with $|\lambda| > R_e(A)$) like in the case of matrices. Such an operator is called a *quasi-compact* operator.

Definition 1.14 (Quasicompact operator). A bounded operator $A: (\mathcal{B}, \|\cdot\|) \rightarrow (\mathcal{B}, \|\cdot\|)$ acting on a Banach space $(\mathcal{B}, \|\cdot\|)$ is said to be quasi-compact if and only if $R_e(A)$ is strictly smaller than $R(A)$.

The most classical example are the so-called compact operators, from where the name originates.

Definition 1.15 (Compact operator). A linear operator $T: X \rightarrow Y$ between two normed spaces is said to be compact if the image (Tx_i) of any bounded sequence $(x_i)_i \subset X$ contains a convergent subsequence.

Note that we have defined compact operators in somewhat more generality than quasi-compact operators. This will come in handy when we state Hennion's theorem, Theorem 1.8 below.

For compact operators we simply have $R_e(A) = 0$, so 0 can be the only accumulation point of eigenvalues. These conditions, for our case, are too restrictive.

It is not possible to get this *spectral gap* when the underlying functional space is too big (e.g., L^1), thus a huge part of the problem lies in choosing the right functional space \mathcal{F} on which \mathbf{H} will act (i.e., $\mathbf{H}[\mathcal{F}] \subset \mathcal{F}$) but for which we will have an spectral gap (and so a ‘‘smaller spectrum’’).

The following theorem by Hennion [Hen93, Corollaire 1] gives sufficient conditions for an operator to be quasi-compact. His result is a reinforcement of a theorem of Ionescu Tulcea and Marinescu [ITM50], by using a formula by Nussbaum [Nus70] for the essential spectral radius in terms of ball-coverings. He goes on to show several examples of applications of his result. The reader is referred to the appendices of [Sar12] for a self-contained proof of Hennion's Theorem.

Theorem 1.8 (Hennion [Hen93]). *Let A be a bounded operator acting on a Banach space $(\mathcal{B}, \|\cdot\|)$. Denote by $R(A)$ its spectral radius. Assume that there is a norm $|\cdot|$ on \mathcal{B} such that*

1. A is a compact operator from $(\mathcal{B}, \|\cdot\|)$ to $(\mathcal{B}, |\cdot|)$.
2. There are sequences (r_n) and (ρ_n) of positive numbers such that $r = \liminf(r_n)^{1/n} < R(A)$ and for all $n \geq 1$ and v

$$\|A^n v\| \leq r_n \|v\| + \rho_n |v|. \quad (1.54)$$

Then the essential radius of A on $(\mathcal{B}, \|\cdot\|)$ satisfies $R_e(A) \leq r < R(A)$, in particular A is quasi-compact.

Note that in the previous theorem we do not ask for the space $(\mathcal{B}, |\cdot|)$ to be complete, just a normed space. The norm $|\cdot|$ is, in a sense, a weaker auxiliary norm.

There is an alternative formulation of Theorem 1.8 which appears in [DMDV05, Theorem C]. The hypothesis are a bit stronger but easily applicable for our context.

Theorem 1.9 (See [DMDV05], theorem C). *Let $(\mathcal{B}, \|\cdot\|)$ be a Banach space. Let A be a bounded operator on $(\mathcal{B}, \|\cdot\|)$. Denote by $R(A)$ its spectral radius. Assume that there exists another norm $|\cdot|$ on \mathcal{B} , also making it a Banach space, which satisfies the following properties*

1. The closed unit ball of $(\mathcal{B}, \|\cdot\|)$ is compact in $(\mathcal{B}, |\cdot|)$.
2. There are sequences (r_n) and (ρ_n) of positive numbers such that $r = \liminf(r_n)^{1/n} < R(A)$ and, for all $n \geq 1$, the following inequality

$$\|A^n v\| \leq r_n \|v\| + \rho_n |v| \quad (\text{Lasota-Yorke})$$

known as a Lasota-Yorke inequality, is fulfilled.

Then the essential radius of A on $(\mathcal{B}, \|\cdot\|)$ satisfies $R_e(A) \leq r < R(A)$, in particular the operator A is quasi-compact.

Proof. We explain how this version derives from Theorem 1.8. In order to do this we must show condition (1) from Theorem 1.8, in other words, show that if (x_n) is a bounded sequence in $(\mathcal{B}, \|\cdot\|)$, then there is a subsequence $(x_{n_j})_j$ that makes $(Tx_{n_j})_j$ converge under the $|\cdot|$ norm. We may assume without loss of generality that $\|x_n\| \leq 1$ for all n .

Since the open ball from $(\mathcal{B}, \|\cdot\|)$ is compact under $|\cdot|$, we know there is a subsequence (x_{m_k}) of (x_n) such that $|x_{m_k}|$ converges, in particular it is bounded. From (2) we derive the inequality $\|Av\| \leq r_1 \|v\| + \rho_1 |v|$ which tells us that $\|Ax_{m_k}\|$ is bounded, hence by our hypothesis (1) it has a convergence subsequence under $|\cdot|$, thus showing the result. \blacksquare

Observation 1.11. We remark that property (2) in Theorem 1.9 can somewhat be relaxed. Indeed, assume that the Lasota-Yorke inequality from property (2) is satisfied for some $s \geq 1$ with $r_s^{1/s} < R(A)$. Then by induction (note $\|A^k\| \leq \|A\|^k$ for all k) we get that

$$\|A^{ns} v\| \leq r_s^n \|v\| + k_s |v|,$$

for some constants $k_s \geq 0$ and all $v \in \mathcal{B}$. Thus we derive a similar inequality for a number of iterations that is not a multiple of s . Indeed, if $0 \leq t < s$, from $\|A^{ns+t} v\| \leq \|A\|^t \|A^{ns} v\|$ we get

$$\|A^{ns+t} v\| \leq \max\{1, \|A\|, \dots, \|A\|^{s-1}\} \times (r_s^n \|v\| + k_s |v|),$$

from which we may apply Theorem 1.9 and get that $R_e(A) \leq r_s^{1/s} < R(A)$. \diamond

Theorem 1.10. *Consider an complete interval dynamical system of class \mathcal{C}^2 (see Definition 1.3) with inverse branches $h \in \mathcal{H}$. Assume further that $\Delta_n := \sup\{\|h'\|_\infty : h \in \mathcal{H}^n\}$ satisfies*

1. (Weak expansion) The quantity Δ_1 satisfies $\Delta_1 \leq 1$.
2. (Strong expansion) There exists an integer m and a real constant $\gamma < 1$ such that $\Delta_m \leq \gamma$.

3. (Bounded distortion) There exists a constant $C > 0$ such that $|h''(x)| \leq C|h'(x)|$ for all $h \in \mathcal{H}$ and $x \in \mathcal{I}$.

Then the associated Perron Frobenius operator \mathbf{H} acting on $BV(\mathcal{I})$ is quasi-compact with $R_e(\mathbf{H}) \leq \gamma^{1/m}$.

Proof. Use the ideas from Observation 1.11 and (1.53) along with the weak norm $\|\cdot\|_1$ and the strong norm $\|\cdot\|_{BV}$. The only thing we have to prove is that the closed unit ball of $(BV(\mathcal{I}), \|\cdot\|_{BV})$ is compact on $(L^1(\mathcal{I}), \|\cdot\|_1)$ but this follows at once from Helly's selection Theorem. ■

Example 1.10 (Continued fractions on \mathcal{C}^1). Let us go back to the Euclidean dynamical system with the Gauss map T_g . We recall that here the Perron Frobenius operator \mathbf{H} is given by

$$\mathbf{H}[g](x) := \sum_{h \in \mathcal{H}} |h'(x)|g(h(x)) = \sum_{m=1}^{\infty} \frac{1}{(m+x)^2} g\left(\frac{1}{m+x}\right),$$

we will show that, over an appropriate space, this operator is quasi-compact.

Of course, this operator acts on $L^1(\mathcal{I})$ and has $R(\mathbf{H}) = 1$, having an eigenvalue $\lambda = 1$ associated with the eigenvector $x \mapsto 1/(1+x)$, corresponding to the Gauss measure. We will show that for a well-chosen space our operator is quasi-compact by applying Hennion's Theorem in the form of Observation 1.11.

We will work with the set of continuously differentiable functions $C^1(\mathcal{I})$. First, it is clear that \mathbf{H} actually acts on this space because the sum defining it converges uniformly and so do the sums of the term-wise derivatives, for any fixed order.

Consider first the norm $|\cdot|$ defined by $|g| = \sup_{x \in \mathcal{I}} |g(x)|$, and the stronger norm $\|\cdot\|$ defined by $\|g\| = |g| + |g'|$. The norm $\|\cdot\|$ makes \mathcal{B} a Banach space. This follows from the fact that the uniform convergence (i.e., in the sup norm) of continuous functions gives a continuous function (see e.g., [Mun00] or [Fol99]). This means at once that $(\mathcal{B}, |\cdot|)$ is a Banach space, but the same holds for $(\mathcal{B}, \|\cdot\|)$ just by showing that if the derivatives converge uniformly to a continuous function this must also be the derivative of the limit function of the functions themselves (upon integration).

The operator \mathbf{H} is a bounded one on $(\mathcal{B}, \|\cdot\|)$. Indeed observe that the derivative of $\mathbf{H}[g]$ is

$$(\mathbf{H}[g])'(x) = - \sum_{m=1}^{\infty} \frac{1}{(m+x)^3} \left(\frac{1}{m+x} g' \left(\frac{1}{m+x} \right) + 2g \left(\frac{1}{m+x} \right) \right),$$

as follows from the uniform convergence of the previous sum. Then $|(\mathbf{H}[g])'| \leq 2\zeta(3)\|g\|$, while it is clear that $|\mathbf{H}[g]| \leq \zeta(2)\|g\|$, thus, as $\|\mathbf{H}[g]\| = |\mathbf{H}[g]| + |(\mathbf{H}[g])'|$, it follows that the \mathbf{H} is a bounded operator.

Now we turn to property (1) of Theorem 1.9. In fact, this property follows at once in this case from the celebrated *Arzelà-Ascoli Theorem*^b (see e.g., [Mun00]).

For property (2) of Theorem 1.9 we study what happens with each particular term $\mathbf{H}_{(m_1, m_2)}[g](x) := |h'_{m_1, m_2}(x)|g(h_{m_1, m_2}(x))$ for each $m_1, m_2 \geq 1$. Differentiating

$$\left| (\mathbf{H}_{(m_1, m_2)}[g])'(x) \right| \leq |h'_{m_1, m_2}(x)|^2 |g'(h_{m_1, m_2}(x))| + |h''_{m_1, m_2}(x)| |h'_{m_1, m_2}(x)| |g(h_{m_1, m_2}(x))|.$$

Notice that $|h''_{m_1, m_2}(x)| \leq 2|h'_{m_1, m_2}(x)|$, thus there exists a constant $C > 0$ such that $|h''_{m_1, m_2}(x)| \leq C|h'_{m_1, m_2}(x)|$. Then observe that $|h'_{m_1, m_2}(x)| \leq 1/4 =: \gamma$, thus

$$\left| (\mathbf{H}_{(m_1, m_2)}[g])'(x) \right| \leq \gamma |h'_{m_1, m_2}(x)| |g'(h_{m_1, m_2}(x))| + C |h'_{m_1, m_2}(x)| |g(h_{m_1, m_2}(x))|.$$

Similarly for any $k \geq 1$

$$\left| (\mathbf{H}_{(m_1, m_2, \dots, m_{2k-1}, m_{2k})}[g])'(x) \right| \leq \gamma^k |h'(x)| |g'(h(x))| + C |h'(x)| |g(h_{m_1, m_2}(x))|,$$

^bThe only non-trivial part is the equicontinuity, and this in turn follows from the Mean Value Theorem for derivatives.

where $h = h_{(m_1, m_2, \dots, m_{2k-1}, m_{2k})}$. Notice that $\|h'\|_\infty \leq 1$ for $h \in \mathcal{H}$ implies $\sum_{h \in \mathcal{H}^{2k}} |h'(x)| \leq \sum_{h \in \mathcal{H}} |h'(x)|$, which is uniformly bounded in our case.

From this we derive

$$\left\| \mathbf{H}^{2k}[g] \right\| \leq D_1 \gamma^k \|g\| + D_0 |g|,$$

for some constants D_0, D_1 (independent from the choice of g) and all $g \in C^1(\mathcal{I})$. It then follows from Observation 1.11 that $R_e(\mathbf{H}) \leq \gamma^{1/2} = 1/2$. \diamond

The previous proof works for all complete interval dynamical systems as long as we add the condition that the image of the constant function $\mathbf{1}(x) = 1$ through \mathbf{H} , that is $\mathbf{H}[\mathbf{1}]$, be well-defined.

Proposition 1.16. *Consider an complete interval dynamical system of class \mathcal{C}^2 (see Definition 1.3) with inverse branches $h \in \mathcal{H}$. Assume further that $\Delta_n := \sup\{\|h'\|_\infty : h \in \mathcal{H}^n\}$ satisfies*

1. (Weak expansion) *The quantity Δ_1 satisfies $\Delta_1 \leq 1$.*
2. (Strong expansion) *There exists an integer m and a real constant $\gamma < 1$ such that $\Delta_m \leq \gamma$.*
3. (Fast decay of the branches) *The series $\sum_{h \in \mathcal{H}} |h'(x)|$ is uniformly bounded by a constant D_1 .*
4. (Bounded distortion) *There is a constant $C > 0$ such that $|h''(x)| \leq C|h'(x)|$ for all $h \in \mathcal{H}$ and $x \in \mathcal{I}$.*

Then the associated Perron Frobenius operator \mathbf{H} acting on $C^1(\mathcal{I})$ is quasi-compact with $R_e(\mathbf{H}) \leq \gamma^{1/m}$. Furthermore, the iterate \mathbf{H}^m satisfies a Lasota-Yorke bound of the kind

$$|\mathbf{H}^{m \times k}[f]|_1 \leq D_1 \gamma^k |f|_1 + D_0 |f|_0, \quad (1.55)$$

for some D_0 independent from f and k .

1.4.4 Spectral gap

Given a quasi-compact operator A , we know that its spectrum, out of some finite ball around the origin, is discrete and consists of eigenvalues of finite multiplicity. When there is just one dominant eigenvalue λ_{\max} , with multiplicity 1, the quasi-compactness tells us that there is a ‘‘gap’’; there is a constant $0 < C < |\lambda_{\max}|$ such that all other elements λ from the spectrum $\text{SP}(A)$ satisfy $|\lambda| < C$.

When our operator A presents such a spectral gap, we may rewrite [Kat95, Sar12] our operator A as the sum $A = \lambda_{\max}P + N$, where P is a projection (i.e. it satisfies $P^2 = P$) with $\dim \text{Im}(P) = 1$, and a bounded linear operator N such that $PN = NP = \mathbf{0}$, with spectral radius $R(N) < C$.

Thus iterating we get $A^k = \lambda_{\max}^k P + N^k$ and we recall that $\|N^k\|^{1/k} \rightarrow R(N)$ by Theorem 1.6, hence

$$A^k(v) = \lambda_{\max}^k P(v) + O(\|v\| \times C^k).$$

In the case of a Hilbert space, i.e., when our Banach space is associated with an inner product $\langle \cdot, \cdot \rangle$, it is not difficult to define a projection P onto a space of finite dimension. Here, however, we deal with Banach spaces with, a priori, no inner-product. This is why we must explain the notion of projection in this context, making the connection with the spectrum.

Theorem 1.11 (Separation of spectrum [Sar12]). *Let A be a bounded operator acting on a Banach space $(\mathcal{B}, \|\cdot\|)$. Suppose $\text{SP}(A) = \Sigma_{in} \uplus \Sigma_{out}$ where $\Sigma_{in}, \Sigma_{out}$ are compact, and let γ be a smooth closed curve which does not intersect $\text{SP}(A)$, and which contains Σ_{in} in its interior, and Σ_{out} in its exterior. Then*

1. *The operator*

$$\mathbf{P} := \frac{1}{2\pi i} \oint_{\gamma} (z\mathbf{I} - A)^{-1} dz$$

is a projection $P^2 = P$, and so $\mathcal{B} = \ker(\mathbf{P}) \oplus \text{Im}(\mathbf{P})$,

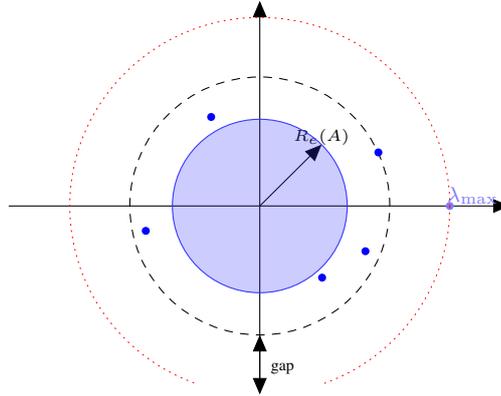


Figure 1.5: An illustration of the spectral gap.

2. The projection commutes with A , i.e., $\mathbf{P}A = A\mathbf{P}$. Therefore

$$A(\ker(\mathbf{P})) \subset \ker(\mathbf{P}), \text{ and } A(\text{Im}(\mathbf{P})) \subset \text{Im}(\mathbf{P}).$$

3. The spectrum of the projection \mathbf{P} is related to that of A by the equalities

$$\text{SP}(A|_{\text{Im}(\mathbf{P})}) = \Sigma_{in}, \text{ and } \text{SP}(A|_{\ker(\mathbf{P})}) = \Sigma_{out}.$$

Of course, what we intend to do is to show that the Perron-Frobenius operator \mathbf{H} has a unique dominant eigenvalue $\lambda_{\max} = 1$ of multiplicity one. This, together with the quasi-compactness will lead to the decomposition $A = \mathbf{P} + \mathbf{N}$ where $\mathbf{P}[f](x) = \varphi(x) \int_{\mathcal{I}} f(t) dt$ where φ is the invariant density.

1.4.5 Eigenvalues and the ergodic properties of the Perron Frobenius operator

Actually there are some deep relations between the eigenvalues of the Perron Frobenius operator \mathbf{H} of an interval dynamical system and the ergodic properties of the system.

Definition 1.16 (Mixing). Let $(\Omega, \mathcal{B}, \mu)$ be a probability space and let $T: \Omega \rightarrow \Omega$ be an ergodic measure-preserving transformation. We say that the map T is mixing if and only if for every pair of measurable sets $A, B \in \mathcal{B}$ we have

$$\lim_{n \rightarrow \infty} \mu(A \cap T^{-n}B) = \mu(A)\mu(B).$$

Remark. Mixing is also sometimes called *strong-mixing* in order to contrast it with *weak-mixing*. We have *weak-mixing* when the means converge

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(A \cap T^{-n}B) = \mu(A)\mu(B).$$

Clearly mixing implies weak-mixing. Note, in turn, that weak-mixing implies ergodicity

We now cite [Sar12]

Proposition 1.17. Consider a complete interval dynamical system of class \mathcal{C}^1 with shift T . Let $(\mathcal{L}, \|\cdot\|)$ be a Banach space with $\mathcal{L} \subset L^1(\mathcal{I})$ and $\|\cdot\|_{\mathcal{L}} \geq \|\cdot\|_{L^1}$.

Suppose the Perron Frobenius operator \mathbf{H} acts on \mathcal{L} and is quasi-compact. If there is an invariant density φ for T , which makes T mixing for the measure defined by $\mu(A) := \int_A \varphi(x) dx$, then \mathbf{H} presents a spectral gap on \mathcal{L} with dominant eigenvalue $\lambda = 1$ and projection

$$P(f) = \varphi \int_{\mathcal{I}} f(x) dx.$$

It turns out that there are converses for \mathbf{H} acting on $C^1(\mathcal{I})$ (see e.g., [Sar12] or [BV03]):

(i) if $\lambda = 1$ is a simple eigenvalue, the map T is ergodic.

(ii) $\lambda = 1$ is the only eigenvalue on $\{|\lambda| = 1\}$ (moreover simple) if and only if the system is mixing.

We underline for (ii) that there is the underlying assumption that \mathbf{H} acts on $C^1(\mathcal{I})$. When we have the weaker assumption that \mathbf{H} acts on $L^1(\mathcal{I})$, it turns out that [Sar12] having $\lambda = 1$ as a simple eigenvalue, and all other eigenvalues of smaller absolute value, implies that the system is *weak-mixing*.

For the case of the Euclidean dynamical system, proving that the system is mixing is not immediate. A useful tool to prove mixing is the so-called *exactness*.

Definition 1.17 (Exactness). Let (Ω, Σ, μ) be a probability space. A measurable map T is said to be *exact* if for every $E \in \bigcap_{n=0}^{\infty} T^{-n}(\Sigma)$, either $\mu(E) = 0$ or $\mu(\Omega \setminus E) = 0$.

Proposition 1.18. *An exact probability preserving map is mixing.*

The proof of this proposition can be found in [Sar12, Appendix A.2] and makes use of the Martingale convergence theorem. Proving that a map is exact is simpler than proving it is mixing directly.

Example 1.11 (Exactness of the Gauss map). Suppose E satisfies $E \in \bigcap_n T^{-n}(B_{\mathcal{I}})$ and $\mu_g(E) > 0$. We must show that $\mu_g(\mathcal{I} \setminus E) = 0$.

Recall that during the proof of Theorem 1.2 we proved (1.33), i.e., that for any measurable A and fundamental interval $\mathcal{I}_{m_1, \dots, m_k}$ we have

$$\mu_g((T^{-k}A) \cap \mathcal{I}_{m_1, \dots, m_k}) \geq C^{-1} \mu_g(A) \mu_g(\mathcal{I}_{m_1, \dots, m_k})$$

for a certain constant $C > 0$ independent from the choice of A and the interval. We note that if $E \in \bigcap_n T^{-n}T^{-n}(B_{\mathcal{I}})$ we have $E = T^{-n}E_n$ for a certain $E_n \in \mathcal{B}_{\mathcal{I}}$. Then it follows, picking $k = n$ and $A = E_n$ that

$$\mu_g(E \cap \mathcal{I}_{m_1, \dots, m_k}) \geq C^{-1} \mu_g(E_n) \mu_g(\mathcal{I}_{m_1, \dots, m_k})$$

But $\mu_g(E_n) = T_g(T^{-n}E_n) = \mu_g(E)$, thus

$$\mu_g(E \cap \mathcal{I}_{m_1, \dots, m_k}) \geq C^{-1} \mu_g(E) \mu_g(\mathcal{I}_{m_1, \dots, m_k}). \quad (1.56)$$

The rest of the corresponding proof in [Sar12] proceeds by the Martingale convergence theorem. We may, however, produce a direct argument, which we give here. The argument is a bit technical but not complicated, and typical of measure theory. The concept is the following: we may approximate any measurable set by a disjoint countable union of fundamental intervals.

As the Gauss measure is clearly regular [Fol99], for every $\epsilon > 0$ there would be an open set O_ϵ containing $\mathcal{I} \setminus E$ such that $\mu_g(O_\epsilon) \leq \mu_g(\mathcal{I} \setminus E) + \epsilon$. The open set O_ϵ can be produced as a countable disjoint union of fundamental intervals $\mathcal{I}_{a_1, \dots, a_k}$ of different depths. For every such interval we have (1.56), thus, summing over these intervals we obtain

$$\mu_g(E \cap O_\epsilon) \geq C^{-1} \mu_g(E) \mu_g(O_\epsilon),$$

but $\mu_g(E \cap O_\epsilon) \leq \mu_g(E \cap (\mathcal{I} \setminus E)) + \epsilon = \epsilon$ and $\mu_g(O_\epsilon) \geq \mu_g(\mathcal{I} \setminus E)$, hence

$$\epsilon \geq C^{-1} \mu_g(E) \mu_g(\mathcal{I} \setminus E).$$

As $\epsilon > 0$ was arbitrary we derive $\mu_g(E) \mu_g(\mathcal{I} \setminus E) = 0$, and therefore $\mu_g(\mathcal{I} \setminus E) = 0$. \diamond

We note that the previous proof generalizes exactly as mentioned in Observation 1.5, giving

Proposition 1.19. *Consider a complete interval dynamical system of class \mathcal{C}^1 with map T , and an invariant density $g \in C^0(\mathcal{I})$ with respect to the Lebesgue measure λ_{Leb} . Suppose*

- (i) the map T^k is expanding for some $k \in \mathbb{N}$, i.e., $\sup_{h \in \mathcal{H}^k} \|h'(x)\|_\infty < 1$ for some k .
- (ii) there is a distortion constant $C > 0$ such that $|h'(y)|/|h'(x)| \leq C$ for all $x, y \in \mathcal{I}$ and inverse branches $h \in \mathcal{H}^*$ of any depth.

Then the map T is exact with respect to the invariant measure $d\mu(x) = g(x)d\lambda_{Leb}(x)$.

This means, in particular, that the CL system is exact, and therefore mixing.

We conclude this subsection by stating the resulting theorem for the spectrum. This is a combination of Proposition 1.19, Theorem 1.7 and Proposition 1.17

Theorem 1.12. Consider a complete interval dynamical system of class \mathcal{C}^1 (see Definition 1.3) with inverse branches $h \in \mathcal{H}$. Assume further that $\Delta_n := \sup\{\|h'\|_\infty : h \in \mathcal{H}^n\}$ satisfies

1. (Weak expansion) The quantity Δ_1 satisfies $\Delta_1 \leq 1$.
2. (Strong expansion) There exists an integer m and a real constant $\gamma < 1$ such that $\Delta_m \leq \gamma$.
3. (Bounded distortion) There exists a constant $C > 0$ such that $|h''(x)| \leq C|h'(x)|$ for all $h \in \mathcal{H}$ and $x \in \mathcal{I}$.

Then the Perron Frobenius operator \mathbf{H} , acting on the space $(BV, \|\cdot\|_{BV})$ admits a spectral gap, with 1 as the simple dominant eigenvalue, associated with the unique invariant density φ of the system (there is one).

The iterates \mathbf{H}^k of the Perron Frobenius operator decomposes as

$$\mathbf{H}^k[f](x) = \left(\int_{\mathcal{I}} f(t) dt \right) \varphi(x) + O\left(\|f\|_{BV} \theta^k\right),$$

for some $\theta < 1$, which depends on the subdominant eigenvalues.

Observation 1.12. It is also possible to substitute BV by $C^1(\mathcal{I})$ when the invariant density is continuous and the series $\sum_{h \in \mathcal{H}} |h'(x)|$ is uniformly bounded.

Why choose BV or $C^1(\mathcal{I})$. When the branches of the dynamical system are not complete, the Perron-Frobenius operator involves characteristic functions which naturally lead to BV rather than $C^1(\mathcal{I})$. For our case, however, the branches will always be complete and hence we may choose either one depending on the target application. When the input functions are naturally discontinuous as in Chapter 5, we pick $BV(\mathcal{I})$. For the case of the CL system in Chapter 7 however, we will prefer $C^1(\mathcal{I})$ as we will also require some analytical properties of the transfer operator (to apply analytical perturbation, see subsection 1.4.7) associated with the system.

1.4.6 An application of the spectral decomposition to the values of the digits

In this subsection we will study the digits $m_k(X)$ of the continued fraction expansion of a random number $X \in \mathcal{I}$ from our “real probabilistic framework” perspective. The random number X is drawn uniformly at random from the unit interval and we study the distribution of m_k and the expected values $\mathbb{E}[f(m_k)]$.

These results were already noted by Kuzmin, who demonstrated the convergence of the iterates of the Perron-Frobenius operator of the Gauss map in a more direct fashion. His proof can be found in [Khi97]. Here we apply our results from the section.

Distribution. We remark that

$$m_k(X) = m \iff T^k(X) \in \left(\frac{1}{m+1}, \frac{1}{m} \right],$$

and this means that

$$\mathbb{P}(m_k(X) = m) = \int_{1/(m+1)}^{1/m} \mathbf{H}^k[\mathbf{1}](x) dx,$$

where $\mathbf{1}(x) = 1$ is the initial uniform density.

As we know that the dynamic system associated with the Gauss map satisfies the hypothesis of Theorem 1.12, we get that $\mathbf{H}^k[\mathbf{1}](x) = \frac{1}{\log 2} \frac{1}{1+x} + O(\theta^k)$ for some $\theta < 1$. Integrating

$$\mathbb{P}(m_k(X) = m) = \frac{1}{\log(2)} \int_{1/(m+1)}^{1/m} \frac{dx}{1+x} + O(\theta^k/m^2), \quad (1.57)$$

which gives the exact same distribution from (1.35) but on a different model!

We remark that these results are independent from the choice of the initial density in $\text{BV}(\mathcal{I})$.

Expected values. For the expected values, from (1.57), we deduce that whenever $S := \sum |f(m)|/m < \infty$ we have

$$\mathbb{E}[f(m_k)] = \frac{1}{\log 2} \sum_{m=1}^{\infty} f(m) \log \left(\frac{1 + 1/m}{1 + 1/(m+1)} \right) + O(S \times \theta^k). \quad (1.58)$$

The expected values $\mathbb{E}[f(m_k)]$ may be produced by using the Perron-Frobenius operator in the lines of the ideas exploited in Chapter 5. This is due to the fact that $m_k(X) = \lfloor q_k(X)/q_{k-1}(X) \rfloor$ where $\lfloor \cdot \rfloor$ denotes the integer part. Then the expected value $\mathbb{E}[f(m_k)]$ can be written as $\mathbf{H}^k[g](0)$ directly, for an appropriate g which equals

$$g(x) = \frac{1}{1+x} f \left(\left\lfloor \frac{1}{x} \right\rfloor \right).$$

With such an analysis it is possible to get slightly better error terms, and hence be able to assert the value of $\mathbb{E}[f(m_k)]$ for a larger class of f . We will not get into details here, for more on this topic, the reader is referred to Chapter 5.

Final comments. It is also worthwhile to mention that, for the case of the Euclidean dynamical system, the value of the subdominant spectral radius (when we take out the dominant eigenvalue) is known [IK02] to be φ^2 where $\varphi = \frac{\sqrt{5}-1}{2} \doteq 0.61803 \dots$. This means that we may pick $\theta = \varphi^2 + \epsilon$ for any $\epsilon > 0$.

1.4.7 The transfer operator

The transfer operator is an extension of the Perron Frobenius operator by adding a complex parameter s . We then obtain the transfer operator \mathbf{H}_s defined by

$$\mathbf{H}_s[f](x) := \sum_{h \in \mathcal{H}} |h'(x)|^s f(h(x)). \quad (1.59)$$

This parameter help us express (Dirichlet) generating functions for our parameters of interest in relation to the dynamical system.

Of course, we consider too the k -th iterate of the transfer operator, which again describes the depth k branches of the dynamical system, given by

$$\mathbf{H}_s^k[f](x) = \sum_{h \in \mathcal{H}^k} |h'(x)|^s f(h(x)). \quad (1.60)$$

When we consider all of the inverse branches we get the so-called *quasi-inverse* of the transfer operator:

$$(\mathbf{I} - \mathbf{H}_s)^{-1} := \mathbf{I} + \mathbf{H}_s^1 + \mathbf{H}_s^2 + \dots \quad (1.61)$$

Spectral decomposition. It turns out that, when the Perron Frobenius operator $\mathbf{H} = \mathbf{H}_1$ presents a spectral gap, the transfer operator inherits the spectral gap locally around $s = 1$ by what is called *analytic perturbation*. This is possible when the operator depends analytically on the variable s (see [Sar12]), meaning that the derivative operator is also part of the functional space.

Thus we have a decomposition as before

$$\mathbf{H}_s = \lambda(s)\mathbf{P}_s + \mathbf{N}_s,$$

where again \mathbf{P}_s is the projection onto the eigenspace of $\lambda(s)$, the dominant eigenvalue. It follows then that for $k \geq 1$

$$\mathbf{H}_s^k = (\lambda(s))^k \mathbf{P}_s + \mathbf{N}_s^k,$$

Note then that, if we can ensure that $|\lambda(s)| < 1$ for $s \neq 1$, as well as $R(\mathbf{N}_s) < C < \lambda(s)$ for a fixed constant C , for all s close enough to $s = 1$, the quasi-inverse satisfies

$$(\mathbf{I} - \mathbf{H}_s)^{-1} := \mathbf{I} + \frac{\lambda(s)}{1 - \lambda(s)} \mathbf{P}_s + \frac{\mathbf{N}_s}{\mathbf{I} - \mathbf{N}_s}. \quad (1.62)$$

Suppose, as will be the case for us, that the dominant eigenvalue for $s = 1$ is $\lambda(1) = 1$. The dependence of $\lambda(s)$ and \mathbf{N}_s on s is analytic for s sufficiently close to 1 by principles of perturbation theory, hence we have the estimate

$$(\mathbf{I} - \mathbf{H}_s)^{-1} \sim -\frac{1}{\lambda'(1)} \frac{1}{s - 1} \mathbf{P}, \quad (1.63)$$

as $s \rightarrow 1$.

The approximation in (1.63) is key to apply the Tauberian Theorem (see Theorem 2.3 and the generalization in [Ten15]) and extract asymptotics for the Dirichlet Generating Functions expressed in terms of the quasi-inverse. The process of using the transfer operator, in particular the quasi-inverse, to produce generating functions related to the system is known as *Dynamical Analysis*. The concepts of generating functions, as well as the need for Dynamical Analysis, are explained in Chapter 2.

The transfer operator and the entropy. The reader may wonder what the curious $\lambda'(1)$ in (1.63) is. The answer is that actually $-\lambda'(1)$ gives the entropy (recall subsection 1.2.7) of the system.

We explain this briefly. It is clear that if

$$H_k = - \sum_{h \in \mathcal{H}^k} \left(\int_0^1 |h'(x)| dx \right) \log \left(\int_0^1 |h'(y)| dy \right),$$

then the entropy H is given by

$$H = - \lim_{k \rightarrow \infty} \frac{1}{k} H_k.$$

Let us write first

$$H_k = - \int_0^1 \left(\sum_{h \in \mathcal{H}^k} |h'(x)| \log \left(\int_0^1 |h'(y)| dy \right) \right) dx.$$

If there is bounded distortion, there is a constant C such that $C^{-1}|h'(x)| \leq |h'(y)| \leq C|h'(x)|$ for all x, y and $h \in \mathcal{H}^k$. As a consequence, applying the bounds for $h'(y)$ we get

$$H_k = - \int_0^1 \left(\sum_{h \in \mathcal{H}^k} |h'(x)| \log |h'(x)| \right) dx + O(1),$$

where the constant in $O(1)$ depends only on C and not on k .

Observe that, if we have a spectral gap by analytic perturbation and invariant density φ , then

$$\sum_{h \in \mathcal{H}^k} |h'(x)| \log |h'(x)| = \frac{\partial}{\partial s} \mathbf{H}_s^k[\mathbf{1}](x) \sim k\lambda'(1)\mathbf{P}[\mathbf{1}](x) = k\lambda'(1)\varphi(x),$$

thus the Entropy follows upon integration.

The entropy of a complete interval dynamical system with expanding map may be computed through a classical formula, known as Rokhlin's entropy formula [PY98, pp.133-134].

Proposition 1.20 (Rokhlin's entropy formula). *Consider a complete interval dynamical system with expanding map T . Then the entropy of the dynamical system is given by*

$$H = \int_0^1 \log |T'(x)|\varphi(x)dx,$$

where φ is the unique invariant density of the system.

Example 1.12. We compute again that the entropy for the Euclidean system (see Example 1.5), associated with the Gauss map $T_g(x) = \{1/x\}$, but using Rokhlin's formula. Notice that the branches of T_g all have derivative $-1/x^2$, hence

$$H = -\frac{2}{\log 2} \int_0^1 \frac{\log x}{1+x} dx,$$

as we recall that the invariant density $\varphi(x)$ is the Gauss density $\varphi(x) = \frac{1}{\log 2} \frac{1}{1+x}$.

Here we write $\frac{1}{1+x} = 1 - x + x^2 \mp \dots (-1)^{k-1}x^{k-1} + \frac{(-x)^k}{1+x}$ so that

$$H = -\frac{2}{\log 2} \sum_{j=0}^{k-1} (-1)^{j+1} \int_0^1 x^j (\log x) dx - 2(-1)^k \int_0^1 x^k \frac{\log x}{1+x} dx.$$

Here by parts $\int_0^1 x^j (\log x) dx = -\frac{1}{(j+1)^2}$ while it is clear that $\int_0^1 x^k \frac{\log x}{1+x} dx$ tends to 0 as $k \rightarrow \infty$ by dominated convergence. Thus

$$H = \frac{2}{\log 2} \sum_{j=1}^{\infty} \frac{(-1)^{j+1}}{j^2} = \frac{\pi^2}{6 \log 2} \cdot \diamond$$

The transfer operator in this dissertation. The *quasi-inverse* will play a key role in Chapter 7 where we will study the algorithm associated with the CL expansion. In order to do this we will actually need to consider all possible truncated expansions of the Continued Logarithm expansion. The k -th iterates of the plain Perron Frobenius operator will play a fundamental role in Chapter 5, where we will work with continued fractions of fixed depth. Finally in Chapter 4 we have a novel selection criteria for the branches from \mathcal{H}^* which, however, cannot be directly written in terms of the operator (see Chapter 6 for the relation between the two models).

1.5 Continued fractions: Rationals and quadratics irrationals

Among the various subfamilies of real numbers we may consider, two of them stand out: the rationals and the quadratic irrationals. In this section we will cite the corresponding results for each, explaining the models and the corresponding sizes but we do not offer proofs.

Rational numbers correspond to finite continued fraction expansions. First, these are precisely the executions of the Euclidean algorithm over the positive integers, so their study takes on great importance. Second, we may naturally wonder how the properties of a finite expansion actually turn into the properties of the reals.

Quadratic irrationals are the irrational numbers α which are the root of a degree-two polynomial over the integers. Due to a famous theorem of Lagrange (see [Fog02, pp.183-184] or [Ten15, pp.156-158]), quadratic irrationals $\alpha \in \mathcal{I}$ correspond exactly to continued fraction expansions which are ultimately periodic, meaning of the form

$$\alpha = [m_1, m_2, \dots, m_k, \overline{m_{k+1}, \dots, m_{k+p}}, m_{k+1}, \dots, m_{k+p}, \dots],$$

which we denote by $[m_1, m_2, \dots, \overline{m_k, m_{k+1}, \dots, m_{k+p}}]$. Thus we may also wonder things such as how do the quadratic irrationals transition to reals as we allow larger and larger periods.

Most importantly, quadratic irrationals naturally turn up in Combinatorics on Words (see [All98], [Lot02, Section 2.3.6]) as fixed points of morphisms, such as the one described in Example 3.2. It is not all quadratic irrationals however, but just the ones called *Sturm numbers*, which have a continued fraction expansion of one of the following forms:

- $\alpha = [1, m_0, \overline{m_1, \dots, m_p}]$ with $m_p \geq m_0$,
- $\alpha = [1 + m_0, \overline{m_1, \dots, m_p}]$ with $m_p \geq m_0$.

1.5.1 Probabilistic model for rational numbers

If we restrict ourselves to the set of rational numbers from the interval \mathcal{I} , we may take sets corresponding to fractions with bounded denominators

$$\Omega_N := \{a/b : 1 \leq a \leq b \leq N\},$$

which are finite sets, and draw elements from Ω_N uniformly at random. Observe that $|\Omega_N| = \varphi(1) + \dots + \varphi(N)$ where φ is Euler's totient function $\varphi(n) := \#\{a : 1 \leq a \leq n, \gcd(a, n) = 1\}$.

On this probabilistic model parametrized on N , it is natural to consider probabilities and expected values, which we denote by \mathbb{P}_N and \mathbb{E}_N respectively for this section.

We remark that studying rational trajectories corresponds to the study of the Euclidean algorithm. Indeed, Ω_N corresponds exactly to all pairs of coprime integers (a, b) with $1 \leq a \leq b \leq N$. As such, these average properties have been studied extensively over the years, due to the crucial importance of the algorithm.

1.5.2 Asymptotic probabilistic properties of digits for rational numbers

A function f is of moderate growth if and only if $f(m) = O(\log m)$. We are interested in the cumulative function F over the *continued fraction development* of a rational $x = a/b$, defined by

$$F(x) = f(m_1) + \dots + f(m_k),$$

where $a/b = [m_1, \dots, m_k]$ and $m_k > 1$.

Theorem 1.13 (Simplified Theorem 4 from [Val06]). *The expectation $\mathbb{E}_N[F]$ of the cumulative function F on the set of rationals with denominator bounded by N is asymptotically,*

$$\mathbb{E}_N[F] \sim \frac{12 \log 2}{\pi^2} \mu(f) \log N, \quad \mathbb{E}_N[F^k] \sim (\mathbb{E}_N[F])^k,$$

where $\mu(f) = \frac{1}{\log 2} \sum_{m=1}^{\infty} f(m) \log \left(\frac{1+1/m}{1+1/(m+1)} \right)$.

The standard deviation is $o(\log N)$, and, consequently, the random variable expressing the cumulative sum F is concentrated around its mean.

By concentration: the normalized random variable $F/\mathbb{E}_N[F]$ tends to 1 in probability as $N \rightarrow \infty$. The reader is referred to [FS09, Proposition III.3, p.162] for the corresponding concentration result.

We remark highlight the case $f \equiv 1$, which gives the average run-time of the Euclidean algorithm. The conclusion is that the average number of steps K over Ω_N is $\sim \frac{12 \log 2}{\pi^2} \log N$.

In turn, this means that if we normalized the expected values by the average number of steps we would get

$$\frac{\mathbb{E}_N[F]}{\mathbb{E}_N[K]} \sim \mu(f),$$

which coincides with the behavior from the “real probabilistic model”, see (1.58).

1.5.3 Probabilistic model for quadratics irrationals

Quadratic irrationals are, of course, countable, yet we need an analogue notion of size, like the denominator of a rational number. In what follows we will begin by studying those that have a **purely periodic** continued fraction expansion for simplicity, only afterwards to delve into the ones that do have a preperiod. In [CV17] the quadratic irrationals that have a purely periodic expansion are said to be *reduced*, and we write rqi in shorthand to mean *reduced quadratic irrational*.

It is important to notice that given a period (m_1, \dots, m_p) for our continued fraction expansion, the repetition $(m_1, \dots, m_p, m_1, \dots, m_p)$ is another possible period, and so on. Thus, when working with quadratic irrationals we must be careful to consider the so-called “primitive” periods. A period is primitive if and only if it is not the repetition of a smaller one.

The size of a quadratic irrational

Let us now introduce the norm. Consider a primitive period $m \in \mathbb{N}^k$, its associated rqi x_m^* is the root of $h_m(X) = X$ with $x_m^* \in \mathcal{I}$. The equation $h_m(X) = X$ clearly translates into the zero of a quadratic polynomial $AX^2 + BX + C = 0$ with (A, B, C) relatively prime integers. Then the size $\epsilon(x_m^*) > 1$ of x_m^* is going to be the fundamental unit of the quadratic field $\mathbb{Q}(\sqrt{\Delta})$, where $\Delta = B^2 - 4AC$.

The size function ϵ may be natural from a number-theoretical point of view, but there is another related notion of size v which relates naturally to the periodicity of the continued fraction expansion. This notion of size may be defined by

$$v(x) := \prod_{i=0}^{p(x)-1} T^i(x) = |h'(x)|^{1/2}, \quad (1.64)$$

where $p(x)$ denotes the period of the continued fraction expansion of x , and h is the LFT associated with x .

In what follows, given a tuple $m = (m_1, \dots, m_k)$, $m^R = (m_k, m_{k-1}, \dots, m_1)$ denotes its mirror. Given two tuples $m \in \mathbb{N}^k$ and $u \in \mathbb{N}^j$ we denote by $m \cdot u$ their concatenation $m \cdot u := (m_1, \dots, m_k, u_1, \dots, u_j)$. Also we will write m^ℓ to denote the repetition ℓ times of m , i.e., $m^\ell := \underbrace{m \cdot m \cdot \dots \cdot m}_{\ell \text{ times}}$.

We now explain why v is a natural notion of size for a quadratic irrational.

Proposition 1.21. *Let $u \in \mathbb{N}^j$ and $w \in \mathbb{N}^k$ then the continuant function satisfies*

$$\lim_{\ell \rightarrow \infty} \frac{q(w^\ell)}{q(w^\ell \cdot u)} = |h'_{u^R}(x_{w^R}^*)|^{1/2}, \quad (1.65)$$

in particular, picking $u = w$ produces

$$\lim_{\ell \rightarrow \infty} \frac{q(w^\ell)}{q(w^{\ell+1})} = |h'_{w^R}(x_{w^R}^*)|^{1/2} = v(x_w^*). \quad (1.66)$$

The classical notion of size ϵ for a reduced quadratic irrational is then defined by

$$\epsilon(x) = v(x)^{-r(x)}, \quad \text{with } r(x) = 1 \text{ for even } p(x), \text{ and } r(x) = 2 \text{ for odd } p(x). \quad (1.67)$$

Now given the size ϵ we define

$$\mathcal{K} := \{x \in \mathcal{I} : x \text{ is a rqi}\}, \quad \mathcal{K}_N := \{x \in \mathcal{K} : \epsilon(x) \leq N\}.$$

We remark that \mathcal{K}_N is clearly a finite set. Indeed, we need only show that there is a finite number of rqi's x with $v(x)^{-1} \leq N$. In turn, to prove this, note that $v(x) = (q_{p(x)}(x) + x q_{p(x)-1}(x))^{-1}$ and hence we must have $q_{p(x)} \leq N$ which trivially implies that the quotients $m_k(x)$ and the period $p(x)$ are bounded.

Fixed N we choose a quadratic irrational uniformly at random from \mathcal{J}_N . On this probabilistic model it is again natural to consider probabilities and expected values, which we denote by \mathbb{P}_N and \mathbb{E}_N respectively.

Asymptotic probabilistic properties of digits for quadratics irrational

Here we follow [CV17], stating a simplified version of their main results for reduced quadratic irrationals.

Again, in this context, we consider a function f of moderate growth (i.e., $f(m) = O(\log m)$). We are interested in the cumulative function F over the *period* of a rqi x , defined by

$$F(x) := f(m_1(x)) + \dots + f(m_{p(x)}(x)).$$

Simple examples are $f \equiv 1$, and $f(m) = \log m$, which are related, respectively, to the length of the minimal period and the cost of storing the quotients of x .

Theorem 1.14 (Simplified Theorem 1 from [CV17]). *Consider the set \mathcal{K} of reduced quadratic irrationals x , endowed with the size ϵ . Given a function f of moderate growth, we consider its cumulative sum F on the subset \mathcal{K} . Then the following holds on the set \mathcal{K}_N of reduced quadratic irrationals x with $\epsilon(x) \leq N$ for $N \rightarrow \infty$:*

(i) *The expectation satisfies*

$$\mathbb{E}_N[F] \sim \mu(f) \log N,$$

$$\text{as } N \rightarrow \infty, \text{ where } \mu(f) = \frac{1}{\log 2} \sum_{m=1}^{\infty} f(m) \log \left(\frac{1+1/m}{1+1/(m+1)} \right)$$

(ii) *There is also a constant $\nu(f)$ such that the variance satisfies*

$$\mathbb{V}_N[F] \sim \nu(f) \log N,$$

$$\text{as } N \rightarrow \infty.$$

(iii) *Moreover, the distribution of F is asymptotically Gaussian,*

$$\mathbb{P}_N \left(x : \frac{C(x) - \mu(f) \log N}{\sqrt{\nu(f) \log N}} \leq t \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-u^2/2} du + O \left(\frac{1}{\sqrt{N}} \right).$$

We remark that the expected length of the minimal period $p(x)$ over \mathcal{K}_N is exactly $\log N$ (note that $\mu(1)$ is telescopic). It follows then that

$$\frac{\mathbb{E}_N[F]}{\mathbb{E}_N[p]} \sim \frac{1}{\log 2} \sum_{m=1}^{\infty} f(m) \log \left(\frac{1+1/m}{1+1/(m+1)} \right),$$

which coincides with the behavior from the ‘‘real probabilistic model’’, see (1.58).

As for part (iii), there is also a similar result for the ‘‘real probabilistic model’’ [Val06, Theorem 7], but the variance may differ.

CHAPTER 2

CONCEPTS FROM ANALYTIC COMBINATORICS

In this chapter we introduce key tools for the probabilistic analysis of algorithms and data structures: generating functions. We give a fast and slightly informal overview of the principles of Analytic Combinatorics. We deal, more specifically, with Dirichlet generating functions and Tauberian theorems, which we use extensively in this dissertation. In the context of dynamical systems, these generating functions are naturally written in terms of transfer operators (see section 1.4.7), giving rise to the so-called Dynamical Analysis.

Generating functions encode sequences in a convenient way. Citing Herbert Wilf “A generating function is a clothesline on which we hang up a sequence of numbers for display.” ([Wil06]). Given a sequence, we consider a formal series (classically a power series from $\mathbb{C}[[z]]$) encoding the terms as coefficients of the series. Then operations between generating functions relate to sequences built up from the original ones. This last property is fundamental and forms the basis of the *symbolic* part of *Analytic Combinatorics*.

Analytic Combinatorics, detailed extensively in [FS09], gives a systematic way to study combinatorial objects which can be constructed by combining simple building blocks by following a set of finite rules. These objects should have an appropriate notion of *size*, which is strongly linked to the choice of the generating function type (we mention here 3 types here!). This diversity stems from the fact that generating functions encode the *counting sequence* of the family of objects according to their size, and different generating functions produce quite different kinds of convolutions (Proposition 2.1, Proposition 2.2, Proposition 2.7), which tell us how many object of a given size are produced when we apply the “product” rule. We introduce these notions in subsection 2.1.1.

At this stage, generating functions are purely formal (algebraic) objects. It is then, interpreting them as functions of a complex variable, that their power fully comes to light. It turns out that the singularities of the generating functions (now really seen as functions!), through their nature, determine the asymptotics of the coefficients. Thus we have two steps in the analysis: a symbolic step and an analytic one. The analytic step for classical generating functions (ordinary and exponential) is briefly discussed in subsection 2.1.3.

Numerous data structures and algorithms can be studied through the principles of Analytic Combinatorics. This is not always direct; our set of construction rules (in order to derive the generating functions) must follow the evolution of the algorithm, actually accumulating the meaningful information along the way. When the algorithm has some sort of finite memory or independence in its evolution, the analysis might be simpler, steps separate into products. In some cases though, as is the case of the Euclidean Algorithm over

the integers, all steps depend strongly on one another. This is where *dynamical analysis* comes into play.

Dynamical analysis mixes the methodology from Analytic Combinatorics with objects coming from dynamical systems, in particular the so-called transfer operator, introduced in subsection 1.4.7, which follows naturally the evolution of the dynamical system. With this operator at hand, we may obtain generating functions for our systems, now *Dirichlet generating functions*, which we describe in subsection 2.1.4. In subsection 2.2.1 we state Delange’s [Del54] Tauberian Theorem, which relates the singularities of a Dirichlet generating function with the asymptotics of the *cumulative sums* of its coefficients. The process of dynamical analysis is introduced informally in subsection 2.2.3.

2.1 Generating functions

In this section we introduce generating functions. In section 2.1.1, we begin by introducing the most classical type of generating functions, known as the “ordinary generating functions”, as well as the closely related “exponential generating functions” in section 2.1.2. We mention, in particular, how these types of generating function are related to different combinatorial contexts. Then we introduce the notion of singularities in section 2.1.3, explaining briefly the role played by the dominant (smallest absolute value) singularities in the growth of the coefficients of the generating functions. In section 2.1.4 we introduce Dirichlet generating functions, explaining how they arise, as well as some very important examples coming from number theory. In this context we highlight the Delange Tauberian theorem in section 2.2.1, which links the behaviour of the dominant (now right-most) singularities to the growth of the *partial sums* of the coefficients. Finally we explain in section 2.2.3 the need for the so-called Dynamical Analysis in the study of parameters stemming from a dynamical system like those introduced in Chapter 1.

2.1.1 Ordinary generating functions

In its most basic form, an ordinary generating function is a power series encoding a given sequence as its coefficients. Their formal interest arises from the fact that operations between generating functions correspond to transformations between the coefficient series, with useful combinatorial interpretations when the coefficients count objects or describe probabilities.

Definition 2.1 (Ordinary generating function). The ordinary generating function (OGF) of a sequence $(a_n)_{n \geq 0}$ is the formal power series $A(z) \in \mathbb{C}[[z]]$ defined by

$$A(z) := \sum_{n \geq 0} a_n z^n. \quad (2.1)$$

Notation 2.1. The coefficient corresponding to z^n in a power series $A(z)$ is denoted by $[z^n]\{A(z)\}$.

We underline some basic properties of ordinary generating functions

Proposition 2.1. Let $(a_n)_{n \geq 0}$ and $(b_n)_{n \geq 0}$ be sequences with OGFs $A(z)$ and $B(z)$ respectively. Then the formal product $A(z)B(z)$ is the OGF associated with the convolution sequence $c_n := \sum_{k=0}^n a_k b_{n-k}$.

Example 2.1. We note that $(1 - z)^{-1} = 1 + z + z^2 + z^3 + \dots$ formally over the ring $\mathbb{C}[[z]]$. The proof is just multiplying the formal power series on the right-hand side by $(1 - z)$.

Here the left-hand side $(1 - z)^{-1}$ was interpreted, for the moment, as the inverse of the unit $(1 - z)$ in the ring. Since the formal operations proving that $(1 - z)^{-1} = 1 + z + z^2 + \dots$ make sense over the complex numbers when $|z| < 1$, we conclude that the equality also remains true over the complex numbers when $|z| < 1$, and then the inverse $(1 - z)^{-1}$ is just $\frac{1}{1-z}$. \diamond

Definition 2.2 (Combinatorial class). A combinatorial class is a pair $(\mathcal{A}, |\cdot|_{\mathcal{A}})$ made up of a finite or countable set \mathcal{A} along with a size function $|\cdot|_{\mathcal{A}}$ satisfying the following conditions:

- the size of an element is a non-negative integer;
- the number of elements of any given size is finite.

Given a combinatorial class \mathcal{A} , we denote by \mathcal{A}_n the set of all elements of size n and by A_n the number of such elements. The sequence (A_n) is known as the *counting sequence* of the class.

The ordinary generating function of a combinatorial class \mathcal{A} is the OGF of its counting sequence, which admits the combinatorial form

$$A(z) = \sum_{\alpha \in \mathcal{A}} z^{|\alpha|}. \quad (2.2)$$

It turns out that certain operations between generating functions correspond to operations between combinatorial classes. Here we mention three key constructions (sum, product and quasi-inverse) which will come in handy in order to get a good grasp of the operations in Dynamical Analysis (where we have sum, composition, playing the role of multiplication, and a quasi-inverse too).

Definition 2.3 (Constructions for unlabelled classes). We denote by \mathcal{E} the combinatorial class consisting of a single element of size 0, and we call this the *neutral class*. By \mathcal{Z} we denote the combinatorial class consisting of a single element of size 1, and we call this the *atom class*.

Sum construction. Given combinatorial classes \mathcal{A} and \mathcal{B} we denote by $\mathcal{A} + \mathcal{B}$ the combinatorial class with underlying set

$$\{(0, a) : a \in \mathcal{A}\} \cup \{(1, b) : b \in \mathcal{B}\}$$

and size $|\cdot|_{\mathcal{A}+\mathcal{B}}$ defined by $|(0, a)| = |a|_{\mathcal{A}}$ for $a \in \mathcal{A}$ and $|(1, b)| = |b|_{\mathcal{B}}$ for $b \in \mathcal{B}$.

When $\mathcal{A} \cap \mathcal{B} = \emptyset$ we may identify $\mathcal{A} + \mathcal{B}$ with $\mathcal{A} \cup \mathcal{B}$ and forget the first “label” entry. Observe also that the sum construction can be extended to countably many terms with a little care.

Product construction. We define the product class $\mathcal{A} \times \mathcal{B}$ with the product of the underlying sets and the sizes being simply the sum $|(\beta, \gamma)|_{\mathcal{A} \times \mathcal{B}} := |\beta|_{\mathcal{A}} + |\gamma|_{\mathcal{B}}$.

The product may easily be extended to a finite number of factor combinatorial classes. For a countably infinite number of factors, on the other hand, we have to be more careful to ensure that the resulting counting sequence is well-defined.

Sequence construction. Given a combinatorial class \mathcal{G} with no element of size 0, we define

$$\text{SEQ}(\mathcal{G}) := \mathcal{E} + \mathcal{G} + \mathcal{G} \times \mathcal{G} + \mathcal{G} \times \mathcal{G} \times \mathcal{G} + \dots$$

which corresponds to all finite “sequences” (g_1, \dots, g_k) of elements from \mathcal{G} , $k \geq 0$, along with the size function $|(g_1, \dots, g_k)| = |g_1|_{\mathcal{G}} + \dots + |g_k|_{\mathcal{G}}$. \diamond

Comments regarding the constructions. There exist several other notable constructions such as the “set” or “multi-set”. We will not get into these here, for more constructions and a more in depth exposition, the reader is referred to [FS09].

It is also possible to define classes recursively using our “grammar” of constructions. Determining systematically the well-foundedness of a recursive specification is a non-trivial matter in general, see [PSS12]. To give an example of a recursive specification, the sequence $\text{SEQ}(\mathcal{G})$ can be defined recursively by

$$\text{SEQ}(\mathcal{G}) = \mathcal{E} + \mathcal{Z} \times \text{SEQ}(\mathcal{G}).$$

Combinatorial constructions and power series. Note that the key role of the power series is easily realized from Proposition 2.1, which relates naturally $\mathcal{F} \times \mathcal{G}$ to $F(z) \cdot G(z)$. On the other hand, $F(z) + G(z)$ is the generating function associated with $\mathcal{F} + \mathcal{G}$. Finally, from these two we derive that the OGF associated with $\text{SEQ}(\mathcal{G})$ is $1 + G(z) + (G(z))^2 + \dots = 1/(1 - G(z))$. This is summarized in Table 2.1.

operation	generating function
$\mathcal{F} + \mathcal{G}$	$F(z) + G(z)$
$\mathcal{F} \cdot \mathcal{G}$	$F(z) \cdot G(z)$
$\text{SEQ}(\mathcal{G})$	$\frac{1}{1-G(z)}$

Table 2.1: A small dictionary between our basic constructions and their corresponding generating functions. Here $F(z)$ and $G(z)$ denote the corresponding ordinary generating functions of the unlabeled classes \mathcal{F} and \mathcal{G} respectively. We remark that for exponential generating functions we have a similar dictionary, substituting the product \cdot between labeled combinatorial classes for the labeled product \star below.

2.1.2 Exponential generating functions

There are other types of useful generating functions, which correspond to a different kind of combinatorial objects. Here we give brief and informal account of the so-called *labeled classes*, for more details the reader is referred again to [FS09]. Whereas unlabeled constructions are naturally associated with OGFs, labeled constructions will be associated with the so called *exponential generating functions* EGFs. What we want to highlight here is how the nature of the objects in question dictates the kind of generating function used.

Definition 2.4 (Exponential generating function). The exponential generating function (EGF) of a sequence $(a_n)_{n \geq 0}$ is the formal power series $A(z) \in \mathbb{C}[[z]]$ defined by

$$A(z) := \sum_{n \geq 0} a_n \frac{z^n}{n!}. \quad (2.3)$$

Proposition 2.2. Let $(a_n)_{n \geq 0}$ and $(b_n)_{n \geq 0}$ be sequences with EGFs $A(z)$ and $B(z)$ respectively. Then the formal product $A(z)B(z)$ is the EGF associated with the convolution sequence $c_n := \sum_{k=0}^n \binom{n}{k} a_k b_{n-k}$.

In the unlabeled universe all atoms \mathcal{Z} are equivalent, indistinguishable. In the labeled universe, on the other hand, they are all different and, what is more, they bear different integer “labels”. A labeled object of size n is a graph whose vertices are labeled with distinct integers. An object of size n is said to be well-labeled if the labels form the set $\{1, \dots, n\}$.

Definition 2.5 (Labeled combinatorial class). A labeled combinatorial class is a combinatorial class of well-labeled objects.

Of course, a labeled object γ of size n may be turned into a well-labeled object $\rho(\gamma)$ of size n by making the label set $\{1, \dots, n\}$ and keeping the relative order of the labels of the nodes. The operation ρ is known as the reduction of the labeled object.

Combinatorial constructions and power series. Even though here we may define the sum construction similarly, the product construction is slightly more complex. Given two labeled combinatorial classes \mathcal{A} and \mathcal{B} we will define a new *labeled product* $\mathcal{A} \star \mathcal{B}$ defined as follows

$$\mathcal{A} \star \mathcal{B} := \{(\beta, \gamma) : (\beta, \gamma) \text{ well-labeled, } \rho(\beta) \text{ from class } \mathcal{A}, \rho(\gamma) \text{ from class } \mathcal{B}\}.$$

That is to say, intuitively, an object from $\mathcal{A} \star \mathcal{B}$ with size n is built by first picking a subset S of $\{1, \dots, n\}$, and then an object of size $|S|$ from \mathcal{A} , which we relabel with the set S , and an object from \mathcal{B} , which we relabel according to the set $\{1, \dots, n\} \setminus S$.

With this definition at hand, it turns out that we still have the same dictionary from Table 2.1, but considering labeled classes, exponential generating functions and labeled products instead. We insist that in the labeled case the sequence construction SEQ may also be defined analogously by considering \star instead of \times .

The formula for the generating function of the product $\mathcal{A} \star \mathcal{B}$ of two labeled classes is intuitively clear from Proposition 2.2. Indeed, the binomial coefficient corresponds to the action of choosing how to distribute the n labels for an object of size n between two labeled objects coming from \mathcal{A} and \mathcal{B} with sizes k and $n - k$.

2.1.3 Power series and singularities: the second step

Once we have characterized the generating functions for our objects of interest, we wish to obtain meaningful information regarding their coefficients. Sometimes, when we are lucky, we may deduce precise expressions for our quantities. More often than not, however, the generating functions involved are not that simple and we may only ever hope to derive asymptotics for their coefficients. This is to be expected in general, as the sequence of coefficients may be quite involved. In this section we explain how to extract asymptotics from the *singularities* of the OGFs or EGFs. The subject is classical and can be found in [FS09, Chapter IV] and a brief summary is included here for the sake of introducing the key link between coefficients of generating functions and their corresponding singularities.

Analytic functions. In what follows, we look at our generating functions as actual functions on the complex plane, $f: \mathcal{D} \rightarrow \mathbb{C}$, for an appropriately chosen open set $\mathcal{D} \subseteq \mathbb{C}$.

The introduction of complex numbers makes for a much stronger notion of differentiability: a complex function is (complex) differentiable on a domain \mathcal{D} if and only if it is analytic at every point of \mathcal{D} . We explain what this means. The function $f(z)$ is analytic at $z_0 \in \mathcal{D}$ if there are complex coefficients a_0, a_1, \dots which make $f(z) = \sum a_k (z - z_0)^k$ for z on some open ball around z_0 . When f is analytic at every point of \mathcal{D} , we say that it is *analytic* on \mathcal{D} .

From the equivalence between being differentiable and analytic, it follows that a complex function that is differentiable once is actually differentiable infinitely many times!

This is not the only remarkable property satisfied by analytic functions; if the open domain \mathcal{D} is connected and the set of zeros of f has an accumulation point in \mathcal{D} , the function f must be identically zero. This property has key consequences. One of them, known as the *uniqueness of analytic continuation*, tells us that an analytic function $f: \mathcal{D} \rightarrow \mathbb{C}$ has only one analytic extension^a $g: \mathcal{D}' \rightarrow \mathbb{C}$ if \mathcal{D}' is a connected open set. Intuitively, the value of g on \mathcal{D} “determines” the rest.

Back to our generating functions, the series actually yield analytic functions whenever the former are convergent. Power series actually have a so-called radius of convergence which is intimately related to the coefficients of the series (see Proposition 2.3). In turn, the radius of convergence is associated with the singularities (see Definition 2.6) of the complex-valued function derived from the GF.

First principle of Coefficient Asymptotics. A first indicator of the order of growth of the coefficients of a power series is the so-called *radius of convergence* (this is introduced in any elementary calculus or analysis book such as). The radius of convergence of a power series $\sum_{n \geq 0} a_n z^n$ is the supremum of all possible $r > 0$ such that the series converges for $|z| < r$. If there is no such r , the radius of convergence is defined to be 0.

Proposition 2.3. *The radius of convergence R of a power series $f(z) = \sum_{n \geq 0} a_n z^n$ is determined by*

$$R = \frac{1}{\limsup |a_n|^{1/n}}, \quad (2.4)$$

whenever the \limsup is non-zero, and it is said to be infinite otherwise.

In other words, given $\epsilon > 0$

^aWe say that $g: \mathcal{D}' \rightarrow \mathbb{C}$ is an extension of $f: \mathcal{D} \rightarrow \mathbb{C}$ when $\mathcal{D} \subset \mathcal{D}'$, $g|_{\mathcal{D}} = f$

- We have $|a_n|^{1/n} < \frac{1}{R} + \epsilon$ for all large enough n
- We have $|a_n|^{1/n} > \frac{1}{R} - \epsilon$ for infinitely many n .

Proposition 2.3 tells us that the radius of convergence largely dictates the asymptotic growth of the coefficients of the power series (more precisely, their “maximal” exponential growth, namely $\limsup |a_n|^{1/n}$), but how do we determine the radius of convergence of a given generating function? This is where the so called *singularities* will come into play: a power series converges until the circle (centered at the origin) of radius $|z|$ “hits” a singularity!

Singularities. We now introduce a key concept: the *singularities*. We briefly comment on how these are related to the radius of convergence, giving a characterization of the radius of convergence.

Definition 2.6 (Singularity). An analytic function $f : \mathcal{D} \rightarrow \mathbb{C}$ defined in an open set \mathcal{D} has a singularity at $z_0 \in \partial\mathcal{D}$ if and only if it is not analytically continuable at z_0 .

The function f function is analytically continuable to a point z_0 if there exists an analytic function g , defined on a larger open set \mathcal{D}' containing \mathcal{D} and z_0 , coinciding with f on \mathcal{D} .

Proposition 2.4. If $f(z)$ is analytic at the origin, and its power series expansion has a radius of convergence R , then there must be a singularity of f in the circle $|z| = R$.

In order to prove this proposition, we would first have to prove the so-called Cauchy’s Integral Theorem and derive Cauchy’s formula for the coefficients, see for instance [Rud87] or [FB09].

The singularities of smallest absolute value, i.e., R , the radius of convergence, are known as the *dominant singularities* of $f(z)$ and play a larger role in the asymptotics of the coefficients.

When $f(z) = \sum_{n \geq 0} f_n z^n$ has non-negative coefficients $f_n \geq 0$, we can say at more with regard to the positions of the singularities, in a similar fashion to the Perron–Frobenius theorem for matrices.

Theorem 2.1 (Pringsheim’s Theorem). Suppose $f(z)$ is analytic at the origin, and that the coefficients in the expansion are all non-negative. If the radius of convergence is $R > 0$, then R is a singularity of f .

Second principle of Coefficient Asymptotics. We have related the radius of convergence R to the analytic properties, namely the position of the singularities, of the complex function $f(z)$ given by the power series. In turn, we know that this radius of convergence determines the maximal exponential growth of the coefficients of its associated power series. We now explain, beginning from an example, how to get precise asymptotics from the knowledge of the dominant singularities in the case of meromorphic functions. The final principle is stated in Proposition 2.6.

Example 2.2 (Euclidean algorithm for polynomials over \mathbb{F}_q). In this example we follow the method of [BLV16] to study the number of steps performed by the Euclidean algorithm over $\mathbb{F}_q[x]$ (see also the example from [FS09, Example IX.15, pp. 662–663] where it is presented from the perspective of continued fractions).

Let us suppose that our input polynomials $(a(x), b(x))$ satisfy $\deg b(x) \leq \deg a(x)$, then the Euclidean algorithm proceeds by performing successive divisions

$$\begin{cases} a(x) = q_1(x)b(x) + r_1(x), & \deg r_1(x) < \deg b(x) \\ b(x) = q_2(x)r_1(x) + r_2(x), & \deg r_2(x) < \deg r_1(x) \\ & \vdots \\ r_{k-1}(x) = q_k(x)r_k(x) + 0. \end{cases}$$

The algorithm stops when the remainder equals zero, i.e., $r_{k+1}(x) = 0$, and outputs the previous one $r_k(x)$. Then we are interested in studying the average number of division steps $k = k(a(x), b(x))$. If, otherwise, $\deg b(x) > \deg a(x)$ the algorithm proceeds as above for the reversed input $(b(x), a(x))$.

We explain what we mean by average: pick the pair $(a(x), b(x))$ uniformly at random from the set

$$\Omega_n := \{(a(x), b(x)) : \deg a(x) + \deg b(x) = n, a(x), b(x) \text{ monic polynomials}\},$$

where we recall that monic means that the leading coefficient is 1. What is the average of $k = k(a(x), b(x))$?

A few comments are in order:

- The whole process remains unchanged if we divided $a(x)$ and $b(x)$ by their gcd; each remainder r_i is just divided by $\gcd(a(x), b(x))$. Thus we will mainly work with the coprime pair $\widehat{a}(x) := a(x)/\gcd(a(x), b(x))$, $\widehat{b}(x) := b(x)/\gcd(a(x), b(x))$, determining the execution of the algorithm.
- In turn, the reduced pair $(\widehat{a}(x), \widehat{b}(x))$ is determined univocally by the sequence of quotients $(q_1(x), \dots, q_k(x))$. Note that $q_1(x)$ is monic by assumption, while the rest $q_2(x), q_3(x), \dots, q_k(x)$ can be arbitrary polynomials.

Thus, the input pair is characterized by the gcd and the sequence of quotients.

Let us consider $U(z) = \frac{1}{1-qz}$, the OGF of monic polynomials, and $G(z) = \frac{(q-1)qz}{1-qz}$, the OGF of the general polynomials of positive degree. These expressions follow from the sequence and product construction.

We “mark” the degree of $a(x)$ by z and the degree of $b(x)$ by t , thus rendering our generating function a bivariate one (in the ring $\mathbb{C}[[z, t]]$). By this we mean that the coefficient of $z^n t^m$ will correspond to a sum over the cases in which $\deg a(x) = n$ and $\deg b(x) = m$. Even though we have not formally introduced multivariate series, we remark that we may realize such a series as a series where the coefficients are themselves series in another variable. The combinatorial operations of sum, multiplication and sequence extend to series in several variables, corresponding to the same operations between generating functions as before.

We have the bivariate identity

$$U(z) \cdot U(t) = \underbrace{U(zt)}_{\text{gcd}} \times \underbrace{(1 + (U(z) - 1) + (U(t) - 1))}_{q_1(x) \Leftrightarrow \text{monic polynomial}} \times \underbrace{\frac{1}{1 - G(zt)}}_{q_2(x), \dots, q_k(x) \Leftrightarrow \text{SEQ}(\text{general polynomials of } \deg \geq 1)}.$$

This expression is justified by the fact that the gcd contributes equally to the degree of both $a(x)$ and $b(x)$, hence the factor $U(zt)$, while the first quotient $q_1(x)$ may contribute to either the degree of $a(x)$ or $b(x)$ depending on which of the two has the smallest degree. Finally, the quotients $q_2(x), \dots, q_k(x)$ contribute equally to the degrees of both $a(x)$ and $b(x)$, hence the identity.

Let us set $z = t$, forgetting the distinction between the two. This makes z mark the sum of the degrees of $a(x)$ and $b(x)$. In order to get the depth, which is just the number of quotients $q_2(x), \dots, q_k(x)$ plus 1, we introduce a variable u marking the number of quotients

$$F(z, u) := \sum_{(a(x), b(x)) \in \Omega} u^{\#\text{steps}(a(x), b(x))} z^{|(a(x), b(x))|} = U(z^2) \times \left((2U(z) - 1) \times \frac{u}{1 - uG(z^2)} \right).$$

Operating with our expressions for $U(z)$ and $G(z)$ we obtain

$$F(z, u) = \frac{(qz + 1)u}{(1 - q^2uz^2 - q(1 - u)z^2)(1 - qz)}.$$

Differentiating in u and evaluating at $u = 1$ produces

$$S(z) := \sum_{(a(x), b(x)) \in \Omega} \#\text{steps}(a(x), b(x)) z^{|(a(x), b(x))|} = \left. \frac{\partial}{\partial u} F(z, u) \right|_{u=1} = \frac{1 - qz^2}{(1 + qz)(1 - qz)^3}.$$

In terms of $F(z, u)$, our expected value reads

$$\mathbb{E}_{\Omega_n}[\#steps] = \frac{\sum_{(a(x), b(x)) \in \Omega_n} \#steps(a(x), b(x))}{\sum_{(a(x), b(x)) \in \Omega_n} 1} = \frac{[z^n]S(z)}{[z^n]F(z, 1)}.$$

At this point we could actually compute the exact coefficients of both $S(z)$ and $F(z, 1)$, as our functions are rational (quotient of polynomials) and apply the identity Equation 2.5 below. However we will do it by looking directly at the nature of the singularities of both generating functions, thus explaining the process of singularity extraction.

For the case of S , the dominant singularities are given by $z = 1/q$, with multiplicity three, and $z = -1/q$ with multiplicity 1. Thus we concentrate on these by applying partial fractions

$$S(z) = \frac{A}{(1 - qz)^3} + \frac{B}{(1 - qz)^2} + \frac{C}{1 - qz} + \frac{D}{1 + qz} + R(z), \quad A = \frac{q-1}{2q},$$

for certain constants B, C, D and $R(z)$ that has a larger radius of convergence. The first four terms actually involve a well-known series because

$$\frac{1}{(1 - z)^n} = \sum_{k=0}^{\infty} \binom{n+k-1}{k} z^k, \quad (2.5)$$

as follows upon differentiating $(n-1)$ times the identity $1/(1-z) = \sum_{k=0}^{\infty} z^k$.

From (2.5) we get, by applying Proposition 2.3 to $R(z)$, that

$$[z^n]S(z) = A \cdot \frac{(n+2)(n+1)}{2} \cdot q^n + B \cdot (n+1) \cdot q^n + Cq^n + D(-q)^n + O((q-\epsilon)^n)$$

for some $\epsilon > 0$.

For $F(z, 1)$ we have $F(z, 1) = (U(z))^2 = 1/(1 - qz)^2$ thus $[z^n]F(z, 1) = (n+1)q^n$ and we derive

$$\mathbb{E}_{\Omega_n}[\#steps] \sim A \frac{n+2}{2}.$$

The constant $A = \frac{q-1}{2q}$ is computed simply by multiplying $S(z)$ by $(1 - qz)^3$ and taking $z \rightarrow 1/q$. \diamond

The previous example showcases several important concepts we now underline.

Probability and bivariate generating functions. First, suppose we had a bivariate generating function

$$F(z, u) = \sum_{\alpha \in \Omega} u^{C(\alpha)} z^{|\alpha|},$$

where Ω is the set of objects, $|\alpha|$ the weight of α and $C(\alpha)$ the cost of α , a function we want to characterize. From $F(z, u)$ we may express at once quantities such as the expected value of $C(\alpha)$ when α is chosen uniformly at random from Ω_n , the set of objects of size n . Indeed

$$\mathbb{E}_{\Omega_n}[C] = \frac{[z^n] \left\{ \frac{\partial}{\partial u} F(z, u) \Big|_{u=1} \right\}}{[z^n]F(z, 1)},$$

as in the previous example but also note that

$$\mathbb{E}_{\Omega_n}[C^k] = \frac{[z^n] \left\{ (u \frac{\partial}{\partial u})^k F(z, u) \Big|_{u=1} \right\}}{[z^n]F(z, 1)},$$

where $(u \frac{\partial}{\partial u})^k$ means we apply the operator $u \frac{\partial}{\partial u}$ exactly k times. This is what Wilf in [Wil06] calls the “ xD operator”. In [FS09] this operation corresponds to a “pointing” construction; the pointing $\Theta\mathcal{C}$ of a combinatorial class \mathcal{C} corresponds to distinguishing one atom from the structure as special.

We may even express the probability distributions in terms of $F(z, u)$

$$\mathbb{P}_{\Omega_n}(C = i) = \frac{[z^n u^i]F(z, u)}{[z^n]F(z, 1)}.$$

Asymptotics and the nature of the singularities. Second, the asymptotic growth is dictated by the nature of the dominant (of smallest absolute value) singularities. In the example above we have rational functions, but the principle applies more generally to the so-called *meromorphic functions*. Meromorphic functions are functions which are locally a quotient of two analytic functions. For that case we just approximate our meromorphic function around each singularity by a rational function.

We remark that the dominant singularities “may” cancel out, and so these may not suffice to get the asymptotic behavior of *all* coefficients. To clarify what we mean, consider the trivial $1/(1 - z^2) + 1/(1 - z/2)$. Here we have $1/(1 - z^2) = 1/(1 - z) + 1/(1 + z)$, and so the contribution of the dominant poles indeed cancels out (as they should!) for odd coefficients.

We state the principle here, for further details see [FS09, Chapter IV].

Proposition 2.5. *Let $F(z) = \sum_n a_n z^n$ be meromorphic on the disc $|z| \leq R$, having poles at the nonzero points $\alpha_1, \dots, \alpha_k$. Assume further that $F(z)$ can be continued analytically to all points on the circle $|z| = R$. Then there are polynomials Π_1, \dots, Π_k such that:*

$$a_n = \sum_{j=1}^k \Pi_j(n) \alpha_j^{-n} + O(R^{-n}).$$

Furthermore, the degree of Π_j is equal to the degree of the pole at α_j minus 1.

When we have a unique dominant (of smallest absolute value) singularity, matters are simpler.

Proposition 2.6. *For a meromorphic generating function $F(z) = \sum a_n z^n$, having a unique dominant (i.e., of smallest absolute value) singularity σ of type $F(z) \sim \frac{G(z)}{(z-\sigma)^{1+a}}$ where $G(z)$ is analytic at $z = \sigma$, we have*

$$a_n = G(\sigma) \frac{(-1)^{1+a}}{\sigma^{1+a}} \binom{n+a}{a} \sigma^{-n} + O\left(\frac{1}{(|\sigma| + \epsilon)^n}\right),$$

for some $\epsilon > 0$, as $n \rightarrow \infty$. Remark that $\binom{n+a}{a}$ is a polynomial of degree a . Further, we have $\binom{n+a}{a} = \frac{n^a}{a!} + O(n^{a-1})$.

Observation 2.1. In our simplified explanation for the singularity analysis, we have not talked about effective error bounds. It is possible to actually bound the size of the coefficients of a power series by considering more precise expressions in terms of contour integrals in the complex plane. Indeed, Cauchy’s integral formula (see [FB09, pp. 96-98]) tells us that for a power series $f(z) = \sum a_n z^n$, we have

$$a_n = \frac{1}{2\pi i} \oint_{\gamma} f(z) \frac{dz}{z^{n+1}},$$

for any simple (and piece-wise smooth) contour γ within the radius of convergence of f . These integral expressions lend themselves to bounding by considering bounds for $|f(z)|$ on specific circles $|z| = r$. That is how the classical *saddle-point bounds* are derived (see [FS09, pp. 546-548]).

2.1.4 Dirichlet generating functions

Problems stemming from number theory many a times involve functions that are called *multiplicative*. An arithmetical function $f: \mathbb{N} \rightarrow \mathbb{C}$ is multiplicative if $f(ab) = f(a)f(b)$ when $\gcd(a, b) = 1$, and completely multiplicative if $f(ab) = f(a)f(b)$ for all $a, b \in \mathbb{N}$.

There are numerous multiplicative functions; for instance the number of divisors $d(n)$, the number of prime divisors $\omega(n)$, Euler's φ function and the celebrated Möbius μ function (see Definition 2.9 below). The appropriate generating functions to these functions is not a power series, but rather a Dirichlet series.

Definition 2.7 (Dirichlet generating function). The Dirichlet generating function (DGF) of a sequence $(a_n)_{n \geq 1}$ is the formal power series

$$A(s) := \sum_{n \geq 1} \frac{a_n}{n^s}. \quad (2.6)$$

Dirichlet generating functions naturally follow the process of multiplication.

Proposition 2.7. Let consider two sequences $(a_n)_{n \geq 1}$ and $(b_n)_{n \geq 1}$ with associated DGFs $A(s)$ and $B(s)$ respectively. Then the product $A(s)B(s)$ is the DGF of the convolution $\left(\sum_{d|n} a_d b_{n/d}\right)_{n \geq 1}$.

Note here that $\sum_{d|n}$ denotes the sum over all positive divisors d of n .

The most basic building block for DGFs is the so-called Riemann Zeta function.

Definition 2.8 (Riemann Zeta function). The Riemann Zeta function $\zeta(s)$, or Zeta function for short, is defined by the series

$$\zeta(s) := \sum_{n \geq 1} \frac{1}{n^s}, \quad \Re s > 1. \quad (2.7)$$

Of course, the series (2.7) converges absolutely for $\Re s > 1$ and so it actually defines an analytic function on this half-plane. It can, however, be extended analytically to the whole plane except $s = 1$.

Example 2.3 (Divisor counting function). Define $d(n) = \sum_{d|n} 1$. The DGF of $d(n)$ is $(\zeta(s))^2$.

A fundamental function in the context of DGFs is the so-called Möbius function $\mu(n)$ defined as follows

Definition 2.9 (Möbius function). Let n be a positive integer. Then $\mu(n)$ is defined as follows

$$\mu(n) := \begin{cases} (-1)^k, & \text{if } n \text{ is free of squares and has exactly } k \text{ prime divisors,} \\ 0, & \text{otherwise.} \end{cases}$$

The fundamental property of the Möbius function is given by the following proposition which tells us that μ actually allows us to perform a “sieving” over the “divisibility property”.

Proposition 2.8. For every positive integer n , the Möbius function satisfies

$$\sum_{d|n} \mu(d) = \begin{cases} 1, & \text{if } n = 1, \\ 0, & \text{otherwise.} \end{cases} \quad (2.8)$$

Proof. The proof is simple enough. We only care about the divisors which are square-free, as the μ -function is 0 for the rest. Write the decomposition of n into product of prime powers $n = p_1^{a_1} \dots p_k^{a_k}$, then all square free divisors of n are given by $d = \prod_{s \in S} p_s$ where $S \subset \{1, \dots, k\}$, as we may not pick a prime number twice. Of course, $\mu(\prod_{s \in S} p_s) = (-1)^{|S|}$ and for each j we have $\binom{k}{j}$ possible subsets S with $|S| = j$. Thus $\sum_{d|n} \mu(d) = \sum_{j=0}^k \binom{k}{j} (-1)^j$ and the result follows from the Binomial Theorem. ■

Observation 2.2. Furthermore

$$F(n) := \sum_{d|n} g(d), \forall n \implies g(n) = \sum_{d|n} \mu(d)F(n/d). \quad (2.9)$$

We are going to briefly explain why μ corresponds to inclusion-exclusion (see [Sta97, pp.64-65], [Wil06, Chapter 4]) with this example. Following the proof of Proposition 2.8, we note that

$$\sum_{d|n} \mu(d)F(n/d) = \sum_{S \subset \{1, \dots, k\}} (-1)^{|S|} F\left(\frac{n}{\prod_{s \in S} p_s}\right).$$

The term $F\left(\frac{n}{\prod_{s \in S} p_s}\right)$ corresponds to a sum of $g(k)$ over all divisors k of n for which the exponent of p_s is not a_s for $s \in S$. Thus we are actually trying to exclude the terms $g(k)$, $k|n$, with the property “there is a prime p_i such that $p_i^{a_i} \nmid k$ ”. Of course, what remains after the cancellation can just be $g(n)$. \diamond

The property from Observation 2.2 is known as the *Möbius inversion formula* [Apo98, p. 32], and can be explained purely in terms of DGFs in quite a direct way, even though the underlying combinatorial interpretation is the inclusion-exclusion above.

Proposition 2.9. *The Dirichlet generating function of the Möbius function $\mu(n)$ is $\frac{1}{\zeta(s)}$.*

Proof. If $F(s) := \sum \mu(n)n^{-s}$, then by Proposition 2.7 and Proposition 2.8 we get $F(s)\zeta(s) = 1$. \blacksquare

The hypothesis Observation 2.2 translates into DGFs as follows

$$\sum F(n)n^{-s} = \zeta(s) \sum g(n)n^{-s},$$

therefore

$$\sum g(n)n^{-s} = \frac{1}{\zeta(s)} \sum F(n)n^{-s}.$$

Now this proves the result because $\frac{1}{\zeta(s)}$ is the DGF of the Möbius function μ (see Proposition 2.9) and the product gives the DGF of the convolution by Proposition 2.7.

Observation 2.3. An important application of the Möbius function which we will use in Chapter 4 is to “filter” the pairs of integers (a, b) with $\gcd(a, b) = 1$. Indeed, suppose we had a positive function $f: \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{R}_{\geq 0}$ and wanted to compute $\sum_{(a,b): \gcd(a,b)=1} f(a, b)$. Then

$$\sum_{(a,b): \gcd(a,b)=1} f(a, b) = \sum_{(a,b) \in \mathbb{N} \times \mathbb{N}} f(a, b) \sum_{d| \gcd(a,b)} \mu(d).$$

Note that $d| \gcd(a, b)$ if and only if both $d|a$ and $d|b$. Thus

$$\sum_{(a,b): \gcd(a,b)=1} f(a, b) = \sum_{(a,b) \in \mathbb{N} \times \mathbb{N}} f(a, b) \sum_{d|a \wedge d|b} \mu(d),$$

and reversing the sums (for this we should add some conditions on f)

$$\sum_{(a,b): \gcd(a,b)=1} f(a, b) = \sum_{d=1}^{\infty} \mu(d) \sum_{(a,b): d|a \wedge d|b} f(a, b),$$

from which we conclude

$$\sum_{(a,b): \gcd(a,b)=1} f(a, b) = \sum_{d=1}^{\infty} \mu(d) \sum_{(a,b) \in \mathbb{N} \times \mathbb{N}} f(a \cdot d, b \cdot d). \quad (2.10)$$

This last equation is extremely useful when f satisfies some homogeneity conditions such as $f(ad, bd) = d^{-r} f(a, b)$ for some $r > 1$ (this already makes the series for f converge absolutely, and the reversing of the sums is valid as well!) and we know the value of $S := \sum_{(a,b)} f(a, b)$, because then we deduce

$$\sum_{(a,b):\gcd(a,b)=1} f(a, b) = S \sum_{d=1}^{\infty} \frac{\mu(d)}{d^r} = \frac{S}{\zeta(r)}. \quad (2.11)$$

A consequence of the previous remark is that the density of coprime pairs is $6/\pi^2$ (see [Ten15])

Proposition 2.10. *The density $D_N := \#\{(a, b) : 1 \leq a, b \leq N, \gcd(a, b) = 1\}/N^2$ of coprime integers on $\{1, \dots, N\}^2$ satisfies*

$$D_N = \frac{6}{\pi^2} + O((\log N)/N).$$

Proof. Consider a function $f := f_N(a, b)$ that is 1 iff $\gcd(a, b) = 1$ with $1 \leq a, b \leq N$ and 0 otherwise. Indeed, from (2.10) it follows that

$$\#\{(a, b) : 1 \leq a, b \leq N, \gcd(a, b) = 1\} = \sum_{d=1}^{\infty} \mu(d) \left[\frac{N}{d} \right]^2 = \frac{N^2}{\zeta(2)} + O\left(N \sum_{d \leq N} \frac{1}{d}\right),$$

and $\zeta(2) = \pi^2/6$ yields the result. ■

2.2 Analytic Combinatorics for Dirichlet generating functions

From the point of view of Analytic Combinatorics, Dirichlet series correspond to integer weights that are multiplicative, i.e. $|\alpha\beta| = |\alpha||\beta|$, for all α, β in the combinatorial class.

Dirichlet generating functions as analytic objects. Ordinary generating and exponential functions have a *radius* of convergence where they present at least one point where they cannot be continued analytically. Dirichlet generating functions, on the other hand, have a *half-plane* of convergence. This is explained by the following classical theorem (see e.g., [Apo98, p. 245]).

Theorem 2.2. *If the Dirichlet series $F(s) = \sum \frac{a_n}{n^s}$ converges for s_0 then it also converges for all s with $\Re s > \Re s_0$. If the series diverges for s_0 , then it diverges for all s with $\Re s < \Re s_0$.*

Dirichlet series in general (if they are not everywhere convergent or everywhere divergent) present two fundamental abscissas: the abscissa of convergence σ_c , for which the series converges if $\Re s > \sigma_c$, and the abscissa of absolute convergence σ_a , for which the series converges absolutely if $\Re s > \sigma_a$. If the terms of the series are positive, as will be our case, then both abscissas coincide $\sigma_a = \sigma_c$. In any case, it is fair to underline that the inequality $0 \leq \sigma_a - \sigma_c \leq 1$ holds (see [Apo98]).

From the point of view of the generating functions as complex-valued functions, the convolutional product from Proposition 2.7 makes sense within the intersection of the half-planes of convergence. By the uniqueness of the analytic continuations (see [FB09, pp.126–128]), we do not pay attention to such details; we first operate formally, deriving our DGFs, and then make sense of them on their corresponding half-planes.

Example 2.4. The series in (2.7), defining the Riemann zeta function, converges for $\Re s > 1$. Thus it defines an analytic function over the half-plane $\Re s > 1$. It turns out that the zeta function can be extended analytically to the whole plane with the sole exception of $s = 1$ (see [WW96, Chapter XIII]) where

$$\zeta(s) = \frac{1}{s-1} + \gamma + \sum_{k \geq 1} \frac{(-1)^k}{k!} \gamma_k \cdot (s-1)^k, \quad (2.12)$$

where γ is the Euler-Mascheroni constant defined by $\gamma := \lim_n \sum_{k=1}^n \frac{1}{k} - \log n$. The coefficients γ_k of the expansion are called the Stieltjes constants. \diamond

2.2.1 Tauberian theorems

Tauberian theorems are the “transfer theorems”, telling us how the information on the singularities translates into asymptotics, corresponding to Dirichlet generating functions. It will be the case for DGF that we will not be able to derive asymptotics for individual coefficients but rather their cumulative sums. This is to be expected as, for instance, DGFs encode irregular functions such as the number of divisors function $d(n)$.

The following theorem is a classical Tauberian theorem from Delange [Del54].

Theorem 2.3 (Delange). *Let $F(s)$ be a Dirichlet series with non negative coefficients such that $F(s)$ converges when the real part of s satisfies $\Re s > \sigma > 0$. Assume that*

- $F(s)$ is analytic on $\Re s = \sigma$, $s \neq \sigma$.
- there is $a > -1$ such that

$$F(s) = \frac{G(s)}{(s - \sigma)^{1+a}} + H(s),$$

where $G(s)$ and $H(s)$ are analytic at $s = \sigma$ with $G(\sigma) \neq 0$.

Then

$$\sum_{n \leq N} a_n \sim \frac{G(\sigma)}{\sigma \Gamma(a+1)} N^\sigma \log^a N,$$

as $N \rightarrow \infty$.

The statement of Theorem 2.3 is to be compared with the singularity analysis for meromorphic functions.

More generally, this holds too for non-integer sizes and we consider more general DGFs of the form $F(s) = \sum_{i \in I} a_i g(i)^{-s}$ where $g(i)$ need not be an integer anymore.

Proposition 2.11. *Let $(a_i)_{i \in I}$ be a family of non-negative numbers indexed on a countable set I , and let $g: I \rightarrow (0, \infty)$. Suppose the series $F(s) = \sum_{i \in I} a_i g(i)^{-s}$ converges when the real part of s satisfies $\Re s > \sigma > 0$. Assume that*

- $F(s)$ is analytic on $\Re s = \sigma$, $s \neq \sigma$.
- there is $a > -1$ such that

$$F(s) = \frac{G(s)}{(s - \sigma)^{1+a}} + H(s),$$

where $G(s)$ and $H(s)$ are analytic at $s = \sigma$ with $G(\sigma) \neq 0$.

Then

$$\sum_{i: h(i) \leq N} a_i \sim \frac{G(\sigma)}{\sigma \Gamma(a+1)} N^\sigma \log^a N,$$

as $N \rightarrow \infty$.

2.2.2 An example of Analytic Combinatorics in arithmetics

In this example we study the gcd of ℓ numbers picked independently and uniformly at random from $\{1, \dots, N\}$. We want to compute its k -th order moment, i.e.,

$$\mathcal{M}_{N,k,\ell}(s) := \frac{1}{N^\ell} \sum_{s_1=1}^N \dots \sum_{s_\ell=1}^N (\gcd(s_1, \dots, s_\ell))^k,$$

as $\ell \rightarrow \infty$. This example originates from a private communication between Joachim von zur Gathen and Brigitte Vallée [Val08].

In order to compute this we consider auxiliary functions $\varphi_\ell(n)$, extending Euler's φ function. Let us consider first the case $\ell = 2$ which corresponds to Euler's φ function.

Case $\ell = 2$. A key identity of Euler's totient function φ states that $\sum_{d|n} \varphi(d) = n$ for all $n \in \mathbb{N}$. This identity can be interpreted combinatorially as follows. First rewrite, by symmetry of the divisors, $\sum_{d|n} \varphi(d) = \sum_{d|n} \varphi(n/d)$ and note that $\varphi(n/d)$ is the number of integers in $\{1, \dots, n/d\}$ that are coprime to n/d . Then $\varphi(n/d)$ corresponds too to the number of integers m from $\{1, \dots, n\}$ that satisfy $\gcd(n, m) = d$ (indeed $\gcd(n, m) = d \iff \gcd(n/d, m/d) = 1$).

Thus the sum $\sum_{d|n} \varphi(n/d)$ counts all integers m from $\{1, \dots, n\}$ by dividing them according to $d = \gcd(n, m)$. Therefore we conclude that $\sum_{d|n} \varphi(n/d) = n$.

It is important to remark that Proposition 2.7 then yields

$$\left(\sum_{n=1}^{\infty} \frac{\varphi(n)}{n^s} \right) \times \zeta(s) = \left(\sum_{n=1}^{\infty} \frac{n}{n^s} \right) = \zeta(s-1) \Rightarrow \sum_{n=1}^{\infty} \frac{\varphi(n)}{n^s} = \frac{\zeta(s-1)}{\zeta(s)}. \quad (2.13)$$

More generally, the above argument yields that

$$\sum_{d|n} \varphi(n/d) d^k = \sum_{m=1}^n (\gcd(n, m))^k.$$

Let us write

$$A_{n,k} := \sum_{d|n} \varphi(n/d) d^k = \sum_{m=1}^n (\gcd(n, m))^k.$$

The convolution defining $A_{n,k}$ implies (thanks to Proposition 2.7)

$$\sum_{n=1}^{\infty} \frac{A_{n,k}}{n^s} = \frac{\zeta(s-1)}{\zeta(s)} \times \zeta(s-k).$$

The cumulative sums of $A_{n,k}$ which yields the Tauberian theorem (Theorem 2.3) give

$$S_{N,k} := \sum_{n \leq N} A_{n,k} \sim \begin{cases} \frac{N^{k+1} \log N}{k+1} \frac{\zeta(2)}{\zeta(k)} & k = 1, \\ \frac{N^{k+1}}{k+1} \frac{\zeta(k)}{\zeta(k+1)} & k > 1. \end{cases}$$

This is due to the nature of the dominant singularity. Indeed, to analyze the singularities we employ (2.7). When $k = 1$ we have a double pole at $s = 2$, due the numerator $(\zeta(s-1))^2$. When $k > 1$ we have a dominant singularity at $s = k+1$ which is a simple pole.

The sum defining $S_{N,k}$ sums $(\gcd(n, m))^k$ over $m \leq n$. To get the other cases, we exploit the symmetry and deduce

$$\sum_{n=1}^N \sum_{m=1}^N (\gcd(n, m))^k = 2S_{N,k} - \sum_{n=1}^N n^k,$$

as we have summed the "diagonal case" $n = m$ twice. Notice here that $\sum_{n=1}^N n^k = \frac{N^{k+1}}{k+1} + O(N^k)$.

When $k > 1$ this means that

$$\frac{1}{N^2} \sum_{n=1}^N \sum_{m=1}^N (\gcd(n, m))^k \sim \left(\frac{2}{k+1} \frac{\zeta(k)}{\zeta(k+1)} - \frac{1}{k+1} \right) N^{k-1}. \quad (2.14)$$

For the case $k = 1$, the diagonal case is of a smaller order than $2S_{N,1}$ therefore we simply get

$$\frac{1}{N^2} \sum_{n=1}^N \sum_{m=1}^N \gcd(n, m) \sim 2S_{N,1}/N^2 \sim \frac{\log N}{\zeta(2)}. \quad (2.15)$$

General case. We extend the φ function by defining φ_ℓ to be

$$\varphi_\ell(n) := \#\{(y_1, \dots, y_{\ell-1}) \in \{1, \dots, n\}^{\ell-1} : \gcd(n, y_1, \dots, y_{\ell-1}) = 1\}, \quad (2.16)$$

in particular $\varphi_2(n) = \varphi(n)$, the classical φ function. As a consequence we get, as for the case $\ell = 2$, that

$$\varphi_\ell(n/d) = \#\{(y_1, \dots, y_{\ell-1}) \in \{1, \dots, n\}^{\ell-1} : \gcd(n, y_1, \dots, y_{\ell-1}) = d\},$$

and therefore

$$\sum_{d|n} \varphi_\ell(n/d) d^k = \sum_{y_1=1}^n \dots \sum_{y_{\ell-1}=1}^n (\gcd(n, y_1, \dots, y_{\ell-1}))^k. \quad (2.17)$$

In particular from the convolution of the case $k = 0$ we deduce its DGF

$$\left(\sum_{n=1}^{\infty} \frac{\varphi_\ell(n)}{n^s} \right) \times \zeta(s) = \sum_{n=1}^{\infty} \frac{n^{\ell-1}}{n^s} = \zeta(s - \ell + 1) \Rightarrow \sum_{n=1}^{\infty} \frac{\varphi_\ell(n)}{n^s} = \frac{\zeta(s - \ell + 1)}{\zeta(s)}. \quad (2.18)$$

Define

$$A_{n,k,\ell} := \sum_{y_1=1}^n \dots \sum_{y_{\ell-1}=1}^n (\gcd(n, y_1, \dots, y_{\ell-1}))^k,$$

thus from the convolution we get

$$\sum_{n=1}^{\infty} \frac{A_{n,k,\ell}}{n^s} = \frac{\zeta(s - \ell + 1)}{\zeta(s)} \times \zeta(s - k). \quad (2.19)$$

From the singularities we get the asymptotics for the cumulative sums,

$$S_{N,k,\ell} = \sum_{n \leq N} A_{n,k,\ell} = \begin{cases} \frac{N^{k+1}}{k+1} \frac{\zeta(k - \ell + 2)}{\zeta(k+1)} & k > \ell - 1, \\ \frac{N^{k+1}}{k+1} \frac{\log N}{\zeta(k+1)} & k = \ell - 1, \\ \frac{N^\ell}{\ell} \frac{\zeta(\ell - k)}{\zeta(\ell)} & k < \ell - 1. \end{cases}$$

This follows from the nature of the singularities, which in turn depends on whether $k > \ell - 1$ (simple pole at $s = k + 1$), $k = \ell - 1$ (pole of order two at $s = k + 1 = \ell$), or $k < \ell - 1$ (simple pole at $s = \ell$).

Here the ‘‘diagonal’’ cases are subtracted through a process of inclusion-exclusion (see [Sta97, pp.64-65] or [Wil06, Ch.4])

$$\sum_{y_0=1}^N \dots \sum_{y_{\ell-1}=1}^N (\gcd(y_0, y_1, \dots, y_{\ell-1}))^k = \binom{\ell}{1} S_{N,k,\ell} - \binom{\ell}{2} S_{N,k,\ell-1} \pm \dots \quad (2.20)$$

Basically, first we sum ℓ times as the maximum of the entries may be anywhere, but we must subtract the cases in which we have that the maximum entry in the tuple appears twice, etc...

When $k > \ell - 1$, this means that

$$\frac{1}{N^\ell} \sum_{y_0=1}^N \cdots \sum_{y_{\ell-1}=1}^N (\gcd(y_0, y_1, \dots, y_{\ell-1}))^k \sim \frac{1}{\zeta(k+1)} \frac{N^{k+1-\ell}}{k+1} \sum_{j=1}^{\ell} \binom{\ell}{j} \zeta(k - \ell + 1 + j) (-1)^{j+1}. \quad (2.21)$$

For $k = \ell - 1$ the “diagonal cases” we subtract are of smaller order, hence we get just the first term of the inclusion-exclusion

$$\frac{1}{N^\ell} \sum_{y_0=1}^N \cdots \sum_{y_{\ell-1}=1}^N (\gcd(y_0, y_1, \dots, y_{\ell-1}))^k \sim \frac{\log N}{\zeta(\ell)}. \quad (2.22)$$

Finally, when $k < \ell - 1$, the first term $\binom{\ell}{1} S_{N,k,\ell}$ with the “largest ℓ ” dominates over the others

$$\frac{1}{N^\ell} \sum_{y_0=1}^N \cdots \sum_{y_{\ell-1}=1}^N (\gcd(y_0, y_1, \dots, y_{\ell-1}))^k \sim \frac{\zeta(\ell - k)}{\zeta(\ell)}. \quad (2.23)$$

We summarize in the following proposition

Proposition 2.12. *Consider random variables $Y_i^{(N)}$, $i = 0, \dots, \ell - 1$, drawn independently and uniformly at random from $\{1, \dots, N\}$. The expected value of the gcd of these variables has 3 different phases depending on whether $k < \ell - 1$, $k = \ell - 1$ or $k > \ell - 1$.*

$$\mathbb{E}[(\gcd(Y_0^{(N)}, \dots, Y_{\ell-1}^{(N)}))^k] \sim \begin{cases} C_{k,\ell} \frac{N^{k+1-\ell}}{\zeta(k+1)} & \text{if } k > \ell - 1, \\ \frac{\log N}{\zeta(\ell)} & \text{if } k = \ell - 1, \\ \frac{\zeta(\ell - k)}{\zeta(\ell)} & \text{if } k < \ell - 1, \end{cases} \quad (2.24)$$

where

$$C_{k,\ell} := \frac{1}{k+1} \sum_{j=1}^{\ell} \binom{\ell}{j} \zeta(k - \ell + 1 + j) (-1)^{j+1}.$$

Observation 2.4. The ideas from Observation 2.3 give a fast way to compute the moments of the greatest common divisor. Indeed

$$\frac{1}{N^\ell} \sum_{y_0=1}^N \cdots \sum_{y_{\ell-1}=1}^N (\gcd(y_0, y_1, \dots, y_{\ell-1}))^k = \sum_{s=1}^{\infty} \mu(s) \sum_{d=1}^{\infty} d^k \left\lfloor \frac{N}{d \times s} \right\rfloor^\ell \frac{1}{N^\ell}. \quad (2.25)$$

This is significantly faster than the calculation on the left-hand side since the number of pairs (s, d) on the right-hand side that are non-zero is at most $\sum_{s=1}^{\infty} \lfloor \frac{N}{s} \rfloor \sim N \log N$.

From (2.25) it is clear that we get (2.23) when $k < \ell - 1$ by estimating $\lfloor x \rfloor = x + O(1)$, but the rest of the cases are more involved because the error due to this approximation of the integer part does not vanish.

Second, from (2.25) we observe that (note simply that k marks the gcd)

$$\lim_{N \rightarrow \infty} \mathbb{P}_N \left(\gcd(Y_0^{(N)}, Y_1^{(N)}, \dots, Y_{\ell-1}^{(N)}) = d \right) = \frac{d^{-\ell}}{\zeta(\ell)},$$

where the random variables $Y_i^{(N)}$ are drawn independently and uniformly at random from $\{1, \dots, N\}$. \diamond

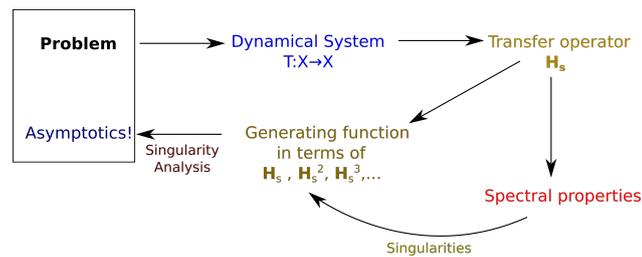


Figure 2.1: The principles of dynamical analysis.

2.2.3 What is Dynamical Analysis? Why is it useful?

In the previous examples coming from number theory, even though it may be non trivial to find expressions for the corresponding DGFs, in the end we did derive an “explicit” generating function in terms of the well-known function ζ . This will not always be the case when we consider DGFs arising in a dynamical system. In this context we make use of the *transfer operator* \mathbf{H}_s , introduced in section 1.4.7, which follows the evolution of the dynamical system. The transfer operator is a building block for our generating functions related to the time evolution of the system. However, there is little hope of extracting an explicit formula for this operator as the dynamical system may be quite correlated from step to step (notice that without correlation, we would have a product of GFs by the principles of Analytic Combinatorics).

After deriving an expression for our DGFs in terms of the transfer operator, we would like to extract asymptotics by means of Delange’s Tauberian Theorem (i.e., Theorem 2.3). In order to do this, we require information concerning the dominant singularities of the operator. This is where the spectral properties of the transfer operator \mathbf{H}_s come into play, dictating the behavior of \mathbf{H}_s^k at time k . The functional space of choice takes on a crucial importance; it must be expressive enough to allow us to produce our generating functions, but, at the same time, we prefer having fewer eigenvalues (we want a spectral gap, see Section 1.4).

Part II

Studies in Word Combinatorics

CHAPTER 3

STURMIAN WORDS

Sturmian words constitute a family of words of primary importance in the field of Word Combinatorics. Introduced by Gustav A. Morse and Marston Hedlund [MH40] in 1940, they represent the simplest non-trivial words in a precise sense we will soon describe. This chapter aims to explain why these objects, which are seemingly purely of combinatorial nature, admit a surprising, and convenient, arithmetical characterization.

We start off from some basic definitions from Combinatorics on Words, in particular that of a Sturmian word (see Definition 3.3), and build up the machinery necessary to characterize Sturmian words as what are called *mechanical sequences*. This equivalent characterization is given by Theorem 3.1 due to Morse and Hedlund [MH40].

With this characterization at hand, we move on to explain why the recurrence function, a fundamental function measuring how much it takes for the factors of a word to reappear, of Sturmian words can be seen purely in terms of continuants deriving from the continued fraction expansion of the frequency of 1s, known also as the “slope” of the word, within the corresponding word. This remarkable property is given in Theorem 3.3 and is due also to Morse and Hedlund [MH40].

3.1 Concepts from Combinatorics on Words

Consider a finite set \mathcal{A} of *symbols* (also *letters*), which we call the *alphabet*. We are interested in the properties of *words* over this alphabet. Words are sequences of symbols from \mathcal{A} , and may be *infinite* $\mathbf{u} = (u_n)_{n \in \mathbb{N}}$ from $\mathcal{A}^{\mathbb{N}}$, or *finite*. We denote by \mathcal{A}^* the set of all finite words over the alphabet \mathcal{A} , while its elements will be denoted by lower-case letters such as w, v and u . The set \mathcal{A}^* contains a special element ϵ , known as the *empty word*.

The length of a finite word $w = w_1 \dots w_n$ where $w_1, \dots, w_n \in \mathcal{A}$ is denoted by $|w| = n$, and the length of the empty word is $|\epsilon| = 0$. An important operation between words is the concatenation \cdot . Given words $u = u_1 \dots u_n$ and $v = v_1 \dots v_m$, their concatenation $u \cdot v$ is the word $u \cdot v = u_1 \dots u_n v_1 \dots v_m$. This operation turns \mathcal{A}^* into a monoid with identity ϵ .

The symbol \cdot is omitted in general. This is coherent, as each individual symbol can be thought of as a word and then $w = w_1 \dots w_n$ can be interpreted as saying that w is the concatenation of n length-1 words.

3.1.1 The complexity function of an infinite word

We study the properties of infinite words, particularly with regard to their *factors*. Let $\mathbf{u} = (u_n)_{n \in \mathbb{N}}$ be an infinite word in $\mathcal{A}^{\mathbb{N}}$. A finite word w of length n is a factor of \mathbf{u} if and only if there exists an index m in \mathbf{u} for which $w = u_m \dots u_{m+n-1}$.

Let $\mathcal{L}_{\mathbf{u}}(n)$ stand for the set of factors of length n of \mathbf{u} , while $\mathcal{L}_{\mathbf{u}}$ stands for the set of all finite factors of \mathbf{u} and is known as its language. Two functions describe the set $\mathcal{L}_{\mathbf{u}}(n)$ of finite factors of the word \mathbf{u} , namely the complexity and the recurrence function.

Definition 3.1 (Complexity function). The *complexity function* of $\mathbf{u} \in \mathcal{A}^*$ is the sequence

$$p_{\mathbf{u}} : n \mapsto |\mathcal{L}_{\mathbf{u}}(n)|.$$

Periodic words are amongst the “simplest” ones in terms of the complexity function, as they clearly satisfy $p_{\mathbf{u}}(n) \leq C$ for some constant C . Sturmian words are, informally, the “next simplest words”.

Let us see some basic properties of this function

Proposition 3.1. *Let $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ be an infinite word. The complexity function $p_{\mathbf{u}}(n)$ satisfies the simple properties*

$$p_{\mathbf{u}}(n) \leq |\mathcal{A}|^n, \quad p_{\mathbf{u}}(n) \leq p_{\mathbf{u}}(n+1), \quad p_{\mathbf{u}}(0) = 1, \quad (3.1)$$

as well as

$$p_{\mathbf{u}}(n+m) \leq p_{\mathbf{u}}(n) \cdot p_{\mathbf{u}}(m), \quad (3.2)$$

for all $n, m \geq 0$.

Inequality (3.2) implies that the complexity function is sub-multiplicative, hence by Fekete’s Lemma [Fek23] (applied to the logarithms) the limit

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log p_{\mathbf{u}}(n), \quad (3.3)$$

exists and is known as the entropy of the word \mathbf{u} . The entropy (3.3) can be thought of as an upper bound for the compressibility of \mathbf{u} , given that the cost per symbol (taking base of the logarithm to be 2) we would have if we simply coded each block of length n as an integer $\{0, \dots, p_{\mathbf{u}}(n) - 1\}$ with fixed length in binary.

Definition 3.2 (Eventually periodic word). A word $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ is said to be eventually periodic if and only if there are finite words $w, v \in \mathcal{A}^*$, with $v \neq \epsilon$ such that $\mathbf{u} = wv^{\mathbb{N}}$ where $v^{\mathbb{N}}$ is the infinite word $v \cdot v \cdot v \dots$. In such a case w is said to be the preperiod of \mathbf{u} , while v is known as the period of \mathbf{u} .

Here is a characterization of the eventually periodic words in terms of the complexity function

Proposition 3.2. *A word $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ is eventually periodic if and only if $p_{\mathbf{u}}(n) \leq n$ holds for some n .*

Proof. It is clear that $p_{\mathbf{u}}(n)$ is bounded when \mathbf{u} is periodic. Indeed, the factors w of length n within \mathbf{u} are always among $u_1 \dots u_n, u_2 \dots u_{n+1}, \dots, u_{i+j} \dots u_{i+j+n-1}$ where i is the length of the preperiod of \mathbf{u} and j the length of the period of v .

Conversely, assume $p_{\mathbf{u}}(n) \leq n$. Then for some k we must have $p_{\mathbf{u}}(k) = p_{\mathbf{u}}(k+1)$. Indeed, it is clear that $p_{\mathbf{u}}(k+1) \geq p_{\mathbf{u}}(k)$, and, if we never had equality, then $p_{\mathbf{u}}(k+1) \geq 1 + p_{\mathbf{u}}(k)$ which implies $p_{\mathbf{u}}(n) \geq n+1$ as we always have $p_{\mathbf{u}}(0) = 1$ because of the empty word ϵ . As this is not the case, there is k such that $p_{\mathbf{u}}(k) = p_{\mathbf{u}}(k+1)$.

This equality means that each factor of length k within \mathbf{u} has exactly one extension to length $k+1$, this means that $w_1 \dots w_k$ determine w_{k+1} . Starting from $u_1 \dots u_k$, which determines univocally $u_2 \dots u_{k+1}$, at some point we must repeat a word of length k as $\mathcal{L}_{\mathbf{u}}(k)$ is finite. Thus \mathbf{u} is eventually periodic. ■

Observation 3.1. As we have seen during the previous proof, a word $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ is eventually periodic

- if and only if $p_{\mathbf{u}}(n) = p_{\mathbf{u}}(n+1)$ holds for some n ,

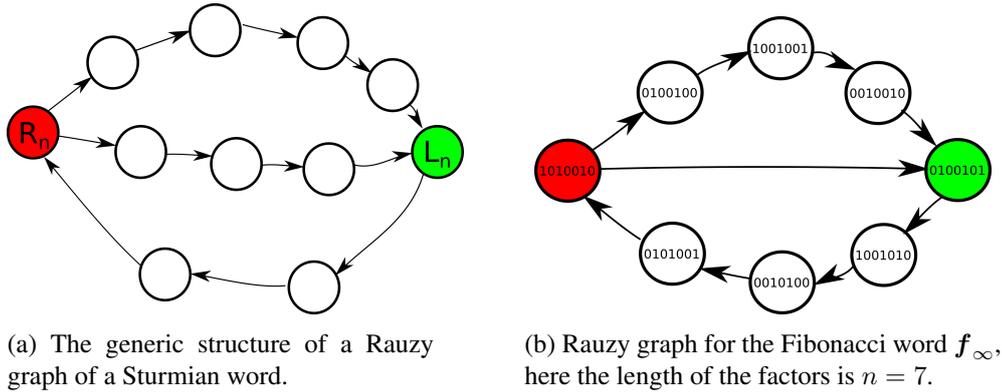


Figure 3.1: The structure of a Rauzy graph of a Sturmian word. Observe that, since each word has exactly one right extension except for R_n , all nodes except from R_n have outdegree 1 while R_n has outdegree 2. Similarly, all nodes save for L_n have indegree 1, while L_n has indegree 2. This leaves us with a structure like the one in the figure: there are two paths from R_n to L_n and one coming back from L_n to R_n .

1. the words 01 and 10 are factors of \mathbf{u} .
2. exactly **one** of the words 00, 11 is a factor of \mathbf{u} .
3. there is only one word $R_n \in \mathcal{L}_{\mathbf{u}}(n)$ that can be extended on the right in two ways $R_n0 \in \mathcal{L}_{\mathbf{u}}(n+1)$ and $R_n1 \in \mathcal{L}_{\mathbf{u}}(n+1)$. The other have only one right extension.
4. there is only one word $L_n \in \mathcal{L}_{\mathbf{u}}(n)$ that can be extended on the left in two ways $0L_n \in \mathcal{L}_{\mathbf{u}}(n+1)$ and $1L_n \in \mathcal{L}_{\mathbf{u}}(n+1)$. The other have only one right extension.

Proof. Property 1 follows from the fact that there are infinitely many 0s and 1s. Now, property 2 follows from the Definition 3.3 which tells us that $p_{\mathbf{u}}(2) = 3$.

For property 3 we observe that both $p_{\mathbf{u}}(n) = n + 1$ and $p_{\mathbf{u}}(n + 1) = n + 2$. Since every factor has some right-extension, this tells us that exactly one must have two right-extensions. Finally, item 4 is analogous as every factor has a left-extension (because by the recurrence we may assume that it is not a prefix of \mathbf{u}). ■

Observation 3.3. Note that R_n is necessarily a suffix of R_{n+1} and L_n is a prefix of L_{n+1} .

The structure of $\mathcal{L}_{\mathbf{u}}(n)$ described in Proposition 3.4 is best seen from its *Rauzy graph* $\mathcal{G}_{\mathbf{u}}(n)$ (see e.g., [Ber96]) : the Rauzy graph $\mathcal{G}_{\mathbf{u}}(n)$ associated with an infinite word \mathbf{u} is a finite automaton with state space $\mathcal{L}_{\mathbf{u}}(n)$ which describes a moving “sliding window” of length n along the word \mathbf{u} ; From a state $b \in \mathcal{L}_{\mathbf{u}}(n)$ upon scanning letter y , the next word is $b \cdot y \in \mathcal{L}_{\mathbf{u}}(n + 1)$: we label the edge with y and the next state is $\tau(b \cdot y)$ where $\tau(f)$ for a word f just erases the leftmost symbol of f . A Sturmian Rauzy graph of order n is thus particularly sparse: it has $n + 1$ vertices and $n + 2$ edges. The special words L_n and R_n play a particular role in the Rauzy graph, this is shown in Figure 3.1

The Rauzy Graph $\mathcal{G}_{\mathbf{u}}(n)$, illustrated in Figure 3.1, is strongly-connected as \mathbf{u} is recurrent. As a consequence its structure is composed of 3 oriented paths, forming two cycles. This is closely related to the 3-distance Theorem (see [Sós58] and [Ber96]), which tells us that the words of $\mathcal{L}_{\mathbf{u}}(n)$ have only 3 possible frequencies in \mathbf{u} , and gives explicit expressions for them.

Thus far we have not explained how to build Sturmian words, which is not immediate from Definition 3.3. In Section 3.2 we give an explicit theorem by Morse and Hedlund [MH40], telling us that Sturmian words occur exactly as a particular coding of the orbit of circle rotations (of irrational angle) introduced in subsection 1.2.2. This not only gives an explicit construction for Sturmian words, it characterizes them all.

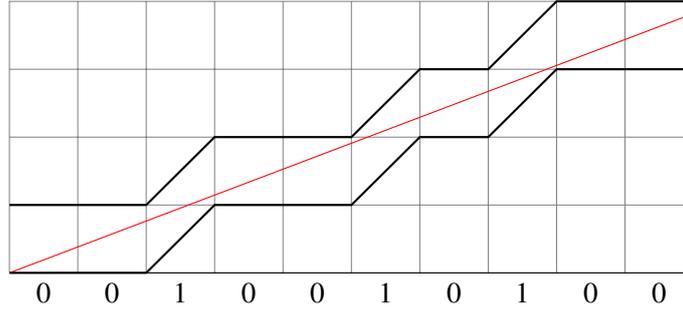


Figure 3.2: In digital geometry $\underline{\mathfrak{S}}$ and $\overline{\mathfrak{S}}$ code discrete lines. In the picture $\underline{\mathfrak{S}}(\alpha, 0)$, where α is the slope.

3.2 The arithmetic nature of Sturmian words

Although Sturmian words may seem purely combinatorial objects, which look difficult to construct, it turns out that all Sturmian words occur as special codings of circle rotations with an irrational angle α . This is the central result of this section, but along the way we shall prove several properties regarding Sturmian words. We cite Morse and Hedlund

“Sturmian trajectories possess certain numerical characteristics, namely, a frequency, a pole, and a type index, and admit mechanical constructions uniquely determined by these characteristics.” [MH40]

This equivalence between Sturmian words and codings of circle rotations was first proved by Morse and Hedlund in [MH40]. They, however, did not originally define Sturmian words as above. They defined Sturmian words to be what in the notation of [Fog02] is called a “balanced word”. This concept that will come in handy during the proof of this characterization of Sturmian words.

Definition 3.4 (Mechanical sequences). Given a pair $(\alpha, \beta) \in [0, 1]^2$, we define two infinite words $\underline{\mathfrak{S}}(\alpha, \beta)$ and $\overline{\mathfrak{S}}(\alpha, \beta)$ whose n -th symbols are respectively $\lfloor \alpha(n+1) + \beta \rfloor - \lfloor \alpha n + \beta \rfloor$ for $\underline{\mathfrak{S}}(\alpha, \beta)$ and $\lceil \alpha(n+1) + \beta \rceil - \lceil \alpha n + \beta \rceil$ for $\overline{\mathfrak{S}}(\alpha, \beta)$.

The words $\underline{\mathfrak{S}}(\alpha, \beta)$ and $\overline{\mathfrak{S}}(\alpha, \beta)$ are depicted in Figure 3.2. In digital geometry, these words code the line $y = \alpha x + \beta$ from above ($\overline{\mathfrak{S}}$) and below ($\underline{\mathfrak{S}}$) by horizontal lines (coded by a 0) and diagonals (coded by a 1). In consequence it is commonplace to call α the *slope* of the word.

Theorem 3.1. [Morse and Hedlund] [MH40] *A word $u \in \{0, 1\}^{\mathbb{N}}$ is Sturmian if and only if it equals $\underline{\mathfrak{S}}(\alpha, \beta)$ or $\overline{\mathfrak{S}}(\alpha, \beta)$ for a pair (α, β) formed by an irrational $\alpha \in (0, 1)$ and a real $\beta \in [0, 1)$.*

We now explain why the sequences from Theorem 3.1 correspond to codings of circle rotations and why they are actually Sturmian when α is irrational. The equivalence with rotations is stated in Proposition 3.5.

A coding of irrational circle rotations. We recall that in circle rotations, we consider the unit circle $\mathbb{T}^1 = \mathbb{R}/\mathbb{Z}$, which is comprised of all reals mod 1. We will consider the representatives in $[0, 1)$, thus the projection of a real number t onto the circle is represented by $t \bmod 1 = \{t\}$, the fractional part of t .

Rotations $\mathcal{R}_\alpha : \mathbb{T}^1 \rightarrow \mathbb{T}^1$ are defined by $\mathcal{R}_\alpha(x) := (x + \alpha) \bmod 1$, and we will be interested in coding the trajectory (or orbit) of an initial point $\beta \in \mathbb{T}^1$, which is given by $(\beta, \mathcal{R}_\alpha(\beta), \mathcal{R}_\alpha^2(\beta), \dots)$.

The words $\underline{\mathfrak{S}}(\alpha, \beta)$ and $\overline{\mathfrak{S}}(\alpha, \beta)$ correspond to special codings of circle rotations. Indeed, note for example that for the word $\underline{\mathfrak{S}}(\alpha, \beta)$ we have the n -th symbol equal to 1 if and only if going through the circle modulo 1 from $\alpha n + \beta$ to $\alpha(n+1) + \beta$ we traverse $0 \equiv 1$, that is, if $\alpha n + \beta \bmod 1$ belongs to $I_1 = [1 - \alpha, 1)$. The coding is illustrated in Figure 3.3.

Proposition 3.5. *The infinite word $u = \underline{\mathfrak{S}}(\alpha, \beta)$ is given in terms of the intervals $I_0 = [0, 1 - \alpha)$ and $I_1 = [1 - \alpha, 1)$ by coding the rotation of β , considering the n -th symbol ($n \geq 0$) to be 1 when $\mathcal{R}_\alpha^n(\beta) \in I_1$.*

$n\alpha + \beta \pmod 1$ belongs to I_1 and 0 otherwise. Analogously, we may consider $I_0 = (0, 1 - \alpha]$ and $I_1 = (1 - \alpha, 1]$ with the identification $0 \equiv 1$, thus giving the word we denote by $\mathfrak{S}(\alpha, \beta)$.

Symbols u_i in $\mathbf{u} = u_0u_1u_2 \dots$ indicate whether $\mathcal{R}_\alpha^i(\beta) \in I_0$ or $\mathcal{R}_\alpha^i(\beta) \in I_1$, this process can be iterated to explain what happens with larger factors $w \in \{0, 1\}^n$ of length n . The result being that distinct factors are in correspondence with the circle intervals delimited by the points [AB98]

$$0, -\alpha, -2\alpha, \dots, -n\alpha,$$

modulo 1. Intuitively, to know what happens for the next symbol we must rotate the borders by $-\alpha$.

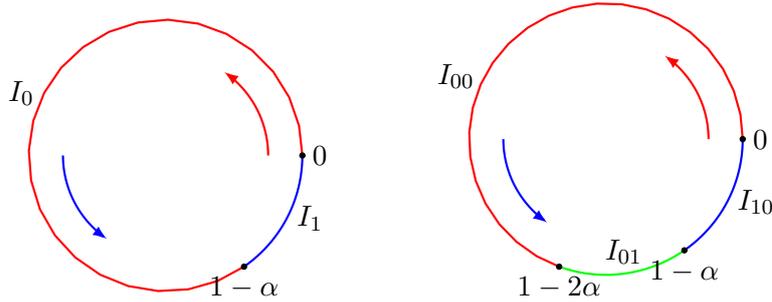


Figure 3.3: Intervals corresponding to the coding of the rotation of the circle.

When α is irrational, factors of length n correspond to exactly distinct $n + 1$ intervals. Thus we find immediately that $p_{\mathbf{u}}(n) \leq n + 1$ for each n . Equality actually holds because the orbit is dense in the circle^a. Thus $\underline{\mathfrak{S}}(\alpha, \beta)$ and $\overline{\mathfrak{S}}(\alpha, \beta)$ are Sturmian words when the slope α is irrational.

Observation 3.4. It is important to remark that the fact that the length of I_1 coincides with the angle of rotation is fundamental to produce Sturmian words.

Indeed, suppose that we coded 1 on the interval $I_1 = [1 - \alpha', 1)$ with $\alpha' > \alpha$ and $\alpha + \alpha' < 1$. Then the intervals corresponding to words of length 2 are necessarily

$$I_{00} = [0, 1 - \alpha - \alpha'), \quad I_{01} = [1 - \alpha - \alpha', 1 - \alpha'), \quad I_{11} = [1 - \alpha', 1 - \alpha), \quad I_{10} = [1 - \alpha, 1),$$

so we have all possible words of length 2, giving a word that is not Sturmian. \diamond

A final important remark is that $\underline{\mathfrak{S}}(\alpha, \beta)$ and $\overline{\mathfrak{S}}(\alpha, \beta)$ coincide unless $\alpha k + \beta$ is an integer for some k (this can happen only for a single integer k), in which case they differ just on two symbols: indeed

$$\lfloor \alpha(k + 1) + \beta \rfloor - \lfloor \alpha k + \beta \rfloor = 0 \neq 1 = \lceil \alpha(k + 1) + \beta \rceil - \lceil \alpha k + \beta \rceil$$

and

$$\lfloor \alpha k + \beta \rfloor - \lfloor \alpha(k - 1) + \beta \rfloor = 1 \neq 0 = \lceil \alpha k + \beta \rceil - \lceil \alpha(k - 1) + \beta \rceil,$$

due to the floor and ceiling functions.

Strategy for the proof of Theorem 3.1. The proof is classical, here we follow [Fog02]. As we have already shown that $\underline{\mathfrak{S}}(\alpha, \beta)$ and $\overline{\mathfrak{S}}(\alpha, \beta)$ are Sturmian, we need only prove the converse, that is: if \mathbf{u} is a Sturmian word, then either $\mathbf{u} = \underline{\mathfrak{S}}(\alpha, \beta)$ or $\mathbf{u} = \overline{\mathfrak{S}}(\alpha, \beta)$.

The proof goes through the following steps:

- (1) the frequency of 1s, as defined in Definition 3.5, exists for every Sturmian word \mathbf{u} , call it α .
- (2) α is irrational.

^aEnough to see that $n \cdot \alpha + m \cdot 1 = \epsilon$ can be arbitrary small (follows from Observation 1.3), as then $|\lfloor \frac{t}{\epsilon} \rfloor n \cdot \alpha + \lfloor \frac{t}{\epsilon} \rfloor m - t| \leq \epsilon$.

- (3) two Sturmian words with the same α have the same language of factors.
- (4) a Sturmian word \mathbf{u} with frequency of 1s α can be “approximated” by taking increasingly long factors from $\underline{\mathfrak{S}}(\alpha, 0)$ and $\overline{\mathfrak{S}}(\alpha, 0)$.
- (5) there is a convergence of this sequence of factors (over $\mathcal{A}^{\mathbb{N}}$) to either $\underline{\mathfrak{S}}(\alpha, \beta)$ or $\overline{\mathfrak{S}}(\alpha, \beta)$ for some β .

We formalize the notion of frequency and discuss the frequency of the words $\underline{\mathfrak{S}}(\alpha, \beta)$ and $\overline{\mathfrak{S}}(\alpha, \beta)$ defined in Definition 3.4 which is both fundamental and motivates our plan.

Definition 3.5 (Frequency of a factor). Let $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ and consider a factor $w \in \mathcal{A}^*$. Then, the limit

$$f_w := \lim_{n \rightarrow \infty} \frac{1}{n} \left| \{ 0 \leq i \leq n - |w| : u_{i+1} \dots u_{i+|w|} = w \} \right|, \quad (3.4)$$

if it exists, is called the frequency of w in \mathbf{u} .

The numerator in (3.4) represents the number of occurrences of w in $u_1 \dots u_n$. It is important to compare this with Birkhoff’s Ergodic Theorem (Theorem 1.1).

For $\underline{\mathfrak{S}}(\alpha, \beta)$ and $\overline{\mathfrak{S}}(\alpha, \beta)$, the frequencies exist for every factor w . Indeed, for $w = w_1 \dots w_n$, consider the fundamental interval $I_w := I_{w_1} \cap \mathcal{R}_\alpha^{-1} I_{w_2} \cap \dots \cap \mathcal{R}_\alpha^{-(n-1)} I_{w_n}$. Then, thanks to Proposition 3.5, the indicator function $\mathbf{1}_{I_w}(\beta)$ tells us whether w occurs at the beginning of the word (index 0). Thus the sum $\mathbf{1}_{I_w}(\beta) + \mathbf{1}_{I_w}(\mathcal{R}_\alpha^1 \beta) \dots + \mathbf{1}_{I_w}(\mathcal{R}_\alpha^{n-1} \beta)$ is number of occurrences of w among the indices $0, \dots, n-1$. Proposition 1.12 tells us that the frequency of w actually exists and equals the length of I_w .

Observe that for $\underline{\mathfrak{S}}(\alpha, \beta)$ and $\overline{\mathfrak{S}}(\alpha, \beta)$, the frequency of 1s is simply α , the angle of rotation/slope. Since the frequency of a factor w , more generally, is $|I_w|$, note that the factors from the language of $\underline{\mathfrak{S}}(\alpha, \beta)$ and $\overline{\mathfrak{S}}(\alpha, \beta)$ are characterized solely by α and are independent from β . Thus for the words $\underline{\mathfrak{S}}(\alpha, \beta)$ and $\overline{\mathfrak{S}}(\alpha, \beta)$ we conclude that there is a bijection between α and the language of the word. This conclusion motivates steps (3) and (4) from the proof plan.

Observation 3.5. From Theorem 3.1 we conclude that the frequency of any factor w of a Sturmian word \mathbf{u} actually exists.

3.2.1 The slope of a Sturmian word

In this subsection we prove steps (1) and (2) from the proof plan for Theorem 3.1, namely, that the frequency of 1s α exists for every Sturmian word \mathbf{u} and that $\alpha \notin \mathbb{Q}$.

Actually, for Sturmian words, not only does the frequency exist, as a matter of fact all factors of \mathbf{u} give excellent approximations of the frequency of 1s (this is given in Corollary 3.1 below). This remarkable property is called the *balance* of a word.

Definition 3.6 (Balanced word). A binary word \mathbf{u} , be it finite or infinite, is said to be balanced if for every $n \geq 0$ and for pair of factors w and v of length n , we have that the number of symbols 1 present in both w and v differ by at most 1.

Notation 3.1. We denote by $|u|_1$ the number of ones in u . Thus, the latter part of the definition can be rewritten as $||w|_1 - |v|_1| \leq 1$. Analogously we write $|u|_0$ for the number of 0s.

We remark that for a balanced word $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$, all factors $w, v \in \mathcal{L}_n(\mathbf{u})$ satisfy $||w|_1/|w| - |v|_1/|v|| \leq 1/n$.

To prove that Sturmian words are balanced (i.e., Proposition 3.6), we require a technical lemma.

Lemma 3.1. *If a binary sequence \mathbf{u} is not balanced (be it finite or not), then it contains two subwords $0w0$ and $1w1$ for some $w \in \{0, 1\}^*$. Further, without loss of generality, we may suppose that the pair satisfies the property that there is no pair of factors $a, b \in \mathcal{L}_u(m)$ with $|a|_1 - |b|_1 \geq 2$ and $m < |w| + 2$.*

Proof. Let a and b satisfy $|a|_1 - |b|_1 \geq 2$ and have length as short as possible. It is clear then that a and b must differ in their borders (else we can eliminate the part that is equal and still get the same difference

$|a|_1 - |b|_1$). Clearly then $a = 1w1$ and $b = 0\tilde{w}0$, for some words w and \tilde{w} . Indeed, the ‘‘crossed’’ cases $a = 1w0$, $b = 0\tilde{w}1$ and $a = 0w1$, $b = 1\tilde{w}0$ are not possible (because $|a|_1 - |b|_1 = |w|_1 - |\tilde{w}|_1$ and the latter are shorter). Then, having $a = 0w0$, $b = 1\tilde{w}1$ implies $|w|_1 - |\tilde{w}|_1 = 2 + |a|_1 - |b|_1 \geq 2$. Now we claim $w = \tilde{w}$. Assume otherwise, then there is a first index for which they differ, that is $w = x\alpha y$, $\tilde{w} = x\tilde{\alpha}y'$ where $\alpha \in \{0, 1\}$, $x, y, y' \in \{0, 1\}^*$. Note then that α cannot be 1 because otherwise we would have $1x1$ and $0x0$ as subwords, and these satisfy $|1x1|_1 - |0x0|_1 = 2$ while being shorter. Hence $\alpha = 0$ and $w = x0y$, $\tilde{w} = x1y'$. But then we have $|a|_1 - |b|_1 = |y|_1 - |y'|_1$ which again leads to an absurd. ■

Proposition 3.6. *A binary word $u \in \{0, 1\}^{\mathbb{N}}$ is Sturmian if and only if it is balanced and non-eventually periodic.*

Proof. Consider a Sturmian word $u \in \{0, 1\}^{\mathbb{N}}$. We already know from Proposition 3.2 that u is not eventually periodic. We now prove it is balanced. Suppose otherwise, then from Lemma 3.1 it follows that there are factors $0w0$ and $1w1$ in u for some $w \in \{0, 1\}^*$, minimal in length. From Proposition 3.4 we know that either we have 00 as a factor or 11 , but not both. This means immediately that $w \neq \epsilon$, but it also tells us that if $w = w_0 \dots w_n$ for some $n \geq 0$, then $w_0 = w_n$. Furthermore, w is a palindrome, i.e., $w_k = w_{n-k}$ for all $k = 0, \dots, n$. Indeed, if it were not, let $k \geq 1$ be the smallest such that $w_k \neq w_{n-k}$, then $0w_0 \dots w_{k-1}0$ and $1w_{n-k+1} \dots w_n1$ form a shorter pair of unbalanced factors.

Now, since w can be extended to the right in two ways, i.e., $w = R_{n+1}$, this means that one of $0w, 1w$ is R_{n+2} . Without loss of generality we assume it to be $0w$. Thus we have that $0w0, 0w1, 1w1$ are factors of u , however note that $1w0$ is not.

Here we remark the following key property:

$$\text{if } 1w1 = u_i \dots u_{i+n+2}, \text{ then } 0w \text{ does not occur within } v := w_i \dots w_{i+2n+3}. \quad (1)$$

This is proved as follows: if it did occur within v , then the word $0w$ must start within the first $n+3$ letters of v (which correspond to a $1w1$), as follows from comparing lengths. Then, as w is a palindrome we must have $0w_0 \dots w_k = w_{n-k}w_{n-k-1} \dots w_01$ for some k , but this is absurd as then $w_k = 1$ and $w_{n-k} = 0$.

Now we may continue with the proof of the proposition, armed with our key property. Property (1) tells us that within $v = w_i \dots w_{i+2n+2}$ at most $n+2$ factors of length $n+2$ may occur. As we have words $w_i \dots w_{i+n+1}, w_{i+1} \dots w_{i+n+2}, \dots, w_{i+n+2} \dots w_{i+n+3}$, for a total of $n+3$, some factor of length $n+2$ must appear at least twice. But all of these factors of length $n+2$ are distinct from $0w = R_{n+2}$, so all of them have one right extension. Thus the factor appearing twice tells us that then this cycle repeats forever, a contradiction, since we know u is not ultimately periodic.

For the converse we must show that $p_u(n) = n+1$. Suppose otherwise, then let n be the smallest positive integer such that $p_u(n) \geq n+2$. Then $p_u(n-1) = n$ and so there are two distinct factors, a and b , of length $(n-1)$ that can be extended right-wards in two ways. Thus our word contains the words $a0, a1, b0, b1$ as factors. We may suppose that the words a and b differ only on their first symbol (else a smaller n would do). Then we have factors of the form $0w0$ and $1w1$, contradicting the balancedness of u . ■

With these results in hand, we can finally prove that the Fibonacci word f_∞ defined in Example 3.1 is a Sturmian word.

Example 3.2 (The Fibonacci word revisited). We show that the Fibonacci word f_∞ defined in Example 3.1 is not eventually periodic, and hence from Proposition 3.6 it will follow that it is a Sturmian word. We prove f_∞ is balanced. In order to prove it, it is more convenient to define the sequence f_k in terms of a morphism $\sigma: \{0, 1\}^* \rightarrow \{0, 1\}^*$ defined by $\sigma(0) = 01$ and $\sigma(1) = 0$. Being a morphism on the monoid $\{0, 1\}^*$ means that σ satisfies $\sigma(\epsilon) = \epsilon$ and $\sigma(a \cdot b) = \sigma(a) \cdot \sigma(b)$, so defining it on $\{0, 1\}$ is enough.

We remark then that $\sigma(f_k) = f_{k+1}$ indeed, by induction we see that $\sigma(f_0) = \sigma(0) = 01 = f_1$ and then if we have $\sigma(f_j) = f_{j+1}$ for all $j \leq k$, we have

$$\sigma(f_{k+1}) = \sigma(f_k \cdot f_{k-1}) = \sigma(f_k) \cdot \sigma(f_{k-1}) = f_{k+1} \cdot f_k = f_{k+2}.$$

Thus in fact, f_∞ is usually defined as the fixed point of the morphism σ [Fog02].

Suppose now that f_∞ were not balanced, then from Lemma 3.1 it follows that there would be factors of the form $0w0$ and $1w1$ within f_m for m large enough. Since f_∞ is clearly recurrent, we may further assume that the two subwords $0w0$ and $1w1$ we will be working with are not at the beginning of f_∞ . Clearly w is not empty as 11 cannot appear in f_∞ by Proposition 3.4. The word w can only start with a 0 and end with a 0, since there cannot be two consecutive 1s (else $1w1$ would not be a subword). Thus $w = 0w'0$ and so $00w'00$ and $10w'01$ are subwords of f_m .

Now, observe that the morphism σ is injective from $\{0, 1\}^*$ to $\{0, 1\}^*$, in fact

- A 1 within the sequence implies that the symbol immediately to its **left** is a 0, and that we may extract the preimage of this pair to be 0.
- A 00 indicates that the first 0 must come from $\sigma(1)$ as otherwise it would be followed by a 1.

In conclusion, we know that the letter to the left of $10w'01$ is a 0 and then f_m contains $010w'01$ which has $0\sigma^{-1}(0w')0$ as its preimage and this is a subword of $\sigma^{-1}(f_m) = f_{m-1}$. Similarly, we note that having $00w'00$ in f_m implies having $1\sigma^{-1}(0w')1$ in f_{m-1} . So both $0\sigma^{-1}(0w')0$ and $1\sigma^{-1}(0w')1$ appear as subwords of f_{m-1} and f_{m-1} is not balanced either. This idea follows inductively until we get m small enough for it to be obvious that f_m is balanced and we have reached an absurd. \diamond

Proposition 3.7. *Let $u \in \mathcal{A}^{\mathbb{N}}$ be a Sturmian word, then for any pair of factors $u \in \mathcal{L}_u(n)$ and $v \in \mathcal{L}_u(m)$ we have*

$$\left| \frac{|u|_1}{n} - \frac{|v|_1}{m} \right| \leq \frac{1}{n} + \frac{1}{m}. \quad (3.5)$$

Proof. This proposition follows from Proposition 3.6. Let g_k be the minimal number of 1s for the words of $\mathcal{L}_u(k)$, so that each word of $\mathcal{L}_u(k)$ has either g_k or $g_k + 1$ ones.

As any factor of length $n \times m$ is composed of n factors of length m , or equivalently m factors of length n , we derive the inequalities

$$mg_n \leq n(g_m + 1), \quad ng_m \leq m(g_n + 1).$$

This means that

$$\frac{g_n}{n} - \frac{g_m}{m} \leq \frac{1}{m}, \quad \frac{g_m}{m} - \frac{g_n}{n} \leq \frac{1}{n},$$

and these imply

$$\frac{|u|_1}{n} - \frac{|v|_1}{m} \leq \frac{g_n + 1}{n} - \frac{g_m}{m} \leq \frac{1}{n} + \frac{1}{m}, \quad \frac{|v|_1}{m} - \frac{|u|_1}{n} \leq \frac{g_m + 1}{m} - \frac{g_n}{n} \leq \frac{1}{m} + \frac{1}{n},$$

thus proving the result. \blacksquare

Corollary 3.1. *Let $u \in \{0, 1\}^{\mathbb{N}}$ be a Sturmian word, then the frequency $\alpha := f_1$ of 1s exists and satisfies*

$$\left| \frac{|w|_1}{|w|} - f_1 \right| \leq \frac{1}{|w|}, \quad (3.6)$$

for any $w \in \mathcal{L}(u)$.

Proof. Proposition 3.7 implies that the sequence $(g_k/k)_{k=1}^{\infty}$, where g_k is the least possible number of ones in a factor of length k in u , is a Cauchy sequence and hence converges to a real number $\alpha \in [0, 1]$.

This number α is equal to f_1 , frequency of 1s, as Sturmian sequences are balanced. Finally (3.6) follows from Proposition 3.7 by choosing V to be a word in $\mathcal{L}_m(u)$ having the least possible number of 1s and then taking the limit $m \rightarrow \infty$. \blacksquare

Following our proof plan described at the beginning of Section 3.2, the frequency $\alpha = f_1$ will end up being the rotation angle in Theorem 3.1. As such, the rotation angle α must be irrational (step (2)).

Proposition 3.8. *Let $\mathbf{u} \in \{0, 1\}^\infty$ be a Sturmian word, then the frequency $\alpha := f_1$ of 1s is irrational.*

Proof. Assume for the sake of contradiction that $\alpha = p/q$ with $\gcd(p, q) = 1$. Let g_k be the minimal number of 1s for the words of $\mathcal{L}_k(\mathbf{u})$. Then we have (just think of \mathbf{u} as a series of blocks of length k)

$$\frac{g_k}{k} \leq \alpha \leq \frac{g_k + 1}{k}.$$

Pick $k = q$, then

$$\frac{g_q}{q} \leq \frac{p}{q} \leq \frac{g_q + 1}{q},$$

thus either $p = g_q$ or $p = g_q + 1$.

Suppose first that $p = g_q$ and consider $k = 2^r q$ for $r \geq 0$. Then $2^r g_q \leq g_{2^r q}$ and therefore $\frac{g_{2^r q}}{2^r q}$ is increasing in r . As it converges to $\alpha = g_q/q$ it we get that $\frac{g_{2^r q}}{2^r q} = g_q/q$ for all $r \geq 0$. We show this is absurd.

As the word \mathbf{u} is not periodic, there is a factor w of length q such that $|w|_1 = g_q + 1$. From Proposition 3.3 we know that w occurs an infinite number of times in \mathbf{u} , hence it must occur at two indices congruent modulo q . Thus for r large enough we can find a factor v of length $2^r q$ for which w occurs at least twice, and hence $|v|_1 \geq 2^r a_q + 2$. This means that $a_{2^r q} \geq 2^r a_q + 1$, thus producing the contradiction.

The proof for the case $p = g_q + 1$ is analogous. ■

3.2.2 The language of a Sturmian word

In this subsection we prove steps (3) and (4) from the proof plan for Theorem 3.1 given at the beginning of Section 3.2, namely, that the language of a Sturmian word \mathbf{u} is determined by its slope α and that we can approximate \mathbf{u} by arbitrarily long factors of $\underline{\mathfrak{S}}(\alpha, 0)$ and $\overline{\mathfrak{S}}(\alpha, 0)$ in a precise sense.

The topology of $\mathcal{A}^\mathbb{N}$. We turn the space of infinite words $\mathcal{A}^\mathbb{N}$ over the alphabet \mathcal{A} into a metric space by introducing the distance

$$d(u, v) = 2^{-m}, \quad m := \min\{k \geq 1 : u_k \neq v_k\},$$

and $d(u, v) = 0$ when $u = v$, which coincides with the intuition “ $\min \emptyset = \infty$ ” and “ $2^{-\infty} = 0$ ”.

Thus a sequence of infinite words $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots$ in $\mathcal{A}^\mathbb{N}$ converges to a word \mathbf{v} if and only if, for each index $k \geq 1$ there is $N_k \geq 0$ for which $u_1^{(n)} \dots u_k^{(n)} = v_1 \dots v_k$ for all $n \geq N_k$. Informally, if the prefixes “stabilize”, becoming constant from some point on.

Observation 3.6. Of course, the topology we have just introduced for $\mathcal{A}^\mathbb{N}$ is non other than the product topology: consider the discrete topology over the finite alphabet \mathcal{A} .

Since the discrete topology makes \mathcal{A} compact, the product $\mathcal{A}^\mathbb{N}$ is compact by Tychonoff’s Theorem (see for instance [Mun00]).

Definition 3.7 (Convergence on $\mathcal{A}^\infty := \mathcal{A}^* \cup \mathcal{A}^\mathbb{N}$). To extend $\mathcal{A}^\mathbb{N}$ to $\mathcal{A}^\infty := \mathcal{A}^* \cup \mathcal{A}^\mathbb{N}$ we may consider that \mathcal{A}^∞ is actually a closed topological subspace of $(\mathcal{A} \cup \{\epsilon\})^\mathbb{N}$, where ϵ denotes the empty word. Thus we may speak of the convergence of finite words in \mathcal{A}^* to infinite words in $\mathcal{A}^\mathbb{N}$.

Example 3.3. This is the case of Example 3.1, where the finite words \mathbf{f}_k converge to \mathbf{f}_∞ . ◇

We may also speak of continuous functions on the space $(\mathcal{A}^\mathbb{N}, d)$. The simple example of such a function being the shift map $S: \mathcal{A} \rightarrow \mathcal{A}$ defined by $S(x_1 x_2 x_3 \dots) = x_2 x_3 \dots$. The map S is clearly continuous as it satisfies the inequality $d(S\mathbf{u}, S\mathbf{v}) \leq 2d(\mathbf{u}, \mathbf{v})$.

We already know that $(\mathcal{A}^\mathbb{N}, d)$ is compact due to Tychonoff’s theorem, however we present a classical proof with a diagonal argument for the sake of completeness.

Proposition 3.9. *The metric space $(\mathcal{A}^{\mathbb{N}}, d)$ is compact.*

Proof. Given a sequence of words $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots$, we will show that there is a convergent subsequence.

Since the alphabet \mathcal{A} is finite, there must be some letter $a_1 \in \mathcal{A}$ that appears infinitely many times as the first letter $\mathbf{u}_0^{(k)} = a$. Let $\mathbf{u}^{(n_1^{(0)})}, \mathbf{u}^{(n_2^{(0)})}, \dots$ be a subsequence such that $\mathbf{u}_0^{(n_k^{(0)})} = a_0$ for all $k \geq 1$.

Repeating the argument we get an increasing subsequence $(n_k^{(1)})_{k=1}^{\infty}$ of $(n_k^{(0)})_{k=1}^{\infty}$ such that the second entry $\mathbf{u}_1^{(n_k^{(1)})}$ is always equal to some symbol $a_1 \in \mathcal{A}$, and so on producing $(n_k^{(2)})_{k=1}^{\infty}, (n_k^{(3)})_{k=1}^{\infty}, \dots$

Now comes the diagonal argument. Consider $m_k := n_k^{(k)}$, then $\mathbf{u}_0^{(m_k)} \dots \mathbf{u}_k^{(m_k)} = a_0 \dots a_k$ for all $k \geq 1$ and we see that $\mathbf{u}^{(n_k^{(k)})}$ converges to $\mathbf{a} = a_0 a_1 \dots$ ■

The slope and the factors of Sturmian words. We denote by $S: \mathcal{A}^{\mathbb{N}} \rightarrow \mathcal{A}^{\mathbb{N}}$ the shift $S(x_1 x_2 x_3 \dots) = x_2 x_3 \dots$. The study of the factors of S comes down to the study of its orbit $\mathcal{O}(\mathbf{u}) := \{\mathbf{u}, S\mathbf{u}, S^2\mathbf{u}, \dots\}$, from a topological perspective. More precisely, we are interested in $X_{\mathbf{u}} := \overline{\mathcal{O}(\mathbf{u})}$, its topological closure.

The set $X_{\mathbf{u}}$ is compact and closed under S (by the continuity). We shall see that the set $X_{\mathbf{u}}$, the subshift of \mathbf{u} , is actually characterized by the fact that every element has the same given frequency of 1s as \mathbf{u} .

Proposition 3.10. *Let $\mathbf{u} \in \{0, 1\}^{\mathbb{N}}$ be a Sturmian word. Then for every $\mathbf{v} \in X_{\mathbf{u}} = \overline{\mathcal{O}(\mathbf{u})}$ we have $\mathcal{L}_{\mathbf{v}} = \mathcal{L}_{\mathbf{u}}$ and the frequency of 1s in \mathbf{v} and \mathbf{u} are equal.*

In order to prove Proposition 3.10, we introduce a new concept: the *uniform recurrency* of a word. This property actually amounts to each elements from $X_{\mathbf{u}}$ having the exact same language as \mathbf{u} . Uniformly recurrent words have, even though they may be aperiodic, a sort of recurrence that we will afterwards measure with the so called *recurrence function*, a central object to this dissertation.

Definition 3.8 (Uniformly recurrent word). A word $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ is said to be uniformly recurrent if and only if each factor w in \mathbf{u} appears infinitely often and with bounded gaps.

That is, given a factor w in \mathbf{u} , there exists a constant $C := C_{\mathbf{u}}(w) > 0$ such that if $u_{i+1} \dots u_{i+n} = w$, there is j with $0 < j - i \leq C$ such that $u_{j+1} \dots u_{j+n} = w$.

Sturmian words are all uniformly recurrent. Thinking from $\underline{\mathfrak{S}}(\alpha, \beta)$ and $\overline{\mathfrak{S}}(\alpha, \beta)$, this happens because $k\alpha \bmod 1$ is dense on $[0, 1]$ and factors correspond to intervals (recall Proposition 3.5).

Proposition 3.11. *Sturmian words are uniformly recurrent.*

Proof. Let \mathbf{u} be a Sturmian word and suppose otherwise. We already know from Proposition 3.3 that \mathbf{u} is recurrent, so for it to be not “uniformly recurrent” we must assume we had a factor w of \mathbf{u} that appears infinitely many times but not with bounded gaps, i.e., for each $C > 0$ there are indices i and j such that $j - i > C$ and $u_{i+1} u_{i+2} \dots u_{i+n} = u_{j+1} u_{j+2} \dots u_{j+n} = w$ and w is not a factor of the finite word $u_{i+2} \dots u_{j+n-1}$.

Thus we may construct an infinite sequence of factors, increasing in length, that do not contain w . This sequence has a fixed point $\mathbf{v} \in \{0, 1\}^{\mathbb{N}}$. Since this new word \mathbf{v} is a limit of factors of \mathbf{u} not containing w , we see that it has at most $|w|$ factors of length $|w|$ (because \mathbf{u} is Sturmian). Thus \mathbf{v} is eventually periodic and its frequency of 1s is rational, but its frequency of 1s should coincide with that of \mathbf{u} by Proposition 3.10, which is irrational because of Proposition 3.8. ■

Corollary 3.2. *Let \mathbf{u} be a Sturmian word. Then $\mathcal{L}_{\mathbf{u}}(n) = \mathcal{L}_{\mathbf{v}}(n)$ for all $\mathbf{v} \in \overline{\mathcal{O}(\mathbf{u})}$.*

Then the following Theorem 3.2 is the final result of this subsection, which states that the slope α characterizes the language of a Sturmian word and summarizes the fact that speaking about language and subshifts is the same thing. This is key to the proof of Theorem 3.1 as this means that a Sturmian word \mathbf{u} with slope α has the same language as $\underline{\mathfrak{S}}(\alpha, 0)$ and $\overline{\mathfrak{S}}(\alpha, 0)$ (clearly of slope α) and hence \mathbf{u} belongs to their subshift; it

can be approximated from arbitrarily long factors of these.

Theorem 3.2. *Let \mathbf{u} and \mathbf{v} be two Sturmian words with the same frequency of 1s. Then the language of factors of both words coincide $\mathcal{L}_{\mathbf{u}} = \mathcal{L}_{\mathbf{v}}$, equivalently $\overline{\mathcal{O}(\mathbf{u})} = \overline{\mathcal{O}(\mathbf{v})}$. Conversely, if $\mathcal{L}_{\mathbf{u}} = \mathcal{L}_{\mathbf{v}}$, then \mathbf{u} and \mathbf{v} have the same slope (frequency of ones).*

Observation 3.7. As a consequence, for Sturmian words we may speak of \mathcal{L}_{α} , where α is the slope.

In order to prove this, we need a technical lemma regarding balanced sets of words. This is related to the characterization of a Sturmian word as an aperiodic balanced word $\mathbf{u} \in \{0, 1\}^{\mathbb{N}}$. In fact, from this technical lemma one can easily derive that a balanced and aperiodic word is a Sturmian word.

Definition 3.9. A set of finite words \mathcal{L} is said to be balanced if and only if for every pair of words $w, v \in \mathcal{L}$ and factors w' and v' , of w and v respectively, of the same length we have $||w'|_1 - |v'|_1| \leq 1$.

Lemma 3.2 (Ex. 6.1.12 [Fog02]). *If a set of finite words \mathcal{L} is balanced, then the number of words of length n is at most $n + 1$.*

Proof. We may assume without loss of generality that our sets \mathcal{L} satisfy that if $u \in \mathcal{L}$ and v is a factor of u , then $v \in \mathcal{L}$, as the addition of v still makes \mathcal{L} balanced. We proceed by induction on n . The result is clear for $n = 0$ and $n = 1$. Now assume the result to be true for all $n < m$, let us prove it for m .

Assume we had $m + 2$ words of length m . Then, as we had only m words of length $m - 1$, two of these u and v must have two right-extensions $u0, u1$ and $v0, v1$ on \mathcal{L} . Write $u = u_1 \dots u_{m-1}$ and $v = v_1 \dots v_{m-1}$ and let $k = \max\{1 \leq i < m : v_i \neq u_i\}$. Of course $k \leq m - 1$ is well-defined as $u \neq v$. Then $U = u_k u_{k+1} \dots u_{m-1} u_k$ and $V = v_k v_{k+1} \dots v_{m-1} v_k$ are words in \mathcal{L} and differ only in the first and last symbol. Thus we get $||U|_1 - |V|_1| = 2$, a contradiction ■

Proof. [Proof of Theorem 3.2] Consider two Sturmian words \mathbf{u} and \mathbf{v} with the same frequency α of 1s. Let $g_k(\mathbf{u})$ and $g_k(\mathbf{v})$ denote the minimal number of ones in a factor of length k in \mathbf{u} and \mathbf{v} respectively, so that any factor of length k in \mathbf{u} has either $g_k(\mathbf{u})$ or $g_k(\mathbf{u}) + 1$ ones, and similarly for \mathbf{v} . Then we know that

$$\frac{g_k(\mathbf{u})}{k} \leq \alpha \leq \frac{g_k(\mathbf{u}) + 1}{k}, \quad \frac{g_k(\mathbf{v})}{k} \leq \alpha \leq \frac{g_k(\mathbf{v}) + 1}{k},$$

but this immediately implies that $g_k(\mathbf{u}) = g_k(\mathbf{v})$, because $k\alpha$ cannot be an integer. In turn, this means that the union $\mathcal{L}_k(\mathbf{u}) \cup \mathcal{L}_k(\mathbf{v})$ is balanced in the sense of Definition 3.9. Thus from Lemma 3.2 we deduce that $|\mathcal{L}_k(\mathbf{u}) \cup \mathcal{L}_k(\mathbf{v})| \leq k + 1$, which means that $\mathcal{L}_k(\mathbf{u}) = \mathcal{L}_k(\mathbf{v}) = \mathcal{L}_k(\mathbf{u}) \cup \mathcal{L}_k(\mathbf{v})$ by comparing the cardinalities following the definition of Sturmian words. ■

3.2.3 End of the proof of the characterization of Morse-Hedlund Theorem 3.1

Proof. Let \mathbf{u} be a Sturmian word of frequency $\alpha = f_1(\mathbf{u})$. We show that there is $\beta \in [0, 1)$ such that either $\mathbf{u} = \underline{\mathfrak{S}}(\alpha, \beta)$ or $\mathbf{u} = \overline{\mathfrak{S}}(\alpha, \beta)$.

Let us consider the centered $\mathbf{v} = \underline{\mathfrak{S}}(\alpha, 0)$, which as we have seen corresponds to the codings of circle rotations in Proposition 3.5. We know from Theorem 3.2 that $\mathbf{u} \in \overline{\mathcal{O}(\mathbf{v})}$ so that \mathbf{u} is seen as the limit of $S^{n_1}\mathbf{v}, S^{n_2}\mathbf{v}, \dots$ for some increasing sequence n_1, n_2, \dots . We note that $S^{n_k}\mathbf{v} = \underline{\mathfrak{S}}(\alpha, \{n_k\alpha\})$ and that the sequence $\{n_k\alpha\}$ on the unit interval mod 1 must have a convergent subsequence. Thus we may assume $\beta = \lim_{k \rightarrow \infty} n_k\alpha$ on \mathbb{T}^1 .

We prove that $\mathbf{u} = \underline{\mathfrak{S}}(\alpha, \beta)$ or $\mathbf{u} = \overline{\mathfrak{S}}(\alpha, \beta)$ by showing that $S^{n_k}\mathbf{v}$ tends to $\underline{\mathfrak{S}}(\alpha, \beta)$ or $\overline{\mathfrak{S}}(\alpha, \beta)$. As $\beta = \lim_{k \rightarrow \infty} n_k\alpha$ on \mathbb{T}^1 , we may assume that the sequence $n_k\alpha$ either increases to β (tends to it counter-clockwise in the sense of Figure 3.3) or it decreases to it (tends to it clockwise). Whether we have $\mathbf{u} = \underline{\mathfrak{S}}(\alpha, \beta)$ or $\mathbf{u} = \overline{\mathfrak{S}}(\alpha, \beta)$ will depend exactly on this property.

We note that $S^{n_k}\mathbf{v} = \underline{\mathfrak{S}}(\alpha, \{n_k\alpha\})$ and $\underline{\mathfrak{S}}(\alpha, \beta)$ differ within the first m symbols if and only if $\{n_k\alpha\}$ and β belong to two different intervals of \mathbb{T}^1 delimited by the points $0, -\alpha, \dots, -m\alpha$ modulo 1. As $\{n_k\alpha\}$

tends to β modulo 1, the only case in which this does not happen for large enough k is when β is one of the points $0, -\alpha, \dots, -m\alpha$ modulo 1.

It is important to picture what happens at the borders of these delimited intervals (as in Proposition 3.5 with $0 \equiv 1$ and $1 - \alpha$). In the case of $\underline{\mathfrak{S}}(\alpha, \beta)$ we note that the delimited intervals (corresponding to factors of length m) partitioning \mathbb{T}^1 . These intervals are $[a_0 = 0, a_1), [a_1, a_2), \dots, [a_m, a_{m+1} = 1)$ where the sequence a_i increases. Moreover, for $\overline{\mathfrak{S}}(\alpha, \beta)$ they read $(0, a_1], (a_1, a_2], \dots, (a_m, 1]$ for the same underlying sequence a_i . Suppose $\beta = a_j$, then if $\{n_k\alpha\}$ increases towards β we have that $\{n_k\alpha\} \in (a_{j-1}, \beta]$, setting $a_{-1} = a_m$, for large enough k and so $\underline{\mathfrak{S}}(\alpha, \{n_k\alpha\})$ coincides with $\underline{\mathfrak{S}}(\alpha, \beta)$ in the first m symbols for large enough k . Else if $\{n_k\alpha\}$ decreases towards β we have that $\{n_k\alpha\} \in (\beta, a_{j+1}]$, setting $a_{m+1} = a_1$, for large enough k and then $\underline{\mathfrak{S}}(\alpha, \{n_k\alpha\})$ coincides with $\overline{\mathfrak{S}}(\alpha, \beta)$ on the first m symbols for large enough k . ■

3.3 A second concept from Word Combinatorics: recurrence

We have already introduced the complexity function and several other key properties concerning Sturmian words, in particular that they are “mechanical sequences” (see Theorem 3.1). Along the way we defined the concept of uniform recurrence, as a convenient combinatorial formulation of the equivalence between having two words \mathbf{u} and \mathbf{v} have equal languages $\mathcal{L}_{\mathbf{u}} = \mathcal{L}_{\mathbf{v}}$ and having \mathbf{v} belonging to the closure $X_{\mathbf{u}}$ of the orbit $\mathcal{O}(\mathbf{u}) = \{\mathbf{u}, S\mathbf{u}, S^2\mathbf{u}, \dots\}$ ^b (equivalently $\mathbf{u} \in X_{\mathbf{v}}$). We recall that an infinite word $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ is *uniformly recurrent* if every factor of \mathbf{u} appears infinitely often and with bounded gaps (see Definition 3.8). We proved in Proposition 3.11 that Sturmian words are uniformly recurrent.

Here we will concentrate on the concept of recurrence for its own sake. We have already discussed the complexity (the number of finite factors) of infinite words. It is also important to study where finite factors occur inside the infinite word \mathbf{u} . For uniformly recurrent words, such as Sturmian words, all factors reappear with bounded gaps. We can then quantify the size of these gaps by using the so-called *recurrence function*.

In this section we introduce the recurrence function, discuss its basic properties and explain a fundamental result, Theorem 3.3, again by Morse and Hedlund [MH40]. This theorem relates the recurrence function of a Sturmian word with slope α to the continuants of α , and its proof depends strongly on Theorem 3.1.

3.3.1 Definitions and basic properties

In this subsection we introduce the recurrence function, and discuss briefly a proposition relating it to the complexity function.

We recall the definition of uniform recurrence.

Definition 3.8 A word $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ is said to be *uniformly recurrent* if and only if each factor w in \mathbf{u} appears infinitely often and with bounded gaps.

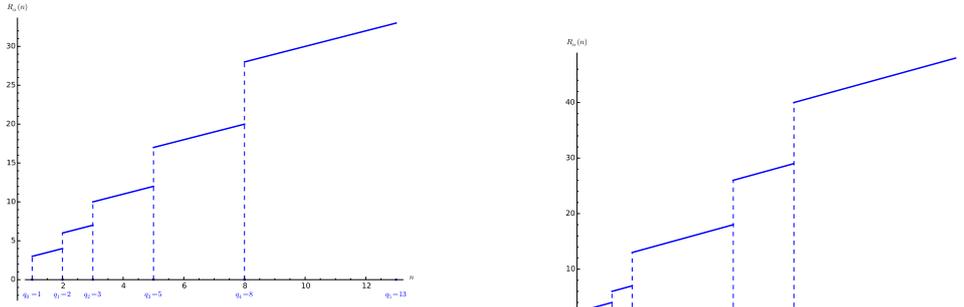
In other words, given a factor w in \mathbf{u} , there exists a constant $C := C_{\mathbf{u}}(w) > 0$ such that if $u_{i+1} \dots u_{i+n} = w$, there is j with $0 < j - i \leq C$ such that $u_{j+1} \dots u_{j+n} = w$.

Denote by $w_{\mathbf{u}}(q, n)$ the minimal number of symbols u_k with $k \geq q$ which have to be inspected to discover the whole set $\mathcal{L}_{\mathbf{u}}(n)$, starting from the index q . Then, the integer $w_{\mathbf{u}}(q, n)$ is a sort of “waiting time” and \mathbf{u} is uniformly recurrent if each set $\{w_{\mathbf{u}}(q, n) \mid q \in \mathbb{N}\}$ is bounded, and we may define the so-called *recurrence function* $n \mapsto R_{\mathbf{u}}(n)$.

Definition 3.10. Recurrence function Let $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ be a uniformly recurrent word, then the recurrence function is defined by

$$R_{\mathbf{u}}(n) := \max\{w_{\mathbf{u}}(q, n) \mid q \in \mathbb{N}\}, \tag{3.7}$$

^bRecall that $S: \mathcal{A}^{\mathbb{N}} \rightarrow \mathcal{A}^{\mathbb{N}}$, the shift map, is defined by $S(x_1x_2x_3\dots) = x_2x_3\dots$



(a) Recurrence function for the Fibonacci word in Example 3.1. Here $\alpha = 1/\phi^2$, with $\phi = \frac{\sqrt{5}+1}{2}$.

(b) Recurrence function for $\alpha = 1/e$.

Figure 3.4: Examples of recurrence function for two Sturmian words. In order to illustrate better the behavior of the recurrence function, we have interpolated linearly between consecutive non-jump points.

where $w_u(q, n) := \min\{m \in \mathbb{N} : \mathbf{u}_q \dots \mathbf{u}_{q+m-1} \text{ contains all of the factors } \mathcal{L}_u(n)\}$.

This is clearly equivalent to the following usual definition: any factor of length $R_u(n)$ from u contains all factors of length n of u , and the length $R_u(n)$ is the smallest integer which satisfies this property.

The complexity function and the recurrence function are related through the following bound

Proposition 3.12. *Let $u \in \mathcal{A}^{\mathbb{N}}$ be an infinite word on a finite alphabet \mathcal{A} , then the complexity function p_u and the recurrence function R_u of u satisfy*

$$R_u(n) \geq p_u(n) + n - 1.$$

Proof. If a factor w of length m in u contains every factor of length n , then for each of the $p_u(n) = |\mathcal{L}_u(n)|$ factors of length n , there is a corresponding index of w where they start. Then we must surely add $n - 1$ characters to complete the factor of length n with the largest starting index on w . ■

Any Sturmian word is uniformly recurrent. Its recurrence function only depends on the slope α (recall Theorem 3.2) and is thus denoted by $n \mapsto R(\alpha, n)$. Figure 3.4 gives the plots for two different slopes. Note that the recurrence is piece-wise affine, a fact that will be explained by Theorem 3.3.

3.3.2 The frequencies of the factors of a Sturmian word

The recurrence function is strongly related to the frequencies of the factors. Intuitively, a factor appearing more rarely should force us to have a larger $R_u(n)$.

We have already explained what happens with the frequencies of the factors of Sturmian words, as these correspond to circle intervals. We recall it here

Proposition 3.13. *Let $u \in \{0, 1\}^{\mathbb{N}}$ be a Sturmian word of slope α and $w = w_1 \dots w_n$ be a factor of u of length n . Then the frequency of w in u is given by $|I_w|$ where*

$$I_w := I_{w_1} \cap \mathcal{R}_\alpha^{-1}(I_{w_2}) \dots \cap \mathcal{R}_\alpha^{n-1}(I_{w_n}),$$

where $I_1 := [1 - \alpha, 1)$ and $I_0 := [0, 1 - \alpha)$, considered as circle intervals on \mathbb{T}^1 .

This would make it seem as if we had to compute the length $|I_w|$ for every single factor w . We underline that this is not the case: for any given length n , there are at most three possible frequencies for the factors of length n . This property is most easily explained from the form of the Rauzy graph (see Figure 3.1).

We recall that the Rauzy Graphs $\mathcal{G}_u(n)$, as shown in Figure 3.1, are strongly-connected (Sturmian words are recurrent). Its structure is composed of 3 oriented paths, forming two cycles, thanks to Proposition 3.4.

The structure of the graph tells us at once that there can be at most 3 possible frequencies for the words in $\mathcal{L}_u(n)$, one of them being the sum of the other two (the frequency of the path back from L_n to R_n).

Proposition 3.14 (Weak three-distance theorem). *Let $\alpha \in \mathcal{I} \setminus \mathbb{Q}$. Consider the circle \mathbb{T}^1 intervals I_0, \dots, I_n delimited by the set of points $0, \alpha, \dots, n\alpha \pmod{1}$ ordered geometrically. Then the lengths of the intervals $(I_k)_{k=0}^n$ attain at most 3 possible values, the largest being the sum of the other two.*

This is a weak version of the three-distance theorem ([S658], [Ber96]), which, in its full form also gives precise formulas for each of the three distances in terms of the continuants $(q_k(\alpha))_k$ of the slope α . We will not prove the complete version of the three-distance theorem, but it is worth highlighting that Lemma 3.4 could be seen as a particular consequence of it.

The smallest and largest of these 3 frequencies play a fundamental role. The smallest frequency, which corresponds to the smallest intervals I_w , are the words we would expect to appear less often, and similarly the largest frequency, which corresponds to the largest intervals I_w .

Definition 3.11 (Smallest distance). Let $\alpha \in [0, 1)$. On the circle $\mathbb{T}^1 = \mathbb{R}/\mathbb{Z}$ consider the intervals delimited by the set of points $0, \alpha, \dots, n\alpha \pmod{1}$. We denote by $\Gamma(\alpha, n)$ the smallest length of these intervals.

Observation 3.8. Note that the smallest distance $\Gamma(\alpha, n)$ corresponds to the smallest possible frequency of a factor from $\mathcal{L}_\alpha(n+1)$.

The smallest distance can be characterized in terms of the convergents of α .

Proposition 3.15. *Let $\alpha \in [0, 1)$ and $n \in \mathbb{N}$. Let k be the unique non-negative integer such that $q_k(\alpha) \leq n < q_{k+1}(\alpha)$. The smallest distance $\Gamma(\alpha, n)$ is then given by*

$$\Gamma(\alpha, n) = M_k(\alpha) := |\alpha q_k - p_k|. \quad (3.8)$$

Proof. We note that if the distance on \mathbb{T}^1 between αi and αj for $0 \leq i < j \leq n$ is $\Gamma(\alpha, n)$, then so is the distance between 0 and $(j-i)\alpha$ where $0 \leq (j-i) \leq n$. The proof then follows from Proposition 1.5. ■

We also give the notation for the largest distance, which will also be involved in the proof.

Definition 3.12 (Largest distance). Let $\alpha \in [0, 1)$. On $\mathbb{T}^1 = \mathbb{R}/\mathbb{Z}$ consider the intervals delimited by the points $0, \alpha, \dots, (n-1)\alpha$ modulo 1. We denote by $\Upsilon(\alpha, n)$ the maximum of the lengths of these intervals.

We remark that the remaining distance equals $\Upsilon(\alpha, n) - \Gamma(\alpha, n)$ due to 3.14.

3.3.3 The Morse-Hedlund formula for recurrence of Sturmian words

Now that we have described the recurrence function and the smallest distance, we come to a key theorem of this section, relating the recurrence function $R_\alpha(n)$ to the continuants of α .

Theorem 3.3 (Morse, Hedlund [MH40]). *Fix an irrational slope $\alpha > 0$, then the recurrence function can be computed in terms of the continuants of α by the formula*

$$R_\alpha(n) = n - 1 + q_{k-1}(\alpha) + q_k(\alpha), \quad (3.9)$$

where the index $k = k(\alpha, n)$ is the only positive integer such that $q_{k-1}(\alpha) \leq n < q_k(\alpha)$.

Observe that Theorem 3.3 implies that the recurrence function is piece-wise affine (when we extend to non-integer points), this is illustrated in Figure 3.4.

Here we rewrite the Morse and Hedlund's proof in our notation. In order to do this, we will first relate the recurrence function to the smallest distance $\Gamma(\alpha, n)$.

Definition 3.13 (First recurrence time). We define the first recurrence time $\xi[\epsilon, \alpha]$ to be the smallest positive integer m such that the maximum length of the circle \mathbb{T}^1 intervals delimited by $0, \alpha, 2\alpha, \dots, (m-1)\alpha$, on the unit circle, is at most $\epsilon > 0$.

It is important to note that in Definition 3.13 we have m points and not $m + 1$ as in the definition of $\Gamma(\alpha, m)$.

Observation 3.9. By symmetry we have $\xi[\epsilon, \alpha] = \xi[\epsilon, -\alpha]$ for any $\alpha \in \mathbb{R}$.

Lemma 3.3. *The recurrence function $R_\alpha(n)$ and the smallest distance $\Gamma(\alpha, n)$ are related through the first recurrence time ξ as follows*

$$R_\alpha(n) = \xi[\Gamma(\alpha, n), \alpha] + n - 1, \quad (3.10)$$

for all n .

Proof. We prove first the inequality $R_\alpha(n) \leq \xi[\Gamma(\alpha, n), \alpha] + n - 1$. Let $m = \xi[\Gamma(\alpha, n), \alpha] + n - 1$, we intend to show that every factor in $\mathbf{u} = \mathbf{u}_{\alpha, \beta} := \underline{\mathfrak{S}}(\alpha, \beta)$ of length m contains all factors of length n . Pick an arbitrary factor of length m within \mathbf{u} , by changing β we may assume that the factor is $u_0 \dots u_{m-1}$.

Thus, to show that $u_0 \dots, u_{m-1}$ contains every factor of length n we have to show that the factors are among

$$u_0 \dots u_{n-1}, \quad u_1 \dots u_n, \quad \dots \quad u_{m-n} \dots u_{m-1}.$$

Factors of length n correspond to the intervals on the circle \mathbb{T}^1 delimited by

$$0, -\alpha, \dots, -n\alpha,$$

modulo 1, and we must show that among the points $\beta, \beta + \alpha, \dots, \beta + (m - n)\alpha$ modulo 1, there is at least one in each of the $n + 1$ circle intervals. Observe that if this were not the case, then the maximum distance between two geometrically consecutive points $\beta, \beta + \alpha, \dots, \beta + (m - n)\alpha$ modulo 1 is greater than $\Gamma(\alpha, n)$, the minimal distance between points $0, -\alpha, \dots, -n\alpha \pmod{1}$, but this is impossible since $m - n + 1 = \xi[\Gamma(\alpha, n), \alpha]$ precisely.

Next, we show the converse inequality $R_\alpha(n) \geq \xi[\Gamma(\alpha, n), \alpha] + n - 1$. Let $m - n + 1 = \xi[\Gamma(\alpha, n), \alpha] - 1$, we will show that there is a factor of length m not containing every factor of length n .

From the definition Definition 3.13 of ξ and Observation 3.9 it follows that, among the intervals delimited by the points $\beta, \beta + \alpha, \dots, \beta + (m - n)\alpha$ on \mathbb{T}^1 , there is an interval I of length greater than $\Gamma(\alpha, n)$. By adjusting β (this does not change the language) we may assume there is one of the intervals delimited by $0, -\alpha, \dots, -n\alpha$, call it J , fully contained within this interval I , as the lengths of these intervals is strictly smaller than that of I . But this means that the word corresponding to J does not occur in $u_0 \dots u_{m-1}$, because it cannot start at any index $0, 1, \dots, m - n$. ■

Next we study the behavior of the largest distance. This is key due to the fact that $\xi[\Gamma(\alpha, n), \alpha]$ is precisely the smallest “time” to have largest distance smaller than $\Gamma(\alpha, n)$.

Lemma 3.4. *Let $\alpha \in \mathcal{I}$ be an irrational and let $k > 1$, then the largest distance Υ defined in Definition 3.12 satisfies*

$$\Upsilon(\alpha, q_{k-1}(\alpha) + q_k(\alpha)) \leq M_{k-1}, \quad (3.11)$$

$$\Upsilon(\alpha, q_{k-1}(\alpha) + q_k(\alpha) - 1) > M_{k-1}, \quad (3.12)$$

where $M_k(\alpha) := |\alpha q_k(\alpha) - p_k(\alpha)|$.

Proof. We start from (3.11). We must show that the maximum length of the non-overlapping intervals of \mathbb{T}^1 delimited by the set of points $0, \alpha, 2\alpha, \dots, (q_{k-1} + q_k - 1)\alpha$ modulo 1, is at most M_{k-1} . To do this we show that for each $\alpha i \pmod{1}$ in our set there is another point $\alpha j \pmod{1}$, from our set, at \mathbb{T}^1 -distance less than M_{k-1} always rotating in the same direction. This indeed implies the first part.

We recall (1.20), i.e., $M_k = (-1)^k(\alpha q_k - p_k)$, note in particular the change of sign with the parity of k , as this will be key to the proof of the first inequality (3.11).

◦ Consider the case $i < q_k$, then the point $\alpha(i + q_{k-1}) \bmod 1$ is still in our set, and observe that $\alpha q_{k-1} \equiv (-1)^{k-1} M_{k-1} \pmod{1}$, hence $\alpha(i + q_{k-1}) \equiv \alpha i + (-1)^{k-1} M_{k-1} \pmod{1}$. Hence the “oriented” distance in \mathbb{T}^1 from αi to $\alpha(i + q_{k-1})$ is M_{k-1} , rotating positively (increasing on $[0, 1)$) for k odd, and negatively (decreasing) otherwise.

◦ Next consider the case $i \geq q_k$, then we write $i = q_k + j$ and note that $\alpha i \equiv \alpha j + (-1)^k M_k \pmod{1}$. Thus $\alpha j \equiv \alpha i + (-1)^{k-1} M_k \pmod{1}$ means that the “oriented” distance in \mathbb{T}^1 from αi to αj is M_k , rotating positively for k odd, and negatively otherwise. In any case, since $M_k < M_{k-1}$, this implies (3.11).

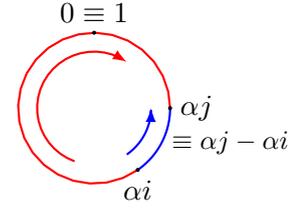


Figure 3.5: The oriented distance is given by the difference modulo 1.

We now move on to proving (3.12). This is the most crucial part of the lemma. In this case we consider the intervals in \mathbb{T}^1 delimited by the points $0, \alpha, \dots, (q_k + q_{k-1} - 2)\alpha$, seen modulo 1. For (3.11) it was enough to exhibit points on the circle that were close together (without necessarily being endpoints of a same interval), for (3.12) however, we must show something slightly stronger: there is a *delimited* interval having at least the given length.

We will show that there is a delimited interval having $\alpha(q_k - 1) \bmod 1$ as an endpoint and length greater than M_{k-1} . This is somewhat to be expected by looking at the preceding proof of (3.11). It is clear that the distance from $\alpha(q_k - 1) \bmod 1$ to each of the points $0, \alpha, \dots, \alpha(q_k - 2) \bmod 1$ is at least $\Gamma(\alpha, q_k - 1) = M_{k-1}$. It is also clear that $\alpha(q_k - 1) \bmod 1$ cannot be at distance M_{k-1} from two distinct points from $0, \alpha, \dots, \alpha(q_k - 2) \bmod 1$ simultaneously as then α would have to be rational^c. Thus we conclude that a \mathbb{T}^1 interval I of length greater than M_{k-1} , having $\alpha(q_k - 1) \bmod 1$ as a border, and containing no point from $0, \alpha, \dots, \alpha(q_k - 2) \bmod 1$. Now note that the smallest circle distance from $\alpha(q_k - 1) \bmod 1$ to $\alpha q_k, \dots, \alpha(q_k + q_{k-1} - 2) \bmod 1$ is greater than M_{k-1} , indeed $\Gamma(\alpha, q_{k-1} - 1) = M_{k-2} > M_{k-1}$.

Thus I contains a subinterval $I^* \subset I$, maybe even itself, having $\alpha(q_k - 1) \bmod 1$ as an endpoint, with length greater than M_{k-1} and no other point from $0, \alpha, \dots, (q_k + q_{k-1} - 2)\alpha \bmod 1$ there. ■

Proof. [Theorem 3.3] From Lemma 3.3 we know that $R_\alpha(n) = \xi[\Gamma(\alpha, n), \alpha] + n - 1$ where we recall that Γ is the smallest distance defined in Definition 3.11. Recall also that $\xi[\Gamma(\alpha, n), \alpha]$ is the waiting time $m \in \mathbb{N}$ to have $\Upsilon(\alpha, m) \leq \Gamma(\alpha, n)$. Since, by Proposition 3.15, $\Gamma(\alpha, n) = M_{k-1}(\alpha)$ for n satisfying $q_{k-1}(\alpha) \leq n < q_k(\alpha)$, we note that Lemma 3.4 implies that $m = q_{k-1}(\alpha) + q_k(\alpha)$, thus proving the result. ■

3.4 The growth of the recurrence function of Sturmian words

The recurrence function $R(\alpha, n) := R_\alpha(n)$ of Sturmian words has been widely studied (e.g., [Cas99, MH40]) on what we call the *worst case*. In particular, in the pioneering work by Morse and Hedlund [MH40]. They studied it first by turning the problem into one about continuants by Theorem 3.3, and then applying Theorem 1.4 and Theorem 1.5 to get information about this worst-case almost everywhere.

Cassaigne [Cas99] introduced a similar notion of recurrence quotient $\rho_u := \limsup_n R(u, n)/n$ (note the lim sup) and studied the spectrum of values attained by it, in particular its topological properties and when it is smallest. In his survey, Cassaigne [Cas01] proves ([Cas01, Theorem 3]) that when u is any not eventually periodic word, then $\rho_u \geq 3$. We will see, however, that from a measure-theoretic point of view, we have $\rho_u = \infty$ for almost every Sturmian word u .

We begin by explaining the results from [MH40] concerning the extreme values of the recurrence function in subsection 3.4.1. These results motivate our work, but unlike them, we study, for the first time, the recurrence quotient $S(\alpha, n) := (R(\alpha, n) + 1)/n$ in a probabilistic setting. We introduce the key ideas in

^cElse there are $i \neq j$ with $0 \leq i, j < q_k - 1$ and $\alpha(q_k - 1 - i) \equiv \alpha(j - q_k + 1) \equiv M_{k-1} \pmod{1}$ which implies $\alpha \in \mathbb{Q}$.

subsection 3.4.3 and develop these extensively in Chapter 4 and Chapter 5.

3.4.1 Classical results: worst-case analysis

Morse and Hedlund [MH40] first studied the extreme values attained by the recurrence function, turning the problem into one about continuants by Theorem 3.3, and then applying Theorem 1.4 and Theorem 1.5. Here is the corresponding result.

Theorem 3.4 (Morse-Hedlund [MH40]). *Let $h(x)$ be a function that is positive and non-decreasing for $x \geq 0$ and such that $\lim_{x \rightarrow \infty} h(x) = \infty$. For almost every $\alpha \in \mathcal{I}$*

$$\limsup_{n \rightarrow \infty} \frac{R(\alpha, n)}{n h(\log n)}$$

is finite or infinite according as to whether the series $\sum_{n \geq 0} 1/h(n)$ is convergent or divergent respectively.

Proof. Note that for n such that $q_{k-1}(\alpha) \leq n < q_k(\alpha)$ we have, by Theorem 3.3, that

$$\frac{R(\alpha, n)}{n} = 1 - \frac{1}{n} + \frac{q_{k-1}(\alpha) + q_k(\alpha)}{n}.$$

Thus the maximum for n subject to $q_{k-1}(\alpha) \leq n < q_k(\alpha)$ is

$$\frac{R(\alpha, q_{k-1}(\alpha))}{q_{k-1}(\alpha)} = 1 - \frac{1}{q_{k-1}(\alpha)} + \frac{q_{k-1}(\alpha) + q_k(\alpha)}{q_{k-1}(\alpha)} = 2 - \frac{1}{q_{k-1}(\alpha)} + m_k(\alpha) + \frac{q_{k-2}(\alpha)}{q_{k-1}(\alpha)},$$

and for $R(\alpha, n)/(nh(\log n))$ we deduce that the maximum for n in $[q_{k-1}(\alpha), q_k(\alpha))$ is

$$\frac{R(\alpha, q_{k-1})}{q_{k-1} h(\log q_{k-1})} = \frac{m_k(\alpha)}{h(\log q_{k-1}(\alpha))} + o(1).$$

It will therefore be enough to analyze what happens with the lim sup of $\frac{m_k(\alpha)}{h(\log q_{k-1}(\alpha))}$ as $k \rightarrow \infty$.

We observe that from Theorem 1.4 and Theorem 1.5 we have that $\limsup \frac{m_k(\alpha)}{h(\log q_{k-1})}$ is finite or infinite according to the convergence or divergence of the sum $\sum 1/h(\log q_{k-1})$.

For almost every α (recall Proposition 1.10) $(1/k) \log q_k(\alpha)$ tends to $K = \pi^2/(12 \log 2)$ which lies between 1 and 2. Thus, for large enough k (depending on α), the value $\log q_{k-1}(\alpha)$ lies between k and $2k$, hence we have $h(k) \leq h(\log q_{k-1}(\alpha)) \leq h(2k)$ for large enough k .

Note that the monotonicity of h implies that $\sum 1/h(n)$ converges if and only if $\sum 1/h(2n)$ converges, thus the convergence of $\sum 1/h(\log q_{k-1})$ is equivalent to the convergence of $\sum 1/h(k)$, this completes the proof. ■

As a consequence we have the following explicit limits:

Proposition 3.16. *Let $\epsilon > 0$. For almost every α the recurrence function satisfies:*

$$\limsup_{n \rightarrow \infty} \frac{R(\alpha, n)}{n \log n} = \infty, \quad \limsup_{n \rightarrow \infty} \frac{R(\alpha, n)}{n (\log n)^{1+\epsilon}} = 0.$$

Informally, we say that the worst-case of $R(\alpha, n)/n$ is roughly of order $\log n$. It is worthwhile to point out that the “best case” is always linear.

Proposition 3.17. *For almost every α the recurrence function satisfies:*

$$\liminf_{n \rightarrow \infty} \frac{R(\alpha, n)}{n} = 2.$$

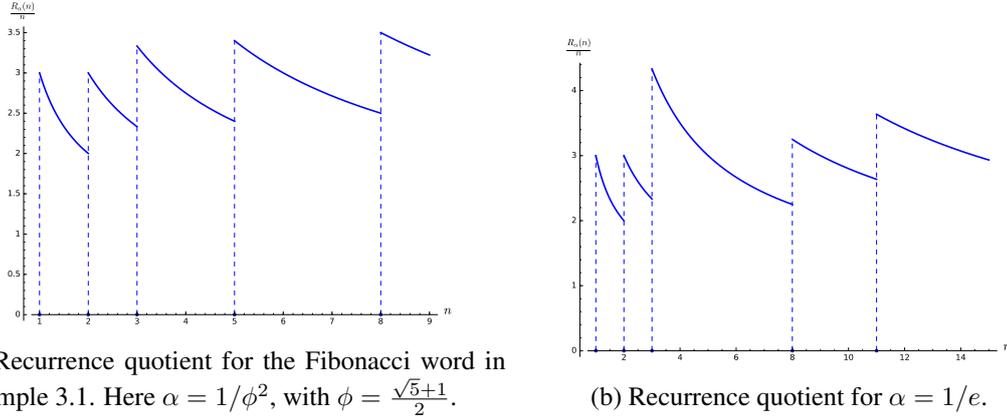


Figure 3.6: Examples of recurrence quotients for two Sturmian words.



Figure 3.7: Here we illustrate the case $q_{k-1}(\alpha) \leq n < q_k(\alpha)$, with a relative position $\mu(\alpha, n) \approx 1/3$.

Proof. Note in the proof of Theorem 3.4 that the best case arrives when $n = q_k(\alpha) - 1$ and that $q_{k-1}(\alpha)/q_k(\alpha)$ is close to 0 infinitely often (i.e., the quotients $m_k(\alpha)$ are unbounded almost surely). ■

It is natural therefore to work with what we call the recurrence quotient

$$S(\alpha, n) := \frac{R(\alpha, n) + 1}{n}. \tag{3.13}$$

The recurrence quotient is illustrated in Figure 3.6. Observe in particular that, as explained during the proof of Theorem 3.4, the recurrence quotient is largest on the left-end of the intervals $[q_{k-1}, q_k)$, while it is smallest on the right-end.

The results above, Theorem 3.4 and Proposition 3.17, tell us that the “best” and “worst” cases for $R(\alpha, n)$ differ widely. This is why the question “what does a *random* Sturmian word look like?” is so relevant. Does the recurrence function of a random Sturmian word behave more like $n \log n$, or is it more linear?

3.4.2 Position parameters

From the proof of Proposition 3.16 and Proposition 3.17, as well as Figure 3.6, it should be evident that the positioning of n within the interval $[q_{k-1}(\alpha), q_k(\alpha))$ plays a huge role when it comes to the magnitude of the recurrence quotient $S(\alpha, n)$ defined in (3.13). Thus we introduce several position parameters which will be ever present in our studies.

First there is the *relative* or *barycentric* position of n within the interval $[q_{k-1}(\alpha), q_k(\alpha))$, where $k = k(\alpha, n)$ is the only positive integer satisfying $q_{k-1}(\alpha) \leq n < q_k(\alpha)$. We therefore define

$$\mu(\alpha, n) = \frac{n - q_{k-1}(\alpha)}{q_k(\alpha) - q_{k-1}(\alpha)}, \tag{3.14}$$

where $k = k(\alpha, n)$ is the only positive integer satisfying $q_{k-1}(\alpha) \leq n < q_k(\alpha)$.

The relative position μ is illustrated in Figure 3.7.

We remark that with this new parameter $\mu = \mu(\alpha, n)$, the recurrence quotient can be rewritten as

$$S(\alpha, n) = 1 + \frac{q_{k-1}(\alpha) + q_k(\alpha)}{n} = 1 + \frac{q_{k-1}(\alpha) + q_k(\alpha)}{q_{k-1}(\alpha) + \mu \cdot (q_k(\alpha) - q_{k-1}(\alpha))},$$

so that, dividing through by $q_k(\alpha)$

$$S(\alpha, n) = 1 + \frac{\frac{q_{k-1}(\alpha)}{q_k(\alpha)} + 1}{\frac{q_{k-1}(\alpha)}{q_k(\alpha)} + \mu \cdot \left(1 - \frac{q_{k-1}(\alpha)}{q_k(\alpha)}\right)} = 1 + \frac{1 + \frac{q_{k-1}(\alpha)}{q_k(\alpha)}}{\mu + \frac{q_{k-1}(\alpha)}{q_k(\alpha)} \cdot (1 - \mu)}. \quad (3.15)$$

The case when $\mu(\alpha, n) \approx 1/2$ is of particular interest. Indeed, we have $S(\alpha, n) \approx 3$, as the numerators and denominators in (3.15) cancel out. This corresponds exactly to Proposition 3.17.

Equation 3.15 makes evident that, to have a large $S(\alpha, n)$ then we must have

$$\mu(\alpha, n) \approx 0, \quad \frac{q_{k-1}(\alpha)}{q_k(\alpha)} \approx 0.$$

This was implicit in the proof of Theorem 3.4, the Morse-Hedlund theorem giving the “worst case”. Indeed, note that we at once picked $\mu(\alpha, n) = 0$ by making the choice $n = q_{k-1}(\alpha)$ while $\frac{q_{k-1}(\alpha)}{q_k(\alpha)} = \left(m_k + \frac{q_{k-2}(\alpha)}{q_{k-1}(\alpha)}\right)^{-1}$, that is small only when m_k is large. Thus we rephrased the principles of Theorem 3.4.

This motivates the introduction of the parameter

$$\rho(\alpha, n) = \frac{q_{k-1}(\alpha)}{q_k(\alpha)}, \quad (3.16)$$

where $k = k(\alpha, n)$ is the only positive integer satisfying $q_{k-1}(\alpha) \leq n < q_k(\alpha)$. The parameter $\rho(\alpha, n)$, which we call *quotient*, measures the relative sizes of the ends of the intervals.

We summarize in the following proposition the bounds relating μ , ρ and S .

Proposition 3.18. *The recurrence quotient $S(\alpha, n)$ satisfies the following inequalities with respect to the relative position $\mu = \mu(\alpha, n)$ and the quotient $\rho = \rho(\alpha, n)$*

$$1 + \frac{1}{\mu + \rho} \leq S(\alpha, n) \leq 1 + \frac{4}{\mu + \rho}. \quad (3.17)$$

Thus the positional parameters $\mu(\alpha, n)$ and $\rho(\alpha, n)$ play a key role in our studies of the recurrence.

3.4.3 Our framework: probabilistic analyses

Motivated by the question “does the recurrence function $R(\alpha, n)$ of a random Sturmian word behave more like $n \log n$ (as in Theorem 3.4), or is it more linear (as in Proposition 3.17)?”, we present two probabilistic models which aim to definitely answer this question. We also seek to explain the worst-case behavior.

In [BCR⁺15] we provided a first probabilistic study concerning the recurrence quotient $S(\alpha, n)$, and in [RV17] we undertook the study of a completely different model which turns out to be somewhat more natural from an algorithmic perspective. In the present text we begin from the latter [RV17] in Chapter 4 and then move on to the model from [BCR⁺15] in Chapter 5, in order to have a gentler introduction to the subject. In any case we explain what the two models are before getting into the heart of the matter. The relation between the two models will be explained throughout.

First, it is fair to recall that, since the language of a Sturmian word \mathbf{u} is determined exclusively by its slope α (see subsection 3.2.1) and, in turn, the recurrence function $R_{\mathbf{u}}(n)$ depends only on the language of \mathbf{u} , we just write $R(\alpha, n)$ in terms of α . Thus we may pick α at random rather than \mathbf{u} , which at first may sound like “picking a random language of a Sturmian word”. We argue, however, that this choice is justified: if we pick a random Sturmian word, which is necessarily of the form $\underline{\mathcal{S}}(\alpha, \beta)$ or $\overline{\mathcal{S}}(\alpha, \beta)$ thanks to Theorem 3.1, to study the resulting recurrence function we need only consider the distribution of the final slope α .

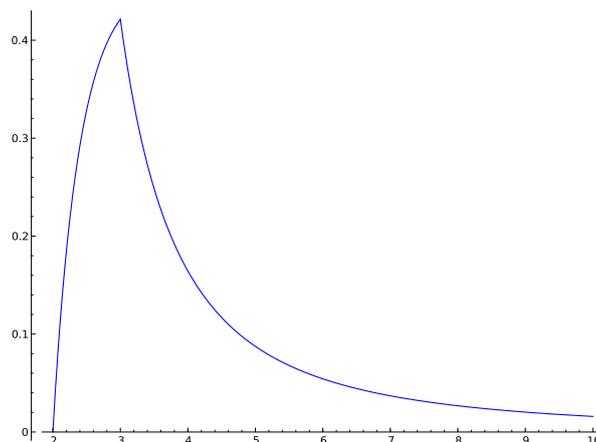


Figure 3.8: The density $\frac{d}{d\lambda} F_S(\lambda)$ of S_n . Observe the cusp at $\lambda = 3$.

Our original models [BCR⁺15] and [RV17] consider α distributed uniformly on $[0, 1]$. This seemingly crucial hypothesis can be slightly relaxed, however, as we later showed that we would get the exact same results of [RV17] if we instead considered α to be a random variable having a density with respect to the usual Lebesgue measure. The reason why this occurs is interesting, and boils down to a sort of independence “on the average” between the fractions p_k/q_k and q_{k-1}/q_k , which is a remarkable fact on its own right.

Here we introduce both models briefly, explaining the basic results and their consequences for Sturmian words. Chapters 4 and 5 provide a more in depth study.

3.4.4 Two probabilistic models

Now we present our probabilistic models in detail.

(i) Fixed $n \rightarrow \infty$ model.

Given a random \mathbf{u} , we wonder how the recurrence $R_{\mathbf{u}}(n)$ behaves as $n \rightarrow \infty$. Our “fixed $n \rightarrow \infty$ ” model follows this idea. Given a random α we study the distribution of the recurrence quotient $S(\alpha, n)$ as $n \rightarrow \infty$.

This is the model from [RV17] and is described in detail in Chapter 4.

Model. Fix the integer n (corresponding to the length of the factors, which will further tend to ∞); the index k of the interval $[q_{k-1}(\alpha), q_k(\alpha))$ which contains n is a random variable $k = k(\alpha, n)$.

The sequence $n \mapsto S_n(\alpha) := S(\alpha, n)$, where S is the recurrence quotient from (3.13), is now a sequence of random variables which we wish to study. It is important to note that then so are $\mu_n(\alpha) := \mu(\alpha, n)$ and $\rho_n(\alpha) := \rho(\alpha, n)$, which are interesting from the point of view of continued fractions in their own right.

Results: a little taste. Our first main theorem in Chapter 4, Theorem 4.1, implies the existence of a limiting distribution for the recurrence quotient $S_n(\alpha) := S(\alpha, n)$ when α is drawn uniformly at random from the unit interval $\mathcal{I} = [0, 1]$.

[Simplified Theorem 4.1 applied to the recurrence quotient]. The sequence $n \mapsto S_n(\alpha)$ as $n \rightarrow \infty$ admits a limit distribution $F_S(\lambda)$ whose derivative is shown in Figure 3.8.

For the recurrence quotient we also derive the following tail-inequality valid for all n ,

$$\mathbb{P}[S_n \geq b] \leq \frac{2}{b-1},$$

and every $b \geq 3$. The limiting distribution actually indicates that the asymptotically (as $n \rightarrow \infty$) tight constant in the above inequality should be $\frac{12}{\pi^2}$, which is closer to 1 than to 2.

Theorem 4.1 actually applies to a more general class of functions that we called \mathcal{Q} -functions (which we introduced in [RV17]). This class includes useful functions such as $S(\alpha, n)$, $\mu(\alpha, n)$ and $\rho(\alpha, n)$.

Our second result concerns the limiting density. It is not necessarily true, in general, that the derivative of the limiting distribution $\frac{d}{d\lambda}F(\lambda)$ will also be the limiting density; one must be careful when speaking about the limiting density. In our case, the variables S_n have actually a *discrete* distribution which tend to a continuous (and differentiable, except for the single point $\lambda = 3$) one.

[Simplified Theorem 4.2] Consider the distribution function $F_n(\lambda)$ of S_n , as well as the corresponding limit distribution $F_S(\lambda)$ derived from Theorem 4.1

For any strictly positive sequence $n \mapsto \epsilon(n)$ which tends to 0 with $n\epsilon(n) \rightarrow \infty$, the secants of the distribution F_n with step $\epsilon(n)$ converge to $F'_S(\lambda)$ uniformly on λ .

Observe then that this result asserts the convergence of the histograms towards the limit density provided that the step $\epsilon(n)$ is not excessively small.

Even though we may study the recurrence quotient in distribution, its expected value $\mathbb{E}[S_n]$ is infinite for each n . This happens because of the contribution of the cases in which the position parameters $\mu(\alpha, n)$ and $\rho(\alpha, n)$, introduced in subsection 3.4.2, are both small. As we wish to characterize the $\log n$ worst-case behavior of $S(\alpha, n)$ given by Proposition 3.16, we condition to events such as $\{\alpha : \mu(\alpha, n) \geq 1/n\}$ or $\{\alpha : \rho(\alpha, n) \geq 1/n\}$, making μ and ρ bounded reasonably far from 0.

[Simplified Theorem 4.3] Consider a sequence $\epsilon(n)$ which is $\Omega(1/(n \log n))$. The conditional expectations of the recurrence quotient S_n with respect to the event $\mu_n \geq \epsilon(n)$ satisfy

$$\mathbb{E} \left[S_n \mid \mu_n \geq \epsilon(n) \right] \sim_{n \rightarrow \infty} \frac{12}{\pi^2} |\log \epsilon(n)| .$$

We will see in Chapter 4 that the choice of $1/n$ is, in a way, optimal and that we get back a $\log n$. This \log arises naturally.

(ii) **Fixed depth $k \rightarrow \infty$ model.**

We introduced this model in [BCR⁺15], and we describe it in detail in Chapter 5. It is seemingly very different from the one described in (i) (we discuss their relation in Chapter 6), and motivated by geometric considerations regarding the interval $[q_{k-1}(\alpha), q_k(\alpha))$ containing n .

In the previous $n \rightarrow \infty$ model, the position parameters from subsection 4.2.1 are random variables which in principle we cannot control. Since α is random, there is no hope of fixing the quotient ρ , but we may try fixing the relative position μ by making n depend on both α and μ .

In our second (ii) model we *fix* the relative position by considering random (depending on α) sequences $(n_k(\alpha))_k$, satisfying $\mu(\alpha, n_k(\alpha)) \approx \mu$, our prescribed constant.

Then we wonder how the recurrence quotient behaves within these sequences!

Model. Fix a depth k (that further tends to ∞), and a real number $\mu \in [0, 1]$. For any slope α , we consider the interval $[q_{k-1}(\alpha), q_k(\alpha))$ delimited by the two successive continuants with indices $k-1$ and k , and we choose there the integer $n := n_k$ at a barycentric position, which now becomes a random variable.

This model may be called the model “large fixed k ”. The sequence $k \mapsto S_k(\alpha) := S(\alpha, n_k(\alpha))$ is a sequence of random variables.

Results: a little taste. Our main theorem in Chapter 5, Theorem 5.1, regards the limiting expected values and densities.

[Part (i) from Theorem 5.1]. For each $\mu \in (0, 1]$, the expected values $\mathbb{E}[S_k]$ satisfy

$$\lim_{k \rightarrow \infty} \mathbb{E}[S_k] = 1 + \frac{1}{\log 2} \frac{|\log \mu|}{1 - \mu},$$

where the convergence can be made uniform provided that $\mu \in [\epsilon, 1]$ for some fixed constant $\epsilon > 0$.

This shows precisely how the size of the recurrence quotient S depends on the relative position μ . It is actually possible to say more, by making $\mu = \mu_k$ a variable of k that tends to 0. This is our second main theorem, Theorem 5.2.

[Part of (i) from Theorem 5.2]. For any α , and for each $\tau \in [\varphi^2, 1]$, there exists a family of increasing subsequences $\mathcal{N}(\alpha, \tau)$, depending on both α and τ , of indices n for which

$$\mathbb{E} \left[\frac{R_\alpha(n)}{n} - \frac{12|\log \tau|}{\pi^2} \log n \right] = O(1) \quad (n \rightarrow \infty). \quad (3.18)$$

This result can be seen as a probabilistic version of the Proposition 3.16 from Morse and Hedlund [MH40].

These, as well as other results concerning the distributions are described extensively in Chapter 5.

CHAPTER 4

A FIRST PROBABILISTIC STUDY OF STURMIAN WORDS AND Q -FUNCTIONS

4.1 Introduction

Two different probabilistic settings.

As mentioned in subsection 3.4.3, we adopt a probabilistic approach and consider a random Sturmian word, associated with a random irrational slope α of the unit interval. We consider two possibilities:

- (i) fix the integer n (corresponding to the length of the factors, which will further tend to ∞); the index k of the interval $[q_{k-1}(\alpha), q_k(\alpha))$ which contains n is a random variable $k = k(\alpha, n)$. This model may be called the model “large fixed n ”. The sequence $n \mapsto S(\alpha, n)$ is now a sequence of random variables.
- (ii) fix a depth k (that further tends to ∞), and a fixed $\mu \in [0, 1]$. For any slope α , we consider the interval $[q_{k-1}(\alpha), q_k(\alpha))$ delimited by the two successive continuants with indices $k - 1$ and k , and we choose there the integer $n := n_k^{(\mu)}(\alpha)$ at a fixed barycentric position μ , which now becomes a random variable. This model may be called the model “large fixed k ”. The sequence $k \mapsto S_k^{(\mu)}(\alpha) := S(\alpha, n_\mu(\alpha, k))$ is a sequence of random variables.

In both cases, we are interested in the same type of questions about the sequence of random variables: does there exist a limit for the expectations? a limit distribution? a limit density?

Main results of the chapter [RV17].

Here we consider the recurrence quotient within model (i), i.e., the model with “a large fixed n ”, while model (ii) will be considered in Chapter 5 and we discuss their relation in Chapter 6. We obtain three results for the recurrence quotient on this model; more precisely, we consider the random variables $\alpha \mapsto S(\alpha, n)$ and study them for large n . We exhibit a limit for their distribution, and prove that there exists a limit density, as $n \rightarrow \infty$. We also study the conditional expectation of the recurrence quotient, when we exclude the possibility of n being too close to the left end of the interval $[q_{k-1}(\alpha), q_k(\alpha))$. More generally, we describe a class of events for which the order of this conditional mean value is exactly $\log n$. This can be



Figure 4.1: Here we illustrate the case $q_{k-1}(\alpha) \leq n < q_k(\alpha)$, with a relative position $\mu(\alpha, n) \approx 1/3$.

viewed as a probabilistic extension of Theorem 3.4.

Our proofs employ elementary methods: they are based on a precise comparison between an integral and its Riemann sum; however, the integral is improper (but convergent) and the Riemann sum is constrained by a coprimality condition, what we call a “coprime Riemann sum”.

We also introduce a general family of functions, called continuant-functions or \mathcal{Q} -functions, which are defined via the sequence of continuants $k \mapsto q_k(\alpha)$. The recurrence quotient is an instance of such a function, but the other “geometric” parameters of interest provide natural examples of such a notion. Our framework is naturally adapted to the study of a general \mathcal{Q} function.

4.2 Framework and results.

This section starts off by introducing several parameters describing the geometry of “continuant intervals” or the position of the integer n inside the continuant interval. Section 4.2.2 defines the class of \mathcal{Q} functions that provides a convenient framework for our study. Then, we state Theorems 4.1 and 4.2 in Sections 4.2.4 and 4.2.5, for a fairly general set of \mathcal{Q} -functions. We return to our specific parameters of interest, notably the recurrence function in Section 4.2.4, with two figures (Figures 4.9 and 4.5b). Finally, Section 4.4 concludes with a study of the conditional expectations.

4.2.1 Position parameters.

Besides the recurrence quotient, there are also three other interesting parameters ν, μ, ρ which describe the geometry of the interval $[q_{k-1}(\alpha), q_k(\alpha))$ which contains n (this is the case for ρ) or the position of n inside this interval (the case for μ and ν)

$$\rho(\alpha, n) = \frac{q_{k-1}(\alpha)}{q_k(\alpha)}, \quad (4.1)$$

$$\mu(\alpha, n) := \frac{n - q_{k-1}(\alpha)}{q_k(\alpha) - q_{k-1}(\alpha)}, \quad \nu(\alpha, n) = \frac{n}{q_k(\alpha)}. \quad (4.2)$$

When n belongs to the interval $[q_{k-1}(\alpha), q_k(\alpha))$, the recurrence quotient is expressed in terms of ρ and ν as

$$S(\alpha, n) = 1 + \frac{1 + \rho(\alpha, n)}{\nu(\alpha, n)}. \quad (4.3)$$

As $\nu(\alpha, n)$ belongs to the interval $[\rho(\alpha, n), 1]$, the following bounds hold

$$2 + \rho(\alpha, n) \leq S(\alpha, n) \leq 2 + \frac{1}{\rho(\alpha, n)} \quad (4.4)$$

(the lower bound holds for n close to $q_k(\alpha)$ whereas the upper bound is attained for $n = q_{k-1}(\alpha)$).

The ratio $\rho(\alpha, n)$ belongs to $(0, 1]$, and the Borel-Bernstein Theorem (Theorem 1.5) proves that $\liminf_{n \rightarrow \infty} \rho(\alpha, n) = 0$ for almost any irrational α . This is implicit in the proof of Theorem 3.4.

Parameter	Function $f(x, y)$	Density $\frac{12}{\pi^2} J_f(\lambda)$
S	$1 + x + y$	$\begin{cases} \frac{12}{\pi^2} \frac{1}{\lambda-1} \log(\lambda-1) & \text{if } 2 \leq \lambda \leq 3 \\ \frac{12}{\pi^2} \frac{1}{\lambda-1} \log(1 + \frac{1}{\lambda-2}) & \text{if } \lambda \geq 3. \end{cases}$
ρ	$\frac{x}{y}$	$\frac{12}{\pi^2} \frac{1}{1+\lambda} \log \lambda $ for $0 \leq \lambda \leq 1$
μ	$\frac{1-x}{y-x}$	$\begin{cases} \frac{12}{\pi^2} \frac{1}{2\lambda-1} \left(2 \log 2 - \frac{\log \lambda}{\lambda-1}\right) & \text{if } \lambda \neq 1/2 \\ \frac{24}{\pi^2} (1 - \log 2) & \text{if } \lambda = 1/2. \end{cases}$
ν	$\frac{1}{y}$	$\frac{12}{\pi^2} \frac{1}{\lambda} \log(1 + \lambda)$ for $0 \leq \lambda \leq 1$

Figure 4.2: Limit densities for the main parameters.

4.2.2 \mathcal{Q} -functions.

More generally, we are interested in functions whose definition depends fundamentally on the partition $\mathcal{Q}(\alpha) = \{[q_{k-1}(\alpha), q_k(\alpha)) : k \geq 1\}$ defined by the continuants of α , and consider the functions $(\alpha, n) \mapsto \Lambda(\alpha, n)$ that are associated with some function f and are written in terms of it as,

$$\Lambda(\alpha, n) = f\left(\frac{q_{k-1}(\alpha)}{n}, \frac{q_k(\alpha)}{n}\right), \quad (4.5)$$

as soon as $n \in [q_{k-1}(\alpha), q_k(\alpha))$.

To get our results in distribution and density, we will have to add several conditions on f , in particular for the results concerning the conditional probabilities, as we will need information concerning the rates of convergence towards the distribution.

A function Λ defined as in (4.5) is what we call a \mathcal{Q} -function (also continuant function), and we demand that the function f associated with Λ satisfy

- (i) it is defined on the unbounded rectangle

$$\mathcal{R} := \{(x, y) : 0 \leq x \leq 1 < y\},$$

- (ii) it is non negative on \mathcal{R} .

To get good error terms when it comes to the convergence in distribution, we shall consider a more specific class of \mathcal{Q} -functions, called \mathcal{LQ} -functions for short, which are the \mathcal{Q} -functions associated with a function f that is the quotient of two linear functions:

$$f(x, y) = \frac{a_1x + b_1y + c_1}{a_2x + b_2y + c_2}. \quad (4.6)$$

Our four parameters of interest, namely the recurrence quotient S , the ratio ρ and the two parameters μ and ν which describe the position of integer n with respect to the interval $[q_{k-1}(\alpha), q_k(\alpha))$ are \mathcal{LQ} -functions, associated to the following functions f

$$f_S(x, y) = 1 + x + y, \\ f_\rho(x, y) = \frac{x}{y}, \quad f_\mu(x, y) = \frac{1-x}{y-x}, \quad f_\nu(x, y) = \frac{1}{y}.$$

The problem of determining the joint limit distribution for the \mathcal{Q} -functions resulting from the projections $(x, y) \mapsto x$ and $(x, y) \mapsto y$ has already been considered by Ustinov in [Ust09], with the addition of a condition $[m_{k+1}(\alpha), m_{k+2}(\alpha), \dots] \leq z$, on the tail of the continued fraction development (here k is the index such that $q_{k-1}(\alpha) \leq n < q_k(\alpha)$). Such a condition does not fundamentally change the problem, as we just need change the probability of the fundamental interval $|\mathcal{I}_{m_1, \dots, m_k}| = |h_{m_1, \dots, m_k}([0, 1])|$ in our current study for

$$|h_{m_1, \dots, m_k}([0, z])| = \frac{z}{q_k(\mathbf{m})(q_k(\mathbf{m}) + zq_{k-1}(\mathbf{m}))},$$

which is the length of the fundamental interval with the added condition that $[m_{k+1}(\alpha), m_{k+2}(\alpha), \dots] \leq z$.

The motivation of Ustinov was mainly answering Sinai and Ulcigrai [SU08], in a question purely concerning the distributions of the stopped continued fractions. Sinai and Ulcigrai showed the existence of the limiting distribution, deriving it from a flow, but did not give an explicit formula.

4.2.3 Probabilistic setting.

We recall the present setting, anticipated in subsection 3.4.4. Consider a fixed integer n , and a random real α in the unit interval $[0, 1]$. The sequence $\Lambda_n(\alpha) := \Lambda(\alpha, n)$ is now a sequence of random variables. We are interested in the limit distribution of the sequence when $n \rightarrow \infty$. Does there exist a limit distribution? a limit density?

4.2.4 Distributions.

In the distributional study, we associate each real $\lambda \geq 0$ to a subdomain of \mathcal{R} ,

$$\Delta_f(\lambda) := \{(x, y) : 0 \leq x \leq 1 < y; f(x, y) \leq \lambda\} \quad (4.7)$$

(which is a convex domain when f is an \mathcal{LQ}), and consider the integral

$$I_f(\lambda) = \iint_{\Delta_f(\lambda)} \omega(x, y) dx dy = I[\omega, \Delta_f(\lambda)], \quad (4.8)$$

which involves the function ω defined on \mathcal{R} by

$$\omega(x, y) = \frac{1}{y(x+y)}, \quad (4.9)$$

whose integral on \mathcal{R} satisfies $I(\omega, \mathcal{R}) = \pi^2/12$. The associated density

$$\psi(x, y) = \frac{12}{\pi^2} \frac{1}{y(x+y)} \quad (4.10)$$

plays a fundamental role in the sequel, as our originally discrete distribution smooths out (converges weakly) to the distribution associated with the density ψ , as the following result shows:

Theorem 4.1. *Consider a \mathcal{LQ} -function Λ associated with a function f . Then the sequence $n \mapsto \Lambda_n(\alpha)$ as $n \rightarrow \infty$ admits a limit distribution, and the sequence*

$$F_n(\lambda) := \mathbb{P}[\Lambda_n \leq \lambda] = \frac{12}{\pi^2} I_f(\lambda) + O\left(\frac{1}{n}\right), \quad (4.11)$$

involves the integral $I_f(\lambda)$ defined in (4.8). Moreover, the constant of the O term can be chosen so that it works for every pair (f, λ)

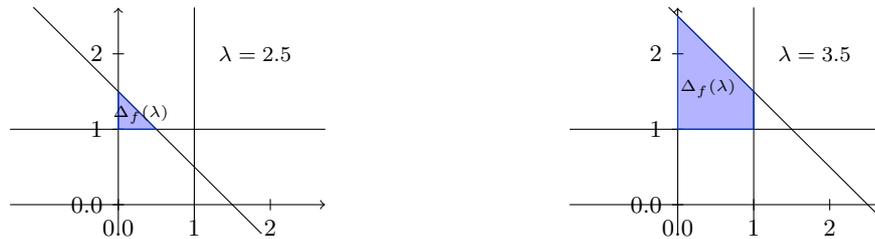


Figure 4.3: Domain of integration $\Delta_f(\lambda)$ for $f(x, y) := f_S(x, y) = 1 + x + y$.

It is worthwhile to notice that \mathcal{R} is the “smooth version” of the domain occupied by the points

$$\left\{ \left(\frac{q_{k-1}(\alpha)}{n}, \frac{q_k(\alpha)}{n} \right) : \alpha \in [0, 1] \setminus \mathbb{Q}, k = k(\alpha, n) \right\},$$

where $k(\alpha, n)$ is the integer $k \geq 1$ such that $q_{k-1}(\alpha) \leq n < q_k(\alpha)$.

The domain $\Delta_f(\lambda)$ in the case of the recurrence quotient S is illustrated in Figure 4.3. In this case $\Delta_f(\lambda)$ consists of all the points of \mathcal{R} lying below the line $f_S(x, y) = \lambda$.

Observe that as λ increases, the line $f_S(x, y) = \lambda$ just gets translated upwards in the y -axis. When $\lambda < 3$ we have the case shown in the first subfigure of Figure 4.3, the domain is delimited by the line $f_S(x, y) = \lambda$, the y -axis and the line $y = 1$. As we reach the vertex $(1, 1)$ for the case $\lambda = 3$, we get to the case of the second subfigure of Figure 4.3, and the domain $\Delta_f(\lambda)$ is now a polygon with 4 sides, delimited by the line $f_S(x, y) = \lambda$, the line $x = 1$, the line $y = 1$ and the y -axis.

The shape of the derivative of the limit distribution (see Figure 4.4), which we will then make sense of as a density in Theorem 4.2, is dictated by the moving frontier of $\Delta_f(\lambda)$. The maximum at $\lambda = 3$ coming from the fact that this moving frontier increases until $\lambda = 3$ and then decreases when we “change sides”. These ideas will be made formal with Theorem 4.2.

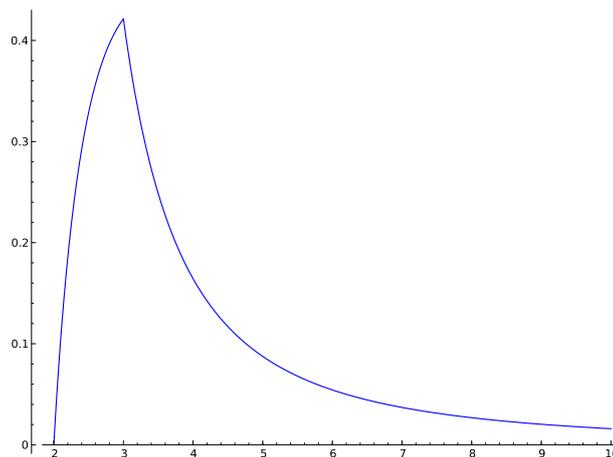


Figure 4.4: Limit density of the recurrence quotient S .

4.2.5 Results - densities.

We further show the convergence to the densities, more precisely, convergence of the histograms to the corresponding densities, and we characterize the points where these densities are not differentiable. An illustrating example is given by the recurrence quotient S , in this case the limit density (see Figure 4.2) is displayed in Figure 4.4, while we see the actual convergence of the experimental histograms towards this density in Figure 4.5.

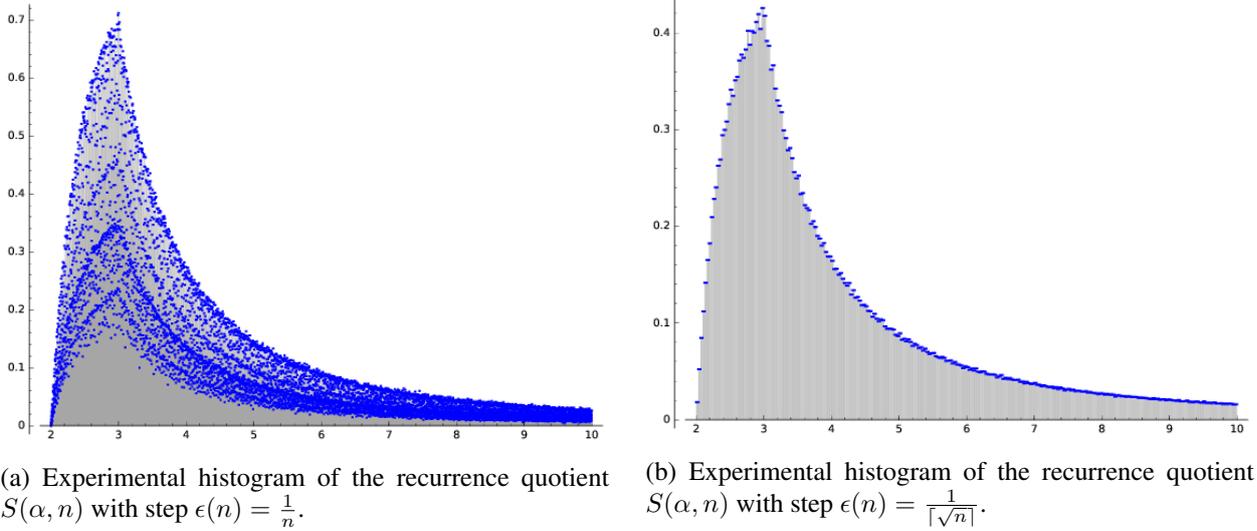


Figure 4.5: Limiting densities $J_{f_S}(\lambda)$ for the sequence $n \mapsto S(\alpha, n)$ as estimated by the scaled histograms. The number of experiments is $M = 10^7$, while $n = 1000$. The histograms are scaled so that they integrate to 1.

Note that the limit density of S is continuously differentiable except for the point $\lambda = 3$, its maximum, where it presents a “cusp”. This will be completely explained in our results.

For this we need to consider the boundary curves $\{(x, y) : f(x, y) = \lambda\}$ and their intersection with \mathcal{R} . We prove the following:

Theorem 4.2. Consider a \mathcal{LQ} -function Λ associated with a function f which is written as in (4.6). Then,

(a) The function $\lambda \mapsto I_f(\lambda)$ and its derivative J_f exist for any λ . The derivative J'_f exists except perhaps on a finite set, consisting of the point b_1/b_2 and two possible other values λ_0 and λ_1 . The following holds:

- (i) At each of the points $\lambda = \lambda_i$, the function J_f admits a left and a right derivative, each of them being finite.
- (ii) When the determinant $r(a, b) := a_1b_2 - a_2b_1$ is zero, the derivative J'_f exists at $\lambda = b_1/b_2$.
- (iii) When the determinant $r(a, b) := a_1b_2 - a_2b_1$ is not zero, the derivative J'_f does not exist at b_1/b_2 and is $O(|b_2\lambda - b_1|^{-1})$ for $\lambda \rightarrow b_1/b_2$.

(b) For any strictly positive sequence $n \mapsto \epsilon(n)$ which tends to 0 with $n\epsilon(n) \rightarrow \infty$, the secants of the distribution F_n with step $\epsilon(n)$ converge to $J_f(\lambda)$ and the following holds

$$\frac{F_n(\lambda + \epsilon(n)) - F_n(\lambda)}{\epsilon(n)} = \frac{12}{\pi^2} J_f(\lambda) + E(\lambda, \epsilon(n)), \quad (4.12)$$

(c) The error term satisfies

$$E(\lambda, \epsilon(n)) = O\left(\frac{1}{\epsilon(n)n}\right) + O\left(|J'_f(\lambda)|\epsilon(n)\right),$$

and the constants in the O -term do not depend on the pair (f, λ) .

4.2.6 Conditional expectations.

In order to control the size of the recurrence quotient $S(\alpha, n)$, we now focus on the position parameters ρ , ν and μ defined in (4.1) and (4.2), and consider the three sequences

$$\rho_n(\alpha) := \rho(\alpha, n), \quad \nu_n(\alpha) := \nu(\alpha, n), \quad \mu_n(\alpha) := \mu(\alpha, n).$$

The largest values of the recurrence quotient arise when ν or μ are small. In particular, the event $[\nu_n \geq \epsilon(n)]$ gathers the reals α for which the integer n is not too close to the left end of the interval $[q_{k-1}(\alpha), q_k(\alpha)]$, and, at the same time, the length of the interval $[q_{k-1}(\alpha), q_k(\alpha)]$ is of the same order as the right end $q_k(\alpha)$. We then consider a sequence $\epsilon(n) \rightarrow 0$, and condition with one of the events

$$[\rho_n \geq \epsilon(n)], \quad [\nu_n \geq \epsilon(n)], \quad [\mu_n \geq \epsilon(n)].$$

Theorem 4.3. *Consider a parameter $\Gamma \in \{\rho, \mu, \nu\}$ defined in (4.1) and (4.2) and a sequence $\epsilon(n)$ which is $\Omega(1/(n \log n))$. Then the conditional expectation of the recurrence quotient S_n with respect to the event $[\Gamma_n \geq \epsilon(n)]$ satisfies*

$$\mathbb{E} \left[S_n \mid \Gamma_n \geq \epsilon(n) \right] \sim_{n \rightarrow \infty} \frac{12}{\pi^2} |\log \epsilon(n)|.$$

This result exhibits a sequence of events, over a space in which the integer n is not too close to the left-end of interval $[q_{k-1}(\alpha), q_k(\alpha)]$. When we are sure not to be too close to this left-end, the recurrence quotient is (on average) of order $\log n$. This can be viewed as a probabilistic counterpart of Theorem 3.4, in the case of particular sequences of the form $\epsilon(n) = 1/(n\psi(n))$, for which the series of general term $\epsilon(n)$ is divergent. We return to this study at the end of Section 4.4.

The log appearing in the expression of Theorem 4.3, we argue, turns up naturally from the expressions for $f_S(x, y)$ and $\psi(x, y)$. Indeed, $\psi(x, y)$ is the “smoothed” density for $\left(\frac{q_{k-1}(\alpha)}{n}, \frac{q_k(\alpha)}{n}\right)$ around the point $(x, y) \in \mathcal{R}$, therefore the quantity $f_S(x, y)\psi(x, y)$ is the one we have to integrate to get expected values (we show this in Theorem 4.4). We point out now that $f_S(x, y)\psi(x, y) = \psi(x, y) + \frac{1}{y}$, and here it is $1/y$ that produces the logarithm. To make $f_S(x, y) = 1 + x + y$ big the only possibility is having a large y , so any condition $[\Gamma_n \geq \epsilon(n)]$ that makes f_S bounded, must bound y accordingly.

4.3 Proofs: distributions and densities.

We first recall some useful properties of continued fraction expansions and introduce coprime Riemann sums. Then, we prove the existence of limit distribution and limit densities for a general \mathcal{LQ} -function. The proof of Theorem 4.1 consists of three main steps, summarized in Proposition 4.1, Proposition 4.2, and Proposition 4.3, and we conclude the proof of Theorem 4.1 in Section 4.3.3. Sections 4.3.4 and 4.3.5. are devoted to the proof of Theorem 4.2.

4.3.1 Continued fractions, fundamental intervals and continuants.

(See Chapter 1 for more details). The continued fraction of an irrational number α of the unit interval $[0, 1]$ is

$$\alpha = \frac{1}{m_1 + \frac{1}{m_2 + \frac{1}{\ddots + \frac{1}{m_k + \frac{1}{\ddots}}}}}$$

Truncated at depth k , it gives rise to a rational number p_k/q_k associated with a coprime integer pair (p_k, q_k) . The numerator $p_k = p_k(\alpha)$ and the denominator $q_k = q_k(\alpha)$ are uniquely defined by the irrational number α . All the irrational numbers α which begin with the same sequence $\mathbf{m} = (m_1, m_2, \dots, m_k) \in \mathbb{N}^k$ belong to an interval, called a fundamental interval of depth k and denoted by $I_k(\mathbf{m})$. As the irrational numbers of $I_k(\mathbf{m})$ have the same convergents of order $\ell \leq k$, we denote their numerator and denominator by $p_\ell(\mathbf{m}), q_\ell(\mathbf{m})$. The ends of the interval $I_k(\mathbf{m})$ are

$$\frac{p_k(\mathbf{m})}{q_k(\mathbf{m})}, \frac{p_k(\mathbf{m}) + p_{k-1}(\mathbf{m})}{q_k(\mathbf{m}) + q_{k-1}(\mathbf{m})}.$$

As the equality $|p_k(\mathbf{m})q_{k-1}(\mathbf{m}) - p_{k-1}(\mathbf{m})q_k(\mathbf{m})| = 1$ holds, the length of the fundamental interval involves the function ω defined in (4.9) under the form

$$|I_k(\mathbf{m})| = \omega(q_{k-1}(\mathbf{m}), q_k(\mathbf{m})). \quad (4.13)$$

This explains why the function ω defined in (4.9) and the associated density ψ are ubiquitous in the study of the \mathcal{Q} -functions.

4.3.2 Distributions. Strategy of the proof.

There are two main steps in the proofs of Theorem 4.1.

- (i) *Discrete step.* We express in Proposition 4.1 the distribution of a \mathcal{Q} function in terms of a variant of a Riemann sum, that is called in the following a ‘‘coprime’’ Riemann sum. This type of ‘‘constrained’’ Riemann sum was already considered in [BCZ01].
- (ii) *Continuous step.* We compare the ‘‘coprime’’ Riemann sum to the associated integral. We begin by the comparison of the ‘‘plain’’ Riemann sum to the integral in Proposition 4.2, then, we take into account the coprimality condition in Proposition 4.3. We extend here the results of [BCZ01] which are only proven for finite domains.

Distributions and Riemann sums.

We begin with the alternative expression of a \mathcal{Q} -function Λ , associated with f , (already defined in (4.5)), which is written with the Iverson bracket^a under the form

$$\Lambda(\alpha, n) = \sum_{k \geq 0} f\left(\frac{q_{k-1}(\alpha)}{n}, \frac{q_k(\alpha)}{n}\right) \mathbb{I}\left[n \in [q_{k-1}(\alpha), q_k(\alpha)]\right].$$

The distribution of a \mathcal{Q} -function associated with f is

$$\mathbb{P}(\Lambda_n \leq \lambda) = \int_0^1 d\alpha \sum_{k \geq 0} \mathbb{I}\left[\left(\frac{q_{k-1}(\alpha)}{n}, \frac{q_k(\alpha)}{n}\right) \in \Delta_f(\lambda)\right].$$

For each k , the family of fundamental intervals $I_k(\mathbf{m})$ defines a pseudo-partition when \mathbf{m} goes through \mathbb{N}^k , and, for any $\alpha \in I_k(\mathbf{m})$, the equality $q_k(\alpha) = q_k(\mathbf{m})$ holds. We deduce

$$\begin{aligned} & \mathbb{P}[\Lambda_n \leq \lambda] \\ &= \sum_{k=0}^{\infty} \sum_{\mathbf{m} \in \mathbb{N}^k} |I_k(\mathbf{m})| \mathbb{I}\left[\left(\frac{q_{k-1}(\mathbf{m})}{n}, \frac{q_k(\mathbf{m})}{n}\right) \in \Delta_f(\lambda)\right]. \end{aligned}$$

^aThe Iverson bracket is a Boolean function defined by $[\mathcal{P}] = 1$ as soon as Property \mathcal{P} is true

Then, with the expression of the length $|I_k(\mathbf{m})|$ in terms of the function ω given in (4.13) and the fact that ω is homogeneous of degree -2 , we obtain

$$|I_k(\mathbf{m})| = \frac{1}{n^2} \omega \left(\frac{q_{k-1}(\mathbf{m})}{n}, \frac{q_k(\mathbf{m})}{n} \right).$$

Now, as we go through all the sequences $\mathbf{m} \in \mathbb{N}^*$, the coprime pairs $(q_{k-1}(\mathbf{m}), q_k(\mathbf{m}))$ give rise to all the coprime pairs (a, b) . Moreover, each coprime pair (a, b) , except the pair $(1, 1)$, appears exactly twice, due to the existence of two continued fraction expansions, the proper one (in which the last digits strictly greater than 1), and the improper one (in which the last digit is equal to 1). Then, the equality holds

$$\mathbb{P}[\Lambda_n \leq \lambda] = \frac{2}{n^2} \sum_{\substack{(a,b) \in \mathbb{Z}^2 \\ (a,b)=1}} \omega \left(\frac{a}{n}, \frac{b}{n} \right) \left[\left[\left(\frac{a}{n}, \frac{b}{n} \right) \in \Delta_f(\lambda) \right] \right].$$

The right member is the Riemann sum of the function 2ω on the domain $\Delta_f(\lambda)$ with step $1/n$, with an extra condition of coprimality. More generally, for a function g integrable on a subset Ω , we are led to the following two Riemann sums with step $1/n$: the first one $R_n(g, \Omega)$ is the usual one,

$$R_n(g, \Omega) = \frac{1}{n^2} \sum_{(a,b) \in \mathbb{Z}^2} g \left(\frac{a}{n}, \frac{b}{n} \right) \left[\left[\left(\frac{a}{n}, \frac{b}{n} \right) \in \Omega \right] \right],$$

whereas the second one $\widehat{R}_n(g, \Omega)$ takes into account the coprimality of (a, b) , and is called the ‘‘coprime’’ Riemann sum,

$$\widehat{R}_n(g, \Omega) = \frac{1}{n^2} \sum_{\substack{(a,b) \in \mathbb{Z}^2 \\ \gcd(a,b)=1}} g \left(\frac{a}{n}, \frac{b}{n} \right) \left[\left[\left(\frac{a}{n}, \frac{b}{n} \right) \in \Omega \right] \right],$$

We summarize:

Proposition 4.1. *Consider a \mathcal{Q} -function Λ associated with a function f . Then the distribution $F_n(\lambda) := \mathbb{P}[\Lambda_n \leq \lambda]$ is expressed with a coprime Riemann sum,*

$$\mathbb{P}[\Lambda_n \leq \lambda] = \widehat{R}_n(2\omega, \Delta_f(\lambda)). \quad (4.14)$$

which involves the density ω defined in (4.9) and the domain $\Delta_f(\lambda)$ defined in (4.7).

The previous result extends if we replace $\Delta_f(\lambda)$ by any other domain $\Omega \subset \mathcal{R}$. In particular, in Section 4, we will deal with two \mathcal{Q} -functions Λ and Γ associated respectively to f and g , together with the domain

$$\underline{\Delta}_{f,g}(\lambda, \epsilon) := \{(x, y) \in \mathcal{R} : f(x, y) \geq \lambda, g(x, y) \geq \epsilon\}, \quad (4.15)$$

and use the equality

$$\mathbb{P}[\Lambda_n \geq \lambda, \Gamma_n \geq \epsilon] = \widehat{R}_n(2\omega, \underline{\Delta}_{f,g}(\lambda, \epsilon)). \quad (4.16)$$

Usual Riemann sums and integrals.

We first deal with the usual Riemann sum, and compare it to its associated integral $I(g, \Omega)$. This is a classical proof, but we consider improper integrals and we wish to have precise error terms.

We now deal (only within this subsection) with

$$\mathcal{S} := [0, 1] \times (0, \infty), \quad (4.17)$$

consider a subset $\Omega \subset \mathcal{S}$ and associate with it the family of subsets

$$\Omega(k) := \Omega \cap ([0, 1] \times [k, k+1]),$$

for any $k \geq 1$, which form a pseudo-partition of Ω . We also consider a positive function g defined on Ω of class \mathcal{C}^1 , bounded on any bounded subset on Ω for which the following two finite bounds^b

$$C_g(\Omega, k) := \sup\{g(x, y) : (x, y) \in \Omega(k)\},$$

$$D_g(\Omega, k) := \sup\left\{\left|\frac{\partial g}{\partial y}(x, y)\right| : (x, y) \in \Omega(k)\right\},$$

define sequences whose associated series are convergent,

$$\sum_{k \geq 0} C_g(\Omega, k) < \infty, \quad \sum_{k \geq 0} D_g(\Omega, k) < \infty.$$

Their sums are denoted by $C_g(\Omega)$ and $D_g(\Omega)$, and we denote by $M_g(\Omega)$ their maximum.

Such a function g is called strongly decreasing on Ω with bound $M_g(\Omega)$. Such a function is integrable on Ω and the inequality $I(g, \Omega) \leq M_g(\Omega)$ holds.

Proposition 4.2. *Consider the domain \mathcal{S} defined in (4.17) and a function g which is strongly decreasing on a convex $\Omega \subset \mathcal{S}$ with bound $M_g(\Omega)$. Then, the Riemann sum of the function g on Ω compares to the integral,*

$$|R_n(g, \Omega) - I(g, \Omega)| \leq \frac{5}{n} M_g(\Omega). \quad (4.18)$$

Proof. We will prove the estimate, for each $k \geq 0$,

$$|R_n(g, \Omega(k)) - I(g, \Omega(k))| \leq \frac{4}{n} (C_g(\Omega, k) + D_g(\Omega, k)).$$

This will entail the result by taking the sum over $k \geq 0$.

We consider the elementary squares of side $1/n$, namely

$$\mathcal{R}_{a,b} = \left[\frac{a}{n}, \frac{a+1}{n}\right] \times \left[\frac{b}{n}, \frac{b+1}{n}\right],$$

and we concentrate on those which meet $\Omega(k)$. There are two cases for such rectangles $\mathcal{R}_{a,b}$, namely

$$(i) \mathcal{R}_{a,b} \subset \Omega(k), \quad \text{or} \quad (ii) \mathcal{R}_{a,b} \cap \Omega(k)^c \neq \emptyset.$$

In the first case (i), the definition of the bound D_g entails the estimate

$$\left|\frac{1}{n^2} g\left(\frac{a}{n}, \frac{b}{n}\right) - I(g, \mathcal{R}_{a,b})\right| \leq I\left(\left|g\left(\frac{a}{n}, \frac{b}{n}\right) - g\right|, \mathcal{R}_{a,b}\right) \leq \frac{1}{n^3} D_g(\Omega, k).$$

As the number of such squares is at most n^2 , the contribution from case (i) is at most $(1/n) D_g(\Omega, k)$.

In the second case (ii), the positivity of g and the definition of the bound C_g entails the estimate

$$\left|\frac{1}{n^2} g\left(\frac{a}{n}, \frac{b}{n}\right) - I(g, \Omega \cap \mathcal{R}_{a,b})\right| \leq \frac{1}{n^2} C_g(\Omega, k).$$

But, the convexity of Ω entails that there are at most $4n$ such squares, and the contribution of the second case is at most $(4/n) C_g(\Omega, k)$.

To see where the constant 4 comes from, we first replace $\Omega(k)$ by a closed convex polygon $\mathcal{C}_n \subset \Omega$, without affecting the bound : in each square $\mathcal{R}_{a,b}$ of the second case, pick a point in $\Omega(k)$ and then take the convex hull. If $\Omega(k)$ is a closed convex polygon, we go through the border in clockwise order and look at the grid rectangles we encounter as explained in Figure 4.6. ■

^bBy convention, we consider that $C_g(\Omega, k)$ and $D_g(\Omega, k)$ are 0 if the set $\Omega(k)$ is empty.

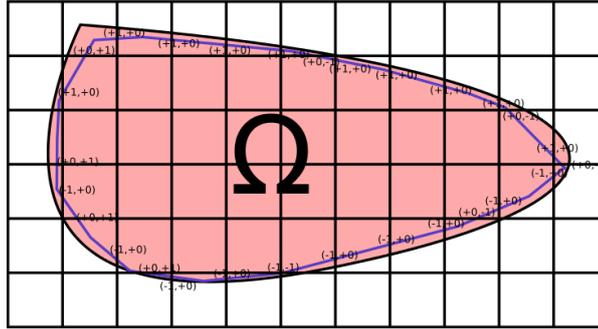


Figure 4.6: The convex domain Ω , and, in blue, a convex polytope \mathcal{P} . These two convex sets have the same grid squares that intersect both themselves and their complement. We traverse the polygon clockwise from the lowest vertex. Each time we intersect a horizontal line we move ± 1 square horizontally in the grid, similarly for the vertical lines, and diagonals. Being the polygon convex, once we stop moving upwards vertically (at most n steps), we can only move downwards (at most n steps) when moving vertically. A similar observation for the horizontal case tells us that there can be at most $2n$ horizontal steps.

Coprime Riemann sums and integrals.

The following result is an extension of the results obtained in [BCZ01], that are only proven for finite domains.

Proposition 4.3. *Consider a positive function g defined on \mathcal{R} , homogeneous of degree $-\beta$ there with $\beta > 1$. Such a function is strictly decreasing on \mathcal{R} . Consider also a convex subset $\Omega \subset \mathcal{R}$. Then, the coprime Riemann sum of the function g on Ω compares to the integral of g on Ω , namely*

$$\left| \widehat{R}_n(g, \Omega) - \frac{6}{\pi^2} I(g, \Omega) \right| \leq \frac{1}{n} (1 + 5\zeta(\beta)) M_g(\mathcal{R}).$$

Proof. To filter the cases in which $\gcd(a, b) > 1$, we use the Möbius function μ which performs “inclusion-exclusion”. The Möbius function $\mu : \mathbb{N} \rightarrow \{-1, 0, +1\}$ satisfies

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n > 1. \end{cases} \quad (4.19)$$

We consider the restricted “coprime” Riemann sum, where the sum is taken over the pairs (a, b) with $\gcd(a, b) = 1$, namely

$$n^2 \widehat{R}_n(g, \Omega) = \sum_{\substack{(a,b) \in \mathbb{Z}^2 \\ \gcd(a,b)=1}} g\left(\frac{a}{n}, \frac{b}{n}\right) \left[\left[\left(\frac{a}{n}, \frac{b}{n}\right) \in \Omega \right] \right].$$

We then “insert” the μ -function inside this restricted “coprime” Riemann sum,

$$n^2 \widehat{R}_n(g, \Omega) = \sum_{(a,b) \in \mathbb{Z}^2} g\left(\frac{a}{n}, \frac{b}{n}\right) \left[\left[\left(\frac{a}{n}, \frac{b}{n}\right) \in \Omega \right] \right] \left(\sum_{d|\gcd(a,b)} \mu(d) \right).$$

As the point $(a/n, b/n)$ belongs to \mathcal{R} with $a > 0$, the inequality $\gcd(a, b) \leq n$ holds. Then, inverting the summations entails the equality

$$n^2 \widehat{R}_n(g, \Omega) = \sum_{d \leq n} \mu(d) \sum_{(a,b) \in \mathbb{Z}^2} g\left(\frac{ad}{n}, \frac{bd}{n}\right) \left[\left[\left(\frac{ad}{n}, \frac{bd}{n}\right) \in \Omega \right] \right].$$

Finally, the following equality holds

$$\widehat{R}_n(g, \Omega) = \sum_{d \leq n} \mu(d) R_n(g_d, \Omega_d), \quad (4.20)$$

and involves the function g_d and the subset Ω_d defined as

$$g_d(x, y) := g(dx, dy), \quad \Omega_d = \frac{1}{d} \Omega.$$

As the inclusion $\Omega_d \subset \mathcal{S}$ holds, we now apply the previous Proposition 4.2 to each (plain) Riemann sum $R_n(g_d, \Omega_d)$ and obtain

$$|R_n(g_d, \Omega_d) - I(g_d, \Omega_d)| \leq \frac{5}{n} M_{g_d}(\Omega_d). \quad (4.21)$$

We now use three properties. We first remark the equality

$$I(g_d, \Omega_d) = \frac{1}{d^2} I(g, \Omega),$$

due to the change of variables $(x', y') = (dx, dy)$. Second, the series of general term $\mu(d)/d^2$ is convergent, and, with the Möbius inversion, its sum equal $1/\zeta(2)$ and

$$\left| \sum_{d \leq n} \frac{\mu(d)}{d^2} - \frac{6}{\pi^2} \right| \leq \frac{1}{n}.$$

Third, we relate the bound $M_{g_d}(\Omega_d)$ to its analogous. As g is homogeneous of degree $-\beta$, its derivative is homogeneous of degree $(-\beta - 1)$ and the two relations

$$g_d(x, y) = g(dx, dy) = \frac{1}{d^\beta} g(x, y),$$

$$\frac{\partial g_d}{\partial y}(x, y) = d \frac{\partial g}{\partial y}(dx, dy) = \frac{1}{d^\beta} \frac{\partial g}{\partial y}(x, y),$$

hold for $(x, y) \in \mathcal{R}$. As g and its derivative are 0 outside \mathcal{R} , the same holds for g_d and its derivative, and

$$M_{g_d}(\Omega_d) = M_{g_d}(\Omega_d \cap \mathcal{R}) = \frac{1}{d^\beta} M_g(\Omega_d \cap \mathcal{R}) \leq \frac{1}{d^\beta} M_g(\mathcal{R}).$$

Then, as $\beta > 1$, one has

$$\sum_{d \leq n} M_{g_d}(\Omega_d) \leq \zeta(\beta) M_g(\mathcal{R}).$$

With the three previous properties, together with Eq. (4.21), we obtain the final result. ■

4.3.3 Distributions. Proof of Theorem 4.1.

Theorem 4.1 is a particular case of the previous Proposition 4.3, when it applies to ω and $\Delta_f(\lambda)$ defined in (4.9) and (4.7). The function ω is homogeneous of degree 2 and the domain $\Delta_f(\lambda)$ is convex, as it is the intersection of the unbounded rectangle \mathcal{R} with the half-plane $\{f(x, y) \leq \lambda\}$. Applying Proposition 4.3 then proves Theorem 4.1.

4.3.4 Proof of Theorem 4.2. First step.

We first prove Assertion (b) assuming Assertion (a). We let

$$F_n(\lambda) := \mathbb{P}[\Lambda_n \leq \lambda], \quad F_\infty(\lambda) = \frac{12}{\pi^2} I_f(\lambda).$$

We know from Assertion (a) that the derivative $J_f(\lambda)$ of $\lambda \mapsto I_f(\lambda)$ exists. This is the same for the function F_∞ and we wish to estimate the difference

$$\left| \frac{F_n(\lambda + \epsilon(n)) - F_n(\lambda)}{\epsilon(n)} - F'_\infty(\lambda) \right|.$$

We begin with the triangle inequality

$$\begin{aligned} & \left| \frac{F_n(\lambda + \epsilon(n)) - F_n(\lambda)}{\epsilon(n)} - F'_\infty(\lambda) \right| \tag{4.22} \\ & \leq \left| \frac{F_n(\lambda + \epsilon(n)) - F_\infty(\lambda + \epsilon(n))}{\epsilon(n)} \right| + \left| \frac{F_\infty(\lambda) - F_n(\lambda)}{\epsilon(n)} \right| \\ & \quad + \left| \frac{F_\infty(\lambda + \epsilon(n)) - F_\infty(\lambda)}{\epsilon(n)} - F'_\infty(\lambda) \right|. \end{aligned}$$

With the special form of function f , the domain $\Delta_f(\lambda)$ is convex, and Theorem 4.1 provides the estimates

$$\begin{aligned} |F_n(\lambda) - F_\infty(\lambda)| &= O(1/n), \\ |F_n(\lambda + \epsilon(n)) - F_\infty(\lambda + \epsilon(n))| &= O(1/n), \end{aligned}$$

where the constant in the O -terms does not depend on λ and $\epsilon(n)$. Then, the first two terms in Inequality (4.22) are $O(1/(n\epsilon(n)))$ and tend to 0 because $n\epsilon(n) \rightarrow \infty$. For the last term in (4.22), we use Taylor expansion of order 2 of the function F_∞ together with Assertion (a).

4.3.5 Proof of Theorem 4.2. Second step.

We now prove Assertion (a).

The set of lines \mathcal{F} . In the set \mathcal{F} of lines, defined as

$$\mathcal{F} := \{f(x, y) = \lambda : \lambda \in \mathbb{R}\},$$

the equation of the line $f(x, y) = \lambda$ is written in terms of coefficients described in (4.6) as

$$(a_1x + b_1y + c_1) - \lambda(a_2x + b_2y + c_2) = 0. \tag{4.23}$$

The case in which the two vectors (a_1, b_1, c_1) and (a_2, b_2, c_2) are colinear is excluded, as in this case $f(x, y)$ is constant. The case $b_1 = b_2 = 0$ is also excluded as we wish that f depend on y . Thus, there is at most one vertical line in \mathcal{F} .

There are two cases for the set \mathcal{F} defined in (4.23)

- (i) the case when the determinant $r(a, b) := a_1b_2 - a_2b_1$ is zero and in this case the determinant $r(a, c) := a_1c_2 - a_2c_1$ is not zero. The set \mathcal{F} is formed by parallel lines of slope $-a_1/b_1$. This is for instance the case of the recurrence quotient with slope -1 or the case of ν with slope 0.

(ii) the case when the determinant $r(a, b) := a_1b_2 - a_2b_1$ is not zero. In this case, we can choose $r(a, b) = 1$ due to the homogeneity of the problem. Then, the set \mathcal{F} is made up of all the lines which contain the point (x_0, y_0) uniquely defined by the relations

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} -c_1 \\ -c_2 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} r(b, c) \\ -r(a, c) \end{pmatrix}.$$

The point (x_0, y_0) is called the basic point of \mathcal{F} . Remark that case (i) can be seen as the limit of the case (ii) when (x_0, y_0) tends to ∞ in the direction a_1/b_1 . The basic points attached to our parameters ρ, μ are $(0, 0)$ for ρ and $(1, 1)$ for μ .

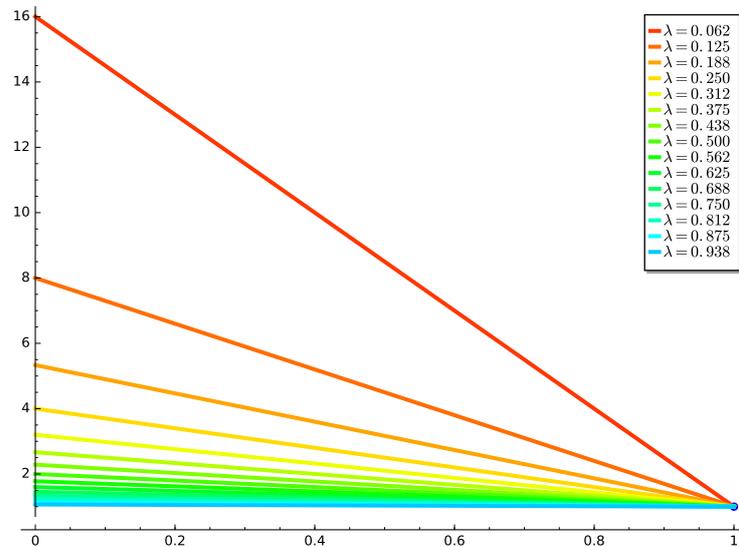


Figure 4.7: The lines $f_\mu(x, y) = \lambda$ for the case of μ . Note the basic point $(1, 1)$ at the bottom-right corner. In the set \mathcal{F} of basic point (x_0, y_0) , the value of λ and the inverse $1/\tau$ of the slope τ of the line $f(x, y) = \lambda$, are related via linear fractional transformations with determinant equal to 1, namely

$$\lambda = F(\tau) = \frac{a_1\tau + b_1}{a_2\tau + b_2} \quad \tau = G(\lambda) = \frac{b_2\lambda - b_1}{a_2\lambda - b_2}.$$

In the set \mathcal{F} of basic point (x_0, y_0) , the parametrization of the line $f(x, y) = \lambda$ of slope $1/\tau$ is thus

$$x = x_0 + \tau(y - y_0), \quad \tau = G(\lambda).$$

Expressions of I_f and its derivative. Consider a function f as in (4.6); denote by $\delta_f(\tau)$ the segment (possibly empty or unbounded) which is the intersection of the line $f(x, y) = \lambda = F(\tau)$ of slope $1/\tau$ with the rectangle \mathcal{R} . Now, the function f is fixed, the point (x_0, y_0) is fixed, and all the indices which involve f are removed. There is an open interval D which gathers the values of τ for which the segment $\delta(\tau)$ is not empty, and we denote by $A(\tau), B(\tau)$ the ordinates of the two ends of the segment $\delta(\tau)$.

As soon as the line $f(x, y) = F(\tau)$ is not horizontal, we consider the natural parametrization h_τ of the line $\delta(\tau)$, namely a map $h_\tau : (A(\tau), B(\tau)) \rightarrow \delta(\tau)$ which associates to y the point

$$h_\tau(y) = h(\tau, y) = (x_0 + \tau(y - y_0), y)$$

of the line $\delta(\tau)$. The map $\tau \mapsto h_\tau$ is of class \mathcal{C}^∞ on D .

Using the change of variables $(\theta, y) \mapsto (h(\theta, y), y)$, and its Jacobian $|(\partial h)/(\partial \theta)(y, \theta)| = |y - y_0|$, the integral $L(\tau) := L_f(\tau) := I_f(F(\tau)) = I_f \circ F(\tau)$ is written as

$$L(\tau) = \int_{-\infty}^{\tau} d\theta \int_{A(\theta)}^{B(\theta)} Q(\theta, y) dy,$$

with $Q(\theta, y) = \omega(x_0 + \theta(y - y_0), y) |y - y_0|$.

(We have used the fact that F is increasing). Then the derivative of L admits the expression

$$L'(\tau) = \int_{A(\tau)}^{B(\tau)} Q(\tau, y) dy. \quad (4.24)$$

The function L' is itself differentiable on the set D , except perhaps on a finite set (as we will see now) and involves the previous functions under the form

$$L''(\tau) = \int_{A(\tau)}^{B(\tau)} \frac{\partial Q}{\partial \tau}(\tau, y) dy \quad (4.25)$$

$$+ B'(\tau) Q(\tau, B(\tau)) - A'(\tau) Q(\tau, A(\tau)), \quad (4.26)$$

$$\text{with } \frac{\partial Q}{\partial \tau}(\tau, y) = \frac{\partial \omega}{\partial x}(x_0 + \tau(y - y_0), y) |y - y_0|^2.$$

We prefer to deal with the function L , as it is easier to “see the geometry”. We will return to the function I and its two derivatives with the relations

$$I'(\lambda) = \frac{L'(\tau)}{F'(\tau)}, \quad I''(\lambda) F'(\tau)^2 = L''(\tau) - L'(\tau) \frac{F''(\tau)}{F'(\tau)}, \quad (4.27)$$

and use the special form of F defined in (4.23).

The role of the corners. The values of τ in D for which I' is *a priori* not differentiable are those for which the line of slope $1/\tau$ is vertical or meets one of the two “corners” of \mathcal{R} , namely the slope $1/\tau_0$ for which it meets the point $(0, 1)$, and the slope $1/\tau_1$ for which it meets the point $(0, 1)$.

There are now two different geometric cases: the generic case (\mathcal{G}) or the exceptional case (\mathcal{E}), described as follows:

(\mathcal{G}) If the point (x_0, y_0) does not belong to the line $y = 1$, there are exactly two lines in \mathcal{F} , each of them containing one corner of \mathcal{R} , associated with two distinct values τ_0 and τ_1 .

(\mathcal{E}) If the point (x_0, y_0) belongs to the line $y = 1$, there is only one value $\tau_0 = \tau_1 = \infty$.

Finally, there are at most three values of τ in the set $\{0, \tau_0, \tau_1\}$ where L' is possibly not differentiable. But, L' possesses at each finite τ_i a left and a right derivative, each of them being finite. This is thus the same for the derivative I' of the function I . At $\tau = 0$, the derivatives $F'(0)$ and $F''(0)$ are finite as soon as $b_2 \neq 0$.

Behavior of $L''(\tau)$ for $\lambda \rightarrow 0$. The ratio $R(\tau) := B(\tau)/A(\tau)$ is important, as the estimates

$$Q(\tau, y) = \Theta(y^{-1}), \quad \frac{\partial Q}{\partial \tau}(\tau, y) = \Theta(y^{-1})$$

entail that $L'(\tau)$ and the first term of $L''(\tau)$ in (4.25) are both $\Theta(\log R(\tau))$.

The bound $B(\tau)$ always tends to $+\infty$ but there are two cases for $A(\tau)$: it remains bounded or not.

(i) The case when $A(\tau)$ remains bounded occurs if and only if the basic point belongs to one of the two vertical lines $x_0 = 1$ or $x_0 = 1$. Then the estimates $R(\tau) = \Theta(\tau^{-1})$ and $L'(\tau) = \Theta(\log \tau)$, directly entail that $L''(\tau)$ is $\Theta(\tau^{-1})$.

(ii) If $A(\tau)$ tends also to ∞ , then the ratio $R(\tau)$ tends to $|x_0 - 1|/|x_0|$, and this limit may be only finite non zero. Then, the derivatives $A'(\tau)$ and $B'(\tau)$ are $\Theta(\tau^{-2})$ whereas $A(\tau)$ and $B(\tau)$ are $\Theta(\tau^{-1})$ and thus $Q(\tau, B(\tau))$ and $Q(\tau, A(\tau))$ are $\Theta(\tau)$ and each term of (4.26) is $\Theta(\tau^{-1})$, whereas the first term in (4.25) tends to a finite limit. More precisely, the estimate

$$\tau \left(B'(\tau) Q(\tau, B(\tau)) - A'(\tau) Q(\tau, A(\tau)) \right) \rightarrow 1$$

ends with (4.27) the proof of Theorem 4.2 (a). ■

4.4 Proofs: conditional expectations.

We now focus on the conditional expectations. Our final purpose is to prove Theorem 4.3 which is devoted to the recurrence quotient. However, we begin by a more general study and we obtain in Section 4.3 a general result regarding the conditional expectations (Theorem 4.5). We then apply it in Section 4.4 to the particular case of the recurrence quotient, and this provides Theorem 4.6, which can be viewed itself as an extension of Theorem 4.3.

4.4.1 Limit expectation of bounded \mathcal{LQ} - functions.

Thus far, we dealt with distributions of \mathcal{LQ} -functions. Now, we consider expected values of an \mathcal{LQ} -function, and use the equality

$$\mathbb{E}[\Lambda_n] = \int_0^\infty \mathbb{P}[\Lambda_n \geq \lambda] d\lambda,$$

valid when $\Lambda \geq 0$, as in our case. We consider here the case of an \mathcal{LQ} -function Λ associated with a bounded function f (which is the case when b_2 is not zero). It is then possible to interchange the limit and the integral and use Theorem 4.1.

When reversing the order of integration, we first integrate with respect to λ , and we are led to the integral

$$\mathbb{E}_\psi[f] := \frac{6}{\pi^2} I(f \cdot 2\omega, \mathcal{R}) \quad (4.28)$$

which is exactly the expectation $\mathbb{E}_\psi[f]$ of the function f on the rectangle \mathcal{R} with respect to the density $\psi := (12/\pi^2)\omega$. We thus obtain the following result which provides an extension of Theorem 4.1:

Theorem 4.4. *Consider an \mathcal{LQ} -function Λ associated with a function f bounded by B_f . Then the sequence $n \mapsto \Lambda_n$ admits a limit expected value as $n \rightarrow \infty$ equal to the expectation $\mathbb{E}_\psi[f]$ of the function f on the rectangle \mathcal{R} with respect to the density $\psi := (12/\pi^2)\omega$, and*

$$\mathbb{E}[\Lambda_n] = \mathbb{E}_\psi[f] + B_f O\left(\frac{1}{n}\right), \quad (4.29)$$

where the constant in the O -term does not depend on f and λ .

4.4.2 Case of the recurrence quotient.

The function f associated with the recurrence quotient $S(\alpha, n)$ is $f_S(x, y) = 1 + x + y$. It is unbounded on \mathcal{R} , and the function f_S is not integrable with respect to ψ . In fact, by the argument of Proposition 4.1 the expected value can be worked out to be

$$\mathbb{E}[S_n] = \widehat{R}_n(2\omega f_S, \mathcal{R}),$$

and here $\widehat{R}_n(2\omega f_S, \mathcal{R})$ is infinite for each n .

This is why we consider the conditional expectations for the sequence S_n with respect to an event $[\Gamma_n \geq \epsilon(n)]$ associated with another \mathcal{LQ} -function Γ , namely

$$\mathbb{E}[S_n | \Gamma_n \geq \epsilon(n)].$$

We will choose in the sequel the \mathcal{LQ} -function Γ from the set $\{\mu, \nu, \rho\}$ and a positive sequence $\epsilon(n)$ tending to 0 not all too quickly.

4.4.3 General conditional expectations.

We consider more general conditional expectations,

$$\mathbb{E}[\Lambda_n | \Gamma_n \geq \epsilon] \quad (\epsilon > 0)$$

when Γ is an \mathcal{LQ} -function associated with a function g which tends to 0 for $y \rightarrow \infty$. (This means that the pair (b_1, b_2) in (4.6) satisfies $b_1/b_2 = 0$). The subset

$$\{(x, y) \in \mathcal{R} : g(x, y) \geq \epsilon\}$$

is bounded for $\epsilon > 0$, and we denote, for $\epsilon > 0$,

$$B_{f|g}(\epsilon) := \sup\{f(x, y) : g(x, y) \geq \epsilon\} < \infty.$$

In this case, the expectation of f with respect to ψ conditioned to the event $[g \geq \epsilon]$ is well defined, and denoted as

$$\mathbb{E}_\psi[f|g \geq \epsilon].$$

The following holds.

Theorem 4.5. *Consider two \mathcal{LQ} -functions Λ and Γ with respective associated functions f and g . Assume that g tends to 0 for $y \rightarrow \infty$. Then the conditional expectation of Λ_n with respect to the event $[\Gamma_n \geq \epsilon]$ satisfies*

$$\begin{aligned} \mathbb{E}[\Lambda_n | \Gamma_n \geq \epsilon] \cdot \mathbb{P}[\Gamma_n \geq \epsilon] &= \mathbb{E}_\psi[f|g \geq \epsilon] \cdot \mathbb{P}_\psi[g \geq \epsilon] \\ &+ B_{f|g}(\epsilon) O\left(\frac{1}{n}\right) \end{aligned}$$

where the constant in the O -term does not depend on either f , g or ϵ .

Proof. The conditional expectation is a ratio; the denominator is $\mathbb{P}[\Gamma_n \geq \epsilon]$ whereas the numerator

$$\int_0^\infty \mathbb{P}[\Lambda_n \geq \lambda, \Gamma_n \geq \epsilon] d\lambda.$$

Associate with the pair (Λ, Γ) its function pair (f, g) and, for any pair (λ, ϵ) of positive real numbers, consider the bounded convex subset already described in (4.15)

$$\underline{\Delta}_{f,g}(\lambda, \epsilon) := \{(x, y) \in \mathcal{R} : f(x, y) \geq \lambda, g(x, y) \geq \epsilon\}.$$

We have remarked in Section 3 that a slight extension of Proposition 4.1 entails the equality

$$\mathbb{P}[\Lambda_n \geq \lambda, \Gamma_n \geq \epsilon] = \widehat{R}_n(2\omega, \underline{\Delta}_{f,g}(\lambda, \epsilon)).$$

Moreover, with the convexity of the domain $\underline{\Delta}_{f,g}(\lambda, \epsilon) \subset \mathcal{R}$, Proposition 4.3 applies, yielding

$$\mathbb{P}[\Lambda_n \geq \lambda, \Gamma_n \geq \epsilon] = \frac{12}{\pi^2} I[\omega, \underline{\Delta}_{f,g}(\lambda, \epsilon)] + O\left(\frac{1}{n}\right).$$

Now we integrate on λ , noticing that we need only integrate from 0 to $B_{f|g}(\epsilon)$

$$\int_0^\infty \mathbb{P}[\Lambda_n \geq \lambda, \Gamma_n \geq \epsilon] d\lambda = \frac{12}{\pi^2} \int_0^\infty I[\omega, \underline{\Delta}_{f,g}(\lambda, \epsilon)] d\lambda + B_{f|g}(\epsilon) O\left(\frac{1}{n}\right).$$

We are led to the integral of ω on the domain of \mathbb{R}^3 defined by

$$\{(x, y, \lambda) \in \mathcal{R} \times \mathbb{R}_{\geq 0} : f(x, y) \geq \lambda, g(x, y) \geq \epsilon\}$$

Parameter Γ	Bound for S	$\mathbb{E}_\psi[f_S f_\Gamma \geq \epsilon(n)] \mathbb{P}_\psi[f_\Gamma \geq \epsilon(n)]$
ρ	$S \leq 2 + 1/\rho \implies B_{f_S f_\rho}(\epsilon) = O(1/\epsilon)$	$A \log(\epsilon(n)) + 1 - A\epsilon(n) \log \epsilon(n) $
μ	$S \leq 1 + 1/\mu \implies B_{f_S f_\mu}(\epsilon) = O(1/\epsilon)$	$A \log \epsilon(n) + \frac{A}{1 - \epsilon(n)}\epsilon(n) \log \epsilon(n) $
ν	$S \leq 1 + 2/\nu \implies B_{f_S f_\nu}(\epsilon) = O(1/\epsilon)$	$A \log \epsilon(n) + 1$

Figure 4.8: In the second column, the bounds for S for each parameter $\Gamma \in \{\rho, \mu, \nu\}$. In the third column, the values of the product $\mathbb{E}_\psi[f_S|f_\Gamma \geq \epsilon(n)] \mathbb{P}_\psi[f_\Gamma \geq \epsilon(n)]$ needed to apply Theorem 4.5. The constant A is $12/\pi^2$.

We interchange the summation, and we first integrate with respect to λ (which provides the term $f(x, y)$), and obtain

$$\begin{aligned} \int_0^\infty I[\omega, \underline{\Delta}_{f,g}(\lambda, \epsilon)] d\lambda &= \iint_{\substack{(x,y) \in \mathcal{R}, \\ g(x,y) \geq \epsilon}} \omega(x, y) \cdot f(x, y) dx dy, \\ &= \frac{\pi^2}{12} \mathbb{E}_\psi[f|g \geq \epsilon] \cdot \mathbb{P}_\psi[g \geq \epsilon]. \quad \blacksquare \end{aligned}$$

4.4.4 Conditional expectation of the recurrence quotient. Proof of Theorem 4.3.

We will prove here a stronger version of Theorem 4.3, where the remainder terms are more precise.

Theorem 4.6. *Consider a parameter $\Gamma \in \{\rho, \mu, \nu\}$ defined in (4.1) and (4.2), and a sequence $n \mapsto \epsilon(n)$ which tends to 0. Then the conditional expectation of the recurrence quotient S_n with respect to the event $[\Gamma_n \geq \epsilon(n)]$ satisfies*

$$\begin{aligned} \mathbb{E} \left[S_n \mid \Gamma_n \geq \epsilon(n) \right] &= \frac{12}{\pi^2} |\log \epsilon(n)| + C(\Gamma) \\ &+ O \left(\frac{1}{\epsilon(n)n} + \epsilon(n) |\log \epsilon(n)|^2 \right). \end{aligned} \quad (4.30)$$

Moreover, the constants $C(\Gamma)$ satisfy

$$C(\nu) = +1, \quad C(\mu) = 0, \quad C(\rho) = +1.$$

Proof. The proof is an application of Theorem 4.5. First, a direct computation with Theorem 4.1 shows that if Γ is one of the \mathcal{Q} -functions ρ, μ or ν , the following estimates hold

$$\mathbb{P}[\Gamma_n \geq \epsilon(n)] = 1 + O(\epsilon(n) + 1/n).$$

Along with the bounds and the integrals provided in Figure 4.8, this implies the result. \blacksquare

Now, Theorem 4.3 is an immediate application of Theorem 4.6. Indeed, in the case when $\epsilon(n) = \Omega\left(\frac{1}{n \log n}\right)$, the remainder term in (4.30) is $o(|\log \epsilon(n)|)$.

4.5 Other applications and extensions

The principles employed above do not limit themselves to \mathcal{Q} -functions, here we present other interesting applications; first to the study of the number of continuants in an interval $[n, cn)$ with fixed $c > 1$, and then to study the minimal distance Γ from Definition 3.11. We also expand on the concept of \mathcal{Q} -functions.

4.5.1 Number of continuants in an interval.

There is an interesting application of the previous ideas, that counts the number of terms of the sequence $k \mapsto q_k(\alpha)$ that belongs to the interval $[n, cn]$, for some fixed $c > 1$. We thus study the function

$$(\alpha, n) \mapsto T(\alpha, n) := \sum_{k \geq 0} \llbracket q_k(\alpha) \in [n, cn] \rrbracket.$$

Proposition 4.4. *Consider the Lévy constant $\kappa := \exp(\pi^2/(12 \log 2))$. Then the mean number of continuants in the interval $[n, \kappa n]$ tends to 1 as $n \rightarrow \infty$*

Proof. Even if T is not a \mathcal{Q} -function, its expectation $\mathbb{E}[T_n]$ is expressed as a Riemann sum of the function 2ω , in a domain \mathcal{T}_c . However the domain \mathcal{T}_c is not a subset of the rectangle \mathcal{R} . We have indeed

$$\begin{aligned} \mathbb{E}[T_n] &= \int_0^1 T(\alpha, n) d\alpha = \int_0^1 d\alpha \sum_k \llbracket q_k(\alpha) \in [n, cn] \rrbracket \\ &= \sum_k \sum_{\mathbf{m} \in \mathbb{N}^k} |I_k(\mathbf{m})| \llbracket q_k(\mathbf{m}) \in [n, cn] \rrbracket \\ &= \frac{1}{n^2} \sum_k \sum_{\mathbf{m} \in \mathbb{N}^k} \omega\left(\frac{q_{k-1}(\mathbf{m})}{n}, \frac{q_k(\mathbf{m})}{n}\right) \llbracket \frac{q_k(\mathbf{m})}{n} \in [1, c] \rrbracket \\ &= 2 \sum_{\substack{(a,b) \in \mathbb{Z}^2 \\ \gcd(a,b)=1}} \omega\left(\frac{a}{n}, \frac{b}{n}\right) \llbracket \frac{a}{n} \leq \frac{b}{n}, 1 \leq \frac{b}{n} \leq c \rrbracket \\ &= \widehat{R}_n(2\omega, \mathcal{T}_c), \quad \text{with } \mathcal{T}_c = \{(x, y) : x \leq y, 1 \leq y \leq c\}. \end{aligned}$$

Even if \mathcal{T}_c is not a subset of \mathcal{R} , Proposition 4.3 applies, and the coprime Riemann series admits a limit equal to the integral

$$\frac{6}{\pi^2} I(2\omega, \mathcal{T}_c) = \frac{12 \log 2}{\pi^2} \log c. \quad \blacksquare$$

Interpretation. The expression in Proposition 4.4 is no coincidence. Given a complete interval dynamical system (see Definition 1.3) with shift map S , consider the inverse branches $h \in \mathcal{H}^k$ rather than the tuple of digits m_1, \dots, m_k . Pick a probability measure μ , which plays the role of the Lebesgue measure above.

Under good hypothesis, the author conjectures that the following limit should exist and equal

$$\lim_{n \rightarrow \infty} \sum_{k \geq 1} \sum_{h \in \mathcal{H}^k} \mu(h(\mathcal{I})) \llbracket n < \frac{1}{\mu(h(\mathcal{I}))} \leq \theta n \rrbracket = \frac{\log \theta}{H_\mu(S)},$$

where $H_\mu(S)$ is the entropy of our system with respect to the probability measure μ , see Section 1.2.7.

We give the philosophical intuition behind this conjecture. If the limit $g(\theta)$ exists for each $\theta > 0$, it should be of the form $g(\theta) = c \log \theta$ for a certain C , as it satisfies $g(\theta_1 \cdot \theta_2) = g(\theta_1) + g(\theta_2)$ and it is increasing. We expect the constant c to be as above because $\theta = \exp(H_\mu(S))$ is the candidate to make the expected number of continuants in $(n, \theta n]$ roughly equal to 1.

4.5.2 The smallest distance and \mathcal{Q} f -functions

Introduction Many functions that are not \mathcal{Q} -functions may *turn into* one after integration. This is the case of the smallest distance $\Gamma(\alpha, n)$ from Chapter 3, Definition 3.11, which constitutes an interesting building example. We shall then show also that $T^{k(\alpha, n)} \alpha$ is another example.

The average case of the smallest distance and its distribution has appeared several times in the literature over the years. First Friedman and Niven [FN59] considered the study of the first recurrence time (see Definition 3.13) on average, introducing several techniques involving the so-called Farey series. Then Kesten [Kes62] built on the work of Friedman and Niven to study the expected values of the smallest distance. It turns out that his proof can be simplified by explaining more deeply the relation between coprime Riemann sums and integrals (see Proposition 4.8 below). Here we shall show that we may achieve the same results, even in distribution, by extending the concepts of this chapter.

It is also worthwhile to point out the works by Knuth [Knu84] and Bosma, Jager and Wiedijk [BJW83]. Knuth has studied the related quantity $\theta_k(\alpha) := q_k(\alpha) \cdot |\alpha q_k(\alpha) - p_k(\alpha)|$ probabilistically, giving a limit density as $k \rightarrow \infty$. Bosma, Jager and Wiedijk, on the other hand, have proved similar results from an Ergodic perspective. We shall give more details on the matter when we state the result for the smallest distance.

Normalized smallest distance Clearly asking for the limit distribution of $\Gamma(\alpha, n)$ does not make much sense: it tends to 0 uniformly, as we know that $\Gamma(\alpha, n) \leq 1/(n+1)$. The inequality follows from the fact that we have partitioned the circle of radius 1 into $n+1$ intervals. The right function would be the *normalized* smallest distance:

$$n\Gamma(\alpha, n) = \Gamma(\alpha, n)/(1/n).$$

Let us recall an important proposition from Chapter 3, which explains the relationship between the smallest distance $\Gamma(\alpha, n)$ and the convergents of α .

Proposition 3.15 *Let $\alpha \in [0, 1)$ and $n \in \mathbb{N}$. Let k be the unique non-negative integer such that $q_k(\alpha) \leq n < q_{k+1}(\alpha)$. The smallest distance $\Gamma(\alpha, n)$ is then given by*

$$\Gamma(\alpha, n) = M_k(\alpha) := |\alpha q_k - p_k|.$$

We recall Proposition 1.4, which tells us that

$$\frac{1}{q_k + q_{k+1}} \leq |\alpha q_k - p_k| \leq \frac{1}{q_{k+1}},$$

and Proposition 3.15. From these we derive that for $q_k(\alpha) \leq n < q_{k+1}(\alpha)$

$$\frac{q_k}{2q_{k+1}} < n |\alpha q_k - p_k| < 1.$$

Thus we derive

Proposition 4.5. *For almost every $\alpha \in (0, 1)$, the normalized smallest distance $n\Gamma(\alpha, n)$ satisfies the bounds*

$$\frac{1}{2} \leq \limsup_{n \rightarrow \infty} n\Gamma(\alpha, n) \leq 1. \quad (4.31)$$

Probabilistic study of the normalized smallest distance We shall prove the following Theorem giving the limit distribution of the normalized smallest distance.

Theorem 4.7. *The normalized smallest distance $n \cdot \Gamma(\alpha, n)$ has a limit distribution*

$$\lim_{n \rightarrow \infty} \mathbb{P}(\alpha : n \cdot \Gamma(\alpha, n) \leq \beta) = \frac{12}{\pi^2} (\beta - G(\beta)), \quad (4.32)$$

where $G(\beta)$ is 0 for $\beta \leq 1/2$, and for $1/2 \leq \beta \leq 1$ we have

$$G(\beta) = -\frac{1}{2} (\log \beta)^2 + (\beta - 1) \log \frac{\beta}{1 - \beta} - \text{Li}_2(\beta) + 2\beta - 1 + \frac{\pi^2}{12},$$

where we recall $Li_2(x) = -\int_0^x \frac{\log(1-u)}{u} du$ for $|x| \leq 1$.

This Theorem was first proved by Kesten in [Kes62] but by using significantly different methods.

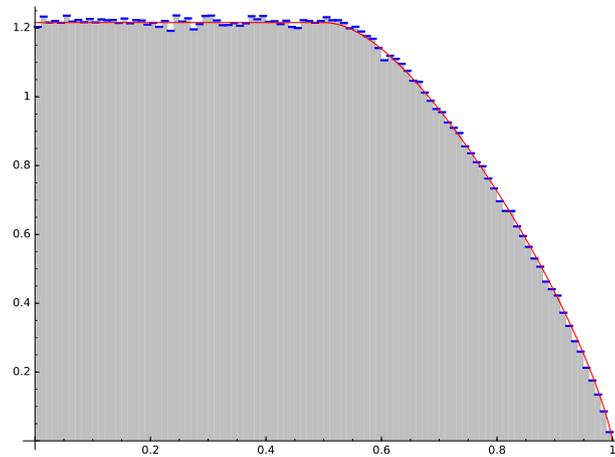


Figure 4.9: Limiting density for $n\Gamma(\alpha, n)$ in red and an escaled experimental histogram in blue.

As a consequence we have the following result for the expectation.

Theorem 4.8. *The expected value of the normalized smallest distance $n \cdot \Gamma(\alpha, n)$ satisfies*

$$\lim_{n \rightarrow \infty} \mathbb{E}[n \cdot \Gamma(\alpha, n)] = \frac{\log 2}{\zeta(2)}. \quad (4.33)$$

Theorem 4.7 has an analog in [BJW83] which we cite here

Theorem 4.9. *Let $b > 0$. For almost every $\alpha \in \mathcal{I}$ we have*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \#\{j : j \leq n, q_j(\alpha) | \alpha q_j(\alpha) - p_j(\alpha) | \leq b\} = \begin{cases} \frac{b}{\log 2} & \text{if } 0 \leq b \leq \frac{1}{2}, \\ \frac{1}{\log 2} (-b + \log(2b) + 1) & \text{if } \frac{1}{2} \leq b \leq 1. \end{cases}$$

Note that in Theorem 4.9 only the n 's that are continuants are considered (i.e., $n = q_j(\alpha)$) and the result is of Ergodic nature, giving the average behavior of the orbit of a fixed α , almost every $\alpha \in \mathcal{I}$. The proof of Theorem 4.9 rewrites the quantity $|\alpha q_j(\alpha) - p_j(\alpha)|$ in terms of $T^j(\alpha)$ and $[m_j(\alpha), \dots, m_1(\alpha)] = q_{j-1}(\alpha)/q_j(\alpha)$, to then apply the Ergodicity of the natural extension $(x, y) \mapsto (T_g(x), 1/(m(x) + y))$.

Knuth in [Knu84] obtains this same limit distribution for $q_k(\alpha) | \alpha q_k(\alpha) - p_k(\alpha) |$, but for k fixed, when considering $\alpha \in \mathcal{I}$ drawn uniformly at random from \mathcal{I} . His strategy, however, is rather different, as he considered methods close to those of Chapter 5.

In order to study the smallest distance, we introduce a new concept.

Definition 4.1 (\mathcal{Q} \int -function). A function $\Lambda : [0, 1] \times \mathbb{N} \rightarrow \mathbb{R}_{\geq 0}$ is said to be a \mathcal{Q} \int -function if and only if for each positive integer n and prefix $\mathbf{m} \in \mathbb{N}^k$ such that $q_{k-1}(\mathbf{m}) \leq n < q_k(\mathbf{m})$, the integral of $\alpha \mapsto \Lambda(\alpha, n)$ on the fundamental interval $\mathcal{I}_{\mathbf{m}}$ associated with \mathbf{m} can be written as

$$\int_{\mathcal{I}_{\mathbf{m}}} \Lambda(\alpha, n) d\alpha = \frac{1}{n^2} f\left(\frac{q_{k-1}(\mathbf{m})}{n}, \frac{q_k(\mathbf{m})}{n}\right), \quad (4.34)$$

for some function $f : (0, 1] \times [1, \infty) \rightarrow \mathbb{R}_{\geq 0}$, independent from both \mathbf{m} and n . In such a case we say that f is the function associated with the \mathcal{Q} \int -function Λ .

Thus \mathcal{Q} \int -functions have averages over the fundamental intervals that are actually \mathcal{Q} -functions. It comes as no surprise that we will be interested in the expected values of \mathcal{Q} \int -functions, even more so due to the following proposition which states that the indicator $\llbracket n \Gamma(\alpha, n) \leq b \rrbracket$, for fixed $b > 0$, is a \mathcal{Q} \int -function.

Proposition 4.6. *Fix $b > 0$. The indicator function $(\alpha, n) \mapsto \llbracket n \Gamma(\alpha, n) \leq b \rrbracket$, is a \mathcal{Q} \int -function, explicitly*

$$\int_{\mathcal{I}_m} \llbracket n \Gamma(\alpha, n) \leq b \rrbracket d\alpha = \begin{cases} \frac{1}{n^2} \frac{1}{y(x+y)} & \text{if } 1 \leq by \\ \frac{1}{n^2} \left(\frac{b}{x} - \frac{1}{x(x+y)} \right) & \text{if } by < 1 \leq b(x+y) \\ 0 & \text{otherwise,} \end{cases}$$

where $x = \frac{q_{k-1}(\mathbf{m})}{n}$ and $y = \frac{q_k(\mathbf{m})}{n}$, so that $q_{k-1}(\mathbf{m}) \leq n < q_k(\mathbf{m})$ implies $0 < x \leq 1 < y$.

Proof. We observe that if k is even, then

$$\mathcal{I}_m = \left[\frac{p_k(\mathbf{m})}{q_k(\mathbf{m})}, \frac{p_k(\mathbf{m}) + p_{k-1}(\mathbf{m})}{q_k(\mathbf{m}) + q_{k-1}(\mathbf{m})} \right],$$

while, if k is odd,

$$\mathcal{I}_m = \left[\frac{p_k(\mathbf{m}) + p_{k-1}(\mathbf{m})}{q_k(\mathbf{m}) + q_{k-1}(\mathbf{m})}, \frac{p_k(\mathbf{m})}{q_k(\mathbf{m})} \right].$$

Here

$$n \Gamma(\alpha, n) \leq b \iff \alpha \leq \frac{p_{k-1}}{q_{k-1}} + \frac{b}{n q_{k-1}}, \quad \text{and } \alpha \geq \frac{p_{k-1}}{q_{k-1}} - \frac{b}{n q_{k-1}}.$$

The first possibility holds trivially when k is even, because then $\alpha \leq \frac{p_{k-1}}{q_{k-1}}$, and similarly the latter one holds trivially occur when k is odd because then $\alpha \geq \frac{p_{k-1}}{q_{k-1}}$.

Case k even. Let us start by supposing that k is even. Then

$$\mathcal{I}_m = \left[\frac{p_k(\mathbf{m})}{q_k(\mathbf{m})}, \frac{p_k(\mathbf{m}) + p_{k-1}(\mathbf{m})}{q_k(\mathbf{m}) + q_{k-1}(\mathbf{m})} \right],$$

and we must compute the length of the interval

$$\mathcal{J}_{m,n} := \mathcal{I}_m \cap \left(\frac{p_{k-1}}{q_{k-1}} - \frac{b}{n q_{k-1}}, \frac{p_{k-1}}{q_{k-1}} + \frac{b}{n q_{k-1}} \right).$$

Since $\alpha \leq \frac{p_{k-1}}{q_{k-1}}$ for k even, we need only check what happens with the left border $\frac{p_{k-1}}{q_{k-1}} - \frac{b}{n q_{k-1}}$.

We compare first $\frac{p_{k-1}}{q_{k-1}} - \frac{b}{n q_{k-1}}$ with $\frac{p_k(\mathbf{m}) + p_{k-1}(\mathbf{m})}{q_k(\mathbf{m}) + q_{k-1}(\mathbf{m})}$. Observe that

$$\frac{p_{k-1}(\mathbf{m})}{q_{k-1}(\mathbf{m})} - \frac{p_k(\mathbf{m}) + p_{k-1}(\mathbf{m})}{q_k(\mathbf{m}) + q_{k-1}(\mathbf{m})} = \frac{1}{q_{k-1}(q_k + q_{k-1})}$$

thus

$$\frac{p_{k-1}}{q_{k-1}} - \frac{b}{n q_{k-1}} \leq \frac{p_k(\mathbf{m}) + p_{k-1}(\mathbf{m})}{q_k(\mathbf{m}) + q_{k-1}(\mathbf{m})} \iff \frac{1}{q_{k-1}(q_{k-1} + q_k)} \leq \frac{b}{n q_{k-1}} \iff n \leq b(1 + x_k) q_k$$

But $\frac{p_{k-1}}{q_{k-1}} - \frac{b}{n q_{k-1}}$ may also exceed the left-hand border of \mathcal{I}_m when

$$\frac{p_{k-1}}{q_{k-1}} - \frac{b}{n q_{k-1}} \leq \frac{p_k}{q_k} \iff \frac{1}{q_k} \leq \frac{b}{n} \iff n \leq b q_k.$$

Thus, for $b q_k < n \leq b(q_k + q_{k-1})$ i.e. $by < 1 \leq b(x+y)$, we have

$$\mathcal{J}_{m,n} = \left[\frac{p_{k-1}}{q_{k-1}} - \frac{b}{n q_{k-1}}, \frac{p_k(\mathbf{m}) + p_{k-1}(\mathbf{m})}{q_k(\mathbf{m}) + q_{k-1}(\mathbf{m})} \right], \quad |\mathcal{J}_{m,n}| = \frac{1}{n^2} \left(\frac{b}{x} - \frac{1}{x(x+y)} \right),$$

while, if $n > b(1 + x_k)q_k$ we have the empty set $\mathcal{J}_{\mathbf{m},n} = \emptyset$, and for $n \leq bq_k$ we get the original interval $\mathcal{J}_{\mathbf{m},n} = \mathcal{I}_{\mathbf{m}}$ in which case $|\mathcal{J}_{\mathbf{m},n}| = \frac{1}{n^2} \frac{1}{y(x+y)}$.

The analysis gives the same results in the end when k is odd, thus

$$\int_{\mathcal{I}_{\mathbf{m}}} \llbracket n \Gamma(\alpha, n) \leq b \rrbracket d\alpha = \begin{cases} \frac{1}{n^2} \frac{1}{y(x+y)} & \text{if } 1 \leq by \\ \frac{1}{n^2} \left(\frac{b}{x} - \frac{1}{x(x+y)} \right) & \text{if } by < 1 \leq b(x+y) \\ 0 & \text{otherwise} \end{cases},$$

where $x = \frac{q_{k-1}(\mathbf{m})}{n}$ and $y = \frac{q_k(\mathbf{m})}{n}$, so that $q_{k-1}(\mathbf{m}) \leq n < q_k(\mathbf{m})$ implies $0 < x \leq 1 < y$. ■

Here is an analogous to Proposition 4.1 for \mathcal{Q} f -functions. Notice that here the density $\omega(x, y)$ does not necessarily intervene, as we actually see from Proposition 4.6.

Proposition 4.7. *Let Λ be a \mathcal{Q} f -function associated with a function $f \geq 0$. Then its expected value is given by*

$$\mathbb{E}_{\alpha}[\Lambda(\alpha, n)] = \frac{2}{n^2} \sum_{\substack{a \leq n < b, \\ \gcd(a,b)=1}} f\left(\frac{a}{n}, \frac{b}{n}\right). \tag{4.35}$$

Proof. For a \mathcal{Q} f -function Λ , associated with f , the expected value takes on the form

$$\mathbb{E}[\Lambda_n] = \sum_{k \geq 1} \sum_{\mathbf{m} \in \mathbb{N}^k} \llbracket q_{k-1}(\mathbf{m}) \leq n < q_k(\mathbf{m}) \rrbracket \int_{\mathcal{I}_{\mathbf{m}}} \Lambda(\alpha, n) d\alpha \tag{4.36}$$

which by definition equals

$$\mathbb{E}[\Lambda_n] = \frac{1}{n^2} \sum_{k \geq 1} \sum_{\mathbf{m} \in \mathbb{N}^k} \llbracket q_{k-1}(\mathbf{m}) \leq n < q_k(\mathbf{m}) \rrbracket f\left(\frac{q_{k-1}(\mathbf{m})}{n}, \frac{q_k(\mathbf{m})}{n}\right). \tag{4.37}$$

Now the rest follows as before by noticing that the pairs $(q_{k-1}(\mathbf{m}), q_k(\mathbf{m}))$ traverse all the pairs of coprime integers (a, b) with $a \leq b$ twice. ■

Observation 4.1. Observe that for the normalized smallest distance, the integrals over the fundamental intervals

$$\int_{\mathcal{I}_{\mathbf{m}}} \llbracket n \Gamma(\alpha, n) \leq b \rrbracket d\alpha$$

can be decomposed, for any $b > 0$, as a sum of homogeneous functions of degree 2 and maybe also 1. So it is not trivial to apply Proposition 4.3 at once.

Here there are two ways forward. In fact, if we admit the case of homogeneous functions of degree -1 in Proposition 4.3, we get a bound of order $O((\log n)/n)$. We feel, however, that it is more illuminating to step back and explain what happens on a much wider context.

The general intuition is explained by the following Proposition 4.8. Given a “filtering” or “weighting” function δ with a natural density $C > 0$ (see the statement of the proposition), any perturbation of the Riemann-integral by filtering (or weighting) the steps by δ just gives the same Riemann integral multiplied by C . For our application we just need $\delta(a, b) = \llbracket \gcd(a, b) = 1 \rrbracket$, which has density $C = 6/\pi^2$ (see Proposition 2.10), but it is important to give the whole intuition behind it.

Proposition 4.8. *Let $f: [0, 1]^k \rightarrow \mathbb{R}$ be bounded, and Riemann integrable. Consider a bounded function $\delta: \mathbb{N}^k \rightarrow \mathbb{R}$ with a natural density*

$$\lim_{N_1, \dots, N_k \rightarrow \infty} \frac{1}{N_1 \dots N_k} \sum_{x_1 \leq N_1, \dots, x_k \leq N_k} \delta(x_1, \dots, x_k) = C,$$

for a certain constant $C > 0$. Then we have

$$\begin{aligned} \lim_{N_1, \dots, N_k \rightarrow \infty} \frac{1}{N_1 \dots N_k} \sum_{x_1 \leq N_1, \dots, x_k \leq N_k} f\left(\frac{x_1}{N_1}, \dots, \frac{x_k}{N_k}\right) \delta(x_1, \dots, x_k) \\ = C \int \dots \int_{[0,1]^k} f(x_1, \dots, x_k) dx_1 \dots dx_k. \end{aligned}$$

Proof. The proof of this Lemma is the archetypical proof in Measure Theory; we prove that a certain class of “easy functions” satisfies the statement, and then move on to prove that these functions approximate well our “target functions”.

The result is trivially true when we pick $f = \mathbf{1}_{(a_1, b_1] \times \dots \times (a_k, b_k]}$. Indeed, by inclusion-exclusion (see e.g., [Sta97, pp.64-65])

$$\sum_{a_1 < \frac{x_1}{N_1} \leq b_1, \dots, a_k < \frac{x_k}{N_k} \leq b_k} \delta(x_1, \dots, x_k) = \sum_{S \subset [k]} (-1)^{|S|} \sum_{\substack{\frac{x_i}{N_i} \leq a_i \text{ for } i \in S, \\ \frac{x_i}{N_i} \leq b_i \text{ for } i \notin S}} \delta(x_1, \dots, x_k), \quad (4.38)$$

because here the properties (sets) that may fail to be satisfied are $a_i < \frac{x_i}{N_i}$ for $i = 1, \dots, k$. Here we observe that

$$\lim_{N_1, \dots, N_k \rightarrow \infty} \frac{1}{N_1 \dots N_k} \sum_{\substack{\frac{x_i}{N_i} \leq a_i \text{ for } i \in S, \\ \frac{x_i}{N_i} \leq b_i \text{ for } i \notin S}} \delta(x_1, \dots, x_k) = C \left(\prod_{i \in S} a_i \right) \cdot \left(\prod_{i \notin S} b_i \right),$$

therefore

$$\lim_{N_1, \dots, N_k \rightarrow \infty} \frac{1}{N_1 \dots N_k} \sum_{a_1 < \frac{x_1}{N_1} \leq b_1, \dots, a_k < \frac{x_k}{N_k} \leq b_k} \delta(x_1, \dots, x_k) = C \sum_{S \subset [k]} (-1)^{|S|} \left(\prod_{i \in S} a_i \right) \cdot \left(\prod_{i \notin S} b_i \right), \quad (4.39)$$

$$= C (b_1 - a_1) \dots (b_k - a_k). \quad (4.40)$$

As a consequence, the result follows at once for step functions (the finite linear combinations of the characteristic functions of rectangles), we have our “easy functions”. We remark that what happens in the borders of the intervals defining the rectangle is not very important, since these cases do not eventually contribute to the limits because $|\delta|$ is bounded.

Let us denote

$$S_f(N_1, \dots, N_k) := \frac{1}{N_1 \dots N_k} \sum_{x_1 \leq N_1, \dots, x_k \leq N_k} f\left(\frac{x_1}{N_1}, \dots, \frac{x_k}{N_k}\right) \delta(x_1, \dots, x_k)$$

for the sake of brevity.

So our “easy functions” are the step functions, now our “target functions” will be all of the Riemann integrable functions. Indeed, the set of step functions is dense in the set of Riemann integrable functions under the norm of L^1 , hence if f is Riemann integrable and $\epsilon > 0$ is arbitrary, there is a step unction g such that $\|f - g\|_1 \leq \epsilon$ and we explain why this, together with the Riemann integrability of $|f - g|$, actually imply the conclusion.

Let K be such that $|\varphi| \leq K$, as it is bounded by hypothesis, then

$$\begin{aligned} & \left| S_f(N_1, \dots, N_k) - S_g(N_1, \dots, N_k) \right| \\ & \leq \frac{1}{N_1 \dots N_k} \sum_{x_1 \leq N_1, \dots, x_k \leq N_k} \left| f\left(\frac{x_1}{N_1}, \dots, \frac{x_k}{N_k}\right) - g\left(\frac{x_1}{N_1}, \dots, \frac{x_k}{N_k}\right) \right| |\delta(x_1, \dots, x_k)| \\ & \leq \frac{K}{N_1 \dots N_k} \sum_{x_1 \leq N_1, \dots, x_k \leq N_k} \left| f\left(\frac{x_1}{N_1}, \dots, \frac{x_k}{N_k}\right) - g\left(\frac{x_1}{N_1}, \dots, \frac{x_k}{N_k}\right) \right|, \end{aligned}$$

and observe here that the right-hand side tends to $K \int |f - g|$, because $|f - g|$ is Riemann integrable.

It follows that for, say $N_1, \dots, N_k \geq M$ we get $|S_f - S_g| \leq 2K\epsilon$. By making M larger if necessary, we may assume $|S_g - \int g| \leq \epsilon$ as we already know the result to hold for g , thus by the triangle inequality

$$|S_f - \int f| \leq |S_f - S_g| + |S_g - \int g| + \left| \int g - \int f \right| \leq (2K + 2)\epsilon,$$

provided that $N_1, \dots, N_k \geq M$, which proves the result because $\epsilon > 0$ was arbitrary. \blacksquare

With Proposition 4.8 at hand, we now explain how to get the distributions from before.

Theorem 4.10. Consider $f \in C^1(\mathcal{R}, \mathbb{R})$, and let $\Lambda(\alpha, n)$ be the \mathcal{Q} -function associated with f , i.e.,

$$\Lambda(\alpha, n) := f\left(\frac{q_{k-1}}{n}, \frac{q_k}{n}\right)$$

for $q_{k-1}(\alpha) \leq n < q_k$. Assume that $\nabla f(x, y) \neq 0$ for almost-every $(x, y) \in \mathcal{R}$, then the limiting distribution of $\Lambda(\alpha, n)$ as $n \rightarrow \infty$ is given by

$$\lim_{n \rightarrow \infty} \mathbb{P}(\alpha : \Lambda(\alpha, n) \leq \lambda) = \frac{12}{\pi^2} \iint_{\Delta_f(\lambda)} \omega(x, y) dx dy,$$

for every $\lambda \geq 0$.

Remark. Observe that the hypothesis hold for the recurrence quotient, where we have $f_S(x, y) = 1 + x + y$, or the relative position, where we have $f_\mu(x, y) = \frac{1-x}{y-x}$. In these cases, however, the fact that $(x, y) \mapsto \mathbf{1}_{f(x, y) \leq \lambda}$ is Riemann-integrable is actually simple enough since the border $f(x, y) = \lambda$ is a line.

Proof. We give a sketch of the proof. There are two parts to it, after noticing that (see Proposition 4.1)

$$\mathbb{P}(\alpha : \Lambda(\alpha, n) \leq \lambda) = \frac{2}{n^2} \sum_{(a, b) \in \mathbb{N}^2} \omega\left(\frac{a}{n}, \frac{b}{n}\right) \left[\left[f\left(\frac{a}{n}, \frac{b}{n}\right) \leq \lambda \right] \left[\frac{a}{n} < \frac{b}{n} \right] \delta(a, b), \right.$$

where $\delta(a, b) = \llbracket \gcd(a, b) = 1 \rrbracket$.

In order to apply Proposition 4.8 we first need to show that $(x, y) \mapsto \llbracket f(x, y) \leq \lambda \rrbracket$ is Riemann-integrable, and then extend the result to the unbounded \mathcal{R} (the result works for the bounded rectangle $[0, 1] \times [0, 1]$).

First we explain why $(x, y) \mapsto \llbracket f(x, y) \leq \lambda \rrbracket$ is Riemann-integrable. This follows from the fact that the points of discontinuity belong to $f^{-1}(\lambda)$, which is a null set, thanks to the inverse function theorem (we may omit the points where $\nabla f = 0$ as this is already a null set, and we show the rest is countable).

Now Proposition 4.8 applies only to bounded rectangles (by re-scaling), but the fast convergence of the sums $\frac{1}{b(a+b)}$ over (a, b) with $a \leq n < b$ gives the result by taking larger and larger rectangles. \blacksquare

We mention here that when one drops the condition that Λ be an \mathcal{LQ} -function, then we cannot really give guarantees with regard to the convergence to the distribution. This is why we gave a different proof in our paper for ANALCO [RV17] with precise error terms, as these are much needed for the results regarding the convergence in density and the conditional expectations.

By using Proposition 4.8 we complete the proof of Theorem 4.7.

Proof. [of Theorem 4.7] We are now ready to prove Theorem 4.7 by using Proposition 4.6. Indeed, fixed $b > 0$ we divide into two cases: (i) $1/b \leq y$ and (ii) $y \leq 1/b$.

When we are in case (ii) we have that the domain (x, y) is contained in the bounded rectangle $[0, 1] \times [1, 1/b]$. Hence we may apply Proposition 4.8 for case (ii) after re-scaling the y -axis. For the case (i) $1/b \leq y$ we note that the formula given by Proposition 4.6 falls into the case of Proposition 4.3. Hence it follows that in the limit we get $12/\pi^2$ times the integral as before:

$$\lim_{n \rightarrow \infty} \mathbb{P}(\alpha : n \cdot \Gamma(\alpha, n) \leq b) = \frac{12}{\pi^2} \iint_{\mathcal{R}} f_b(x, y) dx dy,$$

where

$$f_b(x, y) = \begin{cases} \frac{1}{n^2} \frac{1}{y(x+y)} & \text{if } 1 \leq by \\ \frac{1}{n^2} \left(\frac{b}{x} - \frac{1}{x(x+y)} \right) & \text{if } by < 1 \leq b(x+y) \\ 0 & \text{otherwise} \end{cases}.$$

Thus for $b \leq 1/2$

$$\lim_{n \rightarrow \infty} \mathbb{P}(\alpha : n \cdot \Gamma(\alpha, n) \leq b) = \frac{12}{\pi^2} \int_0^1 \left(\int_{1/b-x}^{1/b} \left(\frac{b}{x} - \frac{1}{x(x+y)} \right) dy + \int_{1/b}^{\infty} \frac{dy}{y(x+y)} \right) dx,$$

since then $1/b - x \geq 1$. It is then a computation to find that this reduces to $\frac{12}{\pi^2} b$.

For $b > 1/2$ it may hold that $1/b - x < 1$ depending on x . Thus subtracting these cases

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(\alpha : n \cdot \Gamma(\alpha, n) \leq b) &= \frac{12}{\pi^2} b \\ &\quad - \frac{12}{\pi^2} \int_{1/b-1}^1 \left(\int_{1/b-x}^1 \left(\frac{b}{x} - \frac{1}{x(x+y)} \right) dy \right) dx, \end{aligned}$$

for $b > 1/2$. Computing the integral gives the result. ■

4.5.3 Independence from the initial distribution

Introduction Thus far we have only considered the uniform (Lebesgue) distribution on the interval $\mathcal{I} = [0, 1]$. It is more generally true that as long as our distribution have a density with respect to the Lebesgue measure (i.e., that our measure μ is absolutely continuous with respect to the Lebesgue measure λ_{Leb} , in symbols $\mu \ll \lambda_{\text{Leb}}$) the limit in Theorem 4.1 will remain the exact same. We do not give guarantees with regard to the convergence speeds, and these may vary according to the nature of the density $g(x) = \frac{d\mu}{d\lambda_{\text{Leb}}}(x)$.

Theorem 4.11 (Independence from the initial distribution). *Consider a probability measure μ that is absolutely continuous with respect to the Lebesgue measure λ_{Leb} , in symbols this reads $\mu \ll \lambda_{\text{Leb}}$. Then, for each fixed $\lambda \in \mathbb{R}$, the limit*

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\mu}(S_n \leq \lambda)$$

exists and is independent from the choice of $\mu \ll \lambda_{\text{Leb}}$.

To motivate the proof, let us consider a density $g \in C^1(\mathcal{I}, \mathbb{R})$. We note that for a fundamental interval $J := \mathcal{I}_{m_1, \dots, m_k}$ we have the following estimate

$$\mu(J) = \int_J g(x) dx = g\left(\frac{p_k}{q_k}\right) |J| + O(|J|^2),$$

by the mean-value Theorem, as $p_k(\mathbf{m})/q_k(\mathbf{m}) \in J$. Recalling that $|J| = 1/(q_k(q_k + q_{k-1}))$ we get

$$\mu(J) = \frac{1}{q_k^2} \frac{g\left(\frac{p_k}{q_k}\right)}{1 + \frac{q_{k-1}}{q_k}} + O(|J|^2),$$

and here we write $\rho = \frac{q_{k-1}}{q_k}$ and $\hat{\rho} = p_k/q_k$. Clearly, if $g \equiv 1$ we get the same result from Theorem 4.10. We will prove that we may indeed get rid of this factor $g(p_k/q_k)$ by showing that p_k/q_k is asymptotically independent (on average) to q_{k-1}/q_k . Then the factor $g(x)$ integrates separately to 1 and we get the result.

The independence of p_k/q_k and q_{k-1}/q_k . We recall that by the mirror property (1.14) we have

$$\rho = [m_k, m_{k-1}, \dots, m_1], \quad \hat{\rho} = [m_1, m_2, \dots, m_k].$$

This property is crucial. We explain intuitively why ρ and $\hat{\rho}$ should be independent. Indeed, given a sequence $\mathbf{m} = (m_1, m_2, \dots)$, the convergents $\hat{\rho} = [m_1, m_2, \dots, m_k]$ converges exponentially fast to $x = [m_1, m_2, \dots]$. Similarly $\rho = [m_k, m_{k-1}, \dots, m_1]$ is mainly determined by the first digits of m_k, m_{k-1}, \dots, m_1 , which present a kind of stationary behavior for almost every x . Thus we see that intuitively ρ and $\hat{\rho}$ should be independent from one another.

Even if we made a seemingly fundamental use of the Riemann sums with $x = q_{k-1}/n$ and $y = q_k/n$ before, this time it will prove significantly simpler to work with the parameters $\rho = q_{k-1}/q_k$, $\hat{\rho} = p_k/q_k$ and $y = q_k/n$. Note we now have three parameters, because the denominators p_k have not intervened in our expressions up to this point. In any case note that $x = \rho \cdot y$ so we may retrieve the expressions from before.

We present a proof that is algebraic in nature and depends fundamentally on the equivalent fact that $x \pmod q$ and $x^{-1} \pmod q$ are asymptotically independent on average. Indeed, we shall prove that this follows from the determinant equation (1.13) which implies the congruence $p_k q_{k-1} \equiv (-1)^{k+1} \pmod{q_k}$. We remark how to get rid of the sign ± 1 : the two signs will occur symmetrically, owing to the fact that each rational has exactly two continued fraction expansions (one of even and one of odd length). To the best of our knowledge, this remark proving that p_k/q_k and q_{k-1}/q_k behave somewhat independently is new, and we would also like to prove this dynamically, but we have not been able to do this yet.

The fact that the modular inverses $x \pmod q$ and $x^{-1} \pmod q$ are asymptotically independent on average might be intuitively clear to many people working in Number Theory, and in order to prove it in the form we want, we cite the survey [Shp12] of Shparlinski, which gives several applications of a precise form of independence inequalities for $x \pmod q$ and $x^{-1} \pmod q$.

For the proof we require [Shp12, Theorem 13] which we cite here. The proof of this Theorem reduces to the application of a bound by Estermann for Kloosterman sums.

$$K_m(r, s) = \sum_{\substack{x: 1 \leq x \leq m, \\ \gcd(x, m) = 1}} \exp\left(\frac{2\pi i}{m} (r x + s (x^{-1} \pmod m))\right),$$

are intuitively a measure of ‘‘correlation’’ between x and $x^{-1} \pmod m$, so it comes as no surprise that they be of use here.

Theorem 4.12 (From [Shp12], Theorem 13). *Let $\mathcal{X} = \{U + 1, \dots, U + X\}$, where $m > X \geq 1$ and $U \geq 0$ are arbitrary integers. Suppose that for every $x \in \mathcal{X}$ we are given a set $\mathcal{Y}_x = \{V_x + 1, \dots, V_x + Y\}$ where $m > Y \geq 1$ and $V_x \geq 0$ are arbitrary integers. Then for any integer $m \geq 1$ and a with $\gcd(a, m) = 1$, we have*

$$\sum_{\substack{(x, y) \in \mathcal{H}_{a, m}, \\ x \in \mathcal{X}, y \in \mathcal{Y}_x}} 1 = \frac{\varphi(m)}{m^2} XY + O(m^{1/2+o(1)}),$$

where $\mathcal{H}_{a, m} = \{(x, y) : xy \equiv a \pmod{m}\}$.

In other words, we know very well (with an error term) what happens over rectangles. It comes as no surprise then, following the previous result for linear combinations of characteristic functions of rectangles, that we have the following Theorem 4.13.

Theorem 4.13. *Let $f: [0, 1] \times [0, 1] \times [0, \infty) \rightarrow \mathbb{R}_{\geq 0}$ be locally Riemann-integrable and compactly supported (i.e., $f(\cdot, \cdot, y) = 0$ for large enough y). Then*

$$\lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{(x,y): \substack{1 \leq x \leq y, \\ \gcd(x,y)=1}} f\left(\frac{x}{y}, \frac{x^{-1} \bmod y}{y}, \frac{y}{N}\right) = \frac{6}{\pi^2} \iiint_{[0,1] \times [0,1] \times [0,\infty)} y f(\rho, \hat{\rho}, y) d\rho d\hat{\rho} dy.$$

It is important to remark that the sum on Theorem 4.13 is bidimensional, yet it yields a triple integral.

Proof. [Sketch] First, we may work on a compact domain $\mathcal{D} := [0, 1] \times [0, 1] \times [0, R]$ containing the support of f . Consider arbitrary step functions g and h with $g \geq f \geq h \geq 0$ on \mathcal{D} .

If the result holds for such arbitrary g and h , it is easy to see it also holds for f on \mathcal{D} (Riemann-integrability on \mathcal{D}). Thus it is enough to prove it for step functions; the functions that are linear combinations of characteristic functions of “rectangles” $[a, b] \times [c, d] \times [A, B]$ (maybe excluding the borders).

For a basis step function $f = \mathbf{1}_{[a,b] \times [c,d] \times [A,B]}$ the result is a direct calculation that follows from Theorem 4.12 and the approximation^c $\sum_{k=1}^N \varphi(k) \sim \frac{3}{\pi^2} N^2$. ■

Now we delve into the relation between modular inverses and the convergents. The formal relation is given by the following theorem.

Theorem 4.14. *Let $f: [0, 1] \times [0, 1] \times [0, \infty) \rightarrow \mathbb{R}_{\geq 0}$ then*

$$\begin{aligned} \sum_{\mathbf{m} \in \mathbb{N}^*} f\left(\frac{q_{k-1}(\mathbf{m})}{q_k(\mathbf{m})}, \frac{p_k(\mathbf{m})}{q_k(\mathbf{m})}, \frac{q_k(\mathbf{m})}{N}\right) &= \sum_{(a,b) \in \mathbb{N}^2: \substack{1 \leq a \leq b, \\ \gcd(a,b)=1}} \left(f\left(\frac{a}{b}, \frac{a^{-1} \bmod b}{b}, \frac{b}{N}\right) \right. \\ &\quad \left. + f\left(\frac{a}{b}, 1 - \frac{a^{-1} \bmod b}{b}, \frac{b}{N}\right) \right) \\ &\quad - f\left(1, 1, \frac{1}{N}\right). \end{aligned}$$

Proof. Let $\Upsilon: \{\mathbf{m} \in \mathbb{N}^+ : m_1 > 1\} \rightarrow \mathbb{N}^+$ be defined by $\mathbf{m} = (m_1, \dots, m_k) \mapsto (1, m_1 - 1, m_2, \dots, m_k)$, then we may write

$$\begin{aligned} \sum_{k \geq 1} \sum_{\mathbf{m} \in \mathbb{N}^k} f\left(\frac{q_{k-1}(\mathbf{m})}{q_k(\mathbf{m})}, \frac{p_k(\mathbf{m})}{q_k(\mathbf{m})}, \frac{q_k(\mathbf{m})}{N}\right) &= \sum_{k \geq 1} \sum_{\substack{\mathbf{m} \in \mathbb{N}^k, \\ m_1 > 1}} \left(f\left(\frac{q_{k-1}(\mathbf{m})}{q_k(\mathbf{m})}, \frac{p_k(\mathbf{m})}{q_k(\mathbf{m})}, \frac{q_k(\mathbf{m})}{N}\right) \right. \\ &\quad \left. + f\left(\frac{q_k(\Upsilon(\mathbf{m}))}{q_{k+1}(\Upsilon(\mathbf{m}))}, \frac{p_{k+1}(\Upsilon(\mathbf{m}))}{q_{k+1}(\Upsilon(\mathbf{m}))}, \frac{q_{k+1}(\Upsilon(\mathbf{m}))}{N}\right) \right) \\ &\quad + f\left(1, 1, \frac{1}{N}\right), \end{aligned}$$

as every (w_1, \dots, w_{k+1}) with $w_1 = 1$ is exactly $\Upsilon(w_2 + 1, \dots, w_{k+1})$. Notice the term $f(1, 1, \frac{1}{N})$ on the right-hand side that is due to the fact that for (1) we may not apply Υ .

Observe that the mirror property (1.14) reads $q_k(\Upsilon(\mathbf{m})) = q_{k-1}(\mathbf{m})$ and $q_{k+1}(\Upsilon(\mathbf{m})) = q_k(\mathbf{m})$. This follows from the equality $[m_1, \dots, m_k] = [m_1, \dots, m_k - 1, 1]$, and the mirror property (1.14).

^cRecall e.g., that the Dirichlet Generating Function of the Euler totient function φ is $\zeta(s-1)/\zeta(s)$, hence this follows from the Tauberian Theorem in Theorem 2.3. See Section 2.2 on DGFs for more details.

Let $\theta: \mathbb{N}^* \rightarrow \mathbb{N}^*$ be the mirror $(m_1, \dots, m_p) \mapsto (m_p, \dots, m_1)$, then we recall that the mirror property reads

$$\begin{pmatrix} p_{k-1}(\theta(\mathbf{m})) & p_k(\theta(\mathbf{m})) \\ q_{k-1}(\theta(\mathbf{m})) & q_k(\theta(\mathbf{m})) \end{pmatrix} = \begin{pmatrix} p_{k-1}(\mathbf{m}) & p_k(\mathbf{m}) \\ q_{k-1}(\mathbf{m}) & q_k(\mathbf{m}) \end{pmatrix}^T. \quad (4.41)$$

Let us continue. Due to the change of parity in the depth

$$\{p_k(\mathbf{m}), p_{k+1}(\Upsilon(\mathbf{m}))\} = \{q_{k-1}^{-1} \bmod q_k, (-q_{k-1}^{-1}) \bmod q_k\},$$

where, in fact, $(-q_{k-1}^{-1}) \bmod q_k = q_k - (q_{k-1}^{-1} \bmod q_k)$. Thus

$$\begin{aligned} \sum_{k \geq 1} \sum_{\mathbf{m} \in \mathbb{N}^k} f\left(\frac{q_{k-1}(\mathbf{m})}{q_k(\mathbf{m})}, \frac{p_k(\mathbf{m})}{q_k(\mathbf{m})}, \frac{q_k(\mathbf{m})}{N}\right) &= \sum_{k \geq 1} \sum_{\substack{\mathbf{m} \in \mathbb{N}^k, \\ m_1 > 1}} \left(f\left(\frac{q_{k-1}}{q_k}, 1 - \frac{q_{k-1}^{-1} \bmod q_k}{q_k}, \frac{q_k}{N}\right) \right. \\ &\quad \left. + f\left(\frac{q_{k-1}}{q_k}, \frac{q_{k-1}^{-1} \bmod q_k}{q_k}, \frac{q_k}{N}\right) \right) \\ &\quad + f\left(1, 1, \frac{1}{N}\right). \end{aligned}$$

We recall that the mirror property tells us that

$$p_k(\theta(\mathbf{m})) = q_{k-1}(\mathbf{m}), \quad q_{k-1}(\theta(\mathbf{m})) = p_k(\mathbf{m}), \quad q_k(\theta(\mathbf{m})) = q_k(\mathbf{m}),$$

thus by applying the mirror θ on the right-hand side (note that this turns $m_1 > 1$ into $m_k > 1$) we have

$$\begin{aligned} \sum_{\mathbf{m} \in \mathbb{N}^*} f\left(\frac{q_{k-1}(\mathbf{m})}{q_k(\mathbf{m})}, \frac{p_k(\mathbf{m})}{q_k(\mathbf{m})}, \frac{q_k(\mathbf{m})}{N}\right) &= \sum_{\mathbf{m} \in \mathbb{N}^k} \sum_{\substack{\mathbf{m} \in \mathbb{N}^k, \\ m_k > 1}} \left(f\left(\frac{p_k}{q_k}, 1 - \frac{p_k^{-1} \bmod q_k}{q_k}, \frac{q_k}{N}\right) \right. \\ &\quad \left. + f\left(\frac{p_k}{q_k}, \frac{p_k^{-1} \bmod q_k}{q_k}, \frac{q_k}{N}\right) \right) + f\left(1, 1, \frac{1}{N}\right). \end{aligned}$$

The coprime pairs (p_k, q_k) go through each reduced fraction $a/b = p_k/q_k$ exactly once (except for $a/b = 1$), thanks to condition $m_k > 1$, thus we obtain

$$\begin{aligned} f\left(1, 1, \frac{1}{N}\right) + \sum_{\mathbf{m} \in \mathbb{N}^*} f\left(\frac{q_{k-1}(\mathbf{m})}{q_k(\mathbf{m})}, \frac{p_k(\mathbf{m})}{q_k(\mathbf{m})}, \frac{q_k(\mathbf{m})}{N}\right) &= \sum_{(a,b) \in \mathbb{N}^2: \substack{1 \leq a \leq b, \\ \gcd(a,b)=1}} \left(f\left(\frac{a}{b}, \frac{a^{-1} \bmod b}{b}, \frac{b}{N}\right) \right. \\ &\quad \left. + f\left(\frac{a}{b}, 1 - \frac{a^{-1} \bmod b}{b}, \frac{b}{N}\right) \right), \end{aligned}$$

and we are done. ■

The following result solves the problem for the distribution of S_n as the event $S_n(\alpha) \leq \lambda$ makes the function we need compactly supported (we clearly have the bound $\frac{q_k}{N} \leq \lambda$ when $S_n \leq \lambda$). The result still holds for the distributions of other \mathcal{LQ} -functions the same argument.

Theorem 4.15. *Let $f: [0, 1] \times [0, 1] \times [0, \infty) \rightarrow \mathbb{R}$ be locally Riemann-integrable and compactly supported. Then*

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{k \geq 0} \sum_{\mathbf{m} \in \mathbb{N}^*} f\left(\frac{q_{k-1}}{q_k}, \frac{p_k}{q_k}, \frac{q_k}{N}\right) &= \frac{12}{\pi^2} \iiint_{[0,1] \times [0,1] \times [0,\infty)} y f(\rho \hat{\rho}, \rho, \hat{\rho}, y) d\rho d\hat{\rho} dy. \end{aligned}$$

Observation 4.2. If we wanted to consider also p_{k-1} for our approximations, we could too. Indeed, the equality $p_{k-1}q_k - p_kq_{k-1} = (-1)^k$ implies

$$\frac{p_{k-1}}{q_k} = \frac{p_k}{q_k} \cdot \frac{q_{k-1}}{q_k} + \frac{(-1)^k}{q_k^2},$$

thus it is easy to extend the result for a C^1 function. The rest follows from approximation with C^1 functions.

Corollary 4.1. *Let $f: [0, 1] \times [0, 1] \times [0, 1] \times [0, \infty) \rightarrow \mathbb{R}$ be locally Riemann-integrable and compactly supported. Then*

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{k \geq 0} \sum_{m \in \mathbb{N}^*} f \left(\frac{p_{k-1}}{q_k}, \frac{q_{k-1}}{q_k}, \frac{p_k}{q_k}, \frac{q_k}{N} \right) \\ = \frac{12}{\pi^2} \iiint_{[0,1] \times [0,1] \times [0,\infty)} y f(\rho\hat{\rho}, \rho, \hat{\rho}, y) d\rho d\hat{\rho} dy. \end{aligned}$$

Concluding remarks. We did not originally think it was possible to extend the results for the limit distributions to more general “initial” densities. The real motivation for this extension was the study of the recurrence function over pure quadratic irrationals (see subsection 6.3.2). When studying the average behavior of the recurrence function when $q_{k-1}(\alpha) \leq n < q_k(\alpha)$ for an index k within the *first* period of the continued fraction expansion of α , the distribution of the recurrence quotient coincides with the classical case from this chapter, but changing the Lebesgue measure for the Gauss measure. The results from this subsection tell us that we retrieve the same limit distributions.

4.6 Conclusions

Beginning from the question “what does the recurrence function of a random Sturmian word look like?”, we define and work within a model that is natural at least from an algorithmic standpoint: pick a large integer n and let the slope of the word be drawn at random from $[0, 1]$. We are led to the notion of the so-called \mathcal{Q} -functions: functions that, given n and a slope α , place n within the sequence of continuants $k \mapsto q_k(\alpha)$ of α , namely consider the index k for which $n \in [q_{k-1}(\alpha), q_k(\alpha))$, and then return a value depending only on the two ratios $(1/n)q_{k-1}(\alpha)$ and $(1/n)q_k(\alpha)$. The recurrence quotient of Sturmian words defines such a \mathcal{Q} function, via a Theorem of Morse and Hedlund, where n is the length of the factors and α the slope of the word.

Then, we study the distribution of a general \mathcal{Q} -function. It defines in fact a sequence of distributions, and we prove that the limit distribution and the limit densities exist. They all involve, as a sort of reference density, the density ψ defined in (4.10), which plays a similar role to that of the Gauss density (defined in (1.8)) when one studies functions that depend on the ratio $q_{k-1}(\alpha)/q_k(\alpha)$, and appears in our study [BCR⁺15] which we explain extensively in Chapter 5.

Our results apply in particular to the recurrence quotient of Sturmian words; we exhibit the limit distribution (and the limit density) of such a quotient. We compare our probabilistic study to the results of Morse and Hedlund (see Theorem 3.4), which exhibit extreme behaviors, attained when n is close to the left border $q_{k-1}(\alpha)$ of the interval $[q_{k-1}(\alpha), q_k(\alpha))$ containing the integer n . That is why we also consider conditional expectations, where the conditional events are related to the various parameters which describe the position of the integer n inside $[q_{k-1}(\alpha), q_k(\alpha))$. We then compare this “constrained probabilistic” behaviors to the extreme behaviors, in a precise manner.

We had already performed a similar study in [BCR⁺15] under a different probabilistic model, which we called “fixed depth $k \rightarrow \infty$ ” model in subsection 3.4.3. We present this model in detail in Chapter 5. In

this model it is rather the index k of the interval $[q_{k-1}(\alpha), q_k(\alpha)[$ the integer n belongs to that is fixed. Then for $k \rightarrow \infty$, we exhibited limit distribution and limit densities all of which involve, as a sort of reference density, the Gauss density. The two models are clearly different, but the two types of results show certain similarities which we discuss in Chapter 6.

Further studies. In Chapter 6 we also present two variants of this model, which we consider in order to tackle the study of the recurrence function for two special families of α .

Rational Numbers. This type of slope gives rise to periodic words, and occurs for Christoffel words. For a bound N , we restrict α to the set of rationals with denominator at most N , endowed with the uniform distribution, and we wish to observe the transition when $N \rightarrow \infty$. We wish to explain how a periodic word “becomes” Sturmian.

Quadratic Irrationals. This type of slope α , which we presented in Section 1.5, occurs for substitutive Sturmian words. There is a natural notion of size ($\epsilon(x)$ from Section 1.5) associated with such numbers, closely related to the period of their continued fraction expansion, and we wish to observe the transition when the size becomes large.

CHAPTER 5

THE RECURRENCE FUNCTION AND THE RELATIVE POSITION

5.1 Introduction

The recurrence function, introduced in Chapter 3, measures the “complexity” of an infinite word and describes the possible occurrences of finite factors inside it together with the maximal gaps between successive occurrences. This recurrence function has been widely studied, notably in the case of Sturmian words (see [Cas99, MH40]). We recall that, due to Theorem 3.1, Sturmian words are strongly characterized by their slope α . This is in particular the case for the recurrence function $n \mapsto R_\alpha(n)$, as explained by Theorem 3.3), where the integer $m = R_\alpha(n)$ is the length of the smallest “window” needed to discover the whole set $\mathcal{L}_\alpha(n)$ of finite factors of length n . The set of factors $\mathcal{L}_\alpha(n)$ is widely used in many applications of Sturmian words (for instance quasicrystals, or digital geometry), and therefore the function $n \mapsto R_\alpha(n)$ intervenes very often as a pre-computation cost, hence the importance of better understanding it “on average”, when the real α is randomly chosen in the unit interval.

We have presented a random model, the “large fixed n ” model, in Chapter 4. The “large fixed n ” answers the question “given a random slope α and a large n of our choice, how big is $R(\alpha, n)$ on average?”, which makes sense from an algorithmic point of view. In the current chapter we will present another model, the “large fixed k ” model, we fix the interval $[q_{k-1}(\alpha), q_k(\alpha))$ between two continuants of α which contains n , and prescribe a geometric description of the positioning of n within the interval. Thus the integer n is not fixed anymore, becoming a random variable. This model is more adept to answering questions about the incidence of the position of n within the interval and hence yield nicely worst-case families of n .

The expression of the recurrence function is recalled in Section 5.2. Our viewpoint and our main results are given in Section 5.3. Proofs are provided in Section 5.4.

5.2 The recurrence function of Sturmian words

Notation. In the sequel $\varphi = (\sqrt{5} - 1)/2 = 0.6180339\dots$ stands for the inverse of the golden ratio, and for two integers a, b , the set of integers n that satisfy $a \leq n \leq b$ is denoted by $\llbracket a, b \rrbracket := [a, b] \cap \mathbb{N}$.

We recall that the simplest words that are not eventually periodic satisfy the equality $p_u(n) = n + 1$ for

each $n \geq 0$. Such words do exist: they are called *Sturmian words*. Morse and Hedlund provided a powerful arithmetic description of Sturmian words (see Chapter 3 for more details).

Proposition. 3.1 Associate with a pair $(\alpha, \beta) \in [0, 1]^2$ the two infinite words $\underline{\mathfrak{S}}(\alpha, \beta)$ and $\overline{\mathfrak{S}}(\alpha, \beta)$ whose n -th symbols are respectively

$$\begin{aligned} \underline{u}_n &= \lfloor \alpha(n+1) + \beta \rfloor - \lfloor \alpha n + \beta \rfloor, \\ \overline{u}_n &= \lceil \alpha(n+1) + \beta \rceil - \lceil \alpha n + \beta \rceil. \end{aligned}$$

Then a word $u \in \{0, 1\}^{\mathbb{N}}$ is Sturmian if and only if it equals $\underline{\mathfrak{S}}(\alpha, \beta)$ or $\overline{\mathfrak{S}}(\alpha, \beta)$ for a pair (α, β) formed with an irrational $\alpha \in (0, 1)$ and a real $\beta \in [0, 1)$.

It is also important to study where finite factors occur inside the infinite word u . This is where the recurrence function (see subsection 3.3.1) comes in, giving a measure of how often factors reappear on the worst case.

More precisely, let $w_u(q, n)$ be the minimal number of symbols u_k with $k \geq q$ which have to be inspected for discovering the whole set $\mathcal{L}_u(n)$ from the index q . Then u is uniformly recurrent if each set $\{w_u(q, n); q \in \mathbb{N}\}$ is bounded, and the *recurrence function* $n \mapsto R_u(n)$ is defined by

$$R_u(n) := \max\{w_u(q, n); q \in \mathbb{N}\}.$$

Any Sturmian word is uniformly recurrent (see Proposition 3.11). Its recurrence function only depends on the slope α and is thus denoted by $n \mapsto R(\alpha, n) = R_\alpha(n)$. Moreover, it only depends on α via its *continuants*. We now recall this notion which plays a central role in the paper. Consider the *continued fraction expansion* (see Chapter 1 for more details) of the irrational α

$$\alpha = \frac{1}{m_1 + \frac{1}{\ddots + \frac{1}{m_k + \frac{1}{\ddots}}}}$$

The positive integers m_k are called the *partial quotients*. The truncated expansion $[m_1, \dots, m_k]$ at depth k defines a rational, and the *continuant* $q_k(\alpha)$ is the denominator of this rational. The continuant sequence satisfies $q_{-1} = 0, q_0 = 1$ and for any $k \geq 1$ the recurrence $q_k = m_k q_{k-1} + q_{k-2}$ for all k .

We recall a fundamental result (Theorem 3.3) by Morse and Hedlund [MH40] :

Theorem. 3.3 For any Sturmian word of slope α , the recurrence function $n \mapsto R_\alpha(n)$ is piece-wise affine and satisfies

$$R_\alpha(n) = n - 1 + q_k(\alpha) + q_{k-1}(\alpha),$$

for $n \in \mathbb{N}$ satisfying $q_{k-1}(\alpha) \leq n < q_k(\alpha)$.

It is thus natural to study the quotient $S(\alpha, n) = (R_\alpha(n)+1)/n$. When n belongs to the interval $\llbracket q_{k-1}(\alpha), q_k(\alpha) - 1 \rrbracket$, this quotient depends itself on two quotients: the quotient $\rho_k(\alpha) := q_{k-1}(\alpha)/q_k(\alpha)$, and the quotient $\nu_k(\alpha) := n/q_k(\alpha)$, and

$$S(\alpha, n) := \frac{R_\alpha(n) + 1}{n} = 1 + \frac{1 + \rho_k(\alpha)}{\nu_k(\alpha)}. \tag{5.1}$$

As $\nu_k(\alpha)$ belongs to the interval $[\rho_k(\alpha), 1]$, the following bounds hold

$$2 + \rho_k(\alpha) \leq \frac{R_\alpha(n) + 1}{n} \leq 2 + \frac{1}{\rho_k(\alpha)} \tag{5.2}$$

(the lower bound holds for n close to $q_k(\alpha)$ whereas the upper bound is attained for $n = q_{k-1}(\alpha)$). The ratio $\rho_k(\alpha)$ belongs to $(0, 1]$, and the Borel-Bernstein Theorem (see e.g., [IK02]) proves that $\liminf_{k \rightarrow \infty} \rho_k(\alpha) = 0$ for almost every irrational α . More precisely:

Proposition. 3.16 and 3.17. *Let $\epsilon > 0$. For almost every α the recurrence function satisfies:*

$$\limsup_{n \rightarrow \infty} \frac{R(\alpha, n)}{n \log n} = \infty, \quad \limsup_{n \rightarrow \infty} \frac{R(\alpha, n)}{n (\log n)^{1+\epsilon}} = 0,$$

as well as

$$\liminf_{n \rightarrow \infty} \frac{R(\alpha, n)}{n} = 2.$$

Finally we recall that, thanks to the previous proposition, we work, not with the recurrence function $R(\alpha, n)$, but rather with the recurrence quotient

$$S(\alpha, n) := \frac{R(\alpha, n) + 1}{n},$$

for which we will sometimes write $S_\alpha(n)$ when we want to fix α and highlight the dependence on n .

5.3 Probabilistic model and main results

The notable feature of the model we are about to introduce, which was mentioned in subsection 3.4.3, is that it highlights the relationship between the size of the recurrence function $R(\alpha, n)$ and the barycentric position $\mu(\alpha, n)$ of n , within its corresponding interval $[q_{k-1}(\alpha), q_k(\alpha))$.

In the first model presented here [RV17] (which was actually our second model historically, see Chapter 4 for more details), the size n of the factors was a fixed (and very large) positive integer, and then the position $\mu_n(\alpha) := \mu(\alpha, n)$ becomes a random variable which can be studied in its own right (see its density in 4.2).

In the model from [BCR⁺15], which we are about to present in detail, we do not have the freedom to choose n anymore (in fact, we will have an analogous random variable $n_k^{(\mu)}$) but we have full freedom to fix the relative position μ as a parameter. We thus begin by revisiting the geometric parameters μ and ρ that will interest us in this model, then we introduce the model and finally the results along with their proofs.

5.3.1 Position parameter μ .

We consider a fixed sequence $(\mu_k)_k$ with values in $[0, 1)$, and for each $\alpha \in \mathcal{I} := [0, 1]$, and each $k \in \mathbb{N}$, we consider the real number at (barycentric) position μ_k inside the interval $\llbracket q_{k-1}(\alpha), q_k(\alpha) - 1 \rrbracket$, namely

$$\tilde{n}_k^{(\mu_k)}(\alpha) := q_{k-1}(\alpha) + \mu_k(q_k(\alpha) - q_{k-1}(\alpha)),$$

together with its integer part (which belongs to $\llbracket q_{k-1}(\alpha), q_k(\alpha) - 1 \rrbracket$),

$$n_k^{(\mu_k)}(\alpha) = \lfloor \tilde{n}_k^{(\mu_k)}(\alpha) \rfloor = q_{k-1}(\alpha) + \lfloor \mu_k(q_k(\alpha) - q_{k-1}(\alpha)) \rfloor.$$

The subsequence $(n_k^{(\mu_k)}(\alpha))_k$ is *the subsequence associated with the positions μ_k* .

We are interested in the subsequence of $n \mapsto S(\alpha, n)$ associated with the subsequence $\{n_k^{(\mu_k)}(\alpha), k \in \mathbb{N}\}$, and we then let $S_k^{(\mu_k)}(\alpha) := S(\alpha, n_k^{(\mu_k)}(\alpha))$, namely

$$S_k^{(\mu_k)}(\alpha) = 1 + \frac{q_{k-1}(\alpha) + q_k(\alpha)}{n_k^{(\mu_k)}(\alpha)} = 1 + \frac{q_{k-1}(\alpha) + q_k(\alpha)}{q_{k-1}(\alpha) + \lfloor \mu_k(q_k(\alpha) - q_{k-1}(\alpha)) \rfloor}. \quad (5.3)$$

If we drop the integer part in the expression of $S_k^{\langle \mu_k \rangle}$, we deal with the sequence $\tilde{S}_k^{\langle \mu_k \rangle}(\alpha) := S(\alpha, \tilde{n}_k^{\langle \mu_k \rangle}(\alpha))$, namely,

$$\tilde{S}_k^{\langle \mu_k \rangle}(\alpha) = 1 + \frac{q_{k-1}(\alpha) + q_k(\alpha)}{\tilde{n}_k^{\langle \mu_k \rangle}(\alpha)} = 1 + \frac{q_{k-1}(\alpha) + q_k(\alpha)}{q_{k-1}(\alpha) + \mu_k(q_k(\alpha) - q_{k-1}(\alpha))}, \quad (5.4)$$

which is expressed with the two sequences $(\rho_k(\alpha))_k$ and (μ_k) as

$$\tilde{S}_k^{\langle \mu_k \rangle}(\alpha) = f_{\mu_k}(\rho_k(\alpha)) \quad \text{with} \quad f_\mu(x) := 1 + \frac{1+x}{x + \mu(1-x)}. \quad (5.5)$$

The study of the function f_μ provides a precise knowledge on the sequence $\tilde{S}_k^{\langle \mu_k \rangle}(\alpha)$, that may be “transferred” to the sequence $S_k^{\langle \mu_k \rangle}(\alpha)$ since the two sequences are “close enough”. The following result provides such a first instance of this strategy:

Proposition 5.1. *Consider a sequence $(\mu_k)_k$ with $\mu_k \in [0, 1]$, and let $\alpha \in [0, 1] \setminus \mathbb{Q}$.*

(i) *Denote by m_k the k -th partial quotient of α . Then, $\rho_k(\alpha) \leq 1/(m_k + 1)$ and*

$$\tilde{S}_k^{\langle \mu_k \rangle}(\alpha) \in \left[1 + \frac{m_k + 2}{\mu_k m_k + 1}, 3 \right] \quad \text{or} \quad \tilde{S}_k^{\langle \mu_k \rangle}(\alpha) \in \left[3, 1 + \frac{m_k + 2}{\mu_k m_k + 1} \right]$$

depending whether $\mu_k \in [1/2, 1]$ or $\mu_k \in [0, 1/2]$.

(ii) *The sequence $S_k^{\langle \mu_k \rangle}(\alpha)$ is bounded if α has bounded partial quotients or if the sequence (μ_k) admits a strictly positive lower bound.*

Proof. The map $f_\mu : [0, 1] \rightarrow \mathbb{R}$ is strictly decreasing when $\mu \in (0, 1/2)$, and strictly increasing when $\mu \in (1/2, 1)$. This is the constant function equal to 3 when $\mu = 1/2$. For any $a \in (0, 1)$, the image $f_\mu([a, 1])$ is the interval with endpoints 3 and $f_\mu(a)$. This proves Assertion (i).

With the two inequalities

$$\tilde{n}_k^{\langle \mu \rangle} \geq n_k^{\langle \mu \rangle} \geq q_{k-1} \geq \varphi^{1-k}, \quad 0 \leq \tilde{n}_k^{\langle \mu \rangle} - n_k^{\langle \mu \rangle} \leq 1,$$

we obtain the inequality

$$0 \leq S_k^{\langle \mu \rangle} - \tilde{S}_k^{\langle \mu \rangle} = \frac{q_k + q_{k-1}}{n_k^{\langle \mu \rangle} \cdot \tilde{n}_k^{\langle \mu \rangle}} (\tilde{n}_k^{\langle \mu \rangle} - n_k^{\langle \mu \rangle}) \leq \frac{1}{q_{k-1}} \frac{q_k + q_{k-1}}{\tilde{n}_k^{\langle \mu \rangle}} \leq \varphi^{k-1} \tilde{S}_k^{\langle \mu \rangle}, \quad (5.6)$$

and we apply (i). ■

5.3.2 Probabilistic model

Let us describe now our *probabilistic model*. We choose a sequence $(\mu_k)_k$ of positions that will be *fixed*. This defines, for each real α , a sequence of indices $n_k := n_k^{\langle \mu_k \rangle}$, and then a sequence of real numbers $k \mapsto S_k^{\langle \mu_k \rangle}(\alpha)$. When the real α is random, and uniformly drawn in the unit interval $\mathcal{I} = [0, 1]$, the sequence $k \mapsto S_k^{\langle \mu_k \rangle}$ becomes a sequence of random variables, and we study the mean value and the distribution of the sequence $k \mapsto S_k^{\langle \mu_k \rangle}$ for $k \rightarrow \infty$.

For any position, the index $n_k^{\langle \mu_k \rangle}$ belongs to the interval $[[q_{k-1}, q_k - 1]]$. Then, as the expectations for $\alpha \in [0, 1]$ of the two extreme sequences $k \mapsto \log q_{k-1}(\alpha)$, $k \mapsto \log q_k(\alpha)$ satisfy the same estimates (see [L36]), it is also the case for the expectation for $\alpha \in [0, 1]$ of the sequence $k \mapsto \log n_k^{\langle \mu_k \rangle}(\alpha)$. It thus satisfies

$$\mathbb{E}[\log n_k^{\langle \mu_k \rangle}] = \frac{\pi^2}{12 \log 2} k + O(1), \quad (5.7)$$

and it is of linear growth with respect to k .

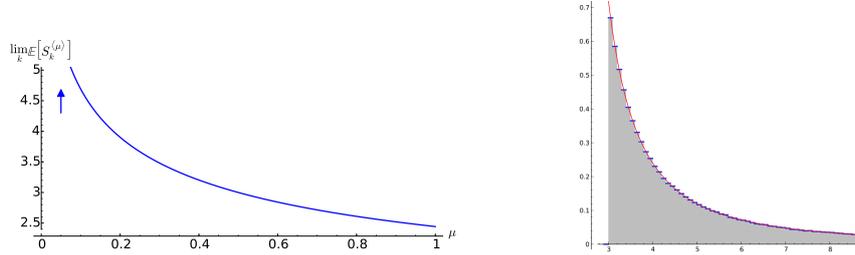


Figure 5.1: On the left, the graph of $\lim_{k \rightarrow \infty} \mathbb{E}[S_k^{(\mu)}]$ as a function of μ . On the right, the graph of the density $s^{(0)}$.

5.3.3 Results for a constant position μ

We first consider the case in which the sequence $(\mu_k)_k$ is a constant, taking a fixed value μ , and we study the expectation and the distribution of the sequence $k \mapsto S_k^{(\mu)}$ of random variables, when $k \rightarrow \infty$, as a function of the position μ . Theorem 5.1 below shows that there are two main cases:

- (a) the case when $\mu = 0$; here, the expectations are infinite, but the functions $k \mapsto S_k^{(0)}$ admit a limit density;
- (b) the case when $\mu \neq 0$; here, both the expectations and the densities have a finite limit; the case $\mu = 1/2$ is particular, as the limit density is a Dirac measure, concentrated at the value 3.

For indices n associated with parameters μ satisfying $\mu \geq \mu_0 > 0$, we exhibit a behavior for the sequence $n \mapsto R_\alpha(n)$ which is thus “linear on average”; the “log n ” behavior of Theorem 3.4 does not occur in this case.

Theorem 5.1. [Fixed position μ] *Let $\varphi = (\sqrt{5} - 1)/2 < 1$. The following results hold for the random variables $S_k^{(\mu)}$.*

- (i) [Expectations] *For each $\mu \in]0, 1]$, their expected values $\mathbb{E}[S_k^{(\mu)}]$ satisfy*

$$\mathbb{E}[S_k^{(\mu)}] = 1 + \frac{1}{\log 2} \frac{|\log \mu|}{1 - \mu} + O\left(\frac{\varphi^{2k}}{\mu}\right) + O\left(\varphi^k \frac{|\log \mu|}{1 - \mu}\right), \quad (5.8)$$

where the constants in the O -terms are uniform with respect to μ and k .

- (ii) [Limit density] *For each $\mu \in [0, 1]$ with $\mu \neq 1/2$, they admit a limit density $s^{(\mu)}$ equal to*

$$s^{(\mu)}(x) = \frac{1}{\log 2} \left(\frac{1}{(x-1)|x(1-\mu) + \mu - 2|} \right) \mathbf{1}_{\mathcal{I}_\mu}(x), \quad (5.9)$$

where \mathcal{I}_μ is the real interval with endpoints 3 and $1 + 1/\mu$.

More precisely, for any $b \in \mathcal{I}_\mu$, one has

$$\mathbb{P}\left(S_k^{(\mu)} \leq b\right) = \int_0^b s^{(\mu)}(x) dx + \frac{1}{b} O\left(\varphi^k\right),$$

where the constant of the O -term is uniform with respect to b and k . It is also uniform with respect to μ when μ satisfies $|\mu - 1/2| \geq \mu_0$ for any $\mu_0 > 0$.

5.3.4 Results when the sequence $\mu_k \rightarrow 0$

We now focus on the difficult case, when the sequence $(\mu_k)_k$ is no longer constant, and we consider a sequence $(\mu_k)_k$ of positions which tends to 0. Theorem 5.2 below tackles this case, first considering sequences $(\mu_k)_k$ which tend exponentially fast to 0, for which we observe that the expectations are of order k . We then consider general sequences $(\mu_k)_k$ which tend to 0, and we show that the associated random variables admit a limit density, with a speed of convergence which depends upon the sequence $(\mu_k)_k$.

Theorem 5.2. [Sequence $\mu_k \rightarrow 0$] *The following holds for the random variables $S_k^{(\mu)}$ associated with a sequence $\mu_k \rightarrow 0$.*

(i) [Expectations] *Consider the sequence $\mu_k = \tau^k$, with $\tau \in [\varphi^2, 1)$. Then*

$$\mathbb{E}[S_k^{(\tau^k)}] = k \frac{|\log \tau|}{\log 2} + O(1), \quad (5.10)$$

where the constant hidden in the O -term is uniform with respect to τ and k .

For any α , and for each $\tau \in [\varphi^2, 1]$, there exists a family of increasing subsequences $\mathcal{N}(\alpha, \tau)$, depending on both α and τ , of indices n for which

$$\mathbb{E} \left[\frac{R_\alpha(n)}{n} - \frac{12|\log \tau|}{\pi^2} \log n \right] = O(1) \quad (n \rightarrow \infty). \quad (5.11)$$

For any $\tau < 1$, if μ_k is drawn uniformly from $[0, 1]$, the conditional expectation with respect to the event $[\mu_k \geq \tau^k]$ satisfies

$$\lim_{k \rightarrow \infty} \mathbb{E} \left[S_k^{(\mu_k)} \mid [\mu_k \geq \tau^k] \right] = 1 + \frac{\pi^2}{6 \log 2}.$$

(ii) [Limit density] *For any sequence $\mu_k \rightarrow 0$, the random variables $S_k^{(\mu_k)}$ admit as limit density the density $s^{(0)}$ equal to*

$$s^{(0)}(x) = \frac{1}{\log 2} \frac{1}{(x-1)(x-2)} \mathbf{1}_{[3, \infty)}(x).$$

More precisely, for any $b \geq 3$, the probability $\mathbb{P}[S_k^{(\mu_k)} \geq b]$ satisfies

$$\mathbb{P}[S_k^{(\mu_k)} \geq b] = \frac{1}{\log 2} \log \left(\frac{b-1}{b-2} \right) + O(\mu_k) + \frac{1}{b} O(\varphi^k),$$

where the constants hidden in the O -term are uniform to respect to b and k . If now the sequences $(b_k)_k$ and $(\mu_k)_k$ satisfy the following three conditions ($b_k \rightarrow \infty$, $\mu_k \rightarrow 0$ with $b_k \mu_k \rightarrow 0$), then

$$\lim_{k \rightarrow \infty} b_k \cdot \mathbb{P} \left(S_k^{(\mu_k)} \geq b_k \right) = \frac{1}{\log 2}.$$

Observation 5.1. The estimate (5.7) together with (5.10) yields (5.11). We have then exhibited a $\log n$ behavior “on average” for the ratio $R_\alpha(n)/n$ for (an infinity of) particular subsequences n (which depend on α). On the contrary, when the position is not too small, the ratio $R_\alpha(n)/n$ remains bounded (on average).

5.4 Strategy for the proofs.

We begin with Theorem 5.1 which deals with a fixed position μ . There are three main steps in the proof of Theorem 5.1.

- (i) We drop the integer part in the expression of $S_k^{(\mu)}$ and consider the sequence $\tilde{S}_k^{(\mu)}(\alpha)$ which can be written as $f_\mu(\rho_k(\alpha))$ (see (5.5)). This is an instance of a smooth sequence (as defined in Section 5.4.1 below). We express its mean value and distribution in terms of the k -th iterate of the Perron Frobenius operator \mathbf{H} .
- (ii) With the spectral properties of the operator \mathbf{H} (described in Section 5.4.2), when acting on the Banach space $BV(\mathcal{I})$ of the functions of bounded variation on the unit interval \mathcal{I} , we obtain the asymptotics of the expectations and the expression of the limit distribution, always for the sequence $\tilde{S}_k^{(\mu)}$.
- (iii) We return to the initial sequence $S_k^{(\mu)}$ with the following estimates

$$\mathbb{E}[S_k^{(\mu)}] = \mathbb{E}[\tilde{S}_k^{(\mu)}] \left(1 + O(\varphi^k)\right), \quad \mathbb{P}(S_k^{(\mu)} \leq b) - \mathbb{P}(\tilde{S}_k^{(\mu)} \leq b) = O\left(\frac{\varphi^k}{b}\right), \quad (5.12)$$

which are refinements of Eq.(5.6).

Since the probabilistic estimates obtained in Theorem 5.1 are uniform with respect to μ and k , we may extend them to the case where μ depends on k , and we may study the interesting case where the sequence $(\mu_k)_k$ tends to 0 for $k \rightarrow \infty$. We then obtain the results of Theorem 5.2.

Observation 5.2. There are two error terms in the asymptotic estimates (5.8) of the expectations. The first one comes from the spectral gap of the Perron-Frobenius operator and the second one arises when one takes into account integer parts in the definition of $S_k^{(\mu)}$.

5.4.1 Smooth sequences

The sequence $\tilde{S}_k^{(\mu)}$ provides an instance of a smooth sequence, defined as follows:

Definition 5.1. A sequence of random variables (T_k) defined on the unit interval $\mathcal{I} = [0, 1]$ is a *smooth sequence* if there exists a function $f \in BV(\mathcal{I})$ for which

$$T_k(\alpha) = f(\rho_k(\alpha)) \quad \text{with} \quad \rho_k(\alpha) = \frac{q_{k-1}(\alpha)}{q_k(\alpha)} \quad \text{for all } \alpha \in \mathcal{I}.$$

Here, we deal with the function f_μ defined in (5.5), whose inverse map g_μ is

$$g_\mu : f_\mu(\mathcal{I}) \mapsto [0, 1], \quad g_\mu(x) = \frac{-1 - \mu + \mu x}{2 - \mu - x(1 - \mu)}.$$

For $\mu \in (0, 1)$, the function f_μ is integrable on \mathcal{I} , and its L^1 -norm satisfies

$$\|f_\mu\|_{L^1} = 1 + \frac{1}{1 - \mu} + \frac{1 - 2\mu}{(1 - \mu)^2} |\log \mu|, \quad \|f_1\|_{L^1} = 5/2.$$

Moreover, for every $\mu \in (0, 1)$, the function f_μ is monotonic and thus of bounded variation, with a total variation equal to $(1/\mu)|1 - 2\mu|$, hence $\|f_\mu\|_{BV} = O(1/\mu)$. Remark that f_0 does not belong to $BV(\mathcal{I})$.

We now recall some basic facts on the underlying dynamical system, together with the Perron-Frobenius operator, that will be useful in the sequel.

5.4.2 The dynamical system and the Perron-Frobenius operator

The underlying dynamical system. The underlying dynamical system is the Euclidean system from Chapter 1. We recall it briefly here.

We consider the dynamical system (\mathcal{I}, T_g) associated with the unit interval \mathcal{I} and the Gauss map T_g , defined by

$$T_g(x) = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor = \left\{ \frac{1}{x} \right\} \quad \text{for } x \neq 0, \quad T_g(0) = 0.$$

The map T_g builds the continued fraction expansion of α , via the function $m(\alpha) := \lfloor 1/\alpha \rfloor$, as

$$\alpha = [m_1, m_2, \dots, m_k, \dots] \quad \text{with} \quad m_{k+1}(\alpha) = m(T_g^k(\alpha)) \text{ for all } k \geq 0.$$

The inverse branches of T_g belong to the set

$$\mathcal{H} := \left\{ h_m : x \mapsto \frac{1}{m+x}; \quad m \geq 1 \right\},$$

and the inverse branches of T_g^k belong to the set

$$\mathcal{H}^k = \{ h_{m_1} \circ h_{m_2} \circ \dots \circ h_{m_k} \quad : \quad m_1, \dots, m_k \geq 1 \}.$$

For a k -uple $\mathbf{m} = (m_1, m_2, \dots, m_k)$, let $h_{\mathbf{m}} := h_{m_1} \circ h_{m_2} \circ \dots \circ h_{m_k}$. The linear fractional transformation $h_{\mathbf{m}}$ is expressed in terms of the sequences (p_k) and (q_k) , of successive numerators and denominators of the convergents, under the form

$$h_{\mathbf{m}}(x) = h_{m_1} \circ h_{m_2} \circ \dots \circ h_{m_k}(x) = \frac{1}{m_1 + \frac{1}{\dots + \frac{1}{m_k + x}}} = \frac{p_{k-1}x + p_k}{q_{k-1}x + q_k}.$$

We remark that q_k, p_k depend only on the k -uple $\mathbf{m} = (m_1, m_2, \dots, m_k)$. We insist that there is no conflict with the notation $q_k(\alpha)$. Indeed, when α is an irrational which belongs to the interval $h_{\mathbf{m}}(\mathcal{I})$, the equality $q_k(\alpha) = q_k(\mathbf{m})$ holds as the digits are uniquely determined from α .

The mirror property (1.14) (in Corollary 1.3) relates the coefficients of $h = h_{m_1} \circ h_{m_2} \circ \dots \circ h_{m_k}$ and those of its mirror $\widehat{h} := h_{m_k} \circ h_{m_{k-1}} \circ \dots \circ h_{m_1}$:

$$h(y) = \frac{p_{k-1}y + p_k}{q_{k-1}y + q_k} \quad \Longrightarrow \quad \widehat{h}(y) = \frac{p_{k-1}y + q_{k-1}}{p_k y + q_k}.$$

Perron-Frobenius operator. We recall that when the unit interval \mathcal{I} is endowed with a density f , after one iteration of T_g , it is endowed with the density

$$\mathbf{H}[f](x) := \sum_{h \in \mathcal{H}} |h'(x)| \cdot f \circ h(x),$$

and after k iterations of T_g , with the density

$$\mathbf{H}^k[f](x) = \sum_{h \in \mathcal{H}^k} |h'(x)| \cdot f \circ h(x).$$

The operator \mathbf{H} is exactly the Perron Frobenius operator from Definition 1.7.

Now, at $x = 0$, the two maps h and \widehat{h} satisfy $|h'(0)| = |\widehat{h}'(0)| = 1/q_k^2$, and the equality $q_{k-1}/q_k = \widehat{h}(0)$ holds. With this remark, the k -th iterate \mathbf{H}^k generates the continuants q_k ,

$$\mathbf{H}^k[f](0) = \sum_{h \in \mathcal{H}^k} \frac{1}{q_k^2} f \left(\frac{p_k}{q_k} \right) = \sum_{h \in \mathcal{H}^k} \frac{1}{q_k^2} f \left(\frac{q_{k-1}}{q_k} \right). \quad (5.13)$$

We now summarize some classical spectral properties of the operator \mathbf{H} (see e.g. [IK02] or [BDV02]). When acting on the Banach space $BV(\mathcal{I})$ of functions of bounded variation, the operator \mathbf{H} admits a unique dominant eigenvalue $\lambda = 1$, with an eigenfunction proportional to $\psi(x) = 1/(1+x)$, as follows from Theorem 1.12, and it has a subdominant spectral radius equal to φ^2 . Moreover, the adjoint \mathbf{H}^* has an eigenmeasure proportional to the Lebesgue measure. Then, for any $g \in BV(\mathcal{I})$ and $\epsilon > 0$, the iterate $\mathbf{H}^k[g]$ decomposes as

$$\mathbf{H}^k[g](x) = \frac{1}{\log 2} \frac{1}{1+x} \cdot \int_{\mathcal{I}} g(x) dx + O((\varphi + \epsilon)^{2k}) \|g\|_{BV}. \quad (5.14)$$

5.4.3 Smooth random variables and Perron-Frobenius operator

We now perform Step (i) in the proof of Theorem 5.1. The following lemma (inspired by [FV98]) expresses the expectation and distribution of smooth sequences in terms of the Perron-Frobenius operator \mathbf{H} .

Lemma 5.1. *Assume that (T_k) is a smooth sequence associated with the function f . Then, the expected value $\mathbb{E}[T_k]$ and the distribution of the random variable (T_k) are both expressed with the k -th iterate of the Perron-Frobenius operator \mathbf{H} :*

$$\mathbb{E}[T_k] = \mathbf{H}^k \left[f(x) \cdot \frac{1}{1+x} \right] (0), \quad \mathbb{P}[T_k \in J] = \mathbf{H}^k \left[\mathbf{1}_J \circ f(x) \cdot \frac{1}{1+x} \right] (0),$$

for any subinterval J of \mathcal{I} .

Proof. For each index k , consider the family of linear fractional transformations $h \in \mathcal{H}^k$. The intervals $h(\mathcal{I})$ form a partition of the interval \mathcal{I} , and the length of the interval $h(\mathcal{I})$ is expressed as a function of the continuants q_k , as

$$|h(\mathcal{I})| = \frac{1}{q_k(q_k + q_{k-1})} = \frac{1}{q_k^2} \left(\frac{1}{1 + \frac{q_{k-1}}{q_k}} \right).$$

Moreover $T_k(\alpha)$ is constant on the interval $h(\mathcal{I})$, and equal to $f(q_{k-1}/q_k)$. Finally

$$\mathbb{E}[T_k] := \int_{\mathcal{I}} T_k(\alpha) d\alpha = \sum_{h \in \mathcal{H}^k} \frac{1}{q_k^2} \ell \left(\frac{q_{k-1}}{q_k} \right) \quad \text{with} \quad \ell(x) = \frac{1}{1+x} f(x).$$

With Relation (5.13), the last expression is exactly $\mathbf{H}^k[\ell](0)$.

We now consider, for any $J \subset \mathbb{R}$, the probability $\mathbb{P}(T_k \in J) = \mathbb{E}[\mathbf{1}_J \circ T_k]$. Using the same transforms as above (now applied to the function $\mathbf{1}_J \circ f(x)$) yields

$$\mathbb{P}[T_k \in J] = \mathbf{H}^k \left[\mathbf{1}_J \circ f(x) \cdot \frac{1}{1+x} \right] (0). \quad \blacksquare$$

5.4.4 Asymptotic study of smooth variables

We now perform Step (ii) in the proof of Theorem 5.1. Since the probabilistic characteristics of the random variable T_k are expressed with the k -th iterate of the Perron Frobenius operator \mathbf{H} , their asymptotics will be related to the dominant spectral properties of this operator when it acts on the Banach space $BV(\mathcal{I})$ of the functions of bounded variation on the unit interval, and we use the decomposition (7.9).

Lemma 5.2. *The following asymptotics hold, for any smooth sequence (T_k) relative to a function $f \in BV(\mathcal{I})$:*

$$\begin{aligned}\mathbb{E}[T_k] &= \frac{1}{\log 2} \int_{\mathcal{I}} f(x) \cdot \frac{1}{1+x} dx + O(\varphi^{2k} \|f\|_{BV}), \\ \mathbb{P}[T_k \in J] &= \frac{1}{\log 2} \int_{\mathcal{I}} \mathbf{1}_J \circ f(x) \cdot \frac{1}{1+x} dx + O(\varphi^{2k}),\end{aligned}$$

for any subinterval J of \mathcal{I} . If moreover the function f is of class \mathcal{C}^1 and monotonic, with an inverse function g , the random variable T_k admits a limit density; for any interval $[a, b] \subset f(\mathcal{I})$, one has

$$\mathbb{P}[T_k \in [a, b]] = \frac{1}{\log 2} \int_a^b \frac{|g'(u)|}{1+g(u)} du + O(\varphi^{2k}) = \frac{1}{\log 2} \left| \log \frac{1+g(a)}{1+g(b)} \right| + O(\varphi^{2k}).$$

Proof. This is just an easy application of the decomposition (7.9). For the distribution, the norm $\|\mathbf{1}_J \circ f \cdot \psi\|_{BV}$ admits an upper bound which does not depend on the choice of f or J . ■

The previous lemma entails the following asymptotics for the probabilistic characteristics of the sequence $\tilde{S}_k^{(\mu)}$. Recall that the density $s^{(\mu)}$ is defined in (5.9).

Lemma 5.3. *For $\mu \in]0, 1]$, the following two asymptotic estimates, namely*

$$\mathbb{E}[\tilde{S}_k^{(\mu)}] = 1 + \frac{1}{\log 2} \frac{|\log \mu|}{1-\mu} + O\left(\frac{\varphi^{2k}}{\mu}\right), \quad \mathbb{P}\left(\tilde{S}_k^{(\mu)} \in J\right) = \int_J s^{(\mu)}(x) dx + O(\varphi^{2k}).$$

The second estimate also holds for $\mu = 0$.

Proof. This is just the application of the previous lemma for $f := f_\mu$. The function f_μ belongs to $BV(\mathcal{I})$ for $\mu > 0$, with norm $\|f_\mu\|_{BV} = O(1/\mu)$. ■

This ends Step (ii) of the proof of Theorem 5.1. The estimates (5.12) needed in Step (iii).

5.4.5 Third Step of the proof of Theorem 5.1.

We first describe Step (iii) of the proof of Theorem 5.1 that explains how to return to the initial sequence $S_k^{(\mu)}$ (involving integer parts). We also give hints to prove Theorem 5.2.

This technical lemma is useful to compare the distributions of two random variables.

Lemma 5.4. (i) *Consider two positive random variables X and Y defined on the unit interval \mathcal{I} . Assume that there exists $\varepsilon > 0$ for which one has $\|X - Y\|_0 < \varepsilon Y$. Then, for any $b > 0$, the following holds*

$$|\mathbb{P}(X \leq b) - \mathbb{P}(Y \leq b)| \leq \max\left(\mathbb{P}(b(1-\varepsilon) \leq X \leq b), \mathbb{P}(b \leq X \leq b(1+\varepsilon))\right).$$

(ii) *Consider two sequences (X_k) and (Y_k) of random variables defined on the unit interval \mathcal{I} , with values in $[0, +\infty[$. Assume the following*

- (a) *there is a real sequence $(a_k)_k$ for which $\|X_k - Y_k\|_0 \leq a_k Y_k$,*
- (b) *the sequence $(X_k)_k$ admits a limit density s with a speed of convergence c_k .*

Then the sequence Y_k satisfies

$$\mathbb{P}(Y_k \leq b_k) = \mathbb{P}(X_k \leq b_k) + O(c_k) + O\left(b_k a_k \sup\{s(x); x \in [b_k(1-a_k), b_k(1+a_k)]\}\right).$$

Proof.

Assertion (i). The inequality

$$|\mathbb{P}(B) - \mathbb{P}(A)| \leq \max \{\mathbb{P}(B \setminus A), \mathbb{P}(A \setminus B)\}$$

is an immediate consequence of the equalities

$$\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B), \quad \mathbb{P}(A) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B).$$

The inequality $Y(1 - \varepsilon) \leq X \leq Y(1 + \varepsilon)$ entails the inclusions between events

$$\{X \leq b\} \setminus \{Y \leq b\} \subset \{b(1 - \varepsilon) \leq X \leq b\}, \quad \{Y \leq b\} \setminus \{X \leq b\} \subset \{b \leq X \leq b(1 + \varepsilon)\},$$

and thus *Assertion (i)* with the choices $A = \{X \leq b\}$ and $B = \{Y \leq b\}$.

Assertion (ii). Using (a) and *Assertion (i)* entails the inequality

$$\begin{aligned} & |\mathbb{P}(X_k \leq b_k) - \mathbb{P}(Y_k \leq b_k)| \\ & \leq \max \{\mathbb{P}(b_k(1 - a_k) \leq X_k \leq b_k), \mathbb{P}(b_k \leq X_k \leq b_k(1 + a_k))\}. \end{aligned} \quad (5.15)$$

With (b), the following estimate

$$\mathbb{P}(c \leq X_k \leq d) = \int_c^d s(t) dt + O(c_k)$$

holds for any interval $[c, d]$ with a O -constant uniform with respect to the interval $[c, d]$. Then each of the two integrals in (5.15) is of the form

$$\int_{b_k - a_k b_k}^{b_k} s(t) dt + O(c_k), \quad \int_{b_k}^{b_k + a_k b_k} s(t) dt + O(c_k),$$

and is $O(b_k a_k \sup \{s(x); x \in [b_k(1 - a_k), b_k(1 + a_k)]\}) + O(c_k)$. ■

Lemma 5.5. For any $\mu \in [0, 1]$, the probabilistic characteristics of the random variable $S_k^{(\mu)}$ and its smoothed version $\tilde{S}_k^{(\mu)}$ are related as follows:

(i) The difference between the two random variables satisfies

$$0 \leq S_k^{(\mu)} - \tilde{S}_k^{(\mu)} \leq \varphi^{k-3} \tilde{S}_k^{(\mu)}.$$

(ii) The following holds for the expectations

$$\mathbb{E}[S_k^{(\mu)}] = \mathbb{E}[\tilde{S}_k^{(\mu)}](1 + O(\varphi^k)).$$

(iii) The following holds for the distributions, for any $b \in I_\mu$,

$$\mathbb{P}(S_k^{(\mu)} \leq b) - \mathbb{P}(\tilde{S}_k^{(\mu)} \leq b) = O\left(\frac{\varphi^k}{b}\right) + O(\varphi^{2k}).$$

Proof.

Assertion (i). Denote by Φ the Golden ratio $\Phi = \frac{1+\sqrt{5}}{2}$. The inequalities (see Observation 1.2)

$$0 \leq \tilde{n}_k^{(\mu)} - n_k^{(\mu)} \leq 1 \quad \text{and} \quad \tilde{n}_k^{(\mu)} \geq n_k^{(\mu)} \geq q_{k-1} \geq \Phi^{k-3}$$

show that

$$0 \leq S_k^{(\mu)} - \tilde{S}_k^{(\mu)} = \frac{q_k + q_{k-1}}{n_k^{(\mu)} \cdot \tilde{n}_k^{(\mu)}} (\tilde{n}_k^{(\mu)} - n_k^{(\mu)}) \leq \frac{1}{q_{k-1}} \frac{q_k + q_{k-1}}{\tilde{n}_k^{(\mu)}} \leq \varphi^{k-3} \tilde{S}_k^{(\mu)}.$$

Assertion (ii). This is a clear consequence of *Assertion (i)*.

Assertion (iii). It follows from *Assertion (i)* of the present Lemma together with *Assertion (ii)* of the previous Lemma 5.4. ■

5.4.6 Comparison between densities $s^{(\mu)}$ and $s^{(0)}$.

Lemma 5.6. *Consider $\mu \in [0, 1/2[$. The difference between the densities $s^{(\mu)}$ and $s^{(0)}$ satisfies on the interval $[3, \infty[$*

$$|s^{(\mu)}(x) - s^{(0)}(x)| \leq \mu \frac{1}{[x(1-\mu) + \mu - 2]^2} \quad \text{for } x \in I_\mu$$

$$|s^{(\mu)}(x) - s^{(0)}(x)| = s^{(0)}(x) \quad \text{for } x \geq 1 + \frac{1}{\mu}$$

The difference between the two distributions satisfies:

$$\left| \int_3^b s^{(\mu)}(x) dx - \int_3^b s^{(0)}(x) dx \right| = O(\mu)$$

where the O -constant does not depend on b .

5.4.7 End of the proof of Theorem 5.2

For Theorem 5.2, we now consider a sequence of positions (μ_k) , and as our estimates of Theorem 5.1 are uniform with respect to μ and k , it is possible to deal with a μ which depends on k . Moreover, the previous lemma is useful to compare the distributions of the variables $S_k^{(\mu)}$ and $S_k^{(0)}$.

CHAPTER 6

COMPARISON BETWEEN THE MODELS AND SPECIAL FAMILIES OF SLOPES

In this chapter we take a step back and seek to compare and find connections between the models and methods from chapters 4 and 5, namely the models with large fixed n and with large fixed k . Further, we study two families of special slopes: the rationals and the quadratic irrationals. This is ongoing work, and we shall point out the expected results and research directions taken. These studies have shown connections with the previous two models, and in particular to the link between the methods employed.

6.1 Relation between the two models.

We wish to relate the two (asymptotic) models: the model “with fixed large n ” and the model “with fixed large k ”? Of course, these two models should be close if the behavior of the sequence $k \mapsto q_k(\alpha)$ does not depend too strongly on α , and we know that it is not the case. However, the behavior of the sequence $k \mapsto \log q_k(\alpha)$ is much more regular, as it is well known (see for instance [Khi97]) that

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log q_k(\alpha) = L = \frac{\pi^2}{12 \log 2} \quad \text{for almost all } \alpha. \quad (6.1)$$

Consider first the present model “with n fixed”, and a sequence $\ell \mapsto n(\ell) = \tau^\ell$. Then Theorem 4.3 reads

$$\mathbb{E}[S_{n(\ell)} \mid \mu_{n(\ell)} \geq \tau^{-\ell}] \sim \left[\frac{12}{\pi^2} \log \tau \right] \ell. \quad (6.2)$$

Furthermore, as $n(\ell)$ belongs to the interval $[q_{k-1}(\alpha), q_k(\alpha)]$, the existence of the limit for the quotient $q_k(\alpha)/n$, that holds for almost any α , and is recalled in (6.1) entails the relation between the index ℓ and the index $k := k(\alpha, n(\ell))$, that holds for almost any α , namely

$$\log n(\ell) = \ell \log \tau \sim \frac{\pi^2}{12 \log 2} k(\alpha, n(\ell)). \quad (6.3)$$

Now, we deal with the model “with k fixed”, and we consider that the index $k(\alpha, n(\ell))$ satisfies (6.3) *everywhere*. Then, the application of the result in the model “with k fixed”, described in (5.10) should entail

$$\mathbb{E}[S_k^{(\tau^k)}] \sim \left[\frac{1}{\log 2} \log \tau \right] k \sim \left[\frac{12}{\pi^2} \log^2 \tau \right] \ell. \quad (6.4)$$

Remark that the conditional events are not the same in the two equations (6.2) and (6.4):

- in (6.2), the event is $\{\alpha \mid \mu(\alpha, n(\ell)) \geq \tau^{-\ell}\}$,
- in (6.4) the event is $\{\alpha \mid \mu(\alpha, n(\ell)) \sim \tau^{-\ell}\}$.

This (heuristic) comparison exhibits in both cases a linear growth with respect to ℓ . However, the events of interest are not the same, and we have considered that the index $k(\alpha, n(\ell))$ satisfies (6.3) everywhere.

6.2 Relationship between the techniques employed

The techniques may seem fundamentally distinct at first sight, we claim, however, that they are closely related to one another.

6.2.1 From Riemann sums to the Transfer Operator

It may come away as a surprise that we may express the results with fixed $n \rightarrow \infty$ from Chapter 4 (our article [RV17]) in terms of the transfer operator. At first sight the results and tools are fundamentally different: a variable depth for the continued fraction expansion against a fixed depth.

This surprising link is given by the use of the Mellin transform [FS09] [FGD⁺95]. The Mellin transform of our coprime Riemann sums will be given in terms of the quasi-inverse of the transfer operator of the Euclidean system. The appearance of the quasi-inverse is natural: we must filter (in terms of n) from all possible branches.

We recall the definition of the Mellin transform [FGD⁺95], and later mention the key properties that interest us. For the Mellin transform method to work, we require several conditions regarding the behavior of the quasi-inverse in a band around its dominant singularity at $s = 1$. We expect, expect that this should be doable by an application of the Dolgopyat-Baladi-Vallée estimates (see Theorem 6.1) we cite below. These estimates give, precisely, a good understanding of the behavior of the quasi-inverse on a band around $s = 1$.

Definition 6.1 (Mellin Transform). Let $f(x)$ be locally Lebesgue integrable over $(0, \infty)$. The Mellin transform of $f(x)$ is defined by

$$\mathcal{M}[f(x); s] := \int_0^\infty f(x)x^{s-1}dx.$$

Once we will have calculated the Mellin transform of our target function, we will explain briefly how to exploit this knowledge to get back the the distributions from Chapter 4.

Context and notation. In this chapter we write $\rho_k = q_{k-1}/q_k$ and $y = q_k/n$, where $t = 1/n$.

Consider a \mathcal{Q} -function Λ , we may write

$$\Lambda(\alpha, n) = f_\Lambda \left(\frac{q_{k-1}(\alpha)}{q_k(\alpha)}, \frac{q_k(\alpha)}{n} \right)$$

for a certain f_Λ , whenever $q_{k-1}(\alpha) \leq n < q_k(\alpha)$. We recall from Chapter 4 that we write $\Lambda_n(\alpha) := \Lambda(\alpha, n)$ to denote the random variable for a fixed n .

Then, following Chapter 4, we are interested in the behaviour of

$$F_\Lambda(t) := t^2 \sum_{k \geq 0} \sum_{\mathbf{m} \in \mathbb{N}^k} \frac{1}{(tq_k)^2(1 + \rho_k)} f_\Lambda(\rho_k, tq_k)$$

as $t \rightarrow 0$. The reason being that this sum gives the expectation of Λ_n when α is distributed uniformly, namely

$$F(1/n) = \mathbb{E}[\Lambda_n].$$

Notice that here $t \rightarrow 0$ is a continuous version of $1/n$.

That is why we are interested, more generally, in sums of the form

$$F(t) := t^2 \sum_{k \geq 0} \sum_{\mathbf{m} \in \mathbb{N}^k} f(\rho_k, tq_k). \quad (6.5)$$

In particular, we consider the choice

$$f(\rho, y) = \frac{1}{y^2(\rho + 1)} \mathbb{I}[f_\Lambda(\rho, y) \leq \lambda, \rho y \leq 1 < y], \quad (6.6)$$

which gives the probability density at λ , namely $F(1/n) = \mathbb{P}(\Lambda_n \leq \lambda)$.

The connection with \mathbf{H}_s . We begin by operating formally. Taking Mellin Transform of $F(t)$ in (6.5) and exchanging the sums and the integrals we have

$$\int_0^\infty F(t)t^{s-1} dt = \sum_k \sum_{\mathbf{m}} \int_0^\infty f(\rho_k, tq_k)t^{s+1} dt.$$

We make a change of variables $t \mapsto tq_k$ in each integral, getting

$$\int_0^\infty F(t)t^{s-1} dt = \sum_k \sum_{\mathbf{m}} q_k^{-(s+2)} \int_0^\infty f(\rho_k, t)t^{s+1} dt.$$

At this point we recall from Chapter 5 that we have identity (5.13), transforming a sum in p_k/q_k into a sum on q_{k-1}/q_k . For our case this gives

$$\sum_{\mathbf{m}} q_k^{-(s+2)} \int_0^\infty f(\rho_k, t)t^{s+1} dt = \mathbf{H}_{s/2+1}^k \left[x \mapsto \int_0^\infty f(x, t)t^{s+1} dt \right] (0),$$

as the variable s accompanies $q_k^{-2} = |h'(0)|$.

Thus, summing in k , we get our Mellin transform in terms of the quasi-inverse

$$\mathcal{M}[F(t); s] = \int_0^\infty F(t)t^{s-1} dt = (\mathbf{I} - \mathbf{H}_{s/2+1})^{-1} \left[x \mapsto \int_0^\infty f(x, t)t^{s+1} dt \right] (0). \quad (6.7)$$

This is our key relation; the Mellin transform takes us from our coprime Riemann sums from Chapter 4 to a quasi-inverse of the transfer operator, at least formally. This is in principle a moral relation between the two, we explain below why we expect this to also be an effective link.

Analytic properties. We observe in (6.7) that the integral is well-defined for any s when we take f of the form (6.6) (which yields the densities), because the domain of integration is limited to the compact $[1, 1/x]$. Moreover, the function

$$x \mapsto \int_0^\infty f(x, t)t^{s+1} dt$$

is of class \mathcal{C}^1 as soon as the frontier $\{(\rho, y) : f_\Lambda(\rho, y) = \lambda, \rho y \leq 1 < y\}$ is of class \mathcal{C}^1 .

Now we turn to the quasi-inverse itself. For this we cite the following result (see [BV05]).

Theorem 6.1 (Dolgopyat-Baladi-Vallée estimates). *Consider the transfer operator \mathbf{H}_s associated with the Euclidean dynamical system when it acts on the functional space $\mathcal{C}^1(\mathcal{I})$ endowed with the family of norms $\|\cdot\|_{1,t}$ defined for $t \neq 0$ as*

$$\|f\|_{1,t} := \sup |f| + (1/|t|) \sup |f'|.$$

There is a real neighborhood Σ_1 of 1, an exponent $\xi > 0$, a real $t_0 > 0$ and a constant $M > 0$ such that, for $s = \sigma + it$, the operator $(I - \mathbf{H}_s)^{-1}$ satisfies the following :

(i) *For $\sigma \in \Sigma_1$, the operator $(I - \mathbf{H}_s)^{-1}$ has a unique pole, simple, and located at $s = 1$.*

(ii) *For $\sigma \in \Sigma_1$ and $|t| > t_0$, the operator $(I - \mathbf{H}_s)^{-1}$ satisfies*

$$\|(I - \mathbf{H}_s)^{-1}\|_{1,t} \leq M \cdot |t|^\xi$$

With this estimates, we expect the Mellin transform $\mathcal{M}[F(t); s]$ to be meromorphic on $\Re s > -a$ for some $a > 0$, and a unique pole at $s = 0$ where

$$\mathcal{M}[F(t); s] \sim \frac{1}{s} \frac{12}{\pi^2} \int_0^1 \left(\int_0^\infty f(x, t) t dt \right) dx, \quad (6.8)$$

due to (1.63).

The relation between the behavior of the Mellin transform on the singularities on the left of the convergence strip, and the asymptotics for $F(t)$ as $t \rightarrow 0$, are given precisely in Theorem 4 and Figure 4 from [FGD⁺95]. In short, a singularity of type $\mathcal{M}[F(t); s] \sim \frac{C}{s^{k+1}}$ as $s \rightarrow 0$ will translate into an asymptotic of type $F(t) \sim C \frac{(-1)^k}{k!} (\log t)^k$ as $t \rightarrow 0$. This is true provided that $\mathcal{M}[F(t); s]$ can be extended meromorphically left-wise to a strip $\langle -a, \infty \rangle$ where $a > 0$ and $s = 0$ is the sole singularity, and the transform also decreases fast enough on the lines $c \pm i\infty$ for c on a strip around 0 (which we expect due to Theorem 6.1).

Then we would deduce

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_k \sum_{m \in \mathbb{N}^k} f \left(\frac{q_{k-1}}{q_k}, \frac{q_k}{n} \right) = \lim_{t \rightarrow 0} F(t) = \frac{12}{\pi^2} \iint_{[0,1] \times [0,\infty)} f(\rho, y) y d\rho dy, \quad (6.9)$$

which would yield again the results from Chapter 4 when f is of the form (6.6).

6.2.2 Riemann sums resembling the Perron-Frobenius operator

When considering the fixed $n \rightarrow \infty$ model, the depth^a $k(\alpha, n)$ of the continued fraction expansion needed is apparently “free”, contrary to the fixed $k \rightarrow \infty$ model where we actually prescribe it. This is why a priori we expect a relation with the quasi-inverse of the transfer operator.

It is not completely true, however, that the depth is “free”; we know that for almost every α there is the limit

$$\lim \frac{1}{k} \log q_k(\alpha) = \frac{\pi^2}{12 \log 2},$$

hence we derive that for $k = k(\alpha, n)$ satisfying $q_{k-1}(\alpha) \leq n < q_k(\alpha)$, we have

$$\lim_{n \rightarrow \infty} \frac{\log n}{k(\alpha, n)} = \frac{\pi^2}{12 \log 2}$$

almost surely.

This motivates the intuition that $k(\alpha, n)$ is, in general, fairly close to $\frac{12 \log 2}{\pi^2} \log n$.

Hence, the idea now is to reverse the construction of Chapter 5. By choosing an interval $[\log n, \log n + \theta]$ to which $\log q_k$ must belong, we “control” the size of k through n . With this in mind we have the following

^aRecall that $k = k(\alpha, n)$ is the unique positive integer satisfying $q_{k-1}(\alpha) \leq n < q_k(\alpha)$.

Proposition 6.1. *Let f be Riemann-integrable and $\theta > 0$, then*

$$\lim_{N \rightarrow \infty} \sum_{k \geq 1} \sum_{m_1, \dots, m_k \geq 1} \frac{1}{q_k} f\left(\frac{q_{k-1}}{q_k}\right) \llbracket N < \log q_k \leq N + \theta \rrbracket = \frac{12\theta}{\pi^2} \int_0^1 f(x) dx.$$

Proof. This is just a generalization of Proposition 4.4. ■

Observe how the expression from Proposition 6.1 resembles that of the Perron-Frobenius operator \mathbf{H} . When we choose θ to be the entropy $\theta = \frac{\pi^2}{12 \log 2}$, which makes the expected number of continuants q_k with $q_k \in (N, N + \theta]$ be 1 by Proposition 4.4, we retrieve

$$\lim_{N \rightarrow \infty} \sum_{k \geq 1} \sum_{m_1, \dots, m_k \geq 1} \frac{1}{q_k} f\left(\frac{q_{k-1}}{q_k}\right) \llbracket N < \log q_k \leq N + \frac{\pi^2}{12 \log 2} \rrbracket = \frac{1}{\log 2} \int_0^1 f(x) dx.$$

6.3 Special families of slopes

Now we present the study of two important families of slopes: the rationals and the quadratic irrationals. It turns out that the method employed here for rationals and quadratic irrationals are very much related, and in fact, related to those of section 6.2.1. This is a subject of ongoing research and we hope to be able to develop a unified approach for the study of \mathcal{Q} -functions for all 3 families of slopes (real, rational and quadratic irrational).

6.3.1 Eventually periodic words: α rational

We consider the rectangle $\mathcal{R} := [0, 1] \times [1, \infty)$ and we recall that a \mathcal{Q} -function Λ is associated with a function $f_\Lambda : \mathcal{R} \rightarrow \mathbb{R}_{\geq 0}$ when, given an irrational slope $\alpha \in [0, 1]$,

$$\Lambda(\alpha, n) := f_\Lambda\left(\frac{q_{k-1}(\alpha)}{n}, \frac{q_k(\alpha)}{n}\right), \quad \text{for } n \in [q_{k-1}(\alpha), q_k(\alpha)).$$

The definition extends to a rational α of the form $\alpha = c/d$ with a coprime pair (c, d) , as soon as the integer n satisfies $n < d$.

Given a fixed an integer n , we are interested in the distribution of the function $(c/d) \mapsto \Lambda(c/d, n)$. More precisely, we consider the subsets

$$\Omega_D = \{(c, d) : 1 \leq c \leq d \leq D; \gcd(c, d) = 1\}, \quad R_D(\varrho, n) = \left\{ (c, d) \in \Omega_D : \Lambda\left(\frac{c}{d}, n\right) \leq \varrho \right\},$$

and we wish to study the distribution of the function Λ on the set Ω_D

$$\text{Pr}_D \left((c, d) \in \Omega_D \mid \Lambda\left(\frac{c}{d}, n\right) \leq \varrho \right) = \frac{|R_D(\varrho, n)|}{|\Omega_D|}.$$

Our main expected result for the rationals is as follows:

Expected Theorem 6.2. *Consider a continuant function Λ associated with a function f . Let the density ψ and the domain $\Delta_f(\varrho)$ be defined by*

$$\psi : \mathcal{R} \rightarrow \mathbb{R}, \quad \psi(x, y) = \frac{2}{\zeta(2)} \frac{1}{y(x+y)}, \quad \Delta_f(\varrho) = \{(x, y) \in \mathcal{R} \mid f(x, y) \leq \varrho\}.$$

Then, the distribution of the function $\Lambda(\cdot, n)$ on the set Ω_D involves the coprime Riemann sum (see in Section 4.3.2) $\widehat{R}_n[\psi; \Delta_f(\varrho)]$ of step $1/n$ of the function ν over the domain $\Delta_f(\varrho)$ under the form

$$\Pr_D[\Lambda(\cdot, n) \leq \varrho] = \zeta(2) \widehat{R}_n[\psi; \Delta_f(\varrho)] + \epsilon_D(n, \varrho) \quad \epsilon_D(n, \varrho) \xrightarrow{D \rightarrow \infty} 0.$$

Moreover, the coprime Riemann sum is related to the integral $I[\psi; \Delta_f(\varrho)]$ of the function ψ over the domain $\Delta_f(\varrho)$ with the estimate

$$\zeta(2) \widehat{R}_n[\psi, \Delta_f(\varrho)] = I[\psi; \Delta_f(\varrho)] + O\left(\frac{1}{n^2}\right).$$

Note that this means that, as the bound D for the denominators increases, the distribution approaches the same exact distribution we have for the real numbers.

One further question we are interested in would be the following: could we get good *mixed* error bounds in D and n which would allow us to make both tend to infinite at the same time, but in some prescribed way?

Proof elements. We consider the subsets

$$\omega_d := \{(c, d) : 1 \leq c \leq d; \gcd(c, d) = 1\}, \quad r_d(\varrho, n) := \left\{ (c, d) \in \omega_d : \Lambda\left(\frac{c}{d}, n\right) \leq \varrho \right\}$$

and the two generating functions

$$F_{n, \varrho}(s) = \sum_{d \geq 1} \frac{1}{d^s} |r_d(\varrho)|, \quad F(s) := \sum_{d \geq 1} \frac{1}{d^s} |\omega_d| = \frac{\zeta(s-1)}{\zeta(s)}. \quad (6.10)$$

With the two relations

$$\Omega_D = \bigcup_{d \leq D} \omega_d, \quad R_D(\varrho, n) = \bigcup_{d \leq D} r_d(\varrho),$$

we remark that the probability of interest is expressed as a ratio, where the numerator involves the coefficients of the series $F_{n, \varrho}(s)$ whereas the denominator involves the coefficients of the series $\zeta(s-1)/\zeta(s)$.

Here comes the key step of the argument. Once we have determined the partial quotients $\mathbf{m} = (m_1, \dots, m_k)$ which make $q_{k-1}(\mathbf{m}) \leq n < q_k(\mathbf{m})$, we have all we need to compute our \mathcal{Q} -function Λ ; that is, any element $\alpha \in \mathcal{I}_{m_1, \dots, m_k}$ will give the same value of $\Lambda(\alpha, n)$. Thus, any element in the image of h_{m_1, \dots, m_k} gives the same value of Λ_n (provided that $q_{k-1}(\mathbf{m}) \leq n < q_k(\mathbf{m})$).

A number from $h_{\mathbf{m}}(\mathcal{I})$ is rational if and only if it is of the form $c/d = h_{m_1, \dots, m_k}(\lambda/\mu)$ for some integers $1 \leq \lambda \leq \mu$. We note that for $\mathbf{m} = (m_1, \dots, m_k)$

$$h_{m_1, \dots, m_k}(\lambda/\mu) = \frac{\mu p_k(\mathbf{m}) + \lambda p_{k-1}(\mathbf{m})}{\mu q_k(\mathbf{m}) + \lambda q_{k-1}(\mathbf{m})},$$

and the right-hand side fraction is reduced if and only if λ/μ is reduced (i.e., $\gcd(\lambda, \mu) = 1$). We are only interested in the reduced denominator for our Dirichlet series (which is marked with s).

Therefore we may rewrite

$$F_{n,\varrho}(s) = \sum_{k \geq 1} \sum_{\mathbf{m} \in \mathbb{N}^k} \left[\left[q_{k-1}(\mathbf{m}) \leq n < q_k(\mathbf{m}), f_\Lambda \left(\frac{q_{k-1}(\mathbf{m})}{n}, \frac{q_k(\mathbf{m})}{n} \right) \leq \varrho \right] \right. \\ \left. \sum_{(\lambda, \mu): \lambda \leq \mu, \gcd(\lambda, \mu)} \frac{1}{(\mu q_k(\mathbf{m}) + \lambda q_{k-1}(\mathbf{m}))^s} \right]$$

Here we could actually use the same ideas as before with coprime Riemann sums, and that was actually our first approach. There would be one coming from the pairs (q_{k-1}, q_k) which traverse all coprime pairs with $q_{k-1} < q_k$, and another one for (μ, λ) inside.

We now explain an approach following principles similar to those of section 6.2.1. The principle is that our whole Dirichlet sum may be rewritten in terms of the quasi-inverse operator of the Euclidean dynamical system. Then we get a situation similar to that of section 6.2.1 but with Dirichlet series instead of a Mellin transform. The analogous asymptotics are derived from an application celebrated Perron formula (see [Apo98] or [Ten15]).

We seek to rewrite the inner sum in terms of the quasi-inverse operator of the Euclidean dynamical system. In order to do this, the expressions involved will be rewritten in terms of inverse branches.

Each fraction λ/μ is produced uniquely as $h(0)$ for an inverse branch $g \in \mathcal{H}^* \cdot \{h_1\}$ (i.e., ending in a quotient 1, or we could also work with all inverse branches not ending in a quotient 1). For simplicity in the explanation we consider all branches $g \in \mathcal{H}^*$ (we are going twice over every fraction except for 1/1). Next, we consider the inverse branch $h = h_{\mathbf{m}}$.

Notice that

$$\frac{1}{(\mu q_k(\mathbf{m}) + \lambda q_{k-1}(\mathbf{m}))^s} = |(h \circ g)'(0)|^{s/2}.$$

Finally, to get everything in terms of inverse branches, we must introduce an auxiliary function b working on the set of inverse branches (notice that this is well-defined) taking h_{m_1, \dots, m_k} to $h_{m_1, \dots, m_{k-1}}$, i.e.,

$$b(h_{m_1, \dots, m_k}) := h_{m_1, \dots, m_{k-1}}.$$

With this auxiliary function, we notice that for $\mathbf{m} \in \mathbb{N}^k$ and $g = h_{\mathbf{m}} \in \mathcal{H}^k$, we have

$$q_k(\mathbf{m}) = |g'(0)|^{-1/2}, \text{ and } q_{k-1}(\mathbf{m}) = |(b(g))'(0)|^{-1/2}.$$

For short, write

$$A_g(x, n, \varrho) := \left[\left[|(b(g))'(x)|^{-1/2} \leq n < |g'(x)|^{-1/2}, f_\Lambda \left(\frac{|(b(g))'(x)|^{-1/2}}{n}, \frac{|g'(x)|^{-1/2}}{n} \right) \leq \varrho \right] \right]. \quad (6.11)$$

Thus we get (forgetting about the $(\lambda, \mu) = (1, 1)$, hence the \sim)

$$2F_{n,\varrho}(s) \sim \sum_{h \in \mathcal{H}^*} A_h(0, n, \varrho) \sum_{g \in \mathcal{H}^*} |(h \circ g)'(0)|^{s/2}. \quad (6.12)$$

Here we notice that $h \circ g$ traverses all inverse branches of the form $h \cdot \mathcal{H}^*$, and hence corresponds to the transfer operator $(\mathbf{I} - \mathbf{H}_{s/2})^{-1} \circ \mathbf{H}_{[h], s/2}$ with $\mathbf{H}_{[h], s}[f](x) := |h'(x)|^s f(h(x))$.

Thus we write

$$2F_{n,\varrho}(s) \sim \sum_{h \in \mathcal{H}^*} A_h(0, n, \varrho) (\mathbf{I} - \mathbf{H}_{s/2})^{-1} \circ \mathbf{H}_{[h], s/2}[1](0).$$

One step further

$$2F_{n,\varrho}(s) \sim (\mathbf{I} - \mathbf{H}_{s/2})^{-1} \left[\sum_{h \in \mathcal{H}^*} A_h(0, n, \varrho) \mathbf{H}_{[h],s/2}[1] \right] (0). \quad (6.13)$$

Here the pole is found at $s = 2$, where the quotient with the pole of $F(s)$ gives us

$$\sum_{h \in \mathcal{H}^*} A_h(0, n, \varrho) \int_{\mathcal{I}} \mathbf{H}_{[h],1}[1](t) dt = \sum_{h \in \mathcal{H}^*} A_h(0, n, \varrho) |\mathcal{I}_h|, \quad (6.14)$$

the same expression from the “real case” from Chapter 4. We expect that it should be possible to combine the Perron formula with the Dolgopyat-Baladi-Vallée estimates (see Theorem 6.1) to assert that this actually is the limit of the probabilities as $\mathbb{P}_D[\Lambda(\cdot, n) \leq \varrho]$ as $D \rightarrow \infty$.

6.3.2 Sturmian words and morphisms: α quadratic irrational

Now we get back to the quadratic irrationals introduced in Section 1.5, in particular to the reduced quadratic irrationals α (rqi for short), which are those whose continued fraction expansion is purely periodic. We recall that there is a natural notion of size $\epsilon(\alpha)$ associated with such numbers α . Our objective is to study, in distribution, what happens with the recurrence quotient $S(\alpha, n)$ defined in (3.13), when α is drawn uniformly at random from the set $\mathcal{K}_N = \{\alpha \text{ rqi} : \epsilon(\alpha) \leq N\}$ with large N .

6.3.3 The size of a quadratic irrational and the model

The classical notion of size ϵ given in Equation 1.67 depends strongly on the more basic notion of size $\nu(\alpha)^{-1}$ defined in (1.64), hence we start by reminding the definition of ν .

We recall that

$$\nu(\alpha) := \prod_{i=0}^{k-1} T^i(\alpha) = |h'(\alpha)|^{1/2},$$

where $h \in \mathcal{H}^k$ is the inverse branch associated with the minimal period $m = (m_1, \dots, m_k)$ of α , which has length $k = p(\alpha)$. We recall that ν is related to the growth of the continuants of α as we accumulate periods

$$\nu(\alpha) = \lim_{\ell \rightarrow \infty} \frac{q(m^\ell)}{q(m^{\ell+1})}.$$

Defined ν , note that

$$\epsilon(\alpha) = \nu(\alpha)^{-r(\alpha)}, \quad \text{with } r(\alpha) = 1 \text{ for even } p(\alpha), \text{ and } r(\alpha) = 2 \text{ for odd } p(\alpha).$$

Therefore, it will be enough to do the analysis for ν , as this will then translate into the results for ϵ . We consider, then, rather than \mathcal{K}_N , the sets

$$\mathcal{M}_N := \left\{ \alpha \text{ rqi} : \frac{1}{\nu(\alpha)} \leq N \right\},$$

and in our new model we shall pick α uniformly at random from \mathcal{M}_N rather than \mathcal{K}_N . We shall use henceforth \mathbb{P}_N and \mathbb{E}_N for the corresponding probabilities and expected values.

6.3.4 The generating functions

Primitive periods. As we consider all possible periods $m = (m_1, \dots, m_k)$, there are clearly periods that define the same periodic word. This is why we pick the minimal, so-called “primitive”, periods: periods $m = (m_1, \dots, m_k)$ that are not themselves the power (repetition) of a smaller period.

We observe then that if \mathcal{P} is the set of primitive periods, we have the decomposition

$$\mathbb{N}^* = \bigsqcup_{\ell=1}^{\infty} \mathcal{P}^\ell, \quad (6.15)$$

where $\mathcal{P}^\ell := \{w^\ell : w \in \mathcal{P}\}$. Indeed, any period m is a repetition of its smaller period, which is unique.

The expression (6.15) will be key, because working with the generating function over all periods in \mathbb{N}^* is in general easier much easier than working over \mathcal{P} .

Working over all possible periods. So far we have defined v over the r.q.i. α , but for the sake of having “nice” expressions for our GFs, it will be useful to extend it to words. We could think of a r.q.i. α as its primitive period (m_1, \dots, m_k) . Given now an arbitrary period $m = (m_1, \dots, m_k)$ which need not be primitive, we still define

$$v(m) := |h'_m(x_m^*)|^{1/2},$$

and the previous properties that held for a r.q.i. α still hold for this new extension, namely

$$v(m) = \lim_{\ell \rightarrow \infty} \frac{q(m^\ell)}{q(m^{\ell+1})}, \quad v(m) = \prod_{i=0}^{k-1} T_g^i(x_m^*).$$

We remark then the key multiplicative property

$$v(m^\ell) = (v(m))^\ell, \quad (6.16)$$

for every $m \in \mathbb{N}^*$ and $\ell \geq 1$.

Target generating functions. We wish to study the occurrence of $S(n, x_m^*) \leq \lambda$ over all periods $m \in \mathbb{N}^*$ that are primitive and $\frac{1}{v(x_m^*)} \leq N$. To study such an event, due to the Tauberian Theorem, we consider in first instance (we shall later adapt it!) the DGF

$$P_n(s) := \sum_{m \in \mathcal{P}} (v(m))^s \mathbb{I}[S(x_m^*, n) \leq \lambda].$$

The decomposition in Equation 6.15 works its way into our final generating functions, giving

$$G_n(s) := \sum_{m \in \mathbb{N}^*} (v(m))^s \mathbb{I}[S(x_m^*, n) \leq \lambda] = P_n(s) + \sum_{\ell \geq 2} P_n(\ell s).$$

The power to the ℓ in (6.15) then shifts the dominant singularities of our DGFs to the left, thanks to (6.16). From this fact we deduce that the leading terms of the asymptotics for the DGF over \mathbb{N}^* and \mathcal{P} coincide. Thus, generally, it will be enough to take all possible periods from \mathbb{N}^* .

6.3.5 The number of complete cycles ℓ

Our current plan, i.e., working with $P_n(s)$, actually is slightly “flawed”. Let us recall briefly how $S(\alpha, n)$ is computed:

$$S(\alpha, n) = 1 + \frac{q_{k-1}(\alpha)}{n} + \frac{q_k(\alpha)}{n},$$

where $k = k(\alpha, n)$ is such that $q_{k-1}(\alpha) \leq n < q_k(\alpha)$. Since $q_k(\alpha) \geq 2^{(k-1)/2}$ for every α , then this means that $k = k(\alpha, n) \leq 2 \log_2 n + 2$, which is fixed when n is fixed. On the other hand, whenever $S(\alpha, n) \leq \lambda$, then $q_k(\alpha) \leq \lambda n$ and this implies that the quotients are bounded $m_i \leq \lambda n$ for all $i = 1, \dots, k(\alpha, n)$.

Hence only a finite number of primitive tuples $m \in \mathcal{P}$ which produce $S(x_m^*, n) \leq \lambda$ satisfy that $k(x_m^*, n)$ is larger than the period $|m|$ (the length of the tuple m). Thus for most of the periods $m \in \mathcal{P}$, satisfying $S(x_m^*, n) \leq \lambda$, we never make use of the periodicity of the continued fraction expansion of x_m^* in order to determine the value of $S(x_m^*, n)$. In such a case we say that n occurs during the first period of x_m^* . This is not a desirable situation as this prevents us from observing the incidence of the periodic nature of the expansion on the probabilistic study.

We formalize this notion by introducing an appropriate index ℓ that gives the number of *complete* “cycles” or periods of m that are needed in the calculation of $S(x_m^*, n)$. To define what we mean by “needed in the calculation of $S(\alpha, n)$ ”, we say that a prefix m of the continued fraction expansion of α is needed in the calculation of $S(\alpha, n)$ if $k(\alpha, n) \leq |m|$.

Definition 6.2 (Number of cycles ℓ). Given a tuple $m = (m_1, \dots, m_p) \in \mathbb{N}^p$ and an integer $n \in \mathbb{N}$ we define the number of cycles $\ell = \ell(m, n)$ of m around n to be the unique non-negative integer ℓ such that

$$q(m^\ell) \leq n < q(m^\ell m_1 \dots m_{p-1})$$

is satisfied.

Suppose $\alpha = x_m^*$, and that $v = w_1 \dots w_k$ is the prefix of $m \cdot m \cdot \dots$ needed to compute $S(\alpha, n)$. Then we may decompose $v = m^\ell u$ where $u \neq \epsilon$ is a prefix of m and we then write $u \preceq m$.

We distinguish two cases:

1. When $\ell = 0$, the word v is a prefix of m and we may then complete the rest of the period after v as we like without having this change the value of S .
2. When $\ell > 0$, the word v has interdependencies given by the period of m .

6.3.6 Generating function for the first cycle: $\ell = 0$

Let us start by studying the probabilities with the added condition that $\ell = 0$. This case is still interesting and introduces several tools used in the study of quadratic irrationals.

We recall that $\ell = 0$ is equivalent to the condition $k(x_m^*, n) \leq |m|$. Adding this condition to the definition of $G_n(s)$ gives a new DGF

$$F_n(s) := \sum_{m \in \mathbb{N}^*} (v(m))^s \llbracket S(x_m^*, n) \leq \lambda, \quad k(x_m^*, n) \leq |m| \rrbracket. \quad (6.17)$$

Here we introduce a w that will represent the actual (maximal) prefix of $m \cdot m \cdot m \dots$ needed to compute $k(x_m^*, n)$, i.e., $k(x_m^*, n) = |w|$. Thus we write

$$F_n(s) = \sum_{m \in \mathbb{N}^*} (v(m))^s \sum_{w: \epsilon \preceq w \preceq m} \llbracket S(x_w^*, n) \leq \lambda, \quad k(x_m^*, n) = |w| \rrbracket,$$

by our extra condition $|w| = k(x_m^*, n) \leq |m|$ (which implies $S(x_w^*, n) = S(x_m^*, n)$).

Note that reversing the sums gives

$$F_n(s) = \sum_{w:\epsilon \not\geq w} \llbracket S(x_w^*, n) \leq \lambda, \quad k(x_w^*, n) = |w| \rrbracket \sum_{m:w \leq m} (v(m))^s.$$

It turns out that one may actually express the sums with $v(m)$ in terms of the so-called *trace* of the transfer operator. On the other hand, the condition in the Iverson bracket may be rewritten in a more familiar form to get

$$F_n(s) = \sum_{k \geq 1} \sum_{w \in \mathbb{N}^k} \llbracket 1 + \frac{q_{k-1}(w) + q_k(w)}{n} \leq \lambda, \quad q_{k-1}(w) \leq n < q_k(w) \rrbracket \sum_{m:w \leq m} (v(m))^s. \quad (6.18)$$

This equation is reminiscent of (6.12) from the rational case. The difference between the two is the presence of $v(m) = |h'_m(x_m^*)|^{1/2}$ (this case!) instead of $q(m) = |h'_m(0)|^{1/2}$ (rational case!).

We are now going to rewrite this expression again in terms of a quasi-inverse. In order to do this we introduce the notion of the trace of an operator.

Generating the sizes: the trace of an operator In this section we cite [Val98b] and [CV17] for the main properties of the *trace* of the transfer operator, as well as its components, over an appropriate space $\mathcal{A}_\infty(\mathcal{V})$. The space $\mathcal{A}_\infty(\mathcal{V})$ consists of functions that are analytic on a complex disc \mathcal{V} and are continuous on its closure $\bar{\mathcal{V}}$.

Producing $v(m)$. We begin by explaining how to produce $v(m)$ for a fixed $m \in \mathbb{N}^*$. Consider the operator

$$\mathbf{H}_{[m],s}[g](x) := |h'_m(x)|^s g(h_m(x)).$$

It turns out that this “component operator” has the following eigenvalues (with multiplicity one) when acting on the space $\mathcal{A}_\infty(\mathcal{V})$

$$|h'_m(x_m^*)|^s, (-1)^{|m|} |h'_m(x_m^*)|^{s+1}, (-1)^{2|m|} |h'_m(x_m^*)|^{s+2}, \dots$$

To motivate the above eigenvalues, notice that that if λ is an eigenvalue of $\mathbf{H}_{[m],s}$ corresponding to the eigenfunction g_1 , and if $g_1(x_m^*) \neq 0$ we have

$$\lambda g_1(x_m^*) = \mathbf{H}_{[m],s}[g_1](x_m^*) = |h'_m(x_m^*)|^s g_1(h_m(x_m^*)) = |h'_m(x_m^*)|^s g_1(x_m^*) \Rightarrow \lambda = |h'_m(x_m^*)|^s.$$

If $g_1(x_m^*) = 0$ but $g'_1(x_m^*) \neq 0$, differentiating we get

$$\lambda g'_1(x_m^*) = (-1)^{|m|} |h'_m(x_m^*)|^{s+1} g'_1(h_m(x_m^*)) = (-1)^{|m|} |h'_m(x_m^*)|^{s+1} g'_1(x_m^*) \Rightarrow \lambda = (-1)^{|m|} |h'_m(x_m^*)|^{s+1}.$$

and so on...

On the space $\mathcal{A}_\infty(\mathcal{V})$, these operators are what is called “trace class”: they have a trace that is the sum of the eigenvalues. Thus the trace equals:

$$\mathrm{Tr} \mathbf{H}_{[m],s} = \frac{|h'_m(x_m^*)|^s}{1 - (-1)^{|m|} |h'_m(x_m^*)|},$$

and we deduce

$$\mathrm{Tr} \mathbf{H}_{[m],s} = (v(m))^{2s} + O((v(m))^{2s+2}). \quad (6.19)$$

Producing the sum over the suffixes. Now we get back to (6.18). Summing over all suffixes m with $w \preceq m$ produces

$$\mathrm{Tr}((\mathbf{I} - \mathbf{H}_s)^{-1} \circ \mathbf{H}_{[w],s}) = \sum_{m:w \preceq m} (v(m))^{2s} + \text{smaller order series}, \quad (6.20)$$

meaning that the ‘‘remainder’’ series has a larger half-plane of convergence (due to the exponent $2s + 2$).

The argument $(\mathbf{I} - \mathbf{H}_s)^{-1} \circ \mathbf{H}_{[w],s}$ of the trace in (6.20) is the series of all terms $\mathbf{H}_{[m],s}$ with $w \preceq m$. Indeed, this follows by composition: we start with v and the component $(\mathbf{I} - \mathbf{H}_s)^{-1}$ tells us that afterwards we may complete as we please.

The singularity. We briefly mention how to get the dominant singularity of $\mathrm{Tr}((\mathbf{I} - \mathbf{H}_s)^{-1} \circ \mathbf{H}_{[w],s})$ from (6.20).

Since the sum in $\mathbf{H}_{[w],s}$ consists of just one term, it is enough to look for the dominant singularity of $(\mathbf{I} - \mathbf{H}_s)^{-1}$ which is found at $s = 1$ where we have just the dominant eigenvector $\psi_g(x) := \frac{1}{\log 2} \frac{1}{1+x}$ (the Gauss density)

$$(\mathbf{I} - \mathbf{H}_s)^{-1} \circ \mathbf{H}_{[w],1}[\psi_g](x) \sim \frac{1}{s-1} \frac{12 \log 2}{\pi^2} \psi_g(x) \int_0^1 \mathbf{H}_{[w],1}[\psi_g] dx = \left(\frac{1}{s-1} \frac{12 \log 2}{\pi^2} \int_{\mathcal{I}_w} \psi_g(x) dx \right) \psi_g(x),$$

as $s \rightarrow 1$. Thus $\frac{1}{s-1} \frac{12 \log 2}{\pi^2} \int_{\mathcal{I}_w} \psi_g(x) dx$ approximates the dominant eigenvalue in the trace, the rest we expect to be of much smaller absolute value as $s \rightarrow 1$.

Going back to (6.18), each $\sum_{m:w \preceq m} (v(m))^s$ is approximated by $\frac{1}{s/2-1} \frac{12 \log 2}{\pi^2} \int_{\mathcal{I}_w} \psi_g(x) dx$ around their pole $s = 1$. This gives a pole for the whole series

$$F_n(s) \sim \frac{1}{s-2} \frac{24 \log 2}{\pi^2} \sum_{k \geq 1} \sum_{w \in \mathbb{N}^k} \left[\left[1 + \frac{q_{k-1}(w) + q_k(w)}{n} \leq \lambda, \quad q_{k-1}(w) \leq n < q_k(w) \right] \int_{\mathcal{I}_w} \psi_g(x) dx, \right. \quad (6.21)$$

where we remark that $\int_{\mathcal{I}_w} \psi_g(x) dx = \mathbb{P}_{\mu_g}(\alpha \in \mathcal{I}_w)$, where \mathbb{P}_{μ_g} is the probability when the input slope is chosen according to the Gauss measure $d\mu_g(x) = dx/(1+x)$ from the whole of \mathcal{I} .

6.3.7 Results for $\ell = 0$

We state our main expected result (unpublished) concerning the first cycle $\ell = 0$.

Expected Theorem 6.3. *Consider the set \mathcal{K} of reduced quadratic irrationals α , endowed with the size $\alpha \mapsto \frac{1}{v(\alpha)}$. Given N , we consider the uniform distribution on the set \mathcal{M}_N of reduced quadratic irrationals α with $1/v(\alpha) \leq N$, giving a probability measure which we denote \mathbb{P}_N . Let $S_n(\alpha) = S(\alpha, n)$ be the recurrence quotient defined in Chapter 3, Equation 3.13, $k(\alpha, n)$ be the sole positive integer k such that $q_{k-1}(\alpha) \leq n < q_k(\alpha)$, and $p(\alpha)$ be the period of the quadratic irrational α . Then, as $N \rightarrow \infty$ we have*

$$\lim_{N \rightarrow \infty} \mathbb{P}_N(S(\alpha, n) \leq \lambda, \quad k(\alpha, n) \leq p(\alpha)) = \mathbb{P}_{\mu_g}(S_n \leq \lambda),$$

where \mathbb{P}_{μ_g} is the probability when the input slope is chosen according to the Gauss measure $d\mu_g(x) = dx/(1+x)$ from the whole of \mathcal{I} .

Furthermore, as $n \rightarrow \infty$ we have that $\mathbb{P}_{\mu_g}(S_n \leq \lambda)$ converges to the limit distribution from Theorem 4.1.

We comment briefly on the proof elements. From (6.21) we have the dominant singularity for $F_n(s)$, defined in (6.17), at $s = 2$. Then we recall that the dominant singularity of $F_n(s)$, where we sum over all possible prefixes $m \in \mathbb{N}^*$ coincides with that of the same sum but over the primitive periods in \mathcal{P} . At this point we

expect to prove the result by showing a combination of the Perron formula and the Dolgopyat-Baladi-Vallée estimates (see Theorem 6.1).

Finally, the fact that $\mathbb{P}_{\mu_g}(S_n \leq \lambda)$ converges to the limit distribution from Theorem 4.1 follows from the “independence from the initial distribution” explained in Section 4.5.3. In fact, this result for the quadratic irrationals was the main motivation for our study of Section 4.5.3.

Of course, we should be able to derive analogous results for other \mathcal{LQ} -functions.

6.3.8 The case $\ell \rightarrow \infty$ and future work

Our study for the case $\ell \rightarrow \infty$ is very much work in progress. Our experiments, and intuitions, suggest strongly that there is some kind of stationary behavior as we will shortly explain.

In this section we consider a \mathcal{Q} -function $\Lambda(\alpha, n)$ with associated function f . Recall that this means that $\Lambda(\alpha, n) = f(q_{k-1}(\alpha)/q_k(\alpha), q_k(\alpha)/n)$ where $k = k(\alpha, n)$ is the sole positive integer satisfying $q_{k-1}(\alpha) \leq n < q_k(\alpha)$.

In the case $\ell \rightarrow \infty$ we wish to study the averages

$$A_\ell(m) := \frac{1}{q(m^{\ell+1}) - q(m^\ell)} \sum_{n=q(m^\ell)}^{q(m^{\ell+1})-1} \Lambda(x_m^*, n) \quad (6.22)$$

over a period $m \in \mathbb{N}^*$. Note that the condition on n makes the index from Section 6.3.5 be fixed and equal to ℓ , and further, these are all the possible n giving the number of complete turns ℓ . There are sound intuitive reasons for choosing to study these averages, which we now explain.

Scaling properties as $\ell \rightarrow \infty$. We recall that

$$v(m) = \lim_{\ell \rightarrow \infty} \frac{q(m^\ell)}{q(m^{\ell+1})},$$

moreover, we can show that for $u \in \mathbb{N}^*$ we have

$$\lim_{\ell \rightarrow \infty} \frac{q(m^\ell)}{q(m^\ell u)} = |h'_{u^R}(x_{m^R}^*)|^{1/2}, \quad (6.23)$$

where we recall that u^R and m^R denote the mirror image of u and m . All of this means that, in a way, the continuants q behave almost multiplicatively.

Example 6.1. We begin by exploiting the regularity of Λ over each interval $[q_k(m), q_{k+1}(m))$, thus decompose

$$\frac{1}{q(m^{\ell+1}) - q(m^\ell)} \sum_{n=q(m^\ell)}^{q(m^{\ell+1})-1} \Lambda(x_m^*, n) = \frac{1}{q(m^{\ell+1}) - q(m^\ell)} \sum_{k=\ell|m|}^{(\ell+1)|m|-1} \sum_{n=q_k}^{q_{k+1}-1} \Lambda(x_m^*, n),$$

and we substitute using our formula for Λ

$$\sum_{n=q_k}^{q_{k+1}-1} \Lambda(x_m^*, n) = \sum_{n=q_k}^{q_{k+1}-1} f\left(\rho_k(x_m^*), \frac{q_{k+1}}{n}\right) \approx q_{k+1}(x_m^*) \int_1^{1/\rho_{k+1}(x_m^*)} f(\rho_{k+1}(x_m^*), y) \frac{dy}{y^2},$$

where $\rho_k(x) := \frac{q_{k-1}(x)}{q_k(x)}$.

The previous integral approximation allows us to estimate the result for a given w fairly fast (for the experiments in particular), but we hope that it will also allow for a nice alternative expression for our final dirichlet generating functions.

At this point we expect

$$\begin{aligned} \lim_{\ell \rightarrow \infty} \frac{1}{q(m^{\ell+1}) - q(m^\ell)} \sum_{n=q(m^\ell)}^{q(m^{\ell+1})-1} \Lambda(x_m^*, n) \\ = \frac{v(m)}{1 - v(m)} \sum_{\epsilon \prec u \preceq m} |h'_{uR}(x_{mR}^*)|^{-1/2} \int_1^{\frac{1}{h_{uR}(x_{mR}^*)}} f\left(\frac{1}{h_{uR}(x_{mR}^*)}, y\right) \frac{dy}{y^2}, \end{aligned}$$

due to (6.23). Thus the averages for a fixed $m \in \mathbb{N}^*$ would converge.

Closing comments. There are several potential generating functions we have considered.

For conciseness, let us write, for $h, g \in \mathcal{H}^*$, that $h \leq g$ if and only if $h = h_{m_1}$ and $g = h_{m_2}$ where $m_1 \preceq m_2$.

Our current conjecture is that a good choice of a DGF should be

$$G_{n,\lambda}^{[\infty]}(s) = \sum_{h \in \mathcal{H}^*} \sum_{\underline{g} \leq h} B_{\underline{g}}(0, nx_{\widehat{h}}^*, \lambda) (v(h))^s, \quad (6.24)$$

where $v(h_m) := v(\mathbf{m})$, \widehat{h} is the mirror of h , and $B_{\underline{g}}(x, n, \lambda)$ is an analog of $A_{\underline{g}}(x, n, \lambda)$ from (6.11), defined by

$$B_{\underline{g}}(x, n, \lambda) := \left[\left[|(e(\underline{g}))'(x)|^{-1/2} \leq n < |g'(x)|^{-1/2}, f_{\Lambda} \left(\frac{|(e(\underline{g}))'(x)|^{-1/2}}{n}, \frac{|g'(x)|^{-1/2}}{n} \right) \leq \varrho \right] \right],$$

where e is the “ending” (if b was beginning in (6.11)) defined by $e(h_{m_1, \dots, m_k}) = h_{m_2, \dots, m_k}$.

There is a sound reason for this choice. The functions $A_{\underline{g}}$ and $B_{\underline{g}}$ are sort of dual, if $g = h^\ell \circ \underline{g}$ with $\underline{g} \leq h$, then we may consider the mirror to consider \underline{g} first, through the remarkable identity

$$A_{\underline{g}}(0, n |(h^\ell)'(0)|^{1/2}, \lambda) = B_{\underline{g}}(\widehat{h}^\ell(0), n, \lambda).$$

Thus, the natural generating function for the ℓ -th tour, namely

$$G_{n,\lambda}^{[\ell]}(s) = \sum_{h \in \mathcal{H}^*} \sum_{h^\ell \leq \underline{g} \leq h^{\ell+1}} A_{\underline{g}}(0, n |(h^\ell)'(0)|^{1/2}, \lambda) (v(h))^s, \quad (6.25)$$

notice the scaling $|(h^\ell)'(0)|^{1/2} = q(h^\ell)^{-1}$, gives naturally rise to (6.24) as $\ell \rightarrow \infty$.

Then we expect the generating function from (6.24) to be dealt with as in the previous cases, being expressed as the sum of traces of expressions involving the quasi-inverse. Our conjecture is that we will get the same dominant pole as before, thus giving a similar result following (maybe) a combined application of the Perron formula and Theorem 6.1. This is part of the ongoing work towards a unified vision of all of the three cases (real, rational and quadratic irrational).

Part III

Studies in Arithmetics

CHAPTER 7

THE CONTINUED LOGARITHM ALGORITHM

In Chapter 1, Example 1.3, we introduced the so-called *continued logarithm expansion* as an example of an interesting interval dynamical system (see Example 1.3, Observation 1.5). Here we study its associated gcd-algorithm in detail, explaining the origins of the algorithm, what was known about it and we present our novel study of its average performance.

7.1 Introduction

In an unpublished manuscript, Gosper [Gos78] introduced a new kind of continued fraction expansion, called the “Continued Logarithm”[BCLM17]. He writes

There is a mutation of continued fractions, which I call continued logarithms, which have several advantages over regular continued fractions, especially for computational hardware. (..)

The primary advantage is the conveniently small information parcel. The restriction to integers of regular continued fractions makes them unsuitable for very large and very small numbers. The continued fraction for Avogadro’s number, for example, cannot even be determined to one term, since its integer part contains 23 digits, only 6 of which are known. (...) By contrast, the continued logarithm of Avogadro’s number begins with its binary order of magnitude, and only then begins the description equivalent to the leading digits – a sort of recursive version of scientific notation. (..)

Although these operations are not as nice on paper, they are beautifully suited to binary machines, requiring only shift, add, subtract, exchange, and compare-for-greater.

The continued logarithm expansion of a real number $x \in (0, 1)$ (it can be easily extended to the whole positive numbers) is of the form

$$x = \frac{2^{-a_1}}{1 + \frac{2^{-a_2}}{1 + \ddots}}$$

where the exponents a_1, a_2, a_3, \dots are non-negative integers. Observe that indeed the first coefficient a_1 gives the logarithmic size of x in base 2, thus why Gosper called it a “sort of recursive version of scientific

notation”.

The Continued Logarithm expansion has been studied from an Ergodic point of view by Chan [Cha05, Cha06], which parallels the classical study of continued fractions presented in Chapter 1. In particular, he shows that the associated dynamical system is ergodic and exhibits its invariant density. The motivation of Chan is somewhat different: he shows that for almost every sequence $\{a_1, a_2, \dots\}$ of natural numbers (we will see what he meant by this), if we define a Fibonacci-like sequence f_k by $f_{-1} = 0, f_0 = 1, a_0 = 0$ and $f_k = 2^{a_k} f_{k-1} + 2^{a_{k-1}} f_{k-2}$ for $k \geq 1$, then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log f_n = 1.30022988 \dots$$

This property is in fact analogous to Proposition 1.10 and we give an explicit constant for the above limit. More recently, the continued logarithm has been considered by Borwein et al. [BHL17, BCLM17]

The idea of Gosper gives rise to an algorithm for computing the gcd of two integers. Shallit describes this algorithm in [Sha16], considering purely as an algorithm for computing the continued logarithm expansion of a *rational* number, and not a gcd-algorithm, and studies the worst case (when the expansion is longest). This algorithm has two advantages: first, it can be calculated starting from the most representative bits, and uses very simple operations (subtractions and shifts) as explained in the above citation of Gosper; it does not employ divisions. Second, as the quotients intervening in the associated continued fraction are powers of two 2^a , we can store each of them with $\log_2 a$ bits. This gives a compelling argument in favor of the small complexity of the algorithm, both in terms of computation and storage.

Shallit [Sha16] performs the worst-case analysis of the algorithm, and studies the number of steps $K(p, q)$, and the total number of shifts $S(p, q)$ that are performed on an integer input (p, q) with $p < q$: he proves the inequalities

$$K(p, q) \leq 2 \log_2 q + 2, \quad S(p, q) \leq (2 \log_2 q + 2) \log_2 q,$$

and exhibits instances, the family $(1, 2^n - 1)$, which show that the previous bounds are asymptotically optimal, namely $K(1, 2^n - 1) = 2n - 2$, $S(1, 2^n - 1) = n(n - 1)/2 + 1$.

Following Shallit’s study of the worst-case [Sha16], he proposed to us to perform the average-case analysis of the algorithm (equivalently, the expansion over rationals), and we answered his question in [RVV18].

We considered the set Ω_N which gathers the integer pairs (p, q) with $0 \leq p \leq q \leq N$, endowed with the uniform probability, and we study the mean values $\mathbb{E}_N[K]$ and $\mathbb{E}_N[S]$ as $N \rightarrow \infty$. We prove that these mean values are asymptotically linear in the size $\log N$, and exhibit their precise asymptotics for $N \rightarrow \infty$,

$$\mathbb{E}_N[K] \sim \frac{2}{H} \log N, \quad \mathbb{E}_N[S] \sim \frac{\log 3 - \log 2}{2 \log 2 - \log 3} \mathbb{E}_N[K].$$

The constant H is related to the entropy of an associated dynamical system and

$$H = \frac{1}{2 \log 2 - \log 3} \left[\frac{\pi^2}{6} + 2 \sum_{k \geq 1} \frac{(-1)^k}{k^2 2^k} - (\log 2)(3 \log 3 - 4 \log 2) \right]. \quad (7.1)$$

This entails numerical estimates (validated by experiments) for the mean values

$$\mathbb{E}_N[K] \sim 1.49283 \log N, \quad \mathbb{E}_N[S] \sim 1.40942 \log N,$$

and the mean number of pseudo-divisions is about half the maximum^a from Shallit [Sha16].

^aNote $2/\log 2 \doteq 2.885 \dots$

7.1.1 The continued logarithm expansion

To motivate the choice of the algorithm we begin rather from the associated continued fraction expansion which gives origin to it. The reason behind this is that, otherwise, the algorithm may come away as somewhat less natural (but we will also explain why it is reasonable from an algorithmic point of view) as there are several details that may be done in other ways.

The continued logarithm associates to each $x \in (0, 1)$ a formal expansion

$$x = \frac{2^{-a_1}}{1 + \frac{2^{-a_2}}{1 + \frac{2^{-a_3}}{1 + \ddots}}}, \quad (7.2)$$

for certain non-negative integer digits $a_1(x), a_2(x), \dots$. We denote the RHS expression by $[a_1, a_2, \dots]$, which can be shown to represent a real number [Khi97], when interpreted as the appropriate limit

$$[a_1, a_2, \dots] := \lim_{k \rightarrow \infty} \frac{2^{-a_1}}{1 + \frac{2^{-a_2}}{1 + \ddots \frac{2^{-a_k}}{1}}}$$

As for continued fractions (see Chapter 1), for rational numbers this expansion is going to be finite

$$\frac{p}{q} = \frac{2^{-a_1}}{1 + \frac{2^{-a_2}}{1 + \frac{2^{-a_3}}{1 + \ddots \frac{2^{-a_k}}{1}}}}, \quad (7.3)$$

but not unique. There are two representations

$$\frac{2^{-a_1}}{1 + \frac{2^{-a_2}}{1 + \ddots \frac{2^{-a_k}}{1}}} = \frac{2^{-a_1}}{1 + \frac{2^{-a_2}}{1 + \ddots \frac{2^{-(a_k-1)}}{1 + \frac{2^{-0}}{1}}}}, \quad (7.4)$$

when $a_k > 0$. This, however, is the only possible redundancy in the representation.

To define the process of computing the continued logarithm expansion in a definite way, we compute the coefficients simply by considering that $a_1 = a_1(x)$ is determined uniquely by

$$2^{-a_1-1} < x \leq 2^{-a_1},$$

and so on. This is the coding of the CL dynamical system we defined in Example 1.3, with the shift map T_c , which we call the *CL map*^b, which satisfies

$$x = \frac{2^{-a_1(x)}}{1 + T_c(x)}.$$

We recall the CL dynamical system here, which is an interval dynamical system (see Definition 1.3).

^bIt is called the ‘‘Hei-Chi Chan map’’ in [AS17] in honor of Chan who studied its Ergodic properties in [Cha05].

Definition 7.1 (*CL system*). Let $\mathcal{I} := (0, 1)$ be the unit interval, we define the shift map $T_c: \mathcal{I} \rightarrow \mathcal{I}$, called the CL map, by

$$T_c(x) = \frac{2^{-a_1(x)}}{x} - 1, \quad a_1(x) = \left\lfloor \log_2 \frac{1}{x} \right\rfloor, \tag{7.5}$$

where $\{\cdot\}$ denotes the fractional part $\{y\} := y - \lfloor y \rfloor$. Then the inverse branches are given by

$$h_a(x) = \frac{1}{2^a(1+x)}, \quad a \geq 0. \tag{7.6}$$

We further define the digits $a_k(x)$ by

$$a_k(x) = a_1(T_c^{k-1}x), \tag{7.7}$$

for $k \geq 2$.

Observation 7.1. We note that for $x = p/q$ this gives

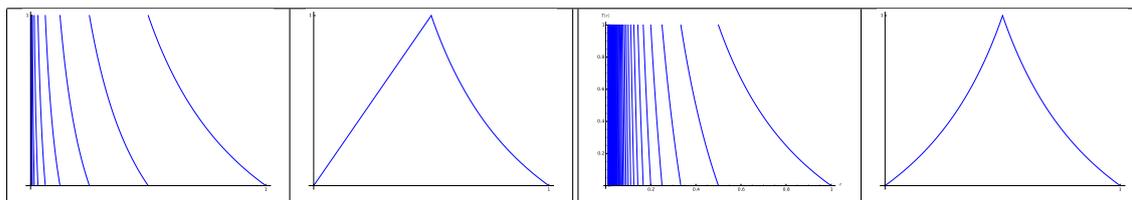
$$T_c\left(\frac{p}{q}\right) = \frac{q - 2^{a_1}p}{2^{a_1}p},$$

so, for fractions, we may think of T_c as a map on the pairs (p, q) which maps

$$(p, q) \mapsto (p', q') := (q - 2^{a_1}p, 2^{a_1}p),$$

where $a_1(p/q) = \max\{k \in \mathbb{N} : 2^k p \leq q\}$. This is how Shallit in [Sha16] computes the coefficients of the continued fraction expansion. It may be seen as a gcd-algorithm as we will soon explain in Section 7.1.2.

The *CL system* (\mathcal{I}, T_c) is displayed on the left of the figure below, along with the shift $S : \mathcal{I} \rightarrow \mathcal{I}$ which gives rise to the *CL system* by induction on the first branch. The map S is a mix of the Binary and Farey maps, as its first branch comes from the Binary system, and the second one from the Farey system. On the right, the usual Euclid dynamical system (defined from the Gauss map T_g) is derived from the Farey shift by induction on the first branch.



7.1.2 The continued logarithm algorithm

In Observation 7.1 we explained how the CL map can be seen as a map on pairs (p, q) , representing respectively the successive numerators and denominators. In this section we describe the gcd algorithm derived from this.

The algorithm, described by Shallit in [Sha16], is a sequence of (pseudo)–divisions: each division associates a pair $(p, q)^c$ with $p < q$ to a new pair (r, p') (where r stands for “remainder”) defined as follows

$$q = 2^a p + r, \quad q' = 2^a p, \quad \text{with } a = a(p, q) := \max\{k \geq 0 \mid 2^k p \leq q\}.$$

This is a gcd algorithm; indeed $\gcd(r, q') = \gcd(q - 2^a p, 2^a p)$ is equal to $\gcd(2^a p, q)$, and this latter gcd may differ from $\gcd(p, q)$ only in a power of 2. Thus the *CL algorithm* determines the odd part of $\gcd(p, q)$ whereas the even part is directly determined by the dyadic valuations of p and q , which amounts to reading the trailing 0s in binary.

Here we show a first example of the execution of the algorithm for $(13, 31)$

^cOur notations are not the same as in the paper of Shallit as we reverse the roles of p and q .

i	a_i	$2^{a_i} q_i$	q_{i+1}	$(2^{a_i} q_i)_2$	$(q_{i+1})_2$	$\delta(2^{a_i} q_i)$	$\delta(q_{i+1})$	$\delta(\widehat{g}_i)$
0	—	75	31	1001011	11111	0	0	0
1	1	62	13	0111110	1101	1	0	0
2	2	52	10	110100	1010	2	1	1
3	2	40	12	101000	1100	3	2	2
4	1	24	16	11000	10000	3	4	3
5	0	16	8	10000	1000	4	3	3
6	0	8	8	1000	1000	3	3	3
7	—	8	0	1000	0	3	∞	3

Figure 7.1: Execution for the input pair $(p, q) = (31, 75)$. Here $\widehat{g}_i = \gcd(2^{a_i} q_i, q_{i+1})$. The dyadic valuation $\delta(\widehat{g}_i)$ seems to linearly increase with i , with the asymptotic $\delta(\widehat{g}_i) \sim \delta(q_{i+1}) \sim i/2$ as $i \rightarrow \infty$.

a	p	q	r	$2^a p$
1	13	31	5	26
2	5	26	6	20
1	6	20	8	12
0	8	12	4	8
1	4	8	0	8

Matrix form. This transformation is conveniently written in matrix form: the pair (p, q) gives a new one $(r, 2^a p)$ satisfying

$$\begin{pmatrix} p \\ q \end{pmatrix} = N_a \begin{pmatrix} r \\ 2^a p \end{pmatrix}, \quad \text{with } N_a = \begin{pmatrix} 0 & 2^{-a} \\ 1 & 1 \end{pmatrix} = 2^{-a} M_a, \quad M_a = \begin{pmatrix} 0 & 1 \\ 2^a & 2^a \end{pmatrix}. \quad (7.8)$$

The CL algorithm begins from an input pair (p, q) with $p < q$. It lets the initial pair be $(q_1, q_0) := (p, q)$, and then performs a sequence of divisions

$$(q_{i+1}, q_i)^T = N_{a_{i+1}} (q_{i+2}, 2^{a_{i+1}} q_{i+1})^T,$$

and stops after $k = K(p, q)$ steps on a pair of the form $(0, 2^{a_k} q_k)$. The complete execution of the algorithm uses the set of matrices N_a defined in (7.8), and expands the input as

$$(p, q)^T = N_{a_1} \cdot N_{a_2} \cdots N_{a_k} (0, 2^{a_k} q_k)^T.$$

The rational input p/q is then written as a continued fraction according to the linear fractional transformations (LFTs for short) h_a associated with matrices N_a or M_a ,

$$\frac{p}{q} = \frac{2^{-a_1}}{1 + \frac{2^{-a_2}}{1 + \frac{2^{-a_3}}{1 + \frac{2^{-a_k}}{1 + \frac{2^{-a_k}}{1}}}}} = h_{a_1} \circ h_{a_2} \circ \cdots \circ h_{a_k}(0), \quad \text{with } h_a : x \mapsto \frac{2^{-a}}{1+x}. \quad (7.9)$$

We note that h_a , for $a \geq 0$, are the inverse branches of the CL map as in Definition 1.3.

Moreover, thanks to the equality (7.4), it is possible to change our definition of $a(p, q)$ to make the last exponent a_k be 0 (and the last quotient to be 1). Indeed, according to our definition of $a(p, q)$, the last exponent is necessarily $a_k > 0$, and through (7.4) we make it into the form on the right-hand side.

Figure 7.1 describes the execution of the algorithm for the initial pair $(31, 75)$. Note that this example is related to our previous example $(13, 31)$. Indeed, we get the pair $(62, 13) = (2 \cdot 31, 31)$ after one iteration.

Worst case analysis. It is simple enough to see that the algorithm terminates; indeed if $2^a p = q$ the algorithm surely ends, else the next second entry of the pair, $q' = 2^a p$, is less than q . It is not trivial, however, to find its worst case behavior. Shallit [Sha16] proved that the number of steps $K(p, q)$ is at most $2 \log_2 q + O(1)$ by proving that the size of the entries at least half every two steps. This is also the idea in a classical simplified proof that Euclid's algorithm terminates in $O(\log q)$ steps [DPV08], but it is more involved in the case of the CL algorithm: Shallit proves his result by showing that $(p, q) \mapsto p^2 + q^2$ is decreasing, roughly dividing its size by 2 for each step, when one reduces (p, q) dividing by the gcd when possible.

The worst case for the Euclidean algorithm is given by the pairs of the form (f_k, f_{k+1}) where the (f_k) are the Fibonacci numbers, and the final depth is k , see for example [CLRS09, pp.935-936]. This follows from choosing the quotients m_i so as to make the continuants q_i as small as possible. As $f_k \sim \varphi^k / \sqrt{5}$ we see that the worst case for the Euclidean algorithm is actually $(\log q) / \log \varphi$ rather than $2 \log_2 q + O(1)$. In the case of the CL algorithm, however, the bound $2 \log_2 q + O(1)$ is asymptotically tight as Shallit demonstrated by considering the family $(p, q) = (1, 2^n - 1)$. Indeed, every two steps we have

$$(1, 2^n - 1) \mapsto_{a=n-1} (2^{n-1} - 1, 2^{n-1}) \mapsto_{a=0} (1, 2^{n-1} - 1).$$

We remark that here the worst-case is not given by minimizing the quotients (or exponents) 2^a at each step. This is due to the fact that there are other cancellations taking place: the natural continuants $q_k(x)$ for the CL expansion do not naturally produce reduced fractions as we shall see in Section 7.2.1 below.

7.2 The CL dynamical system

In this section we point out several properties of the CL dynamical system.

7.2.1 The continuants of the CL expansion

Each number $x \in \mathcal{I}$ admits an infinite continued fraction expansion derived from the dynamical system, called its *CL expansion*. When truncated at depth k , the expansion of x becomes finite, as in (7.9), and represents a rational p/q , which we assume to be irreducible. As x belongs to a unique fundamental interval \mathcal{J} of depth k , of the form $\mathcal{J} = h(\mathcal{I})$ with $h \in \mathcal{H}^k$, the pair (k, x) determines a unique LFT $h := h_{\mathbf{a}}$ of depth k , and the rational p/q equals $h(0)$.

The k -tuple \mathbf{a} defines a matrix $M_{\mathbf{a}}$ (see equation (7.8)) and an integer pair (P, Q) , called the *continuant pair*, defined by $(P, Q)^T := M_{\mathbf{a}}(0, 1)^T$. The equality $\frac{P}{Q} = \frac{p}{q}$ clearly holds, but the pair (P, Q) is not necessarily coprime. Nevertheless, $\gcd(P, Q)$ is a power of 2, as follows from taking determinants $|\det(M_{\mathbf{a}})| = 2^{a_1 + \dots + a_k}$. Moreover, the integer $R(Q) := Q / \gcd(P, Q)$, called the *reduced continuant*, is an important parameter actually dictating the quality of the rational approximation of x given by the truncation of its *CL expansion*.

The recurrence of the continuant pairs. The continuant pairs satisfy a recurrence analogous to that of Proposition 1.1 for the convergents of the Euclidean Algorithm. To avoid ambiguity, we will write $P(\mathbf{a})$ and $Q(\mathbf{a})$ for the continuant pair associated with a tuple $\mathbf{a} = (a_1, \dots, a_k)$. In an abuse of notation, we will also write $P_k(\mathbf{a})$ and $Q_k(\mathbf{a})$ to denote $P(a_1, \dots, a_k)$ and $Q(a_1, \dots, a_k)$, respectively, when \mathbf{a} is a tuple of length at least k (maybe an infinite sequence). Similarly, when given an irrational x we will also write $P_k(x)$ and $Q_k(x)$ for $P(a_1(x), \dots, a_k(x))$ and $Q(a_1(x), \dots, a_k(x))$.

The following result is actually the starting point definition used by Chan [Cha05].

Proposition 7.1. *Consider non-negative integers a_1, \dots, a_k then we have the following recurrence for the continuant pairs of the CL expansion*

$$\begin{aligned} P(a_1, \dots, a_k) &= 2^{a_k} P(a_1, \dots, a_{k-1}) + 2^{a_{k-1}} P(a_1, \dots, a_{k-2}), & (k \geq 2) \\ Q(a_1, \dots, a_k) &= 2^{a_k} Q(a_1, \dots, a_{k-1}) + 2^{a_{k-1}} Q(a_1, \dots, a_{k-2}), & (k \geq 1) \end{aligned}$$

where we take $P_0 = 0, P_1 = 1$ and $Q_{-1} = 0, Q_0 = 1$.

Proof. The definition $(P(\mathbf{a}), Q(\mathbf{a}))^T := M_{\mathbf{a}}(0, 1)^T$ is not the simplest when it comes to proving this recurrence, but if we extend it slightly to 2×2 matrices

$$\begin{pmatrix} 2^{a_k} P(a_1, \dots, a_{k-1}) & P(a_1, \dots, a_k) \\ 2^{a_k} Q(a_1, \dots, a_{k-1}) & Q(a_1, \dots, a_k) \end{pmatrix} = M_{\mathbf{a}} = M_{a_1} \cdot M_{a_2} \cdot \dots \cdot M_{a_k}, \quad (7.10)$$

we easily get the result.

Now we explain (7.10). It is clear, from our definition, that the second column in the left-hand side matrix is correct. As for the first column, notice that the equalities

$$M_{\mathbf{a}} = (M_{a_1} \dots M_{a_{k-1}}) M_{a_k}, \quad M_{a_k} = \begin{pmatrix} 0 & 1 \\ 2^{a_k} & 2^{a_k} \end{pmatrix},$$

imply that the first column of $M_{\mathbf{a}}$ is the second column of $(M_{a_1} \dots M_{a_{k-1}})$ multiplied by 2^{a_k} , hence yielding the first column on the left-hand side of (7.10). ■

Note that from (7.10) we see at once that

$$|P(a_1, \dots, a_{k-1})Q(a_1, \dots, a_k) - Q(a_1, \dots, a_{k-1})P(a_1, \dots, a_k)| = 2^{a_1 + \dots + a_{k-1}}. \quad (7.11)$$

The continuants and the inverse branches. Another important result noted by Chan in [Cha05] concerns the inverse branches of the CL system. The following proposition tells us how to express the inverse branches in terms of the continuant pairs. The reader familiar with linear fractional transforms (LFT) will note at once that, since $h_{\mathbf{a}}$ is the LFT associated with $M_{\mathbf{a}}$, this result follows at once from the expression (7.10) as the multiplication of matrices corresponds to the composition of LFTs.

Proposition 7.2. *Consider non-negative integers a_1, \dots, a_k . Then, the inverse branch*

$$h_{\mathbf{a}}(x) = h_{a_1} \circ \dots \circ h_{a_k}(x),$$

can be written in terms of the continuant pairs as follows

$$h_{\mathbf{a}}(x) = \frac{P(a_1, \dots, a_k) + x \cdot 2^{a_k} P(a_1, \dots, a_{k-1})}{Q(a_1, \dots, a_k) + x \cdot 2^{a_k} Q(a_1, \dots, a_{k-1})}. \quad (7.12)$$

The expression in (7.12) tells us a lot regarding the CL system. In particular it gives us an expression for the length of the fundamental intervals and the speed of convergence of the truncated expansions.

Corollary 7.1. *Consider non-negative integers a_1, \dots, a_k . Then, the length of the fundamental interval $\mathcal{I}_{\mathbf{a}} := \mathcal{I}_{a_1, \dots, a_k}$, associated with the CL system, is given by*

$$|\mathcal{I}_{\mathbf{a}}| = \frac{2^{a_1 + \dots + a_k}}{Q(a_1, \dots, a_k) \cdot (Q(a_1, \dots, a_k) + 2^{a_k} Q(a_1, \dots, a_{k-1}))}.$$

The gcd of the continuant pair. Contrary to the case of classical continued fractions, the continuant pair, seen as a rational number, is not reduced. In fact, it follows at once from (7.11) that the greatest common divisor $\gcd(P_k, Q_k)$ divides $2^{a_1+\dots+a_{k-1}}$, hence being a power of two, but in general it is not 1. Indeed, when $a_k(x), a_{k-1}(x) > 0$, the gcd is not one, and this happens infinitely many often for almost every x .

In fact, we have the following conjecture regarding the continuant pairs of a real number

Conjecture. For almost every $x \in \mathcal{I}$ the limit

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log_2 \gcd(P_k(x), Q_k(x)) = \frac{1}{2} \tag{7.13}$$

holds. ◇

This conjecture is based on experimental evidence, with the convergence to $1/2$ being apparently faster for the expected values

$$\lim_{k \rightarrow \infty} \frac{1}{k} \mathbb{E}_x[\log_2 \gcd(P_k(x), Q_k(x))] = \frac{1}{2}.$$

For the moment we can offer the following bounds, showing that the greatest common divisor increases at least at an exponential rate almost surely.

Proposition 7.3. For almost every $x \in \mathcal{I}$

$$\liminf_{k \rightarrow \infty} \frac{1}{k} \log_2 \gcd(P_k(x), Q_k(x)) \geq \frac{1}{2} \mathbb{E}_\mu[A] \doteq 0.30542\dots,$$

where $A(x) = \min\{a_1(x), a_1(T_c x)\}$ and μ is the probability measure on \mathcal{I} given by the density $\psi_c(x) = \frac{1}{\log(4/3)} \frac{1}{(x+1)(x+2)}$ with respect to the Lebesgue Measure.

This proposition demonstrates that a big gcd is the rule rather than the exception.

Corollary 7.2. For almost every $x \in \mathcal{I}$ we have

$$\limsup_{k \rightarrow \infty} \frac{1}{k} \log_2 \gcd(P_k(x), Q_k(x)) \leq \frac{\log(3/2)}{\log(4/3)} - \frac{1}{2} \mathbb{E}_\mu[A] \doteq 1.1040\dots,$$

where $A(x) = \min\{a_1(x), a_1(T_c x)\}$ and μ is the probability measure on \mathcal{I} given by the density $\psi_c(x) = \frac{1}{\log(4/3)} \frac{1}{(x+1)(x+2)}$ with respect to the Lebesgue Measure.

The proof of Proposition 7.3 employs Birkhoff’s Ergodic Theorem, and relies strongly on two observations that we will now explain. On the other hand, Corollary 7.2 follows directly from (7.11) by Proposition 1.11 and Proposition 7.3.

Let $\delta(n)$ denote the dyadic valuation of n , i.e., the greatest integer such that $2^{\delta(n)} | n$. We know that the greatest common divisor of P_k and Q_k is a power of two (similarly for $\gcd(Q_k, Q_{k-1})$ and $\gcd(P_k, P_{k-1})$ in fact), therefore it suffices to study the evolution of $\delta(P_k)$ and $\delta(Q_k)$. Then $\log_2 \gcd(P_k, Q_k) = \min\{\delta(P_k), \delta(Q_k)\}$.

The first key observation for Proposition 7.3 is that the continuant pairs of an irrational $x \in \mathcal{I}$ satisfy

$$P_k(x) = Q_{k-1}(T_c x), \tag{7.14}$$

for all $k \geq 1$.

Equation 7.14 implies that, if we could prove that for almost every x we had $\lim_{k \rightarrow \infty} \frac{1}{k} \delta(Q_k(x)) = \frac{1}{2}$, then the same limit would hold with P_k instead of Q_k . Therefore if $\lim_{k \rightarrow \infty} \frac{1}{k} \delta(Q_k(x)) = \frac{1}{2}$ holds for almost every x , then $\lim_{k \rightarrow \infty} \frac{1}{k} \log_2 \gcd(P_k(x), Q_k(x)) = \frac{1}{2}$ for almost every x . This observation also holds when taking lim sup or lim inf instead of a limit, and changing $1/2$ for another constant.

The second observation is given by the following lemma.

Lemma 7.1. *Suppose that for a given $h \geq 0$ we had $\delta(Q_{k-1}) \geq h$ and $\delta(Q_{k-2}) \geq h$. If $a_{k-1}, a_k \geq t$ we have $\delta(Q_k) \geq h + t$ and $\delta(Q_{k+1}) \geq h + t$, and further $\delta(Q_j) \geq h + t$ for all $j \geq k$.*

Proof. Follows at once from Proposition 7.1. ■

From Lemma 7.1 we see that

$$\delta(Q_k(x)) \geq \min\{a_k(x), a_{k-1}(x)\} + \min\{a_{k-2}(x), a_{k-3}(x)\} + \dots,$$

from which the result follows by applying Birkhoff's Theorem to $T_c^2 = T_c \circ T_c$, recall $a_k(x) = a_1(T_c^{k-1}x)$.

7.2.2 The Perron Frobenius operator

The Perron Frobenius operator (recall subsection 1.2.4) for the CL dynamical system is given explicitly by

$$\mathbf{H}[f](x) := \sum_{h \in \mathcal{H}} |h'(x)| f(h(x)) = \left(\frac{1}{1+x} \right)^2 \sum_{a \geq 0} 2^{-a} f\left(\frac{2^{-a}}{1+x} \right). \quad (7.15)$$

This operator describes the evolution of densities : if f is the initial density, $\mathbf{H}[f]$ is the density after one iteration of the system (\mathcal{I}, T_c) . The invariant density ψ_c is a fixed point for \mathbf{H} and satisfies the functional equation

$$\psi_c(x) = \left(\frac{1}{1+x} \right)^2 \sum_{a \geq 0} 2^{-a} \psi_c\left(\frac{2^{-a}}{1+x} \right). \quad (7.16)$$

We recall that in Observation 1.5 we pointed out that the CL map T_c is Ergodic with respect to the Lebesgue measure, a result proved by Chan in [Cha05], who showed the explicit invariant *CL density* $\psi_c(x)$ given by

$$\psi_c(x) = \frac{1}{\log(4/3)} \frac{1}{(x+1)(x+2)}.$$

However, Chan did not provide an explicit expression for the entropy. We obtain here such an expression, with a precise study of the transfer operator of the system.

We introduce two (complex) parameters t, v in (7.15), and deal with a perturbation of the operator \mathbf{H} , defined by

$$\mathbf{H}_{t,v}[f](x) := \sum_{h \in \mathcal{H}} |h'(x)|^t d(h)^v f(h(x)) = \left(\frac{1}{1+x} \right)^{2t} \sum_{a \geq 0} 2^{a(v-t)} f\left(\frac{2^{-a}}{1+x} \right). \quad (7.17)$$

Such an operator $\mathbf{H}_{t,v}$ is called a transfer operator. When (t, v) satisfies $\Re(t - v) > 0$, we prove the following: the operator $\mathbf{H}_{t,v}$ acts nicely on the space $\mathcal{C}^1(\mathcal{I})$ endowed with the norm $|\cdot|_{1,1}$, defined by $|f|_{1,1} := |f|_0 + |f'|_0$, where $|\cdot|_0$ denotes the sup norm. In particular, it has a dominant eigenvalue $\lambda(t, v)$ separated from the remainder of the spectrum by a spectral gap, for (t, v) close to $(1, 0)$. The Taylor expansion of $\lambda(t, v)$ near $(1, 0)$

$$\lambda(t, v) \approx 1 - A(t - 1) + Dv$$

involves the two constants $A = -\partial\lambda/\partial t(1, 1, 0)$, $D = \partial\lambda/\partial v(1, 1, 0)$, that are expressed as mean values with respect to the invariant density ψ ,

$$A = E - D, \quad E = \mathbb{E}_\psi[2|\log x|], \quad D = (\log 2) \mathbb{E}_\psi[\underline{a}], \quad (7.18)$$

(here, the function \underline{a} associates with x the integer defined with the Iverson bracket $\underline{a}(x) := a \cdot \llbracket x \in h_a(\mathcal{I}) \rrbracket$). The constants A is the entropy of the system, and E, D admit explicit expressions

$$E = \frac{1}{\log(4/3)} \left[\frac{\pi^2}{6} + 2 \sum_{k \geq 1} \frac{(-1)^k}{k^2 2^k} \right], \quad D = (\log 2) \frac{\log(3/2)}{\log(4/3)} \quad (7.19)$$

Then, with (7.18) and (7.19), there is an explicit value for the entropy A , and

$$A \doteq 1.62352\dots, \quad D \doteq 0.97693\dots, \quad E \doteq 2.60045\dots.$$

There are two contexts in which a finite CL expansion arises: in the real context, where it comes from the truncation at a fixed depth of an infinite CL expansion – and in the rational context, where it is finite *per se*, and the depth is no longer fixed. Our main study deals with the second context, but it is also central to understand the former, real context.

7.3 Costs and model for the algorithm

The main interesting costs associated with a finite expansion, as in (7.9), are defined via the associated inverse branch (an LFT) $h \in \mathcal{H}^k$ and mainly involve the continuant pair (P, Q) , associated with h , together with the absolute value of the determinant of the LFT h , denoted by $d(h)$.

7.3.1 Generating functions for our main costs

We recall that for the Euclidean dynamical system, if $h(0) = p/q$ is the reduced fraction, then $|h'(0)| = q^{-2}$, giving back our reduced denominator. For the CL dynamical system this is not so simple. We want to be able to produce the reduced continuants $R(Q)$, which correspond to the size of our input, in terms of h' and h so as to use the transfer operator.

Let us look at an inverse branch $h \in \mathcal{H}^k$. We note that by Proposition 7.2 its derivative gives $|h'(0)| = d(h)/Q^2$. Here we wish to obtain $R(Q) = Q/\gcd(P, Q)$. Since the \gcd is a power of two, we have $\gcd(P, Q) = 2^{\min\{\delta(P), \delta(Q)\}}$ where $\delta(q)$ denotes the dyadic valuation of q . Also, we recall, $d(h)$ is a power of two. Hence we introduce the dyadic numbers \mathbb{Q}_2 .

For an integer q , $\delta(q)$ denotes the dyadic valuation, i.e., is the greatest integer k for which 2^k divides q . The dyadic norm $|\cdot|_2$ is defined on \mathbb{Q} with the equality $|a/b|_2 := 2^{\delta(b) - \delta(a)}$. The dyadic field \mathbb{Q}_2 is the completion of \mathbb{Q} for this norm. See [Kob84] for more details about the dyadic field \mathbb{Q}_2 .

Dyadic numbers can be realized as series of the form $y = 2^{a_1} + 2^{a_2} + \dots$ where $a_1 < a_2 < \dots$ are integers (maybe negative). For instance we remark that $-1 = 1 + 2 + 4 + 8 + \dots$. Indeed, notice that $A_n := (1 + 2 + \dots + 2^n)$ is a Cauchy sequence because $|A_n - A_m|_2 = 2^{-\min\{n, m\}}$ and hence the limit $A = \lim A_n$ exists. Then $1 + A_n = 2^{n+1}$ which tends to 0, thus $A = -1$.

The next result describes provides alternative expressions for these costs.

Proposition 7.4. *Consider the function $G_2 : \mathbb{Q}_2 \rightarrow \mathbb{R}^+$ (called the \gcd map) equal to $G_2(y) = \min(1, |y|_2^{-2})$, namely*

$$G_2(y) = 1 \quad \text{for } |y|_2 \leq 1, \quad G_2(y) = |y|_2^{-2} \quad \text{for } |y|_2 > 1. \quad (7.20)$$

The main costs associated with the CL expansion of a rational $h(0)$

$$Q, \quad g(P, Q) := \gcd(P, Q), \quad R(P, Q) = Q/\gcd(P, Q), \quad |Q|_2,$$

are all expressed in terms of the quadruple $(|h'(0)|, |h'(0)|_2, d(h), G_2[h(0)])$ as follows

$$Q^{-2} = |h'(0)|/d(h), \quad |Q|_2^{-2} = d(h) |h'(0)|_2,$$

$$R^{-2}(Q) = |h'(0)| |h'(0)|_2 G_2[h(0)], \quad g^2(P, Q) = d(h) |h'(0)|_2 G_2[h(0)].$$

Proof. One has (by definition)

$$P/Q = h(0), \quad Q^{-2} = |h'(0)|/d(h), \quad r(Q)^{-2} = g^2(P, Q)/Q^2.$$

Cost C	Cost $c = \log C$	Quadruple γ_C	Constant $M(c)$	Numerical value of $M(c)$
$d(h)$	σ	$(0, 0, 1, 0)$	D	$\doteq 0.97693\dots$
Q^2	q	$(-1, 0, 1, 0)$	$A + D$	$\doteq 2.6004\dots$
$g^2(P, Q)$	ϱ	$(0, 1, 1, 1)$	$B + D$	$\doteq 1.26030\dots$
$R^2(P, Q)$	r	$(-1, -1, 0, -1)$	$A - B$	$\doteq 1.34015\dots$
$ Q _2^{-2}$	q_2	$(0, 1, 1, 0)$	$B + D$	$\doteq 1.26030\dots$

Figure 7.2: Main costs of interest, with their quadruple, and the constant which intervenes in the analysis of their mean value. (see Thm 1).

As $g(P, Q)$ is a power of 2, and using the function G_2 defined in (7.20), one has

$$g^2(P, Q) = \min(|P|_2, |Q|_2)^{-2} = |Q|_2^{-2} \min(1, |P/Q|_2^{-2}) = |Q|_2^{-2} G_2(P/Q)$$

We conclude with the equalities : $|Q|_2^{-2} = |h'(0)|_2 / |d(h)|_2$, $d(h) \cdot |d(h)|_2 = 1$. ■

Any cost C from Proposition 7.4 admits an expression of the form

$$|h'(0)|^t |h'(0)|_2^u d(h)^v G_2[h(0)]^z.$$

The quadruple (t, u, v, z) associated with the cost C is denoted as γ_C . Moreover, as these costs C are expected to be of exponential growth with respect to the depth of the CF , we will work with their logarithms $c = \log C$. Figure 7.2 summarizes the result.

7.3.2 Probabilistic model for the rational case

We now change our context and deal with sets of coprime^d integer pairs

$$\Omega := \{(p, q) : 0 < p < q, \gcd(p, q) = 1\}, \quad \Omega_N := \{(p, q) : 0 < p < q \leq N, \gcd(p, q) = 1\}.$$

The set Ω_N is endowed with the uniform measure, and we wish to study the mean values $\mathbb{E}_N[C]$ of the main parameters C on Ω_N . We focus on parameters which describe the execution of the algorithm, that can be “read” from the $CF(p/q)$ built by the algorithm as in (7.9).

We apply standard analytic combinatorics methodology (see Chapter 2) and we work with (Dirichlet) generating functions (DGFs in short). We first consider the plain Dirichlet generating function (recall (2.13))

$$S(s) := \sum_{(p,q) \in \Omega} \frac{1}{q^{2s}} = \sum_{q=1}^{\infty} \frac{\varphi(q)}{q^{2s}} = \frac{\zeta(2s-1)}{\zeta(2s)}, \tag{7.21}$$

but we also associate a cost $C : \Omega \rightarrow \mathbb{R}^+$ with two generating functions, the bivariate DGF and the cumulative DGF, namely

$$S_C(s, w) := \sum_{(p,q) \in \Omega} \frac{e^{wC(p,q)}}{q^{2s}}, \quad S_C(s) := \sum_{(p,q) \in \Omega} C(p, q) \frac{1}{q^{2s}} = \left. \frac{\partial}{\partial w} S_C(s, w) \right|_{w=0}. \tag{7.22}$$

The expectation $\mathbb{E}_N[C]$ is now expressed as a ratio which involves the sums $\Phi_N(S)$, $\Phi_N(S_C)$ of the first N coefficients of the Dirichlet series $S(s)$ and $S_C(s)$, namely,

$$E_N[C] = \Phi_N[S_C] / \Phi_N[S]. \tag{7.23}$$

^d This restriction can be easily removed and our analysis extends to the set of all the integer pairs.

We know from principles of Analytic Combinatorics that the dominant singularity of a DGF (here its singularity of largest real part) provides precise information about its coefficients, in the case of DGFs for the cumulative $\Phi_N[C]$. Here this transfer from the analytic behavior of the DGF to the asymptotics of its coefficients is provided by Theorem 2.3, due to Delange.

We thus need an alternative expression of these series, from which it is possible to obtain information regarding the dominant singularity, which will be transferred to the asymptotics of coefficients.

Proposition 7.5. *The Dirichlet generating $S(s)$ and its bivariate version $S_C(s, w)$ relative to a cost $C : \Omega \rightarrow \mathbb{R}$, admit alternative expressions^e involving the gcd map G_2 from Proposition 7.4*

$$S(s) = \sum_{h \in \mathcal{H}^* \cdot h_0} |h'(0)|^s |h'(0)|_2^s G_2^s \circ h(0), \quad S_C(s, w) = \sum_{h \in \mathcal{H}^* \cdot h_0} e^{wC(h)} |h'(0)|^s |h'(0)|_2^s G_2^s \circ h(0).$$

For any cost $C \in \{\sigma, q, \varrho, r, q_2\}$ studied in Proposition 7.4, the general term of the bivariate DGF is

$$|h'(0)|^t |h'(0)|_2^u d(h)^v G_2^z \circ h(0).$$

The quadruple $(t, u, v, z) = \gamma_C(s, w)$, corresponding to a cost C , linearly depends on the two exponents s and w . The following table describes the quadruple $\gamma_C(s, w)$ for our costs, together with its derivative with respect to w at $w = 0$ denoted as γ_C .

Cost C	Quadruple $\gamma_C(s, w)$	Quadruple γ_C
σ	(s, s, w, s)	$(0, 0, 1, 0)$
q	$(s - w, s, w, s)$	$(-1, 0, 1, 0)$
ϱ	$(s, s + w, w, s + w)$	$(0, 1, 1, 1)$
r	$(s - w, s - w, 0, s - w)$	$(-1, -1, 0, -1)$
q_2	$(s, s - w, w, s)$	$(0, -1, 1, 0)$

Dynamical analysis. At this point we look for an alternative form for our generating functions in terms of the transfer operator of the dynamical system which underlies the algorithm. Here, it is not possible to obtain such an alternative expression if we stay in the real “world”. This is why we will add a component to our system which allows us to express parameters with a dyadic flavor. It will be possible to express our DGFs in term of a (quasi-inverse) of an (extended) transfer operator, and relate their dominant singularity to the dominant eigenvalue of this extended transfer operator.

We then obtain our main result: We will prove that the mean values $E_N[C]$ of our costs of interest are all of order $\Theta(\log N)$, and satisfy precise asymptotics that involve three constants A, B, D : the constants A and D come from the real world, and have been previously defined in (7.18) and (7.19), but there arises a new constant B that comes from the dyadic world and describes the behavior of the logarithm of the gcd. See the precise statement in Theorem 4.1.

7.4 The extended dynamical system.

In this section, we extend the CL dynamical system, adding a new component to study the dyadic nature of our costs. We then introduce the corresponding transfer operators. It is then possible to express the generating functions in terms of the quasi-inverses of this transfer operator.

7.4.1 Extension of the dynamical system

We will work with a two-component dynamical system: its first component is the initial CL system, to which we add a second (new) component which is used to “follow” the evolution of dyadic phenomena during the

^eWe recall that the last exponent is 0 by convention.

execution of the first component.

We consider the extended interval $\underline{\mathcal{I}} := \mathcal{I} \times \mathbb{Q}_2$. We define a new shift $\underline{T}: \underline{\mathcal{I}} \rightarrow \underline{\mathcal{I}}$ from the characteristics of the old shift T defined in (7.5). As the branches T_a, h_a are LFTs with rational coefficients, they are well-defined on \mathbb{Q}_2 , and they are moreover bijections from $\mathbb{Q}_2 \cup \{\infty\}$ to $\mathbb{Q}_2 \cup \{\infty\}$.

Finally, each branch \underline{T}_a of the new shift \underline{T} is defined via the equality $\underline{T}_a(x, y) := (T_a(x), T_a(y))$ on the fundamental domain $\underline{\mathcal{I}}_a := \mathcal{I}_a \times \mathbb{Q}_2$, and the shift \underline{T}_a is a bijection from $\underline{\mathcal{I}}_a$ to $\underline{\mathcal{I}} := \mathcal{I} \times \mathbb{Q}_2$ whose inverse branch $\underline{h}_a: (x, y) \mapsto (h_a(x), h_a(y))$ is a bijection from $\underline{\mathcal{I}}$ to $\underline{\mathcal{I}}_a$.

7.4.2 Measures on \mathbb{Q}_2

To properly speak of the transfer operator, or even the Perron Frobenius operator, we must first define a base measure which extends the Lebesgue measure from the one-component CL system. By “base” measure we mean the measure with respect to which the densities (the Perron Frobenius operator works on densities!) are to be integrated to define probabilities. This base measure is key, as it leads to the form of the Perron Frobenius operator of the system, and it is fundamental that there be a change of variable formula.

The Haar measure. Given an appropriate (Hausdorff locally compact) abelian topological group $(G, +, 0)$ there is a unique measure ν (up to scalars) that is translation invariant, i.e., $\nu_G(A + v) = \nu_G(A)$ for every Borel set $A \subset G$ and $v \in A$. This measure is called the Haar measure of the group and is finite on each compact set of G . For a proof of this fact see [RV99].

We consider three basic subdomains of \mathbb{Q}_2

$$\mathcal{B} := \mathbb{Q}_2 \cap \{|y|_2 < 1\}, \quad \mathcal{U} := \mathbb{Q}_2 \cap \{|y|_2 = 1\}, \quad \mathcal{C} := \mathbb{Q}_2 \cap \{|y|_2 > 1\}, \quad (7.24)$$

and we will denote by \mathbb{Z}_2 the closed unit ball $\mathbb{Z}_2 := \mathcal{U} \uplus \mathcal{B}$.

Since \mathbb{Q}_2 is locally compact, there is such a Haar measure ν_0 on \mathbb{Q}_2 . We may normalize our Haar measure ν_0 further so that $\nu_0(\mathcal{U}) = 1/3$ (see [RV99]).

If our measure is to be translation invariant, we must have

$$\nu_0(2^k \mathcal{U}) = 2\nu_0(2^{k+1} \mathcal{U}) \quad (7.25)$$

for every $k \in \mathbb{Z}$.

Observation 7.2. The measure ν_0 satisfies $\nu_0(2^k \mathcal{U}) = (1/3)2^{-k}$ for any $k \in \mathbb{Z}$. Since the sets $\{2^k \mathcal{U} : k \in \mathbb{Z}\}$ actually form a basis for the dyadic topology, this characterizes the measure ν_0 uniquely.

Thus, our measure ν_0 cannot be finite and does not give a probability on \mathbb{Q}_2 . The measure ν_0 actually satisfies a change of variable formula akin to that of the Lebesgue integral. Here we first explain it for quotients of linear functions, which in particular tells us that the change of variables holds for the inverse branches h of the CL system.

Lemma 7.2. *Let ν_0 be a translation invariant measure on \mathbb{Q}_2 , and let h be a non-constant rational function $h(y) = \frac{ay+b}{cy+d}$. Then for any measurable $F: \mathbb{Q}_2 \rightarrow \mathbb{C}$ with $F \geq 0$ or $F \in L^1(\nu_0)$, we have the change of variables formula*

$$\int_{\mathbb{Q}_2} |h'(y)|_2 F(h(y)) d\nu_0(y) = \int_{\mathbb{Q}_2} F(y) d\nu_0(y).$$

Proof. It is enough, by compositions, to prove the result when h is of the form $h(y) = y + b$, $h(y) = ay$ or $h(y) = 1/y$. By working with linear combinations of simple sets, it is enough to prove the results for the case in which F is the characteristic function of a set of the form $B := t + 2^k \mathbb{Z}_2$ with $2^{-k} < |t|_2$, as these generate the σ -algebra. This is a classical strategy in measure theory (see e.g., [Fol99]).

The case $h(y) = y + b$ is just the translation invariance of the measure ν_0 . Consider then $h(y) = ay$ first.

Then

$$\int_{\mathbb{Q}_2} |h'(y)|_2 F(h(y)) d\nu_0(y) = |a|_2 \int_{\mathbb{Q}} \mathbf{1}_B(ay) d\nu_0(y) = |a|_2 \int_{\mathbb{Q}} \mathbf{1}_{B/a}(y) d\nu_0(y) = |a|_2 \nu_0(B/a).$$

Note that $B/a = t/a + (2^k/a)\mathbb{Z}_2 = t/a + (2^k|a|_2)\mathbb{Z}_2$ which by (7.25) and the translation invariance has measure $\nu_2(B/a) = |a|_2^{-1} \nu(B)$. Thus we get our change of variables for this case.

Lastly consider the case $h(y) = 1/y$. We claim that it is enough to prove the result for $B = 2^k\mathbb{Z}_2$, i.e., taking $t = 0$. Indeed, assume the result holds for $t = 0$ and any k . Consider $t \neq 0$, then we will have

$$\int_{\mathbb{Q}_2} |y|_2^{-2} F(1/y) d\nu_0(y) = \sum_{j=-\infty}^{\infty} 2^{2j} \int_{2^j\mathcal{U}} \mathbf{1}_{t+2^k\mathbb{Z}_2}(1/y) d\nu_0(y) = \sum_{j=-\infty}^{\infty} 2^{2j} \int_{2^{-j}\mathcal{U}} |y|_2^{-2} \mathbf{1}_{t+2^k\mathbb{Z}_2}(y) d\nu_0(y),$$

where we first decomposed \mathbb{Q}_2 into the disjoint union of $(2^j\mathcal{U})_j$, and then applied our assumption. Now $(t + 2^k\mathbb{Z}_2) \cap (2^{-j}\mathcal{U})$ is not empty only when $|t|_2 = 2^j$ and so the intersection is the whole $t + 2^k\mathbb{Z}_2$. Thus, considering the only non-zero term, for which $|t|_2 = 2^j$, we have

$$\int_{\mathbb{Q}_2} |y|_2^{-2} F(1/y) d\nu_0(y) = \sum_{j=-\infty}^{\infty} 2^{2j} \times \left(\int_{2^{-j}\mathcal{U}} |y|_2^{-2} \mathbf{1}_{t+2^k\mathbb{Z}_2}(y) d\nu_0(y) \right) = |t|_2^2 \times (|t|_2^{-2} \nu_0(t + 2^k\mathbb{Z}_2)),$$

thereby proving the result for $t \neq 0$.

Now we prove that the result holds for $B = 2^k\mathbb{Z}_2$ with $k \in \mathbb{Z}$. Indeed

$$\int_{\mathbb{Q}_2} |y|_2^{-2} F(1/y) d\nu_0(y) = \sum_{j=-\infty}^{\infty} 2^{2j} \int_{2^j\mathcal{U}} \mathbf{1}_{2^k\mathbb{Z}_2}(1/y) d\nu_0(y) = \sum_{j=-\infty}^{\infty} 2^{2j} \nu_0\left((2^j\mathcal{U}) \cap (2^{-k}\mathbb{Z}_2)\right),$$

and here $(2^j\mathcal{U}) \cap (2^{-k}\mathbb{Z}_2) = 2^j\mathcal{U}$ when $j \leq -k$ and is empty otherwise. Thus, applying (7.25) we deduce

$$\int_{\mathbb{Q}_2} |y|_2^{-2} F(1/y) d\nu_0(y) = \sum_{j=-\infty}^{-k} 2^{2j} \nu_0(2^j\mathcal{U}) = \sum_{j=k}^{\infty} 2^{-2j} 2^j \nu_0(\mathcal{U}) = \sum_{j=k}^{\infty} \nu_0(2^j\mathcal{U}) = \nu_0(2^k\mathbb{Z}_2),$$

as desired. ■

Lemma 7.2 can be generalized to quite an extent, giving the classical change of variables formula but for p -adics. We refer the reader to the paper [Eva06, see Proposition 2.3], which states the multivariable change of variables over the p -adics. The book [Sch84, see Appendix A.7] gives a very readable account of such results for complete (locally compact) non-archimedean fields that are non-trivially valued. Bourbaki [Bou07], chapter 10, p.36. gives a general version of this result for Haar measures and manifolds.

A function $f: \mathbb{Q}_2 \rightarrow \mathbb{Q}_2$ is continuously differentiable if and only if there is a continuous $R: \mathbb{Q}_2 \times \mathbb{Q}_2 \rightarrow \mathbb{Q}_2$ such that $f(x) - f(y) = R(x, y)(x - y)$ for all $x, y \in \mathbb{Q}_2$. Then we define $f'(x) := R(x, x)$. We remark that for the case of the rational functions from Lemma 7.2 both notions coincide.

Theorem 7.1. *Let $g: \mathbb{Q}_2 \rightarrow \mathbb{Q}_2$ be a continuously differentiable bijection satisfying $g'(y) \neq 0$ for all y . Then for a function $F \in L^1(\mathbb{Q}_2, \nu_0)$, we have the change of variable formula*

$$\int_{\mathbb{Q}_2} |g'(y)|_2 F(g(y)) d\nu_0(y) = \int_{\mathbb{Q}_2} F(y) d\nu_0(y).$$

The proof of this result, given in [Sch84] for ultrametrics, goes through two steps. First, there is a key local invertibility theorem stating that for sufficiently small balls B around a point a , the image $g(B)$ is a ball of radius $|g'(a)|_2 \text{diam}(B)$ around $g(a)$. Then the proof concludes by noticing that $|g'|_2$ is locally constant.

Thus the proof relies strongly on the ‘‘ultrametric’’ properties of \mathbb{Q}_2 . It is true, however, that even though the proofs look different from the surface, the main proof steps are similar to those of the Lebesgue measure (see e.g., in [Apo74]).

Probability measure on \mathbb{Q}_2 . We now introduce our probability measure ν , defined by

$$\nu(A) := \int_A G_2(y) d\nu_0(y), \quad (7.26)$$

for every measurable $A \subset \mathcal{B}_{\mathbb{Q}_2}$.

It is simple enough to see that this defines probability measure on \mathbb{Q}_2 .

Proposition 7.6. *The measure ν is a probability measure on \mathbb{Q}_2 .*

Proof. We need to verify that $\nu(\mathbb{Q}_2) = 1$. Note that $\nu_0(\mathbb{Z}_2) = 1/3(\nu_0(\mathcal{U}) + \nu_0(2\mathcal{U}) + \dots) = 2/3$.

Of course, for $|y|_2 \leq 1$ we have $G_2(y) = 1$, hence $\nu(\mathbb{Z}_2) = 2/3$ too. Now we must show that $\nu(\mathcal{C}) = 1/3$.

Note that by undoing the change of variables $y \mapsto 1/y$ we have

$$\nu(\mathcal{C}) = \int_{\mathcal{C}} G_2(y) d\nu_0(y) = \int_{\mathcal{C}} |y|_2^{-2} d\nu_0(y) = \int_{\mathcal{B}} d\nu_0(y) = \nu(\mathcal{B}).$$

Observe that, as $\nu = \nu_0$ over \mathbb{Z}_2 ,

$$\nu(\mathcal{B}) = \nu(2\mathcal{U}) + \nu(2^2\mathcal{U}) + \dots = 2^{-1}\nu(\mathcal{U}) + 2^{-2}\nu(\mathcal{U}) + \dots = \nu(\mathcal{U}),$$

hence $\nu(\mathcal{B}) = \nu(\mathcal{U}) = 1/3$, and the proposition follows. ■

Observation 7.3. We remark that the measure ν satisfies $\nu(2^k\mathcal{U}) = (1/3)2^{-|k|}$ for any $k \in \mathbb{Z}$.

From Lemma 7.2 and $d\nu = G_2 d\nu_0$, we deduce the following change of variables formula, valid for any $F \in L^1(\mathbb{Q}_2, \nu)$,

$$\int_{\mathbb{Q}_2} |h'(y)|_2 F(h(y)) \left[\frac{G_2(h(y))}{G_2(y)} \right] d\nu(y) = \int_{\mathbb{Q}_2} F(y) d\nu(y). \quad (7.27)$$

Measure for the extended system Now, given the probability ν on \mathbb{Q}_2 we can define our probability for the extended system. This will be the product measure

$$d\rho(x, y) := d\lambda_{\text{Leb}}(x) \times d\nu(y). \quad (7.28)$$

7.4.3 Density transformer and transfer operator

We now consider the ‘‘Perron Frobenius operator’’ $\underline{\mathbf{H}}$ associated with the extended system. The density transformer reads as follows: given a function $F \in L^1(\underline{\mathcal{I}}, \rho)$, it returns a new function defined by

$$\underline{\mathbf{H}}[F](x, y) := \sum_{h \in \mathcal{H}} |h'(x)| |h'(y)|_2 F(h(x), h(y)) \left[\frac{G_2(h(y))}{G_2(y)} \right]. \quad (7.29)$$

When F is a density in $L^1(\underline{\mathcal{I}}, \rho)$, then $\underline{\mathbf{H}}[F]$ is indeed the new density on $\underline{\mathcal{I}}$ after one iteration of the shift \underline{T} . This follows easily from the change of variables formula (7.27) applied to each inverse branch $h \in \mathcal{H}$.

Proposition 7.5 leads us to a new operator that depends on a quadruple (t, u, v, z) ,

$$\mathbf{H}_{t,u,v,z}[F](x, y) := \sum_{h \in \mathcal{H}} |h'(x)|^t |h'(y)|_2^u d(h)^v F(h(x), h(y)) \left[\frac{G_2(h(y))}{G_2(y)} \right]^z. \quad (7.30)$$

We will focus on costs described in Figure 7.2: we thus deal with operators associated with quadruples $\gamma_{\mathcal{C}}(s, w)$ defined in Proposition 7.5, and in particular with the quadruple $(s, s, 0, s)$, and its associated operator $\underline{\mathbf{H}}_s := \mathbf{H}_{s,s,0,s}$.

7.4.4 Alternative expressions of the Dirichlet generating functions

We start from the expressions in Proposition 7.4. Consider the three types of DGF defined in (7.21) and (7.22), use the equality $G_2(0) = 1$, and consider the operator $\underline{\mathbf{J}}_s$ relative to the branch J used in the last step. For the plain DGF in (7.21), we obtain

$$S(s) = \sum_{h \in \mathcal{H}^* \cdot J} |h'(0)|^s |h'(0)|_2^s G_2^s \circ h(0) = \underline{\mathbf{J}}_s \circ (I - \underline{\mathbf{H}}_s)^{-1}[1](0, 0), \quad (7.31)$$

We now consider the bivariate DGF's defined in (7.22). For the depth K , one has

$$S_K(s, w) = e^w \underline{\mathbf{J}}_s \circ (I - e^w \underline{\mathbf{H}}_s)^{-1}[1](0, 0);$$

For costs C of Figure 7.2, the bivariate DGF involves the quasi-inverse of $\mathbf{H}_{\gamma_C(s, w)}$,

$$S_C(s, w) = \mathbf{J}_{\gamma_C(s, w)} \circ (I - \mathbf{H}_{\gamma_C(s, w)})^{-1}[1](0, 0),$$

except for $C = |Q|_2^{-2}$, where the function 1 is replaced by the function G_2^w .

The DGF $\widehat{S}_C(s)$ defined in (7.22) is obtained with taking the derivative of the bivariate DGF with respect to w (at $w = 0$); it is thus written with a double^f quasi inverse which involves the plain operator $\underline{\mathbf{H}}_s$, separated “in the middle” by the cumulative operator $\underline{\mathbf{H}}_{s, (C)}$, namely

$$\widehat{S}_C(s) \asymp \underline{\mathbf{J}}_s \circ (I - \underline{\mathbf{H}}_s)^{-1} \circ \underline{\mathbf{H}}_{s, (C)} \circ (I - \underline{\mathbf{H}}_s)^{-1} [1](0, 0), \quad (7.32)$$

and the cumulative operator is itself defined by $\underline{\mathbf{H}}_{s, (C)} := \left. \frac{\partial}{\partial w} \mathbf{H}_{\gamma_C(s, w)} \right|_{w=0}$.

7.5 Functional Analysis

This section is devoted to the study of the quasi-inverses $(I - \underline{\mathbf{H}}_s)^{-1}$ intervening in the expressions of the generating functions of interest. We deal with a delicate context: even though the inverse branches are contracting on the interval \mathcal{I} , they are not contracting on \mathbb{Q}_2 , just contracting on average. We first define an appropriate functional space on which the operators are proven to act and admit dominant spectral properties. This will prove that the quasi-inverse $(I - \underline{\mathbf{H}}_s)^{-1}$ admits a pole at $s = 1$, and we study its residue, which gives rise to the constants that appear in the expectations of our main costs.

7.5.1 Lasota-Yorke for the classical CL dynamical system

We kick off this section by explaining why the CL dynamical system, that is, the system with just the **real** component, is so well-behaved. This good behavior will end up extending to the extended system.

The following Lemma shows that the transfer operator \mathbf{H}_t operating over $C^1(\mathcal{I})$, defined by

$$\mathbf{H}_t : f(x) \mapsto \sum_{a \geq 0} |h'_a(x)|^t f(h_a(x)),$$

satisfies *Lasota-Yorke bound* like that in Theorem 1.9. Over the reals we consider the norm $|\cdot|_0$ and the semi-norm $|\cdot|_1$, related respectively to $\|\cdot\|_0$ and $\|\cdot\|_1$ for the extended case with space \mathcal{F} , defined by

$$|f|_0 := \sup_{x \in \mathcal{I}} |f(x)|, \quad |f|_1 := \sup_{x \in \mathcal{I}} |f'(x)|.$$

^fThere is another term which involves only a quasi-inverse. It does not intervene in the analysis.

This Lasota-Yorke bound, along with the compactness of the closed unit ball

$$B = \{f \in C^1(\mathcal{I}) : |f|_0 + |f|_1 \leq 1\}$$

of the Banach space $(C^1(\mathcal{I}), |\cdot|_0 + |\cdot|_1)$, over $(C^1(\mathcal{I}), |\cdot|_0)$, imply that the operator \mathbf{H}_t is quasi-compact due to Hennion's Theorem (see Theorem 1.9). The compactness of the ball B over $(C^1(\mathcal{I}), |\cdot|_0)$ follows at once from the celebrated *Arzelà-Ascoli Theorem*[§].

Lemma 7.3. *The component operator $\mathbf{H}_{t,(a)}$ associated to $a \in \mathbb{N}$ defined as*

$$\mathbf{H}_{t,(a)}[f](x) = |h'_a(x)|^t f(h_a(x))$$

acts on $C^1(\mathcal{I})$ and the integral

$$K_a[f, t] := \int_{\mathcal{I}} [\mathbf{H}_{t,(a)}[f](x) + [\mathbf{H}_{t,(a)}[f](x)]'] dx$$

satisfies

$$|K_a[f](t)| \leq I_a(t)|f|_0(1 + 2|t|) + |f|_1 I_a(t + 1), \quad \text{with} \quad I_a(t) := \int_{\mathcal{I}} |h'_a(x)|^t dx.$$

Moreover, the integral $I_a(t)$ satisfies

$$I_a(1) = 2^{-1} 2^{-at}, \quad I_a(t) \leq 2^{-1} 2^{-at} \quad (t \geq 1), \quad I_a(t) \leq 2^{1-2t} 2^{-at} \quad (t \leq 1)$$

This admits a generalization to k -uples $\mathbf{a} := (a_1, a_2, \dots, a_k)$ and leads to the Lasota-Yorke bound,

$$|\mathbf{H}_t^k[f]|_1 \leq (1 + 2k|t|)|f|_0 + \varrho^k |f|_1$$

that holds for $\Re t = 1$. Here, ϱ is the contraction ratio, and we have used the fact that $|h''(x)| \leq 2k|h'(x)|$ for any branch h of depth k .

7.5.2 Functional space

The delicate point of the dynamical analysis is finding a good functional space. It must surely be a subspace of $L^1(\underline{\mathcal{I}}, \rho)$. Here, we know that, in the initial *CL* system, the transfer operator \mathbf{H}_s acts nicely on the space of continuously differentiable functions $C^1(\mathcal{I})$. As in our case it is the real part of the system that leads the evolution, we will just ask for some sort of integrability condition over \mathbb{Q}_2 .

For a function F defined on $\underline{\mathcal{I}}$, the main role will be played by the family of “sections” $F_y : x \mapsto F(x, y)$ which will be asked to belong to $C^1(\mathcal{I})$, with the following norm $|\cdot|_{1,1}$,

$$|F_y|_{1,1} := |F_y|_0 + |F_y|_1, \quad \text{with} \quad |F_y|_0 := \sup_{x \in \mathcal{I}} |F_y(x)|, \quad |F_y|_1 := \sup_{x \in \mathcal{I}} \left| \frac{\partial}{\partial x} F_y(x) \right|.$$

We then just ask for the function $y \mapsto |F_y|_{1,1}$ to be bounded on \mathbb{Q}_2 and deal with the Banach space

$$\mathcal{F} := \{F : \underline{\mathcal{I}} \rightarrow \mathbb{C} : F_y \in C^1(\mathcal{I}), \quad y \mapsto F_y \text{ in } L^1(\mathbb{Q}_2, \nu)\}$$

endowed with the norm $\|F\| := \|F\|_0 + \|F\|_1$, with

$$\|F\|_0 := \int_{\mathbb{Q}_2} |F_y|_0 d\nu(y), \quad \|F\|_1 := \int_{\mathbb{Q}_2} |F_y|_1 d\nu(y). \quad (7.33)$$

[§]For $f \in \mathcal{B} : |f(x) - f(x_0)| = \left| \int_{x_0}^x f'(t) dt \right| \leq |x - x_0| |f|_1 \leq |x - x_0|$, which proves the equicontinuity.

7.5.3 Action of the operator $\mathbf{H}_{t,u,v,z}$ on \mathcal{F}

We are interested in determining the (or a sufficiently large) set of quadruples (t, u, v, z) for which the operator $\mathbf{H}_{t,u,v,z}$ defined in (7.30) is bounded on \mathcal{F} and, moreover, is analytic with respect to the quadruple (for the analytic perturbation!).

As we will deal mainly with integrals arising from a change of variables, it will be important to underline the following property of the inverse branches h

Observation 7.4. Consider an inverse branch $h_a(x) = \frac{1}{2^a(1+x)}$ of depth 1. Then

$$|h'_a(x)|^t = |h'_a(x)|^{t-1} |h'_a(x)| = |2^a (h_a(x))^2|^{t-1} |h'_a(x)|, \quad (7.34)$$

which of course holds for the dyadic norm $|\cdot|_2$ as well. The factor $|h'_a(x)|$ on the right-hand side of (7.34) will serve to perform the change of variables when working with integrals. Thus

$$\int_{\mathcal{I}} |h'_a(x)|^t dx = \int_{h_a(\mathcal{I})} |2^a u^2|^{t-1} du, \quad (7.35)$$

even when changing $|\cdot|$ for the norm $|\cdot|_2$ and the Lebesgue measure for the Haar measure ν_0 on \mathbb{Q}_2 . \diamond

Before moving on to the corresponding results, we motivate the constraint $u = z$ which we will impose on our domain for (t, u, v, z) . Such a constraint is not a problem (and we claim it is necessary) when it comes to the concrete cases of our GFs (recall Proposition 7.5).

Given $R: \mathbb{Q}_2 \rightarrow [0, \infty)$, we consider the integral

$$J_{a,u,z}[R] := \int_{\mathbb{Q}_2} G_2^{-z}(y) |h'_a(y)|_2^u R(h_a(y)) G_2^z(h_a(y)) G_2(y) d\nu_0(y), \quad (7.36)$$

which we wish to compare with $\int_{\mathbb{Q}_2} R(y) G_2(y) d\nu_0(y)$.

We perform the change of variables $x = h_a(y)$ and we note the two following equalities, following from (7.34) and the previous discussion, according as to whether $y \in \mathcal{C}$

$$G_2(y)^{1-z} |h'_a(y)|_2^u = \begin{cases} 2^{-a(u-1)} |x|_2^{2(u-1)} |h'_a(y)|_2, & y \notin \mathcal{C} \\ 2^{-a(u-2z+1)} |x|_2^{2(u-z)} |h'_a(y)|_2, & y \in \mathcal{C} \end{cases}.$$

Then, the integral decomposes into three integrals

$$\begin{aligned} J_{a,u,z}[R] &= 2^{-a(u-1)} \int_{h_a(\mathcal{B})} |x|_2^{2(u-1)} R(x) G_2^z(x) d\nu_0(x) \\ &\quad + 2^{-a(u-1)} \int_{h_a(\mathcal{U})} |x|_2^{2(u-1)} R(x) G_2^z(x) d\nu_0(x) \\ &\quad + 2^{-a(u-2z+1)} \int_{h_a(\mathcal{C})} |x|_2^{2(u-z)} R(x) G_2^z(x) d\nu_0(x). \end{aligned}$$

Why consider $u = z$. Assume, for the sake of motivating the need for the equality $u = z$, that both u and z were real numbers (else we would take the real parts).

First consider the case $u > z$. Then we would have $h_a(\mathcal{U}) \subset \mathcal{C}$. This means that for the integral over $h_a(\mathcal{U})$ we have $G_2(x) = |x|_2^{-2}$ and get an integral which is $|x|_2^{2(u-z)}$ times our target $R(x)G_2(x)$, but the factor $|x|_2^{2(u-z)}$ is unbounded!

If we had $u < z$, then the factor multiplying the integral over $h_a(\mathcal{C})$ will give us a divergent series, even if we could compare the corresponding integral with $\int_{\mathbb{Q}_2} R(x)G_2(x)d\nu_0(x)$.

Comparison when $u = z$. When $z = u \in \mathbb{C}$ notice that we have $|x|_2^{-2} = G_2(x)$ for the first two integrals, whence $|x|_2^{-2(u-1)} = G_2^{z-1}(x)$. Indeed, whenever $x \in h_a(\mathcal{B}) \cup h_a(\mathcal{U})$ we have $|x|_2 \geq 2^a \geq 1$.

For the third, the factor $|x|_2^{2(u-z)}$ vanishes, and we may have a large $(G_2(x))^{z-1}$ only when $x \in h_a(\mathcal{C}) \cap \mathcal{C}$ and $\Re z - 1$ is negative. By definition of h_a we have $1 < |x|_2 < 2^a$, whence $|G_2^{z-1}(x)| \leq \max\{1, 2^{2a\Re(1-z)}\}$.

We conclude the ensuing discussion with the following lemma.

Lemma 7.4. *Given a function $R: \mathbb{Q}_2 \rightarrow [0, \infty)$ and a real u we define*

$$J_{a,u}[R] := \int_{\mathbb{Q}_2} G_2^{-u}(y) |h'_a(y)|_2^u R(h_a(y)) (G_2(h_a(y)))^u G_2(y) d\nu_0(y).$$

Then $J_{a,u}[R]$ satisfies

$$\begin{aligned} J_{a,u}[R] &= 2^{-a(u-1)} \int_{h_a(\mathcal{B})} R(x) G_2(x) d\nu_0(x) \\ &\quad + 2^{-a(u-1)} \int_{h_a(\mathcal{U})} R(x) G_2(x) d\nu_0(x) \\ &\quad + 2^{-a(1-u)} \int_{h_a(\mathcal{C})} R(x) G_2^{u-1}(x) G_2(x) d\nu_0(x), \end{aligned}$$

as well as the bounds

$$|J_{a,u}[R]| \leq 2^{a|1-u|} \int_{\mathbb{Q}_2} R(y) G_2(y) d\nu_0(y), \quad \left| \frac{\partial}{\partial u} J_{a,u}[R] \right| \leq 4 \cdot 2^{a|1-u|} \cdot (1+a) \int_{\mathbb{Q}_2} R(y) G_2(y) d\nu_0(y).$$

Proof. The only thing we have not explained is the second bound. For this we use our expression for $J_{a,u}[R]$ as a sum of three integrals. Notice that differentiating in u the first 2 terms produces a factor $-a \times (\log 2)$. For the third integral we have that $|\log G_2(x)| \leq 1 + 2(\log 2)a$ is satisfied, while the derivative of $2^{-a(1-u)}$ also gives a term with a factor a . This yields the inequality. \blacksquare

We are now ready to state the key result concerning our domain.

Proposition 7.7. *For a triple $(t, u, v) \in \mathbb{C}^3$ satisfying the conditions $\Re(t) > 0$ and $\Re(t - v - |u - 1|) > 0$, the operator $\mathbf{H}_{t,u,v,u}$ acts on \mathcal{F} and is analytic with respect to the triple (t, u, v) .*

Proof. We divide the proof into two parts: the proof that the operator acts on \mathcal{F} , and the proof that it depends analytically on the triples (t, u, v) .

The operator acts on the space. Observe that by the triangle inequality

$$\sup_{x \in \mathcal{I}} |\mathbf{H}_{t,u,v,u}[F](x, y)| \leq \sum_{a=0}^{\infty} |h'_a(x)|^{\Re t} |h_a(y)|^{\Re u} d(h_a)^{\Re v} \sup_{x \in \mathcal{I}} |F(x, h_a(y))| \left[\frac{G_2(h_a(y))}{G_2(y)} \right]^{\Re u}.$$

Consider then Lemma 7.4 with $R(y) := |F_y|_0$, we get that

$$\|\mathbf{H}_{t,u,v,u}[F]\|_0 \leq \sum_{a=0}^{\infty} 2^{-a(\Re(t-v-|1-u|))} \|F\|_0$$

when $\Re t > 0$, as $|h'(x)| \leq 2^{-a}$. For our given conditions, the series $\sum_a 2^{-a(\Re(t-v-|1-u|))}$ converges.

We must now consider $\|\mathbf{H}_{t,u,v,u}[F]\|_1$. By differentiating we find that $\frac{\partial}{\partial x} \mathbf{H}_{t,u,v,u}[F]$ satisfies

$$\begin{aligned} \left| \frac{\partial}{\partial x} \mathbf{H}_{t,u,v,u}[F](x, y) \right| &\leq \sum_{a=0}^{\infty} |t| |h'_a(x)|^{\Re t - 1} |h''_a(x)| |h'_a(y)|_2^{\Re u} d(h_a)^{\Re v} F(h_a(x), h_a(y)) \left[\frac{G_2(h_a(y))}{G_2(y)} \right]^{\Re u} \\ &\quad + \sum_{a=0}^{\infty} |h'_a(x)|^{\Re t + 1} |h'_a(y)|_2^{\Re u} d(h_a)^{\Re v} \frac{\partial F}{\partial x}(h_a(x), h_a(y)) \left[\frac{G_2(h_a(y))}{G_2(y)} \right]^{\Re u}. \end{aligned}$$

The inverse branches h_a are of bounded distortion $|h_a''(x)| \leq 2|h_a'(x)|$, hence we may write

$$\left| \frac{\partial}{\partial x} \mathbf{H}_{t,u,v,u}[F](x,y) \right| \leq 2|t| \sum_{a=0}^{\infty} |h_a'(x)|^{\Re t} |h_a'(y)|_2^{\Re u} d(h_a)^{\Re v} F(h_a(x), h_a(y)) \left[\frac{G_2(h_a(y))}{G_2(y)} \right]^{\Re u} \\ + \sum_{a=0}^{\infty} |h_a'(x)|^{\Re t+1} |h_a'(y)|_2^{\Re u} d(h_a)^{\Re v} \frac{\partial F}{\partial x}(h_a(x), h_a(y)) \left[\frac{G_2(h_a(y))}{G_2(y)} \right]^{\Re u}.$$

Now we note again that $|h_a'(x)| \leq 2^{-a}$, take \sup_x and apply Lemma 7.4. After the integration on $y \in \mathbb{Q}_2$ with the measure $d\nu(y) = G_2(y)d\nu_0(y)$, we get

$$\|\mathbf{H}_{t,u,v,u}[F]\|_1 \leq 2|t| \sum_{a=0}^{\infty} 2^{-a\Re(t-v-|1-u|)} \|F\|_0 + \sum_{a=0}^{\infty} 2^{-a\Re(t-v-|1-u|+1)} \left\| \frac{\partial}{\partial x} F \right\|_0.$$

This last inequality implies that the operator is bounded, as then

$$\|\mathbf{H}_{t,u,v,u}[F]\| \leq (2|t| + 1) \left(\sum_{a=0}^{\infty} 2^{-a\Re(t-v-|1-u|)} \right) \times \|F\|,$$

and the series converges over our domain.

The operator depends analytically on the triple. The analyticity will follow from the fact that the series of the term-wise derivatives belong to \mathcal{F} .

Observe that differentiating $\mathbf{H}_{t,u,v,u}[F](x,y)$ with respect to t gives

$$\frac{\partial}{\partial t} \mathbf{H}_{t,u,v,u}[F](x,y) = \sum_{a=0}^{\infty} (\log |h_a'(x)|) |h_a'(x)|^t |h_a'(y)|_2^u d(h_a)^v F(h_a(x), h_a(y)) \left[\frac{G_2(h_a(y))}{G_2(y)} \right]^u.$$

Of course, $\log |h_a'(x)| = -a \cdot (\log 2) + O(1)$ in the variable a , and the rest proceeds as before, showing that this partial derivative gives a bounded operator. Indeed, the additional factor a does not change the convergence/divergence region of the series $\sum_a 2^{a(\Re v + |1-u| - \Re(t))}$.

With respect to v we have an analogous situation. Thus let us get to the interesting case: differentiating with respect to u . It is here that we need the second inequality from Lemma 7.4.

From Lemma 7.4 we deduce

$$\left\| \frac{\partial}{\partial u} \mathbf{H}_{t,u,v,u}[F](x,y) \right\|_0 = \sum_{a=0}^{\infty} 2^{-a\Re(t-v)} \left| \frac{\partial}{\partial u} J_{a,u}[y \mapsto |F_y|_0] \right| \leq 4 \sum_{a=0}^{\infty} (1+a) 2^{-a(\Re(t-v)-|1-u|)} \|F\|_0,$$

where the series converges again on the same domain. And for the other norm $\|\cdot\|_1$ we have, as in the previous part when we considered $\frac{\partial}{\partial x}$,

$$\left\| \frac{\partial}{\partial u} \mathbf{H}_{t,u,v,u}[F](x,y) \right\|_1 \leq 2|t| \sum_{a=0}^{\infty} 2^{-a\Re(t-v)} \left| \frac{\partial}{\partial u} J_{a,u}[y \mapsto |F_y|_0] \right| + \sum_{a=0}^{\infty} 2^{-a\Re(1+t-v)} \left| \frac{\partial}{\partial u} J_{a,u}[y \mapsto |F_y|_1] \right|,$$

and the boundedness follows again from Lemma 7.4 and the convergence of the series. \blacksquare

Proposition 7.8. *The operator $\underline{\mathbf{H}}_s := \mathbf{H}_{s,s,0,s}$ acts on \mathcal{F} for $\Re s > 1/2$. Moreover for each s with $\Re s > 1$ the spectral radius of $\underline{\mathbf{H}}_s$ is strictly less than 1.*

7.5.4 Dominant spectral properties of the operator

The next result describes some of the main spectral properties of the operator on the space \mathcal{F} . Assertion (a) entails that the k -th iterate of the operator behaves as a true k -th power of its dominant eigenvalue. Then, as stated in (c), its quasi-inverse behaves as a true quasi-inverse which involves its dominant eigenvalue.

Proposition 7.9. *The following properties hold for the operator $\mathbf{H}_{t,u,v,u}$, when the triple (t, u, v) belongs to a neighborhood \mathcal{V} of $(1, 1, 0)$.*

- (a) *At $(t, u, v, u) = (1, 1, 0, 1)$, the operator $\mathbf{H}_{t,u,v,u}$ coincides with the density transformer $\underline{\mathbf{H}}_1$. The operator $\underline{\mathbf{H}}_1$ has a unique dominant eigenvalue equal to 1, the corresponding eigenvector is the invariant density Ψ and the projection is given by the integration with respect to the measure ρ .*
- (b) *There is a unique dominant eigenvalue with multiplicity one, separated from the remainder of the spectrum by a spectral gap, and denoted as $\lambda(t, u, v)$, with a (normalized) dominant eigenfunction $\Psi_{t,u,v}$ and a dominant eigenmeasure $\rho_{t,u,v}$ for the dual operator.*
- (c) *The estimate holds for any function $F \in \mathcal{F}$ with $\rho[F] \neq 0$,*

$$(I - \mathbf{H}_{t,u,v,u})^{-1}[F](x, y) \sim \frac{\lambda(t, u, v)}{1 - \lambda(t, u, v)} \Psi_{t,u,v}(x, y) \rho_{t,u,v}[F],$$

as $(t, u, v) \rightarrow (1, 1, 0)$.

- (d) *For $\Re s = 1, s \neq 1$, the spectral radius of $\mathbf{H}_{s,s,0,s}$ is strictly less than 1.*

The third result describes the Taylor expansion of $\lambda(t, u, v)$ at $(1, 1, 0)$, and makes precise the behavior of the quasi-inverse described in (c).

Proof.

(a) We already know that the Perron Frobenius operator \mathbf{H} of the CL system presents a spectral gap, with a dominant eigenvalue 1 that is simple, and no other eigenvalue on the unit circle. We know that $\underline{\mathbf{H}}_1$ inherits the Lasota–Yorke bound from that of \mathbf{H} . Then, since $\underline{\mathbf{H}}_1$ is actually an extension of \mathbf{H} to two-variable functions, the density transformer $\underline{\mathbf{H}}_1$ of the extended system will inherit this spectral gap.

(b) There is a Lasota–Yorke bound, inherited from Lemma 7.3, for the $\mathbf{H}_{1,1,0,1}$ in \mathcal{F} with the two norms $\|\cdot\|_0$ and $\|\cdot\|_1$. The unit ball in $(\mathcal{F}, \|\cdot\|_1)$ is precompact on $(\mathcal{F}, \|\cdot\|_0)$ again by the Arzelà–Ascoli Theorem, and we apply Hennion’s Theorem, yielding the quasi-compactness of $\mathbf{H}_{t,u,v,u}$.

(c) Follows from classical principles as in subsection 1.4.7.

(d) We assume that there exists a complex number s of the form $s = 1 + it, t \neq 0$, with t real, for which $\underline{\mathbf{H}}_s$ has a spectral radius equal to 1. Then, due to the Lasota–Yorke inequality satisfied by $\underline{\mathbf{H}}_s$, the operator $\underline{\mathbf{H}}_s$ is quasi-compact, and therefore it admits an eigenvalue of modulus 1. We follow the same first steps as in “Dynamical Sources in Information Theory: Fundamental intervals and Word Prefixes” [Val01] Proposition 9 [basic case] that prove that the following holds

$$\left| \frac{1}{2^a(1+x_a)^2} \right|^{it} = 1 \quad \text{for any } a \geq 0$$

and involves the quadratic irrational $x_a \in [0, 1]$ that is the fixed point of the branch h_a .

Remark two facts:

- (i) each x_a is a non rational element of the quadratic field $\mathbb{Q}(\sqrt{1 + 2^{2-a}})$, and
- (ii) we have the inequality $0 < x_a < 2^{-a}$.

Then, for any a large enough, there exists an integer k_a for which

$$\left| k_a \frac{2\pi}{t \log 2} - a \right| = \log_2(1 + x_a), \quad 0 < \log_2(1 + x_a) \leq K2^{-a} \quad \text{for some constant } K.$$

This proves that the sequence $k_{2a} - 2k_a$ tends to 0; it is thus equal to 0 for a large enough. This now entails that the two irrational quadratics x_a^2 and x_{2a} are equal. This is not possible since $\mathbb{Q}(\sqrt{1 + 2^{2-a}}) \cap \mathbb{Q}(\sqrt{1 + 2^{2-2a}}) = \mathbb{Q}$, at least when $a > 1$ is odd.

This is intuitively obvious, but the proof is not immediate. This is actually significantly simpler to prove by contradiction for a odd, which can be done as we may pick any sufficiently large odd a . The contradiction arrives after squaring and comparing dyadic valuations.

Going back, now knowing that $\alpha, \gamma \neq 0$, and squaring

$$\alpha^2(1 + 2^{2-a}) = \gamma^2(1 + 2^{2-2a}) + 2\gamma(\delta - \beta)\sqrt{1 + 2^{2-2a}} + (\delta - \beta)^2,$$

and therefore $\delta = \beta$, as $\sqrt{1 + 2^{2-2a}} \notin \mathbb{Q}$. Thus let us look at the possibility $\delta = \beta$. Then

$$\alpha\sqrt{1 + 2^{2-a}} = \gamma\sqrt{1 + 2^{2-2a}},$$

for nonzero integers α, γ . Then

$$\alpha^2(2^a + 4)2^a = \gamma^2(2^{2a} + 4),$$

and this is nonsense when a is odd and $a > 1$ as the left hand side has an odd dyadic valuation, while the right hand side has an even valuation. \blacksquare

We have precise information concerning the eigenvalue $\lambda(t, u, v)$ when (t, u, v) is near $(1, 1, 0)$.

Proposition 7.10. *The Taylor expansion of the eigenvalue $\lambda(t, u, v)$ at $(1, 1, 0)$, written as $\lambda(t, u, v) \sim 1 - A(t - 1) + B(u - 1) + Dv$, involves the constants*

$$A = -\partial\lambda/\partial t(1, 1, 0), \quad B = \partial\lambda/\partial u(1, 1, 0), \quad D = \partial\lambda/\partial v(1, 1, 0)$$

(a) *The constants A and D already appear in the context of the plain dynamical system, and are precisely described in (7.19) and (7.18). In particular $A - D$ is equal to the integral $E := \mathbb{E}_\Psi[2 \log x]$;*

(b) *The constant B is defined in terms of the extension of the dynamical system and its invariant density $\Psi = \Psi_{1,1,0}$. The constant $B + D$ is equal to the dyadic analog E_2 of the integral E , namely, $B + D = E_2 := \mathbb{E}_\Psi[2 \log |y|_2]$;*

(c) *The constant $A - B$ is the entropy of the extended dynamical system.*

About the constant B . The invariant density Ψ –more precisely the function $\widehat{\Psi} := \Psi \cdot G_2$ – satisfies a functional equation of the same type as the invariant function ψ , (described in Equation 7.16), namely,

$$\widehat{\Psi}(x, y) = \left(\frac{1}{1+x} \right)^2 \left| \frac{1}{1+y} \right|_2^2 \sum_{a \geq 0} \widehat{\Psi} \left(\frac{2^{-a}}{1+x}, \frac{2^{-a}}{1+y} \right).$$

Comparing to Equation 7.16, we “lose” the factor 2^{-a} in the sum, and so we have not succeeded in finding an explicit formula for Ψ . We do not know how to evaluate the integral E_2 defined in Proposition 7.10(b). However, we conjecture the equality $D - B = \log 2$, from experiments of the same type as those described in Figure 7.1. This would entail an explicit value for the entropy H of the extended system,

$$H = \frac{1}{2 \log 2 - \log 3} \left[\frac{\pi^2}{6} + 2 \sum_{k \geq 1} \frac{(-1)^k}{k^2 2^k} - (\log 2)(3 \log 3 - 4 \log 2) \right] \doteq 1.33973 \dots$$

7.6 Final result for the analysis of the CL algorithm

We then obtain our final result:

Theorem 7.2. *The mean values $E_N[c]$ for $c \in \{K, \sigma, q, \varrho, r, q_2\}$ on the set Ω_N are all of order $\Theta(\log N)$ and admit the precise following estimates,*

$$E_N[K] \sim \frac{2}{H} \log N, \quad \mathbb{E}_N[c] \sim M(c) \cdot \mathbb{E}_N[K], \quad \text{for } c \in \{\sigma, q, \varrho, r, q_2\}.$$

The constant H is the entropy of the extended system. The constants H and $M(c)$ are expressed with a scalar product that involves the gradient $\nabla\lambda$ of the dominant eigenvalue at $(1, 1, 0)$ and the beginning $\widehat{\gamma}_C$ of the quadruple γ_C associated with the cost c . More precisely

$$H = \langle \nabla\lambda, (-1, -1, 0) \rangle, \quad M(c) = \langle \nabla\lambda, \widehat{\gamma}_C \rangle.$$

The constants $M(c)$, in terms of the derivatives of $\lambda(t, u, v)$ as defined in Proposition 7.10, are exhibited in Figure 7.2 and we recall them here

Cost C	Cost $c = \log C$	Beginning $\widehat{\gamma}_C$	Constant $M(c)$	Numerical value of $M(c)$
$d(h)$	σ	$(0, 0, 1)$	D	$\doteq 0.97693 \dots$
Q^2	q	$(-1, 0, 1)$	$A + D$	$\doteq 2.6004 \dots$
$g^2(P, Q)$	ϱ	$(0, 1, 1)$	$B + D$	$\doteq 1.26030 \dots$
$R^2(P, Q)$	r	$(-1, -1, 0)$	$A - B$	$\doteq 1.33973 \dots$
$ Q _2^{-2}$	q_2	$(0, 1, 1)$	$B + D$	$\doteq 1.26072 \dots$

Proof. Now, the Tauberian Theorem comes into play, relating the behavior of a Dirichlet series $F(s)$ near its dominant singularity with asymptotics for the sum $\Phi_N(F)$ of its first N coefficients. Delange's Tauberian Theorem (Theorem 2.3)

We now show that the two DGFs $S(s)$ and $\widehat{S}_C(s)$ satisfy the hypotheses of the Tauberian theorem. The two expressions obtained in (7.31) and (7.32) involve quasi-inverses $(I - \underline{\mathbf{H}}_s)^{-1}$, a simple one in (7.31), a double one in (7.32).

First, Propositions 7.8 with 7.9(b), and 7.9(d) prove that such quasi-inverses are analytic on $\Re s \geq 1, s \neq 1$. Then Proposition 7.9(c), together with Equation 7.32, shows that $S(s)$ and $\widehat{S}_C(s)$ have a pole at $s = 1$, of order 1 for $S(s)$, of order 2 for $\widehat{S}_C(s)$.

We now evaluate the dominant constants: first, the estimate holds,

$$1 - \lambda(s, s, 0) \sim (A - B)(s - 1) = H(s - 1).$$

Second, with Proposition 7.9(c), the DGFs $S(s)$ and $\widehat{S}_C(s)$ admit the following estimates which both involve the constant $a = \underline{\mathbf{J}}[\Psi](0, 0)$, namely,

$$S(s) \sim \frac{a}{H(s - 1)}, \quad \widehat{S}_C(s) \sim \frac{a}{H^2(s - 1)^2} \rho \left[\underline{\mathbf{H}}_{1,(C)}[\Psi] \right].$$

We now explain the occurrence of the constant $M(c)$: we use the definition of the triple $\widehat{\gamma}_C(s, w)$, the definition of the cumulative operator $\underline{\mathbf{H}}_{1,(C)}$ as the derivative of the bivariate operator $\mathbf{H}_{\gamma_C(s,w)}$ at $(s, w) = (1, 0)$, and the fact that $\mathbf{H}_{1,1,0,1} = \underline{\mathbf{H}}_1$ is the density transformer. This entails the sequence of equalities,

$$\rho \left[\underline{\mathbf{H}}_{1,(C)}[\Psi] \right] = \frac{\partial}{\partial w} \lambda(\widehat{\gamma}_C(1, w)) \Big|_{w=0} = \langle \nabla\lambda, \widehat{\gamma}_C \rangle = M(c). \quad \blacksquare$$

CONCLUSIONS

In this dissertation we have presented several instances of Dynamical Analysis to study objects coming from Combinatorics on Words (namely Sturmian words) and Number Theory (the Continued Logarithm algorithm). In both cases there had been preceding studies on the worst case (see [MH40] and [Sha16]), and we, on the other hand, strive to study them probabilistically. This is where Dynamical Analysis actually comes in, as the objects we consider can be described in a common framework mixing continued fractions and dynamical systems.

Studies in Combinatorics on Words

We have first studied the recurrence function $R_\alpha(n)$ of Sturmian words, a fundamental quantity dictating the worst “waiting time” to find all factors of a given length n , in two different probabilistic models. The fundamental link between Sturmian words and continued fractions is given by Morse and Hedlund in [MH40], and this means that the Euclidean system is the dynamical system underlying here. For our first model, given a Sturmian word of slope α , and a sequence (μ_k) , we have associated a sequence of indices n_k (lengths of factors) defined by their barycentric position μ_k inside $[q_{k-1}(\alpha), q_k(\alpha))$. We then have elucidated the role played by the position in the behavior of the recurrence of a random Sturmian word when α is drawn uniformly at random. In this case the dynamical analysis involves the powers of the density transformer of the underlying system, and constitutes a starting point to our work. Our second model constitutes another take on the problem, with a somewhat orthogonal viewpoint. We fix the length n of the factors and pick α at random, then we study the distribution of the recurrence function. This study a priori differs widely from the previous one; it employs elementary properties of continued fractions as well as coprime Riemann sums over unbounded domain. This idea led us to the concept of \mathcal{Q} -functions, of which we have shown several examples, other than the recurrence quotient, in relation to continued fractions. This study involves the generalization and adaptation of ideas which have appeared in [BCZ03] (for coprime Riemann sums) and [Ust09].

Work in progress. We have since considered the possibility of the slope α belonging to a particular (but important) set, such as the rational numbers (yielding periodic words) or the quadratic irrationals (related to words produced by morphisms). These lead to models mixing the ideas from our two previous studies and requiring several other key concepts such as the transfer operator of the system and the trace. We have obtained partial results in this direction, and individuated the need to introduce an index ℓ , counting the number of complete periods of the continued fraction expansion of α , for the case of quadratic irrationals. The stationary behavior when $\ell \rightarrow \infty$ is an attractive question we are working on. Furthermore, the work

may lead to a unification of all three contexts (real, rational and quadratic irrational).

Studies in Number Theory

We have studied the Continued Logarithm Algorithm on average and analyzed in particular the number of pseudo divisions, and the total number of shifts. Thus we have answered a question posed by Shallit [Sha16] following his analysis on the worst-case. Our work makes crucial use of Dynamical Analysis, but requires a twist: the introduction of a dyadic component to the transfer operator. This dyadic component is necessary to account for the complex dyadic behavior of the algorithm; even if the input pair is coprime, the greatest common divisor of the working pair increases as the algorithm progresses, always being a power of two. We have shown that this system possess the spectral properties we require to carry out our dynamical analysis, namely, there is a spectral gap which follows from the fact that the one-component original system is well-behaved. Thus we complete our analysis deriving constants which depend on the entropy of the system.

Work in progress and open questions. We would like to obtain an explicit expression for the elusive invariant density of the extended system. This would entail a proven expression for the entropy of the dynamical system, for which we have provided a conjectural value. This conjectural value is sound in the sense that it related to a more basic conjecture (see (7.13)), stating that $\frac{1}{k} \log_2 \gcd(q_k, q_{k+1}) \rightarrow \frac{1}{2}$, which seems reasonable, and fits in very well with the experiments performed.

It is also surely possible to analyze the bit complexity of the algorithm, notably in the case when one eliminates the rightmost zeroes when shared by the two q_i 's (as suggested by Shallit). Such a version of this algorithm may have a competitive bit complexity that merits a further study.

There exist two other gcd algorithm that are based on binary shifts, all involving a dyadic point of view: the Binary Algorithm, and “the Tortoise and the Hare” algorithm, already analyzed in [Val98a] and [DMDV05]; however, the role of the binary shifts is different in each case. The strategy of the present algorithm is led by the most significant bits, whereas the strategy of the “Tortoise and the Hare” is led by the least significant bits. The Binary algorithm adopts a mixed strategy, as it performs both right-shifts and subtractions. We have the project to unify the analysis of these three algorithms, and better understand the role of the dyadic component in each case.

Finally, we are working on the so-called “real case” for the Continued Logarithm expansion. We consider a random real in the unit interval and, a given number k of steps in its continued fraction expansion, we wish to describe the evolution of the main parameters, notably $\gcd(q_{k-1}, q_k)$, associated with this expansion, when the depth k tends to ∞ . We wish to prove that generic reals have behavior akin to that of rationals p/q from the previous model on the rationals.

BIBLIOGRAPHY

- [AA07] Boris Adamczewski and Jean-Paul Allouche. Reversals and palindromes in continued fractions. *Theoretical Computer Science*, 380(3):220 – 237, 2007. Combinatorics on Words.
- [AB98] Pascal Alessandri and Valerie Berthe. Three distance theorems and combinatorics on words. 44, 01 1998.
- [All98] Cyril Allauzen. Une caractérisation simple des nombres de Sturm. *J. Théor. Nombres Bordeaux*, 10(2):237–241, 1998.
- [Apo74] Tom M. Apostol. *Mathematical analysis*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., second edition, 1974.
- [Apo98] T.M. Apostol. *Introduction to Analytic Number Theory*. Undergraduate Texts in Mathematics. Springer New York, 1998.
- [AS17] P. Arnoux and T. A. Schmidt. Natural extensions and Gauss measures for piecewise homographic continued fractions. *ArXiv e-prints*, September 2017.
- [Bal00] V. Baladi. *Positive Transfer Operators And Decay Of Correlations*. Advanced Series In Non-linear Dynamics. World Scientific Publishing Company, 2000.
- [BCLM17] Jonathan M. Borwein, Neil J. Calkin, Scott B. Lindstrom, and Andrew Mattingly. Continued logarithms and associated continued fractions. *Exp. Math.*, 26(4):412–429, 2017.
- [BCR⁺15] Valérie Berthé, Eda Cesaratto, Pablo Rotondo, Brigitte Vallée, and Alfredo Viola. Recurrence function on sturmian words: A probabilistic study. In Giuseppe F Italiano, Giovanni Pighizzini, and Donald T. Sannella, editors, *Mathematical Foundations of Computer Science 2015*, pages 116–128, Berlin, Heidelberg, 2015. Springer Berlin Heidelberg.
- [BCZ01] Florin P. Boca, Cristian Cobeli, and Alexandru Zaharescu. A conjecture of R. R. Hall on Farey points. *J. Reine Angew. Math.*, 535:207–236, 2001.
- [BCZ03] Florin P. Boca, Cristian Cobeli, and Alexandru Zaharescu. On the distribution of the Farey sequence with odd denominators. *Michigan Math. J.*, 51(3):557–573, 2003.
- [BDV02] Jérémie Bourdon, Benoit Daireaux, and Brigitte Vallée. Dynamical analysis of α -Euclidean algorithms. *J. Algorithms*, 44(1):246–285, 2002. Analysis of algorithms.
- [Ber96] Valérie Berthé. Fréquences des facteurs des suites sturmiennes. *Theoret. Comput. Sci.*, 165(2):295–309, 1996.
- [BHL17] Jonathan M. Borwein, Kevin G. Hare, and Jason G. Lynch. Generalized continued logarithms

- and related continued fractions. *J. Integer Seq.*, 20(5):Art. 17.5.7, 51, 2017.
- [Bil65] P. Billingsley. *Ergodic Theory and Information*. Wiley series in probability and mathematical statistics. J. Wiley, 1965.
- [BJW83] W. Bosma, H. Jager, and F. Wiedijk. Some metrical observations on the approximation by continued fractions. *Nederl. Akad. Wetensch. Indag. Math.*, 45(3):281–299, 1983.
- [BLRS08] Jean Berstel, Aaron Lauve, Christophe Reutenauer, and Franco Saliola. Combinatorics on words: Christoffel words and repetitions in words, 12 2008.
- [BLV16] Valérie Berthé, Loïck Lhote, and Brigitte Vallée. Probabilistic analyses of the plain multiple gcd algorithm. *J. Symb. Comput.*, 74(C):425–474, May 2016.
- [Bou07] N. Bourbaki. *Variétés différentielles et analytiques: Fascicule de résultats*. Bourbaki, Nicolas. Springer Berlin Heidelberg, 2007.
- [BV03] B. Baladi and B. Vallée. Systèmes dynamiques et algorithmique. Lecture Notes ALEA 2002, written in collaboration with Frédéric Chazal and Véronique Maume-Deschamps, Comptes-Rendus du Seminaire ALGO, Rapport de Recherches INRIA 5003, pp 121-150., November 2003.
- [BV05] Viviane Baladi and Brigitte Vallée. Euclidean algorithms are Gaussian. *J. Number Theory*, 110(2):331–386, 2005.
- [Cas99] Julien Cassaigne. Limit values of the recurrence quotient of sturmian sequences. *Theoretical Computer Science*, 218(1):3 – 12, 1999.
- [Cas01] Julien Cassaigne. Recurrence in infinite words. In Afonso Ferreira and Horst Reichel, editors, *STACS 2001*, pages 1–11, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg.
- [Cha05] Hei-Chi Chan. The asymptotic growth rate of random Fibonacci type sequences. *Fibonacci Quart.*, 43(3):243–255, 2005.
- [Cha06] Hei-Chi Chan. The asymptotic growth rate of random Fibonacci type sequences. II. *Fibonacci Quart.*, 44(1):73–84, 2006.
- [CLRS09] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to algorithms*. MIT Press, Cambridge, MA, third edition, 2009.
- [CT06] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2006.
- [CV17] Eda Cesaratto and Brigitte Valle. Gaussian behavior of quadratic irrationals. 07 2017.
- [Del54] Hubert Delange. Généralisation du théorème de Ikehara. *Ann. Sci. Ecole Norm. Sup. (3)*, 71:213–242, 1954.
- [DMDV05] Benoît Daireaux, Véronique Maume-Deschamps, and Brigitte Vallée. The Lyapunov tortoise and the dyadic hare. In *2005 International Conference on Analysis of Algorithms*, Discrete Math. Theor. Comput. Sci. Proc., AD, pages 71–94. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2005.
- [DPV08] Sanjoy Dasgupta, Christos H. Papadimitriou, and Umesh V. Vazirani. *Algorithms*. McGraw-Hill, 2008.
- [Eva06] Steven N. Evans. The expected number of zeros of a random system of p -adic polynomials. *Electron. Comm. Probab.*, 11:278–290, 2006.
- [EW11] Manfred Einsiedler and Thomas Ward. *Ergodic theory with a view towards number theory*, volume 259 of *Graduate Texts in Mathematics*. Springer-Verlag London, Ltd., London, 2011.
- [FB09] Eberhard Freitag and Rolf Busam. *Complex analysis*. Universitext. Springer-Verlag, Berlin,

- second edition, 2009.
- [Fek23] M. Fekete. über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten. *Math. Z.*, 17(1):228–249, 1923.
- [FGD⁺95] Philippe Flajolet, Xavier Gourdon, Philippe Dumas, Dedicated To Don Knuth, N. G. De Bruijn, and Hjalmar Mellin. Mellin transforms and asymptotics: Harmonic sums. *Theoretical Computer Science*, 144:3–58, 1995.
- [FN59] Bernard Friedman and Ivan Niven. The average first recurrence time. *Trans. Amer. Math. Soc.*, 92:25–34, 1959.
- [Fog02] N. Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.
- [Fol99] Gerald B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.
- [FS09] Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.
- [FV98] Philippe Flajolet and Brigitte Vallée. Continued fraction algorithms, functional operators, and structure constants. *Theoretical Computer Science*, 194(1):1 – 34, 1998.
- [Gos78] Bill Gosper. Continued fraction arithmetic. Unpublished manuscript, ca 1978.
- [Hen93] H Hennion. Sur un thorme spectral et son application aux noyaux lipchitziens. *Proceedings of the American Mathematical Society*, 118:627–634, 06 1993.
- [HW08] G. H. Hardy and E. M. Wright. *An introduction to the theory of numbers*. Oxford University Press, Oxford, sixth edition, 2008. Revised by D. R. Heath-Brown and J. H. Silverman, With a foreword by Andrew Wiles.
- [IK02] Marius Iosifescu and Cor Kraaikamp. *Metrical theory of continued fractions*, volume 547 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, 2002.
- [ITM50] C. T. Ionescu Tulcea and G. Marinescu. Théorie ergodique pour des classes d’opérations non complètement continues. *Ann. of Math. (2)*, 52:140–147, 1950.
- [Kat95] Tosio Kato. *Perturbation theory for linear operators*. Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the 1980 edition.
- [Kes62] Harry Kesten. Some probabilistic theorems on Diophantine approximations. *Trans. Amer. Math. Soc.*, 103:189–217, 1962.
- [Khi97] A. Ya. Khinchin. *Continued fractions*. Dover Publications, Inc., Mineola, NY, russian edition, 1997. With a preface by B. V. Gnedenko, Reprint of the 1964 translation.
- [KMS16] Marc Kesseböhmer, Sara Munday, and Bernd Otto Stratmann. *Infinite ergodic theory of numbers*. De Gruyter Graduate. De Gruyter, Berlin, 2016.
- [KN74] L. Kuipers and H. Niederreiter. *Uniform distribution of sequences*. Wiley-Interscience [John Wiley & Sons], New York-London-Sydney, 1974. Pure and Applied Mathematics.
- [Knu84] Donald E. Knuth. The distribution of continued fraction approximations. *Journal of Number Theory*, 19(3):443 – 448, 1984.
- [Kob84] Neal Koblitz. *p-adic numbers, p-adic analysis, and zeta-functions*, volume 58 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1984.
- [L36] Paul Lévy. Sur le développement en fraction continue d’un nombre choisi au hasard. *Compo-*

- sitio Math.*, 3:286–303, 1936.
- [Lot02] M. Lothaire. *Algebraic combinatorics on words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2002. A collective work by Jean Berstel, Dominique Perrin, Patrice Seebold, Julien Cassaigne, Aldo De Luca, Steffano Varricchio, Alain Lascoux, Bernard Leclerc, Jean-Yves Thibon, Veronique Bruyere, Christiane Frougny, Filippo Mignosi, Antonio Restivo, Christophe Reutenauer, Dominique Foata, Guo-Niu Han, Jacques Desarmenien, Volker Diekert, Tero Harju, Juhani Karhumaki and Wojciech Plandowski, With a preface by Berstel and Perrin.
- [MH40] Marston Morse and Gustav A. Hedlund. Symbolic dynamics II. Sturmian trajectories. *Amer. J. Math.*, 62:1–42, 1940.
- [Mun00] J.R. Munkres. *Topology*. Featured Titles for Topology Series. Prentice Hall, Incorporated, 2000.
- [Nus70] Roger D. Nussbaum. The radius of the essential spectrum. *Duke Math. J.*, 37:473–478, 1970.
- [PSS12] Carine Pivoteau, Bruno Salvy, and Michèle Soria. Algorithms for combinatorial structures: well-founded systems and Newton iterations. *J. Combin. Theory Ser. A*, 119(8):1711–1773, 2012.
- [PY98] Mark Pollicott and Michiko Yuri. *Dynamical systems and ergodic theory*, volume 40 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1998.
- [Rud87] Walter Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.
- [RV99] Dinakar Ramakrishnan and Robert J. Valenza. *Fourier analysis on number fields*, volume 186 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1999.
- [RV17] Pablo Rotondo and Brigitte Vallée. The recurrence function of a random Sturmian word. In *2017 Proceedings of the Fourteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, pages 100–114. SIAM, Philadelphia, PA, 2017.
- [RVV18] Pablo Rotondo, Brigitte Vallée, and Alfredo Viola. Analysis of the continued logarithm algorithm. In Michael A. Bender, Martin Farach-Colton, and Miguel A. Mosteiro, editors, *LATIN 2018: Theoretical Informatics - 13th Latin American Symposium, Buenos Aires, Argentina, April 16-19, 2018, Proceedings*, volume 10807 of *Lecture Notes in Computer Science*, pages 849–863. Springer, 2018.
- [Sar12] Omri Sarig. Introduction to the transfer operator method. Lecture Notes of the Second Brazilian School on Dynamical Systems, October 2012.
- [Sch84] W. H. Schikhof. *Ultrametric calculus*, volume 4 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1984. An introduction to p -adic analysis.
- [Sha16] J. Shallit. Length of the continued logarithm algorithm on rational inputs. *ArXiv e-prints*, June 2016.
- [Shp12] Igor E. Shparlinski. Modular hyperbolas. *Japanese Journal of Mathematics*, 7(2):235–294, Nov 2012.
- [Sós58] Vera T Sós. On the distribution mod 1 of the sequence $n\alpha$. *Ann. Univ. Sci. Budapest, Eötvös Sect. Math.*, 1:127–134, 1958.
- [Sta97] Richard P. Stanley. *Enumerative combinatorics. Vol. 1*, volume 49 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1997. With a foreword by Gian-Carlo Rota, Corrected reprint of the 1986 original.
- [SU08] Yakov G. Sinai and Corinna Ulcigrai. Renewal-type limit theorem for the Gauss map and

- continued fractions. *Ergodic Theory Dynam. Systems*, 28(2):643–655, 2008.
- [Ten15] Gérald Tenenbaum. *Introduction to analytic and probabilistic number theory*, volume 163 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, third edition, 2015. Translated from the 2008 French edition by Patrick D. F. Ion.
- [Ust09] A. V. Ustinov. On the statistical properties of elements of continued fractions. *Dokl. Akad. Nauk*, 424(4):459–461, 2009.
- [Val98a] B. Vallée. Dynamics of the binary euclidean algorithm: Functional analysis and operators. *Algorithmica*, 22(4):660–685, Dec 1998.
- [Val98b] Brigitte Vallée. Dynamique des fractions continues contraintes périodiques. *Journal of Number Theory*, 72(2):183 – 235, 1998.
- [Val01] Brigitte Vallée. Dynamical sources in information theory: fundamental intervals and word prefixes. *Algorithmica*, 29(1-2):262–306, 2001. Average-case analysis of algorithms (Princeton, NJ, 1998).
- [Val06] Brigitte Vallée. Euclidean dynamics. *Discrete Contin. Dyn. Syst.*, 15(1):281–352, 2006.
- [Val08] Brigitte Vallée. Private communication between brigitte vallée and joachim von zur gathen., 2008.
- [Wil06] Herbert S. Wilf. *generatingfunctionology*. A K Peters, Ltd., Wellesley, MA, third edition, 2006.
- [WW96] E. T. Whittaker and G. N. Watson. *A course of modern analysis*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1996. An introduction to the general theory of infinite processes and of analytic functions; with an account of the principal transcendental functions, Reprint of the fourth (1927) edition.
- [Yos95] Kōsaku Yosida. *Functional analysis*. Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the sixth (1980) edition.