Automorphic coefficient sums and the quantum ergodicity question

Scott A. Wolpert*

April 5, 2000

1 Introduction

The arithmetic functions of elementary number theory have statistical distributions, [7]. In 1849 Dirichlet showed that the divisor function satisfies

$$\sum_{1 \leq n \leq x} d(n) = x \log x + (2\gamma - 1)x + O(x^{1/2})$$

for $\gamma$ Euler’s constant; more generally for $\sigma_\alpha(n) = \sum_{d|n} d^\alpha$ then for $\alpha$ positive, real

$$\sum_{1 \leq n \leq x} \sigma_\alpha(n) = \zeta(\alpha + 1) \frac{x^{\alpha + 1}}{\alpha + 1} + O(x^{\max\{1,\alpha\}}).$$

In 1915 Ramanujan [7] presented the formula

$$\sum_{1 \leq n \leq x} d(n)^2 \sim \frac{x}{\pi^2} (\log x)^3$$

which when combined with the formula of Ingham [4, 12] provides that the coefficient sum

$$S(t, \hat{x}) = \sum_{1 \leq n \leq t} d(n)e^{2\pi in\hat{x}}$$

satisfies

$$|S(t, \hat{x})|^2 \sim \frac{t}{\pi^2} (\log t)^3$$

*Research supported in part by NSF Grants DMS-9504176 and DMS-9800701.
as a positive measure in $\hat{x}$. The formula will serve as our paradigm for coefficient sums. The sums are associated with automorphic eigenfunctions. The multiplicative arithmetic function $\sigma_\alpha(|n|)$ occurs as the Fourier coefficients of the modular Eisenstein series

$$E(z; s) = \frac{1}{2} \sum_{(c,d)=1} \frac{y^s}{|cz+d|^{2s}}$$

for $\Re s > 1$ and $z = x + iy, \ y > 0$, [2]. The Eisenstein series provides a basic example of an automorphic (non-square integrable) eigenfunction for the Laplace-Beltrami operator associated to the upper half plane $\mathbb{H}$.

We are interested in the statistical properties of automorphic eigenfunctions, particularly of ensembles of Fourier coefficients. The statistics of a large-eigenvalue limit of eigenfunctions presents a model for the transition between quantum and classical mechanics, [3, 6, 8, 10, 15, 17, 18, 23, 24, 25]. The geodesic flow represents time evolution for the classical mechanical system; the flow is ergodic for quotients of hyperbolic space. The quantum ergodicity question is to understand the transition between quantum and classical mechanics in the presence of a classical ergodic flow, [1, 3, 9, 16, 17, 18, 23, 24].

We are intrigued by the transition mechanism on the upper half plane. The mechanism involves automorphic eigenfunctions, coefficient sums and geodesic flow. The correspondence principle provides that high-energy eigenfunctions of the hyperbolic Laplacian concentrate along geodesics. Egorov's Theorem provides that a high-energy eigenfunction on a quotient $\Gamma \backslash \mathbb{H}$ gives rise to an almost measure (a distribution) on the unit (co)tangent bundle of the quotient, that is almost geodesic flow invariant, [6, 18]. For $\Gamma$ a cofinite, non-cocompact, subgroup of $SL(2; \mathbb{R})$ a square integrable automorphic eigenfunction has a Fourier expansion

$$\phi(z) = \sum_{n \neq 0} a_n (y \sinh \pi r)^{1/2} K_{ir} (2\pi |n| y) e^{2\pi inx}$$

for $z = x + iy \in \mathbb{H}$, eigenvalue $-\lambda = -\left(\frac{1}{4} + r^2\right) < -\frac{1}{4}$ for the hyperbolic Laplacian $y^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)$ and $K_{ir}$ the Macdonald-Bessel function, [19, 21].

A theme of our investigations is that at high-energy the Egorov concentration measure on the space of geodesics is approximately $|S_\phi|^2$ for the coefficient sum

$$S_\phi(t, \hat{x}) = r^{-1/2} \sum_{1 \leq |n| \leq rt(2\pi)^{-1}} a_n e^{2\pi in\hat{x}}$$

for $(\hat{x}, t)$ certain elementary coordinates on the space of geodesics for $\mathbb{H}$, [22]. We describe applications of our results to coefficients sums.
2 The $SL(2; \mathbb{R})$ formalism

An element $B \in SL(2; \mathbb{R})$ has the unique Iwasawa decomposition

$$B = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$$

which provides for an equivalence of $SL(2; \mathbb{R}) = NAK$ with $S^*(\mathbb{H})^{1/2}$ the square root (double cover) of the unit cotangent bundle to the upper half plane by the rule

$$x + iy = y^{1/2}e^{i\theta}(ai + b), \quad y^{-1/2}e^{i\theta} = d - ic$$

for $z = x + iy \in \mathbb{H}$ and $\theta$ the argument for the root cotangent vector measured from the positive vertical, [14]. A symmetric $k$-tensor $f(z)dz^k$ on $\mathbb{H}$ is lifted to $SL(2; \mathbb{R})$ by first considering the balanced tensor $f(z)y^k dz^{k/2}d\bar{z}^{-k/2}$ (the hyperbolic metric is $ds = y^{-1/2}dz^{1/2}d\bar{z}^{1/2}$) and then associating the function $\tilde{f}(B) = \tilde{f}(z, \theta) = f(z)y^k e^{2ik\theta}$ on $SL(2; \mathbb{R})$. The complex exterior differential $\partial$ maps forms of type $dz^k$ to forms of type $d\bar{z}^k$; the product $\partial_{hyp} = y^2\partial$ maps forms of type $dz^k$ to forms of type $d\bar{z}^{k-1}$; $\partial_{hyp}$ commutes with the action of $SL(2; \mathbb{R})$ translation. A generalization of the setup is as follows. Functions or symmetric tensors on $\mathbb{H}$ lift to functions on $SL(2; \mathbb{R})$; $SL(2; \mathbb{R})$ acts on functions on $SL(2; \mathbb{R})$ by left translation, the Lie algebra $sl(2; \mathbb{R})$ (containing generalizations of $\partial_{hyp}$ and $\partial_{hyp}$) acts by right translation. The action of $H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, $V = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $W = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, $X = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $E^\pm = H \pm iV$ are basic to our considerations. The infinitesimal generator of geodesic flow is $H = \frac{1}{2}(E^+ + E^-)$; $W$ is the infinitesimal generator of $K$, the fiber rotations of $S^*(\mathbb{H})^{1/2}$. In terms of the coordinates $(x, y, \theta)$ for $SL(2; \mathbb{R})$ the operator $E^+$ is simply $E^+ = 4iye^{2i\theta}\frac{\partial}{\partial x} - ie^{2i\theta}\frac{\partial}{\partial \theta}$ ($E^+$ is the raising operator and is closely related to the derivative $\partial_{hyp}$). We will also consider the Casimir operator $C = E^-E^+ - W^2 - 2iW$ and the $SL(2; \mathbb{R})$-invariant volume form $d\mathcal{M} = y^{-2}dxdy\theta$ (Haar measure).

3 Helgason’s Fourier representation and Zelditch’s equation

Helgason’s representation theorem for eigenfunctions of the hyperbolic Laplacian is readily presented in terms of the Klein disc model $\mathbb{D}$ [11, 24]. For $z$ in the unit disc $\mathbb{D}$ and $b$ on the boundary $\mathbb{B}$ let $<z, b>$ denote the signed distance from the origin to the horocycle joining $z$ to $b$. The functions $e^{(2ir+1)<z, b>}$ give a complete set of generalized eigenfunctions for $L^2(\mathbb{D})$.
as \((r, b)\) ranges over \(\mathbb{R}^+ \times \mathbb{B}\). A smooth eigenfunction \(u\) of the hyperbolic Laplacian with eigenvalue \(-\left(\frac{1}{4} + r^2\right)\) is represented as

\[
u(z) = \int_{\mathbb{R}} e^{(2ir+1)\langle z, b \rangle} dT(b)
\]

for a distribution \(T \in \mathcal{D}(\mathbb{B})\). Zelditch observed \([24]\) that the integrand can be factored

\[
e^{(2ir+1)\langle z, b \rangle} dT = u^\infty(z, b)e^{2\langle z, b \rangle} db
\]

with the distribution \(u^\infty\) having special properties:

1. \(u^\infty\) is \(A\)-invariant if and only if \(u\) is \(A\)-invariant for \(A \in SL(2; \mathbb{R})\);
2. \(Hu^\infty = (2ir - 1)u^\infty\);
3. \(Xu^\infty = 0\);
4. \(u^\infty\) has the \(K\)-expansion \(u^\infty = \sum_m u_{2m}\) where \(Wu_{2m} = 2imu_{2m}, u_0 = u\) (the original eigenfunction) and \(E^\pm u_{2m} = (2ir \pm 2m + 1)u_{2m\pm 2}\).

The distribution \(u^\infty\) encodes the oscillation of \(u\). Modulo powers of the hyperbolic metric for \(m\) positive the term \(u_{2m}\) is simply the derivative \(\partial^m_{hyp} u\) with the conjugate derivative for \(m\) negative. Motivated by considerations of the calculus of pseudo-differential operators Zelditch introduced the following, \([24]\).

**Definition 1** For \(u, v\) eigenfunctions of the hyperbolic Laplacian on \(\mathbb{H}\) with eigenvalue \(-\left(\frac{1}{4} + r^2\right)\) set \(Q(u, v) = uv^\infty\), the microlocal lift of the pair \((u, v)\).

Zelditch discovered that \(Q\) satisfies a second-order differential equation \((H^2+4X^2+4irH)Q(u, v) = 0, [24]\). A proof based on the above properties is given in \([22]\) (the argument does not involve growth conditions on \(u\) or \(v\)).

**Lemma 2** \(Q(u, v)\) is a distribution on \(SL(2; \mathbb{R})\). The geodesic flow derivative \(HQ(u, v)\) has magnitude \(O(r^{-1})\).

In particular for \(\chi \in C_c(SL(2; \mathbb{R}))\) from Zelditch’s equation and integration by parts

\[
\int_{SL(2; \mathbb{R})} HQ(u, v)\chi d\mathcal{M} = \frac{i}{4r} \int_{SL(2; \mathbb{R})} Q(u, v)(H^2 + 4X^2)\chi d\mathcal{M}.
\]

For the right hand integrand \((H^2+4X^2)\chi\) is smooth and the contribution of \(Q(u, v)\) is bounded by \(\|u\|_2\|v\|_2\) for \(L^2\)-norms over a neighborhood of \(supp(\chi)\); the right hand side is \(O(r^{-1})\).
4 The Macdonald-Bessel functions and the geodesic-indicator measure

We introduce the microlocal lift of the normalized Macdonald-Bessel functions.

**Definition 3** For \( t = 2\pi nr^{-1}, \ n \in \mathbb{Z} \), set

\[
K(z, t) = (yr \sinh \pi r)^{1/2} K_{ir}(2\pi |n| y)e^{2\pi inx}
\]

and

\[
K_{\infty, \text{even}} = \sum_{m \ \text{even}} K(z, t)_{2m}
\]

and for \( \Delta t = 2\pi r^{-1} \), set

\[
Q(t) = \sum_{k \in \mathbb{Z}} K(z, t + k\Delta t)K_{\infty, \text{even}}(z, t).
\]

\( Q(t) \) is the microlocal lift of the Macdonald-Bessel function; \( Q \) encodes the oscillation of \( K \) by the sequence of all derivatives; \( Q \) satisfies the Zelditch differential equation.

**Lemma 4** \( Q(t) \) is an order-four tempered distribution on \( SL(2; \mathbb{R}) \). Given \( 0 < t_0 < t_1 \), \( Q(t) \) is uniformly bounded for \( t_0 \leq |t| \leq t_1 \) and \( r \) large.

Pre compactness plays a structural role in our considerations. The way is prepared to consider limits.

We first consider geodesic flow invariant measures. To each point of a complete geodesic on \( \mathbb{H} \) are associated the forward and backward directed (co)tangents. Each cotangent vector in turn has two square roots in \( S^*(\mathbb{H})^{1/2} \approx SL(2; \mathbb{R}) \). The association of the four root cotangent vectors to a point of a geodesic provides a lift of the geodesic to \( SL(2; \mathbb{R}) \) (the lift consists of four complete right action orbits of the subgroup \( A \) of the \( NAK \)-decomposition).

**Definition 5** For \( \tilde{\alpha}\beta \) a geodesic on \( \mathbb{H} \) let \( \Delta_{\tilde{\alpha}\beta} \) be the Dirac delta measure of flow-time (lifted arc-length) integration over the four root cotangent fields of \( \tilde{\alpha}\beta \).

A non-vertical geodesic on the upper half plane is a Euclidean circle. For \( y = (t^{-2} - (x - \hat{x})^2)^{1/2} \) with center \( (\hat{x}, 0) \) and radius \( t^{-1} \), then \( (\hat{x}, t) \) provides a coordinate for \( \mathbb{G} \) the space of non-vertical geodesics. The \( SL(2; \mathbb{R}) \) invariant area element on \( \mathbb{G} \), the full space of geodesics on \( \mathbb{H} \), is simply \( d\hat{x}dt \). We combine a sequence of integral identities and an induction argument on the \( K \)-weight using Zelditch’s equation to obtain the main result, [22]. Denote the group of integer translations by \( \Gamma_{\mathbb{Z}} = \{ (\frac{1}{n} \hat{y}) | n \in \mathbb{Z} \} \).
Theorem 6 Given $0 < t_0 < t_1$ for the geodesic $\bar{\alpha}\bar{\beta}$: $y = (t^{-2} - x^2)^{1/2}$ on $\mathbb{H}$

$$\lim_{r \to \infty} Q(t) d\mathcal{V} = \frac{\pi^2}{8} \sum_{\gamma \in \Gamma} \Delta_{\gamma(\bar{\alpha}\bar{\beta})}$$

in the sense of tempered distributions on $SL(2; \mathbb{R})$. The convergence is uniform for $t_0 \leq |t| \leq t_1$ as $r$ becomes large.

The Macdonald-Bessel functions with the proper scaling of parameters concentrate along a single geodesic. The result provides an instance of the correspondence principle independent of the calculus of pseudo-differential operators. The uniform convergence will be used to consider sums of Macdonald-Bessel functions.

5 Applications

We consider four applications of the results to automorphic eigenfunctions. Let $\Gamma$ be a cofinite non-cocompact subgroup of $SL(2; \mathbb{R})$ containing $\Gamma \mathbb{Z}$ as the stabilizer of infinity.

5.1 General equivalences

For automorphic eigenfunctions $\phi$ the coefficient sums (3) play a basic role. We begin with a coefficient summation scheme for studying quantities quadratic in the eigenfunction. The scheme provides a positive measure. For $(\hat{x}, t)$ in $\mathbb{R} \times \mathbb{R}_+$ introduce the measure (a tempered distribution)

$$\Omega_{\phi,N} = d_t \mathcal{F}_N * |S_\phi(t, \hat{x})|^2$$

for $d_t$ the Lebesgue-Stieljes derivative in $t$ and for convolution in $\hat{x}$ with the Fejér kernel $\mathcal{F}_N$. The tempered distribution $\Omega_{\phi,N}$ is bounded by $\|\phi\|_2^2$. In fact for $\{\phi_j\}$ a sequence of unit-norm automorphic eigenfunctions the sequences $\{Q(\phi_j, \phi_j)\}$ and $\{\Omega_{\phi_j,N}\}$ are precompact. We can consider a convergent sequence $\{\phi_j\}$ and write $Q_{\text{limit}} = \lim_j Q(\phi_j, \phi_j) d\mathcal{V}$ and $\mu_{\text{limit}} = \lim_n \lim_j \Omega_{\phi_j,N}$. For $(\hat{x}, t)$ the above described coordinates the distributions $\Omega_{\phi,N}$ and $\mu_{\text{limit}}$ are given on $\mathbb{G}$ the space of non-vertical geodesics. An application of our formula is the following relationship

$$Q_{\text{limit}} = \frac{\pi^2}{8} \int_G \Delta_{\alpha\beta} \mu_{\text{limit}}$$

in the sense of tempered distributions on $\Gamma \backslash SL(2; \mathbb{R})$, [22]. A consequence of the construction of $\Omega_{\phi,N}$ is that $\mu_{\text{limit}}$ is positive; $\Delta_{\alpha\beta}$ is positive and it follows that $Q_{\text{limit}}$ is a positive measure. $Q_{\text{limit}}$ is $\Gamma$-invariant; it follows that the extension (by vertical geodesics forming a null set) of
\(\mu_{\text{limit}}\) to \(\hat{G}\) is \(\Gamma\)-invariant. The action of \(\Gamma\) on \(\hat{G}\) is ergodic relative to the \(SL(2; \mathbb{R})\)-invariant measure; the \(\Gamma\)-invariance of \(\mu_{\text{limit}}\) is a strong condition on the limits of coefficient sums. The equality \(Q_{\text{limit}} = d\mathfrak{W}\) is equivalent to the weak* convergence of \(|S_{\phi_j}(t, \hat{x})|^2\) to \(4\pi^{-2} t\). The quantum unique ergodicity conjecture poses that every \(Q_{\text{limit}}\) is indeed a constant multiple of \(d\mathfrak{W}\) \([17]\); in particular that high-energy microlocal approximate the uniform density.

Good has a related result on coefficient sums for a fixed automorphic form, \([5]\). For Ramanujan’s function \(\tau(n)\) defined by the weight 12 modular form

\[
\Delta = (2\pi)^{12} q \prod_{n=1}^\infty (1 - q^n)^{24} = (2\pi)^{12} \sum_{n=1}^\infty \tau(n) q^n, \quad q = e^{2\pi i z}
\]

Good considered the coefficient sums

\[
S(t, \hat{x}) = \sum_{1 \leq n \leq t} \tau(n) e^{2\pi in \hat{x}}
\]

and found an explicit form of weak* convergence involving the Petersson inner product

\[
\int_{\hat{x}_1}^{\hat{x}_2} |S(t, \hat{x})|^2 d\hat{x} = \frac{3}{12! \pi^{13}} < \Delta, \Delta >_p (\hat{x}_2 - \hat{x}_1)t^{12} + O(t^{12-1/3+\epsilon}).
\]

The form is fixed and so the limit in the sum length \(t\) is the analog of the high-energy limit. The limit of \(|S|^2\) is the uniform density in \(\hat{x}\).

### 5.2 The modular Eisenstein series

Luo and Sarnak followed by Jakobson considered the high-energy limit of the analytic continuation of the modular Eisenstein series on the spectral line, \([13, 15]\). Even though the Eisenstein series is not square integrable the analysis can be effected. From the Maass-Selberg relation the microlocal lift \(Q_E(r) = E(z; \frac{1}{2} + ir) E(z; \frac{1}{2} + ir)^{\infty, \text{even}}\) has magnitude \(\log |r|\). Luo-Sarnak and Jakobson used the hard-analysis estimates available for \(L\)-functions and for Kloosterman (exponential) sums to obtain their results. The authors found that \((\log |r|)^{-1} Q_E(r)\) weak* converges to \(48\pi^{-1} d\mathfrak{W}\). For limits \(Q_{E,\text{limit}} = \lim_j (\log |r_j|)^{-1} Q_E(r_j) d\mathfrak{W}\) and \(\mu_{E,\text{limit}} = \lim_N \lim_j (\log |r_j|)^{-1} d\mathfrak{W} * |S_E(t, \hat{x})|^2\) we find the relationship

\[
Q_{E,\text{limit}}^\text{symm} = \frac{\pi^2}{8} \int_G \Delta_{\alpha\beta}^\text{symm} \mu_{E,\text{limit}}.
\]

We then find that \((\log |r|)^{-1} Q_E(r)\) converging to \(48\pi^{-1} d\mathfrak{W}\) is equivalent to

\[
(|\zeta(1 + 2ir)|^2 |r| \log |r|)^{-1} \sum_{1 \leq n \leq rt} \sigma_{2ir}(n) n^{-ir} e^{2\pi in \hat{x}} |^2 \text{ converging to } 48\pi^{-2} t
\]
weak* in $\hat{x}$ for each positive $t$. The normalization of the sum by the Riemann zeta function is significant since $|\zeta(1+2ir)|$ is known to at least vary between $(\log \log |r|)^{-1}$ and $\log \log |r|$, [20]. The formula can be compared to the Ramanujan and Ingham formulas. The convergence is also suggestive of Good’s formula and of the residue formula at $s = 1$ for the Ramanujan identity [20]

$$
\sum_{n=1}^{\infty} \frac{|\sigma_{2ir}(n)|^2}{n^s} = \frac{\zeta^2(s)\zeta(s + 2ir)\zeta(s - 2ir)}{\zeta(2s)},
$$

5.3 The spectral average of modular eigenfunctions

Zelditch considers for $\Gamma$ a cofinite, non-cocompact subgroup of $SL(2; \mathbb{R})$ with orthonormal basis for $L^2(\Gamma \backslash \mathbb{H})$-eigenfunctions $\{(\phi_j, \lambda_j)\}$ and a basis of Eisenstein series $\{E_k\}$ the joint spectral average

$$
\sigma_T = \sum_{0 \leq r_j \leq T} Q(\phi_j, \phi_j) + \frac{1}{4\pi} \sum_k \int_{-T}^{T} Q_{E_k}(r) dr [25].
$$

From the Selberg-Weyl law the spectral contribution in the interval $[−T,T]$ is given as $(4\pi)^{-1} Area(\Gamma \backslash \mathbb{H}) T^2$. Zelditch shows that

$$
\sigma_T \sim T^2
$$

in the sense of distributions [25, Theorem 5.1]. For the case of congruence subgroups the spectral contribution of the Eisenstein series has a smaller order of magnitude and thus effectively $\sigma_T$ is given by the sum $\sum_{r_j \leq T} Q(\phi_j, \phi_j)$. It follows from Zelditch’s result and our considerations that for congruence subgroups the spectral average of the coefficient sums

$$
T^{-2} \sum_{0 \leq r_j \leq T} |S_{\phi_j}(t, \hat{x})|^2 converges to \frac{4t}{\pi^2}
$$

weak* in $\hat{x}$ for each positive $t$. The spectral average of the coefficient sums is the uniform density in $\hat{x}$.

5.4 Renormalization of semi-classical limits

It is an open question if high-energy limits are necessarily non trivial; in the absence of unique quantum ergodicity all mass could in the limit escape into the cusps resulting in $Q_{\text{limit}}$ and $\mu_{\text{limit}}$ being trivial. To investigate this possibility we renormalize the eigenfunctions to unit $L^2$-mass on a compact set and consider the limit of corresponding microlocal lifts [22, Chapter 5]. The first matter is to compare normalizations: we show that the resulting $L^2$-norms are bounded
by $\log \lambda$. The bound is used to establish pre compactness for a sequence of renormalized microlocal lifts and that a high-energy limit has the expected basic properties. The limit is necessarily non trivial. The limit of square coefficient sums is found to be the zeroth Fourier-Stieltjes coefficient of a $\Gamma$-invariant measure on the space of geodesics. The action of $\Gamma$ on the space of geodesics is noted to have a compact fundamental set. A lower bound for square coefficient sums results. Given $\Gamma$ there exist positive constants $t_0 < t_1$ such that for large eigenvalues the mapping $\phi \mapsto (S_\phi(t_1, \theta) - S_\phi(t_0, \theta))$ from eigenfunctions to linear coefficient sums twisted by an additive character is a uniform quasi-isometry relative to the $L^2$-norms for a suitable compact set and the unit circle. In particular the mapping from eigenfunctions to twisted linear coefficient sums $S_\phi(t_1, \theta)$ is an injection.

References


