QUANTUM UNIQUE ERGODICITY AND NUMBER THEORY

KANNAN SOUNDARARAJAN

1. Introduction

In this course I will describe recent progress on the "Quantum Unique Ergodicity" conjecture of Rudnick and Sarnak in a special arithmetic situation. To explain what this conjecture is about, let $\mathbb{H}$ denote the upper half plane $\{x+iy : y>0\}$. The group $SL_2(\mathbb{R})$ acts on $\mathbb{H}$ by Möbius transformations, and let $\Gamma = SL_2(\mathbb{Z})$. In number theory it is of great interest to study functions on $\mathbb{H}$ that are either invariant under the action of Γ, or transform in some nice way under this action. The classical theory of modular forms of weight k (an even positive integer) considers holomorphic functions f satisfying

$$f(\gamma z) = (cz+d)^k f(z)$$

for all $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z})$. If we also require f to be holomorphic and decay rapidly "at the cusp at infinity" then we get the theory of cusp forms, and the most famous example of this is Ramanujan's Δ-function. In the 1940's and 50's Maass and Selberg developed a nice theory of functions satisfying $f(\gamma z) = f(z)$ for all $\gamma \in SL_2(\mathbb{Z})$. These functions are no longer holomorphic but are real analytic eigenfunctions of the Laplace operator $\Delta = -y^2(\frac{d^2}{dx^2} + \frac{d^2}{dy^2})$. If we also require that these eigen-functions should decay rapidly at ∞, then we get the theory of *Maass* cusp-forms; even their existence is not easy to demonstrate, and was first established by Selberg using his trace formula.

Let ϕ denote such a Maass cusp form, and let λ denote its Laplace eigenvalue, and let ϕ be normalized so that $\int_X |\phi(z)|^2 \frac{dx\ dy}{y^2} = 1$ (where $X = SL_2(\mathbb{Z})\backslash\mathbb{H}$). From work of Zelditch [61] it follows that as $\lambda \to \infty$, for a *typical* Maass form ϕ the measure $\mu_\phi := |\phi(z)|^2 \frac{dx\ dy}{y^2}$ approaches the uniform distribution measure $\frac{3}{\pi}\frac{dx\ dy}{y^2}$. That is, typically the L^2-mass is uniformly spread out over a fundamental domain for $\Gamma\backslash\mathbb{H}$. This statement is referred to as "Quantum Ergodicity." Rudnick and Sarnak [49] have conjectured that an even stronger result holds. Namely, that as

$\lambda \to \infty$, for *every* Maass form ϕ the measure μ_ϕ approaches the uniform distribution measure. This is a special case of their "Quantum Unique Ergodicity" conjecture. Lindenstrauss [36] made great progress towards this conjecture, showing that, for Maass cusp forms that are eigenfunctions of the Laplacian and all the Hecke operators,[1] the only possible limiting measures are of the form $\frac{3}{\pi}c\frac{dx\ dy}{y^2}$ with $0 \le c \le 1$. Recently I showed [56] that $c = 1$, completing the proof of the QUE conjecture for Hecke-Maass forms on $SL_2(\mathbb{Z})\backslash\mathbb{H}$. Lindenstrauss's work is based on ergodic theory and measure rigidity, and these will be explained in Einsiedler's lectures. In fact Lindenstrauss's work starts with a micro-local lift of ϕ to a function on $SL_2(\mathbb{Z})\backslash SL_2(\mathbb{R})$ and he demonstrates the equi-distribution on this larger space (except for escape of mass, which again is ruled out by my work).

In this course, I will explain the proof (due to Holowinsky and me [29]) of the analog of the quantum unique ergodicity conjecture for classical holomorphic modular forms. Let f be a holomorphic modular cusp form of weight k (an even integer) for $SL_2(\mathbb{Z})$. Associated to f we have the measure

$$\mu_f := y^k|f(z)|^2\frac{dx\ dy}{y^2},$$

which is invariant under the action of $SL_2(\mathbb{Z})$, and we suppose that f has been normalized so that

$$\int_X y^k|f(z)|^2\frac{dx\ dy}{y^2} = 1.$$

The space $S_k(SL_2(\mathbb{Z}))$ of cusp forms of weight k for $SL_2(\mathbb{Z})$ is a vector space of dimension about $k/12$, and contains elements such as $\Delta(z)^{k/12}$ (if $12|k$, and where Δ is Ramanujan's cusp form) for which the measure will not tend to uniform distribution. Therefore one restricts attention to a particularly nice set of cusp forms, namely those that are eigenfunctions of all the Hecke operators. The Rudnick-Sarnak conjecture in this context states that as $k \to \infty$, for *every* Hecke eigencuspform f the measure μ_f tends to the uniform distribution measure. For simplicity, we have restricted ourselves to the full modular group, but the conjecture could be formulated just as well for holomorphic newforms of level N. Luo and Sarnak [39] have shown that equidistribution holds for most Hecke eigenforms, and Sarnak [51] has shown that it holds in

[1]The spectrum of the Laplacian is expected to be simple, so that any eigenfunction of the Laplacian would automatically be an eigenfunction of all Hecke operators. This is far from being proved.

the special case of dihedral forms. It does not seem clear how to extend Lindenstrauss's work to the holomorphic setting.[2]

The proof of this holomorphic QUE combines two different approaches developed independently by Holowinsky [28] and myself [57]. At their heart, both approaches rely on an understanding of mean-values of multiplicative functions, and I will explain some of the key results in that area. Either of these approaches is capable of showing that there are very few possible exceptions to the conjecture, and under reasonable hypotheses either approach would show that there are no exceptions. However, it seems difficult to show unconditionally that there are no exceptions using just one of these approaches. Fortunately, as we shall explain below, the two approaches are complementary, and the few rare cases that are untreated by one method fall easily to the other method. Both approaches use in an essential way that the Hecke eigenvalues of a holomorphic eigencuspform satisfy the Ramanujan conjecture (Deligne's theorem). The Ramanujan conjecture remains open for Maass forms, and this is the (only) barrier to using our methods in the non-holomorphic setting.[3]

We end this quick introduction by explaining where the name "quantum unique ergodicity" comes from. The classical dynamics of geodesics on $SL_2(\mathbb{Z})\backslash\mathbb{H}$ is known to be chaotic: two nearby geodesics deviate from each other rapidly, and a generic geodesic will fill the region $SL_2(\mathbb{Z})\backslash\mathbb{H}$ uniformly. Thus a (generic) classical particle moving on this surface will have a complicated trajectory, and will be equally likely to be in any part of the surface. Consider now a quantum analog of this situation. Here one is interested in wave-functions $\Psi(z,t)$ which evolve according to Schrödinger's equation $i\frac{d}{dt}\Psi(z,t) = \Delta\Psi(z,t)$ (where Δ denotes the hyperbolic Laplacian in the z-variable) . One is especially interested in the time independent (or standing wave) solutions where $|\Psi(z,t)|^2$ is independent of the time t. An important class of such solutions are of the form $\Psi(z,t) = \Psi(z)e^{-it\lambda}$, and then we see that $\Psi(z)$ is an eigenfunction of the Laplacian with eigenvalue λ. The physical interpretation of λ is that it corresponds to the energy. The QUE conjecture thus states that for large energy the probability of finding the quantum particle in any region depends only on the area of that region. So the conjecture for Maass forms has a nice physical interpretation.

[2]The difficulty from the ergodic point of view concerns the invariance under the geodesic flow of the quantum limits of the micro-local lifts associated to holomorphic forms.

[3]Assuming the Ramanujan conjecture for Maass forms, our methods would obtain the stronger micro-local version of QUE. Moreover our methods would then be able to quantify the rate at which equi-distribution is attained.

The analogous conjecture for holomorphic forms doesn't have such a nice physical interpretation, but it does imply a striking Corollary. A cusp form of weight k has about $k/12$ zeros inside a fundamental domain. How are these zeros distributed? If we take a large power of Ramanujan's Δ function, then there is only one zero of multiplicity $k/12$ at the cusp at ∞. However, if the L^2-mass of f is equidistributed on the fundamental domain, then Rudnick [48] showed that the zeros are also equidistributed (with the measure $\frac{3}{\pi}\frac{dx\ dy}{y^2}$). In particular, since we know that the mass of Hecke eigenforms is equidistributed, we deduce that so are the zeros of a Hecke eigenform.

In these lectures we have restricted attention just to the case of $SL_2(\mathbb{Z})$. The techniques developed here also extend to congruence subgroups of $SL_2(\mathbb{Z})$. There are other arithmetic groups $\tilde{\Gamma}$, arising from quaternion algebras, which have a compact quotient $\tilde{\Gamma}\backslash\mathbb{H}$ for which QUE is known for Maass forms (from the work of Lindenstrauss), but the holomorphic analog remains open. The same is true for Hecke eigenforms on the sphere; see [4]. More generally, the original Rudnick-Sarnak conjecture was formulated for compact Riemannian surfaces of (strictly) negative curvature. In this generality, the problem remains wide open, although recently Anantharaman [1] made important partial progress. One can also consider QUE problems for billiards in a stadium (or other domains), and here the answer can be negative: see the recent work of Hassell [22].

Suggested reading The books by Iwaniec [30] and [31], and Iwaniec and Kowalski [32] give good introductions to the analytic theory of modular and Maass forms. The article [37] and Einsiedler's lecture notes give accounts of the ergodic theoretic approach to QUE. The articles by Sarnak [52] and [53] give motivated accounts of the QUE conjecture and progress towards it. Finally, for more on *quantum chaos* consult [42] and [3].

2. Preliminaries

Recall that for two smooth bounded functions g_1 and g_2 on $X = SL_2(\mathbb{Z})\backslash\mathbb{H}$ we may define the Petersson inner product

$$\langle g_1, g_2\rangle = \int_X g_1(z)\overline{g_2(z)}\frac{dx\ dy}{y^2}.$$

In this definition we could allow for one of the g_1 or g_2 to be unbounded, so long as the other function decays appropriately for the integral to converge.

If f is a modular cusp form of weight k then we shall let $F_k(z)$ denote $y^{k/2}f(z)$ where $z = x + iy$. Note that F_k is not a function on X, but $|F_k|$ is. Therefore, we can talk sensibly of $\langle F_k, F_k\rangle$ although this is an abuse of notation.

The space of modular forms of weight k comes equipped with a large family of commuting operators which are self-adjoint with respect to the Petersson inner product. These are the Hecke operators, which are given explicitly by (note our normalization which may be a little different from other sources)

$$(T_n f)(z) = \frac{1}{n^{\frac{k+1}{2}}} \sum_{ad=n} a^k \sum_{b \pmod d} f\Big(\frac{az+b}{d}\Big).$$

As noted above, the Hecke operators commute and in fact

$$T_m T_n = \sum_{d|(m,n)} T_{\frac{mn}{d^2}}.$$

Since the Hecke operators are also self-adjoint we may simultaneously diagonalize the space of cusp forms with respect to all Hecke operators, and we call such forms Hecke eigenforms (or Hecke eigencuspforms).

A Hecke eigencuspform f of weight k has a Fourier expansion of the shape

$$f(z) = C \sum_{n=1}^{\infty} \lambda_f(n) n^{\frac{k-1}{2}} e(nz),$$

where $e(z) = e^{2\pi i z}$, and $\lambda_f(n)$ denote the Hecke eigenvalues which have been normalized so that Deligne's theorem (formerly the Ramanujan conjecture) reads $|\lambda_f(n)| \leq d(n)$, the number of divisors of n. The quantity $C = C(f)$ above represents a constant, and we may normalize f (by choosing C) so that the Petersson norm $\|f\|^2 = \langle F_k, F_k\rangle = 1$.

For a prime number p we may write the Hecke eigenvalue $\lambda_f(p)$ as $\alpha_p + \beta_p$ where $\alpha_p\beta_p = 1$ and $|\alpha_p| = |\beta_p| = 1$. The L-function associated to f is then

$$L(s,f) = \sum_{n=1}^{\infty} \frac{\lambda_f(n)}{n^s} = \prod_p \Big(1 - \frac{\alpha_p}{p^s}\Big)^{-1} \Big(1 - \frac{\beta_p}{p^s}\Big)^{-1},$$

where the series and product above are absolutely convergent in $\sigma > 1$, and $L(s,f)$ extends analytically to $\mathbb{C}$ with a functional equation connecting the values at s and $1-s$. In our work an important quantity is the related symmetric square L-function.

The symmetric square L-function is

$$L(s, \mathrm{sym}^2 f) = \sum_{n=1}^{\infty} \frac{\lambda_f^{(2)}(n)}{n^s} = \prod_p \Big(1 - \frac{\alpha_p^2}{p^s}\Big)^{-1} \Big(1 - \frac{1}{p^s}\Big)^{-1} \Big(1 - \frac{\beta_p^2}{p^s}\Big)^{-1}.$$

The series and product above converge absolutely in $\mathrm{Re}(s) > 1$, and by the work of Shimura [54], we know that $L(s, \mathrm{sym}^2 f)$ extends analytically to the entire complex plane, and satisfies the functional equation

$$\begin{aligned} \Lambda(s, \mathrm{sym}^2 f) &= \Gamma_{\mathbb{R}}(s+1)\Gamma_{\mathbb{R}}(s+k-1)\Gamma_{\mathbb{R}}(s+k) L(s, \mathrm{sym}^2 f) \\ &= \Lambda(1-s, \mathrm{sym}^2 f), \end{aligned}$$

where $\Gamma_{\mathbb{R}}(s) = \pi^{-s/2}\Gamma(s/2)$.

The symmetric square L-function appears naturally in the normalization of f to have L^2-norm 1. Precisely, the constant $C = C(f)$ appearing in the Fourier expansion of f is given by

$$|C|^2 = \frac{(4\pi)^{k-1}}{\Gamma(k)} \frac{2\pi^2}{L(1, \mathrm{sym}^2 f)}.$$

In our work we shall require some understanding of this value $L(1, \mathrm{sym}^2 f)$; we require specifically a good lower bound for this quantity. This is known due to the work of many authors: Shimura [54], Gelbart and Jacquet [12], Hoffstein and Lockhart [26], and Goldfeld, Hoffstein and Lieman [13]. Gelbart and Jacquet [12] have shown that $L(s, \mathrm{sym}^2 f)$ arises as the L-function of a cuspidal automorphic representation of $GL(3)$. Therefore, invoking the Rankin-Selberg convolution for $\mathrm{sym}^2 f$, one can establish a classical zero-free region for $L(s, \mathrm{sym}^2 f)$. For example, from Theorem 5.42 (or Theorem 5.44) of Iwaniec and Kowalski [32] one obtains that for some constant $c > 0$ the region

$$\mathcal{R} = \Big\{ s = \sigma + it : \quad \sigma \geq 1 - \frac{c}{\log k(1+|t|)} \Big\}$$

does not contain any zeros of $L(s, \mathrm{sym}^2 f)$ except possibly for a simple real zero. The work of Hoffstein and Lockhart [26] (see the appendix by Goldfeld, Hoffstein and Lieman [13]) shows that $c > 0$ may be chosen so that there is no real zero in our region $\mathcal{R}$. Thus $L(s, \mathrm{sym}^2 f)$ has no zeros in $\mathcal{R}$. Moreover the work of Goldfeld, Hoffstein, and Lieman [13] shows that

$$L(1, \mathrm{sym}^2 f) \gg \frac{1}{\log k}.$$

To be precise, the work of Goldfeld, Hoffstein, and Lieman considers symmetric square L-functions of Maass forms in the eigenvalue aspect, but our case is entirely analogous, and follows upon making minor modifications to their argument.

Suggested reading. For general information on modular forms we again refer to Iwaniec's books [30] and [31]. For more on the symmetric square L-function and its zero-free region, in addition to [13] and [26] you should look at the paper by Hoffstein and Ramakrishnan [27], and also [38] and [6].

3. Spectral expansions, and expansions into incomplete Eisenstein and Poincare series

Let h denote a smooth bounded function on X. Considering h as fixed, and letting $k \to \infty$, the Rudnick-Sarnak conjecture asserts that for every Hecke eigencuspform f of weight k we have

$$\langle hF_k, F_k\rangle \to \frac{3}{\pi}\langle h, 1\rangle, \tag{1}$$

with the rate of convergence above depending on the function h.

To attack the conjecture (1), it is convenient to decompose the function h in terms of a basis of smooth functions on X. There are two natural ways of doing this, and both decompositions play important roles in the proof of the Rudnick-Sarnak conjecture.

First we could use the spectral decomposition of a smooth function on X in terms of eigenfunctions of the Laplacian. The spectral expansion will involve (i) the constant function $\sqrt{3/\pi}$, (ii) Maass cusp forms ϕ that are also eigenfunctions of all the Hecke operators, and (iii) Eisenstein series on the $\frac{1}{2}$ line. Recall that the Eisenstein series is defined for $\mathrm{Re}(s) > 1$ by

$$E(z, s) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} \mathrm{Im}(\gamma z)^s,$$

where $\Gamma = SL_2(\mathbb{Z})$ and Γ_∞ denotes the stabilizer group of the cusp at infinity (namely the set of all translations by integers). The Eisenstein series $E(z, s)$ admits a meromorphic continuation, with a simple pole at $s = 1$, and is analytic for s on the line $\mathrm{Re}(s) = \frac{1}{2}$. For more on the spectral expansion see Iwaniec's book [30]. Note that (1) is trivial when h is the constant eigenfunction. To establish (1) using the spectral decomposition, we would need to show that for a fixed Maass eigencuspform ϕ, and for a fixed real number t that

$$\langle \phi F_k, F_k\rangle, \qquad \text{and} \qquad \langle E(\cdot, \tfrac{1}{2} + it)F_k, F_k\rangle \to 0,$$

as $k \to \infty$. The above statement should be thought of as an analog of Weyl's equidistribution criterion. The inner products above may be related to values of L-function, and we shall discuss this connection in the next section.

Alternatively, one could expand the function h in terms of incomplete Poincare and Eisenstein series. Let ψ denote a smooth function, compactly supported in $(0,\infty)$. For an integer m the incomplete Poincare series is defined by

$$P_m(z \mid \psi) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} e(m\gamma z)\psi(\mathrm{Im}(\gamma z)).$$

In the special case $m = 0$ we obtain incomplete Eisenstein series $E(z \mid \psi) = P_0(z \mid \psi)$. For an account on approximating a smooth function h using incomplete Poincare series see the paper of Luo and Sarnak [40]; essentially it amounts to taking a Fourier expansion of $h(x+iy)$ for each fixed value of y. Now conjecture (1) can be reformulated (again analogously to Weyl's equidistribution criterion) as saying that

$$\langle F_k, F_k P_m(\cdot \mid \psi)\rangle \to 0,$$

for $m \neq 0$ (considered to be fixed), and any given smooth function ψ. In the case $m = 0$ we want that

$$\langle F_k, F_k E(\cdot \mid \psi)\rangle \to \frac{3}{\pi}\langle 1, E(\cdot, \psi)\rangle,$$

for any fixed ψ and as $k \to \infty$. The Rankin-Selberg unfolding method can be used to handle these inner products. For example the inner product with Poincare series (for $m \neq 0$) was related by Luo and Sarnak [39] to the problem of estimating the shifted convolution sums (for m fixed, and as $k \to \infty$)

$$\sum_{n \asymp k} \lambda_f(n)\lambda_f(n+m),$$

where the sum is over n of size k, and $\lambda_f(n)$. We will discuss the inner products with these incomplete Poincare and Eisenstein series in more detail in section 5.

Suggested reading The spectral expansion of nice functions on X is discussed nicely (and in greater generality) in Iwaniec's book [30]. The idea of approximating h by incomplete Poincare series amounts for each fixed y to taking a Fourier expansion in x of $h(x+iy)$. For details see Luo and Sarnak's paper [40].

4. Relation to L-functions and the subconvexity problem

In the approach to the Rudnick-Sarnak conjecture via a spectral expansion, we need to estimate $\langle F_k, F_k E(\cdot, \frac{1}{2} + it)\rangle$ for fixed t and as $k \to \infty$, and also $\langle F_k, F_k \phi\rangle$ where ϕ is a fixed Hecke-Maass cusp form, and $k \to \infty$. Both of these inner products are linked to L-functions.

In the case of Eisenstein series this is the classical work of Rankin and Selberg. The unfolding method (starting with $E(z,s)$ in the domain of absolute convergence, and extending to $s = 1/2 + it$ by analytic continuation) leads to

$$|\langle E(\cdot, \tfrac{1}{2}+it)F_k, F_k\rangle| = \Big|\pi^{\frac{3}{2}} \frac{\zeta(\frac{1}{2}+it)L(\frac{1}{2}+it, \mathrm{sym}^2 f)}{\zeta(1+2it)L(1,\mathrm{sym}^2 f)} \frac{\Gamma(k-\frac{1}{2}+it)}{\Gamma(k)}\Big|.$$

Since $|\Gamma(k-\frac{1}{2}+it)| \leq \Gamma(k-\frac{1}{2})$, $|\zeta(\frac{1}{2}+it)| \ll (1+|t|)^{\frac{1}{4}}$, and $|\zeta(1+2it)| \gg 1/\log(1+|t|)$, using Stirling's formula it follows that

$$|\langle E(\cdot, \tfrac{1}{2}+it)F_k, F_k\rangle| \ll \frac{(1+|t|)^2 |L(\frac{1}{2}+it, \mathrm{sym}^2 f)|}{k^{\frac{1}{2}} L(1, \mathrm{sym}^2 f)}.$$

As noted in §2, the term $L(1, \mathrm{sym}^2 f)$ is $\gg 1/(\log k)$, and hence the inner product with Eisenstein series tends to zero provided we can establish an upper bound for $|L(\frac{1}{2}+it, \mathrm{sym}^2 f)|$ which is better than $k^{\frac{1}{2}}/\log k$.

The problem of bounding L-functions on the critical line has a long history, going back to work of Weyl, Hardy and Littlewood in the case of the Riemann zeta-function. In general one has a bound for L-functions of the form $\ll C^{\frac{1}{4}}$, where C is an object called the *analytic conductor* (defined below in §6) which measures the complexity of the L-function. Such a bound is called the convexity bound; usually the convexity bound is stated as $\ll C^{\frac{1}{4}+\epsilon}$, and the refined bound we have stated is a recent observation of Heath-Brown. For example, for the zeta-function the convexity bound states that $|\zeta(\frac{1}{2}+it)| \ll |t|^{\frac{1}{4}}$ and the work of Weyl-Hardy-Littlewood furnished improvements over this, leading for example to $|\zeta(\frac{1}{2}+it)| \ll |t|^{\frac{1}{6}}$. Here the truth is expected to be the Lindelöf bound $|\zeta(\frac{1}{2}+it)| \ll |t|^{\epsilon}$, and this bound is a consequence of the Riemann hypothesis. The problem of obtaining a bound for L-values of the shape $C^{\frac{1}{4}-\delta}$ for some $\delta > 0$ is known as the *subconvexity problem*, and is an important outstanding problem in number theory. The subconvexity problem is now resolved for L-functions arising from $GL(1)$ or $GL(2)$, and a handful of other cases, but in general the problem is wide open. One of the most striking applications of subconvexity is to the problem of representing integers by ternary quadratic forms (see [8]).

Returning to our case, we need a bound for $|L(\frac{1}{2}+it, \mathrm{sym}^2 f)|$. The analytic conductor for this L-function is about $(1+|t|)^3 k^2$, and so the convexity bound gives $|L(\frac{1}{2}+it, \mathrm{sym}^2 f)| \ll k^{\frac{1}{2}}(1+|t|)^{\frac{3}{4}}$. Using this in our inner product with Eisenstein series, we realize that this is barely

insufficient to show that decay of this inner product, and any subconvexity bound would be sufficient. The Generalized Riemann Hypothesis implies such a bound (in fact that the L-value is $\ll k^{\epsilon}(1+|t|)^{\epsilon}$), but unconditionally subconvexity for symmetric square L-functions is not known. Recently, X. Li [35] obtained a subconvexity bound for k fixed and $t \to \infty$, but for our application we want the opposite case of t fixed and $k \to \infty$. In a general context I established a *weak subconvexity* bound (described in §6) which shows that

$$|L(\tfrac{1}{2}+it, \mathrm{sym}^2 f)| \ll \frac{k^{\frac{1}{2}}(1+|t|)^{\frac{3}{4}}}{(\log k)^{1-\epsilon}}. \tag{2}$$

Since we only know that $L(1, \mathrm{sym}^2 f) \gg 1/\log k$ we see that weak subconvexity also fails (now only by $(\log k)^{\epsilon}$) to show the decay of inner products with Eisenstein series. However one can show that $L(1, \mathrm{sym}^2 f)$ is very rarely less than $(\log k)^{-\delta}$ for any $\delta > 0$ (there are at most K^{ϵ} exceptional Hecke eigenforms with weight below K), and so for the vast majority of cases weak subconvexity suffices. On GRH we also know that $L(1, \mathrm{sym}^2 f) \gg 1/\log\log k$, but improving lower bounds for L-functions on the 1-line is unconditionally a very difficult problem connected with widening the zero-free region for that L-function (and so quite likely harder than subconvexity!).

Now let us turn to the inner product with a fixed Hecke-Maass cusp form. Here there is a deep and beautiful formula of Tom Watson (see Theorem 3 of [59]) shows that (here ϕ has been normalized so that $\langle \phi, \phi \rangle = 1$) which is exactly analogous to the easier Eisenstein series case:

$$|\langle \phi F_k, F_k \rangle|^2 = \frac{1}{8} \frac{L_{\infty}(\frac{1}{2}, f \times f \times \phi) L(\frac{1}{2}, f \times f \times \phi)}{\Lambda(1, \mathrm{sym}^2 f)^2 \Lambda(1, \mathrm{sym}^2 \phi)}$$

where $L(s, f \times f \times \phi)$ is the triple product L-function, and L_{∞} denotes its Gamma factors, whose definitions we now recall. Also, in the formula above,

$$\Lambda(s, \mathrm{sym}^2 f) = \Gamma_{\mathbb{R}}(s+1)\Gamma_{\mathbb{R}}(s+k-1)\Gamma_{\mathbb{R}}(s+k) L(s, \mathrm{sym}^2 f),$$

and

$$\Lambda(s, \mathrm{sym}^2 \phi) = \Gamma_{\mathbb{R}}(s)\Gamma_{\mathbb{R}}(s+2it_{\phi})\Gamma_{\mathbb{R}}(s-2it_{\phi}) L(s, \mathrm{sym}^2 \phi),$$

where we have written the Laplace eigenvalue of ϕ as $\lambda_{\phi} = \frac{1}{4} + t_{\phi}^2$, where[4] $t_{\phi} \in \mathbb{R}$.

[4]This is true since we are working on the full modular group. For a congruence subgroup, we could use Selberg's bound that the least eigenvalue is $\geq \frac{3}{16}$ which gives that $|\mathrm{Im}(t_{\phi})| \leq \frac{1}{4}$; see [30].

Write the p-th Hecke eigenvalue of f as $\alpha_f(p)+\beta_f(p)$ where $\alpha_f(p)\beta_f(p) = 1$ and $|\alpha_f(p)| = |\beta_f(p)| = 1$. Write the p-th Hecke eigenvalue of ϕ as $\alpha_\phi(p) + \beta_\phi(p)$ where $\alpha_\phi(p)\beta_\phi(p) = 1$, but we do not know here the Ramanujan conjecture that these are both of size 1. The triple product L-function $L(s, f\times f\times\phi)$ is then defined by means of the Euler product of degree 8 (absolutely convergent in $\mathrm{Re}(s) > 1$)

$$\prod_p \quad \Big(1-\frac{\alpha_f(p)^2\alpha_\phi(p)}{p^s}\Big)^{-1}\Big(1-\frac{\alpha_\phi(p)}{p^s}\Big)^{-2}\Big(1-\frac{\beta_f(p)^2\alpha_\phi(p)}{p^s}\Big)^{-1}$$
$$\times \quad \Big(1-\frac{\alpha_f(p)^2\beta_\phi(p)}{p^s}\Big)^{-1}\Big(1-\frac{\beta_\phi(p)}{p^s}\Big)^{-2}\Big(1-\frac{\beta_f(p)^2\beta_\phi(p)}{p^s}\Big)^{-1}.$$

This L-function is not primitive and factors as $L(s,\phi)L(s,\mathrm{sym}^2 f\times\phi)$. The archimedean factor $L_\infty(s, f\times f\times\phi)$ is defined as the product of eight Γ-factors

$$\prod_{\pm}\Gamma_{\mathbb{R}}(s+k-1\pm it_\phi)\Gamma_{\mathbb{R}}(s+k\pm it_\phi)\Gamma_{\mathbb{R}}(s\pm it_\phi)\Gamma_{\mathbb{R}}(s+1\pm it_\phi).$$

From the work of Garrett [11], it is known that the *completed L-function* $L(s, f\times f\times\phi)L_\infty(s, f\times f\times\phi)$ is an entire function in $\mathbb{C}$, and its value at s equals its value at $1-s$.

Using Stirling's formula we deduce that

$$|\langle\phi F_k, F_k\rangle|^2 \ll_\phi \frac{L(\frac{1}{2}, f\times f\times\phi)}{kL(1,\mathrm{sym}^2 f)^2}.$$

Now the analytic conductor of $L(\frac{1}{2}, f\times f\times\phi)$ is about k^4, and again we see that the convexity bound ($\ll_\phi k$) is insufficient to show that the triple product above tends to zero, but any subconvexity bound would suffice. In particular GRH again gives that these triple products tend to zero as $k\to\infty$ (and so the Rudnick-Sarnak conjecture is implied by GRH). In this case also we have a weak subconvexity bound

$$L(\tfrac{1}{2}, f\times f\times\phi) \ll_{\phi,\epsilon} \frac{k}{(\log k)^{1-\epsilon}} \tag{3}$$

and so if $L(1,\mathrm{sym}^2 f) \geq (\log k)^{-\frac{1}{2}+\delta}$ for some $\delta > 0$ then we would be done. Such a bound holds in all but a very small number of exceptional cases, but establishing such a lower bound in all cases seems extremely difficult: even for the zeta-function we only know that $|\zeta(1+it)| \gg (\log|t|)^{-\frac{2}{3}-\epsilon}$, and the methods of Vinogradov that achieve this are unavailable for general L-functions.

Suggested reading. The approach to QUE via the spectral expansion was carried out by Luo and Sarnak [40] in the case of Eisenstein series (even though these are not in L^2); their work and the paper by

Jakobson [34] make instructive reading. Here the subconvexity problems that arise can be solved.

5. Inner products with Poincare series and the shifted convolution problem

Now we turn to the approach to the Rudnick-Sarnak conjecture via incomplete Eisenstein and Poincare series. First let us consider the inner product with Poincare series $P_m(z \mid \psi)$ with $m \neq 0$. The inner product $\langle F_k, F_k P_m(\cdot \mid \psi)\rangle$ can be evaluated by the Rankin-Selberg unfolding method. This was carried out by Luo and Sarnak [39], and we quickly recall the argument. We have (recall $X = SL_2(\mathbb{Z})\backslash\mathbb{H}$)

$$\begin{aligned}\langle F_k, F_k P_m(\cdot \mid \psi)\rangle &= \int_X y^k |f(z)|^2 \sum_{\gamma\in\Gamma_\infty\backslash\Gamma} e(m\gamma z)\psi(\mathrm{Im}(\gamma z))\frac{dx\ dy}{y^2}\\ &= \int_0^1\int_0^\infty y^k|f(z)|^2\psi(y)e(mz)\frac{dxdy}{y^2}\end{aligned}$$

and by Parseval this equals

$$C^2\sum_{r=1}^\infty \lambda_f(r)\lambda_f(r+m)(r(m+r))^{\frac{k-1}{2}}\int_0^\infty y^{k-1}\psi(y)e^{-4\pi(r+m)y}\frac{dy}{y},$$

where we set the Hecke eigenvalues at negative integers to be zero.

Now it is easy to analyze the integral over y above. The term $y^{k-1}e^{-4\pi(r+m)y}$ attains its maximum for $y = (k-1)/(4\pi(r+m))$, and is sharply peaked at that maximum. Note also that $\int_0^\infty y^{k-1}e^{-4\pi(r+m)y}\frac{dy}{y} = (4\pi(r+m))^{-(k-1)}\Gamma(k-1)$. From these remarks, and using from §2 the formula for $|C|^2$, we obtain that $\langle F_k, F_k P_m(\cdot \mid \psi)\rangle$ is

$$\sim \frac{2\pi^2}{(k-1)L(1,\mathrm{sym}^2 f)}\sum_{r\geq 1}\Big(\frac{r}{r+m}\Big)^{\frac{k-1}{2}}\lambda_f(r)\lambda_f(r+m)\psi\Big(\frac{k-1}{4\pi(r+m)}\Big).$$

Since ψ is a fixed smooth function compactly supported in $(0,\infty)$ we may think of the above sum as essentially being

$$\frac{1}{kL(1,\mathrm{sym}^2 f)}\sum_{r\asymp k}\lambda_f(r)\lambda_f(r+m),$$

where r runs over a range of values of size k. Finding cancellation in such sums is known as the *shifted convolution problem*. If $m\neq 0$ then we expect that the terms $\lambda_f(r)$ and $\lambda_f(r+m)$ behave independently and cancel out on average. If that were so, then we would reach the desired conclusion that the triple product with Poincare series tends to zero. For fixed m and k, and as $x\to\infty$ it is known that there is cancellation

in $\sum_{r\leq x}\lambda_f(r)\lambda_f(r+m)$, however in our case we are interested in the delicate range where x is of size k, and such cancellation remains unknown. Holowinsky's ingenious idea is to forego cancellation in shifted convolution sums, and instead just bound $\sum_{r\asymp k}|\lambda_f(r)\lambda_f(r+m)|$. The insight is that the Hecke eigenvalues tend to be small in size, and we will explain this in more detain in §7.

One can also carry out the above argument for $m=0$ when we have the incomplete Eisenstein series $E(z\mid\psi)$. The only difference is that here we have a main term to deal with. Here we want to show that

$$\langle F_k, F_kE(\cdot\mid\psi)\rangle\to\frac{3}{\pi}\langle 1, E(\cdot\mid\psi)\rangle=\frac{3}{\pi}\int_0^\infty\psi(y)\frac{dy}{y^2},$$

where the equality above follows by unfolding. Arguing as above we find that the LHS above is

$$\sim\frac{2\pi^2}{(k-1)L(1,\mathrm{sym}^2f)}\sum_{r=1}^\infty|\lambda_f(r)|^2\psi\Big(\frac{k-1}{4\pi r}\Big),$$

and so the problem here is to show that

$$\frac{2\pi^2}{(k-1)L(1,\mathrm{sym}^2f)}\sum_{r=1}^\infty|\lambda_f(r)|^2\psi\Big(\frac{k-1}{4\pi r}\Big)\sim\frac{3}{\pi}\int_0^\infty\psi(y)\frac{dy}{y^2}. \tag{4}$$

In this context we recall that by Rankin-Selberg theory we have

$$\sum_{n\leq x}|\lambda_f(n)|^2\sim L(1,\mathrm{sym}^2f)x,$$

for $x\geq k^{1+\epsilon}$. This makes our asymptotic above plausible, but just out of reach. A subconvexity bound for the symmetric square L-function would give our desired asymptotic, but as noted earlier this remains unknown.

Here Holowinsky introduces an important refinement of evaluating the above inner product. The idea is to use the Siegel domain $\{0\leq x\leq 1, y>1/Y\}$ for some parameter Y. This Siegel domain contains essentially $3Y/\pi$ copies of the fundamental domain for $SL_2(\mathbb{Z})\backslash\mathbb{H}$, and further it is relatively easy to compute inner products on this Siegel domain. In this manner we can reduce the problem of establishing (4) to proving asymptotics for $\sum_{n\leq x}|\lambda_f(n)|^2$ for x of size kY. In this argument Y will be chosen to be a power of $(\log k)$, and this small extra flexibility allows the use of weak subconvexity to resolve this problem.

Suggested reading. Luo and Sarnak [39], and Holowinsky [28].

6. Mean values of multiplicative functions and weak subconvexity

We saw in §4 how the Rudnick-Sarnak conjecture is related to obtaining subconvex bounds for values of certain L-functions on the critical line. We now describe a general result which obtains a weak subconvexity bound for such L-values; special cases of this result were described already in §4.

A fundamental problem in number theory is to estimate the values of L-functions at the center of the critical strip. The Langlands program predicts that all L-functions arise from automorphic representations of $GL(N)$ over a number field, and moreover that such L-functions can be decomposed as a product of primitive L-functions arising from irreducible cuspidal representations of $GL(n)$ over $\mathbb{Q}$. The L-functions that we consider will either arise in this manner, or will be the Rankin-Selberg L-function associated to two irreducible cuspidal representations. Note that such Rankin-Selberg L-functions are themselves expected to arise from automorphic representations, but this is not known in general.

Given an irreducible cuspidal automorphic representation π (normalized to have unitary central character), we denote the associated standard L-function by $L(s,\pi)$, and its analytic conductor (whose definition we shall recall shortly) by $C(\pi)$. There holds generally a convexity bound of the form $L(\frac{1}{2},\pi) \ll_\epsilon C(\pi)^{\frac{1}{4}+\epsilon}$ (see Molteni [45]). [5] The Riemann hypothesis for $L(s,\pi)$ implies the Lindelöf hypothesis: $L(\frac{1}{2},\pi) \ll C(\pi)^\epsilon$. In several applications it has emerged that the convexity bound barely fails to be of use, and that any improvement over the convexity bound would have significant consequences. Obtaining such subconvexity bounds has been an active area of research, and estimates of the type $L(\frac{1}{2},\pi) \ll C(\pi)^{\frac{1}{4}-\delta}$ for some $\delta > 0$ have been obtained for several important classes of L-functions. However in general the subconvexity problem remains largely open. For comprehensive accounts on L-functions and the subconvexity problem we refer to Iwaniec and Sarnak [33], Michel [43], and for example the papers [2], [7], [35], [44] and [58].

We now describe an axiomatic framework (akin to the Selberg class) for the class of L-functions that we consider. The properties of L-functions that we assume are mostly standard, and we have adopted this framework in order to clarify the crucial properties needed for our

[5] Recently, Roger Heath-Brown [23] has pointed out an elegant application of Jensen's formula for strips that leads generally to the stronger convexity bound $L(\frac{1}{2},\pi) \ll C(\pi)^{\frac{1}{4}}$.

method. In addition to the usual assumptions of a Dirichlet series with an Euler product and a functional equation, we will need an assumption on the size of the Dirichlet series coefficients. We call this a *weak* Ramanujan hypothesis, as the condition is implied by the Ramanujan conjectures. The reader may prefer to ignore our conditions below and restrict his attention to automorphic L-functions satisfying the Ramanujan conjectures, but our framework allows us to deduce results even in cases where the Ramanujan conjectures are not known.

Let $m \geq 1$ be a fixed natural number. Let [6] $L(s,\pi)$ be given by the Dirichlet series and Euler product

$$L(s,\pi) = \sum_{n=1}^{\infty} \frac{a_\pi(n)}{n^s} = \prod_p \prod_{j=1}^{m} \Big(1 - \frac{\alpha_{j,\pi}(p)}{p^s}\Big)^{-1}, \tag{5}$$

and we suppose that both the series and product are absolutely convergent in $\mathrm{Re}(s) > 1$. We write

$$L(s,\pi_\infty) = N^{\frac{s}{2}} \prod_{j=1}^{m} \Gamma_{\mathbb{R}}(s+\mu_j) \tag{6}$$

where $\Gamma_{\mathbb{R}}(s) = \pi^{-s/2}\Gamma(s/2)$, N denotes the conductor, and the μ_j are complex numbers. The completed L-function $L(s,\pi)L(s,\pi_\infty)$ has an analytic continuation[7] to the entire complex plane, and has finite order. Moreover, it satisfies a functional equation

$$L(s,\pi_\infty)L(s,\pi) = \kappa L(1-s,\tilde{\pi}_\infty)L(1-s,\tilde{\pi}), \tag{7}$$

where κ is the root number (a complex number of magnitude 1), and

$$L(s,\tilde{\pi}) = \sum_{n=1}^{\infty} \frac{\overline{a_\pi(n)}}{n^s}, \qquad \text{and} \qquad L(s,\tilde{\pi}_\infty) = N^{\frac{s}{2}} \prod_{j=1}^{m} \Gamma_{\mathbb{R}}(s+\overline{\mu_j}). \tag{8}$$

We define the analytic conductor $C = C(\pi)$ (see [21]) by

$$C(\pi) = N \prod_{j=1}^{m} (1 + |\mu_j|). \tag{9}$$

Our goal is to obtain an estimate for $L(\frac{1}{2},\pi)$ in terms of the analytic conductor $C(\pi)$.

Properties (5), (6), (7), (8), (9) are standard features of all interesting L-functions. We now need an assumption on the size of the numbers $\alpha_{j,\pi}(p)$. The Ramanujan conjectures, which are expected to

[6]Here the notation is meant to suggest that π corresponds to an automorphic representation, but this is not assumed.

[7]Thus we are not allowing $L(s,\pi)$ to have any poles. It would not be difficult to modify our results to allow the completed L-function to have poles at 0 and 1.

hold for all L-functions, predict that $|\alpha_{j,\pi}(p)| \le 1$ for all p. Further, it is expected that the numbers μ_j appearing in (6) all satisfy $\mathrm{Re}(\mu_j) \ge 0$. Towards the Ramanujan conjectures it is known (see [41]) that if π is an irreducible cuspidal representation of $GL(m)$ then $|\alpha_{j,\pi}(p)| \le p^{\frac{1}{2}-\delta_m}$ for all p, and that $\mathrm{Re}(\mu_j) \ge -\frac{1}{2} + \delta_m$ where $\delta_m = 1/(m^2+1)$. We will make the following weak Ramanujan hypothesis.

Write

$$-\frac{L'}{L}(s,\pi) = \sum_{n=1}^{\infty} \frac{\lambda_\pi(n)\Lambda(n)}{n^s}, \tag{10}$$

where $\lambda_\pi(n) = 0$ unless $n = p^k$ is a prime power when it equals $\sum_{j=1}^m \alpha_{j,\pi}(p)^k$. We assume that for some constants $A_0, A \ge 1$, and all $x \ge 1$ there holds

$$\sum_{x<n\le ex} \frac{|\lambda_\pi(n)|^2}{n}\Lambda(n) \le A^2 + \frac{A_0}{\log ex}. \tag{11}$$

Note that the Ramanujan conjecture would give (11) with $A = m$, and $A_0 \ll m^2$. Analogously for the parameters μ_j we assume that[8]

$$\mathrm{Re}(\mu_j) \ge -1 + \delta_m, \qquad \text{for some } \delta_m > 0, \text{ and all } 1 \le j \le m. \tag{12}$$

If $L(s,\pi)$ is an L-function satisfying the above criteria, then for any $\epsilon > 0$ we have

$$L(\tfrac{1}{2},\pi) \ll \frac{C(\pi)^{\frac{1}{4}}}{(\log C(\pi))^{1-\epsilon}}. \tag{13}$$

The L-functions of §4 satisfy the criteria above (for the symmetric square L-function the weak Ramanujan criterion follows from the known estimates of Deligne, for the triple product L-function one uses the Rankin-Selberg theory for the *fixed* Maass form ϕ in order to check the weak Ramanujan criterion), and hence the estimates stated in §4 hold.

We now discuss the main ideas behind the proof of (13). For an L-function satisfying the above criteria, we may use the convexity bound

[8]This assumption is very weak: from [41] we know that it holds for all automorphic L-functions, and also for the Rankin-Selberg L-function associated to two automorphic representations.

to establish that [9]

$$\sum_{n\leq x} a_\pi(n) \ll \frac{x}{\log x}, \tag{14}$$

provided $x \geq C^{\frac{1}{2}}(\log C)^B$ for some positive constant B. Our main idea is to show that similar cancellation holds even when $x = C^{\frac{1}{2}}(\log C)^{-B}$ for any constant B: For any $\epsilon > 0$, any positive constant B, and all $x \geq C^{\frac{1}{2}}(\log C)^{-B}$ we have

$$\sum_{n\leq x} a_\pi(n) \ll \frac{x}{(\log x)^{1-\epsilon}}. \tag{15}$$

The implied constant may depend on A, A_0, m, δ_m, B and ϵ.

Once (15) is established, (13) will follow from a standard partial summation argument using an approximate functional equation. In (15) and (13), by keeping track of the various parameters involved, it would be possible to quantify ϵ. However, the limit of our method would be to obtain a bound $C^{\frac{1}{4}}/\log C$ in (13), and $x/\log x$ in (15).

Why does the extrapolation (15) hold? At the heart of its proof is the fact that mean values of multiplicative functions vary slowly. Knowing (14) in the range $x \geq C^{\frac{1}{2}}(\log C)^B$, this fact will enable us to extrapolate (14) to the range $x \geq C^{\frac{1}{2}}(\log C)^{-B}$.

The possibility of obtaining such extrapolations was first considered by Hildebrand [24], [25]. If f is a multiplicative function, we shall denote by $S(x) = S(x; f)$ the partial sum $\sum_{n\leq x} f(n)$. Hildebrand [25] showed that if $-1 \leq f(n) \leq 1$ is a real valued multiplicative function then for $1 \leq w \leq \sqrt{x}$

$$\frac{1}{x}\sum_{n\leq x} f(n) = \frac{w}{x}\sum_{n\leq x/w} f(n) + O\Big(\Big(\log\frac{\log x}{\log 2w}\Big)^{-\frac{1}{2}}\Big). \tag{16}$$

In other words, the mean value of f at x does not change very much from the mean-value at x/w. Hildebrand [24] used this idea to show that from knowing Burgess's character sum estimates[10] for $x \geq q^{\frac{1}{4}+\epsilon}$ one may obtain some non-trivial cancellation even in the range $x \geq q^{\frac{1}{4}-\epsilon}$.

[9] We recall here that $L(s, \pi)$ was assumed not to have any poles. If we alter our framework to allow a pole at $s = 1$, say, then (14) would be modified to an asymptotic formula with a main term of size x. Then (15) would extrapolate that asymptotic formula to a wider region.

[10] For simplicity, suppose that q is cube-free.

Elliott [9] generalized Hildebrand's work to cover complex valued multiplicative functions with $|f(n)| \le 1$, and also strengthened the error term in (16). Notice that a direct extension of (16) for complex valued functions is false. Consider $f(n) = n^{i\tau}$ for some real number $\tau \neq 0$. Then $S(x; f) = x^{1+i\tau}/(1+i\tau) + O(1)$, and $S(x/w; f) = (x/w)^{1+i\tau}/(1+i\tau) + O(1)$. Therefore (16) is false, and instead we have that $S(x)/x$ is close to $w^{i\tau}S(x/w)/(x/w)$. Building on the pioneering work of Halasz [19], [20] on mean-values of multiplicative functions, Elliott showed that for a multiplicative function f with $|f(n)| \le 1$, there exists a real number $\tau = \tau(x)$ with $|\tau| \le \log x$ such that for $1 \le w \le \sqrt{x}$

$$S(x) = w^{1+i\tau}S(x/w) + O\Big(x\Big(\frac{\log 2w}{\log x}\Big)^{\frac{1}{19}}\Big). \tag{17}$$

In [16], Granville and Soundararajan give variants and stronger versions of (2.2), with $\frac{1}{19}$ replaced by $1 - 2/\pi - \epsilon$.

In order to establish (15), we require similar results when the multiplicative function is no longer constrained to the unit disc. The situation here is considerably more complicated, and instead of showing that a suitable linear combination of $S(x)/x$ and $S(x/w)/(x/w)$ is small, we will need to consider linear combinations involving several terms $S(x/w^j)/(x/w^j)$ with $j = 0, \ldots, J$. In order to motivate our main result, it is helpful to consider two illustrative examples.

Example 1. Let k be a natural number, and take $f(n) = d_k(n)$, the k-th divisor function. Then, it is easy to show that $S(x) = xP_k(\log x) + O(x^{1-1/k+\epsilon})$ where P_k is a polynomial of degree $k-1$. If $k \ge 2$, it follows that $S(x)/x - S(x/w)/(x/w)$ is of size $(\log w)(\log x)^{k-2}$, which is not $o(1)$. However, if $1 \le w \le x^{1/2k}$, the linear combination

$$\begin{aligned}\sum_{j=0}^{k}(-1)^j\binom{k}{j}\frac{S(x/w^j)}{x/w^j} &= \sum_{j=0}^{k}(-1)^j\binom{k}{j}P_k(\log x/w^j) + O(x^{-\frac{1}{2k}}) \\ &= O(x^{-\frac{1}{2k}})\end{aligned}$$

is very small.

Example 2. Let $\tau_1, \ldots, \tau_R$ be distinct real numbers, and let $k_1, \ldots, k_R$ be natural numbers. Let f be the multiplicative function defined by $F(s) = \sum_{n=1}^{\infty} f(n)n^{-s} = \prod_{j=1}^{R}\zeta(s - i\tau_j)^{k_j}$. Consider here the linear

combination (for $1 \le w \le x^{1/(2(k_1+\dots+k_R))}$)

$$\frac{1}{x}\sum_{j_1=0}^{k_1}\cdots\sum_{j_R=0}^{k_R} (-1)^{j_1+\dots+j_R}\binom{k_1}{j_1}\cdots\binom{k_R}{j_R} w^{j_1(1+i\tau_1)+\dots+j_R(1+i\tau_R)} \times S\Big(\frac{x}{w^{j_1+\dots+j_R}}\Big).$$

By Perron's formula we may express this as, for $c > 1$,

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\prod_{j=1}^{R}\zeta(s-i\tau_j)^{k_j}(1-w^{1+i\tau_j-s})^{k_j}x^{s-1}\frac{ds}{s}.$$

Notice that the poles of the zeta-functions at $1+i\tau_j$ have been cancelled by the factors $(1-w^{1+i\tau_j-s})^{k_j}$. Thus the integrand has a pole only at $s=0$, and a standard contour shift argument shows that this integral is $\ll x^{-\delta}$ for some $\delta > 0$.

Fortunately, it turns out that Example 2 captures the behavior of mean-values of the multiplicative functions of interest to us. In order to state our result, we require some notation. Let f denote a multiplicative function and recall that

$$S(x) = S(x; f) = \sum_{n\le x} f(n).$$

We shall write

$$F(s) = \sum_{n=1}^{\infty}\frac{f(n)}{n^s},$$

and we shall assume that this series converges absolutely in $\mathrm{Re}(s) > 1$. Moreover we write

$$-\frac{F'}{F}(s) = \sum_{n=1}^{\infty}\frac{\lambda_f(n)\Lambda(n)}{n^s} = \sum_{n=1}^{\infty}\frac{\Lambda_f(n)}{n^s},$$

where $\lambda_f(n) = \Lambda_f(n) = 0$ unless n is the power of a prime p. We next assume the analog of the weak Ramanujan hypothesis (11). Namely, we suppose that there exist constants A, $A_0 \ge 1$ such that for all $x \ge 1$ we have

$$\sum_{x<n\le ex}\frac{|\lambda_f(n)|^2\Lambda(n)}{n} \le A^2 + \frac{A_0}{\log(ex)}. \tag{18}$$

Let R be a natural number, and let $\tau_1, \dots, \tau_R$ denote R real numbers. Let $\underline{\ell} = (\ell_1, \dots, \ell_R)$ and $\underline{j} = (j_1, \dots, j_R)$ denote vectors of non-negative

integers, with the notation $\underline{j} \le \underline{\ell}$ indicating that $0 \le j_1 \le \ell_1$, ..., $0 \le j_R \le \ell_R$. Define

$$\binom{\underline{\ell}}{\underline{j}} = \binom{\ell_1}{j_1} \cdots \binom{\ell_R}{j_R}.$$

Finally, we define a measure of the oscillation of the mean-values of f by setting

$$\begin{aligned} \mathcal{O}_{\underline{\ell}}(x,w) &= \mathcal{O}_{\underline{\ell}}(x,w;\tau_1,\dots,\tau_R) \\ &= \sum_{\underline{j}\le\underline{\ell}} (-1)^{j_1+\dots+j_R} \binom{\underline{\ell}}{\underline{j}} w^{j_1(1+i\tau_1)+\dots+j_R(1+i\tau_R)} S\Big(\frac{x}{w^{j_1+\dots+j_R}}\Big). \end{aligned}$$

With the above notations, the estimate (15) follows from the following result: Let $X \ge 10$ and $1 \ge \epsilon > 0$ be given. Let $R = [10A^2/\epsilon^2]+1$ and put $L = [10AR]$, and $\underline{L} = (L,\dots,L)$. Let w be such that $0 \le \log w \le (\log X)^{\frac{1}{3R}}$. There exist real numbers τ_1, ..., τ_R with $|\tau_j| \le \exp((\log\log X)^2)$ such that for all $2 \le x \le X$ we have

$$|\mathcal{O}_{\underline{L}}(x,w;\tau_1,\dots,\tau_R)| \ll \frac{x}{\log x}(\log X)^{\epsilon}. \tag{19}$$

The implied constant above depends on A, A_0 and ϵ.

Deducing (15) from (19). Let $x_0 = C^{\frac12}(\log C)^B$ be such that the convexity bound gives cancellation in $\sum_{n\le x} a_\pi(n)$ for $x \ge x_0$ as mentioned in(14). Let $x_0 \ge x \ge C^{\frac12}/(\log C)^B$. Take $w = x_0/x$ and $X = xw^{LR}$. Applying (19) to the multiplicative function a_π (note that (11) gives the assumption (18)) we find that for an appropriate choice of τ_1, ..., τ_R that

$$|\mathcal{O}_{\underline{L}}(X,w)| \ll \frac{X}{(\log X)^{1-\epsilon}}. \tag{20}$$

But, by definition, the LHS above is

$$w^{LR}\Big|\sum_{n\le X/w^{LR}} a_\pi(n)\Big| + O\Big(\sum_{j=0}^{LR-1} w^j \Big|\sum_{n\le X/w^j} a_\pi(n)\Big|\Big). \tag{21}$$

Now $X/w^{LR} = x$, and for $0 \le j \le LR-1$ we have $X/w^j \ge xw = x_0$ so that the bound of (14) applies. Therefore (21) equals

$$w^{LR}\Big|\sum_{n\le x} a_\pi(n)\Big| + O\Big(\frac{X}{\log X}\Big),$$

From (20) we conclude that

$$\Big|\sum_{n\le x} a_\pi(n)\Big| \ll w^{-LR}\frac{X}{(\log X)^{1-\epsilon}} \ll \frac{x}{(\log x)^{1-\epsilon}},$$

which proves (15).

For a general multiplicative function, we cannot hope for any better bound for the oscillation than $x/\log x$. To see this, suppose $w \geq 2$, and consider the multiplicative function f with $f(n) = 0$ for $n \leq x/2$ and $f(p) = 1$ for primes $x/2 < p \leq x$. Then $S(x) \gg x/\log x$ whereas $S(x/w^j) = 1$ for all $j \geq 1$, and therefore for any choice of the numbers $\tau_1, \ldots, \tau_R$ we would have $\mathcal{O}_{\underline{L}}(x, w) \gg x/\log x$.

Our proof of (19) builds both on the techniques of Halasz (as developed in [9] and [16]), and also the idea of *pretentious* multiplicative functions developed by Granville and Soundararajan (see [17] and [18]). We describe just a couple of the main ideas used: how the numbers in τ_j in (19) are defined, and more generally what is special about mean-values of multiplicative functions?

Definition of the successive maxima τ_j. From now on, we shall write $T = \exp((\log\log X)^2)$. We define τ_1 to be that point t in the compact set $\mathcal{C}_1 = [-T, T]$ where the maximum of $|F(1+1/\log X+it)|$ is attained. Now remove the interval $(\tau_1-(\log X)^{-\frac{1}{R}}, \tau_1+(\log X)^{-\frac{1}{R}})$ from $\mathcal{C}_1 = [-T, T]$, and let $\mathcal{C}_2$ denote the remaining compact set. We define τ_2 to be that point t in $\mathcal{C}_2$ where the maximum of $|F(1+1/\log X+it)|$ is attained. Next remove the interval $(\tau_2 - (\log X)^{-\frac{1}{R}}, \tau_2 + (\log X)^{-\frac{1}{R}})$ from $\mathcal{C}_2$ leaving behind the compact set $\mathcal{C}_3$. Define τ_3 to be the point where the maximum of $|F(1 + 1/\log X + it)|$ for $t \in \mathcal{C}_3$ is attained. We proceed in this manner, defining the successive maxima $\tau_1, \ldots, \tau_R$, and the nested compact sets $\mathcal{C}_1 \supset \mathcal{C}_2 \supset \ldots \supset \mathcal{C}_R$. Notice that all the points $\tau_1, \ldots, \tau_R$ lie in $[-T, T]$, and moreover are well-spaced: $|\tau_j - \tau_k| \geq (\log X)^{-\frac{1}{R}}$ for $j \neq k$.

It is easy to see that $|F(1 + 1/\log X + it)|$ by $\ll (\log X)^A$. For $t \in [-T, T]$ we will show that a much better bound holds, unless t happens to be near one of the points $\tau_1, \ldots, \tau_R$. This is the content of (22) below which is is inspired by the ideas in [17] and [18].

Let $1 \leq j \leq R$ and let t be a point in $\mathcal{C}_j$. Then

$$|F(1 + 1/\log X + it)| \ll (\log X)^{A\sqrt{1/j+(j-1)/(jR)}}. \tag{22}$$

In particular if $t \in \mathcal{C}_R$ we have $|F(1 + 1/\log X + it)| \ll (\log X)^{\epsilon/2}$.

Proof of (22) If $t \in \mathcal{C}_j$ then for all $1 \leq r \leq j$

$$|F(1+1/\log X+it)| \leq |F(1+1/\log X+i\tau_j)| \leq |F(1+1/\log X+i\tau_r)|.$$

Therefore,

$$\begin{aligned}|F(1+1/\log X+i\tau_j)| &\le \Big(\prod_{r=1}^{j}|F(1+1/\log X+i\tau_r)|\Big)^{\frac{1}{j}}\\ &\le \exp\Big(\mathrm{Re}\frac{1}{j}\sum_{n\ge 2}\frac{\lambda_f(n)\Lambda(n)}{n^{1+1/\log X}(\log n)}(n^{-i\tau_1}+\ldots+n^{-i\tau_j})\Big).\end{aligned}$$

By Cauchy-Schwarz

$$\begin{aligned}\sum_{n\ge 2}\frac{|\lambda_f(n)|\Lambda(n)}{n^{1+1/\log X}\log n}\Big|\sum_{r=1}^{j}n^{-i\tau_r}\Big| &\le \Big(\sum_{n\ge 2}\frac{|\lambda_f(n)|^2\Lambda(n)}{n^{1+1/\log X}\log n}\Big)^{\frac12}\\ &\quad\times\Big(\sum_{n\ge 2}\frac{\Lambda(n)}{n^{1+1/\log X}\log n}\Big|\sum_{r=1}^{j}n^{-i\tau_r}\Big|^2\Big)^{\frac12}.\end{aligned}$$

By (18) the first factor above is $\le (A^2\log\log X+O(1))^{\frac12}$. To handle the second factor, we expand out the square and obtain

$$\begin{aligned}&\sum_{n\ge 2}\frac{\Lambda(n)}{n^{1+1/\log X}\log n}\Big|\sum_{r=1}^{j}n^{-i\tau_r}\Big|^2\\ =\;& j\sum_{n\ge 2}\frac{\Lambda(n)}{n^{1+1/\log X}\log n}+2\mathrm{Re}\sum_{1\le r<s\le j}\sum_{n\ge 2}\frac{\Lambda(n)}{n^{1+1/\log X+i(\tau_r-\tau_s)}\log n}\\ =\;& j(\log\log X+O(1))+2\sum_{1\le r<s\le j}\log|\zeta(1+1/\log X+i(\tau_r-\tau_s))|.\end{aligned}$$

Now note that $(\log X)^{-\frac{1}{R}}\le|\tau_r-\tau_s|\le 2T$ and hence $|\zeta(1+1/\log X+i(\tau_r-\tau_s))|\le(\log X)^{\frac{1}{R}}+O(1)$. Using this above, (22) follows.

Multiplicative functions and integral equations. Write $S_1(x)=S_1(x;f)=\sum_{n\le x}f(n)\log n$. Then

$$S_1(x)=\sum_{d\le x}\Lambda_f(d)S(x/d).$$

If we suppose that $|\Lambda_f(d)|\le k\Lambda(d)$ for some fixed k then the above is bounded in magnitude by

$$\le k\sum_{d\le x}\Lambda(d)|S(x/d,f)|.$$

By a "partial summation" argument (this needs some elaboration and is not obvious) we find that this is

$$\ll\int_1^x|S(x/t,f)|dt=x\int_1^x|S(t,f)|\frac{dt}{t^2}.$$

We may also check that (in the case $|f(n)| \leq 1$, and more care is needed for the general case)

$$S_1(x,f) = S(x,f)\log x + O(x),$$

and thus we conclude that

$$|S(x,f)| \ll \frac{x}{\log x} + \frac{x}{\log x}\int_1^x |S(t,f)|\frac{dt}{t^2}. \tag{23}$$

The relation above is crucial, and it shows how the mean value of a multiplicative function is dominated by an average of such mean values. This forces a smoother structure of these mean values than one would have expected. Wirsing's pioneering result (on mean-values of real valued multiplicative functions) and Halasz's work on complex valued multiplicative functions both exploit this feature very nicely. See also the work of Granville and Soundararajan on the "spectrum of multiplicative functions" where the analogy with integral equations is made precisely.

We have mentioned several times Halasz's theorem without stating it properly. We now describe the result in general terms. Consider a multiplicative function f with $|f(n)| \leq 1$. If f is real valued then Wirsing, proving a conjecture of Erdos and Wintner, showed that $\lim_{x\to\infty} \frac{1}{x}\sum_{n\leq x} f(n)$ exists. Moreover the limit is non-zero if and only if $\sum_{p\leq x}(1-f(p))/p$ converges; that is f looks like the function that is 1 always. To see that this result is non-trivial, just consider $f(n) = \mu(n)$. Halasz generalized Wirsing's result to complex valued multiplicative functions with $|f(n)| \leq 1$. If we consider the example $f(n) = n^{i\alpha}$ where $\sum_{n\leq x} f(n) \sim \frac{x^{1+i\alpha}}{1+i\alpha}$ we see that the limiting mean-value need no longer exist. Halasz realized that this example is the only obstruction, and the limiting mean-value tends to zero (and he quantified this nicely) unless it happens that $\sum_p (1 - \text{Re } f(p)p^{-i\alpha})/p$ converges for some α; that is, f is pretending to be the function $n^{i\alpha}$. When f is no longer restricted to the unit circle, matters are more complicated. But, extending Halasz's insight we may look for functions of the form $n^{-i\alpha_j}$ which f correlates with (or pretends to be). This is the motivation for the *successive maxima* that we identified earlier, and the oscillation result shows that we can handle the effect of those bounded number of functions that f can pretend to be.

Suggested reading. For work on mean-values of multiplicative functions see [60], [19], [20]. A nice account of *pretentiousness* and its applications is given in [14]. Full details of weak subconvexity may be found in [56]. Accounts of the subconvexity problem (together with other applications) can be found in [33] and [43].

7. Sieve methods and Holowinsky's work

Here we describe Holowinsky's approach to bounding the shifted convolution sums that arose in §5. We only deal with the inner products with Poincare series $P_m(z \mid \psi)$ with $m \neq 0$. We begin by explaining why we might expect the size of Hecke eigenvalues to be small on average; such a result goes back to work of Elliott, Moreno and Shahidi [10] in the context of Ramanujan's τ-function.

A general result on non-negative multiplicative functions Suppose we are given a non-negative multiplicative function f. We assume a stronger form of non-negativity: namely, that the coefficients $\lambda_f(n)$ appearing in the series for $\log F(s)$ are non-negative. Note that

$$f(n)\log n = \sum_{d|n} \Lambda_f(d) f(n/d),$$

and so

$$\begin{aligned}\sum_{n\le x} f(n)\log n &= \sum_{n=md\le x} \Lambda_f(d) f(m)\\ &= \sum_{m\le x} f(m) \sum_{d\le x/m} \Lambda_f(d)\\ &= x\Big(\max_{2\le y\le x} \frac{1}{y}\sum_{d\le y}\Lambda_f(d)\Big)\sum_{m\le x}\frac{f(m)}{m}.\end{aligned}$$

Further, since $\log(x/n) \le x/n$,

$$\sum_{n\le x} f(n)\log n \ge \log x \sum_{n\le x} f(n) - x\sum_{n\le x}\frac{f(n)}{n},$$

and so we conclude that

$$S(x;f) \le \frac{x}{\log x}\Big(1 + \max_{2\le y\le x}\frac{1}{y}\sum_{d\le y}\Lambda_f(d)\Big)\sum_{n\le x}\frac{f(n)}{n}. \tag{24}$$

This easy bound is extremely useful. If we now suppose that on the prime powers $\Lambda_f(d) \le A\Lambda(d)$ for some constant A, then we deduce that

$$S(x;f) \ll \frac{x}{\log x}\sum_{n\le x}\frac{f(n)}{n} \ll \frac{x}{\log x}\exp\Big(\sum_{p\le x}\frac{f(p)}{p}\Big). \tag{25}$$

Application: Elliott-Moreno-Shahidi [10]. Take $f(n) = \tau(n)n^{-11/2}$ where τ denotes Ramanujan's function. By Deligne's theorem $|f(n)| \le$

$d(n)$, and $|\Lambda_f(p^k)| \leq k+1$. By (25) we obtain that

$$\sum_{n\leq x}|f(n)| \ll \frac{x}{\log x}\sum_{n\leq x}\frac{|f(n)|}{n} \ll \frac{x}{\log x}\exp\Big(\sum_{p\leq x}\frac{|\tau(p)p^{-11/2}|}{p}\Big).$$

By Rankin-Selberg theory we know that

$$\sum_{p\leq x}\frac{f(p)^2}{p} \sim \log\log x.$$

Using Rankin-Selberg for the $GL(3)$ automorphic form associated to $f(p^2) = f(p)^2 - 1$ we obtain that

$$\sum_{p\leq x}\frac{(f(p)^2-1)^2}{p} \sim \log\log x.$$

But $(f(p)^2-1)^2 \leq 9(|f(p)|-1)^2$ so that

$$\sum_{p\leq x}\frac{(|f(p)|-1)^2}{p} \geq \frac{1}{9}\log\log x + O(1),$$

and we deduce that

$$\sum_{p\leq x}\frac{|f(p)|}{p} \leq \frac{17}{18}\log\log x + O(1).$$

Consequently

$$\sum_{n\leq x}|f(n)| \ll x(\log x)^{-\frac{1}{18}}.$$

This shows that on average the values of $|f(n)|$ are somewhat small.

Here is an explanation of why we might expect $|f(n)|$ to be small. By Rankin-Selberg we know that

$$\sum_{n\leq x} f(n)^2 \sim cx,$$

for a positive constant c (which is related to the symmetric square L-function of Δ evaluated at 1). So by Cauchy-Schwarz we know that $\sum_{n\leq x}|f(n)| \ll x$. For this estimate to be tight, one would need that the $f(n)$ should be of constant size, and since f is multiplicative this means that most $|f(p)|$ should be close to 1. However the distribution of $f(p)$ is governed by the Sato-Tate law (now known thanks to the work of Taylor and others), and so there is considerable fluctuation in the sizes of $|f(n)|$. The mean square is dominated by the large values of $|f(n)|$, and so naturally we would expect the average of $|f(n)|$ to be small. Our argument above uses information about the first four

symmetric powers of Δ, which were known for a while, whereas Sato-Tate amounts to using information about all symmetric powers.

Nair's theorem. In Holowinsky's work we need an estimate for (essentially) $\sum_{n\le k}|\lambda_f(n)\lambda_f(n+m)|$ where $m\neq 0$ (and we have returned to the notation of letting $\lambda_f(n)$ denote the Hecke eigenvalues of an eigenform f). We need now an analog of (25) for these shifted convolution sums. There is a lovely result of Mohan Nair [46] which establishes such an analog for general classes of multiplicative functions evaluated on polynomials. Nair's work extends work of Peter Shiu [55] who had considered such estimates for multiplicative functions in short intervals and arithmetic progressions.

We don't describe Nair's result in full generality, but restrict ourselves to the special case at hand. The basic point is that if m is a fixed non-zero integer then the multiplicative structure of the integers n and $n+m$ should have very little in common (e.g. if $m=1$ the two numbers are coprime), and hence the values $|\lambda_f(n)|$ and $|\lambda_f(n+m)|$ should behave independently of each other. In other words we may expect the average of $|\lambda_f(n)\lambda_f(n+m)|$ to be have like the product of the average of $|\lambda_f(n)|$ and the average of $|\lambda_f(n+m)|$; i.e. like the square of the average of $|\lambda_f(n)|$. Such an analog of (25) is what's guaranteed by Nair's theorem: we have for $n\neq 0$

$$\sum_{n\le k}|\lambda_f(n)\lambda_f(n+m)|\ll_m k\exp\Big(\sum_{p\le k}\frac{2|\lambda_f(p)|-2}{p}\Big). \tag{26}$$

Holowinsky [28] gives an independent proof of a slightly weaker result using a simple Selberg sieve argument.

Using the bound (26) in our work in §5 we find that

$$\langle F_k, F_kP_m(\cdot\mid\psi)\rangle\ll\frac{1}{L(1,\mathrm{sym}^2 f)}\exp\Big(\sum_{p\le k}\frac{2|\lambda_f(p)|-2}{p}\Big). \tag{27}$$

We'll analyze what this means in the next section, where we show how this estimate together with the weak subconvexity bounds of §4 together give a proof of the Rudnick-Sarnak conjecture.

Suggested reading. The papers [10], [28], [46], [47], [55], all make for interesting reading.

8. Proof of mass equidistribution

We first describe the result on the *cuspidal space*. The cuspidal space is that spanned by the Hecke-Maass cusp forms, or equivalently that spanned by the incomplete Poincare series $P_m(z\mid\psi)$ with $m\neq 0$.

First we observe that

$$L(1, \mathrm{sym}^2 f) \gg \exp\Big(\sum_{p\leq k} \frac{\lambda_f(p^2)}{p}\Big). \tag{28}$$

In fact $L(1, \mathrm{sym}^2 f)$ is of the same size as the RHS above, and the RHS is essentially the Euler product defining $L(s, \mathrm{sym}^2 f)$. This can be established generally for any L-function at the edge of the critical strip, provided there are no Siegel zeros. In the symmetric square case, we already noted that the work of Hoffstein-Lockhart and Goldfeld-Hoffstein-Lieman rules out the existence of Siegel zeros. A slightly weaker version of this bound is described in Lemma 2 of [29].

Using this bound, and noting that $\lambda_f(p^2) = \lambda_f(p)^2 - 1$, in (27) we deduce that

$$\langle F_k, F_k P_m(\cdot \mid \psi)\rangle \ll \exp\Big(-\sum_{p\leq k} \frac{(|\lambda_f(p)| - 1)^2}{p}\Big).$$

Thus Holowinsky's argument would give the decay of inner products with Poincare series unless it so happened that

$$\sum_{p\leq k} \frac{(|\lambda_f(p)| - 1)^2}{p} \ll 1.$$

But if the above holds then

$$\sum_{p\leq k} \frac{\lambda_f(p^2)}{p} = \sum_{p\leq k} \frac{(\lambda_f(p)+1)(\lambda_f(p)-1)}{p} \geq -3\sum_{p\leq k} \frac{||\lambda_f(p)| - 1|}{p},$$

and using Cauchy-Schwarz we have

$$\sum_{p\leq k} \frac{||\lambda_f(p)| - 1|}{p} \ll \sqrt{\log\log k}.$$

Using this in (28) we have in this case $L(1, \mathrm{sym}^2 f) \gg (\log k)^{-\epsilon}$. But then the weak subconvexity bound immediately gives that the inner products of F_k with fixed Maass cusp forms is small. In other words, if Holowinsky's method fails then weak subconvexity succeeds! A variant of this argument is described in [29] and these argument show that the inner product of F_k with an element in the cuspidal space is always small.

The argument on the space of Eisenstein series is a little more complicated, but the principle remains the same: the two different methods work in complementary cases and together give the result. We noted earlier in §6 that here Holowinsky introduced a refinement of the Luo-Sarnak criterion, and that refinement together with weak subconvexity

for Rankin-Selberg L-functions can be used to show that

$$\Big|\langle F_k, F_k E(\cdot \mid \psi)\rangle - \frac{3}{\pi}\langle 1, E(\cdot \mid \psi)\rangle\Big| \ll (\log k)^{\epsilon} \exp\Big(-\frac{1}{2}\sum_{p\le k}\frac{(|\lambda_f(p)|-1)^2}{p}\Big).$$

Now the proof proceeds as in the cuspidal case.

Suggested reading. Complete details of this proof, together with a quantification of the rate of convergence, are given in [29].

9. The escape of mass argument

In this last section we give a description of the argument in [57] which eliminates the possibility of escape of mass for Hecke-Maass cusp forms, and thus completes Lindenstrauss's proof of QUE for $SL_2(\mathbb{Z})\backslash\mathbb{H}$.

As remarked in the introduction, Lindenstrauss has shown that any weak-$*$ limit of the micro-local lifts of Hecke-Maass forms is a constant c (in $[0,1]$) times the normalized volume measure on $Y = SL_2(\mathbb{R})\backslash SL_2(\mathbb{Z})$. Projecting these measures down to the modular surface X, we see that any weak-$*$ limit of the measures μ_ϕ associated to Hecke-Maass forms is of the shape $c\frac{3}{\pi}\frac{dx\ dy}{y^2}$. Our aim is to show that in fact $c = 1$, and there is no escape of mass. If on the contrary $c < 1$ for some weak-$*$ limit, then we have a sequence of Hecke-Maass forms ϕ_j with eigenvalues λ_j tending to infinity such that for any fixed $T \ge 1$ and as $j \to \infty$

$$\int_{\substack{z\in\mathcal{F}\\ y\le T}} |\phi_j(z)|^2 \frac{dx\ dy}{y^2} = (c+o(1))\frac{3}{\pi}\int_{\substack{z\in\mathcal{F}\\ y\le T}} \frac{dx\ dy}{y^2} = (c+o(1))\Big(1-\frac{3}{\pi T}\Big);$$

here $\mathcal{F} = \{z = x+iy : |z| \ge 1,\ -1/2 \le x \le 1/2, y > 0\}$ denotes the usual fundamental domain for $SL_2(\mathbb{Z})\backslash\mathbb{H}$. It follows that as $j \to \infty$

$$\int_{\substack{|x|\le \frac{1}{2}\\ y\ge T}} |\phi_j(z)|^2 \frac{dx\ dy}{y^2} = 1 - c + \frac{3}{\pi T}c + o(1). \tag{29}$$

Now uniformly for any Hecke-Maass form of eigenvalue $\lambda = \frac{1}{4} + r^2$ (and normalized to have Petersson norm 1) we may show that

$$\int_{\substack{|x|\le \frac{1}{2}\\ y\ge T}} |\phi(z)|^2 \frac{dx\ dy}{y^2} \ll \frac{\log(eT)}{\sqrt{T}}. \tag{30}$$

Clearly (30) contradicts (29) if $c < 1$ for suitably large T, and this establishes that $c = 1$.

Now let us explain why (30) holds. Letting $\lambda(n)$ denote the n-th Hecke eigenvalue of the form ϕ, we recall that ϕ has a Fourier expansion

of the form

$$\phi(z) = C\sqrt{y}\sum_{n=1}^{\infty}\lambda(n)K_{ir}(2\pi ny)\cos(2\pi nx),$$

or

$$\phi(z) = C\sqrt{y}\sum_{n=1}^{\infty}\lambda(n)K_{ir}(2\pi ny)\sin(2\pi nx),$$

where C is a constant (normalizing the L^2 norm), K denotes the usual K-Bessel function, and we have cos or sin depending on whether the form is even or odd.

Using Parseval we find that

$$\int_{\substack{|x|\leq\frac{1}{2}\\ y\geq T}}|\phi(x+iy)|^2\frac{dx\ dy}{y^2} = \frac{C^2}{2}\int_T^{\infty}\sum_{n=1}^{\infty}|\lambda(n)|^2|K_{ir}(2\pi ny)|^2\frac{dy}{y}.$$

By a change of variables we may write this as

$$\frac{C^2}{2}\sum_{n=1}^{\infty}|\lambda(n)|^2\int_{nT}^{\infty}|K_{ir}(2\pi t)|^2\frac{dt}{t} = \frac{C^2}{2}\int_1^{\infty}|K_{ir}(2\pi t)|^2\sum_{n\leq t/T}|\lambda(n)|^2\frac{dt}{t}.$$

Now for $t \geq 1$ if we know that

$$\sum_{n\leq t/T}|\lambda(n)|^2 \leq 10^8\frac{\log eT}{\sqrt{T}}\sum_{n\leq t}|\lambda(n)|^2, \tag{31}$$

then the above is

$$\begin{aligned} &\ll \frac{\log eT}{\sqrt{T}}\frac{C^2}{2}\int_1^{\infty}|K_{ir}(2\pi t)|^2\sum_{n\leq t}|\lambda(n)|^2\frac{dt}{t} \\ &= \frac{\log eT}{\sqrt{T}}\int_{\substack{|x|\leq\frac{1}{2}\\ y\geq 1}}|\phi(x+iy)|^2\frac{dx\ dy}{y^2} \ll \frac{\log eT}{\sqrt{T}}, \end{aligned}$$

since the region $|x| \leq \frac{1}{2}$, $y \geq 1$ is contained inside a fundamental domain for $SL_2(\mathbb{Z})\backslash\mathbb{H}$. This would prove (30).

Lastly it remains to justify (31). In fact this statement is a general fact about a large class of multiplicative functions that we will call *Hecke-multiplicative*. We say that a function f is Hecke-multiplicative if it satisfies the Hecke relation

$$f(m)f(n) = \sum_{d|(m,n)}f(mn/d^2),$$

and $f(1) = 1$. If f is Hecke-multiplicative, then for all $1 \le y \le x$ we have

$$\sum_{n \le x/y} |f(n)|^2 \le 10^8 \Big(\frac{1 + \log y}{\sqrt{y}} \Big) \sum_{n \le x} |f(n)|^2. \tag{32}$$

Clearly this statement proves (31).

We won't go into the proof of (32), but just mention that it is based on elementary analytic and combinatorial arguments. It is noteworthy that (32) makes no assumptions on the size of the function f. Hecke-multiplicative functions satisfy $f(p^2) = f(p)^2 - 1$, so that at least one of $|f(p)|$ or $|f(p^2)|$ must be bounded away from zero; this observation plays a crucial role in our proof. We also remark that apart from the $\log y$ factor, (32) is best possible: Consider the Hecke-multiplicative function f defined by $f(p) = 0$ for all primes p. The Hecke relation then mandates that $f(p^{2k+1}) = 0$ and $f(p^{2k}) = (-1)^k$. Therefore, in this example, $\sum_{n \le x} |f(n)|^2 = \sqrt{x} + O(1)$ and $\sum_{n \le x/y} |f(n)|^2 = \sqrt{x/y} + O(1)$.

Suggested reading. For the proof of (32) see [56].

References

[1] N. Anantharaman *Entropy and localization of eigenfunctions* Ann. of Math. **168** (2008) 435–475.

[2] J. Bernstein and A. Reznikov, *Periods, subconvexity of L-functions and representation theory*, J. Differential Geom. **70** (2005), 129–141.

[3] S. de Bievere *Quantum chaos: a brief first visit* , available from his website.

[4] S. Böcherer, P. Sarnak, and R. Schulze-Pillot, *Arithmetic and equidistribution of measures on the sphere*, Comm. Math. Phys. **242**, (2003), 67–80.

[5] F. Brumley, *Effective multiplicity one on GL_N and narrow zero-free regions for Rankin-Selberg L-function*, Amer. J. Math., **128** (2006), 1455-1474.

[6] J. Cogdell and P. Michel, *On the complex moments of symmetric power L-functions at $s = 1$*, IMRN (2004), 1561–1617.

[7] W. Duke, J. Friedlander, and H. Iwaniec *The subconvexity problem for Artin L-functions* Invent. Math. **149** (2002) 489–577.

[8] W. Duke and R. Schulze-Pillot *Representation of integers by positive ternary quadratic forms and the equidistribution of lattice points on ellipsoids* Invent. Math. **99** (1990) 49–57.

[9] P.D.T.A. Elliott, *Extrapolating the mean-values of multiplicative functions*, Indag. Math., **51** (1989), 409–420.

[10] P.D.T.A. Elliott, C. Moreno, and F. Shahidi, *On the absolute value of Ramanujan's τ-function*, Math. Ann. **266** (1984) 507–511.

[11] P. Garrett, *Decomposition of Eisenstein series: Rankin triple products*, Ann. of Math. **125**, (1987) 209–235.

[12] S. Gelbart and H. Jacquet *A relation between automorphic representations of* GL(2) *and* GL(3) Ann. Sci. Ecole Norm. Sup. **11** (1978), 471–542.

[13] D. Goldfeld, J. Hoffstein and D. Lieman *Appendix to the paper by Hoffstein and Lockhart* Ann. of Math. **140** (1994) 161–181.
[14] A. Granville *Pretentiousness in analytic number theory*, J. Theor. Nombres Bordeaux **21** (2009) 159–173.
[15] A. Granville and K. Soundararajan *The distribution of values of* $L(1, \chi_d)$ Geom. funct. anal. **13**, (2003), 992-1028.
[16] A. Granville and K. Soundararajan *Decay of mean-values of multiplicative functions*, Can. J. Math. **55** (2003) 1191-1230.
[17] A. Granville and K. Soundararajan *Pretentious multiplicative functions and an inequality for the zeta-function*, CRM Proceedings and Lecture Notes, **46** (2008) 191–197.
[18] A. Granville and K. Soundararajan *Large character sums: Pretentious characters and the Polya-Vinogradov theorem* J. Amer. Math. Soc. **20** (2007), 357–384.
[19] G. Halasz, *On the distribution of additive and mean-values of multiplicative functions* Studia Sci. Math. Hungar. **6** (1971) 211-233.
[20] G. Halasz *On the distribution of additive arithmetic functions* Acta Arith. **27** (1975) 143–152.
[21] G. Harcos *Uniform approximate functional equation for principal L-functions*, Int. Math. Res. Not. (2002) 923–932.
[22] A. Hassell *Ergodic billiards that are not quantum unique ergodic* Ann. of Math. **171** (2010) 605–618.
[23] D. R. Heath-Brown *Convexity bounds for L-functions* Acta Arith. **136** (2009) 391–395.
[24] A. J. Hildebrand *A note on Burgess' character sum estimate* C. R. Math. Rep. Acad. Sci. Canada **8** (1986) 35–37.
[25] A. J. Hildebrand *On Wirsing's mean value theorem for multiplicative functions* Bull. London Math. Soc. **18** (1986) 147–152.
[26] J. Hoffstein and P. Lockhart *Coefficients of Maass forms and the Siegel zero* Ann. of Math. **140** (1994) 161–181.
[27] J. Hoffstein and D. Ramakrishnan *Siegel zeros and cusp forms* IMRN (1995) 279–308.
[28] R. Holowinsky *Sieving for mass equidistribution* Ann of Math. to appear, preprint, available as `arxiv.org:math/0809.1640`.
[29] R. Holowinsky and K. Soundararajan *Mass equidistribution of Hecke eigenforms* Ann. of Math., to appear, preprint, available as `arxiv.org:math/0809.1636`.
[30] H. Iwaniec *Spectral methods of automorphic forms*, AMS Grad. Studies in Math. **53** (2002).
[31] H. Iwaniec *Topics in classical automorphic forms*. Grad. Studies in Math. AMS **17**.
[32] H. Iwaniec and E. Kowalski *Analytic number theory* AMS Coll. Publ. **53** (2004).
[33] H. Iwaniec and P. Sarnak *Perspectives on the analytic theory of L-functions*, Geom. Funct. Analysis Special Volume (2000) 705-741.
[34] D. Jakobson *Quantum unique ergodicity for Eisenstein series on* $PSL_2(\mathbb{Z})\backslash PSL_2(\mathbb{R})$ Ann. Inst. Fourier **44** (1994) 1477–1504.

[35] X. Li *Bounds for $GL(3) \times GL(2)$ L-functions and $GL(3)$ L-functions*, Ann. of Math., to appear.
[36] E. Lindenstrauss *Invariant measures and arithmetic quantum unique ergodicity*, Ann. of Math. **163** (2006) 165–219.
[37] E. Lindenstrauss *Adelic dynamics and arithmetic quantum unique ergodicity* Curr. Developments in Math. (2004) 111-139.
[38] W. Luo *Values of symmetric square L-functions at* 1 J. Reine angew. Math. **506** (1999) 215–235.
[39] W. Luo and P. Sarnak *Mass equidistriution for Hecke eigenforms* Comm. Pure Appl. Math. **56** (2003) 874–891.
[40] W. Luo and P. Sarnak *Quantum ergodicity for $SL_2(\mathbb{Z})/\mathbb{H}^2$* IHES
[41] W. Luo, Z. Rudnick, and P. Sarnak *On Selberg's eigenvalue conjecture*, Geom. and Funct. Anal. **5** (1995) 477-502.
[42] J. Marklof *Arithmetic quantum chaos* Encyclopedia of Math. Phys. (Eds: J.-P. Francoise, G.L. Naber and Tsou S.T.) Elsevier (2006) **1** 212–220.
[43] P. Michel *Analytic number theory and families of automorphic L-functions* Automorphic forms and applications 181–295. IAS/Park City Math. Ser. 12, Amer. Math. Soc., Providence, RI (2007)
[44] P. Michel and A. Venkatesh *The subconvexity problem for $GL(2)$* preprint, available on arxiv (2009).
[45] G. Molteni *Upper and lower bounds at $s = 1$ for certain Dirichlet series with Euler product* Duke Math. J. **111** (2002) 133–158.
[46] M. Nair *Multiplicative functions of polynomial values in short intervals* Acta Arith. **62** (1992) 257–269.
[47] M. Nair and G. Tenenbaum *Short sums of certain arithmetic functions* Acta Math. **180** (1998) 119–144.
[48] Z. Rudnick *On the asymptotic distribution of zeros of modular forms* IMRN (2005) 2059–2074.
[49] Z. Rudnick and P. Sarnak *The behaviour of eigenstates of arithmetic hyperbolic manifolds* Comm. Math. Phys. **161** (1994) 195–213.
[50] Z. Rudnick and P. Sarnak *Zeros of principal L-functions and random matrix theory* Duke Math. J. **81** (1996) 269–322.
[51] P. Sarnak *Estimates for Rankin-Selberg L-functions and Quantum Unique Ergodicity* J. Funct. Anal. **184** (2001) 419–453.
[52] P. Sarnak *Recent progress on QUE* preprint available on his website.
[53] P. Sarnak *Arithmetic quantum chaos* Israel Math. Conf. Proc., Bar-Ilan Univ., Ramat Gan, **8** (1995) 183–236.
[54] G. Shimura *On the holomorphy of certain Dirichlet series* Proc. London Math. Soc. **31** (1975) 79–98.
[55] P. Shiu *A Brun-Titchmarsh theorem for multiplicative functions*, J. Reine Angew. Math. **313** (1980) 161–170.
[56] K. Soundararajan *Quantum unique ergodicity for $SL_2(\mathbb{Z})\backslash\mathbb{H}$*, Ann. of Math., to appear, available as `arxiv.org/abs/0901.4060`
[57] K. Soundararajan *Weak subconvexity for central values of L-functions* Ann. of Math., to appear, preprint available at `http://arxiv.org//abs/0809.1635`.
[58] A. Venkatesh *Sparse equidistribution problems, period bounds, and subconvexity* Ann. of Math., to appear, available as `arxiv.org:math/0506224`

[59] T. Watson *Rankin triple products and quantum chaos* Ph. D. Thesis, Princeton University (eprint available at: `http://www.math.princeton.edu/~tcwatson/watson_thesis_final.pdf`) (2001).

[60] E. Wirsing *Das asymptotische Verhalten von Summen über multiplikative Funktionen* Acta. Math. Acad. Sci. Hungar. **18** (1967) 411–467.

[61] S. Zelditch *Mean Lindelöf hypothesis and equidistribution of cusp forms* J. Funct. Anal. **97** (1991) 1–49.

Stanford University, 450 Serra Mall, Building 380, Stanford, CA 94305-2125

E-mail address: `ksound@math.stanford.edu`