Spectral form factors of rectangle billiards

Jens Markloff[*†]

Division de Physique Théorique, Institut de Physique Nucléaire
F-91406 Orsay Cedex, France

and

Institut des Hautes Études Scientifique, 35, Route de Chartres
F-91440 Bures-sur-Yvette, France

January 1998

Abstract. The Berry-Tabor conjecture asserts that local statistical measures of the eigenvalues λ_j of a "generic" integrable quantum system coincide with those of a Poisson process. We prove that, in the case of a rectangle billiard with random ratio of sides, the sum $N^{-1/2} \sum_{j \leq N} \exp(2\pi \mathrm{i}\, \lambda_j \tau)$ behaves for τ random and N large like a random walk in the complex plane with a non-Gaussian limit distribution. The expectation value of the distribution is zero; its variance, which is essentially the average pair correlation function, is one, in accordance with the Berry-Tabor conjecture, but all higher moments (≥ 4) diverge. The proof of the existence of the limit distribution uses the mixing property of a dynamical system defined on a product of hyperbolic surfaces. The Berry-Tabor conjecture and the existence of the the limit distribution for a *fixed* generic rectangle are related to an equidistribution conjecture for long horocycles on this product space.

[*]Supported by the European Post-Doctoral Institute for the Mathematical Sciences, the Engineering and Physical Sciences Research Council (Grant #GR K99015), the European Commission (TMR Marie Curie Grant), and the Basic Research Institute in the Mathematical Sciences, Hewlett-Packard Laboratories, Bristol.

[†]Permanent address: School of Mathematics, University of Bristol, Bristol BS8 1TW, U.K. (Email: J.Marklof@Bristol.ac.uk)

Contents

1 Introduction **3**

1.1 Billiards in a rectangle 7

1.2 The main results 7

1.3 Basic definitions and notations 11

2 Invariance properties of the form factor **11**

3 Regime I: $|\tau| \ll \lambda^{-1/2-\epsilon}$ **17**

4 Regime II: $\lambda^{-1/2-\epsilon} \ll |\tau| \ll \lambda^{-1/2+\epsilon}$ **18**

5 Ergodicity, mixing and equidistribution **19**

6 Regime III: $\tau \sim$ const **24**

6.1 The expectation value 24

6.2 The limit distribution — smooth cut-off functions 28

6.3 The limit distribution — general cut-off functions 33

6.4 Random walks 34

7 Rational α **35**

7.1 The expectation value 37

7.2 The limit distribution 37

1 Introduction

Let $\lambda_1 \le \lambda_2 \le \lambda_3 \le \ldots \to \infty$ be a sequence of numbers satisfying

$$\#\{j : \lambda_j \le \lambda\} \sim \lambda, \qquad \lambda \to \infty, \tag{1}$$

which means that the average spacing between adjacent levels is asymptotically unity. One quantity measuring the "randomness" of the deterministic sequence $\{\lambda_j\}_j$ is the *consecutive level spacing distribution*, which is defined by

$$P(s,N) = \frac{1}{N}\sum_{j=1}^{N} \delta(s - \lambda_{j+1} + \lambda_j), \tag{2}$$

where $\delta(x)$ is the Dirac mass. The limit distribution of $P(s,N)$ for $N \to \infty$ (if it exists) shall be denoted by $P(s)$. That is, for any sufficiently nice test function h,

$$\lim_{N\to\infty} \int_0^\infty P(s,N)\, h(s)\, ds = \int_0^\infty P(s)\, h(s)\, ds. \tag{3}$$

Berry and Tabor conjectured [5] that, when the sequence $\{\lambda_j\}_j$ is a sequence of eigenvalues of a quantum Hamiltonian, whose classical dynamics is integrable, then the limit distribution $P(s)$ should in general coincide with the one for a random sequence generated by a Poisson process, i.e.,

$$P_{\mathrm{Poisson}}(s) = \exp(-s).$$

This is particularly interesting, because the limit distribution for systems, which are not integrable but chaotic, is expected to be the Gaudin distribution for the eigenvalues of random matrices [18], which is approximately described by Wigner's surmise. For the GUE ensemble, say, it reads

$$P_{\mathrm{GUE}}(s) \approx \frac{32}{\pi^2}\, s^2\, \mathrm{e}^{-\frac{4}{\pi}s^2}\ .$$

Obvious examples for "non-generic" integrable systems, which violate the Berry-Tabor conjecture, are two-dimensional harmonic oscillators, as was already noted by Berry and Tabor [5] and later studied in more detail by Pandey et al. [49], Bleher [7, 8] and Greenman [32].[1]

Further negative examples are Zoll surfaces, where, like on the sphere, all geodesics are closed and have the same length. In the case of the sphere

[1] The spacings of two-dimensional harmonic oscillators are directly related to the spacings between the fractional parts of the sequence $n\theta$, which had been studied earlier, see [60] for a survey.

the eigenvalues of the (negative) Laplacian $-\Delta$ are (after rescaling) $E_{l,m} = l(l+1)$, $m = -l, \ldots, l$, hence with multiplicity $2l+1$. Label these numbers in increasing order by $\lambda_1, \lambda_2, \ldots$; due to the high multiplicity one thus has

$$P(s) = \delta(s). \tag{4}$$

The same result holds for all other Zoll surfaces where the eigenvalues are extremely clustered around the values $l(l+1)$, compare Duistermaat and Guillemin [26], Weinstein [65], Colin de Verdière [23].[2]

Results in favor of the Berry-Tabor conjecture are rare and can so far be only proved for the pair correlation density[3],

$$R_2(s,N) = \frac{1}{N} \sum_{j,k=1}^{N} \delta(s - \lambda_j + \lambda_k), \tag{5}$$

which measures the spacings between *all* elements of the sequence and is therefore not a probability distribution. In the case of a random sequence from a Poisson process, $R_2(s,N)$ converges to the limiting density[4]

$$R_{2\ \mathrm{Poisson}}(s) = \delta(s) + 1,$$

which is consequently the expectation for integrable systems. In other words,

$$\lim_{N\to\infty} \int_{-\infty}^{\infty} R_2(s,N)\, h(s)\, ds = h(0) + \int_{-\infty}^{\infty} h(s)\, ds, \tag{6}$$

for a suitable class of test functions h. This variant of the Berry-Tabor conjecture with respect to $R_2(s)$ was verified by Sarnak [54] for the eigenvalues of the Laplacian on almost every flat torus (almost every with respect to Lebesgue measure in the moduli space of two-dimensional flat tori), but he simultaneously disproved the conjecture for a set of second Baire category[5]. His result was recently extended to four-dimensional tori by VanderKam [63, 64]. Similar studies in this direction are due to Rudnick and Sarnak [52], whose results can be related to the eigenvalues of boxed oscillators, and

[2] The local correlations of the eigenvalues of *each individual* cluster are studied by Uribe and Zelditch [62].

[3] Sinai [59] and Major [45] showed that the statistics of lattice points in certain generic strips follow Poisson statistics in *all* moments. The boundary of these domains is not twice differentiable and looks like a trajectory for Brownian motion. Spectra of integrable systems like integrable geodesic flows are, however, related via EBK quantization to lattice points in domains with piecewise *smooth* boundary [22].

[4] The delta mass $\delta(s)$ is a result of our definition which counts spacings between equal elements whose spacing is trivially zero. The interesting part is the "1".

[5] Sets of second Baire category are all sets which are not of first Baire category, and the latter are sets, which are countable unions of nowhere dense sets, so pretty sparse in the topological sense.

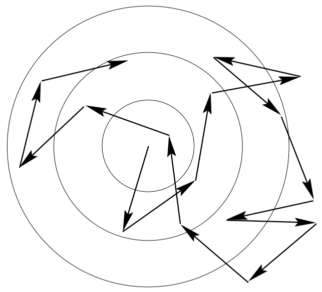

Figure 1: A walk of $N = 16$ unit steps in the complex plane.

Zelditch [66], who considers the level spacings for quantum maps in genus zero. It should be pointed out that, although the above results hold almost everywhere in the corresponding parameter spaces, the Berry-Tabor conjecture could not be proved for a specific example.[6] For a more detailed up-to-date review on these topics see Sarnak's lectures [55, 56].

The Berry-Tabor conjecture can only be expected to hold for *local* statistics such as $P(s)$ or $R_2(s)$, i.e. statistics which only measure the independence of eigenvalues which are separated by at most a finite distance. *Non-local* statistics like the number variance $\Sigma^2(L)$ are well known to violate the Poisson prediction due to non-universal long-range correlations, see Berry [6] and Bleher and Lebowitz [16, 17]. A different type of non-local statistics is connected with the fluctuations of the energy-level counting function ("the spectral staircase") around its mean value. It was shown that for certain integrable systems[7] these fluctuations are non-Gaussian, even though the local statistics follow in general the Poisson prediction.[8]

In the present article, we shall study similar *non-local* statistical properties, which are, however much closer linked to the *local* level spacing statistics. The central object of our investigation will be the *spectral form factor* $K_2(\tau, N)$ (also called *pair correlation form factor*), which is defined as the Fourier transform of the pair correlation density,

$$K_2(\tau, N) = \int_{-\infty}^{\infty} R_2(s, N)\, e(\tau\, s)\, ds$$

$$R_2(s, N) = \int_{-\infty}^{\infty} K_2(\tau, N)\, e(-s\, \tau)\, d\tau$$

[6]During the completion of this manuscript I have learned from A. Eskin that it is possible to prove relation (6) for rectangle billiards with ratio of sides $\alpha^{1/2}$ and α diophantine (e.g. $\alpha = \sqrt{2}$), see [31]. This remarkable result is consistent with Conjecture 1.2 (Section 1.2).

[7]e.g. billiards in a rectangle and tori [34, 9, 10, 14, 13], Liouville surfaces [41, 15], other surfaces of revolution [11], and Zoll surfaces [57]. For a survey see [12].

[8]In contrast, for chaotic systems the fluctuations of the spectral staircase are conjectured to be Gaussian [4, 3], based mainly on numerical evidence.

with $e(z) \equiv \mathrm{e}^{2\pi \mathrm{i} z}$, hence,

$$K_2(\tau, N) = \big| N^{-1/2} \sum_{j=1}^{N} e(\lambda_j \tau) \big|^2. \tag{7}$$

The sum

$$\sum_{j=1}^{N} e(\lambda_j \tau) \tag{8}$$

may be viewed as a walk in the complex plane (Figure 1), consisting of N steps of unit length, whose direction is determined by the phases $\xi_j = 2\pi\lambda_j\tau$. In the case when the λ_j come from a Poisson process, the probability of finding the end point after N steps outside a disk of radius $\sqrt{NR}$ (R is a fixed constant) is given in the limit $N \to \infty$ by a Gaussian limit distribution in the complex plane. This means in particular that

$$\lim_{N\to\infty} \mathrm{Prob}^{\mathrm{Poisson}}\{K_2(\tau, N) > R\} = \int_R^\infty \mathrm{e}^{-r}\, dr = \mathrm{e}^{-R}. \tag{9}$$

If, however, the λ_j are given by a deterministic sequence, we can test the "independence" of the λ_j by considering the distribution of the sum (8), i.e. the distribution of endpoints of the corresponding walk, for different values of τ. That is, we throw τ at random with probability density ρ, and ask if, as above,

$$\lim_{N\to\infty} \mathrm{Prob}^{\rho}\{K_2(\,\cdot\,, N) > R\} = \mathrm{e}^{-R}. \tag{10}$$

In this probabilistic set-up, relation (6) can be formulated as

$$\lim_{N\to\infty} \mathbb{E}^{\rho} K_2(\,\cdot\,, N) = 1 + \rho(0), \tag{11}$$

where $\mathbb{E}^{\rho}$ denotes the expectation, and $h = \hat{\rho}$ and ρ are related by Fourier transformation,

$$\hat{\rho}(s) = \int_{-\infty}^{\infty} \rho(\tau)\, e(s\,\tau)\, ds.$$

The statistical properties of the form factor $K_2(\tau, N)$ have received great attention in the quantum chaos literature, see e.g. [43, 2, 44, 27, 50, 1] and references therein. It is generally believed that the normalized fluctuations of the form factor for a chaotic system are of Gaussian nature [50, 42], cf. also footnote 28 in [2]. The situation for integrable systems seems to be more subtle. We shall see that, even though a generic rectangle billiard is likely to follow Poisson statistics locally [5, 21, 17], the fluctuations of the form factor are not Gaussian but have a limit distribution with algebraic-logarithmic tail.[9] On the other hand, numerical studies of the circle billiard exhibit Gaussian fluctuations of the form factor [61]. If these are truly Gaussian for the circle or other generic integrable cases (e.g. Liouville surfaces), remains an interesting open problem.

1.1 Billiards in a rectangle

The quantum energies of a rectangle billiard are given by the eigenvalues of the negative Laplacian

$$-\Delta = -\frac{\partial^2}{\partial q_1^2} - \frac{\partial^2}{\partial q_2^2}$$

with Dirichlet conditions on the boundary of the rectangle. The eigenvalues scale trivially with the area of the rectangle, so the only parameter which will enter the theory is the side ratio $\alpha^{1/2}$, $0 < \alpha \leq 1$. (The reason why we work with α rather than the side ratio itself will become apparent later.)

Taking the area to be 4π we have eigenvalues

$$E_{m,n}^{(\alpha)} = \frac{\pi}{4}(\alpha^{1/2} m^2 + \alpha^{-1/2} n^2), \quad m, n \in \mathbb{N},$$

which we label in increasing order and with multiplicity by $0 < \lambda_1^{(\alpha)} < \lambda_2^{(\alpha)} \leq \lambda_3^{(\alpha)} \leq \ldots$. This sequence clearly satisfies (1), since the number of lattice points in an ellipse of area λ is asymptotically λ. More precisely, we have

$$\#\{j : \lambda_j^{(\alpha)} \leq \lambda\} = \lambda - \pi^{-1/2}(\alpha^{1/4} + \alpha^{-1/4})\, \lambda^{1/2} + O_\gamma(\lambda^\gamma), \tag{12}$$

where the relation is conjectured to hold for any $\gamma > \frac{1}{4}$; the best bound so far, $\gamma > \frac{23}{73}$, is due to Huxley [36].

1.2 The main results

Let us view τ as a random variable, which is distributed on the compact interval $I \subset \mathbb{R}$ with a piecewise continuous probability density ρ, and α as a random variable with piecewise continuous probability density σ on the compact interval $A \subset \mathbb{R}^+ - \{0\}$. The expectation $\mathbb{E}^{\rho,\sigma}$ for the random variable $f(\tau, \alpha)$ is then defined as

$$\mathbb{E}^{\rho,\sigma} f = \int_I \int_A f(\tau, \alpha)\, \rho(\tau)\ \sigma(\alpha)\, d\tau\, d\alpha,$$

[9]Similar deviations from a Gaussian distribution have been observed by Casati et al. [21], who calculated (numerically) the fluctuations of the difference $P(s,N) - P_{\text{Poisson}}(s)$. Other non-local statistics for rectangle billiards , such as the number variance and the fluctuations of the spectral staircase have been considered e.g. by Casati et al. [20, 21], Berry [6], Bleher et al. [10, 14, 13, 16, 17].

and its probability to be greater than R by

$$\mathrm{Prob}^{\rho,\sigma}\{f > R\} = \int_I \int_A D_R(\tau,\alpha)\,\rho(\tau)\ \sigma(\alpha)\,d\tau\,d\alpha,$$

with the distribution function

$$D_R(\tau,\alpha) = \begin{cases} 1 & \text{if } f(\tau,\alpha) > R \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 1.1 *There exists a decreasing function* $\Psi : \mathbb{R}^+ \to \mathbb{R}^+$ *with*

$$\Psi(0) = 1, \qquad \int_0^\infty \Psi(R)\,dR = 1,$$

$$\Psi(R) \sim c\,R^{-2}\log R \quad \textit{(for } R\to\infty\textit{)},$$

discontinuous only for at most countably many R*, such that,*

$$\lim_{N\to\infty} \mathrm{Prob}^{\rho,\sigma}\{K_2(\,\cdot\,,N) > R\} = \Psi(R),$$

except possibly at the discontinuities of $\Psi(R)$*.*

The constant c is given by the expression

$$c = \frac{2}{\pi^6}\int_{\mathbb{R}^2}\Big|\int_{t_1^2+t_2^2\leq 1} e(t_1^2u_1 + t_2^2u_2)\,d^2t\Big|^4 d^2u. \tag{13}$$

Remarks. **(A)** The condition $\int_0^\infty \Psi(R)\,dR = 1$ is is consistent with a Gaussian limit distribution and thus consistent with the fact that the pair correlation function is the one for Poisson random numbers.

(B) The condition $\Psi(R) \sim c\,R^{-2}\log R$ implies that the limit distribution is non-Gaussian. In particular, the higher moments $\int_0^\infty R^k d\Psi(R)$ $(k > 2)$ diverge. This divergence has the following origin. Just as the moment $k = 1$ is related to the pair correlation function $R_2(s,N)$ (i.e. counting spacings), the moment $k = 2$ is related to the density

$$\frac{1}{N^2}\sum_{j_1,j_2,k_1,k_2=1}^{N} \delta(s - \lambda_{j_1} - \lambda_{j_2} + \lambda_{k_1} + \lambda_{k_2}), \tag{14}$$

and thus to (with the above test function h)

$$\frac{1}{N^2}\sum_{j_1,j_2,k_1,k_2=1}^{N} h(\lambda_{j_1} + \lambda_{j_2} - \lambda_{k_1} - \lambda_{k_2}), \tag{15}$$

which amounts to counting the number of quadruples of eigenvalues such that

$$\lambda_{j_1} + \lambda_{j_2} - \lambda_{k_1} - \lambda_{k_2}$$

is in the interval $[-a, a]$, say. In the case of a rectangle billiard, this number grows like $\gg (N \log N)^2$, due to number-theoretic degeneracies: recall that in this case we count essentially the integers $m_1, \ldots, m_4, n_1, \ldots, n_4 \leq \sqrt{N}$ with

$$(m_1^2 + \alpha n_1^2) + (m_2^2 + \alpha n_2^2) - (m_3^2 + \alpha n_3^2) - (m_4^2 + \alpha n_4^2) \tag{16}$$

in some interval. The number of solutions of

$$m_1^2 + m_2^2 = m_3^2 + m_4^2, \qquad n_1^2 + n_2^2 = n_3^2 + n_4^2$$

with $m_1, \ldots, m_4, n_1, \ldots, n_4 \leq \sqrt{N}$ grows like $\gg (N \log N)^2$, due to Landau's classical result on the number of ways of writing an integer as a sum of two squares. Hence the number of solutions of (16) in any arbitrarily small interval $[-\epsilon, \epsilon]$ grows like $\gg (N \log N)^2$, hence the divergence of the second moment, at least for test functions h with $h(0) \neq 0$. But since h is the Fourier transform of a probability density, we in fact have $h(0) = 1$. This explains the divergence of the moments $k = 2$ and higher. Consequently, we cannot employ the method of moments to prove the above theorem. We shall instead use another approach based on the transformation formulas of theta functions, relating the existence of the limit distribution to the mixing property of certain flows, see Section 5 for details. This is essentially the same idea as in the proofs of the limit theorems of theta sums

$$\sum_{n=1}^{N} e(n^2 x)$$

in [46, 47, 48], which is now generalized to theta sums with more variables. The methods presented here could be further generalized to Siegel theta sums of arbitrary quadratic forms $Q(\xi)$ of d variables,

$$\sum_{\xi \in \mathbb{Z}^d \cap \Lambda N} e(Q(\xi)\, \tau).$$

Here, ΛN denotes some suitable domain $\Lambda \subset \mathbb{R}^d$, which is magnified by a factor of N. It would be interesting to see, in which cases the values of the above theta sums have a limit distribution for τ random. The expectation value of this limit distribution is related to the quantitative version of the Oppenheim conjecture for $Q(\xi)$, cf. Eskin at al. [29, 30] and Borel's survey [19].

(C) Theorem 1.1 implies that there cannot be a set of α of non-zero Lebesgue measure, for which (for fixed α and random τ) $\mathrm{Prob}^\rho\{K_2(\,\cdot\,, N) >$

$R\}$ converges to e^{-R}. We instead have good reasons to believe the following to be true for those α, which are diophantine, i.e., badly approximable by rationals.[10]

Conjecture 1.2 *Theorem 1.1 even holds when α is not random but fixed, as long as α is diophantine. That means in particular*

$$\lim_{N\to\infty} \mathrm{Prob}^{\rho}\{K_2^{(\alpha)}(\,\cdot\,,N) > R\} = \Psi(R),$$

with the same function $\Psi(R)$ as before.

The truth of this conjecture is related to the equidistribution of horocycles in the product space $\Gamma\backslash \mathrm{PSL}(2,\mathbb{R}) \times \Gamma\backslash \mathrm{PSL}(2,\mathbb{R})$ (for details see Section 5, in particular Conjecture 5.6, and Section 6).

The conjecture is obviously false for rational $\alpha = \frac{p}{q}$. The following theorem follows from the results in [39, 46], see Section 7 for details.

Theorem 1.3 *Let ψ be piecewise continuous and of compact support. Then there exists a decreasing function $\Psi^{(\frac{p}{q})} : \mathbb{R}^+ \to \mathbb{R}^+$ with*

$$\Psi^{(\frac{p}{q})}(0) = 1, \qquad \Psi^{(\frac{p}{q})}(R) \sim c^{(\frac{p}{q})}\, R^{-1} \quad (\textit{for } R \to \infty),$$

discontinuous for at most countably many R, such that,

$$\lim_{\lambda\to\infty} \mathrm{Prob}^{\rho}\{K_2^{(\frac{p}{q})}(\,\cdot\,,\lambda) > R\} = \Psi^{(\frac{p}{q})}(R),$$

except possibly at the discontinuities of $\Psi^{(\frac{p}{q})}(R)$.

In the special case $\alpha = \frac{p}{q} = 1$ the constant $c^{(\frac{p}{q})}$ reads $c^{(1)} = 1/\pi$.

Remark. The fact that now $\int_0^\infty \Psi(R)\, dR = \infty$ diverges is due to the well known degeneracy of values of rational quadratic forms at integers (cf. previous remark B). In particular, one has for the pair correlation function (Proposition 7.1)

$$\mathbb{E}^{\rho} K_2^{(\frac{p}{q})}(\,\cdot\,,N) \sim b^{(\frac{p}{q})} \log N \qquad (N \to \infty). \tag{17}$$

For $\alpha = \frac{p}{q} = 1$ we have $b^{(1)} = 1/\pi$. The logarithmic divergence resembles the average number of ways to write an integers as a sum of two squares, which is a classical result by Landau, cf. the previous remark (B). As a consequence, the consecutive level spacing distribution is $P(s) = \delta(s)$ for all rational α. For more information on two-level statistics for the square billiard, cf. also Connors and Keating [24].

[10] The precise definition of a diophantine number will be given later (Section 5). The set of diophantine numbers is of full Lebesgue measure in $\mathbb{R}$.

1.3 Basic definitions and notations

The expressions $x \ll_a y$ and $x = O_a(y)$ both mean *there exists a constant* C_a (which may depend on some additional parameter a) *such that* $|x| \leq C_a|y|$. The notation $x = O(y^{-\infty})$ is an abbreviation for $x = O_M(y^{-M})$ *for every* $M \geq M_0$*, for some suitably large constant* M_0. A *piecewise continuous* function is discontinuous only on a set of measure zero, and bounded on all compact sets. We denote by $\mathcal{S}(\mathbb{R}^n)$ the Schwartz class on $\mathbb{R}^n$, i.e. the space of smooth functions $f(t_1, \ldots, t_n)$ which decrease rapidly when $t_1^2 + \ldots + t_n^2 \to \infty$. The same must hold for all derivatives of f.

2 Invariance properties of the form factor

The pair correlation density $R_2(s, \lambda)$ and the form factor $K_2(\tau, \lambda)$ measure the statistics of levels in an energy window $[0, \lambda]$. For technical reasons it is convenient to smooth this window, i.e., choose a smooth cut-off function $\psi \in \mathrm{C}^\infty(\mathbb{R}^+)$, which is rapidly decreasing at ∞ and consider the smoothed (in λ) pair correlation density

$$R_{2,\psi}(s, \lambda) = \frac{1}{\lambda} \sum_{\lambda_j, \lambda_k} \psi(\frac{\lambda_j}{\lambda})\, \psi(\frac{\lambda_k}{\lambda})\, \delta(s - \lambda_j + \lambda_k), \tag{18}$$

and the smoothed form factor

$$K_{2,\psi}(\tau, \lambda) = \big|\lambda^{-1/2} \sum_{\lambda_j} \psi(\frac{\lambda_j}{\lambda})\, e(\lambda_j\, \tau)\big|^2. \tag{19}$$

In the case of the rectangle billiard the form factor has the explicit expression

$$K_{2,\psi}^{(\alpha)}(\tau, \lambda) = \big|\lambda^{-1/2} \sum_{m,n \in \mathbb{N}} \psi(\frac{\frac{\pi}{4}(\alpha^{1/2} m^2 + \alpha^{-1/2} n^2)}{\lambda}) \times \\ \times e(\frac{\pi}{4}(\alpha^{1/2} m^2 + \alpha^{-1/2} n^2)\, \tau)\big|^2. \tag{20}$$

For symmetry reasons, we can write the sum as

$$4 \sum_{m,n \in \mathbb{N}} = \sum_{m,n \in \mathbb{Z}} - \sum_{m=0, n \in \mathbb{Z}} - \sum_{m \in \mathbb{Z}, n=0} + \sum_{m=0, n=0},$$

The form factor can now be expressed in terms of the theta functions

$$\Theta_f(z) = y^{1/4} \sum_{n \in \mathbb{Z}} f(n\, y^{1/2})\, e(n^2 x), \tag{21}$$

and

$$\Theta_f(z_1;z_2)=y_1^{1/4}y_2^{1/4}\sum_{m,n\in\mathbb{Z}} f(m\,y_1^{1/2},n\,y_2^{1/2})\,e(m^2x_1+n^2x_2), \tag{22}$$

where $z_j=x_j+\mathrm{i}\,y_j$ is a complex variable. Setting

$$f(t)=\psi(t^2), \qquad f(t_1,t_2)=\psi(t_1^2+t_2^2),$$

and

$$z_1=\frac{\pi}{4}\alpha^{1/2}(\tau+\mathrm{i}\,\lambda^{-1}), \qquad z_2=\frac{\pi}{4}\alpha^{-1/2}(\tau+\mathrm{i}\,\lambda^{-1}),$$

we have

$$\begin{aligned} K_{2,\psi}^{(\alpha)}(\tau,\lambda)=\frac{1}{16}\Big|(\frac{\pi}{4})^{-1/2}\,\Theta_f(z_1,z_2) \\ -\lambda^{-1/4}\big[(\frac{\pi}{4}\alpha^{1/2})^{-1/4}\,\Theta_f(z_1)+(\frac{\pi}{4}\alpha^{-1/2})^{-1/4}\,\Theta_f(z_2)\big] \\ +\lambda^{-1/2}\psi(0)\Big|^2. \end{aligned} \tag{23}$$

It is intuitively clear that in the limit $\lambda\to\infty$ the most important contributions should come from the two-variable theta sum. However, for special values of α and τ this need not be the case, but the set of exceptions is fortunately of measure zero. The following lemma follows from standard estimates on theta sums [33].

Lemma 2.1 *For almost all α and τ (with respect to Lebesgue measure) and any $\epsilon>0$ we have*

$$K_{2,\psi}^{(\alpha)}(\tau,\lambda)=\hat{K}_{2,\psi}^{(\alpha)}(\tau,\lambda)+O_{\alpha,\tau,\epsilon}(\lambda^{-1/4+\epsilon}),$$

with

$$\hat{K}_{2,\psi}^{(\alpha)}(\tau,\lambda)=\frac{1}{4\pi}\big|\Theta_f(z_1,z_2)\big|^2.$$

The transformation formulas of theta functions (21) in one complex variable were the starting point of the studies in [46, 47, 48], and can be readily generalized to the two-variable sum $\Theta_f(z_1;z_2)$, by considering each variable separately.

Before stating the transformation formulas, we have to introduce some geometry. Every element g in the Lie group $\mathrm{PSL}(2,\mathbb{R})=\mathrm{SL}(2,\mathbb{R})/\{\pm 1\}$ has a unique Iwasawa decomposition

$$g=\begin{pmatrix}1 & x\\ 0 & 1\end{pmatrix}\begin{pmatrix}y^{1/2} & 0\\ 0 & y^{-1/2}\end{pmatrix}\begin{pmatrix}\cos\phi & -\sin\phi\\ \sin\phi & \cos\phi\end{pmatrix},$$

where $z = x + \mathrm{i}\, y$ is a point in the upper half plane

$$\mathfrak{H} = \{z = x + \mathrm{i}\, y : x, y \in \mathbb{R},\ y > 0\},$$

and $\phi \in [0, \pi)$ parameterizes the circle S^1, so the underlying manifold of $\mathrm{PSL}(2, \mathbb{R})$ may be identified with the manifold $\mathfrak{H} \times \mathrm{S}^1$. The invariant volume element (Haar measure) reads in this choice of coordinates

$$d\mu(g) = \frac{dx\, dy\, d\phi}{y^2}.$$

By virtue of the relation

$$\begin{aligned} &\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \\ &\qquad = \begin{pmatrix} 1 & x' \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y'^{1/2} & 0 \\ 0 & y'^{-1/2} \end{pmatrix} \begin{pmatrix} \cos\phi' & -\sin\phi' \\ \sin\phi' & \cos\phi' \end{pmatrix}, \end{aligned} \tag{24}$$

where

$$x' + \mathrm{i}\, y' = \frac{a(x + \mathrm{i}\, y) + b}{c(x + \mathrm{i}\, y) + d}, \qquad \phi' = \phi + \arg[c(x + \mathrm{i}\, y) + d],$$

the action of $\mathrm{PSL}(2, \mathbb{R})$ on $\mathfrak{H} \times \mathrm{S}^1$ is canonically given by

$$g\,(z, \phi) = \big(gz, \phi + \arg(cz + d)\big), \qquad g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{25}$$

where g acts on $z \in \mathfrak{H}$ by fractional linear transformations, i.e.,

$$g\, z = \frac{az + b}{cz + d}.$$

The *theta group* Γ_θ, which is generated by the elements

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & -1/2 \\ 2 & 0 \end{pmatrix},$$

corresponding to the transformations

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} (z, \phi) = \big(z + 1, \phi\big), \qquad \begin{pmatrix} 0 & -1/2 \\ 2 & 0 \end{pmatrix} (z, \phi) = \big(-\frac{1}{4z}, \phi + \arg z\big),$$

is an example of a discrete subgroup of $\mathrm{PSL}(2, \mathbb{R})$, such that the quotient manifold

$$\mathcal{M}_\theta = \Gamma_\theta \backslash \mathrm{PSL}(2, \mathbb{R}) = \{\Gamma_\theta h \,:\, h \in \mathrm{PSL}(2, \mathbb{R})\}$$

has finite volume $\mu(\mathcal{M}_\theta) = \pi^2$, but is not compact (with respect to the measure introduced above). A fundamental region of Γ_θ in $\mathfrak{H}$ is (Figure 2)

$$\mathcal{F}_\theta = \{z \in \mathfrak{H} \,:\, |x| < 1/2,\ |z| > 1/2\}. \tag{26}$$

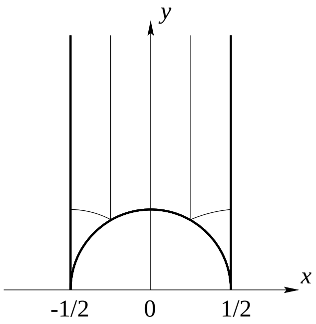

Figure 2: The fundamental region $\mathcal{F}_\theta = \{z \in \mathfrak{H} : |x| < 1/2, |z| > 1/2\}$ in $\mathfrak{H}$ for the theta group Γ_θ. The cusps are represented by the points at ∞ and $\frac{1}{2}$. The thin lines indicate the symmetries of the surface.

There are two cusps which are represented by the points at ∞ and $\frac{1}{2}$ ($-\frac{1}{2}$ is equivalent to $\frac{1}{2}$). It is well known that every rational point on the boundary $\operatorname{Im} z = 0$ of $\mathfrak{H}$ is Γ_θ-equivalent to one of the cusp points, i.e. there is always an element

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_\theta$$

such that

$$\frac{a\frac{p}{q} + b}{c\frac{p}{q} + d} = \infty \text{ or } \frac{1}{2}. \tag{27}$$

We first review some properties of one-variable theta sums, cf. [46, 47, 48]. For $f \in \mathcal{S}(\mathbb{R})$, define the theta function Θ_f as a function on $\mathfrak{H} \times [0, 4\pi)$ by

$$\Theta_f(z, \phi) = y^{1/4} \sum_{n \in \mathbb{Z}} f_\phi(n\, y^{1/2})\, e(n^2 x), \tag{28}$$

where

$$f_\phi(t) = \int_{\mathbb{R}} G_\phi(t, t')\, f(t')\, dt', \tag{29}$$

with the harmonic oscillator Green function

$$G_\phi(t, t') = 2^{1/2} e(-\sigma_\phi/8) |\sin\phi|^{-1/2} e\left[\frac{(t^2 + t'^2)\cos\phi - 2t\,t'}{\sin\phi}\right],$$

where $\sigma_\phi = 2k + 1$ when $k\pi < \phi < (k+1)\pi$ with $k \in \mathbb{Z}$.

Proposition 2.2 *Let $f \in \mathcal{S}(\mathbb{R})$. Then $\Theta_f(z, \phi)$ is infinitely differentiable and satisfies the following functional relations:*

$$\Theta_f(z + 1, \phi) = \Theta_f(z, \phi),$$

$$\Theta_f(-\frac{1}{4z}, \phi + \arg z) = \mathrm{e}^{-\mathrm{i}\pi/4}\, \Theta_f(z, \phi),$$

The function $|\Theta_f|^2$ *may thus be viewed as an infinitely differentiable function on the manifold*

$$\mathcal{M}_\theta = \Gamma_\theta \backslash \mathrm{PSL}(2, \mathbb{R}).$$

The second relation implies $\Theta_f(z, \phi + \pi) = -\mathrm{i}\, \Theta_f(z, \phi)$.

Since our theta function is smooth, it is bounded except in the cusps. In order to state the asymptotic behaviour of the function in the cusp at $\frac{1}{2}$, we introduce a set of new coordinates,

$$(w, \theta) = (-(4z - 2)^{-1}, \phi + \arg(z - 1/2)),$$

in which the cusp at $\frac{1}{2}$ is represented as a cusp at infinity. That is, $v = \operatorname{Im} w$ is the coordinate pointing into the cusp, and $u = \operatorname{Re} w$ the one orthogonal to it.

Proposition 2.3 *Let* $f \in \mathcal{S}(\mathbb{R})$. *Then*

$$\Theta_f(z, \phi) = \begin{cases} y^{1/4} f_\phi(0) + O(y^{-\infty}) & (y > \frac{1}{100}) \\ O(v^{-\infty}) & (v > \frac{1}{100}). \end{cases}$$

The ranges for which the above relations hold, cover the entire fundamental region $\{z \in \mathfrak{H} \,:\, 0 < x < 1,\, |z| > 1/2,\, |z - 1| > 1/2\}$. (This new region is obtained from the old one $\mathcal{F}_\theta$ by shifting the left half by $x \mapsto x + 1$.) Within these ranges, the relations are uniform in (z, ϕ) (1st rel.), (w, θ) (2d rel.).

For $f \in \mathcal{S}(\mathbb{R}^2)$, define the theta function Θ_f as a function on $\mathfrak{H} \times [0, 4\pi) \times \mathfrak{H} \times [0, 4\pi)$ by

$$\Theta_f(z_1, \phi_1; z_2, \phi_2) = y_1^{1/4} y_2^{1/4} \sum_{m,n \in \mathbb{Z}} f_{\phi_1,\phi_2}(m\, y_1^{1/2}, n\, y_2^{1/2})\, e(m^2 x_1 + n^2 x_2), \tag{30}$$

where

$$f_{\phi_1,\phi_2}(t_1, t_2) = \iint_{\mathbb{R}^2} G_{\phi_1}(t_1, t_1')\, G_{\phi_2}(t_2, t_2')\, f(t_1', t_2')\, dt_1' dt_2', \tag{31}$$

with the same harmonic oscillator Green function $G_\phi(t, t')$ as before. It can be readily verified that $f_{\phi_1,\phi_2}(t_1, t_2)$ is again $\in \mathcal{S}(\mathbb{R}^2)$ (use partial integration in t_1 and t_2), so the sum defining Θ_f is rapidly convergent for every (ϕ_1, ϕ_2).

Proposition 2.4 *Let $f \in \mathcal{S}(\mathbb{R}^2)$. Then $\Theta_f(z_1,\phi_1;z_2,\phi_2)$ is infinitely differentiable and satisfies the following functional relations:*

$$\Theta_f(z_1+1,\phi_1;z_2,\phi_2)=\Theta_f(z_1,\phi_1;z_2,\phi_2),$$

$$\Theta_f(-\frac{1}{4z_1},\phi_1+\arg z_1;z_2,\phi_2)=\mathrm{e}^{-\mathrm{i}\pi/4}\,\Theta_f(z_1,\phi_1;z_2,\phi_2),$$

$$\Theta_f(z_2,\phi_2;z_1,\phi_1)=\Theta_f(z_1,\phi_1;z_2,\phi_2).$$

The function $|\Theta_f|^2$ may thus be viewed as an infinitely differentiable function on the manifold

$$\mathcal{M}_\theta^2=\Gamma_\theta\backslash\operatorname{PSL}(2,\mathbb{R})\times\Gamma_\theta\backslash\operatorname{PSL}(2,\mathbb{R}).$$

The second relation implies $\Theta_f(z_1,\phi_1+\pi;z_2,\phi_2)=-\mathrm{i}\,\Theta_f(z_1,\phi_1;z_2,\phi_2)$.

As mentioned above, the proof of the proposition follows exactly the lines of the analogous proposition for one-variable theta functions, compare e.g. [47].

As a fundamental region of $\Gamma_\theta\times\Gamma_\theta$ we choose the set

$$\mathcal{F}=\mathcal{F}_\theta\times[0,\pi)\times\mathcal{F}_\theta\times[0,\pi). \tag{32}$$

The four cusps of codimension two are represented by

$$(\infty,\phi_1;z_2,\phi_2),\quad(\frac{1}{2},\phi_1;z_2,\phi_2),\quad(z_1,\phi_1;\infty,\phi_2),\quad(z_1,\phi_1;\frac{1}{2},\phi_2);$$

every point of the form

$$(\frac{p}{q},\phi_1;z_2,\phi_2),\quad(z_1,\phi_1;\frac{p}{q},\phi_2),\qquad\frac{p}{q}\in\mathbb{Q}$$

is equivalent to one of those four, compare (27).

Proposition 2.5 *Let $f \in \mathcal{S}(\mathbb{R}^2)$. Then*

$$\Theta_f(z_1,\phi_1;z_2,\phi_2)=$$
$$=\begin{cases}(y_1\,y_2)^{1/4}f_{\phi_1,\phi_2}(0,0)+O((y_1\,y_2)^{-\infty}) & (y_1>\frac{1}{100},\ y_2>\frac{1}{100})\\ O((v_1\,v_2)^{-\infty}) & (v_1>\frac{1}{100},\ v_2>\frac{1}{100})\\ O(y_1^{1/4}\,v_2^{-\infty}) & (y_1>\frac{1}{100},\ v_2>\frac{1}{100})\\ O(v_1^{-\infty}\,y_2^{1/4}) & (v_1>\frac{1}{100},\ y_2>\frac{1}{100}).\end{cases}$$

These relations are uniform in $(z_1, \phi_1; z_2, \phi_2)$ (1st rel.), $(w_1, \theta_1; w_2, \theta_2)$ (2d rel.), $(z_1, \phi_1; w_2, \theta_2)$ (3d rel.), $(w_1, \theta_1; z_2, \phi_2)$ (4th rel.). The proof follows directly from Proposition 2 in [46].

The inner products and norms of the L^2 spaces in question are defined as usual by

$$(f, g)_{\mathrm{L}^2(\mathbb{R}^2)} = \iint_{\mathbb{R}^2} f\,\overline{g}\; d^2x, \qquad \|f\|_{\mathrm{L}^2(\mathbb{R}^2)} = \sqrt{(f, f)}_{\mathrm{L}^2(\mathbb{R}^2)} \tag{33}$$

and

$$(F, G)_{\mathrm{L}^2(\mathcal{M}_\theta^2)} = \iint_{\mathcal{M}_\theta^2} F\,\overline{G}\; d^2\mu, \qquad \|F\|_{\mathrm{L}^2(\mathcal{M}_\theta^2)} = \sqrt{(F, F)}_{\mathrm{L}^2(\mathcal{M}_\theta^2)}. \tag{34}$$

By virtue of the last proposition, it is clear that $|\Theta_f|$ is in $\mathrm{L}^2(\mathcal{M}_\theta^2)$. If f is even, i.e. $f(t_1, t_2) = f(-t_1, t_2) = f(t_1, -t_2)$, we have the relation

$$\frac{1}{2\pi^2}\|\Theta_f\|_{\mathrm{L}^2(\mathcal{M}_\theta^2)} = \|f\|_{\mathrm{L}^2(\mathbb{R}^2)}, \tag{35}$$

compare [46], Proposition 4.

3 *Regime I:* $|\tau| \ll \lambda^{-1/2-\epsilon}$

We begin with the regime where $|\tau| \leq C_I \lambda^{-\gamma}$ with $\gamma > \frac{1}{2}$, and C_I an arbitrary constant. The behaviour of the form factor is entirely determined by the values of the theta function in the cusp at ∞, since

$$\hat{K}_{2,\psi}^{(\alpha)}(\tau, \lambda) = \frac{1}{4\pi}|\Theta_f(z_1, 0; z_2, 0)|^2 = \frac{1}{4\pi}|\Theta_f(-\frac{1}{4z_1}, \arg z_1; -\frac{1}{4z_2}, \arg z_2)|^2$$

with

$$-\frac{1}{4z_1} = \frac{1}{\pi\alpha^{1/2}}\,\frac{-\tau + \mathrm{i}\,\lambda^{-1}}{\tau^2 + \lambda^{-2}}, \qquad \arg z_1 = \arg(\tau + \mathrm{i}\,\lambda^{-1}),$$

$$-\frac{1}{4z_2} = \frac{1}{\pi\alpha^{-1/2}}\,\frac{-\tau + \mathrm{i}\,\lambda^{-1}}{\tau^2 + \lambda^{-2}}, \qquad \arg z_2 = \arg(\tau + \mathrm{i}\,\lambda^{-1}).$$

For $\lambda \to \infty$, the imaginary part of $-1/4z_j$ becomes infinitely large, so Proposition 2.5 is applicable, yielding

$$\hat{K}_{2,\psi}^{(\alpha)}(\tau, \lambda) = \frac{1}{4\pi^2}|f_{\arg(\tau+\mathrm{i}\,\lambda^{-1}),\arg(\tau+\mathrm{i}\,\lambda^{-1})}(0, 0)|^2\,\frac{\lambda^{-1}}{\tau^2 + \lambda^{-2}} + O(\lambda^{-\infty}) \tag{36}$$

which holds uniformly for $|\tau| \leq C_I \lambda^{-\gamma}$. It is easy to see that

$$
\begin{aligned}
&|f_{\arg(\tau+\mathrm{i}\,\lambda^{-1}),\arg(\tau+\mathrm{i}\,\lambda^{-1})}(0,0)|^2 \\
&\qquad = 4\lambda^2(\tau^2+\lambda^{-2})\,\Big|\iint e[(t_1^2+t_2^2)\tau\lambda]\, f(t_1,t_2)\, dt_1 dt_2\Big|^2 \\
&\qquad\qquad = 4\lambda^2(\tau^2+\lambda^{-2})\,\Big|\pi\int_0^\infty e(r\tau\lambda)\,\psi(r)\,dr\Big|^2. \qquad (37)
\end{aligned}
$$

Therefore

$$
\hat{K}^{(\alpha)}_{2,\psi}(\tau,\lambda) = \Big|\int_0^\infty e(r\tau\lambda)\,\psi(r)\,dr\Big|^2\,\lambda + O(\lambda^{-\infty}) \qquad (38)
$$

for $\lambda \to \infty$ uniformly in $|\tau| \leq C_I\lambda^{-\gamma}$. In this regime, the form factor looks therefore asymptotically as a delta mass at zero: upon applying a test function ρ we have

$$
\int_{-C_I\lambda^{-\gamma}}^{C_I\lambda^{-\gamma}} \rho(\tau)\hat{K}^{(\alpha)}_{2,\psi}(\tau,\lambda)\,d\tau \sim \int_{-C_I\lambda^{1-\gamma}}^{C_I\lambda^{1-\gamma}} \rho(\tau/\lambda)\Big|\int_0^\infty e(r\tau)\,\psi(r)\,dr\Big|^2 d\tau,
$$

and with $\rho(\tau/\lambda) \sim \rho(0)$ for $\tau \in [-C_I\lambda^{1-\gamma}, C_I\lambda^{1-\gamma}]$ the above reduces to

$$
\sim \rho(0)\int_{-\infty}^{\infty}\Big|\int_0^\infty e(r\tau)\,\psi(r)\,dr\Big|^2 d\tau = \rho(0)\int_0^\infty \psi(r)^2 dr
$$

by Parseval's equality. In summary,

$$
\begin{aligned}
\int_{-C_I\lambda^{-\gamma}}^{C_I\lambda^{-\gamma}} \rho(\tau)K^{(\alpha)}_{2,\psi}(\tau,\lambda)\,d\tau &\sim \int_{-C_I\lambda^{-\gamma}}^{C_I\lambda^{-\gamma}} \rho(\tau)\hat{K}^{(\alpha)}_{2,\psi}(\tau,\lambda)\,d\tau \\
&\sim \rho(0)\int_0^\infty \psi(r)^2 dr. \qquad (39)
\end{aligned}
$$

Regime I is usually referred to as the *saturation regime*, since the number variance $\Sigma^2(L)$ saturates in this regime, in contrast to the prediction for a Poisson process [6, 16].

4 *Regime II:* $\lambda^{-1/2-\epsilon} \ll |\tau| \ll \lambda^{-1/2+\epsilon}$

Regime II is defined as the region where $C_I\lambda^{-1/2-\epsilon} \leq |\tau| \leq C_{II}\lambda^{-1/2+\epsilon}$, with C_I and C_{II} arbitrary positive constants. From the discussion of Regime I it is clear that the form factor in Regime II can at most grow like

$$
K_{2,\psi}(\tau,\lambda) = O_\epsilon(\lambda^{2\epsilon}), \qquad (40)
$$

which is obtained from (38) for $\tau = \lambda^{-1/2-\epsilon}$. The other τ-values in Regime II are bounded farther away from the cusp, so that the above indeed gives an upper bound. The most interesting part of Regime II is when $C_I\lambda^{-1/2} \leq |\tau| \leq C_{II}\lambda^{-1/2}$. Let us put $\omega = \tau\lambda^{1/2}$. Then we have for large λ

$$-\frac{1}{4z_j} = \frac{1}{\pi\alpha^{\pm 1/2}}\,\frac{-\omega\lambda^{-1/2} + \mathrm{i}\,\lambda^{-1}}{\omega^2\lambda^{-1} + \lambda^{-2}} \sim \frac{1}{\pi\alpha^{\pm 1/2}}\,(-\omega^{-1}\lambda^{1/2} + \mathrm{i}\,\omega^{-2})$$

and

$$\arg z_j = \arg(\omega\lambda^{-1/2} + \mathrm{i}\,\lambda^{-1}) \to 0^+.$$

Since the series defining $\Theta_f(z_1,\phi_1;z_2,\phi_2)$ converges uniformly for y_1 and y_2 in compacta, we finally obtain

$$K_{2,\psi}^{(\alpha)}(\tau,\lambda) \sim \hat{K}_{2,\psi}^{(\alpha)}(\tau,\lambda) \sim \hat{K}_{2,\psi}^{(1/\alpha)}\Big(-\frac{4}{\pi^2}\,\frac{\lambda^{1/2}}{\omega},\frac{4}{\pi^2}\,\frac{1}{\omega^2}\Big) \qquad (\lambda\to\infty) \tag{41}$$

uniformly for $|\omega| = |\tau|\lambda^{1/2} \in [C_I, C_{II}]$.

Before turning to Regime III ($\tau \sim$ const.), where we shall study the fluctuations of the form factor around its mean, we will have to prove some facts about the equidistribution of certain sets in $\mathcal{M}_\theta^2$.

5 Ergodicity, mixing and equidistribution

In this this section we discuss some ergodic properties of the geodesic flow on hyperbolic surfaces of finite area, whose unit tangent bundle is represented by the quotient $\mathcal{M} = \Gamma\backslash\,\mathrm{PSL}(2,\mathbb{R})$, where Γ is a discrete subgroup, such as the theta group Γ_θ. The surface should only have a finite number of cusps (≥ 0), and we assume that, if there is at least one cusp, the half-plane coordinates are chosen in such a way that one of the cusps appears as the standard cusp of unit width at ∞.

Consider the following three flows on $\mathcal{M} = \Gamma\backslash\,\mathrm{PSL}(2,\mathbb{R})$, defined by right translation,

$$g \mapsto g\Phi^t, \qquad \Phi^t = \begin{pmatrix} \mathrm{e}^{t/2} & 0 \\ 0 & \mathrm{e}^{-t/2} \end{pmatrix}, \tag{42}$$

and

$$g \mapsto g\Psi_\pm^t, \qquad \Psi_+^t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \qquad \Psi_-^t = \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}. \tag{43}$$

These flows actually represent the geodesic and the (positive and negative) horocycle flow on the unit tangent bundle of the surface $\Gamma\backslash\mathfrak{H}$, which can be identified with $\Gamma\backslash\,\mathrm{PSL}(2,\mathbb{R})$ (the angle ϕ is in fact $-\frac{1}{2}$ times the orientation angle of the tangent vector). These flows are well known to be ergodic and mixing [25]. The mixing property can be stated as follows.

Proposition 5.1 *Let $F, G \in \mathrm{L}^2(\mathcal{M})$. Then*

$$\lim_{t\to\pm\infty} \int_{\mathcal{M}} F(g)\, G(g\Phi^t)\, d\mu(g) = \frac{1}{\mu(\mathcal{M})} \int_{\mathcal{M}} F\, d\mu \int_{\mathcal{M}} G\, d\mu.$$

The mixing property has an interesting consequence for the asymptotic distribution of long arcs of horocycles, which is stated in the following corollary. In fact, the investigation of measures concentrated along unstable fibers (which in our case are the horocycles) is a central issue in the theory of dynamical systems. Our proof will follow an idea of Eskin and McMullen [28], Theorem 7.1.; for related methods, cf. Kleinbock and Margulis [40], Section 2.2.1. and references therein.

Corollary 5.2 *Let F be bounded and piecewise continuous on $\mathcal{M}$, and h be piecewise continuous and of compact support on $\mathbb{R}$. Then, for any $g_0 \in \mathcal{M}$,*

$$\lim_{t\to-\infty} \int_{\mathbb{R}} h(u)\, F(g_0 \Psi_+^u \Phi^t)\, du = \frac{1}{\mu(\mathcal{M})} \int_{\mathcal{M}} F\, d\mu \int_{\mathbb{R}} h\, du.$$

Proof. Every element $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{R})$ with $d \neq 0$ can be written as a product

$$g = \pm \Psi_+^u \Phi^a \Psi_-^b. \tag{44}$$

Let F be continuous on $\mathrm{PSL}(2,\mathbb{R})$, left-invariant under Γ and compactly supported when viewed as a function on $\mathcal{M} = \Gamma \backslash \mathrm{PSL}(2,\mathbb{R})$. Let furthermore H be the function on $\mathrm{PSL}(2,\mathbb{R})$ defined for some fixed $\epsilon > 0$ by

$$H(g) = \begin{cases} h(u)\, \frac{1}{\epsilon}\chi(\frac{a}{\epsilon})\, \chi(b) & \text{for } g = \Psi_+^u \Phi^a \Psi_-^b \\ 0 & \text{for } g = \left(\begin{smallmatrix} a & b \\ c & 0 \end{smallmatrix}\right), \end{cases}$$

where χ is the characteristic function of the interval $[-\frac{1}{2}, \frac{1}{2}]$. Consider the integral

$$\int_{\mathrm{PSL}(2,\mathbb{R})} H(g)\, F(g_0 g \Phi^t)\, d\mu(g) = \int_{\mathrm{PSL}(2,\mathbb{R})} H(\Psi_+^u \Phi^a \Psi_-^b)\, F(g_0 \Psi_+^u \Phi^a \Psi_-^b \Phi^t)\, d\mu. \tag{45}$$

The bi-invariant Haar measure in these coordinates is given by

$$d\mu = \mathrm{e}^{-a}\, da\, db\, du. \tag{46}$$

Step A. Now using the relation $\Psi_-^b \Phi^t = \Phi^t \Psi_-^{b\,\mathrm{e}^t}$ the integral transforms to

$$\int_{\mathrm{PSL}(2,\mathbb{R})} H(\Psi_+^u \Phi^a \Psi_-^b)\, F(g_0 \Psi_+^u \Phi^{a+t} \Psi_-^{b\,\mathrm{e}^t})\, d\mu. \tag{47}$$

The distance (with respect to the invariant metric on $\mathrm{PSL}(2,\mathbb{R})$) between the points $g_0\Psi_+^u\Phi^{a+t}\Psi_-^{b\,\mathrm{e}^t}$ and $g_0\Psi_+^u\Phi^{a+t}$ is $b\,\mathrm{e}^t$, hence

$$F(g_0\Psi_+^u\Phi^{a+t}\Psi_-^{b\,\mathrm{e}^t}) = F(g_0\Psi_+^u\Phi^{a+t}) + O(b\,\mathrm{e}^t), \tag{48}$$

where the implied constant does not depend on u, a, b or t, for F is uniformly continuous. For $-t$ large we thus have

$$\int_{\mathrm{PSL}(2,\mathbb{R})} H(\Psi_+^u\Phi^a\Psi_-^b)\, F(g_0\Psi_+^u\Phi^{a+t}\Psi_-^{b\,\mathrm{e}^t})\, d\mu \\ = \int_{\mathbb{R}^2} h(u)\,\frac{1}{\epsilon}\chi(\frac{a}{\epsilon})\, F(g_0\Psi_+^u\Phi^{a+t})\ \mathrm{e}^{-a}\, da\, du + O(\mathrm{e}^t). \tag{49}$$

Step B. We can rewrite (45) as

$$\int_{\mathrm{PSL}(2,\mathbb{R})} H(g_0^{-1}g)\, F(g\Phi^t)\, d\mu(g) = \int_{\mathcal{M}} \big(\sum_{\gamma\in\Gamma} H(g_0^{-1}\gamma g)\big)\, F(g\Phi^t)\, d\mu(g) \tag{50}$$

since F and $d\mu$ are Γ-left-invariant. The latter, by the mixing property, has the limit (notice that the function $G(g) = \sum_{\gamma\in\Gamma} H(g_0^{-1}\gamma g)$ is Γ-left-invariant)

$$\frac{1}{\mu(\mathcal{M})}\int_{\mathcal{M}} F\ d\mu \int_{\mathcal{M}} \sum_{\gamma\in\Gamma} H(g_0^{-1}\gamma g)\ d\mu(g) \\ = \frac{1}{\mu(\mathcal{M})}\int_{\mathcal{M}} F\ d\mu \int_{\mathrm{PSL}(2,\mathbb{R})} H\ d\mu \\ = \frac{1}{\mu(\mathcal{M})}\int_{\mathcal{M}} F\ d\mu \int_{\mathbb{R}^2} h(u)\,\frac{1}{\epsilon}\chi(\frac{a}{\epsilon})\ \mathrm{e}^{-a}\, da\, du. \tag{51}$$

Step C. Combining Steps A and B we conclude that

$$\lim_{t\to-\infty}\int_{\mathbb{R}^2} h(u)\,\chi(a)\, F(g_0\Psi_+^u\Phi^{\epsilon a+t})\ \mathrm{e}^{-\epsilon a}\, da\, du \\ = \frac{1}{\mu(\mathcal{M})}\int_{\mathcal{M}} F\ d\mu \int_{\mathbb{R}^2} h(u)\,\chi(a)\ \mathrm{e}^{-\epsilon a}\, da\, du. \tag{52}$$

By the uniform continuity of F, given any $\delta > 0$, we find an $\epsilon > 0$ such that

$$\int_{\mathbb{R}^2} h(u)\,\chi(a)\, F(g_0\Psi_+^u\Phi^{\epsilon a+t})\ \mathrm{e}^{-\epsilon a}\, da\, du - \delta \\ < \int_{\mathbb{R}^2} h(u)\,\chi(a)\, F(g_0\Psi_+^u\Phi^{t})\, da\, du \\ < \int_{\mathbb{R}^2} h(u)\,\chi(a)\, F(g_0\Psi_+^u\Phi^{\epsilon a+t})\ \mathrm{e}^{-\epsilon a}\, da\, du + \delta. \tag{53}$$

Since the limits $t \to -\infty$ on the left and right hand side exist for every fixed ϵ, and differ by 2δ, which can be made arbitrarily small, the limit

$$\lim_{t\to-\infty} \int_{\mathbb{R}} h(u)\, F(g_0 \Psi_+^u \Phi^t)\, du = \frac{1}{\mu(\mathcal{M})} \int_{\mathcal{M}} F\, d\mu \int_{\mathbb{R}} h(u)\, du \tag{54}$$

must exist as well. In order to relax the condition of compact support for F notice that the assertion trivially holds for $F \equiv \text{const}$. One can then again use the inclusion principle to relax both the compact support hypothesis and the continuity hypothesis. □

Remarks. **(A)** It is crucial that the limit is $t \to -\infty$. The limit $t \to +\infty$ diverges, since in this case the support of the measure converges towards a single limit point on the boundary of $\mathfrak{H}$.

(B) For $g_0 = 1$ the assertion of the corollary can be stated in the (z, ϕ)-coordinates as

$$\lim_{y\to 0} \int_{\mathbb{R}} h(x) F(z, 0)\, dx = \frac{1}{\mu(\mathcal{M})} \int_{\mathcal{M}} F\, d\mu \int_{\mathbb{R}} h\, dx. \tag{55}$$

In the case when the average is taken over a *closed* horocycle and $h \equiv 1$, this was proved by Sarnak [53]. Hejhal [35] extended the result to h of the above class, but still for closed horocycles and for F which are independent of ϕ, i.e. functions on $\Gamma\backslash\mathfrak{H}$. His proof involves estimates on Poincaré series of weight zero.

The observation of the previous corollary shall now be extended from averages over translates of horocycles to averages over translates of more general curves, which are given by the equation (in suitable coordinates) $y = \mathrm{e}^{f(x)}$, where f is a bounded real function. The case $f \equiv \text{const}$ corresponds to horocycles.

Corollary 5.3 *Let F, h be as before, and let $f : \mathbb{R} \to \mathbb{R}$ be continuous on the support of h. Then, for any $g_0 \in \mathcal{M}$,*

$$\lim_{t\to-\infty} \int_{\mathbb{R}} h(u)\, F(g_0 \Psi_+^u \Phi^{t+f(u)})\, du = \frac{1}{\mu(\mathcal{M})} \int_{\mathcal{M}} F\, d\mu \int_{\mathbb{R}} h\, du.$$

That is, for $g_0 = 1$,

$$\lim_{y\to 0} \int_{\mathbb{R}} h(x) F(x + \mathrm{i}\, y\, \mathrm{e}^{f(x)}, 0)\, dx = \frac{1}{\mu(\mathcal{M})} \int_{\mathcal{M}} F\, d\mu \int_{\mathbb{R}} h\, dx.$$

Proof. Repeat the proof of the previous corollary with

$$H(g) = \begin{cases} h(u)\, \frac{1}{\epsilon} \chi\!\left(\frac{a - f(u)}{\epsilon}\right) \chi(b) & \text{for } g = \Psi_+^u \Phi^a \Psi_-^b \\ 0 & \text{for } g = \left(\begin{smallmatrix} a & b \\ c & 0 \end{smallmatrix}\right), \end{cases}$$

which still has compact support, since f is bounded on the support of h. The only main difference to the previous proof is in Step C where we have instead

$$\frac{1}{\epsilon}\int_{\mathbb{R}^2} h(u)\,\chi\Big(\frac{a-f(u)}{\epsilon}\Big)\,F(g_0\Psi_+^u\Phi^{a+t})\,da\,du \\ = \int_{\mathbb{R}^2} h(u)\,\chi(a)\,F(g_0\Psi_+^u\Phi^{\epsilon a+t+f(u)})\,da\,du, \quad (56)$$

and the conclusion is the same as before, for f is bounded on the domain of integration. □

The next step is to define flows on the product space

$$\mathcal{M}^2 = \Gamma\backslash\,\mathrm{PSL}(2,\mathbb{R}) \times \Gamma\backslash\,\mathrm{PSL}(2,\mathbb{R})$$

by the diagonal action

$$\Gamma\backslash\,\mathrm{PSL}(2,\mathbb{R}) \times \Gamma\backslash\,\mathrm{PSL}(2,\mathbb{R}) \to \Gamma\backslash\,\mathrm{PSL}(2,\mathbb{R}) \times \Gamma\backslash\,\mathrm{PSL}(2,\mathbb{R}), \quad (57)$$

$$(g_1,g_2) \mapsto (g_1,g_2)\Phi^t := (g_1\Phi^t, g_2\Phi^t)$$

and

$$(g_1,g_2) \mapsto (g_1,g_2)\Psi_\pm^t := (g_1\Psi_\pm^t, g_2\Psi_\pm^t).$$

We shall still call these flows *geodesic* or *horocyclic*, respectively. Each one is the direct product of two mixing dynamical systems, and thus mixing itself [25]:

Proposition 5.4 *Let $F, G \in \mathrm{L}^2(\mathcal{M}^2)$. Then*

$$\lim_{t\to\pm\infty}\int_{\mathcal{M}^2} F(g_1,g_2)\,G((g_1,g_2)\Phi^t)\;d^2\mu(g_1,g_2) = \frac{1}{\mu(\mathcal{M}^2)}\int_{\mathcal{M}^2} F\;d^2\mu\int_{\mathcal{M}^2} G\;d^2\mu.$$

The analogue of Corollaries 5.2 and 5.5 is the following:

Corollary 5.5 *Let F be bounded and piecewise continuous on $\mathcal{M}^2$, h be piecewise continuous and of compact support on $\mathbb{R}^2$, and $f=(f_1,f_2):\mathbb{R}^2\to\mathbb{R}^2$ continuous on the support of h. Then, for any $(g_1,g_2)\in\mathcal{M}^2$,*

$$\lim_{t\to-\infty}\int_{\mathbb{R}^2} h(u_1,u_2)\,F(g_1\Psi_+^{u_1}\Phi^{t+f_1(u_1,u_2)}, g_2\Psi_+^{u_2}\Phi^{t+f_2(u_1,u_2)})\;d^2u \\ = \frac{1}{\mu(\mathcal{M}^2)}\int_{\mathcal{M}^2} F\;d^2\mu\int_{\mathbb{R}^2} h\,d^2u.$$

The proof of Corollary 5.2 obviously generalizes to this case. For $(g_1, g_2) = 1$ we have

$$\lim_{y\to 0} \int_{\mathbb{R}^2} h(x_1,x_2) F(x_1 + \mathrm{i}\, y\, \mathrm{e}^{f_1(x_1,x_2)}, 0; x_2 + \mathrm{i}\, y\, \mathrm{e}^{f_2(x_1,x_2)}, 0)\, d^2x$$
$$= \frac{1}{\mu(\mathcal{M}^2)} \int_{\mathcal{M}^2} F\, d^2\mu \int_{\mathbb{R}^2} h\, d^2x. \quad (58)$$

There are reasons to believe that the above equidistribution results do not only hold for two-dimensional averages over (x_1, x_2), but even for averages over one-dimensional lines given by $x_2 = \eta x_1$. In the case when the lattice group Γ is a congruence subgroup of $\mathrm{PSL}(2,\mathbb{R})$ (such as the theta group Γ_θ), it has to be assumed, however, that η is badly approximable by rationals, i.e.

$$\left| \eta - \frac{p}{q} \right| \geq \frac{C}{q^\kappa}$$

for all rationals $\frac{p}{q}$ and some $\kappa \geq 2$. Numbers η satisfying this condition are called *diophantine*. The reason for the necessity of this condition will become clearer in Section 7.

Conjecture 5.6 *Suppose $\mathcal{M} = \Gamma\backslash \mathrm{PSL}(2,\mathbb{R})$ with Γ a congruence subgroup of $\mathrm{PSL}(2,\mathbb{R})$. Let F be bounded and piecewise continuous on $\mathcal{M}^2$, h be piecewise continuous and of compact support on $\mathbb{R}$. If η is diophantine, we have for all $y_1, y_2 > 0$*

$$\lim_{y\to 0} \int_{\mathbb{R}} h(x) F(x + \mathrm{i}\, y_1 y, 0; \eta x + \mathrm{i}\, y_2 y, 0)\, dx = \frac{1}{\mu(\mathcal{M}^2)} \int_{\mathcal{M}^2} F\, d^2\mu \int_{\mathbb{R}} h\, dx.$$

The above observations on the equidistribution of certain sets will now be applied to understand the value distribution of the form factor in the most subtle Regime III, where the order of magnitude of τ is independent of λ.

6 Regime III: $\tau \sim$ const

6.1 The expectation value

The following proposition is the analogue of Sarnak's Proposition 2.1(A) on the weak convergence of the pair correlation function in the case of a two-dimensional family of tori [54]. Here, the weak convergence is with respect to our one-dimensional family of rectangles, parameterized by α. (Recall the definition of ρ and σ in Section 1.2.)

Proposition 6.1 *Let ψ be piecewise continuous and of compact support. Assume furthermore ρ is continuous at 0. Then*

$$\lim_{\lambda\to\infty} \mathbb{E}^{\rho,\sigma} K_{2,\psi}(\,\cdot\,,\lambda) = \big(1+\rho(0)\big) \int_0^\infty \psi(r)^2 dr.$$

Proof. We have to estimate the integral

$$\iint \Big|\lambda^{-1/2} \sum_{m,n\in\mathbb{N}} \psi\Big(\frac{\frac{\pi}{4}(\alpha^{1/2}m^2+\alpha^{-1/2}n^2)}{\lambda}\Big) \times \\ \times e\Big(\frac{\pi}{4}(\alpha^{1/2}m^2+\alpha^{-1/2}n^2)\,\tau\Big)\Big|^2 \rho(\tau)\ \sigma(\alpha)\, d\tau\, d\alpha. \quad (59)$$

Let us first assume all functions involved are infinitely differentiable. For later purposes it is furthermore convenient to assume that the functions ρ and σ are not necessarily probability densities, i.e., have averages also different from one. This freedom will be needed when we approximate piecewise continuous probability densities from above/below by smooth ρ and σ. To be able to use the inclusion principle for ψ, we have to replace (59) by the slightly more general expression

$$\frac{1}{\lambda}\iint \sum_{m_1,n_1,m_2,n_2\in\mathbb{N}} \psi_1\Big(\frac{\frac{\pi}{4}(\alpha^{1/2}m_1^2+\alpha^{-1/2}n_1^2)}{\lambda}\Big)\ \psi_2\Big(\frac{\frac{\pi}{4}(\alpha^{1/2}m_2^2+\alpha^{-1/2}n_2^2)}{\lambda}\Big) \times \\ \times e\Big(\frac{\pi}{4}(\alpha^{1/2}(m_1^2-m_2^2)+\alpha^{-1/2}(n_1^2-n_2^2))\,\tau\Big)\rho(\tau)\ \sigma(\alpha)\, d\tau\, d\alpha. \quad (60)$$

We get back to (59) by setting $\psi_1=\psi_2=\psi$. In the sequel, ψ_1 and ψ_2 are taken to be C^∞ with compact support. Let us split the domain of integration over τ into the three parts corresponding to Regime I, Regime II and Regime III, that is,

$$\int_0^\infty = \int_0^{C_I\lambda^{-1/2-\epsilon}} + \int_{C_I\lambda^{-1/2-\epsilon}}^{C_{II}\lambda^{-1/2+\epsilon}} + \int_{C_{II}\lambda^{-1/2+\epsilon}}^\infty$$

and similarly for the integral over the negative axis. The first integral is similar to the one calculated in Section 3 (where $\psi_1=\psi_2=\psi$), see (39), and we have for large λ

$$\int_{-C_I\lambda^{-1/2-\epsilon}}^{C_I\lambda^{-1/2-\epsilon}} \sim \rho(0)\int_0^\infty \psi_1(r)\,\psi_2(r)dr \int \sigma(\alpha)\,d\alpha. \quad (61)$$

This result can be obtained most easily by replacing the the modulus squared of the theta function, $|\Theta_f(z_1,\phi_1;z_2,\phi_2)|^2$, by the product

$$\Theta_{f_1}(z_1,\phi_1;z_2,\phi_2)\ \overline{\Theta_{f_2}(z_1,\phi_1;z_2,\phi_2)}, \quad (62)$$

with $f_\nu(t_1,t_2)=\psi_\nu(t_1^2+t_2^2)$. The function in (62) can still be viewed as a function on $\mathcal{M}_\theta^2$, cf. Proposition 2.4.

The second integral vanishes to leading order, due to the bound analogous to bound (40) valid in Regime II (Section 4), so

$$\int_{-C_{II}\lambda^{-1/2+\epsilon}}^{-C_I\lambda^{-1/2-\epsilon}}+\int_{C_I\lambda^{-1/2-\epsilon}}^{C_{II}\lambda^{-1/2+\epsilon}} \ll \lambda^{2\epsilon}(C_{II}\lambda^{-1/2+\epsilon}-C_I\lambda^{-1/2-\epsilon}) \ll \lambda^{-1/2+3\epsilon}. \tag{63}$$

We are left with the third integral, where $|\tau|\geq C_{II}\lambda^{-1/2+\epsilon}$, which we split into two parts, the *diagonal* part

$$\frac{1}{\lambda}\iint_{|\tau|\geq C_{II}\lambda^{-1/2+\epsilon}}\sum_{m,n\in\mathbb{N}}\psi_1\Big(\frac{\frac{\pi}{4}(\alpha^{1/2}m^2+\alpha^{-1/2}n^2)}{\lambda}\Big)\times \\ \times\psi_2\Big(\frac{\frac{\pi}{4}(\alpha^{1/2}m^2+\alpha^{-1/2}n^2)}{\lambda}\Big)\,\rho(\tau)\;\sigma(\alpha)\,d\tau\,d\alpha, \tag{64}$$

which clearly converges for $\lambda\to\infty$ to the desired expression[11]

$$\int_0^\infty\psi_1(r)\,\psi_2(r)\,dr\int\rho(\tau)\,d\tau\int\sigma(\alpha)\,d\alpha,$$

and the *off-diagonal* part,

$$\frac{1}{\lambda}\sum_{(m_1,n_1)\neq(m_2,n_2)\in\mathbb{N}^2}\iint_{|\tau|\geq C_{II}\lambda^{-1/2+\epsilon}}\times \\ \times\psi_1\Big(\frac{\frac{\pi}{4}(\alpha^{1/2}m_1^2+\alpha^{-1/2}n_1^2)}{\lambda}\Big)\,\psi_2\Big(\frac{\frac{\pi}{4}(\alpha^{1/2}m_2^2+\alpha^{-1/2}n_2^2)}{\lambda}\Big)\times \\ \times e\Big(\frac{\pi}{4}(\alpha^{1/2}(m_1^2-m_2^2)+\alpha^{-1/2}(n_1^2-n_2^2))\,\tau\Big)\;\rho(\tau)\;\sigma(\alpha)\,d\tau\,d\alpha. \tag{65}$$

Substituting $x_1=\frac{\pi}{4}\alpha^{1/2}\tau$, $x_2=\frac{\pi}{4}\alpha^{-1/2}\tau$ the last expression equals

$$\frac{1}{\lambda}\sum_{(m_1,n_1)\neq(m_2,n_2)\in\mathbb{N}^2}\iint_{|\frac{2}{\pi}\sqrt{x_1x_2}|\geq C_{II}\lambda^{-1/2+\epsilon}}\times \\ \times\psi_1\Big(\frac{\pi}{4}\frac{\sqrt{\frac{x_1}{x_2}}m_1^2+\sqrt{\frac{x_2}{x_1}}n_1^2}{\lambda}\Big)\,\psi_2\Big(\frac{\pi}{4}\frac{\sqrt{\frac{x_1}{x_2}}m_2^2+\sqrt{\frac{x_2}{x_1}}n_2^2}{\lambda}\Big)\times \\ \times e(x_1(m_1^2-m_2^2)+x_2(n_1^2-n_2^2))\;h(x_1,x_2)\,dx_1\,dx_2, \tag{66}$$

where h is related to ρ, σ and the Jacobian of the substitution in the obvious way,

$$h(x_1,x_2)=\frac{4}{\pi}\sqrt{\frac{x_1}{x_2}}\,x_2^{-1}\;\rho\Big(\frac{4}{\pi}\sqrt{x_1x_2}\Big)\,\sigma\Big(\frac{x_1}{x_2}\Big). \tag{67}$$

---

[11]notice that the sum over m,n converges to the corresponding Riemann integral.

Let us divide (66) into the sums

$$\sum_{(m_1,n_1)\neq(m_2,n_2)\in\mathbb{N}^2} = \sum_{m_1\neq m_2, n_1\neq n_2\in\mathbb{N}} + \sum_{m_1\neq m_2, n_1=n_2\in\mathbb{N}} + \sum_{m_1=m_2, n_1\neq n_2\in\mathbb{N}}.$$

Partial integration with respect to x_1 and x_2 shows the first sum converges independent of λ, and, after some manipulations, this part of (66) turns out to be bounded by (use partial integration in x_1, x_2, and then re-substitute the old variables τ and α)

$$\ll \frac{1}{\lambda} \sum_{m_1\neq m_2, n_1\neq n_2\in\mathbb{N}} \iint_{|\tau|\geq C_{II}\lambda^{-1/2+\epsilon}} \frac{\rho(\tau)\,\sigma(\alpha)\,d\tau\,d\alpha}{\tau^2\,|m_1^2-m_2^2|\,|n_1^2-n_2^2|}$$
$$\ll \frac{1}{\lambda}\int_{|\tau|\geq\lambda^{-1/2+\epsilon}} \frac{\rho(\tau)}{\tau^2}\,d\tau \ll \lambda^{-1/2-\epsilon}\int_{|\tau|\geq 1}\frac{\rho(\tau\lambda^{-1/2+\epsilon})}{\tau^2}\,d\tau \ll \lambda^{-1/2-\epsilon}. \quad (68)$$

The remaining two sums can be shown to be of non-leading order by similar means.

We have thus shown so far that for $\lambda\to\infty$

$$\frac{1}{\lambda}\sum_{(m_1,n_1),(m_2,n_2)\in\mathbb{N}^2}\int \psi_1\Big(\frac{\frac{\pi}{4}(\alpha^{1/2}m_1^2+\alpha^{-1/2}n_1^2)}{\lambda}\Big)\,\psi_2\Big(\frac{\frac{\pi}{4}(\alpha^{1/2}m_2^2+\alpha^{-1/2}n_2^2)}{\lambda}\Big)\times$$
$$\times\,\hat\rho\Big(\frac{\pi}{4}(\alpha^{1/2}(m_1^2-m_2^2)+\alpha^{-1/2}(n_1^2-n_2^2))\Big)\,\sigma(\alpha)\,d\alpha$$
$$\longrightarrow \int\psi_1(r)\psi_2(r)\,dr\left(\rho(0)+\int\rho(\tau)\,d\tau\right)\int\sigma(\alpha)\,d\alpha. \quad (69)$$

Taking finite linear combinations $H(r_1,r_2)=\sum\psi_{j_1}(r_1)\,\psi_{j_2}(r_2)$ of functions of the above type, we have

$$\frac{1}{\lambda}\sum_{(m_1,n_1),(m_2,n_2)\in\mathbb{N}^2}\int H\Big(\frac{\frac{\pi}{4}(\alpha^{1/2}m_1^2+\alpha^{-1/2}n_1^2)}{\lambda},\frac{\frac{\pi}{4}(\alpha^{1/2}m_2^2+\alpha^{-1/2}n_2^2)}{\lambda}\Big)\times$$
$$\times\,\hat\rho\Big(\frac{\pi}{4}(\alpha^{1/2}(m_1^2-m_2^2)+\alpha^{-1/2}(n_1^2-n_2^2))\Big)\,\sigma(\alpha)\,d\alpha$$
$$\longrightarrow \int H(r,r)\,dr\left(\rho(0)+\int\rho(\tau)\,d\tau\right)\int\sigma(\alpha)\,d\alpha. \quad (70)$$

Let us now see how the smoothness condition on ψ can be relaxed to piecewise continuous ψ with compact support. Following the lines of the proof of Theorem 3.2 in [51], $\psi(r_1)\psi(r_2)$ can now be approximated from above/below by smooth test functions $H(r_1,r_2)$, which are admissible for the previous derivation and for which

$$\int|H(r,r)-\psi(r)^2|\,dr<\epsilon. \quad (71)$$

By the inclusion principle the statement of the proposition holds thus for piecewise continuous ψ. The result can be further extended to piecewise continuous ρ and σ by approximating both sides of the relation

$$\lim_{\lambda\to\infty} \mathbb{E}^{\rho,\sigma} K_{2,\psi}(\,\cdot\,,\lambda) = \big(1+\rho(0)\big) \int_0^\infty \psi(r)^2 dr.$$

from above and below using smooth functions ρ_ϵ and σ_ϵ, which are ϵ-close[12] to ρ and σ, respectively. □

6.2 The limit distribution — smooth cut-off functions

Let us first consider the simpler case when ψ is a smooth cut-off function. The results will later be extended to more general cut-off functions. We denote by $\mathcal{S}(\mathbb{R}^+)$ the space of even functions in $\mathcal{S}(\mathbb{R})$, restricted to the positive half line $\mathbb{R}^+$.

Theorem 6.2 *Let $\psi \in \mathcal{S}(\mathbb{R}^+)$. Then there exists a decreasing function $\Psi : \mathbb{R}^+ \to \mathbb{R}^+$ with*

$$\Psi_\psi(0) = 1, \qquad \int_0^\infty \Psi_\psi(R)\, dR = \int_0^\infty \psi(r)^2 dr,$$

discontinuous only for at most countably many R, such that,

$$\lim_{\lambda\to\infty} \mathrm{Prob}^{\rho,\sigma}\{K_{2,\psi}(\,\cdot\,,\lambda) > R\} = \Psi_\psi(R),$$

except possibly at the discontinuities of $\Psi_\psi(R)$.

Theorem 6.3 *For $R \to \infty$, we have the asymptotic relation*

$$\Psi_\psi(R) = c_\psi R^{-2} \log R + d_\psi R^{-2} + O_\psi(R^{-\infty}),$$

where

$$c_\psi = \frac{2}{\pi^6} \int_{\mathbb{R}^2} \Big| \int_{\mathbb{R}^2} e(t_1^2 u_1 + t_2^2 u_2)\, \psi(t_1^2 + t_2^2)\, d^2 t \Big|^4 d^2 u.$$

The constant d_ψ is given by a more complicated expression, see the proof of this theorem.

The proofs of these theorems will be given below. The following conjecture is a direct consequence of the equidistribution hypothesis of Conjecture 5.6, compare the proof of Theorem 6.2.

[12] in the L^1 sense but also in the sense that $|\rho(0) - \rho_\epsilon(0)| < \epsilon$, which is where the continuity of ρ at 0 is required.

Conjecture 6.4 *Theorem 6.2 even holds when α is not random but fixed, as long as α is diophantine. That means in particular*

$$\lim_{\lambda\to\infty} \mathrm{Prob}^{\rho}\{K_{2,\psi}^{(\alpha)}(\,\cdot\,,\lambda) > R\} = \Psi_\psi(R),$$

with the same function $\Psi_\psi(R)$ as before.

This conjecture does not hold when α is rational or well approximable by rationals, as we shall see in the next section.

Proof of Theorem 6.2. Apply Corollary 5.5, i.e. relation (58), where $\chi_{\mathcal{D}}$ is taken to be the characteristic function of the set

$$\mathcal{D} = \{(z_1,\phi_1;z_2,\phi_2) \in \mathcal{M}_\theta^2 \,:\, \frac{1}{4\pi}|\Theta_f(z_1,\phi_1;z_2,\phi_2)|^2 > R\}.$$

The boundary of $\mathcal{D}$ can only have positive measure for a countable number of R, otherwise this would be a contradiction to the measurability of Θ_f. Hence $\chi_{\mathcal{D}}$ is piecewise continuous except for countably many R. We chose the function $h(x_1,x_2)$ in a way that

$$\begin{aligned} h(x_1,x_2)\,\chi_{\mathcal{D}}(x_1 + \mathrm{i}\,y\,\mathrm{e}^{f_1(x_1,x_2)},0;x_2+\mathrm{i}\,y\,\mathrm{e}^{f_2(x_1,x_2)},0)\,d^2x \\ = D_R(\tau,\alpha,\lambda)\,\rho(\tau)\,\sigma(\alpha)\,d\tau\,d\alpha, \end{aligned} \quad (72)$$

where

$$x_1 = \frac{\pi}{4}\alpha^{1/2}\tau, \qquad x_2 = \frac{\pi}{4}\alpha^{-1/2}\tau,$$

$$y = \frac{\pi}{4}\lambda^{-1}, \qquad f_1(x_1,x_2) = \log\sqrt{\frac{x_1}{x_2}}, \qquad f_2(x_1,x_2) = \log\sqrt{\frac{x_2}{x_1}},$$

and

$$D_R(\tau,\alpha,\lambda) = \begin{cases} 1 & \text{if } K_{2,\psi}^{(\alpha)}(\tau,\lambda) > R \\ 0 & \text{otherwise.} \end{cases}$$

Since the compact interval A does not contain 0, the functions f_1 and f_2 are bounded on the domain of integration. Hence, except for at most countably many R,

$$\lim_{\lambda\to\infty} \mathrm{Prob}^{\rho,\sigma}\{\hat{K}_{2,\psi}(\,\cdot\,,\lambda) > R\} = \Psi_\psi(R). \quad (73)$$

To obtain the same result for $K_{2,\psi}$, recall that by virtue of (23) we can express the difference between $K_{2,\psi}$ and $\hat{K}_{2,\psi}$ as the sum

$$\lambda^{-1/4}F_1(z_1,\phi_1;z_2,\phi_2) + \lambda^{-1/2}F_2(z_1,\phi_1;z_2,\phi_2) \\ + \lambda^{-3/4}F_3(z_1,\phi_1;z_2,\phi_2) + \lambda^{-1}\psi(0)^2,$$

where the F_j are sums of products of theta functions and their modulus is therefore majorized by functions on $\mathcal{M}_\theta^2$; it is therefore clear that

$$\lim_{\lambda\to\infty} \mathrm{Prob}^{\rho,\sigma}\{|K_{2,\psi}(\,\cdot\,,\lambda) - \hat{K}_{2,\psi}(\,\cdot\,,\lambda)| > \epsilon\} = 0 \tag{74}$$

for arbitrarily small (but fixed) $\epsilon > 0$. Hence for any given $\epsilon > 0$, we have

$$\begin{aligned} &\mathrm{Prob}^{\rho,\sigma}\{\hat{K}_{2,\psi}(\,\cdot\,,\lambda) > R+\epsilon\} \\ &\qquad\qquad \le \mathrm{Prob}^{\rho,\sigma}\{K_{2,\psi}(\,\cdot\,,\lambda) > R\} \\ &\qquad\qquad\qquad\qquad \le \mathrm{Prob}^{\rho,\sigma}\{\hat{K}_{2,\psi}(\,\cdot\,,\lambda) > R-\epsilon\}, \end{aligned} \tag{75}$$

for every large enough λ. The limits $\Psi_\psi(R+\epsilon)$ and $\Psi_\psi(R-\epsilon)$ of the left and right hand side exist except for countably many R, ϵ. Since

$$\{(z_1,\phi_1;z_2,\phi_2) \in \mathcal{M}_\theta^2 \,:\, \frac{1}{4\pi}|\Theta_f(z_1,\phi_1;z_2,\phi_2)|^2 = R\}.$$

has positive measure only for countably many R, the difference $\Psi_\psi(R-\epsilon) - \Psi_\psi(R+\epsilon)$ can be made arbitrarily small for suitable small $\epsilon > 0$, except for countably many R.

Finally, the relation $\Psi_\psi(0) = 1$ holds by definition,[13] and the integral of Ψ_ψ can be calculated as follows (notice that $\mu(\mathcal{M}_\theta^2) = \pi^4$),

$$\begin{aligned} &\int_0^\infty \Psi_\psi(R)\, dR \\ &= \frac{1}{\mu(\mathcal{M}_\theta^2)} \int_0^\infty \mu\{(z_1,\phi_1;z_2,\phi_2) \in \mathcal{F} \,:\, \frac{1}{4\pi}|\Theta_f(z_1,\phi_1;z_2,\phi_2)|^2 > R\}\, dR \\ &\qquad = \frac{1}{4\pi^5}\int_{\mathcal{M}_\theta^2} |\Theta_f|^2\, d^2\mu = \frac{1}{\pi}\int_{\mathbb{R}^2} |f(t_1,t_2)|^2\, d^2t = \int_0^\infty \psi(r)^2 dr, \end{aligned} \tag{76}$$

by virtue of relation (35). □

Proof of Theorem 6.3. It follows from the proof of Theorem 6.2 that for all but countably many R (recall the definition of the fundamental region $\mathcal{F}$ in (32))

$$\begin{aligned} \Psi_\psi(R) &= \frac{1}{\pi^4}\mu\{(z_1,\phi_1;z_2,\phi_2) \in \mathcal{F} \,:\, \frac{1}{4\pi}|\Theta_f(z_1,\phi_1;z_2,\phi_2)|^2 > R\} \\ &= \frac{2}{\pi^4}\mu\{(z_1,\phi_1;z_2,\phi_2) \in \mathcal{F} \,:\, y_1 > y_2 \text{ and } \frac{1}{4\pi}|\Theta_f(z_1,\phi_1;z_2,\phi_2)|^2 > R\}. \end{aligned} \tag{77}$$

The asymptotics of $\Psi_\psi(R)$ is clearly determined by the large values of the theta function $\Theta_f(z_1,\phi_1;z_2,\phi_2)$ in the cusps. By Proposition 2.5 and the refined asymptotic relation

$$\Theta_f(z_1,\phi_1;z_2,\phi_2) = y_1^{1/4}\Theta_f(\phi_1;z_2,\phi_2) + O(y_1^{-\infty}) \tag{78}$$

[13] It might happen that $\Psi_\psi(\epsilon) \le C < 1$ for arbitrarily small ϵ, but since we allow $\Psi_\psi(R)$ to be discontinuous at countably many R, it is most sensible to normalize $\Psi_\psi(0) = 1$.

with the theta function

$$\Theta_f(\phi_1; z_2, \phi_2) = y_2^{1/4} \sum_{n\in\mathbb{Z}} f_{\phi_1,\phi_2}(0, n\, y_2^{1/2})\, e(n^2 x_2),$$

we thus have

$$\begin{aligned}\Psi_\psi(R) = {} & \frac{2}{\pi^4}\mu\{(z_1,\phi_1;z_2,\phi_2)\in\mathcal{F} \,:\, y_1 > y_2, \\ & \quad y_1 > 10 \text{ and } \frac{1}{4\pi}\, y_1^{1/2}|\Theta_f(\phi_1;z_2,\phi_2)|^2 > R\} + O(R^{-\infty}) \\ = {} & \frac{2}{\pi^4}\int_0^\pi \int_{\mathcal{F}_\theta}\int_0^\pi \min\big\{\frac{1}{y_2}, \frac{1}{10}, (4\pi R)^{-2}|\Theta_f(\phi_1;z_2,\phi_2)|^4\big\} d\phi_1\, \frac{dx_2\, dy_2\, d\phi_2}{y_2^2} \\ & + O(R^{-\infty}). \quad (79)\end{aligned}$$

Let us first discuss the integral over the range

$$\frac{1}{y_2} < (4\pi R)^{-2}|\Theta_f(\phi_1;z_2,\phi_2)|^4, \qquad \text{i.e. } (4\pi R)^2 < y_2|\Theta_f(\phi_1;z_2,\phi_2)|^4,$$

which, for large R, requires y_2 to be large as well. From the asymptotic relation

$$\Theta_f(\phi_1;z_2,\phi_2) = \begin{cases} y_2^{1/4} f_{\phi_1,\phi_2}(0,0) + O(y_2^{-\infty}) & (y_2 > \frac{1}{100}) \\ O(v_2^{-\infty}) & (v_2 > \frac{1}{100}), \end{cases} \qquad (80)$$

compare Proposition 2.5, it follows that the integral over the range in concern is bounded from above and below by the same integral over the ranges $y_2|f_{\phi_1,\phi_2}(0,0)|^2 > 4\pi R \mp C_M R^{-M}$ (for any $M > 2$ and a suitable constant C_M) which now can be worked out to give, for R large enough,

$$\begin{aligned}&\frac{2}{\pi^4}\int_0^\pi\int_0^\pi\int_{(4\pi R\mp C_M R^{-M})/|f_{\phi_1,\phi_2}(0,0)|^2}^\infty \int_0^1 \min\big\{\frac{1}{y_2},\frac{1}{10}\big\}\frac{dx_2\,dy_2\,d\phi_2}{y_2^2}\, d\phi_1 \\ &\quad = \frac{2}{\pi^4}\int_0^\pi\int_0^\pi\int_{(4\pi R\mp C_M R^{-M})/|f_{\phi_1,\phi_2}(0,0)|^2}^\infty \int_0^1 \frac{dx_2\,dy_2\,d\phi_2}{y_2^3}\, d\phi_1 \\ &\quad = \frac{1}{\pi^4}\,(4\pi R)^{-2}\int_0^\pi\int_0^\pi |f_{\phi_1,\phi_2}(0,0)|^4\, d\phi_1\, d\phi_2 + O_M(R^{-M}). \quad (81)\end{aligned}$$

As to the remaining range $(4\pi R)^2 > y_2|\Theta_f(\phi_1;z_2,\phi_2)|^4$, the same reasoning as before permits to give the upper and lower bounds

$$\begin{aligned}&\frac{2}{\pi^4}\int_0^\pi\int_0^\pi\int_{\mathcal{F}_\theta((4\pi R\pm C_M R^{-M})/|f_{\phi_1,\phi_2}(0,0)|^2)} \times \\ &\qquad \times \min\big\{\frac{1}{10}, (4\pi R)^{-2}|\Theta_f(\phi_1;z_2,\phi_2)|^4\big\}\frac{dx_2\,dy_2\,d\phi_2}{y_2^2}\, d\phi_1 \quad (82)\end{aligned}$$

with the truncated fundamental region

$$\mathcal{F}_\theta(T) = \{z \in \mathcal{F}_\theta \;:\; y < T\}.$$

The above integral, restricted to the range $\frac{1}{10} < (4\pi R)^{-2} |\Theta_f(\phi_1; z_2, \phi_2)|^4$, equals

$$\frac{2}{10\pi^4} \int_0^\pi \int_0^\pi \int_{(*)} \frac{dy_2}{y_2^2}\, d\phi_2\, d\phi_1 + O(R^{-\infty}) = O(R^{-\infty}), \tag{83}$$

where the range $(*)$ of integration of the inner integral is

$$\frac{(4\pi R)^2}{10\, |f_{\phi_1,\phi_2}(0,0)|^4} \leq y_2 \leq \frac{4\pi R}{|f_{\phi_1,\phi_2}(0,0)|^2}.$$

In order to work out the integral over the range $\frac{1}{10} > (4\pi R)^{-2} |\Theta_f(\phi_1; z_2, \phi_2)|^4$ let us define the truncated function

$$H_f(\phi_1; z_2, \phi_2) = \begin{cases} |\Theta_f(\phi_1; z_2, \phi_2)|^4 - y_2 |f_{\phi_1,\phi_2}(0,0)|^4 & \text{for } y_2 > 1 \\ |\Theta_f(\phi_1; z_2, \phi_2)|^4 & \text{otherwise,} \end{cases} \tag{84}$$

which is rapidly decreasing in all cusps. The integral over the range under consideration can now be expressed as

$$\begin{aligned} &\frac{1}{8\pi^6 R^2} \int_0^\pi \int_0^\pi \int_{\mathcal{F}_\theta} H_f(\phi_1; z_2, \phi_2) \frac{dx_2\, dy_2\, d\phi_2}{y_2^2}\, d\phi_1 \\ &\quad + \frac{1}{8\pi^6 R^2} \int_0^\pi \int_0^\pi \int_1^{4\pi R/|f_{\phi_1,\phi_2}(0,0)|^2} y_2 |f_{\phi_1,\phi_2}(0,0)|^4 \frac{dy_2}{y_2^2}\, d\phi_2\, d\phi_1 \\ &\qquad + O(R^{-\infty}). \end{aligned} \tag{85}$$

The second integral yields

$$\begin{aligned} &\frac{1}{8\pi^6 R^2} \log R \int_0^\pi \int_0^\pi |f_{\phi_1,\phi_2}(0,0)|^4\, d\phi_1\, d\phi_2 \\ &\quad + \frac{1}{8\pi^6 R^2} \int_0^\pi \int_0^\pi |f_{\phi_1,\phi_2}(0,0)|^4 \log\left(4\pi/|f_{\phi_1,\phi_2}(0,0)|^2\right) d\phi_1\, d\phi_2. \end{aligned} \tag{86}$$

Collecting all leading-order terms in the above estimates, we obtain

$$\begin{aligned} \Psi_\psi(R) = {}& \frac{1}{8\pi^6} \left\{ \int_0^\pi \int_0^\pi |f_{\phi_1,\phi_2}(0,0)|^4\, d\phi_1\, d\phi_2 \right\} R^{-2} \log R \\ &+ \frac{1}{8\pi^6} \left\{ \int_0^\pi \int_0^\pi |f_{\phi_1,\phi_2}(0,0)|^4 \log\left(4\pi/|f_{\phi_1,\phi_2}(0,0)|^2\right) d\phi_1\, d\phi_2 \right. \\ &\quad + \int_0^\pi \int_0^\pi \int_{\mathcal{F}_\theta} H_f(\phi_1; z_2, \phi_2) \frac{dx_2\, dy_2\, d\phi_2}{y_2^2}\, d\phi_1 \\ &\qquad \left. + \frac{1}{2} \int_0^\pi \int_0^\pi |f_{\phi_1,\phi_2}(0,0)|^4\, d\phi_1\, d\phi_2 \right\} R^{-2} + O(R^{-\infty}). \end{aligned} \tag{87}$$

The coefficients c_ψ and d_ψ are thus determined. By virtue of (31) we have, after some change of variables,

$$\int_0^\pi \int_0^\pi |f_{\phi_1,\phi_2}(0,0)|^4 \, d\phi_1 \, d\phi_2 = 2^4 \int_{\mathbb{R}^2} \Big| \int_{\mathbb{R}^2} e(t_1^2 u_1 + t_2^2 u_2) \, f(t_1,t_2) \, d^2 t \Big|^4 d^2 u, \tag{88}$$

which concludes the proof of Theorem 6.3. □

6.3 The limit distribution — general cut-off functions

We shall now assume only that f (and thus ψ) is piecewise continuous and of compact support.

Theorem 6.5 *Let ψ be piecewise continuous and of compact support. Then there exists a decreasing function $\Psi : \mathbb{R}^+ \to \mathbb{R}^+$ with*

$$\Psi_\psi(0) = 1, \qquad \int_0^\infty \Psi_\psi(R) \, dR = \int_0^\infty \psi(r)^2 dr,$$

discontinuous only for at most countably many R, such that,

$$\lim_{\lambda\to\infty} \mathrm{Prob}^{\rho,\sigma}\{K_{2,\psi}(\,\cdot\,,\lambda) > R\} = \Psi_\psi(R),$$

except possibly at the discontinuities of $\Psi_\psi(R)$.

Proof. The proof follows along the lines of the proof of Theorem 4 in [46]. For a given $\epsilon > 0$ choose a function $\psi_\epsilon \in \mathrm{C}^\infty(\mathbb{R}^+)$ of compact support, such that

$$\int_0^\infty |\psi(r) - \psi_\epsilon(r)|^2 \, dr < \epsilon. \tag{89}$$

The crucial observation is that, for λ large enough, we have

$$\int_I \int_A K^{(\alpha)}_{2,\psi-\psi_\epsilon}(\tau,\lambda) \, d\tau \, d\alpha \le C\epsilon, \tag{90}$$

for some constant C independent of λ and ϵ. This fact is a consequence of Proposition 6.1 and condition (89).

Consider the set

$$S_y^\epsilon = \{\tau \in I, \alpha \in A : \; K^{(\alpha)}_{2,\psi-\psi_\epsilon}(\tau,\lambda) < \epsilon^{1/2}\}.$$

The integral over the complement of this set must satisfy

$$C\epsilon > \iint_{(I\times A)-S_y^\epsilon} K^{(\alpha)}_{2,\psi-\psi_\epsilon}(\tau,\lambda)\,d\tau\,d\alpha \geq \iint_{(I\times A)-S_y^\epsilon} \epsilon^{1/2}\,d\tau\,d\alpha,$$

hence

$$|S_y^\epsilon| > |I|\,|A| - C\epsilon^{1/2}. \tag{91}$$

Define the distribution function

$$D_{R,\psi}(\tau,\alpha,\lambda) = \begin{cases} 1 & \text{if } K^{(\alpha)}_{2,\psi}(\tau,\lambda) > R \\ 0 & \text{otherwise,} \end{cases} \tag{92}$$

and the probability

$$\Psi_{\psi,y}(R) = \int_I \int_A D_{R,\psi}(\tau,\alpha,\lambda)\,\rho(\tau)\ \sigma(\alpha)\,d\tau\,d\alpha. \tag{93}$$

By virtue of (91) we have the inclusions

$$\Psi_{\psi_\epsilon,y}(R+\epsilon^{1/2}) - C'\epsilon^{1/2} \leq \Psi_{\psi,y}(R) \leq \Psi_{\psi_\epsilon,y}(R-\epsilon^{1/2}) + C'\epsilon^{1/2}, \tag{94}$$

where C' does not depend on ϵ, y, R. By Theorem 6.2, for $y\to 0$, the left and right hand side have the limits

$$\lim_{y\to 0}\Psi_{\psi_\epsilon,y}(R\pm\epsilon^{1/2}) = \Psi_{\psi_\epsilon}(R\pm\epsilon^{1/2}) \tag{95}$$

except for countably many R, ϵ. Some analysis shows (for details compare [46]) that for every $\delta>0$ there is an $\epsilon>0$ such that

$$\Psi_{\psi_\epsilon}(R-\epsilon^{1/2}) - \Psi_{\psi_\epsilon}(R+\epsilon^{1/2}) < \delta \tag{96}$$

(except for countably many R). Hence there is a function $\Psi_\psi(R)$ such that

$$\lim_{y\to 0}\Psi_{\psi,y}(R) = \Psi_\psi(R), \tag{97}$$

which proves the claim. □

6.4 Random walks

Proof of Theorem 1.1. The task is to show that the difference between

$$\mathrm{Prob}^{\rho,\sigma}\left\{\left|\frac{1}{\sqrt{N}}\sum_{j=1}^N e(\lambda_j^{(\alpha)}\tau)\right|^2 > R\right\}$$

and

$$\mathrm{Prob}^{\rho,\sigma}\left\{\left|\frac{1}{\sqrt{N}}\sum_{\lambda_j^{(\alpha)}\leq N} e(\lambda_j^{(\alpha)}\tau)\right|^2 > R\right\}$$

vanishes for $N \to \infty$, since then Theorem 1.1 will follow from Theorem 6.5. First notice that

$$\left|\frac{1}{\sqrt{N}}\sum_{j=1}^{N} e(\lambda_j^{(\alpha)}\tau) - \frac{1}{\sqrt{N}}\sum_{\lambda_j^{(\alpha)}\leq N} e(\lambda_j^{(\alpha)}\tau)\right|^2 = \left|\frac{1}{\sqrt{N}}\sum_{j\in X_N^{(\alpha)}} e(\lambda_j^{(\alpha)}\tau)\right|^2 \tag{98}$$

where $X_N^{(\alpha)}$ is the set of j defined by

$$X_N^{(\alpha)} = \{j \leq N \,:\, \lambda_j^{(\alpha)} > N\} \cup \{j > N \,:\, \lambda_j^{(\alpha)} \leq N\}.$$

From the asymptotic relation (12) it can be readily seen that the number of elements in $X_N^{(\alpha)}$ is bounded by

$$\# X_N^{(\alpha)} \ll \sqrt{N}, \tag{99}$$

where the implied constant does not depend on $\alpha \in A$ (A fixed). It is therefore clear from the calculation done to prove Proposition 6.1 that

$$\int_I \int_A \left|\frac{1}{\sqrt{N}}\sum_{j\in X_N^{(\alpha)}} e(\lambda_j^{(\alpha)}\tau)\right|^2 d\tau\, d\alpha \tag{100}$$

vanishes for large N. Hence we can use the inclusion principle in the same fashion as in the proof of Theorem 6.5 (the vanishing of (100) is the analogue of relation (90)) to prove the existence of the limit $\Psi(R)$, which furthermore has to equal $\Psi(R) = \Psi_\chi(R)$, with χ the characteristic function of the interval $[0,1]$. □

7 *Rational* α

The simplest case is $\alpha = 1$, since $z_1 = z_2$ and so the form factor is related to the theta function $\Theta_f(z,\phi;z,\phi)$ by

$$\hat{K}_{2,\psi}^{(1)}(\tau,\lambda) = \frac{1}{4\pi}|\Theta_f(z,0;z,0)|^2.$$

The function $F(z,\phi) = |\Theta_f(z,\phi;z,\phi)|^2$ can now be viewed as a function on the manifold $\mathcal{M}_\theta$, which is embedded as a three-dimensional submanifold in $\mathcal{M}_\theta^2$, and we can apply the theorems which were developed in the beginning of Section 5. This observation holds in a similar way for all rational $\alpha = \frac{p}{q}$;

however, the corresponding embedded three-manifold $\mathcal{M}_{\frac{p}{q}}$ becomes densely distributed in $\mathcal{M}_\theta^2$ when the sequence of rationals $\frac{p}{q}$ approaches an irrational. Let us discuss this in more detail.

For a given integer N we define the congruence subgroups $\Gamma_0(N)$ of $\mathrm{SL}(2,\mathbb{Z})$ by

$$\Gamma_0(N)=\left\{\begin{pmatrix} a & b\\ c & d\end{pmatrix}\in\mathrm{SL}(2,\mathbb{Z})\;:\; c\equiv 0 \mod N\right\}. \tag{101}$$

With

$$\eta_N=\begin{pmatrix} N & 0\\ 0 & 1\end{pmatrix}$$

we find [58] that

$$\Gamma_0(N)=\eta_N^{-1}\,\mathrm{SL}(2,\mathbb{Z})\eta_N\cap\mathrm{SL}(2,\mathbb{Z}). \tag{102}$$

The index of these subgroups in $\mathrm{SL}(2,\mathbb{Z})$ is finite, more precisely [58],

$$[\mathrm{SL}(2,\mathbb{Z}):\Gamma_0(N)]=N\prod_{\substack{p\ \mathrm{prime}\\ p|N}}(1+p^{-1}). \tag{103}$$

Consider now the form factor for rational $\alpha=\frac{p}{q}$,

$$\hat{K}_{2,\psi}^{(\frac{p}{q})}(\tau,\lambda)=\frac{1}{4\pi}|\Theta_f(z_1,0;z_2,0)|^2=\frac{1}{4\pi}|\Theta_f(pz,0;qz,0)|^2, \tag{104}$$

with

$$z=\frac{z_1}{p}=\frac{z_2}{q}=\frac{\pi}{4\sqrt{pq}}\,(\tau+\mathrm{i}\,\lambda^{-1}).$$

For $\tilde{\gamma}=\eta_p^{-1}\gamma\eta_p$ with $\gamma=\left(\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right)\in\Gamma_0(4)$ we have the functional relations

$$\begin{aligned}&|\Theta_f(p\tilde{\gamma}z,\phi+\arg(pc\,z+d);qz,\phi)|^2\\ &\qquad=|\Theta_f(\gamma pz,\phi+\arg(c\,pz+d);qz,\phi)|^2=|\Theta_f(pz,\phi;qz,\phi)|^2,\end{aligned} \tag{105}$$

and similarly

$$|\Theta_f(pz,\phi;q\tilde{\gamma}z,\phi+\arg(qc\,z+d))|^2=|\Theta_f(pz,\phi;qz,\phi)|^2 \tag{106}$$

for $\tilde{\gamma}=\eta_q^{-1}\gamma\eta_q$ with $\gamma\in\Gamma_0(4)$. Therefore the function

$$|\Theta_f^{(\frac{p}{q})}(z,\phi)|^2=|\Theta_f(pz,\phi;qz,\phi)|^2 \tag{107}$$

is invariant under the group

$$\Gamma^{(\frac{p}{q})}=\eta_p^{-1}\Gamma_0(4)\eta_p\cap\eta_q^{-1}\Gamma_0(4)\eta_q,$$

which, by virtue of (102), contains the congruence group $\Gamma_0(4pq) \subset \Gamma^{(\frac{p}{q})}$; hence $\Gamma^{(\frac{p}{q})}$ is of finite index in $\mathrm{SL}(2,\mathbb{Z})$ and $\mathcal{M}_{\frac{p}{q}} = \Gamma^{(\frac{p}{q})} \backslash \mathrm{SL}(2,\mathbb{R})$ has finite volume.

Using the theory developed in [46, 48] and the equidistribution of horocycles (Corollary 5.2) we can now prove the analog statements of the previous sections, but now for rational α.

7.1 The expectation value

Proposition 7.1 *Let ψ be piecewise continuous and of compact support. Then, for $\lambda \to \infty$,*

$$\mathbb{E}^\rho K_{2,\psi}^{(\frac{p}{q})}(\,\cdot\,,\lambda) \sim b_\psi^{(\frac{p}{q})} \log \lambda$$

for some constant $b_\psi^{(\frac{p}{q})}$. In the case $\frac{p}{q} = 1$ we have in particular

$$b_\psi^{(1)} = \frac{1}{\pi} \int_0^\infty \psi(r)^2 \, dr.$$

Sketch of the proof. The function $|\Theta_f^{(\frac{p}{q})}(z,\phi)|^2$ on $\mathcal{M}_{\frac{p}{q}} = \Gamma^{(\frac{p}{q})} \backslash \mathrm{SL}(2,\mathbb{R})$ is not bounded so Corollary 5.2 is not directly applicable. However, there is a way of resolving this difficulty using a regularization with Eisenstein series, see [46, 48] for details. Compare alternatively Jurkat and van Horne [39] for a different approach. □

7.2 The limit distribution

Theorem 7.2 *Let ψ be piecewise continuous and of compact support. Then there exists a decreasing function $\Psi_\psi^{(\frac{p}{q})}$ with $\Psi_\psi^{(\frac{p}{q})}(0) = 1$, discontinuous for at most countably many R, such that,*

$$\lim_{\lambda \to \infty} \mathrm{Prob}^\rho \{ K_{2,\psi}^{(\frac{p}{q})}(\,\cdot\,,\lambda) > R \} = \Psi_\psi^{(\frac{p}{q})}(R),$$

except possibly at the discontinuities of $\Psi_\psi^{(\frac{p}{q})}(R)$. For large R,

$$\Psi_\psi^{(\frac{p}{q})}(R) \sim c_\psi^{(\frac{p}{q})} R^{-1},$$

with some constant $c_\psi^{(\frac{p}{q})}$, which in the special case $\frac{p}{q} = 1$ reads

$$c_\psi^{(1)} = \frac{1}{\pi} \int_0^\infty \psi(r)^2 \, dr.$$

Sketch of the proof. Simply use Corollary 5.2, and proceed as in the proof of Theorems 6.2 and 6.5. Compare also the corresponding theorems in [46, 47, 48] and Jurkat and van Horne's results [37, 38, 39]. □

Acknowledgments

I thank G. Casati, A. Eskin, M.R. Haggerty, D.A. Hejhal, J.P. Keating, R.E. Prange, Z. Rudnick, U. Smilansky, G. Tanner and S. Zelditch for helpful discussions. The research was carried out at the Basic Research Institute in the Mathematical Sciences, Hewlett-Packard Laboratories, Bristol, and at the Programme on Disordered Systems and Quantum Chaos, Isaac Newton Instiute, Cambridge. I would like to thank both Institutes for their hospitality.

References

[1] H. Alt, H.-D. Gräf, T. Guhr, H.L. Harney, R. Hofferbert, H. Rehfeld, A. Richter and P. Schardt, Correlation-hole method for the spectra of superconducting microwave billiards, *Phys. Rev.* E **55** (1997) 6674-6683.

[2] N. Argaman, Y. Imry and U. Smilansky, Semiclassical analysis of spectral correlations in mesoscopic systems, *Phys. Rev.* B **47** (1993) 4440-4457.

[3] R. Aurich, A. Bäcker and F. Steiner, Mode fluctuations as fingerprints of chaotic and non-chaotic systems, *Int. J. Mod. Phys.* B **11** (1997) 805-.

[4] R. Aurich, J. Bolte and F. Steiner, Universal signatures of quantum chaos, *Phys. Rev. Lett.* **73** (1994) 1356-.

[5] M.V. Berry and M. Tabor, Level clustering in the regular spectrum, *Proc. Roy. Soc.* A **356** (1977) 375-394.

[6] M.V. Berry, Semiclassical theory of spectral rigidity, *Proc. Roy. Soc.* A **400** (1985) 229-251.

[7] P.M. Bleher, The energy level spacing for two harmonic oscillators with golden mean ratio of frequencies, *J. Statist. Phys.* **61** (1990) 869-876.

[8] P.M. Bleher, The energy level spacing for two harmonic oscillators with generic ratio of frequencies, *J. Statist. Phys.* **63** (1991) 261-283.

[9] P.M. Bleher, On the distribution of the number of lattice points inside a family of convex ovals, *Duke Math. J.* **67** (1992) 461-481.

[10] P.M. Bleher, Distribution of the error term in the Weyl asymptotics for the Laplace operator on a two-dimensional torus and related lattice problems, *Duke Math. J.* **70** (1993) 655-682.

[11] P.M. Bleher, Distribution of energy levels of a quantum free particle on a surface of revolution, *Duke Math. J.* **74** (1994) 45-93.

[12] P.M. Bleher, Trace formula for quantum integrable systems, lattice point problem, and small divisors, in: D. Hejhal, F. Chung, J. Friedman, M. Gutzwiller and A. Odlyzko (eds.), *Emerging Applications of Number Theory* (Springer, to appear).

[13] P.M. Bleher, Z. Cheng, F.J. Dyson and J.L. Lebowitz, Distribution of the error term for the number of lattice points inside a shifted circle, *Comm. Math. Phys.* **154** (1993) 433-469.

[14] P.M. Bleher, F.J. Dyson and J.L. Lebowitz, Non-Gaussian energy level statistics for some integrable systems, *Phys. Rev. Lett.* **71** (1993) 3047-3050.

[15] P.M. Bleher, D.V. Kosygin and Ya.G. Sinai, Distribution of energy levels of quantum free particle on the Liouville surface and trace formulae, *Comm. Math. Phys.* **170** (1995) 375-403.

[16] P.M. Bleher and J.L. Lebowitz, Energy-level statistics of model quantum systems: universality and scaling in a lattice-point problem, *J. Statist. Phys.* **74** (1994) 167-217.

[17] P.M. Bleher and J.L. Lebowitz, Variance of number of lattice points in random narrow elliptic strip, *Ann. Inst. H. Poincaré Probab. Statist.* **31** (1995) 27-58.

[18] O. Bohigas, M.-J. Giannoni and C. Schmit, Characterization of chaotic quantum spectra and universality of level fluctuation laws, *Phys. Rev. Lett.* **52** (1984) 1-4.

[19] A. Borel, Values of indefinite quadratic forms at integral points and flows on spaces of lattices, *Bull. Amer. Math. Soc.* **32** (1995) 184-204.

[20] G. Casati, I. Guarneri and F. Valz-Gris, Degree of randomness of the sequence of eigenvalues, *Phys. Rev.* A **30** (1984) 1586-1588.

[21] G. Casati, B.V. Chirikov and I. Guarneri, Energy-level statistics of integrable quantum systems, *Phys. Rev. Lett.* **54** (1985) 1350-1353.

[22] Y. Colin de Verdière, Quasi-modes sur les varietes Riemanniennes, *Invent. Math.* **43** (1977) 15-52.

[23] Y. Colin de Verdière, Sur le spectre des opérateurs elliptiques à bicaractéristiques toutes periodique, *Comment. Math. Helvetici* **54** (1979) 508-522.

[24] R.D. Connors and J.P. Keating, Two-point spectral correlations for the square billiard, *J. Phys.* A **30** (1997) 1817-1830.

[25] I.P. Cornfeld, S.V. Fomin and Ya.G. Sinai, *Ergodic Theory* (Springer, 1982).

[26] J.J. Duistermaat and V.W. Guillemin, The spectrum of positive elliptic operators and periodic bicharacteristics, *Invent. Math.* **29** (1975) 39-79.

[27] B. Eckhardt and J. Main, Semiclassical form factor of matrix element fluctuations, *Phys. Rev. Lett.* **75** (1995) 2300-2303.

[28] A. Eskin and C. McMullen, Mixing, counting, and equidistribution in Lie groups, *Duke Math. J.* **71** (1993) 181-209.

[29] A. Eskin, G. Margulis and S. Mozes, On a quantitative version of the Oppenheim conjecture, *Electron. Res. Announc. Amer. Math. Soc.* **1** (1995) 124-130.

[30] A. Eskin, G. Margulis and S. Mozes, Upper bounds and asymptotics in a quantitative version of the Oppenheim conjecture, *Ann. Math.*, to appear.

[31] A. Eskin, G. Margulis and S. Mozes, Eigenvalue spacings on square 2-tori, in preparation.

[32] C. Greenman, The generic spacing distribution of the two-dimensional harmonic oscillator, *J. Phys.* A **29** (1996) 4065-4081.

[33] G.H. Hardy and J.E. Littlewood, Some problems in diophantine approximation. II, *Acta Math.* **37** (1914) 193-239.

[34] D.R. Heath-Brown, The distribution and moments of the error term in the Dirichlet divisor problem, *Acta Arith.* **60** (1992) 389-415.

[35] D. Hejhal, On value distribution properties of automorphic functions along closed horocycles, *XVIth Rolf Nevanlinna Colloquium (Joensuu, 1995)*, 39-52 (de Gruyter, 1996).

[36] M.N. Huxley, Exponential sums and lattice points II, *Proc. London Math. Soc.* **66** (1993) 279-301, Corrigenda *ibid.* **68** (1994) 264.

[37] W.B. Jurkat and J.W. van Horne, The proof of the central limit theorem for theta sums, *Duke Math. J.* **48** (1981) 873-885.

[38] W.B. Jurkat and J.W. van Horne, On the central limit theorem for theta series, *Michigan Math. J.* **29** (1982) 65-77.

[39] W.B. Jurkat and J.W. van Horne, The uniform central limit theorem for theta sums, *Duke Math. J.* **50** (1983) 649-666.

[40] D.Y. Kleinbock and G.A. Margulis, Bounded orbits of nonquasiunipotent flows on homogeneous spaces, in: L.A. Bunimovich et al. (ed.), *Sinai's Moscow Seminar on Dynamical Systems*, 141-172, AMS Transl. (2) **171**, Amer. Math. Soc., Providence, RI, 1996.

[41] D.V. Kosygin, A.A. Minasov and Ya.G. Sinai, Statistical properties of the spectra of Laplace-Beltrami operators on Liouville surfaces, *Russian Math. Surveys* **48** (1993) 1-142

[42] P. Lebœuf and G. Iacomelli, Statistical properties of the time evolution of complex systems. I (Preprint, Orsay 1997) `cond-mat/970970`.

[43] L. Leviandier, M. Lombardi, R. Jost and J.P. Pique, Fourier transform: A tool to measure statistical level properties in very complex spectra, *Phys. Rev. Lett.* **56** (1986) 2449-2452.

[44] M. Lombardi and T.H. Seligman, Universal and nonuniversal statistical properties of levels and intensities for chaotic Rydberg molecules, *Phys. Rev.* A **47** (1993) 3571-3586.

[45] P. Major, Poisson law for the number of lattice points in a random strip with finite area, *Probab. Theory Related Fields* **92** (1992) 423-464.

[46] J. Marklof, Limit theorems for theta sums, *Duke Math. J.*, to appear.

[47] J. Marklof, Theta sums, Eisenstein series, and the semiclassical dynamics of a precessing spin, in: D. Hejhal, F. Chung, J. Friedman, M. Gutzwiller and A. Odlyzko (eds.), *Emerging Applications of Number Theory* (Springer, to appear).

[48] J. Marklof, *Limit Theorems for Theta Sums with Applications in Quantum Mechanics*, Dissertation, Universität Ulm, 1997 (Shaker Verlag, Aachen, 1997).

[49] A. Pandey, O. Bohigas and M.-J. Giannoni, Level repulsion in the spectrum of two-dimensional harmonic oscillators, *J. Phys.* A **22** (1989) 4083-4088.

[50] R.E. Prange, The spectral form factor is not self-averaging, *Phys. Rev. Lett.* **78** (1997) 2280-2283.

[51] Z. Rudnick and P. Sarnak, Zeros of principal L-functions and random matrix theory, *Duke Math. J.* **81** (1996) 269-322.

[52] Z. Rudnick and P. Sarnak, The pair correlation function for fractional parts of polynomials, *Comm. Math. Phys.*, to appear.

[53] P. Sarnak, Asymptotic behavior of periodic orbits of the horocycle flow and Eisenstein series, *Comm. Pure Appl. Math.* **34** (1981) 719-739.

[54] P. Sarnak, Values at integers of binary quadratic forms, *Harmonic analysis and number theory* (Montreal, 1996), 181-203, CMS Conf. Proc. **21**, Amer. Math. Soc., Providence, RI, 1997.

[55] P. Sarnak, Quantum chaos, symmetry and zeta functions. Lecture I: Quantum chaos, *Curr. Dev. Math.* (1997) 84-101.

[56] P. Sarnak, Quantum chaos, symmetry and zeta functions. Lecture II: Zeta functions, *Curr. Dev. Math.* (1997) 102-115.

[57] R. Schubert, The trace formula and the distribution of eigenvalues of Schrödinger operators on manifolds all of whose geodesics are closed, *DESY-Report* 95-090 (May 1995).

[58] G. Shimura, *Introduction to the Arithmetic Theory of Automorphic Functions* (Princeton Univ. Press, 1971).

[59] Ya.G. Sinai, Poisson distribution in a geometric problem, *Dynamical systems and statistical mechanics* (Moscow, 1991), 199-214, Adv. Soviet Math. **3**, Amer. Math. Soc., Providence, RI, 1991.

[60] N.B. Slater, Gaps and steps for the sequence $n\theta \mod 1$, *Proc. Cambridge Philos. Soc.* **63** (1967) 1115-1123.

[61] G. Tanner, private communication.

[62] A. Uribe and S. Zelditch, Spectral statistics on Zoll surfaces, *Comm. Math. Phys.* **154** (1993) 313-346.

[63] J.M. VanderKam, Values at integers of homogeneous polynomials, *Duke Math. J.*, to appear.

[64] J.M. VanderKam, Pair correlation of four-dimensional flat tori, (Preprint, Princeton University, 1997).

[65] A. Weinstein, Asymptotics of eigenvalue clusters for the Laplacian plus a potential, *Duke Math. J.* **44** (1977) 883-892.

[66] S. Zelditch, Level spacings for quantum maps in genus zero (Preprint, Isaac Newton Institute, Cambridge 1997).