Maass cusp forms

(L function/Teichmüller space)

J.-M. DESHOUILLERS*, H. Iwaniec†, R. S. PHILLIPS‡, and P. SARNAK‡

*Université de Bordeaux I, F-33405 Talence, France; †Institute for Advanced Study, Princeton University, Princeton, NJ 08540; and ‡Department of Mathematics, Stanford University, Stanford, CA 94305

Communicated by Peter D. Lax, January 24, 1985

ABSTRACT It is shown that, under certain standard assumptions, such as extended Riemann hypotheses, the scattering matrix $\phi(s)$ for generic $\Gamma \leq \operatorname{SL}(2, \mathbb{R})$ is unexpectedly of order 2. This leads to the conjecture that the generic cofinite $\Gamma$ has very few Maass cusp forms.

In his 1954 Göttingen lectures on the theory of Eisenstein series, Selberg (1) raises the question of the asymptotic density of the poles of the Eisenstein series. For $\Gamma$ a cofinite (noncompact) discrete subgroup of $\operatorname{PSL}(2, \mathbb{R})$, let $\phi(s)$ be the determinant of the scattering matrix (2, 3). Selberg shows that as $R \to \infty$

$$M(R) = - \int_{-\infty}^{\infty} \frac{\phi'(v + it) dt}{\phi(v + it)} = O(R^2).$$

[1]

The poles of $\phi$ coincide with those of the Eisenstein series and $M$ gives a measure of the density of these poles. Selberg also shows that when $\Gamma$ is a congruence subgroup then $M(R)$ is $O(R \log R)$. He asks the question as to what is the true size of $M(R)$ for a general $\Gamma$.

The importance of this question lies in its intimate connection with the problem of the existence of Maass cusp forms for $\Gamma$. Denote by $\mathfrak{H}$ the upper half plane and by $\Delta$ the Laplacian for $\Gamma \mathfrak{H}$ with its Poincaré metric. Let $\nu_1, \nu_2, \ldots$ be a basis for the Maass cusp forms for $\Gamma \mathfrak{H}$ (2). Thus $\nu_j$ is a cusp form and satisfies

$$\Delta \nu_j + \frac{1}{4} \nu_j = 0.$$  

[2]

It is known (1-3) that if $N(R) = \# \{ j : |r_j| \leq R \}$ then

$$M(R) + N(R) \sim \frac{\operatorname{Vol}(\Gamma \mathfrak{H})}{4\pi} R^2.$$  

[3]

In particular it follows from 3 and the previous remarks that for a congruence subgroup there are a lot of cusp forms, in fact

$$N(R) \sim \frac{\operatorname{Vol}(\Gamma \mathfrak{H})}{4\pi} R^2.$$  

[4]

In general, however, it is not known which, if either, of the two terms in 3 is responsible for the asymptotics. On the basis of the above result for congruence subgroups, Selberg conjectured that $N(R)$ is always dominant. In its strongest form his conjecture is that for every $\Gamma$ there is an $\epsilon_0 > 0$ such that $M(R) = O(R^{2-\epsilon_0})$.

Our main result may be stated as follows: Under certain hypotheses (such as an extended Riemann hypothesis), the generic $\Gamma$ does not satisfy the strong form of the above conjecture.

We state this more precisely: Let $\Gamma_0(q)$ be the Hecke congruence subgroup of level $q$; i.e., $\Gamma_0(q) = \{ \left( \begin{smallmatrix} a & b \\ c & d \end{smallmatrix} \right) \in \operatorname{SL}(2, \mathbb{Z}) : q \mid d \}$. For simplicity we assume that $q$ is prime. Let $L(z)$ be a holomorphic cusp form of weight 4 for $\Gamma_0(q)$. Assume further that $q$ is a Hecke eigenform. Let $u_j$ be, as above, an orthonormal basis for the Maass cusp forms for $\Gamma_0(q)$. These too are assumed to be Hecke eigenforms. Denote by $L(Q \otimes u_j, s)$ the Rankin–Selberg $L$ function of $Q$ and $u_j$. We assume that we have normalized this function so that it has a functional equation $s \rightarrow 1 - s$, it is entire, and of course it has an Euler product [it has a factor of $\zeta_2(2s)$ where $\zeta_2(s) = (1 - q^{-2s})(s)$ to make it entire]. The extended Riemann hypothesis referred to earlier is that $L(Q \otimes u_j, s)$ satisfies the analogue of the Riemann hypothesis. Actually it is not quite this that we need but rather the analogue of the Lindelöf hypothesis (5).

The Extended Lindelöf Hypothesis

For $\varepsilon > 0$,

$$L(Q \otimes u_j, \frac{1}{2} + it) < (1 + |t|)(1 + |r_j|)^{\varepsilon} \text{exp}(\varepsilon \pi r_j^{1/2}).$$  

[5]

The factor $(\cosh \pi r_j)^{-\varepsilon}$ comes from the normalization $||u_j||_2 = 1$. As in the case of the classical zeta function the extended Lindelöf hypothesis follows from the extended Riemann hypothesis and the Ramanujan conjectures (unpublished work).

For a given real number $r$, we define $\xi$, to be the subspace of Maass cusp forms of $\Gamma_0(q)$ with eigenparameter $r$ and we set $m(r)$ equal to the dimension of $\xi$. At present all that is known for sure is that $m(r) < 1/r \log r$. In the case of the modular group it is conjectured that all eigenvalues are simple; i.e., $m(r) \leq 1$. For a congruence subgroup the analogous conjecture would be that $m(r)$ is bounded. We shall say that $\Gamma_0(q)$ has a cusp form degeneracy of order $\beta$ if

$$m(r) < r^\beta.$$  

[6]

Finally for any $\Gamma$ let $T(\Gamma)$ denote its Teichmüller space (6). The assertion that a given property holds generically in a Teichmüller space means that it holds except on a set of first Baire category.

Theorem 1. Fix $q$ and set $T = T(\Gamma_0(q))$. Assuming the extended Lindelöf hypothesis for the $L$ functions $L(Q \otimes u_j, s)$ and that the degeneracy of cusp forms for $\Gamma_0(q)$ is of order $\beta$, then for generic $\Gamma$ in $T$ we have

$$M(R) \gg R^{2-\beta - \varepsilon} \quad \forall \varepsilon > 0.$$  

[7]

As mentioned above, we expect that $\beta$ may in fact be chosen arbitrarily small for $\Gamma_0(q)$. In this case, we would have for generic $\Gamma$ in $T$ the lower bound

$$M(R) \gg R^{2 - \varepsilon}, \quad \forall \varepsilon > 0,$$
contrary to the strong form of Selberg's conjecture.

The proof of Theorem 1 is based on the methods developed in refs. 4 and 7. In ref. 7, it was shown that destroying cusp forms under a deformation is closely related to the vanishing or not of the numbers \( L(Q \otimes \mathbf{u}_j, \frac{1}{2} + ir_j) \) (these are special values for these \( L \) functions). On the other hand using the methods in ref. 4 we can prove Theorem 2.

**Theorem 2.** There is \( C > 0 \) depending on \( Q \) such that

\[
\sum_{\chi | \chi + \frac{1}{2} \text{ and with } \chi} \left| L(Q \otimes \mathbf{u}_j, \frac{1}{2} + ir_j) \right|^2 \sim CR^2 \log R. \tag{8}
\]

This shows that on average these numbers are not zero. However, in applying this result, the presence of multiple eigenvalues introduces an added complication. In this case another basis for the Maass cusp forms, which we call the Kato basis and denote by \( \{ \psi_k \} \), plays a more central role than the \( \{ \mathbf{u}_j \} \) (see below). In any case Theorem 2 allows us to prove, without any assumptions at all, that an infinite number of zeros of the Selberg zeta function \( Z_{\Gamma \phi(q)}(s) \) move off the line \( \Re(s) = \frac{1}{2} \) under a deformation in \( T \). In fact at least \( R^2 \epsilon^{-1} \) of the \( R^2 \) zeros whose height is less than \( R \) move off. (See ref. 2 for the definition and properties of \( Z(s) \).) Theorem 1 together with the methods of this paper further support the possibility alluded to in ref. 7, which we now state as a precise conjecture.

**Conjecture.** For generic \( \Gamma \in \mathcal{H}(\Gamma_0(q)) \) there are only a finite number of Maass cusp forms; that is, \( N(R) \) in 3 is a bounded function.

We remark that this conjecture is not that far out of reach. If we ignore the technical difficulties of degeneracy of the eigenvalues then the conjecture would follow from knowing that all but a finite number of \( L(Q \otimes \mathbf{u}_j, \frac{1}{2} + ir_j) \) are not zero. All evidence points to this being the case.

We turn to a brief outline of the main steps in the proof of Theorem 1. For simplicity assume first that the cusp spectrum of \( \Gamma_0(q) \) is simple (or that we are dealing with the simple part of it). For \( \gamma \), a real analytic curve in \( T \) with tangent vector \( \mathbf{Q} \) at \( \Gamma_0(q) \), it is shown in ref. 7 that if \( L(Q \otimes \mathbf{u}_j, \frac{1}{2} + ir_j) \neq 0 \), then for all but a countable number of \( r \), the corresponding \( \mathbf{u}_j(e) \) is not a cusp form. (Here \( \mathbf{u}_j(e) \) depend analytically on \( e \) and are eigenfunctions of the pseudo-Laplacian \( \Delta_e \) (see refs. 8 and 9 for the definition of \( \Delta_e \).) It follows that the set of \( e \) for which all of the \( \mathbf{u}_j(e) \) are not cusp forms whenever \( L(Q \otimes \mathbf{u}_j, \frac{1}{2} + ir_j) \neq 0 \) is all but a countable set of \( \gamma \). The destruction of cusp forms is readily related to the growth of \( M(R) \) (see ref. 3). It follows from Theorem 2 and 5 that

\[
\#\{|r| \leq R \mid L(Q \otimes \mathbf{u}_j, \frac{1}{2} + ir_j) \neq 0 \} \gg R^{2-\epsilon} \tag{9}
\]

\[ \forall \epsilon > 0. \]

Parenthetically we note that assuming only the Ramanujan conjectures, one can prove

\[
\#\{|r| \leq R \mid L(Q \otimes \mathbf{u}_j, \frac{1}{2} + ir_j) \neq 0 \} \gg R^{3-\epsilon} \tag{10}
\]

\[ \forall \epsilon > 0. \]

(unpublished work).

The estimate 9, together with the previous remarks and Lemma 1 below, are the main ingredients in the proof of Theorem 1 for simple eigenvalues.

Let \( \rho_j(e) \) denote the eigenparameters of the pseudo-Laplacian \( \Delta_e(e) \); these include the cusp eigenvalues \( r_j(e) \). As shown in ref. 7, \( \rho_j(e) \) depend analytically on \( e \).

**Lemma 1.** For any \( \eta \) there is \( \varepsilon > 0 \) such that

\[
|\rho_j(e)| \leq 2|\rho_j(0)| + 1 \text{ for } |e - \eta| < \varepsilon. \]

In the case that \( \Gamma_0(q) \) has a multiple eigenvalue \( r \), we can choose a basis \( \mathbf{v}_1(0), \mathbf{v}_2(0), \ldots, \mathbf{v}_k(0) \) of the \( r \) eigenspace \( \mathbf{E}_r \) so that \( \mathbf{v}_1(e), \ldots, \mathbf{v}_k(e) \) are analytic in \( e \). We call this basis of \( \mathbf{E}_r \), the Kato basis (10). Unfortunately we have little control over this basis and it may have no relationship with the more natural Hecke eigenbasis of \( \mathbf{E}_r \). In particular we can no longer rely only on the estimate 5 and it is this that forces us to make the assumption on the order of degeneracies stated in Theorem 1. Complete details of the proofs will appear in ref. 11 and elsewhere.