Vol. 97, No. 1 DUKE MATHEMATICAL JOURNAL

LIMIT THEOREMS FOR THETA SUMS

JENS MARKLOF

Contents

1. Introduction. The classical theta sum is defined by

$$S_N(x) = N^{-1/2}\sum_{n=1}^{N} e\big(n^2 x\big), \tag{1}$$

where $e(t) = \exp(2\pi i t)$. We are interested in the asymptotic behaviour of $S_N(x)$ as N tends to infinity, for arbitrary values of x in $\mathbb{R}$. The case when x is rational is the easiest. It is not hard to see that here $S_N(x) = A(x)N^{1/2} + O(N^{-1/2})$ with some constant $A(x)$ (which can be zero for certain x), for $S_N(x)$ reduces to ordinary Gauss sums. The more difficult part of giving estimates for generic values of x was first discussed by Hardy and Littlewood [3], [4], using diophantine approximation. Their methods were later refined in a number of publications, some of which we will mention here. In contrast to these approaches, we investigate the asymptotic behaviour of theta sums by means of ergodic theory, exploiting a connection to the geodesic flow and the horocycle flow on the unit tangent bundle of a certain hyperbolic surface.

In order to simplify the presentation of the main ideas, we replace the sharp cutoff in the sum (1) by a smooth one; that is, we take a C^∞-function f (with compact support, say) and consider the sum

$$\widetilde{S}_N(x) = N^{-1/2}\sum_{n\in\mathbb{Z}} f(n/N)e\big(n^2 x\big). \tag{2}$$

Choosing f to be the characteristic function of the interval $(0,1]$ would clearly lead

Received 4 September 1996. Revision received 31 October 1997.

1991 *Mathematics Subject Classification*. Primary 11F27; Secondary 11K60, 11L15, 58F17.

Author's research supported by Gottlieb Daimler- und Karl Benz-Stiftung under grant number 2.94.36.

back to (1). There is no need to restrict N to the integers. It is convenient to put $y = N^{-2}$ and to view $\widetilde{S}_N(x)$ as a function

$$\Theta_f(z) = y^{1/4} \sum_{n\in\mathbb{Z}} f\big(ny^{1/2}\big) e\big(n^2 x\big) \tag{3}$$

on the complex upper half-plane $\mathfrak{H} = \{z = x + \mathrm{i}\, y : y > 0\}$. In the following, we only assume that f is of Schwartz class, that is, not necessarily of compact support, but still smooth and rapidly decreasing. By that, the important case of the classical theta series

$$\theta_0(z) = y^{1/4} \sum_{n\in\mathbb{Z}} e\big(n^2 z\big), \tag{4}$$

which corresponds to a cutoff function $f(t) = \exp(-2\pi t^2)$, is also included in our considerations.

The main result is concerned with the distribution of values of $\Theta_f(z)$ in the complex plane $\mathbb{C}$. Depending on the choice of x as rational, real quadratic, or generic, different kinds of patterns are created in the complex plane by the values of $\Theta_f(x + \mathrm{i}\, y)$, as $y \to 0$ and x remains fixed. For $S_N(x)$ these "curlicues" are studied, for example, by Dekking and Mendès France [2] and by Berry and Goldberg [1] (see also references therein). Theorem 5.2 explains how the points of such a pattern are distributed in the plane, provided the cutoff function f is smooth. For any open subset $\mathcal{B} \subset \mathbb{C}$ with nice boundary ("nice" is defined in Section 5), we prove the existence of the limit

$$\lim_{T\to\infty} \frac{1}{T} \Big| \big\{ t \in [0, T] : \Theta_f\big(x + \mathrm{i}\, \mathrm{e}^{-t}\big) \in \mathcal{B} \big\} \Big| \tag{5}$$

for almost all $x \in \mathbb{R}$ (with respect to Lebesgue measure). $|\cdot|$ denotes the standard Lebesgue measure of a subset of $\mathbb{R}$. The limit (5) has the same value for all generic x and is given by Theorem 5.2. We see later on that the limiting distribution is not a normal Gaussian distribution. Theorem 5.2 is proved by identifying the set $\{x + \mathrm{i}\, \mathrm{e}^{-t} : t \in [0, T]\}$ with a generic geodesic on the unit tangent bundle $\mathrm{T}_1 M$ (or, rather, a fourfold cover of it) of the noncompact hyperbolic surface $M = \Gamma_1(4)\backslash\mathfrak{H}$, where $\Gamma_1(4)$ is a congruence subgroup of $\mathrm{SL}(2, \mathbb{Z})$, and then by applying the ergodic theorem for the geodesic flow on $\mathrm{T}_1 M$, going back to Hopf [7].

If the cut-off function $f(t)$ is real valued and monotonically increasing for $t < 0$ and is monotonically decreasing for $t > 0$, we show that the curlicue represented by the set

$$\big\{ \Theta_f\big(x + \mathrm{i}\, \mathrm{e}^{-t}\big) : t \in [0, T] \big\}$$

becomes densely distributed in the complex plane, as T tends to infinity compare with Proposition 5.4.

The limit (5) does also exist for all real quadratic $x \notin \mathbb{Q}$, but has a different, nonuniversal value because here the above set can be identified with a geodesic that approximates a closed geodesic, as $T \to \infty$. In the case $x \in \mathbb{Q}$, the relevant geodesics are scattering orbits that run into a cusp of M.

Instead of keeping x fixed and varying y, we can also look at the curves created by the values of $\Theta_f(x+\mathrm{i}\,y)$, when x runs from 0 to 1 and y remains fixed. These curves are clearly closed since $\Theta_f(x+\mathrm{i}\,y)$ is periodic in x with period 1. Theorem 5.3 states the existence of the limit (with $\mathcal{B}$ as above)

$$\lim_{y\to 0}\left|\{x \in [0,1] : \Theta_f(x+\mathrm{i}\,y) \in \mathcal{B}\}\right|, \tag{6}$$

which is proved this time by identifying the set $\{x+\mathrm{i}\,y : x \in [0,1]\}$ with a closed horocycle and applying the equidistribution theorem of long, closed horocycles due to Sarnak [13]. Besides that, we also determine the asymptotic behaviour of the αth moment

$$\int_0^1 |\Theta_f(x+\mathrm{i}\,y)|^\alpha\, dx;$$

compare with Theorem 6.1.

As in the case of the curlicues considered first, the set

$$\{\Theta_f(x+\mathrm{i}\,y) : x \in [0,1]\}$$

becomes densely distributed in $\mathbb{C}$ as y tends to zero, when $f(t)$ is real valued and monotonically increasing for $t<0$ and is monotonically decreasing for $t>0$; compare with Proposition 5.5. The limiting distributions for (5) and (6) are, in fact, the same.

In Section 7 we give a density argument that shows that the limit (6),

$$\lim_{y\to 0}\left|\left\{x \in [0,1] : N^{-1/2}\sum_{n=1}^{N} e\big(n^2 x\big) \in \mathcal{B}\right\}\right|,$$

exists also in the case of the classical theta sum. The extension of Theorem 5.2 seems harder in this respect and is not discussed in this article.

Theorems 5.3, 6.1, and 7.3 imply results of Jurkat and van Horne [8], [9], who studied the functions $\theta_0(z)$ and $S_N(x)$ by significantly different methods based on diophantine approximation. Their uniform limit theorem in [10] may be derived in an analogous way as a consequence of the equidistribution of arcs of long, closed horocycles (compare [6]), but we do not treat this case here.

2. Basic definitions. The square root $z^{1/2}$ of a complex number z is to be chosen such that $-\pi/2 < \arg z^{1/2} \le \pi/2$. Also, let $z^{m/2} := (z^{1/2})^m$ for any integer $m \in \mathbb{Z}$.

The action of $\mathrm{SL}(2,\mathbb{R})$ on the upper half-plane $\mathfrak{H} = \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$ is defined via fractional linear transformations; that is,

$$g : z \mapsto gz = \frac{az+b}{cz+d}, \qquad g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{R}). \tag{7}$$

Let

$$\widetilde{\mathrm{SL}}(2,\mathbb{R})=\{[g,\beta_g]: g\in \mathrm{SL}(2,\mathbb{R}),\ \beta_g \text{ a function on } \mathfrak{H} \text{ s.t. } \mathrm{e}^{\mathrm{i}\beta_g(z)}=\epsilon_g(z)\},$$

$$\epsilon_g(z)=(cz+d)/|cz+d|, \tag{8}$$

be the universal covering group of $\mathrm{SL}(2,\mathbb{R})$, with the multiplication law

$$\big[g,\beta_g^1\big]\big[h,\beta_h^2\big]=\big[gh,\beta_{gh}^3\big], \qquad \beta_{gh}^3(z)=\beta_g^1(hz)+\beta_h^2(z). \tag{9}$$

The inverse of $[g,\beta_g]$ is then $[g,\beta_g]^{-1}=\big[g^{-1},\beta'_{g^{-1}}\big]$ with $\beta'_{g^{-1}}(z)=-\beta_g(g^{-1}z)$.

For an integer m, put

$$Z_m=\big\langle[-1,\beta_{-1}]^m\big\rangle, \quad \text{with } \beta_{-1}(z)=\pi, \tag{10}$$

that is, the cyclic subgroup generated by $[-1,\beta_{-1}]^m$. Z_m is contained in the center of $\widetilde{\mathrm{SL}}(2,\mathbb{R})$, and in particular $\mathrm{PSL}(2,\mathbb{R})\simeq\widetilde{\mathrm{SL}}(2,\mathbb{R})/Z_1$, $\mathrm{SL}(2,\mathbb{R})\simeq\widetilde{\mathrm{SL}}(2,\mathbb{R})/Z_2$, and so on.

We may identify $\widetilde{\mathrm{SL}}(2,\mathbb{R})$ with $\mathfrak{H}\times\mathbb{R}$ via

$$\big[g,\beta_g\big]\mapsto(z,\phi)=\big(g\,\mathrm{i},\beta_g(\mathrm{i})\big). \tag{11}$$

The action of $\widetilde{\mathrm{SL}}(2,\mathbb{R})$ on $\mathfrak{H}\times\mathbb{R}$ is then canonically defined by

$$\big[g,\beta_g\big](z,\phi)=\big(gz,\phi+\beta_g(z)\big). \tag{12}$$

The congruence subgroups

$$\Gamma_0(N)=\left\{\begin{pmatrix}a&b\\c&d\end{pmatrix}\in\mathrm{SL}(2,\mathbb{Z}): c\equiv 0 \bmod N\right\}, \tag{13}$$

$$\Gamma_1(N)=\left\{\begin{pmatrix}a&b\\c&d\end{pmatrix}\in\Gamma_0(N): a\equiv d\equiv 1 \bmod N\right\}, \tag{14}$$

where N is a positive integer, play a central role in our investigation. Furthermore, put

$$\Delta_1(4)=\left\{\big[g,\beta_g\big]: g\in\Gamma_1(4),\ \beta_g \text{ a function on } \mathfrak{H} \text{ s.t. } \mathrm{e}^{\mathrm{i}\beta_g(z)/2}=\left(\frac{c}{d}\right)\epsilon_g(z)^{1/2}\right\}, \tag{15}$$

which is a discrete subgroup of $\widetilde{\mathrm{SL}}(2,\mathbb{R})$, as shown in the next section. For an integer a and an odd integer b, the symbol $\left(\frac{a}{b}\right)$ is characterized by the following properties (cf. [16]):

(i) $\left(\frac{a}{b}\right)=0$ if $(a,b)\neq 1$;

(ii) if b is an odd prime, $\left(\frac{a}{b}\right)$ coincides with the ordinary quadratic residue symbol;
(iii) if $b > 0$, $\left(\frac{\cdot}{b}\right)$ defines a character modulo b;
(iv) if $a \neq 0$, $\left(\frac{a}{\cdot}\right)$ defines a character modulo a divisor of $4a$, whose conductor is the conductor of $\mathbb{Q}(\sqrt{a})$ over $\mathbb{Q}$;
(v) $\left(\frac{a}{-1}\right) = \operatorname{sgn} a$;
(vi) $\left(\frac{0}{\pm 1}\right) = 1$.

In particular, $\left(\frac{a}{b}\right)^2$=1, if $(a,b) = 1$.

3. Theta functions as functions on $\Delta_1(4)\backslash\widetilde{\mathrm{SL}}(2,\mathbb{R})$**.** It is well known that the hyperbolic manifold $\Gamma_1(4)\backslash\mathfrak{H}$ is not compact but is of finite measure with respect to the canonical Riemann measure $dx\,dy/y^2$. We see that $\Delta_1(4)$ is a discrete group containing the subgroup $Z_4 = \{[1,\beta_1] : \beta_1(z) = 4\pi n, n \in \mathbb{Z}\}$; hence $\mathcal{M} = \Delta_1(4)\backslash\widetilde{\mathrm{SL}}(2,\mathbb{R})$ is of finite measure, too, with respect to the invariant measure $dx\,dy\,d\phi/y^2$. A fundamental region of $\Delta_1(4)$ in $\mathfrak{H}\times\mathbb{R}$ is, for instance, $\mathcal{F}_{\Delta_1(4)} = \mathcal{F}_{\Gamma_1(4)} \times [0,4\pi)$, if $\mathcal{F}_{\Gamma_1(4)}$ is a fundamental region of $\Gamma_1(4)$ in $\mathfrak{H}$. One may take, for example,

$$\mathcal{F}_{\Gamma_1(4)} = \{z \in \mathfrak{H} : x \in (-1/2, 1/2], |z+1/4| > 1/4, |z-1/4| \geq 1/4\}. \tag{16}$$

Denote by $\mathcal{S}(\mathbb{R})$ the Schwartz class of $\mathrm{C}^\infty(\mathbb{R})$ functions of rapid decrease.

PROPOSITION 3.1. *For every $f \in \mathcal{S}(\mathbb{R})$, there exists a function $\Theta_f \in \mathrm{C}^\infty(\mathcal{M})$ such that*

$$\Theta_f(z,0) = y^{1/4}\sum_{n\in\mathbb{Z}} f\big(ny^{1/2}\big)e\big(n^2x\big).$$

Proof. The kth Hermite function h_k reads

$$h_k(t) = \big(2^{k-1}k!\big)^{-1/2} H_k\big(2\pi^{1/2}t\big)\,\mathrm{e}^{-2\pi t^2}, \tag{17}$$

with the Hermite polynomial

$$H_k(t) = (-1)^k\,\mathrm{e}^{t^2}\,\frac{d^k}{dt^k}\,\mathrm{e}^{-t^2}.$$

The h_k form an orthonormal basis of $\mathrm{L}^2(\mathbb{R})$ with respect to the standard scalar product, that is, $(h_j, h_k) = \delta_{jk}$. Let $\hat{f}(k) = (f, h_k)$ be the kth Hermite coefficient. Since $f \in \mathcal{S}(\mathbb{R})$, the Hermite expansion

$$\sum_{k=0}^{N} \hat{f}(k)h_k(t) \to f(t) \tag{18}$$

converges uniformly in t due to the bounds

$$\big|\hat{f}(k)\big| \leq C_m(2k+1)^{-m}, \quad \text{for any } m, \tag{19}$$

and

$$\left|h_k(t)\right| \leq \begin{cases} C\left((2k+1)^{1/3} + \left|4\pi t^2 - (2k+1)\right|\right)^{-1/4}, & 2\pi t^2 \leq 2k+1, \\ C\,\mathrm{e}^{-\gamma t^2}, & 2\pi t^2 > 2k+1, \end{cases} \tag{20}$$

with positive constants C_m, C, γ. (For this and more, see [19].)

Let us consider the theta function

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

with the transformation property (see [17])

$$\theta_k(gz) = j_g(z)^{2k+1}\theta_k(z), \quad \text{for every } g \in \Gamma_0(4), \tag{22}$$

$$j_g(z) = \epsilon_d^{-1}\left(\frac{c}{d}\right)\left(\frac{cz+d}{|cz+d|}\right)^{1/2}.$$

This is the usual automorphic factor divided by its modulus, and $\epsilon_d = 1$ or i if $d \equiv 1$ or $3 \pmod 4$. We can forget about ϵ_d if we restrict ourselves to the subgroup $\Gamma_1(4)$, which is of index 2 in $\Gamma_0(4)$. To check that $\Delta_1(4)$ is a discrete subgroup of $\widetilde{\mathrm{SL}}(2,\mathbb{R})$, observe that $\mathrm{e}^{\mathrm{i}\beta_g(z)/2} = j_g(z)$, for $g \in \Gamma_1(4)$, and

$$j_{gh}(z) = \frac{\theta_0(ghz)}{\theta_0(z)} = \frac{\theta_0(ghz)}{\theta_0(hz)}\frac{\theta_0(hz)}{\theta_0(z)} = j_g(hz)\,j_h(z), \tag{23}$$

which is consistent with the multiplication law (9) of $\widetilde{\mathrm{SL}}(2,\mathbb{R})$. We conclude that the function

$$\theta_k(z,\phi) = \theta_k(z)\,\mathrm{e}^{-\mathrm{i}(2k+1)\phi/2} \tag{24}$$

is invariant under $\Delta_1(4)$, that is, it is a function on $\mathcal{M}$.

Put

$$\Theta_f(z,\phi) = y^{1/4} \sum_{n\in\mathbb{Z}} f_\phi\big(ny^{1/2}\big) e\big(n^2 x\big), \tag{25}$$

where

$$f_\phi(t) = \sum_{k=0}^{\infty} \hat{f}(k)\,\mathrm{e}^{-\mathrm{i}(2k+1)\phi/2}\,h_k(t),$$

since then $\Theta_f(z) = \Theta_f(z,0)$. The series $f_\phi(ny^{1/2})$ in (25) converges uniformly in n (z and ϕ are fixed). Therefore, we can exchange the order of summation to obtain

$$\Theta_f(z,\phi) = \sum_{k=0}^{\infty} \hat{f}(k)\theta_k(z,\phi), \tag{26}$$

which is a C^∞ function on $\mathcal{M}$. □

Remarks. $f_\phi(t)$ satisfies the Schrödinger equation for the harmonic oscillator,

$$\left(-\frac{1}{4\pi}\frac{\partial^2}{\partial t^2}+4\pi t^2\right) f_\phi(t) = 2\,\mathrm{i}\,\frac{\partial}{\partial\phi} f_\phi(t), \tag{27}$$

from which we obtain the integral representation

$$f_\phi(t) = \int_{-\infty}^{\infty} G_\phi\big(t,t'\big) f\big(t'\big)\, dt', \tag{28}$$

with the Green function

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

where $\sigma_\phi = 2k+1$ when $k\pi < \phi < (k+1)\pi$.

Our manifold $\mathcal{M}$ has a cyclic automorphism group of order 8, which is generated by the transformation $(z,\phi) \mapsto (-(4z)^{-1}, \phi + \arg z)$ corresponding to

$$\left[\begin{pmatrix} 0 & -1/2 \\ 2 & 0 \end{pmatrix}, \arg\right] \in \widetilde{\mathrm{SL}}(2,\mathbb{R}).$$

To see how Θ_f transforms under this map, let us apply Poisson's summation formula, which yields

$$\Theta_f(z,\phi) = y^{-1/4} \sum_{n\in\mathbb{Z}} \int_{-\infty}^{\infty} e\left(t^2\frac{x}{y} - \frac{nt}{y^{1/2}}\right) f_\phi(t)\, dt. \tag{29}$$

By virtue of the relation

$$G_{\arg z}\left(n\frac{y^{1/2}}{2|z|}, t\right) = \mathrm{e}^{-\mathrm{i}\pi/4} y^{-1/4} \left(\frac{y}{4|z|^2}\right)^{-1/4} e\left(n^2\frac{x}{4|z|^2} + t^2\frac{x}{y} - \frac{nt}{y^{1/2}}\right), \tag{30}$$

one easily finds

$$\Theta_f\big(-(4z)^{-1}, \phi + \arg z\big) = \mathrm{e}^{-\mathrm{i}\pi/4}\, \Theta_f(z,\phi). \tag{31}$$

Notice that repeated application yields $\Theta_f(z,\phi+\pi) = -\mathrm{i}\,\Theta_f(z,\phi)$.

$\mathcal{M}$ has three cusps of codimension 2 at $(0,\phi)$, $(1/2,\phi)$, and (∞,ϕ). We now study the asymptotic behaviour of $\Theta_f(z,\phi)$ in these cusps. To this end, let us choose the following convenient coordinates:

$$\begin{cases} (z_0,\phi_0) = \big(-(4z)^{-1}, \phi + \arg z\big), \\ (z_{1/2},\phi_{1/2}) = \big(-(4z-2)^{-1}, \phi + \arg(z-1/2)\big), \\ (z_\infty,\phi_\infty) = (z,\phi). \end{cases} \tag{32}$$

PROPOSITION 3.2. *Let $f \in \mathcal{S}(\mathbb{R})$. Then*

$$\Theta_f(z,\phi) = \begin{cases} e^{i\pi/4} f_{\phi_0}(0) y_0^{1/4} + O_N\big(y_0^{-N}\big), & y_0 \to \infty, \\ O_N\big(y_{1/2}^{-N}\big), & y_{1/2} \to \infty, \\ f_{\phi_\infty}(0) y_\infty^{1/4} + O_N\big(y_\infty^{-N}\big), & y_\infty \to \infty, \end{cases}$$

for any N.

Proof. Since $f_\phi \in \mathcal{S}(\mathbb{R})$, the asymptotic relation in the cusp at (∞,ϕ) is evident, and the behaviour at $(0,\phi)$ is clear from transformation formula (31). In order to understand the second relation, we observe that

$$\begin{aligned} \Theta_f(z,\phi) &= 2^{1/2}\Theta_f(4z-2,\phi) - \Theta_f(z-1/2,\phi) \\ &= 2^{1/2} e^{i\pi/4}\Theta_f(z_{1/2}/4,\phi_{1/2}) - e^{i\pi/4}\Theta_f(z_{1/2},\phi_{1/2}) \qquad (33) \\ &= O_N\big(y_{1/2}^{-N}\big). \end{aligned}$$

□

Denote by $\mathcal{M}^*$ the compactification of the manifold $\mathcal{M}$, that is,

$$\mathcal{M}^* = \mathcal{M} \cup \bigcup_{j\in\{0,1/2,\infty\}} \{(j,\phi) : \phi \in [0,4\pi)\}.$$

If the cutoff function f satisfies the condition $f_\phi(0) \neq 0$ for all $\phi \in [0,4\pi)$, the map $\Theta_f : \mathcal{M} \to \mathbb{C}$ can be extended to a continuous map $\Theta_f^* : \mathcal{M}^* \to \mathbb{C}\cup\{\infty\}$ by defining $\Theta_f^*|_{\mathcal{M}} = \Theta_f$, $\Theta_f^*(0,\phi) = \infty$, $\Theta_f^*(1/2,\phi) = 0$, and $\Theta_f^*(\infty,\phi) = \infty$.

PROPOSITION 3.3. *Let $f \in \mathcal{S}(\mathbb{R})$, and assume that $f_\phi(0) \neq 0$ for all $\phi \in [0,4\pi)$. Then the map*

$$\Theta_f^* : \mathcal{M}^* \to \mathbb{C}\cup\{\infty\}$$

is onto.

Proof. We think of $\mathbb{C}\cup\{\infty\}$ as the Riemann sphere S^2, where the north pole is identified with ∞ and the south pole with zero. In the first step of this proof, we show that the closed curve

$$\gamma : [0,4\pi] \to S^2 - \{0,\infty\}, \qquad t \mapsto f_t(0),$$

is not null-homotopic on $S^2 - \{0,\infty\}$. To this end, split γ into

$$\gamma_1 : [0,2\pi] \to S^2 - \{0,\infty\}, \qquad t \mapsto f_t(0),$$

connecting $f_0(0) = f(0)$ with $f_{2\pi}(0) = -f(0)$, and

$$\gamma_2 : [0,2\pi] \to S^2 - \{0,\infty\}, \qquad t \mapsto f_{t+2\pi}(0),$$

connecting $f_{2\pi}(0) = -f(0)$ with $f_{4\pi}(0) = f(0)$. Since $\gamma_2 = -\gamma_1$, it is plain to see that γ cannot be contracted to a point without crossing a north or south pole.

Now consider the family of closed curves

$$\tilde{\gamma}_y : [0, 4\pi] \to \mathrm{S}^2 - \{\infty\}, \qquad t \mapsto \Theta_f(1/2 + \mathrm{i}\, y, t), \tag{34}$$

parameterized by y. From Proposition 3.2 we know $\Theta_f(1/2+\mathrm{i}\,y,t) = f_t(0) y^{1/4} + O_N(y_\infty^{-N})$, as $y \to \infty$. So for all $y \geq y_0$, with y_0 large enough, the curve $\tilde{\gamma}_y$ has the same property as γ— it is not null-homotopic on $\mathrm{S}^2 - \{0, \infty\}$.

We now contract the closed curve $\tilde{\gamma}_{y_0}$ on $\mathrm{S}^2 - \{0\}$ to the north pole by considering the family (34), where y runs from y_0 to ∞. This deformation is clearly continuous in y since Θ_f^* is. Notice in particular that $\Theta_f^*(1/2+\mathrm{i}\,y,t) \neq 0$ for $y \geq y_0$, and hence $\tilde{\gamma}_y$ does not pass the south pole.

Alternatively, we can as well contract $\tilde{\gamma}_{y_0}$ continuously on $\mathrm{S}^2 - \{\infty\}$ to the south pole, by letting y run from y_0 to zero. Here we have $\Theta_f^*(1/2 + \mathrm{i}\,y, t) \neq \infty$ for $y \in [0, y_0]$, and hence $\tilde{\gamma}_y$ does not pass the north pole.

Therefore, given any point $P \in \mathrm{S}^2$, the closed curve $\tilde{\gamma}_y$ must pass P on its way either to the north or to the south pole, so

$$\left\{\Theta_f^*(1/2+\mathrm{i}\,y,t) : (y,t) \in [0,\infty] \times [0,4\pi]\right\} \cong \mathrm{S}^2. \tag{35}$$

This proves the map Θ_f^* is onto. □

Remark. A sufficient condition for the cutoff function f to guarantee $f_\phi(0) \neq 0$ for all ϕ is, for example, that $f(t)$ is real valued and monotonically increasing for $t < 0$ and is monotonically decreasing for $t > 0$. (Here and in the following, it is of course always assumed that f is not the trivial zero-function.) This can be seen as follows. Recall from (28) that $f_\phi(0) \neq 0$ is equivalent to $f(0) \neq 0$ and

$$\int_{-\infty}^{\infty} e\big(t^2 w\big) f(t)\, dt \neq 0$$

for all $w \in \mathbb{R}$. If $f(t)$ is assumed to be real valued and monotonically increasing (resp., decreasing) for $t < 0$ (resp., $t > 0$), then the first condition is clearly satisfied, because f has to vanish at $\pm\infty$ and thus must be positive. The monotone increase (resp., decrease) implies also that

$$\int_{-\infty}^{\infty} \cos\big(2\pi t^2 w\big) f(t)\, dt > 0,$$

which in turn gives the second condition.

We now provide some information on Eisenstein series of the theta group Γ_θ, which we use to prove the upcoming proposition.

The theta group Γ_θ is generated by the transformations $z \mapsto z+1$ and $z \mapsto -1/4z$. A fundamental region is given by

$$\mathscr{F}_{\Gamma_\theta} = \{z \in \mathfrak{H} : x \in (-1/2, 1/2], |z| > 1/2\}, \tag{36}$$

so the two cusps of $\Gamma_\theta\backslash\mathfrak{H}$ are identified with the points $1/2$ and ∞. The Eisenstein series associated with the ith cusp is defined as

$$E_i(z,s)=\sum_{\gamma\in\Gamma_i\backslash\Gamma_\theta}\left[\operatorname{Im}(\gamma_i\gamma z)\right]^s,\quad \operatorname{Re}s>1, \tag{37}$$

where Γ_i is the stabilizer of the ith cusp (e.g., $\Gamma_\infty=\{\gamma : z\mapsto z+n, n\in\mathbb{Z}\}$). γ_i is the transformation mapping the ith cusp onto the standard cusp at ∞ with unit width. We put $z_i=\gamma_i z$ (compare with definition (32)). Fourier expansion in x_j yields

$$E_i(z,s) = \delta_{ij}y_j^s+\varphi_{ij}(s)y_j^{1-s}+2\sum_{m=1}^{\infty}\varphi_{ij}^{(m)}(s)y_j^{1/2}K_{s-1/2}(2\pi m y_j)\cos(2\pi m x_j), \tag{38}$$

where $K_\nu(z)$ is the K-Bessel function. The functions φ_{ij} and $\varphi_{ij}^{(m)}$ can be expressed in terms of number-theoretic functions, in particular,

$$\varphi_{ij}(s)=\frac{\xi(2s-1)}{\xi(2s)}\eta_{ij}(s) \tag{39}$$

with $\xi(s)=\pi^{-s/2}\Gamma(s/2)\zeta(s)$, where $\zeta(s)$ is the Riemann zeta function, and

$$\big(\eta_{ij}(s)\big)=\frac{1}{2^{2s}-1}\begin{pmatrix}1 & 2^s-2^{1-s}\\ 2^s-2^{1-s} & 1\end{pmatrix}.$$

See Hejhal [5, Chapter 11, Section 3].

The expansion (38) can be used to find a meromorphic continuation of $E_i(z,s)$ to the whole complex plane. Let us also note that

$$E_i(z,s)=\delta_{ij}y_j^s+\varphi_{ij}(s)y_j^{1-s}+O\big(\mathrm{e}^{-2\pi y_j}\big),\quad y_j\to\infty, \tag{40}$$

which holds uniformly on compacta in the half-plane $\operatorname{Re}s>1/2$, since in this domain the only singularity of $E_i(z,s)$ is the simple pole of $\varphi_{ij}(s)$ at $s=1$.

PROPOSITION 3.4. *Let $f\in\mathrm{L}^2(\mathbb{R})$. Then there exists a function $\Theta_f\in\mathrm{L}^2(\mathcal{M})$ such that*

$$\left\|\Theta_f-\sum_{k=0}^{N}\hat{f}(k)\theta_k\right\|_{\mathrm{L}^2(\mathcal{M})}\to 0$$

as $N\to\infty$, where $\hat{f}(k)$ are the Hermite coefficients of f.

Proof. It is sufficient to show that the functions θ_{2k} are orthonormal. In fact, we shall see that

$$(4\pi)^{-2}\int_{\mathcal{M}}\theta_{2j}(z,\phi)\overline{\theta_{2k}(z,\phi)}\,\frac{dx\,dy\,d\phi}{y^2}=\delta_{jk}. \tag{41}$$

Proposition 3.2 guarantees the convergence of the integral. The ϕ-integration can be carried out easily, and we are left with

$$(4\pi)^{-1}\delta_{jk}\int_{\Gamma_1(4)\backslash\mathfrak{H}}|\theta_{2k}(z)|^2\frac{dx\,dy}{y^2}=(2\pi)^{-1}\delta_{jk}\int_{\Gamma_\theta\backslash\mathfrak{H}}|\theta_{2k}(z)|^2\frac{dx\,dy}{y^2}, \tag{42}$$

since by formula (31) the modulus of $\theta_{2k}(z)$ is invariant under the theta group Γ_θ, which is of index 2 over $\Gamma_1(4)/\{\pm1\}$. Define

$$I_k(s)=\int_{\Gamma_\theta\backslash\mathfrak{H}}\Big[|\theta_{2k}(z)|^2E_\infty(z,s)-h_{2k}(0)^2E_\infty(z,s+1/2)\Big]\frac{dx\,dy}{y^2}, \tag{43}$$

which is convergent due to (40). The Eisenstein series has a simple pole at $s=1$ with residue $\operatorname{Res}E_\infty(z,s=1)=\pi^{-1}$ and is regular for $\operatorname{Re}s>1$, thus

$$\operatorname{Res}I_k(s=1)=\pi^{-1}\int_{\Gamma_\theta\backslash\mathfrak{H}}|\theta_{2k}(z)|^2\frac{dx\,dy}{y^2}. \tag{44}$$

On the other hand, we can unfold the integral $I_k(s)$ using the definition (37). This gives, for $\operatorname{Re}s>1$,

$$\begin{aligned}I_k(s)&=\int_0^\infty\int_0^1\Big[|\theta_{2k}(z)|^2y^s-h_{2k}(0)^2y^{s+1/2}\Big]\frac{dx\,dy}{y^2}\\&=4\int_0^\infty\sum_{n=1}^\infty h_{2k}\big(ny^{1/2}\big)^2y^{s-1/2}\frac{dy}{y}\\&=4\zeta(2s-1)\int_{-\infty}^\infty h_{2k}(t)^2|t|^{2s-2}dt.\end{aligned} \tag{45}$$

In the last step we exchanged the order of integration and summation. This is justified, as the integral converges uniformly, in the sense that for every $\epsilon>0$ we find a C such that

$$\left|\int_0^C\sum_{n=1}^N h_{2k}(nt)t^{2s-2}dt\right|<\epsilon \tag{46}$$

uniformly for all $N\in\mathbb{N}$. To see this notice that (46) is dominated by

$$\left|\sum_{n=1}^K n^{1-2s}\int_0^{nC}h_{2k}(t)t^{2s-2}dt\right|+\left|\sum_K^\infty n^{1-2s}\right|\int_0^\infty\left|h_{2k}(t)t^{2s-2}\right|dt,$$

where we choose K such that the second sum is less than $\epsilon/2$, and then we choose C such that the first sum is less than $\epsilon/2$.

Finally, we obtain the desired $\operatorname{Res}I_k(s=1)=2$. □

Remark. It is a direct consequence of (41) that if f is even (i.e., $f(t) = f(-t)$), then

$$(4\pi)^{-2}(\Theta_f, \theta_k)_{\mathrm{L}^2(\mathcal{M})} = \hat{f}(k), \tag{47}$$

$$(4\pi)^{-1}\|\Theta_f\|_{\mathrm{L}^2(\mathcal{M})} = \|f\|_{\mathrm{L}^2(\mathbb{R})}. \tag{48}$$

It is worthwhile to mention that the content of this section may be rephrased in terms of the Shale-Weil representation of $\widetilde{\mathrm{SL}}(2,\mathbb{R})$. A nice introduction to the representation theory of this group is, for instance, contained in [11]. In particular, it is interesting that Θ_f, $f \in \mathcal{S}(\mathbb{R})$, satisfies the eigenvalue equation

$$\mathrm{C}\,\Theta_f = -\frac{3}{16}\Theta_f, \tag{49}$$

where C is the Casimir operator

$$\mathrm{C} = y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) + y\frac{\partial^2}{\partial x\,\partial\phi}$$

of $\widetilde{\mathrm{SL}}(2,\mathbb{R})$. Equation (49) may be verified by a straightforward calculation using (27); compare also with Lemma 1.5 of Shintani [18].

4. Geodesic flows and horocycle flows. The unit tangent bundle $\mathrm{T}_1 M$ of a hyperbolic manifold $M = \Gamma\backslash\mathfrak{H}$, with Γ a discrete subgroup of $\mathrm{PSL}(2,\mathbb{R})$, is isomorphic to $\Gamma\backslash(\mathrm{T}_1\mathfrak{H})$. $\mathrm{T}_1\mathfrak{H}$ denotes the unit tangent bundle of $\mathfrak{H}$, which is usually identified with $\mathfrak{H}\times(\mathbb{R}/2\pi\mathbb{Z})$, and the action of $\mathrm{PSL}(2,\mathbb{R})$ on $\mathrm{T}_1\mathfrak{H}$ is given by

$$g : (z,\theta) \mapsto \big(gz, \theta - 2\beta_g(z)\big); \tag{50}$$

see [13]. To view this in the framework of $\widetilde{\mathrm{SL}}(2,\mathbb{R})$, which we have developed so far, let us substitute $\theta = -2\phi$, and identify $\mathrm{T}_1\mathfrak{H}$ with $\mathfrak{H}\times(\mathbb{R}/\pi\mathbb{Z})$. Now the action of $\mathrm{PSL}(2,\mathbb{R})$ looks more like (12),

$$g : (z,\phi) \mapsto \big(gz, \phi + \beta_g(z)\big), \tag{51}$$

in particular, if we view $\mathrm{PSL}(2,\mathbb{R})$ as $\widetilde{\mathrm{SL}}(2,\mathbb{R})/Z_1$.

If we take $M = \Gamma_1(4)\backslash\mathfrak{H}$ and denote by Λ the group generated by the elements of $\Delta_1(4)$ and Z_1, we find that

$$\mathrm{T}_1 M \simeq \Lambda\backslash\widetilde{\mathrm{SL}}(2,\mathbb{R}). \tag{52}$$

$\Delta_1(4)$ is normal in Λ, we have the disjunct decomposition

$$\Lambda = \Delta_1(4)\cup\Delta_1(4)[-1,\pi]\cup\Delta_1(4)[1,2\pi]\cup\Delta_1(4)[-1,3\pi], \tag{53}$$

and thus we have

$$\Delta_1(4)\backslash\Lambda \simeq \mathbb{Z}_4.$$

This means that $\mathcal{M}$ is a fourfold cover of $\mathrm{T}_1 M$.

Let $\{\Phi^t\}_{t\in\mathbb{R}}$ be a 1-parameter subgroup in $\widetilde{\mathrm{SL}}(2,\mathbb{R})$, that is, $\Phi^s\Phi^t = \Phi^{s+t}$. Assume $\operatorname{tr}\Phi^t = 2\cosh(t/2)$. (By the trace of an element in $\widetilde{\mathrm{SL}}(2,\mathbb{R})$, we mean the trace of its $\mathrm{SL}(2,\mathbb{R})$ component.) Then Φ^t is conjugate to

$$\Phi_0^t = \left[\begin{pmatrix} \mathrm{e}^{-t/2} & 0 \\ 0 & \mathrm{e}^{t/2} \end{pmatrix}, 0\right];$$

that is, we find a $\tilde{g}$ such that $\Phi^t\tilde{g} = \tilde{g}\Phi_0^t$. We define the *geodesic flow* on $\widetilde{\mathrm{SL}}(2,\mathbb{R})$ as the right translation by Φ_0^t,

$$\Phi^t : \widetilde{\mathrm{SL}}(2,\mathbb{R}) \to \widetilde{\mathrm{SL}}(2,\mathbb{R}), \quad \tilde{g} \mapsto \tilde{g}\Phi_0^t. \tag{54}$$

In the same fashion, we define the *horocycle flow*

$$\Psi^t : \widetilde{\mathrm{SL}}(2,\mathbb{R}) \to \widetilde{\mathrm{SL}}(2,\mathbb{R}), \quad \tilde{g} \mapsto \tilde{g}\Psi_0^t, \tag{55}$$

with

$$\Psi_0^t = \left[\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, 0\right].$$

The associated 1-parameter subgroups $\{\Psi^t\}_{t\in\mathbb{R}}$ are characterized by the condition $\operatorname{tr}\Psi^t = 2$.

We call the set $\{\tilde{g}\Phi_0^t\}_{t\in[0,T]}$ a *geodesic of length* T. We call $\{\tilde{g}\Psi_0^t\}_{t\in[0,T]}$ a *horocycle of length* T.

For a discrete subgroup Δ, we define the geodesic flow on the quotient by

$$\Phi^t : \Delta\backslash\widetilde{\mathrm{SL}}(2,\mathbb{R}) \to \Delta\backslash\widetilde{\mathrm{SL}}(2,\mathbb{R}), \quad \Delta\tilde{g} \mapsto \Delta\tilde{g}\Phi_0^t. \tag{56}$$

A geodesic $\{\Delta\tilde{g}\Phi_0^t\}_{t\in[0,T]}$ is closed if and only if $\Phi^T = \tilde{g}\Phi_0^T\tilde{g}^{-1}$ is in Δ. T is then the *period* of the closed geodesic, and the smallest $t > 0$ for which $\Phi^t \in \Delta$ is its *primitive period*. The same terminology applies to closed horocycles.

The ergodicity of the geodesic flow on the unit tangent bundle of a hyperbolic surface of finite volume, hence for quotients $\Gamma\backslash\mathrm{PSL}(2,\mathbb{R})$, is proved by Hopf [7]. The generalization to quotients $\Delta\backslash\widetilde{\mathrm{SL}}(2,\mathbb{R})$ of finite measure is obvious. We state the result for our special case $\mathcal{M} = \Delta_1(4)\backslash\widetilde{\mathrm{SL}}(2,\mathbb{R})$.

Proposition 4.1. *Let $F \in \mathrm{L}^1(\mathcal{M})$. Then for almost all $\tilde{g} \in \mathcal{M}$,*

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T F\big(\tilde{g}\Phi_0^t\big)dt = \frac{1}{\mu(\mathcal{M})}\int_{\mathcal{M}} F(\tilde{g})d\mu(\tilde{g})$$

with respect to Haar measure $d\mu(\tilde{g}) = dx\,dy\,d\phi/y^2$.

Remarks. In the case when F is uniformly continuous, we note that if equidistribution holds for some $\tilde{g}$, it holds as well for all other points on the stable manifold of $\tilde{g}$, that is, for all points $\tilde{h}$ of the form

$$\tilde{h} = \tilde{g}\Phi_0^v\left[\begin{pmatrix}1 & 0\\ u & 1\end{pmatrix}, \beta\right]$$

with $\beta(z) = \arg(uz+1)$, for any (fixed) $u, v \in \mathbb{R}$. In particular, the relation

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T F\big(x+\mathrm{i}\,\mathrm{e}^{-t}, 0\big)dt = \frac{1}{\mu(\mathcal{M})}\int_{\mathcal{M}} F(\tilde{g})d\mu(\tilde{g}) \tag{57}$$

holds for almost all x, with respect to Lebesgue measure on $\mathbb{R}$.

This statement remains valid if F is taken to be a characteristic function.

COROLLARY 4.2. *Let $\chi_{\mathcal{D}}$ be the characteristic function of an open subset $\mathcal{D} \subset \mathcal{M}$ with boundary of measure $\mu(\partial\mathcal{D}) = 0$. Then*

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T \chi_{\mathcal{D}}\big(x+\mathrm{i}\,\mathrm{e}^{-t}, 0\big)dt = \frac{\mu(\mathcal{D})}{\mu(\mathcal{M})} \tag{58}$$

holds for almost all x.

Proof. We apply a standard density argument: For any given $\epsilon > 0$, there exist uniformly continuous functions F_1, $F_2 \in \mathrm{L}^1(\mathcal{M})$ such that $F_1 \le \chi_{\mathcal{D}} \le F_2$ and

$$\frac{1}{\mu(\mathcal{M})}\int_{\mathcal{M}}\big[F_2(\tilde{g}) - F_1(\tilde{g})\big]d\mu(\tilde{g}) < \epsilon.$$

Thus for $T \to \infty$, the left-and right-hand sides of

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\int_0^T F_1\big(\tilde{g}\Psi_0^t\big)dt &\le \lim_{T\to\infty}\frac{1}{T}\int_0^T \chi_{\mathcal{D}}\big(\tilde{g}\Psi_0^t\big)dt\\ &\le \lim_{T\to\infty}\frac{1}{T}\int_0^T F_2\big(\tilde{g}\Psi_0^t\big)dt\end{aligned} \tag{59}$$

have limits that differ by less than ϵ. Since ϵ can be arbitrarily small, the corollary follows. □

The ergodicity of the geodesic flow is intimately related to the ergodicity of the horocycle flow. What is more, Sarnak [13] even shows individual equidistribution of long *closed* horocycles on the unit tangent bundle of noncompact hyperbolic surfaces of finite volume (in the compact case no horocycle is closed). His proof still holds in our case if the Eisenstein series of even integral weight are replaced by Eisenstein series of half-integral weight

$$E_\infty(z; s, m) = \sum_{\gamma\in\Gamma_\infty\backslash\Gamma_1(4)}\big[\operatorname{Im}(\gamma z)\big]^s j_\gamma(z)^{-m},$$

with the transformation property

$$E_\infty(gz;s,m) = j_g(z)^m E_\infty(z;s,m)$$

for any $g \in \Gamma_1(4)$. The use of these Eisenstein series would allow us to give an asymptotic expansion in inverse powers of the horocycle length, as in [13], provided the test functions are smooth and of compact support (see [12] for details). In order to permit a wider class of test functions, we restrict ourselves to the leading-order contribution.

PROPOSITION 4.3. *Let F be continuous on $\mathcal{M}$, and assume $F(z,\phi) = O(y_i^\sigma)$ for some $\sigma < 1$, as $y_i \to \infty$ $(i = 0, 1/2, \infty)$. Then for*

$$\tilde{g} = \left[\begin{pmatrix} T^{-1/2} & 0 \\ 0 & T^{1/2} \end{pmatrix}, 0\right],$$

we have

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T F\big(\tilde{g}\Psi_0^t\big)dt = \frac{1}{\mu(\mathcal{M})} \int_{\mathcal{M}} F(\tilde{g})d\mu(\tilde{g}).$$

In coordinates of $\mathfrak{H} \times \mathbb{R}$, the last relation reads

$$\lim_{y\to 0} \int_0^1 F(z,0)dx = \frac{1}{\mu(\mathcal{M})} \int_{\mathcal{M}} F(z,\phi) \frac{dx\,dy\,d\phi}{y^2}. \tag{60}$$

Proof. Let us first have a look at functions that are independent of ϕ and invariant under, say, Γ_θ, that is, functions on $\Gamma_\theta\backslash\mathfrak{H}$. In a recent paper, Hejhal [6] showed that for such functions f satisfying

$$f \in \mathrm{C}^2(\Gamma_\theta\backslash\mathfrak{H}) \cap \mathrm{L}^2(\Gamma_\theta\backslash\mathfrak{H}), \qquad \Delta f \in \mathrm{L}^2(\Gamma_\theta\backslash\mathfrak{H}),$$

one has

$$\int_0^1 f(z)dx = \frac{1}{\pi} \int_{\mathcal{F}_{\Gamma_\theta}} f(z) \frac{dx\,dy}{y^2} + O\big(y^{1/2}\big), \tag{61}$$

for $0 < y \leq 1$. Δ denotes the Laplacian $y^2(\partial_x^2 + \partial_y^2)$ of the upper half-plane $\mathfrak{H}$. Now, by a standard approximation argument (similar to the one given above) we have, for continuous functions $f \in \mathrm{C}(\Gamma_\theta\backslash\mathfrak{H})$ with $f(z) = O(y_i^\sigma)$ $(\sigma < 1/2)$ in the cusps,

$$\lim_{y\to 0} \int_0^1 f(z)dx = \frac{1}{\pi} \int_{\mathcal{F}_{\Gamma_\theta}} f(z) \frac{dx\,dy}{y^2}. \tag{62}$$

Suppose next that f is continuous and can be written in the form

$$f(z) = C y_i^\sigma + O\big(y_i^{1/2-\epsilon}\big) \tag{63}$$

in the neighbourhood of the ith cusp, with constants C, $1/2<\sigma<1$, and $\epsilon>0$. The function

$$g(z)=f(z)-C\sum_{i=1/2,\infty}E_i(z,\sigma) \tag{64}$$

satisfies $g(z)=O(y_i^{1-\sigma})+O(y_i^{1/2-\epsilon})$, so relation (62) is applicable. Together with the fact that the zeroth Fourier coefficient of the Eisenstein series $E_i(z,\sigma)$ vanishes in the limit $y\to 0$ when $1/2<\sigma<1$ (compare with (38)), we obtain

$$\lim_{y\to 0}\int_0^1 f(z)dx=\frac{1}{\pi}\int_{\mathcal{F}_{\Gamma_\theta}} f(z)\frac{dx\,dy}{y^2}-\frac{C}{\pi}\sum_{i=1/2,\infty}\int_{\mathcal{F}_{\Gamma_\theta}} E_i(z,\sigma)\frac{dx\,dy}{y^2}.$$

The integrals over the Eisenstein series vanish, since

$$\int_{\mathcal{F}_A} E_i(z,\sigma)\frac{dx\,dy}{y^2}=\frac{1}{\sigma(\sigma-1)}\int_{\mathcal{F}_A}\Delta E_i(z,\sigma)\frac{dx\,dy}{y^2}, \tag{65}$$

where the integration is performed over the truncated fundamental domain $\mathcal{F}_A=\{z\in\mathcal{F}_{\Gamma_\theta}: y_i<A, i=1/2,\infty\}$. By Green's theorem, (65) equals

$$\sum_{j=1/2,\infty}\left(\frac{A^{\sigma-1}}{\sigma-1}-\frac{\varphi_{ij}(\sigma)A^{-\sigma}}{\sigma}\right)\to 0, \tag{66}$$

which clearly vanishes for $1/2<\sigma<1$ in the limit $A\to\infty$.

Proposition 4.3 is now a consequence of the following observation. Let F satisfy the hypothesis of Proposition 4.3. Then, given any $\epsilon>0$, we find functions $F_1,F_2\in \mathrm{C}^\infty(\mathcal{M})$ with compact support and we find a continuous function $f\geq 0$, independent of ϕ, invariant under Γ_θ, satisfying (63), such that $F_1-f\leq F\leq F_2+f$ on all of $\mathcal{M}$, and such that

$$\frac{1}{\mu(\mathcal{M})}\int_{\mathcal{M}}\big[F_2(\tilde g)-F_1(\tilde g)+2f(\tilde g)\big]d\mu(\tilde g)<\epsilon. \tag{67}$$

Since we know that the statement of Proposition 4.3 holds for the three functions F_1, F_2, f, it must as well hold for F. This proves the proposition. □

COROLLARY 4.4. *Let $\chi_{\mathcal{D}}$ be the characteristic function of an open subset $\mathcal{D}\subset\mathcal{M}$ with boundary of measure $\mu(\partial\mathcal{D})=0$. Then*

$$\lim_{y\to 0}\int_0^1\chi_{\mathcal{D}}(z,0)dx=\frac{\mu(\mathcal{D})}{\mu(\mathcal{M})}. \tag{68}$$

Proof. The proof is the same as for Corollary 4.2. □

The limit theorems stated in the following sections are now a simple consequence of the ergodic properties discussed above.

5. The distribution of values in the complex plane. In order to study the distribution of values of $\Theta_f(z,\phi)$ in the complex plane, let us take an open set $\mathcal{B} \subset \mathbb{C}$ and define the distribution function

$$D_{f,\mathcal{B}}(z,\phi) = \begin{cases} 1, & \text{if } \Theta_f(z,\phi) \in \mathcal{B}, \\ 0, & \text{if } \Theta_f(z,\phi) \notin \mathcal{B}. \end{cases} \tag{69}$$

$D_{f,\mathcal{B}}(z,\phi)$ is the characteristic function of the preimage $\Theta_f^{-1}(\mathcal{B})$ of $\mathcal{B}$ in $\mathcal{M}$. We say $\mathcal{B}$ has a "nice" boundary $\partial\mathcal{B}$ if the boundary of $\Theta_f^{-1}(\mathcal{B})$ has measure zero in $\mathcal{M}$. A sufficient condition for $\partial\mathcal{B}$ to be nice is that $\partial\mathcal{B}$ is of Lebesgue measure zero and contains no critical values of Θ_f. By Sard's theorem, the set of critical values of a smooth map $\mathcal{M} \to \mathbb{C}$ is itself of measure zero in $\mathbb{C}$.

Although the definition of "nice" depends on Θ_f, we indicate by the following example that most sets indeed have a nice boundary.

LEMMA 5.1. *Suppose $f \in \mathcal{S}(\mathbb{R})$. Let $\mathcal{B} \subset \mathbb{C}$ be an open convex set containing zero and with smooth boundary, and let furthermore*

$$\mathcal{B}(w,R) = \{Rz + w : z \in \mathcal{B}\}$$

be its magnified (by $R > 0$) and translated (by $w \in \mathbb{C}$) copy. Fix w. Then, except for countably many R, the boundary of $\mathcal{B}(w,R)$ is nice.

Proof. The measure of the set

$$\mathcal{X}(R) = \big\{(z,\phi) \in \mathcal{M} : \Theta_f(z,\phi) \in \mathcal{B}(w,R)\big\}$$

tends to $\mu(\mathcal{M}) = 8\pi^2$, when $R \to \infty$, since Θ_f is measurable. The sets

$$\mathcal{I}(R) = \big\{(z,\phi) \in \mathcal{M} : \Theta_f(z,\phi) \in \partial\mathcal{B}(w,R)\big\}$$

are disjoint for different values of R (this is due to the convexity of $\mathcal{B}(w,R)$). For these two reasons, there can only be countably many R for which $\mathcal{I}(R)$ has positive measure. Since Θ_f is continuous, the boundary of $\mathcal{X}(R)$ is contained in the set $\mathcal{I}(R)$, which proves the lemma. □

THEOREM 5.2. *Let $\mathcal{B} \subset \mathbb{C}$ be open with nice boundary, and let $f \in \mathcal{S}(\mathbb{R})$. Then for almost all $x \in \mathbb{R}$ (with respect to Lebesgue measure),*

$$\lim_{T\to\infty} \frac{1}{T} \Big| \big\{ t \in [0,T] : \Theta_f\big(x + \mathrm{i}\,\mathrm{e}^{-t}\big) \in \mathcal{B} \big\} \Big| = \frac{\mu\big(\Theta_f^{-1}(\mathcal{B})\big)}{8\pi^2}.$$

THEOREM 5.3. *Let $\mathcal{B} \subset \mathbb{C}$ be open with nice boundary, and let $f \in \mathcal{S}(\mathbb{R})$. Then*

$$\lim_{y\to 0} \big| \{ x \in [0,1] : \Theta_f(x + \mathrm{i}\,y) \in \mathcal{B} \} \big| = \frac{\mu\big(\Theta_f^{-1}(\mathcal{B})\big)}{8\pi^2}.$$

Proof of Theorems 5.2 and 5.3. First observe that

$$\Big|\big\{t\in[0,T]:\Theta_f\big(x+\mathrm{i}\,\mathrm{e}^{-t}\big)\in\mathcal{B}\big\}\Big| = \int_0^T D_{f,\mathcal{B}}\big(x+\mathrm{i}\,\mathrm{e}^{-t},0\big)\,dt$$

and

$$\big|\{x\in[0,1]:\Theta_f(x+\mathrm{i}\,y)\in\mathcal{B}\}\big| = \int_0^1 D_{f,\mathcal{B}}(x+\mathrm{i}\,y,0)\,dx.$$

Since $\Theta_f\in \mathrm{C}^\infty(\mathcal{M})$ (Proposition 3.1) and the boundary of $\mathcal{B}$ is nice, the preimage of $\mathcal{B}$ is open with boundary of measure zero. The statements now follow from Corollaries 5.2 and, 5.3 respectively. □

To give a little illustration, let us have a look at the distribution of the absolute values of $\Theta_f(z,\phi)$ on the real line; that is, we choose $\mathcal{B}=\mathcal{B}_R := \{z : |z| > R\}$. In this case it is easy to obtain an asymptotic expression for $\mu(\Theta_f^{-1}(\mathcal{B}_R))$, R large. Notice that, due to Proposition 3.2, the measure of the set

$$\mathcal{D}_R = \big\{(z,\phi)\in\mathcal{F}_{\Delta_1(4)} : y_0 > |f_{\phi_0}(0)|^{-4}R^4\big\} \\ \cup\big\{(z,\phi)\in\mathcal{F}_{\Delta_1(4)} : y_\infty > |f_{\phi_\infty}(0)|^{-4}R^4\big\} \quad (70)$$

($\mathcal{F}_{\Delta_1(4)}$ is the fundamental region defined in Section 3) converges rapidly, for $R\to\infty$, to what we seek:

$$\mu\big(\Theta_f^{-1}(\mathcal{B}_R)\big) = \mu(\mathcal{D}_R)+O_M\big(R^{-M}\big). \quad (71)$$

Elementary integration yields (provided R is so large that the sets in (70) are disjoint)

$$\mu(\mathcal{D}_R) = 2R^{-4}\int_0^{4\pi}|f_\phi(0)|^4 d\phi, \quad (72)$$

where

$$|f_\phi(0)|^4 = 4(\sin\phi)^{-2}\left|\int_{-\infty}^{\infty} e\big(t^2\cot\phi\big)f(t)dt\right|^4;$$

compare to equation (28). After some change of variables, we eventually obtain

$$\mu(\mathcal{D}_R) = 32R^{-4}\int_{-\infty}^{\infty}\left|\int_{-\infty}^{\infty} e\big(t^2u\big)f(t)dt\right|^4 du. \quad (73)$$

It is worthwhile mentioning that the statement in Theorem 5.2 does not hold when x is rational or real quadratic. First notice that every $x\in\mathbb{Q}$ is $\Gamma_1(4)$-equivalent to one of the cusps 0, 1/2, or ∞ of the hyperbolic surface $\Gamma_1(4)\backslash\mathfrak{H}$ (see, e.g., [15, p. 14]). The asymptotic behaviour of $\Theta_f(z,0)$ for $x\in\mathbb{Q}$, $y\to 0$, is therefore dictated by Proposition 3.2. Suppose next that x is a root of the equation

$$ax^2+bx+c=0 \quad (74)$$

with integral coefficients satisfying $(a,b,c)=1$ and $b^2-4ac>0$. There is a well-known one-to-one correspondence between primitive quadratic forms $Q(x,y)=ax^2+bxy+cy^2$ and primitive hyperbolic elements of $\mathrm{PSL}(2,\mathbb{Z})$ (see, e.g., [14]). Hence every solution of a quadratic equation (74) is associated with a closed geodesic on the modular surface $\mathrm{PSL}(2,\mathbb{Z})\backslash\mathfrak{H}$, which lifts to a closed geodesic (as defined in Section 4) on $\mathcal{M}=\Delta_1(4)\backslash\widetilde{\mathrm{SL}}(2,\mathbb{R})$. Since this closed geodesic is approached exponentially quickly by the geodesic $\{(x+\mathrm{i}\,\mathrm{e}^{-t},0):t\in[0,T]\}$, as $T\to\infty$, the theta function $\Theta_f(x+\mathrm{i}\,\mathrm{e}^{-T})$ is asymptotically periodic in T. Hence a limit distribution exists, but, of course, it depends on the corresponding closed geodesic and thus on x.

Proposition 5.4. *Let $f\in\mathcal{S}(\mathbb{R})$, and assume that $f_\phi(0)\neq 0$ for all $\phi\in[0,4\pi)$. Then for almost all $x\in\mathbb{R}$ (with respect to Lebesgue measure) the set*

$$\left\{\Theta_f\left(x+\mathrm{i}\,\mathrm{e}^{-t},0\right):t\in[0,T]\right\}$$

becomes densely distributed in $\mathbb{C}$ as T tends to infinity.

Proposition 5.5. *Let $f\in\mathcal{S}(\mathbb{R})$, and assume that $f_\phi(0)\neq 0$ for all $\phi\in[0,4\pi)$. Then the set*

$$\{\Theta_f(x+\mathrm{i}\,y,0):x\in[0,1]\}$$

becomes densely distributed in $\mathbb{C}$ as y tends to zero.

Proof of Propositions 5.4 and 5.5. Apply Proposition 3.3 and Corollaries 4.2 and 4.4. □

We recall the remark after Proposition 3.3 where $f_\phi(0)\neq 0$ holds for all ϕ, if $f(t)$ is assumed to be real valued and monotonically increasing for $t<0$ and is monotonically decreasing for $t>0$.

6. The asymptotic behaviour of moments. For some π-periodic continuous function $h(\phi)$ and $\mathrm{Re}\,s>1$, let us define the Eisenstein series

$$E_\infty(z,\phi;s)=\sum_{\gamma\in\Gamma_\infty\backslash\Gamma_\theta}h\left(\phi+\beta_\gamma(z)\right)\left[\mathrm{Im}(\gamma z)\right]^s, \tag{75}$$

which we may view as a function on $\Gamma_\theta\backslash\mathrm{PSL}(2,\mathbb{R})$ if we identify $\mathrm{PSL}(2,\mathbb{R})$ with $\mathfrak{H}\times[0,\pi)$. Its behaviour in the cusps at $(1/2,\phi)$ and (∞,ϕ) is (for $\mathrm{Re}\,s>1$)

$$E_\infty(z,\phi;s)=\begin{cases}O\left(y_{1/2}^{1-\mathrm{Re}\,s}\right), & y_{1/2}\to\infty,\\ h(\phi)y^s+O\left(y^{1-\mathrm{Re}\,s}\right), & y\to\infty,\end{cases} \tag{76}$$

since

$$|E_\infty(z,\phi;s)|\leq C\,E_\infty(z,\mathrm{Re}\,s),$$
$$\left|E_\infty(z,\phi;s)-h(\phi)y^s\right|\leq C\left[E_\infty(z,\mathrm{Re}\,s)-y^{\mathrm{Re}\,s}\right],$$

for some constant C. $E_\infty(z,s)$ is the classical Eisenstein series we have already discussed. We could easily improve (76), but we do not really need it at this place. The zeroth Fourier coefficient reads

$$\int_0^1 E_\infty(z,\phi;s)dx = h(\phi)y^s + \frac{1}{2^{2s}-1}\frac{\zeta(2s-1)}{\zeta(2s)}H(\phi;s)y^{1-s}, \tag{77}$$

where

$$H(\phi;s) = \int_0^\pi h(\phi+\beta)(\sin\beta)^{2(s-1)}d\beta.$$

For $h(\phi)=1$, we get back to the classical case, that is,

$$H(\phi;s) = \pi^{1/2}\frac{\Gamma(s-1/2)}{\Gamma(s)}.$$

Furthermore, it is useful to have a function that grows linearly in the cusp at (∞,ϕ). Such a function is, for example, given by

$$G_\infty(z,\phi;\epsilon) = \frac{\mathrm{i}\epsilon}{\pi}\int_{1+\epsilon-\mathrm{i}\infty}^{1+\epsilon+\mathrm{i}\infty}\frac{E_\infty(z,\phi;s)}{(s-1)(s-1-2\epsilon)}ds, \tag{78}$$

for some $\epsilon>0$, since

$$\frac{\mathrm{i}\epsilon}{\pi}\int_{1+\epsilon-\mathrm{i}\infty}^{1+\epsilon+\mathrm{i}\infty}\frac{y^s}{(s-1)(s-1-2\epsilon)}ds = \begin{cases} y^{1+2\epsilon}, & \text{if } 0<y<1,\\ y, & \text{if } y\geq 1,\end{cases} \tag{79}$$

and thus, as a consequence of (76),

$$G_\infty(z,\phi;\epsilon) = \begin{cases} O\big(y_{1/2}^{-\epsilon}\big), & y_{1/2}\to\infty,\\ h(\phi)y+O\big(y^{-\epsilon}\big), & y\to\infty.\end{cases} \tag{80}$$

The zeroth Fourier coefficient can be calculated from (77). In particular, for $0<y<1$ and $0<\epsilon<1/2$, one has

$$\begin{aligned}\int_0^1 G_\infty(z,\phi;\epsilon)dx &= h(\phi)y^{1+\epsilon} + Q_\phi(1)\log y^{-1} + \frac{1}{2\epsilon}Q_\phi(1) + Q'_\phi(1)\\ &\quad + \frac{\mathrm{i}\epsilon}{\pi}\int_{1-\epsilon-\mathrm{i}\infty}^{1-\epsilon+\mathrm{i}\infty} Q_\phi(s)\frac{y^{1-s}}{(s-1)^2(s-1-2\epsilon)}ds, \end{aligned}\tag{81}$$

with the functions

$$Q_\phi(s) = \frac{s-1}{2^{2s}-1}\frac{\zeta(2s-1)}{\zeta(2s)}H(\phi;s), \qquad Q_\phi(1) = \frac{1}{\pi^2}H(0;1) \tag{82}$$

being holomorphic in the half-plane $\operatorname{Re} s > 1/2$. $Q'_\phi(s)$ means derivative with respect to s. Notice that

$$Q'_\phi(1) = 2\left[\gamma - \frac{4\log 2}{3} - \frac{\zeta'}{\zeta}(2)\right] Q_\phi(1) + \frac{2}{\pi^2}\int_0^\pi h(\phi+\beta)\log(\sin\beta)\,d\beta, \qquad (83)$$

where γ denotes Euler's constant.

We are now prepared to prove the following theorem.

THEOREM 6.1. *Let $f \in \mathcal{S}(\mathbb{R})$ and $\alpha \geq 0$ be some real number. Then, for $y \to 0$,*

$$\int_0^1 |\Theta_f(z)|^\alpha dx = \begin{cases} b_\alpha + o(1), & \text{if } \alpha < 4, \\ b_\alpha \log y^{-1} + c_\alpha + o(1), & \text{if } \alpha = 4, \\ b_\alpha y^{1-\alpha/4} + c_\alpha + o(1), & \text{if } \alpha > 4, \end{cases}$$

with the constants

$$b_\alpha = \begin{cases} \dfrac{1}{\pi^2}\displaystyle\int_{\Gamma_\theta\backslash \mathrm{PSL}(2,\mathbb{R})} |\Theta_f(z,\phi)|^\alpha \frac{dx\,dy\,d\phi}{y^2}, & \text{if } \alpha < 4, \\ \dfrac{4}{\pi^2}\displaystyle\int_{-\infty}^{\infty}\left|\int_{-\infty}^{\infty} e\big(t^2u\big) f(t)\,dt\right|^4 du, & \text{if } \alpha = 4, \\ \dfrac{1}{1-2^{-\alpha/2}}\dfrac{\zeta(\alpha/2-1)}{\zeta(\alpha/2)}\displaystyle\int_{-\infty}^{\infty}\left|\int_{-\infty}^{\infty} e(t^2u) f(t)\,dt\right|^\alpha du, & \text{if } \alpha > 4, \end{cases}$$

and

$$c_\alpha = \begin{cases} \dfrac{1}{\pi^2}\displaystyle\int_{\Gamma_\theta\backslash \mathrm{PSL}(2,\mathbb{R})} \left[|\Theta_f(z,\phi)|^4 - R_f^4(z,\phi;\epsilon)\right]\frac{dx\,dy\,d\phi}{y^2} + \frac{1}{2\epsilon}b_4 \\ \qquad + 2\left[\gamma - \dfrac{4\log 2}{3} - \dfrac{\zeta'}{\zeta}(2)\right] b_4 \\ \qquad - \dfrac{4}{\pi^2}\displaystyle\int_{-\infty}^{\infty}\left|\int_{-\infty}^{\infty} e\big(t^2u\big) f(t)\,dt\right|^4 \log\big(1+u^2\big)\,du, & \text{if } \alpha = 4, \\ \dfrac{1}{\pi^2}\displaystyle\int_{\Gamma_\theta\backslash \mathrm{PSL}(2,\mathbb{R})} \left[|\Theta_f(z,\phi)|^\alpha - R_f^\alpha(z,\phi)\right]\frac{dx\,dy\,d\phi}{y^2}, & \text{if } \alpha > 4. \end{cases}$$

Here, for $\alpha > 4$,

$$R_f^\alpha(z,\phi) = \sum_{\gamma\in\Gamma_\infty\backslash\Gamma_\theta} \left| f_{\phi+\beta_\gamma(z)}(0)\big[\operatorname{Im}(\gamma z)\big]^{1/4}\right|^\alpha,$$

while, for $\alpha = 4$,

$$R_f^4(z,\phi;\epsilon) = \frac{\mathrm{i}\epsilon}{\pi}\int_{1+\epsilon-\mathrm{i}\infty}^{1+\epsilon+\mathrm{i}\infty} \frac{R_f^{4s}(z,\phi)}{(s-1)(s-1-2\epsilon)}\,ds,$$

for any $0 < \epsilon < 1/2$, where the constant c_α is independent of a particular choice of ϵ.

Proof. In the case $\alpha < 4$, we can directly apply Proposition 4.3. When $\alpha > 4$, we apply Proposition 4.3 to the function $|\Theta_f(z,\phi)|^\alpha - R_f^\alpha(z,\phi)$, which is bounded (so that the hypothesis of Proposition 4.3 is satisfied), since the divergent parts of both terms cancel. To see this, first notice that $R_f^\alpha(z,\phi)$ is just the Eisenstein series $E_\infty(z,\phi;\alpha/4)$ defined in (75), with $h(\phi) = |f_\phi(0)|^\alpha$, and then use the asymptotic relation (76). The leading-order term $b_\alpha y^{1-\alpha/4}$ is essentially the mean value of $R_f^\alpha(z,0)$ with respect to x,

$$\int_0^1 R_f^\alpha(z,0)dx = |f(0)|^\alpha y^{\alpha/4} + \frac{y^{1-\alpha/4}}{2^{\alpha/2}-1}\frac{\zeta(\alpha/2-1)}{\zeta(\alpha/2)}\int_0^\pi |f_\phi(0)|^\alpha (\sin\phi)^{\alpha/2-2}d\phi, \tag{84}$$

where the integral can be further simplified using (28). The case $\alpha = 4$ can be handled in a similar way, with obvious modifications. □

7. The classical theta sum. In this section we show how Theorem 5.3 can be extended to theta sums $\Theta_f(z)$ such as the classical theta sum $S_N(x)$. In fact, it is sufficient to assume that the cutoff function f, which determines Θ_f, is Riemann-integrable. The main task is to relate the sum $\Theta_f(z)$ to the theta function $\Theta_f(z,\phi)$, now defined only as an $L^2(\mathcal{M})$ function in the sense of Proposition 3.4. In particular, the relation $\Theta_f(z) = \Theta_f(z,0)$, derived for Schwartz functions f, is not valid anymore.

Let $\mathcal{B} \subset \mathbb{C}$ be an open convex set containing zero and with smooth boundary. Furthermore, let

$$\mathcal{B}(w,R) = \{Rz + w : z \in \mathcal{B}\}$$

be its magnified (by $R > 0$) and translated (by $w \in \mathbb{C}$) copy. In the following, keep w fixed.

THEOREM 7.1. *Let f be Riemann-integrable. Then, except for countably many R, we have*

$$\lim_{y\to 0}\big|\{x \in [0,1] : \Theta_f(x+\mathrm{i}\,y) \in \mathcal{B}(w,R)\}\big| = \Psi_{\mathcal{B}}(w,R),$$

where

$$\Psi_{\mathcal{B}}(w,R) = \frac{1}{8\pi^2}\mu\{(z,\phi) \in \mathcal{M} : \Theta_f(z,\phi) \in \mathcal{B}(w,R)\}.$$

Proof. We know that the theorem holds when $f \in \mathcal{S}(\mathbb{R})$; compare to Theorem 5.3 and Lemma 5.1. In the following, assume f is Riemann-integrable and $f_\epsilon \in \mathcal{S}(\mathbb{R})$, such that

$$\int_{-\infty}^{\infty} |f(t) - f_\epsilon(t)|^2 dt < \epsilon. \tag{85}$$

Without loss of generality, both f and f_ϵ are taken to be even. Because f is Riemann-integrable, we have

$$|\Theta_f(z)| \le y^{1/4} \sum_{n\in\mathbb{Z}} \left| f\left(ny^{1/2}\right)\right| < \infty$$

for every *finite* y. For y small enough, we have the relation

$$\int_0^1 \left|\Theta_f(z) - \Theta_{f_\epsilon}(z)\right|^2 dx = y^{1/2}\left|f(0) - f_\epsilon(0)\right|^2 + 2y^{1/2} \sum_{n\in\mathbb{Z}^\times} \left| f\left(ny^{1/2}\right) - f_\epsilon\left(ny^{1/2}\right)\right|^2 < 3\epsilon, \tag{86}$$

since the sum converges to twice the Riemann integral of (85). Consider the sets

$$A_y^\epsilon = \left\{x \in [0,1] : \left|\Theta_f(z) - \Theta_{f_\epsilon}(z)\right| < \epsilon^{1/4}\right\},$$

$$X_y(R) = \left\{x \in [0,1] : \Theta_f(x + \mathrm{i}\, y) \in \mathcal{B}(w,R)\right\},$$

and

$$X_y^\epsilon(R) = \left\{x \in [0,1] : \Theta_{f_\epsilon}(x + \mathrm{i}\, y) \in \mathcal{B}(w,R)\right\}.$$

The integral over the complement of A_y^ϵ has to satisfy

$$3\epsilon > \int_{[0,1]-A_y^\epsilon} \left|\Theta_f(z) - \Theta_{f_\epsilon}(z)\right|^2 dx \ge \int_{[0,1]-A_y^\epsilon} \epsilon^{1/2}\, dx,$$

and hence we have

$$\left|A_y^\epsilon\right| > 1 - 3\epsilon^{1/2}. \tag{87}$$

Consequently,

$$\left|X_y(R)\right| < \left|X_y(R) \cap A_y^\epsilon\right| + 3\epsilon^{1/2} \tag{88}$$

and

$$\left|X_y^\epsilon(R)\right| < \left|X_y^\epsilon(R) \cap A_y^\epsilon\right| + 3\epsilon^{1/2}. \tag{89}$$

We have the inclusions

$$\left(X_y^\epsilon\left(R - C\epsilon^{1/4}\right) \cap A_y^\epsilon\right) \subset X_y(R) \tag{90}$$

and

$$\left(X_y(R) \cap A_y^\epsilon\right) \subset X_y^\epsilon\left(R + C\epsilon^{1/4}\right) \tag{91}$$

if the constant C is chosen large enough (C depends solely on the set $\mathcal{B}(w,R)$). So,

$$\left|X_y^\epsilon\bigl(R-C\epsilon^{1/4}\bigr)\right|-3\epsilon^{1/2}\le\left|X_y(R)\right|\le\left|X_y^\epsilon\bigl(R+C\epsilon^{1/4}\bigr)\right|+3\epsilon^{1/2}. \tag{92}$$

Let us now consider the sets

$$\mathcal{A}^\epsilon=\left\{(z,\phi)\in\mathcal{M}:\left|\Theta_f(z)-\Theta_{f_\epsilon}(z)\right|<\epsilon^{1/4}\right\},$$

$$\mathcal{X}(R)=\left\{(z,\phi)\in\mathcal{M}:\Theta_f(z,\phi)\in\mathcal{B}(w,R)\right\},$$

and

$$\mathcal{X}^\epsilon(R)=\left\{(z,\phi)\in\mathcal{M}:\Theta_{f_\epsilon}(z,\phi)\in\mathcal{B}(w,R)\right\}.$$

By virtue of Proposition 3.4, we see that

$$(4\pi)^{-2}\int_{\mathcal{M}}\left|\Theta_f(z,\phi)-\Theta_{f_\epsilon}(z,\phi)\right|^2 d\mu(z,\phi)=\int_{-\infty}^{\infty}\left|f(t)-f_\epsilon(t)\right|^2 dt<\epsilon. \tag{93}$$

Thus

$$\epsilon>(4\pi)^{-2}\int_{\mathcal{M}-\mathcal{A}^\epsilon}\left|\Theta_f(z,\phi)-\Theta_{f_\epsilon}(z,\phi)\right|^2 d\mu(z,\phi)\ge(4\pi)^{-2}\int_{\mathcal{M}-\mathcal{A}^\epsilon}\epsilon^{1/2}d\mu(z,\phi)$$

implies that

$$\mu\bigl(\mathcal{A}^\epsilon\bigr)>8\pi^2\bigl(1-2\epsilon^{1/2}\bigr), \tag{94}$$

and hence

$$\mu\bigl(\mathcal{X}(R)\bigr)<\mu\bigl(\mathcal{X}(R)\cap\mathcal{A}^\epsilon\bigr)+16\pi^2\epsilon^{1/2} \tag{95}$$

and

$$\mu\bigl(\mathcal{X}^\epsilon(R)\bigr)<\mu\bigl(\mathcal{X}^\epsilon(R)\cap\mathcal{A}^\epsilon\bigr)+16\pi^2\epsilon^{1/2}. \tag{96}$$

Corresponding to (90) and (91), one has

$$\bigl(\mathcal{X}^\epsilon\bigl(R-C\epsilon^{1/4}\bigr)\cap\mathcal{A}^\epsilon\bigr)\subset\mathcal{X}(R) \tag{97}$$

and

$$\bigl(\mathcal{X}(R)\cap\mathcal{A}^\epsilon\bigr)\subset\mathcal{X}^\epsilon\bigl(R+C\epsilon^{1/4}\bigr), \tag{98}$$

with the same constant C. So,

$$\mu\bigl(\mathcal{X}^\epsilon\bigl(R-C\epsilon^{1/4}\bigr)\bigr)-16\pi^2\epsilon^{1/2}\le\mu\bigl(\mathcal{X}(R)\bigr)\le\mu\bigl(\mathcal{X}^\epsilon\bigl(R+C\epsilon^{1/4}\bigr)\bigr)+16\pi^2\epsilon^{1/2}. \tag{99}$$

By Theorem 5.3, for all but countably many R, we have

$$\lim_{y\to0}\left|X_y^\epsilon\bigl(R\pm C\epsilon^{1/4}\bigr)\right|=\frac{1}{8\pi^2}\mu\bigl(\mathcal{X}^\epsilon\bigl(R\pm C\epsilon^{1/4}\bigr)\bigr).$$

Therefore, if the difference between the measures $\mu(\mathscr{X}^\epsilon(R-C\epsilon^{1/4}))$ and $\mu(\mathscr{X}^\epsilon(R+C\epsilon^{1/4}))$ vanishes along some sequence of $\epsilon \to 0$, because of (92) and (99) the sequence $|X_y(R)|$ $(y \to 0)$ then must converge to the desired $\mu(\mathscr{X})$.

What thus remains to be shown is that, given any $\delta > 0$, there is an $\epsilon > 0$ such that

$$\mu\big(\mathscr{X}^\epsilon\big(R+C\epsilon^{1/4}\big)\big)-\mu\big(\mathscr{X}^\epsilon\big(R-C\epsilon^{1/4}\big)\big) < \delta. \tag{100}$$

To this end, observe that by virtue of (99)

$$\begin{aligned}\mu\big(\mathscr{X}^\epsilon\big(R+C\epsilon^{1/4}\big)\big)&-\mu\big(\mathscr{X}^\epsilon\big(R-C\epsilon^{1/4}\big)\big)\\ &< \mu\big(\mathscr{X}\big(R+2C\epsilon^{1/4}\big)\big)-\mu\big(\mathscr{X}\big(R-2C\epsilon^{1/4}\big)\big)+32\pi^2\epsilon^{1/2}.\end{aligned} \tag{101}$$

Now suppose that R is such that

$$\mu\big(\mathscr{X}\big(R+2C\epsilon^{1/4}\big)\big)-\mu\big(\mathscr{X}\big(R-2C\epsilon^{1/4}\big)\big) \geq c(R)$$

for arbitrarily small ϵ, where $c(R)$ is some positive constant. This means, however, that

$$\begin{aligned}\mu\Big(\Big\{(z,\phi)\in\mathscr{M} : \Theta_f(z,\phi)\in\mathscr{B}\big(w, R+2C\epsilon^{1/4}\big),&\\ \Theta_f(z,\phi)\notin\mathscr{B}\big(w, R-2C\epsilon^{1/4}\big)\Big\}\Big) &\geq c(R),\end{aligned} \tag{102}$$

for arbitrarily small ϵ, which in turn implies

$$\mu\Big(\Big\{(z,\phi)\in\mathscr{M} : \Theta_f(z,\phi)\in\partial\mathscr{B}(w,R)\Big\}\Big) \geq c(R).$$

This can only happen for countably many R, since Θ_f is measurable (compare with Lemma 5.1). □

We have the following corollaries.

COROLLARY 7.2. *There exists a function* $\Phi(a,b)$ *such that for all (except for countably many)* $a, b \in \mathbb{R}$*,*

$$\lim_{N\to\infty}\left|\left\{x\in[0,1] : a < N^{-1/2}\sum_{n=1}^{N}\cos\big(2\pi n^2 x\big) < b\right\}\right| = \Phi(a,b)$$

and

$$\lim_{N\to\infty}\left|\left\{x\in[0,1] : a < N^{-1/2}\sum_{n=1}^{N}\sin\big(2\pi n^2 x\big) < b\right\}\right| = \Phi(a,b).$$

The reason why both limits are equal to $\Phi(a,b)$ is that, due to equation (31), which holds for smooth functions, we have $\Theta_f(z,\phi+\pi) \sim -\mathrm{i}\,\Theta_f(z,\phi)$ in the L^2 sense. Therefore, the limit distribution in the complex plane should be invariant a.e. under rotations by $-\mathrm{i}$.

COROLLARY 7.3. *There exists a function* $\Psi(a,b)$ *such that for all (except for countably many)* $a, b \in \mathbb{R}$,

$$\lim_{N\to\infty} \left| \left\{ x \in [0,1] : a < N^{-1/2} \left| \sum_{n=1}^{N} e\big(n^2 x\big) \right| < b \right\} \right| = \Psi(a,b).$$

Corollary 7.3 was first proved by Jurkat and van Horne [8], as mentioned before. It follows from the algebraic decay of the limit distribution, which was discussed in Section 5 (see also [8]), that it is not a normal distribution; in particular,

$$\Phi(a,b) \neq \frac{1}{\sqrt{2\pi}} \int_a^b \mathrm{e}^{-(1/2)t^2}\, dt. \tag{103}$$

Acknowledgments. I thank Professor Peter Sarnak for suggesting that I carry out these studies. His help and advice are gratefully acknowledged. I would also like to thank Professor Zeév Rudnick for helpful discussions.

REFERENCES

[1] M. V. BERRY AND J. GOLDBERG, *Renormalisation of curlicues*, Nonlinearity **1** (1988), 1–26.

[2] F. M. DEKKING AND M. MENDÈS FRANCE, *Uniform distribution modulo one: A geometrical viewpoint*, J. Reine Angew. Math. **329** (1981), 143–153.

[3] G. H. HARDY AND J. E. LITTLEWOOD, *Some problems in diophantine approximation, II*, Acta Math. **37** (1914), 193–239.

[4] ———, *A further note on the trigonometrical series associated with the elliptic* ϑ*-functions*, Proc. Cambridge Philos. Soc. **21** (1923), 1–5.

[5] D. HEJHAL, *The Selberg Trace Formula for* $\mathrm{PSL}(2,\mathbb{R})$*, Vol. 2*, Lect. Notes in Math. **1001**, Springer-Verlag, Berlin, 1983.

[6] ———, "On value distribution properties of automorphic functions along closed horocycles" in *XVIth Rolf Nevanlinna Colloquium (Joensuu, 1995)*, de Gruyter, Berlin, 1996, 39–52.

[7] E. HOPF, *Statistik der geodätischen Linien in Mannigfaltigkeiten negativer Krümmung*, Ber. Verh. Sächs. Akad. Wiss. Leipzig **91** (1939), 261–304.

[8] 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.

[9] ———, *On the central limit theorem for theta series*, Michigan Math. J. **29** (1982), 65–77.

[10] ———, *The uniform central limit theorem for theta sums*, Duke Math. J. **50** (1983), 649–666.

[11] G. LION AND M. VERGNE, *The Weil Representation, Maslov Index and Theta Series*, Progr. in Math. **6**, Birkhäuser, Boston, 1980.

[12] J. MARKLOF, "Theta sums, Eisenstein series, and the semiclassical dynamics of a precessing spin" in *Emerging Applications of Number Theory*, ed. D. Hejhal et al., IMA Vol. Math. Appl. **109**, Springer-Verlag, New York, 1998, 405–450.

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

[14] ———, *Class numbers of indefinite binary quadratic forms*, J. Number Theory **15** (1982), 229–247.

[15] G. SHIMURA, *Introduction to the Arithmetic Theory of Automorphic Functions*, Kanô Memorial Lectures **1**, Publ. Math. Soc. Japan **11**, Iwanami Shoten, Tokyo; Princeton Univ. Press, Princeton, 1971.

[16] ———, *On modular forms of half integral weight*, Ann. of Math. (2) **97** (1973), 440–481.
[17] ———, *On the transformation formulas of theta series*, Amer. J. Math. **115** (1993), 1011–1052.
[18] T. Shintani, *On construction of holomorphic cusp forms of half integral weight*, Nagoya Math. J. **58** (1975), 83–126.
[19] S. Thangavelu, *Lectures on Hermite and Laguerre Expansions*, Math. Notes **42**, Princeton Univ. Press, Princeton, 1993.

Department of Mathematics, Princeton University, Fine Hall, Washington Road, Princeton, New Jersey 08544, USA

Current: Institut des Hautes Études Scientifiques, 35 Route de Chartres, F-91440 Bures-sur-Yvette, France, marklof@ihes.fr; Division de Physique Théorique, Institute de Physique Nucléaire, F-91406 Orsay Cedex, France