



Jordan, T. M., & Sahlsten, T. V. A. (2016). Fourier transforms of Gibbs measures for the Gauss map. *Mathematische Annalen*, 364(3), 983-1023. <https://doi.org/10.1007/s00208-015-1241-9>

Peer reviewed version

Link to published version (if available):  
[10.1007/s00208-015-1241-9](https://doi.org/10.1007/s00208-015-1241-9)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

"The final publication is available at Springer via [http://dx.doi.org/\[DOI 10.1007/s00208-015-1241-9\]](http://dx.doi.org/[DOI 10.1007/s00208-015-1241-9])

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/pure/about/ebr-terms>

# FOURIER TRANSFORMS OF GIBBS MEASURES FOR THE GAUSS MAP

THOMAS JORDAN AND TUOMAS SAHLSTEN\*

ABSTRACT. We investigate under which conditions a given invariant measure  $\mu$  for the dynamical system defined by the Gauss map  $x \mapsto 1/x \bmod 1$  is a Rajchman measure with polynomially decaying Fourier transform

$$|\widehat{\mu}(\xi)| = O(|\xi|^{-\eta}), \quad \text{as } |\xi| \rightarrow \infty.$$

We show that this property holds for any Gibbs measure  $\mu$  of Hausdorff dimension greater than  $1/2$  with a natural large deviation assumption on the Gibbs potential. In particular, we obtain the result for the Hausdorff measure and all Gibbs measures of dimension greater than  $1/2$  on badly approximable numbers, which extends the constructions of Kaufman and Queffélec-Ramaré. Our main result implies that the Fourier-Stieltjes coefficients of the Minkowski's question mark function decay to 0 polynomially answering a question of Salem from 1943. As an application of the Davenport-Erdős-LeVeque criterion we obtain an equidistribution theorem for Gibbs measures, which extends in part a recent result by Hochman-Shmerkin. Our proofs are based on exploiting the nonlinear and number theoretic nature of the Gauss map and large deviation theory for Hausdorff dimension and Lyapunov exponents.

## 1. INTRODUCTION AND THE MAIN RESULTS

1.1. **Rajchman measures.** Given a Borel probability measure  $\mu$  on the unit circle  $\mathbb{T}$ , we define the *Fourier transform* of  $\mu$  at the frequency  $\xi \in \mathbb{R}$  by the quantity

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

It is a central question in Fourier analysis and its applications to understand how the behaviour or decay of  $\widehat{\mu}$  affect for example the absolute continuity, singularity, geometric or arithmetic structure of the measure  $\mu$ . If  $\widehat{\mu} \rightarrow 0$  at infinity, then  $\mu$  is called a *Rajchman measure*. Rajchman measures have their root and motivation in the study of sets of uniqueness and multiplicity of Fourier series; see for example [14] and the survey of Lyons [21] for a more detailed historical background.

When we know the *rate* of decay for the Fourier transform of a Rajchman measure, we can recover a lot of information from the structure of the measure. For example, a classical bound coming from potential theoretic methods (see for example [22] and references therein) shows that if the Fourier transform  $\widehat{\mu} \rightarrow 0$  *polynomially*, that is,  $|\widehat{\mu}(\xi)| = O(|\xi|^{-\eta})$  as  $|\xi| \rightarrow \infty$  for some  $\eta > 0$ , then the Hausdorff dimension of the measure  $\mu$  satisfies  $\dim \mu \geq 2\eta$ . This gives rise to the notion of *Salem sets*, that is, a set  $A \subset \mathbb{R}$  with Hausdorff dimension  $s = \dim A$  that supports a Rajchman measure  $\mu$  with Fourier transform  $|\widehat{\mu}(\xi)| = O(|\xi|^{-\eta})$  as  $|\xi| \rightarrow \infty$  with  $\eta$  arbitrarily close to  $s/2$ . Thanks to Plancherel's theorem, if the Fourier transform  $\widehat{\mu} \in L^2$ , then  $\mu$  must be absolutely continuous with  $L^2$  density, and moreover, if  $\widehat{\mu} \in L^1$ , then  $\mu$  is a continuous function. With the help of the convolution formula  $\widehat{\mu * \nu} = \widehat{\mu} \cdot \widehat{\nu}$  this kind of powerful information can be linked to the structure and size of the sumsets and distance sets of the supports of Rajchman measures with rapidly enough decaying Fourier transform; for a detailed background, see for example the survey of Mattila [23].

---

2010 *Mathematics Subject Classification.* 42A38 (Primary), 11K50, 37C30, 60F10 (Secondary).

*Key words and phrases.* Fourier transform, Gibbs measures, Gauss map, Diophantine approximation, normal numbers, large deviation theory, thermodynamical formalism.

\*Corresponding author. T.S. acknowledges the partial support from University of Bristol, the Finnish Centre of Excellence in Analysis and Dynamics Research, Emil Aaltonen Foundation and the European Union (ERC grant #306494).

Rajchman measures with polynomially decaying Fourier transforms share many similar properties with Lebesgue measure. The *Erdős-Turán inequality* yields a universal bound for the measure  $\mu(I)$  of an interval  $I$  compared to its length  $|I|$  via the Fourier coefficients  $\widehat{\mu}(k)$ . Lebesgue measure also has the following characteristic property that almost every number is normal and a similar criterion can be deduced for Rajchman measures with polynomial decay as well. Recall that a given a sequence  $x_1, x_2, \dots$ , of non-negative real numbers *equidistributes modulo 1*, if for any interval  $I \subset [0, 1]$  the frequency of  $x_k$  hitting  $I$  is the length  $|I|$ . In other words, there exists

$$\lim_{N \rightarrow \infty} \frac{|\{1 \leq k \leq N : x_k \bmod 1 \in I\}|}{N} = |I|.$$

A particularly important case is the sequence  $x_k = n^k x$ ,  $k \in \mathbb{N}$ , where  $n \in \mathbb{N}$  is a fixed base. If the sequence  $(n^k x)$  equidistributes modulo 1, the number  $x$  is called *n-normal*. Moreover, a measure  $\mu$  on  $[0, 1]$  is *n-normal* if  $\mu$  almost every  $x$  is *n-normal*. Invoking the famous *Weyl's criterion* for equidistribution with exponential sums, we obtain the following

**Theorem 1.1** (Davenport-Erdős-LeVeque). *Let  $\mu$  be a probability measure on  $[0, 1]$  and  $(s_k)_{k \in \mathbb{N}}$  a sequence of natural numbers. If any  $p \neq 0$  satisfies*

$$\sum_{N=1}^{\infty} \frac{1}{N^3} \sum_{k,m=1}^N \widehat{\mu}(p(s_k - s_m)) < \infty, \quad (1.1)$$

*then the sequence  $(s_k x)_{k \in \mathbb{N}}$  equidistributes modulo 1 at  $\mu$  almost every  $x$ .*

*When  $\mu$  is a Rajchman measure with  $\widehat{\mu} \rightarrow 0$  polynomially, then (1.1) holds for every strictly increasing sequence  $(s_k)_{k \in \mathbb{N}}$ . In particular,  $\mu$  is *n-normal* for any  $n \in \mathbb{N}$ .*

This was proved by Davenport, Erdős and LeVeque in [5] and for the proof of the statement in this form, see for example [29, Theorem 7.2]. Moreover, see also [29, Corollary 7.4] for the deduction for the corollary on Rajchman measures.

Theorem 1.1 can be applied to find normal numbers in small sets which still support Rajchman measures with polynomially decaying Fourier coefficients. This approach has been particularly useful in Diophantine approximation of irrational numbers after the seminal works of Kaufman [12, 13]. A classical result of Dirichlet says that for any irrational number  $x \in [0, 1]$  we can find infinitely many rationals  $p/q$  with  $|x - p/q| \leq q^{-2}$ . A number is *well approximable* if the rate of approximation Dirichlet's theorem gives can be improved: for example, if  $\alpha \geq 2$ , the class of  *$\alpha$ -well approximable numbers*  $W(\alpha)$  consists of those irrational  $x$  with  $|x - p/q| \leq q^{-\alpha}$  for infinitely many rationals  $p/q$ . The set  $W(\alpha)$  is when  $\alpha > 2$  a fractal with Hausdorff dimension  $\dim W(\alpha) = 2/\alpha$  by a classical result of Jarník-Besicovitch. In [13] Kaufman proved that there is a Rajchman measure  $\mu$  on  $W(\alpha)$  with polynomial decay

$$|\widehat{\mu}(\xi)| = o(\log |\xi|) |\xi|^{-1/\alpha},$$

which shows that  $W(\alpha)$  is a Salem set. The papers by Bluhm [2, 3] contain a more detailed proof of Kaufman's result and demonstrate that the set of Liouville numbers (numbers in every  $W(\alpha)$ ) also supports a Rajchman measure; note that as the Liouville numbers have Hausdorff dimension 0 the decay cannot be polynomial.

A polar opposite to well approximability is the collection of *badly approximable numbers*, that is, those irrational  $x \in [0, 1]$  for which we can find a constant  $c(x) > 0$  such that  $|x - p/q| \geq c(x)/q^2$  for all rationals  $p/q$ . All such numbers are exactly those numbers for which the *continued fraction expansion*  $a_1(x), a_2(x), \dots$ , is bounded (see Section 2). Thus this gives rise to the sets

$$B_N = \{x \in [0, 1] \setminus \mathbb{Q} : a_i(x) \leq N \text{ for all } i \in \mathbb{N}\}$$

whose union is the set of all badly approximable numbers. As  $B_N$  can be interpreted as an attractor to a self-conformal iterated function system  $\{x \mapsto 1/(x + a) : a = 1, 2, \dots, N\}$  (see Section 2), it is possible to estimate the Hausdorff dimension of  $B_N$ . In [8] it was estimated

$\dim B_N \geq \dim B_2 = 0.531\dots$  for any  $N \in \mathbb{N}$  and Jarník [11] proved that  $\dim B_N \nearrow 1$  as  $N \rightarrow \infty$ . These sets support Rajchman measures:

**Theorem 1.2** (Kaufman & Queffélec-Ramaré). *Given  $N \geq 2$ , there exists a Rajchman measure  $\mu$  on  $B_N$  such that  $\widehat{\mu} \rightarrow 0$  polynomially.*

The measure constructed in this proof is now widely known as the *Kaufman measure*. Theorem 1.2 was proved by Kaufman [12] for  $N \geq 3$  and Queffélec-Ramaré [29] later extended this result for  $N \geq 2$  by modifying the proof of Kaufman and with more careful analysis of continuants. Thus this yields by Theorem 1.1 that there are infinitely many badly approximable  $n$ -normal numbers for any  $N \geq 2$  and  $n \in \mathbb{N}$ . For  $N \geq 3$  this was already settled by Baker, who pointed out this could be deduced from Kaufman's result (see for example the lecture notes by Montgomery [26]), but for  $B_2$  the Kaufman measure constructed by Queffélec-Ramaré finally settled this problem for all  $N \geq 2$ .

Kaufman measures and their construction have since become useful tool in Diophantine approximation. See for example the paper by Pollington and Velani [28] with a connection to the Littlewood conjecture.

**1.2. Fourier transforms of invariant measures.** The motivation for this paper is to investigate under which conditions on a dynamical system a given invariant measure  $\mu$  is a Rajchman measure with a polynomial decaying Fourier transform. It turns out the construction of the Kaufman measure on badly approximable numbers in Theorem 1.2 provides a good reference to understand this problem.

Kaufman's original proof relied on constructing a measure  $\mu$  with large enough dimension supported on  $B_N$  which satisfies suitable  $\mu$  independence assumption for the measures of cylinders. The independence in the construction allowed the use of law of large numbers to obtain  $\mu$  generic growth rates for the *continuants* of the continued fraction expansions and that the generic continuants do not *deviate* much from this generic growth. This controlled growth of generic continuants is crucial in the proof when studying the oscillations of  $\mu$  with  $|\widehat{\mu}(\xi)|$  for large frequencies  $\xi$ .

The growth rate of continuants can be explained by the evolution of the dynamical system  $([0, 1] \setminus \mathbb{Q}, T)$  on irrationals, where  $T : [0, 1] \setminus \mathbb{Q} \rightarrow [0, 1] \setminus \mathbb{Q}$  is the *Gauss map*, defined by

$$T(x) = \frac{1}{x} \bmod 1.$$

The Gauss map acts as a shift for the continued fraction expansions of irrational numbers and the growth rate of the derivative  $(T^n)'(x)$  as  $n \rightarrow \infty$  of the composition  $T^n$  is comparable to the growth of the  $n$ th continuant of an irrational  $x$  (see the end of Section 2). Thus implicitly Kaufman's construction already relies heavily on the dynamics of the Gauss map, which leads to the problem; is the decay of Fourier coefficients in general just a generic property of the dynamical system?

We show that Rajchman measures are indeed quite common for the Gauss map and the classical *Gibbs condition* is enough (see Section 3 for a definition). Gibbs measures arise in thermodynamical formalism as equilibrium states for some potential functions and they enjoy a weak form of independence; for example, they are always ergodic  $T$  invariant measures. All Bernoulli measures on badly approximable numbers are Gibbs measures and there is a Gibbs measure which is equivalent to the Hausdorff measure  $\mathcal{H}^s$  of dimension  $s = \dim B_N$  when restricted to  $B_N$ . Gibbs measures tend to satisfy strong statistical properties, in particular in this setting we can show they satisfy large deviation results on the generic growth of the Gauss map, which enable us to adapt Kaufman's approach.

However, as the alphabet generated by the Gauss map is infinite, the potential associated to the Gibbs measure  $\mu$  may not in general satisfy the large deviation principle, so we need to impose a weak condition on the decay in the tail of the Gibbs measure. More precisely, that there exists  $\delta > 0$  such that when  $n \rightarrow \infty$  the probability

$$\mu(\{x \in [0, 1] \setminus \mathbb{Q} : a_1(x) \geq n\}) = O(n^{-\delta}) \tag{1.2}$$

This condition on the distribution of the first continued fraction digit of  $x$  is equivalent to the Gibbs measure  $\mu$  satisfying the large deviation theory bounds for the Kaufman type arguments we use in the proof; see Proposition 4.4 and Remark 4.5 below. For a finite alphabet, which would for example cover the situation of  $B_N$  as symbolically it is  $\{1, 2, \dots, N\}^{\mathbb{N}}$ , such an assumption is unnecessary so the only assumption on the Gibbs measure required is that the dimension is greater than  $\frac{1}{2}$ . More generally, we can just consider any finite  $\mathcal{A} \subset \mathbb{N}$  and define an irrational  $x$  to be  $\mathcal{A}$ -badly approximable, if the digits of the continued fraction expansion of  $x$  are only in  $\mathcal{A}$ , and write  $x \in B(\mathcal{A})$ . Then  $B_N = B(\{1, 2, \dots, N\})$ .

**Theorem 1.3.** *We have the following properties:*

- (1) *If  $\mu$  is any Gibbs measure for the Gauss map which satisfies (1.2) and Hausdorff dimension  $\dim \mu > 1/2$ , then  $\mu$  is a Rajchman measure with Fourier transform*

$$\widehat{\mu}(\xi) = O(|\xi|^{-\eta}), \quad |\xi| \rightarrow \infty, \quad (1.3)$$

*for some  $\eta > 0$ .*

- (2) *If  $\mathcal{A} \subset \mathbb{N}$  is finite and  $\mu$  is any Gibbs measure for the Gauss map restricted to  $B(\mathcal{A})$  with dimension  $\dim \mu > 1/2$ , then (1.3) holds.*

It is worth pointing out the contrast between this result and the case for the map  $\times n$  map  $x \mapsto nx \bmod 1$  on  $\mathbb{T}$ , where a measure  $\mu$  is  $\times n$  invariant if and only if  $\widehat{\mu}(nk) = \widehat{\mu}(k)$  for all  $k \in \mathbb{Z}$  so the only  $\times n$  invariant measure with whose Fourier transform has a power decay is Lebesgue measure.

After a slight adjustment of the proof of Theorem 1.3, it also applies to Hausdorff measures  $\mathcal{H}^s$  on the badly approximable numbers  $B(\mathcal{A})$  with large enough dimension:

**Corollary 1.4.** *If  $\mathcal{A} \subset \mathbb{N}$  is finite  $s = \dim B(\mathcal{A}) > 1/2$ , then the Hausdorff measure  $\mathcal{H}^s$  restricted to  $B(\mathcal{A})$  is a Rajchman measure with polynomially decaying Fourier transform. In particular, the Hausdorff measure on  $B_N$  satisfies this property for any  $N \geq 2$ .*

This is due to the fact that Hausdorff measure  $\mathcal{H}^s$  is an  $s$ -conformal measure in our setting and equivalent to a Gibbs measure with uniformly positive and finite densities; see Section 6.7 for more details and references.

Another application of Theorem 1.3 concerns the Fourier-Stieltjes coefficients of singular monotonic functions  $f : [0, 1] \rightarrow [0, 1]$ , that is, monotonic  $f$  with zero derivative Lebesgue almost everywhere. Such functions arise naturally from singular measures. If  $\mu$  is a singular probability measure with respect to Lebesgue measure, with support  $[0, 1]$  and without atoms, then the cumulative distribution function  $x \mapsto \mu([0, x])$  is an increasing singular monotonic function. Moreover, if we have an increasing singular monotonic function  $f : [0, 1] \rightarrow [0, 1]$ , then defining a Stieltjes measure  $\mu$  by the formula  $\mu([x, y]) = f(y) - f(x)$  for all reals  $0 \leq x < y \leq 1$  has  $f$  as the cumulative distribution function. The *Fourier-Stieltjes coefficients* of  $f$  are then precisely the Fourier coefficients  $\widehat{\mu}(k)$ ,  $k \in \mathbb{Z}$ , of the measure  $\mu$ . There has been much research to understand the asymptotics of these coefficients for various singular functions, see Salem's paper [30] for references.

All Gibbs measures  $\mu$  for the Gauss map that are fully supported and not equal to the *Gauss measure* (see Remark 3.4(2) for a definition) give rise to a singular monotonic function by the cumulative distribution function  $x \mapsto \mu([0, x])$ . Indeed, any ergodic measure for the Gauss map not equal to the *Gauss measure* (see Remark 3.4(2)) has to be singular with respect to Gauss measure and hence to Lebesgue measure. Thus if a singular monotonic function  $f$  gives rise to a Gibbs measure  $\mu$  for a Gauss map with the requirement (1.2) is on the tail of  $\mu$  and  $\dim \mu > \frac{1}{2}$  satisfied, we can apply Theorem 1.3 to have that the Fourier-Stieltjes coefficients of  $f$  decay to 0 at infinity with a polynomial rate. An example of this application is given in the following.

A famous construction of a singular monotonic function given by Minkowski in [25] is the *Minkowski's question mark function*, which is defined for irrationals  $x \in [0, 1]$  by

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

and for rationals  $x \in [0, 1]$  with continued fraction expansion  $[a_1(x), \dots, a_m(x)]$  by replacing the sum just up to  $m$  and setting  $\mu(0) = 0$ . The original motivation for the construction of  $\mu$  was to relate the continued fraction expansions of quadratic irrational numbers to their binary expansions. It turns out that  $\mu$  has many fractal features, see for example the work by Kesseböhmer and Stratmann [15] on multifractal analysis and the references therein. Let  $\mu$  is the Stieltjes measure defined by  $\mu$ . It dates back to the work of Salem [30] in 1943 where he proved that the Fourier coefficients of  $\mu$  converge 'in average' to 0, that is,

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

as  $n \rightarrow \infty$  for a suitable  $\eta > 0$ . In the same paper Salem asked can this be strengthened to  $\widehat{\mu} \rightarrow 0$ , i.e. whether  $\mu$  is Rajchman or not. This problem has been re-visited over the years, for example it was recently mentioned in [4].

We see that  $\mu$  is indeed a Gibbs measure for the Gauss map; in particular, it is a *Bernoulli measure* (see Remark 3.4(1) for a definition) associated to the weights  $p_a = 2^{-a}$ ,  $a \in \mathbb{N}$ . The tail event  $\{a_1(x) \geq n\}$  has an exponential decay:

$$\mu(\{x \in [0, 1] \setminus \mathbb{Q} : a_1(x) \geq n\}) = 2^{-n+1}$$

so (1.2) is satisfied. Finally in [19] the Hausdorff dimension of  $\mu$  is found. It is shown that

$$\dim \mu = \frac{\log 2}{2 \int_0^1 \log(1+x) d\mu(x)} > \frac{1}{2}.$$

Thus the measure  $\mu$  satisfies the assumptions of Theorem 1.3 and we obtain a corollary, which answers to Salem's original question and also gives a bound for the rate for the decay:

**Corollary 1.5.** *The Fourier-Stieltjes coefficients of the Minkowski's question mark function decay to 0 at infinity with a polynomial rate.*

This application of Theorem 1.3 was pointed out by Persson in [27]. Note an alternative approach to solving Salem's problem was presented in [35].

Furthermore, the main Theorem 1.3 together with the equidistribution Theorem 1.1 yields that suitable Gibbs measures will give us normal numbers in their support:

**Corollary 1.6.** (1) *If  $\mu$  is any Gibbs measure for the Gauss map which satisfies (1.2) and Hausdorff dimension  $\dim \mu > 1/2$ , then  $(s_k x)_{k \in \mathbb{N}}$  equidistributes modulo 1 at  $\mu$  almost every  $x$  for any strictly increasing sequence of natural numbers  $(s_k)_{k \in \mathbb{N}}$ . In particular,  $\mu$  is  $n$ -normal for any  $n \in \mathbb{N}$ .*

(2) *If  $\mathcal{A} \subset \mathbb{N}$  is finite and  $\mu$  is any Gibbs measure for the Gauss map restricted to  $B(\mathcal{A})$  with dimension  $\dim \mu > 1/2$ , then the same conclusion holds for  $\mu$ .*

Recently, in [9] Hochman and Shmerkin presented a general criterion for the  $n$ -normality of a measure  $\mu$ , which adapts to many general nonlinear iterated function systems including the one determined by the Gauss map. In particular, they obtained the following in [9, Theorem 1.12].

**Theorem 1.7** (Hochman-Shmerkin). *Any Gibbs measure for the Gauss map supported on  $B(\mathcal{A})$  for a finite  $\mathcal{A} \subset \mathbb{N}$  is  $n$ -normal for any  $n \in \mathbb{N}$ .*

In their proof Hochman and Shmerkin did not rely on large deviations, Fourier transforms or the Davenport-Erdős-LeVeque criterion on equidistribution but rather they used a completely different approach by studying the ergodic properties of the so called *scenery flow* invariant distributions generated by the Gibbs measure. However, the results of Hochman and Shmerkin do not apply to give the more general statement on equidistribution of  $(s_k x)_{k \in \mathbb{N}}$  for *any* strictly increasing sequence  $(s_k)_{k \in \mathbb{N}}$  obtained from Davenport-Erdős-LeVeque, so our result extends their results in this respect even in the case of finite alphabets with the caveat of the curious Hausdorff dimension assumption. The  $n$ -normality part of Corollary 1.6(1) was conjectured also in [9]: "...It seems very likely that

the result holds also for Gibbs measures when  $\mathcal{A} \subset \mathbb{N}$  is infinite, under standard assumptions on the Gibbs potential...”, so dropping the dimension assumption  $\dim \mu > 1/2$  in Corollary 1.6(1) should be possible at least for the  $n$ -normality.

We finish this section by outlining how the rest of the paper is structured. In Sections 2 and 3 we describe the basic results from the theory of continued fractions and thermodynamic formalism that we use. In Section 4 we prove the large deviation results we need and in Section 5 we describe how we can use these results to decompose the Gibbs measure, where we can apply Kaufman type arguments. We can then complete the proof of Theorem 1.3 in Section 6 and we finish the paper by commenting on some possible extensions of our work beyond the system  $([0, 1] \setminus \mathbb{Q}, T)$  to more general nonlinear dynamical systems.

## 2. DIOPHANTINE APPROXIMATION AND THE GAUSS MAP

During the course of the proof of Theorem 1.3 we will use heavily the language of continued fractions. In this section we will introduce our notation for continued fractions and state the elementary properties of continued fraction that we make use of. For the majority of the results presented here, we refer to the book by Khinchin [16]. Write  $\mathbb{N} = \{1, 2, \dots\}$ . Given digits  $a_1, \dots, a_n \in \mathbb{N}$ , where  $n \in \mathbb{N}$ , a *continued fraction* is the rational number

$$[a_1, a_2, \dots, a_n] = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\ddots + \frac{1}{a_n}}}}$$

For each irrational  $x \in [0, 1]$  we can find unique numbers  $a_i(x) \in \mathbb{N}$  such that

$$x = [a_1(x), a_2(x), \dots] := \lim_{n \rightarrow \infty} [a_1(x), a_2(x), \dots, a_n(x)].$$

This expression is the *continued fraction expansion* of  $x$ . Thus we obtain a coding of  $[0, 1] \setminus \mathbb{Q}$  via the words  $\mathbf{a} = (a_1, a_2, \dots) \in \mathbb{N}^{\mathbb{N}}$ . For notational purposes, let  $\mathbb{N}^n$  and  $\mathbb{N}^*$  be the collections of length  $n$  and any finite length words of natural numbers, and in the proceedings  $\mathbf{a}$  can mean either finite word  $(a_1, a_2, \dots, a_n)$  or an infinite word  $(a_1, a_2, \dots)$  depending on the context.

**Definition 2.1** (Continuants). For any finite word  $\mathbf{a} = (a_1, a_2, \dots, a_n) \in \mathbb{N}^n$ , there exists integers  $p_n(\mathbf{a})$  and  $q_n(\mathbf{a})$  such that

$$[a_1, a_2, \dots, a_n] = \frac{p_n(\mathbf{a})}{q_n(\mathbf{a})}.$$

The denominator here  $q_n(\mathbf{a})$  called the *continuant* of the continued fraction  $[\mathbf{a}]$ . We define the numbers  $q_k(\mathbf{a})$  also for  $k < n$  even if the word  $\mathbf{a}$  has length bigger than  $k$ . Then  $q_k(\mathbf{a})$  is just the continuant  $q_k(\mathbf{a}|_k)$ , where  $\mathbf{a}|_k$  is the restriction  $(a_1, \dots, a_k)$ .

Continuants satisfy by construction the following important relations:

*Remark 2.2.* Given  $\mathbf{a} \in \mathbb{N}^n$ , we have

- (1) the recurrence relation

$$q_n(\mathbf{a}) = a_n q_{n-1}(\mathbf{a}) + q_{n-2}(\mathbf{a}); \tag{2.1}$$

- (2) relation between numerators and denominators

$$q_n(\mathbf{a}) p_{n-1}(\mathbf{a}) - q_{n-1}(\mathbf{a}) p_n(\mathbf{a}) = (-1)^n; \tag{2.2}$$

- (3) invariance and recovery under *mirroring*

$$q_n(\mathbf{a}) = q_n(\mathbf{a}^{\leftarrow}) \quad \text{and} \quad q_{n-1}(\mathbf{a}) = p_n(\mathbf{a}^{\leftarrow}) \tag{2.3}$$

where  $\mathbf{a}^{\leftarrow} = (a_n, \dots, a_1)$  is the *mirror image* of  $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{N}^n$ .

Continuants can be used to estimate the rate of convergence of continued fractions to a given irrational number. In particular, if  $x \in [0, 1] \setminus \mathbb{Q}$  and  $\mathbf{a} \in \mathbb{N}^n$  is its continued fraction digits up to  $n$ , then *Dirichlet's theorem* tells us

$$\left| x - \frac{p_n(\mathbf{a})}{q_n(\mathbf{a})} \right| \leq \frac{1}{q_n(\mathbf{a})^2}.$$

Continuants themselves have already a lower bound for their growth: in particular, if

$$\theta = \frac{\sqrt{5} + 1}{2}$$

is the *Golden ratio*, then we have the following

**Lemma 2.3.** *If  $\mathbf{a} \in \mathbb{N}^n$ , then  $q_n(\mathbf{a}) \geq c_0 \theta^n$  for some constant  $c_0 > 0$  independent of  $\mathbf{a}$ .*

*Proof.* We have that each  $a_i \geq 1$  so we have by the recurrence relation (2.1), that the word  $1^n = (1, 1, \dots, 1) \in \mathbb{N}^n$  satisfies  $q_n(\mathbf{a}) \geq q_n(1^n)$ . The recurrence relation (2.1) gives the relation

$$q_n(1^n) = q_{n-1}(1^n) + q_{n-2}(1^n)$$

with  $q_1(1^1) = 1$  and for convention  $q_0(1^0) := 0$ , so the sequence  $(q_n(1^n))_{n \in \mathbb{N}}$  is the Fibonacci sequence. By Binet's formula this means that it has exponential growth rate given by  $c_0 \theta^n$ .  $\square$

Another way to understand the continuants  $q_n$  when  $n$  is large, is to interpret the problem with the growth rate of orbits for a suitable dynamical system. This gives rise to the notion of Gauss map:

**Definition 2.4** (Gauss map and inverse branches). Let  $T : [0, 1] \setminus \mathbb{Q} \rightarrow [0, 1] \setminus \mathbb{Q}$  be the *Gauss map* defined by

$$T(x) = \frac{1}{x} \bmod 1.$$

Moreover, let  $T_a : [0, 1] \setminus \mathbb{Q} \rightarrow I_a := [\frac{1}{a+1}, \frac{1}{a}] \setminus \mathbb{Q}$  be the inverse branch  $(T|_{I_a})^{-1}$  of the Gauss map given by

$$T_a(x) = \frac{1}{x+a}, \quad x \in [0, 1] \setminus \mathbb{Q}.$$

The Gauss map acts as a shift for the continued fraction digits of  $x \in [0, 1] \setminus \mathbb{Q}$ :

$$T(x) = [a_2(x), a_3(x), \dots] \in [0, 1] \setminus \mathbb{Q}.$$

In the language of dynamical systems this means that the Gauss map is conjugate to the full shift  $\sigma : \mathbb{N}^{\mathbb{N}} \rightarrow \mathbb{N}^{\mathbb{N}}$ .

The growth rate of the continuants under compositions of the inverse branches

$$T_{\mathbf{a}} := T_{a_1} \circ T_{a_2} \circ \dots \circ T_{a_n}$$

for  $\mathbf{a} \in \mathbb{N}^n$  is about of the order  $q_n(\mathbf{a})^{-2}$  as the *construction intervals*

$$I_{\mathbf{a}} := T_{\mathbf{a}}([0, 1] \setminus \mathbb{Q}) = \{x \in [0, 1] \setminus \mathbb{Q} : a_i(x) = a_i \text{ for } 1 \leq i \leq n\}$$

are of the size  $q_n(\mathbf{a})^{-2}$ . More precisely

**Lemma 2.5.** *Given  $\mathbf{a} \in \mathbb{N}^n$ , then we have*

$$\frac{1}{4} q_n(\mathbf{a})^{-2} \leq |T'_{\mathbf{a}}| \leq q_n(\mathbf{a})^{-2} \tag{2.4}$$

*In particular, the same bounds hold for the length  $|I_{\mathbf{a}}|$ .*

*Proof.* By construction the inverse branch  $T_{\mathbf{a}}$  at  $x \in [0, 1]$  and  $\mathbf{a} \in \mathbb{N}^n$  is precisely

$$T_{\mathbf{a}}(x) = [a_1, a_2, \dots, a_n + x] = \frac{p_{n-1}(\mathbf{a})x + p_n(\mathbf{a})}{q_{n-1}(\mathbf{a})x + q_n(\mathbf{a})} = \frac{p_n(\mathbf{a})}{q_n(\mathbf{a})} + \frac{(-1)^n x}{(q_{n-1}(\mathbf{a})x + q_n(\mathbf{a}))q_n(\mathbf{a})}. \tag{2.5}$$



This yields that the derivative has the following representation

$$T_{\mathbf{a}}'(x) = \frac{(-1)^n}{(q_{n-1}(\mathbf{a})x + q_n(\mathbf{a}))^2}. \quad (2.6)$$

By the recurrence relation (2.1) we have  $q_{n-1}(\mathbf{a}) \leq q_n(\mathbf{a})$  so

$$\frac{1}{4q_n(\mathbf{a})^2} \leq |T_{\mathbf{a}}'(x)| \leq \frac{1}{q_n(\mathbf{a})^2}.$$

By computing  $T_{\mathbf{a}}$  at the end points or using mean value theorem we obtain the same bounds for the length  $|I_{\mathbf{a}}|$ .  $\square$

Using this, we can give a useful quasi-independence for the continuants:

**Lemma 2.6.** *Given  $\mathbf{a} \in \mathbb{N}^n$  and  $1 \leq j < n$ , we have*

$$\frac{1}{2} \leq \frac{q_n(\mathbf{a})}{q_{n-j}(a_1, \dots, a_{n-j})q_j(a_{n-j+1}, \dots, a_n)} \leq 4.$$

*Proof.* Write  $\mathbf{b} = a_1, \dots, a_{n-j}$  and  $\mathbf{c} = a_{n-j+1}, \dots, a_n$ . Given  $x \in [0, 1]$  we have by the chain rule

$$T_{\mathbf{a}}'(x) = T_{\mathbf{b}}'(T_{\mathbf{c}}(x))T_{\mathbf{c}}'(x).$$

Then by Lemma 2.5, we have

$$\frac{1}{16q_{n-j}(\mathbf{b})^2q_j(\mathbf{c})^2} \leq |T_{\mathbf{b}}'(T_{\mathbf{c}}(x))||T_{\mathbf{c}}'(x)| \leq \frac{1}{q_{n-j}(\mathbf{b})^2q_j(\mathbf{c})^2}$$

so applying Lemma 2.5 again for  $T_{\mathbf{a}}(x)$  and taking a root we have

$$\frac{1}{2}q_{n-j}(\mathbf{b})q_j(\mathbf{c}) \leq q_n(\mathbf{a}) \leq 4q_{n-j}(\mathbf{b})q_j(\mathbf{c}).$$

$\square$

### 3. THERMODYNAMICAL FORMALISM AND GIBBS MEASURES

The fact that the Gauss map is conjugate to the full shift on  $\mathbb{N}^{\mathbb{N}}$  enables us to use the thermodynamical formalism for countable shifts developed by Mauldin and Urbański in [24] and Sarig in the papers [31, 32]. This also relates to earlier work of Walters in [33]. Let us now recall some classical notation and definitions related to thermodynamical formalism and Gibbs measures.

For the rest of the section, we write  $X = [0, 1] \setminus \mathbb{Q}$  on which the Gauss map is conjugated to the full shift on  $\mathbb{N}^{\mathbb{N}}$ . Let  $\mathcal{M}$  be the set of all Borel probability measures on  $[0, 1]$  and write  $\mathcal{M}_T \subset \mathcal{M}$  as the collection of  $T$  invariant measures  $\mu$ , that is,

$$\mu(A) = \mu(T^{-1}A) = \sum_{a \in \mathbb{N}} \mu(T_a A)$$

for a Borel set  $A \subset X$ . Given a function  $\varphi : X \rightarrow \mathbb{R}$  (which we often call a *potential* or an *observable*) and  $n \in \mathbb{N}$  let  $S_n\varphi$  be the *Birkhoff sum* defined at  $x \in X$  by

$$S_n\varphi(x) = \sum_{k=0}^{n-1} \varphi(T^k(x)).$$

**Definition 3.1** (Entropy and Lyapunov exponents). Given  $\mu \in \mathcal{M}_T$ , the *Kolmogorov-Sinai entropy*  $h_{\mu}$  and *Lyapunov-exponent*  $\lambda_{\mu}$  of  $\mu$  with respect to  $T$  are defined by

$$h_{\mu} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\mathbf{a} \in \mathbb{N}^n} -\mu(I_{\mathbf{a}}) \log \mu(I_{\mathbf{a}}) \quad \text{and} \quad \lambda_{\mu} = \int \log |T'| d\mu.$$

We say a function  $\varphi : X \rightarrow \mathbb{R}$  is *locally Hölder* if there exists constants  $C > 0$  and  $\delta > 0$  such that for all  $n \in \mathbb{N}$  we have

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

For a finite word  $\mathbf{a} \in \mathbb{N}^n$  we will let  $\mathbf{a}^\infty$  denote the infinite periodic word of period  $n$ , which repeats the word  $\mathbf{a}$ . For a given function  $\varphi$  we can define its pressure as follows

**Definition 3.2** (Pressure and equilibrium states). The *pressure* of a potential  $\varphi$  is the quantity

$$P(\varphi) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \left( \sum_{\mathbf{a} \in \mathbb{N}^n} \exp(S_n \varphi(\mathbf{a}^\infty)) \right).$$

Alternatively, we have the variational principle

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

and any measure assuming this supremum is called an *equilibrium state* for  $\varphi$ .

For a given potential  $\varphi$  there often exists a *unique* invariant measure  $\mu = \mu_\varphi$ , which satisfies a suitable regularity condition with respect to  $e^{S_n(\varphi(x)) - nP(\varphi)}$  known as the *Gibbs condition*:

**Definition 3.3** (Gibbs measures). A measure  $\mu \in \mathcal{M}_T$  is a *Gibbs measure* with potential  $\varphi$ , if there is a constant  $C \geq 1$  such that for any  $n \in \mathbb{N}$  and  $\mathbf{a} \in \mathbb{N}^n$  we have

$$C^{-1} e^{S_n \varphi(T_{\mathbf{a}}(x)) - nP(\varphi)} \leq \mu(I_{\mathbf{a}}) \leq C e^{S_n \varphi(T_{\mathbf{a}}(x)) - nP(\varphi)}$$

at any  $x \in X$ . Notice that by definition  $T_{\mathbf{a}}(x)$  is always an element of the construction interval  $I_{\mathbf{a}}$ .

*Remark 3.4.* Here are a few well-known examples of Gibbs measures:

- (1) Take some numbers  $0 \leq p_a \leq 1$ ,  $a \in \mathbb{N}$ , with  $\sum_{a \in \mathbb{N}} p_a = 1$ , then the measure  $\mu$  giving mass  $p_{a_1} p_{a_2} \cdots p_{a_n}$  to the interval  $I_{\mathbf{a}}$  for any  $\mathbf{a} = a_1 \dots a_n$  is called a *Bernoulli measure* for  $T$ . This measure is a Gibbs measure associated to the potential  $\varphi$  defined by  $\varphi(x) = \log p_{a_1(x)}$ , where  $a_1(x)$  is the first continued fraction digit of  $x$ . Indeed, the pressure  $P(\varphi) = 0$ , and so the Birkhoff sum  $S_n \varphi(x)$  reduces the Gibbs comparison to equality:

$$\mu(I_{\mathbf{a}}) = p_{a_1} p_{a_2} \cdots p_{a_n} = e^{S_n \varphi(T_{\mathbf{a}}(x))}.$$

- (2) More delicate examples can be found from taking the potential  $\varphi = -\log |T'|$  for which *Gauss measure*  $\mu$  defined by

$$d\mu(x) = \frac{1}{\log 2} \frac{1}{1+x} dx$$

is Gibbs. Gauss measure is the unique  $T$  invariant measure which is equivalent to Lebesgue measure.

- (3) If we take some finite  $\mathcal{A} \subset \mathbb{N}$  with at least 2 elements and let  $s = \dim B(\mathcal{A})$ , then the  $s$ -dimensional Hausdorff measure  $\mathcal{H}^s$  restricted to  $B(\mathcal{A})$  is equivalent to a Gibbs measure with the potential  $\varphi = -s \log |T'|$  with densities bounded away from 0 and infinity. This follows for example from [7, Theorem 5.3] when interpreting  $B(\mathcal{A})$  as a cookie-cutter set.

Gibbs measures are often uniquely determined as invariant measures for the Ruelle transfer operator  $\mathcal{L}_\varphi$  determined by the potential. Let  $C(X)$  be the space of bounded continuous maps  $f : X \rightarrow \mathbb{C}$ .

**Definition 3.5** (Ruelle transfer operator). The Ruelle *transfer operator* for a potential  $\varphi$  is the action  $\mathcal{L}_\varphi : C(X) \rightarrow C(X)$ , defined for  $f \in C(X)$  and  $x \in X$  by

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

The dual operator of  $\mathcal{L}_\varphi$  on  $\mathcal{M}_T$  is the action  $\mathcal{L}_\varphi^* : \mathcal{M} \rightarrow \mathcal{M}$ , defined for  $\nu \in \mathcal{M}$  and  $f \in C(X)$  by

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

The following result relates the concepts of Gibbs measures, transfer operators and the pressure function.

**Proposition 3.6.** *Let  $\varphi : X \rightarrow \mathbb{R}$  be a locally Hölder potential. We have*

$$P(\varphi) < \infty \quad \text{if and only if} \quad \sum_{n=1}^{\infty} \exp(\varphi((n)^\infty)) < \infty.$$

Moreover if  $P(\varphi) = 0$  then the following results hold:

- (1) *There exists a Hölder continuous function  $h$  bounded away from 0 and a measure  $\mu \in \mathcal{M}$  such that  $\mathcal{L}_\varphi h = h$  and  $\mathcal{L}_\varphi^* \mu = \mu$ . Moreover the measure  $\mu_\varphi$  defined by*

$$d\mu_\varphi(x) = h(x) d\mu(x)$$

*is the unique Gibbs measure for  $\varphi$ .*

- (2) *If  $\int \varphi d\mu_\varphi > -\infty$ , then  $\mu_\varphi$  is the unique equilibrium state for the potential  $\varphi$ .*

*Proof.* This follows from Proposition 2.1.9, Theorem 2.7.3 and Corollary 2.7.5 in [24]. □

For a locally Hölder potential  $\varphi$  with  $P(\varphi) = 0$  we will use  $\mu_\varphi$  to denote the unique Gibbs measure for  $\varphi$ . One advantage of the transfer operator in our setting is what happens under iteration. Let  $n \in \mathbb{N}$  and  $f$  be a continuous map we have at any  $x \in [0, 1] \setminus \mathbb{Q}$  that

$$\mathcal{L}_\varphi^n f(x) = \sum_{y: T^n(y)=x} e^{S_n \varphi(y)} f(y) = \sum_{\mathbf{a} \in \mathbb{N}^n} e^{S_n(\varphi(T_{\mathbf{a}}(x)))} f(T_{\mathbf{a}}(x)).$$

Notice that since  $T_{\mathbf{a}}(x) \in I_{\mathbf{a}}$  and the pressure  $P(\varphi) = 0$ , the definition of Gibbs measure gives

$$w_{\mathbf{a}}(x) := e^{S_n(\varphi(T_{\mathbf{a}}(x)))}$$

is comparable to  $\mu_\varphi(I_{\mathbf{a}})$  up to the Gibbs constant  $C \geq 1$ . We now define the family of potentials which we consider.

**Definition 3.7.** We will consider the class of locally Hölder potentials  $\varphi$  which satisfy the following properties:

- (1)  $P(\varphi) = 0$ ,  $\mathcal{L}_\varphi 1 = 1$  and  $\varphi \leq 0$ ,  
(2) There exists  $t_c > 0$  such that for all  $t \in (-2t_c, 2t_c)$  we have that

$$P(-t \log |T'| + \varphi) < \infty \quad \text{and} \quad P(t(\varphi - s \log |T'|) + \varphi) < \infty,$$

where  $s$  is the fraction  $h_\mu/\lambda_\mu$  of the entropy  $h_\mu$  and Lyapunov exponent  $\lambda_\mu$  of  $\mu = \mu_\varphi$ .

*Remark 3.8.* We note that it follows from the Birkhoff ergodic theorem that Gibbs measures are exact dimensional and the number  $s = h_\mu/\lambda_\mu$  in Definition 3.7 is in fact the Hausdorff dimension  $\dim \mu$  of the Gibbs measure  $\mu = \mu_\varphi$ , see for example [24, Theorem 4.4.2].

It should be noted that our first assumption here is not restrictive at all due to the following proposition.

**Proposition 3.9.** *Let  $\psi$  be a locally Hölder function such that  $P(\psi) < \infty$ . We can find a locally Hölder function  $\varphi$  such that  $\varphi \leq 0$ ,  $P(\varphi) = 0$ ,  $\mathcal{L}_\varphi 1 = 1$  and*

$$\mu_\varphi = \mu_\psi.$$

*Proof.* It follows from Lemma 1 in [31] that if  $P(\psi) < \infty$  then we can find a locally Hölder  $\varphi$  such that  $\varphi \leq 0$ ,  $P(\varphi) = 0$ ,  $\mathcal{L}_\varphi 1 = 1$  and  $\varphi - \psi$  is cohomologous to a constant in the class of bounded Hölder potentials. It follows from Theorem 2.2.7 in [24] that in this case  $\mu_\varphi = \mu_\psi$ . □

The following proposition summarises the main results from the thermodynamic formalism on countable Markov shifts we need:

**Proposition 3.10.** *Let  $\varphi$  be a potential satisfying Definition 3.7. Then*

(1) *for all bounded continuous functions  $f$  we have that*

$$\int \mathcal{L}_\varphi f d\mu_\varphi = \int f d\mu_\varphi;$$

(2) *let  $t_0 > 0$  and  $\psi$  be a locally Hölder continuous potential such that  $P(t\psi + \varphi) < \infty$  for all  $t \in (-t_0, t_0)$ . If we define  $Q : (-t_0, t_0) \rightarrow \mathbb{R}$  by*

$$Q(t) = P(t\psi + \varphi),$$

*then  $Q$  is analytic and convex on  $(-t_0, t_0)$  and  $Q'(t) = \int \psi d\mu_{t\psi + \varphi}$ .*

*Proof.* The first part of the proposition is immediate from the condition that  $\mathcal{L}_\varphi 1 = 1$  and the fact that  $\mu_\varphi$  is then a fixed point for the dual operator. The second part of the proposition follows from Corollary 4 in Sarig [32].  $\square$

*Remark 3.11.* We also include the case where we restrict  $T$  to  $\mathcal{B}(\mathcal{A})$  for a finite  $\mathcal{A} \subset \mathbb{N}$ . Then we let  $\varphi : \mathcal{B}(\mathcal{A}) \rightarrow \mathbb{R}$  be a Hölder continuous potential satisfying the conditions of Definition 3.7. All the above results hold in this setting with the added advantage that the pressure is never infinite.

#### 4. LARGE DEVIATION BOUNDS FOR GIBBS MEASURES

Let  $X_1, X_2, \dots$  be identically distributed independent random variables with expectation 0. In probability theory large deviations in their simplest form examine the rate at which

$$\mathbb{P}(|X_1 + \dots + X_n| > \varepsilon n)$$

decays as  $n \rightarrow \infty$  for some  $\varepsilon > 0$ . *Cramér's theorem* states that if the moment generating function is always finite then this decay is exponential, see for example Section 1.9 in [6]. Large deviations also appear in dynamics and it turns out that the theory can be developed for certain invariant measures for hyperbolic dynamical systems, see for example the paper by Young [36] in the compact setting. In this section we prove results of this form for the Gauss map we need for the Rajchman property.

In the finite state space case the results follow from the much more general results in [17, 20, 36] so we concentrate on proving the results in the countable state space. This is by no means the first time large deviations have been considered for the countable shift, for example see Section 6 of the paper Kifer, Peres and Weiss [18] or Yuri [37]. However, they consider observables which are bounded whereas we are interested in the observable  $-\log |T'|$  which is unbounded.

Let  $\varphi$  be a locally Hölder potential such that  $\varphi \leq 0$  and  $P(\varphi) = 0$ . We denote by  $\mu_\varphi$  the Gibbs measure for  $\varphi$ . For simplicity of the notation, given a locally Hölder observable  $f$ , write

$$\alpha(f) := \int f d\mu_\varphi.$$

Our main large deviation result is the following

**Theorem 4.1.** *Let  $f$  be any locally Hölder observable. Suppose*

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

*contains a neighbourhood of the origin. Then for any  $\varepsilon > 0$  there exists  $\delta = \delta(\varepsilon) > 0$  and  $n_1 = n_1(\varepsilon) \in \mathbb{N}$  such that for  $n \geq n_1$*

$$\mu_\varphi(\{x \in [0, 1] : |S_n f(x) - n\alpha(f)| > n\varepsilon\}) \leq e^{-n\delta}.$$

As a corollary to this we have the large deviation result for the observables we use. Write  $\psi = -\log |T'|$  for the observable corresponding to the Lyapunov exponent. That is, we have

$$\lambda = \lambda_{\mu_\varphi} = -\alpha(\psi).$$

We let  $s = h_{\mu_\varphi}/\lambda_{\mu_\varphi}$ , which is the Hausdorff dimension of  $\mu_\varphi$ .

**Corollary 4.2.** (1) *Suppose that  $\{t \in \mathbb{R} : P(t\psi + \varphi) < \infty\}$  contains a neighbourhood of the origin. Then for any  $\varepsilon > 0$  there exists  $\delta = \delta(\varepsilon) > 0$  and  $n_1 = n_1(\varepsilon) \in \mathbb{N}$  such that for all  $n \geq n_1$  we have*

$$\mu_\varphi(\{x \in [0, 1] : |S_n\psi(x) + n\lambda| \geq n\varepsilon\}) \leq e^{-n\delta}.$$

(2) *Suppose that  $\{t \in \mathbb{R} : P(t(\varphi - s\psi) + \varphi) < \infty\}$  contains a neighbourhood of the origin. Then for any  $\varepsilon > 0$  and for all  $n \geq n_1$  we have*

$$\mu_\varphi\left(\left\{x \in [0, 1] : \left|\frac{S_n\varphi(x)}{S_n\psi(x)} - s\right| > \varepsilon\right\}\right) \leq e^{-n\delta}.$$

*Proof.* Part 1 is Theorem 4.1 with  $f = \psi$ . To prove part 2 we let  $f = \varphi - s\psi$  and note that here  $\alpha(f) = 0$ . We then have that for  $\varepsilon > 0$  if  $x$  satisfies that

$$\left|\frac{S_n\varphi(x)}{S_n\psi(x)} - s\right| > \varepsilon$$

giving  $|S_nf(x)| > \varepsilon|S_n\psi(x)|$ . By Lemma 2.5 and Lemma 2.3 we have that

$$|(T^k)'(x)| \geq c_0^2\theta^{2k}$$

where, recall,  $\theta > 1$  was the Golden ratio and  $c_0 > 0$  a universal constant. Then we obtain by the definition of the Birkhoff sum  $S_n\psi(x)$  that

$$|S_nf(x)| > \varepsilon|S_n\psi(x)| = \varepsilon \cdot \sum_{k=0}^{n-1} \log |(T^k)'(x)| \geq \varepsilon n \log(c_0^2\theta^{n+1}).$$

Choosing  $n_1 = n_1(\varepsilon)$  larger in Theorem 4.1 such that  $\log(c_0^2\theta^{n+1}) \geq 1$  holds for all  $n \geq n_1$ , the result is now an application of Theorem 4.1 since  $\alpha(f) = 0$ .  $\square$

We now give the proof of Theorem 4.1. We start with the following simple lemma which exploits the convexity of the pressure function.

**Lemma 4.3.** *If there exists  $t_0 > 0$  such that  $(-t_0, t_0) \subset \{t \in \mathbb{R} : P(tf + \varphi) < \infty\}$  and  $\alpha \neq \alpha(f)$ , then there exists  $t \in (-t_0, t_0)$  such that  $P(t(f - \alpha) + \varphi) < 0$ .*

*Proof.* For simplicity we assume that  $\alpha < \alpha(f)$  the case where  $\alpha > \alpha(f)$  can be handed analogously. We define the function  $Q : (-t_0, t_0) \rightarrow \mathbb{R}$  by

$$Q(t) = P(tf + \varphi)$$

and note that by Proposition 3.10 this function will be analytic, convex,  $Q'(t) = \int f d\mu_t$  where  $\mu_t$  is the unique Gibbs state for  $tf + \varphi$  and  $Q(0) = 0$  with  $Q'(0) = \alpha(f)$ . Thus there exists  $\alpha_1 \in [\alpha, \alpha(f)]$  and  $-t_0 < t < 0$  such that  $Q'(t) = \alpha_1$ . Now consider the case  $\alpha_1 = \alpha(f)$ . Because  $\alpha < \alpha(f) = \alpha_1$  we have that

$$P(t(f - \alpha) + \varphi) = Q(t) - \alpha t < Q(t) - \alpha_1 t = Q(0) = 0$$

by the convexity of  $Q$ . Moreover, in the case  $\alpha_1 < \alpha(f)$  we have by the convexity of  $Q$  that

$$P(t(f - \alpha) + \varphi) = Q(t) - \alpha t \leq Q(t) - \alpha_1 t < Q(0) = 0.$$

$\square$

*Proof of Theorem 4.1.* By Lemma 4.3, we can fix  $\alpha \neq \alpha(f)$  and  $t \in \mathbb{R}$  such that

$$\delta_1 := -P(t(f - \alpha) + \varphi) > 0.$$

Let us first assume  $\alpha < \alpha(f)$ ; the other case is symmetric. Note that in this case we must have  $t < 0$ . Indeed, the map  $W$  defined by  $W(v) = P(v(f - \alpha) + \varphi)$  is convex by  $\alpha < \alpha(f)$ , the value  $W(0) = P(\varphi) = 0$ , the derivative  $W'(0) = \alpha(f) - \alpha > 0$  and that  $W(t) = -\delta_1 < 0$ . As  $f$  is locally Hölder, we may fix  $n_1 \in \mathbb{N}$  such that for any  $n \geq n_1$ ,  $\mathbf{a} \in \mathbb{N}^n$  and any  $x, y \in I_{\mathbf{a}}$  we have that  $|S_n f(x) - S_n f(y)| \leq \frac{n\delta_1}{4t}$  and also that

$$\left| \frac{1}{n} \log \left( \sum_{\mathbf{a} \in \mathbb{N}^n} \exp(S_n(t(f - \alpha) + \varphi)(\mathbf{a}^\infty)) \right) - P(t(f - \alpha) + \varphi) \right| \leq \frac{\delta_1}{4},$$

by the definition of the pressure, where recall  $\mathbf{a}^\infty \in \mathbb{N}^{\mathbb{N}}$  is the periodic infinite word repeating  $\mathbf{a}$  thought as a point in  $X$ . We also have that by the Gibbs property of  $\mu_\varphi$  we have for all  $n \in \mathbb{N}$  that

$$\mu_\varphi(I_{\mathbf{a}}) \leq C \exp(S_n \varphi(\mathbf{a}^\infty))$$

where  $C$  is the Gibbs constant of  $\mu_\varphi$ . We let

$$\mathcal{C}_n = \{\mathbf{a} \in \mathbb{N}^n : S_n f(x) \leq n\alpha \text{ for some } x \in I_{\mathbf{a}}\}.$$

We can then calculate for  $n \geq n_1$  that

$$\begin{aligned} \mu_\varphi(\{x \in [0, 1] : S_n f(x) \leq n\alpha\}) &\leq \sum_{\mathbf{a} \in \mathcal{C}_n} \mu(I_{\mathbf{a}}) \\ &\leq C \sum_{\mathbf{a} \in \mathcal{C}_n} \exp(S_n \varphi(\mathbf{a}^\infty)) \\ &\leq C e^{n\delta_1/4} \sum_{\mathbf{a} \in \mathcal{C}_n} \exp(S_n(t(f - \alpha) + \varphi)(\mathbf{a}^\infty)) \\ &\leq C e^{n\delta_1/4} \sum_{\mathbf{a} \in \mathbb{N}^n} \exp(S_n(t(f - \alpha) + \varphi)(\mathbf{a}^\infty)) \\ &\leq C e^{n\delta_1/2} \exp(nP(t(f - \alpha) + \varphi)) \\ &= C e^{-n\delta_1/2}. \end{aligned}$$

In the case  $\alpha > \alpha(f)$  we obtain a symmetric computation with  $t > 0$  we use to define  $\delta_1$ . Thus for any  $0 < \varepsilon < |\alpha - \alpha(f)|$ , we obtain

$$\mu_\varphi(\{x \in [0, 1] : |S_n f(x) - n\alpha(f)| > \varepsilon n\}) \leq 2C e^{-n\delta_1/2}$$

which completes Theorem 4.1 as now any  $\delta < \delta_1/2$  is enough for the claim.  $\square$

Moreover, the second assumption of Definition 3.7 can be characterised by the fatness of the tail. This will be shown in the following result, which explains why the assumption (1.2) is necessary and sufficient for  $\varphi$  to satisfy the conditions of parts (1) and (2) of Corollary 4.2

**Proposition 4.4.** *Let  $\varphi : X \rightarrow \mathbb{R}$  be a locally Hölder potential with pressure  $P(\varphi) = 0$ ,  $\mathcal{L}_\varphi 1 = 1$  and  $\varphi \leq 0$  with  $s \in [0, 1]$  the Hausdorff dimension of  $\varphi$ . Then the following are equivalent*

- There exists  $\delta > 0$  such that

$$\mu_\varphi(\{x \in X : a_1(x) \geq n\}) = O(n^{-\delta})$$

- There exists  $t_0 > 0$  such that for all  $-t_0 < t < t_0$  we have that

$$P(t\psi + \varphi) < \infty \text{ and } P(t(\varphi - s\psi) + \varphi) < \infty.$$

*Proof.* Firstly suppose there exists  $\delta > 0$  such that

$$\mu_\varphi(\{x \in X : a_1(x) \geq n\}) = O(n^{-\delta}).$$

Thus by the Gibbs property of  $\mu_\varphi$  we have that

$$\sum_{k=n}^{\infty} e^{\varphi((k)^\infty)} = O(n^{-\delta}),$$

which implies that for any  $0 < t < \delta/2$  we have

$$\sum_{k=1}^{\infty} e^{\varphi((k)^\infty)} k^{2t} < \infty \quad (4.1)$$

and

$$\sum_{k=1}^{\infty} e^{(t+1)(\varphi((k)^\infty)) - st\psi((k)^\infty)} < \infty. \quad (4.2)$$

Since by Lemma 2.5 we have  $k^2 \leq |T'((k)^\infty)| \leq 4k^2$  so  $e^{-st\psi((k)^\infty)} \leq 4^{st} k^{2st} \leq 4^{st} k^{2t}$ . By Proposition 3.6 it follows from (4.1) and (4.2) that for  $t$  sufficiently small

$$P(t\psi + \varphi) < \infty \text{ and } P(t(\varphi - s\psi) + \varphi) < \infty$$

as claimed. For the other direction if for  $t > 0$  sufficiently small  $P(t\psi + \varphi) < \infty$  we have that

$$\sum_{k=1}^{\infty} e^{\varphi((k)^\infty)} k^{2t} < \infty$$

from which it follows that for that

$$\sum_{k=n}^{\infty} e^{\varphi((k)^\infty)} = O(n^{-2t})$$

as  $n \rightarrow \infty$ . Indeed if  $e_k := e^{\varphi((k)^\infty)}$  and  $c := \sum_{k=1}^{\infty} e_k k^{2t} < \infty$ , then for all  $n \in \mathbb{N}$  we have  $\sum_{k=n}^{\infty} e_k k^{2t} \leq c$  and so as  $t > 0$  we have

$$\sum_{k=n}^{\infty} e_k \leq \sum_{k=n}^{\infty} e_k k^{2t} n^{-2t} \leq c n^{-2t}.$$

Thus by the Gibbs property of  $\mu_\varphi$  we have

$$\mu_\varphi(\{x \in X : a_1(x) \geq n\}) \leq \sum_{k=n}^{\infty} C e^{\varphi((k)^\infty)} = O(n^{-2t})$$

and the result follows. One can observe that in this direction of the equivalence we do not need the second condition on the pressure.  $\square$

*Remark 4.5.* It should be noted that if  $\mu_\varphi$  did not satisfy condition (1.2) then the large deviation result could not hold. In this case we would have that

$$\limsup_{n \rightarrow \infty} n^\delta \mu_\varphi(\{x \in X : a_1(x) \geq n\}) = \infty$$

then as  $e^{\varphi(x)}$  is at least  $4a_1(x)^2$  by Lemma 2.5 we have that Corollary 4.2 could not hold as  $\mu_\varphi(\{x \in X : |S_n \psi(x) + n\lambda| \geq n\varepsilon\})$  would decay subexponentially.

*Remark 4.6.* We also remark that in the proofs of the results in this section, such as Theorem 4.1, we do not use any specific properties of the Gauss map itself, only the symbolic coding with the countable Markov shift, so the results could be extended to other Markov maps with similar coding.

## 5. DECOMPOSITION OF THE GIBBS MEASURE

In this section we provide an important decomposition of the Gibbs measure on each generation  $n$ . This will be done by creating cylinders from the ‘good sets’ for the large deviations, that is, where we can control the exponential decay of  $|T'_{\mathbf{a}}(x)|$  and the measure  $\mu(I_{\mathbf{a}})$  with respect to the length  $|I_{\mathbf{a}}|$ . Thanks to the large deviation bounds, the exceptional sets to these regular parts is exponentially small. Fix  $\varepsilon > 0$  from now on and suppress it from the notation. We will only make assumptions on  $\varepsilon$  when dealing with the proof of Theorem 1.3 in Section 6.

We begin by choosing  $n_0$  sufficiently large such that the large deviation bounds in Corollary 4.2 and that some technical conditions are true. It should be noted that we keep the Gibbs measure  $\mu$  and its potential  $\varphi$  fixed.

**Choice of  $n_0$ .** Choose an *even* number  $n_0 \in \mathbb{N}$  such that the following are true

- (1) If  $n_1 = n_1(\varepsilon/2)$  is the threshold from large deviations result Corollary 4.2 with  $\varepsilon/2$  in the place of  $\varepsilon$ , then

$$n_0/2 > n_1. \quad (5.1)$$

- (2) If  $C$  is the Gibbs constant for  $\mu$ , and  $\theta$  is the Golden ratio, and  $c_0$  is the constant from Lemma 2.3, we assume

$$\frac{\log 4}{n_0/2} < \varepsilon/2 \quad \text{and} \quad \frac{\log 4C^2}{\log(c_0^2\theta^{n_0})} < \varepsilon/2. \quad (5.2)$$

- (3) If  $\delta = \delta(\varepsilon/2) > 0$  is the rate function for the large deviations Corollary 4.2, then

$$\frac{e^{-\delta n_0/2}}{1 - e^{-\delta}} < e^{-\delta n_0/4}. \quad (5.3)$$

- (4) If  $\lambda$  is the Lyapunov exponent of  $\mu$ , then

$$e^{n_0/2} \geq 2\pi e^{\lambda+2\lambda\varepsilon}, \quad e^{\varepsilon\lambda n_0} \geq 2C\pi e^{(1+2s)\lambda} \quad \text{and} \quad \frac{1}{192} e^{(\frac{\lambda}{2}-2\varepsilon)n_0} \geq 1. \quad (5.4)$$

The reason for (1) in this choice is that we will want to obtain large deviation bounds at the halved generation  $n/2$  for some  $n \geq n_0$ .

**Construction** (Regular intervals). We will now construct a collection  $\mathcal{R}_n$  of words  $\mathbf{a} \in \mathbb{N}^n$  where we have good control on the growth of the continuants  $q_n(\mathbf{a})$  and the weights  $\mu(I_{\mathbf{a}})$ . With the definition of  $n_0$  in mind, for  $n \geq n_0$  write the  $n$ -regular set for large deviations

$$A_n(\varepsilon) := \left\{ x \in [0, 1] : \left| \frac{1}{n} S_n \psi(x) + \lambda \right| < \varepsilon, \left| \frac{S_n \varphi(x)}{S_n \psi(x)} - s \right| < \varepsilon \right\}$$

and define the collection of  $n$ -regular words by

$$\mathcal{R}_n := \bigcap_{k=\lfloor n/2 \rfloor}^n \{ \mathbf{a} \in \mathbb{N}^n : I_{\mathbf{a}|_k} \subset A_k(\varepsilon) \}$$

and let  $R_n \subset [0, 1]$  be the corresponding union of intervals in  $\mathcal{R}_n$ .

The reason to take intersections in the construction is to have the  $n$ -regular intervals from  $\mathcal{R}_n$  nested in the sense that if  $\mathbf{a} \in \mathcal{R}_n$ , then we know the ‘ $k$ -regularity’ for the interval  $I_{\mathbf{a}|_k}$  for smaller  $k$  up to  $n/2$ . This space is useful when dealing with large intervals for example in the proof of Theorem 1.3.

The advantage of large deviations for Lyapunov exponents and Hausdorff dimension is that we obtain bounds for the continuants  $q_n(\mathbf{a})$  and the weights  $w_{\mathbf{a}}(x) = e^{S_n \varphi(T_{\mathbf{a}}(x))}$  in the terms of  $e^{\lambda n}$  when  $\mathbf{a} \in \mathcal{R}_n$ . The following lemma summarises all the comparisons we have:

**Lemma 5.1.** *Fix  $n \geq n_0$ ,  $\mathbf{a} \in \mathcal{R}_n$  and  $\lfloor n/2 \rfloor \leq k \leq n$ . Then the following comparisons hold*



(1) the continuant  $q_k(\mathbf{a})$  satisfies

$$e^{(\lambda-\varepsilon)k} \leq q_k(\mathbf{a})^2 \leq 4e^{(\lambda+\varepsilon)k};$$

(2) the length  $|I_{\mathbf{a}|_k}|$  and the derivative  $T'_{\mathbf{a}|_k}$  satisfies

$$\frac{1}{16}e^{(-\lambda-\varepsilon)k} \leq |I_{\mathbf{a}|_k}| \leq e^{(-\lambda+\varepsilon)k} \quad \text{and} \quad \frac{1}{16}e^{(-\lambda-\varepsilon)k} \leq |T'_{\mathbf{a}|_k}| \leq e^{(-\lambda+\varepsilon)k};$$

(3) if  $x \in [0, 1]$ , then the weight  $w_{\mathbf{a}|_k}(x) = e^{S_k \varphi(T_{\mathbf{a}|_k}(x))}$  satisfies

$$e^{(-s\lambda-3\lambda\varepsilon)k} \leq w_{\mathbf{a}|_k}(x) \leq e^{(-s\lambda+3\lambda\varepsilon)k};$$

(4) if  $C \geq 1$  is the Gibbs constant of  $\mu$ , then the measure  $\mu(I_{\mathbf{a}|_k})$  satisfies

$$C^{-1}e^{(-s\lambda-3\lambda\varepsilon)k} \leq \mu(I_{\mathbf{a}|_k}) \leq Ce^{(-s\lambda+3\lambda\varepsilon)k}.$$

*Proof.* We do this just in the case  $k = n$  as the proof is the same for the other cases. Fix  $x \in [0, 1]$  and  $\mathbf{a} \in \mathbb{N}^n$ . Then by the chain rule

$$\log |T'_{\mathbf{a}}(x)| = \sum_{j=0}^{n-1} \log \frac{1}{|T'(T^j(T_{\mathbf{a}}(x)))|} = S_n \psi(T_{\mathbf{a}}(x)). \quad (5.5)$$

Lemma 2.5 gives  $\frac{1}{4}q_n(\mathbf{a})^{-2} \leq |T'_{\mathbf{a}}(x)| \leq q_n(\mathbf{a})^{-2}$  so

$$e^{S_n \psi(T_{\mathbf{a}}(x))} \leq q_n(\mathbf{a})^{-2} \leq 4e^{S_n \psi(T_{\mathbf{a}}(x))}. \quad (5.6)$$

If  $T_{\mathbf{a}}(x) \in A_n(\varepsilon)$ , then

$$e^{(-\lambda-\varepsilon)n} \leq e^{S_n \psi(T_{\mathbf{a}}(x))} \leq e^{(-\lambda+\varepsilon)n}.$$

Thus by (5.6), we obtain

$$e^{(\lambda-\varepsilon)n} \leq q_n(\mathbf{a})^2 \leq 4e^{(\lambda+\varepsilon)n},$$

which proves (1). Now (2) follows when recalling the distortion bounds for the lengths  $|I_{\mathbf{a}|_k}|$  and  $T'_{\mathbf{a}|_k}$  from Lemma 2.5. Moreover, as  $S_n \psi(T_{\mathbf{a}}(x)) \leq 0$  we have

$$e^{(s+\varepsilon)S_n \psi(T_{\mathbf{a}}(x))} \leq e^{S_n \varphi(T_{\mathbf{a}}(x))} \leq e^{(s-\varepsilon)S_n \psi(T_{\mathbf{a}}(x))}.$$

Thus by using crude bounds  $\varepsilon, s \leq 1$  and  $\lambda > 1$  we obtain

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

which is (3). This also gives (4) by the Gibbs property that links  $\mu(I_{\mathbf{a}})$  to  $w_{\mathbf{a}}(x)$ .  $\square$

As we are intersecting with sets  $A_k(\varepsilon)$  whose  $\mu$  measure is exponentially close to one, we do not lose much in the measure:

**Lemma 5.2.** *For all  $n \geq n_0$  we have*

$$\mu([0, 1] \setminus R_n) \leq e^{-\delta n/4}$$

where  $\delta = \delta(\varepsilon/2) > 0$  is the rate function from the large deviations with  $\varepsilon/2$ .

*Proof.* We wish to show that

$$\bigcap_{k=\lfloor n/2 \rfloor}^n A_k(\varepsilon/2) \subset R_n. \quad (5.7)$$

This is enough for us as it yields by the geometric series formula that

$$\mu([0, 1] \setminus R_n) \leq \sum_{k=\lfloor n/2 \rfloor}^n \mu([0, 1] \setminus A_k(\varepsilon/2)) \leq \frac{e^{-\delta n/2}}{1 - e^{-\delta}} < e^{-\delta n/4},$$

where  $\delta := \delta(\varepsilon/2)/2$  since  $\lfloor n/2 \rfloor \geq n_1(\varepsilon/2)$  by (5.1) of the choice of  $n_0$  since  $n_0$  is even.

To prove (5.7) let  $\mathbf{a} \in \mathbb{N}^n$  be a word such that  $T_{\mathbf{a}}(x) \in A_k(\varepsilon/2)$  for all  $k \in \lfloor n/2 \rfloor, \dots, n$ . We wish to prove that  $I_{\mathbf{a}|_k} \subset A_k(\varepsilon)$  for all these  $k$ . Fix  $k$  and a point  $y \in [0, 1]$  and let us prove that  $T_{\mathbf{a}|_k}(y) \in A_k(\varepsilon)$ . Since  $T_{\mathbf{a}}(x) \in A_k(\varepsilon/2)$  we have by Lemma 2.5 that

$$\left| \frac{1}{k} S_k \psi(T_{\mathbf{a}|_k}(y)) + \lambda \right| < \varepsilon/2 + \frac{\log 4}{k} < \varepsilon$$

and for the Gibbs constant  $C \geq 1$  in addition the Lemma 2.3 shows that

$$\left| \frac{S_k \varphi(T_{\mathbf{a}|_k}(y))}{S_k \psi(T_{\mathbf{a}|_k}(y))} - s \right| < \varepsilon/2 + \frac{\log 4C^2}{|S_k \psi(T_{\mathbf{a}|_k}(y))|} \leq \varepsilon/2 + \frac{\log 4C^2}{\log(c_0^2 \theta^{2k})} < \varepsilon$$

by the choice of  $n_0$  (the property (5.2)). More precisely, we used the triangle inequality after applying (5.5) to obtain  $S_k \psi(T_{\mathbf{a}}(x)) = \log |T'_{\mathbf{a}|_k}(z)|$ , where  $z = T_{\sigma^k \mathbf{a}}(x)$  and  $\sigma$  is the shift, which yields

$$S_k \psi(T_{\mathbf{a}|_k}(y)) - S_k \psi(T_{\mathbf{a}}(x)) = \log \frac{|T'_{\mathbf{a}|_k}(y)|}{|T'_{\mathbf{a}|_k}(z)|}$$

and finishing the proof with Lemma 2.5 which allows us to compare the derivatives at points  $y$  and  $z$ . Thus  $I_{\mathbf{a}|_k} \subset A_k(\varepsilon)$  and we are done as  $k = \lfloor n/2 \rfloor, \dots, n$  is arbitrary.  $\square$

## 6. PROOF OF THE MAIN RESULT

**6.1. Overview.** We will now give some heuristics and general strategy we follow during the proof.

- (1) We first choose a natural number  $n \geq n_0$  so large that  $|\xi|^{-1}$  is approximately the exponential number  $e^{-c\lambda n}$  for some constant  $c > 0$ , where  $\lambda$  is the Lyapunov exponent of the Gibbs measure  $\mu = \mu_\varphi$ . Then the proof reduces to verify that the Fourier transform  $|\widehat{\mu}(\xi)|$  has an exponential bound  $e^{-\eta c \lambda n}$  for some constant  $\eta > 0$ .
- (2) The first step is to use the invariance of the Gibbs measure  $\mu = \mu_\varphi$  under the dual operator  $\mathcal{L}_\varphi^*$  and apply the Ruelle transfer operator  $n$  times to the oscillation  $x \mapsto e^{-2\pi i \xi x}$ . This, by the definition of the transfer operator, yields that  $\widehat{\mu}(\xi)$  becomes a summation over  $n$ th pre-images of  $T$ , that is, a summation over the words  $\mathbf{a} \in \mathbb{N}^n$  of certain oscillative  $\mu$  integrals.
- (3) This is the point where we invoke the decomposition of the Gibbs measure to the regular and irregular part using the  $n$ -regular cylinders  $\mathcal{R}_n$  given by the large deviations. Since not much  $\mu$  mass enters outside of  $\mathcal{R}_n$ , we obtain an exponential decay  $e^{-\delta n/2}$  for that part from the large deviations and Lemma 5.2; see Section 6.3 below. Thus we are left to just study the summation over the regular words  $\mathbf{a} \in \mathcal{R}_n$  of oscillative  $\mu$  integrals.
- (4) We then exploit the Hölder bounds for the measure  $\mu(I_{\mathbf{a}})$  on regular intervals  $I_{\mathbf{a}}$  and the decay rate for  $T'_{\mathbf{a}}(x)$  to transfer the estimates from the oscillative  $\mu$  integrals for the Gibbs measure to oscillative  $L^2$  Lebesgue integrals. The price we pay is that there is an exponentially increasing term fighting against the Lebesgue integrals, so we need a rapid enough decay to counter this multiplicative error.
- (5) When we deal with the Lebesgue measure and  $n$ -regular words  $\mathbf{a}$  for which we can control the growth of the continuants  $q_n(\mathbf{a})$ , we can then rely on *stationary phase* type inequalities that Kaufman also used, which, when invoking the rapid expansive nature of the Gauss map, yield an exponential decay bound  $e^{-\lambda n/2}$  in our case. This decays rapidly enough for us to counter the error produced in the step (4).

**6.2. Fixing parameters.** We begin the proof by fixing the parameters  $\varepsilon > 0$  for large deviations and the number  $n \geq n_0$  related to  $|\xi|^{-1}$ , recall choice of  $n_0$  in the beginning of Section 5. We will invoke the convention  $f(x) \ll g(x)$  if there is a constant independent of  $x$  such that  $f(x) \leq Cg(x)$ .

Let  $\mu = \mu_\varphi$  be Gibbs measure with entropy  $h = h_\mu$ , Lyapunov exponent  $\lambda = \lambda_\mu$  and Hausdorff-dimension  $s = h/\lambda = \dim \mu > 1/2$ . Then the number

$$\eta_s := \frac{2s^2 - s}{(4-s)(1+2s)} > 0.$$

Let us fix  $\varepsilon > 0$  such that

$$\varepsilon < 1/26 \quad \text{and} \quad \eta_s - 38\varepsilon > 0. \quad (6.1)$$

This now fixes  $n_0 = n_0(\varepsilon) \in \mathbb{N}$  which we defined in the Section 5 before decomposing the Gibbs measure.

To prove the main Theorem 1.3 it is enough to show that

$$|\widehat{\mu}(\xi)| \ll |\xi|^{-\eta_s + 38\varepsilon} + |\xi|^{-\frac{\delta}{12\lambda}}, \quad (6.2)$$

where  $\delta = \delta(\varepsilon/2) > 0$  is the exponential rate from the large deviations with  $\varepsilon/2$ . Depending on how small  $\varepsilon$  is, one of the terms in 6.2 will dominate the other but in both cases we obtain polynomial decay for the Fourier transform.

Let us write for the rest of the paper  $u := |\xi|$ . Assume  $u$  is initially so large that we can choose  $n \geq n_0$  such that

$$e^{(1+2s)\lambda n} \leq u < e^{(1+2s)\lambda(n+1)}. \quad (6.3)$$

**6.3. Splitting the transfer operator.** Define  $h : [0, 1] \rightarrow \mathbb{C}$  by  $h(y) = \exp(-2\pi i \xi y)$ . Since  $\mu$  is a Gibbs measure with potential  $\varphi$ , we have that  $\mu$  satisfies invariance condition  $\mathcal{L}_\varphi^* \mu = \mu$  and if applied  $n$  times to  $h$  gives

$$\widehat{\mu}(\xi) = \int \mathcal{L}_\varphi^n h(x) d\mu(x).$$

By the definition of the transfer operator

$$\mathcal{L}_\varphi^n h(x) = \sum_{\mathbf{a} \in \mathbb{N}^n} w_{\mathbf{a}}(x) \exp(-2\pi i \xi T_{\mathbf{a}}(x))$$

where, recall, the weight  $w_{\mathbf{a}}(x) := e^{S_n \varphi(T_{\mathbf{a}}(x))}$ . Now decompose the summation

$$\mathcal{L}_\varphi^n h(x) = f(x) + g(x)$$

to the *regular part*  $f(x)$  and the *irregular part*  $g(x)$  according to  $\mathcal{R}_n$  as follows

$$f(x) = \sum_{\mathbf{a} \in \mathcal{R}_n} w_{\mathbf{a}}(x) \exp(-2\pi i \xi T_{\mathbf{a}}(x)) \quad \text{and} \quad g(x) = \sum_{\mathbf{a} \in \mathbb{N}^n \setminus \mathcal{R}_n} w_{\mathbf{a}}(x) \exp(-2\pi i \xi T_{\mathbf{a}}(x)). \quad (6.4)$$

Hence in order to estimate  $|\widehat{\mu}(\xi)|$  we need to estimate  $\int |f| d\mu$  and  $\int |g| d\mu$ . Both terms will contribute to the decay rate of the Fourier transform, depending on the value  $\varepsilon > 0$ . For the irregular part, we can immediately use the measure bound from Lemma 5.2. Indeed since  $e^{(1+2s)\lambda n} \leq u$  and by the Gibbs property of  $\mu$  we have  $w_{\mathbf{a}}(x) \leq C\mu(I_{\mathbf{a}})$  so inputting  $e^{(1+2s)\lambda n} \leq u$  from (6.3) we have

$$\int |g(x)| d\mu(x) \leq C\mu([0, 1] \setminus \mathcal{R}_n) \leq Ce^{-\delta n/4} \ll u^{-\frac{1}{4(1+2s)\lambda}\delta} \ll u^{-\frac{\delta}{12\lambda}} \quad (6.5)$$

since  $\lambda > 1$  and  $s \leq 1$ . For the regular part, on the other hand, we will obtain the bound

$$\int |f(x)| d\mu(x) \ll u^{-\eta_s + 38\varepsilon} + u^{-\frac{\delta}{12\lambda}} \quad (6.6)$$

with the same term on the right. Since  $u = |\xi|$ , this yields the desired decay and Theorem 1.3. The proof of this bound is now split into the Sections 6.4, 6.5 and 6.6 below.

**6.4. From Gibbs to Lebesgue.** To estimate the oscillative integral  $\int |f(x)| d\mu(x)$  we rely on a method to transfer the problem on integrating over the Gibbs measure  $\mu$  back to Lebesgue measure. In the case of discrete sums of point masses, this kind of approach reduces to *large sieve inequalities*, see for example inequality (4) in Baker's paper [1]. The analytic version used for example by Baker requires classically Hölder bounds from the original measure  $\mu$ , see for example the inequalities used by Kaufman [12, Section 5]. In our case, Gibbs measures do not have Hölder bounds but on the  $n$ -regular cylinders we do have control for them. Thus we will give the following version using the  $L^2$  norm  $\|f\|_2$  with respect to Lebesgue measure. Note that here the proof applies to any  $C^1$  function  $f$  instead of the specific  $f$  we consider:

**Lemma 6.1.** *Suppose we are given numbers  $\alpha, \beta > 0$  such that  $\gamma = \alpha + \beta \geq 1$  and let  $f : [0, 1] \rightarrow \mathbb{C}$  be  $C^1$  with  $\|f\|_\infty \leq 1$  and  $\|f'\|_\infty \leq e^{\alpha\lambda n}$ . Then*

$$\int |f(x)| d\mu(x) \ll e^{(-\beta\lambda + \varepsilon\gamma)n} + \|f\|_2^2 e^{((1-s)\alpha\lambda + (3-s)\beta\lambda + 3\lambda\varepsilon\gamma)n} + e^{-\delta\gamma n/4}.$$

*Proof.* Write  $m = \lfloor \gamma n \rfloor$  and

$$\mathcal{R}'_m = \{\mathbf{a} \in \mathcal{R}_m : \sup\{|f(x)| : x \in I_{\mathbf{a}}\} \geq 2e^{\alpha\lambda n} e^{(-\lambda + \varepsilon)m}\}.$$

By the condition  $\alpha + \beta \geq 1$  and  $n \geq n_0$ , we have  $m \geq n_0$  so we may apply all the large deviation theory for the words in  $\mathcal{R}'_m$ . Let  $\mathbf{a} \in \mathcal{R}'_m$ . Since  $f$  is  $C^1$  and by Lemma 5.1(2) we have  $|I_{\mathbf{a}}| \leq e^{(-\lambda + \varepsilon)m}$  then by the mean value theorem

$$|f(x)| \geq 2e^{\alpha\lambda n} e^{(-\lambda + \varepsilon)m} - \|f'\|_\infty |I_{\mathbf{a}}| \geq e^{\alpha\lambda n} e^{(-\lambda + \varepsilon)m}$$

for all  $x \in I_{\mathbf{a}}$ . Thus by Lemma 5.1(2) we have

$$\begin{aligned} \int |f(x)|^2 dx &\geq \sum_{\mathbf{a} \in \mathcal{R}'_m} e^{2\alpha\lambda n + 2(-\lambda + \varepsilon)m} |I_{\mathbf{a}}| \\ &\geq |\mathcal{R}'_m| \cdot e^{2\alpha\lambda n + (-2\lambda + 2\varepsilon)m} \cdot \frac{1}{16} e^{(-\lambda - \varepsilon)m} \\ &\geq \frac{1}{16} |\mathcal{R}'_m| \cdot e^{2\alpha\lambda n + (-3\lambda + \varepsilon)m}. \end{aligned}$$

This yields by Lemma 5.1(4) and  $\|f\|_\infty \leq 1$  that for the Gibbs constant  $C > 0$  the integral

$$\begin{aligned} \int_{R'_m} |f(x)| d\mu(x) &\leq \sum_{\mathbf{a} \in \mathcal{R}'_m} \mu(I_{\mathbf{a}}) \leq 16C \|f\|_2^2 e^{-2\alpha\lambda n + (3\lambda - \varepsilon)m} e^{(-s\lambda + 3\lambda\varepsilon)m} \\ &\leq 16C \|f\|_2^2 e^{-2\alpha\lambda n + (3-s)\lambda m + 3\lambda\varepsilon m} \\ &\ll \|f\|_2^2 e^{(-2\alpha\lambda + (3-s)\lambda\gamma + 3\lambda\varepsilon\gamma)n} \\ &= \|f\|_2^2 e^{((1-s)\alpha\lambda + (3-s)\beta\lambda + 3\lambda\varepsilon\gamma)n}. \end{aligned}$$

Moreover, by the definition of  $\mathcal{R}'_m$  we have as  $\alpha + \beta = \gamma$  that

$$\int_{R_m \setminus R'_m} |f(x)| d\mu(x) \leq 2e^{\alpha\lambda n + (-\lambda + \varepsilon)m} \ll e^{(-\beta\lambda + \varepsilon\gamma)n}.$$

Finally, the remaining part we use  $\|f\|_\infty \leq 1$  to obtain a measure bound

$$\int_{[0,1] \setminus R_m} |f(x)| d\mu(x) \leq \mu([0,1] \setminus R_m) \leq e^{-\delta m/4} \ll e^{-\delta\gamma n/4}$$

by Lemma 5.2. □

We apply Lemma 6.1 for the regular part  $f(x) = \sum_{\mathbf{a} \in \mathcal{R}_n} w_{\mathbf{a}}(x) \exp(-2\pi i \xi T_{\mathbf{a}}(x))$ , recall (6.4). The first term of the bound in Lemma 6.1 already decays exponentially as  $n \rightarrow \infty$ . However, the larger  $\beta > 0$  we choose, the faster the exponential term in the front of  $\|f\|_2^2$  grows. Thus we need to find a suitable choice of  $\beta$  so the growth will not terminate the decay of  $\|f\|_2^2$  as  $n \rightarrow \infty$ . The number  $\alpha$  we will fix from the following

**Lemma 6.2.** *For the function  $f$  defined in (6.4) we have*

$$\|f'\|_\infty \leq e^{\alpha\lambda n}$$

with  $\alpha := 2s + 2\varepsilon$ .

*Proof.* Differentiating term-wise, we obtain

$$|f'(x)| \leq 2\pi u \sum_{\mathbf{a} \in \mathcal{R}_n} w_{\mathbf{a}}(x) |T'_{\mathbf{a}}(x)|.$$

For  $\mathbf{a} \in \mathcal{R}_n$ , we have by Lemma 5.1(2) that  $|T'_{\mathbf{a}}(x)| \leq e^{(-\lambda+\varepsilon)n}$ . Now inputting  $u < e^{(1+2s)\lambda(n+1)}$  from (6.3), the Gibbs property  $w_{\mathbf{a}}(x) \leq C\mu(I_{\mathbf{a}})$  and  $\lambda > 1$  we obtain

$$|f'(x)| \leq 2C\pi u e^{(-\lambda+\varepsilon)n} \sum_{\mathbf{a} \in \mathbb{N}} \mu(I_{\mathbf{a}}) \leq 2C\pi e^{(1+2s)\lambda} e^{(2s\lambda+\varepsilon\lambda)n} \leq e^{\alpha\lambda n}$$

since by the choice (5.4) of  $n_0$  we have  $2C\pi e^{(1+2s)\lambda} \leq e^{\varepsilon\lambda n_0}$  (recall  $n \geq n_0$ ).  $\square$

**6.5. Bounding the  $L^2$  norm.** Thanks to Lemma 6.1, we are now reduced to just bound the  $L^2$  norm of  $f$ . We dedicate this section to prove the following

**Proposition 6.3.** *For the function  $f$  defined in (6.4) we have*

$$\|f\|_2^2 \ll u^{-1/2} e^{(\frac{\lambda}{2}+15\lambda\varepsilon)n} + e^{(-s\lambda+3\lambda\varepsilon)n} \quad (6.7)$$

This is the part where we use heavily the nonlinear nature of the Gauss map and much of the number theoretical tools in continued fractions appear. This is also the part where we adapt the mirroring argument presented by Queffélec and Ramaré, which allows us to only assume that the Hausdorff dimension of  $\mu$  is at least  $1/2$  in contrast to Kaufman's bound  $2/3$ . The main idea of bounding the  $L^2$  norm is to study the distribution of the differences  $T_{\mathbf{a}}(x) - T_{\mathbf{b}}(x)$  for different  $n$ -regular words  $\mathbf{a}$  and  $\mathbf{b}$  and exploit *stationary phase* type inequalities (see Section 6 of [34]) to bound oscillative Lebesgue integrals in the terms of derivatives of the oscillations that are under control for words in  $\mathcal{R}_n$ . The key lemma is the following modification of the lemma used by Kaufman in [12].

**Lemma 6.4** (Stationary phase). *Let  $\psi : [0, 1] \rightarrow \mathbb{R}$  be a  $C^2$  map such that the derivative has the form*

$$\psi'(x) = \varphi(x) \cdot (\alpha_1 x + \alpha_2)$$

for some  $C^1$  map  $\varphi : [0, 1] \rightarrow \mathbb{R}$  with  $a \leq |\varphi| \leq b$  and  $|\varphi'| \leq b$ , where  $b > a > 0$ .

(1) *If  $\alpha_1 \neq 0$ , then*

$$\left| \int e^{2\pi i \psi(x)} dx \right| \leq 6ba^{-3/2} |\alpha_1|^{-1/2}.$$

(2) *If  $|\alpha_1| \leq |\alpha_2|/2$  and  $\alpha_2 \neq 0$ , then*

$$\left| \int e^{2\pi i \psi(x)} dx \right| \leq 8|\alpha_2|^{-1} (a^{-1} + ba^{-2}).$$

*Proof.* The proofs were given in Section 4 of [12] and they follow immediately from partial integration since we know the growth bounds for  $\psi$ . However, the inequality presented in [12] for the case (2) has a slightly different form. There, precisely speaking, it was proved that for any  $C^2$  map  $\psi : [0, 1] \rightarrow \mathbb{R}$  with  $|\psi'| \geq r$  and  $|\psi''| \leq R$  we have

$$\left| \int e^{2\pi i \psi(x)} dx \right| \leq r^{-1} + Rr^{-2}.$$

However, this yields immediately the case (2) since by the assumptions on  $\varphi$  and  $|\alpha_1| \leq |\alpha_2|/2$  the derivatives

$$|\psi'(x)| = |\varphi(x)| |\alpha_1 x + \alpha_2| \geq a|\alpha_2|/2$$

and

$$|\psi''(x)| = |\varphi'(x)| (|\alpha_1| + |\alpha_2|) + |\varphi(x)| |\alpha_1| \leq 2b|\alpha_2|,$$

which gives the claim with  $r = a|\alpha_2|/2$  and  $R = 2b|\alpha_2|$ .  $\square$

Now, let us look at the proof of Proposition 6.3, and how we can reduce the setting to a situation where we can apply the stationary phase. Given  $\mathbf{a}, \mathbf{b} \in \mathcal{R}_n$ , define the  $C^2$  mapping  $\psi_{\mathbf{a}, \mathbf{b}} : [0, 1] \rightarrow \mathbb{R}$  at  $x \in [0, 1]$  by

$$\psi_{\mathbf{a}, \mathbf{b}}(x) = \xi(T_{\mathbf{a}}(x) - T_{\mathbf{b}}(x)).$$

After expanding the square, we obtain

$$|f(x)|^2 = \sum_{\mathbf{a}, \mathbf{b} \in \mathcal{R}_n} w_{\mathbf{a}}(x)w_{\mathbf{b}}(x)e^{2\pi i\psi_{\mathbf{a}, \mathbf{b}}(x)},$$

where we have that  $w_{\mathbf{a}}(x) = e^{S_n\varphi(T_{\mathbf{a}}(x))}$ . Writing

$$I(\mathbf{a}, \mathbf{b}) := \left| \int \exp(2\pi i\psi_{\mathbf{a}, \mathbf{b}}(x)) dx \right| \quad \text{and} \quad S(\mathbf{a}) := \sum_{\mathbf{b} \in \mathcal{R}_n} \mu(I_{\mathbf{b}})I(\mathbf{a}, \mathbf{b}),$$

the Gibbs property of  $\mu$  yields

$$\int |f(x)|^2 dx \ll \sum_{\mathbf{a}, \mathbf{b} \in \mathcal{R}_n} \mu(I_{\mathbf{a}})\mu(I_{\mathbf{b}})I(\mathbf{a}, \mathbf{b}) = \sum_{\mathbf{a} \in \mathcal{R}_n} \mu(I_{\mathbf{a}})S(\mathbf{a}). \quad (6.8)$$

Thus we end up estimating the integrals  $I(\mathbf{a}, \mathbf{b})$  for possible choices of  $\mathbf{a}$  and  $\mathbf{b}$ . For this purpose, we denote the difference of the  $n$ th continuants by

$$d_n(\mathbf{a}, \mathbf{b}) := |q_n(\mathbf{a}) - q_n(\mathbf{b})|.$$

Similarly, define  $d_{n-1}$  using the  $(n-1)$ th continuants of  $\mathbf{a}$  and  $\mathbf{b}$ . Notice that these differences attain only integer values. Then with the stationary phase we obtain a relationship between  $I(\mathbf{a}, \mathbf{b})$  and the differences  $d_n$  and  $d_{n-1}$  in the following key lemma:

**Lemma 6.5.** *Let  $\mathbf{a}, \mathbf{b} \in \mathcal{R}_n$ . If  $d_{n-1}(\mathbf{a}, \mathbf{b}) \neq 0$ , then*

$$I(\mathbf{a}, \mathbf{b}) \ll \frac{u^{-1/2}e^{(3\lambda/4+7\epsilon)n}}{d_{n-1}(\mathbf{a}, \mathbf{b})^{1/2}}. \quad (6.9)$$

and if  $d_{n-1}(\mathbf{a}, \mathbf{b}) \leq d_n(\mathbf{a}, \mathbf{b})/2$  and  $d_n(\mathbf{a}, \mathbf{b}) \neq 0$  we have

$$I(\mathbf{a}, \mathbf{b}) \ll \frac{u^{-1/2}e^{(3\lambda/4+7\epsilon)n}}{d_n(\mathbf{a}, \mathbf{b})^{1/2}}. \quad (6.10)$$

*Proof.* Fix  $\mathbf{a}, \mathbf{b} \in \mathcal{R}_n$  and write  $\psi = \psi_{\mathbf{a}, \mathbf{b}}$  as defined above. Recall that the derivative of  $T_{\mathbf{a}}$  has the form

$$T'_{\mathbf{a}}(x) = \frac{(-1)^n}{(q_{n-1}(\mathbf{a})x + q_n(\mathbf{a}))^2}.$$

so writing  $\alpha_1 := q_{n-1}(\mathbf{a}) - q_{n-1}(\mathbf{b})$ ,  $\alpha_2 := q_n(\mathbf{a}) - q_n(\mathbf{b})$  and

$$\varphi(x) := (-1)^n \xi \cdot \frac{(q_{n-1}(\mathbf{a}) + q_{n-1}(\mathbf{b}))x + q_n(\mathbf{a}) + q_n(\mathbf{b})}{(q_{n-1}(\mathbf{a})x + q_n(\mathbf{a}))^2 (q_{n-1}(\mathbf{b})x + q_n(\mathbf{b}))^2} \quad (6.11)$$

we obtain  $\psi'(x) = \varphi(x) \cdot (\alpha_1 x + \alpha_2)$ . Thus we will apply Lemma 6.4 using bounds for  $|\varphi|$  and  $|\varphi'|$ . Since we have  $q_{n-1}(\mathbf{a}) \leq q_n(\mathbf{a})$  and

$$e^{(\lambda/2-\epsilon/2)n} \leq q_n(\mathbf{a}) \leq 4e^{(\lambda/2+\epsilon/2)n}$$

and the same bounds for the continuant  $q_n(\mathbf{b})$ , we obtain the bound

$$a := \frac{1}{8}ue^{(-3\lambda-5\epsilon)n/2} \leq |\varphi(x)| \leq 16ue^{(-3\lambda+5\epsilon)n/2} \leq 40ue^{(-3\lambda+5\epsilon)n/2} := b.$$

Differentiating again (6.11), we obtain as well

$$|\varphi'(x)| \leq b.$$

Note that here we did the crude bound for  $|\varphi(x)|$  above as we wanted the same bound with  $b$  as for  $\varphi'$  to use Lemma 6.4. Then  $0 < a < b$  and moreover

$$ba^{-3/2} \ll u^{-1/2} e^{(3\lambda+25\varepsilon)n/4}.$$

Recall that by the choice (6.3) of  $n$ , we have  $e^{(1+2s)\lambda n} \leq u$ . Moreover,  $\lambda > 1$  and  $s > 1/2$  so

$$u^{-1} e^{(3\lambda+25\varepsilon)n/2} \leq e^{(\frac{1}{2}-2s+13\varepsilon)\lambda n} \leq e^{(-\frac{1}{2}+13\varepsilon)\lambda n} < 1$$

since, recall we chose initially  $\varepsilon < 1/26$  in (6.1). Thus by taking square roots, we have the same bounds for  $ba^{-2}$  and  $a^{-1}$  as well:

$$ba^{-2} \ll u^{-1} e^{(3\lambda+15\varepsilon)n/2} \leq u^{-1/2} e^{(3\lambda+25\varepsilon)n/4} < u^{-1/2} e^{(3\lambda/4+7\varepsilon)n}$$

and

$$a^{-1} \ll u^{-1} e^{(3\lambda+5\varepsilon)n/2} \leq u^{-1/2} e^{(3\lambda+25\varepsilon)n/4} < u^{-1/2} e^{(3\lambda/4+7\varepsilon)n}.$$

An application of the case (1) of Lemma 6.4 with  $\alpha_1 = q_{n-1}(\mathbf{a}) - q_{n-1}(\mathbf{b})$  and  $\alpha_2 = q_n(\mathbf{a}) - q_n(\mathbf{b})$  then yields in the case  $d_{n-1}(\mathbf{a}, \mathbf{b}) \neq 0$  that

$$I(\mathbf{a}, \mathbf{b}) \leq 6ba^{-3/2} |\alpha_1|^{-1/2} = 6ba^{-3/2} (d_{n-1}(\mathbf{a}, \mathbf{b}))^{-1/2} \ll \frac{u^{-1/2} e^{(3\lambda/4+7\varepsilon)n}}{d_{n-1}(\mathbf{a}, \mathbf{b})^{1/2}}.$$

In the case when  $d_{n-1}(\mathbf{a}, \mathbf{b}) \leq d_n(\mathbf{a}, \mathbf{b})/2$  and  $d_n(\mathbf{a}, \mathbf{b}) \neq 0$  we use the case (2) of Lemma 6.4 to obtain

$$I(\mathbf{a}, \mathbf{b}) \leq 8|\alpha_2|^{-1} (a^{-1} + ba^{-2}) \ll \frac{u^{-1/2} e^{(3\lambda/4+7\varepsilon)n}}{d_n(\mathbf{a}, \mathbf{b})^{1/2}}$$

since  $d_n(\mathbf{a}, \mathbf{b})^{1/2} \leq d_n(\mathbf{a}, \mathbf{b})$  as  $d_n(\mathbf{a}, \mathbf{b}) \geq 1$ . □

Now as we noted in (6.8) to prove Proposition 6.3 we are left to estimate the sum

$$S(\mathbf{a}) = \sum_{\mathbf{b} \in \mathcal{R}_n} \mu(I_{\mathbf{b}}) I(\mathbf{a}, \mathbf{b})$$

for a fixed  $\mathbf{a} \in \mathcal{R}_n$ . Write

$$r_0 := \frac{1}{192} e^{(\frac{\lambda}{2}-2\varepsilon)n},$$

which, by the choice (5.4) of  $n_0$ , is larger than or equal to 1. We decompose the sum  $S(\mathbf{a})$  termwise to four smaller sums  $A_1, A_2, A_3$  and  $A_4$  depending on the values of  $d_{n-1}(\mathbf{a}, \mathbf{b})$  and  $d_n(\mathbf{a}, \mathbf{b})$  with respect to  $r_0$  and to each other; see their definitions in the Lemmas 6.6, 6.7, 6.8 and 6.9 below. Then in these lemmas we will apply Lemma 6.5 to obtain bounds for the sums  $A_1, \dots, A_4$  that are independent of  $\mathbf{a}$  as follows:

$$A_1 \ll e^{(-s\lambda+3\lambda\varepsilon)n} \quad \text{and} \quad A_2, A_3, A_4 \ll u^{-1/2} e^{(\frac{\lambda}{2}+15\lambda\varepsilon)n}.$$

Thus if we apply the bound (6.8) we have for  $\|f\|_2^2$  and use the fact that  $\mu$  is a probability measure, we obtain

$$\int |f(x)|^2 dx \ll \sum_{\mathbf{a} \in \mathcal{R}_n} \mu(I_{\mathbf{a}}) S(\mathbf{a}) \ll e^{(-s\lambda+3\lambda\varepsilon)n} + u^{-1/2} e^{(\frac{\lambda}{2}+15\lambda\varepsilon)n}$$

as claimed in Proposition 6.3. Therefore, we are just left to describe the decomposition to  $A_1, \dots, A_4$  and prove the desired bounds for them.

Let us first we consider words in the summation  $S(\mathbf{a})$  for which  $d_n$  and  $d_{n-1}$  agree:

**Lemma 6.6.** *We have*

$$A_1 := \sum_{\mathbf{b} \in \mathcal{R}_n: d_{n-1}(\mathbf{a}, \mathbf{b}) = d_n(\mathbf{a}, \mathbf{b}) = 0} \mu(I_{\mathbf{b}}) I(\mathbf{a}, \mathbf{b}) \ll e^{(-s\lambda+3\lambda\varepsilon)n}.$$

*Proof.* In this case we notice that whenever  $\mathbf{b} \in \mathcal{R}_n$  with

$$d_n(\mathbf{a}, \mathbf{b}) = d_{n-1}(\mathbf{a}, \mathbf{b}) = 0,$$

then in fact we have  $\mathbf{b} = \mathbf{a}$ . Indeed, if  $q_n(\mathbf{a}) = q_n(\mathbf{b})$  and  $q_{n-1}(\mathbf{a}) = q_{n-1}(\mathbf{b})$  and recall

$$q_n(\mathbf{a})p_{n-1}(\mathbf{a}) - q_{n-1}(\mathbf{a})p_n(\mathbf{a}) = (-1)^n = q_n(\mathbf{b})p_{n-1}(\mathbf{b}) - q_{n-1}(\mathbf{b})p_n(\mathbf{b})$$

by (2.2), so we obtain

$$q_{n-1}(\mathbf{a})(p_n(\mathbf{a}) - p_n(\mathbf{b})) \equiv 0 \pmod{q_n(\mathbf{a})}.$$

As  $q_{n-1}(\mathbf{a})$  and  $q_n(\mathbf{a})$  are coprime, we obtain  $p_n(\mathbf{a}) \equiv p_n(\mathbf{b}) \pmod{q_n(\mathbf{a})}$ . However,

$$p_n(\mathbf{b})/q_n(\mathbf{a}) = p_n(\mathbf{b})/q_n(\mathbf{b}) \leq 1,$$

so this is only possible if  $p_n(\mathbf{a}) = p_n(\mathbf{b})$  yielding  $\mathbf{b} = \mathbf{a}$  whenever  $d_n(\mathbf{a}, \mathbf{b}) = d_{n-1}(\mathbf{a}, \mathbf{b}) = 0$ . Thus by Lemma 5.1(4) we can bound

$$\sum_{\mathbf{b}: d_{n-1}(\mathbf{a}, \mathbf{b})=d_n(\mathbf{a}, \mathbf{b})=0} \mu(I_{\mathbf{b}})I(\mathbf{a}, \mathbf{b}) = \mu(I_{\mathbf{a}}) \ll e^{(-s\lambda+3\lambda\varepsilon)n}.$$

□

In the second case we have words for which  $d_n$  or  $d_{n-1}$  is bounded below by  $r_0$ :

**Lemma 6.7.** *We have*

$$A_2 := \sum_{\mathbf{b} \in \mathcal{R}_n: d_n(\mathbf{a}, \mathbf{b}) > r_0 \text{ or } d_{n-1}(\mathbf{a}, \mathbf{b}) > r_0} \mu(I_{\mathbf{b}})I(\mathbf{a}, \mathbf{b}) \ll u^{-1/2} e^{(\frac{\lambda}{2}+15\lambda\varepsilon)n}.$$

*Proof.* First of all, using (6.10) of Lemma 6.5 with Lemma 5.1(4) we obtain

$$\sum_{\mathbf{b}: \substack{d_n(\mathbf{a}, \mathbf{b}) > r_0 \\ d_{n-1}(\mathbf{a}, \mathbf{b}) \leq d_n(\mathbf{a}, \mathbf{b})/2}} \mu(I_{\mathbf{b}})I(\mathbf{a}, \mathbf{b}) \ll u^{-1/2} r_0^{-1/2} e^{(-s\lambda+3\lambda\varepsilon)n} e^{(3\lambda/4+7\varepsilon)n} \ll u^{-1/2} e^{(\frac{\lambda}{2}-s\lambda+12\lambda\varepsilon)n},$$

by the definition of  $r_0 = \frac{1}{192} e^{\frac{n\lambda}{2}-2\varepsilon n}$ . Secondly, using (6.9) of Lemma 6.5 yields the same bound for the rest

$$\begin{aligned} \sum_{\mathbf{b}: \substack{d_n(\mathbf{a}, \mathbf{b}) > r_0 \\ d_{n-1}(\mathbf{a}, \mathbf{b}) > d_n(\mathbf{a}, \mathbf{b})/2}} \mu(I_{\mathbf{b}})I(\mathbf{a}, \mathbf{b}) &\leq \sum_{\mathbf{b}: d_{n-1}(\mathbf{a}, \mathbf{b}) > r_0/2} \mu(I_{\mathbf{b}})I(\mathbf{a}, \mathbf{b}) \\ &\ll u^{-1/2} (r_0/2)^{-1/2} e^{(-s\lambda+3\lambda\varepsilon)n} e^{(3\lambda/4+7\varepsilon)n} \ll u^{-1/2} e^{(\frac{\lambda}{2}-s\lambda+12\lambda\varepsilon)n} \end{aligned}$$

Similarly, (6.9) of Lemma 6.5 gives us

$$\sum_{\mathbf{b}: d_{n-1}(\mathbf{a}, \mathbf{b}) > r_0} \mu(I_{\mathbf{b}})I(\mathbf{a}, \mathbf{b}) \ll u^{-1/2} r_0^{-1/2} e^{(-s\lambda+3\lambda\varepsilon)n} e^{(3\lambda/4+7\varepsilon)n} \ll u^{-1/2} e^{(\frac{\lambda}{2}-s\lambda+12\lambda\varepsilon)n}.$$

This number is always less than the desired bound  $u^{-1/2} e^{(\frac{\lambda}{2}+15\lambda\varepsilon)n}$ . □

Thus we are just left with the part of the sum  $S(\mathbf{a})$  where both  $d_n$  and  $d_{n-1}$  are bounded from above by  $r_0$  and where they are not simultaneously 0. In this part we have two sums  $A_3$  and  $A_4$  defined depending on whether  $d_n \leq 2d_{n-1}$  or not, and bounds for them in Lemmas 6.8 and 6.9 below. In the next lemma we have  $d_n \leq 2d_{n-1}$  which allows us to exclude the case  $d_{n-1} = 0$  as then it would force  $d_n = 0$ , which is already handled in Lemma 6.6 above.

**Lemma 6.8.** *We have*

$$A_3 := \sum_{\mathbf{b} \in \mathcal{R}_n: \substack{d_n(\mathbf{a}, \mathbf{b}) \leq 2d_{n-1}(\mathbf{a}, \mathbf{b}) \\ 1 \leq d_{n-1}(\mathbf{a}, \mathbf{b}) \leq r_0, d_n(\mathbf{a}, \mathbf{b}) \leq r_0}} \mu(I_{\mathbf{b}})I(\mathbf{a}, \mathbf{b}) \ll u^{-1/2} e^{(\frac{\lambda}{2}+15\lambda\varepsilon)n}.$$



*Proof.* For now on, write  $m := \lceil \frac{n}{2} \rceil$ . Given  $j = 0, 1, \dots, m$ , let us write

$$r_j := \frac{1}{192} e^{(\frac{n}{2}-j)\lambda-2\varepsilon n}.$$

For  $j = 0$  this agrees with the definition of  $r_0$  given above and  $r_0 \geq \frac{1}{192} e^{(\frac{\lambda}{2}-2\varepsilon)n_0} \geq 1$  by the choice (5.4) of  $n_0$  and because by the choice of  $\varepsilon$  in (6.1) we have  $\varepsilon < 1/26 < \lambda/4$  by  $\lambda > 1$ . Note that  $r_0 > r_1 > \dots > r_m$  and  $r_m < 1$ . For  $j = 0, 1, \dots, m-1$  define the annulus

$$\mathcal{A}_j = \{\mathbf{b} \in \mathcal{R}_n : d_n(\mathbf{a}, \mathbf{b}) \leq 2r_j \text{ and } r_{j+1} \leq d_{n-1}(\mathbf{a}, \mathbf{b}) \leq r_j\}.$$

Now these annuli allow us to decompose the summation we consider. Since  $\mu(I_{\mathbf{b}}) \ll e^{(-s\lambda+3\lambda\varepsilon)n}$  by Lemma 5.1(4) we have using the bound (6.9) of Lemma 6.5 that

$$\begin{aligned} \sum_{\mathbf{b}: \substack{d_n(\mathbf{a}, \mathbf{b}) \leq 2d_{n-1}(\mathbf{a}, \mathbf{b}) \\ 1 \leq d_{n-1}(\mathbf{a}, \mathbf{b}) \leq r_0, d_n(\mathbf{a}, \mathbf{b}) \leq r_0}} \mu(I_{\mathbf{b}}) I(\mathbf{a}, \mathbf{b}) &\leq \sum_{j=1}^{m-1} \sum_{\mathbf{b} \in \mathcal{A}_j} \mu(I_{\mathbf{b}}) I(\mathbf{a}, \mathbf{b}) \\ &\ll u^{-1/2} e^{(3\lambda/4+7\varepsilon)n} \sum_{j=0}^{m-1} r_j^{-1/2} \sum_{\mathbf{b} \in \mathcal{A}_j} e^{(-s\lambda+3\lambda\varepsilon)n} \\ &= u^{-1/2} e^{(3\lambda/4-s\lambda+10\lambda\varepsilon)n} \sum_{j=0}^{m-1} r_j^{-1/2} |\mathcal{A}_j| \end{aligned}$$

Thus our estimate reduces to estimating the cardinality of  $\mathcal{A}_j$ . Fix  $\mathbf{b} \in \mathcal{A}_j$ . Recall that the *mirror images* of  $\mathbf{a}$  and  $\mathbf{b}$  are defined by

$$\mathbf{a}^{\leftarrow} = (a_n, a_{n-1}, \dots, a_1) \quad \text{and} \quad \mathbf{b}^{\leftarrow} = (b_n, b_{n-1}, \dots, b_1).$$

By the mirroring property (2.3) of continuants we obtain

$$d_{n-1}(\mathbf{a}, \mathbf{b}) = |p_n(\mathbf{a}^{\leftarrow}) - p_n(\mathbf{b}^{\leftarrow})| \quad \text{and} \quad d_n(\mathbf{a}, \mathbf{b}) = d_n(\mathbf{a}^{\leftarrow}, \mathbf{b}^{\leftarrow})$$

so  $r_{j+1} \leq |p_n(\mathbf{a}^{\leftarrow}) - p_n(\mathbf{b}^{\leftarrow})| \leq r_j$  and  $d_n(\mathbf{a}^{\leftarrow}, \mathbf{b}^{\leftarrow}) \leq 2r_j$ , which yields

$$\left| \frac{p_n(\mathbf{a}^{\leftarrow})}{q_n(\mathbf{a}^{\leftarrow})} - \frac{p_n(\mathbf{b}^{\leftarrow})}{q_n(\mathbf{b}^{\leftarrow})} \right| \leq \frac{|p_n(\mathbf{a}^{\leftarrow}) - p_n(\mathbf{b}^{\leftarrow})|}{q_n(\mathbf{a}^{\leftarrow})} + \frac{|q_n(\mathbf{a}^{\leftarrow}) - q_n(\mathbf{b}^{\leftarrow})| p_n(\mathbf{b}^{\leftarrow})}{q_n(\mathbf{a}^{\leftarrow}) q_n(\mathbf{b}^{\leftarrow})} \leq \frac{3r_j}{q_n(\mathbf{a})},$$

where we used the mirroring which gives  $q_n(\mathbf{a}) = q_n(\mathbf{a}^{\leftarrow})$ .

By the quasi-independence Lemma 2.6 we have

$$q_n(\mathbf{a}) \geq \frac{1}{2} q_{n-j}(a_1, \dots, a_{n-j}) q_j(a_{n-j+1}, \dots, a_n).$$

On the other hand, by the mirroring property, we have

$$q_j(a_{n-j+1}, \dots, a_n) = q_j(a_n, \dots, a_{n-j+1})$$

so we can bound

$$\frac{q_{n-j}(a_1, \dots, a_{n-j})^2}{4q_n(\mathbf{a})^2} \leq q_j(a_n, \dots, a_{n-j+1})^{-2}.$$

On the other hand, the definition  $r_j = \frac{1}{192} e^{(\frac{n}{2}-j)\lambda-2\varepsilon n}$  gives us

$$3r_j = \frac{1}{64} e^{(\frac{n}{2}-j)\lambda-2\varepsilon n} < \frac{1}{64} \frac{e^{(\lambda-\varepsilon)(n-j)}}{e^{(\lambda+\varepsilon)n/2}} \leq \frac{1}{64} \frac{q_{n-j}(a_1, \dots, a_{n-j})^2}{q_n(\mathbf{a})}$$

using the fact that  $\lfloor n/2 \rfloor \leq n-j < n$  and Lemma 5.1(1). Thus we have shown that the annulus  $\mathcal{A}_j \subset \mathcal{B}_j$ , where

$$\mathcal{B}_j := \left\{ \mathbf{b} \in \mathcal{R}_n : \left| \frac{p_n(\mathbf{a}^{\leftarrow})}{q_n(\mathbf{a}^{\leftarrow})} - \frac{p_n(\mathbf{b}^{\leftarrow})}{q_n(\mathbf{b}^{\leftarrow})} \right| < \frac{1}{16} q_j(a_n, \dots, a_{n-j+1})^{-2} \right\}.$$

We wish to show that the cardinality of  $\mathcal{B}_j$  is at most  $2|\mathcal{C}_j|$ , where

$$\mathcal{C}_j := \{(b_1, \dots, b_{n-j+1}) : \mathbf{b} \in \mathcal{R}_n\}.$$

Note that  $\mathcal{C}_j$  is precisely the collection of all generation  $n - j + 1$  subwords of elements in  $\mathcal{R}_n$ . Since  $\lfloor n/2 \rfloor < n - j + 1 \leq n$ , we may use Lemma 5.1 for the words in  $\mathcal{C}_j$ . To prove this consider the interval  $I = I_{a_n, \dots, a_{n-j+1}}$ .

- (a) If  $a_{n-j+1} \neq 1$ , then  $I$  is neighboured by the intervals  $I_{a_n, \dots, a_{n-j+1}-1}$  and  $I_{a_n, \dots, a_{n-j+1}+1}$  both of which must have diameter greater than  $\frac{1}{16}q_j(a_n, \dots, a_{n-j+1})^{-2}$ . Indeed, by the recurrence relation for the continuants

$$\begin{aligned} q_j(a_n, \dots, a_{n-j+1} + 1) &= q_j(a_n, \dots, a_{n-j+1}) + q_{j-1}(a_n, \dots, a_{n-j+1}) \\ &\leq 2q_j(a_n, \dots, a_{n-j}) \end{aligned}$$

and

$$\begin{aligned} q_j(a_n, \dots, a_{n-j+1} - 1) &= q_j(a_n, \dots, a_{n-j}) - q_{j-1}(a_n, \dots, a_{n-j+1}) \\ &\leq q_j(a_n, \dots, a_{n-j}), \end{aligned}$$

which thanks to Lemma 2.5 gives the size bound  $\frac{1}{16}q_j(a_n, \dots, a_{n-j+1})^{-2}$  for the intervals from below. Thus if we require  $\mathbf{b} \in \mathcal{B}_j$  we would need  $(a_n, \dots, a_{n-j+2}) = (b_n, \dots, b_{n-j+2})$  as otherwise the point  $p_n(\mathbf{b}^\leftarrow)/q_n(\mathbf{b}^\leftarrow)$  would be too far from the point  $p_n(\mathbf{a}^\leftarrow)/q_n(\mathbf{a}^\leftarrow)$ .

- (b) If  $a_{n-j+1} = 1$ , then depending on the sign of the derivative of  $T'_{a_n, \dots, a_{n-j+2}}$ , the two neighbouring cylinders are the  $j$ th generation interval  $I_{a_n, \dots, a_{n-j+1}+1}$  and the  $(j-1)$ th generation interval  $I_{a_n, \dots, a_{n-j+2}+1}$  and both of these have diameter greater than  $\frac{1}{16}q_j(a_n, \dots, a_{n-j+1})^{-2}$  again by the recurrence relation for the continuants. Thus there are at most two choices for  $(b_n, \dots, b_{n-j+2})$ .

The cases (a) and (b) together prove that a given  $\mathbf{b} \in \mathcal{B}_j$  has at most two possibilities for the segment  $b_n, \dots, b_{n-j+2}$ . Moreover, as  $\mathbf{b} \in \mathcal{R}_n$ , the initial segment  $(b_1, \dots, b_{n-j+1}) \in \mathcal{C}_j$ . Hence  $|\mathcal{B}_j| \leq 2|\mathcal{C}_j|$  as we wanted.

Now, we can just bound the cardinality of  $\mathcal{C}_j$ . Given  $(b_1, \dots, b_{n-j+1}) \in \mathcal{C}_j$  we have by Lemma 5.1(4) that

$$\mu(I_{b_1, \dots, b_{n-j+1}}) \geq C^{-1}e^{-(n-j+1)(\lambda s + 3\lambda\varepsilon)}.$$

Since  $\mu$  is a probability measure this gives us a cardinality bound

$$|\mathcal{C}_j| \leq Ce^{(n-j+1)(\lambda s + 3\lambda\varepsilon)} \leq Ce^{(s\lambda + 3\lambda\varepsilon)n} e^{-s\lambda j}.$$

Therefore, we can bound the cardinality of  $\mathcal{A}_j$  as follows

$$|\mathcal{A}_j| \leq 2|\mathcal{C}_j| \ll e^{(s\lambda + 3\lambda\varepsilon)n} e^{-s\lambda j}.$$

Now recall that we are interested in the sum  $\sum_{j=0}^{m-1} r_j^{-1/2} |\mathcal{A}_j|$ . Inputting the definition of  $r_j = \frac{1}{192}e^{(\frac{n}{2}-j)\lambda - 2\varepsilon n}$ , we see that

$$r_j^{-1/2} \ll e^{(-\frac{\lambda}{4} + 2\lambda\varepsilon)n} e^{\frac{\lambda}{2}j}$$

using a crude bound  $\sqrt{\lambda} \leq \lambda$  as  $\lambda > 1$ . Thus the sum

$$\sum_{j=0}^{m-1} r_j^{-1/2} |\mathcal{A}_j| \ll e^{(s\lambda + 3\lambda\varepsilon)n} e^{(-\frac{\lambda}{4} + 2\lambda\varepsilon)n} \sum_{j=0}^{m-1} e^{(\frac{1}{2}-s)\lambda j} \ll e^{(s\lambda + 3\lambda\varepsilon)n} e^{(-\frac{\lambda}{4} + 2\lambda\varepsilon)n} = e^{(-\frac{\lambda}{4} + s\lambda + 5\lambda\varepsilon)n} \quad (6.12)$$

where we used  $s > 1/2$  so that the geometric sum is dominated by a constant. Inputting this to what we obtained previously, we have our desired claim

$$\sum_{\mathbf{b}: \substack{d_n(\mathbf{a}, \mathbf{b}) \leq 2d_{n-1}(\mathbf{a}, \mathbf{b}) \\ 1 \leq d_{n-1}(\mathbf{a}, \mathbf{b}) \leq r_0, d_n(\mathbf{a}, \mathbf{b}) \leq r_0}} \mu(I_{\mathbf{b}}) I(\mathbf{a}, \mathbf{b}) \ll u^{-1/2} e^{(3\lambda/4 - s\lambda + 10\lambda\varepsilon)n} \sum_{j=0}^{m-1} r_j^{-1/2} |\mathcal{A}_j| \ll u^{-1/2} e^{(\frac{\lambda}{2} + 15\lambda\varepsilon)n}.$$

□

Finally we are left with the case  $d_n > 2d_{n-1}$ . In this case we can exclude those words with  $d_n = 0$  as then it would force  $d_{n-1} < 0$ , which is absurd as the differences are non-negative.

**Lemma 6.9.** *We have*

$$A_4 := \sum_{\mathbf{b} \in \mathcal{R}_n: \substack{d_n(\mathbf{a}, \mathbf{b}) > 2d_{n-1}(\mathbf{a}, \mathbf{b}) \\ d_{n-1}(\mathbf{a}, \mathbf{b}) \leq r_0, 1 \leq d_n(\mathbf{a}, \mathbf{b}) \leq r_0}} \mu(I_{\mathbf{b}})I(\mathbf{a}, \mathbf{b}) \ll u^{-1/2} e^{(\frac{\lambda}{2} + 15\lambda\varepsilon)n}.$$

*Proof.* In this case we can use the part (6.10) of Lemma 6.5, which gives

$$\sum_{\mathbf{b}: \substack{d_n(\mathbf{a}, \mathbf{b}) > 2d_{n-1}(\mathbf{a}, \mathbf{b}) \\ d_{n-1}(\mathbf{a}, \mathbf{b}) \leq r_0, 1 \leq d_n(\mathbf{a}, \mathbf{b}) \leq r_0}} \mu(I_{\mathbf{b}})I(\mathbf{a}, \mathbf{b}) \ll u^{-1/2} e^{(3\lambda/4 - s\lambda + 10\lambda\varepsilon)n} \sum_{\mathbf{b}: \substack{d_n(\mathbf{a}, \mathbf{b}) > 2d_{n-1}(\mathbf{a}, \mathbf{b}) \\ d_{n-1}(\mathbf{a}, \mathbf{b}) \leq r_0, 1 \leq d_n(\mathbf{a}, \mathbf{b}) \leq r_0}} d_n(\mathbf{a}, \mathbf{b})^{-1/2}$$

again by Lemma 5.1(4). Now consider the summation on the right-hand side. We proceed similarly as in the proof of Lemma 6.8: now we split the summation to the annuli determined by the conditions  $d_{n-1}(\mathbf{a}, \mathbf{b}) \leq r_j/2$  and  $r_{j+1} \leq d_n(\mathbf{a}, \mathbf{b}) \leq r_j$  with the same radii  $r_j$ . By the mirroring property (2.3) of continuants these conditions yield

$$|p_n(\mathbf{a}^{\leftarrow}) - p_n(\mathbf{b}^{\leftarrow})| \leq r_j/2 \quad \text{and} \quad |q_n(\mathbf{a}^{\leftarrow}) - q_n(\mathbf{b}^{\leftarrow})| \leq r_j$$

so we can make a same bound as in the proof of Lemma 6.8 for the difference:

$$\left| \frac{p_n(\mathbf{a}^{\leftarrow})}{q_n(\mathbf{a}^{\leftarrow})} - \frac{p_n(\mathbf{b}^{\leftarrow})}{q_n(\mathbf{b}^{\leftarrow})} \right| \leq \frac{|p_n(\mathbf{a}^{\leftarrow}) - p_n(\mathbf{b}^{\leftarrow})|}{q_n(\mathbf{a}^{\leftarrow})} + \frac{|q_n(\mathbf{a}^{\leftarrow}) - q_n(\mathbf{b}^{\leftarrow})| |p_n(\mathbf{b}^{\leftarrow})|}{q_n(\mathbf{a}^{\leftarrow})q_n(\mathbf{b}^{\leftarrow})} < \frac{3r_j}{q_n(\mathbf{a})}.$$

Thus the proof after this is exactly as in Lemma 6.8 since this allows us to reduce the estimation to a cardinality bound for  $\mathcal{B}_j$  and summation over  $j$ . Hence we obtain the same bound up to a constant

$$\sum_{\mathbf{b}: \substack{d_n(\mathbf{a}, \mathbf{b}) > 2d_{n-1}(\mathbf{a}, \mathbf{b}) \\ d_{n-1}(\mathbf{a}, \mathbf{b}) \leq r_0, 1 \leq d_n(\mathbf{a}, \mathbf{b}) \leq r_0}} \mu(I_{\mathbf{b}})I(\mathbf{a}, \mathbf{b}) \ll u^{-1/2} e^{(\frac{\lambda}{2} + 15\lambda\varepsilon)n}.$$

□

**6.6. Completion of the proof of the main theorem.** Recall that in Section 6.3 we deduced that the proof of Theorem 1.3 is complete if we are able to prove that the main integral bound (6.6) holds for the regular part, that is,

$$\int |f(x)| d\mu(x) \ll u^{-\eta_s + 38\varepsilon} + u^{-\frac{\delta}{12\lambda}},$$

where the exponent

$$\eta_s = \frac{2s^2 - s}{(4 - s)(1 + 2s)}$$

and  $\delta = \delta(\varepsilon/2) > 0$  is the constant from large deviations with  $\varepsilon/2$ . To verify this, write

$$\alpha := 2s + 2\varepsilon \quad \text{and} \quad \beta := (1 + 2s)\eta_s.$$

Then  $\gamma := \alpha + \beta \geq 1$  since  $s > 1/2$  and crudely bounding also  $\gamma \leq 7$ . By Lemma 6.2 we have  $\|f'\|_{\infty} \leq e^{\alpha\lambda n}$ . Thus the function  $f$  satisfies the assumptions of Lemma 6.1. This gives us

$$\int |f(x)| d\mu(x) \ll e^{(-\beta\lambda + \varepsilon\gamma)n} + \|f\|_2^2 e^{((1-s)\alpha\lambda + (3-s)\beta\lambda + 3\lambda\varepsilon\gamma)n} + e^{-\delta\gamma n/4} =: t_1 + t_2 + t_3.$$

Recall that we chose in (6.3) the number  $n \geq n_0$  such that

$$e^{(1+2s)\lambda n} \leq u < e^{(1+2s)\lambda(n+1)}. \quad (6.13)$$

By the definition of  $\beta$  as  $\lambda > 1$  and  $\gamma \leq 7$  we obtain

$$t_1 = e^{-(1+2s)\eta_s \lambda n + \gamma \varepsilon n} \ll u^{-\eta_s + \frac{\gamma}{(1+2s)\lambda} \varepsilon} < u^{-\eta_s + 7\varepsilon}$$

and similarly as  $\gamma > 1$  and  $s \leq 1$  we have

$$t_3 \ll u^{-\frac{\gamma}{4(1+2s)\lambda} \delta} < u^{-\frac{\delta}{12\lambda}}.$$

Thus we are just left with bounding  $t_2$  with respect to  $u$ .

Recall the  $L^2$  bound from Proposition 6.3:

$$\|f\|_2^2 \ll u^{-1/2} e^{(\lambda/2+15\lambda\varepsilon)n} + e^{(-s\lambda+3\lambda\varepsilon)n}.$$

Inputting this to the definition of  $t_2$  yields

$$t_2 \ll u^{-1/2} e^{(\lambda/2+15\lambda\varepsilon)n} e^{((1-s)\alpha\lambda+(3-s)\beta\lambda+3\lambda\varepsilon\gamma)n} + e^{(-s\lambda+3\lambda\varepsilon)n} e^{((1-s)\alpha\lambda+(3-s)\beta\lambda+3\lambda\varepsilon\gamma)n}.$$

Now define

$$\varrho_s := \frac{1}{2} - \frac{1/2 + 2s(1-s) + (3-s)(1+2s)\eta_s}{1+2s} = \frac{s - 2s(1-s) - (3-s)(1+2s)\eta_s}{1+2s}.$$

If we plug-in the definition of  $\alpha = 2s + 2\varepsilon$  and  $\beta = (1+2s)\eta_s$  to the first term bounding  $t_2$  and use (6.13), we obtain a bound

$$\begin{aligned} u^{-1/2} e^{(\lambda/2+15\lambda\varepsilon)n} e^{((1-s)\alpha\lambda+(3-s)\beta\lambda+3\lambda\varepsilon\gamma)n} &= u^{-1/2} e^{(\lambda/2+2s(1-s)\lambda+(3-s)(1+2s)\eta_s\lambda+(3\gamma+17)\lambda\varepsilon)n} \\ &\ll u^{-\varrho_s + \frac{3\gamma+17}{1+2s}\varepsilon} \\ &\ll u^{-\varrho_s + 38\varepsilon} \end{aligned}$$

using  $\gamma \leq 7$ . Moreover, for the second term bounding  $t_2$  we obtain similarly

$$e^{(-s\lambda+2s(1-s)\lambda+(3-s)(1+2s)\eta_s\lambda+(3\gamma+5)\lambda\varepsilon)n} \ll u^{-\varrho_s + \frac{3\gamma+5}{1+2s}\varepsilon} \ll u^{-\varrho_s + 26\varepsilon}.$$

On the other hand, by the definition of  $\eta_s = \frac{2s^2-s}{(4-s)(1+2s)}$ , the exponent

$$\varrho_s = \frac{(4-s)[s - 2s(1-s) - (3-s)(1+2s)\eta_s]}{(4-s)(1+2s)} = \frac{8s^3 + 10s^2 - s}{(4-s)(1+2s)} > \eta_s.$$

Thus the polynomial decay coming from the term  $t_1$  involving  $\eta_s$  dominates the decay we obtain from  $t_2$  since  $\eta_s - 38\varepsilon > 0$  by the choice of  $\varepsilon$  in (6.1). This finishes the final claim and the proof of Theorem 1.3.

**6.7. The case of Hausdorff measure.** To finish the section, let us make some remarks on the proof of Corollary 1.4 and see how the proof of the main Theorem 1.3 can be modified to have the Rajchman property for the Hausdorff measure  $\nu = \mathcal{H}^s|_{B(\mathcal{A})}$  for a finite  $\mathcal{A} \subset \mathbb{N}$ . The crucial observation is that even though  $\nu$  is not  $T$  invariant, it is an  $s$ -conformal measure, that is, a fixed point for the dual operator  $\mathcal{L}_\varphi^*$  with potential  $\varphi = -s \log |T'|$ ; see for example [10, Theorem 6]. In Section 6.3, when we apply  $\mathcal{L}_\varphi$  to the oscillation  $x \mapsto e^{-2\pi i \xi x}$ , we only need the measure to be a fixed point for  $\mathcal{L}_\varphi^*$ . The  $T$  invariance is only needed in the deduction of the large deviation bounds, but by [7, Theorem 5.3] we know that  $\nu$  is equivalent to a  $T$  invariant measure  $\mu$  on  $B(\mathcal{A})$  with uniformly positive and finite densities. This allows us to use the same large deviation bounds or measure comparisons to the lengths  $|I_{\mathbf{a}}|$  for  $\nu$ , which we already established for  $\mu$  up to a fixed multiplicative error constant, which does not affect the decay rate of the Fourier transform.

## 7. PROSPECTS

**7.1. Dimension and decay.** An immediate question arising from our main result is that why is there a dimension assumption  $\dim \mu > 1/2$  for the Gibbs measures  $\mu$ ? The result by Hochman-Shmerkin [9] on finding normal numbers in the supports of Gibbs measures does not have dimension restrictions, so it could be possible that the dimension requirement is just an artifact of the method we use. The dimension assumption is important in the summation of the geometric series in the  $L^2$  estimations on the lemmas Lemma 6.8 and 6.9 of the proof of Proposition 6.3 in Section 6.5. In particular, in the estimate (6.12) the important term  $r_0$  may not dominate if  $s$  is less than  $1/2$ .

Another direction is to estimate the decay rate we have for the Fourier transform. It is an open problem whether the set of badly approximable numbers is a Salem set. As we noted in the introduction, the Hausdorff dimension of this set is 1, so we would need to find Rajchman measures on badly approximable numbers with  $|\widehat{\mu}(\xi)| = O(|\xi|^{-\eta})$  with  $\eta$  arbitrarily close to  $1/2$ . The rates

we obtain for the Gibbs measures when the Hausdorff dimension  $s$  is large are not close to  $1/2$ . The proofs in this paper are designed to give polynomial decay but we have not considered how to optimise this decay and it is probable that new methods would need to be developed to do this.

**7.2. Markov maps.** It is a natural question whether the results here could be extended to dynamical systems other than  $([0, 1] \setminus \mathbb{Q}, T)$ . A possible problem could be to consider general expanding Markov maps on the circle  $\mathbb{T}$  and the Gibbs measures related to these maps. Nonlinearity of the system seems to be crucial in the proof and in fact it is straightforward to see that the results cannot hold for the maps  $x \mapsto nx \bmod 1$ . However, even if we assume nonlinearity, another obstacle for extending the results are some of the number theoretical properties of the Gauss map we use. The mirroring property which basically guarantees that the interval  $I_{\mathbf{a}}$  has a comparable length to  $I_{\mathbf{a}^{\leftarrow}}$  we used in bounding the  $L^2$  norm of the regular part  $f$  in Section 6.5 is rarely available to general Markov maps. This can probably be overcome by slightly strengthening the assumption on the dimension of  $\mu$  since this property is only needed to weaken this assumption. However in part 1 of the proof of Proposition 6.3 we estimate the set for which  $d_{n-1}(\mathbf{a}, \mathbf{b}) = 0$  and  $d_n(\mathbf{a}, \mathbf{b}) = 0$ . This makes important use of properties of the continuants of the continued fraction expansion and so this part of the proof would need to be overhauled to extend the result to more general Markov maps.

## 8. ACKNOWLEDGEMENTS

The authors are grateful to Alan Haynes, Andrew Ferguson and all the other members of the “*Dynamical Systems and Diophantine Approximation*” reading group in the University of Bristol during Spring 2013 where this collaboration was initiated. TS also thanks Sanju Velani and all the participants of the minicourse “*An Invitation to Geometric and Diophantine Fourier Analysis*” in the University of York during December 2013 for many comments and suggestions on this paper. We thank Tuomas Orponen, Michał Rams and Mike Todd for useful comments on a preliminary version of this work, Tomas Persson for pointing out the application to Salem’s problem, and the anonymous referee for comments and suggestions.

## REFERENCES

- [1] R. Baker. Metric number theory and the large sieve. *J. London Math. Soc. (2)*, 24(1):34–40, 1981.
- [2] C. Bluhm. On a theorem of Kaufman: Cantor-type construction of linear fractal Salem sets. *Ark. Mat.*, 36(2):307–316, 1998.
- [3] C. Bluhm. Liouville numbers, Rajchman measures, and small Cantor sets. *Proc. Amer. Math. Soc.*, 128(9):2637–2640, 2000.
- [4] F. Canto-Martín, H. Hedenmalm, and A. Montes-Rodríguez. Perron-Frobenius operators and the Klein-Gordon equation. *J. Eur. Math. Soc. (JEMS)*, 16(1):31–66, 2014.
- [5] H. Davenport, P. Erdős, and W. J. LeVeque. On Weyl’s criterion for uniform distribution. *Michigan Math. J.*, 10:311–314, 1963.
- [6] R. Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fourth edition, 2010.
- [7] K. J. Falconer. *Techniques in fractal geometry*. John Wiley & Sons Ltd., Chichester, 1997.
- [8] I. Good. The fractional dimensional theory of continued fractions. *Math. Proc. Cambridge Philos. Soc.*, 37(3):199–228, 1941.
- [9] M. Hochman and P. Shmerkin. Equidistribution from fractals. *Invent. Math.*, 2013. To appear, Preprint at <http://arxiv.org/abs/1302.5792>.
- [10] F. Hofbauer. Hausdorff and conformal measures for expanding piecewise monotonic maps of the interval. *Stud. Math.*, 103(2):191–206, 1992.
- [11] V. Jarník. Zur metrischen theorie der diophantischen approximationen. *Prace Mat.-Fiz.*, 36(1):91–106, 1928–1929.
- [12] R. Kaufman. Continued fractions and Fourier transforms. *Mathematika*, 27(2):262–267, 1980.
- [13] R. Kaufman. On the theorem of Jarník and Besicovitch. *Acta Arith.*, 39(3):265–267, 1981.
- [14] R. Kaufman.  $M$ -sets and measures. *Ann. of Math. (2)*, 135(1):125–130, 1992.
- [15] M. Kesseböhmer and B. O. Stratmann. Fractal analysis for sets of non-differentiability of Minkowski’s question mark function. *J. Number Theory*, 128(9):2663–2686, 2008.

- [16] A. Khinchin. *Continued fractions*. Dover Publications Inc., Mineola, NY, russian edition, 1997. With a preface by B. V. Gnedenko, Reprint of the 1964 translation.
- [17] Y. Kifer. Large deviations in dynamical systems and stochastic processes. *Trans. Amer. Math. Soc.*, 321(2):505–524, 1990.
- [18] Y. Kifer, Y. Peres, and B. Weiss. A dimension gap for continued fractions with independent digits. *Israel J. Math.*, 124:61–76, 2001.
- [19] J. R. Kinney. Note on a singular function of Minkowski. *Proc. Amer. Math. Soc.*, 11:788–794, 1960.
- [20] A. Lopes. Entropy and large deviation. *Nonlinearity*, 3(2):527–546, 1990.
- [21] R. Lyons. Seventy years of Rajchman measures. *Proceedings of the Conference in Honor of Jean-Pierre Kahane (Orsay, 1993)*, *J. Fourier Anal. Appl.*, pages 363–377, 1995.
- [22] P. Mattila. *Geometry of Sets and Measures in Euclidean Spaces: Fractals and Rectifiability*. Cambridge University Press, Cambridge, 1995.
- [23] P. Mattila. Hausdorff dimension, projections, and the Fourier transform. *Publ. Mat.*, 48(1):3–48, 2004.
- [24] R. D. Mauldin and M. Urbański. *Graph directed Markov systems*, volume 148 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2003. Geometry and dynamics of limit sets.
- [25] H. Minkowski. Geometrie der zahlen. *Gesammelte Abhandlungen*, 2, 1911.
- [26] H. Montgomery. *Ten lectures on the interface between analytic number theory and harmonic analysis*, volume 84 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1994.
- [27] T. Persson. On a problem by R. Salem concerning minkowski’s question mark function. Preprint, <http://arxiv.org/abs/1501.00876>, 2015.
- [28] A. Pollington and S. Velani. On a problem in simultaneous Diophantine approximation: Littlewood’s conjecture. *Acta Math.*, 185(2):287–306, 2000.
- [29] M. Queffélec and O. Ramaré. Analyse de Fourier des fractions continues à quotients restreints. *Enseign. Math.* (2), 49(3-4):335–356, 2003.
- [30] R. Salem. On some singular monotonic functions which are strictly increasing. *Trans. Amer. Math. Soc.*, 53:427–439, 1943.
- [31] O. Sarig. Phase transitions for countable Markov shifts. *Comm. Math. Phys.*, 217(3):555–577, 2001.
- [32] O. Sarig. Existence of Gibbs measures for countable Markov shifts. *Proc. Amer. Math. Soc.*, 131(6):1751–1758 (electronic), 2003.
- [33] P. Walters. Invariant measures and equilibrium states for some mappings which expand distances. *Trans. Amer. Math. Soc.*, 236:121–153, 1978.
- [34] T. H. Wolff. *Lectures on harmonic analysis*, volume 29 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2003. With a foreword by Charles Fefferman and preface by Izabella Laba, Edited by Laba and Carol Shubin.
- [35] S. Yakubovich. The affirmative solution to Salem’s problem revisited. Preprint, <http://arxiv.org/abs/1501.00141>, 2015.
- [36] L.-S. Young. Large deviations in dynamical systems. *Trans. Amer. Math. Soc.*, 318(2):525–543, 1990.
- [37] M. Yuri. Large deviations for countable to one Markov systems. *Comm. Math. Phys.*, 258(2):455–474, 2005.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF BRISTOL, UNIVERSITY WALK, CLIFTON, BRISTOL, BS8 1TW, ENGLAND

*E-mail address*: `thomas.jordan@bristol.ac.uk`

EINSTEIN INSTITUTE OF MATHEMATICS, THE HEBREW UNIVERSITY OF JERUSALEM, GIVAT RAM, JERUSALEM 91904, ISRAEL

*E-mail address*: `tsahlsten@math.huji.ac.il`