

Diophantine approximations, Lagrange and Markov spectra, and dynamical Cantor sets

Carlos Matheus *

Carlos Gustavo Moreira †

December 31, 2020

1 Diophantine approximations

The seminal works of Diophantus of Alexandria (circa AD 250) on rational approximations to the solutions of certain algebraic equations began the important subfield of Number Theory called *Diophantine approximations*. Among the basic problems in this topic, one has the question of finding rational numbers $p/q \in \mathbb{Q}$ approximating a given real number $\alpha \in \mathbb{R}$ in such a way that the denominator q is not “big” and the error is $|\alpha - p/q|$ is “small”.

1.1 Rational approximations of π

The first few decimal digits of the number π are well-known: $\pi = 3.1415926\dots$. By definition, this provides some rational approximations of π like $314/100$ and $3141592/10^6$. Nonetheless, these fractions are certainly *not* answers to the Diophantine problem posed above because we can get *better* approximations with *smaller* denominators: for instance, Archimedes (circa 250 BC) knew that

$$\left| \pi - \frac{22}{7} \right| < \frac{1}{700} < \left| \pi - \frac{314}{100} \right|$$

and it is possible to check that

$$\left| \pi - \frac{355}{113} \right| < \frac{1}{3 \cdot 10^6} < \left| \pi - \frac{3141592}{10^6} \right|.$$

*The first author is a CNRS researcher affiliated to École Polytechnique (France). His email address is carlos.matheus@math.cnrs.fr.

†The second author is a full professor at IMPA (Brazil). His email address is gugu@impa.br.

1.2 Dirichlet’s pigeonhole principle

The example of the number π makes us wonder how small $|\alpha - p/q|$ can be when the denominator q varies in a fixed range $1 \leq q \leq Q$. A preliminary answer comes from the following elementary remark. Recall that any real number α lies between two consecutive integers, namely $[\alpha] \leq \alpha < [\alpha] + 1$ where $[\alpha] \in \mathbb{Z}$ is the *integer part* of α . Therefore, given $\alpha \in \mathbb{R}$ and $q \in \mathbb{N}$, we can find $p \in \mathbb{Z}$ such that $|q\alpha - p| \leq 1/2$, i.e.,

$$\left| \alpha - \frac{p}{q} \right| < \frac{1}{2q}.$$

In 1841, Dirichlet used his famous *pigeonhole principle* to significantly improve upon the elementary statement in the previous paragraph: more concretely, for any irrational number α , one has that

$$\# \left\{ \frac{p}{q} \in \mathbb{Q} : \left| \alpha - \frac{p}{q} \right| < \frac{1}{q^2} \right\} = \infty.$$

Indeed, given $Q \in \mathbb{N}$, Dirichlet considered how the $Q+1$ numbers $\{i\alpha\} := i\alpha - [i\alpha] \in [0, 1)$, $i = 0, \dots, Q$, are distributed across the elements of the partition $[0, 1) = \bigcup_{n=1}^Q [(n-1)/Q, n/Q)$ into Q intervals. By the pigeonhole principle, two fractional parts, say $\{i\alpha\}, \{j\alpha\}$, $0 \leq i < j \leq Q$, must lie in the same interval, say $[(n-1)/Q, n/Q)$, so that $|\{j\alpha\} - \{i\alpha\}| < 1/Q$ and, *a fortiori*, there exists $p \in \mathbb{Z}$ such that

$$\left| \alpha - \frac{p}{j-i} \right| < \frac{1}{(j-i)Q} \leq \frac{1}{(j-i)^2}.$$

1.3 Hurwitz theorem

In 1891, Hurwitz proved¹ that Dirichlet's theorem is essentially optimal as far as *all* irrational numbers are concerned: one has

$$\# \left\{ \frac{p}{q} \in \mathbb{Q} : \left| \alpha - \frac{p}{q} \right| < \frac{1}{\sqrt{5} \cdot q^2} \right\} = \infty$$

for all irrational number α , and²

$$\# \left\{ \frac{p}{q} \in \mathbb{Q} : \left| \frac{1 + \sqrt{5}}{2} - \frac{p}{q} \right| < \frac{1}{(\sqrt{5} + \varepsilon) \cdot q^2} \right\} < \infty$$

for all $\varepsilon > 0$.

2 Classical spectra

Despite the almost optimality of Dirichlet theorem, we can ask whether it can be improved for *individual* irrational numbers α by inquiring about the nature of the *best* constant $\ell(\alpha)$ among the quantities c such that

$$\# \left\{ \frac{p}{q} \in \mathbb{Q} : \left| \alpha - \frac{p}{q} \right| < \frac{1}{c \cdot q^2} \right\} = \infty,$$

i.e., $\ell(\alpha) := \limsup_{p,q \rightarrow \infty} \frac{1}{q^2 |\alpha - p/q|} = \limsup_{p,q \rightarrow \infty} \frac{1}{|q(q\alpha - p)|}$.

The *Lagrange spectrum* L is the collection of *finite*³ best constants of Diophantine approximation, i.e.,

$$L := \{ \ell(\alpha) < \infty : \alpha \in \mathbb{R} \setminus \mathbb{Q} \}.$$

In this setting, Hurwitz theorem says that the minimum of L is $\min L = \sqrt{5}$.

¹Actually, this result was first established by Korkine and Zolotarev in 1873.

²Let $\phi := (1 + \sqrt{5})/2$. If $p/q \in \mathbb{Q}$ and $|q\phi - p| \leq 1$, then $1 \leq |q^2 + pq - p^2| = |q\phi - p| \cdot |q\phi^{-1} + p|$ and $|q\phi^{-1} + p| = |q(\phi^{-1} + \phi) - q\phi + p| \leq q\sqrt{5} + 1$. Thus, $\frac{1}{|q(q\phi - p)|} \leq \sqrt{5} + \frac{1}{q}$.

³It is possible to show that $\ell(\alpha) = \infty$ for Lebesgue almost every α . Hence, the Lagrange spectrum tries to encode Diophantine properties of irrational numbers beyond the probabilistic dominant regime.

2.1 Beginning of the classical spectra

The Lagrange spectrum was systematically studied⁴ in connection with the theory of binary quadratic forms by Markov in 1879. In fact, the quantity $q(q\alpha - p)$ is the value of the binary quadratic form $h_\alpha(x, y) = \alpha y^2 - xy$ at the integral point $(p, q) \in \mathbb{Z}^2$, so that the Lagrange spectrum is somewhat related to the *Markov spectrum* M of *finite* best constants

$$m(h) := \sup_{(p,q) \in \mathbb{Z}^2 \setminus \{(0,0)\}} \frac{\sqrt{\Delta(h)}}{|h(p,q)|}$$

of Diophantine approximations of real, indefinite, binary⁵ quadratic forms $h(x, y) = ax^2 + bxy + cy^2$ with positive discriminant $\Delta(h) = b^2 - 4ac > 0$. In this context, Markov proved that

$$L \cap [\sqrt{5}, 3) = M \cap [\sqrt{5}, 3) = \left\{ \sqrt{9 - \frac{4}{z_n^2}} : n \in \mathbb{N} \right\},$$

where z_n is a *Markov number*, i.e., the largest coordinate of a solution $(x_n, y_n, z_n) \in \mathbb{N}^3$ of the Markov–Hurwitz equation

$$x_n^2 + y_n^2 + z_n^2 = 3x_n y_n z_n.$$

2.1.1 Fermat's descent on Markov's cubic

The Markov–Hurwitz equation determines a cubic surface S whose integral points are called *Markov triples*. Since the Markov–Hurwitz equation is quadratic on a given variable (when we freeze the other two variables), the cubic surface S has a rich group of automorphisms: besides permuting the coordinates, we can replace (x, y, z) by $(3yz - x, y, z)$, $(x, 3xz - y, z)$ or $(x, y, 3xy - z)$ without leaving S . The last three automorphisms are called *Vieta involutions* and they were used by Markov to produce a *descent argument* showing that *any* Markov triple

⁴This work was part of Markov's M.Sc. dissertation (directly motivated by the results of Korkine and Zolotarev), while Markov's PhD thesis focused on the theory of Markov chains in Probability Theory.

⁵The theory of n -ary quadratic forms is also very interesting, but we will skip it here because the dynamical techniques involved in Margulis' solution of Oppenheim conjecture are dramatically different from the methods appearing in this text.

$(x, y, z) \in N^3$ can be obtained from the fundamental solution $(1, 1, 1)$ after applying a sequence of permutations of coordinates and Vieta involutions.

In fact, it is not hard to see that a Markov triple $(x, y, z) \in N^3$ with $x \leq y \leq z$ falls into two categories: either $x = y$ or $x < y < z$. In the first case, $z^2 = (3z - 2)x^2$, so that $x = y$ divides z , say $z = nx$ with $n^2 = 3nx - 2$; this means that n divides 2, and, thus, $n = 1$ and $(x, y, z) = (1, 1, 1)$ or $n = 2$ and $(x, y, z) = (1, 1, 2)$. In the second case, we apply Vieta involution $(x, y, z) \mapsto (x, y, z')$ with $z' = 3xy - z$ and we study the quantity $a := (y - z)(y - z')$; since $z + z' = 3xy$ and $zz' = x^2 + y^2$, we have

$$a = 2y^2 + x^2 - 3xy^2 = (2y^2 - 2xy^2) + (x^2 - xy^2) < 0$$

as $2y^2 \leq 2xy^2$ and $x^2 < xy^2$ thanks to our assumption $1 \leq x < y$; because $y - z < 0$ (by hypothesis), $a < 0$ means that $z' < y < z$.

In summary, given a Markov triple $(x, y, z) \in N^3$ with $x \leq y \leq z$, we saw that either $x = y$ and $(x, y, z) \in \{(1, 1, 1), (1, 1, 2)\}$, or $x < y < z$ and we can apply a Vieta involution to convert (x, y, z) into a Markov triple (x, y, z') with $x, z' < y < z$. By repeating this argument finitely many times, we see that a sequence of Vieta involutions and permutations of coordinates allows to convert the Markov triple (x, y, z) into $(1, 1, 1)$, as desired.

2.1.2 Markov tree

The descent argument above permits to organize all ordered Markov triples $(x, y, z) \in N^3$, $x \leq y \leq z$, into the so-called *Markov tree* whose branches connect ordered Markov triples deduced from each other by a Vieta involution (up to permutation of coordinates).

markoffCIRM.pdf

The knowledge of Markov's tree permits to write down the first few elements of $M \cap [\sqrt{5}, 3)$: since the first few Markov triples are $(1, 1, 1)$, $(1, 1, 2)$, $(1, 2, 5)$, $(1, 5, 13)$ and $(2, 5, 29)$, we have that the first few

Markov numbers⁶ are 1, 2, 5, 13 and 29, so that

$$M \cap [\sqrt{5}, 3) = \left\{ \sqrt{5}, \sqrt{8}, \frac{\sqrt{221}}{5}, \frac{\sqrt{1517}}{13}, \frac{\sqrt{7565}}{29}, \dots \right\}$$

2.1.3 Beyond Markov tree ...

The Markov tree and numbers are fascinating objects. For instance, it was conjectured by Frobenius in 1913 that Markov triples $(x, y, z) \in N^3$, $x \leq y \leq z$ are actually determined by the Markov number z (cf. Bombieri's survey article [?MR2345177]).

Also, Zagier [?MR669663] showed that the number $M(x)$ of Markov number below x is

$$M(x) = c(\log x)^2 + O(\log x(\log \log x)^2)$$

where $c = 0.18071704711507\dots$ is an *explicit* constant⁷, and, more recently, Baragar [?MR1295950] and Gamburd–Magee–Ronan [?MR4024562] studied the general problem of counting integral points on the Markov–Hurwitz varieties of the form

$$x_1^2 + \dots + x_n^2 = ax_1 \dots x_n + k$$

where $n \geq 3$, $a \geq 1$ and k are integers.

Moreover, the Markov triples are related to lengths of simple closed geodesics on a certain *hyperbolic* once-punctured torus: in fact, the commutator subgroup Γ of $SL(2, Z)$ is an index 12 subgroup generated by $A_0 = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$ and $B_0 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$; the quotient $SL(2, R)/\Gamma$ is the unit cotangent bundle of a hyperbolic once-punctured torus whose simple closed geodesics correspond to the elements $A \in \Gamma$ in a pair (A, B) of generators of Γ ; the hyperbolic lengths of these geodesics have the form $2 \cosh^{-1}(tr(A)/2)$, so that they are related to Markov triples because Fricke proved that any generating pair (A, B) of Γ satisfies

$$tr(A)^2 + tr(B)^2 + tr(AB)^2 = tr(A)tr(B)tr(AB),$$

⁶The attentive reader certainly noticed that some of these numbers are part of Fibonacci's sequence $(F_n)_{n \in N}$ and this is not a coincidence: it is possible to check that $(1, F_{2m-1}, F_{2m+1})$ is a Markov triple for all $m \in N$.

⁷Furthermore, Zagier conjectured that $M(x) = c(\log(3x))^2 + o(\log x)$, so that n -th Markov number z_n is $z_n \sim \frac{1}{3}A\sqrt{n}$ where $A = e^{1/\sqrt{c}} = 10.5101504\dots$

i.e., $(tr(A)/3, tr(B)/3, tr(AB)/3)$ is a Markov triple.

Furthermore, it is known (cf. [?MR2026539]) that the level sets of the function $\kappa(x, y, z) = x^2 + y^2 + z^2 - xyz - 2$ parametrize the elements of the $SL(2, R)$ -character variety⁸ of once-punctured torii and each Markov triple (x, y, z) produces an integral point $(3x, 3y, 3z)$ of $\kappa^{-1}(-2)$.

Finally, Bourgain–Gamburd–Sarnak [?MR3456887] investigated the family of graphs (\mathcal{G}_p) (indexed by the set of prime numbers p) obtained by applying Vieta involutions and permutation of coordinates to the solutions in $F_p^3 \setminus \{(0, 0, 0)\}$ to Markov–Hurwitz equation $x^2 + y^2 + z^2 = 3xyz$. In this setting, they showed \mathcal{G}_p has a giant component \mathcal{C}_p in the sense that $\#(\mathcal{G}_p \setminus \mathcal{C}_p) = O_\varepsilon(p^\varepsilon)$ for all $\varepsilon > 0$, and they used the technology involved in the proof of this statement to establish that almost all Markov numbers are composite, i.e.,

$$\frac{\#\{p \text{ prime Markov number} \leq T\}}{\#\{z \text{ Markov number} \leq T\}} \rightarrow 0$$

as $T \rightarrow \infty$. Also, they conjectured that the graphs \mathcal{G}_p are connected⁹ and they form an expander family¹⁰.

2.2 Continued fractions

The definition of the Lagrange spectrum suggests that we can study L provided there is a method to find the best rational approximations of a given irrational number α (such as $22/7$ and $355/113$ for π).

As it turns out, one can guess the best rational approximations for α out of its *continued fraction expansion*. More precisely, given an irrational number $\alpha_0 = \alpha$, let $a_0 = \lfloor \alpha \rfloor$, so that $\alpha_0 - a_0 \in (0, 1)$. We define recursively $\alpha_1 = \frac{1}{\alpha_0 - a_0}$ and $a_n = \lfloor \alpha_n \rfloor \in N^*$, $\alpha_{n+1} = \frac{1}{\alpha_n - a_n}$ for all $n \in N^*$. In this context, we

⁸Naively speaking, the G -character variety of a topological surface $S_{g,n}$ of genus g with n punctures is the set of equivalence classes of representations $\rho : \pi_1(S_{g,n}) \rightarrow G$ modulo the natural action of G by conjugation.

⁹The connectedness of \mathcal{G}_p for all large p was very recently established by Chen: cf. the arxiv preprint arXiv:2011.12940.

¹⁰That is, there is an *uniform* spectral gap for the adjacency matrices of these graphs.

say that α has continued fraction expansion

$$\alpha = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\ddots}}}$$

and we denote by

$$\frac{p_n}{q_n} = a_0 + \frac{1}{a_1 + \frac{1}{\ddots + \frac{1}{a_n}}}$$

the *convergents* of α . For example, $\pi = [3; 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, 1, 14, 2, 1, \dots]$, so that $\frac{p_1}{q_1} = 22/7$ and $\frac{p_3}{q_3} = 355/113$.

It is possible to prove that (p_n/q_n) provides the best rational approximations for α because

- if $\frac{p}{q} \in Q$ and $|\alpha - \frac{p}{q}| < \frac{1}{2q^2}$, then $\frac{p}{q} = \frac{p_n}{q_n}$ for some $n \in N$;
- for each $m \in N$, $|\alpha - \frac{p_m}{q_m}| < 1/q_m^2$ and $\min\{q_m|q_m\alpha - p_m|, q_{m+1}|q_{m+1}\alpha - p_{m+1}|\} < 1/2$;
- for each $m \in N$, $|\alpha - p/q| > |\alpha - p_m/q_m|$ for all $p/q \neq p_m/q_m$ with $q < q_{m+1}$.

In particular, the best constant $\ell(\alpha) = \limsup_{p,q \rightarrow \infty} \frac{1}{|q(q\alpha - p)|}$ of Diophantine approximation for α depends only on its convergents, i.e.,

$$\ell(\alpha) = \limsup_{n \rightarrow \infty} \frac{1}{|q_n(q_n\alpha - p_n)|}$$

2.3 Perron’s definition of the spectra

The basic formula

$$\frac{1}{q_n(q_n\alpha - p_n)} = (-1)^n(\alpha_{n+1} + \beta_{n+1})$$

where $\beta_{n+1} := q_{n-1}/q_n = [0; a_n, \dots, a_1]$ led Perron to propose in 1921 the following *dynamical* interpretation of L . Let $\Sigma = (N^*)^Z$ be the (non-compact) symbolic space of bi-infinite sequences of non-zero natural numbers. The shift map $\sigma : \Sigma \rightarrow \Sigma$ is the dynamical system given by

$$\sigma((a_n)_{n \in Z}) = (a_{n+1})_{n \in Z}.$$

In this language, the Lagrange spectrum is the set of finite *asymptotic* records of heights of the orbits of σ with respect to the (proper) height function $f : \Sigma \rightarrow R$, $f((a_n)_{n \in \mathbb{Z}}) := [a_0; a_1, a_2, \dots] + [0; a_{-1}, a_{-2}, \dots]$, i.e.,

$$L = \left\{ \limsup_{n \rightarrow \infty} f(\sigma^n(x)) < \infty : x \in \Sigma \right\}.$$

Interestingly enough, one can use the classical reduction theory of binary quadratic forms (due to Lagrange and Gauss) to prove that the Markov spectrum is the set of finite *absolute* records of heights of the orbits of σ with respect to f , i.e.,

$$M = \left\{ \sup_{n \in \mathbb{Z}} f(\sigma^n(x)) < \infty : x \in \Sigma \right\}.$$

From these dynamical characterizations of L and M , Perron deduced that

- $\sup_{n \in \mathbb{Z}} f(\sigma^n(x)) \leq \sqrt{12}$ if and only if $x \in \{1, 2\}^{\mathbb{Z}}$;
- $\sqrt{12}, \sqrt{13}, \frac{9\sqrt{3}+65}{22} \in L$;
- $M \cap (\sqrt{12}, \sqrt{13}) = M \cap (\sqrt{13}, \frac{9\sqrt{3}+65}{22}) = \emptyset$.

Moreover, one can use¹¹ this dynamical point of view to prove that

$$L = \overline{\left\{ \sup_{n \rightarrow \infty} f(\sigma^n(y)) : y \in \Sigma \text{ is periodic} \right\}}$$

and

$$M = \overline{\left\{ \sup_{n \in \mathbb{Z}} f(\sigma^n(z)) : z \in \Sigma \text{ is eventually periodic} \right\}}.$$

Thus, $L \subset M$ are *closed* subsets of the real line.

2.4 Dynamics on the modular surface

The shift map $\sigma : \Sigma \rightarrow \Sigma$ can be thought as an invertible map extending the *Gauss map* $G : (0, 1] \rightarrow [0, 1)$, $G(x) = \{1/x\}$. Indeed, the definitions imply that the Gauss map acts on continued fraction expansions by

¹¹In the sequel, *eventually periodic* means eventually periodic on *both* sides (perhaps with different periods).

left-shift on *half-infinite* sequences of natural numbers:

$$G([0; a_1, a_2, \dots]) = [0; a_2, \dots]. \quad (1)$$

Using the well-known link (due to Artin, Cohn, Series, Arnoux, ...) between the Gauss map and the geodesic flow g_t on the unit cotangent bundle $SL(2, R)/SL(2, Z)$ to the modular surface (cf. [MR1279059]), one can also describe the Lagrange spectrum as the set of finite asymptotic records of the heights of the orbits of a continuous-time, smooth dynamical system, namely,

$$L = \left\{ \limsup_{t \rightarrow \infty} H(g_t(x)) < \infty : x \in SL(2, R)/SL(2, Z) \right\}$$

where $H : SL(2, R)/SL(2, Z) \rightarrow R$ is a certain (proper) function¹².

2.5 Ending of the classical spectra

The expression of the height function $f : \Sigma \rightarrow R$ in Perron's definition of the spectra suggests that L and M are related to *arithmetic sums* of Cantor sets of real numbers whose continued fraction expansions have restricted digits.

In other terms, the study of *projections* of products of certain Cantor sets under the function $\pi : R^2 \rightarrow R$, $\pi(x, y) := x + y$, should provide some insights into the fine structures of L and M .

This idea was explored by Hall in 1947 to show that L contains the half-line $[6, \infty)$. For this sake, Hall considered the Cantor set $C(4) = \{[0; a_1, a_2, \dots] : 1 \leq a_i \leq 4 \forall i \in \mathbb{N}\}$ and he established that

$$\begin{aligned} C(4) + C(4) &:= \{x + y : (x, y) \in C(4) \times C(4)\} \\ &= [\sqrt{2} - 1, 4(\sqrt{2} - 1)] \end{aligned}$$

is an interval of length > 1 . This fact implies that given $\ell \in [6, \infty)$, one can find $c_0 \in \mathbb{N}$ such that $5 \leq c_0 \leq \ell$ and $\ell - c_0 \in C(4) + C(4)$, say

$$\ell = c_0 + [0; a_1, a_2, \dots] + [0; b_1, b_2, \dots]$$

¹²By thinking of $SL(2, R)/SL(2, Z)$ as the space of unimodular lattices in R^2 , one has $H(x) = 2/\text{sys}(x)^2$, where $\text{sys}(x)$ is the systole of $x \simeq g(Z^2)$, $g \in SL(2, R)$.

with $1 \leq a_i, b_i \leq 4$ for all $i \in N$. Thus, the irrational number α with continued fraction expansion

$$\alpha = [0; \underbrace{b_1, c_0, a_1, \dots}_{1^{st} \text{ block}}, \underbrace{b_n, \dots, b_1, c_0, a_1, \dots, a_n, \dots}_{n^{th} \text{ block}}]$$

satisfies $\ell = \ell(\alpha) \in L$. Since $\ell \geq 6$ was arbitrary, we conclude that $L \supset [6, \infty)$.

The largest half-line of the form $[c_F, \infty)$ included in L is called *Hall ray*. In 1975, Freiman famously claimed that

$$c_F = 4 + \frac{253589820 + 283798\sqrt{462}}{491993569} = 4.527829566\dots$$

2.6 Intermediate portions of L and M

We saw above that $L \subset M$ are closed subsets of the real line such that $L \cap (-\infty, 3) = M \cap (-\infty, 3) = \{\sqrt{5}, \sqrt{8}, \dots\}$, $L \cap [c_F, \infty) = M \cap [c_F, \infty) = [c_F, \infty)$, and $\sqrt{12}, \sqrt{13} \in L$, but $(\sqrt{12}, \sqrt{13}) \cap L = (\sqrt{12}, \sqrt{13}) \cap M = \emptyset$.

In particular, L and M coincide in the portions $(-\infty, 3)$, $[\sqrt{12}, \sqrt{13}]$, and $[c_F, \infty)$. Nonetheless, it was discovered by Freiman in 1968 that $M \setminus L \neq \emptyset$: more concretely, Freiman found a *countable* subset of $M \setminus L$ located near 3.11.

Subsequently, Freiman discovered (in 1973) a new element of $M \setminus L$ near 3.29, and Flahive showed¹³ in 1977 that this element is the limit of an explicit sequence of elements of $M \setminus L$ near 3.29.

This “concentration” of examples of elements of $M \setminus L$ between 3 and 3.3 led Cusick to *conjecture* in 1975 that L and M should coincide above $\sqrt{12}$.

In any case, the previous paragraphs hint that the intermediate portions of the spectra (between 3 and c_F) might have a *complicated* structure in comparison to their beginning and ending.

In 1971, Hall gave an upper bound on the *fractal complexity* of $M \cap [\sqrt{5}, \sqrt{10}]$. More precisely, he used Perron’s definition of the spectra to establish that $M \cap [3, \sqrt{10}] \subset 2 + U + U$ where

$$U = \{[0; a_1, a_2, \dots] : (a_i a_{i+1} a_{i+2}) \neq (121) \forall i \in N^*\}.$$

¹³Actually, in his unpublished PhD thesis from 1976, Y.-C. You found an *uncountable* subset of $M \setminus L$ near 3.29 which is bi-Lipschitz homeomorphic to Cantor set of continued fraction expansions obtained by concatenations of the words 11 and 22.

After that, he analysed the sizes of all intervals covering U with extremities of the form $[0; a_1, \dots, a_n, \overline{2122}]$ and $[0; a_1, \dots, a_n, \overline{1222}]$ (where \overline{w} stands for the periodic sequence obtained by infinite concatenation of a string w) in order to show that the *Hausdorff dimension*¹⁴ of U is at most 0.465. From this estimate, it is possible to infer that the Hausdorff dimension of $U + U = \pi(U \times U)$ has Hausdorff dimension at most $2 \cdot \dim(U) < 0.93$ and, *a fortiori*, $M \cap [\sqrt{5}, \sqrt{10}] \subset 2 + U + U$ has zero Lebesgue measure.

On the other hand, it is believed that L and M should contain non-trivial intervals between t_1 and c_F : for instance, a folklore question (appearing in page 71 of Cusick–Flahive book [?MR1010419]) asks whether $L \cap [\sqrt{5}, \sqrt{12}]$ has non-empty interior, and Berstein conjectured in 1973 that $[4.1, 4.52] \subset L$. Here, it is worth to point out that the inspiration for the first (folklore) question comes from:

- Perron’s result that $M \cap [\sqrt{5}, \sqrt{12}]$ is closely related to the arithmetic sum $C(2) + C(2)$ where $C(2) := \{[0; a_1, a_2, \dots] : 1 \leq a_i \leq 2 \forall i\}$;
- the expectation¹⁵ that $C(2) + C(2)$ contains intervals because $C(2) + C(2)$ is the projection $\pi(C(2) \times C(2))$ of a planar “non-linear” Cantor set $C(2) \times C(2) \subset R^2$ with Hausdorff dimension $2 \cdot \dim(C(2)) > 1$.

Also, Berstein thinks that $[4.1, 4.52] \subset L$ because of Freiman’s work on the computation of the beginning c_F of Hall’s ray.

¹⁴Recall that the Hausdorff dimension $0 \leq \dim(X) \leq m$ of $X \subset R^m$ measures how hard it is to efficiently cover X :

$$\text{in fact, } \dim(X) = \inf \left\{ s > 0 : \inf_{X \subset \bigcup_{n \in N} B(x_n, r_n)} \sum_{n \in N} r_n^s = 0 \right\}.$$

Thus, a countable set has zero Hausdorff dimension and any set $X \subset R^m$ with $\dim(X) < m$ has zero m -dimensional Lebesgue measure.

¹⁵The typical projections of generic planar Cantor sets with Hausdorff dimension > 1 are expected to contain intervals thanks to the combination of Marstrand’s theorem asserting that the projections in almost all directions of a subset of R^2 with Hausdorff dimension > 1 have positive Lebesgue measures and the work of Yoccoz and the second author on the stable intersections of Cantor sets via renormalization techniques.

2.7 Recent results about $M \cap (3, c_F)$

Despite the strong belief (expressed by the conjectures and questions in the previous subsection) that $M \cap (3, c_F)$ must have an intricate structure, the first rigorous result in this direction was obtained only in 2018 by the second author [?MR3815461]. In fact, he showed that:

- for each $t \in R$, the Hausdorff dimension $d(t)$ of $L \cap (-\infty, t)$ coincides with the Hausdorff dimension of $M \cap (-\infty, t)$, i.e., $M \setminus L$ is *not* big enough to create jumps in dimension between L and M ;
- $d(t)$ is a *continuous*, non-Hölder function of t such that $d(3 + \varepsilon) > 0$ for all $\varepsilon > 0$ and $d(\sqrt{12}) = 1$.

The second item above implies that L and M necessarily contain complicated fractal sets. For instance, it is not difficult to check that if $K \subset R$ is a Cantor set defined by simple interactive (dynamical) rules like Cantor’s middle-third set¹⁶, then the function $t \mapsto \dim(K \cap (-\infty, t))$ is always a piecewise constant and discontinuous:

$$\dim(K \cap (-\infty, t)) = \begin{cases} 0 & \text{if } t \leq \min K \\ \dim(K) & \text{if } t > \min K \end{cases}$$

More recently, we investigated in [?MR3860483], [?MR3925086] and [?MR4152626] the fine structure of $M \setminus L$ and we proved that it is *richer* than conjectured by Cusick. More precisely, there are three open intervals I_1 , I_2 and I_3 near 3.11, 3.29 and 3.7 such that:

- the sizes of I_1, I_2, I_3 are $\sim 2 \cdot 10^{-10}, 2 \cdot 10^{-7}, 10^{-10}$;
- the extremities of I_j belong to L , but $L \cap I_j = \emptyset$ for each $1 \leq j \leq 3$;
- $(M \setminus L) \cap I_j$, $1 \leq j \leq 3$, are *closed* subsets with Hausdorff dimensions $> 0.26, 0.353, 0.531$ resp.;

¹⁶Recall that Cantor’s middle-third set is given by breaking $[0, 1]$ into three intervals of equal lengths, removing the subinterval in the middle, and then successively repeating this operation to each of the remaining subintervals.

- I_1 and I_2 contain the examples of elements of $M \setminus L$ previously found by Freiman and Flahive, and the elements of $(M \setminus L) \cap I_3$ provide a *negative* answer to Cusick’s conjecture that L and M should coincide above $\sqrt{12}$.

On the other hand, we proved that $M \setminus L$ is *not* very rich¹⁷ because $\dim(M \setminus L) < 0.987$.

Lagrange2.pdf

Besides the metric results discussed above, our techniques also give *topological* consequences for L and M : for example, the subset L' of non-isolated points of L is *perfect*, i.e., $L' = L''$, and the *interiors* of L and M coincide. However, we *ignore* whether M' is a perfect set, and, contrary to the initial impression given by the fact that $(M \setminus L) \cap I_j$, $1 \leq j \leq 3$, are closed subsets, Lima, Vieira and the authors proved that $M \setminus L$ is *not* closed¹⁸ by establishing that $1 + 3/\sqrt{2} \in L \cap \overline{(M \setminus L)}$.

In an attempt to further investigate interesting questions about the structure of $M \cap (3, c_F)$, Delecroix and the authors [?MR4109576] developed an algorithm providing $1/Q$ -approximations¹⁹ to L and M after a running time $O(Q^{2.367})$. This algorithm was implemented (on Sage) by Delecroix to produce the figure below of $L_2 := L \cap [\sqrt{5}, \sqrt{12}]$, but unfortunately, we could not use this algorithm yet to get definite ideas about Berstein’s conjecture that $[4.1, 4.52] \subset L$. Nevertheless, we hope that some variant of this algorithm will be helpful in the future because its running time is not very big in comparison²⁰ with the “naive” algorithm stemming from the characterisations of L and M via the closures of the

¹⁷We also offered some *heuristic* evidence towards an upper bound of the form $\dim(M \setminus L) < 0.888$ and, as it turns out, Pollicott–Vytona proved that $\dim(M \setminus L) < 0.8822195$ in a preprint available at arXiv:2012.07083.

¹⁸We also gave some *evidence* towards the possibility that $3 \in L \cap \overline{(M \setminus L)}$.

¹⁹Here, we mean close in *Hausdorff topology*, i.e., $A, B \subset R$ are δ -close if and only if for each $a \in A$ and $b \in B$ there are $c \in B$ and $d \in A$ such that $|a - c|, |b - d| \leq \delta$.

²⁰In fact, a back-of-the-envelope calculation reveals that the naive algorithm requires computations with continued fractions associated to $\sim 4Q^4$ strings of lengths $\leq Q^4$ to rigorously produce a $1/Q$ -approximation to L and M .

values of height records of periodic and eventually periodic elements of $\Sigma = (N^*)^{\mathbb{Z}}$.

L2.pdf

The common theme behind all results described in this subsection is the study of portions of L and M via the fractal geometry of certain types of *dynamically defined Cantor sets*. In the next section, we will briefly discuss these objects and we will make some comments about their applications to the proofs of our results on the Hausdorff dimensions of $L \cap (-\infty, t)$, $M \cap (-\infty, t)$ and $M \setminus L$.

3 Dynamical Cantor sets

A *dynamically defined* Cantor set $K \subset R$ is

$$K = \bigcap_{n \in \mathbb{N}} \psi^{-n}(I_1 \cup \dots \cup I_k),$$

where $\psi : I_1 \cup \dots \cup I_k \rightarrow I$ is a smooth map from a collection of disjoint compact intervals I_1, \dots, I_k onto the convex hull I of their union such that

- ψ is *expanding*: $|\psi'(x)| > 1 \forall x \in I_1 \cup \dots \cup I_k$;
- $\{I_1, \dots, I_k\}$ is a *Markov partition*: each $\psi(I_i)$ is the convex hull of the union of some of the intervals I_j ;
- ψ is *topologically mixing*: there exists $n_0 \in \mathbb{N}$ with $\psi^{n_0}(K \cap I_i) = K$ for all $1 \leq i \leq k$.

The most famous example of dynamically defined Cantor set is arguably Cantor's middle-third set

$$K_{1/3} = \left\{ \sum_{n=1}^{\infty} \frac{a_n}{3^n} : (a_n)_{n \in \mathbb{N}} \in \{0, 2\}^{\mathbb{N}} \right\}$$

obtained by dividing $[0, 1]$ into three intervals with equal lengths, removing the middle one $(1/3, 2/3)$, dividing the remaining intervals into three equal subintervals, removing the middle ones, repeating this procedure *ad infinitum*, and keeping only the points of $[0, 1]$ never falling into an excluded subinterval.

In fact, $K_{1/3} = \bigcap_{n \in \mathbb{N}} \phi^{-n}(I_1 \cup I_2)$ where $I_1 = [0, 1/3]$, $I_2 = [2/3, 1]$, and

$$\phi(x) = \begin{cases} 3x & \text{if } x \in I_1, \\ 3x - 2 & \text{if } x \in I_2. \end{cases}$$

The geometry of dynamical Cantor sets was studied in depth by several authors over the last 100 years. In particular, we dispose nowadays of several methods to compute the Hausdorff dimension of dynamical Cantor sets. For instance, given a dynamical Cantor set K generated by $\psi : I_1 \cup \dots \cup I_k \rightarrow I$, for each $m \in \mathbb{N}$, we can consider the collection \mathcal{R}_m of intervals consisting of the connected components of $\bigcap_{n=1}^m \psi^{-n}(I_1 \cup \dots \cup I_k)$. By definition, \mathcal{R}_m is a cover of K and it can be used to compute $\dim(K)$: if we denote by $\lambda(J) = \min |\psi'|_J$, $\Lambda(J) = \max |\psi'|_J$ and we fix a mixing time n_0 for ψ (i.e., $\psi^{n_0}(I_j \cap K) = K$ for each $1 \leq j \leq k$), then it is possible to prove that $\alpha_m \leq \dim(K) \leq \beta_m$, where

$$\sum_{J \in \mathcal{R}_m} \frac{1}{\lambda(J)^{\beta_m}} = 1 \text{ and } \sum_{J \in \mathcal{R}_m} \frac{1}{\Lambda(J)^{\alpha_m}} = \max |(\psi^{n_0-1})'|$$

(cf. pages 68 to 70 of Palis–Takens book [?MR1237641]). These bounds allow an *exact* calculation of the Hausdorff dimension of dynamical Cantor sets associated to piecewise *affine* maps ψ with full branches (i.e., mixing time $n_0 = 1$) such as Cantor's middle-third set $K_{1/3}$: indeed, $K_{1/3}$ is defined by a map $\phi : I_1 \cup I_2 \rightarrow I$ such that $\phi' \equiv 3$ (and $n_0 = 1$), so that $\alpha_1 \leq \dim(K_{1/3}) \leq \beta_1$, where

$$2(1/3)^{\alpha_1} = 1 = 2(1/3)^{\beta_1},$$

i.e., $\dim(K_{1/3}) = \log 2 / \log 3$.

Unfortunately, the elementary technique described in the previous paragraph doesn't permit to compute the dimension of dynamical Cantor sets K associated to *non-essentially affine*²¹ maps $\psi : I_1 \cup \dots \cup I_k \rightarrow I$. In fact, one can check that $\beta_m - \alpha_m = O(1/m)$ in general, so that the convergence of α_m and β_m to

²¹In the sense that there is no smooth change of coordinates $h : I \rightarrow J$ making $h \circ \psi \circ h^{-1}$ into a piecewise affine map.

$\dim(K)$ might be quite slow²² in the non-essentially affine situations related to the classical spectra L and M .

Nevertheless, Bowen [?MR556580] discovered in 1979 a famous formula for the Hausdorff dimension of dynamical Cantor sets K which was subsequently explored by several authors (including Falk, Hensley, Jenkinson, McMullen, Nussbaum, Pollicott, Vyt-nova) for a *fast* computation of *several* digits of $\dim(K)$. Roughly speaking, Bowen's formula for a dynamical Cantor set K associated to an expanding map $\psi : I_1 \cup \dots \cup I_k \rightarrow I$ starts with a family $(\mathcal{L}_t)_{t \in (0,1)}$ of *Ruelle–Perron–Frobenius transfer operators*

$$\mathcal{L}_t f(x) = \sum_{y \in \psi^{-1}(x)} f(y) |\psi'(y)|^{-t}$$

acting on adequate spaces of smooth functions $f : I \rightarrow \mathbb{R}$. This kind of operator was originally introduced by Ruelle in his study of ergodic theoretical properties of dynamical systems via an analogy with the so-called *thermodynamical formalism* in statistical mechanics.²³ The operators \mathcal{L}_t are *quasi-compact* (i.e., their spectral theories share many parallels with matrices in finite-dimensional vector spaces) and they possess a leading eigenvalue λ_t with multiplicity one. In this setting, Bowen showed that the Hausdorff dimension of K is the unique parameter $\dim(K) \in (0, 1)$ such that

$$\lambda_{\dim(K)} = 1.$$

In particular, Cantor's middle-third set $K_{1/3}$ has a Hausdorff dimension $\log 2 / \log 3$ such that $\mathcal{L}_{\log 2 / \log 3}(1) = 1$.

The spectral characterisation of $\dim(K)$ provided by Bowen can be reformulated by saying that $s =$

²²Note that the calculations of α_m and β_m require to manipulate the intervals in \mathcal{R}_m and, hence, our $O(1/m)$ -approximation to $\dim(K)$ comes from the computations with a quantity $\#\mathcal{R}_m$ of intervals growing exponentially with m .

²³Transfer operators describe the action of ψ on the densities (Radon–Nykodym derivatives) of probability measures: by fixing a probability measure λ which is non-singular for ψ (in the sense that $\psi_*(\lambda) = \lambda \circ \psi^{-1}$ is absolutely continuous with respect to λ), we get a transfer operator $\mathcal{L}_{(\lambda)}$ by looking at the Radon–Nykodym derivative $\mathcal{L}_{(\lambda)}(f) = \frac{d\psi_*(f\lambda)}{d\lambda}$ of the push-forward $\psi_*(f\lambda)$ under ψ of a probability $\nu = f\lambda$ which is absolutely continuous with respect to λ .

$\dim(K)$ is the unique parameter such that

$$\det(\text{Id} - \mathcal{L}_s) = 0.$$

In other words, we are interested in the value of s such that $z = 1$ is a zero of $\det(\text{Id} - z\mathcal{L}_s)$. As it turns out, we can efficiently compute this value thanks to the *Fredholm determinant expansion*²⁴

$$\det(\text{Id} - z\mathcal{L}_s) = \exp\left(-\sum_{n=1}^{\infty} \text{tr}(\mathcal{L}_s^n) \frac{z^n}{n}\right).$$

Indeed, it is possible to check that the traces $\text{tr}(\mathcal{L}_s^n)$ have nice expressions in terms of the *periodic points* of ψ :

$$\text{tr}(\mathcal{L}_s^n) = \sum_{\psi^n(p)=p} \frac{|(\psi^n)'(p)|^{-s}}{1 - (\psi^n)'(p)^{-1}}.$$

Thus, we can write $\det(\text{Id} - z\mathcal{L}_s) = 1 + \sum_{n=1}^{\infty} d_n(s) z^n$, where $d_n(s)$ are *explicit* functions of s and the periodic points of ψ . This provides a practical scheme to compute $\dim(K)$ when ψ is piecewise real-analytic because it is possible to prove that in this situation the solutions s_M of the truncations $1 + \sum_{n=1}^M d_n(s) = 0$ of Bowen's formula

$$\det(\text{Id} - \mathcal{L}_{\dim(K)}) = 1 + \sum_{n=1}^{\infty} d_n(\dim(K)) = 0$$

converge quickly to $\dim(K)$, namely,

$$|s_M - \dim(K)| = O(\theta^{M^2})$$

for some $\theta < 1$.

In particular, the method in the previous paragraph was successfully explored by Jenkinson and

²⁴This expansion generalizes the case of finite-dimensional vector space: if K is a $n \times n$ matrix with eigenvalues $\lambda_1, \dots, \lambda_n$, then $\det(\text{Id} + K) = \prod_{j=1}^n (1 + \lambda_j)$, so that $\log \det(\text{Id} + K) = \sum_{j=1}^n \log(1 + \lambda_j) = \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} \left(\sum_{j=1}^n \lambda_j^m \right) = \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} \text{tr}(K^m)$.

Pollicott in 2018 to compute the first 100 decimal digits of the Hausdorff dimension $\dim(C(2))$ of the Cantor set $C(2) = \{[0; a_1, a_2, \dots] : 1 \leq a_i \leq 2 \forall i\}$ which is dynamically defined by the piecewise real-analytic map given by the restriction of the Gauss map to the intervals $[[0; 2\overline{1}2], [0; 2, \overline{2}1]]$ and $[[0; 1\overline{1}2], [0; 1, \overline{2}1]]$. The outcome of their calculations is that

$$\dim(C(2)) = 0.531280506277205141624468647368\dots$$

3.1 Dimension of Gauss–Cantor sets

For our purpose of studying the classical spectra, the relevant class of dynamical Cantor sets are the so-called *Gauss–Cantor sets* defined as follows.

Let $B \subset \bigcup_{n \geq 1} (N^*)^n$ be a finite set of finite words which is *primitive* in the sense that none of its elements is a prefix of another one. The corresponding *complete Gauss–Cantor set* is

$$K(B) = \{[0; \beta_1, \beta_2, \dots] : \beta_i \in B \forall i \geq 1\}.$$

The simplest examples²⁵ of complete Gauss–Cantor sets are the sets

$$C(k) := \{[0; a_1, a_2, \dots] : 1 \leq a_i \leq k \forall i\}$$

for $k \geq 2$.

In general, a *Gauss–Cantor set* is a set of the type

$$K(\gamma, B) = \{[0; \gamma, \beta_1, \beta_2, \dots] : \beta_i \in B \forall i \geq 1\},$$

where $\gamma \in \bigcup_{n \geq 1} (N^*)^n$ is a finite word.

Notice that $K(\gamma, B)$ is the image of $K(B)$ under the bi-Lipschitz homeomorphism

$$[0; \gamma, x] \mapsto [0; x] = G^{|\gamma|}([0; \gamma, x]),$$

where G is the Gauss map and $|\gamma|$ is the size of γ . In particular, $K(B)$ and $K(\gamma, B)$ have the same Hausdorff dimension.

Complete Gauss–Cantor sets are dynamical Cantor sets defined by iterates of the Gauss map G . Indeed,

²⁵Note that $C(4)$ and $C(2)$ already appeared in our discussions of the ending and the intermediate portions of the classical spectra L and M .

$K(B)$ is a dynamical Cantor set $\psi : \bigcup_{\beta \in B} I(\beta) \rightarrow I$

where $\psi|_{I(\beta)} = G^{|\beta|}$ and $I(\beta)$ are intervals with extremities of the form $[0; \beta, x_\beta]$ and $[0; \beta, y_\beta]$ for adequate choices of $x_\beta, y_\beta \in B^N$.

Similarly, one can verify that, in general, Gauss–Cantor sets are also dynamical Cantor sets.

It is not difficult to show that any Gauss–Cantor set is *non-essentially affine*, i.e., it is dynamically defined by a map $\psi : I_1 \cup \dots \cup I_k \rightarrow I$ such that there is no smooth change of coordinates h making the second derivative of $h \circ \psi \circ h^{-1}$ to vanish identically on $h(K)$. Thus, a Gauss–Cantor set is geometrically more intricate than dynamical Cantor sets given by piecewise affine maps (such as Cantor’s middle-third set) and, hence, we do not expect to get an *exact formula* for its Hausdorff dimension (but only some high precision approximation coming from variants of Bowen’s formula, for example).

Nonetheless, the second author discovered²⁶ that the renormalization techniques (including the so-called *scale recurrence lemma*) introduced by Yoccoz and him [?MR1865980] in their study of stable intersections of dynamical Cantor sets can be used to prove that if K is a non-essentially affine dynamical Cantor set and K' is an arbitrary dynamical Cantor set, then the projection $f(K \times K') = K + K'$ of $K \times K'$ under $f(x, y) = x + y$ has *expected* Hausdorff dimension

$$\dim(K + K') = \min\{1, \dim(K) + \dim(K')\}.$$

In particular, Gauss–Cantor sets satisfy the following *dimension formula*: for any finite words γ, γ' and finite sets of finite words B, B' , one has

$$\dim(K(\gamma, B) + K(\gamma', B')) = \min\{1, \dim(K(B)) + \dim(K(B'))\}.$$

Also, it is worth to point out the following²⁷ useful “symmetry” on the Hausdorff dimension of Gauss–Cantor sets. Given a finite set of finite words B , denote by $B^T = \{\beta^T : \beta \in B\}$ the *transpose* of

²⁶A version of this formula was obtained by Hochman and Shmerkin.

²⁷The dynamical explanation for this symmetry is the fact that the Gauss map has a smooth, area-preserving, natural extension.

B , where $\beta^T := (a_n, \dots, a_1)$ stands as usual for the transpose of $\beta = (a_1, \dots, a_n)$. Then, the Hausdorff dimension of the Gauss–Cantor sets associated to B and B^T are equal:

$$\dim(K(B)) = \dim(K(B^T)).$$

In a certain sense, the proof of this fact goes back to Euler: indeed, he proved that for any finite word β , if $[0; \beta] = p_n/q_n$, then $[0; \beta^T] = r_n/q_n$. Since the lengths of the intervals $I(\beta_1\beta_2\dots\beta_k)$ in the k -th step of the construction of $K(B)$ depend only on the denominators of the convergents of $[0; \beta_1\beta_2\dots\beta_k]$, Euler’s result says that $K(B)$ and $K(B^T)$ are Cantor sets constructed from small intervals with comparable lengths, and, *a fortiori*, they have the same Hausdorff dimension.

3.2 Dimension across the spectra

As we told in the end of §2.7, we want to study portions of L and M via dynamical Cantor sets. In particular, a central idea towards the main theorem of [MR3815461] about the continuity of the Hausdorff dimension across the classical spectra is to approximate $L \cap (-\infty, t)$ and $M \cap (-\infty, t)$ from inside and outside by arithmetic sums of Gauss–Cantor sets.

More precisely, let $D(t) = \dim(\ell^{-1}(-\infty, t))$ (where $\ell(\alpha)$ is the best constant²⁸ of Diophantine approximation of α). Given t such that $D(t) > 0$ and $\epsilon > 0$, the second author proved the existence of

- a parameter $\delta > 0$,
- a finite set of positive integers $(a_j)_{1 \leq j \leq m+1}$,
- a finite set of finite prefixes $\{\gamma_j\}_{j=1}^{2m+2} \subset \bigcup_{n \geq 1} (N^*)^n$, and
- two finite sets B, B' of finite words

such that the translated arithmetic sum of Gauss–Cantor sets

$$a_{m+1} + K(\gamma_{2m+1}, B') + K(\gamma_{2m+2}, (B')^T)$$

²⁸Cf. the beginning of §2.

is contained in $L \cap (-\infty, t - \delta)$, the union of translated arithmetic sums of Gauss–Cantor sets

$$\bigcup_{j \leq m} (a_j + K(\gamma_{2j-1}, B) + K(\gamma_{2j}, B^T))$$

contains $M \cap (-\infty, t + \delta)$, and

$$D(t) - \epsilon < \dim(K(B')) \leq D(t) \leq \dim(K(B)) < D(t) + \epsilon.$$

In particular, these facts together with the dimension formula and Euler’s symmetry imply that

$$\begin{aligned} \min\{1, 2D(t) - 2\epsilon\} &\leq \min\{1, 2 \dim(K(B'))\} \\ &= \min\{1, \dim(K(B')) + \dim(K((B')^T))\} \\ &\leq \dim(L \cap (-\infty, t - \delta)) \\ &\leq \dim(L \cap (-\infty, t)) \end{aligned}$$

and

$$\begin{aligned} \dim(M \cap (-\infty, t)) &\leq \dim(M \cap (-\infty, t + \delta)) \\ &\leq \min\{1, \dim(K(B)) + \dim(K(B^T))\} \\ &= \min\{1, 2 \dim(K(B))\} \\ &\leq \min\{1, 2D(t) + 2\epsilon\}. \end{aligned}$$

Since $\epsilon > 0$ was arbitrary, we conclude that

$$\dim(L \cap (-\infty, t)) = \dim(M \cap (-\infty, t)) = \min\{1, 2D(t)\}$$

is a continuous function of t .

In 1982, Bumby extended Hall’s ideas in §2.6 to give a computer-assisted argument *indicating* that $D(3.33437) < 1/2$ and $D(3.3344) > 1/2$, so that

$$3.33437 < \inf\{t : \dim(M \cap [\sqrt{5}, t]) = 1\} < 3.3344$$

in view of our current discussion.

3.3 Bounds on $\dim(M \setminus L)$

In [MR4152626], we explored the geometry of the intersections of the so-called local *stable* and *unstable* sets²⁹ of the shift dynamics σ to prove that

²⁹The local stable set of the left-shift map $\sigma : (N^*)^Z \rightarrow (N^*)^Z$ at $x = (x_n)_{n \in Z}$ is the subset of $y = (y_n)_{n \in Z}$ such that $y_m = x_m$ for all $m \geq 0$. This nomenclature is justified by the fact that the distance between $\sigma^k(y)$ and $\sigma^k(x)$ goes exponentially fast to zero as $k \rightarrow +\infty$ when y belongs to the local stable set of x . Analogously, we get local unstable sets after replacing σ by σ^{-1} in the previous definition.

$M \setminus L \supset a + K(\gamma, B)$, where $a = [3; \overline{3, 2, 1, 2, 2, 2, 3, 3}]$, $\gamma = (2, 2, 2, 1, 2, 3, 3, 2, 2, 2, 1, 2, 2, 1, 2, 1, 2, 1, 2)$ and $B = \{1, 2\}$. Thus, $\dim(M \setminus L) \geq \dim(K(\gamma, B)) = \dim(K(B)) = \dim(C(2)) > 0.531$.

After this brief discussion of lower bounds on $\dim((M \setminus L))$, let us now sketch³⁰ a proof of the upper bound $\dim((M \setminus L) \cap [\sqrt{5}, \sqrt{12}]) < 0.93$.

We saw in §2.3 that if $\theta \in (N^*)^{\mathbb{Z}}$, then

$$m(\theta) := \sup_{n \in \mathbb{Z}} f(\sigma^n(\theta)) \leq \sqrt{12} \iff \theta \in \{1, 2\}^{\mathbb{Z}}.$$

Moreover, we saw in §2.6 that $m(x) \leq \sqrt{10}$ implies that $x \in \{1, 2\}^{\mathbb{Z}}$ doesn't contain 121 and, *a fortiori*,

$$\dim(M \cap [3, \sqrt{10}]) < 0.93.$$

Since $(M \setminus L) \cap [3, \sqrt{12}] \subset (M \cap [3, \sqrt{10}]) \cup ((M \setminus L) \cap [\sqrt{10}, \sqrt{12}])$, our claim about the dimension of $(M \setminus L) \cap [\sqrt{5}, \sqrt{12}]$ is reduced to prove³¹ that

$$\dim((M \setminus L) \cap [\sqrt{10}, \sqrt{12}]) < 0.93 \quad (2)$$

The crucial idea at this point is to put restrictions on the past or future shift dynamics of sequences $\theta \in \{1, 2\}^{\mathbb{Z}}$ leading to Markov values in $(M \setminus L) \cap [\sqrt{10}, \sqrt{12}]$. For this sake, we take inspiration from the so-called the *shadowing lemma*, a classical result from the theory of uniformly hyperbolic dynamical systems asserting that pseudo-orbits are tracked by genuine orbits (cf. Hasselblatt–Katok book [?MR1326374]). In our setting, this lemma essentially says that if $m(\theta) \in (M \setminus L) \cap [\sqrt{10}, \sqrt{12}]$, then, up to transposition, the future dynamics of θ lives in the gaps of the *symmetric block* $B = \{11, 22\}^{\mathbb{Z}} \subset \{x \in \{1, 2\}^{\mathbb{Z}} : x \text{ not containing } 121\}$. Let us now try to make this statement more precise.

We have a setting which can also be used in the analysis of other portions of $M \setminus L$: more concretely, we dispose of a symmetric block C , which here

³⁰These ideas can be adapted to give nontrivial upper estimates for $\dim(M \setminus L)$ by analyzing other regions of the spectra.

³¹Contrary to the crude bound on the dimension of $(M \setminus L) \cap [3, \sqrt{10}]$ via the dimension of $M \cap [3, \sqrt{10}]$, the estimate $\dim((M \setminus L) \cap [\sqrt{10}, \sqrt{12}]) < 0.93$ is non-trivial because, as we already mentioned, $\dim(M \cap [\sqrt{10}, \sqrt{12}]) = 1$ thanks to the fact that $\dim(C(2)) > 1/2$.

is $\{1, 2\}^{\mathbb{Z}}$, together with its corresponding Gauss–Cantor set $K(C) = C(2)$, and a smaller symmetric block B giving rise to the Gauss–Cantor set $K(B) = K(\{11, 22\})$. The elements of $(M \setminus L) \cap [\sqrt{10}, \sqrt{12}]$ will be of the form $f(\theta)$ for some $\theta \in C$ such that $m(\theta) = f(\theta)$, i.e., let us consider

$$Y = \{\theta \in C : m(\theta) = f(\theta) \in (M \setminus L) \cap [\sqrt{10}, \sqrt{12}]\}.$$

In order to prove that $\dim((M \setminus L) \cap [\sqrt{10}, \sqrt{12}]) \leq d$, it is enough (by taking countable subcoverings) to prove that, for every $\theta = (\theta_j)_{j \in \mathbb{Z}} \in Y$, there is $N \in \mathbb{N}$ such that

$$\dim(f(V_N(\theta) \cap Y)) \leq d,$$

where $V_N(\theta) = \{\tilde{\theta} = (\tilde{\theta}_j)_{j \in \mathbb{Z}}; \tilde{\theta}_j = \theta_j, -N \leq j \leq N\}$.

We say that a finite sequence $\omega = (\theta_j)_{-N \leq j \leq n}$ has an *allowed continuation* τ , where τ is a finite string of 1 and 2, if there are sequences $\alpha, \beta \in \{1, 2\}^{\mathbb{N}}$ such that $\beta^T \omega \tau \alpha \in Y$.

We *claim* that a shadowing lemma type argument implies that, if $\theta \in Y$, then (perhaps replacing θ by its transpose θ^T) there is $N \in \mathbb{N}$ such that, for every $n \geq N$, the bifurcation tree (of allowed continuations) of such a θ is severely constrained:

- either $\theta_{-N} \dots \theta_n$ has a unique 1-element continuation θ_{n+1} ,
- or its (3-elements) allowed continuations are 112 and 221.

In order to show this, observe that if $\theta \in Y$, then, since L is closed, there is $N \in \mathbb{N}$ such that the distance of $m(\theta)$ to L is larger than $\frac{1}{2^{N-3}}$. Also, if $\tilde{\theta} \in V_N(\theta) \cap Y$, then $|m(\tilde{\theta}) - m(\theta)| = |f(\tilde{\theta}) - f(\theta)| < \frac{1}{2^{N-2}}$ and $\lambda_j(\tilde{\theta}) := f(\sigma^j(\tilde{\theta})) \leq f(\tilde{\theta}), \forall j \in \mathbb{Z}$.

If the claimed statement above is not true for this N , there are $n \geq N$, $\tau = (1, a, b) \neq (1, 1, 2)$, $\tau' = (2, c, d) \neq (2, 2, 1)$, $\alpha, \alpha', \beta, \beta' \in \{1, 2\}^{\mathbb{N}}$ such that $\beta^T \omega \tau \alpha, \beta'^T \omega \tau' \alpha' \in Y$, where $\omega = (\theta_j)_{-N \leq j \leq n}$. Since the smallest continued fractions in $C(2)$ beginning by $[0; 1]$ begin indeed by $[0; 1, 1, 2]$, the largest continued fractions in $C(2)$ beginning by $[0; 2]$ begin by $[0; 2, 2, 1]$, and there are elements of $K(B)$ beginning by $[0; 1, 1, 2]$ and by $[0; 2, 2, 1]$, it follows that there is $\tilde{\alpha} \in \{11, 22\}^{\mathbb{N}}$ such that

$$[0; \tilde{\alpha}] \in K(B) \cap ([0; \tau \alpha], [0; \tau' \alpha']).$$

By monotonicity of maps of the type $h(y) = [c_0, c_1, \dots, c_k + y]$, it follows that, for $-N \leq j \leq n+1$,

$$\lambda_j(\beta^T \omega \tilde{\alpha}) < m(\theta) + \frac{1}{2^{N-2}} + \frac{1}{2^{N-1}} < m(\theta) + \frac{1}{2^{N-3}}.$$

On the other hand, since $\tilde{\alpha} \in \{11, 22\}^N$, we have, for $j > n+1$, $\lambda_j(\beta^T \omega \tau \tilde{\alpha}) \leq [2; \overline{1, 1}] + [0; 2, \overline{2, 1}] < 3.0407 < \sqrt{10} \leq m(\theta)$.

Hence, if the claimed statement above on allowed continuations fails both for θ and θ^T , we may find $n, n' \geq N$ and $\tilde{\alpha}, \hat{\alpha} \in \{11, 22\}^N$ such that $\lambda_j(\hat{\alpha}^T \hat{\omega} \tilde{\alpha}) < m(\theta) + \frac{1}{2^{N-3}}$ for all $j \in Z$, where $\hat{\omega} = (\theta_j)_{-n' \leq j \leq n}$.

This leads to a contradiction: if $\tilde{\alpha} = (\tilde{a}_1, \tilde{a}_2, \dots)$ and $\hat{\alpha} = (\hat{a}_1, \hat{a}_2, \dots)$, and $\tilde{\alpha} = (\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_{2N}, \hat{a}_{2N}, \dots, \hat{a}_2, \hat{a}_1) \in \{1, 2\}^{4N}$, then, if $\gamma = \tilde{\omega} \tilde{\alpha}$ is the periodic sequence with period $\tilde{\omega} \tilde{\alpha}$, then $|m(\gamma) - m(\theta)| < \frac{1}{2^{N-3}}$ and $m(\gamma) \in L$, an absurd.

The result above on allowed continuations can be used to get that the set $\{[\theta_0; \theta_1, \theta_2, \dots]\}$ related to those allowed continuations of $\theta_{-N} \dots \theta_N$ is contained in a small ‘‘Cantor set of gaps’’ K_G and, *a fortiori*, we have $(M \setminus L) \cap [\sqrt{10}, \sqrt{12}] \subset C(2) + K_G$.

The non-trivial control of the bifurcation tree allows to infer that

$$\dim(K_G) \leq s_0$$

where $s_0 \in (0, 1)$ is any number such that

$$\begin{aligned} |I(\theta_1, \dots, \theta_n, 1, 1, 2)|^{s_0} + |I(\theta_1, \dots, \theta_n, 2, 2, 1)|^{s_0} \\ \leq |I(\theta_1, \dots, \theta_n)|^{s_0} \end{aligned}$$

for all $(\theta_0, \dots, \theta_n) \in \{1, 2\}^n$. On the other hand, it is possible to derive³² from the inclusion $(M \setminus L) \cap [\sqrt{10}, \sqrt{12}] \subset C(2) + K_G$ that

$$\dim((M \setminus L) \cap [\sqrt{10}, \sqrt{12}]) \leq \dim(C(2)) + \dim(K_G).$$

These facts permit to give a good upper bound on $\dim((M \setminus L) \cap [\sqrt{10}, \sqrt{12}])$ because $\dim(C(2)) <$

³²Here, we are using two useful facts: the box dimension of a Gauss–Cantor set (such as $C(2)$) coincides with its Hausdorff dimension and $\dim(X \times Y) \leq \dim_{\text{box}}(X) + \dim(Y)$ where $\dim_{\text{box}}(X)$ is the box dimension of X .

0.531281 and some elementary estimates on continued fractions yield that $s_0 = 0.174813$ satisfies the above requirement, so that

$$\begin{aligned} \dim((M \setminus L) \cap [\sqrt{10}, \sqrt{12}]) &< 0.531281 + 0.174813 \\ &= 0.706094. \end{aligned}$$

This completes the proof of e.MM-upper-bound.

4 Beyond the classical spectra

Partly inspired by Perron’s characterization of the classical spectra, several authors (including Maucourant, Paulin, Parkkonen and the second author) proposed dynamical generalizations of the Markov and Lagrange spectra. In a nutshell, *dynamical Lagrange and Markov spectra* are obtained after replacing σ by a *general* dynamical system and f by a general height function: for instance, given a homeomorphism $\varphi : M \rightarrow M$ of a topological space M , a compact φ -invariant subset Λ of M , and a continuous function $f : M \rightarrow \mathbb{R}$, the *dynamical Lagrange and Markov spectra*³³ associated to (f, Λ) are

$$L(f, \Lambda) := \left\{ \limsup_{n \rightarrow \infty} f(\varphi^n(x)) : x \in \Lambda \right\}$$

and

$$M(f, \Lambda) = \left\{ \sup_{n \in \mathbb{Z}} f(\varphi^n(x)) : x \in \Lambda \right\}.$$

A direct generalization of the classical spectra is provided by the dynamical spectra associated to a diffeomorphism $\varphi : M \rightarrow M$ of a surface M acting on a *horseshoe*³⁴ and a typical differentiable real function f . In fact, arbitrarily large compact parts of the classical Markov and Lagrange spectra can be viewed as dynamical Markov and Lagrange spectra associated to horseshoes of *conservative* (i.e., area-preserving) diffeomorphisms. More precisely, as it is

³³An analogous definition can be made when the discrete time dynamical system $\varphi : M \rightarrow M$ is replaced by a continuous time dynamical system $(\varphi^t)_{t \in \mathbb{R}}$.

³⁴This is a compact, φ -invariant, uniformly hyperbolic set of saddle type: cf. Hasselblatt–Katok book [MR1326374].

explained in [?MR1279059], for each $m \geq 2$, the map $T_1 : (0, 1) \times (0, 1) \rightarrow [0, 1) \times (0, 1)$ given by

$$T_1(x, y) = \left(\left\{ \frac{1}{x} \right\}, \frac{1}{y + [1/x]} \right)$$

preserves a smooth area-form near the horseshoe $\Lambda(m) = C(m) \times C(m)$ corresponding to the maximal invariant set of $(\frac{1}{m+1}, 1) \times (0, 1)$. In particular, since

$$\begin{aligned} T_1([0; a_0, a_1, a_2, \dots], [0; b_1, b_2, b_3, \dots]) &= \\ &= ([0; a_1, a_2, a_3, \dots], [0; a_0, b_1, b_2, \dots]), \end{aligned}$$

we see that T_1 is a (piecewise) smooth, conservative realization of the natural extension of the Gauss map (compare with e.Gauss-shift). The dynamical Markov and Lagrange spectra of $(T_1, \Lambda(m))$ with respect to the function $f(x, y) = y + \frac{1}{x}$ have the same intersections with $(-\infty, m+1]$ as the classical Markov and Lagrange spectra.

In 2018, Cerqueira and the authors established the continuity of the Hausdorff dimension across the dynamical Lagrange and Markov spectra of typical *thin* horseshoes of *conservative* surface diffeomorphisms with respect to typical smooth functions (in analogy with the main continuity result in §2.7 for L and M). More precisely, let φ_0 be a smooth diffeomorphism of a surface M^2 preserving an area-form ω . Suppose that φ_0 possesses a thin horseshoe Λ_0 in the sense that its Hausdorff dimension is $\dim(\Lambda_0) < 1$. Denote by \mathcal{U} a small C^∞ neighborhood of φ_0 in the space $\text{Diff}_\omega^\infty(M)$ of smooth area-preserving diffeomorphisms of M such that Λ_0 admits a *continuation*³⁵ Λ for every $\varphi \in \mathcal{U}$. If $\mathcal{U} \subset \text{Diff}_\omega^\infty(M)$ is sufficiently small, then there exists a Baire residual subset $\mathcal{U}^{**} \subset \mathcal{U}$ with the following property. For every $\varphi \in \mathcal{U}^{**}$ and $r \geq 2$, there exists a C^r -open and dense subset $\mathcal{R}_{\varphi, \Lambda} \subset C^r(M, R)$ such that

$$\dim(L(\Lambda, f) \cap (-\infty, t)) = \dim(M(\Lambda, f) \cap (-\infty, t))$$

is a continuous function of t whenever $f \in \mathcal{R}_{\varphi, \Lambda}$.

³⁵I.e., if U_0 is a neighborhood of Λ_0 such that $\Lambda_0 = \bigcap_{n \in \mathbb{Z}} \varphi_0^n(U_0)$, then \mathcal{U} is taken small enough so that $\Lambda = \bigcap_{n \in \mathbb{Z}} \varphi^n(U_0)$ still is a horseshoe for any $\varphi \in \mathcal{U}$.

Still concerning the beginning of the dynamical spectra, the second author [?MR4142446] proved that, for typical pairs (f, Λ) as above, the minima of the corresponding Lagrange and Markov dynamical spectra coincide and are given by the image of a periodic point of the dynamics by the real function, solving a question by Yoccoz.

Recently, Lima and the second author³⁶ proved that, for typical pairs (f, Λ) as in the previous paragraphs,

$$\begin{aligned} \sup\{t \in R : \dim(M(f, \Lambda) \cap (-\infty, t)) < 1\} &= \\ &= \inf\{t \in R : \text{int}(L(f, \Lambda)) \cap (-\infty, t) \neq \emptyset\}, \end{aligned}$$

and, inspired by this result, they conjectured that the classical Lagrange spectrum must have non-empty interior right after the transition point where the classical Markov spectrum acquires Hausdorff dimension one, i.e.,

$$\text{int}(L) \cap (t_1, t_1 + \epsilon) \neq \emptyset$$

for all $\epsilon > 0$, where $t_1 = \inf\{t : \dim(M \cap [\sqrt{5}, t]) = 1\}$.

Concerning the ending of dynamical spectra associated to horseshoes, Romaña and the second author proved that if Λ is a (not necessarily conservative) horseshoe associated to a C^2 -diffeomorphism φ such that $\dim(\Lambda) > 1$, then there is, arbitrarily close to φ a diffeomorphism φ_0 and a C^2 -neighborhood W of φ_0 such that, if Λ_ψ denotes the continuation of Λ associated to $\psi \in W$, there is an open and dense set $H_\psi \subset C^1(M, R)$ such that for all $f \in H_\psi$, we have

$$\text{int } L(f, \Lambda_\psi) \neq \emptyset \text{ and } \text{int } M(f, \Lambda_\psi) \neq \emptyset.$$

Another direct generalization of the classical spectra for geodesic flows on negatively curved manifolds and moduli spaces of translation surfaces (along the lines of §2.4) was studied by Maucourant, Paulin, Parkkonen, Artigiani, Delecroix, Hubert, Lelièvre, Marchese, and Ulcigrai (among others): in their respective settings, these authors showed that their spectra shared some properties of the classical spectra such as isolated minima and a Hall's ray. We refer

³⁶See the article *Phase transitions on the Markov and Lagrange dynamical spectra*, to appear in Ann. Inst. H. Poincaré Anal. Non Linéaire. <https://doi.org/10.1016/j.anihpc.2020.11.007>

the reader to our recent book with Lima and Romãna
[?doi:10.1142/11965] (and the references therein)
for more details about dynamical generalizations of
the classical spectra.