FOURIER DECAY IN NONLINEAR DYNAMICS

A THESIS SUBMITTED TO THE UNIVERSITY OF MANCHESTER
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE FACULTY OF SCIENCE AND ENGINEERING

2021

Connor Michael Stevens
Department of Mathematics
School of Natural Sciences

Contents

Word count 36,902

The University of Manchester

Connor Michael Stevens
Doctor of Philosophy
Fourier Decay in Nonlinear Dynamics
November 29, 2021

In 2017, Bourgain–Dyatlov [8] prove that Patterson–Sullivan measures on the limit set of convex cocompact Fuchsian groups have polynomial Fourier decay. We begin by proving that their main tool, Bourgain's exponential sum theory, can be used to prove polynomial Fourier decay for Gibbs measures for sufficiently nonlinear Markov maps. We follow up by proving a remark of Bourgain–Dyatlov which stated that a technical dimension assumption can be removed from a Fourier decay theorem of Jordan–Sahlsten [26] by proving that the Gauss map is sufficiently nonlinear.

We move on to prove an analogous theorem for a much more general class of (finite) nonlinear Markov maps with a strong separation condition. We do so using the complex transfer operator theory of Naud [44] as recommended to us by Jialun Li and Frédéric Naud. All aforementioned work is joint with Tuomas Sahlsten.

To finish, we go back to the ground-breaking work of Bourgain–Dyatlov, and ask whether we can prove their main Fourier decay result for Gibbs measures on limit sets of convex cocompact Fuchsian groups. A corollary of the aforementioned theorem on general nonlinear Markov maps with strong separation is that we can obtain polynomial Fourier decay for such measures. Alternatively, we can use the combinatorial large deviation theory of Jordan–Sahlsten to prove a polynomial Fourier decay theorem for a class of measures which are defined using Schottky structures. This avoids the need for complex transfer operator theory to prove that the Fuchsian groups are sufficiently nonlinear; we can just use the distortion factor analysis of Bourgain–Dyatlov. We conclude by proving a fractal uncertainty principle for Gibbs measures for Markov maps with (eventual) polynomial Fourier decay by slightly adapting a proof of Bourgain–Dyatlov [8].

Declaration

The results presented in Chapters 2, 3, and 5 appear in the paper [54], all versions of which are publicly available on arXiv. The results presented in Chapter 4 appear in the paper [53] which is available on arXiv. Both [54] and [53] were written in collaboration with Tuomas Sahlsten.

I declare that the material presented in Chapters 1, 6, and 7 is my own work, unless otherwise indicated in the text.

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

I dedicate the hard work that I have put into this thesis to the memory of Nanna and Grandad.

Acknowledgements

To begin, I first must express my gratitude to my supervisor Tuomas Sahlsten. It is hard to know where to begin because I owe him a lot over the past five years of working together. I have achieved far beyond my own expectations by producing theorems together that we can both be proud of. In Tuomas I have gained a mentor, coauthor, and friend. I hope that my contribution means that he is a step closer to becoming one of the biggest names in Fourier transforms for measures, because he certainly deserves such prestige.

I would like to thank Charles Walkden for his support and guidance. Charles is clearly an expert in his role of supervisor, from his vast knowledge of ergodic theory, to his humour and enjoyable conversation. I hope that he soon takes on more PhD students so he can continue to pass on his knowledge, and I look forward to him publishing his books.

Over the course of my PhD I have had the pleasure of meeting and discussing with many brilliant mathematicians. I would like to thank Semyon Dyatlov, Thomas Jordan, Jialun Li, and Frédéric Naud for their contributions which led me to be able to study some beautiful mathematics. The main results of this thesis were made possible because of the brilliant work of the late Jean Bourgain. I am also very thankful to many teachers from the Minster School and the Dukeries College for their dedication to my learning.

I would like to thank my family for all their love and support. In particular I would like to thank Mum, Dad, Chris and Louise for always welcoming me home, even if I am mostly just a pain. My academic journey would have ended much sooner if it wasn't for their praise, encouragement and love. I should also mention Char, Nins, Tabat, Craig, and Kyle for similar reasons.

I would like to thank my friends, old and new, for welcoming me back home to

Nottinghamshire and making my time studying the PhD an unforgettable experience.

The work presented in this thesis was supported by a Doctoral Training Award from the EPSRC and the University of Manchester.

Chapter 1

Introduction

1.1 A Brief History of Time and Frequencies

The Fourier transform is an incredibly useful operator when studying functions. It is well known that the Fourier transform of an integrable function $f : \mathbb{R} \to \mathbb{C}$ at a frequency $\xi \in \mathbb{R}$ is defined by

$$\widehat{f}(\xi) = \int_{\mathbb{R}} f(x)e^{-2\pi i x\xi}\, dx.$$

They are used heavily in Quantum mechanics to define wave functions of the momentum of a free particle using its position wave function (see subsection 1.1.6). Most notably, Joseph Fourier in 1807 was the first person to introduce the idea of Fourier series to solve the Heat equation [12]. His intention was to mathematically model the flow of heat along a strip of metal. He assumed that the strip:

(i) is infinitesimally thick (two dimensional);

(ii) is infinitely long (think of this strip as starting from an axis $y = 0$ and going to $y \to \infty$);

(iii) has finite width (consider edges at $x = -1$ and $x = 1$) and the edges will always have zero temperature;

(iv) is in a vacuum (or more specifically, there is no loss of heat from the strip by any means).

Initially, Fourier considered the case of applying constant heat 1 (ignoring units) to the strip across $y = 0$. This already causes issues because we want zero heat at the edges $x = -1$ and $x = 1$. This was resolved by assuming that 1 could be written as an infinite sum of cosine functions. This indeed gave a solution, but his methods were mostly left unproven, including the use of a convergence theorem for integration and the use of an infinite sum of cosine functions to represent the constant function 1. His solution matched findings in experiments, but the mathematical community did not accept his result. Fourier's findings were only proven to be true 22 years later [12].

There is an analogous operator to the Fourier transform when studying measures. For a Borel measure μ with support in $\mathbb{R}^n$, we can define the Fourier transform of the measure at frequency $\xi \in \mathbb{R}^n$ to be

$$\widehat{\mu}(\xi) = \int e^{-2\pi i \xi \cdot x} \, d\mu(x)$$

where we use the Euclidean inner product in the exponent. We have a uniqueness property by Fubini's theorem which says if two measures have the same Fourier transform, then they must be the same measure. The Fourier transform as a function is typically easier to analyse than the measure itself, and it can allude to many properties of the measure. The behaviour of $\widehat{\mu}$ towards infinity in particular can tell us how uniform the measure is. In the Lebesgue measure case, one can use the fundamental theorem of calculus to show that the Fourier transform of Lebesgue measure is identically zero (except at the origin, where the Fourier transform of a probability measure is always equal to one). If the Fourier transform is very close to zero, one should think of it as being 'close to Lebesgue', or uniform.

1.1.1 Uniqueness of Fourier Series

So Fourier used these Fourier series represent the constant function 1 as a sum of cosine functions. We know that the Fourier series of a function is unique; the coefficients are just given by the Fourier transform. However, the Fourier series is not always a perfect construction, which is dictated by how it converges (e.g. pointwise or with respect to the L^2 norm) to the original function. We could ask whether some given coefficients could give a Fourier series. A more general statement is whether we can have trigonometric series which are not unique, as they are in the case of Fourier series. We know that for sequences $a_m, b_m \in \mathbb{R}$ with $m \in \mathbb{Z}$, if we suppose that the two Fourier series are equal on $x \in [0,1]$, i.e. that

$$\sum_{m\in\mathbb{Z}} a_m e^{2\pi i m x} = \sum_{m\in\mathbb{Z}} b_m e^{2\pi i m x}$$

then we must have that $a_m = b_m$ for all $m \in \mathbb{Z}$ as proved by Riemann. The set $[0,1]$ is 'too big' for this not to be true. We can ask how large the set of values of x has to be to be able to say that $a_m = b_m$ for all $m \in \mathbb{Z}$ [29]. This motivates the definition of a set of uniqueness.

Definition 1.1.1. *We say that $E \subset [0,1]$ is a set of uniqueness if there does not exist a sequence $(a_m)_{m\in\mathbb{Z}}$ which is non-zero such that for all $x \in E^c$ we have that*

$$\sum_{m\in\mathbb{Z}} a_m e^{2\pi i m x} = 0.$$

If E is not a set of uniqueness, we call it a set of multiplicity.

Unsurprisingly this is a big problem in Fourier analysis, and we can use Fourier transforms for measures to answer some questions. Salem [56] proved that if there exists a probability measure μ such that $\operatorname{supp}\mu \subset E$ and the measure has Fourier decay, then E must be a set of multiplicity. This is one of the many reasons why Fourier decay is an interesting property.

1.1.2 Dimension Theory

One of the main motivations of studying Fourier transforms of measures comes from dimension theory. We can heuristically think of the dimension of a set as the exponential growth of the set at infinitesimal scales. When trying to find the Hausdorff dimension of a set $A \subset \mathbb{R}^n$, there are known techniques to find upper bounds for the dimension. The most common method is the use of covering sets. Finding the lower bound for the Hausdorff dimension of a set is typically where the most work is involved. One can consider the following definition of Hausdorff dimension which alludes to how the Fourier transform can help [39].

Definition 1.1.2. *For $s \in \mathbb{R}$ we define the s-energy of μ to be*

$$I_s(\mu) := \int\int |x-y|^{-s}\, d\mu(x) d\mu(y).$$

When the s-energy is finite, one can think of this as telling us that the measure does not concentrate on some ball $B(x,r) \subset \mathbb{R}^n$ with measure approximately r^s. This tells us that the measure is not supported on a set with dimension bigger than s. We can therefore motivate the following definition [39].

Definition 1.1.3. *We define the Hausdorff dimension of A to be*

$$\dim_H A := \sup\{s \leq n : I_s(\mu) < \infty \text{ for some } \mu \in \mathcal{M}(A)\}.$$

We can bound the s-energy of μ from above by

$$(2\pi)^{-n} \int |x|^{s-n} |\widehat{\mu}(x)|^2 \, dx$$

up to some constant [39]. As a result, if we can find a measure μ such that $\widehat{\mu}(\xi) = O(|\xi|^{-s/2})$, then we must have that s is a lower bound for $\dim_H A$. In fact, we can define another notion of dimension by trying to find lower bounds for Hausdorff dimension in such a way as follows [18].

Definition 1.1.4. *We define the Fourier dimension of A to be*

$$\dim_F A := \sup\{s \leq n : |\widehat{\mu}(\xi)| = O(|\xi|^{-s/2}) \text{ for some } \mu \in \mathcal{M}(A) \text{ with } \mu(A) > 0\}.$$

If the aforementioned method for finding lower bounds of Hausdorff dimension using s-energies were ideal, we would have that $\dim_F A = \dim_H A$. When this property does hold, we call A a Salem set. Fourier dimension has been well studied, so there are many examples and non-examples of Salem sets. For $\alpha > 2$, one collection of examples is the α-well approximable numbers

$$W(\alpha) := \bigcap_{n=1}^{\infty} \bigcup_{q=n}^{\infty} \bigcup_{p \in \mathbb{Z}} \{x \in (0,1) \setminus \mathbb{Q} : |x - p/q| \leq q^{-\alpha}\}$$

which are numbers where the bounds in Dirichlet's theorem for approximable numbers can be improved. This is a fractal set arising in number theory. Jarník and Besicovitch proved that the Hausdorff dimension of $W(\alpha)$ is $2/\alpha$ [25] [6]. Kaufman then proved that the Fourier dimension is the same by constructing a measure with sufficient decay, proving that $W(\alpha)$ is Salem. This Salem property is not typical, however.

A non-example for Salem sets is the graph of a typical Wiener process. Wiener processes are somewhat nice in the context of Salem sets; level sets [65] [46] and images of compact sets [41] for typical Wiener processes are Salem [28][27][29]. However, the graph of a typical Wiener process produces a non-example. A typical graph has Hausdorff dimension 3/2 [64], but the graph of any continuous function from the real line to itself can be shown to have Fourier dimension bounded from above by one [19].

Another non-example is a line in $\mathbb{R}^2$. By considering frequencies proportional to the normal vector for the line, it can be shown that a measure on the line can never have Fourier decay. On the other hand, we have the following theorem for surfaces with curvature, and it justifies why Salem sets are sometimes referred to as Round sets [39].

Theorem 1.1.5. *Consider a smooth hypersurface $S \subset \mathbb{R}^n$. Let σ be the surface measure on S. Let $U \subset \mathbb{R}^{n-1}$ be an open and bounded set, and let $\zeta : \mathbb{R}^n \to \mathbb{R}$ be a smooth function with support contained in $U \times \mathbb{R}$. For some point $p \in S$, assume that $\operatorname{spt} \zeta \cap S$ is a graph of a smooth function $\varphi : T_pS \to \mathbb{R}$. Assume that S has non-zero curvature everywhere. Then measures μ satisfying $d\mu := \zeta d\sigma$ have polynomial Fourier decay with rate $(n-1)/2$.*

So the Salem property is not typical for sets, but it certainly is desirable; this is because there are plenty of nice properties of Salem sets, for example when dealing with sumsets. If two sets $A, B \subset \mathbb{R}^n$ are Salem, then there exist two measures μ_A and μ_B on A and B respectively which eventually have polynomial Fourier decay of order $\dim A/2$ and $\dim B/2$ respectively. As a result, by the convolution formula for measures, $A+B$ is also Salem. Fourier dimension does have its shortcomings however. As shown by Ekström–Persson–Schmeling [18], the classical definition of Fourier dimension is not countably stable. In their paper, they show that the definition of Fourier dimension can be modified to make it countably stable by only consider measures that give non-zero mass to the set A. This is why we include the non-zero measure condition in the definition of Fourier dimension, which is not included in the classical definition.

1.1.3 Diophantine Approximation

Measures invariant under the Gauss map are of particular interest because they support subsets of the set of badly approximable numbers [26]. You can think of them as a complimentary notion to well-approximability. That being said, we always have

Dirichlet's theorem for approximable numbers, so these numbers can only be so bad [26].

Definition 1.1.6. *We say that an irrational number x is badly approximable if there exists a constant $c > 0$ such that for all proper rationals $p/q \in \mathbb{Q}$ we have that*

$$\left|x - \frac{p}{q}\right| > \frac{c}{q^2}.$$

We shall define $\mathcal{B} \subset (0,1)$ to be the set of all badly approximables.

Equivalenty, a number is badly approximable if all its continued fraction entries are bounded. So if we let $a_j(x)$ be the j-th continued fraction entry for x, and define $\mathcal{B}_N := \{x \in (0,1) \setminus \mathbb{Q} : \forall j \in \mathbb{N}, a_j(x) \leq N\}$, we can say that

$$\mathcal{B} = \bigcup_{N \in \mathbb{N}} \mathcal{B}_N.$$

The set of badly approximable numbers is a self-conformal set with respect to the Gauss map, and hence has a fractal structure. Jarník proved that $\dim_H \mathcal{B}_N \to 1$ as $N \to \infty$ [25]. We can proceed by asking a long standing problem in the world of number theory [26]: is the set of badly approximables Salem? This turns out to be a very complicated problem, and not many mathematicians in the field would conjecture with confidence that $\mathcal{B}$ is Salem or not. That being said, there is plenty of research which slowly approaches the idea that $\mathcal{B}$ is Salem.

Kaufman proved that for $N \geq 3$ there is a measure on $\mathcal{B}_N$ with polynomial Fourier decay [30]. This proof used a probabilistically constructed measure which allowed for the use of a law of large numbers to control continuants q for the Gauss map (defined more formally in Chapter 4). Such a measure is referred to as a Kaufman measure, which you can further think of as being an invariant measure with respect to T^n, that is the Gauss map to some power n. Queffélec and Ramaré used the ideas of Kaufman along with some more careful continuant analysis to prove that Kaufman's work can be extended to $N = 2$ [50]. So is this decay a property of the measures constructed, or statistically defined measures with respect to the Gauss map in general? Much progress was made towards this question in the last decade, notably beginning with Jordan–Sahlsten in 2015 [26].

1.1.4 Gibbs measures for the Gauss map

The following theorem is a precursor to all of the main theorems which we shall explore.

Theorem 1.1.7 (Jordan–Sahlsten, 2015 [26]). *Consider a Gibbs (statistical) measure μ with respect to the Gauss map T with finite Lyapunov exponent. Assume that* $\dim_H \mu > 1/2$*, and that there exists a $\delta > 0$ such that*

$$\mu(\{x \in (0,1) : x < 1/n\}) = O(n^{-\delta})$$

which we refer to as the tail condition. Then we have that μ has polynomial Fourier decay.

The tail condition here can be thought of as making sure that extreme events are unlikely, so that nice statistical ergodic theory can be used. The measure dimension condition is far less natural, and occurs because of an multiplicative approximation error when turning the Fourier transform as an integral with respect to μ into an integral with respect to Lebesgue measure. Jordan–Sahlsten use large deviation theory to control continuants, which you could think of as a dynamical analogue of Kaufman's use of a law of large numbers. Further, large deviation theory works for all Gibbs measures, and not just Kaufman measures.

It turns out that the main theorem of Jordan–Sahlsten solves a long standing problem posed by Salem. This problem concerns the Fourier–Stieltjes measure defined with respect to the Minkowski Question Mark function $? : (0,1) \to \mathbb{R}$ defined by

$$?(x) := 2\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{2^{a_1(x)+\ldots+a_n(x)}}.$$

The corresponding Fourier–Stieltjes measure can be defined by extending the condition $\mu_?([x,y]) :=?(x)-?(y)$ for intervals $[x,y] \subset (0,1)$. This measure is equivalently a Bernoulli measure invariant under the Gauss map with weights defined for each $n \in \mathbb{N}$ by

$$\mu_?\Big(\Big[\frac{1}{n+1}, \frac{1}{n}\Big]\Big) = 2^{-n}.$$

Salem proved that there exists an $\eta > 0$ such that

$$\frac{1}{2n+1}\sum_{k=-n}^{n} |\widehat{\mu_?}(k)| = O(n^{-\eta})$$

which lead Salem to conjecture that μ has Fourier decay. Persson pointed out that Jordan–Sahlsten solved Salem's problem by showing that $\dim_H \mu_? > 1/2$ and that the tail condition holds, and further this gave polynomial decay rate [47].

It is worth noting that the $\dim_H \mu > 1/2$ assumption was believed to be unnecessary. A result of Hochman–Shmerkin proved that if you consider some finite set of at least two elements $\mathcal{A} \subset \mathbb{N}$ and

$$\mathcal{B}_{\mathcal{A}} := \{x \in (0,1) \setminus \mathbb{Q} : \forall j \in \mathbb{N}, a_j(x) \in \mathcal{A}\}$$

then any Gibbs measure for the Gauss map which is supported on $\mathcal{B}_{\mathcal{A}}$ is normal (that is, n-normal for every $n \in \mathbb{N}$). Normality of a measure is weaker than polynomial Fourier decay, but Hochman–Shmerkin's result works for measures with non-zero dimension, so their work supports the idea that Jordan–Sahlsten's dimension assumption could potentially be removed. The Gauss map assumption however is very important, and its non-linearity is crucial in being able to prove the results of Jordan–Sahlsten and Hochman–Shmerkin.

1.1.5 Fuchsian groups

In Bourgain–Dyatlov [8], they prove a polynomial Fourier decay theorem for a class of measures defined on the limit set of convex cocompact Fuchsian groups. We begin by defining a generalised Fourier transform [8].

Definition 1.1.8. *Consider functions $\Phi \in C^2(\mathbb{R},\mathbb{R})$ and $g \in C^1(\mathbb{R},\mathbb{C})$. We define the generalised Fourier transform of a measure μ to be*

$$\widehat{\mu}_{\Phi,g}(\xi) = \int e^{i\xi\Phi(x)} g(x)\, d\mu.$$

So here we have the usual Fourier transform of a measure, but with a nonlinear *phase function* Φ and a varying weight function g. This generalisation is very useful, especially when considering wave functions (see subsection 1.1.6). Bourgain and Dyatlov prove the following result for Patterson–Sullivan measures.

Theorem 1.1.9 (Bourgain–Dyatlov, 2017 [8])**.** *Consider a hyperbolic surface $\mathbb{H}/\Gamma$ defined by a convex cocompact Fuchsian group. Let $\delta > 0$ be the dimension of the limit set Λ_Γ. Let μ be the Patterson–Sullivan measure on this limit set. Assume that there* *exists some $C_{\Phi,g} > 0$ such that the functions $\Phi \in C^2(\mathbb{R},\mathbb{R})$ and $g \in C^1(\mathbb{R},\mathbb{C})$ given in Definition 1.1.8 satisfy*

$$||\Phi||_{C^2} + ||g||_{C^1} \leq C_{\Phi,g} \quad \text{and} \quad \inf_X |\Phi'| \geq C_{\Phi,g}^{-1}.$$

Then there exists a $\varepsilon_\delta > 0$ and a $C_{\Gamma,\Phi,g} > 0$ such that for all $\xi \in \mathbb{R}$ such that $|\xi| > 1$,

$$|\widehat{\mu}_{\Phi,g}(\xi)| \leq C_{\Gamma,\Phi,g}|\xi|^{-\varepsilon_\delta}.$$

Such a result was possible because of the nonlinearity of the Möbius transformations that form Γ [8]. This strongly suggested that a similar result should be true for measures invariant under nonlinear maps (noting that Patterson–Sullivan measures are also equivariant under their corresponding Fuchsian group) as suggested in [8]. In Bourgain–Dyatlov [8], they use theory on Schottky structures to be able to write the limit set Λ_Γ as a self-conformal set

$$\Lambda_\Gamma := \bigcap_{n\in\mathbb{N}} \bigcup_{\mathbf{a}\in\mathcal{W}_n} I_{\mathbf{a}}$$

where I_a are construction intervals defined by the Schottky structure consisting of Möbius transformations γ_a for $a \in \mathcal{A}$, and $\mathcal{W}_n$ is the set of words of length n using an alphabet $\mathcal{A}$ (without any inverse cancellations). So you can think of the Schottky structure as a set of transformations which contract the construction intervals to the limit set, in a similar way to an iterated function system. Furthermore, for some generation n, a reduced set of words $Z(\tau)$ can be defined whose construction intervals sufficiently cover the limit set, whilst maintaining that for all words $\mathbf{a} \in Z(\tau)$,

$$|I_{\mathbf{a}}| \leq \tau < |I_{\mathbf{a}'}|$$

where $\mathbf{a}'$ is the word $\mathbf{a}$ with the last letter deleted. They prove more important results about words in $Z(\tau)$, including that there exist $C > 0$ such that

$$C^{-1}|I_{\mathbf{a}}|^\delta \leq \mu(I_{\mathbf{a}}) \leq C|I_{\mathbf{a}}|^\delta, \; C^{-1}|I_{\mathbf{a}}| \leq |I_{\bar{\mathbf{a}}}| \leq C|I_{\mathbf{a}}|$$

where $\bar{\mathbf{a}}$ is the inverse of the word $\mathbf{a}$ in the sense that $\gamma_{\bar{\mathbf{a}}} = \gamma_{\mathbf{a}}^{-1}$. The first bounds are down to the strong properties of the Patterson–Sullivan measure. You can almost think of this measure as a Hausdorff measure of dimension equal to the dimension of the limit set. The second bounds we call 'control of regular words under inversion', which turns out to be a desired property when a tree structure exists like in the Schottky structure case. This is because a lot of information about a derivative $\gamma'_{\mathbf{a}}$ can be retrieved by the point $\gamma_{\mathbf{a}}^{-1}(\infty)$, which is equal to $\gamma_{\bar{\mathbf{a}}}(\infty)$. You can think of this as the point where the transformation $\gamma_{\mathbf{a}}$ distorts the boundary of the hyperbolic space. It is also the sole reason why the transformation is nonlinear; without it, the transformation would just be a translation.

The aim of Bourgain–Dyatlov was to be able to use the exponential sum theory of Bourgain to prove their desired result. To do so, they reduce to a combinatorial bound on words, which basically says that for a given generation n, the difference in derivatives of transformations $\gamma_{\mathbf{a}}$ for $\mathbf{a} \in Z(\tau)$ are sufficiently different. This condition is equivalent to the condition that a basic counting measure on derivative values is Ahlfors–David regular [39], which basically says that the measure of a ball can be controlled by its diameter to some power. Bourgain–Dyatlov prove this property using the tree structure of the Schottky group, and the control of words under inversion/inverting. The inverting words condition in particular is a very strong property which is almost unique to Möbius transformations. However, in a remark, Bourgain–Dyatlov state that their methods can be used to prove a Fourier decay result when considering statistical measures for the Gauss map. The key question is how this result could be obtained without this inverting property. It turned out that the analogous property would be reversing, to be defined later.

1.1.6 Fractal Uncertainty

The Heisenberg Uncertainty principle is familiar to all people in the field Quantum mechanics. The principle basically states that you cannot easily measure position and momentum of a free particle at the same time. In particular, if σ_x and σ_v are the standard deviations of position and momentum respectively, then we have that

$$\sigma_x \sigma_v \geq h/4\pi$$

where h is the Planck constant. Many mathematicians and physicists will tell you that this is one of the most important statements in Quantum theory. You can consider this inequality as an application of the following more general statement for Fourier transforms of functions in Harmonic analysis [62].

Theorem 1.1.10. *For $f \in L^2(\mathbb{R})$ we have that for any position $t_0 \in \mathbb{R}$ and frequency $\xi_0 \in \mathbb{R}$,*

$$\int (t-t_0)^2 |f(t)|^2 \, dt \cdot \int (\xi - \xi_0)^2 |\widehat{f}(\xi)|^2 \, d\xi \geq ||f||_2^4 / 16\pi^2.$$

So this theorem says that subject to the variance of a function at a point and the variance of its Fourier transform at some frequency, one must be large with respect to the other. Here we will consider "Fractal Uncertainty", so how could we define this?

The first important thing to note is that the Heisenberg principle and Theorem 1.1.10 are subject to the variance of a function and its transform around a pair (t_0, ξ_0). When looking at Fractal Uncertainty principles, we will consider them about many points, in particular locally to fractal sets. Hence we are no longer interested in a statement about a pair (t_0, ξ_0) and the behaviour of a function and its transform about this point. In our case, we will require a more global quantity to observe that will be able to simultaneously consider the functions at all points local to a fractal set. Hence instead of variance about a point/mean, we will just look at the square integral of these functions, namely

$$\int |\widehat{f}(\xi)|^2 \, d\mu(\xi) \cdot \int |f(x)|^2 \, d\mu(x)$$

and the behaviour local to fractal sets. Note that we need not use the same measure in these two integrals, but we will do so here to smoothly transition from Theorem 1.1.10. A Fractal uncertainty principle will tell us that a function cannot be localised close to a fractal set in both position and frequency [14]. So how would we translate such a statement into an equation? We can first restrict our attention to functions supported on the neighbourhood of a fractal set X. Let $0 < h \ll 1$ and let X_h be the h-neighbourhood of a fractal X. So our aim is to construct a mathematical statement that says "for two fractals X, Y, if a function f is supported on X_h, then $\widehat{f}$ does not have much of the L^2-mass of f on Y_h". We can formalise this statement using norms, and we will quantify 'not much' using h. We first define our Fourier transform in this setting [8].

Definition 1.1.11. *For fixed $h \in (0,1)$, define the Fourier transform operator $\mathcal{F}_h : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ by*

$$\mathcal{F}_h(\xi) = \frac{1}{\sqrt{2\pi h}} \int e^{-ix\xi/h} f(x) \, d\mu(x).$$

So this is the generalised Fourier transform with a fixed weight and phase function determined by the Planck constant. This operator is used to recover the wave function of momentum from the position of a quantum particle. We define our uncertainty principle as follows [8].

Definition 1.1.12. *Consider two fractal sets $X, Y \subset \mathbb{R}$. Let X_h and Y_h be their h-neighbourhoods respectively. We say that X and Y satisfy a Fractal Uncertainty principle if there exists a $\beta > 0$ such that*

$$||\mathbb{1}_{Y_h}\mathcal{F}_h\mathbb{1}_{X_h}||_{L^2(\mathbb{R})\to L^2(\mathbb{R})} \leq h^\beta.$$

Using the definition of the norm of an operator in L^2, we see that this definition seems to formalise the statement given earlier about localisation. So we have some sort of generalisation of Heisenberg to fractal sets. An important question is, can we work back from such a statement to recover Heisenberg? This is indeed the case (see [14]). To begin, we can reduce to the case when a particle's expected position and momentum are zero. So in this situation, our fractal is $\{0\}$, and $\{0\}_h = [0, h]$ is the h neighbourhood of our fractal. To model position of a quantum particle, wave functions are used. A wave function describes the quantum state of a particle by giving a probability distribution for its position. Given a wave function $f(x, t)$, for position x and time t, its square will be our probability distribution. So firstly by definition of a probability function,

$$\int_{-\infty}^{\infty} f^2\,dx = 1$$

since there is a 100% probability that the particle is somewhere at any time t. So we define the probability that the particle is at a position in the set $[a, b]$ at time t by

$$\mathbb{P}^x_{[a,b]}(t) = \int_{[a,b]} |f(x,t)|^2\,dt.$$

Momentum of a particle is similarly modelled using the probability distribution given by the Fourier transform of f. So we define the probability that the particle is moving with momentum in $[a, b]$ by

$$\mathbb{P}^m_{[a,b]}(t) = \int_{[a,b]} |\mathcal{F}_h f(\xi)|^2\,d\xi.$$

So in the setting of Definition 1.1.12, $X_h = [0, h]$ and $Y_h = [0, h]$. We will begin by assuming that f and its Fourier transform are both very large on $[0, h]$, and for heuristic proof, completely supported on $[0,h]$. This corresponds to the standard deviations of position and momentum being large, hence we are assuming that Heisenberg is false when given the fractal uncertainty principle with $\beta = 1/2$ (for a contradiction). So we have that

$$\mathbb{P}^x_{[0,h]}(t) = 1 = \mathbb{P}^m_{[0,h]}(t)$$

as our assumption. Hence by Definition 1.1.12,

$$1 = \Big(\int_{[0,h]} |\mathcal{F}_h(f|_{[0,h]})(\xi)|^2 \, d\xi\Big)^2 \leq h||f|_{[0,h]}||_2^2 = h.$$

So $1 \leq h$ (up to some heuristic approximation) for a contradiction. Hence the fractal uncertainty principle heuristically agrees with Heisenberg for $\beta = 1/2$. For a complete proof, see [14, Section 2.1].

There are two major methods for proving Fractal Uncertainty principles. One of these methods is by Fourier decay of $\mathbb{1}_{X_h}$. This lead to the Uncertainty principle of Bourgain–Dyatlov on the limit set of a convex cocompact Fuchsian group [8, Proposition 4.1]. Another method is by the additive energy of $\mathbb{1}_{Y_h}$, which we do not discuss here, but refer the reader to [14].

1.1.7 Linear map invariance

The following theorem of Mosquera–Shmerkin gives us a basis to study the case of linear Markov map invariant measures.

Theorem 1.1.13 (Mosquera–Shmerkin, 2018 [42])**.** *Consider $F \in C^2(\mathbb{R}, \mathbb{R})$ with positive second derivative. Let μ be a homogenous positive-entropy Bernoulli measure with respect to a self-similar IFS. Then the measure $F\mu$ has polynomial Fourier decay.*

In a later section we show that this corresponds to a class of *not* totally nonlinear maps (section 5.5). More specifically, we get Fourier decay when the dynamics T is such that there exists some Lipschitz function $g : X \to \mathbb{R}$ and constant function $c > 0$ with

$$\log |T'| = c - g \circ T + g.$$

Mosquera–Shmerkin focus on homogenous self-similar measures, so this constant function cannot be any locally constant function. The general locally constant case would correspond to studying piecewise linear Markov maps whose branches have different gradients. There has been much progress towards this case, especially in 2019 with the works of Li–Sahlsten and Solomyak. In the one dimensional setting, Solomyak proved the following promising result.

Theorem 1.1.14 (Solomyak, 2019 [60]). *Let μ be a self-similar Bernoulli measure with non-zero Bernoulli weights for an IFS with contractions given by the entries of a vector $\lambda \in (0,1)^m$. Then μ has polynomial Fourier decay for all $\lambda \notin \mathcal{E}$, where $\mathcal{E}$ is a set of Hausdorff dimension zero.*

This theorem suggests that in the linear Markov map invariant case, we should be able to obtain polynomial Fourier decay most of the time. This theorem does not give specific examples, and in many known examples the decay rate is weaker. The following examples exhibit logarithmic decay.

Theorem 1.1.15 (Li–Sahlsten, 2019 [38]). *Consider a self-similar non-atom measure μ on $\mathbb{R}$ whose dynamics has two distinct contracting branches f_i and f_j whose gradients r_i and r_j are such that*

$$R_{i,j} = \frac{\log r_i}{\log r_j} \notin \mathbb{Q}.$$

Then μ has Fourier decay. If we can further say that some $R_{i,j}$ is badly approximable, we get logarithmic Fourier decay.

An immediate question is whether this decay rate can be improved to polynomial decay, and to more general examples. Li–Sahlsten made further progress in a higher dimensional analogue with some algebraic conditions arising from renewal theory.

Theorem 1.1.16 (Li–Sahlsten, 2019 [37]). *Let μ be a self-affine measure on $\mathbb{R}^d$ for $d \geq 2$ with positive entropy and associated contractions $f_j = A_j + b_j$. If $\Gamma := \langle A_j : j \in \mathcal{A}\rangle < GL(d,\mathbb{R})$ and Γ is proximal and totally irreducible, then μ has Fourier decay.*

Li–Sahlsten believe that the proximal condition is unnecessary for Fourier decay, the definition for which is in their paper. Total irreducibility is the main condition, which requires that Γ does not fix a finite union of proper subspaces of $\mathbb{R}^d$. They further prove that given some topological conditions on Γ, they can get polynomial Fourier decay. They conjecture that the topological and proximal conditions are not required for μ to have polynomial Fourier decay. Proving this conjecture would require improving their tools from renewal theory.

Recent work of Rapaport [52] gives necessary and sufficient conditions for a Bernoulli measure for an affinely irreducible IFS to be non-Rajchman. Here affinely irreducible means that the IFS doesn't fix an affine subspace. This theorem also uses a renewal theoretic proof idea first coined by Li.

1.1.8 Nonlinear map invariance

Of course we have already discussed many examples of Fourier decay of measures defined using nonlinear maps, such as the Gauss map and Fuchsian groups. Much of this theory is very actively researched, meaning that many more general theorems have been made possible. Three such theorems will be our focus later on, but here we will discuss a very recent paper which is strongly related to the main general nonlinear map theorem to presented.

Theorem 1.1.17 (Algom–Hertz–Wang [1], 2020)**.** *For an interval $J \subset \mathbb{R}$ and some $\gamma > 0$, let*

$$\Phi = \{f_1, \dots, f_n \mid f_i : J \to \mathbb{R} \; \forall i = 1, \dots, n\}$$

be a $C^{1+\gamma}(J)$ IFS which is uniformly contracting and never has zero derivative. Assume that for every $t, r \in \mathbb{R}$ the set

$$\{-\log |f'(y)| : y \text{ is the fixed point of } f \in \Phi\}$$

is not contained in the set $t + r\mathbb{Z}$.

Then we have that any non-atomic self conformal measure with respect to Φ has Fourier decay.

How they prove their result seems to have appealing links with much of the literature we have discussed. Their method involves a stopping time argument as in Li–Sahlsten [38] to be able to control the derivatives of the IFS under iteration. To extract a 'regular' part of the measure, they use a Central Limit theorem which seems to be analogous to the use of a large deviation theorem as in Jordan–Sahlsten [26]. It may be more directly relatable to Kaufman's [30] use of a law of large numbers (as discussed in subsection 1.1.3). Their aim is to be able to use a Fourier decay type lemma of Hochman [21], which itself is proved using integration by parts (where again we can see a parallel to Kaufman's methods [30]).

This theorem of Algom–Hertz–Wang will turn out to be very similar to a general theorem on Markov maps to be proved later (see Theorem 1.3.4 ahead). The uniform contraction, non-zero derivative, and non-lattice derivative conditions are analogous to uniform expansion, non-infinite derivative, and total nonlinearity conditions to be explored in the Markov map settings. There are three major differences between the two theorems. Firstly, the main theorem of Algom–Hertz–Wang does not assume any separation conditions on the IFS (in contrast with a very strong separation condition in the main theorem to be proved). One can expect this because the theorem is a result of the IFS being nonlinear, not the position of its construction intervals. On the other hand, their methods do not present a decay rate, unlike our main theorem to be presented which gives (eventual) polynomial Fourier decay. The final major difference is methods used which seem very independent of each other; it would be an interesting prospect to compare the ideas used.

1.2 Proving Fourier decay

1.2.1 Transfer Operators

Thermodynamical formalism lies at the heart of so much modern work in the theory of (statistical) nonlinear-invariant measures. In the paper of Jordan–Sahlsten, the theory of transfer operators are heavily used to be able to make use of the large deviation theory which is central to achieving their proof. We will define them in the same way as Jordan–Sahlsten [26].

Definition 1.2.1. *Given a dynamical system $T : X \to X$ and a (real) potential φ, we define the transfer operator on $C(X)$ by*

$$\mathcal{L}_\varphi f(x) := \sum_{y \in T^{-1}x} e^{\varphi(y)} f(y).$$

This operator is intended to extend the core theory of finite linear dynamical systems, which use matrices, to infinite/nonlinear dynamics. A fundamental example of such core theory is apparent in the Perron–Frobenius theorem (see Theorem 6.3.2), which can by generalised to transfer operators as follows [26, Proposition 3.10].

Theorem 1.2.2 (Ruelle–Perron–Frobenius)**.** *There exists a positive eigenfunction $w \in C(X)$ with eigenvalue $\lambda > 0$ for the transfer operator $\mathcal{L}_\varphi$.*

Furthermore, there exists a Gibbs measure μ on X such that for all $f \in C(X)$ we have that

$$\int (\mathcal{L}_\varphi f)\, d\mu = \lambda \int f\, d\mu.$$

Transfer operators are a way of studying the fine properties of integrals under invariant measures. By studying equilibrium states for Gibbs measures, we get that the Ruelle–Perron–Frobenius theory holds for $\lambda = 1$ for the measures that we will consider (under some normalisation [26]). So by using the invariance of the transfer operator under integration, we have a way to separate integrals using the dynamics. In the case of expanding Markov maps, we can essentially split integrals into a sum of integrals on construction intervals which are defined by the dynamics. This is a widely used idea in the theory of symbolic dynamics. By iterating the transfer operator, we get the same result when considering n-th generation construction intervals, so we can make use of the combinatorial theory of large deviations. Furthermore, we can iterate $\mathcal{L}_\varphi^n$ to be able to consider integrals on blocks of length n words, which allows us to impose the necessary conditions for Bourgain–Dyatlov's [8] exponential sum theory, which we will use to prove Fourier decay results. Moving forward to proving the main (nonlinearity) assumption of this exponential sum theory in the context of dynamical systems, it becomes clear that a complex analogue of transfer operators could be advantageous (to be discussed in subsection 1.2.5).

1.2.2 Large Deviations

Birkhoff's Ergodic Theorem [66] says that we expect the average of a function over orbits to be dictated by the mean when we consider an ergodic system. Consider a bounded space $X \subset \mathbb{R}$ with an expanding dynamical system $T : X \to X$. Assume that T has branches which can be indexed by an alphabet $\mathcal{A}$. Let $\mathbf{a} := a_1 a_2 \dots a_n a_{n+1} \dots \in \mathcal{A}^{\mathbb{N}}$ and $\mathbf{a}|_n := a_1 \dots a_n$ be the first n entries of $\mathbf{a}$ and $f_{\mathbf{a}|_n} := f_{a_1} \circ \dots \circ f_{a_n}$ be the corresponding inverse branch for T^n. Further assume that T is conjugate to the shift map σ on the infinite word space $\mathcal{A}^{\mathbb{N}}$. This will mean that an infinite word $\mathbf{a} \in \mathcal{A}^{\mathbb{N}}$ will give us a point $y \in X$ in the limit set of T where

$$y := \lim_{n\to\infty} f_{\mathbf{a}|_n}(x_{a_{n+1}})$$

where $x_{a_{n+1}}$ is defined to be the centre point of the construction interval $I_{a_{n+1}} := f_{a_{n+1}}(X)$. Given the point $y \in X$, we will define $y_{n+1} := T^n y \in X$ so that $y = f_{\mathbf{a}|_n}(y_{n+1})$.

Let μ be a Gibbs measure with respect to T. When considering the potential function $\psi := -\log|T'|$, let λ be its mean with respect to μ. So Birkhoff's Ergodic theorem tells us that there exists some set D of measure zero such that for all $y = f_{\mathbf{a}|_n}(y_{n+1}) \in X \setminus D$ we have that

$$\lim_{n\to\infty} \frac{1}{n} S_n\psi(y) = \lim_{n\to\infty} \frac{1}{n} \log|f'_{\mathbf{a}|_n}(y_{n+1})| = -\lambda$$

where we use the inverse theorem for differentiable functions and the chain rule to get the first equality. Breaking down the limiting statement, we can say that for μ almost every $y \in X$ and for every word $\mathbf{a}|_n \in \mathcal{A}^n$ which is the initial coding of such a point y, for every $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that for all $n > N$,

$$|\log f'_{\mathbf{a}|_n}(y_{n+1}) + \lambda n| < \varepsilon n.$$

We can get a very similar statement for any point $x \in X$ rather than y_{n+1} by assuming that $f_{\mathbf{a}|_n}$ (or T itself) is sufficiently smooth.

Large deviation theory analyses the set of points y where we deviate from this averaging statement, and how the set of deviating points behave in measure as N increases (as we make ε smaller). Call this deviating set $X \setminus R_n(\varepsilon)$, the compliment of the non-deviating (regular) set with respect to ε and n. The aim of large deviation theory is to say something like that there exists some $\delta > 0$ and $C > 1$ such that

$$\mu(X \setminus R_n(\varepsilon)) \leq e^{-n\delta}.$$

so that we can ignore the deviating set when dealing with expressions such as the Fourier transform. In the countable alphabet case, this will also mean we only have to consider finitely many words in $\mathbb{N}^n$ when studying nth level construction intervals. The following theorem of Jordan–Sahlsten proves a Large deviation theorem for (eventually) expanding Markov maps.

Theorem 1.2.3. *Consider a Hölder continuous observable f and suppose that the set*

$$\{t \in \mathbb{R} : P(tf + \varphi) < \infty\}$$

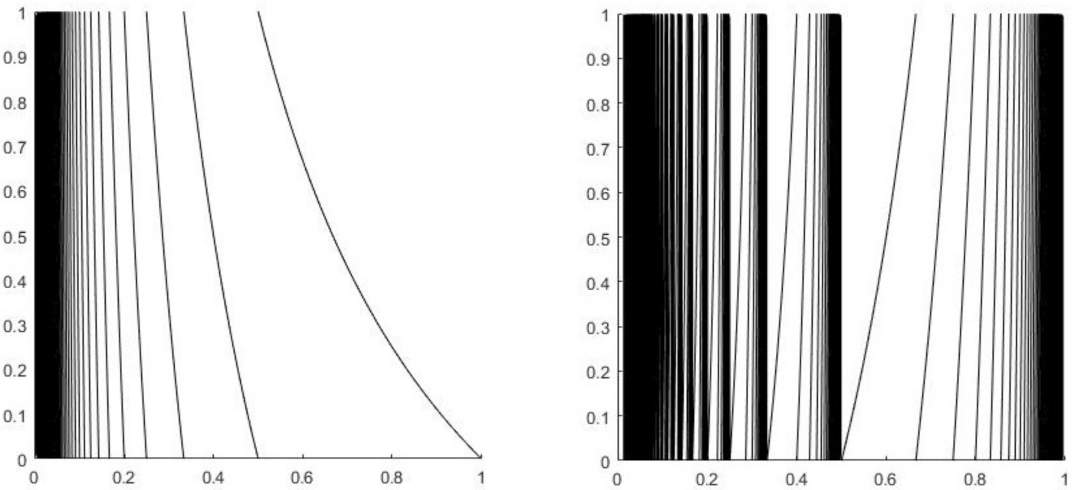

Figure 1.1: The first and second iterations of the Gauss map (pictures created in MATLAB).

contains a neighbourhood of zero. Then for any $\varepsilon > 0$, there exists a $\delta > 0$ and an $n_1 \in \mathbb{N}$ both depending on ε such that for $n \geq n_1$ we have that

$$\mu_\varphi\Big(\Big\{x \in X : \Big|S_n f(x) - n\int f\, d\mu\Big| > n\varepsilon\Big\}\Big) = O(e^{-n\delta}).$$

Jordan–Sahlsten bound the deviating sets using pressure [26]. They then bound pressure using its properties such as analyticity and convexity. In the countable alphabet case as studied in Jordan–Sahlsten [26], the pressure can be infinite (unlike the finite alphabet case). In this situation, the alphabet can be truncated (made finite), and then the behaviour of the pressure can be analysed when the alphabet approaches its original countable state before truncation. In any case, the tools of Bourgain–Dyatlov [8] and large deviation theory are central to the proofs of the main results to be presented.

1.2.3 Renewal

We will look at the renewal theory in the context of the work of Jialun Li on stationary measures [33]. Some technical details will be omitted, because we will mainly be interested in the parallels to the dynamical context here.

For $m \geq 1$ we will consider the action of the matrix group $G = SL_{m+1}(\mathbb{R})$ on the projective space of lines $\mathbb{PR}^{m+1}$. We can define a random walk on the projective space using G by defining a measure on it. We consider a Borel measure μ on G and let Γ_μ be the subgroup generated by the support of μ. By considering Γ_μ rather than G, we can ensure that μ doesn't trap random walks in parts of the projective space. This is important if we want to consider some notion of 'mixing'. To get mixing, we assume that Γ_μ is a (Zariski) dense subgroup of G. So given a point $x \in \mathbb{PR}^{m+1}$, we can define a walk $x, g_1x, g_2g_1x, g_3g_2g_1x, \ldots$ for $g_i \in \Gamma_\mu$ by selecting each g_i randomly with respect to the measure μ. We then get a stationary measure ν on $\mathbb{PR}^{m+1}$ satisfying the μ-stationary equation

$$\nu = \mu * \nu := \int_{\Gamma_\mu} g_* \nu \, d\mu(g)$$

where $g_*\nu$ is the standard push-forward measure [33]. Stationary measures are the first examples Li studied when proving Fourier decay using renewal theory, initially for $m = 1$.

Theorem 1.2.4 (Li, 2017 [33])**.** *Consider a Borel probability measure μ on Γ_μ which we assume to be (Zariski) dense. Assume further μ has finite exponential moment. Then we have that μ-stationary measures have Fourier decay.*

The finite exponential moment condition is analogous to the finite pressure condition of Theorem 1.2.3. This is not surprising, because the random walk theorem of Li [33, Proposition 2.8] controls the random elements of G in a similar way as Jordan–Sahlsten control the inverse branches for the dynamics using their large deviation theory. A higher dimensional analogue of Theorem 1.2.4 was proven by Li [35] by proving a higher dimensional sum-product theorem [34] and employing the exponential sum theory of Bourgain–Dyatlov [8]. Note that polynomial Fourier decay was achieved using this method. We will now consider the renewal theoretic methods in the context of Li–Sahlsten [38], because they are slightly simpler.

Consider some scaling constant $\tau > 0$. Consider words denoted by $\mathbf{a}$ and let $\mathbf{a}'$ be the word $\mathbf{a}$ with the last letter deleted. Let's say that we want to consider a set of words $Z(\tau)$ such that

$$|I_{\mathbf{a}}| \leq \tau < |I_{\mathbf{a}'}|$$

where $I_{\mathbf{a}}$ is the image of an inverse branch $f_{\mathbf{a}}$. Then $Z(\tau)$ gives a partition of the space X using construction intervals. In Bourgain–Dyatlov's [8] case of Patterson–Sullivan measures of dimension $\delta > 0$, they get that the measure of $I_{\mathbf{a}}$ can be controlled by τ^δ due to the strong geometric properties of the measures. Li–Sahlsten use a slightly weaker control for Bernoulli measures.

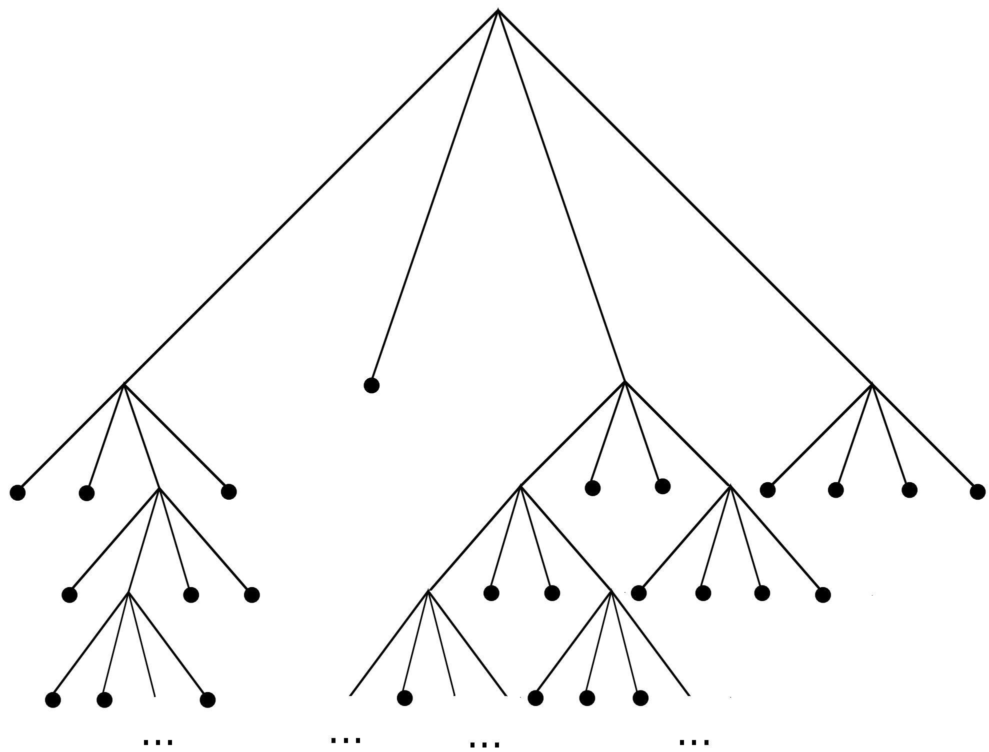

Figure 1.2: A visual representation of $Z(\tau)$ for some $\tau > 0$ with an alphabet of size four, hence four branches going down from each node. Each end of the branches represents some word with length equal to its stage down the tree (with the top of the tree being the empty word). Each black dot is where the stopping time kicks in; going further, the corresponding construction interval would be too small.

In the work of Li–Sahlsten [38], it is shown that the Fourier transform of Bernoulli measures can be written as the expectation of a sum over a random walk. This then allows the application of renewal theory to bound the Fourier transform. Li–Naud–Pan [36] on the other hand use renewal theory to prove the main (nonlinearity) assumption of Bourgain's sum-product theory [8, Lemma 2.16]. This allows the use of Bourgain–Dyatlov's exponential sum theory.

One can ask whether there is a dynamical analogue of renewal theory which could be used to prove results like Jordan–Sahlsten. Complex transfer operators theory is an example of such an analogue, as pointed out to Sahlsten by Li in conversations with Frédéric Naud (also see [33, Remark 1.10]).

1.2.4 Exponential Sum Theory

The exponential sum theory of Bourgain–Dyatlov [8] will be the key that unlocks all of the main results to be proven in the later chapters. In this section we will explore a few of these exponential sum theorems; in particular those which have applications to dynamical systems. When considering Fourier transforms, the study of exponential sum theory is easily motivated when you consider approximating the integral with a sum. This will be the philosophy of most of the examples to be presented, although the details of this approximation can be quite intricate to fit its purpose (for example, to achieve a strong rate of Fourier decay). We begin by first looking at possibly the most famous exponential sum theorem [66].

Theorem 1.2.5 (Weyl's Criterion). *A real sequence $(x_n)_{n\in\mathbb{N}}$ is uniformly distributed modulo 1 if and only if for each $l \in \mathbb{Z}\setminus\{0\}$ we have that*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} e^{2\pi i l x_j} = 0.$$

Using this criterion, one can prove that $x_n := n\alpha$ is uniformly distributed if and only if α is irrational. The following theorem was obtained using Weyl's criterion [26].

Theorem 1.2.6 (Davenport–Erdős–LeVeque). *Consider a probability measure $\mu \in \mathcal{M}([0,1])$ and let $(s_k)_{k\in\mathbb{N}}$ be a strictly increasing sequence. If for any non-zero $p \in \mathbb{R}$ we have that*

$$\sum_{N=1}^{\infty} \frac{1}{N^3} \sum_{k,m=1}^{N} \widehat{\mu}(p(s_k - s_m)) < \infty$$

then $(s_k x)_{k\in\mathbb{N}}$ is uniformly distributed modulo 1 for μ almost every $x \in [0,1]$. In particular, if μ has logarithmic Fourier decay (or better), then the finite sum condition holds.

This theorem can in particular be used to prove that if μ has logarithmic Fourier decay, then μ almost every $x \in [0,1]$ is normal (i.e. a normal number in every natural base). This theory further motivates the study of Fourier decay, and exponential sum theory itself can be used to prove strong decay theorems. The following multiplicative convolution theorem of Bourgain is not an exponential sum theorem, at least in its general form [7]. Note that the multiplicative convolution of two measures μ and ν on $\mathbb{R}$ is defined in an analogous manner to additive convolutions as follows: for $f \in C_0(\mathbb{R})$,

$$\int f\, d(\mu \otimes \nu) = \iint f(xy)\, d\mu(x)\, d\nu(y).$$

We will say that $|\xi| \sim N$ if there exists a constant $C > 0$ independent of ξ and N such that $C^{-1}N \leq |\xi| \leq CN$. The following theorem is the most essential ingredient necessary to obtain the main proofs to be presented.

Theorem 1.2.7 (Bourgain's Multiplicative Convolution Theorem [7])**.** *For all $\delta_1 > 0$, there exist $\varepsilon_3, \varepsilon_4 > 0$ and large $k \in \mathbb{N}$ such that the following holds.*

Let μ be a probability measure on $[\frac{1}{2}, 1]$ let and N be a large integer. Assume for all $N^{-1} < \sigma < N^{-\varepsilon_3}$ that

$$\max_{a\in\mathbb{R}} \mu(B(a,\sigma)) < \sigma^{\delta_1}. \tag{1.1}$$

Then for all $\xi \in \mathbb{R}$ with $|\xi| \sim N$,

$$|\widehat{\mu^{\otimes k}}(\xi)| < N^{-\varepsilon_4}. \tag{1.2}$$

To prove Fourier decay in the main results to come, we will apply this theorem to counting measures on derivatives of dynamics, making it an exponential sum theorem. Most of the work to be presented will either be attempting to generalise this exponential sum theorem, or proving its main assumptions. In particular, we will see that (1.1) can prove polynomial Fourier decay in many nonlinear dynamical settings. We will call it the non-concentration condition, or in our dynamical context, non-concentration of derivatives (of T).

Sum-Product Theory

As previously mentioned, Bourgain's theorem on multiplicative convolutions will be crucial in obtaining the main results to be presented. Theorem 1.2.7 is proved using a so called 'sum-product theorem'. In general, the sum-product phenomenon states that a finite set cannot simultaneously have additive and geometric structure. We begin by stating the sum-product theorem of Erdős–Szemerédi. Note that for a set $A \subset \mathbb{R}$ we define sumsets by

$$A + A := \{a_1 + a_2 : a_1, a_2 \in A\}$$

and product sets by

$$A \cdot A := \{a_1 a_2 : a_1, a_2 \in A\}.$$

Theorem 1.2.8. *There exist constants $c > 0$ and $\varepsilon > 0$ such that for any finite set $A \subset \mathbb{N}$ we have that*

$$\max\{|A + A|, |A \cdot A|\} \geq c|A|^{1+\varepsilon}.$$

If you assume that A has a small sumset, in particular that there exists $C > 0$ independent of $|A|$ such that

$$|A + A| \leq C|A|$$

then Freiman's theorem for reals states that A is contained in a generalised arithmetic progression. In this case, A cannot have strong geometric structure (geometric progressions), because you cannot fit many geometric lattices in arithmetic ones (think of trying to fit many/long geometric progressions in an arithmetic subset of the integers). For $A \in \mathbb{R}^+$, if we assume that $|A \cdot A| \leq C|A|$, we can use Freiman's theorem applied to the set $\log A$ to show that A must be contained in a generalised geometric progression. Again, we can say that in this case, A cannot have strong additive structure in a similar way as we reasoned earlier. The combination of these cases is the idea which the sum-product philosophy represents. Bourgain's discretized sum-product theorem gives more quantitative sum-product bounds for sets $A \subset [1/2, 1]$ such that there is an $N > 1$ such that A is N^{-1}-separated. The sum-product bound is $N^\varepsilon \cdot |A|$, where this specific bound is important to be able to get (1.2).

So why should sum-product theory suggest a multiplicative convolution theorem like Bourgain's? The idea is that the Fourier transform will be large on additive structures, and the measure will be defined so that it will be large on multiplicative structures. To explain this, let us first consider the Fourier transform on arithmetic progressions. Consider a measure which has large mass on an arithmetic progression $AP := \{a, a+\theta, \dots, a+\theta n\}$ for some $a, \theta \in \mathbb{R}$ and some integer $n > 1$. On AP, the Fourier transform of the measure at frequencies $\xi = j/\theta$ for $j \in \mathbb{Z}$ would be

$$\widehat{\mu_{AP}}(\xi) = \int_{AP} e^{2\pi i x\xi}\, d\mu(x) = \int_{a+\theta k \in AP} e^{2\pi i a j/\theta} e^{2\pi i k j}\, d\mu(a+\theta k) = e^{2\pi i a j/\theta}\mu(AP).$$

So on the arithmetic progression AP, the size of the Fourier transform is the mass of the progression. So if the measure concentrates too much on arithmetic progressions, the Fourier transform will be bounded from below by the mass of the progressions for arbitrarily large frequencies. This will mean that the arithmetic progressions will stop the Fourier transform from decaying. Bourgain's multiplicative convolution theorem will first tell us that in the case that we don't have additive structure, we can get decay of the Fourier transform.

Now we try to consider how we can make sure that the Fourier transform decays when we have no multiplicative structure. We can do so by forcing a multiplicative structure on the measure. Consider a measure μ assigning large mass to a geometric progression $\{a, a\theta, \dots, a\theta^n\}$. Define the multiplicative convolution $\mu \otimes \mu$ as before, where for an integrable complex function f we have that

$$\int f\, d\mu \otimes \mu = \int \int f(xy)\, d\mu(x) d\mu(y).$$

There will be many ways to get the same result when multiplying two terms in the geometric progression, for example if you want to multiply two terms to get $a^2\theta^n$,

$$a^2\theta^n = (a)(a\theta^n) = (a\theta)(a\theta^{n-1}) = \dots$$

So in this example, the measure $\mu \otimes \mu$ will give large mass to $a^2\theta^n$, and this is true for similar values defined using the geometric progression. By convolving the measure more times, we can concentrate the mass on points relating to the geometric progression. So this will mean that we cannot expect the Fourier transform of a many-times self-convolved measure to be small. When considering a lack of geometric structure, Bourgain's theorem tells us that integrals with respect to the multiplicative convolution will be small; how small is dictated by the sum-product bound. The reader is referred to [7, Lemma 8.43] for precise arguments using a delicate combination of the Balog–Szemerédi–Gowers theorem and Freiman's theorem.

1.2.5 Complex Transfer Operators

In Naud's paper [44], he focuses on the study of Selberg Zeta functions. They are also the primary focus of the paper of Bourgain–Dyatlov [8]. Let $\mathcal{P}$ be the set of primitive (trace out their image exactly once) closed geodesics for a given surface M, and let $l(\gamma)$ be the length of $\gamma \in \mathcal{P}$. Given a compact Riemann surface M of constant negative curvature -1, we define the Zeta function by

$$Z_M(s) := \prod_{k=0}^{\infty} \prod_{\gamma \in \mathcal{P}} (1 - e^{-(s+k)l(\gamma)}).$$

This function turns out to be very useful in finding asymptotic bounds for the number of geodesics in $\mathcal{P}$. This is done by studying non-trivial zeros of Z_M. Naud's first main theorem in this paper states that there exists an $\varepsilon > 0$ such that $Z_m(s)$ is analytic and non-vanishing on $\{\Re(s) > \delta - \varepsilon\}$ where δ is the Poincaré exponent (defined in the same way as Bourgain–Dyatlov), excluding the simple zero at $s = \delta$. This theorem is used to prove the following.

Theorem 1.2.9 (Naud, 2005 [44]). *If $N(T)$ is the number of primitive closed geodesics of length less than $T > 0$, then there exists $\alpha \in (0, \delta)$ such that*

$$N(T) = li(e^{\delta T}) + O(e^{\alpha T})$$

where $li(x) = \int_2^x \frac{dt}{\log t}$.

This result is proven using the central tool of this paper, which is on the spectral gap of complex transfer operators. Such a property can be used to prove many more results. In the paper of Araújo–Melbourne [2], they prove the exponential decay of correlations for a certain class of flows using spectral theory for complex transfer operators.

Definition 1.2.10. *Consider a flow F_t and let R be its suspension. Let μ be the (normalised) product of an R-invariant probability measure and Lebesgue measure. We define correlations as*

$$\varrho_{v,w}(t) := \int vw \circ F_t \, d\mu - \int v \, d\mu \int w \, d\mu$$

The flow is said to be mixing if the correlations decay to zero as t goes to infinity.

The paper of Araújo–Melbourne has a particularly interesting corollary on correlations. To explain, we first define robust mixing.

Definition 1.2.11. *We define a flow Φ_t to be robustly mixing if there exists an open set of functions containing Φ_t such that all functions in this set are mixing (not just Φ_t).*

Their main corollary on the Lorenz attractor flows is as follows, obtained by using their main theorem on exponential decay of correlations.

Theorem 1.2.12 (Araújo–Melbourne, 2016 [2])**.** *The Lorenz attractor is robustly exponentially mixing.*

The main assumption of this paper on the suspension dynamics is important when considering the application to more general dynamics. They assume a uniform non-integrability condition, which states that there exists some $C > 0$ such that if R_n is the nth Birkhoff sum for R and h_i are two inverse branches for the map R^n, then

$$\inf D(|R_n \circ h_1 - R_n \circ h_2|) \geq C > 0. \tag{1.3}$$

This turns out to be very similar to the main assumption of Naud's theorems. In Naud's theorem, this is used as an assumption for nonlinear dynamics in general (not just flows). In the context of dynamics on a one-dimensional space, the non-integrability condition will be the same as saying that there exists two n-generation inverse branches for the dynamics T such that there difference in derivatives is bounded below from zero. We will later show that this is related to the non-concentrated derivative assumption.

1.3 Main Results

The main aim of the work to be presented is to be able to generalise the result of Jordan–Sahlsten. Their result on polynomial Fourier decay has several assumptions, some of which we do not believe to be necessary. By eliminating some of these assumptions, we may bring our field closer to understanding what is necessary for polynomial decay of the Fourier transform of a measure. The initial progress, in joint work with Tuomas Sahlsten, proved that the main theorem of Jordan–Sahlsten without the $\dim_H \mu > 1/2$ condition.

Theorem 1.3.1 (Sahlsten–S, 2018 [53]). *Consider a Gibbs measure μ with respect to the Gauss map T with locally Hölder potential. Assume that $\dim_H \mu \neq 0$, and that there exists some $\delta > 0$ such that*

$$\mu(\{x \in (0,1) : x < 1/n\}) = O(n^{-\delta}).$$

Then we have that μ has polynomial Fourier decay.

This was done by combining the techniques of Bourgain–Dyatlov with the large deviation theory presented in Jordan–Sahlsten.

The second assumption that we want to remove is the dependence on the Gauss map. A folklore conjecture [8] is that a Gibbs measure for a nonlinear map should also have polynomial Fourier decay. This conjecture is reasonable when looking at comments made by Jordan–Sahlsten [26] and Bourgain–Dyatlov [8]. Results of Li [33] and Li–Naud–Pan [36] also reinforce such a statement. So, how would we begin to try and remove the Gauss map assumption and replace it with a nonlinear map? In [53], we prove that it is sufficient to prove that the map has "non-concentrated derivative" which we define as follows [54].

Definition 1.3.2. *Let $\tau : \mathbb{N} \to \mathbb{R}$ be a positive function that decreases exponentially and consider some $\varepsilon > 0$. Define $J_{\tau(n)}(\varepsilon) := \{\eta \in \mathbb{R} : \tau(n)^{-1/4} \leq |\eta| \leq e^{\varepsilon n}\tau(n)^{-1/2}\}$. Let $T : X \to X$ be an expanding map and μ be a Gibbs measure for T with Lyapunov exponent $\lambda > 1$ and Gibbs constant $C > 0$. For $n \in \mathbb{N}$ define $R(n,\varepsilon) := 16^2 C e^{3\varepsilon\lambda n}$.*

We say that T has non-concentrated derivative with respect to μ if there exists $c_0 > 0$, $\varepsilon_3 > 0$ and $\kappa_0 > 0$ such that for all sufficiently small $\varepsilon > 0$, all sufficiently large $n \in \mathbb{N}$, all $\eta \in J_{\tau(n)}(\varepsilon)$, $\sigma \in [R(n,\varepsilon)^{-2}|\eta|^{-1}, |\eta|^{-\varepsilon_3}]$ and all $x \in I$,

$$\sharp\{(\mathbf{a},\mathbf{b},\mathbf{c}) \in \mathcal{R}_n(\varepsilon)^3 : |e^{2\lambda n} f'_{\mathbf{ab}}(x) - e^{2\lambda n} f'_{\mathbf{ac}}(x)| \leq \sigma\} \leq e^{\kappa_0 \varepsilon n}\sigma^{c_0}\sharp\mathcal{R}_n(\varepsilon)^3.$$

This is a dynamical analogue of a condition proven by Bourgain–Dyatlov in the context of convex cocompact Fuchsian groups, and is a major key we can use to unlock polynomial Fourier decay. We prove this property for a wide class of nonlinear maps known as totally nonlinear maps [3].

Definition 1.3.3. *Consider a space $X \subset \mathbb{R}$ and a map $T : X \to \mathbb{R}$. Let Δ be a large interval containing the domain and range of T. We say that T is totally-nonlinear if* *there does not exist a C^1 diffeomorphism $g : \Delta \to \mathbb{R}$ and some locally constant function $c : X \to \mathbb{R}$ such that*

$$\log |T'| = c + g \circ T - g$$

on the set X (not necessarily the entirety of Δ).

We prove the following theorem for totally nonlinear maps, where Λ_T will be the repeller for T [51, Section 4.2].

Theorem 1.3.4 (Sahlsten–S, 2020 [54])**.** *Assume that X is a disjoint union of intervals and $T : X \to \mathbb{R}$ is real-analytic and (eventually) expanding on each of these intervals. Further assume that T is totally nonlinear and satisfies the Markov property. Then there exists a $\delta_0 > 0$ such that the following holds. Let μ be a Gibbs measure for T with locally Hölder potential and that*

$$\dim_H \Lambda_T - \delta_0 \leq \dim_H \mu \leq \dim_H \Lambda_T.$$

Then μ has polynomial Fourier decay.

Remark 1.3.5. *It is important to note that this theorem is from a previous version of the arxiv paper [54], but all previous versions are available to view on arxiv in the standard way. The latest version on arxiv removes the dimension assumption using a paper of Stoyanov [63]. We do not cover the extension here for various reasons, but see Remark 5.3.5 for some heuristic details for the extension. Alternatively see [54] for full details.*

Since proving Theorem 1.3.1 in 2018, it was unclear to Sahlsten and I as to whether we could prove a theorem like Theorem 1.3.4. This is because the non-concentrated derivative assumption is quite difficult to prove. What we did know at the time was that we could prove this assumption for the Gauss map [53] and convex cocompact Fuchsian groups [8]. In particular, the result of Bourgain–Dyatlov was only proven for Patterson–Sullivan measures. One can ask whether the large deviation theory of Jordan–Sahlsten can be used to extend the result of Bourgain–Dyatlov to statistically defined measures on the limit set of Fuchsian groups. The proof of the non-concentrated derivative property given by Bourgain–Dyatlov does take a lot of advantage of the strong properties of the Patterson–Sullivan measures. A particular property called 'inversion' (defined in Chapter 6) can be proven to be sufficient, but this property is not typical for general statistical measures. We prove that a small, but uncountable class of measures has this inversion property, and hence get the following result.

Theorem 1.3.6 (S, 2021). *Consider a Gibbs measure μ defined on the limit set of a convex cocompact Fuchsian group such that the dimension of the limit set is non-zero. Assume further that μ has the inversion property (which holds for an uncountable class of Gibbs measures), and that μ has locally Hölder potential. Then μ has polynomial Fourier decay.*

1.4 Why we get the main results

The reduction of the Fourier transform to exponential sums in the Markov map case [53] turned out to follow a similar method to that of Bourgain–Dyatlov [8] by replacing some of their main tools when dealing with Fuchsian groups with the large deviation theory of Jordan–Sahlsten [26]. In the context of Jordan–Sahlsten when dealing with statistical measures μ for nonlinear dynamics T, they define a set of nice words using large deviations. Let $\lambda > 1$ be the Lyapunov exponent for the Markov map T, let $s = \dim_H \mu > 0$, and consider $\varepsilon > 0$ small and $C_{\varepsilon,n} := e^{\varepsilon n}$. They define the set of regular words $\mathbf{a}$ of length n using properties which give us that

$$C^{-1} C_{\varepsilon,n}^{-3\lambda} e^{-\lambda s n} \leq \mu(I_{\mathbf{a}}) \leq C C_{\varepsilon,n}^{3\lambda} e^{-\lambda s n}.$$

Essentially we have that $\mu(I_{\mathbf{a}})$ is controlled by $e^{-s\lambda n}$, so we compare $e^{-\lambda n}$ with τ in Bourgain–Dyatlov (see Subsection 1.1.5). The main work here is to check that the theory of Bourgain–Dyatlov still works when the ε exponentially growing terms are introduced.

In the nonlinear markov map case, we can reduce proving polynomial Fourier decay to proving the non-concentrated derivative assumption. Bourgain–Dyatlov prove this assumption using a tree structure inherited by the definition of their 'regular words' $Z(\tau)$ and distortion factor properties for Schottky groups. In the large deviation context of Jordan–Sahlsten, we do not have such a tree structure, because the structure of the regular set is unknown (as is typical when using statistical theorems). However, we are able to define regular words in a way such that they are regular at many generations, not just the generation n, which will aid us without a tree structure.

It seems that when trying to prove that a map has non-concentrated derivative, in both the Gauss map case and Fuchsian group setting, we have a strong geometric structure which allows us to analyse derivatives. In both cases, what we are essentially studying is $f''_{\mathbf{a}}/f'_{\mathbf{a}} = D(\log f'_{\mathbf{a}})$ for indexes $\mathbf{a} \in \mathcal{A}^n$, where $f_{\mathbf{a}}$ are iterated inverse branches for an (eventually) expanding map. In the Mobius transformation case, this is related to studying the preimage of infinity of the transformation, notably $\gamma_{\mathbf{a}}^{-1}(\infty)$. We can then use inverting (or 'reversal' in Bourgain–Dyatlov [8]) to say that $\gamma_{\mathbf{a}}^{-1}(\infty)$ is a point $\gamma_{\bar{\mathbf{a}}}(\infty)$ with initial coding $\bar{\mathbf{a}}$, where $\bar{\mathbf{a}}$ will be the 'inverse' of $\mathbf{a}$. We then use the measure in question to study how these points distribute. It also makes sense heuristically as to why we study $\gamma_{\mathbf{a}}^{-1}(\infty)$, because this point tells you where derivatives of $\gamma_{\mathbf{a}}$ get large. You can think of $\gamma_{\mathbf{a}}^{-1}$ as the point that causes the real line to stretch under the transformation, hence knowing how they distribute helps us to understand derivatives of these transformations.

In the Gauss map case, we find out that $f''_{\mathbf{a}}/f'_{\mathbf{a}}$ corresponds to the point with finite continued fraction expansion given by $\mathbf{a}^{\leftarrow}$, the reverse of the word $\mathbf{a} := a_1 \dots a_{n-1}a_n$, namely

$$[\mathbf{a}^{\leftarrow}] = \frac{p_n(\mathbf{a}^{\leftarrow})}{q_n(\mathbf{a}^{\leftarrow})} := \cfrac{1}{a_n + \cfrac{1}{a_{n-1} + \cfrac{1}{\dots + \cfrac{1}{a_1}}}}.$$

The differences of such points is studied by Queffélec and Ramaré again using the measure itself. Jordan–Sahlsten use their arguments to count such words $\mathbf{a}$, and hence we could almost directly use their argument in [53].

In the case of general nonlinear maps, we were recommended by Jialun Li after his discussions with Frédéric Naud to try using the theory of complex transfer operators to prove non-concentration. One main reason behind this recommendation is that these operators are historically used in counting problems. Indeed, they are well known for counting primitive geodesics on hyperbolic surfaces (see Theorems 1.2.9 and 1.2.12). This lead to the progress that we have now, which proves that these operators can indeed be used to count derivatives, and prove the non-concentration assumption.

Another reason why it seemed that complex transfer operators could be relevant to this problem is because of the common assumptions to get the relevant spectral gap theorems. The main assumption, which typically is something similar to a map being totally-nonlinear, can be shown to be equivalent to a condition on distortions $f''_{\mathbf{a}}/f'_{\mathbf{a}}$, see (1.3). In particular, the assumption heuristically says that sometimes these distortions can be far apart. This suggests a strong link between the methods of Bourgain–Dyatlov, Jordan–Sahlsten (and Queffélec–Ramaré), and complex transfer operator theory, because all of these ideas consider control over the the distortions when they are close together.

Chapter 2

Thermodynamical Formalism

We will assume that the first generation inverse branches of the dynamics $T : X \to X$ are monotonic increasing. We could study a less restrictive condition, such as in the Gauss map case the inverse branches are decreasing, which usually just means taking note of whether the generation n being analysed is odd or even. However, for simplicity, we restrict to the easiest case to manage which is increasing branches.

Throughout this thesis we shall study dynamical systems given by Markov maps which are conjugate to a subshift of finite type on some symbolic space $\mathcal{A} \subset \mathbb{N}$. In the Hyperbolic surface setting, $\mathcal{A} = \{1, \ldots, 2r\}$ for some r will be a finite set. However, there will be a condition on concatenating letters to form words when studying points $y \in T^{-1}\{x\}$ where T is the Markov map in this setting (to be defined). If we can say that the concatenation condition is satisfied, we say that the resulting concatenation of letters (or word) is admissible. Given two finite words $\mathbf{a}$ and $\mathbf{b}$, we will say that $\mathbf{a} \to \mathbf{b}$ if the concatenated word $\mathbf{ab}$ is admissible. In the Gauss map case, the alphabet will be $\mathbb{N}$ and hence countable, but there is no condition on concatenating letters. The aim of this section is to give a general framework which can be used throughout this thesis. There will be some slight differences in the framework when we study the countable Gauss map case, but the reader will be referred to the details given in the corresponding paper in this case [53].

We consider intervals $(I_j)_{j\in\mathcal{A}}$ which are closed and bounded intervals in $\mathbb{R}$. We will let X be the interior of the union of these intervals. We will consider a Markov map $T : X \to X$, and define its branches $T_j := T|_{I_j} \in C^2(I_j)$ for each interval I_j. We will denote the inverse branches of T by $f_j := (T|_{I_j})^{-1}$ for each $j \in \mathcal{A}$. We will require that T:

(1) is eventually expanding, i.e. that there exists some $\theta > 1$, $C_E > 0$, and $N_E \geq 1$ such that for all $n \geq N_E$ and all $x \in T^{-n}(X)$,

$$|(T^n)'(x)| \geq C_E \theta^n;$$

(2) satisfies the Markov property, i.e. for all $i, j \in \mathcal{A}$, if $T(I_j) \cap I_i^o$ is non-empty, then $I_i \subset T(I_j)$.

This gives us a limit set

$$\Lambda := \cap_{j=0}^{\infty} T^{-j}(X)$$

which is such that $T(\Lambda) \subset \Lambda$. To analyse the points in the limit set, we will use a symbolic coding given by the Markov map.

We also require a condition that we refer to as finite distortion. The assumption says that for all iterated inverse branches $f_{\mathbf{a}}$, we have that there exists some $B > 0$ such that for all $z \in I_b$ where $b \in \mathcal{A}$ is such that $\mathbf{a} \to b$, we have that

$$\Big|\frac{f''_{\mathbf{a}}(z)}{f'_{\mathbf{a}}(z)}\Big| \leq B.$$

This assumption can be significantly weakened by only considering maps T whose inverse branches are C^2, and the size of the image of the branches are bounded from above and below. This also uses the fact that we only consider finitely many branches for simplicity. This assumption will allow us to control an inverse branch evaluated on different points.

Lemma 2.0.1. *Given $b \in \mathcal{A}$, consider $\mathbf{a} \in \mathcal{A}_n$ such that $\mathbf{a} \to b$ and $x, y \in I_b$. Assume that there exists some $B > 0$ such that for all $z \in I_b$*

$$\Big|\frac{f''_{\mathbf{a}}(z)}{f'_{\mathbf{a}}(z)}\Big| \leq B.$$

Then we have that

$$\frac{f'_{\mathbf{a}}(x)}{f'_{\mathbf{a}}(y)} \leq \exp(B|x-y|).$$

Proof. By the mean value theorem we have that

$$\begin{aligned}\frac{f'_{\mathbf{a}}(x)}{f'_{\mathbf{a}}(y)} &= \exp\Big(\log\frac{f'_{\mathbf{a}}(x)}{f'_{\mathbf{a}}(y)}\Big) \leq \exp|\log f'_{\mathbf{a}}(x) - \log f'_{\mathbf{a}}(y)| \\ &= \exp(|(\log f'_{\mathbf{a}})'(z)| \cdot |x-y|) \leq \exp(B|x-y|).\end{aligned}$$

□

We define our transition matrix $A = (a_{i,j})$ which is size $(\#\mathcal{A}) \times (\#\mathcal{A})$ (which can be infinite) by

$$a_{i,j} = \begin{cases} 1 & \text{when } I_j \subset T(I_i) \\ 0 & \text{otherwise.} \end{cases}$$

We will assume that A is aperiodic, so that there exists a p such that A^p has no zero entries. If $a_{i,j} = 1$, we will write $i \to j$. The matrix A will then define our admissible words

$$\Sigma := \{\mathbf{a} := a_1 a_2 \ldots \in \mathcal{A}^{\mathbb{N}} : a_{i,i+1} = 1 \, \forall i \geq 1\}$$

We will let $\mathcal{A}_n$ be the corresponding set of finite words of length n (which are admissible with respect to A). We can define the corresponding inverse branches $f_{\mathbf{a}} := f_{a_1} \circ \ldots \circ f_{a_n}$ for $\mathbf{a} \in \mathcal{A}_n$. On the sequence space Σ, we define the shift map $\sigma : \Sigma \to \Sigma$ by $(\sigma \mathbf{a})_n := a_{n+1}$ for all $n \geq 1$. We define our coding map $\Pi : \Sigma \to \Lambda$ by

$$\Pi(\mathbf{a}) := \lim_{n \to \infty} f_{a_1} \circ \ldots \circ f_{a_{n-1}}(x_{a_n})$$

where x_{a_n} is any point in I_{a_n} (for example, the centre point). Under Π, the shift on Σ is conjugate to T acting on Λ so that $T \circ \Pi = \Pi \circ \sigma$.

We can now define (Gibbs) measures on such spaces. We do so using a potential $\varphi \in C^2(X)$. We define the nth Birkhoff sum of φ on X by

$$S_n \varphi := \sum_{j=0}^{n-1} \varphi \circ T^j.$$

We write $\mathcal{M}$ as the space of all Borel probability measures on X and $\mathcal{M}_T \subset \mathcal{M}$ as the space of all T-invariant ones X. We will only consider potentials $\varphi : X \to \mathbb{R}$ which are locally Hölder, that is there exists $C > 0$ and $\delta < 1$ such that for all $n \in \mathbb{N}$,

$$\sup_{\mathbf{a} \in \mathcal{A}_n} \sup\{\varphi(x) - \varphi(y) : x, y \in I_{\mathbf{a}}\} \leq C \delta^n.$$

So these potentials behave nicely on construction intervals, which will be a useful property when considering the following constructions.

Definition 2.0.2. *(1) Let $C(X)$ be the space of all bounded continuous functions $h : X \to \mathbb{R}$. The transfer operator associated to a potential φ is the map $\mathcal{L}_\varphi : C(X) \to C(X)$, defined for $h \in C(X)$ and $x \in X$ by*

$$\mathcal{L}_\varphi f(x) := \sum_{y \in T^{-1}\{x\}} e^{\varphi(y)} h(y).$$

The dual operator of $\mathcal{L}_\varphi$ *on* $\mathcal{M}$ *is the map* $\mathcal{L}_\varphi^* : \mathcal{M} \to \mathcal{M}$*, defined by*

$$\mathcal{L}_\varphi^* \nu(h) := \int \mathcal{L}_\varphi h \, d\nu$$

at $\nu \in \mathcal{M}$ *and* $h \in C(X)$.

(2) *Let* $\mu \in \mathcal{M}_T$*. Then the Kolmogorov-Sinai entropy of* μ *is defined by*

$$h_\mu := \lim_{n\to\infty} \frac{1}{n} \sum_{\boldsymbol{a} \in \mathcal{A}_A^n} -\mu(I_{\boldsymbol{a}}) \log \mu(I_{\boldsymbol{a}})$$

and the Lyapunov exponent of μ *is*

$$\lambda_\mu := \int \log |T'| \, d\mu.$$

(3) *Given a potential* $\varphi : X \to \mathbb{R}$*, define the pressure associated to* φ *by*

$$P(\varphi) := \sup_{\mu \in \mathcal{M}_T} \left\{ h_\mu + \int \varphi \, d\mu : \int \varphi \, d\mu > -\infty \right\}$$

Any potential attaining this supremum is called an equilibrium state for φ.

It will be useful to represent the transfer operator using the symbolic coding. We will have that for $b \in \mathcal{A}$,

$$\mathcal{L}_\varphi h(x) = \sum_{a \in \mathcal{A}:\, a \to b} e^{\varphi(f_a x)} h(f_a x) \quad x \in I_b.$$

In the context of countable Markov maps, the equilibrium state is not unique, and also without further assumptions on the potential, it fails to have many nice properties. The following theorem follows from Sarig [57], but written down in e.g. [26] gives natural assumptions that we will impose and then gives us the statistical theorems needed to get enough regularity for equilibrium states. Note that this theory also holds for the case of Markov maps with finitely many branches because the tail condition is satisfied.

Theorem 2.0.3. *Suppose* $\varphi : X \to \mathbb{R}$ *is a locally Hölder continuous potential and* μ *an equilibrium state associated to* φ*. Assume* μ *has at most polynomial tail, that is there exists some* $p > 1$ *such that*

$$\mu\Big(\bigcup_{j=n}^{\infty} I_j\Big) = O(n^{-p}) \quad n \to \infty.$$

Then there exists a locally Hölder continuous potential $\varphi_0 : X \to \mathbb{R}$ *such that*

(1) $\varphi_0 \leq 0$, $P(\varphi_0) = 0$ *and* $\mathcal{L}_{\varphi_0} 1 = 1$*;*

(2) the potential φ_0 *has a corresponding measure* μ_{φ_0} *satisfying*

$$\mathcal{L}^*_{\varphi_0} \mu_{\varphi_0} = \mu_{\varphi_0}$$

and the Gibbs condition: there exists $C > 0$ *such that for* μ *almost every* $x \in X$ *we have for all* $n \in \mathbb{N}$ *that*

$$C^{-1} \exp(S_n \varphi_0(x)) \leq \mu_{\varphi_0}(I_{\mathbf{a}}) \leq C \exp(S_n \varphi_0(x)).$$

(3) the equilibrium state μ_{φ_0} *associated to* φ_0 *is the same measure as* μ*:*

$$\mu = \mu_{\varphi_0}.$$

Thanks to part (3) of Theorem 2.0.3, we may, from the beginning assume that $\varphi \leq 0$, $P(\varphi) = 0$, $\mathcal{L}_\varphi 1 = 1$, μ is the unique equilibrium state associated to φ satisfying the invariance under the transfer operator

$$\mathcal{L}^*_\varphi \mu = \mu$$

and the Gibbs condition: there exists some $C > 0$ such that for all $n \in \mathbb{N}$ and all $x \in X$,

$$C^{-1} \exp(S_n \varphi(x)) \leq \mu(I_{\mathbf{a}}) \leq C \exp(S_n \varphi(x)).$$

We shall assume $C > 1$ without loss of generality.

By assuming that the potential $-\log |T'|$ is locally Hölder, we will get that inverse branches 'linearise' under iteration. We will use a weak form of this linearisation. Note that given a word $\mathbf{a} \in \mathcal{A}_n$, we will say that a point $x \in X$ is such that $\mathbf{a} \to x$ if $x \in I_b$ for some $b \in \mathcal{A}$ and $\mathbf{a} \to b$.

Lemma 2.0.4. *Assume that* $-\log |T'|$ *is locally Hölder. For* $\mathbf{a} \in \mathcal{A}_n$*,* $x, y \in X$ *such that* $\mathbf{a} \to x, y$*, we have that*

$$e^{-C} |f'_{\mathbf{a}}(x)| \leq |f'_{\mathbf{a}}(y)| \leq e^C |f'_{\mathbf{a}}(x)|.$$

Proof. Since the potential $-\log |T'|$ is assumed to be locally Hölder we have that

$$|\log |T'(f_{\mathbf{a}}(x))| - \log |T'(f_{\mathbf{a}}(y))| \leq C\delta^n < C$$

where we weakly use $\delta < 1$. We can conclude using the inverse rule for differentiable functions. □

For two numbers $a, b \in \mathbb{R}$, we will say that $a \sim b$ if there exists a $C \geq 1$ independent of a and b such that $C^{-1}a \leq b \leq Ca$. We use the following basic results to compare words of similar length.

Lemma 2.0.5. *For a word* $\mathbf{a} \in \mathcal{A}_n$*, consider* $b \in \mathcal{A}$ *such that* $\mathbf{a} \to b$*. Then we have that for any* $x, y \in X$ *such that* $\mathbf{a} \to x$ *and* $\mathbf{a}' \to y$

$$|f'_{\mathbf{a}}(x)| \sim |f'_{\mathbf{a}'}(y)|.$$

From this we will also have that for any $\mathbf{b} \in \mathcal{A}_m$ *such that* $\mathbf{a} \to \mathbf{b}$ *the following hold*

$$\begin{aligned} |f'_{\mathbf{a}}(y)| &\sim |I_{\mathbf{a}}| \\ |I_{\mathbf{a}b}| &\sim |I_{\mathbf{a}}| \\ |I_{\mathbf{ab}}| &\sim |I_{\mathbf{a}}| \cdot |I_{\mathbf{b}}|. \end{aligned}$$

Proof. By the chain rule we have that

$$|f'_{\mathbf{a}}(x)| = |f'_{\mathbf{a}'}(f_{a_n}x)| \cdot |f'_{a_n}(x)|.$$

By assuming that T has bounded derivative we have that $|f'_{a_m}x| \sim 1$ by using the inverse rule for differentiable functions. We can therefore get our first $\sim$ relation of the lemma by using Lemma 2.0.4.

To relate derivatives to construction intervals, we use the mean value theorem to see that

$$|I_{\mathbf{a}}| = |f'_{\mathbf{a}'}(\xi)| \cdot |I_{a_n}|$$

for some $\xi \in I_{a_n}$, so $|I_{\mathbf{a}}| \sim |f'_{\mathbf{a}'}(y)|$ by Lemma 2.0.4. We therefore get the second $\sim$ relation. We follow by noticing that

$$|f'_{\mathbf{a}}(x)| \sim |f'_{\mathbf{a}'}(y)| \sim |I_{\mathbf{a}}|$$

giving the third $\sim$ relation. In a similar manner we get that

$$|I_{\mathbf{a}b}| \sim |f'_{\mathbf{a}}(x)| \sim |f'_{\mathbf{a}'}(y)| \sim |I_{\mathbf{a}}|$$

and

$$|I_{\mathbf{ab}}| \sim |f'_{\mathbf{ab}'}(\xi)| \sim |f'_{\mathbf{a}}(f_{\mathbf{b}'}\xi)| \cdot |f_{\mathbf{b}'}(\xi)| \sim |I_{\mathbf{a}}| \cdot |I_{\mathbf{b}}|$$

for some $\xi \in I_{b_m}$, giving the final two $\sim$ relations. □

We will require a similar concatenation property for measuring construction intervals, known as the quasi-Bernoulli property.

Lemma 2.0.6. *For* $\mathbf{a} \in \mathcal{A}_n$ *and* $\mathbf{b} \in \mathcal{A}_m$ *with* $\mathbf{a} \to \mathbf{b}$*, we have that*

$$\mu(I_{\mathbf{ab}}) \sim \mu(I_{\mathbf{a}}) \cdot \mu(I_{\mathbf{b}}).$$

Proof. If we define the Bernoulli weights

$$w_{\mathbf{a}|_k}(x) := e^{S_k\varphi(f_{\mathbf{a}|_k}x)}$$

for $x \in X$ such that $\mathbf{a} \to x$, then $\mu(I_{\mathbf{a}}) \sim w_{\mathbf{a}}(x)$ by the Gibbs condition. We get the required result by using the fact that

$$w_{\mathbf{ab}}(y) = e^{S_{n+m}\varphi(f_{\mathbf{ab}}y)} = e^{S_n\varphi(f_{\mathbf{ab}}(y))+S_m\varphi(f_{\mathbf{b}}x)} = w_{\mathbf{a}}(f_{\mathbf{b}}y) \cdot w_{\mathbf{b}}(x).$$

□

2.1 Iterating Transfer Operators

We will define blocks of words with respect to the transition matrix. For some $k \in \mathbb{N}$, we define

$$\mathcal{A}_n^k := \{\mathbf{B} := \mathbf{b}_1 \dots \mathbf{b}_k : \mathbf{b}_j \in \mathcal{A}_n\}$$

to be the set of k-blocks of words of length n, noting that we do not add an admissibility condition here for now. We will always consider $\mathbf{A} \in \mathcal{A}_n^{k+1}$ and $\mathbf{B} \in \mathcal{A}_n^k$, at least until we later impose regularity (using $\mathcal{R}_n$ in Lemma 2.2.3) and distributive conditions (using $\mathcal{W}$ in Lemma 3.2.2) on these objects. They will remain the same lengths however, namely $n(k+1)$ and nk respectfully. We define the following concatenation operations (as in [8]):

$$\mathbf{A} * \mathbf{B} := \mathbf{a}_0\mathbf{b}_1\mathbf{a}_1\mathbf{b}_2 \dots \mathbf{a}_{k-1}\mathbf{b}_k\mathbf{a}_k \quad \text{and} \quad \mathbf{A}\#\mathbf{B} := \mathbf{a}_0\mathbf{b}_1\mathbf{a}_1\mathbf{b}_2 \dots \mathbf{a}_{k-1}\mathbf{b}_k.$$

For these operations to be well-defined, we must have that $(\mathbf{a}_j)_n \to (\mathbf{b}_{j+1})_1$ for all $j = 0, \dots, k$, and additionally $(\mathbf{b}_k)_n \to (\mathbf{a}_k)_1$ for the $*$ operation. If these conditions hold, we write $\mathbf{A} \leftrightarrow \mathbf{B}$. We can start to study the dynamics using these blocks of words by using the transfer operator (see proof of [45, Proposition 2.1]).

Lemma 2.1.1. *We have that*

$$\mathcal{L}_\varphi^n h(x) = \sum_{\mathbf{a}\in\mathcal{A}_n : a_n \to b} e^{S_n\varphi(f_\mathbf{a}x)} h(f_\mathbf{a}x), \quad x \in I_b$$

Proof. First note that this result is true for $n = 1$ by definition of the transfer operator. We shall assume that this result is true for $n = k$. For $n = k + 1$ we have that

$$\begin{aligned}
\mathcal{L}_\varphi^{k+1} h(x) &= \mathcal{L}_\varphi\Big(\sum_{\mathbf{a}\in\mathcal{A}_k : a_k \to b} e^{S_k\varphi(f_\mathbf{a})} h \circ f_\mathbf{a}\Big)(x) \\
&= \sum_{\mathbf{a}\in\mathcal{A}_k : a_k \to a_{k+1}} \sum_{a_{k+1}\in\mathcal{A}} e^{\varphi(f_{a_{k+1}}x)}\Big(e^{S_k\varphi(f_\mathbf{a} f_{a_{k+1}}x)} h(f_\mathbf{a} f_{a_{k+1}}x)\Big) \\
&= \sum_{\mathbf{a}a_{k+1}\in\mathcal{A}_{k+1}} \exp\Big(\varphi(f_{a_{k+1}}x) + \sum_{i=0}^{k-1} \varphi(T^i f_\mathbf{a} f_{a_{k+1}}x)\Big) h(f_{\mathbf{a}a_{k+1}}x)
\end{aligned}$$

We get the result for $n = k + 1$ by noting that the power in the exponent of the above line is equal to $S_{k+1}\varphi(x)$ by definition of the Birkhoff sum. □

To consider blocks of words, we can repeatedly use Lemma 2.1.1 to see that

$$(\mathcal{L}_\varphi^n)^{2k+1} h(x) = \sum_{\substack{\mathbf{A}\leftrightarrow\mathbf{B} \\ \mathbf{A}\to b}} w_{\mathbf{A}*\mathbf{B}}(x) h(f_{\mathbf{A}*\mathbf{B}}x) \quad x \in I_b$$

where we say $\mathbf{A} \to b$ if $(\mathbf{a}_k)_n \to b$. So we will now be able to consider how functions behave locally on construction intervals rather than the whole space X. The advantage of this is that we will be able to ignore any construction intervals on which the dynamics behave badly by defining deviating sets.

2.2 Large Deviations for Expanding Markov Maps

Cramér's theorem states that for a sum of n independent identically distributed random variables with finite moment generating function, for some $\varepsilon > 0$, the probability that the sum is bounded away from its mean by εn will decay exponentially as $n \to \infty$. We can reach an analogue for dynamical systems by considering Birkhoff sums. In the dynamical context, we will have assumptions on the pressure which are analogous to the finite moment assumptions in Cramér's theorem. The reason we need the specific conditions on pressure and the tail condition is that we need to find a large regular part of the measure μ in terms of the Lyapunov exponent and Hausdorff dimension, which allow us to prove good estimates on the Fourier transforms. In Bourgain–Dyatlov [8] they dealt with Patterson–Sullivan measures which automatically are Ahlfors–David regular, which is stronger than the Gibbs condition, so they do not need a large deviation theorem to control the measure. Large deviations allow us to extract a "large part" of the support with similar Ahlfors–David regular behaviour for μ. Here is also where we need the finite Lyapunov exponent for μ. The following theorem is given in the paper of Jordan–Sahlsten [26], and we refer the reader there for a proof.

Theorem 2.2.1 (Jordan–Sahlsten, Large deviations for (μ, T) [26])**.** *Let μ be the equilibrium state associated to φ having at most polynomial tail, that is there exists some $p > 1$ such that*

$$\mu\Big(\bigcup_{a=n}^{\infty} I_a\Big) = O(n^{-p}) \quad n \to \infty.$$

Let λ be the Lyapunov exponent of μ and s the Hausdorff dimension. Write

$$\psi = -\log |T'|.$$

Then we have that for any $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\mu\Big(\Big\{x \in X : \Big|\frac{1}{n} S_n\psi(x) + \lambda\Big| \geq \varepsilon \quad \text{or} \quad \Big|\frac{S_n\varphi(x)}{S_n\psi(x)} - s\Big| \geq \varepsilon\Big\}\Big) = O(e^{-n\delta}).$$

This theory is crucial for us to apply the ideas from additive combinatorics as often this type of initial regularisation is needed. We can now define the parts of the dynamics which are regular with respect to large deviations.

Definition 2.2.2 (Regular words and blocks)**.** *Fix now $\varepsilon, \varepsilon_r > 0$ and $n \in \mathbb{N}$. Write*

$$A_n(\varepsilon) = \Big\{x \in X : \Big|\frac{1}{n} S_n\psi(x) + \lambda\Big| < \varepsilon \quad \text{and} \quad \Big|\frac{S_n\varphi(x)}{S_n\psi(x)} - s\Big| < \varepsilon\Big\}.$$

Define for a fixed $n \in \mathbb{N}$ the set of regular words:

$$\mathcal{R}_n = \bigcap_{k=\lfloor \varepsilon_r n \rfloor}^{n} \{\mathbf{a} \in \mathbb{N}^n : I_{\mathbf{a}|_k} \subset A_k(\varepsilon)\}$$

Using the definitions of regular words, we define a regular block of length k to be the concatenation of k regular words of length n. We denote the set of such words by $\mathcal{R}_n^k$. Note that we can equivalently define this set as

$$\mathcal{R}_n^k = \{\mathbf{A} \in (\mathbb{N}^n)^k : I_{(\sigma^n)^i \mathbf{A}} \subset A_n(\varepsilon), \forall i = 0, 1, \ldots, n-1\}$$

where σ is the shift mapping and $A_n(\varepsilon)$ is the n-regular set. We shall consider the corresponding geometric points to be

$$R_n^k := \bigcup_{\mathbf{A}\in\mathcal{R}_n^k} I_{\mathbf{A}} \subset X$$

for $\varepsilon > 0$ small enough.

Note that unlike the choice made by Jordan–Sahlsten [26], we will require $\lfloor \varepsilon_r n \rfloor$-regularity as opposed to $\lfloor n/2 \rfloor$ for some $\varepsilon_r > 0$ to be defined. One can achieve $\lfloor \varepsilon_r n \rfloor$-regularity for any desired $\varepsilon_r > 0$ by choosing large enough n so that

$$\lfloor \varepsilon_r n \rfloor - \varepsilon_r n/2 \geq \frac{1}{\delta} \log \frac{4}{1 - e^{-\delta}}$$

in the paper of [26]. To make our choice of ε_r, one must look at the exponential sum theorem we shall use. In Lemma 3.1.2, we define some $\varepsilon_3 > 0$ using only the dimension of the measure that we consider. The lemma requires us to consider derivatives of inverse branches of size up to $e^{-\varepsilon_3 \lambda n}$. To be sure that we can do this, we choose $\varepsilon_r = \varepsilon_3/8$. For this definition to make sense, we must make sure that ε does not dominate ε_r. We can do so by demanding that $0 < \varepsilon < \varepsilon_3/9^9$, which is valid because ε_3 only depends on the measure. Note we require ε significantly smaller than ε_r so that we can consider many multiples of ε during the proof and still remain less than ε_r so it remains the dominant power.

Lemma 2.2.3. *Define $C_{\varepsilon,j} := e^{\varepsilon j}$, and assume that n is chosen large enough so that*

$$\frac{\log 4}{\varepsilon_r n} < \varepsilon/2, \quad \frac{\log 4C^2}{\log(\theta^{2\varepsilon_r n})} < \varepsilon/2 \quad \text{and} \quad \frac{e^{-\delta\varepsilon_r n}}{1 - e^{-\delta}} < e^{-\delta\varepsilon_r n/2}$$

recalling that $\theta > 1$. Given an n-regular word $\mathbf{a} \in \mathcal{R}_n$ and any $j \in \{\lfloor \varepsilon_r n \rfloor, \ldots, n\}$ we have that the following hold:

(i) the size of the derivative $|f'_{\mathbf{a}|_j}|$ satisfies

$$\frac{1}{16} C_{\varepsilon,j}^{-1} e^{-\lambda j} \leq |f'_{\mathbf{a}|_k}| \leq C_{\varepsilon,j} e^{-\lambda j}$$

and hence so does the length $|I_{\mathbf{a}|_j}|$;

(ii) The measure satisfies

$$C^{-1} \cdot C_{\varepsilon,j}^{-3\lambda} e^{-s\lambda j} \leq \mu(I_{\mathbf{a}|_j}) \leq C \cdot C_{\varepsilon,j}^{3\lambda} e^{-s\lambda j};$$

(iii) The Birkhoff weights satisfy

$$C_{\varepsilon,j}^{-3\lambda} e^{-s\lambda j} \leq w_{\mathbf{a}|_j}(x) \leq C_{\varepsilon,j}^{3\lambda} e^{-s\lambda j}.$$

(iv) The cardinality

$$\tfrac{1}{2} C^{-1} C_{\varepsilon,n}^{-3\lambda} e^{\lambda sn} \leq \sharp \mathcal{R}_n \leq C C_{\varepsilon,n}^{3\lambda} e^{\lambda sn}$$

Recall we assume that T is eventually expanding. As a result, for $k \in \mathbb{N}$ we have that if $n \to \infty$,

$$\mu(X \setminus R_n^k) = O(e^{-\delta \varepsilon_r n/2})$$

where $\delta = \delta(\varepsilon/2)$ is given to us Theorem 2.2.1.

Remark 2.2.4 (Regularity at many scales)**.** *So this multi-regularity will allow us to say for a regular word $\mathbf{a}$, $\mu(I_{\mathbf{a}|_j}) \approx e^{-s\lambda j}$ and $|I_{\mathbf{a}|_j}| \approx |f_{\mathbf{a}|_j}| \approx e^{-\lambda j}$ for $\varepsilon_r n \leq j \leq n$. This will have a great deal of uses in the proofs of all the main theorems to be presented.*

It will allow us to control the measure of sets I with $|I| \approx e^{-\lambda j}$, meaning we can exclude 'bad' sets that we don't like. For example, when dealing with exponential sums with exponents of order $\eta > 0$, we can ignore frequencies of order smaller than $e^{\lambda j}$, which will be necessary to use the exponential sum theory.

This will also be useful when proving non-concentration of derivatives, i.e. when considering two iterated inverse branches evaluated on the same regular construction interval $I_{\mathbf{d}}$, and their difference being approximately σ. We will be able to say that the set of such points x is in an interval J of length approximately $\sigma^{1/2}$. So it will be useful to be sure that there is a valid choice of j such that $\sigma^{1/2} \approx e^{-\lambda j}$. We will then be able to count the number of construction intervals $I_{\mathbf{d}}$ contained in J by using regular measure bounds on the intervals $I_{\mathbf{d}}$ and the interval J. This will drive the non-concentration proofs.

We will need ε_r to be at least $1/2$ in the Gauss map setting to use the proof of Jordan–Sahlsten where they modify some continuant analysis of Queffélec–Ramaré [26][50]. This analysis is pivotal in proving the non-concentration condition for the Gauss map, which gives Fourier decay.

In the totally-nonlinear map setting, we will define ε_r with respect to the spectral gap of a complex transfer operator so that the corresponding theory can be used to *count derivatives. In particular, we will want to restrict the inverse branches so that their derivative is approximately the spectral gap itself.*

When attempting to generalise the proof of Bourgain–Dyatlov [8] to statistical measures, we will see that the Patterson–Sullivan measures that they consider will have the incredibly useful property that $\mu(I) \approx |I|^{\dim \mu}$ for intervals I. We will desire the same property in the statistical measure case for sufficiently small I. This will correspond to ensuring that ε_r is made sufficiently small.

The following proof is very similar to that of Jordan–Sahlsten [26]. The difference is that we must prove a decay result for blocks of words rather than just words themselves. The result still gives exponential decay, but the bound will depend on the length of the blocks required as you would expect. The main difference in the original proof is that we must prove multiregularity of each word in certain blocks. This comes down to some simple formalisation by isolating words in the blocks.

Proof. Parts (i), (ii), and (iii) are done in Jordan–Sahlsten [26] and the part (iv) follows from the bounds for $\mu(I_{\mathbf{a}})$ and combining with the measure bound for $\mu(X \setminus R_n)$.

It should be noted for part (i) that we need to take admissibility of the symbolic space into account. Our definition of construction intervals is $I_{\mathbf{a}} := f_{\mathbf{a}'}(I_{a_n})$, so by the mean value theorem there exists some $\xi \in I_{a_n}$ such that

$$|I_{\mathbf{a}}| = |f_{\mathbf{a}'}(\xi)||I_{a_n}|$$

which holds since we consider inverse branches to be monotonic. So we can get regularity bounds of $|I_{\mathbf{a}}|$ if $f'_{\mathbf{a}}$ has regularity bounds.

For the measure bound for $\mu(X \setminus R_n^k)$, it is sufficient to prove that

$$\bigcap_{i=0}^{k-1} (T^{-1})^{ni} \Big(\bigcap_{j=\lfloor \varepsilon_r n \rfloor}^{n} A_j(\varepsilon/2) \Big) \subset R_n^k$$

since we have that

$$\begin{aligned} \mu(X \setminus R_n^k) &\leq \mu\Big(X \setminus \bigcap_{i=0}^{k-1} (T^{-1})^{ni} \Big(\bigcap_{j=\lfloor \varepsilon_r n \rfloor}^{n} A_j(\varepsilon/2) \Big)\Big) \\ &\leq \sum_{i=0}^{k-1} \mu\Big(X \setminus (T^{-1})^{ni} \Big(\bigcap_{j=\lfloor \varepsilon_r n \rfloor}^{n} A_j(\varepsilon/2) \Big)\Big) \\ &\leq k\mu\Big(X \setminus \Big(\bigcap_{j=\lfloor \varepsilon_r n \rfloor}^{n} A_j(\varepsilon/2) \Big)\Big) \leq k e^{-\delta \varepsilon_r n/2} \end{aligned}$$

where the details of the last inequality are given in Jordan–Sahlsten [26].

We now prove the claim. Let $\mathbf{B} \in (\mathbb{N}^n)^k$ be a word such that $T^{ni} f_{\mathbf{B}} x \in A_j(\varepsilon/2)$ for all $i = 0, 1, \dots, k-1$ and all $j = \lfloor \varepsilon_r n \rfloor, \dots, n$. We want to prove that $f_{\mathbf{B}} x \in R_n^k$. By definition of R_n^k, it is enough for us to prove that $f_{\mathbf{B}} x \in I_{\mathbf{A}}$ for some $\mathbf{A} \in \mathcal{R}_n^k$. So we can just prove that $\mathbf{B} \in \mathcal{R}_n^k$. By definition of $\mathcal{R}_n^k$, we need to prove that $I_{(\sigma^n)^i \mathbf{B}|_j} \subset A_j(\varepsilon)$ for all $i = 0, 1, \dots, k-1$ and $j = \lfloor \varepsilon_r n \rfloor, \dots, n$. If we have $y \in X \setminus Q$, then $T_{(\sigma^n)^i \mathbf{B}|_j} y$ is a general point in $I_{(\sigma^n)^i \mathbf{B}|_j}$ (we may equivalently consider the point $T^{ni} f_{\mathbf{B}|_j} y$). So we want to prove that $f_{(\sigma^n)^i \mathbf{B}|_j} y \in A_j(\varepsilon)$. Using the assumptions on $\mathbf{B}$ we have that

$$\begin{aligned}\Big|\frac{1}{j} S_j \psi(f_{(\sigma^n)^i \mathbf{B}|_j} y) + \lambda\Big| &\leq \Big|\frac{1}{j} S_j \psi(f_{(\sigma^n)^i \mathbf{B}|_j} y) - \frac{1}{j} S_j \psi(f_{(\sigma^n)^i \mathbf{B}} x)\Big| + \frac{\varepsilon}{2} \\ = \frac{\varepsilon}{2} + \frac{1}{j} \log \frac{|f'_{(\sigma^n)^i \mathbf{B}|_j}(f_{(\sigma^n)^{i+1} \mathbf{B}} y)|}{|f'_{(\sigma^n)^i \mathbf{B}|_j}(f_{(\sigma^n)^{i+1} \mathbf{B}} x)|} &\leq \frac{\varepsilon}{2} + \frac{\exp(B \max_{b \in \mathcal{A}} |I_b|)}{j} \leq \varepsilon\end{aligned}$$

by choice of n, and using the fact that $\varepsilon < \varepsilon_r$. We can get the second to last inequality by using Lemma 2.0.1. Now for the second condition we see that

$$\begin{aligned}\Big|\frac{S_j \varphi(f_{(\sigma^n)^i \mathbf{B}|_j} y)}{S_j \psi(f_{(\sigma^n)^i \mathbf{B}|_j} y)} - s\Big| &\leq \Big|\frac{S_j \varphi(f_{(\sigma^n)^i \mathbf{B}|_j} y)}{S_j \psi(f_{(\sigma^n)^i \mathbf{B}|_j} y)} - \frac{S_j \varphi(f_{(\sigma^n)^i \mathbf{B}} x)}{S_j \psi(f_{(\sigma^n)^i \mathbf{B}} x)}\Big| + \frac{\varepsilon}{2} \\ &\leq \frac{\log 4C^2}{\log(c\theta^{2j})} + \frac{\varepsilon}{2} < \frac{\log 4C^2}{j \log \theta} + \frac{\varepsilon}{2} < \varepsilon\end{aligned}$$

by following the proof of Lemma 5.2 in [26]. □

Lemma 2.2.5. *For* $\mathbf{a}, \mathbf{b} \in \mathcal{R}_n$*,*

$$16^{-2} C_{\varepsilon,n}^{-2} e^{-2\lambda n} \leq |I_{\mathbf{ab}}| \leq C_{\varepsilon,n}^2 e^{-2\lambda n}.$$

Proof. This again follows by the mean value theorem. We get that for some $\xi \in I_{b_n}$,

$$|I_{\mathbf{ab}}| = |f'_{\mathbf{ab}'}(\xi)||I_{b_n}|.$$

By the chain rule, we get the result by the regularity bounds in Lemma 2.2.3. □

Chapter 3

Multiplicative Convolutions

In the work of Jordan–Sahlsten, they bound the Fourier transform of the measure with a Lebesgue integral. Making the integral Lebesgue means that the bounds of Kaufman can be used, which arise from the use of integration by parts. However, there is a multiplicative error applied to the Lebesgue integral which towards the end of the proof forces the $\dim \mu > 1/2$ assumption. In the proof of Bourgain–Dyatlov, they do not need their integrals to be Lebesgue; they bound the transform with exponential sums. The exponential sums can then be bounded using the multiplicative convolution theory of Bourgain [7]. Here we investigate whether this theory of multiplicative convolutions can be applied to wider classes of measures using large deviation theory.

3.1 Multiplicative Convolutions and Exponential Sums

To control exponential sums as in Bourgain and Dyatlov [8], we will use the following Fourier decay theorem for multiplicative convolutions proved in this form by Bourgain [7, Lemma 8.43] that follows from the discretized sum-product theorem. In [8] Bourgain and Dyatlov showed that by taking linear combinations of measures μ_j, one can prove an analogous statement for multiplicative convolutions of several measures μ_j with the growth assumption (1.1) on $\mathbb{R}$ replaced with a growth assumption for $\mu_j \times \mu_j$ on $\mathbb{R}^2$. Then in the case of discrete measures μ_j, this implies the following decay theorem for exponential sums:

Lemma 3.1.1 (Bourgain–Dyatlov [8])**.** *Fix $\delta_0 > 0$. Then there exist $k \in \mathbb{N}$, $\varepsilon_3, \varepsilon_2 > 0$*

depending only on δ_0 *such that the following holds. Let* $C_0, N \geq 0$ *and for* $j = 1, \dots, k$ *let* $\mathcal{Z}_j$ *be finite sets such that* $\sharp \mathcal{Z}_j \leq C_0 N$. *For* $\eta \in \mathbb{R}$ *with* $|\eta| > 1$, *assume that for each* j *we have a map* $\zeta_j : \mathcal{Z}_j \to [C_0^{-1}, C_0]$ *such that for each* $\sigma \in [|\eta|^{-1}, |\eta|^{-\varepsilon_3}]$,

$$\#\{(\mathbf{b}, \mathbf{c}) \in \mathcal{Z}_j^2 : |\zeta_j(\mathbf{b}) - \zeta_j(\mathbf{c})| \leq \sigma\} \leq C_0 N^2 \sigma^{\delta_0}.$$

Then for some constant C_1 *depending only on* C_0 *and* δ_0 *we have that*

$$\Big| N^{-k} \sum_{\mathbf{b}_1 \in \mathcal{Z}_1, \dots, \mathbf{b}_k \in \mathcal{Z}_k} \exp(2\pi i \eta \zeta_1(\mathbf{b}_1) \dots \zeta_k(\mathbf{b}_k)) \Big| \leq C_1 |\eta|^{-\varepsilon_2}.$$

However, in our case, due to the fluctuations arising from large deviations of the

$$\psi = -\log |T'|$$

potential, the maps ζ_j we obtain do not map the sets $\mathcal{Z}_j$ into a fixed interval $[C_0^{-1}, C_0]$, but when we increase $|\eta|$, the C_0 will change and will actually blow-up polynomially in $|\eta|$. Since the constant C_1 in Lemma 3.1.1 depends on C_0, it could cause problems when we increase $|\eta|$. For this reason we will open up the argument of Bourgain and Dyatlov (Proposition 3.2 of [8]) to give a more precise dependence on the constant C_1 and C_0 and have the following quantitative version:

Lemma 3.1.2. *Fix* $\varepsilon_0 > 0$. *Then there exist* $k \in \mathbb{N}$, $\varepsilon_2 > 0$, $\varepsilon_3 > 0$ *depending only on* ε_0 *such that the following holds. Let* $R, N > 1$ *and* $\mathcal{Z}_1, \dots, \mathcal{Z}_k$ *be finite sets such that* $\sharp \mathcal{Z}_j \leq RN$. *Consider* $\eta \in \mathbb{R}$ *with* $|\eta|$ *sufficiently large. Suppose* ζ_j, $j = 1, \dots, k$, *on the sets* $\mathcal{Z}_j$ *satisfy for all* $j = 1, \dots, k$ *that*

(1) the range

$$\zeta_j(\mathcal{Z}_j) \subset [R^{-1}, R];$$

(2) for all $\sigma \in [R^{-2}|\eta|^{-1}, |\eta|^{-\varepsilon_3}]$

$$\sharp\{(\mathbf{b}, \mathbf{c}) \in \mathcal{Z}_j^2 : |\zeta_j(\mathbf{b}) - \zeta_j(\mathbf{c})| \leq \sigma\} \leq N^2 \sigma^{\varepsilon_0}.$$

Then there exists a constant $c > 0$ *depending only on* k *such that we have that*

$$\Big| N^{-k} \sum_{\mathbf{b}_1 \in \mathcal{Z}_1, \dots, \mathbf{b}_k \in \mathcal{Z}_k} \exp(2\pi i \eta \zeta_1(\mathbf{b}_1) \dots \zeta_k(\mathbf{b}_k)) \Big| \leq c R^k |\eta|^{-\varepsilon_2}.$$

Proof. Define a measure μ_j on $\mathbb{R}$ by

$$\mu_j(A) = N^{-1}\sharp\{\mathbf{b} \in \mathcal{Z}_j : \zeta_j(\mathbf{b}) \in A\}, \quad A \subset \mathbb{R}.$$

We begin by altering assumption (2). We have that

$$\mu_j([x-\sigma, x+\sigma]) \leq \sigma^{\varepsilon_0/2}$$

for $\sigma \in [R^{-2}|\eta|^{-1}, |\eta|^{-\varepsilon_2}/2]$ by using (2). Then $\mu_j(\mathbb{R}) \leq R$ and by the assumptions (1) and (2) of the lemma we are about to prove, we have that the measure μ_j is a Borel measure on $[R^{-1}, R]$ and that

$$(\mu_j \times \mu_j)(\{(x,y) \in \mathbb{R}^2 : |x-y| \leq \sigma\}) \leq \sigma^{\varepsilon_0}$$

for all $\sigma \in [R^2|\eta|^{-1}, |\eta|^{-\varepsilon_2}]$. Then to prove the claim, we just need to check that the Fourier transform of the multiplicative convolutions of μ_j satisfies:

$$|(\mu_1 \otimes \cdots \otimes \mu_k)\widehat{\ }(\eta)| \leq R^k|\eta|^{-\varepsilon_2}. \tag{3.1}$$

The rate of decay to be found will be given by

$$\varepsilon_2 := \frac{1}{10}\min(\varepsilon_4, \varepsilon_3)$$

where ε_3 and ε_4 are given in Theorem 1.2.7.

Fix $\ell \in \mathbb{N}$ such that $2^\ell < R \leq 2^{\ell+1}$. Then $\operatorname{supp}\mu_j \cap [R^{-1}, R]$ can be covered by intervals of the form $I^{[i]} := [2^{i-1}, 2^i]$ for $i = -l, \dots, l, l+1$. Let $\mu_j^{[i]}$ be μ_j restricted to $I^{[i]}$. Thus writing the re-scaling map

$$S_i(x) = 2^{-i}x, \quad x \in \mathbb{R},$$

we have that the measure $\nu_j^{[i]} = S_i(\mu_j^{[i]})$ is supported on $[\frac{1}{2}, 1]$. Moreover, it satisfies

$$\begin{aligned}(\nu_j^{[i]} \times \nu_j^{[i]})(\{(x,y) \in \mathbb{R}^2 : |x-y| \leq \sigma\}) &\leq (\mu_j \times \mu_j)(\{(x,y) \in \mathbb{R}^2 : |x-y| \leq 2^i\sigma\}) \\ &\leq (2^i\sigma)^{\varepsilon_0} \leq 2R\sigma^{\varepsilon_0} \leq \sigma^{\varepsilon_0/2}\end{aligned}$$

where we use the fact that $2R \leq \sigma^{-1/2}$ which holds by assuming that $|\eta| > 4$. We know that the main assumption is satisfied for $\sigma \in [2^i|\eta|^{-1}, 2^i|\eta|^{\varepsilon_3}]$, so for the rescaled measure, we get the main assumption for the required range $\sigma \in [|\eta|^{-1}, |\eta|^{\varepsilon_3}]$. We will use that fact that

$$(\nu_1^{[i_1]} \otimes \cdots \otimes \nu_k^{[i_k]})\widehat{\ }\Big(\eta\prod_{j=1}^k 2^{-i_j}\Big) = (\mu_1^{[i_1]} \otimes \cdots \otimes \mu_k^{[i_k]})\widehat{\ }(\eta).$$

Each μ_j is a sum of at most $2l+2$ of the restricted measures $\mu_j^{[i]}$, in particular

$$\mu_j = \sum_{i=-l}^{l+1} \mu_j^{[i]}.$$

So the Fourier transform $(\mu_1 \otimes \cdots \otimes \mu_k)\widehat{\ }(\eta)$ decomposes into at most $(2l+2)^k$ terms consisting of Fourier transforms $(\mu_1^{[i_1]} \otimes \cdots \otimes \mu_k^{[i_k]})\widehat{\ }(\eta)$ going through all the possible restrictions $\mu_j^{[i]}$. More formally, this is because

$$\mu_1 \otimes \ldots \otimes \mu_k = \sum_{i_k=-l}^{l+1} \cdots \sum_{i_1=-l}^{l+1} \mu_1^{[i_1]} \otimes \ldots \otimes \mu_k^{[i_k]}.$$

Hence if we can prove

$$|(\nu_1^{[i_1]} \otimes \cdots \otimes \nu_k^{[i_k]})\widehat{\ }(\eta)| \leq C^* |\eta|^{-\varepsilon_2} \tag{3.2}$$

for some constant $C^* > 0$ only depending on k, the triangle inequality gives

$$|(\mu_1 \otimes \cdots \otimes \mu_k)\widehat{\ }(\eta)| \leq 2(2l+1)^k C^* |2^{-lk}\eta|^{-\varepsilon_2} \lesssim R^k C^* |\eta|^{-\varepsilon_2}.$$

Thus let us assume from the start that μ_j is supported on $[\frac{1}{2}, 1]$, which we can do since we reduce proving (3.1) to proving (3.2). As in [8], let us first argue that it is enough to consider the case $\mu_1 = \mu_2 = \cdots = \mu_k$. Given $\lambda = (\lambda_1, \ldots, \lambda_k) \in [0,1]^k$, write

$$G(\lambda) := (\mu_\lambda \otimes \cdots \otimes \mu_\lambda)\widehat{\ }(\eta) = \widehat{\mu_\lambda^{\otimes k}}(\eta).$$

and the linear combination

$$\mu_\lambda = \lambda_1 \mu_1 + \cdots + \lambda_k \mu_k.$$

Consider each λ_j to be variables in $[0,1]$. Expanding $\widehat{\mu_\lambda^{\otimes k}}(\eta)$ using the definition of μ_λ as a weighted sum of μ_j's, we see that it is a sum of k^k multiplicative convolutions of the μ_j's with coefficients given by products of $\lambda_1, \ldots, \lambda_k$. More formally, this is because

$$\mu_\lambda^{\otimes k} = \sum_{i_k=1}^{k} \cdots \sum_{i_1=1}^{k} \lambda_{i_1} \ldots \lambda_{i_k} \mu_{i_1} \otimes \ldots \otimes \mu_{i_k}$$

and using linearity of the Fourier transform operator. Then if we know that (3.1) holds for $\mu_1 = \cdots = \mu_k$, then we can apply it to μ_λ and obtain that

$$\sup_{\lambda \in [0,1]^k} |G(\lambda)| = \sup_{\lambda \in [0,1]^k} |\widehat{\mu_\lambda^{\otimes k}}(\eta)| \leq C^* |\eta|^{-\varepsilon_2}$$

since the claim of Fourier decay holds for each $\lambda \in [0,1]^k$. From this we see that the map G is a polynomial of degree k. We can use the fact that the set of polynomials of degree less than or equal to k, in k variables, is a vector space V. For $G \in V$, let $\mathcal{C}(G)$ be the set of coefficients of the polynomial G. we can define two norms on this space:

(i) $||G||_1 := \sup_{\lambda \in [0,1]^k} |G(\lambda)|$;

(ii) $||G||_2 := \max_{c \in \mathcal{C}(G)} |c|$.

These norms are equivalent because the vector space is finite dimensional. Hence we can say that for some $C^* > 0$,

$$|\partial_{\lambda_1} \dots \partial_{\lambda_k} G(\lambda)|_{\lambda=0} \leq ||G||_2 \leq C^* ||G||_1$$

where the first inequality holds because the partial derivative is the coefficient of $\lambda_1 \dots \lambda_k$ in G. So there is a constant $C^* > 0$ such that

$$\frac{1}{k!} |\partial_{\lambda_1} \dots \partial_{\lambda_k} G(\lambda)|_{\lambda=0}| \leq C^* |\eta|^{-\varepsilon_2}.$$

However,

$$|(\mu_1 \otimes \cdots \otimes \mu_k)\hat{}\,(\eta)| = \frac{1}{k!} |\partial_{\lambda_1} \dots \partial_{\lambda_k} G(\lambda)|_{\lambda=0}|,$$

so this gives the claim.

As for the case $\mu_1 = \mu_2 = \cdots = \mu_k$, depending on the amount of mass μ_1 has, we have two cases.

If $\mu_1(\mathbb{R}) \geq |\eta|^{-\varepsilon_3 \varepsilon_0/10}$, choose an integer N such that $N/2 \leq |\eta| \leq N$. The probability measure

$$\mu_0 = \frac{\mu_1}{\mu_1(\mathbb{R})}$$

on $\mathbb{R}$ satisfies

$$\sup_x \mu_0(B(x, \sigma)) < \sigma^{\varepsilon_0/2}$$

for all $\sigma \in [4R^{-2}N^{-1}, N^{-\varepsilon_3}]$. Similarly we have by applying the above for $\sigma := 4R^2\sigma$ (when $R > 1$), we obtain this for $\sigma \in [N^{-1}, 4R^2N^{-\varepsilon_3}]$ by monotonicity of μ, which holds for $|\eta|^{1-\varepsilon_3} \geq 16R^4$. Hence Theorem 1.2.7 proves the claim. Note that here the constant dependence does not change.

If $\mu_1(\mathbb{R}) \leq |\eta|^{-\varepsilon_3 \varepsilon_0/10}$, then one can use a trivial bound on exponential function in the integral convolution to obtain the claim. The desired decay can be achieved by noting $k \geq 1$ in this final case. $\square$

3.2 Non-Concentrated Derivatives of Markov Maps

We will now see how we can apply the multiplicative convolution theorem to the case of nonlinear dynamical systems. The following lemma is an application of the mean value theorem which will allow us to use distortion factor analysis in the Fuchsian group setting, and (Queffélec and Ramaré type-) continuant analysis in the Gauss map case.

Lemma 3.2.1. *Consider some interval I_d and words $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathcal{A}_n$ such that $\mathbf{a} \to \mathbf{b} \to d$ and $\mathbf{a} \to \mathbf{c} \to d$. Consider a closed interval $L \subset \mathbb{R}$. Then the set of points $x \in I_d$ such that*

$$f(x) := \log \frac{f'_{\mathbf{ab}}(x)}{f'_{\mathbf{ac}}(x)} \in L$$

is contained in some interval whose length is bounded above by

$$|L| \cdot \max_{x \in I_d} \left| \frac{f''_{\mathbf{ab}}(x)}{f'_{\mathbf{ab}}(x)} - \frac{f''_{\mathbf{ac}}(x)}{f'_{\mathbf{ac}}(x)} \right|^{-1}.$$

Proof. We want to cover the set $f^{-1}(L)$ with an interval. Since f is continuous, there exist points $x_1, x_2 \in I_d$ such that x_1 and x_2 are the end points of an interval that contains the closed set $f^{-1}(L)$. By the Mean Value Theorem, we have that

$$\frac{|f(x_1) - f(x_2)|}{|x_1 - x_2|} = f'(\xi) \text{ for some } \xi \in [x_1, x_2] \subset I_d.$$

Hence it follows that

$$|f^{-1}(L)| \leq |x_1 - x_2| = |f'(\xi)|^{-1} \cdot |f(x_1) - f(x_2)| \leq |L| \cdot \max_{x \in I_d} \left| \frac{f''_{\mathbf{ab}}(x)}{f'_{\mathbf{ab}}(x)} - \frac{f''_{\mathbf{ac}}(x)}{f'_{\mathbf{ac}}(x)} \right|^{-1}$$

as required. □

In this section we will prove the following other distribution property for general Markov maps in the regularised tree given by large deviations, which is similar to what Bourgain and Dyatlov [8] employed. Recall the definition of blocks of words $\mathbf{A}$ and $\mathbf{B}$ defined in the beginning of Section 2.1. For $\mathbf{a} \in \mathcal{A}_n$, we will define $x_{\mathbf{a}}$ to be the centre point of the interval $I_{\mathbf{a}}$, and to study how the derivatives of inverse branches distribute, we will define

$$\zeta_{j,\mathbf{A}}(\mathbf{b}) := \frac{f'_{\mathbf{a}_{j-1}\mathbf{b}}(x_{\mathbf{a}_j})}{e^{-2\lambda n}}.$$

Recall that we define $R := 16^2 C C_{\varepsilon,n}^{3\lambda}$ where $C > 0$ is the Gibbs constant.

Proposition 3.2.2. *Assuming that T has non-concentrated derivative, then the following holds. Write $\mathcal{W} \subset \mathcal{R}_n^{k+1}(\varepsilon)$ to be the set of "non-concentrated word blocks" $\mathbf{A}$ defined such that for all $j = 1, \dots, k$, $\eta \in J_{\tau(n)}(\varepsilon)$ and $\sigma \in [R(n,\varepsilon)^{-2}|\eta|^{-1}, |\eta|^{-\varepsilon_3}]$, where we have that*

$$\sharp\{(\mathbf{b},\mathbf{c}) \in \mathcal{R}_n(\varepsilon)^2 : |\zeta_{j,\mathbf{A}}(\mathbf{b}) - \zeta_{j,\mathbf{A}}(\mathbf{c})| \leq \sigma\} \leq \sharp\mathcal{R}_n(\varepsilon)^2 \sigma^{c_0/2}.$$

Then most blocks are non-concentrated, so for some $\kappa_0 > 0$,

$$e^{-\lambda(k+1)\delta n} |\mathcal{R}_n^{k+1}(\varepsilon) \setminus \mathcal{W}| \leq C_{\varepsilon,n}^{2\kappa_0} \sigma^{c_0/4}.$$

It is at this stage that we need to use blocks of words. This is so that we can put the distribution condition on these blocks, so that we can satisfy the main assumption of the exponential sum theorems. To prove Proposition 3.2.2, we need to split the analysis into two parts depending on the distortion distance.

Lemma 3.2.3. *Consider $\mathbf{a} \in \mathcal{R}_n$ fixed. We have that for $e^{-2\lambda n} \leq \sigma \leq C_{\varepsilon,n}^{-4} e^{-2\varepsilon_r \lambda n}$, the set*

$$D_2(\mathbf{a}) := \left\{ (\mathbf{b},\mathbf{c},\mathbf{d}) \in \mathcal{R}_n^3 : \left| \frac{f''_{\mathbf{ab}}(x_{\mathbf{d}})}{f'_{\mathbf{ab}}(x_{\mathbf{d}})} - \frac{f''_{\mathbf{ac}}(x_{\mathbf{d}})}{f'_{\mathbf{ac}}(x_{\mathbf{d}})} \right| \geq \frac{1}{2}\sqrt{\sigma}, |f'_{\mathbf{ab}}(x_{\mathbf{d}}) - f'_{\mathbf{ac}}(x_{\mathbf{d}})| \leq e^{-2\lambda n}\sigma \right\}$$

has size less than or equal to $96C^2 e^{\lambda} C_{\varepsilon,n}^{10\lambda} e^{3\lambda s n} \sigma^{s/2}$.

Remark 3.2.4. *It is worth pointing out the range of σ considered in this lemma. This lemma is not proved for general η here, unlike most other proofs. This is because this proposition will only be used in the Gauss map and convex cocompact Fuchsian group settings. In these situations, we consider $\eta \in J_{\tau(n)}(\varepsilon)$ where $\tau(n) = e^{-\lambda n}$. This will mean that the range of σ considered in Proposition 3.2.2 is sufficient, and moreso the lower bound for σ considered is overkill. To be sure that the upper bound is sufficient, we just have to consider $\varepsilon > 0$ small enough with respect to $\varepsilon_r > 0$, and $\varepsilon_r > 0$ small enough with respect to $\varepsilon_3 > 0$.*

The following proof does get rather combinatorial and messy towards the end, but the idea is straightforward. We end up having to count the number of construction intervals given by regular words that are contained in some interval of a known maximum length. We want a bound from above, so we assume the worst case that the given interval only contains regular intervals. We then compare the upper bound of the length of the interval to the approximate value of a j-regular interval, namely $e^{-\lambda j}$. We then cover this interval in the maximum number of j-regular intervals, and then count these intervals. This will give a maximum number of n-regular intervals.

Proof of Lemma 3.2.3. Consider words $\mathbf{b}$ and $\mathbf{c}$ also to be fixed. In this situation, we only want to count the number of centre points $x_{\mathbf{d}}$.

We begin by rewriting the second condition in the definition of $D_2(\mathbf{a})$. By Lemma 2.2.5, we have that

$$16^{-2}C_{\varepsilon,n}^{-2}e^{-2\lambda n} \leq |I_{\mathbf{ab}}|, |I_{\mathbf{ac}}| \leq C_{\varepsilon,n}^{2}e^{-2\lambda n}$$

and since by the chain rule $|f'_{\mathbf{ab}}(x)| = |f'_{\mathbf{a}}(f_{\mathbf{b}}(x))||f'_{\mathbf{b}}(x)|$ we must have the same bounds for $|f'_{\mathbf{ab}}(x)|$ and $|f'_{\mathbf{ac}}(x)|$. By the Mean Value Theorem, we have that

$$\frac{|\log f'_{\mathbf{ab}}(x_{\mathbf{d}}) - \log f'_{\mathbf{ac}}(x_{\mathbf{d}})|}{|f'_{\mathbf{ab}}(x_{\mathbf{d}}) - f'_{\mathbf{ac}}(x_{\mathbf{d}})|} = \frac{1}{|\xi|} \text{ for some } \xi \in [f'_{\mathbf{ab}}(x_{\mathbf{d}}), f'_{\mathbf{ac}}(x_{\mathbf{d}})] \cup [f'_{\mathbf{ac}}(x_{\mathbf{d}}), f'_{\mathbf{ab}}(x_{\mathbf{d}})].$$

So we have that $|\xi| \in [16^{-2}C_{\varepsilon,n}^{-2}e^{-2\lambda n}, C_{\varepsilon,n}^{2}e^{-2\lambda n}]$. As a result we see that

$$|\log f'_{\mathbf{ab}}(x_{\mathbf{d}}) - \log f'_{\mathbf{ac}}(x_{\mathbf{d}})| \leq \frac{1}{|\xi|}|f'_{\mathbf{ab}}(x_{\mathbf{d}}) - f'_{\mathbf{ac}}(x_{\mathbf{d}})| \leq 16^2C_{\varepsilon,n}^{2}e^{2\lambda n}e^{-2\lambda n}\sigma = 16^2C_{\varepsilon,n}^{2}\sigma.$$

So we have that $D_2(\mathbf{a}) \subset D_2(\mathbf{a},\mathbf{b},\mathbf{c})' \times \mathcal{R}_n^2$ where

$$D_2(\mathbf{a},\mathbf{b},\mathbf{c})' := \left\{\mathbf{d} \in \mathcal{R}_n : \left|\frac{f''_{\mathbf{ab}}(x_{\mathbf{d}})}{f'_{\mathbf{ab}}(x_{\mathbf{d}})} - \frac{f''_{\mathbf{ac}}(x_{\mathbf{d}})}{f'_{\mathbf{ac}}(x_{\mathbf{d}})}\right| \geq \frac{1}{2}\sqrt{\sigma}, \left|\log\frac{f'_{\mathbf{ab}}(x_{\mathbf{d}})}{f'_{\mathbf{ac}}(x_{\mathbf{d}})}\right| \leq 4^4C_{\varepsilon,n}^{2}\sigma\right\}$$

By Lemma 3.2.1, we have that the set of centre points $x_{\mathbf{d}}$ corresponding to the regular words in $D_2(\mathbf{a},\mathbf{b},\mathbf{c})'$ must be contained in an interval J of length at most $4^5C_{\varepsilon,n}^{2}\sqrt{\sigma}$. Note that instead of counting the centre points $x_{\mathbf{d}}$, we can instead count the number of corresponding intervals $I_{\mathbf{d}}$. However, it is important to note that there might exist at most two intervals $I_{\mathbf{d}}$ whose centre points do lie in J, but the intervals themselves are not entirely contained in J. If we were to cover J with j-parents of n-regular intervals who length are at least $e^{(-\lambda-\varepsilon)j}/4$, then in the 'worst case' (when J does not contain any irregular geometric points), then the number K of j-parent covering sets would satisfy the last inequality in the following

$$4^5C_{\varepsilon,n}^{2}\sqrt{\sigma} \leq \frac{4^5}{96}e^{-\lambda j-3\varepsilon n/2} \leq 11e^{-\lambda j} \leq \frac{K}{4}e^{-\lambda j-\varepsilon j}.$$

So we can sufficiently choose a $K \geq 44e^{\varepsilon\lambda n} \geq 44e^{\varepsilon j}$, for example $K = \lceil 44e^{\varepsilon\lambda n}\rceil \leq 48C_{\varepsilon,n}^{\lambda}$ where the inequality is true if we assume that n is large enough so that $e^{\varepsilon\lambda n} \geq 1$.

Given a j-parent $I_{\mathbf{d}_j}$ in the cover, we now approximate the number of n-regular intervals $I_{\mathbf{d}}$ (corresponding to the number of regular words which we wanted originally) contained in this set. We will say that $\mathbf{a} \prec \mathbf{b}$ if and only if $I_{\mathbf{b}} \subset I_{\mathbf{a}}$, that is there exists a finite word $\mathbf{c}$ such that $\mathbf{b} = \mathbf{ac}$. We see that by Lemma 2.2.3

$$\#\{\mathbf{d} \in \mathcal{R}_n : \mathbf{d}_j \prec \mathbf{d}\} C^{-1} e^{-\lambda sn} C_{\varepsilon,n}^{-3\lambda} \leq \mu\Big(\bigcup_{\mathbf{d} \in \mathcal{R}_n : \mathbf{d}_j \prec \mathbf{d}} I_{\mathbf{d}} \Big) \leq \mu(I_{\mathbf{d}_j}) \leq C e^{(-s\lambda + 3\lambda\varepsilon)j}$$

so we get that

$$\#\{\mathbf{d} \in \mathcal{R}_n : \mathbf{d}_j \prec \mathbf{d}\} \leq C_{\varepsilon,n}^{3\lambda} C^2 e^{\lambda sn} e^{(-s\lambda + 3\lambda\varepsilon)j}.$$

So to conclude, we get that

$$\#D_2(\mathbf{a}, \mathbf{b}, \mathbf{c})' \leq K C_{\varepsilon,n}^{3\lambda} C^2 e^{\lambda sn} e^{(-s\lambda + 3\lambda\varepsilon)j} \leq 96^2 C^2 e^{\lambda} C_{\varepsilon,n}^{11\lambda} e^{\lambda sn} \sigma^{s/2}$$

as required. □

Now using Lemma 3.2.3, and assuming that T has non-concentrated derivative, we can prove Proposition 3.2.2.

Proof of Proposition 3.2.2. Consider $l \in \mathbb{Z}$ such that $e^{-\lambda n} \leq 2^{-l} \leq 2e^{-\lambda\varepsilon_3 n/4}$, noting that only finitely many such l exist. Define $\mathcal{R}_l^*$ to be the set of n-regular pairs $(\mathbf{a}, \mathbf{d}) \in \mathcal{R}_n^2$ such that

$$e^{-2\lambda sn} \#\{(\mathbf{b}, \mathbf{c}) \in \mathcal{R}_n^2 : |f'_{\mathbf{ab}}(x_{\mathbf{d}}) - f'_{\mathbf{ac}}(x_{\mathbf{d}})| \leq e^{-2\lambda n} 2^{-l}\} \leq 2^{-(l+1)s/4}.$$

For every $\sigma \in [e^{-\lambda n}, e^{-\lambda\varepsilon_3 n/4}]$ there is a unique l such that $2^{-l-1} \leq \sigma \leq 2^{-l}$. In this setting, if we have a block $\mathbf{A}$ such that $(\mathbf{a}_{j-1}, \mathbf{a}_j) \in \mathcal{R}_l^*$ for every $j = 1, \ldots, k$ and every l, then by definition of $\mathcal{R}_l^*$ and by definition of $\zeta_{j,\mathbf{A}}(\mathbf{b})$ we have that

$$\begin{aligned} &e^{-2\lambda sn} \#\{(\mathbf{b}, \mathbf{c}) \in \mathcal{R}_n^2 : |\zeta_{j,\mathbf{a}}(\mathbf{b}) - \zeta_{j,\mathbf{a}}(\mathbf{c})| \leq \sigma\} \\ &\leq e^{-2\lambda sn} \#\{(\mathbf{b}, \mathbf{c}) \in \mathcal{R}_n^2 : |f'_{\mathbf{a}_{j-1}\mathbf{b}}(\mathbf{a}_j) - f'_{\mathbf{a}_{j-1}\mathbf{c}}(\mathbf{a}_j)| \leq e^{-2\lambda n} 2^{-l}\} \leq 2^{-(l+1)s/4} \leq \sigma^{s/4}. \end{aligned}$$

This therefore tells us that

$$\bigcap_j \bigcap_l \{\mathbf{A} : (\mathbf{a}_{j-1}, \mathbf{a}_j) \in \mathcal{R}_l^*\} \subset \mathcal{W}.$$

From this containment, we can say that a $k+1$ block $\mathbf{A}$ is not in $\mathcal{W}$ if there exists at least one position j in the block and a scale l such that the pair $(\mathbf{a}_{j-1}, \mathbf{a}_j) \notin \mathcal{R}^*$. So to prove the lemma, it is enough to show that $e^{-2\lambda sn} \#\{\mathcal{R}_n^2 \setminus \mathcal{R}_l^*\} \leq a C_{\varepsilon,n}^b e^{-\lambda\varepsilon_3 sn/16}$ for some $a, b > 0$. We achieve this bound by considering the counting measure $\sharp$ on pairs in $\mathcal{R}_n^2$ and use Chebychev's inequality to get an upper bound on $\#\mathcal{R}_n^2 \setminus \mathcal{R}_l^*$. We apply Chebyshev's inequality to the counting function defined by

$$f(\mathbf{a}, \mathbf{d}) = e^{-2\lambda sn} \#\{(\mathbf{b}, \mathbf{c}) \in \mathcal{R}_n^2 : |f'_{\mathbf{ab}}(x_{\mathbf{d}}) - f'_{\mathbf{ac}}(x_{\mathbf{d}})| \leq e^{-2\lambda n} 2^{-l}\}$$

which gives us that

$$\begin{aligned} |\{\mathcal{R}_n^2 \setminus \mathcal{R}_l^*\}| &= \sharp\{(\mathbf{a}, \mathbf{d}) \in \mathcal{R}_n^2 : |f(\mathbf{a}, \mathbf{d})| \geq 2^{-(l+1)s/4}\} \leq 2^{(l+1)s/4} \int_{\mathcal{R}_n^2} |f| \, d\sharp \\ &= e^{-2\lambda sn} 2^{(l+1)s/4} \sharp\{(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) \in \mathcal{R}_n^4 : |f'_{\mathbf{ab}}(x_{\mathbf{d}}) - f'_{\mathbf{ac}}(x_{\mathbf{d}})| \leq e^{-2\lambda n} 2^{-l}\}. \end{aligned}$$

By using the bound for $\sharp\mathcal{R}_n$ (recall Lemma 2.2.3(iv)) and the assumption that T has non-concentrated derivative, we have the claim. □

3.3 From Fourier transforms to Exponential Sums

To get a desired bound on the Fourier transform of μ, we use the ideas of Bourgain–Dyatlov [8, Chapter 3] to be able to reduce our problem to studying exponential sums:

Lemma 3.3.1. *Consider $\tau : \mathbb{N} \to \mathbb{R}$, an exponentially decreasing function with $\tau > 0$. Define*

$$J_{\tau(n)}(\varepsilon) := \{\eta \in \mathbb{R} : \tau(n)^{-1/4} \leq |\eta| \leq C_{\varepsilon,n} \tau(n)^{-1/2}\}$$

to be a corresponding set of frequencies. Assume that there exists some $b > 0$ such that

$$\mu \times \mu(\{(x, y) \in X^2 : |x - y| \leq C_{\varepsilon,n} \tau(n)^{1/4}\}) \leq \tau(n)^b.$$

Then we have that the size of the generalised Fourier transform squared is bounded from above by

$$C_{\varepsilon,n}^{9\lambda(2k+2)} e^{-\lambda(2k+1)sn} \sum_{\mathbf{A} \in \mathcal{R}_n^{k+1}} \sup_{\eta \in J_{\tau(n)}(\varepsilon)} \left| \sum_{\mathbf{B} \in \mathcal{R}_n^k} e^{2\pi i \eta \zeta_{1,\mathbf{A}}(\mathbf{b_1}) \dots \zeta_{k,\mathbf{A}}(\mathbf{b}_k)} \right| + C_R \mu(X \setminus R_n^{k+1})^2 + E(n)$$

for some $C_R > 0$, where $E(n)$ is an approximation term which decays exponentially with respect to n for small enough ε, namely

$$E(n) := C_1 \delta^n + 64 C' e^{\lambda} C_0 C_{\varepsilon,n} e^{-\lambda sn/4} + \tau(n)^b$$

for some $C_1 > 0$.

Remark 3.3.2. *We shall present the proof for the case when we have a finite alphabet. The Gauss map case is presented in [53]. The proof presented there can also be extended to countable Markov maps whose transition matrix has no entries equal to zero. In the Gauss map case, the transfer operator is global over the space X, and does not differ over construction intervals I_b for $b \in \mathcal{A}$ because all words are admissible. In the finite alphabet case, the always-admissible property is unnecessary. This will be proven to be the case by slightly modifying the proof presented in [54], which follows the framework of Bourgain–Dyatlov [8].*

Fix a frequency $\xi \in \mathbb{R}$. Assume we are given some $\tau(n) > 0$ which is exponentially decreasing with n. Let $n \in \mathbb{N}$ be the number such that

$$\tau(n)^{-1/2} e^{(2k+1)\lambda n} \leq \xi \leq \tau(n+1)^{-1/2} e^{(2k+1)\lambda(n+1)}.$$

Recall that

$$\#\mathcal{R}_n \leq C C_{\varepsilon,n}^{3\lambda} e^{\lambda s n}$$

and if $\mathbf{a} \in \mathcal{R}_n$, we have that for all $x \in X$

$$w_{\mathbf{a}}(x) \leq C_{\varepsilon,n}^{3\lambda} e^{-\lambda s n}.$$

Write

$$h(x) := \exp(2\pi i \xi x).$$

Given a word $\mathbf{a}$, we shall define $x_{\mathbf{a}} \in I_{\mathbf{a}}$ to be the centre point of this construction interval. Recall the definition of blocks of words $\mathbf{A}$ and $\mathbf{B}$ defined in the beginning of Section 2.1. Given $j \in \mathbb{N}_k$ and some regular word $\mathbf{b} \in \mathcal{R}_n$, we define the real number

$$\zeta_{j,\mathbf{A}}(\mathbf{b}) := e^{2\lambda n} f'_{\mathbf{a_{j-1}} \mathbf{b}}(x_{\mathbf{a_j}}).$$

By the chain rule we have that

$$\zeta_{j,\mathbf{A}}(\mathbf{b}) = e^{2\lambda n} f'_{\mathbf{a_{j-1}}}(f_{\mathbf{b}} x_{\mathbf{a_j}}) f'_{\mathbf{b}}(x_{\mathbf{a}_j}).$$

Hence by Lemma 2.2.3 we have that

Lemma 3.3.3.

$$\zeta_{j,\mathbf{A}}(\mathbf{b}) \in [16^{-2} C_{\varepsilon,n}^{-2}, C_{\varepsilon,n}^{2}]$$

This holds because $f'_{\mathbf{a}_{j-1}}$ and $f'_{\mathbf{b}}$ must both be either positive or negative because they are defined by words of the same length. Later, we will use the fact that

$$\zeta_{j,\mathbf{A}}(\mathbf{b}) \in [16^{-2}C^{-1}C_{\varepsilon,n}^{-3\lambda}, 16^2 C C_{\varepsilon,n}^{3\lambda}]$$

where C is the Gibbs constant. This is basically so that we have a universal C_0 that can be used in our extension of Bourgain–Dyatlov's exponential sum theorem.

Since μ is invariant under the transfer operator $\mathcal{L}_\varphi^*$:

$$\mu = \mathcal{L}_\varphi^* \mu$$

we obtain immediately the following estimate

Lemma 3.3.4. *Recall the definition of R_n^k in Definition 2.2.2. For $h(x) := e^{-2\pi i \xi \Phi(x)} g(x)$ we have that*

$$\Big|\int h(x)\, d\mu(x)\Big|^2 \leq 2\Big|\sum_{\mathbf{A},\mathbf{B}} \sum_{b:\mathbf{A}\to b} \int h(f_{\mathbf{A}*\mathbf{B}}(x)) w_{\mathbf{A}*\mathbf{B}}(x)\, d\mu(x)\Big|^2 + 4C^2 C_g^2 \mu(X \setminus R_n^{k+1})^2$$

where the sum is over $\mathbf{A} \in \mathcal{R}_n^{k+1}$ and $\mathbf{B} \in \mathcal{R}_n^k$.

Proof. By the invariance of the transfer operator

$$\int_X h(x)\, d\mu(x) = \int_X \mathcal{L}_\varphi^{(2k+1)n} h(x)\, d\mu(x) = \sum_{b \in \mathcal{A}} \int_{I_b} (\mathcal{L}_\varphi^n)^{2k+1} h(x)\, d\mu(x).$$

This splits using $\mathcal{R}_n^k$ and $(\mathcal{A}_n)^k \setminus \mathcal{R}_n^k$ to

$$\Big|\sum_{b:\mathbf{A}\to b} \int_{I_b} \sum_{\substack{\mathbf{A}\in\mathcal{R}_n^{k+1} \\ \mathbf{B}\in\mathcal{R}_n^k}} w_{\mathbf{A}*\mathbf{B}}(x) h(f_{\mathbf{A}*\mathbf{B}}x)\, d\mu\Big| + \Big|\sum_{b:\mathbf{A}\to b} \int_{I_b} \sum_{\substack{\mathbf{A}\in(\mathcal{A}_n)^{k+1}\setminus\mathcal{R}_n^{k+1} \\ \text{or } \mathbf{B}\in(\mathcal{A}_n)^k\setminus\mathcal{R}_n^k}} w_{\mathbf{A}*\mathbf{B}}(x) h(f_{\mathbf{A}*\mathbf{B}}x)\, d\mu\Big|.$$

We shall bound the right hand side by considering that

$$\begin{aligned}
&\Big|\sum_{b:\mathbf{A}\to b} \int_{I_b} \sum_{\substack{\mathbf{A}\in(\mathcal{A}_n)^{k+1}\setminus\mathcal{R}_n^{k+1} \\ \text{or } \mathbf{B}\in(\mathcal{A}_n)^k\setminus\mathcal{R}_n^k}} w_{\mathbf{A}*\mathbf{B}}(x) h(f_{\mathbf{A}*\mathbf{B}}x)\, d\mu\Big| \\
&\leq \sum_{b:\mathbf{A}\to b} \int_{I_b} \sum_{\substack{\mathbf{A}\in(\mathcal{A}_n)^{k+1}\setminus\mathcal{R}_n^{k+1} \\ \text{or } \mathbf{B}\in(\mathcal{A}_n)^k\setminus\mathcal{R}_n^k}} w_{\mathbf{A}*\mathbf{B}}(x) |g(f_{\mathbf{A}*\mathbf{B}}x)|\, d\mu(x) \\
&\leq \sum_{b:\mathbf{A}\to b} \sum_{\substack{\mathbf{A}\in(\mathcal{A}_n)^{k+1}\setminus\mathcal{R}_n^{k+1} \\ \text{or } \mathbf{B}\in(\mathcal{A}_n)^k\setminus\mathcal{R}_n^k}} C C_g \mu(I_{\mathbf{A}*\mathbf{B}}) \\
&\leq (\#\mathcal{A}) C C_g \mu(X \setminus R_n^{k+1}) + (\#\mathcal{A}) C C_g \mu(X \setminus R_n^k) \\
&\leq 2(\#\mathcal{A}) C C_g \mu(X \setminus R_n^{k+1})
\end{aligned}$$

where we use the assumption that $||g||_{C^1} \leq C_g$. We get the required result by noting that $R_n^{k+1} \subset R_n^k$, which follows by the fact that for any $\mathbf{A} \in \mathcal{R}_n^{k+1}$ we have that there exists $\mathbf{B} \in \mathcal{R}_n^k$ such that $\mathbf{B} \prec \mathbf{A}$. Conclude using $|a+b|^2 \leq 2|a|^2 + 2|b|^2$ for complex numbers. □

From now on we will always consider blocks $\mathbf{A} \in \mathcal{R}_n(\varepsilon)^{k+1}$ and $\mathbf{B} \in \mathcal{R}_n(\varepsilon)^k$. This is because all other blocks can be ignored by Lemma 3.3.4. To control the above sums, we will rely on the local variation assumption of the potential φ defining the Gibbs measure and the bounded distortion assumption on T. First of all, since the distortion $|(\log f_{\mathbf{a}}')'(z)|$ is uniformly bounded over $\mathbf{a} \in \mathbb{N}^n$ and $z \in X$, we obtain the following:

Lemma 3.3.5. *For the value $\delta > 0$ coming from the local variation assumption of the potential φ, we have that*

$$\Big| \int h \, d\mu \Big|^2 \leq C C_{\varepsilon,n}^{9\lambda(2k+1)} e^{-(2k-1)\lambda sn} \sum_{b:\mathbf{A}\to b} \sum_{\mathbf{A},\mathbf{B}} \Big| \int_{I_b} e^{i\xi\Phi(f_{\mathbf{A}*\mathbf{B}}(x))} w_{\mathbf{a}_k}(x) d\mu(x) \Big|^2 + E_1$$

where

$$E_1 := C_1 C_{\varepsilon,n}^{9\lambda(2k+2)} \delta^n + 64 C' e^{\lambda} C_0 C_{\varepsilon,n} e^{-\lambda sn/4} + C_R \mu(X \setminus R_n^{k+1})^2$$

for some constants $C', C_R, C_1 > 0$ and sufficiently small $\varepsilon > 0$.

Part of the proof requires us to compare a Birkhoff weight evaluated at two difference points in the same regular interval. We do this using Hölder regularity of the measure's potential.

Proof. Choose a point $y \in X$ such that $x_{\mathbf{a}_k} = f_{\mathbf{a}_k}(y)$. Then we have that

$$\frac{w_{\mathbf{A}\#\mathbf{B}}(f_{\mathbf{a}_k} x)}{w_{\mathbf{A}\#\mathbf{B}}(x_{\mathbf{a}_k})} = \exp(S_{2kn}\varphi(f_{\mathbf{A}*\mathbf{B}}(x)) - S_{2kn}\varphi(f_{\mathbf{A}*\mathbf{B}}(y))).$$

Since φ is locally Hölder, we know that there exists a constant $C > 0$ and $0 < \delta < 1$ such that for any $m \in \mathbb{N}$ we have

$$\sup_{\mathbf{w} \in \mathbb{N}^m} \sup\{|\varphi(u) - \varphi(v)| : u, v \in I_{\mathbf{w}}\} \leq C\delta^m.$$

This gives as $|\mathbf{A} * \mathbf{B}| = (2k+1)n$ that

$$|S_{2kn}\varphi(f_{\mathbf{A}*\mathbf{B}}(x)) - S_{2kn}\varphi(f_{\mathbf{A}*\mathbf{B}}(y))| \leq \sum_{j=0}^{2kn-1} C\delta^{2kn+n-i} \leq \frac{C}{1-\delta}\delta^{n+1} =: C_0\delta^{n+1}.$$

Hence

$$\exp(-C_0\delta^n) \leq \frac{w_{\mathbf{A}\#\mathbf{B}}(f_{\mathbf{a}_k}x)}{w_{\mathbf{A}\#\mathbf{B}}(x_{\mathbf{a}_k})} \leq \exp(C_0\delta^n).$$

Rearranging this result we have that

$$|w_{\mathbf{A}\#\mathbf{B}}(f_{\mathbf{a}_k}x) - w_{\mathbf{A}\#\mathbf{B}}(x_{\mathbf{a}_k})| \leq \max\{|\exp(\pm C_0\delta^n) - 1|\}w_{\mathbf{A}\#\mathbf{B}}(x_{\mathbf{a}_k}).$$

Since $|e^{i\xi\varphi(x)}| \leq 1$ we have by the triangle inequality

$$\begin{aligned}
&|h(f_{\mathbf{A}*\mathbf{B}}x) - e^{i\xi\varphi(f_{\mathbf{A}*\mathbf{B}}x)}g(x_{\mathbf{a}_0})| \leq |g(f_{\mathbf{A}*\mathbf{B}}x) - g(x_{\mathbf{a}_0})| \\
&\leq C_g|f_{\mathbf{A}*\mathbf{B}}(x) - x_{\mathbf{a}_0}| \leq C_g C_{\varepsilon,n} e^{-\lambda n}
\end{aligned}$$

where the last inequality holds since $f_{\mathbf{A}*\mathbf{B}}(x), x_{\mathbf{a}_0} \in I_{\mathbf{a}_0}$, and the previous holds because g is Lipschitz (by assumption). Therefore by defining the constant weight $g_{\mathbf{A}\#\mathbf{B}} := w_{\mathbf{A}\#\mathbf{B}}(x_{\mathbf{a}_k})g(x_{\mathbf{a}_0})$ we have that for some $C > 0$,

$$\begin{aligned}
&|h(f_{\mathbf{A}*\mathbf{B}}x)w_{\mathbf{A}*\mathbf{B}}(x) - g_{\mathbf{A}\#\mathbf{B}}e^{i\xi\Phi(f_{\mathbf{A}*\mathbf{B}}x)}w_{\mathbf{a}_k}(x)| \\
&\leq |h(f_{\mathbf{A}*\mathbf{B}}x) - e^{i\xi\Phi(f_{\mathbf{A}*\mathbf{B}}x)}g(x_{\mathbf{a}_0})w_{\mathbf{A}*\mathbf{B}}(x)| \\
&+ |g(x_{a_0})| \cdot |w_{\mathbf{a}_k}(x)| \cdot |w_{\mathbf{A}\#\mathbf{B}}(f_{\mathbf{a}_k}x) - w_{\mathbf{A}\#\mathbf{B}}(x_{\mathbf{a}_k})| \\
&\leq CC_{\varepsilon,n}^{3\lambda(2k+1)+1}e^{-(2k+1)\lambda sn - \lambda n} + CC_{\varepsilon n}^{3(2k+1)\lambda}e^{-(2k+1)\lambda sn}\delta_{LH}^n
\end{aligned}$$

where we use that fact that $w_{\mathbf{A}*\mathbf{B}}(x) = w_{\mathbf{A}\#\mathbf{B}}(f_{\mathbf{a}_k}(x))w_{\mathbf{a}_k}(x)$. Comparing this with the integral on the right hand side of Lemma 3.3.4 we see that

$$\begin{aligned}
&\Big|\sum_{\mathbf{A},\mathbf{B}}\sum_{b:\mathbf{A}\to b}\int_{I_b} h(f_{\mathbf{A}*\mathbf{B}}(x))w_{\mathbf{A}*\mathbf{B}}(x)\,d\mu(x) - \sum_{\mathbf{A},\mathbf{B}} g_{\mathbf{A}\#\mathbf{B}}\sum_{b:\mathbf{A}\to b}\int_{I_b} h(f_{\mathbf{A}*\mathbf{B}}x)w_{\mathbf{a}_k}(x)\,d\mu(x)\Big| \\
&\leq \sum_{\mathbf{A},\mathbf{B}}\sum_{b:\mathbf{A}\to b}\int_{I_b} |h(f_{\mathbf{A}*\mathbf{B}}x)w_{\mathbf{A}*\mathbf{B}}(x) - g_{\mathbf{A}\#\mathbf{B}}h(f_{\mathbf{A}*\mathbf{B}}x)w_{\mathbf{a}_k}(x)|\,d\mu(x) \\
&\leq (\#\mathcal{A})\sum_{\mathbf{A},\mathbf{B}} CC_{\varepsilon,n}^{9\lambda}e^{-(2k+1)\lambda sn}\cdot(e^{-\lambda n} + \delta^n) \leq C_1 C_{\varepsilon,n}^{9\lambda(2k+2)}(\delta^{n/2} + e^{-\lambda n/2})
\end{aligned}$$

by choice of ε with respect to δ. Using Hölder's inequality we get that

$$\begin{aligned}
&\Big|\sum_{\mathbf{A},\mathbf{B}} w_{\mathbf{A}\#\mathbf{B}}(x_{\mathbf{a}_k})\sum_{b:\mathbf{A}\to b}\int_{I_b} h(f_{\mathbf{A}*\mathbf{B}}x)w_{\mathbf{a}_k}(x)\,d\mu(x)\Big|^2 \\
&\leq CC_{\varepsilon,n}^{9\lambda(2k+1)}e^{-(2k-1)\lambda sn}\sum_{\mathbf{A},\mathbf{B}}\sum_{b:\mathbf{A}\to b}\Big|\int_{I_b} h(f_{\mathbf{A}*\mathbf{B}}x)w_{\mathbf{a}_k}(x)\,d\mu(x)\Big|^2
\end{aligned}$$

Using $|a+b|^2 \leq 2|a|^2 + 2|b|^2$ for $a, b \in \mathbb{C}$, we get the result. □

Now we are ready to finish the proof of Lemma 3.3.1. The end of the following proof again uses the idea of approximating a set using regular intervals to be able to get a sufficiently decaying bound on the measure of the set.

Proof of Lemma 3.3.1. First of all,

$$|\widehat{\mu}(\xi)|^2 \le CC_{\varepsilon,n}^{9\lambda(2k+1)}e^{-\lambda(2k-1)sn}\sum_{\mathbf{A},\mathbf{B}}\sum_{b:\mathbf{A}\to b}\Big|\int_{I_b} h(f_{\mathbf{A}*\mathbf{B}}(x))w_{\mathbf{a}'_k}(x)\,d\mu(x)\Big|^2 + E_1.$$

The first term on the right-hand side of the above inequality is

$$\sum_{\mathbf{A},\mathbf{B}}\sum_{b:\mathbf{A}\to b}\Big|\int_{I_b} h(f_{\mathbf{A}*\mathbf{B}}(x))w_{\mathbf{a}'_k}(x)\,d\mu(x)\Big|^2,$$

which, when opening up, is equal to

$$\sum_{\mathbf{A}}\sum_{b:\mathbf{A}\to b}\int_{I_b^2} w_{\mathbf{a}'_k}(x)w_{\mathbf{a}'_k}(y)\sum_{\mathbf{B}} e^{2\pi i\xi(\Phi\circ f_{\mathbf{A}*\mathbf{B}}(x)-\Phi\circ f_{\mathbf{A}*\mathbf{B}}(x))}\,d\mu(x)\,d\mu(y).$$

Taking absolute values, and using the bound for $w_{\mathbf{a}'_k}(x)$, this is bounded from above by

$$CC_{\varepsilon,n}^{6\lambda}e^{-2\lambda sn}\sum_{\mathbf{A}}\sum_{b:\mathbf{A}\to b}\int_{I_b^2}\Big|\sum_{\mathbf{B}} e^{2\pi i\xi(\Phi\circ f_{\mathbf{A}*\mathbf{B}}(x)-\Phi\circ f_{\mathbf{A}*\mathbf{B}}(x))}\Big|\,d\mu(x)\,d\mu(y).$$

Consider a fixed block $\mathbf{A}$. Given $x, y \in X$, define $\widehat{x} := f_{\mathbf{a}_k}(x)$ and $\widehat{y} := f_{\mathbf{a}_k}(y)$ both of which are in $I_{\mathbf{a}_k}$. We also have that $f_{\mathbf{A}*\mathbf{B}}(x) = f_{\mathbf{A}\#\mathbf{B}}(\widehat{x})$ and $f_{\mathbf{A}*\mathbf{B}}(y) = f_{\mathbf{A}\#\mathbf{B}}(\widehat{y})$. By the Fundamental Theorem of Calculus we have that

$$\Phi\circ f_{\mathbf{A}*\mathbf{B}}(y) - \Phi\circ f_{\mathbf{A}*\mathbf{B}}(x) = \int_{\widehat{x}}^{\widehat{y}}(\Phi\circ f_{\mathbf{A}\#\mathbf{B}})'(t)dt.$$

By applying the Chain rule k times, we have that there exists $t_i \in I_{\mathbf{a}_i}$ for $i = 1,\dots,k$ such that

$$(\Phi\circ f_{\mathbf{A}\#\mathbf{B}})'(t) = \Phi'(f_{\mathbf{A}\#\mathbf{B}}t)f'_{\mathbf{a}_0\mathbf{b}_1}(t_1)f'_{\mathbf{a}_1\mathbf{b}_2}(t_2)\dots f'_{\mathbf{a}_{k-1}\mathbf{b}_k}(t_k)$$

where $t_k = t$. Lemma 2.0.1 gives us that

$$\exp(-2|x_{\mathbf{a}_i} - t_i|) \le \frac{f'_{\mathbf{a}_{i-1}\mathbf{b}_i}(t_i)}{e^{-2\lambda n}e^{2\lambda n}f'_{\mathbf{a}_{i-1}\mathbf{b}_i}(x_{\mathbf{a}_i})} \le \exp(2|t_i - x_{\mathbf{a}_i}|)$$

where the upper bound is direct, but the lower bound is achieved by swapping $x_{\mathbf{a}_i}$ and t_i in the lemma. We also have that $|x_{\mathbf{a}_i} - t_i| \le C_{\varepsilon,n}e^{-\lambda n}$ because both points are in $I_{\mathbf{a}_i}$. Hence using the definition of $\zeta_{i,\mathbf{A}}(\mathbf{b}_i)$ we have that

$$\exp(-2kC_{\varepsilon,n}e^{-\lambda n}) \le \frac{(\Phi\circ f_{\mathbf{A}\#\mathbf{B}})'(t)}{\Phi'(x_{\mathbf{a}_0})e^{-2k\lambda n}\zeta_{1,\mathbf{A}}(\mathbf{b}_1)\dots\zeta_{k,\mathbf{A}}(\mathbf{b}_k)} \le \exp(2kC_{\varepsilon,n}e^{-\lambda n}).$$

We shall denote the denominator of the above fraction by P_k to see that by rearranging we have

$$[\exp(-2kC_{\varepsilon,n}e^{-\lambda n}) - 1]P_k \leq (\Phi \circ f_{\mathbf{A}\#\mathbf{B}})'(t) - P_k \leq [\exp(2kC_{\varepsilon,n}e^{-\lambda n}) - 1]P_k.$$

So by integrating between $\widehat{y}$ and $\widehat{x}$ we get that

$$\begin{aligned}[\exp(-2kC_{\varepsilon,n}e^{-\lambda n}) - 1]P_k(\widehat{y} - \widehat{x}) &\leq \Phi \circ f_{\mathbf{A}*\mathbf{B}}(x) - \Phi \circ f_{\mathbf{A}*\mathbf{B}}(y) - P_k(\widehat{y} - \widehat{x}) \\ &\leq [\exp(2kC_{\varepsilon,n}e^{-\lambda n}) - 1]P_k(\widehat{y} - \widehat{x}).\end{aligned}$$

Since $\widehat{y}, \widehat{x} \in I_{\mathbf{a}_k}$ and $\zeta_{i,\mathbf{A}} \in [C_{\varepsilon,n}^{-2}, C_{\varepsilon,n}^{2}]$, we have that $|P_k| \leq C_{\varepsilon,n}^{k} e^{-2k\lambda n}$ and so

$$|\Phi \circ f_{\mathbf{A}*\mathbf{B}}(x) - \Phi \circ f_{\mathbf{A}*\mathbf{B}}(y) - P_k(\widehat{y} - \widehat{x})| \leq e^{2k} C_{\varepsilon,n}^{k+2} e^{-(2k+2)\lambda n}.$$

We define

$$\eta := \frac{\xi}{2\pi} \cdot e^{-2k\lambda n}\Phi'(x_{\mathbf{a}_0})(\widehat{x} - \widehat{y}).$$

By the Mean Value Theorem and using the regularity bounds on $|f'_{\mathbf{a}_k}|$ we get that $C_{\varepsilon,n}^{-1}e^{-\lambda n}|x-y| \leq |\widehat{x} - \widehat{y}| \leq C_{\varepsilon,n}e^{-\lambda n}|x-y|$ and hence we have that

$$C_{\varepsilon,n}^{-1}\tau^{-1/2}|x-y| \leq |\eta| \leq C_{\varepsilon,n}\tau^{-1/2}$$

where we use the fact that I is bounded for the upper bound. Using the fact that the map $x \to e^{ix}$ is Lipschitz, we get that

$$\begin{aligned}&\Big|\sum_{\mathbf{B}} e^{2\pi i\xi(\Phi\circ f_{\mathbf{A}*\mathbf{B}}(x) - \Phi\circ f_{\mathbf{A}*\mathbf{B}}(x))}\Big| \\ &\leq \Big|\sum_{\mathbf{B}} e^{2\pi i\zeta_{1,\mathbf{a}}(\mathbf{b_1})\ldots\zeta_{k,\mathbf{a}}(\mathbf{b}_k)}\Big| + \Big|\sum_{\mathbf{B}} e^{2\pi i\xi(\Phi\circ f_{\mathbf{A}*\mathbf{B}}(x) - \Phi\circ f_{\mathbf{A}*\mathbf{B}}(x))} - e^{2\pi i\zeta_{1,\mathbf{a}}(\mathbf{b_1})\ldots\zeta_{k,\mathbf{a}}(\mathbf{b}_k)}\Big| \\ &\leq \Big|\sum_{\mathbf{B}} e^{2\pi i\zeta_{1,\mathbf{a}}(\mathbf{b_1})\ldots\zeta_{k,\mathbf{a}}(\mathbf{b}_k)}\Big| \\ &+ \sum_{\mathbf{B}} |2\pi\xi(\Phi \circ f_{\mathbf{A}*\mathbf{B}}(x) - \Phi \circ f_{\mathbf{A}*\mathbf{B}}(x)) - 2\pi\eta\zeta_{1,\mathbf{a}}(\mathbf{b_1})\ldots\zeta_{k,\mathbf{a}}(\mathbf{b}_k)| \\ &\leq \Big|\sum_{\mathbf{B}} e^{2\pi i\zeta_{1,\mathbf{a}}(\mathbf{b_1})\ldots\zeta_{k,\mathbf{a}}(\mathbf{b}_k)}\Big| + \sum_{\mathbf{B}} 2\pi e^{-\lambda n/2} \\ &\leq \Big|\sum_{\mathbf{B}} e^{2\pi i\zeta_{1,\mathbf{a}}(\mathbf{b_1})\ldots\zeta_{k,\mathbf{a}}(\mathbf{b}_k)}\Big| + 2\pi C^k C_{\varepsilon,n}^{3k\lambda} e^{(ks-1/2)\lambda n}\end{aligned}$$

This gives us that the generalised transform is bounded from above by

$$C_{\varepsilon,n}^{9\lambda(2k+2)}e^{-\lambda(2k+1)sn} \sum_{\mathbf{A}\in\mathcal{R}_n^{k+1}} \sum_{b:\mathbf{A}\to b} \int_{I_b^2} \Big|\sum_{\mathbf{B}} e^{2\pi i\eta\zeta_{1,\mathbf{a}}(\mathbf{b_1})\ldots\zeta_{k,\mathbf{a}}(\mathbf{b}_k)}\Big| \, d\mu(x)\, d\mu(y) + E_3$$

where

$$E_3 := C' C_{\varepsilon,n}^{(2k+1)\lambda} e^{-\lambda n/2} + C_1 \delta^n + 64 C' e^{\lambda} C_0 C_{\varepsilon,n} e^{-\lambda s n/4} + C_R \mu(X \setminus R_n^{k+1})^2.$$

By the assumption on $\mu \times \mu$, we can just consider our double integral where $|x-y| \geq C_{\varepsilon,n} e^{-\lambda n/4}$, which in turn gives us that $\eta \in J_{\tau(n)}(\varepsilon)$. We conclude using the crude containment which says that $R_n^{k+1} \subset R_n$, and the finiteness of $\mathcal{A}$. □

Chapter 4

The Gauss Map

4.1 Preliminaries

Given a finite word consisting of natural numbers $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{N}^n$ for some $n \in \mathbb{N}$, define its *continued fraction* to be

$$[\mathbf{a}] := [a_1, a_2, \ldots, a_n] := \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ldots + \cfrac{1}{a_n}}}}.$$

We use the fact given to us by number theory that for each irrational number $x \in [0, 1] \setminus \mathbb{Q}$, we can find a unique sequence of numbers $a_i(x) \in \mathbb{N}$ such that

$$x = \lim_{n \to \infty} [a_1(x), a_2(x), \ldots, a_n(x)].$$

We obtain an identification of the set of irrational numbers in $[0, 1] \setminus \mathbb{Q}$ and countable words consisting of natural numbers given by $(a_1, a_2, \ldots) \in \mathbb{N}^{\mathbb{N}}$.

Definition 4.1.1 (The Gauss Map)**.** *Define the Gauss map $T : [0, 1] \to [0, 1]$ as follows*

$$T(x) = \begin{cases} \frac{1}{x} \bmod 1 & x \in (0, 1] \\ 0 & x = 0. \end{cases}$$

For fixed $n \in \mathbb{N}$, the Gauss map is bijective on the intervals $I_n = (\frac{1}{n+1}, \frac{1}{n}]$, so we can consider the inverse of $T|_{I_n}(x) = \frac{1}{x} - n$; the inverse $f_n : [0,1] \to I_n$ is given by

$$f_n(x) := \frac{1}{x+n}.$$

We call the graph of the functions f_n *the inverse branches of the Gauss map* T. From now on, consider the Gauss map and its inverse branches given in Definition 4.1.1 on the set of irrationals $X := [0,1] \setminus \mathbb{Q}$. The Gauss map corresponds to the shift map σ in $\mathbb{N}^{\mathbb{N}}$.

Definition 4.1.2. *We can rewrite the continued fraction for a finite word* $\boldsymbol{a} = (a_1, \dots, a_n)$ *in the following way:*

$$[a_1, \dots, a_n] =: \frac{p_n(\boldsymbol{a})}{q_n(\boldsymbol{a})}$$

where $p_n(\boldsymbol{a}), q_n(\boldsymbol{a}) \in \mathbb{N}$ *are coprime. We call the denominator* $q_n(\boldsymbol{a})$ *the continuant of* $[\boldsymbol{a}]$*. For* $k < n$*, we define* $q_k(\boldsymbol{a}) := q_k(\boldsymbol{a}|_k)$ *where* $\mathbf{a}|_k := (a_1, \dots, a_k)$ *is the word consisting of the first* k *letters of* $\boldsymbol{a}$*.*

Below are some useful relations about continuants. Define the mirror of a word $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{N}^n$ to be $\mathbf{a}^{\leftarrow} = (a_n, \dots, a_1)$.

Proposition 4.1.3. *For a word* $\boldsymbol{a} \in \mathbb{N}^n$ *we have that the following hold:*

(i) $q_n(\boldsymbol{a}) = a_n q_{n-1}(\boldsymbol{a}) + q_{n-2}(\boldsymbol{a})$ *(the recurrence relation for continuants);*

(ii) $p_n(\boldsymbol{a}) = a_n p_{n-1}(\boldsymbol{a}) + p_{n-2}(\boldsymbol{a})$ *(the recurrence relation for numerators);*

(iii) $q_n(\mathbf{a}) = q_n(\mathbf{a}^{\leftarrow})$ *and* $q_{n-1}(\mathbf{a}) = p_n(\mathbf{a}^{\leftarrow})$ *(invariance and recovery under mirroring);*

(iv) $q_n(\boldsymbol{a})p_{n-1}(\boldsymbol{a}) - q_{n-1}(\boldsymbol{a})p_n(\boldsymbol{a}) = (-1)^n$.

We control the derivatives of inverse branches by continuants using the following lemma [26].

Lemma 4.1.4. *For* $\boldsymbol{a} \in \mathbb{N}^n$ *we have that*

$$\frac{1}{4} q_n(\boldsymbol{a})^{-2} \leq |f'_{\boldsymbol{a}}| \leq q_n(\boldsymbol{a})^{-2}.$$

This gives us the same bounds for the length of the intervals $I_{\boldsymbol{a}}$*.*

We can therefore control continuants for regular words in the same way as derivatives using the following [26].

Lemma 4.1.5. *Suppose* $\mathbf{a} \in \mathcal{R}_n$ *and* $j = \varepsilon_r, \dots, n$. *Then*

$$C_{\varepsilon,n}^{-1} e^{\lambda j} \leq q_j(\mathbf{a})^2 \leq 4 C_{\varepsilon,n} e^{\lambda j};$$

We will use the following quasi-independence lemma on continuants to be able to analyse them as in [26] and [50].

Lemma 4.1.6. *For* $\boldsymbol{a} \in \mathbb{N}^n$, *let* $\boldsymbol{b} := (a_1, \dots, a_{n-k})$ *be the first* $n-k$ *digits of* $\boldsymbol{a}$, *and* $\boldsymbol{c} := (a_{n-k+1}, \dots, a_n)$ *be the last* k *digits for any* $1 \leq k < n$. *We have that*

$$\frac{1}{2} \leq \frac{q_n(\boldsymbol{a})}{q_{n-k}(\boldsymbol{b}) q_k(\boldsymbol{c})} \leq 4.$$

A key property of the Gauss map we shall use is a sort of 'invariance under reversing words'. This turns out to be an analogue of Bourgain–Dyatlov considering inverses of words.

Lemma 4.1.7. *For* $\mathbf{a} \in \mathcal{R}_n$ *we have that*

$$\frac{1}{16} C_{\varepsilon,n}^{-1} e^{-\lambda n} \leq |I_{\mathbf{a}^{\leftarrow}}| \leq C_{\varepsilon,n} e^{-\lambda n}.$$

Proof. By invariance of continuants under mirroring we have that

$$\frac{1}{16} C_{\varepsilon,n}^{-1} e^{-\lambda n} \leq \frac{1}{4} q_n(\mathbf{a})^{-2} = \frac{1}{4} q_n(\mathbf{a}^{\leftarrow})^{-2} \leq |I_{\mathbf{a}^{\leftarrow}}| \leq q_n(\mathbf{a}^{\leftarrow})^{-2} = q_n(\mathbf{a})^{-2} \leq C_{\varepsilon,n} e^{-\lambda n}.$$

□

4.2 Distortion control

We begin by establishing that the Gauss map has non-concentrated derivative. In particular, we will be looking at the distribution of distortions of the form

$$\frac{f''_{\mathbf{b}}(x)}{f'_{\mathbf{b}}(x)}$$

so that we can prove the following:

Proposition 4.2.1. *The Gauss map has non-concentrated derivative.*

The main tool that allows us to reduce the proof of Proposition 4.2.1 to properties of continued fractions is that it turns out that the distortion differences between iterated branches represented by words $\mathbf{b}$ and $\mathbf{c}$ can be controlled using the difference between two corresponding geometric points represented by the reverse of these words:

Lemma 4.2.2. *Let* $\mathbf{b}, \mathbf{c} \in \mathbb{N}^n$. *Then we have for all* $x \in [0,1]$ *that*

$$\frac{1}{2}\Big|\frac{p_n(\mathbf{b}^{\leftarrow})}{q_n(\mathbf{b}^{\leftarrow})} - \frac{p_n(\mathbf{c}^{\leftarrow})}{q_n(\mathbf{c}^{\leftarrow})}\Big| \leq \Big|\frac{f''_{\mathbf{b}}(x)}{f'_{\mathbf{b}}(x)} - \frac{f''_{\mathbf{c}}(x)}{f'_{\mathbf{c}}(x)}\Big| \leq 2\Big|\frac{p_n(\mathbf{b}^{\leftarrow})}{q_n(\mathbf{b}^{\leftarrow})} - \frac{p_n(\mathbf{c}^{\leftarrow})}{q_n(\mathbf{c}^{\leftarrow})}\Big|.$$

The above lemma is central to identifying a proof of the non-concentration condition, but the proof is quite straightforward. We will need a simple lemma on the upper bound for distortion:

Lemma 4.2.3. *For* $\mathbf{a} \in \mathbb{N}^n$, *for all* $z \in X$ *we have that*

$$\Big|\frac{f''_{\mathbf{a}}(z)}{f'_{\mathbf{a}}(z)}\Big| \leq 2.$$

Proof. We have that

$$\begin{aligned} |(\log|f'_{\mathbf{a}}(x)|)'| &= \Big|\Big(\log\frac{1}{(q_{n-1}(\mathbf{a})x + q_n(\mathbf{a}))^2}\Big)'\Big| = |(-2\log(q_{n-1}(\mathbf{a})x + q_n(\mathbf{a})))'| \\ &= \Big|\frac{2q_{n-1}(\mathbf{a})}{q_{n-1}(\mathbf{a})x + q_n(\mathbf{a})}\Big| \leq \frac{2q_{n-1}(\mathbf{a})}{q_n(\mathbf{a})} \leq 2. \end{aligned}$$

$\square$

Proof of Lemma 4.2.2. By the formula for $f'_{\mathbf{b}}$ and $f'_{\mathbf{c}}$ in terms of continuants and using the reversal property, we have that

$$\begin{aligned} \frac{1}{2}\Big|\frac{p_n((\mathbf{b})^{\leftarrow})}{q_n((\mathbf{b})^{\leftarrow})} - \frac{p_n((\mathbf{c})^{\leftarrow})}{q_n((\mathbf{c})^{\leftarrow})}\Big| &= \Big|\frac{2q_{n-1}(\mathbf{b})q_n(\mathbf{c}) - 2q_n(\mathbf{b})q_{n-1}(\mathbf{c})}{2q_n(\mathbf{b}) \cdot 2q_n(\mathbf{c})}\Big| \\ &\leq \Big|\frac{2q_{n-1}(\mathbf{b})q_n(\mathbf{c}) - 2q_n(\mathbf{b})q_{n-1}(\mathbf{c})}{(q_{n-1}(\mathbf{b}) + q_n(\mathbf{b})) \cdot (q_{n-1}(\mathbf{c}) + q_n(\mathbf{c}))}\Big| \\ &\leq \Big|\frac{2q_{n-1}(\mathbf{b})q_n(\mathbf{c}) - 2q_n(\mathbf{b})q_{n-1}(\mathbf{c})}{(q_{n-1}(\mathbf{b})x_{\mathbf{d}} + q_n(\mathbf{b})) \cdot (q_{n-1}(\mathbf{c})x_{\mathbf{d}} + q_n(\mathbf{c}))}\Big| \\ &= \Big|\frac{2q_{n-1}(\mathbf{b})}{(q_{n-1}(\mathbf{b})x_{\mathbf{d}} + q_n(\mathbf{b}))} - \frac{2q_{n-1}(\mathbf{c})}{(q_{n-1}(\mathbf{c})x_{\mathbf{d}} + q_n(\mathbf{c}))}\Big| \\ &= \Big|\frac{f''_{\mathbf{b}}(x)}{f'_{\mathbf{b}}(x)} - \frac{f''_{\mathbf{c}}(x)}{f'_{\mathbf{c}}(x)}\Big| \leq \Big|\frac{2q_{n-1}(\mathbf{b})q_n(\mathbf{c}) - 2q_n(\mathbf{b})q_{n-1}(\mathbf{c})}{q_n(\mathbf{b}) \cdot q_n(\mathbf{c})}\Big| \\ &= 2\Big|\frac{q_{n-1}(\mathbf{b})}{q_n(\mathbf{b})} - \frac{q_{n-1}(\mathbf{c})}{q_n(\mathbf{c})}\Big| = 2\Big|\frac{p_n((\mathbf{b})^{\leftarrow})}{q_n((\mathbf{b})^{\leftarrow})} - \frac{p_n((\mathbf{c})^{\leftarrow})}{q_n((\mathbf{c})^{\leftarrow})}\Big| \end{aligned}$$

as required. $\square$

We will often consider concatenated words, but to use a proof of Jordan–Sahlsten, we will need to remove some concatenations.

Lemma 4.2.4 (Distance bounds for concatenating regular words)**.** *Given a word* $\mathbf{a} \in \mathbb{N}^n$ *of length* n*, we have that*

$$|[\mathbf{b}^{\leftarrow}] - [\mathbf{c}^{\leftarrow}]| \leq |[(\mathbf{ab})^{\leftarrow}] - [(\mathbf{ac})^{\leftarrow}]| + 2C_{\varepsilon,n}e^{-\lambda n/2}.$$

Proof.

$$\begin{aligned} |[\mathbf{b}^{\leftarrow}] - [\mathbf{c}^{\leftarrow}]| &= |[\mathbf{b}^{\leftarrow}] - [(\mathbf{ab})^{\leftarrow}] + [(\mathbf{ab})^{\leftarrow}] - [(\mathbf{ac})^{\leftarrow}] + [(\mathbf{ac})^{\leftarrow}] - [\mathbf{c}^{\leftarrow}]| \\ &\leq C_{\varepsilon,n}e^{-\lambda n} + |[(\mathbf{ab})^{\leftarrow}] - [(\mathbf{ac})^{\leftarrow}]| + C_{\varepsilon,n}e^{-\lambda n} \\ &\leq |[(\mathbf{ab})^{\leftarrow}] - [(\mathbf{ac})^{\leftarrow}]| + 2C_{\varepsilon,n}e^{-\lambda n/2}. \end{aligned}$$

□

Proof of Proposition 4.2.1. For $\mathbf{a} \in \mathcal{R}_n$, define

$$D_1(\mathbf{a}) := \left\{(\mathbf{b},\mathbf{c},\mathbf{d}) \in \mathcal{R}_n^3 : \left|\frac{f''_{\mathbf{ab}}(x_{\mathbf{d}})}{f'_{\mathbf{ab}}(x_{\mathbf{d}})} - \frac{f''_{\mathbf{ac}}(x_{\mathbf{d}})}{f'_{\mathbf{ac}}(x_{\mathbf{d}})}\right| \leq \frac{1}{2}\sqrt{\sigma}, |f'_{\mathbf{ab}}(x_{\mathbf{d}}) - f'_{\mathbf{ac}}(x_{\mathbf{d}})| \leq e^{-2\lambda n}\sigma\right\}$$

and recall the definition of $D_2(\mathbf{a})$ in Lemma 3.2.3. To prove non-concentration, it will be sufficient to bound the cardinality of $D_1(\mathbf{a}) \cup D_2(\mathbf{a})$. Define

$$R := \{(\mathbf{a},\mathbf{b},\mathbf{c}) \in \mathcal{R}_n : |[(\mathbf{ab})^{\leftarrow}] - [(\mathbf{ac})^{\leftarrow}]| \leq \sqrt{\sigma}\}.$$

By Lemma 4.2.2, we have that $D_2(\mathbf{a}) \subset R$. Note that

$$R = \mathcal{R}_n^2 \times \{\mathbf{b} \in \mathcal{R}_n : |[(\mathbf{ab})^{\leftarrow}] - [(\mathbf{ac})^{\leftarrow}]| \leq \sqrt{\sigma}\} =: \mathcal{R}_n^2 \times R'$$

So it will be sufficient to get a cardinality bound for R'. By Lemma 4.2.4, and since $e^{-\lambda n/2} \leq \sqrt{\sigma}$ we have that

$$R' \subset \{\mathbf{b} \in \mathcal{R}_n : |[\mathbf{b}^{\leftarrow}] - [\mathbf{c}^{\leftarrow}]| \leq 3C_{\varepsilon,n}\sqrt{\sigma}\} =: R''.$$

By demanding that

$$3\sqrt{\sigma} \leq \frac{3}{96}e^{-\lambda j - \frac{7}{2}\varepsilon n} \leq \frac{3r_j}{q_n(\mathbf{c})}$$

we have that $R'' \subset \mathcal{B}_j$ in the proof of Lemma 6.8 in the paper of Jordan–Sahlsten [26] (see the sentence before their Lemma 2.6 is used in the middle of page 24), and they get that $\#\mathcal{B}_j \le 2C_{\varepsilon,n}^{3\lambda} e^{\lambda sn} e^{-s\lambda j}$ (they do so using their bound for a set $\mathcal{C}_j$). For the given σ, choose j such that

$$\frac{1}{\lambda}\Big(-\lambda - \frac{7}{2}\varepsilon n + \log\Big(\frac{1}{96\sqrt{\sigma}}\Big)\Big) \le j \le \frac{1}{\lambda}\Big(\log\Big(\frac{1}{96\sqrt{\sigma}}\Big) - \frac{7}{2}\varepsilon n\Big).$$

This upper bound for j is equivalent to the demand made to have that $R'' \subset \mathcal{B}_j$. Also, the lower bound is such that j is contained in an interval of length 1, so there is at least one choice for j (if the left hand bound is negative, 0 is a valid choice for j since the upper bound is non-negative). The lower bound for j is equivalent to

$$\frac{1}{96} e^{-\lambda} e^{-\lambda j - \frac{7}{2}\varepsilon n} \le \sqrt{\sigma}.$$

So we have that

$$|\mathcal{B}_j| \le 192 e^{\lambda} C_{\varepsilon,n}^{7\lambda} e^{\lambda sn} \sigma^{s/2}.$$

Following the containment arguments given at the start of this proof, we get that

$$\#D_1(\mathbf{a}) \le \alpha C_{\varepsilon,n}^{6\lambda} e^{3\lambda sn} \sigma^{s/2}$$

where α is some positive constant. Using this fact in conjunction with Lemma 3.2.3 gives a sufficient bound for the cardinality of $D_1(\mathbf{a}) \cup D_2(\mathbf{a})$. □

Remark 4.2.5. *Proving polynomially decaying Fourier transform for measures invariant under a generalised continued fraction maps should indeed be possible. The central reason for this is due to there being a direct analogue of continuants which obey a recurrence relation. This means that we should be able to get a proof using the Gauss map case. The main work involved would be dealing with the fact that most of these maps have both strictly increasing and decreasing branches. I don't think this would be a problem for most of the proof, but the ideas of Queffélec and Ramaré used in Jordan–Sahlsten's paper may be bothersome to formalise.*

4.3 Polynomial Fourier decay

Our aim will be to prove the following improvement of Jordan–Sahlsten [26]:

Theorem 4.3.1. *Let μ be an equilibrium state for the Gauss map with positive Hausdorff dimension satisfying the thin tail condition: there exists $p > 0$ such that*

$$\mu(x \in X : a_n(x) \ge n) = O(n^{-p}), \quad n \to \infty.$$

Then the Fourier transform $\widehat{\mu}(\xi) \to 0$ *as* $|\xi| \to \infty$ *at a polynomial rate.*

To be able to apply relevant large deviation results, we need to make sure that the values of n that we consider are sufficiently large. The conditions that will be laid out now are analogous of those given in Section 5 page 15 of Jordan–Sahlsten [26]. Assuming that $\varepsilon_r n$-level regularity is required, we first assume that $\varepsilon_r = m^{-1} > 0$ for some $m \in \mathbb{N}$. If this is not the case, we can simply make ε_r smaller so that this is so. We begin by choosing n_0 so that $m | n_0$, as well as the following:

1. If n_1 is the generation that arises from the main large deviation theorem, then we require
$$n_0 \varepsilon_r > n_1$$
to ensure we have valid regularity at each scale that we need.

2. If θ is the rate of expansion of $(T^n)'$ with respect to n, and C is the Gibbs constant for μ, we require
$$\frac{\log 4}{\varepsilon_r n_0} < \varepsilon_2, \quad \frac{\log 4C^2}{\log(\theta^{2\varepsilon_r n_0})} < \varepsilon/2 \quad \text{and} \quad \frac{e^{-\delta \varepsilon_r n_0}}{1 - e^{-\delta}} < e^{-\delta \varepsilon_r n_0/2}$$
to ensure that we get decay on multiregular blocks of words.

3. Finally we require that
$$\frac{1}{192} e^{(\lambda/2 - 2\varepsilon) n_0} \geq 1$$
to use part of the proof of Lemma 6.8 in the paper of Jordan–Sahlsten [26] as used in the proof of Proposition 4.2.1.

Let us now begin the proof of Theorem 4.3.1. Let $s = \dim_{\mathrm{H}} \mu$ and λ the Lyapunov exponent of μ. Write $s_0 = \kappa/2$ from the nonlinearity assumption for T and μ. Let $k \in \mathbb{N}$ and $\varepsilon_2 > 0$ from Lemma 3.1.1.

Fix a frequency $\xi \in \mathbb{R}$. Let $n \in \mathbb{N}$ be the number such that $\xi = \operatorname{sgn} \xi \cdot \varrho e^{(2k+3/2)n}$ where $\varrho \in [1, e^{2k+3/2}]$. Recall that
$$\# \mathcal{R}_n \leq C C_{\varepsilon,n}^{3\lambda} e^{\lambda s n}$$
and if $\mathbf{a} \in \mathcal{R}_n$, we have for all $x \in X$,
$$w_{\mathbf{a}}(x) \leq C_{\varepsilon,n}^{3\lambda} e^{-\lambda s n}.$$
Write
$$h(x) := e^{-2\pi i \xi \Phi(x)} g(x).$$

4.3.1 Applying the decay of exponential sums

It is important to recall what was mentioned in Remark 3.3.2 here, because we actually use a different proof of Lemma 3.3.1. The complete proof is given in [53], and is almost identical, but works for the infinitely branched Gauss map. We begin by proving the assumption required to use Lemma 3.3.1.

Proposition 4.3.2. *The main assumption of Lemma 3.3.1 about $\mu \times \mu$ holds for the Gibbs measures for the Gauss map with $\tau = e^{-\lambda n}$.*

Proof. By covering the n-regular part of the following set with $\lfloor n/4 \rfloor$-generation parent intervals, for fixed $y \in X$ we have that

$$\begin{aligned}
&\mu(\{x \in X : |x-y| \leq C_0 e^{-\lambda n/4}\}) \\
&\leq \mu(X \setminus R_n) + \mu(\{x \in R_n : |x-y| \leq C_0 e^{-\lambda n/4}\}) \\
&\leq \mu(X \setminus R_n) + \left\lceil \frac{2C_0 e^{-\lambda n/4}}{C_{\varepsilon,n}^{-1} e^{-\lambda \lfloor n/4 \rfloor}/16} \right\rceil e^{-\lambda s \lfloor n/4 \rfloor} \\
&\leq \mu(X \setminus R_n) + 64 C_0 C_{\varepsilon,n} e^{-\lambda s(n/4-1)} \\
&\leq \mu(X \setminus R_n) + 64 e^{\lambda} C_0 C_{\varepsilon,n} e^{-\lambda s n/4}.
\end{aligned}$$

Hence we have that

$$\mu \times \mu(\{(x,y) \in X^2 : |x-y| \leq C_0 e^{-\lambda n/4}\}) \leq \mu(X \setminus R_n) + 64 e^{\lambda} C_0 C_{\varepsilon,n} e^{-\lambda s n/4}$$

as required. □

We move on to remove 'concentrated' blocks from consideration in Lemma 3.3.1. We see that for some $a' > 0$,

$$\begin{aligned}
&C_{\varepsilon,n}^{9\lambda(2k+2)} e^{-(2k+1)\lambda sn} \sum_{\mathbf{a} \in \mathcal{R}_n^{k+1} \setminus \mathcal{W}} \sup_{\eta \in J_n} \left| \sum_{\mathbf{B}:\mathbf{A} \leftrightarrow \mathbf{B}} e^{2\pi i \eta \zeta_{1,\mathbf{A}}(\mathbf{b_1}) \ldots \zeta_{k,\mathbf{A}}(\mathbf{b}_k)} \right| \\
&\leq C_{\varepsilon,n}^{9\lambda(2k+2)} e^{-(2k+1)\lambda sn} \sum_{\mathbf{A} \in \mathcal{R}_n^{k+1} \setminus \mathcal{W}} \sup_{\eta \in J_n} \sum_{\mathbf{B}:\mathbf{A} \leftrightarrow \mathbf{B}} 1 \\
&\leq C_{\varepsilon,n}^{9\lambda(2k+2)} e^{-(2k+1)\lambda sn} \sum_{\mathbf{A} \in \mathcal{R}_n^{k+1} \setminus \mathcal{W}} C^k C_{\varepsilon,n}^{3\lambda k} e^{k\lambda sn} \\
&\leq a C^k C_{\varepsilon,n}^{11(2k+2)\lambda} e^{-(k+1)\lambda sn} e^{(k+1)\lambda sn} e^{\varepsilon_2 \lambda sn/20} \\
&= a' C_{\varepsilon,n}^{33\lambda k} e^{-\varepsilon_2 \lambda sn/20}.
\end{aligned}$$

Hence we have that

$$\Big|\int h(x)\,d\mu(x)\Big|^2 \le C_{\varepsilon,n}^{9\lambda(2k+2)}e^{-k\lambda sn}\max_{\mathbf{A}\in\mathcal{W}}\sup_{\eta\in J_n}\Big|\sum_{\mathbf{B}:\mathbf{A}\leftrightarrow\mathbf{B}} e^{2\pi i\eta\zeta_{1,\mathbf{A}}(\mathbf{b_1})\dots\zeta_{k,\mathbf{A}}(\mathbf{b}_k)}\Big| + a'C_{\varepsilon,n}^{33\lambda k}e^{-\varepsilon_2\lambda sn/20} + 2^7C^2C_{\varepsilon,n}^2e^{\lambda}(k+1)^2e^{-\delta n/8} + C_1\delta^{2n}.$$

Note that by the regularity bounds for the measure of construction intervals, since μ is a probability measure, we have that $|\mathcal{R}_n| \le CC_{\varepsilon,n}^{3\lambda}e^{s\lambda n}$. Let $\eta\in J_n$. Recall that

$$s_0 = \min\{\kappa, s\}/4.$$

By the definition of J_n and the definition of $\mathcal{W}$, we have

$$\sharp\{(\mathbf{b},\mathbf{c})\in\mathcal{R}_n^2 : |\zeta_{j,\mathbf{A}}(\mathbf{b}) - \zeta_{j,\mathbf{A}}(\mathbf{c})| \le \sigma\} \le e^{2\lambda sn}\sigma^{s_0}$$

Note that also $\zeta_{j,\mathbf{A}}(\mathbf{b}) \in [16^{-2}C^{-1}C_{\varepsilon,n}^{-3\lambda}, 16^2CC_{\varepsilon,n}^{3\lambda}]$. Thus we may apply Lemma 3.1.2 to the maps $\zeta_{j,\mathbf{A}}$. It implies that there exists some $d>0$ depending only on s and k such that for all $\mathbf{A}\in\mathcal{W}$ and $\eta\in J_n$,

$$C_{\varepsilon,n}^{9\lambda(2k+2)}e^{-k\lambda sn}\Big|\sum_{\mathbf{B}:\mathbf{A}\leftrightarrow\mathbf{B}} e^{2\pi i\eta\zeta_{1,\mathbf{a}}(\mathbf{b_1})\dots\zeta_{k,\mathbf{a}}(\mathbf{b}_k)}\Big| \le dC_{\varepsilon,n}^{3k\lambda}|\eta|^{-\varepsilon_2} \le dC_{\varepsilon,n}^{3k\lambda}e^{-\varepsilon_2\lambda n/4}.$$

Thus we have proved

$$\Big|\int h(x)\,d\mu(x)\Big|^2 \le dC_{\varepsilon,n}^{3k\lambda}e^{-\varepsilon_2\lambda n/4} + a'C_{\varepsilon,n}^{33k\lambda}e^{-\varepsilon_3\lambda sn/20} + DC_{\varepsilon,n}^2e^{-\delta n/8} + C_1\delta^{2n}.$$

By making sure that ε is chosen such that $33\lambda k\varepsilon \le \varepsilon_2 s/20$, the proof of Theorem 4.3.1 is complete.

Chapter 5

General Nonlinear Maps

5.1 Naud's Theory for Cantor Sets

5.1.1 Non-integrability condition

Let $\mathcal{A} \subset \mathbb{N}$ be a finite alphabet for the Markov map $T \in C^2$. Assume that the sets I_a for $a \in \mathcal{A}$ are pairwise disjoint. Assume that the map is piecewise increasing for simplicity. Our potential of interest is $-\psi = \log |T'|$ which is differentiable if we make the aforementioned assumption. For $\xi, \eta \in \mathcal{A}^{\mathbb{N}}$, define the temporal distance function to be

$$\varphi_{\xi,\eta}(u, v) := \Delta_\xi(u, v) - \Delta_\eta(u, v)$$

where

$$\Delta_\xi(u, v) := \lim_{n\to\infty} \sum_{j=0}^{n-1} \log |T'(T^j f_{\xi|_n} u)|.$$

To simplify $\Delta_\xi(u, v)$, we use the fact that

$$\sum_{j=0}^{n-1} \log |T'(T^j f_{\xi|_n} u)| = -\log |f'_{\xi|_n} u|$$

by the inverse and chain rule for differentiable functions. Hence we have that

$$\Delta_\xi(u, v) = \lim_{n\to\infty} \Big(-\log f'_{\xi|_n} u + \log f'_{\xi|_n} v \Big) =: -\log f'_\xi u + \log f'_\xi v$$

where we assume that the map is piecewise increasing. The main assumption of Naud's paper [44] is the following.

Definition 5.1.1. *Let K be the repeller of T [51]. We say that $-\psi$ is not locally integrable NLI if there exists $j \in \{1, \ldots, k\}$, $\xi, \eta \in \mathcal{A}^{\mathbb{N}}$ such that for some $u_0, v_0 \in K \cap I_j$,*

$$\frac{\partial \varphi_{\xi,\eta}}{\partial u}(u_0, v_0) \neq 0.$$

There is a simpler definition if T and $-\psi$ are real analytic which Dolgopyat uses in a symbolic setting where they only need check if $\varphi_{\xi,\eta}$ is identically zero. Naud also mentions that NLI is related to the following.

Definition 5.1.2. *The potential $-\psi$ is non-lattice if there is no function $L : K \to m\mathbb{Z}$ for some $m > 0$ and $f : K \to \mathbb{R}$ where f is Lipschitz on K such that for all $x \in K$*

$$\log |T'(x)| = f \circ T(x) - f(x) + L(x).$$

Naud notes that NLI implies non-lattice, but not necessarily the other way around. The non-lattice assumption is a weaker assumption than non-conjugacy to a linear map because of the condition that L only not map to $m\mathbb{Z}$ rather than any set of positive numbers (e.g. $\{\pi, 3\}$ which isn't a subset of any $m\mathbb{Z}$).

Definition 5.1.1 can be simplified in the case we are interested in by noting that

$$\frac{\partial \varphi_{\xi,\eta}}{\partial u}(u_0, v_0) = -\frac{f''_\xi u_0}{f'_\xi u_0} + \frac{f''_\eta u_0}{f'_\eta u_0}$$

which is the same as the distortion differences studied in [53], but here η, ξ are infinite words. It should be noted that Proposition 5.5 for Naud [44] looks friendlier in our setting due to dealing with finite words. Using Naud's proof, we can prove an if and only if version of [44, Proposition 5.5].

Proposition 5.1.3. *The NLI holds for $-\psi$ if and only if there exist $M, m > 0$ and N_0 such that for all $N > N_0$ there exists words $\mathbf{a}, \mathbf{b} \in \mathcal{A}^N$ such that for all $u \in X$,*

$$m \leq \left| \frac{d}{du}(-S_N\psi \circ f_{\mathbf{a}} + S_N\psi \circ f_{\mathbf{b}})(u) \right| \leq M.$$

Proof. For the direction not covered by Naud, assume that NLI does not hold to prove a contrapositive statement. Then the temporal distance function is identically vanishing, that is, $\varphi_{\mathbf{w},\mathbf{v}}(x, y) \equiv 0$ for all $\mathbf{w}, \mathbf{v} \in \mathcal{A}^\infty$, $x, y \in X$, since T and $-\psi$

are real analytic. Define $\tau := -\psi$. We have for all $a \in \mathcal{A}$ and $\mathbf{w}, \mathbf{v} \in \mathcal{A}^\infty$ with $T(I_{\mathbf{v}_1}) \cap T(I_{\mathbf{w}_1}) \supset I_a$ and $x, y \in I_a$ that the derivative

$$\begin{aligned}\frac{\partial \varphi_{\mathbf{v},\mathbf{w}}}{\partial x}(x,y) =& \frac{d}{dx}(S_n\tau(f_{\mathbf{w}|n}(x)) - S_n\tau(f_{\mathbf{v}|n}(x))) \\ &+ \sum_{k\geq n} \frac{\tau'(f_{\mathbf{w}|n}(x))}{T'(f_{\mathbf{w}|1}(x))\dots T'(f_{\mathbf{w}|(p-1)}(x))} \\ &- \sum_{k\geq n} \frac{\tau'(f_{\mathbf{v}|n}(x))}{T'(f_{\mathbf{v}|1}(x))\dots T'(f_{\mathbf{v}|(p-1)}(x))}\end{aligned}$$

so we see that the vanishing temporal distance function $\varphi_{\mathbf{w},\mathbf{v}}(x,y) \equiv 0$ implies that for all $n \in \mathbb{N}$, $\mathbf{a}, \mathbf{b} \in \mathcal{A}^n$ and $x \in X$:

$$\Big|\frac{d}{dx}(S_n\tau(f_{\mathbf{a}}(x)) - S_n\tau(f_{\mathbf{b}}(x)))\Big| \leq \frac{D}{\gamma^n}\frac{2\|\tau'\|_\infty}{\gamma - 1},$$

where $\gamma > 1$ and $D > 0$ are the constants from the expansion assumption of T and $\|\tau'\|_\infty \leq B < \infty$ by the bounded distortions assumption of T. Then by letting $n \to \infty$ we realise the required result. □

Observe the main inequalities on inverse branches given in Proposition 5.1.3. They do not involve a limit, and so is precisely that same as

$$m \leq \Big| -\frac{f''_{\mathbf{a}}u}{f'_{\mathbf{a}}u} + \frac{f''_{\mathbf{b}}u}{f'_{\mathbf{b}}u}\Big| \leq M$$

which is our normal definition of distortion differences. This difference is what we refer to as the distortion difference for words **a** and **b**. In Bourgain–Dyatlov, these can be analysed nicely using distortion factors of Möbius transformations. In [53], we use the tools from diophantine approximation presented in Jordan–Sahlsten using the ideas of Queffélec and Ramaré. This reinforces the fact that complex transfer operators do indeed seem to be a very reasonable tool for tackling this problem.

5.2 Totally Nonlinear Maps

It should also be noted that Araújo–Melbourne [2] uses a very similar statement of the above proposition (for suspension flows) to define their uniform non-integrability condition rather than Naud's NLI. This is also very close to Proposition 7.5 of Avila–Gouëzel–Yoccoz [3] except there are some direction fields involved. However, this proposition is for Markov maps and does prove that non-conjugacy to a linear map implies the conclusion of Proposition 5.1.3. The following is part of Proposition 7.5 of Avila–Gouëzel–Yoccoz.

Proposition 5.2.1. *Let Δ be a Riemannian manifold considered with Lebesgue measure. Let T be a uniformly expanding Markov map on an open partition $\{\Delta^{(l)}\}$ whose union is a full Lebesgue measure subset of Δ. Let $\mathcal{H}^n$ be the set of inverse branches of T^n, and $\mathcal{H} := \mathcal{H}^1$. Let $r : \Delta \to \mathbb{R}$ be a C^1 function on each set $\Delta^{(l)}$ with $\sup_{h\in\mathcal{H}} ||D(r \circ h)||_{C^0} < \infty$. The following are equivalent:*

(1) T is totally nonlinear. This means that is is not possible to write

$$r = \psi + \varphi \circ T - \varphi$$

on a set X of full measure (not necessarily the entirety of Δ) where $\psi : X \to \mathbb{R}$ is constant on each set $\Delta^{(l)}$ and $\varphi : \Delta \to \mathbb{R}$ is measurable.

(2) There exists $C > 0$ such that there exists $n \in \mathbb{N}$, two inverse branches $h, k \in \mathcal{H}^n$, and a continuous unitary vector field $x \mapsto y(x)$ such that for all $x \in \Delta$,

$$|D(r^{(n)} \circ h)(x) \cdot y(x) - D(r^{(n)} \circ k)(x) \cdot y(x)| > C$$

where $r^{(n)}$ is the nth Birkhoff sum for r.

Remark 5.2.2. *Note that in the setting of Naud to be considered, $\{\Delta^{(l)}\}$ will be the domain of the dynamics which will not be a full Lebesgue measure subset of Δ. However, it does have full measure when considering our measure supported on the limit set. The proof of Proposition 5.2.1 can be modified to instead use this measure.*

For our application, we will consider Δ to be a large interval containing the domain and range of T. We note that the only continuous unitary vector fields on Δ are constant ± 1, hence the second condition in the proposition will be about our desired distortion difference for $r = \log |T'|$.

5.2.1 Fourier Decay in the case of Totally-Nonlinear Dynamics

We follow the setting and notations of Naud [44] and use the same notation. Let $I_1, \ldots, I_N$, $N \geq 2$, be closed, disjoint and bounded intervals in X, and write $X =$

$\bigcup_{a=1}^{N} I_a$. Let $T : X \to \mathbb{R}$ be a mapping such that each restriction $T_a := T|_{I_a}$ is a real analytic map and assume T is conjugated to the full shift on $\mathcal{A}^{\mathbb{N}}$, where $\mathcal{A} = \{1, \dots, N\}$. Moreover, we need that T also satisfies

(1) *Uniform expansion*: There exists $\gamma > 1$ and $D > 0$ such that for all $n \in \mathbb{N}$ and all $x \in X$ we have

$$|(T^n)'(x)| \geq D^{-1}\gamma^n.$$

(2) *Markov property*: For all $a, b = 1, \dots, N$, if $T(I_b) \cap \mathrm{Int}(I_a) = \varnothing$, then $T(I_b) \supset I_a$.

(3) *Bounded distortions*: Define the *distortion function* as

$$\tau := \log |T'|.$$

Then there exists $B < \infty$ such that

$$\|\tau'\|_\infty = \|T''/T'\|_\infty \leq B.$$

(4) *Total non-linearity*: $\tau : X \to \mathbb{R}$ is *not C^1 cohomologous to a locally constant function*, that is it is not possible to write

$$\tau = \psi_0 + g \circ T - g$$

on the set X where $\psi_0 : X \to \mathbb{R}$ is constant on every $I_a \subset X$, $a \in \mathcal{A}$, and $g \in C^1(\Delta)$.

Theorem 5.2.3 (Sahlsten–S, 2020 [54])**.** *Suppose K satisfies conditions* (1)*,* (2)*,* (3) *and* (4) *and let μ be an equilibrium state on K associated to a potential φ with exponentially vanishing variations. If the Hausdorff dimension* $\dim_{\mathrm{H}} \mu$ *is close enough to* $\dim_{\mathrm{H}} K$*, then the Fourier coefficients of μ tend to zero with a polynomial rate.*

To prove this theorem, we require a slightly different version of Lemma 3.3.1. This will reflect the fact that in Bourgain–Dyatlov, they achieve non-concentrated derivatives of Fuchsian groups with $c_0 = \delta/2$ (see definition 1.3.2) where $\delta > 0$ is the Hausdorff dimension of the limit set of the Fuchsian group. This is a strong condition and reflects the fact that they can prove the non-concentrated derivative assumption by using distortion factor analysis. The same goes for the Gauss map case, where we can get $\kappa = \dim \mu/2$ by using the continuant analysis presented in Jordan–Sahlsten

[26]. In the general nonlinear setting, we do not have such tools to give these bounds. The complex transfer operator theory will present a κ which could be arbitrarily small, and depends only on the map T, not the dimension of the measure. This results in us having to change the frequencies ξ that we consider, and so we use Lemma 3.3.1 with $\tau = e^{-2\varepsilon_0 n}$.

5.3 Total non-linearity and non-concentration

In order to apply Lemma 3.1.2 in our setting, we will need to verify the non-concentration assumption for the maps $\zeta_j = \zeta_{j,\mathbf{A}}$ which are defined by

$$\zeta_{j,\mathbf{A}}(\mathbf{b}) = e^{2\lambda n} f'_{\mathbf{a}_{j-1}\mathbf{b}}(x_{\mathbf{a}_j})$$

for $\mathbf{b} \in \mathcal{R}_n(\varepsilon)$ where $\mathbf{A} = \mathbf{a}_1 \dots \mathbf{a}_k \in \mathcal{R}_n^{k+1}$ and $x_{\mathbf{a}_j}$ is the centre point of the interval $I_{\mathbf{a}_j}$. Before we do this, we need to fix the parameters R and the range of σ we consider needed for Lemma 3.1.2:

Remark 5.3.1. *(1) For $n \in \mathbb{N}$ and $\varepsilon > 0$ the number*

$$R = R(n, \varepsilon) := 16^2 C C_{\varepsilon,n}^{3\lambda},$$

where $C > 0$ is the constant satisfying the Gibbs condition of μ, recall $\lambda = \int \tau \, d\mu$ is thee Lyapunov exponent of μ. Then for the map

$$\zeta_{j,\mathbf{A}}(\mathbf{b}) = e^{2\lambda n} f'_{\mathbf{a_{j-1}b}}(x_{\mathbf{a_j}}).$$

we see that

$$\zeta_{j,\mathbf{A}}(\mathbf{b}) \in [R^{-1}, R].$$

Indeed, the chain rule gives

$$\zeta_{j,\mathbf{A}}(\mathbf{b}) = e^{2\lambda n} f'_{\mathbf{a_{j-1}}}(f_{\mathbf{b}} x_{\mathbf{a_j}}) f'_{\mathbf{b}}(x_{\mathbf{a}_j})$$

so we can apply Lemma 2.2.3 and the fact that $f'_{\mathbf{a}_{j-1}}$ and $f'_{\mathbf{b}}$ must both be either positive or negative because they are defined by words of the same length.

(2) Let $s_0 > 0$ be the unique solution to $P(-s_0\tau) = 0$ for the distortion function $\tau(x) = \log |T'(x)|$. Suppose $0 < \delta = \dim_{\mathrm{H}} \mu < s_0 = \dim_{\mathrm{H}} K$. Then choose $\Xi \in$

$(0,1)$ *such that* $\delta > s_0 - \delta_0(\Xi)$ *and that* $\delta_1(\Xi)/4 < \lambda/2$, *the Lyapunov exponent of* μ *and* $\delta_0(\Xi)$ *is from Theorem 5.3.2. Such* Ξ *exists as* $0 < \delta = \dim_{\mathrm{H}} \mu < s_0$ *is close enough to* s_0. *Define*

$$\varepsilon_0 := \delta_1(\Xi)/4 > 0$$

Then also $\varepsilon_0 < \lambda$ *and now fixes our* $\mathcal{R}_n(\varepsilon)$, $\mathcal{R}_n^k(\varepsilon)$ *and*

$$J_n(\varepsilon) = \{\eta \in \mathbb{R} : e^{\varepsilon_0 n/2} \leq |\eta| \leq C_{\varepsilon,n} e^{\varepsilon_0 n}\}$$

which all implicitly depend on $\varepsilon_0 > 0$.

Let $C^1(I)$ be the set of all complex valued C^1 functions g on X with the norm

$$\|g\|_{C^1} := \|g\|_\infty + \|g'\|_\infty.$$

Given $\psi \in C^1(I)$, define the transfer operator $\mathcal{L}_\psi$ on the Banach space $C^1(I)$ by

$$\mathcal{L}_\psi g(x) := \sum_{y : T(y) = x} e^{\psi(y)} g(y)$$

If $s \in \mathbb{C}$, then in the case of the potential $\psi = -s\tau$ for the distortion function $\tau = \log |T'|$, the following was proved by Naud in [44, Theorem 2.3]:

Theorem 5.3.2 (C^1-contraction of transfer operators)**.** *Under the assumptions of Theorem 5.2.3, the following holds. Let* $\Xi > 0$. *Then there exists* $C_\Xi > 0$, $\delta_0(\Xi) > 0$, $\delta_1(\Xi) > 0$, $t_0(\Xi) > 0$ *such that for all* $\operatorname{Re} s \in (s_0 - \delta_0(\Xi), s_0]$ *and* $|\operatorname{Im} s| \geq t_0(\Xi)$ *we have for all* $f \in C^1(X)$ *and* $m \in \mathbb{N}$ *that*

$$\|\mathcal{L}_{-s\tau}^m f\|_{C^1} \leq C_\Xi |\operatorname{Im} s|^{1+\Xi} e^{-\delta_1(\Xi) m} \|f\|_{C^1},$$

where $s_0 > 0$ *is the unique real number satisfying* $P(-s_0\tau) = 0$ *and* P *is the topological pressure on* K.

We will now give the key non-concentration estimate for distortions as a consequence of Theorem 5.3.2.

Lemma 5.3.3 (Non-concentration)**.** *Let* $s_0 > 0$ *be the unique solution to* $P(-s_0\tau) = 0$ *for the distortion function* $\tau(x) = \log |T'(x)|$. *Suppose* $0 < \delta = \dim_{\mathrm{H}} \mu < s_0$. *Then there exists* $c_0 > 0$ *and* $\kappa_0 > 0$ *such that for all* $\varepsilon > 0$, $n \in \mathbb{N}$, $\eta \in J_n(\varepsilon)$, $\sigma \in [R(n,\varepsilon)^{-2}|\eta|^{-1}, |\eta|^{-\varepsilon_3}]$, $x \in X$ *we have*

$$\sharp\{(\mathbf{a},\mathbf{b},\mathbf{c}) \in \mathcal{R}_n(\varepsilon)^3 : |e^{2\lambda n} f_{\mathbf{ab}}'(x) - e^{2\lambda n} f_{\mathbf{ac}}'(x)| \leq \sigma\} \lesssim C_{\varepsilon,n}^{\kappa_0} \sigma^{c_0} \sharp \mathcal{R}_n(\varepsilon)^3,$$

where $R(n, \varepsilon)$ and ε_0 be the parameters fixed in Remark 5.3.1 and $\varepsilon_3 > 0$ is from Lemma 3.1.2.

Remark 5.3.4 (Idea of the proof)**.** *The following proof is very technical, but follows a basic idea. In this remark we will explain the idea with limited technical details. Most of the technicalities here will be in the set-up, which is necessary to be able to present a believable proof idea.*

We will define m so that $\sigma \approx e^{-\varepsilon_0 m}$. We can then define $\mathbf{d}$ to be the last m entries of $\mathbf{b}$, i.e. $\mathbf{d} := \sigma^{n-m}\mathbf{b}$ noting an abuse of notation with σ being a number as well as the shift mapping on words. We will say that $\mathbf{d} \in \widetilde{\mathcal{R}_m}$, which you can just think of as being regular words of length m. In reality, they are words of length m which have the same regularity bounds as a word in $\mathcal{R}_m$. We cannot say precisely that $\mathbf{d} \in \mathcal{R}_m$, but we do know that $\mathbf{b} \in \mathcal{R}_n$ and $\mathbf{b}|_{n-m}$ has $(n-m)$-regular bounds. We can then use the quasi-Bernoulli property (Lemma 2.0.6) and the length concatenation property for construction intervals (Lemma 2.0.5) to be able to say that $\mathbf{d}$ also has some m-regular bounds, hence we say $\mathbf{d} \in \widetilde{\mathcal{R}_m}$.

For the proof, it is sufficient to bound

$$\#\{\mathbf{d} \in \widetilde{\mathcal{R}_m} : e^{2\lambda n} f'_{\mathbf{ed}}(x) \in B(y, \sigma)\}$$

where $\mathbf{e} := \mathbf{ab}|_{n-m}$, by choosing $y = e^{2\lambda n} f'_{\mathbf{ac}}(x)$. We can then just bound

$$\{\mathbf{d} \in \widetilde{\mathcal{R}_m} : -\log |f'_{\mathbf{ed}}(x)| \in J\}$$

where J is some interval whose length can be approximated using the mean value theorem. The length of J turns out to be approximately σ. We create a smooth 'indicator function' h on J, where we make it smooth to be able to use Fourier analysis on it.

So what we will want to bound is

$$\sum_{\mathbf{d} \in \widetilde{\mathcal{R}_m}} h(-\log |f'_{\mathbf{ed}}(x)|)^{1/2}$$

where the power of a half is only included for a nice use of the Cauchy–Schwarz inequality. We use Cauchy–Schwarz here to add weights to the sum, so we study

$$\sum_{\mathbf{d} \in \widetilde{\mathcal{R}_m}} |f'_{\mathbf{ed}}(x)|^{\delta} h(-\log |f'_{\mathbf{ed}}(x)|)^{1/2}$$

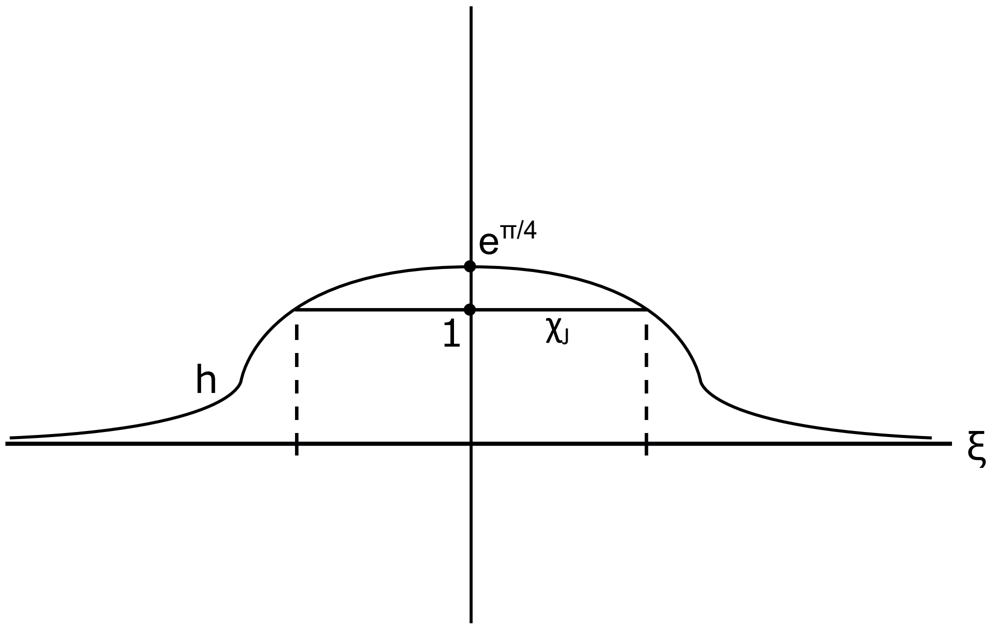

Figure 5.1: The function h, notably smooth and bigger than the indicator for J.

which will allow us to compare this sum to a complex transfer operator. It's worth noting that these weights will just contribute to the $\#\mathcal{R}_n^3$ *bound in the end. Now we apply Fourier inversion and recovery to* h *to get that our sum is*

$$\int_{\mathbb{R}} \widehat{h}(\xi) \sum_{\mathbf{d}\in\widetilde{\mathcal{R}_m}} |f'_{\mathbf{ed}}(x)|^{\delta} e^{2\pi i\xi \log|f'_{\mathbf{ed}}(x)|}\, d\xi$$

which is valid by smoothness properties of h*. We now split this integral into two regions of frequencies.*

The first region, where $|\xi| < t_0/2\pi$ *is where we cannot use the main complex transfer operator theorem by definition of some* $t_0 > 0$*. This region is small however, so we can just use basic bounds, along with* $|\widehat{h}(\xi)| \leq ||h||_{L^1} \leq 3|J|$*, which arises from definable properties of* h*. This bound will be sufficient to get the desired non-concentration bound.*

For the second region, $|\xi| > t_0/2\pi$*, we can use our transfer operator bound. We compare the sum in our integral to*

$$\mathcal{L}^n_{-s\log|T'|} g_{\mathbf{e}}(x) = \sum_{\mathbf{d}\in\mathcal{A}_m} |f'_{\mathbf{ed}}(x)|^{\delta} e^{-2\pi i\xi \log|f'_{\mathbf{ed}}(x)|}$$

where $g_{\mathbf{e}}(z) := |f'_{\mathbf{e}}(z)|^{\delta}$ *for* $z \in X$*. We can then use the bound in Theorem 5.3.2 to conclude.*

Remark 5.3.5. *Theorem 5.3.2 is the key result that will give us the main theorem for totally nonlinear maps. As such, the* $\mathrm{Re}\, s \in (s_0 - \delta_0(\Xi), s_0]$ *condition carries through, and eventually leaves us with requiring us to assume that* $\dim_H \mu =: \mathrm{Re}\, s \in (s_0 - \delta_0(\Xi), s_0]$*. We will say that this means that the dimension of the measure must be 'large enough'. We believe that this assumption can be fully reduced to requiring that* $\dim_H \mu > 0$*. A proof has been presented in the latest version of [54]. There we use a different contraction theorem used by Stoyanov [63] which gives more flexibility in the real part of* s*. In particular, it can be defined using a potential* φ*, so we can consider that to be the potential of the measure. The idea of the proof does not change so much. Essentially the Cauchy–Schwarz inequality has to be applied slightly differently to include the weights* $w_{\mathbf{d}}(x)$ *arising from the use of the potential of the measure. The norm for Stoyanov's contraction theorem is also a Lipschitz norm, not the* C^1 *norm. The Lipschitz norm requires using a slightly stronger application of integration by parts when bounding the Fourier transform of the smooth indicator function. We* *do not consider this more general theorem here, because this addition was mainly done by Tuomas Sahlsten (coauthor of [54]), and the use of Stoyanov's contraction theorem was pointed out by Jialun Li. It is also much easier to see how Naud's contraction theorem can be applied to our setting.*

Proof. Recall the definition of $A_n(\varepsilon)$ in Definition 2.2.2. Choose $m \in \mathbb{N}$ such that $e^{-\varepsilon_0(m-1)} \leq \sigma \leq e^{-\varepsilon_0 m}$. We will first prove that for all $y \in \mathbb{R}$ with $y \pm e^{-\varepsilon_0 m} \in [R^{-1}, R]$, $\mathbf{e} \in \tilde{\mathcal{R}}_{2n-m}(\varepsilon)$ and $x \in I$ we have

$$\sharp\{\mathbf{d} \in \tilde{\mathcal{R}}_m(\varepsilon) : e^{2\lambda n} f'_{\mathbf{ed}}(x) \in B(y, e^{-\varepsilon_0 m})\} \lesssim C^{\kappa_0}_{\varepsilon,m} e^{-c_0 m} \sharp\tilde{\mathcal{R}}_m(\varepsilon), \tag{5.1}$$

where

$$\tilde{\mathcal{R}}_m(\varepsilon) := \Big\{\mathbf{d} \in \mathcal{A}^m : I_{\mathbf{d}} \subset A_m(2\varepsilon)\Big\}$$

and

$$\tilde{\mathcal{R}}_{2n-m}(\varepsilon) := \Big\{\mathbf{e} \in \mathcal{A}^{2n-m} : I_{\mathbf{e}} \subset A_{2n-m}(2\varepsilon)\Big\}.$$

Then up to $C_{2\varepsilon,n}$ multiplicative error $\sharp\tilde{\mathcal{R}}_m(\varepsilon) \sim \sharp\mathcal{R}_m(\varepsilon)$ and $\sharp\tilde{\mathcal{R}}_{2n-m}(\varepsilon) \sim \sharp\mathcal{R}_{2n-m}(\varepsilon)$ by Lemma 2.2.3 for the properties of regular words, where by $a \sim b$ we mean $b/c \leq a \leq cb$ with multiplicative error $c > 0$. Indeed, Lemma 5.3.3 follows now from (5.1) by first setting $\mathcal{P}$ to be the set of pairs $(\mathbf{e}, y)$ such that $y \pm e^{-\varepsilon_0 m} \in [R^{-1}, R]$ and $\mathbf{e} \in \tilde{\mathcal{R}}_{2n-m}(\varepsilon)$, and bounding

$$\begin{aligned}
&\sharp\{(\mathbf{a}, \mathbf{b}, \mathbf{c}) \in \mathcal{R}_n(\varepsilon)^3 : |e^{2\lambda n} f'_{\mathbf{ab}}(x) - e^{2\lambda n} f'_{\mathbf{ac}}(x)| \leq \sigma\} \\
&\leq \sharp\mathcal{R}_n(\varepsilon) \sup_{\mathbf{c} \in \mathcal{R}_n(\varepsilon)} \sharp\{(\mathbf{a}, \mathbf{b}) \in \mathcal{R}_n(\varepsilon)^2 : |e^{2\lambda n} f'_{\mathbf{ab}}(x) - e^{2\lambda n} f'_{\mathbf{ac}}(x)| \leq \sigma\} \\
&\leq \sharp\mathcal{R}_n(\varepsilon) \sharp\tilde{\mathcal{R}}_{2n-m}(\varepsilon) \sup_{(\mathbf{e}, y) \in \mathcal{P}} \sharp\{\mathbf{d} \in \tilde{\mathcal{R}}_m(\varepsilon) : e^{2\lambda n} f'_{\mathbf{ed}}(x) \in B(y, \sigma)\}
\end{aligned}$$

since every $\mathbf{ab}$, for $\mathbf{a}, \mathbf{b} \in \mathcal{R}_n(\varepsilon)$ splits into a word $\mathbf{ab} = \mathbf{ed}$ with $\mathbf{e} := \mathbf{ab}|_{2n-m} \in \tilde{\mathcal{R}}_{2n-m}(\varepsilon)$ and $\mathbf{d} := \sigma^{2n-m}(\mathbf{ab}) \in \tilde{\mathcal{R}}_m(\varepsilon)$ using the quasi-Bernoulli property of the Gibbs measure μ: since $\mathbf{ab} = \mathbf{ed}$, we have

$$\mu(I_{\mathbf{e}})\mu(I_{\mathbf{d}}) \lesssim \mu(I_{\mathbf{ab}}) \lesssim \mu(I_{\mathbf{e}})\mu(I_{\mathbf{d}})$$

and that the lengths

$$|I_{\mathbf{e}}||I_{\mathbf{d}}| \lesssim |I_{\mathbf{ab}}| \lesssim |I_{\mathbf{e}}||I_{\mathbf{d}}|.$$

Then fix $(\mathbf{e}, y) \in \mathcal{P}$. Since $B(y, \sigma) \subset B(y, e^{-\varepsilon_0 m})$ we have by (5.1) that

$$\sharp\{\mathbf{d} \in \tilde{\mathcal{R}}_m(\varepsilon) : e^{2\lambda n} f'_{\mathbf{ed}}(x) \in B(y, \sigma)\} \lesssim C^{\kappa_0}_{\varepsilon,m} e^{-c_0 m} \sharp\tilde{\mathcal{R}}_m(\varepsilon)$$

Note that $e^{-c_0 m} = e^{-\frac{c_0}{\varepsilon_0}\varepsilon_0 m} \lesssim \sigma^{c_0/\varepsilon_0}$ so the claim follows by the cardinality bounds of $\tilde{\mathcal{R}}_{2n-m}(\varepsilon)$ and $\tilde{\mathcal{R}}_m(\varepsilon)$ by setting in the statement of the lemma the exponent $c_0 > 0$ as c_0/ε_0.

Let us now verify the non-concentration estimate (5.1) we need.

Step 1. Write $r := e^{-\varepsilon_0 m}$. Since $y - r \geq R^{-1} > 0$, we know that $e^{2\lambda n} f'_{\mathbf{ed}}(x) \in B(y, r)$ if and only if

$$-\log|f'_{\mathbf{ed}}(x)| \in J := [2\lambda n - \log(y+r), 2\lambda n - \log(y-r)].$$

Note that the interval J has length $|J| = \log\frac{y+r}{y-r}$. By the mean value theorem, we have

$$2R^{-1}r \leq \frac{2r}{y+r} \leq |J| \leq \frac{2r}{y-r} \leq 2Rr. \tag{5.2}$$

Hence

$$\sharp\{\mathbf{d} \in \tilde{\mathcal{R}}_m(\varepsilon) : e^{2\lambda n} f'_{\mathbf{ed}}(x) \in B(y, e^{-\varepsilon_0 m})\} = \sharp\{\mathbf{d} \in \tilde{\mathcal{R}}_m(\varepsilon) : -\log|f'_{\mathbf{ed}}(x)| \in J\}$$

Step 2. Now let us approximate the indicator of χ_J by a mollifier $h \in C^4(\mathbb{R})$ satisfying

(1) $\chi_J \leq h$

(2) $\|h\|_1 \lesssim |J|$

(3) $\|h''''\|_{L^1} \lesssim \frac{1}{|J|^3}$.

This function can be obtained, for example using a scaled and translated Gaussian function

$$h(x) := e^{\frac{\pi}{4}} g_0\Big(\frac{x - x_J}{|J|}\Big), \quad \text{where} \quad g_0(x) := e^{-\pi x^2}.$$

where x_J is the central point of J. Then Since $\chi_J \leq h$, we have

$$\sharp\{\mathbf{d} \in \tilde{\mathcal{R}}_m(\varepsilon) : -\log|f'_{\mathbf{ed}}(x)| \in J\} \leq \sum_{\mathbf{d} \in \tilde{\mathcal{R}}_n(\varepsilon)} h(-\log|f'_{\mathbf{ed}}(x)|)^{1/2}$$

We use Cauchy-Schwartz here, that is

$$\Big(\sum_{\mathbf{d} \in \tilde{\mathcal{R}}_m(\varepsilon)} h(-\log|f'_{\mathbf{ed}}(x)|)^{1/2}\Big)^2$$
$$\leq \Big(\sum_{\mathbf{d} \in \tilde{\mathcal{R}}_m(\varepsilon)} |f'_{\mathbf{ed}}(x)|^\delta h(-\log|f'_{\mathbf{ed}}(x)|)\Big)\Big(\sum_{\mathbf{d} \in \tilde{\mathcal{R}}_m(\varepsilon)} \frac{1}{|f'_{\mathbf{ed}}(x)|^\delta}\Big)$$

Step 3. Taking the inverse Fourier transform of $\widehat{h}$ gives us for all $x \in X$, $\mathbf{d} \in \mathcal{A}^m$ and $m \in \mathbb{N}$ that

$$h(-\log |f'_{\mathbf{ed}}(x)|) = \int e^{-2\pi i \xi \log |f'_{\mathbf{ed}}(x)|} \widehat{h}(\xi)\, d\xi$$

Therefore

$$\begin{aligned}\sum_{\mathbf{d} \in \tilde{\mathcal{R}}_m(\varepsilon)} |f'_{\mathbf{ed}}(x)|^\delta h(-\log |f'_{\mathbf{ed}}(x)|) &= \sum_{\mathbf{d} \in \tilde{\mathcal{R}}_m(\varepsilon)} |f'_{\mathbf{ed}}(x)|^\delta \int \widehat{h}(\xi) e^{-2\pi i \xi \log |f'_{\mathbf{ab}}(x)|}\, d\xi \\ &= \int \widehat{h}(\xi) \sum_{\mathbf{d} \in \tilde{\mathcal{R}}_m(\varepsilon)} |f'_{\mathbf{ed}}(x)|^\delta e^{-2\pi i \xi \log |f'_{\mathbf{ed}}(x)|}\, d\xi\end{aligned}$$

Split the integration now over $|\xi| > t_0/2\pi$ and $|\xi| \leq t_0/2\pi$.

Step 4. If $|\xi| > t_0/2\pi$, we will estimate as follows. Inside the integral, use the estimate

$$\sum_{\mathbf{d} \in \tilde{\mathcal{R}}_m(\varepsilon)} |f'_{\mathbf{ed}}(x)|^\delta h(-\log |f'_{\mathbf{ed}}(x)|) \leq \sum_{\mathbf{d} \in \mathcal{A}^m} |f'_{\mathbf{ed}}(x)|^\delta h(-\log |f'_{\mathbf{ed}}(x)|)$$

We can iterate the definition of the complex transfer operator applied for $g \in C^1(X)$ defined by

$$g_{\mathbf{e}}(z) := |f'_{\mathbf{e}}(z)|^s, \quad z \in X$$

to obtain with $s = \delta - 2\pi\xi i$ that

$$\mathcal{L}^n_{-s\tau} g_{\mathbf{e}}(x) = \sum_{\mathbf{d} \in \mathcal{A}^m} g_{\mathbf{e}}(f_{\mathbf{d}}(x)) e^{s \log |f'_{\mathbf{d}}(x)|} = \sum_{\mathbf{d} \in \mathcal{A}^m} |f'_{\mathbf{ed}}(x)|^\delta e^{-2\pi i \xi \log |f'_{\mathbf{ed}}(x)|} \tag{5.3}$$

whenever $x \in X$ with $\mathbf{d} \to x$ since $|f'_{\mathbf{e}}(z)|^s = |f'_{\mathbf{e}}(z)|^\delta e^{-2\pi i \xi \log |f'_{\mathbf{e}}(z)|}$ and using the chain rule. Here we will employ the C^1 contraction, Theorem 5.3.2 with first a fixed $0 < \Xi < 1$ to obtain

$$\operatorname{Re} s = \delta \in (s_0 - \delta_0(\Xi), s_0]$$

as long as $\delta = \dim_{\mathrm{H}} \mu \geq s_0 - \delta_0(\Xi)$ and

$$|\operatorname{Im} s| = 2\pi|\xi| \geq t_0(\Xi)$$

by the choice of ξ. Thus we have for some $\delta_1(\Xi) > 0$ that

$$\begin{aligned}
&\int_{|\xi|>t_0/2\pi} \widehat{h}(\xi) \sum_{\mathbf{d}\in\mathcal{A}^m} |f'_{\mathbf{ed}}(x)|^\delta e^{-2\pi i\xi \log|f'_{\mathbf{ed}}(x)|}\, d\xi \\
&\le \int_{|\xi|>t_0/2\pi} |\widehat{h}(\xi)| \cdot ||\mathcal{L}^n_{-s\tau} g_{\mathbf{e}}||_{C^1}\, d\xi \\
&\le \int_{|\xi|>t_0/2\pi} |\widehat{h}(\xi)| \cdot C_\Xi |\mathrm{Im}\, s|^{1+\Xi} e^{-\delta_1(\Xi)n} \|g_{\mathbf{e}}\|_{C^1}\, d\xi \\
&\lesssim_\Xi C_{\varepsilon,n} e^{-\lambda\delta(2n-m)} e^{-\delta_1(\Xi)n} \int_{|\xi|>t_0/2\pi} |\widehat{h}(\xi)| \cdot |\xi|^{2+\Xi}\, d\xi
\end{aligned}$$

where χ_I is the constant function on I. Here we used the bounded distortion for T to bound the C^1 norm of $g_{\mathbf{e}}$. First of all, we have

$$|g_{\mathbf{e}}(z)| = |f_{\mathbf{e}}(z)|^\delta \le C_{\varepsilon,n} e^{-\lambda\delta(2n-m)} \quad \text{and} \quad |g'_{\mathbf{e}}(z)| \lesssim |\xi||f_{\mathbf{e}}(z)|^\delta \le C_{\varepsilon,n} e^{-\lambda\delta(2n-m)} |\xi|$$

so

$$\|g_{\mathbf{e}}\|_{C^1} \lesssim C_{\varepsilon,n} e^{-\lambda\delta(2n-m)} |\xi|.$$

Indeed, after fixing a branch of the logarithm, using bounded distortion $|f''_{\mathbf{e}}(z)| \le B|f'_{\mathbf{e}}(z)|$, and $|s| \lesssim |\xi|$, we obtain that

$$|g_{\mathbf{e}}(z)| = \frac{|s|}{|f'_{\mathbf{e}}(z)|} |\exp(s \log|f'_{\mathbf{e}}(z)|)||f''_{\mathbf{e}}(z)| = \frac{|s||T''(f_{\mathbf{e}} z)|}{|T'(f_{\mathbf{e}} z)|^2} |g_{\mathbf{e}}(z)| \lesssim |\xi||g_{\mathbf{e}}(z)|$$

where we use the inverse rule for differentiable functions and the chain rule.

Using integration by parts, we have for the Fourier transform $\widehat{g}$ that for all $\xi \in \mathbb{R}$ the following estimate holds:

$$|\widehat{h}(\xi)| \le \frac{1}{1+|2\pi\xi|^4} (\|h\|_{L^1} + \|h''''\|_{L^1}).$$

Then in particular as $\Xi < 1$, we have

$$\int_{|\xi|>t_0/2\pi} |\widehat{h}(\xi)| \cdot |\xi|^{2+\Xi}\, d\xi \le \int \frac{\xi^{2+\Xi}}{1+|2\pi\xi|^4} (\|h\|_{L^1} + \|h''''\|_{L^1})\, d\xi \lesssim \|h\|_{L^1} + \|h''''\|_{L^1}.$$

Step 5. We are left with the case $|\xi| \le t_0/2\pi$, that is, an estimation for

$$\int_{|\xi|\le t_0/2\pi} \widehat{h}(\xi) \sum_{\mathbf{d}\in\tilde{\mathcal{R}}_m(\varepsilon)} |f'_{\mathbf{ed}}(x)|^\delta e^{-2\pi i\xi \log|f'_{\mathbf{ed}}(x)|}\, d\xi$$

Let us bound

$$\begin{aligned}
\sup_{|\xi|\le t_0/2\pi} \widehat{h}(\xi) \sum_{\mathbf{d}\in\tilde{\mathcal{R}}_m(\varepsilon)} |f'_{\mathbf{ed}}(x)|^\delta e^{-2\pi i\xi \log|f'_{\mathbf{ed}}(x)|} &\le \sup_{|\xi|\le t_0/2\pi} |\widehat{h}(\xi)| \sum_{\mathbf{d}\in\tilde{\mathcal{R}}_m(\varepsilon)} |f'_{\mathbf{ed}}(x)|^\delta \\
&\le C_{\varepsilon,2n-m} e^{-\lambda\delta(2n-m)} |J|
\end{aligned}$$

since $|\widehat{h}(\xi)| \leq \|h\|_{L^1} \leq 3|J|$ and by the chain rule

$$\sum_{\mathbf{d}\in\tilde{\mathcal{R}}_m(\varepsilon)} |f'_{\mathbf{ed}}(x)|^\delta = \sum_{\mathbf{d}\in\tilde{\mathcal{R}}_m(\varepsilon)} |f'_{\mathbf{e}}(f_{\mathbf{d}}(x))|^\delta |f'_{\mathbf{d}}(x)|^\delta$$
$$\lesssim C_{\varepsilon,2n-m} e^{-\lambda\delta(2n-m)} \sum_{\mathbf{d}\in\tilde{\mathcal{R}}_m(\varepsilon)} |f'_{\mathbf{d}}(x)|^\delta$$
$$\lesssim C_{\varepsilon,2n-m} e^{-\lambda\delta(2n-m)}.$$

by the properties of $\tilde{\mathcal{R}}_m(\varepsilon)$.

Step 6. Combining Step 1, 2, 3, 4 and 5 gives us

$$\sharp\{\mathbf{d} \in \tilde{\mathcal{R}}_m(\varepsilon) : e^{2\lambda n} f'_{\mathbf{ed}}(x) \in B(y, e^{-\varepsilon_0 m})\}^2$$
$$\lesssim_\Xi C_{\varepsilon,2n-m} e^{-\lambda\delta(2n-m)} E_{\mathbf{e}}(x)[e^{-\delta_1(\Xi)m}(\|h\|_{L^1} + \|h''''\|_{L^1}) + |J|],$$

where

$$E_{\mathbf{e}}(x) := \sum_{\mathbf{d}\in\tilde{\mathcal{R}}_m(\varepsilon)} \frac{1}{|f'_{\mathbf{ed}}(x)|^\delta}.$$

Finally, let us now analyse all the quantities we have. Lemma 2.2.3 gives that

$$E_{\mathbf{e}}(x) \lesssim C^\delta_{\varepsilon,2m} e^{2\lambda\delta n} \sharp\tilde{\mathcal{R}}_m(\varepsilon)$$

so for some $\kappa > 0$

$$C_{\varepsilon,2n-m} e^{-\lambda\delta(2n-m)} E_{\mathbf{e}}(x) \lesssim C^\kappa_{\varepsilon,n} \sharp\tilde{\mathcal{R}}_m(\varepsilon)^2$$

Moreover, recall that (5.2) gives

$$|J| \leq 2Rr$$

and

$$\frac{1}{|J|^3} \leq \frac{1}{2} R^3 r^{-3}$$

and when inputting $r = e^{-\varepsilon_0 m}$ and $R = 16^2 C C^{3\lambda}_{\varepsilon,m}$, we obtain

$$|J| \lesssim C^{3\lambda}_{\varepsilon,m} e^{-\varepsilon_0 m}$$

and

$$\frac{1}{|J|^3} \lesssim C^{9\lambda}_{\varepsilon,m} e^{3\varepsilon_0 m}$$

Then by the choice of h, we have

$$\|h\|_{L^1} + \|h''''\|_{L^1} \leq |J| + \frac{1}{|J|^3} \lesssim C^{3\lambda}_{\varepsilon,m} e^{-\varepsilon_0 m} + C^{9\lambda}_{\varepsilon,m} e^{3\varepsilon_0 m}.$$

Thus we obtain

$$\begin{aligned}&\sharp\{\mathbf{d} \in \tilde{\mathcal{R}}_m(\varepsilon) : e^{2\lambda n} f'_{\mathbf{ed}}(x) \in B(y, e^{-\varepsilon_0 m})\}^2 \\ &\lesssim C^{\kappa}_{\varepsilon,n} \sharp\tilde{\mathcal{R}}_m(\varepsilon)^2 [e^{-\delta_1(\Xi)m}(C^{3\lambda}_{\varepsilon,m} e^{-\varepsilon_0 m} + C^{9\lambda}_{\varepsilon,m} e^{3\varepsilon_0 m}) + C^{3\lambda}_{\varepsilon,m} e^{-\varepsilon_0 m}]\end{aligned}$$

Here we see that the possible obstacle to the decay would come from the term

$$e^{-\delta_1(\Xi)m} e^{3\varepsilon_0 m} = e^{-(\delta_1(\Xi)-3\varepsilon_0)m}.$$

But since we defined $\varepsilon_0 = \delta_1(\Xi)/4$, we obtain $c_0 := \delta_1(\Xi) - 3\varepsilon_0 > 0$. This completes the proof of the estimate (5.1) and thus the whole lemma. □

5.4 Polynomial Fourier decay

We begin the proof by fixing first $\varepsilon > 0$ small enough that $C^{\kappa}_{\varepsilon,n} = e^{\kappa\varepsilon n}$ have the exponent $\kappa\varepsilon$ small enough in terms of λ, s, ε_0 and c_0. To be able to apply relevant large deviation results, we need to make sure that the values of n that we consider are sufficiently large. We begin by choosing $n_0(\varepsilon)$.

1. If n_1 is the generation that arises from the main large deviation theorem, then we require
$$n_0(\varepsilon)\varepsilon_0 > n_1$$
to ensure we have valid regularity at each scale that we need.

2. If γ is the rate of expansion of $(T^n)'$ with respect to n, and C is the Gibbs constant for μ, we require
$$\frac{\log 4}{\varepsilon_0 n_0} < \varepsilon_2, \quad \frac{\log 4C^2}{\log(\gamma^{2\varepsilon_0 n_0})} < \varepsilon/2 \quad \text{and} \quad \frac{e^{-\delta\varepsilon_0 n_0}}{1-e^{-\delta}} < e^{-\delta\varepsilon_0 n_0/2}$$
to ensure that we get decay on multiregular blocks of words.

Let $s = \dim_{\mathrm{H}} \mu$ and λ the Lyapunov exponent of μ. Let $k \in \mathbb{N}$ and $\varepsilon_2 > 0$ from Lemma 3.1.1.

Fix a frequency $\xi \in \mathbb{R}$ such that $|\xi|$ is large enough. Let $n \in \mathbb{N}$ be the number such that

$$e^{(2k+1)n\lambda} e^{\varepsilon_0 n} \le |\xi| \le e^{(2k+1)(n+1)\lambda} e^{\varepsilon_0(n+1)}.$$

so up to a multiplicative constant depending on k and ε_0, we have:

$$|\xi| \sim e^{(2k+1)n\lambda} e^{\varepsilon_0 n},$$

where $|\xi| \sim N$ means that there exists a constant $c > 0$ such that $c^{-1}N \leq |\xi| \leq cN$. Recall that

$$|\mathcal{R}_n(\varepsilon)| \lesssim C_{\varepsilon,n}^{3\lambda} e^{-\lambda\delta n}$$

and if $\mathbf{a} \in \mathcal{R}_n(\varepsilon)$, we have for all $x \in X$ that

$$w_{\mathbf{a}}(x) \leq C_{\varepsilon,n}^{3\lambda} e^{-\lambda\delta n}.$$

We begin by recalling the estimate from Proposition 3.3.1. Recall that there we have

$$J_n(\varepsilon) = \{\eta \in \mathbb{R} : e^{\varepsilon_0 n/2} \leq |\eta| \leq C_{\varepsilon,n} e^{\varepsilon_0 n}\}.$$

Proposition 5.4.1. *The assumption of Lemma 3.3.1 on $\mu \times \mu$ holds for $\tau = e^{-2\varepsilon_0 n}$.*

Proof. By covering the n-regular part of the following set with $\lfloor n/4 \rfloor$-generation parent intervals, for fixed $y \in X$ we have that

$$\begin{aligned}
&\mu(\{x \in X : |x-y| \leq C_0 e^{-\lambda n/4}\}) \\
&\leq \mu(X \setminus R_n) + \mu(\{x \in R_n : |x-y| \leq C_0 e^{-\lambda n/4}\}) \\
&\leq \mu(X \setminus R_n) + \left\lceil \frac{2C_0 e^{-\lambda n/4}}{C_{\varepsilon,n}^{-1} e^{-\lambda\lfloor n/4 \rfloor}/16} \right\rceil e^{-\lambda s \lfloor n/4 \rfloor} \\
&\leq \mu(X \setminus R_n) + 64 C_0 C_{\varepsilon,n} e^{-\lambda s(n/4-1)} \\
&\leq \mu(X \setminus R_n) + 64 e^{\lambda} C_0 C_{\varepsilon,n} e^{-\lambda s n/4}.
\end{aligned}$$

Hence we have that

$$\mu \times \mu(\{(x,y) \in X^2 : |x-y| \leq C_0 e^{-\lambda n/4}\}) \leq \mu(X \setminus R_n) + 64 e^{\lambda} C_0 C_{\varepsilon,n} e^{-\lambda s n/4}$$

as required. □

We have the following estimate in terms of exponential sums and error terms:

$$\begin{aligned}
|\widehat{\mu}(\xi)|^2 \lesssim_\mu\ & C_{\varepsilon,n}^{(2k+1)\lambda} e^{-\lambda(2k+1)\delta n} \sum_{\mathbf{A} \in \mathcal{R}_n^{k+1}(\varepsilon)} \sup_{\eta \in J_n(\varepsilon)} \Bigg| \sum_{\mathbf{B} \in \mathcal{R}_n^k(\varepsilon)} e^{2\pi i \eta \zeta_{1,\mathbf{A}}(\mathbf{b_1}) \ldots \zeta_{k,\mathbf{A}}(\mathbf{b}_k)} \Bigg| \\
&+ e^{2k} C_{\varepsilon,n}^{k+2} e^{-\lambda n} e^{\varepsilon_0 n} + \mu(X \setminus R_n^{k+1}(\varepsilon))^2 \\
&+ e^{-\varepsilon_2 n/2} + \mu(I \setminus R_n(\varepsilon)) + C_{\varepsilon,n}^2 e^{-\delta\varepsilon_0 n/2},
\end{aligned}$$

where if blocks $\mathbf{A} \in \mathcal{R}_n^{k+1}(\varepsilon), \mathbf{B} \in \mathcal{R}_n^k(\varepsilon)$, $j \in \{1, \dots, k\}$ and $\mathbf{b} \in \mathcal{R}_n$, we defined

$$\zeta_{j,\mathbf{A}}(\mathbf{b}) = e^{2\lambda n} f'_{\mathbf{a_{j-1}b}}(x_{\mathbf{a_j}}).$$

Now, recall that in Proposition 3.2.2, we defined the non-concentrated blocks of words $\mathbf{A} \in \mathcal{W}$ as follows: for all $j = 1, \dots, k$, $\eta \in J_n(\varepsilon)$ and $\sigma \in [R^2|\eta|^{-1}, |\eta|^{-\varepsilon_3}]$, where we have that

$$|\{(\mathbf{b}, \mathbf{c}) \in \mathcal{R}_n(\varepsilon)^2 : |\zeta_{j,\mathbf{A}}(\mathbf{b}) - \zeta_{j,\mathbf{A}}(\mathbf{c})| \le \sigma\}| \le \sharp\mathcal{R}_n(\varepsilon)^2 \sigma^{c_0/2}.$$

Proposition 3.2.2 said most blocks are non-concentrated: there exists $\kappa_0 > 0$,

$$e^{-\lambda(k+1)\delta n} |\mathcal{R}_n^{k+1}(\varepsilon) \setminus \mathcal{W}| \le C_{\varepsilon,n}^{2\kappa_0} \sigma^{c_0/4}.$$

For the exponential sum term

$$C_{\varepsilon,n}^{(2k+1)\lambda} e^{-\lambda(2k+1)\delta n} \sum_{\mathbf{A} \in \mathcal{R}_n^{k+1}(\varepsilon)} \sup_{\eta \in J_n(\varepsilon)} \Big| \sum_{\mathbf{B} \in \mathcal{R}_n^k(\varepsilon)} e^{2\pi i \eta \zeta_{1,\mathbf{A}}(\mathbf{b_1}) \dots \zeta_{k,\mathbf{A}}(\mathbf{b}_k)} \Big|$$

in the estimate for $|\widehat{\mu}(\xi)|$ we begin by removing not non-concentrated blocks, which gives that for some $\kappa_0' > 0$,

$$\begin{aligned}
&C_{\varepsilon,n}^{(2k+1)\lambda} e^{-(2k+1)\lambda\delta n} \sum_{\mathbf{a} \in \mathcal{R}_n^{k+1} \setminus \mathcal{W}} \sup_{\eta \in J_n(\varepsilon)} \Big| \sum_{\mathbf{B} \in \mathcal{R}_n^k(\varepsilon)} e^{2\pi i \eta \zeta_{1,\mathbf{A}}(\mathbf{b_1}) \dots \zeta_{k,\mathbf{A}}(\mathbf{b}_k)} \Big| \\
&\le C_{\varepsilon,n}^{(2k+1)\lambda} e^{-(2k+1)\lambda\delta n} \sum_{\mathbf{A} \in \mathcal{R}_n^{k+1} \setminus \mathcal{W}} \sup_{\eta \in J_n(\varepsilon)} \sum_{\mathbf{B} \in \mathcal{R}_n^k(\varepsilon)} 1 \\
&\le C_{\varepsilon,n}^{(2k+1)\lambda} e^{-(2k+1)\lambda\delta n} \sum_{\mathbf{A} \in \mathcal{R}_n^{k+1} \setminus \mathcal{W}} C^k C_{\varepsilon,n}^{3\lambda k} e^{k\lambda\delta n} \\
&\lesssim C^k C_{\varepsilon,n}^{(5k+7)\lambda} e^{-(k+1)\lambda\delta n} e^{(k+1)\lambda\delta n} \sigma^{c_0/4} \\
&\lesssim C_{\varepsilon,n}^{\kappa_0'} \sigma^{c_0/4}.
\end{aligned}$$

Hence we have that

$$\begin{aligned}
|\widehat{\mu}(\xi)|^2 \lesssim\ & C_{\varepsilon,n}^{k\lambda} e^{-k\lambda\delta n} \max_{\mathbf{A} \in \mathcal{W}} \sup_{\eta \in J_n(\varepsilon)} \Big| \sum_{\mathbf{B} \in \mathcal{R}_n^k(\varepsilon)} e^{2\pi i \eta \zeta_{1,\mathbf{A}}(\mathbf{b_1}) \dots \zeta_{k,\mathbf{A}}(\mathbf{b}_k)} \Big| \\
&+ C_{\varepsilon,n}^{\kappa_0'} \sigma^{c_0/4} + e^{2k} C_{\varepsilon,n}^{k+2} e^{-\lambda n} e^{\varepsilon_0 n} + \mu(X \setminus R_n^{k+1}(\varepsilon))^2 \\
&+ e^{-\varepsilon_2 n/2} + \mu(I \setminus R_n(\varepsilon)) + C_{\varepsilon,n}^2 e^{-\delta\varepsilon_0 n/2}.
\end{aligned}$$

Recall Remark 5.3.1, where we defined $R := R(n, \varepsilon) := 16^2 C C_{\varepsilon,n}^{3\lambda}$. Thus $\zeta_{j,\mathbf{A}}(\mathbf{b}) \in [R^{-1}, R]$. Moreover, if we fix $\eta \in J_n(\varepsilon)$, $A \in \mathcal{W}$ and $\sigma \in [R^{-2}|\eta|^{-1}, |\eta|^{-\varepsilon_3}]$, we have by the definition of $\mathcal{W}$ that

$$\sharp\{(\mathbf{b}, \mathbf{c}) \in \mathcal{R}_n(\varepsilon)^2 : |\zeta_{j,\mathbf{A}}(\mathbf{b}) - \zeta_{j,\mathbf{A}}(\mathbf{c})| \le \sigma\} \le \sharp\mathcal{R}_n(\varepsilon)^2 \sigma^{c_0/2}$$

Thus we may apply Lemma 3.1.2 to the maps $\zeta_{j,\mathbf{A}} : \mathcal{R}_n(\varepsilon) \to [R^{-1}, R]$ with $N = e^{\lambda\delta n}$. It implies that for all $\mathbf{A} \in \mathcal{W}$ and $\eta \in J_n(\varepsilon)$ that

$$C_{\varepsilon,n}^{k\lambda} e^{-k\lambda\delta n} \Big| \sum_{\mathbf{B} \in \mathcal{R}_n^k(\varepsilon)} e^{2\pi i \eta \zeta_{1,\mathbf{a}}(\mathbf{B}_1) \dots \zeta_{k,\mathbf{a}}(\mathbf{b}_k)} \Big| \lesssim R^{2k\lambda} |\eta|^{-\varepsilon_2} \lesssim C_{\varepsilon,n}^{9k\lambda} e^{-\varepsilon_0 \varepsilon_2 n/2}$$

since $|\eta| \geq C_{\varepsilon,n}^{-1} e^{\varepsilon_0 n/2}$ by the definition of $J_n(\varepsilon)$ as $\eta \in J_n(\varepsilon)$. By making sure that $\varepsilon > 0$ is chosen small enough and as $\varepsilon_0 \leq \lambda/2 < \lambda$, recall Remark 5.3.1(2) for the choice of ε_0 using the spectral gap of the transfer operator, which is independent of ξ, we have proved that for some $\alpha > 0$,

$$|\widehat{\mu}(\xi)| = O(|\xi|^{-\alpha})$$

as $|\xi| \to \infty$. The proof of Theorem 5.2.3 is complete.

5.5 The case of not totally-nonlinear dynamics

As previously mentioned, the work of Mosquera–Shmerkin [42] presents an analogous Fourier decay theorem for a large class of cases when the dynamics is not totally-nonlinear. Using their results, we can prove Fourier decay for the case when $\log |T'|$ is cohomologous to a constant function. Note that this is not precisely the compliment to totally-nonlinear dynamics, because that includes locally constant functions, not just constant functions. This corresponds to studying measures invariant under self-similar IFS's and *homogenous* self-similar IFS's (the details will be explained in this subsection).

Let us consider the cohomologous to a constant case. So we assume that there exist some diffeomorphism $g : X \to \mathbb{R}$ and constant function c such that

$$\log |T'| = c - g \circ T + g$$

for our dynamics $T : X \to X$. We will require a function $h : X \to \mathbb{R}$ such that $g = \log h'$, so define

$$h(x) := \int_{y \leq x} e^{g(y)} \, dy$$

which is well-defined since g is Lipschitz. We also note that h' is non-zero, and hence h is a diffeomorphism by the inverse function theorem. We begin by identifying the iterated function system which we will apply to Mosquera–Shmerkin [42].

Lemma 5.5.1. *Consider a diffeomorphism $h : X \to \mathbb{R}$. Then*

$$\log |T'| = c - \log h' \circ T + \log h'$$

if and only if for all $x \in X$ and $a \in \mathcal{A}$

$$(hf_{\mathbf{a}}h^{-1})'(x) = e^{-c}$$

where $f_a : X \to I_a$ represents an inverse branch for the dynamics T.

Proof. Let $y = h^{-1}(x)$. We have that for all $a \in \mathcal{A}$ and all $x \in X$

$$\begin{aligned}(hf_a h^{-1})'(x) = e^{-c} &\iff h'(f_a h^{-1}x) \cdot f_a'(h^{-1}x) \cdot (h^{-1})'(x) = e^{-c} \\ &\iff \log h'(f_a y) - \log |T'(f_a y)| - \log h'(y) = -c \\ &\iff \log |T'(x)| = c + \log h' \circ T(x) - \log h'(x)\end{aligned}$$

as required. □

So we have that

$$\{hf_a h^{-1} : a \in \mathcal{A}\}$$

is a homogenous self-similar IFS with contraction ratio e^{-c}. We now relate a Bernoulli measure μ under the dynamics T to a Bernoulli measure for the given IFS. Assume that

$$\mu = \sum_{a \in \mathcal{A}} p_a f_a \mu$$

for some Bernoulli weights $p_a \geq 0$.

Lemma 5.5.2. *We have that if $\mu = (h^{-1}\nu)$ is Bernoulli under T then ν is Bernoulli for the IFS $\{hf_a h^{-1} : a \in \mathcal{A}\}$.*

Proof. For a test function φ we have that

$$\begin{aligned}\int \varphi \, d\nu &= \int \varphi \circ h(x) \, d\mu(x) \\ &= \sum_{a \in \mathcal{A}} p_a \int \varphi f_a h(x) \, d\mu \\ &= \sum_{a \in \mathcal{A}} p_a \int \varphi h^{-1} f_a h(x) \, dh\mu \\ &= \int \varphi(x) \, d\Big(\sum_{a \in \mathcal{A}} p_a (h^{-1} f_a h)\nu \Big)\end{aligned}$$

so μ and ν have the same Bernoulli weights. □

We can conclude that Mosquera–Shmerkin is directly applicable to this situation with the following theorem.

Theorem 5.5.3. *Let μ be a Bernoulli measure for a map $T : X \to X$. Assume that there exists a diffeomorphism $g \in C^2(\mathbb{R}, \mathbb{R})$ such that for some constant function $c : X \to \mathbb{R}$*

$$\log |T'| = c + g - g \circ T.$$

Then μ has polynomial Fourier decay.

Chapter 6

Convex Cocompact Fuchsian groups

Once we proved the non-concentrated derivatives assumption for Gibbs measures of the Gauss map, the next question was whether we could extend this to more maps. This is quite a difficult problem, and we were not even sure that we had the tools to solve it before our introduction to complex transfer operator theory. The diophantine tools of Queffélec and Ramaré were required for the Gauss map proof, and the Schottky tree structure for the Fuchsian group setting. At this point, we knew it would be interesting to try and further understand why these cases work. Of course the Fuchsian group setting is well understood thanks to the work of Bourgain–Dyatlov [8]. In this setting, they could use the fact that distortions of the form $\gamma_{\mathbf{a}}''(x)/\gamma_{\mathbf{a}}'(x)$ are comparable to the geometric point $\gamma_{\mathbf{a}}^{-1}(\infty)$. The measure and Schottky structure can then be used to analyse these points. This work then lead to the Gauss map setting, where we could similarly say that the differences of the distortions $f_{\mathbf{a}}''(x)/f_{\mathbf{a}}'(x) - f_{\mathbf{b}}''(x)/f_{\mathbf{b}}'(x)$ are comparable to the difference of geometric points $p_n(\mathbf{a}^{\leftarrow})/q_n(\mathbf{a}^{\leftarrow}) - p_n(\mathbf{b}^{\leftarrow})/q_n(\mathbf{b}^{\leftarrow})$. We can then similarly use the measure to analyse these differences using the ideas of Queffélec and Ramaré presented in Jordan–Sahlsten [50][26]. To help understand why these cases work, we decided to look deeper into the Fuchsian group setting. A natural question to ask is whether we can extend the proof of Bourgain–Dyatlov to a wider class of measures. The aim here would be to prove a result for statistical (Gibbs) measures similarly to the Gauss map setting. This means identifying the necessary properties of Patterson–Sullivan measures to get a result. Bourgain–Dyatlov used the fact that the Patterson–Sullivan measures are 'Ahlfors–David' regular, and that they 'preserve inversion' which shall be explored in this chapter. Note that Bourgain–Dyatlov [8] refer to 'inverting' as 'reversing', but to avoid confusion with the Gauss map case where we do actually reverse words (referring to the use of $\mathbf{a}^{\leftarrow}$), we avoid using the word 'reverse' here.

6.1 Schottky Structure

We consider the case when we have a Fuchsian group which is convex and cocompact, i.e. a Fuchsian group giving a hyperbolic surface $\mathbb{H}/\Gamma$ which is geometrically finite (has a fundamental domain which is a finite-sided convex polygon) and does not have any cusps. In this case the surface has infinite area. In this set up, we have a great way to think about the set of transformations Γ.

We begin by considering disjoint closed half disks $D_1, \dots, D_{2r} \subset \bar{\mathbb{H}} := \mathbb{H} \cup \partial\mathbb{H}$ that intersect the real line at right angles. Define an alphabet $\mathcal{A} = \{1, 2, \dots, 2r\}$. For $j \in \{1, \dots, r\}$, γ_j will be defined to be the the transformation which sends $\mathbb{H} \backslash D_{j+r}$ into D_j, and hence for bijectivity, γ_j sends D_{j+r} to $\mathbb{H} \backslash D_j$. The action of this transformation can be extended to the boundary $\partial\mathbb{H}$ in the usual way [67]. We will define $\bar{j} := j + r \mod 2r$ for any $j \in \mathcal{A}$, and we let $\gamma_{\bar{j}} := \gamma_j^{-1}$. We will define the first generation construction intervals $I_a := D_a \cap \partial\mathbb{H}$ for $a \in \mathcal{A}$.

Definition 6.1.1. *We say that a Fuchsian group Γ has a Schottky structure if there exists such a set of half disks $\{D_j : j \in \mathcal{A}\}$ such that the corresponding set of transformations $\{\gamma_j : j \in \mathcal{A}\}$ generate Γ.*

It is worth noting that the set $\mathbb{H} \setminus \cup_{j \in \mathcal{A}} D_j$ is a natural fundamental domain for the surface $\mathbb{H}/\Gamma$, and the disjointness of the half disks means that the surface has infinite surface area. This also means that the Fuchsian group is convex and cocompact. The following theorem is given in Borthwick's book [4].

Theorem 6.1.2. *Suppose the $\mathbb{H}/\Gamma$ is a geometrically finite, infinite area hyperbolic surface without cusps. Then the surface is homeomorphic to another hyperbolic surface whose corresponding Fuchsian group has a Schottky structure.*

We will consider such surfaces from now on. So the Schottky structure without inverses, namely the set $\{\gamma_1, \dots, \gamma_r\}$ will generate the Fuchsian group, so to consider

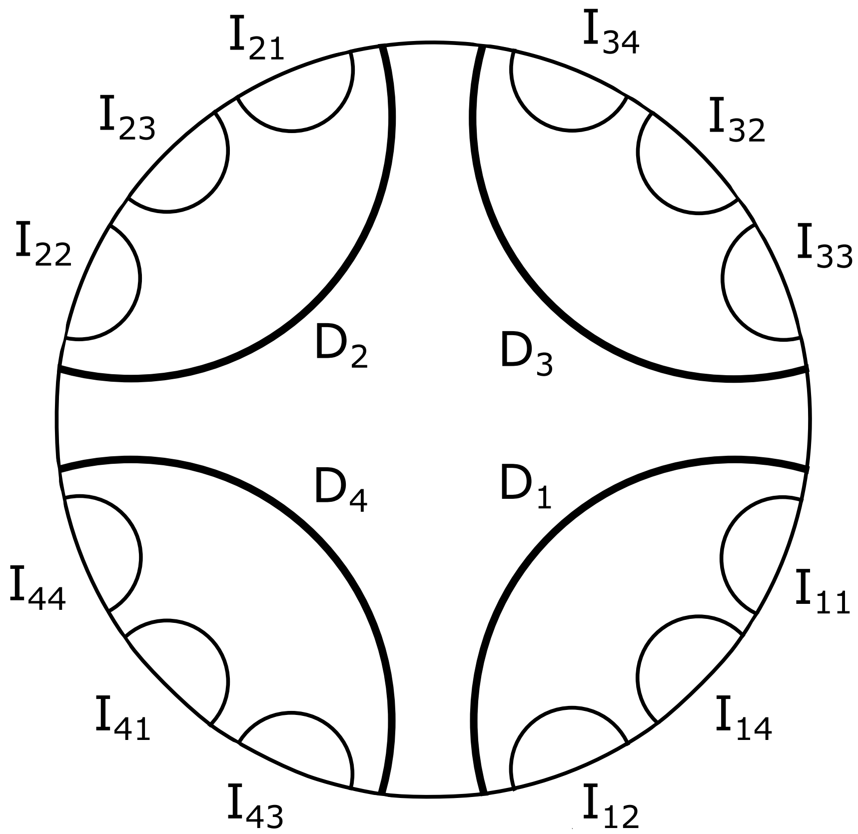

Figure 6.1: A Schottky structure for $r = 2$ when considered in the Poincaré disk model.

admissible words in the alphabet $\mathcal{A}$, we only need require a letter j to not be followed by its inverse $\bar{j}$. We will define words of length n as

$$\mathcal{A}_n := \{a_1 \dots a_n \in \mathcal{A}^n : a_j \neq \overline{a_{j+1}} \, \forall j = 1, \dots, n-1\}$$

and let $\mathcal{W} := \cup_n \mathcal{A}_n$ be the set of all finite length words. Given a word $\mathbf{a} := a_1 \dots a_n \in \mathcal{A}_n$, we will define its inverse to be $\bar{\mathbf{a}} := \bar{a_n} \dots \bar{a_1} \in \mathcal{A}_n$. This is defined as such since $\gamma_{\mathbf{a}}\gamma_{\bar{\mathbf{a}}} = \gamma_{\bar{\mathbf{a}}}\gamma_{\mathbf{a}}$ will be the identity transformation, where we will define $\gamma_{\mathbf{a}} := \gamma_{a_1} \circ \dots \circ \gamma_{a_n}$. We will define $\mathbf{a}' := a_1 \dots a_{n-1} \in \mathcal{A}_{n-1}$ to be $\mathbf{a}$ with the last letter deleted. Given another word $\mathbf{b} := b_1 \dots b_m \in \mathcal{A}_m$, we will denote $\mathbf{a} \to \mathbf{b}$ to mean that $\bar{a_n} \neq b_1$, i.e. that $\mathbf{ab}$ is an admissible word. We will denote $\mathbf{a} \prec \mathbf{b}$ to mean that there exists a word $\mathbf{c} \in \mathcal{W}$ such that $\mathbf{b} = \mathbf{ac}$ i.e. that $\mathbf{a}$ precedes the word $\mathbf{b}$. We will iteratively define the construction intervals $I_{\mathbf{a}} := \gamma_{\mathbf{a}'}(I_{a_n})$. The support of the measures that we will consider is the limit set

$$\Lambda := \bigcap_{n \in \mathbb{N}} \bigsqcup_{\mathbf{a} \in \mathcal{W}} I_{\mathbf{a}}$$

and we call a finite set of words $Z \subset \mathcal{W}$ a partition if the corresponding construction intervals are mutually disjoint and cover the limit set.

Given a transformation $\gamma \in SL(2, \mathbb{R})$ with $\gamma(I) = J$ for $I, J \subset \partial\mathbb{H}$, we define the distortion of γ on $I := [x_0, x_1]$ to be

$$\alpha(\gamma, I) := \log \frac{\gamma^{-1}(\infty) - x_1}{\gamma^{-1}(\infty) - x_0} \in \mathbb{R}.$$

If $\gamma^{-1}(\infty) = \infty$, γ is a translation, so we will say that intervals are not distorted under γ and define $\alpha(\gamma, I) = 0$ for all intervals I. Define $\gamma_I, \gamma_J \in SL(2, \mathbb{R})$ to be the transformations which map the unit interval to I and J respectfully. We then get that if we define $\gamma_\alpha \in SL(2, \mathbb{R})$ such that

$$\gamma_\alpha(x) := \frac{e^{\alpha/2} x}{(e^{\alpha/2} - e^{-\alpha/2})x + e^{-\alpha/2}},$$

then we have that $\gamma = \gamma_J \gamma_{\alpha(\gamma, I)} \gamma_I^{-1}$, [8]. We will use the following lemma to bound distortion [8, Lemma 2.4].

Lemma 6.1.3. *Let* $\mathbf{a} \in \mathcal{W}$ *and* $b \in \mathcal{A}$ *such that* $\mathbf{a} \to b$. *Then the distortion of* $\gamma_{\mathbf{a}}$ *on* I_b *is bounded, i.e.* $|\alpha(\gamma_{\mathbf{a}}, I_b)| \leq C_\Gamma$ *for some constant* $C_\Gamma > 0$ *depending only on the Schottky structure.*

The following lemma slightly adapts Lemma 2.5 of Bourgain–Dyatlov [8] by redefining $\mathbf{a}$ as $\mathbf{a}b$ where $\mathbf{a} \to b$ and using the parent-child ratio (Lemma 2.6 [8]).

Lemma 6.1.4. *For* $\mathbf{a} \in \mathcal{A}_n$ *and all* $b \in \mathcal{A}$ *such that* $\mathbf{a} \to b$*, we have that*

$$C_\Gamma^{-1}|I_\mathbf{a}| \leq \gamma'_\mathbf{a}(x) \leq C_\Gamma|I_\mathbf{a}|$$

and for all $x, y \in I_b$*,*

$$\frac{\gamma'_\mathbf{a}(x)}{\gamma'_\mathbf{a}(y)} \leq \exp(C_\Gamma|x-y|).$$

The following lemma adapts some lemmas of Bourgain–Dyatlov in a similar manner [8, Lemmas 2.6, 2.7, and 2.8].

Lemma 6.1.5. *For* $\mathbf{a} \in \mathcal{W}$*,* $b \in \mathcal{A}$ *such that* $\mathbf{a} \to b$*, and* $\mathbf{b} \in \mathcal{W}$ *such that* $\mathbf{a} \to \mathbf{b}$ *we have that*

$$C_\Gamma^{-1}|I_\mathbf{a}| \leq |I_{\mathbf{a}b}| \leq |I_\mathbf{a}| \tag{6.1}$$

$$C_\Gamma^{-1}|I_\mathbf{a}| \cdot |I_\mathbf{b}| \leq |I_{\mathbf{ab}}| \leq C_\Gamma|I_\mathbf{a}| \cdot |I_\mathbf{b}| \tag{6.2}$$

$$C_\Gamma^{-1}|I_\mathbf{a}| \leq |I_{\bar{\mathbf{a}}}| \leq C_\Gamma|I_\mathbf{a}| \tag{6.3}$$

We call these properties of construction intervals the Parent-Child ratio, the concatenation property, and the inversion property respectfully.

6.2 Large Deviations for the Bowen–Series Map

The Schottky structure gives us a nice way to code the limit set Λ_Γ symbolically. If we define first generation construction intervals $I_j := D_j \cap \partial\mathbb{H}$, let

$$X := \cup_{j\in\mathcal{A}} I_j$$

then we define the Bowen–Series map $T : X \to \mathbb{R}$ by

$$T(x) := \gamma_j^{-1}x \text{ for } x \in I_j$$

This is a self-map on the limit set. So you can think of T removing the first symbol of some coded point. To be able to apply large deviation theory, we need T to be (eventually) expanding. The following proposition is used from Borthwick's book [4].

Proposition 6.2.1. *There exists a finite set of intervals* $\{J_i\}$ *which cover the limit set, such that for some* $n \in \mathbb{N}$*, there exists a* $\beta > 1$ *such that for all* $x \in \cup J_i$*,*

$$|(T^n)'x| > \beta.$$

So the map T is eventually expanding, which will be sufficient for large deviations. We require large deviations to be able to do various analysis, most notably reducing the Fourier transform to exponential sums. The large deviation theorem of Jordan–Sahlsten can be used almost directly. The only big difference we must consider is that the Bowen–Series map is eventually expanding rather than expanding, and the dynamics will be considered in a limit set which is a subset of some large interval, rather than the limit set (badly approximables) as a subset of the unit interval. On the other hand, the measure on the limit set of a Fuchsian group case is made easier because $-\log|T'|$ is a bounded potential (not on X, but it is bounded on the construction intervals for words of length bigger than one) as opposed to the Gauss map analogue, and we also only have a finite alphabet. We will be able to prove that T is locally Hölder by using the fact that the construction intervals are contracting [8, (2.4)].

Proposition 6.2.2. *The potential* $-\log|T'|$ *is locally Hölder, that is there exists a* $C > 0$ *and a* $\delta < 1$ *such that for any* $n \geq 2$ *and any word* $\mathbf{a} \in \mathcal{A}_n$ *we have that for any* $x, y \in I_{\mathbf{a}}$*,*

$$|-\log|T'(x)| + \log|T'(y)|| \leq C\delta^n.$$

Proof. For $z \in I_{\mathbf{a}}$ we have that $T(z) := \gamma_{a_1}^{-1}(z)$. We note that $\gamma_{a_1}(\infty) \notin I_{\mathbf{a}}$ because this otherwise would mean that $\infty \in I_{a_2}$, so T' and T'' are indeed bounded (from above and below). We use the mean value theorem to see that for some $\xi \in I_{\mathbf{a}}$ and some constant $C_\Gamma > 1$ depending only on the Schottky structure for Γ,

$$|\log|T'(x)| - \log|T'(y)|| = \frac{|T''(\xi)|}{|T'(\xi)|}|x-y| < C_\Gamma|x-y|.$$

We know that $|x - y| \leq |I_{\mathbf{a}}|$, so we can conclude using the contractive property of the Schottky structure proved in Bourgain–Dyatlov [8, (2.5) page 6], where $\delta := (1 - C_\Gamma^{-1})$. □

Recall the main large deviation theorem given earlier (Theorem 1.2.3) of Jordan–Sahlsten [26]. This theorem is proven in the context of Gibbs measures for the Gauss map, but can be generalised to any Gibbs measure for some Markov map. The change of domain does not need to be considered in the proof, only the change in symbolic coding. In fact, in the Fuchsian group setting, the finite alphabet will make some technical steps of their proof unnecessary. The assumption on finite pressure is straight forward to prove for the bounded potential and finite alphabet case, because this means that the pressure is always finite for any t. As a result, we can succinctly obtain the large deviation theorem in this setting. This allows us to use large deviation theory in the same way that we have done in the Gauss map and totally-nonlinear map cases.

6.3 Measure Inverting

To continue with the proof of showing that the Bowen–Series map has non-concentrated derivative, we require a strong condition on the measure.

Definition 6.3.1. *We will say that a measure μ is preserved under inversion (or that that μ preserves inverting) if there exists some $C_I > 0$ such that for any $\mathbf{a} \in \mathcal{W}$,*

$$C_I^{-1}\mu(I_{\mathbf{a}}) \leq \mu(I_{\bar{\mathbf{a}}}) \leq C_I\mu(I_{\mathbf{a}}).$$

Such a condition is used because we will meet the situation where we want to study words $\mathbf{a} \in \mathcal{W}$ whose inverse is regular i.e. $\bar{\mathbf{a}} \in \mathcal{R}_n(\varepsilon)$. This condition holds for Patterson–Sullivan measures by Lemma 6.1.5 (6.3). We can also find more measures with inverting by looking at Markov measures. Consider a $(2r) \times (2r)$ non-negative aperiodic (P^n is positive for some n) row-Stochastic (for any row, the sum over entries in that row is one) transition matrix $P = (p_{i,j})$. We consider the following special case of the Perron–Frobenius theorem for the case of row-Stochastic matrices [68, Theorem 8.3].

Theorem 6.3.2 (Perron–Frobenius)**.** *Let P be a non-negative aperiodic row-Stochastic transition matrix. Then we have that*
(1) The column vector of one's is the unique right eigenvector for P;
(2) There exists a unique left eigenvector $\pi = (\pi_1, \ldots, \pi_n)$ such that $\pi P = \pi$ and π is normalised (the sum of its entries is one).

We will consider matrices P that are compatible with the Bowen–Series map, that is $p_{i,\bar{i}} = 0$ for each $i \in \mathcal{A}$. Note that $\bar{i} = i + r \bmod 2r$ by definition of the Schottky structure. This means we only assign mass to admissible words (where a symbol $a \in \mathcal{A}$ is never followed by its inverse $\bar{a}$). For $\mathbf{a} \in \mathcal{W}_n$, we can then define a measure on construction intervals by

$$\mu(I_{\mathbf{a}}) = \pi_{a_1} p_{a_1,a_2} p_{a_2,a_3} \dots p_{a_{n-1},a_n}.$$

Now we ask how we can force that such a measure preserves inverting. It will be sufficient (although maybe not necessary) to assume that $p_{i,j} = p_{\bar{j},\bar{i}}$ for $(i,j) \in \mathcal{A}^2$, noting that i and j are swapped. Hence we will have measure inverting with $C_I = \max_j \pi_j^{-1}$. We also have that this is a Gibbs measure with potential $\varphi(x) = \log p_{a_1(x),a_1(Tx)}$ and Gibbs constant C_I. The measure inverting condition will be equivalent to saying that there exists $r \times r$ matrices A and B such that

$$P = \begin{pmatrix} A & B \\ B^T & A^T \end{pmatrix}$$

because the condition is the same as requiring that $p_{i,j} = p_{j+r \bmod 2r, i+r \bmod 2r}$ (when considering entries in the matrix, you can think of it as taking a transpose, then a vertical and horizontal translation by r). We still require that P is stochastic. We can force this by restricting to the case when A and B are symmetric. Then the sum of row j will equal the sum of row $j + r \bmod 2r$, so we only need check the first r rows. To make sure P is compatible with the Bowen–Series map, we make sure that the diagonal entries of B are zero. We can also restrict to make sure all other entries of P are positive so that P is aperiodic (with $n = 2$). The main consequence of all these restrictions is that P is symmetric, so the unique left eigenvector of P is the column vector of ones (by Perron–Frobenius), so we must have that $\mu(I_a) = \pi_a = 1/2r$ for all $a \in \mathcal{A}$. For specific examples, consider the case when $r = 2$. Then if we apply the aforementioned restrictions, we have that

$$P = \begin{pmatrix} b & c & 0 & a \\ c & b' & a & 0 \\ 0 & a & b & c \\ a & 0 & c & b' \end{pmatrix}$$

where $a, b, c, b' > 0$. Note we must have that $b = b'$ in this situation for the sum over rows to be equal to one. So by choosing some $a, b, c > 0$, we know there are Markov measures which preserve inverting in the case of $r = 2$. This method can be followed to construct measures for larger r.

6.4 Non-Concentrated Derivative

There are two counting lemmas that we will need. The first is summing over regular words such that their construction interval intersects an interval J. We can deal with this as in the proof of Lemma 3.2.3.

The second counting lemma is where we may need a measure inverting assumption. We require to count words $\mathbf{a}$ (not necessarily regular) whose inverse is regular, and the inverse construction interval intersects some interval J. We can proceed with the proof in the same way as the first counting lemma, but the idea is that we can say that the word $\mathbf{a}$ has the properties of a regular word by using the quasi-Bernoulli property and measure inverting to use the properties of the regular word $\bar{\mathbf{a}}$. The proof then follows as in the first lemma, but it actually works slightly better. The first lemma annoyingly requires us to consider regularity all the way to some $\varepsilon_2 n$ level (for some $\varepsilon_2 > 0$ small), so we have to make sure that ε_r is small enough so that this works.

Lemma 6.4.1. *Consider an interval J with $2C_{\varepsilon,n}^{-1}e^{-\lambda n/4} \leq |J| \leq 2C_{\varepsilon,n}^{-1}e^{-\varepsilon_r \lambda n}$. Define*

$$D_2(\mathbf{a},\mathbf{b},\mathbf{c})' := \#\{\mathbf{d} \in \mathcal{R}_n : I_{\mathbf{d}} \cap J \neq \varnothing\}.$$

We have that for some $C > 0$,

$$\#D_2(\mathbf{a},\mathbf{b},\mathbf{c})' \leq Ce^{\lambda s n}|J|^{s/2}.$$

Remark 6.4.2 (Idea of the proof)**.** *The idea of the proof of Lemma 6.4.1 is quite basic. We have an interval J, and we want to count the n-regular intervals $I_{\mathbf{d}}$ inside it. We choose j such that $|J| \approx e^{-\lambda j}$ so that we can instead count the intervals $I_{\mathbf{d}|_j}$ which are approximately the same length as J by regularity of $\mathbf{d} \in \mathcal{R}_n$ at the j-th level (see Figure 6.2). This will give us a small (only growing with order $C_{\varepsilon,n}$) collection of j-parent intervals $I_{\mathbf{d}|_j}$ which contain the intervals $I_{\mathbf{d}}$ that we want to count. We can then count the number of n-regular words $\mathbf{d}$ in each j-parent interval $\mathbf{d}|_j$ using the regularity of $\mathbf{d}$ again to get the result.*

Proof. To begin, choose j such that

$$\frac{1}{48}e^{-\lambda}e^{-\lambda j - \frac{3}{2}\varepsilon n} \leq |J| \leq \frac{1}{48}e^{-\lambda j - \frac{3}{2}\varepsilon n}$$

noting that there is at least one such choice for j in $\{\lfloor \varepsilon_r n \rfloor, \dots, n\}$.

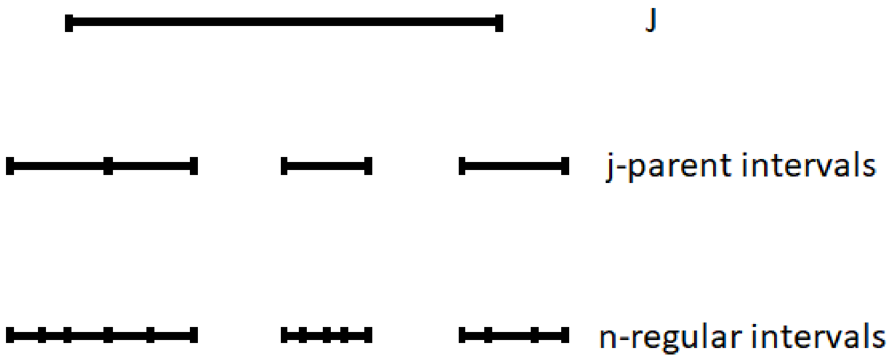

Figure 6.2: The interval J, it's regular j-parent covering, and the n-regular intervals we want to count.

It is important to note that there might exist at most two intervals $I_{\mathbf{d}}$ which are not contained in J, but do intersect J. If we were to cover J with j-parents of n-regular intervals whose length are at least $e^{(-\lambda-\varepsilon)j}/4$, then in the 'worst case' (when J does not contain any irregular geometric points), then the number K of j-parent covering sets would satisfy the last inequality in the following

$$|J| \leq e^{-\lambda j - 3\varepsilon n/2} \leq e^{-\lambda j} \leq \frac{K}{4} e^{-\lambda j - \varepsilon j}.$$

So we can sufficiently choose a $K \geq 4e^{\varepsilon\lambda n} \geq 4e^{\varepsilon j}$, for example $K = \lceil 4e^{\varepsilon\lambda n} \rceil \leq 8C_{\varepsilon,n}^{\lambda}$ where the inequality is true since $e^{\varepsilon\lambda n} \geq 1$.

Given a j-parent $I_{\mathbf{d}|_j}$ in the cover, we now approximate the number of n-regular intervals $I_{\mathbf{d}}$ (corresponding to the number of regular words which we wanted originally) contained in this set. We see that by Lemma 2.2.3

$$\#\{\mathbf{d} \in \mathcal{R}_n : \mathbf{d}_j \prec \mathbf{d}\} C^{-1} e^{-\lambda s n} C_{\varepsilon,n}^{-3\lambda} \leq \mu\Big(\bigcup_{\mathbf{d} \in \mathcal{R}_n : \mathbf{d}_j \prec \mathbf{d}} I_{\mathbf{d}}\Big) \leq \mu(I_{\mathbf{d}_j}) \leq C e^{(-s\lambda + 3\lambda\varepsilon)j}$$

so we get that

$$\#\{\mathbf{d} \in \mathcal{R}_n : \mathbf{d}_j \prec \mathbf{d}\} \leq C_{\varepsilon,n}^{3\lambda} C^2 e^{\lambda s n} e^{(-s\lambda + 3\lambda\varepsilon)j}.$$

So to conclude, we get that

$$\#D_2(\mathbf{a}, \mathbf{b}, \mathbf{c})' \leq K C_{\varepsilon,n}^{3\lambda} C^2 e^{\lambda s n} e^{(-s\lambda + 3\lambda\varepsilon)j} \leq 48 C^2 e^{\lambda} C_{\varepsilon,n}^{8\lambda} e^{\lambda s n} |J|^s \leq 48 e^{\lambda} C^2 e^{\lambda s n} |J|^{s/2}$$

(using the fact that $s \leq 1$, $j \leq n$, and ε is much smaller than ε_r) which is as we require. $\square$

The following lemma will allow is to count inverted regular words.

Lemma 6.4.3. *Consider an interval J with $2e^{-\lambda n/4} \leq |J| \leq 2e^{-\varepsilon_r n}$. We have that*

$$\#\{\mathbf{c} \in \mathcal{A}_n : \bar{\mathbf{c}} \in \mathcal{R}_n,\ I_{\mathbf{c}} \cap J \neq \varnothing\} \leq e^{\lambda s n}|J|^{s/2}$$

noting that we get exponent $s/2$ unlike Bourgain–Dyatlov [8] to get rid of constants and order $C_{\varepsilon,n}$ growing terms.

Remark 6.4.4 (Idea of the proof)**.** *We use the same idea as the proof of Lemma 6.4.1. The only problem here is that the words $\mathbf{c}$ which we count are not necessarily regular. We force regularity properties by assuming that the measure is invertible (see Definition 6.3.1). Also note that this proof only requires regularity up to $n/2$ as used by Jordan–Sahlsten [26].*

Proof. To begin, choose j such that

$$\frac{1}{48}e^{-\lambda}e^{-\lambda j-\frac{3}{2}\varepsilon n} \leq |J| \leq \frac{1}{48}e^{-\lambda j-\frac{3}{2}\varepsilon n}$$

noting that there is at least one such choice for j in $\{0,\dots,\lceil n/2\rceil\}$ since $\varepsilon_r < 1/2$.

It is important to note that there might exist at most two intervals $I_{\mathbf{c}}$ which are not contained in J, but do intersect J.

We first need some facts about the j-parents of words $\mathbf{c}$ which we are counting. By concatenation (6.2), we have that

$$|I_{\mathbf{c}|_j}| \geq C_C^{-1}\frac{|I_{\mathbf{c}}|}{|I_{\sigma^j\mathbf{c}}|}.$$

By (6.3) we get that since $\bar{\mathbf{c}}$ is regular

$$|I_{\mathbf{c}}| \geq C_I^{-1}|I_{\bar{\mathbf{c}}}| \geq \frac{C_I^{-1}}{16}C_{\varepsilon,n}^{-1}e^{-\lambda n}.$$

Then using the fact that $\sigma^j\mathbf{c}$ is the inverse of the word $\bar{\mathbf{c}}|_{n-j}$, we get by (6.3)

$$|I_{\sigma^j\mathbf{c}}| \leq C_I|I_{\bar{\mathbf{c}}|_{n-j}}| \leq C_I C_{\varepsilon,n}e^{-\lambda(n-j)}.$$

Hence we can say that

$$|I_{\mathbf{c}|_j}| \geq \frac{1}{16}C_C^{-1}C_I^{-2}C_{\varepsilon,n}^{-2}e^{-\lambda j}$$

If we were to cover J with intervals of j-parents for words $\mathbf{c}$, then in the 'worst case', the number K of j-parent covering sets would satisfy the last inequality in the following

$$|J| \leq e^{-\lambda j-3\varepsilon n/2} \leq e^{-\lambda j} \leq \frac{K}{16}C_C^{-1}C_I^{-2}C_{\varepsilon,n}^{-2}e^{-\lambda j}.$$

So we can sufficiently choose a $K \geq 32e^{2\varepsilon\lambda n}$, for example $K = \lceil 32e^{2\varepsilon\lambda n} \rceil \leq 32C_{\varepsilon,n}^{2\lambda}$ where the inequality is true if we assume that n is large enough so that $e^{\varepsilon\lambda n/3} \geq \max\{C_C, C_I\}$.

We now use a similar argument but for measure rather than length. We proceed with the proof by first noting that

$$CC_I C_{\varepsilon,n}^{3\lambda} e^{-s\lambda n} \geq C_I \mu(I_{\bar{\mathbf{c}}}) \geq \mu(I_{\mathbf{c}}) \geq C_I^{-1} \mu(I_{\bar{\mathbf{c}}}) \geq C^{-1} C_I^{-1} C_{\varepsilon,n}^{-3\lambda} e^{-s\lambda n}$$

by regularity of $\mathbf{c}$ inverse. Also, by the Bernoulli property,

$$C_B^{-1} \mu(I_{\mathbf{c}|_j}) \mu(I_{\sigma^j \mathbf{c}}) \leq \mu(I_{\mathbf{c}}).$$

As before, we use the fact that the inverse of the word $\bar{\mathbf{c}}|_{n-j}$ is the same as $\sigma^j \mathbf{c}$, hence we get that

$$\mu(I_{\sigma^j \mathbf{c}}) \geq C_I^{-1} \mu(I_{\bar{\mathbf{c}}|_{n-j}}) \geq C_I^{-1} C^{-1} C_{\varepsilon,(n-j)}^{-3\lambda} e^{-s\lambda(n-j)}$$

by multiregularity of $\mathbf{c}$ inverse. So we can say that

$$\mu(I_{\mathbf{c}|_j}) \leq C_B C_I^2 C^2 C_{\varepsilon,n}^{6\lambda} e^{-s\lambda j}.$$

Given a j-parent $I_{\mathbf{c}_j}$ in the cover, we now approximate the number of n-regular intervals $I_{\mathbf{c}}$ (corresponding to the number of inverted regular words which we wanted originally) contained in this set. We see that by Lemma 2.2.3

$$\#\{\mathbf{c} : \bar{\mathbf{c}} \in \mathcal{R}_n, \mathbf{c}_j \prec c\} C^{-1} e^{-\lambda sn} C_{\varepsilon,n}^{-3\lambda} \leq \mu\Big(\bigcup_{\mathbf{c}:\mathbf{c}_j \prec \mathbf{c}} I_{\mathbf{c}} \Big) \leq \mu(I_{\mathbf{c}_j}) \leq C_B C_I^2 C^2 C_{\varepsilon,n}^{6\lambda} e^{-s\lambda j}$$

so we get that

$$\#\{\mathbf{c} : \bar{\mathbf{c}} \in \mathcal{R}_n, \mathbf{c}_j \prec \mathbf{c}\} \leq C_B C_I^2 C^3 C_{\varepsilon,n}^{9\lambda} e^{s\lambda(n-j)}$$

So to conclude, we get that

$$\# D_2(\mathbf{a}, \mathbf{b}, \mathbf{c})' \leq (K+2) C_B C_I^2 C^3 C_{\varepsilon,n}^{9\lambda} e^{s\lambda(n-j)} \leq e^{\lambda sn} |J|^{s/2}$$

(where we use the fact that n is sufficiently large, and ε is small with respect to s) which is as we require. □

We will use the following application of the mean value theorem to obtain the interval J [8].

Lemma 6.4.5. *Consider an interval L and suppose that γ_1 and γ_2 are two Möbius transformations such that $\gamma_j(I) = J_j$ for intervals I, J_j for $j = 1, 2$. Then we have that the set of points $x \in I$ such that*

$$\log \frac{\gamma_1'(x)}{\gamma_2'(x)} \in L$$

is contained in some interval of length smaller than

$$\frac{C \cdot |I| \cdot |L|}{|\alpha(\gamma_1, I) - \alpha(\gamma_2, I)|}$$

for some constant $C > 0$.

Proof. The proof is given in Bourgain–Dyatlov, Lemma 2.3 [8]. The only difference is that we use Lemma 6.1.3 to get the constant $C > 0$. □

Finally, we can prove the non-concentration assumption as required to show Fourier decay.

Lemma 6.4.6. *Consider $\mathbf{a} \in \mathcal{R}_n(\varepsilon)$. For each σ with $e^{-\lambda n} \le \sigma \le C_{\varepsilon,n}^{-1} e^{-\varepsilon_r \lambda n}$ we have that*

$$\#\{(\mathbf{b}, \mathbf{c}, \mathbf{d}) \in \mathcal{R}_n(\varepsilon)^3 : \mathbf{a} \to \mathbf{b} \to \mathbf{d}, \mathbf{a} \to \mathbf{c} \to \mathbf{d}, |\gamma_{\mathbf{ab}}'(x_{\mathbf{d}}) - \gamma_{\mathbf{ac}}'(x_{\mathbf{d}})| \le e^{-2\lambda n}\sigma\}$$

can be bounded from above by $e^{3\lambda s n}\sigma^{s/4}$.

We follow the proof of Bourgain–Dyatlov [8, Lemma 2.16]. The main difference once again is $C_{\varepsilon,n}$ error terms, which in the end will just result in a slightly weaker exponent of σ for the upper bound (namely $s/4$ rather than Bourgain–Dyatlov who get $s/2$). The relationship between distortion factors and the preimage of infinity is explained in detail.

Proof. Consider fixed $\mathbf{b} \in \mathcal{R}_n$. We first consider counting the set

$$B := \{\mathbf{c} \in \mathcal{R}_n : \mathbf{a} \to \mathbf{c}, |\gamma_{\mathbf{ab}}^{-1}(\infty) - \gamma_{\mathbf{ac}}^{-1}(\infty)| \le \sqrt{\sigma}\}.$$

By definition of the Schottky structure we have that $\gamma_{\mathbf{ac}}^{-1}(\infty) = \gamma_{\bar{\mathbf{c}}\bar{\mathbf{a}}}(\infty) \in I_{\bar{\mathbf{c}}}$. Therefore B is contained in the set

$$\{\mathbf{e} \in \mathcal{W}_n : \bar{\mathbf{e}} \in \mathcal{R}_n(\varepsilon), I_{\mathbf{e}} \cap J \neq \varnothing\}$$

where $J := \gamma_{\mathbf{ab}}^{-1}(\infty) + [-\sqrt{\sigma}, \sqrt{\sigma}]$. Hence we can apply Lemma 6.4.3 to bound this set. We can then bound the set of triples in B by using the cardinality bounds given in Lemma 2.2.3.

We complete the proof by considering when a triple satisfies $|\gamma_{\mathbf{ab}}^{-1}(\infty) - \gamma_{\mathbf{ac}}^{-1}(\infty)| \geq \sqrt{\sigma}$. We can count such triples by considering the set

$$D := \{\mathbf{d} \in \mathcal{R}_n(\varepsilon) : \mathbf{b} \to \mathbf{d}, \mathbf{c} \to \mathbf{d}, |\gamma'_{\mathbf{ab}}(x_{\mathbf{d}}) - \gamma'_{\mathbf{ac}}(x_{\mathbf{d}})| \leq e^{-2\lambda n}\sigma\}$$

for fixed $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$. By Lemma 2.2.3 we have that $\frac{1}{16^2}C_{\varepsilon,n}^{-2}e^{-2\lambda n} \leq \gamma'_{\mathbf{ab}}, \gamma'_{\mathbf{ac}} \leq C_{\varepsilon,n}^2 e^{-2\lambda n}$ on the set I_{d_1} which is valid by the conditions on the words. Hence by the mean value theorem applied to the logarithm we have that the set

$$D' := \left\{\mathbf{d} \in \mathcal{R}_n(\varepsilon) : \mathbf{b} \to \mathbf{d}, \mathbf{c} \to \mathbf{d}, \left|\log \frac{\gamma'_{\mathbf{ab}}(x_{\mathbf{d}})}{\gamma'_{\mathbf{ac}}(x_{\mathbf{d}})}\right| \leq C_{\varepsilon,n}^3 \sigma\right\}$$

contains D, which crudely uses $C_{\varepsilon,n} > 16^2$.

Let $I_{d_1} =: [x_0, x_1]$. We have that

$$\begin{aligned} |\alpha(\gamma_{\mathbf{ab}}, I_{d_1}) - \alpha(\gamma_{\mathbf{ac}}, I_{d_1})| &= \left|\log \frac{\gamma_{\mathbf{ab}}^{-1}(\infty) - x_1}{\gamma_{\mathbf{ab}}^{-1}(\infty) - x_0} - \log \frac{\gamma_{\mathbf{ac}}^{-1}(\infty) - x_1}{\gamma_{\mathbf{ac}}^{-1}(\infty) - x_0}\right| \\ &= \frac{1}{|M|}\left|\frac{\gamma_{\mathbf{ab}}^{-1}(\infty) - x_1}{\gamma_{\mathbf{ab}}^{-1}(\infty) - x_0} - \frac{\gamma_{\mathbf{ac}}^{-1}(\infty) - x_1}{\gamma_{\mathbf{ac}}^{-1}(\infty) - x_0}\right| \end{aligned}$$

where the last equality holds using the mean value theorem with

$$M \in \left[\frac{\gamma_{\mathbf{ab}}^{-1}(\infty) - x_1}{\gamma_{\mathbf{ab}}^{-1}(\infty) - x_0}, \frac{\gamma_{\mathbf{ac}}^{-1}(\infty) - x_1}{\gamma_{\mathbf{ac}}^{-1}(\infty) - x_0}\right].$$

So $|M|$ can be bounded from above using a constant depending only on the Schottky data because $\gamma_{\mathbf{ac}}^{-1}(\infty) \in I_{\bar{c_n}}$ and $\gamma_{\mathbf{ab}}^{-1}(\infty) \in I_{\bar{b_n}}$. So using the fact that $I_{d_1} = [x_0, x_1]$ we have that for some constant $C_\Gamma > 0$ depending on the Schottky data,

$$\begin{aligned} |\alpha(\gamma_{\mathbf{ab}}, I_{d_1}) - \alpha(\gamma_{\mathbf{ac}}, I_{d_1})| &= \frac{1}{|M|}\left|\frac{\gamma_{\mathbf{ab}}^{-1}(\infty)|I_{d_1}| - \gamma_{\mathbf{ac}}^{-1}(\infty)|I_{d_1}|}{(\gamma_{\mathbf{ab}}^{-1}(\infty) - x_0)(\gamma_{\mathbf{ac}}^{-1}(\infty) - x_0)}\right| \\ &\geq \frac{1}{C_\Gamma}|\gamma_{\mathbf{ab}}^{-1}(\infty) - \gamma_{\mathbf{ac}}^{-1}(\infty)|. \end{aligned}$$

Hence by Lemma 6.4.5 we know that there is an interval J' of length $C_{\varepsilon,n}^3\sqrt{\sigma}$ such that the points $x_{\mathbf{d}} \in D'$ intersect J'. By using Lemma 6.4.1, we get the required bound. □

So we have proven the non-concentration condition when we have 'invertible' measures. We hence get polynomial Fourier decay for these measures by using the same proof strategy as in the Gauss map case, where we also use the frequency parameter $\tau = e^{-\lambda n}$. See Subsection 4.3.1 for details.

6.5 Fractal Uncertainty Principle

Using Theorem 5.2.3, we can prove polynomial Fourier decay for Gibbs measures on limit sets of convex cocompact Fuchsian groups by the following theorem of Naud.

Lemma 6.5.1 (Naud, 2005 [44]). *For a convex cocompact Fuchsian group* Γ*, the function* $\log |T'|$ *is totally nonlinear on the limit set.*

We hence get a generalisation of Bourgain–Dyatlov [8] for Gibbs measures of large dimension, but we only get polynomial Fourier decay for very large frequencies. The following theorem proves a fractal uncertainty principle when we only get polynomial decay of the Fourier transform for exponentially large frequencies, unlike Bourgain–Dyatlov when their decay theorem works for frequencies with size bigger than one.

Proposition 6.5.2. *For* $\Phi \in C^3([0,1]^2, \mathbb{R})$ *and* $G \in C^1([0,1], \mathbb{C})$*, assume that*

$$||\Phi||_{C^3} + ||G||_{C^1} \le C_{\Phi,G} \text{ and } \inf |\partial^2_{xy}\Phi| \ge C^{-1}_{\Phi,G}.$$

We will assume further that there exists some $b > 0$ *and exponentially decreasing* $\tau(n) > 0$ *such that for all sufficiently large* n *we have that*

$$\mu \times \mu(\{(x,y) \in X^2 : |x-y| \le C_{\varepsilon,n}\tau(n)^{1/4}\}) \le \tau(n)^b$$

which is the same as the assumption in the main Exponential Sum Lemma 3.3.1. Then there exists some $n_0' > 1$ *such that for* $h \in (0, e^{-n_0'})$*, if we define* $B(h) : L^2([0,1],\mu) \to L^2([0,1],\mu)$ *by*

$$B(h)u(x) = \int e^{i\Phi(x,y)/h} G(x,y)u(y)\, d\mu(y).$$

Then there exists some $\varepsilon_F > 0$ *and some* $C = C(T,\Phi,G) > 0$ *such that*

$$||B(h)||_{L^2([0,1],\mu)} \le Ch^{\varepsilon_F}.$$

The following proof modifies the proof given by Bourgain–Dyatlov so that it works in the setting of Large Deviations and eventual (for large frequencies) polynomial Fourier decay. The following is a proof that is more relevant for the Gauss map case, but can easily be modified to the Bowen–Series setting. The main difference is that X is not a single interval in the Fuchsian case, so a partition of unity is necessary as presented in the proof of Bourgain–Dyatlov [8, Proposition 4.1]. This is also true for the totally-nonlinear maps setting, and by extension to Gibbs measures on the limit set with large enough dimension.

Proof. For the given measure μ, we get Fourier decay for $|\xi| > e^{(2k+3/2)n_0}$ for some n_0 defined when studying large deviations of the relevant dynamics. Consider some small $h > 0$. We will have to choose n large enough so that

$$\tau(n)^{-1/4} > \tau(n_0)^{-1/2} e^{(2k+1)\lambda n_0}$$

which will mean that we can use the (eventual) Fourier decay. Choose $n \geq n_0$ such that

$$\tau(n+1)^{1/2} \leq h \leq \tau(n)^{1/2} < 1.$$

It suffices to prove that

$$||B(h)B(h)^*||_{L^2([0,1],\mu)} \leq Ch^{\varepsilon_s/2}$$

where $B(h)B(h)^*$ can be shown to be given by

$$B(h)B(h)^* f(x) = \int K(x,x') f(x')\, d\mu(x')$$

where

$$K(x,x') = \int e^{i(\Phi(x,y)-\Phi(x',y))/h} G(x,y)\overline{G(x',y)}\, d\mu(y).$$

By Schur's test, to prove the required bound we can sufficiently prove that

$$\sup_{x\in[0,1]} \int |K(x,x')|\, d\mu(x') \leq Ch^{\varepsilon_s/2}.$$

For $x, x' \in [0,1]$, define $\varphi_{xx'}$ and $g_{xx'}$ such that

$$\Phi(x,y) - \Phi(x',y) = (x-x')\varphi_{xx'}(y) \ \text{ and } \ g_{xx'}(y) = G(x,y)\overline{G(x',y)}.$$

Then we have that

$$K(x,x') = \int e^{i\xi\varphi_{xx'}(y)} g_{xx'}(y)\, d\mu(y)$$

where $\xi := (x-x')/h$. By the assumptions on G and Φ we get that for some $C > 0$

$$||\varphi_{xx'}||_{C^2} + ||g_{xx'}||_{C^1} \leq C \ \text{ and } \ \inf_{[0,1]} |\partial_y \varphi_{xx'}| \geq C^{-1}.$$

By the main decay theorem, we get that

$$|K(x,x')| \leq C\left|\frac{x-x'}{h}\right|^{\varepsilon_s}$$

when considering the integral over $\{(x,x') : |x-x'| > h^{1/2}\}$. This will contribute a bound of $Ch^{\varepsilon_s/2}$.

For the integral over $\{(x,x') : |x-x'| \leq h^{1/2}\}$, which is contained in $\{(x,x') : |x-x'| < C_{\varepsilon,n}\tau(n)^{1/4}\}$, we can use the $\mu \times \mu$ assumption to bound this part of the integral by some constant times $\tau(n+1)^{b/2}$. Since $\tau(n) > 0$ is exponentially decreasing, there exists some $C > 1$ such that $\tau(n) \leq C\tau(n+1)$ so we can say that this part of the integral is bounded by $h^{b/2}$ for sufficiently small h, concluding the proof. $\square$

Chapter 7

Prospects

We find ourselves in a situation where many questions about the interplay between nonlinear dynamics and Fourier transforms have been answered in recent years, but there are still so many questions yet to be answered.

In all this work we assume that the dimension of the measure is at least non-zero. In the zero case, we have less tools at our disposal. In the infinite Lyapunov exponent case, we lose the control from large deviations, which in turn likely means using the sum-product theory of Bourgain will not be possible. This is also the case for the infinite entropy setting. Vastly different techniques will need to be employed for these cases.

Asking about higher dimensional analogues is a natural question, and one that many people are interested in. The main thing to note is that the paper of Stoyanov [63] works in higher dimensions. However, the difficulty is likely to be in the application of Stoyanov's work to prove non-concentration of derivatives. Li–Naud–Pan [36] proved a higher dimensional version of Bourgain–Dyatlov which took a great deal of work, and took full advantage of the strength of the Schottky structure of Fuchsian groups. It might be that without such a structure where the non-concentrated behaviour of derivatives is strong, it might not be possible to prove results. However, we said the same thing about extending the Gauss map decay theorems to general nonlinear maps, and that turned out to be possible. I have no doubt that if a higher dimensional analogue is possible, it will take the commitment of the best mathematicians in this field to produce a proof.

Large deviation theory has proven to be a very useful tool for solving problems in dynamical systems. It is central in being able to reduce to the sum-product theory of Bourgain in Chapter 3.1. In the Markov map setting, we are interested in proving more large deviation theorems so that we can directly obtain results on Fourier transforms of measures using the proofs of [54]. Results such as Pollicott–Sharp [49] prove that the Manneville–Pomeau map has large deviation theorems that exhibit polynomial decay in the worst case. If such theorems could be proved for totally nonlinear maps, the results in Chapter 3.1 can be used to give logarithmic Fourier decay.

So we have Fourier decay results for the case of Markov map invariance when considering Mosquera–Shmerkin [42] and Theorem 1.3.4 together. We also have a variety of theorems in the linear setting, such as those presented by Li–Sahlsten [38] [37]. In their setting, the best decay rate they can achieve is logarithmic, but in the nonlinear setting we can get polynomial decay. An important consideration is how the polynomial decay rate of nonlinear systems behaves when you make them more linear. In Solomyak's [60] work, he shows that every self-similar measure has polynomial Fourier decay outside a set of measures with dimension zero. This suggests that the decay results obtained by Li–Sahlsten are not the best possible in most cases. One question is whether proving polynomial Fourier decay for specific self-similar measures (linear dynamics) would be possible, or whether the random component of the measure is necessary to use tools like in Mosquera–Shmerkin. This is a question which a lot of people are actively researching (see Subsection 1.1.7).

Bibliography

[1] A. Algom, F. Rodriguez Hertz, and Z. Wang: Pointwise normality and Fourier decay for self-conformal measures, Preprint arXiv:2012.06529, 2020.

[2] V. Araújo and I. Melbourne: Exponential decay of correlations for non-uniformly hyperbolic flows with a $C^{1+\alpha}$ stable foliation, including the classical Lorenz attractor, *Ann. Henri Poincaré* 17 (2016) 2975-3004.

[3] A. Avila, S. Gouëzel and J.-C. Yoccoz: Exponential mixing of the Teichmüller flow, *Publ. Math. Inst. Hautes Études Sci.*, (104):143-211, 2006.

[4] D. Borthwick: Spectral theory of infinite-area hyperbolic surfaces (2nd ed), *Birkhäuser*, 2016.

[5] S. Baker and N. Jurga: Maximising Bernoulli measures and dimension gaps for countable branched systems, *Ergodic Theory Dynam. Systems*, 41(7), 1921-1939, 2018.

[6] A. Besicovitch: Sets of fractional dimension (IV): on rational approximation to real numbers, *J. London Math. Soc.*, 9:126-131, 1934.

[7] J. Bourgain: The discretized sum-product and projection theorems, *J. Anal. Math.* 112(2010), 193-236.

[8] J. Bourgain and S. Dyatlov: Fourier dimension and spectral gaps for hyperbolic surfaces, *Geom. Funct. Anal.* (GAFA) 27 (2017), 744-771.

[9] J. Bourgain and S. Dyatlov: Spectral gaps without the pressure condition, *Ann. of Math.* 187(2018), 825-867.

[10] J. Bourgain and A. Gamburd: A Spectral Gap Theorem in $SU(d)$, *JEMS* 014.5 (2012): 1455-1511.

[11] R. Bowen and C. Series: Markov maps associated with Fuchsian groups, *IHES Publications* 50, 1979, 153-170.

[12] D. Bressoud: A Radical Approach to Real Analysis, *The Mathematical Association of America*, 2007.

[13] H. Davenport, P. Erdös and W. LeVeque: On Weyls criterion for uniform distribution, *Michigan Math. J.*, 10:311-314, 1963.

[14] S. Dyatlov: An introduction to fractal uncertainty principle, *Journal of Mathematical Physics* 60(2019), 081505.

[15] S. Dyatlov and L. Jin: Semiclassical measures on hyperbolic surfaces have full support, *Acta Math.*, 220(2018), 297-339.

[16] S. Dyatlov and J. Zahl: Spectral gaps, additive energy, and a fractal uncertainty principle, *Geom. Funct. Anal.*(GAFA) 26 (2016), 1011-1094.

[17] F. Ekström: The Fourier dimension is not finitely stable, *Real Analysis Exchange* 40(2) p.397-402, 2015.

[18] F. Ekström, T. Persson and J. Schmeling: On the Fourier dimension and a modification, *Journal of Fractal Geometry*, Vol 2, Issue 3, 309-337, 2015.

[19] J. M. Fraser, T. Orponen and T. Sahlsten: On Fourier analytic properties of graphs, *Int. Math. Res. Not.* (IMRN), (2014), 2730-2745.

[20] A. Gamburd, M. Magee and R. Ronan: An asymptotic formula for integer points on Markoff-Hurwitz varieties, *Ann. of Math.* (2) 190 (2019), no. 3, 751809, MR 4024562.

[21] M. Hochman: A short proof of Hosts equidistribution theorem, preprint arXiv:2103.08938, 2021.

[22] M. Hochman and P. Shmerkin: Equidistribution from fractal measures, *Invent Math.* Vol 202, Issue 1, 427-479, 2015.

[23] L. Hörmander: Estimates for translations invariant operators in L^p spaces, *Acta Math.* 104 (1960), 93-140.

[24] S. Igarí: Functions of L^p multipliers, *Tôhoku Math.* J. 21 (1969), 304-320.

[25] V. Jarník: Diophantischen Approximationen und Hausdorffsches Mass, *Matematicheskii Sbornik*, 36(3-4):371-382, 1929.

[26] T. Jordan and T. Sahlsten: Fourier transforms of Gibbs measures for the Gauss map, *Math. Ann.* (2016) Vol 364 (3). 983-1023.

[27] J.-P. Kahane: Some Random series of functions (2nd ed.), *Cambridge Univ. Press*, 1985.

[28] J.-P. Kahane: Ensembles alatoires et dimensions, *In Recent Progress in Fourier Analysis (proceedings of a seminar held in El Escorial)*, 65-121, 1983.

[29] J.-P. Kahane and R. Salem: Distribution modulo 1 and sets of uniqueness, *Bull. Amer. Math. Soc.* 70 (1964), no. 2, 259-261.

[30] R. Kaufman: Continued fractions and Fourier transforms, *Mathematika, 27(2).* 262-267, 1980.

[31] R. Kaufman: On the theorem of Jarník and Besicovitch, *Acta Arith.*. 39(3):265-267, 1981.

[32] I. Laba and M. Pramanik: Arithmetic progressions in sets of fractional dimension, *Geom. Funct. Anal.* (GAFA), 19(2):429-456, 2009.

[33] J. Li: Decrease of Fourier coefficients for stationary measures, *Math. Ann.* Vol 372 (3-4), 2018.

[34] J. Li: Discretized Sum-product and Fourier decay in Rn, *Journal dAnalyse Math* to appear, 2018.

[35] J. Li: Fourier decay, Renewal theorem and Spectral gaps for random walks on split semisimple Lie groups, *Annales Scientifiques de l'NS* to appear, 2018.

[36] J. Li, F. Naud and W. Pan: Kleinian Schottky groups, Patterson–Sullivan measures, and Fourier decay, *Duke Math.* to appear, arXiv:1902.01103, 2019.

[37] J. Li and T. Sahlsten: Fourier transform of self-affine measures, Adv. Math. to appear (2019), arXiv:1903.09601

[38] J. Li and T. Sahlsten: Trigonometric Series and Self-similar Sets, JEMS to appear (2019), arXiv:1902.00426

[39] P. Mattila: Fourier Analysis and Hausdorff Dimension, *Cambridge Univ. Press*, 2015.

[40] P. Mattila: Geometry of Sets and Measures in Euclidean Spaces: Fractals and Rectifiability, *Cambridge Univ. Press*, 1995.

[41] H. McKean: Hausdorff-Besicovitch dimension of Brownian motion paths, *Duke Math J.*, 22, (1955), 229-234.

[42] C. Mosquera and P. Shmerkin: Self-similar measures: asymptotic bounds for the dimension and Fourier decay of smooth images, *Ann. Acad. Sci. Fenn. Math.*, 2018, Vol. 43 Issue 2, 823-834.

[43] H. Nakada: Metrical theory for a class of continued fraction transformations and their natural extensions, *Tokyo J. Math.* 4 (1981).

[44] F. Naud: Expanding maps on Cantor sets and analytic continuation of zeta functions, *Ann ENS* (4) 38(1):116-153, 2005.

[45] W. Parry and M. Pollicott: Zeta functions and the periodic orbit structure of hyperbolic dynamics, *Astérisque* 187-188, Soc. Math. France, 1990.

[46] E. Perkins: The exact Hausdorff measure of the level sets of Brownian motion, Z. Wahrscheinlichkeitstheorie verw. Gebiete, 58, (1981), 373-388.

[47] T. Persson: On a problem by R. Salem concerning Minkowski's question mark function, preprint arXiv:1501.00876, 2015.

[48] H.R. Pitt and N. Wiener: On absolutely convergent Fourier–Stieltjes transforms, *Duke Math. J.* 4 (1938), 420-436.

[49] M. Pollicott and R. Sharp: Large deviations for intermittent maps, preprint `http://homepages.warwick.ac.uk/~masdbl/indifference.pdf`

[50] M. Queffélec and O. Ramaré: Analyse de Fourier des fractions continues á quotients restreints, *Enseign. Math.*(2), 49(3-4):335-356, 2003.

[51] M. Rams: Theory of fractals (lecture notes), `http://ssdnm.mimuw.edu.pl/pliki/wyklady/RamsFraktale.pdf`, 2013.

[52] A. Rapaport: On the Rajchman property for self-similar measures on Rd, preprint arXiv:2104.03955, 2021.

[53] T. Sahlsten and C. Stevens: Fourier decay in nonlinear dynamics, preprint arXiv:1810.01378, 2018.

[54] T. Sahlsten and C. Stevens: Fourier transform and expanding maps on cantor sets, preprint arXiv:2009.01703, 2020.

[55] R. Salem: On some singular monotonic functions which are strictly increasing, *Trans. Amer. Math. Soc.* 53:427-439, 1943.

[56] R. Salem: Sets of uniqueness and sets of multiplicity, *Trans AMS*, 54:218-228, 1943, corrected on pp.595-598, vol 63, 1948.

[57] O. Sarig: Existence of Gibbs measures for countable Markov shifts, *Proc. Amer. Math. Soc., 131(6).* 1751-1758, 2003.

[58] P. Sarnak: Spectra of singular measures as multipliers on L^p, *J. Funct. Anal.* 37 (1980), 302-317.

[59] P. Shmerkin and V. Suomala: Spatially independent martingales, intersections, and applications, *Mem. Amer. Math. Soc.* 251 (2018), no. 1195

[60] B. Solomyak: Fourier decay for self-similar measures, preprint (2019), arXiv:1906.12164

[61] E. Stein: Harmonic analysis on $\mathbb{R}^n$, in Studies in Harmonic Analysis. (J. M. Ash, Ed.), pp. 97-135, Studies in Mathematics, Vol. 13, Mathematical Association of America, 1976.

[62] E. Stein and R. Shakarchi: Fourier Analysis: An introduction, *Princeton University Press*, 2003.

[63] L. Stoyanov: Spectra of Ruelle transfer operators for Axiom A flows, *Nonlinearity* (2011), Vol. 24, Number 4.

[64] S. J. Taylor: The Hausdorff α-dimensional measure of Brownian paths in n-space, *Math. Proc. Cambridge Philos. Soc.*, 49, (1953), 31-39.

[65] S. J. Taylor: The α-dimensional measure of the graph and set of zeros of a Brownian path, *Math. Proc. Cambridge Philos. Soc.*, 51, (1955), 265-274.

[66] C. Walkden: Ergodic Theory (lecture notes), `https://personalpages.manchester.ac.uk/staff/charles.walkden/ergodic-theory/ergodic_theory.pdf`, 2018.

[67] C. Walkden: Hyperbolic Geometry (lecture notes), `https://personalpages.manchester.ac.uk/staff/charles.walkden/hyperbolic-geometry/hyperbolic_geometry_1920.pdf`, 2019.

[68] C. Walkden: Thermodynamical Formalism (lecture notes for a course on Ergodic Theory), `https://personalpages.manchester.ac.uk/staff/charles.walkden/magic/lecture09.pdf`

[69] A. Zygmund: Trigonometric Series, Two vols, *Cambridge Univ. Press*, 1959.