ASIAN J. MATH.
Vol. 4, No. 4, pp. 839–854, December 2000

008

ON THE UNIFORM EQUIDISTRIBUTION OF LONG CLOSED HOROCYCLES*

DENNIS A. HEJHAL†

1. Introduction. I first met Hua during the 1979 Analytic Number Theory Symposium in Durham, England. Hua was quite stimulated by the talk I gave there on pseudo cusp forms for $PSL(2,\mathbb{Z})$ [5] and, for several years thereafter, was a major source of encouragement to me as I began working more systematically with computational spectral theory on Fuchsian groups.

I retain many warm memories from the discussions we had during this period.

Though the spectral techniques in the present paper are decidedly *non*-computational, it is a curious fact that the issue considered here first arose while trying to place certain machine-based heuristics, specifically those of [11], on a more satisfactory geometric footing.

Paper [11] is part of the computational series [6-12] that was first envisioned in conversations with Hua.

In light of this, it seems fair to regard the present paper's main result, theorem A in §3, as having a genesis which springs partially at least from Hua's enthusiasm for computationally-oriented mathematics.

2. Some preliminaries. To get started, we need a bit of notation. Let Γ be any cofinite Fuchsian group acting on the Poincaré upper half-plane H. Assume that $\Gamma \setminus H$ has only one cusp. By making an auxiliary conjugation of Γ, one can position this cusp at $i\infty$ and arrange things so that the isotropy subgroup Γ_∞ is generated by the translation $S(z) = z + 1$. At the same time, it is convenient to equip Γ with a standard fundamental polygon $\mathcal{F}$ which contains the half-strip $[0,1] \times [B,\infty)$ for some $B \geqq 1$. (Cf. [4, pp. 3-5] and [15, pp. 59,61].)

The numerical set-up of both [11] and [12] necessitates looking at the pull-back of $\{0 \leqq x \leqq 1, y = h\}$ inside $\mathcal{F}$ for small values of h. We'll denote this pull-back by $\mathcal{C}_h$.

Since $ds = |dz|/y$, the locus $\mathcal{C}_h$ has hyperbolic length $1/h$. On the other hand, since $S \in \Gamma$, its projection $\pi(\mathcal{C}_h)$ on $\Gamma \setminus H$ is manifestly a closed curve; one calls $\pi(\mathcal{C}_h)$ a *closed horocycle.* (Note that $\pi(\mathcal{C}_h)$ is real-analytic.)

In view of the fact that $1/h \to \infty$ as $h \to 0$, it is only natural to wonder where $\pi(\mathcal{C}_h)$ "goes".

Computer tests show some beautiful patterns and quickly suggest that $\mathcal{C}_h$ becomes everywhere dense in $\mathcal{F}$ as $h \to 0$. Much more, in fact, is true. The set $\mathcal{C}_h$ actually becomes equidistributed with respect to Poincaré area μ as the parameter h decays. (Here $d\mu = y^{-2}dxdy$.) By elementary functional analysis, this is the same thing as saying that

$$(2.1) \qquad \int_0^1 f(x+ih)dx \longrightarrow \frac{1}{\mu(\mathcal{F})} \int_{\mathcal{F}} f(z)d\mu(z)$$

*Received August 7, 2000; accepted for publication October 7, 2000.

†Department of Mathematics, Uppsala University, S-75106 Uppsala, Sweden; and School of Mathematics, University of Minnesota, Minneapolis, Mn. 55455, USA (hejhal@math.uu.se, hejhal@math.umn.edu). While preparing this paper, the author was supported by the Göran Gustafsson Foundation for Research in Natural Science and Medicine, and by grants from the Institute for Advanced Study and NEC Research Institute, Inc.

holds for every compactly supported function $f \in C^2(\Gamma \setminus H)$ as $h \to 0^+$.

Relation (2.1) seems to have been independently considered by any number of people over the years including A. Selberg (unpublished; but, cf. [19, eqs. (15)(17)], [20, (2.16)]), D. Zagier [24, p. 279], and P. Sarnak [17]. Sarnak actually obtains a sharper and more general — phase space — version of (2.1).

Once (2.1) is known, it is natural to go deeper and ask if $\mathcal{C}_h$'s equidistribution takes place in fact more locally (*or* uniformly). Specifically: for numbers $0 \leqq a < b \leqq 1$ satisfying $(b-a)/h \to \infty$, is it true that

$$\text{(2.2)} \qquad \frac{1}{b-a}\int_a^b f(x+ih)dx \longrightarrow \frac{1}{\mu(\mathcal{F})}\int_{\mathcal{F}} f(z)d\mu(z)\ ?$$

In this generality, matters definitely undergo a split.

For fixed a and b, relation (2.2) is a theorem. See [11, p. 44] and the spectral-theoretic proof outlined there. The same assertion can also be obtained using ergodic-theoretic techniques; see [3, p. 206 (sketch)] as well as [22, theorem 1.4]. (I am grateful to Jens Marklof for drawing my attention to this last fact.)

At the other extreme, by looking at elements $T = \left(\begin{smallmatrix} r & s \\ c & d \end{smallmatrix}\right) \in \Gamma$ with $c \neq 0$ and intersecting $T\{\mathbb{R} \times [2B, \infty)\}$ with $\{Im(z) = h\}$, it is evident that there are numerous cases with $b - a \sim$ (const.)$\sqrt{h}$ for which the pull-back $\mathcal{C}_h[a,b]$ is not even close to being equidistributed.

The (Euclidean) length scale $\sqrt{h}$ thus seems to have a special significance.

Our aim in this paper will be to show how the outline given in [11, p. 44] can be strengthened so as to prove that relation (2.2) actually holds uniformly[1] anytime

$$b - a \geqq h^{c(\Gamma)-\varepsilon},$$

where $c(\Gamma)$ is a certain positive constant (less than $\frac{1}{2}$) depending solely on the geometry of $\Gamma \setminus H$.

3. Statement of our main theorem. Let Δ denote the non-Euclidean Laplacian and $0 = \lambda_0 < \lambda_1 \leqq \lambda_2 \leqq \ldots$ be the associated list of discrete eigenvalues of $-\Delta$ for $\Gamma \setminus H$. Write $\lambda_n = s_n(1 - s_n)$ and $M = \max\{j : \lambda_j < \frac{1}{4}\}$ as in [4, p. 472], the multiplier system $(m, \mathcal{W})$ being understood to be trivial. The numbers $s_0, \ldots, s_M$ will thus lie in the semi-open interval $(\frac{1}{2}, 1]$. Finally, to conveniently accomodate functions which tend to a nonzero limit as $y \to \infty$, we introduce:

$$C_b^k(\Gamma \setminus H) = C^k(\Gamma \setminus H) \cap L^\infty(\Gamma \setminus H).$$

THEOREM A. *Let*

$$c(\Gamma) = \left\{ \begin{array}{ll} \frac{1}{3}, & \textit{if } M = 0 \\ \min(\frac{1}{3}, 1 - \frac{1}{3-2s_1}), & \textit{if } M \geqq 1 \end{array} \right\},$$

$0 < \varepsilon < c(\Gamma)$, *and* $f \in C_b^2(\Gamma \setminus H), \Delta f \in L^2(\Gamma \setminus H)$. *Then*

$$\frac{1}{b-a}\int_a^b f(x+ih)dx \longrightarrow \frac{1}{\mu(\mathcal{F})}\int_{\mathcal{F}} f(z)d\mu(z)$$

uniformly as $h \to 0$ *so long as* $b - a$ *remains bigger than* $h^{c(\Gamma)-\varepsilon}$. *The relevant difference will go to zero in fact like a small power of* h.

[1] for given f

Though formulated in a one-cusp setting, exactly the same result holds if $\Gamma \setminus H$ has several cusps. (One of the cusps is distinguished as being $i\infty$.)

COROLLARY. *When Γ is congruence subgroup, uniform equidistribution takes place for $b - a \geqq h^{\frac{1}{3}-\varepsilon}$.*

Indeed, for any congruence subgroup, one knows that $\lambda_1 \geqq \frac{3}{16}$ (i.e., $s_1 \leqq \frac{3}{4}$). Cf. [20, p. 13 (bot)] and [14, p. 184]. The conjecture, of course, is that $\lambda_1 \geqq \frac{1}{4}$. Regrettably, the number $c(\Gamma)$ in Theorem A levels off at $\frac{1}{3}$ once s_1 passes through $\frac{3}{4}$. [2]

4. Laying the groundwork for the proof. Apart from some material about trigonometric polynomials approximating $\chi_{ab}(x)$, the characteristic function of $[a, b]$ (for which we simply refer to [16, pp. 6, 8][3]), the proof of theorem A can basically be seen as a new section of [4] — insertable just after page 709.

To keep matters to the point, we shall assume that the reader already has at least a modest familiarity with the contents of [4] in the case of a trivial multiplier system: specifically chapters 6-8, section 2 of appendix E, and pp. 570, 583, 645, 646, 732(note 2). We shall also be content, as matters progress, to just *indicate* most of the steps (filling in the details being largely pedestrian modulo the aforementioned material from [4]).

Technique-wise, the proof of theorem A will be seen to be a mixture of spectral theory, harmonic analysis, and a couple of very simple L^p estimates. The function ([4, p. 666])

$$P_n(z; s) = \sum_{W \in [S] \setminus \Gamma} (Im Wz)^s e^{-2\pi|n| Im(Wz)} e^{2\pi i n Re(Wz)},$$

which generalizes the Eisenstein series $E(z; s)$, will play the role of a building block (attached in some sense to $e^{2\pi i n x}$). [4]

To expedite matters, we restrict attention to $\{Re(s) > \frac{1}{2}\}$ and agree that all implied constants [in our subsequent "big O" estimates] depend solely on $\Gamma, \mathcal{F}$ unless otherwise indicated by subscripts.

The analytic continuation properties of $P_n(z; s)$ for $n \neq 0$ are most easily visualized by noting that

$$\frac{(\pi|n|)^{s-\frac{1}{2}}}{\Gamma(s + \frac{1}{2})} P_n(z; s)$$

differs from the more standard function ([4, pp. 41-42, 255-257])

$$F_n(z; s) = \sum_{W \in [S] \setminus \Gamma} (Im Wz)^{\frac{1}{2}} I_{s-\frac{1}{2}}(2\pi|n| Im(Wz)) e^{2\pi i n Re(Wz)}$$

by a W-summation which is readily checked to be nicely holomorphic for all $\{Re(s) > \frac{1}{2}\}$. The essential trick here is to keep $z \in \mathcal{F}$ without loss of generality, and then split the W-sum into two parts: $W \in [S], W \notin [S]$. In the latter part, $Im(Wz)$ is uniformly

---

[2] According to a forthcoming result of A. Strömbergsson (Uppsala), $c(\Gamma)$ can be taken to be $\frac{1}{2}$ anytime there are no Eisenstein poles in $(\frac{1}{2}, 1)$. This is the case, for instance, in a congruence subgroup. Strömbergsson's techniques are much different than those we use here.

[3] cf. also [1, pp. 15-19]

[4] See (4.8) and (4.9) below.

bounded, so passage to the series expansion of $I_{s-\frac{1}{2}}(2\pi|n|Im(Wz))$ is very natural. The pertinent estimate is then subsumed by [4, p. 667 (proposition 2)] with $s \to s+1$.

A similar procedure will be used later to get good L^∞ bounds for certain (shifted) linear combinations of P_n. See (5.3).

As mentioned in [4, p. 668],

$$P_n(z;s) = O_s(1)y^{1-s} \tag{4.1}$$

for $n \neq 0, Re(s) > 1, z \in \mathcal{F}$. Uninspired term-by-term differentiation shows that

$$\begin{aligned} &[\text{any first partial of } P_n(z;s) \text{ w.r.t. } x,y] = O_s(1)y^{-s} \\ &[\text{any second partial of } P_n(z;s) \text{ w.r.t. } x,y] = O_s(1)y^{-s-1} \end{aligned} \tag{4.2}$$

under the same conditions. The associated W-sums are uniformly absolutely convergent on compact subsets of $H \times \{Re(s) > 1\}$. One also checks [4, p. 669] that

$$\Delta P_n(z;s) + s(1-s)P_n(z;s) = -4\pi|n|sP_n(z;s+1). \tag{4.3}$$

Simple use of [4, p. 667(proposition 2)] permits one to see that the implied constant in (4.1) can be taken to be

$$\frac{O(1)}{Re(s)-1} \tag{4.1'}$$

for, say, $1 < Re(s) < 100$.

Another function of central importance to us on $\{\mathrm{Re}(s) > \frac{1}{2}\}$ is the resolvent kernel $G_s(z;w)$. For this, cf. [4, pp. 33, 244 (2.4)(2.5), 250 (thm. 3.5)].

There are two facts of particular interest when $s \neq s_0, s_1, \ldots, s_M$. First, that

$$\Delta F + s(1-s)F = Q$$

holds with (given) $F, Q \in C^2(\Gamma \setminus H) \cap L^2(\Gamma \setminus H)$ if and only if

$$F(z) = \int_{\mathcal{F}} G_s(z;w)Q(w)d\mu(w). \tag{4.4}$$

Second, that when F is defined by (4.4) for a Q known only to be in $L^2(\Gamma \setminus H)$, one automatically has $F \in C(\Gamma \setminus H) \cap L^2(\Gamma \setminus H)$ and

$$\int_{\mathcal{F}} |F(z)|^2 d\mu(z) \leqq \frac{1}{\operatorname{dist}[s(1-s), \operatorname{Spec}(-\Delta)]^2} \int_{\mathcal{F}} |Q(w)|^2 d\mu(w). \tag{4.5}$$

The set $\operatorname{Spec}(-\Delta)$ is understood here to include the continuous spectrum $[\frac{1}{4}, \infty)$ as well.

The second assertion is the more basic one; its proof follows from direct substitution of the L_2 spectral expansions of G_s and Q. Cf. [4, pp. 244, 245, 263 (note 16A)].

The first assertion is then proved by writing $F_0(z)$ for the integral in (4.4) and applying $\Delta + s(1-s)I$ to either F_0 or $F - F_0$. For F_0, cf. the reasoning in [4, pp. 97(bot), 98(top), 646(top)].

As in [4, p. 669], the foregoing assertions about (4.4) immediately lead to the identity

$$P_n(z;s) = -4\pi s \int_{\mathcal{F}} G_s(z;w)|n|P_n(w;s+1)d\mu(w), \tag{4.6}$$

first for $Re(s) > 1$, then for $Re(s) > \frac{1}{2}$. Here $n \neq 0$. In particular: $P_n \in C^2(\Gamma \setminus H) \cap L^2(\Gamma \setminus H)$ for $s \neq s_0, s_1, \dots, s_M$.

A second identity of significant interest is that

$$\int_{\mathcal{F}} f(z)P_n(z;s)d\mu = \frac{1}{s(s-1)}\int_{\mathcal{F}}(\Delta f)P_n(z;s)d\mu + \frac{4\pi}{s-1}\int_{\mathcal{F}} f(z)|n|P_n(z;s+1)d\mu \tag{4.7}$$

for $Re(s) > 1$ and any f as in theorem A. The proof consists of applying Green's identity (with $u = f, v = P_n$) on

$$\mathcal{F}_Y = \mathcal{F} \cap \{Im(z) < Y\}.$$

The fact that

$$\iint_{\mathcal{F}} \left(|f_x|^2 + |f_y|^2\right) dxdy < \infty$$

(cf. [4, p. 732]) permits one to conclude that the inequality

$$\int_0^1 |f_y(x+iy)|^2 dx < \frac{1}{y}$$

holds infinitely often as $y \to \infty$. Coupled with the estimates (4.1)+(4.2), this is enough to get

$$s(1-s)\int_{\mathcal{F}} fP_n d\mu + \int_{\mathcal{F}}(\Delta f)P_n d\mu + 4\pi s\int_{\mathcal{F}} f(z)|n|P_n(z;s+1)d\mu = 0,$$

i.e., the desired conclusion. Relation (4.7) actually holds for a wider class of f, but we do not need this.

In proving theorem A, one exploits the counterpart of (4.7) for a generic linear combination

$$\mathcal{B}(z;s) = \sum_{1 \leqq |n| \leqq K} b_n P_n(z;s). \tag{4.8}$$

The motivation for this stems from the fact that $\int_{\mathcal{F}} f\mathcal{B}d\mu$ is nothing but

$$\int_0^\infty y^{s-2}\Big(\int_0^1 f(x+iy)\Big[\sum_{1 \leqq |n| \leqq K} b_n e^{-2\pi|n|y}e^{2\pi i n x}\Big]dx\Big)dy, \tag{4.9}$$

as one sees by unfolding. Since f is bounded, the iterated integral converges absolutely for $Re(s) > 1$.

Finally, it is convenient to recall that, for the Mellin transform

$$\mathcal{M}(s) \equiv \int_0^\infty A(y)y^{s-1}dy,$$

Plancherel's theorem asserts that

$$\int_0^\infty |y^c A(y)|^2 \frac{dy}{y} = \frac{1}{2\pi}\int_{-\infty}^\infty |\mathcal{M}(c+it)|^2 dt. \tag{4.10}$$

The Mellin inversion formula (in the L_2 sense) then becomes

$$\lim_{T\to\infty}\int_0^\infty y^{2c}\left|A(y)-\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\mathcal{M}(s)y^{-s}ds\right|^2\frac{dy}{y}=0.$$

For $y^{c-1}A(y)\in C^1(\mathbb{R}^+)\cap L^1(\mathbb{R}^+)$, one knows that

$$A(y)=\frac{1}{2\pi i}\lim_{T\to\infty}\int_{c-iT}^{c+iT}\mathcal{M}(s)y^{-s}ds \tag{4.11}$$

pointwise.

5. Proof of Theorem A. With the groundwork in place, the way is now clear to proceeding directly into the proof. Write

$$\mathcal{B}(z;s)=\sum_{1\leqq|n|\leqq K}b_nP_n(z;s)$$

as before, and let $s=\sigma+it$. Also write

$$\hat{\mathcal{B}}(z;s)=\sum_{1\leqq|n|\leqq K}b_n|n|P_n(z;s+1).$$

By (4.3) and (4.6), one knows that

$$\Delta\mathcal{B}+s(1-s)\mathcal{B}=-4\pi s\hat{\mathcal{B}} \tag{5.1}$$

$$\mathcal{B}(z;s)=-4\pi s\int_{\mathcal{F}}G_s(z;w)\hat{\mathcal{B}}(w;s)d\mu(w) \tag{5.2}$$

for $s\neq s_0,s_1,\ldots,s_M$. Let $\|f\|_p$ have its usual meaning and put

$$\hat{\mathcal{M}}=\sum_{1\leqq|n|\leqq K}|n|^{\frac{1}{2}}|b_n|.$$

By directly estimating the associated W-sum utilizing [4, p. 667 (proposition 2)] and the fact that

$$0\leqq\sqrt{u}e^{-2\pi u}\leqq e^{-\pi u},$$

one immediately checks that

$$|\hat{\mathcal{B}}(z;s)|=O(1)\hat{\mathcal{M}}\frac{y^{\frac{1}{2}-\sigma}}{\sigma-\frac{1}{2}} \tag{5.3}$$

for $z\in\mathcal{F}$ and (say) $\frac{1}{2}<\sigma<10$. This estimate "gets the ball rolling." The estimate $O(1)\sqrt{K}\hat{\mathcal{M}}y^{-\sigma}$ is *also* available, but the factor $\sqrt{K}$ is troublesome later (completely overpowering $(\sigma-\frac{1}{2})^{-1}$ in our final, optimized set-up).

To treat

$$Q(s)\equiv\int_{\mathcal{F}}f(z)\mathcal{B}(z;s)d\mu(z), \tag{5.4}$$

we need some L_2 estimates for $\mathcal{B}(z;s)$. By coupling (4.5), (5.2), and (5.3) with a mimic of [4, pp. 670-672], one first sees that

$$\int_{\mathcal{F}} |\mathcal{B}(z;s)|^2 d\mu(z) \leqq \frac{16\pi^2|s|^2}{\operatorname{dist}(*)^2} \int_{\mathcal{F}} |\hat{\mathcal{B}}(w;s)|^2 d\mu(w), \tag{5.5}$$

then that

$$\int_{\mathcal{F}} |\mathcal{B}(z;s)|^2 d\mu(z) = \left\{ \begin{array}{ll} O(1)\dfrac{\hat{\mathcal{M}}^2}{(\sigma-\frac{1}{2})^4}, & \text{for } \ \frac{1}{2} < \sigma < 10, |t| \geqq 1 \\ O(1)\dfrac{\hat{\mathcal{M}}^2}{(\sigma-\frac{1}{2})^4|s-\frac{1}{2}|^2}, & \text{for } \ \frac{1}{2} < \sigma < \frac{1}{2}+\beta, |t| \leqq 1 \end{array} \right\}, \tag{5.6}$$

wherein

$$\beta = \frac{1}{2}(s_M - \frac{1}{2}).$$

Concomitantly, by (4.1′),

$$\int_{\mathcal{F}} |\mathcal{B}(z;s)|^2 d\mu(z) = O(1)\frac{\hat{\mathcal{M}}^2}{(\sigma-1)^2} \quad \text{for} \quad 1 < \sigma < 10. \tag{5.7}$$

For any other s in $\{\frac{1}{2} < \sigma < 10\}$, we just use (5.5) as stated.

The function Q is thus analytic for $s \neq s_0, s_1, \ldots, s_M$. By (5.5) and the Cauchy-Schwarz inequality, $Q(s)$ has at most first-order poles [even if some of the λ_j occur with multiplicity]. Since f need not have compact support, a bit of care is necessary in determining the respective singular parts. The simplest approach is to note that the improper integral

$$\int_{\mathcal{F}} f(z)(s-\alpha)\mathcal{B}(z;s)d\mu(z)$$

converges locally uniformly (w.r.t. s) for each $\alpha \in (\frac{1}{2}, 1]$. Cf. (5.5). One can therefore compute $\operatorname{Res}\{Q(s); s = \alpha\}$ as

$$\lim_{Y\to\infty} \operatorname{Res}\{Q_Y(s); s = \alpha\},$$

in the obvious notation. By [4, p. 256 (corollary 4.4)] and our earlier comments about relating P_n to $F_n(z;s)$, one immediately finds that

$$\operatorname{Res}\{Q(s); s = \alpha\} = \frac{\Gamma(\alpha+\frac{1}{2})}{2\alpha-1} \sum_{s_j=\alpha} \Big(\sum_n \frac{b_n \bar{c}_{n,j}}{(\pi|n|)^{\alpha-1/2}}\Big)\Big(\int_{\mathcal{F}} f\varphi_j d\mu\Big) \tag{5.8}$$

where $c_{n,j}$ is the usual Fourier coefficient for $\varphi_j(z)$.

Corresponding to (4.7), we now write

$$Q(s) = \frac{1}{s(s-1)}U(s) + \frac{4\pi}{s-1}W(s) \tag{5.9}$$

for $s \neq s_0, s_1, \ldots, s_M$, with

$$U(s) = \int_{\mathcal{F}} (\Delta f)\mathcal{B}(z;s)d\mu \quad \text{and} \quad W(s) = \int_{\mathcal{F}} f(z)\hat{\mathcal{B}}(z;s)d\mu.$$

For $\frac{1}{2} < \sigma < 10$ and $|t| \geqq 1$, we clearly get

$$\begin{aligned}|Q(\sigma+it)| &= O(t^{-2})\|\Delta f\|_2\|\mathcal{B}\|_2 + O(t^{-1})\|f\|_2\|\hat{\mathcal{B}}\|_2 \\ &= O(t^{-2})\|\Delta f\|_2\frac{\hat{\mathcal{M}}}{(\sigma-\frac{1}{2})^2} + O(t^{-1})\|f\|_2\frac{\hat{\mathcal{M}}}{\sigma-\frac{1}{2}} \\ &= O(t^{-1})\frac{\hat{\mathcal{M}}}{(\sigma-\frac{1}{2})^2}[\|\Delta f\|_2 + \|f\|_2]\end{aligned}$$

as an *a priori* bound, by (5.6) and (5.3).

On the other hand, let us now also write

$$I(y) = \int_0^1 f(x+iy)\Big(\sum_{1\leqq|n|\leqq K} b_n e^{-2\pi|n|y}e^{2\pi inx}\Big)dx. \tag{5.10}$$

Clearly $|I(y)| \leqq \hat{\mathcal{M}}\|f\|_\infty e^{-2\pi y}$. By (4.9) and the Mellin inversion formula with $A(y) = y^{-1}I(y)$,

$$I(y) = \frac{y}{2\pi i}\int_{c-i\infty}^{c+i\infty} Q(s)y^{-s}ds$$

for any $c > 1$. The right-hand side is an improper Riemann integral; cf. (4.11).

By the Cauchy residue theorem and the aforementioned *a priori* bound for $|Q|$, we then get

$$I(y) = \frac{y}{2\pi i}\int_{\frac{1}{2}+\delta-i\infty}^{\frac{1}{2}+\delta+i\infty} Q(s)y^{-s}ds + \sum_{\frac{1}{2}+\delta<\alpha\leqq 1} \mathrm{Res}\{Q(s)y^{1-s}; s=\alpha\} \tag{5.11}$$

for any $\delta \in (0,\beta]$. The s-integral is again interpreted à la (4.11).

The U-portion of the s-integral is immediately seen to be

$$O(1)y^{\frac{1}{2}-\delta}\frac{\hat{\mathcal{M}}\|\Delta f\|_2}{\delta^2}\log\Big(\frac{1}{\delta}\Big) \tag{5.12}$$

by (5.6).

The W-portion is clearly

$$O(1)y^{\frac{1}{2}-\delta}\Big(\int_{-\infty}^{\infty}|W(\frac{1}{2}+\delta+it)|^2dt\Big)^{1/2}.$$

By the analog of (4.9), however,

$$W(s) = \int_0^\infty v^{s-1}\hat{I}(v)dv$$

for $Re(s) > \frac{1}{2}$, where

$$\hat{I}(v) = \int_0^1 f(x+iv)\Big(\sum_{1\leqq|n|\leqq K} |n|b_n e^{-2\pi|n|v}e^{2\pi inx}\Big)dx.^5$$

Accordingly, by (4.10),

$$\int_{-\infty}^{\infty} |W(\sigma+it)|^2 dt = 2\pi \int_0^{\infty} v^{2\sigma-1} |\hat{I}(v)|^2 dv.$$

Since

$$\sum_{1 \leqq |n| \leqq K} |n||b_n| e^{-2\pi|n|v} \leqq \frac{1}{\sqrt{v}} \hat{\mathcal{M}} e^{-\pi v},$$

this trivially yields

$$\int_{-\infty}^{\infty} |W(\sigma+it)|^2 dt = O(1) \hat{\mathcal{M}}^2 \|f\|_\infty^2 \Gamma(2\sigma-1).$$

Almost as easily,

$$|\hat{I}(v)|^2 \leqq \frac{\mathcal{M}^2}{v} \int_0^1 |f(x+iv)|^2 e^{-2\pi v} dx$$

$$\begin{aligned}\int_0^{\infty} v^{2\sigma-1} |\hat{I}(v)|^2 dv &\leqq \hat{\mathcal{M}}^2 \int_0^{\infty} \int_0^1 v^{2\sigma} |f(x+iv)|^2 e^{-2\pi v} dx \frac{dv}{v^2} \\ &= \hat{\mathcal{M}}^2 \int_{\mathcal{F}} |f(z)|^2 \Big[\sum_{W \in [S]\backslash\Gamma} (ImWz)^{2\sigma} e^{-2\pi Im(Wz)} \Big] d\mu(z) \\ &= \hat{\mathcal{M}}^2 \int_{\mathcal{F}} |f(z)|^2 O(1) \frac{v^{1-2\sigma}}{2\sigma-1} d\mu(z)\end{aligned}$$

for $\frac{1}{2} < \sigma < 2$ (say), by the procedure of (4.1)+(4.1′). This gives

$$\int_{\infty}^{\infty} |W(\sigma+it)|^2 dt = O(1) \hat{\mathcal{M}}^2 \|f\|_2^2 \frac{1}{2\sigma-1}.$$

The W-portion of the s-integral can therefore be expressed as either

$$O(1) y^{\frac{1}{2}-\delta} \frac{\hat{\mathcal{M}}\|f\|_\infty}{\sqrt{\delta}} \quad \text{or} \quad O(1) y^{\frac{1}{2}-\delta} \frac{\hat{\mathcal{M}}\|f\|_2}{\sqrt{\delta}}. \tag{5.13}$$

We'll go with the former.

Upon combining (5.11)-(5.13) with (5.8), we get:

$$\begin{aligned} I(y) &= O(1) \frac{y^{\frac{1}{2}-\delta}}{\delta^3} \hat{\mathcal{M}} (\|\Delta f\|_2 + \|f\|_\infty) \\ &+ \sum_{j=0}^{M} \frac{\Gamma(s_j+\frac{1}{2})}{2s_j-1} \langle f, \bar{\varphi}_j \rangle \Big(\sum_{1 \leqq |n| \leqq K} \frac{b_n \bar{c}_{n,j}}{(\pi|n|)^{s_j-1/2}} \Big) y^{1-s_j}. \end{aligned} \tag{5.14}$$

Note here that y is arbitrary.

For $0 < y < \frac{1}{10}$ (say), we now optimize by taking

$$\delta = \frac{\beta}{\log(1/y)}.$$

[5] We use v as a dummy variable to avoid any confusion with the y in (5.11).

This finally gives

$$I(y) = O(1)y^{\frac{1}{2}}(\log\frac{1}{y})^3\hat{\mathcal{M}}(\|\Delta f\|_2 + \|f\|_\infty) \tag{5.15}$$
$$+\sum_{j=0}^{M}\frac{\Gamma(s_j+\frac{1}{2})}{2s_j-1}\langle f,\bar{\varphi}_j\rangle\Big(\sum_{1\leqq|n|\leqq K}\frac{b_n\bar{c}_{n,j}}{(\pi|n|)^{s_j-1/2}}\Big)y^{1-s_j}.$$

To complete the picture, one needs to include $n = 0$, *i.e.*, obtain a similar formula for

$$\int_0^1 f(x+iy)b_0dx.$$

Cf. (5.10). This was done in [11, pp. 41-42]. Our procedure there was simply to expand $f(z)$ in a Hilbert-Schmidt type spectral expansion à la [4, pp. 244-245] and then integrate both sides with respect to x. Cf. [4, p. 732] apropos [4, p. 243 condition (c_0)]. One finds that:

$$\int_0^1 f(x+iy)dx = \sum_{j=0}^{M}\langle f,\varphi_j\rangle A_j y^{1-s_j} + O(1)\sqrt{y}(\|\Delta f\|_2 + \|f\|_2),$$

where

$$\int_0^1 \varphi_j(x+iy)dx = A_j y^{1-s_j}.$$

By replacing f by $\bar{f}$, this can be re-expressed as

$$\int_0^1 f(x+iy)dx = \sum_{j=0}^{M}\langle f,\bar{\varphi}_j\rangle \bar{A}_j y^{1-s_j} + O(1)\sqrt{y}(\|\Delta f\|_2 + \|f\|_2). \tag{5.16}$$

Of course, in this relation,

$$A_0 = \varphi_0(z) = \frac{1}{\sqrt{\mu(\mathcal{F})}} \quad \text{and} \quad s_0 = 1.$$

All told, then,

$$\int_0^1 f(x+iy)\Big(\sum_{|n|\leqq K} b_n e^{-2\pi|n|y}e^{2\pi inx}\Big)dx \tag{5.17}$$
$$= \frac{b_0}{\mu(\mathcal{F})}\int_{\mathcal{F}} f(z)d\mu(z) + \sum_{j=1}^{M} b_0\langle f,\bar{\varphi}_j\rangle\bar{A}_j y^{1-s_j}$$
$$+\sum_{j=1}^{M}\frac{\Gamma(s_j+\frac{1}{2})}{2s_j-1}\langle f,\bar{\varphi}_j\rangle\Big(\sum_{1\leqq|n|\leqq K}\frac{b_n\bar{c}_{n,j}}{(\pi|n|)^{s_j-1/2}}\Big)y^{1-s_j}$$
$$+ O(1)y^{\frac{1}{2}}(\log\frac{1}{y})^3(\hat{\mathcal{M}} + |b_0|)(\|\Delta f\|_2 + \|f\|_\infty)$$

for $0 < y < 1/10$.

Since

$$\Big|\int_0^1 f(x+iy)\Big[\sum_{|n|\leqq K} b_n(1-e^{-2\pi|n|y})e^{2\pi inx}\Big]dx\Big|$$
$$\leqq \|f\|_\infty \int_0^1 \sum_{|n|\leqq K} |b_n|(2\pi)^{\frac{1}{2}}|n|^{\frac{1}{2}}y^{\frac{1}{2}}dx = (2\pi)^{\frac{1}{2}}\|f\|_\infty \hat{\mathcal{M}} y^{\frac{1}{2}}$$

by virtue of the inequality $0 \leqq 1-e^{-u} \leqq \sqrt{u}$, one also has

$$\int_0^1 f(x+iy)\Big(\sum_{|n|\leqq K} b_n e^{2\pi inx}\Big)dx = \{\text{RHS of (5.17)}\} \tag{5.18}$$

for $0 < y < 1/10$.

The simplest way of handling

$$\sum_{1\leqq|n|\leqq K} \frac{b_n \bar{c}_{n,j}}{(\pi|n|)^{s_j-1/2}}$$

is to apply Cauchy-Schwarz and the *a priori* Rankin-Selberg bound developed in [14, pp. 60-61]. (We remark that some minor revisions are called for on p. 61 top.) One concludes that

$$\Big|\sum_{1\leqq|n|\leqq K} \frac{b_n \bar{c}_{n,j}}{(\pi|n|)^{s_j-1/2}}\Big| = O(1)\Big(\sum_{1\leqq|n|\leqq K} \frac{|b_n|^2}{|n|^{2s_j-1}}\Big)^{\frac{1}{2}}\sqrt{K}.$$

We can therefore re-state (5.18) as

$$\begin{aligned} &\int_0^1 f(x+iy)\Big(\sum_{|n|\leqq K} b_n e^{2\pi inx}\Big)dx \\ &= \frac{b_0}{\mu(\mathcal{F})}\int_{\mathcal{F}} f(z)d\mu(z) + \sum_{j=1}^{M} C_j y^{1-s_j} \\ &+ O(1)y^{\frac{1}{2}}(\log\frac{1}{y})^3 \hat{\mathcal{M}}_0(\|\Delta f\|_2 + \|f\|_\infty) \end{aligned} \tag{5.19}$$

for $0 < y < \frac{1}{10}$, where

$$\hat{\mathcal{M}}_0 = |b_0| + \sum_{|n|\leqq K} |n|^{\frac{1}{2}}|b_n|$$

and the numbers C_j satisfy

$$|C_j| \leqq O(1)\|f\|_\infty\Big[|b_0| + \sqrt{K}\Big(\sum_{1\leqq|n|\leqq K} \frac{|b_n|^2}{|n|^{2s_j-1}}\Big)^{\frac{1}{2}}\Big]. \tag{5.20}$$

For large K, the *idea* is to now select b_n so that the L_1-norm

$$\int_0^1 |\chi_{ab}(x) - P_K(x)|dx$$

with

$$P_K(x) = \sum_{|n| \leqq K} b_n e^{2\pi i n x}$$

is essentially minimal. The same type of idea is utilized in modern proofs of the Erdös-Turán inequality (in the theory of uniform distribution mod 1). Cf. [16, pp. 6, 8] and [1, pp. 15-20]. See also [2, 21, 23] for, among other things, the related problem of approximating $\chi_{ab}(x)$ in $L^1(\mathbb{R})$ using band-limited functions

$$F_K(x) = \int_{-K}^{K} b(t) e^{2\pi i t x} dt.$$

The comments in [16, p. 14 (§3)] reflect this.

The pertinent minimum for the original P_K-problem is $\leqq \frac{1}{K}$ for any $[a, b]$. As in [16, pp. 6, 8], this quickly leads to

$$\left\{ \begin{array}{lcl} b_0 & = & b - a + O(\frac{1}{K}) \\ |b_n| & \leqq & O(\frac{1}{K}) + O(1) \min(b-a, \frac{1}{|n|}) \end{array} \right\}. \tag{5.21}$$

Relation (5.19) will then yield

$$\begin{aligned} \frac{1}{b-a} \int_a^b f(x+iy) dx = & \frac{1}{\mu(\mathcal{F})} \int_{\mathcal{F}} f(z) d\mu(z) + O(1) \frac{\|f\|_\infty}{K(b-a)} \\ & + \frac{1}{b-a} \sum_{j=1}^{M} C_j y^{1-s_j} \\ & + O(1) y^{\frac{1}{2}} (\log \frac{1}{y})^3 \frac{\sqrt{K}}{b-a} (\|\Delta f\|_2 + \|f\|_\infty). \end{aligned} \tag{5.22}$$

There is no hope of getting a good remainder term here unless $K(b-a) \to \infty$. We therefore assume $K(b-a) > 1$ and continue. After a bit of calculation, (5.20) and (5.21) give

$$|C_j| = O(1) \|f\|_\infty [b - a + \sqrt{K} (b-a)^{s_j}] = O(1) \|f\|_\infty \sqrt{K} (b-a)^{s_j}.$$

Thus

$$\begin{aligned} \frac{1}{b-a} \int_a^b f(x+iy) dx = & \frac{1}{\mu(\mathcal{F})} \int_{\mathcal{F}} f(z) d\mu(z) + O(1) \frac{\|f\|_\infty}{K(b-a)} \\ & + O(1) \|f\|_\infty \sqrt{K} \Big(\frac{b-a}{y}\Big)^{s_1 - 1} \\ & + O(1) y^{\frac{1}{2}} (\log \frac{1}{y})^3 \frac{\sqrt{K}}{b-a} (\|\Delta f\|_2 + \|f\|_\infty) \end{aligned} \tag{5.23}$$

for $0 < y < \frac{1}{10}$ anytime $b - a > \max(y, K^{-1})$. (The s_1-term is understood to be empty when $M = 0$.) [6]

[6] It also bears mentioning that, on the RHS of (5.23), the 1^{st} and 3^{rd} K's originate in the Beurling/Selberg approximation, while the 2^{nd} stems from Rankin-Selberg.

Put $b - a = y^\omega$ and $K = y^{-\alpha}$, keeping $0 \leqq \omega < 1$ and $\omega < \alpha$ to ensure that $(b-a)/y \to \infty$ and $K(b-a) > 1$. Things are now optimized in (5.23) by considering the intersection points of 2 or 3 obvious lines in the $\alpha\omega$-plane. For $M = 0$, the "key vertex" is $(\alpha, \omega) = (\frac{1}{3}, \frac{1}{3})$; for $M \geqq 1$, it's

$$\left\{ \begin{array}{ll} \left(\frac{1}{3}, \frac{1}{3}\right), & \text{when } \frac{1}{2} < s_1 \leqq \frac{3}{4} \\ \left(\frac{2-2s_1}{3-2s_1}, \frac{2-2s_1}{3-2s_1}\right), & \text{when } \frac{3}{4} < s_1 < 1 \end{array} \right\}.$$

Theorem A follows immediately from (5.23) and the location of the aforementioned vertex; the exponent in the "small power of h" can be taken to be $\frac{9}{10}c\varepsilon$, where

$$c = \left\{ \begin{array}{ll} 1, & \text{if } M = 0 \\ 1 - s_1, & \text{otherwise} \end{array} \right\}.$$

□

6. Concluding remarks.

(I) Our approach to this problem (viz., uniform equidistribution) has been based[7] strictly on $Q(s)$, (4.7), and the inverse Mellin transform. Once the line of integration got shifted to $\sigma = \frac{1}{2} + \delta$ in (5.11), absolute values could be inserted, and some fairly standard analytic machinery then produced the desired conclusion. It is conceivable that $c(\Gamma)$ might be improved, at least when $M = 0$, by somehow trying to push things still further to the left, *i.e.* into $\{Re(s) < \frac{1}{2}\}$. One naturally thinks here of the functional equation for $F_n(z; s)$; cf. [4, p. 256(v)]. The problem, of course, will be poles of the Eisenstein series $E(z; s)$. (See (4.6) above, as well as [4, p. 250 (theorem 3.5)] and [24, p. 279]).

If things can be controlled well enough to get, say,

$$(*) \qquad O_\varepsilon(1)\left[y^{\frac{3}{4}-\varepsilon}\frac{\sqrt{K}}{b-a} + y^{\frac{1}{2}}\left(\log\frac{1}{y}\right)^N \frac{K^\varepsilon}{b-a}\right](\|\Delta f\|_2 + \|f\|_\infty)$$

in place of the final term in (5.23), the relation $c(\Gamma) = \frac{1}{2}$ would follow immediately (assuming, still, that $M = 0$). In this connection, note too that, since $b_n = O(1)$ in (5.21), the contribution stemming from any bounded set of n-values is automatically subsumed by the K^ε-portion of (*). One can therefore restrict attention to large $|n|$, if desired.

(II) As mentioned earlier, all of this reasoning extends to groups with several cusps. The case of a general multiplier system $(m, \mathcal{W})$ can also be considered — but, when $m \neq 0$, the calculations are more cumbersome (*i.e.* daunting). Familiarity with [4, chapter 9] becomes more-or-less essential. Cf., e.g., [4, p. 372 (5.34)(5.37)]. We remark that, in this setting, (4.3) becomes

$$\Delta_m P_{nj} + s(1-s)P_{nj} = -4\pi|n + \alpha_j|(s - mH_{n\alpha_j})P_{nj}(z; s+1).$$

Cf. [4, p. 701]; here $H_{n\alpha_j} = \frac{1}{2}sgn(n + \alpha_j)$. In passing to $\mathcal{B}_j(z; s)$, it is convenient to write $\mathcal{B}_j = \mathcal{B}'_j + \mathcal{B}''_j$ corresponding to

$$b'_n = b_n \frac{1 + sgn(n + \alpha_j)}{2} \quad \text{and} \quad b''_n = b_n \frac{1 - sgn(n + \alpha_j)}{2},$$

[7](apart from (5.16))

and then work separately with $\mathcal{B}'_j$ and $\mathcal{B}''_j$.

(III) A standard approximation argument [cf. near (2.1)] shows that the limit formula in theorem A actually holds for any $f \in C_b(\Gamma \setminus H)$. Likewise for any (sensible) piecewise continuous $f \in L^\infty$. In both cases, however, one deletes the subsequent remark about "the relevant difference".

(IV) In regard to the s_1-term in (5.23), it is interesting to consider a situation wherein — akin to [13] — $\Gamma \setminus H$ is variable and "is having its neck pinched along a certain dividing cycle". Let ℓ be the neck length. The ideas of [18] and [13, pp. 92 (theorem 7.2), 93 (line −10), 99 (lines −10 to −8)] show that $1 - s_1 \approx (\text{constant})\ell$ and that φ_1 corresponds to an Eisenstein pole. Cf. also [4, p. 736 (following c)].

As $\ell \to 0$, it intuitively becomes harder and harder for the closed horocycles associated with $i\infty$ to penetrate the chunk being pinched off. Cf., e.g., [13, p. 71 (4.6)].

At $\ell = 0$, the surface finally becomes noded with *two* components — one now being completely *in*accessible to any closed horocycle which originates on the other component.

The techniques used in arriving at (5.18), (5.22), (5.23) manifest a certain robustness as Γ is varied. Cf. [4, pp. 214 (note 30), 572 (7.11), 574 (7.12)] and [13, pp. 54, 59, 64, 86 (theorem 6.6), 97 (remark 7.5)]. One can presumably use Γ-*uniform* versions of (5.18)+(5.22) to formulate some kind of (asymptotically sharper) variant of theorem A in which the interplay between $h \to 0$ and the general pinching process (e.g. avoidance of "pinched off" regions) is more properly addressed.

We prefer to leave this matter for another occasion, however.

It is worth mentioning though that, in the present setting, where only one neck is being pinched, some preliminary information about frequency of entrance into the "receding chunk" can already be extracted from (5.16) by use of the techniques of [13, pp. 92-95, 98-100], especially pp. 93 (line −10) and 100 (7.11).

The situation for $[a, b]$ can naturally be expected to be more delicate.

(V) The author is grateful to Andreas Strömbergsson for a number of insightful comments on an earlier version of this work.

REFERENCES

[1] R. C. Baker, *Diophantine Inequalities*, Oxford Univ. Press, 1986.
[2] A. Beurling, *On functions with a spectral gap*, in The Collected Works of Arne Beurling, Vol. 2: Harmonic Analysis, Birkhäuser, 1989, pp. 370–372. (Notes from a 1942 Uppsala seminar.)
[3] A. Eskin and C. McMullen, *Mixing, counting, and equidistribution in Lie groups*, Duke Math. J., 71 (1993), pp. 181–209.
[4] D. A. Hejhal, *The Selberg Trace Formula for $PSL(2,\mathbb{R})$, vol. 2*, Lecture Notes in Mathematics 1001, Springer-Verlag, 1983.
[5] D. A. Hejhal, *Some observations concerning eigenvalues of the Laplacian and Dirichlet L-series*, in Recent Progress in Analytic Number Theory, vol. 2, H. Halberstam and C. Hooley, ed., Academic Press, 1981, pp. 95–110.
[6] D. A. Hejhal, *Eigenvalues of the Laplacian for $PSL(2,\mathbb{Z})$: some new results and computational techniques*, in International Symposium in Memory of Hua Loo-Keng, vol. 1, S. Gong, et al., ed., Science Press and Springer-Verlag, pp. 59–102. (Reprinted with [7].)
[7] D. A. Hejhal, *Eigenvalues of the Laplacian for Hecke Triangle Groups*, Memoirs Amer. Math. Soc. 469, 1992.

[8] D. A. HEJHAL, *On eigenvalues of the Laplacian for Hecke triangle groups*, in Zeta Functions in Geometry, N. Kurokawa and T. Sunada, ed., Adv. Studies in Pure Math. 21, 1992, pp. 359–408.

[9] D. A. HEJHAL AND S. ARNO, *On Fourier coefficients of Maass waveforms for* $PSL(2,\mathbb{Z})$, Math. of Comp., 61 (1993), pp. 245–267 and pp. S11–S16.

[10] D. A. HEJHAL AND B. RACKNER, *On the topography of Maass waveforms for* $PSL(2,\mathbb{Z})$, Exper. Math., 1 (1992), pp. 275–305.

[11] D. A. HEJHAL, *On value distribution properties of automorphic functions along closed horocycles*, in XVI-th Rolf Nevanlinna Colloquium, I. Laine and O. Martio, ed., deGruyter, 1996, pp. 39–52.

[12] D. A. HEJHAL, *On eigenfunctions of the Laplacian for Hecke triangle groups*, in Emerging Applications of Number Theory, D. Hejhal, et al., ed., IMA Vol. 109, Springer-Verlag, 1999, pp. 291–315.

[13] D. A. HEJHAL, *Regular b-Groups, Degenerating Riemann Surfaces, and Spectral Theory*, Memoirs Amer. Math. Soc. 437, 1990.

[14] H. IWANIEC, *Introduction to the Spectral Theory of Automorphic Forms*, Biblioteca de la Revista Mathemática Iberoamericana, Madrid, 1995.

[15] I. KRA, *Automorphic Forms and Kleinian Groups*, W. A. Benjamin, 1972.

[16] H. L. MONTGOMERY, *Ten Lectures on the Interface between Analytic Number Theory and Harmonic Analysis*, Amer. Math. Soc. CBMS Series 84, 1994.

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

[18] R. SCHOEN, S. WOLPERT, AND S.-T. YAU, *Geometric bounds on the low eigenvalues of a compact surface*, AMS Proc. Symp. Pure Math., 36 (1980), pp. 279–285.

[19] A. SELBERG, *Beweis eines Darstellungssatzes aus der Theorie der ganzen Modulformen*, Archiv för Math. og Naturvid., B44 (1941), pp. 33–44.

[20] A. SELBERG, *On the estimation of Fourier coefficients of modular forms*, AMS Proc. Symp. Pure Math., 8 (1965), pp. 1–15.

[21] A. SELBERG, *Lectures on sieves*, in Collected Papers, vol. 2, Springer-Verlag, 1991, pp. 65–247, especially §§20, 21.

[22] N. SHAH, *Limit distributions of expanding translates of certain orbits on homogeneous spaces*, Proc. Indian Acad. Sci. (Math. Sci.), 106 (1996), pp. 105–125. (Available electronically at: www.arXiv.org.)

[23] J. VAALER, *Some extremal functions in Fourier analysis*, Bull. Amer. Math. Soc., 12 (1985), pp. 183–216.

[24] D. ZAGIER, *Eisenstein series and the Riemann zeta function*, in Automorphic Forms, Representation Theory and Arithmetic, Tata Institute (Bombay), Springer-Verlag, 1981, pp. 275–301.