Functoriality and Small Eigenvalues of Laplacian on Riemann Surfaces

Freydoon Shahidi

Abstract. The purpose of this article is to survey the recent progress made on estimating positive eigenvalues of Laplacian on hyperbolic Riemann surfaces in the case of congruence subgroups in connection with the Selberg conjecture, as well as certain related ones. The results are obtained as consequences of establishing certain important cases of Langlands’ functoriality conjecture.

Contents

1. Introduction 385
2. Ramanujan Conjecture 387
3. The Hyperbolic Circle Problem 388
4. Functoriality and Maass Forms 390
5. The Method 394
References 397

1. Introduction

Hyperbolic Riemann surfaces are one dimensional complex connected manifolds whose universal coverings are the upper half plane $H$ (conformally equivalent to the open disc $\Delta$ of radius 1). They comprise most Riemann surfaces and are uniformized as $M = \Gamma \backslash H$, where $\Gamma$ is a freely acting Fuchsian group, i.e., a discrete subgroup of $PSL_2(\mathbb{R})$, the group of conformal automorphisms of $H$, acting without fixed points on $H$ through fractional linear transformations

\begin{equation}
\gamma \cdot z = (az + b)/(cz + d) \quad (\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma).
\end{equation}

Then $\Gamma = \pi_1(M)$, the fundamental group of $M$.

2000 Mathematics Subject Classification. Primary 11F70, 30F35; Secondary 11R39, 11R42, 34B25.

Key words and phrases. Hyperbolic Riemann surfaces, Laplace operators, Selberg and Ramanujan conjectures, Langlands functoriality conjecture.

The author was partially supported by NSF grant DMS-0200325.
Throughout this paper we are interested only in those $M$ which are parametrized by Fuchsian groups of the first kind. These are simply Fuchsian groups for which $\text{Vol}(\Gamma \setminus H) < \infty$ and they are usually called “finite volume type groups.” Here the volume is calculated with respect to the hyperbolic measure $dxdy/y^2$ on $H$.

Let us recall that a congruence subgroup is a subgroup of $SL_2(\mathbb{Z})$ containing a principal congruence subgroup $\Gamma(N)$, the subgroup of all $\gamma \in SL_2(\mathbb{Z})$ satisfying $\gamma \equiv I \pmod{N}$ for some positive integer $N$. In view of the fact that the congruence subgroup problem (cf. [60]) is not valid for $SL_2(\mathbb{R})$, not every arithmetic subgroup (i.e., finite index) of $SL_2(\mathbb{Z})$ is a congruence subgroup. This was first observed in 1887 by Fricke [16] and Pick [52]. We refer to [44], page 251, for a discussion of this and examples of non–congruence arithmetic subgroups of $SL_2(\mathbb{Z})$.

The Laplace operator on $M = \Gamma \setminus H$ is simply

\begin{equation}
\Delta = -y^2 \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right).
\end{equation}

This is a symmetric and non–negative operator which has a self–adjoint extension to all of $L^2(\Gamma \setminus H)$. Spectral decomposition of $\Delta$ on $L^2(\Gamma \setminus H)$ decomposes $L^2(\Gamma \setminus H)$ to the direct sum of its discrete and continuous spectrum.

It is traditional to write $\lambda = s(1 - s), \ s \in \mathbb{C}$, to denote an eigenvalue for $\Delta$. We note that the continuous part $s = \frac{1}{2} + it, \ t \in \mathbb{R}$, giving $\lambda = \frac{1}{4} + t^2$. For the discrete spectrum, we denote the distinct eigenvalues as $0 = \lambda_0 < \lambda_1 < \lambda_2 < \ldots$. If $0 < \lambda_i < 1/4$, we call $\lambda_i$ exceptional; meaning in particular, that they are of a different nature than those greater than or equal to $1/4$ (cf. Selberg’s conjecture below). They are finite in number.

One of the central problems in the theory of Riemann surfaces is how small $\lambda_1$ can get. When $\Gamma$ is cocompact, there are many examples in which $\lambda_1 < 1/4$. In fact, if $\Gamma \setminus H$ is of “signature $(g,0,0)$”, then Schoen, Wolpert and Yau [58] have shown that $\lambda_{2g-3}$ can be made as small as one wishes, allowing exceptional eigenvalues below $1/4$, while Buser (cf. [3]) has proved

$$\lambda_{4g-2} \geq 1/4,$$

i.e., $\lambda_{4g-2}$ is never exceptional, but $\lambda_{2g-3}$ can be exceptional for arbitrary $g \geq 2$. Remarkably enough $\lambda_1$ has a universal upper bound due to Yang and Yau [72]:

$$\lambda_1 \leq 2 \frac{g + 1}{g - 1} \leq 6.$$ 

As for non–compact Riemann surfaces, one can find examples where $\lambda_1(\Gamma \setminus H) < 1/4$ ([59, 73]). But they are not congruence subgroups. In fact, for a congruence subgroup Selberg made the following remarkable conjecture.

**Conjecture 1.1 (Selberg [59], 1965).** There are no exceptional eigenvalues for congruence subgroups, i.e., one has

$$\lambda_1 \geq 1/4.$$ 

In fact, in [59], Selberg proved:

**Theorem 1.2.** For any congruence subgroup $\Gamma$

$$\lambda_1 = \lambda_1(\Gamma \setminus H) \geq 3/16.$$ 

With all the examples discussed earlier, it is clear that the conjecture is very much of arithmetic nature and quite deep. In fact, most of the progress in improving
these lower bounds have come from the theory of automorphic forms. The purpose of this article is to report on recent progress on this important conjecture and other related ones such as that of Ramanujan. We refer to [1, 4, 5, 6, 20, 25, 27, 55, 56, 69, 70] for some recent excellent expository articles on different aspects of these conjectures and more.

I would like to thank the editors of “Surveys in Differential Geometry” and in particular Professor S.T. Yau for their invitation that I prepare such a report for inclusion in the volume. I would also like to thank the referees for their comments towards the improvement of this exposition.

2. Ramanujan Conjecture

To study the eigenvalues of Δ one needs to concentrate on eigenfunctions. These are now functions on \( \Gamma \backslash H \) and those for eigenvalue \( \lambda = s(1 - s) \) can be written as

\[
(2.1) \quad f(x + iy) = f_0(y) + \sum_{n \neq 0} a_n(|n|y)^{1/2} K_{s - \frac{1}{2}}(2\pi |n|y)e^{2\pi inx},
\]

where \( K_\nu(z) \) is the Whittaker–Bessel function bounded at infinity, i.e., the solution to the differential equation

\[
(2.2) \quad t^2 K''_\nu + tK'_\nu - (t^2 + \nu^2)K_\nu = 0
\]

satisfying

\[
(2.3) \quad K_\nu(t) \sim \sqrt{\frac{\pi}{2t}} e^{-t}
\]
as \( t \) goes to \( +\infty \).

The complex numbers \( a_n \) are the corresponding Fourier coefficients. Moreover, when \( \Gamma \) is a congruence subgroup the eigenfunctions for exceptional eigenvalues, if any, will all be cuspidal, i.e., \( f_0(y) \equiv 0 \).

When \( f_0(y) = 0 \), the functions defined by (2.1) are the so called Maass cusp forms and although they exist in plenty (Weyl’s Law), no explicit examples are known, unless one resorts to congruence subgroups in which case there are explicit constructions of these forms using Galois–Weil representations [47, 54].

There is an analogue of the Selberg conjecture for the Fourier coefficients \( a_n \) when \( f \) is an eigenfunction for all the Hecke operators (cf. [17, 47, 71]). Let us further normalize \( f \) by assuming \( a_1 = 1 \). Then

**Conjecture 2.1 (Ramanujan–Petersson).** For every Maass cusp form \( f \) which is an eigenfunction for all the Hecke operators with \( a_1 = 1 \), and every prime number \( p \),

\[
|a_p| \leq 2p^{-1/2}.
\]

Henceforth, following the traditional terminology, we shall call this the Ramanujan conjecture.

Both Ramanujan and Selberg conjectures are quite hard and are still out of reach. But remarkable progress has been recently made in finding better bounds [32, 34, 35, 36, 37] which we shall now begin to explain. Before taking on this task, let us explain a problem of interest to both number theorists and geometers whose solution requires only partial improvements of existing bounds towards these conjectures, and which has now been completely resolved as a consequence of this
progress [37]. These new bounds are consequences of the recent progress in establishing new cases of Langlands’ functoriality conjecture all of which were considered out of reach even a few years ago.

Remark 2.2. There is an analogue of the Ramanujan–Petersson conjecture for normalized holomorphic cuspidal eigenforms of weight $k$, stating $|a_p| \leq 2p^{(k-1)/2}$. This was proved by Deligne in 1973, as a consequence of his proof of Weil’s conjecture [14]. The holomorphic structure of these forms play a central role in his proof, something that Maass forms lack.

3. The Hyperbolic Circle Problem

The hyperbolic distance function $\rho(z, w)$ on $H$ is simply given by

$$
\rho(z, w) = \log \frac{|z - w| + |z - w|}{|z - w| - |z - w|} \quad (z, w \in H)
$$

from which one can deduce a function $u(z, w)$

$$
u(z, w) = \frac{|z - w|^2}{4ImzImw}
$$

through

$$
\cosh \rho(z, w) = 1 + 2u(z, w)
$$

which is easier to work with. Given a positive real number $X$, the hyperbolic circle problem demands an estimate or asymptotic for the number of lattice points inside a hyperbolic circle of radius $X$ centered at a point $w \in H$ and generated by the $\Gamma$–orbit of another point $z \in H$ ($\Gamma$–lattice points). More precisely, one wants to estimate

$$
P(X) = \# \{ \gamma \in \Gamma | 4u(\gamma \cdot z, w) + 2 \leq X \}.
$$

Since $\Gamma \backslash H$ has a negative constant curvature ($\kappa = -1$), the euclidean circle packing arguments cannot be used. In fact, as the area and the length are of the same order of magnitude, Gauss’s circle problem estimate which appeals to an area calculation for the circle to estimate the lattice points inside it, fails. Instead one uses spectral theory of $\Delta$ on $L^2(\Gamma \backslash H)$ by cleverly choosing a kernel function for counting these points, and then estimating it by using spectral theory [26].

We refer to [43] for the first published version of these results. They have been known to Selberg in our setting, i.e., with comparable error estimates, but were never published (cf. [13]). We refer to [51] for an earlier result on this problem. The error estimates in [51] are not as good as those in [13, 43]. We should finally mention [3] for a detailed discussion of the problem and its history in the compact case.

To explain the result, let $\{u_j(z)\}$ be a complete set of orthonormal eigen–cusp forms for $\Delta$ attached to complex parameters $1/2 < \lambda_j \leq 1$, i.e., $0 \leq \lambda_j < 1/4$, then one can show [26]:

Let $\Gamma$ be a congruence subgroup. Then for $X \geq 2$, we have:

$$
P(X) = \sum_{1/2 < \lambda_j \leq 1} c\pi^{1/2} \frac{\Gamma(s_j - \frac{1}{2})}{\Gamma(s_j + 1)} u_j(z)\pi_j(w)X^{s_j} + O(X^{2/3}),
$$

where $c = 2$ or 1 according as $-1$ is in $\Gamma$ or not.
We refer to [26] for a detailed discussion of this problem and many other issues in the theory of Maass forms.

**Remark 3.1.** This is in fact true for any finite volume group $\Gamma$, if we include also the residual eigenfunctions of $\Delta$, i.e., the non–cuspidal discrete eigenfunctions. They can be constructed as residues of Eisenstein series.

One can then quickly show that if $F = \text{Vol}(\Gamma \backslash H)$, then

$$ P(X) = \sum_{\frac{2}{3} < s_j < 1} c\pi^{1/2} \frac{\Gamma(s_j - \frac{1}{2})}{\Gamma(s_j + 1)} u_j(z) \overline{u}_j(w) X^{s_j} $$

$$ + c\pi F^{-1} X + O(X^{2/3}) \quad (c = 1 \text{ or } 2) $$

and conclude:

**Proposition 3.2.** Suppose there are no exceptional eigenvalues in the range $\frac{2}{3} < s_j < 1$, i.e., $\lambda_1 \geq 2/9 = 0.222\ldots$ Then

$$ P(X) = \frac{c\pi}{F} X + O(X^{2/3}) \quad (c = 1 \text{ or } 2). $$

It is amusing to see that when this is applied to $\Gamma = SL_2(\mathbb{Z})$ and $z = w = \sqrt{-1}$ for which $c = 2$ and $F = \pi/3$ one gets

$$ \text{Card}\{(a, b, c, d) \in \mathbb{Z}^4|ad - bc = 1 \text{ and } a^2 + b^2 + c^2 + d^2 \leq X\} = 6X + O(X^{2/3}). $$

The fact that there are no exceptional eigenvalues for $SL_2(\mathbb{Z})$ is well known and old (cf. Section 11.3 of [26]).

Many other arithmetic approximations can be deduced from this for which we refer to Section 12 of [26]. For example, if $r(m)$ is the number of integral points inside a Euclidean circle of radius $\sqrt{m}$, $m \in \mathbb{N}$, then by applying the proposition to an appropriate conjugate of $\Gamma_0(2)$, one can show

$$ \sum_{m \leq X} r(m) r(m + 1) = 8X + O(X^{2/3}). $$

While for such large congruence subgroups Selberg’s conjecture has been known for sometime (in fact for subgroups of level $\leq 7$), the general case remains unavailable and it is for this reason and for the complete resolution of certain problems such as the hyperbolic circle problem (cf. [26, 37] for other examples), that partial improvements are quite desirable and highly appreciated by experts.

As for the Selberg conjecture the best result established so far is:

**Theorem 3.3.** (Kim–Sarnak [34]).

$$ \lambda_1 \geq \frac{1}{4} - \left(\frac{7}{64}\right)^2 = \frac{975}{4096} \approx 0.2380371. $$

While this is more than enough to remove the assumption on $\lambda_1$ in Proposition 3.2, i.e., $\lambda_1 \geq 0.222\ldots$, the unconditional resolution of the hyperbolic circle problem was first proved in [37] as:

**Proposition 3.4.** (Kim–Shahidi [37]). Suppose $X \geq 2$. Then

$$ P(X) = \frac{c\pi}{F} X + O(X^{2/3}), $$

where $c = 2$ or 1 according as $-1 \in \Gamma$ or not.
The proposition is a consequence of the following estimate proved in [37].

**Theorem 3.5 [37].** $\lambda_1 \geq \frac{1}{4} - (\frac{5}{34})^2 = \frac{66}{289} = 0.2283737\ldots$

As for the Ramanujan conjecture the best estimate at present is [34]:

$$p^{-\frac{1}{2}}(p^{7/64} - p^{-7/64}) \leq |a_p| \leq p^{-1/2}(p^{7/64} + p^{-7/64})$$

### 4. Functoriality and Maass Forms

The recent striking improvements towards the Ramanujan and the Selberg conjectures are consequences of certain special cases of Langlands’ functoriality conjecture [1, 42] proved recently [32, 35, 37].

The functoriality conjecture is one of the central components of a vast program (the Langlands program) which deals with automorphic forms on general reductive groups and considering our limitation it is better to only discuss it in the present context. We refer to [18] for an elementary introduction to the Langlands program.

A Maass (cusp) form $f$ is simply a real analytic eigenfunction of $\Delta$ in $L^2(\Gamma\backslash H)$ orthogonal to all the Eisenstein series and their residues. We will further assume that $f$ is an eigenfunction for all the Hecke operators (cf. [17, 26, 47, 71]) and normalize it so that $a_1 = 1$.

To explain functoriality it is best to introduce the adeles. This is simply a ring defined as the restricted product of all the completions of the field $\mathbb{Q}$ of rational numbers, with respect to their ring of integers. In fact, if $\mathbb{Q}_p$ and $\mathbb{Z}_p$ are respectively the field of $p$–adic numbers and its ring of integers, i.e., those whose $p$–adic absolute values are less than or equal to 1, then

$$x = (x_p) \in \prod_{p \leq \infty} \mathbb{Q}_p$$

is an adele if and only if $x_p \in \mathbb{Z}_p$ for almost all finite primes $p$. Here by convention $\mathbb{Q}_\infty = \mathbb{R}$, the field of real numbers for which $\mathbb{Z}_\infty$ is not a ring! The ring of adeles of $\mathbb{Q}$ is denoted by $\mathbb{A}_\mathbb{Q}$. It is a locally compact ring under the direct limit topology. The group $\mathbb{A}_\mathbb{Q}^\times = \mathbb{I}_\mathbb{Q}$ is called the group of *ideles*. It is a locally compact group if one takes the topology induced by

$$\mathbb{I}_\mathbb{Q} \simeq \{(x, x^{-1})|x \in \mathbb{I}_\mathbb{Q}\} \subset \mathbb{A}_\mathbb{Q} \times \mathbb{A}_\mathbb{Q}.$$

This whole notion can be extended to any number field $F$, i.e., a finite field extension of $\mathbb{Q}$; and one defines its ring of adeles $\mathbb{A}_F$ and group of ideles $\mathbb{I}_F = \mathbb{A}_F^\times$ in the same manner. One notes that $F$ (resp. $F^*$) is a discrete subgroup of $\mathbb{A}_F$ (resp. $\mathbb{I}_F$).

It is quite standard [17] to attach to an eigen–cusp Maass form $f$ an irreducible subrepresentation $\pi$ of

$$L^2(GL_2(\mathbb{Q})\backslash GL_2(\mathbb{A}_\mathbb{Q}), \chi),$$

the space of square integrable functions on $\mathbb{A}_\mathbb{Q}^\times GL_2(\mathbb{Q})\backslash GL_2(\mathbb{A}_\mathbb{Q})$ transforming under $\mathbb{A}_\mathbb{Q}^\times$ according to a grössencharacter $\chi$, i.e., a complex (unitary) character of $\mathbb{Q}^\times \backslash \mathbb{A}_\mathbb{Q}^\times$. There is a non–unique way of factorizing $\pi$ to $\pi = \otimes_p \pi_p$, where each $\pi_p$ is an irreducible preunitary representation of $GL_2(\mathbb{Q}_p)\backslash GL_2(\mathbb{R})$ if $p = \infty$. But the class of each $\pi_p$ is unique. Moreover, almost all of them are unramified or spherical, i.e., each have a vector fixed by $GL_2(\mathbb{Z}_p)$. One can then realize such $\pi_p$ as the full space of all the locally constant complex functions

$$\varphi: GL_2(\mathbb{Q}_p) \to \mathbb{C}$$
satisfying
\begin{equation}
\varphi\left(\begin{bmatrix} a & x \\ 0 & b \end{bmatrix} g\right) = \mu_p(a)\nu_p(b)|a/b|^{1/2}\varphi(g),
\end{equation}
for \(a, b \in \mathbb{Q}_p^*, x \in \mathbb{Q}_p\), where \(\mu_p\) and \(\nu_p\) are a pair of unramified characters of \(\mathbb{Q}_p^*\).

The class of the representation \(\pi_p\) is then determined by the conjugacy class of the diagonal element
\begin{equation}
t_p = \begin{pmatrix} \mu_p(p) & 0 \\ 0 & \nu_p(p) \end{pmatrix}
\end{equation}
in \(GL_2(\mathbb{C})\). Let \(\alpha_p = \mu_p(p)\) and \(\beta_p = \nu_p(p)\). Then the Ramanujan conjecture demands
\begin{equation}
|\alpha_p| = |\beta_p| = 1.
\end{equation}

For a Maass form, \(\pi_\infty\) is also fully induced from a pair of characters \((\mu_\infty, \nu_\infty)\) of \(\mathbb{R}^*\) with
\begin{equation}
\mu_\infty/\nu_\infty = |s_\infty| \quad (s_\infty \in \mathbb{C}),
\end{equation}
and the corresponding eigenvalue \(\lambda\), defined by \(\Delta f = \lambda f\), is given by
\begin{equation}
\lambda = \frac{1 - s_\infty^2}{4},
\end{equation}
in which \(s_\infty \in (-1, 1) \cup i\mathbb{R}\), \(i = \sqrt{-1}\).

The Selberg conjecture is then equivalent to
\begin{equation}
s_\infty \in i\mathbb{R}.
\end{equation}

Remark 4.1. Representations of these forms satisfying (4.3) and (4.6) are among a class of representations which are called “tempered”, and the generalized Ramanujan conjecture [57] requires that for a cuspidal representation \(\pi = \otimes'_p \pi_p\), every \(\pi_p\) be tempered (not necessarily only of the above forms). Therefore the Selberg conjecture is a special case of the generalized Ramanujan conjecture at \(p = \infty\), while the Ramanujan is one at \(p < \infty\).

The partial estimates in [34] are that
\begin{equation}
p^{-7/64} \leq |\alpha_p|, |\beta_p| \leq p^{7/64}
\end{equation}
and
\begin{equation}
s_\infty \in [-7/32, 7/32] \cup i\mathbb{R}
\end{equation}
or
\begin{equation}
\lambda \geq \lambda_1 \geq \frac{1}{4} - \left(\frac{7}{64}\right)^2 = \frac{975}{4096} \approx 0.2380371.
\end{equation}

Langlands’ functoriality conjecture can be formulated for any pair of arbitrary connected reductive groups. But for our purposes we will restrict ourselves to the case of general linear groups.

Given \(m \in \mathbb{N}\), a natural number, we define a homomorphism
\begin{equation}
\text{Sym}^m : GL_2(\mathbb{C}) \longrightarrow GL_{m+1}(\mathbb{C})
\end{equation}
as follows. For a form \(P(x, y)\) (homogeneous polynomial) of degree \(m\) in variables \(x\) and \(y\), and a \(g \in GL_2(\mathbb{C})\), we let \(\text{Sym}^m g \in GL_{m+1}(\mathbb{C})\) be the matrix which expresses the coefficients of \(P((x, y)g)\) in terms of those of \(P(x, y)\). Then \(\text{Sym}^m\) defines a homomorphism from \(GL_2(\mathbb{C})\) into \(GL_{m+1}(\mathbb{C})\).
For $\pi = \otimes_p \pi_p$ a cuspidal representation of $GL_2(\mathbb{A}_\mathbb{Q})$, we let, as before,

$$t_p = \begin{pmatrix} \alpha_p & 0 \\ 0 & \beta_p \end{pmatrix} \in GL_2(\mathbb{C})$$

parametrize $\pi_p$ for almost all $p < \infty$. Observe that

$$\text{Sym}^m t_p = \text{diag}(\alpha_p^m, \alpha_p^{m-1} \beta_p, \ldots, \beta_p^m) \in GL_{m+1}(\mathbb{C})$$

can be used to define a semisimple conjugacy class in $GL_{m+1}(\mathbb{C})$.

It is well–known that each $\text{Sym}^m t_p$ determines an irreducible admissible representation of $GL_{m+1}(\mathbb{Q}_p)$, denoted by $\text{Sym}^m \pi_p$, which has a vector fixed by $GL_{m+1}(\mathbb{Z}_p)$.

Langlands’ functoriality conjecture in this case then demands the existence of representations $\text{Sym}^m \pi_p$ at all other places $p$ such that

$$\text{Sym}^m \pi = \otimes'_p \text{Sym}^m \pi_p$$

is an automorphic representation of $GL_{m+1}(\mathbb{A}_\mathbb{Q})$, i.e., it appears in

$$L^2(GL_{m+1}(\mathbb{Q}) \backslash GL_{m+1}(\mathbb{A}_\mathbb{Q}), \omega_{\pi}^{m(m+1)/2}),$$

where $\omega_{\pi}$ is the central character of $\pi$, i.e., $\pi|_{\mathbb{A}_\mathbb{Q}^*}$.

The appearance of $\text{Sym}^m \pi$ does not need to be as a discrete subspace. It could appear through an Eisenstein series as part of the continuous spectrum (by means of incomplete Eisenstein series [26, 41, 48]). But the fact that it appears is very deep as it is evident from the consequences of the validity of the conjecture even for $m = 3$ and 4 (c.f. [32, 34, 36, 37, 53, 54, 67]).

**Theorem 4.2 ([32, 37]).** Let $\pi$ be a cuspidal representation of $GL_2(\mathbb{A}_\mathbb{Q})$. Then $\text{Sym}^m \pi$ is automorphic for not only $m = 1$ and 2 but also for $m = 3$ and 4. (The case $m = 1$ is trivial and the case $m = 2$ is due to Gelbart–Jacquet [19].)

**Remark 4.3.** The theorem is valid for cuspidal representations of $GL_2(\mathbb{A}_F)$ for any number field $F$.

There are many consequences of these results in automorphic forms and number theory. But let us only point out how (4.7), (4.8) and (4.9) are obtained from Theorem 4.2 here.

It follows from a result of Luo–Rudnick–Sarnak [46] that one can bound $\alpha_p^m$ and $\beta_p^m$ by

$$p^{-(\frac{1}{2} - \frac{1}{(m+1)^2+1})} \leq |\alpha_p|^m, \quad |\beta_p|^m \leq p^{\frac{1}{2} - \frac{1}{(m+1)^2+1}},$$

or

$$p^{-(\frac{1}{m} - \frac{1}{(m+1)^2+1})} \leq |\alpha_p|, \quad |\beta_p| \leq p^{\frac{1}{m} - \frac{1}{(m+1)^2+1}},$$

As for the Selberg conjecture, using [46] one gets

$$s_\infty \in [-\ell, \ell] \cup i\mathbb{R},$$

where

$$\ell = \frac{2}{m} \left( \frac{1}{2} - \frac{1}{(m+1)^2+1} \right).$$
Remark 4.4. The result in [46] is quite foundational and must be considered as a breakthrough. It provides us with the first non-local estimates for Hecke-eigenvalues of cuspidal representations of $GL_n(\mathbb{A}_F)$. Most of the improvements on the Selberg and Ramanujan Conjectures are consequences of combining functorial transfers of forms on $GL_2(\mathbb{A}_F)$ to appropriate $GL_m(\mathbb{A}_F)$, which we will discuss in the next section, with these estimates.

Using case $m = 3$ of Theorem 4.2 we have

**Corollary 4.5.** $s_\infty \in [-5/17, 5/17] \cup i\mathbb{R}$ or equivalently

(4.19) \[ \lambda_1 \geq \frac{1}{4} - \left(\frac{5}{34}\right)^2 = \frac{66}{289} \simeq 0.2283737 > 2/9. \]

Consequently, if $\Gamma$ is a congruence subgroup

(4.20) \[ P(X) = \frac{c\pi}{F} X + O(X^{2/3}), \]

where $c = 2$ or $1$ according as $-1 \in \Gamma$ or not, i.e., the hyperbolic circle estimates are valid.

A direct appeal to (4.15) and (4.16) for $m = 4$ will provide us with the estimates (cf. [32])

(4.21) \[ p^{-3/26} \leq |\alpha_p|, \quad |\beta_p| \leq p^{3/26} \]

and

(4.22) \[ s_\infty \in [-3/13, 3/13] \cup i\mathbb{R} \]

or

(4.23) \[ \lambda_1 \geq \frac{1}{4} - \left(\frac{3}{26}\right)^2 = \frac{40}{169} \simeq 0.2366839, \]

which although still quite striking, are even weaker than

(4.24) \[ p^{-1/9} < |\alpha_p|, \quad |\beta_p| < p^{1/9} \]

and

(4.25) \[ s_\infty \in (-2/9, 2/9) \cup i\mathbb{R} \]

or equivalently

(4.26) \[ \lambda_1 > \frac{1}{4} - \frac{1}{81} = \frac{77}{324} \simeq 0.23765432 \]

obtained in [36] (cf. [33] for the archimedean estimate), a result which although slightly weaker than (3.10) and (3.12), is valid over every number field.

The estimates (3.10) and (3.12) proved in [34] require further appeal to the theory of automorphic $L$–functions (that of $L(s, \text{Sym}^4\pi, \text{Sym}^2)$, to be precise, which is now available using the same machinery [32, 34, 63, 64, 66] since $\text{Sym}^4\pi$ is automorphic) and methods of analytic number theory [2, 15]. We refer to [34] for details.

**Remark 4.6.** It is important to put the recent progress in perspective. The earlier estimates on both conjectures were in the range of exponent $1/5$ (over arbitrary number fields) and the slightly better exponent $\frac{1}{5} + \varepsilon$ for all $\varepsilon > 0$, over $\mathbb{Q}$. They were obtained in [65] (cf. [68] for an exposition) and [2], respectively. For the Selberg conjecture $\frac{1}{5} + \varepsilon$, all $\varepsilon > 0$, was the best one proved in [45]. It gave the lower bound $\lambda_1 \geq 0.21$. They were all consequences of the automorphy of $\text{Sym}^2\pi$
proved in \([19]\) and techniques and estimates such as those in \([15, 46]\) (cf. Remark 4.4).

For quite a long time experts struggled to obtain an estimate in the range of 1/6 as that would already simplify and improve a number of results in number theory and automorphic forms. It was therefore quite surprising when the estimate as that would already simplify and improve a number of results in number theory and automorphic forms. It was therefore quite surprising when the estimate was 1/5 over any number field, which was only slightly weaker than 1/7, was announced by the author and Kim. The striking estimates \((3.10), (3.12), (4.24)\) and \((4.25)\) were obtained a few months after that \([33, 34, 36]\) as soon as the automorphy of \(\text{Sym}^4\pi\) was also ready at hand \([32]\). The techniques in establishing \((3.10), (3.12), (4.24)\) and \((4.25)\) using the automorphy of \(\text{Sym}^3\pi\) and \(\text{Sym}^4\pi\) are similar to those for \(5/28 + \varepsilon\) and 1/5.

5. The Method

It is now clear that these estimates are consequences of the existence of \(\text{Sym}^3\pi\) and \(\text{Sym}^4\pi\) as automorphic forms on \(\text{GL}_2(\mathbb{A}_F)\) and \(\text{GL}_5(\mathbb{A}_Q)\), respectively. In this section we will try to briefly explain the machinery behind it.

The automorphy of \(\text{Sym}^3\pi\) and \(\text{Sym}^4\pi\) are consequences of applying converse theorems of Cogdell and Piatetski–Shapiro \([10, 11]\) to analytic properties of certain \(L\)–functions proved by the Langlands–Shahidi method \([21, 31, 40, 41, 62, 63, 64, 65, 66]\). At present, they cannot be obtained from other methods which have been developed to prove functoriality.

To start off, let us note that in our approach, and more generally in the Langlands program, the choice of the number field is of no concern and one may assume \(\pi\) is a cuspidal representation of \(\text{GL}_2(\mathbb{A}_F)\), where \(\mathbb{A}_F\) is the ring of adeles of an arbitrary number field. Beside the fact that functoriality needs to be proved for groups over all global fields (as well as local ones), many applications, even those in algebraic number theory \([6, 12]\), require estimates and results over arbitrary number fields. The reader should now appreciate that in this approach and context there will no longer be any need to study Hilbert modular forms as a different entity, at least when these questions arise, than those over \(\mathbb{Q}\) or any other number field. We may and will therefore assume \(\pi\) is an infinite dimensional irreducible admissible subrepresentation of \(L^2(\text{GL}_2(F)\backslash \text{GL}_2(\mathbb{A}_F), \chi)\) for some grössencharacter \(\chi\). It will automatically be cuspidal.

We can again write \(\pi = \otimes_v \pi_v\), where each \(\pi_v\) is an irreducible unitary representation of \(\text{GL}_2(F_v)\), where \(F_v\) is the completion of \(F\) in a place \(v\). There are obvious generalizations of all the notions visited in the previous section such as the fact that almost all \(\pi_v\) are spherical, i.e., have a vector fixed by \(\text{GL}_2(O_v)\), and that they are given by a pair of unramified characters of \(F_v^*\). (Here \(O_v\) is the ring of integers of \(F_v\).) In particular, for almost all \(v\), \(\pi_v\), or more specifically its class, is given by the conjugacy class of a diagonal element \(t_v = \text{diag}(\alpha_v, \beta_v) \in \text{GL}_2(\mathbb{C})\).

As explained earlier, the automorphy of \(\text{Sym}^2\pi = \otimes_v \text{Sym}^2\pi_v\) was established more than 25 years ago \([19]\). When \(\pi_v\) is spherical, it is attached to the diagonal element \(\text{Sym}^2 t_v \in \text{GL}_3(\mathbb{C})\) whose entries define the unramified characters which determine the class of \(\text{Sym}^2\pi_v\). The diagonal element \(t_v \otimes \text{Sym}^2 t_v \in \text{GL}_6(\mathbb{C})\) will then determine a spherical representation \(\pi_v \boxtimes \text{Sym}^2\pi_v\) of \(\text{GL}_6(F_v)\). The operation \(\boxtimes\) mirrors that of tensor products when one parameterizes these representations by means of two and three dimensional representations of \(W_{F_v}^\prime\), the corresponding Deligne–Weil group. The parameterization problem for irreducible admissible
representations of $GL_n(F_v)$ by means of $n$–dimensional complex representations of $W_{F_v}$ has now been completely resolved by Harris–Taylor [23] and Henniart [24]. This is a particularly deep and important result and of particular interest to us. In fact, it allows us to extend the operation to any pair of representations and not only spherical ones (cf. [42] for archimedean places). In particular, one can define $\pi_v \boxtimes \text{Sym}^2 \pi_v$ for all $v$. One of the special cases of the main result of [37] is

Theorem 5.1. a) $\pi \boxtimes \text{Sym}^2 \pi = \otimes_v' (\pi_v \boxtimes \text{Sym}^2 \pi_v)$ is an automorphic representation of $GL_6(\mathbb{A}_F)$.

b) $\pi \boxtimes \text{Sym}^2 \pi$ appears in the continuous spectrum (cf. [30]) of $GL_6(\mathbb{A}_F)$ and is defined by an “Eisenstein series of type (2,4)” attached to $(\pi \otimes \omega_\pi, \text{Sym}^2 \pi)$, where $\omega_\pi$ is the central character of $\pi$. In particular, $\text{Sym}^3 \pi$ is an automorphic representation of $GL_4(\mathbb{A}_F)$.

The main result of [37] (Theorem 5.1) proves the automorphy of $\pi_1 \boxtimes \pi_2 = \otimes_v' (\pi_{1v} \boxtimes \pi_{2v})$ for any pair of automorphic cuspidal representations $\pi_1$ and $\pi_2$ of $GL_2(\mathbb{A}_F)$ and $GL_3(\mathbb{A}_F)$, respectively.

This is a very fine result, incorporating both local and global functoriality, and consequently its proof is quite complicated. Using the Langlands–Shahidi method, one attaches a triple product $L$–function $L(s, \pi_1 \times \pi_2 \times (\tau \otimes \eta))$ to $\pi_1, \pi_2$ and $\tau$ cuspidal representations of $GL_2(\mathbb{A}_F)$, $GL_3(\mathbb{A}_F)$ and $GL_n(\mathbb{A}_F)$, $1 \leq n \leq 4$, respectively. Here $\sigma$ is assumed to be unramified at every place where $v < \infty$ and either $\pi_{1v}$ or $\pi_{2v}$ is ramified. Moreover, $\eta$ is a grössencharacter of $\mathbb{A}_F^*$, i.e., one of $F^* \backslash \mathbb{A}_F^*$, which we shall assume to be highly ramified at least at one place where one of $\pi_{iv}$’s, $i = 1, 2$, are ramified. Observe that now all the archimedean places at which the Selberg conjecture is rooted, are put outside the bad primes, enabling us to conclude deep results on them.

To utilize the method one notes that these $L$–functions appear in constant terms of certain Eisenstein series [22, 40, 41, 48, 50, 65, 66] defined on exceptional groups $\text{Spin}(10), E_6$ and $E_7$, granting the case $n = 1$ as a special case of Rankin–Selberg product $L$–functions on $GL_2(\mathbb{A}_F) \times GL_3(\mathbb{A}_F)$. The Langlands–Shahidi method then proves the necessary analytic properties of these $L$–functions. The twist by the highly ramified grössencharacter destroys all the possible symmetries that could lead to the existence of poles, proving they are entire (cf. [31]). One can also conclude that these entire functions are bounded in vertical strips of finite width [21]. Moreover, the technology developed in [63, 64, 65, 66] allows us to define root numbers $\varepsilon(s, \pi_1 \times \pi_2 \times (\sigma \otimes \eta))$ by means of which one establishes the functional equation

\begin{equation}
L(s, \pi_1 \times \pi_2 \times (\sigma \otimes \eta)) = \varepsilon(s, \pi_1 \times \pi_2 \times (\sigma \otimes \eta))L(1-s, \pi_1 \times \pi_2 \times (\sigma \otimes \eta^{-1}))
\end{equation}

where $\sim$ signifies appropriately defined duals.

Finally, we have

\begin{equation}
L_S(s, \pi_1 \times \pi_2 \times (\sigma \otimes \eta)) = L_S(s, (\pi_1 \boxtimes \pi_2) \times (\sigma \otimes \eta)),
\end{equation}

where the $L$–function on the right is that of Rankin–Selberg studied in [28, 29, 30, 49, 62, 63, 64] defined by an infinite product of local factors in which for all $v \notin S$, $S$ a finite set of places of $F$, either $\pi_v$ is unramified or $v = \infty$.

It is to the $L$–functions on the right hand side of (5.2) that one can apply converse theorems of Cogdell and Piatetski–Shapiro [10, 11] which states that if they are entire, bounded in vertical strips of finite width and satisfying (5.1),
then there exists an automorphic representation $\Pi = \otimes_v \Pi_v$ of $GL_6(\mathbb{A}_F)$ for which $\Pi_v \simeq \pi_{1v} \boxtimes \pi_{2v}$ whenever both $\pi_{iv}$ are unramified or $v$ is archimedean.

While we know nothing about the analytic properties of $L_S(s, (\pi_1 \boxtimes \pi_2) \times (\sigma \otimes \eta))$ (as they are defined only by means of an infinite product of local $L$-functions), their equality with $L_S(s, \pi_1 \times \pi_2 \times (\sigma \otimes \eta))$ provides us with the knowledge needed for applying the converse theorem as explained earlier.

Quite a bit more technical work is needed to show that in fact $\Pi_v \simeq \pi_{1v} \boxtimes \pi_{2v}$ for all $v$ and therefore $\pi_1 \boxtimes \pi_2$ is automorphic. We refer to [37] for details and appropriate references to both local and global difficulties.

The converse theorem [11] of Cogdell and Piatetski–Shapiro that we have used is non–standard and fairly recent (1999). It is designed precisely to handle $L$–functions for which holomorphy is obtained only up to a twist, something that one can provide from our method as it was observed by Kim in [31].

The automorphy of $\text{Sym}^4 \pi$ is proved again inductively. One applies the same machinery to certain Eisenstein series on groups of type $\text{Spin}(2n)$, $4 \leq n \leq 7$, to transfer $\text{Sym}^3 \pi$ to an automorphic form $\text{Sym}^3 \pi$ on $GL_6(\mathbb{A}_F)$ of type $(1, 5)$ attached to $(\omega_3, \text{Sym}^4 \pi \otimes \omega_3)$ from which automorphy of $\text{Sym}^4 \pi$ follows. We refer the reader to [32] for details.

Having established the automorphy of $\text{Sym}^3 \pi$ and $\text{Sym}^4 \pi$, one can then apply techniques of [65] (Section 5 and in particular Lemma 5.8 of [65]) to prove:

**Theorem 5.2.** [36] Let $\pi = \otimes_v' \pi_v$ be a cuspidal representation of $GL_2(\mathbb{A}_F)$. For each unramified $\pi_v$ let $t_v = \text{diag}(\alpha_v, \beta_v) \in GL_2(\mathbb{C})$ represent the corresponding semisimple conjugacy class. Then

$$q_v^{-1/9} < |\alpha_v|, \quad |\beta_v| < q_v^{1/9}.$$  

The archimedean analogue of this result is also valid. It can be formulated as in (4.25) for the corresponding $s_\infty$ defined by (4.4) at each archimedean place of $F$. We observe that unless $F = \mathbb{Q}$, this is the best bound available for $|\alpha_v|$ and $|\beta_v|$ at present (cf. (3.10) and (3.12) for $F = \mathbb{Q}$). We refer to [7, 12] for an application of this result to number theory.

There are many other applications of the existence of $\text{Sym}^3 \pi$ and $\text{Sym}^4 \pi$. But covering them all will become a bulky task and outside the scope of this paper. We refer the reader to [27, 32, 36, 37, 53, 54] for some of these applications. As has been the case with $\text{Sym}^2 \pi$ for the past 25 years we expect many consequences of the automorphy of $\text{Sym}^3 \pi$ and $\text{Sym}^4 \pi$ in several important subjects, ranging from number theory to automorphic forms and arithmetic geometry. At present the automorphy of $\text{Sym}^3 \pi$ remains out of reach!

**Remark 5.3.** As explained in Remark 4.1 the Ramanujan and Selberg conjectures are equivalent to each $\pi_v$ being a tempered representation, i.e., one appearing in $L^2(F_v^* \backslash GL_2(F_v))$, either discretely or continuously, the latter case implying that $\mu_v$ and $\nu_v$ are unitary for every $v$, finite or infinite.

A generalization of these conjectures to the effect that all the local components of a cuspidal representation of a general reductive group are tempered is one of the central problems in the theory of automorphic forms. Although this is false in general, one expects its validity for general linear groups.

There are instances of functoriality through which the study of automorphic forms on (split) classical groups $SO(2n)$, $SO(2n + 1)$ and $Sp(2n)$ is reduced to that of $GL(N)$ with $N = 2n$, except in the symplectic case $Sp(2n)$ for which $N =
For the generic spectrum of these classical groups, i.e., those with a non-zero Fourier coefficient of highest rank, the functoriality has now been established in [8, 9] and consequently the Ramanujan conjecture for these groups is reduced to that for $GL(m)$, for all $m \leq N$. We refer to [8, 9] where the same techniques as those used to establish existence of symmetric powers are used to prove these new results.

When our method is fully developed (converse theorems already are developed for any global field) and these transfers are established for the generic spectrum of cuspidal representations of classical groups over function fields, this should lead to a proof of the Ramanujan conjecture for generic cuspidal representations of these groups. This can be done using the validity of the conjecture for $GL(N)$ which has been proved by Lafforgue [38] for general linear groups over function fields. The case of number fields even for $GL(2)$ is still out of reach.

References


